

Protein mass spectrometry for bioprocess development & monitoring

A thesis submitted to University College London for the degree of
DOCTOR OF ENGINEERING

Alex Berrill
September, 2009

I, Alex Berrill confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

ABSTRACT

Bioprocesses for therapeutic protein production typically require significant resources to be invested in their development. This development could be improved with technologies that can elucidate the physicochemical properties of process stream components with small sample volumes in a rapid and readily performed manner. This is especially true in early phase development when material and established analytical methods are limiting.

This thesis has investigated various process materials but was focussed mainly on those existing in an ApolipoproteinA-IM (ApoA-IM) process, produced using an *Eschericia coli* (*E. coli*) host.

Using a mass spectrometric technique this project began by monitoring the product and contaminant during the ApoA-IM process and how this analytical approach compares to traditional analytical methods such as high performance liquid chromatography (HPLC). Results showed that, unlike many other analytical methods, surface enhanced laser desorption ionisation mass spectrometry (SELDI-MS) can handle early process samples that contain complex mixtures of biological molecules with limited sample pre-treatment and thereby provide meaningful process-relevant information.

The change in material during the flocculation/centrifugation stage of the process was then examined. When only a change in cellular debris was observed an existing methodology developed at University College London (UCL) was implemented to maximise cellular debris removal. The predictive scale down methodology enabled rapid optimization of the operating conditions for a flocculation followed with a centrifugation step using only small volumes (20mL) of a high solids (~20% w/w) *E. coli* heat extract. These experiments suggested that adding a higher level of a cationic polymer could substantially increase the strength of the flocculated particles produced, thereby enhancing overall clarification performance in a large scale centrifuge. This was subsequently validated at pilot scale.

The proteins remaining from this flocculation/centrifugation stage were then compared using the mass spectrometric technique to calculate the difficulty of removing each protein contaminant from the ApoA-IM product and suggested conditions for future sorbent scouting runs.

ACKNOWLEDGEMENTS

This work would not have been possible without the invaluable advice, and support of my supervisors both at UCL and Pfizer: Daniel Bracewell, Nigel Titchener-Hooker & Sa Ho. I particularly want to thank my supervisors for the opportunity they gave me to do part of my EngD at Pfizer in St. Louis, an experience that I will never forget. I would also like to thank my family and friends for all their encouragement and support over the course of my studies.

This project would also not have been possible without the support from the Engineering and the Physical Sciences Research Council (EPSRC) under the Innovative Manufacturing Research Initiative along with the support from Pfizer Inc.

CONTENTS

| | |
|--|-----------|
| ABSTRACT | 3 |
| ACKNOWLEDGEMENTS | 5 |
| LIST OF FIGURES | 11 |
| LIST OF TABLES | 14 |
| LIST OF EQUATIONS | 15 |
| CHAPTER 1 INTRODUCTION | 16 |
| 1.1 PROTEIN PRODUCTION PROCESSES | 16 |
| 1.1.1 Protein recovery | 18 |
| 1.1.2 Protein purification | 19 |
| 1.2 ULTRA SCALE-DOWN (USD) TECHNOLOGIES TO AID PROCESS DEVELOPMENT | 26 |
| 1.2.1 Determining the shear sensitivity of process material | 27 |
| 1.2.2 Determining the physicochemical properties of proteins | 28 |
| 1.2.3 Sorbent screening | 34 |
| 1.3 SELDI-MS | 37 |
| 1.3.1 Surface enhanced affinity capture | 38 |
| 1.3.2 Laser desorption/ionisation | 42 |
| 1.3.3 Time of flight mass spectrometry | 43 |
| 1.3.4 Potential uses | 46 |
| 1.3.5 Advantages | 48 |
| 1.3.6 Limitations | 49 |
| 1.4 HOST ORGANISM | 49 |
| 1.5 PROCESS MATERIALS | 49 |
| 1.5.1 Hen egg white | 50 |
| 1.5.2 Antibody fragment process material | 50 |
| 1.5.3 ApolipoproteinA-IM dimer process material | 51 |
| 1.6 PROJECT OBJECTIVES | 53 |
| 1.6.1 Objective 1: Develop the use of a mass spectrometry technology for bioprocess materials | 54 |
| 1.6.2 Objective 2: Application of the approach to an industrial process in conjunction with a USD device | 54 |
| 1.6.3 Objective 3: Develop novel approaches for acquisition and analysis of mass spectra to expedite bioprocess development and monitoring | 54 |
| CHAPTER 2 CONFIGURATION OF A PROTEIN MASS SPECTROMETER FOR BIOPROCESS MATERIALS | 56 |
| 2.1 INTRODUCTION | 56 |
| 2.2 MATERIALS & METHODS | 57 |

| | | |
|--|--|-----------|
| 2.2.1 | Protein samples | 57 |
| 2.2.2 | Mass spectrometry | 58 |
| 2.3 | RESULTS | 60 |
| 2.3.1 | Sample analysis with mass spectrometry | 60 |
| 2.3.2 | Width of mass spectra peaks | 62 |
| 2.3.3 | Alignment of mass spectra peaks | 64 |
| 2.3.4 | The effect of protein concentration and laser strength on mass spectra peak resolution and intensity | 65 |
| 2.4 | SUMMARY | 71 |
| CHAPTER 3 PROTEIN MASS SPECTROMETRY FOR MONITORING THE APOLIPOPROTEINA-IM PROCESS | | 73 |
| 3.1 | INTRODUCTION | 73 |
| 3.2 | MATERIALS & METHODS | 75 |
| 3.2.1 | Protein samples | 75 |
| 3.2.2 | Mass spectrometry | 75 |
| 3.2.3 | Reversed Phase HPLC..... | 77 |
| 3.2.4 | Size Exclusion HPLC..... | 77 |
| 3.2.5 | Host cell protein analysis | 77 |
| 3.3 | RESULTS | 78 |
| 3.3.1 | Initial capture chromatography | 78 |
| 3.3.2 | Ion-exchange chromatography..... | 83 |
| 3.3.3 | Oxidation | 84 |
| 3.3.4 | Anion-exchange chromatography & final step | 84 |
| 3.3.5 | Host cell protein removal | 85 |
| 3.3.6 | Mass spectrometry analysis during flocculation process | 86 |
| 3.4 | SUMMARY | 88 |
| CHAPTER 4 DEFINING & IMPROVING THE FLOCCULATION & CENTRIFUGATION STAGE OF THE APOLIPOPROTEINA-IM PROCESS USING A USD DEVICE | | 90 |
| 4.1 | INTRODUCTION | 90 |
| 4.2 | MATERIALS & METHODS | 91 |
| 4.2.1 | Bench-top studies | 91 |
| 4.2.2 | Pilot plant studies | 93 |
| 4.2.3 | Microscope images | 94 |
| 4.3 | THEORETICAL CONSIDERATIONS..... | 94 |
| 4.3.1 | Sigma theory..... | 94 |
| 4.3.2 | Hindered settling | 95 |
| 4.4 | RESULTS | 97 |
| 4.4.1 | Evaluation of shear | 97 |

| | | |
|--|--|------------|
| 4.4.2 | Adoption of a hindered settling correction factor | 99 |
| 4.4.3 | Comparing % clarification between pilot scale and USD mimic | 99 |
| 4.4.4 | Varying flocculating agent addition rate for both scales | 100 |
| 4.4.5 | Varying flocculating agent addition ratio with USD mimic | 103 |
| 4.4.6 | Validation of USD flocculating agent addition ratio at pilot scale | 106 |
| 4.5 | SUMMARY | 107 |
| CHAPTER 5 DEVELOPING AN EXPERIMENTAL METHODOLOGY TO DEFINE PRODUCT AND CONTAMINANT RELATIONSHIPS | | 109 |
| 5.1 | INTRODUCTION | 109 |
| 5.2 | MATERIALS & METHODS | 110 |
| 5.2.1 | Reducing pH below the pI of the ApoA-IM monomer | 110 |
| 5.2.2 | Mass spectrometry | 110 |
| 5.3 | RESULTS | 112 |
| 5.3.1 | Selection of wash step solutions for hydrophobic ProteinChips | 112 |
| 5.3.2 | Selection of final wash step solutions for all of the selective ProteinChips | 117 |
| 5.3.3 | Final decision on which buffers to use for selective ProteinChips | 120 |
| 5.3.4 | Selective adsorption profiles for the ApoA-IM post centrifuge supernatant proteins | 120 |
| 5.4 | SUMMARY | 123 |
| CHAPTER 6 DATABASE FOR HANDLING PROTEIN MASS SPECTRA | | 125 |
| 6.1 | INTRODUCTION | 125 |
| 6.2 | MATERIALS & METHODS | 126 |
| 6.2.1 | Mass spectrometry | 126 |
| 6.2.2 | Data treatment | 127 |
| 6.3 | RESULTS | 127 |
| 6.3.1 | Organisation & storage of mass spectrometry data | 127 |
| 6.3.2 | Program execution..... | 129 |
| 6.3.3 | Treatment of raw mass spectrometry data | 129 |
| 6.3.4 | Reproducibility..... | 135 |
| 6.4 | SUMMARY | 139 |
| CHAPTER 7 MASS SPECTROMETRY TO DESCRIBE PRODUCT AND CONTAMINANT RELATIONSHIPS FOR BIOPROCESS DEVELOPMENT..... | | 140 |
| 7.1 | INTRODUCTION | 140 |
| 7.2 | MATERIALS & METHODS | 141 |
| 7.2.1 | Protein samples | 141 |
| 7.2.2 | Mass spectrometry | 141 |
| 7.2.3 | Chromatography..... | 142 |
| 7.2.4 | Data visualisation techniques | 143 |

| | | |
|--|--|------------|
| 7.3 | RESULTS | 145 |
| 7.3.1 | Measurement of contaminants during purification | 145 |
| 7.3.2 | Generating basic physicochemical properties using mass spectrometry | 147 |
| 7.3.3 | Visualising nearest neighbour contaminants using mass spectrometry data | 149 |
| 7.3.4 | Identifying conditions for separating nearest neighbour contaminants using mass spectrometry | 152 |
| 7.3.5 | Nearest neighbour removal analysis | 154 |
| 7.4 | SUMMARY | 157 |
| CHAPTER 8 CONCLUSIONS & RECOMMENDATIONS FOR FUTURE WORK | | 159 |
| 8.1 | REVIEW OF PROJECT OBJECTIVES | 159 |
| 8.1.1 | Objective 1: Develop the use of a mass spectrometry technology for bioprocess materials | 159 |
| 8.1.2 | Objective 2: Application of the approach to an industrial process in conjunction with a USD device | 159 |
| 8.1.3 | Objective 3: Develop novel approaches for acquisition and analysis of mass spectra to expedite bioprocess development and monitoring..... | 160 |
| 8.2 | PUBLICATIONS | 160 |
| 8.3 | FINAL COMMENTS AND FUTURE WORK | 161 |
| 8.3.1 | Related to objective 1: Develop the use of a mass spectrometry technology for bioprocess materials..... | 161 |
| 8.3.2 | Related to objective 2: Application of the approach to an industrial process in conjunction with a USD device | 162 |
| 8.3.3 | Related to objective 3: Develop novel approaches for acquisition and analysis of mass spectra to expedite bioprocess development and monitoring..... | 162 |
| CHAPTER 9 VALIDATION OF BIOPROCESSES: IMPLICATION OF SCALE-DOWN METHODOLOGIES | | 166 |
| 9.1 | PROCESS VALIDATION | 166 |
| 9.2 | SCALE-DOWN MODELS..... | 166 |
| 9.2.1 | Process validation scenario: Effect of contaminant properties on purification | 167 |
| 9.2.2 | Process validation scenario: A new protein candidate for a process development team | 167 |
| ABBREVIATIONS | | 169 |
| NOMENCLATURE | | 171 |
| APPENDIX - MATLAB CODE | | 173 |
| | READING SPECTRA DATA..... | 173 |
| | Reading processed data | 180 |
| | Reading raw data..... | 181 |
| | PEAK SELECTION | 187 |
| | NEAREST NEIGHBOUR PREDICTOR..... | 188 |

REFERENCES 192

LIST OF FIGURES

| | |
|---|----|
| FIGURE 1.1 TYPICAL PROTEIN BIOPROCESS OPTIONS (LIENQUEO AND ASENJO,2000). | 16 |
| FIGURE 1.2 OPERATIONAL SEQUENCE SELECTION FOR PROTEIN PURIFICATION (ASENJO AND ANDREWS,2004). | 20 |
| FIGURE 1.3 A THEORETICAL MECHANISM FOR HYDROPHOBIC INTERACTIONS. | 24 |
| FIGURE 1.4 THE THEORY BEHIND GEL FILTRATION (GE GEL FILTRATION PRINCIPLES AND METHODS,2003). | 26 |
| FIGURE 1.5 MAPPING THE PHYSICOCHEMICAL PROPERTIES OF PROTEINS (WALL ET AL. 2002). | 28 |
| FIGURE 1.6 PREDICTING DIMENSIONLESS RETENTION TIME (LIENQUEO ET AL. 2007). | 31 |
| FIGURE 1.7 A TWO DIMENSIONAL POLYACRYLAMIDE GEL (OU ET AL. 2001). | 33 |
| FIGURE 1.8 EXPERIMENTAL PROCEDURE FOR MICRO-BATCH ADSORPTION. | 36 |
| FIGURE 1.9 A MICRO-PIPETTE COLUMN (WENGER ET AL. 2007). | 36 |
| FIGURE 1.10 COMPARISON OF DIFFERENT CHROMATOGRAPHY SCALES (WENGER ET AL. 2007). | 37 |
| FIGURE 1.11 SELDI SELECTIVE PROTEINCHIP PREPARATION. | 39 |
| FIGURE 1.12 USING THE SELDI BIOPROCESSOR. | 40 |
| FIGURE 1.13 THE SELDI PROTEINCHIP READER. | 43 |
| FIGURE 1.14 THE SELDI MASS SPECTROMETER (MERCHANT AND WEINBERGER,2000). | 44 |
| FIGURE 1.15 TIME OF FLIGHT IN A MASS SPECTROMETER (MERCHANT AND WEINBERGER,2000). | 44 |
| FIGURE 1.16 FAB' PRODUCTION PROCESS. | 51 |
| FIGURE 1.17 APOA-IM AMINO ACID SEQUENCE. | 52 |
| FIGURE 1.18 THE STRUCTURE OF TRUNCATED HUMAN APOLIPOPROTEINA-I (APOA-I) (BORHANI ET AL. 1997) | 52 |
| FIGURE 1.19 APOA-IM DIMER PRODUCTION PROCESS. | 53 |
| FIGURE 2.1 REPRESENTATIVE MASS SPECTRA FOR ADSORBED PROTEIN SAMPLES. | 61 |
| FIGURE 2.2 RELATIONSHIP BETWEEN M/Z AND PEAK WIDTH. | 63 |
| FIGURE 2.3 THE EFFECT OF PEAK ALIGNMENT ON PARTITION REPRODUCIBILITY. | 65 |
| FIGURE 2.4 THE EFFECT OF PROTEIN CONCENTRATION ON PEAK INTENSITY. | 67 |
| FIGURE 2.5 THE EFFECT OF LASER STRENGTH ON PEAK INTENSITY. | 69 |
| FIGURE 2.6 THE EFFECT OF PROTEIN CONCENTRATION AND LASER STRENGTH ON RESOLUTION. | 70 |
| FIGURE 3.1 APOA-IM DIMER PRODUCTION PROCESS SAMPLES. | 74 |
| FIGURE 3.2 CALCULATION OF SUMMED PEAK INTENSITY. | 76 |

| | |
|--|-----|
| FIGURE 3.3 MASS SPECTRA FOR APOA-IM DIMER PRODUCTION PROCESS SAMPLES. | 79 |
| FIGURE 3.4 APOA-IM MONOMER QUANTIFICATION USING MASS SPECTROMETRY AND RP-HPLC. | 81 |
| FIGURE 3.5 MASS SPECTRA OF USD FLOCCULATION SAMPLES..... | 87 |
| FIGURE 4.1 MICROSCOPE IMAGES OF FLOCCULATED APOA-IM MATERIAL EXPOSED TO SHEAR. | 98 |
| FIGURE 4.2 DEVELOPING A MIMIC OF THE APOA-IM DIMER RECOVERY PROCESS..... | 98 |
| FIGURE 4.3 COMPARING THE USD % CLARIFICATION WITH PILOT SCALE | 100 |
| FIGURE 4.4 VARYING FLOCCULATING AGENT ADDITION RATE AT BOTH SCALES. | 102 |
| FIGURE 4.5 USING THE USD MIMIC TO STUDY THE EFFECT OF VARYING THE RATIO OF FLOCCULATING AGENT ADDITION ON THE LABORATORY CENTRIFUGE INDEX. | 104 |
| FIGURE 4.6 USING THE USD MIMIC TO STUDY THE EFFECT OF VARYING THE RATIO OF FLOCCULATING AGENT ADDITION ON THE % CLARIFICATION. | 105 |
| FIGURE 4.7 VALIDATING AT PILOT SCALE THE USD PREDICTIONS ON THE EFFECT OF FLOCCULATING AGENT ADDITION RATIO..... | 106 |
| FIGURE 5.1 STEPS FOR SELECTIVE PROTEINCHIP PREPARATION..... | 111 |
| FIGURE 5.2 COMPETITIVE ADSORPTION OF BSA AND LYSOZYME TO A HYDROPHOBIC PROTEINCHIP USING REVERSE PHASE SOLVENTS. | 113 |
| FIGURE 5.3 NON-COMPETITIVE ADSORPTION OF BSA TO A HYDROPHOBIC PROTEINCHIP USING TFA OR AMMONIUM SULPHATE. | 114 |
| FIGURE 5.4 COMPETITIVE ADSORPTION OF APOA-IM POST CENTRIFUGE SUPERNATANT PROTEINS TO A HYDROPHOBIC PROTEINCHIP WITH ACN OR AMMONIUM SULPHATE. | 116 |
| FIGURE 5.5 NON-COMPETITIVE ADSORPTION OF BSA TO A HYDROPHOBIC PROTEINCHIP BY VARYING AMMONIUM SULPHATE CONCENTRATION AND INCUBATION TIME. | 118 |
| FIGURE 5.6 VARYING THE FINAL WASH STEP FOR ADSORBED APOA-IM POST CENTRIFUGE SUPERNATANT PROTEINS. .. | 119 |
| FIGURE 5.7 CONDITION ARRAY FOR PROFILING POST CENTRIFUGE SUPERNATANT | 121 |
| FIGURE 5.8 MASS SPECTRA FOR ADSORBED APOA-IM POST CENTRIFUGE SUPERNATANT PROTEINS USING CONDITIONS IN FIGURE 5.7..... | 122 |
| FIGURE 6.1 OPTIONS FOR MASS SPECTRA DATA TREATMENT..... | 126 |
| FIGURE 6.2 IDENTIFICATION CODE FOR MASS SPECTRA FILES. | 128 |
| FIGURE 6.3 UPLOADING MASS SPECTRA DATA INTO THE MATLAB ENVIRONMENT..... | 129 |
| FIGURE 6.4 PROCESSING MASS SPECTRA DATA WITHIN THE MATLAB ENVIRONMENT. | 130 |
| FIGURE 6.5 RAW MASS SPECTRA FOR ADSORBED APOA-IM POST CENTRIFUGE SUPERNATANT PROTEINS. | 131 |
| FIGURE 6.6 MASS SPECTRA SIGNAL NOISE FOR ADSORBED APOA-IM POST CENTRIFUGE SUPERNATANT PROTEINS. | 132 |

FIGURE 6.7 DENOISED MASS SPECTRA FOR ADSORBED APOA-IM POST CENTRIFUGE SUPERNATANT PROTEINS. 133

FIGURE 6.8 DENOISED & BASELINED MASS SPECTRA FOR ADSORBED APOA-IM POST CENTRIFUGE SUPERNATANT PROTEINS..... 134

FIGURE 6.9 PEAK REPRODUCIBILITY BETWEEN SPOT PARTITIONS FOR ADSORBED APOA-IM POST CENTRIFUGE SUPERNATANT PROTEINS..... 136

FIGURE 6.10 PEAK REPRODUCIBILITY BETWEEN PROTEINCHIPS FOR ADSORBED APOA-IM POST CENTRIFUGE SUPERNATANT PROTEINS..... 138

FIGURE 7.1 A STAR PLOT EXAMPLE FROM THE MATHWORKS INC. (NATICK, MA, U.S.A.)..... 143

FIGURE 7.2 DENDROGRAM FORMATION 145

FIGURE 7.3 MASS SPECTRA FOR PROCESS CHROMATOGRAPHY SAMPLES..... 146

FIGURE 7.4 NON-COMPETITIVE ADSORPTION OF OVALBUMIN TO ELECTROSTATICALLY CHARGED SURFACES. 148

FIGURE 7.5 COMPETITIVE ADSORPTION OF HEN EGG WHITE PROTEINS TO ELECTROSTATICALLY CHARGED SURFACES. ... 149

FIGURE 7.6 STAR PLOTS FOR THIRTY PROTEINS CONTAINED WITHIN POST CENTRIFUGE SUPERNATANT. 151

FIGURE 7.7 DENDROGRAM TO SHOW RELATIONSHIPS BETWEEN POST CENTRIFUGE SUPERNATANT PROTEINS. 152

FIGURE 7.8 SELECTIVITY MAPS FOR PRODUCT AND NEAREST NEIGHBOUR CONTAMINANTS. 154

FIGURE 7.9 IMPURITY REMOVAL ANALYSIS FROM CHROMATOGRAPHY ELUTION FRACTIONS. 156

FIGURE 8.1 DEFINING CUT-OFF TO CLUSTER HARD TO REMOVE PROTEINS 164

FIGURE 9.1 THEORETICAL FRAMEWORK FOR THE RAPID EVALUATION OF PROTEIN CANDIDATES AT EARLY STAGE PROCESS DEVELOPMENT..... 168

LIST OF TABLES

| | |
|--|-----|
| TABLE 1.1 SEPARATION PRINCIPLES AND UNIT OPERATIONS USED IN DOWNSTREAM PROCESSING (LIENQUEO AND ASENJO,2000)..... | 17 |
| TABLE 1.2 INFLUENCE OF PROTEIN STABILITY ON ITS PURIFICATION (GE PROTEIN PURIFICATION,2003)..... | 21 |
| TABLE 1.3 COMPARING CHROMATOGRAPHY TECHNIQUES (GE PROTEIN PURIFICATION,2003)..... | 22 |
| TABLE 1.4 TECHNIQUES FOR DETERMINING PHYSICOCHEMICAL PROPERTIES..... | 29 |
| TABLE 1.5 CHROMATOGRAPHY MODE SELECTION BASED ON NEW PROTEOMIC TECHNIQUES OR CONVENTIONAL CHROMATOGRAPHY (WENGER ET AL. 2007)..... | 35 |
| TABLE 1.6 FACTORS IN SPA FORMULATION AND APPLICATION (CORDINGLEY ET AL. 2003)..... | 41 |
| TABLE 1.7 THE FOUR MAIN PROTEINS IN HEN EGG WHITE (VACHIER ET AL. 1995)..... | 50 |
| TABLE 2.1 SETTINGS USED WHEN PROCESSING DATA WITH CIPHERGEN EXPRESS 3.0..... | 60 |
| TABLE 3.1 PEAKS MANUALLY DETECTED IN MASS SPECTRA..... | 76 |
| TABLE 3.2 QUANTIFICATION OF MONOMER AND TRUNCATED SPECIES USING MASS SPECTROMETRY AND RP-HPLC..... | 80 |
| TABLE 3.3 QUANTIFICATION OF MONOMER, DIMER AND AGGREGATES USING MASS SPECTROMETRY AND SE-HPLC..... | 82 |
| TABLE 3.4 QUANTIFICATION OF HOST CELL PROTEINS USING MASS SPECTROMETRY AND ELISA..... | 83 |
| TABLE 3.5 RP-HPLC ANALYSIS BEFORE AND AFTER THE PILOT SCALE FLOCCULATION..... | 88 |
| TABLE 5.1 FACTORS IN SELDI PROTEINCHIP PREPARATION..... | 109 |
| TABLE 5.2 EQUILIBRATION AND FINAL WASH BUFFERS FOR THE SELECTIVE PROTEINCHIPS..... | 120 |

LIST OF EQUATIONS

| | |
|--|-----|
| EQUATION 1.1 EMPIRICAL DIMENSIONLESS RETENTION TIME (DRT)..... | 30 |
| EQUATION 1.2 THEORETICAL DIMENSIONLESS RETENTION TIME (DRT) (LIENQUEO ET AL. 2007). | 31 |
| EQUATION 1.3 EMPIRICAL PARTITION COEFFICIENTS, K_i | 32 |
| EQUATION 1.4 TIME OF FLIGHT MASS SPECTROMETER CALIBRATION..... | 45 |
| EQUATION 2.1 WIDTH OF MASS SPECTRA PEAKS. | 63 |
| EQUATION 4.1 % CLARIFICATION FOR CENTRIFUGATION. | 92 |
| EQUATION 4.2 SIGMA FOR A SWING-OUT LABORATORY CENTRIFUGE..... | 94 |
| EQUATION 4.3 SIGMA FOR A DISC-STACK CENTRIFUGE..... | 95 |
| EQUATION 4.4 COMPARING TWO DIFFERENT CENTRIFUGES. | 95 |
| EQUATION 4.5 HINDERED SETTLING CORRECTION FACTOR..... | 96 |
| EQUATION 7.1 EUCLIDEAN DISTANCE..... | 144 |
| EQUATION 7.2 CALCULATING THE NEW DISTANCE FROM CLUSTER AB TO PEAK MASS C USING UPGMA. | 144 |

CHAPTER 1 INTRODUCTION

1.1 Protein production processes

There are many different routes in the biopharmaceutical industry for recovering and purifying protein products. Whether the protein is produced extracellular or intracellular is critical to the route chosen (Figure 1.1). Extracellular proteins are generally easier to purify because they can be separated from whole cells without the need for cell breakage. This avoids the release of many host cell proteins that would be additional impurities to remove downstream. Methods for separation include centrifugation, microfiltration or expanded bed adsorption (EBA). Intracellular proteins require cell disruption to be released which could involve lysis or homogenisation (Table 1.1).

Image removed
due to copyright

Figure 1.1 Typical protein bioprocess options (Lienqueo and Asenjo,2000).

Table removed
due to copyright

Table 1.1 Separation principles and unit operations used in downstream processing (Lienqueo and Asenjo,2000).

One of the notable difficulties with process development is the inherent complexity caused by interactions between unit operations. Optimization of biopharmaceutical processes tend to be based on yield or product quality (Groep *et al.* 1997). Biopharmaceutical processing steps are highly interactive and optimising individual steps is unlikely to lead to an optimal process (Dyr and Suttner,1997;Groep *et al.* 1997). Groep *et al.* demonstrated this by varying conditions in the aerobic fermentation and homogenization for an alcohol dehydrogenase (ADH) production process grown intracellularly in *Saccharomyces cerevisiae* (Groep *et al.* 1997). An interaction was demonstrated between growth conditions, cell strength and cell disruption.

1.1.1 Protein recovery

There are many varied operations for protein recovery but this thesis will focus on flocculation and centrifugation. Flocculation allows cellular debris to form strong, stable aggregates which can then be removed from the liquid using a single unit operation such as centrifugation.

1.1.1.1 Flocculation

Flocculation describes the aggregation of individual particles in a suspension due to the addition of a flocculating agent. Early work in 1969 by McGregor *et al.* highlighted the factors that affect the flocculation of homogenous cultures. These were (i) bacteria genus; (ii) suspending media; (iii) temperature; (iv) physiological age; (v) flocculating agent (type, concentration & rate of addition); (vi) surface shear (McGregor and Finn,1969). Most of these variables were shown to be related to the release of proteins, polysaccharides, or nucleic acids.

Flocculation with natural polymers from micro-organisms, mineral flocculating agents such as aluminium sulphate or synthetic polymers are traditionally used in waste water treatment to remove microorganisms from the effluent of activated sludge (McGregor and Finn,1969). Additives such as aluminium sulphate have also been found to aid protein recovery from pure cultures such as those in the biopharmaceutical industry (Richardson *et al.* 1990). Gasner *et al.* found that strong polyelectrolytes as well as mineral hydrocolloids were the most effective (Gasner and Wang,1970). These mineral flocculating agents however tend to form low mechanical strength flocs rendering these agents useless for centrifugation (Shan *et al.* 1996).

Synthetic flocculating agents, such as the one used in this thesis consist of long polymeric chains that adsorb different particles in a suspension, aiding their aggregation. These carry various charged groups that cause adsorption of the particles. Once adsorbed the particles either bridge together or their charges are neutralised. The polymer used in this thesis was cationic, binding to negatively charged cellular debris to form flocs.

Dosing of these agents is very important and there is usually an optimum level of addition where the maximum amount of flocculation occurs. If insufficient agent is added only some of the particles will flocculate. An over dose of flocculant, however means many of the particles are coated with flocculant, causing the particles to have the same charge, leading to particle repulsions and preventing flocculation.

Floc particle strength is heavily influenced by surface shear during its formation. During floc formation a steady-state between floc growth and breakage is achieved. Mixing increases floc growth by increasing collision frequency but also breaks down the large floc particles due to shear force effects (Kim *et al.* 2001).

The main downside of adding a flocculating agent to a bioprocess is that regulators such as the FDA require evidence that the flocculant has been removed downstream prior to drug formulation. Theoretically, if the flocculation is carried out correctly most of the flocculant will be removed with the flocs during clarification. However, undoubtedly some flocculant will remain in the product stream and analytical strategies have to be devised to demonstrate its removal.

1.1.1.2 Centrifugation

Centrifugation uses centrifugal force to separate mixtures such as flocculated process material. It's frequently used in biopharmaceutical processes mainly due to its operational robustness and low running costs. Separation is achieved by density differences between heavy components (flocs) and lighter components (liquor) in the process material. Centrifuge designs include multi-chamber, tubular bowl and disc-stack. Disc-stack centrifuges have particular benefits as they allow intermittent solids discharge removing the need for manual dismantling of the centrifuge bowl to remove solids.

1.1.2 Protein purification

Purification typically defines the profitability of a process consuming around 60-70% of the total process operating costs (Dyr and Suttner,1997). The extent of

purification in a biopharmaceutical process is ultimately dictated by the product and its end use. An injectable drug will necessitate a great deal more purification than a food product. This is reflected in their respective manufacturing costs, which ultimately dictate their selling price. Hence, a biopharmaceutical company is constantly trying to meet drug purity requirements while minimising costs. This results in a constant trade-off between resolution, capacity and speed. Product will be lost at each step of the process so it is important to limit the number of steps to keep the overall yield high (Bonnerjea *et al.* 1986). The route with the least steps to give a higher yield is rarely found and non-optimal conditions are chosen in favour of increased speed to market (Dyr and Suttner,1997).

Limited knowledge of the molecular properties of process material and how to handle such data has made the creation of a rational method to separate target proteins from impurities difficult (Asenjo and Andrews,2004). There are many factors involved in the decision on which combination of techniques (Figure 1.2) to purify a target protein including the properties of the process material (Table 1.2). Chromatography techniques are frequently used due to their high purification power and relatively low cost.

Image removed
due to copyright

Figure 1.2 Operational sequence selection for protein purification (Asenjo and Andrews,2004).

Usually a chromatography sequence is designed so that minimal treatment is required between chromatography steps. For example, during ion-exchange the target protein is usually eluted in high salt conditions. It makes sense therefore to have a hydrophobic column following this step since binding in these columns also occurs with high salt concentrations.

Chromatography resins now exist that allow mixed-mode separation. These resins have both hydrophobic and electrostatic properties, allowing hydrophobic binding and electrostatic elution. This reduces the number of steps in a process by having a single step that can take the place of both hydrophobic interaction and ion-exchange chromatography.

| Sample and target protein properties | Influence on purification strategy |
|--------------------------------------|--|
| Temperature stability | Need to work rapidly at lowered temperature |
| pH stability | Selection of buffers for extraction and purification. Selection of conditions for ion exchange, affinity or reversed phase chromatography |
| Organic solvents stability | Selection of conditions for reversed phase chromatography |
| Detergent requirement | Consider effects on chromatographic steps and the need for detergent removal. Consider choice of detergent. |
| Salt (ionic strength) | Selection of conditions for precipitation techniques, ion exchange and hydrophobic interaction chromatography |
| Co-factors for stability or activity | Selection of additives, pH, salts, buffers |
| Protease sensitivity | Need for fast removal of proteases or addition of inhibitors |
| Sensitivity to metal ions | Need to add EDTA or EGTA to buffers |
| Redox sensitivity | Need to add reducing agents |
| Molecular weight | Selection of gel filtration media |
| Charge | Selection of ion exchange conditions |
| Biospecific affinity | Selection of ligand for affinity medium |
| Post-translational modifications | Selection of group-specific affinity medium |
| Hydrophobicity | Selection of medium for hydrophobic interaction chromatography |

Table 1.2 Influence of protein stability on its purification (GE protein purification, 2003).

1.1.2.1 Chromatography

Chromatography exploits differences in the physicochemical properties of molecules present in process feeds. The three most common types of chromatography utilise differences in charge (ion-exchange chromatography), hydrophobicity (hydrophobic interaction chromatography) and size (size exclusion chromatography). The choice of which type to use and at which stage of the process is dependant on a number of factors some of which are highlighted in Table 1.3.

| Technique | Main features | Capture | Intermediate | Polish | Sample start condition | Sample end condition |
|-----------|--|---------|--------------|--------|---|---|
| IEX | High resolution High capacity High speed | ☆☆☆ | ☆☆☆ | ☆☆☆ | Low ionic strength Sample volume Not limiting | High ionic strength or pH change concentrated |
| HIC | High resolution High capacity High speed | ☆☆ | ☆☆☆ | ☆ | High ionic strength Sample volume not limiting | Low ionic strength concentrated |
| SEC | High resolution | | ☆ | ☆☆☆ | Limited sample volume (<5% total column volume) and flow rate range | Buffer exchanged (if required) Diluted |

Table 1.3 Comparing chromatography techniques (GE protein purification,2003).

1.1.2.1.1 Ion-exchange

Ion-exchange (IEX) chromatography is potentially a very high resolution separation that has a high sample loading capacity (GE protein purification,2003). The separation is based on the reversible adsorption of a charged protein onto an oppositely charged chromatographic medium. Proteins are loaded onto a column and then conditions are altered so that bound substances are eluted differentially. This elution is usually performed by increases in salt concentration or changes in pH. Changes are made stepwise or with a continuous gradient. Most commonly,

samples are eluted with salt (NaCl), using a gradient elution. Target proteins are concentrated during adsorption and collected in a purified, concentrated form.

The net surface charge of proteins varies according to the surrounding pH. When above its isoelectric point (pI) a protein will bind to an anion-exchanger, when below its pI a protein will bind to a cation-exchanger. Typically, IEX chromatography is used to adsorb the target molecule, but can also be used to bind impurities. IEX chromatography can be repeated at different pH values to separate several proteins which have distinctly different charge properties.

1.1.2.1.2 Hydrophobic interaction

Protein binding in hydrophobic interaction chromatography (HIC) is promoted by moderately high concentrations of anti-chaotropic salts, which also have a stabilizing influence on protein structure (Figure 1.3). Elution can be achieved by a linear or stepwise decrease in the salt concentration of the mobile phase.

HIC works well in combination with ion-exchange and size exclusion chromatography because its principle of protein adsorption is complementary. Since it requires a minimum of sample pre-treatment steps it can also be well utilised in processes with preceding precipitation steps. Its sensitivity is exhibited by the fact that even non-polar groups deep within the tertiary structure of a molecule can interact if they are incorrectly folded or damaged. This is particularly valuable when it is necessary to purify the native form of a molecule from close derivatives.

Early developed adsorbents appeared to show a mixed mode of hydrophobic and ionic interaction but these were later improved by removing charged groups present in the sorbents allowing high binding of neutral proteins at high salt concentrations to be seen. Elution was also as expected by using a salt-free buffer or by decreasing the polarity of the eluent.

HIC sorbents are usually composed of cellulose, dextran, or cross-linked agarose. Sorbents on the market include Phenyl and Octyl Sepharose fast flow,

Sepharose high performance and Superose. These meet various demands on chromatographic productivity, selectivity and efficiency.

Possible theories for hydrophobic interaction:

1. A “salting-out” effect in hydrophobic adsorption with the driving force being the entropy gained from structure changes in the water surrounding the interacting hydrophobic groups. The displacement of the ordered water molecules surrounding the hydrophobic ligands and the proteins leads to an increase in entropy (ΔS) resulting in a negative value for the change in free energy (ΔG) of the system. This implies that the hydrophobic ligand-protein interaction is thermodynamically favourable (Figure 1.3).
2. The increase in surface tension of the water arising from the structure – forming salts dissolved in it.
3. Van der Waals attraction forces between protein and ligand increase as the ordered structure of water increases in the presence of salting out salts.

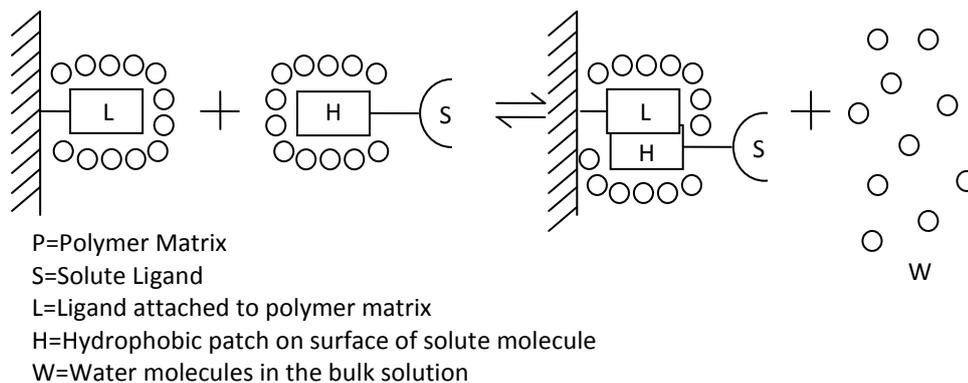


Figure 1.3 A theoretical mechanism for hydrophobic interactions.

Close to the surface of the hydrophobic ligand and solute (L and H), the water molecules are more highly ordered than in the bulk water and appear to “shield off” the hydrophobic ligand and solute molecules. Added salt interacts strongly with the water molecules leaving less water available for the “shielding off” effect, which is

the driving force for L and H to interact with each other (GE HIC principle and methods,2003).

1.1.2.1.3 Immobilised metal affinity

Immobilized metal affinity is when amino acids covalently bind to immobilised metals ions. An amino acid that does this readily is histidine. In a chromatography format the immobilised metal affinity sorbent is first charged with metal ions such as nickel (Ni^{2+}) and copper (Cu^{2+}). Proteins with an affinity for these ions interact with them and are retained by the sorbent while proteins with low affinity flow through. Changes in pH or competitive molecules like imidazole can then elute these proteins from the sorbent.

Proteins that do not possess an affinity for metal ions can be engineered using recombinant DNA techniques in order to acquire this property. The DNA coding the protein of interest is inserted in a vector adjacent to a polyhistidine coding DNA sequence. During translation, a polyhistidine tag is therefore added to the C-terminal or N-terminal of the protein of interest giving it an affinity for metal ions.

1.1.2.1.4 Size exclusion

Size exclusion (SE) chromatography separates molecules according to differences in molecular weight (MW) (Figure 1.4) (GE gel filtration principles and methods,2003). Size exclusion is ideal for bio-molecules that maybe sensitive to pH, metal ion or co-factor concentration, and harsh environmental conditions. In contrast to IEX and HIC, there is no adsorption between molecules and the media, so buffer composition does not affect resolution. This means SE can be used for buffer exchange and separations can be performed in the presence of essential ions, cofactors, detergents, urea, guanidine hydrochloride, under variable ionic strengths and temperatures.

The media for size exclusion are porous spherical particles selected for their physicochemical stability, and inertness.

A major disadvantage of using size exclusion chromatography in a process is that only the small volumes can be loaded, typically less than 10%, meaning larger volumes of resin/buffers and longer operating times are required. For this reason, SEC is usually only used as a final step in a process.

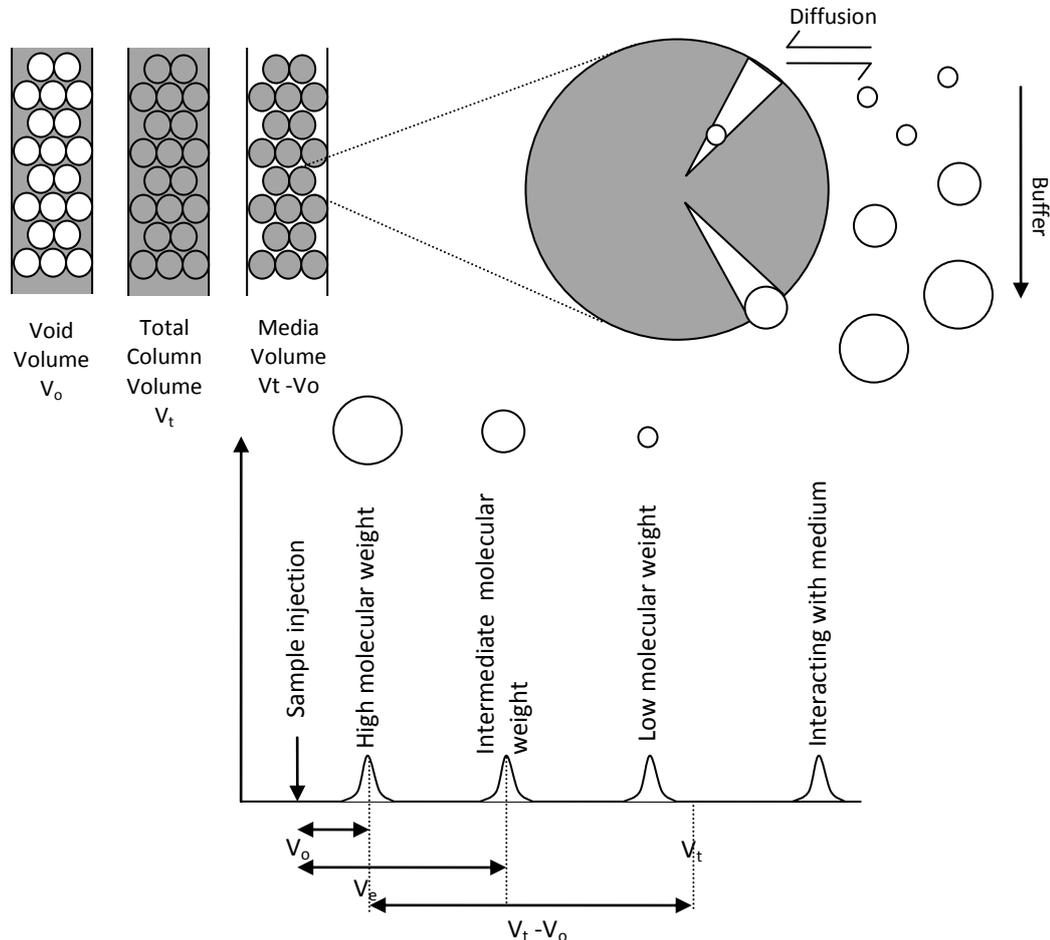


Figure 1.4 The theory behind gel filtration (GE gel filtration principles and methods,2003).

Equilibration is carried out with buffer which fills the inter- and intra-particle spaces. Intra-particle fluid (stationary phase) is in equilibrium with inter-particle fluid (mobile phase). Samples are eluted isocratically (constant buffer composition).

1.2 Ultra scale-down (USD) technologies to aid process development

USD technologies are beneficial as they can simulate large scale equipment while only requiring millilitre volumes of material. Thus, predictive performance of

large scale operations can be obtained at a fraction of the time and volume of material required compared with conventional approaches (Boychyn et al. 2004; Boychyn et al. 2000; Neal et al. 2003; Reynolds et al. 2003; Willoughby et al. 2004). USD technologies are particularly beneficial in early development when typically only small quantities of material are available.

The USD approach for determining the sensitivity of the flocculated suspensions used in this thesis was a rotating disc device developed at UCL. The USD technology used for determining physicochemical properties of proteins used in this thesis was surface enhanced laser desorption ionisation mass spectroscopy (SELDI-MS).

1.2.1 Determining the shear sensitivity of process material

Biopharmaceutical processes typically exert a lot of stress on process particulates which can lead to cell breakage or protein aggregation (Cromwell *et al.* 2006). Determining the shear sensitivity of these particulates can guide equipment selection and the conditions used during process development. Shear cells are ultra-scale down devices that can determine shear sensitivity (Biddlecombe et al. 2007; Boychyn et al. 2001; Boychyn et al. 2004; Boychyn et al. 2000; Hutchinson et al. 2006; Maybury et al. 2000; Salte et al. 2005; Tustian et al. 2007). The bioprocess material is placed in the shear device and exposed to different levels of shear. The amount of particulate damage is then analysed using techniques such as UV spectroscopy. This can then be modelled in the pilot scale.

A pilot scale centrifuge can be modelled in the laboratory by using a shear device to simulate the shear experienced by particulates as they enter the pilot scale centrifuge feed zone followed by their separation with a laboratory centrifuge.

1.2.2 Determining the physicochemical properties of proteins

It has been the aim of many researchers to predict the physicochemical properties of protein populations (Wall et al. 2002; Xu and Glatz, 2009). Typical properties investigated are size, charge and hydrophobicity (Figure 1.5).

Image removed
due to copyright

Figure 1.5 Mapping the physicochemical properties of proteins (Wall et al. 2002).

The common properties are charge, molecular weight and hydrophobicity measured respectively as the pI, MW and the percentage of acetonitrile required to elute each protein from an RP-HPLC column. The relative shades of the data points are related to electrospray ionisation-time of flight-mass spectrometry (ESI-TOF-MS) intensity. ESI-TOF-MS can give relative quantification between peaks in different samples.

There are many techniques for determining physicochemical properties of proteins (table 1.4).

| Method | Derived | Throughput | Material required per run | Impurity ID | Cost per run | Reproducibility |
|------------------|---------------|------------|---------------------------|----------------------|--------------|-----------------|
| Structure models | Theoretically | High | N/A | Yes | Low | N/A |
| SELDI | Empirically | High | Low | Yes | High | Medium |
| Gel Techniques | Empirically | Low | Medium | Yes (low resolution) | Medium | Medium |
| Microwell /Resin | Empirically | Medium | Medium | No | Medium | Medium |
| Packed Column | Empirically | Low | High | No | Low | High |

Table 1.4 Techniques for determining physicochemical properties.

1.2.2.1 Theoretically derived

Many groups are working on linking a proteins known structure with its physicochemical properties. In 2002, Berggren *et al.*, investigated the link between the number and type of surface exposed amino acids in a protein with its partitioning in an aqueous two-phase system (Berggren *et al.* 2002). The partitioning was between two EO30PO70-dextran aqueous phases of varying concentration. It was found that aromatic amino acids had the biggest effect, moving to the upper EO30PO70 phase and charged proteins tended to move into the lower dextran phase.

Cramer *et al.* have looked at quantitative structure property relationship (QSPR) models (Chen *et al.* 2007;Chen *et al.* 2008;Chen and Cramer,2007;Cramer and Jayaraman,1993;Ladiwala *et al.* 2005;Ladiwala *et al.* 2006;Yang *et al.* 2007a;Yang *et al.* 2007b). This approach uses non-linear support vector machine (SVM) techniques to correlate experimental data with physicochemical attributes. This method has had some success in predicting the retention of proteins in linear gradient hydrophobic interaction chromatography (Chen *et al.* 2008) and high salt binding on ion exchange resins (Yang *et al.* 2007a). The technique generates descriptors of a protein by its three dimensional structure. These descriptors are then grouped and ranked on their ability to predict experimental data for a training set of proteins. A priori predictions are then made based on these descriptors for a

test set of proteins not included in the models. A group at GE Healthcare has also used QSPR to predict retention times (Malmquist *et al.* 2006).

Asenjo *et al.* used structural data to predict the empirical protein retention time (Equation 1.1) for hydrophobic interaction chromatography (Lienqueo *et al.* 2002;Lienqueo *et al.* 2003;Lienqueo *et al.* 2006;Lienqueo *et al.* 2007;Mahn *et al.* 2004;Mahn *et al.* 2005;Mahn *et al.* 2007;Mahn and Asenjo,2005;Salgado *et al.* 2005a;Salgado *et al.* 2005b;Salgado *et al.* 2006a;Salgado *et al.* 2006b;Salgado *et al.* 2008).

$$DRT = \frac{t_r - t_0}{t_f - t_0}$$

Equation 1.1 Empirical dimensionless retention time (DRT).

t_r is the time corresponding to the retention time of the target protein; *t₀* is the time corresponding to the start of the elution gradient; *t_f* is the time corresponding to the end of the salt gradient. If a protein is not retained by the resin, DRT is equal to 0, and if a protein elutes only after the gradient has been completed, its DRT is equal to 1.

They calculated various surface hydrophobicity measurements based on the three dimensional structure and the hydrophobicity of the exposed amino acid residues to predict the dimensionless retention time (DRT). Depending on the information available and the hydrophobicity distribution, different models were used for calculating theoretical DRT (Figure 1.6). An example of one of the formulas used is shown in Equation 1.2.

Image removed
due to copyright

Figure 1.6 Predicting dimensionless retention time (Lienqueo et al. 2007).

The different methodologies for computational experiments to determine the suitable purification of a target protein. Where LH is local hydrophobicity; HCA is hydrophobic contact area; HI is hydrophobic imbalance; DRT III is the linear estimation of the amino acid surface composition.

Image removed
due to copyright

Equation 1.2 Theoretical dimensionless retention time (DRT) (Lienqueo et al. 2007).

Where $\Phi_{surface}$ is the average surface hydrophobicity value; s_{aai} is the solvent accessible area occupied by amino acid 'i'; φ_{aai} is the hydrophobicity value for amino

acid 'i'; s_p is the total solvent accessible area of the protein; A , B and C are constants for each set of operating conditions (e.g. varying resins).

1.2.2.2 Empirically derived

Techniques to derive physicochemical properties experimentally include aqueous biphasic systems (ABS), two dimensional-polyacrylamide gel electrophoresis (2D-PAGE) and reverse phase (RP) HPLC.

1.2.2.2.1 Aqueous Biphasic Systems

Aqueous biphasic systems are produced when two polymers, one polymer and a kosmotropic salt, or a chaotrope and kosmotrope are mixed together at an appropriate concentration or temperature. The proteins to be separated are then added. The greater a protein's hydrophobicity, the greater its abundance in the more hydrophobic phase and hence the greater its partition coefficient, K (Equation 1.3). Unlike non-polar organic phase systems these two immiscible phases are water based and avoid the introduction of volatile organic compounds which can damage and denature biological molecules. Aqueous systems maintain protein stability; can partition individual proteins in mixtures; and give hydrophobic information (Gu and Glatz,2007).

$$K_i = \frac{C_i^{top}}{C_i^{bottom}}$$

Equation 1.3 Empirical partition coefficients, K_i .

Where C is the concentration of protein i in the top and bottom phases. Relatively hydrophobic proteins have $\text{Log}(K) > 0$.

Glatz *et al.* used these systems to determine hydrophobic descriptors for a number of proteins (Xu and Glatz,2009). The upper phase they used was formed by polyethylene glycol (PEG) which was more hydrophobic than their lower phase solution of sodium sulphate and sodium chloride.

1.2.2.2.2 Two dimensional gel electrophoresis

Glatz *et al.*, used 2D-PAGE (an example of a 2D-PAGE gel is shown in Figure 1.7) to determine the MW and pI of proteins. This information along with aqueous biphasic hydrophobicity data was used to construct a three dimensional property map of all the proteins in a corn extract sample.

Typically, SDS-PAGE alone can resolve about one hundred distinct proteins but with the addition of a second dimension (charge) many more proteins can be observed. One of the main advantages of 2D-PAGE analysis is its ability to resolve proteins in very complex materials (Figure 1.7).

The 2D-PAGE approach relies on proteins first undergoing isoelectric focussing using a tube or strip gel with a pH gradient to deduce the first dimension of pI. This gel is then laid horizontally next to an SDS slab gel and an electric current applied to get the second dimension of MW.

Image removed
due to copyright

Figure 1.7 A two dimensional polyacrylamide gel (Ou et al. 2001).

The 2D-PAGE gel shows the protein complexity of a human hepatocellular carcinoma cell line.

1.2.2.2.3 RP-HPLC

The elution time for proteins in RP-HPLC is related to their hydrophobicity. The percent acetonitrile at time of elution is related to the ratio of non-polar to polar amino acids (Wall et al. 2002).

1.2.3 Sorbent screening

Instead of determining physicochemical properties to predict chromatography behaviour, proteins can be run directly on a variety of sorbents and conditions using USD methods. This can be done using micro-batch adsorption or micro-pipette columns (Table 1.5).

Table removed
due to copyright

Table 1.5 Chromatography mode selection based on new proteomic techniques or conventional chromatography (Wenger et al. 2007).

1.2.3.1 Micro-batch adsorption

Companies use many different sorbents and comparison studies are timely due to the running of many columns, analysis of many fractions and the cost of the sorbents used (Weinberger et al. 2002a). Filter plates can be filled with various sorbents and mixed with the sample under various binding and elution conditions

(Figure 1.8) (Guerrier et al. 2007;Rege et al. 2006). Flow-throughs, washes and eluates can be extracted through the filter using a vacuum or centrifuge and then analysed using UV spectroscopy or HPLC (Weinberger et al. 2002a).

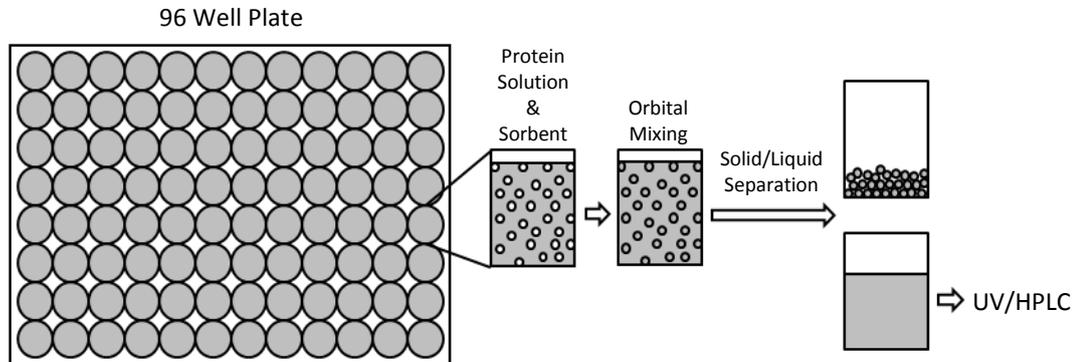


Figure 1.8 Experimental procedure for micro-batch adsorption.

1.2.3.2 Micro-pipette column

Small scale columns (less than 1mL) are available that can improve the throughput of sorbent scouting while maintaining dynamic conditions. An adaptation of this technique includes pipette tips packed with sorbent (PhyTips™) (Figure 1.9). Using these tips in conjunction with a robotic device, Wenger *et al.*, achieved a 10-fold increase in throughput (Wenger et al. 2007).

Image removed
due to copyright

Figure 1.9 A micro-pipette column (Wenger et al. 2007).

A driving force for this 1000-fold reduction in scale was to increase throughput and reduce labour. Conditions in a *Saccharomyces cerevisiae* fermentation for a recombinant human papillomavirus vaccine were altered and a

following cation exchange chromatography step carried out in a laboratory column and micro-pipette tip format. The elution fractions were analysed and fermentation productivity was expressed as the total protein recovered after chromatography per input of cell weight. There was a good correlation for productivity between the laboratory column and the micro-pipette tips (Figure 1.10).

Image removed
due to copyright

Figure 1.10 Comparison of different chromatography scales (Wenger et al. 2007).

Cation exchange purification was used with an automated micro-scale and a laboratory purification to assess the fermentation productivity (total protein recovery after chromatography per cell weight input) across a wide range of cell growth and induction conditions.

1.3 SELDI-MS

Surface enhanced laser desorption/ionisation time of flight mass spectrometry (SELDI-TOF-MS) is a new high throughput analytical method developed by Hutchens and Yip (Hutchens and Yip,1993). It promotes the revolutionary concept of derivatizing the mass spectrometric probe surface with functional groups to allow surface enhanced affinity capture (SEAC) of components

in a sample (Merchant and Weinberger,2000). These retained components are then analysed using mass spectrometry. This allows a simple all-in-one platform for reactions along with a highly sensitive identifiable detection (Caputo et al. 2003).

1.3.1 Surface enhanced affinity capture

The probe surface plays an active role in the extraction, presentation, structural modification, and/or amplification of components in the sample (Merchant and Weinberger,2000). Biomolecules bind to these surfaces due to hydrophobic, electrostatic, coordinate covalent bond or Lewis-acid/base interactions (Merchant and Weinberger,2000). Creation of specific biochemical surfaces is possible for SEAC, including antibodies, receptors, enzymes, DNA, small molecules, ligands and lectins (Merchant and Weinberger,2000). This allows compounds with shared physicochemical properties to be retained; thus, the various analytes are likely to have a more equal probability of becoming incorporated in the matrix crystal, which in turn reduces analyte signal suppression (Merchant and Weinberger,2000). Each ProteinChip has several of these active surfaces, all possessing the same functional group (Barzaghi et al. 2004;Weinberger et al. 2002a). ProteinChips can mimic chromatography media by altering their selectivity using different binding buffers for application and washing (Figure 1.11).

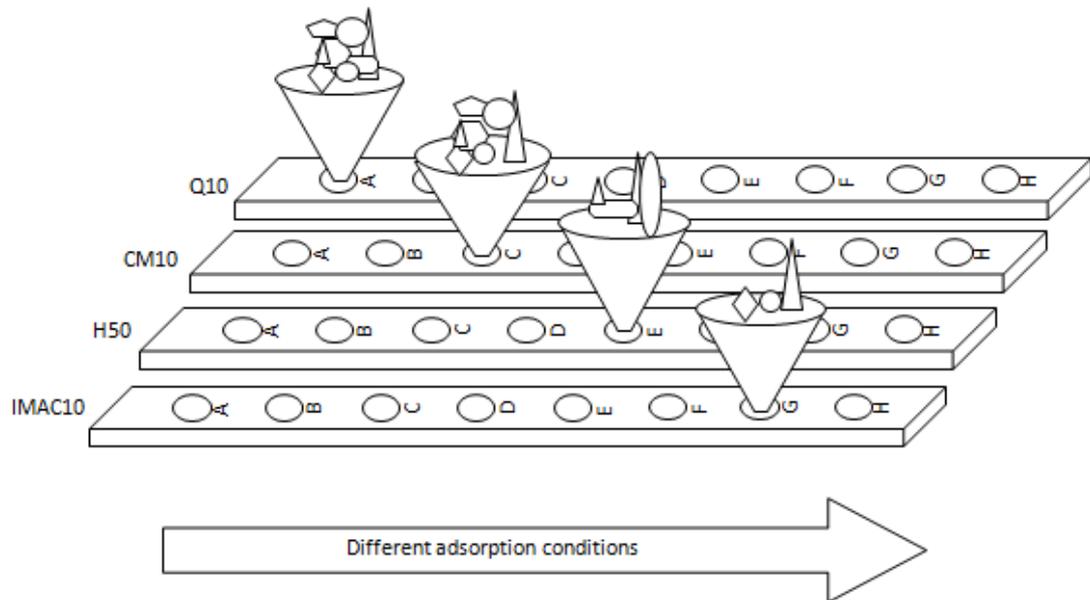


Figure 1.11 SELDI selective ProteinChip preparation.

Four ProteinChip arrays, each with distinct affinity capture properties are aligned side by side. Each array has eight specific regions or “spots” with defined surface chemistry allowing eight different experimental conditions. Different proteins (shapes) are retained on different arrays due to the specific surface chemistries. Different adsorption conditions (e.g. pH and/or [NaCl]) influence selectivity on a given array (Merchant and Weinberger,2000).

The SELDI ProteinChip arrays are produced by Bio-Rad Laboratories Inc. (Hercules, CA, U.S.A.) and consist of thin pieces of stainless steel with chemically active 1mm diameter spots (Figure 1.12). Cation-exchange (CM10, a carboxy methyl surface), anion-exchange (Q10, a quaternary amine surface), hydrophobic (H50) and immobilised metal affinity (IMAC10) are known as selective ProteinChips and the normal phase (NP20) as non-selective. The normal phase (NP20) arrays have a silicon oxide surface that hydrophilic and charged protein residues can bind to by direct pipetting of the sample to the surface (Favre-Kontula et al. 2008). The selective ProteinChips are prepared by exposing them to the sample in a bioprocessor (Figure 1.12).

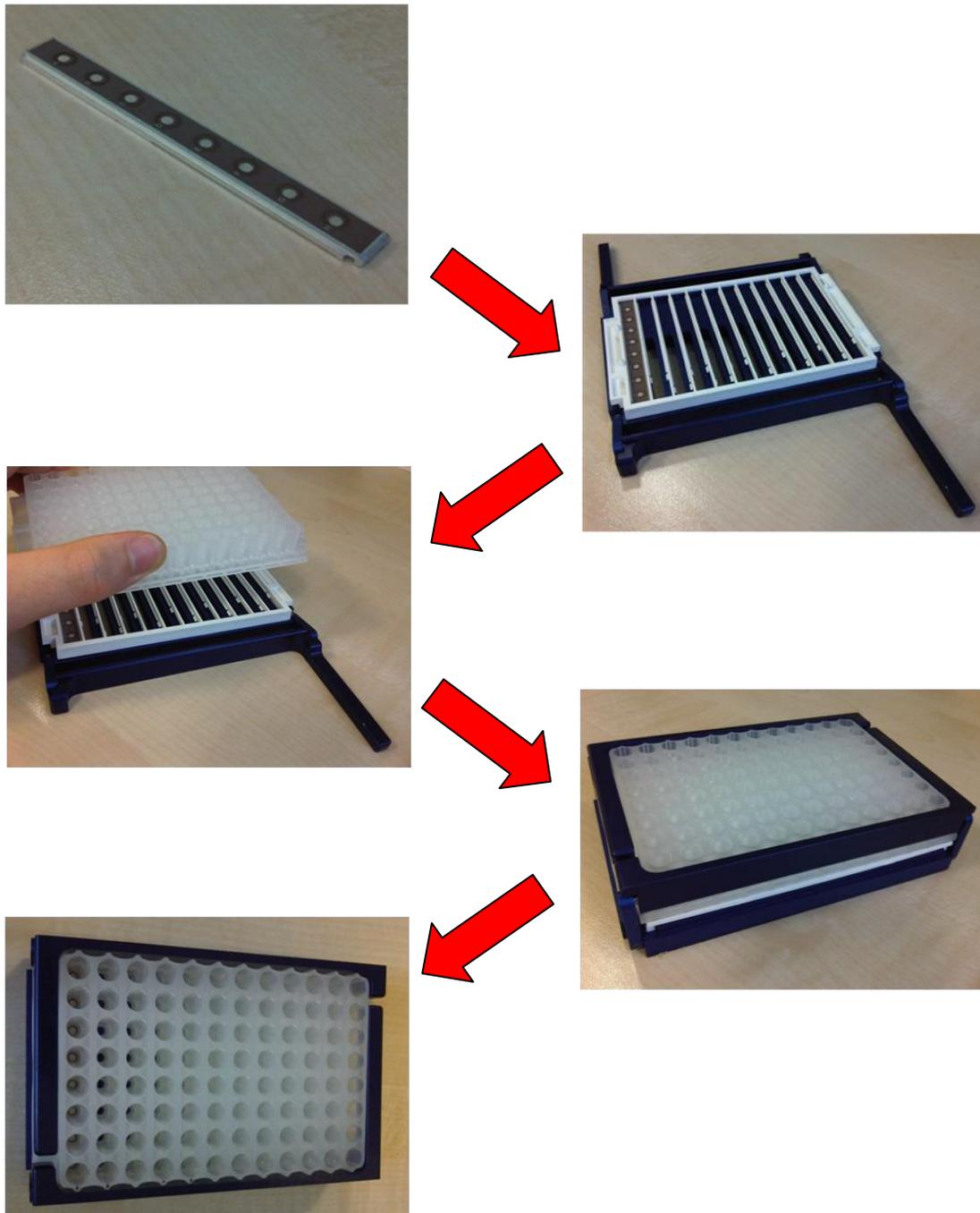


Figure 1.12 Using the SELDI bioprocessor.

Selective ProteinChips are inserted into a plastic cassette. A bottomless plastic 96 well reservoir is then placed over the top that fits neatly over each spot on the ProteinChip. An aluminium frame clips the whole assembly into place. This complete unit is called a bioprocessor and it allows the isolation of each spot for the application of different binding buffers.

Washing steps on the ProteinChip allow the removal of unbound or weakly bound molecules and salts that can interfere with ProteinChip analysis. Once the ProteinChips have undergone washing they can be air-dried ready for the addition of the energy absorbing molecule (EAM). The EAM is added in excess to the ProteinChips and causes crystallization of proteins present allowing better protein ionisation during the laser desorption/ionisation (LDI) stage.

Selection of the EAM depends on the sample under investigation but it is commonly sinapinic acid (SPA) dissolved in trifluoroacetic (TFA) acid and acetonitrile (ACN). The application of EAM allows better protein ionisation, enhances signal sensitivity, reproducibility and quantification (Weinberger et al. 2002a).

Work by Cordingley *et al.* looked at the effects of factors involved in EAM addition in a fractional factorial design of experiments approach on several developed metrics reflecting SELDI trace quality and reproducibility (Cordingley et al. 2003). This allowed them to investigate each factor and the interactions between factors in order to select the optimised level of each for the best possible traces. The factors chosen are shown in Table 1.6

Table removed
due to copyright

Table 1.6 Factors in SPA formulation and application (Cordingley et al. 2003).

They found that the best conditions for SPA addition were increased time between last wash and SPA application, TFA/ACN mixture made just prior to

application, high precision pipetting, and two applications of a relatively low volume of SPA. The age of the SPA did not seem to have a significant effect.

1.3.2 Laser desorption/ionisation

LDI analysis has been around since the 1960s. Initially it was used to study small inorganic salts and organic molecules but was later extended to large biopolymers with the arrival of matrix assisted LDI (MALDI) (Merchant and Weinberger,2000;Tang et al. 2004;Weinberger et al. 2002b;Weinberger et al. 2002a;Weinberger et al. 2002c). Unlike SELDI, the sample probe used with MALDI is passive, its sole purpose is to facilitate presentation of all of the molecules in the sample to the mass spectrometer (Merchant and Weinberger,2000).

LDI occurs when the sample directly absorbs energy from the laser and heats up via direct or secondary thermal changes, thus producing desorbed gaseous ions (Merchant and Weinberger,2000). Energy is typically provided for desorption as a pulsed UV laser (Merchant and Weinberger,2000). Commonly a nitrogen laser ($\lambda = 337\text{nm}$) is used, due to a simple footprint and relatively low cost (Merchant and Weinberger,2000). An optical attenuator controls laser fluence, and an optical beam splitter is used to divert some of the laser beam to a high speed photodetector that is used to trigger time of flight (TOF) measurement (Merchant and Weinberger,2000).

A ProteinChip spot is divided into pixels and these are grouped into partitions. Typically the spot is divided into four of these partitions allowing different acquisition settings to be used for each. This means, for the first partition the laser fires at every fourth pixel starting with the first and for the second partition the laser fires at every fourth pixel starting with the second and so on. The advantage of partitioning is that four different acquisition settings can be used to get four data sets representative of the whole spot surface. For instance, two partitions can be used with different laser strengths to get low and high m/z data leaving two partitions for other acquisition settings if deemed necessary.

1.3.3 Time of flight mass spectrometry

The SELDI ProteinChip reader used in this thesis is shown in Figure 1.13 along with schematics on its operation (Figure 1.14 & Figure 1.15). Once the laser has generated the gaseous ions they enter an acceleration stage (Figure 1.15). This acceleration stage consists of an ion acceleration region containing an ion optic assembly, and an ion-free flight region comprised of an ion drift tube (Merchant and Weinberger,2000). Two different ions with respective masses, m_1 and m_2 ($m_2 > m_1$) are created at the same time and location with the same charge, z (Merchant and Weinberger,2000). The repeller is raised to some potential V while the ground aperture is held at ground potential (Merchant and Weinberger,2000). If m_1 and m_2 are cations then the repeller plate is raised to a positive potential and a negative potential if they are anions (Merchant and Weinberger,2000). In this regard, only a single polarity of ions can be analysed at any one time (Merchant and Weinberger,2000). Due to the creation of this electric field, ions are accelerated to a constant final energy along distance s , through the opening in the ground aperture, and out into the drift tube region of the mass spectrometer (Merchant and Weinberger,2000). Ions continue to travel through free flight distance x prior to striking the system's detector (Merchant and Weinberger,2000).



Figure 1.13 The SELDI ProteinChip Reader

Image removed
due to copyright

Figure 1.14 The SELDI mass spectrometer (Merchant and Weinberger,2000).

Image removed
due to copyright

Figure 1.15 Time of flight in a mass spectrometer (Merchant and Weinberger,2000).

Signal processing is promoted by a high-speed analog-to-digital converter (A/D) typically linked to a personal computer running a system control and data

reduction software program (Merchant and Weinberger,2000). Detected analyte is displayed as a peak whose amplitude or area is proportional to abundance. Time of flight (TOF) is related to the ion analyte mass-to-charge ratio (m/z) (Merchant and Weinberger,2000). The individual mass spectral peaks generated correspond to the molecular weights of components retained on each spot; the effective molecular mass range of the instrument is 1-500kDa (Weinberger et al. 2002a).

Calibration is achieved by correlating the flight time for a number of well-characterized analytes with their established m/z values (Merchant and Weinberger,2000). A least-squares fit algorithm is applied to determine slope and intercept values for the linear relationship between m/z and the square of total flight time (Merchant and Weinberger,2000). A calibration expression similar to that of Equation 1.4 is subsequently used to determine the m/z for all unknown samples.

$$\frac{m}{z} = at_i^2 + /- b$$

Equation 1.4 Time of flight mass spectrometer calibration.

Where m is the mass of the ion; z is the ion charge; t is the time of flight; a and b are calibration constants.

Two calibration strategies are typically employed: internal and external standard calibration (Merchant and Weinberger,2000). For internal calibration, the sample of interest is spiked with at least one calibrant and the sample of interest is simultaneously analysed with the added calibrants in a single experiment (Merchant and Weinberger,2000). The resultant spectrum from this measurement is calibrated using known m/z values for each calibrant, allowing a highly accurate m/z determination for the unknown (Merchant and Weinberger,2000). During external calibration, calibrants and unknowns are analyzed in independent experiments (Merchant and Weinberger,2000). The calibration function derived during the calibration experiment is applied to the unknown spectrum to provide an accurate determination of m/z (Merchant and Weinberger,2000). Due to finite

differences in ion acceleration potential, sample location, and free flight differences, internal calibration generally provides an error of less than 100ppm compared to an error of 500-1000ppm with the external calibration(Merchant and Weinberger,2000).

1.3.4 Potential uses

The majority of SELDI-MS publications are related to biomarker discovery but there is also literature on its use in fermentation/cell-culture optimization (Park et al. 2006); purification development (Merchant and Weinberger,2000); process monitoring (Kumar *et al.* 2008) and product analysis (Woolley and Al Rubeai,2009) including on-chip digestion (Caputo et al. 2003;Lodish et al. 1999).

1.3.4.1 Fermentation/cell culture optimisation

SELDI-MS has been used to measure the production of an anti-botulinum neurotoxin antibody fragment (bt-Fab) during its fermentation (Park et al. 2006). Linked with a statistical experimental design IMAC (Ni²⁺) ProteinChips could be used to find fermentation conditions that could improve bt-Fab production. An immuno-affinity assay corroborated these conclusions. SELDI-MS has also been used to select the best strategy for the optimal expression of a recombinant S-LAT mouse protein and were able to discriminate between a 23.9 and a 24.1 kDa -N and -C terminal construct of the protein (Brenac *et al.* 2006).

1.3.4.2 Process monitoring and product analysis

SELDI-MS can be used for analysis of proteins in process liquid streams (Weinberger et al. 2002a). Normal phase ProteinChips have silicate surfaces that retain proteins through hydrogen bonding and/or Van der Waals forces (Weinberger et al. 2002a). During a process, SELDI-MS gives a highly sensitive analysis of process components and can also reveal important information about protein integrity (e.g. proteolytic degradation, glycosylation) that can be exploited

for the purpose of process optimisation, lot acceptance criteria, or product specifications (Weinberger et al. 2002a).

One group has investigated using SELDI-MS in parallel with SDS-PAGE to analyze fractions in order to optimize chromatography runs (Brenac, V et al. 2008). In this work it was possible to track both the target protein and host cell proteins (Brenac, V et al. 2008;Weinberger et al. 2002a). Other groups have also used SELDI-MS to evaluate purification steps (Fortis et al. 2006;Guerrier et al. 2005;Guerrier et al. 2007;Wierling et al. 2007).

1.3.4.3 Purification development

Adsorption to SELDI surfaces may reveal the biochemical property of the analyte, including chemical properties such as hydrophobicity, total charge, pI, phosphorylation, glycosylation, and primary composition (Merchant and Weinberger,2000). In situ clean-up diminishes sample loss by eliminating non-specific binding and dilutions inherent with traditional column chromatography (Merchant and Weinberger,2000). SELDI-MS also eliminates fractional analysis present with conventional chromatography due to proteins on the ProteinChips being selected and analysed directly (Weinberger et al. 2002a).

Chromatography mode decisions can be made and scaled up to commercially available sorbents with similar chemical properties as the ProteinChips (Weinberger *et al.* 2002a). A study by Weinberger *et al.*, used SELDI ProteinChips to predict the chromatography step for endostatin purification from *Pichia pastoris* (Weinberger *et al.* 2002a). They predicted that cation exchange chromatography at pH 5 promoted optimal binding of endostatin and full desorption occurred at greater or equal to 300mM of NaCl. Scale up of these conditions to a lab-scale column with CM HyperZ sorbent showed that endostatin purity was increased from less than 10% in the initial feedstock to greater than 90% after the column (Weinberger et al. 2002a). They also found polishing conditions with an IMAC (Cu²⁺) array that could reduce the small molecular weight components that were still present after cation exchange (Weinberger et al. 2002a).

1.3.5 Advantages

Historically, protein analysis has been carried out with 2D-PAGE. However 2D-PAGE lacks reproducibility, fails to resolve most proteins outside 5-100kDa and the molecular weight measurement is typically only accurate to $\pm 5\%$. Many protein spots need further analysis with MS and the technique has difficulty in separating membrane proteins.

The advantage of using SELDI over other mass spectrometry techniques such as MALDI is selectivity. This allows easy wash removal of salts, detergent, buffer, and organics so interference common with MALDI is reduced with minimal loss of sample (Caputo et al. 2003). This also means that samples incompatible with MALDI can still be used with SELDI with very little sample modification (Caputo et al. 2003; Favre-Kontula et al. 2008; Weinberger et al. 2002a).

The energy absorbing molecule (EAM) is applied to the proteins already homogeneously presented on the ProteinChips surface (Weinberger et al. 2002a). MALDI however mixes the EAM with the sample before sample application to an inert surface, this inherently causes non-homogeneity of the sample (Weinberger et al. 2002a). This results in less efficient desorption and consequently makes reproducible detection, identification and quantification of the proteins in complex mixtures difficult (Weinberger et al. 2002a). SELDI-MS is sensitive enough for detection and quantification of proteins in the range of 1-50 femtomoles per protein with remarkably fine molecular weight based resolution (Weinberger et al. 2002a).

SELDI-MS data acquisition and analysis can be as little as 30 minutes, providing an analytical technique that can match high throughput sorbent screening reducing the time for chromatography sorbent evaluation (Brenac, V et al. 2008). Weinberger *et al.*, took less than 5 days and 10mL of sample to screen 200 sorbents some of which showed target protein or impurity capture (Weinberger et al. 2002a).

1.3.6 Limitations

Ion suppression, sample matrix, coeluting compounds, and cross-talk can all affect the quantitative performance of a mass detector (Annesley,2003). Ion suppression is caused by non-volatile material such as salts, ion-pairing agents, endogenous compounds and drugs/metabolites. These materials change the efficiency of droplet formation and/or surface ionization. Consequently, this affects the amount of charged ions in the gas phase that ultimately reach the detector.

Only relative quantification with SELDI-MS can be achieved unless the ProteinChip is spiked with known concentrations of the protein under investigation.

1.4 Host organism

The recombinant host used in this thesis was *Escherichia coli* (*E.coli*). *E.coli* is gram negative, motile and rod-shaped belonging to the *Enterobacteriaceae* family. The bacterium is a well established host within the pharmaceutical industry and has become one of the most comprehensively studied free-living, single celled organisms. *E.coli* is used as the host for many bioprocesses due to it being less complex and easier to grow and maintain in minimal media compared to other cells such as those derived from mammals.

An *E.coli* population can double in less than an hour allowing rapid population growth. As a bacterium, *E.coli* has an additional segment of DNA known as plasmid DNA that complements the main segment of DNA and is found in the cytoplasm. A desired gene can be inserted into the plasmid to enable *E.coli* to produce recombinant proteins.

1.5 Process materials

A variety of different protein solutions were used in this thesis including single protein solutions, hen egg white, a fab lysate and ApoA-IM process samples. The latter two materials were produced in *E.coli*.

1.5.1 Hen egg white

Lysozyme is important in the pharmaceutical and food industry due to its effective anti-microbial properties (Vachier et al. 1995). Ovotransferrin is an iron transport glycoprotein which exhibits anti-microbial activity (Vachier et al. 1995). Ovalbumin is a glycoprotein that is the main cause of egg white gelling and has nutritional value as a protein supplement (Vachier et al. 1995). The abundance of the four main proteins in HEW, along with their pI and MW are shown in Table 1.7.

Table removed
due to copyright

Table 1.7 The four main proteins in hen egg white (Vachier et al. 1995).

Laboratory techniques to purify these proteins include precipitation by salts or solvents, ionic strength reduction or chromatography (Vachier et al. 1995). Chromatography particularly with ion-exchange has been the most promising due to other techniques results in protein denaturation and low purities (Vachier et al. 1995).

1.5.2 Antibody fragment process material

An antibody fragment (fab') was produced in a high cell density batch/fed-batch industrial fermentation of a recombinant *E.coli* strain W3110. The *E.coli* was harvested from the fermentation broth by centrifugation before undergoing an alkali heat lysis step to extract the fab' from the periplasm. The cells were then removed by centrifugation and the remaining periplasmic extract purified using ion-exchange chromatography (Figure 1.16).

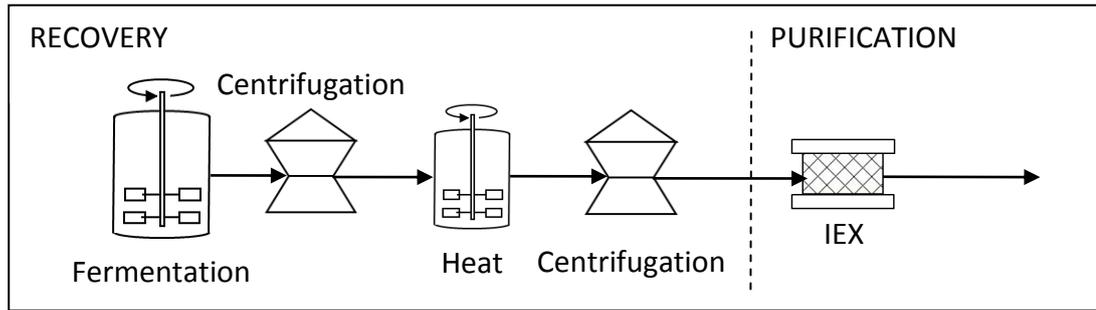


Figure 1.16 Fab' production process.

1.5.3 ApolipoproteinA-IM dimer process material

ApolipoproteinA-IM (ApoA-IM) is a naturally occurring variant of ApoA-I that appears to confer protection against cardiovascular disease to those that carry the mutated gene (Suurkuusk and Hallen,1999). Apolipoprotein A-I (ApoA-I) is a 28 kDa protein present in human plasma in low concentration. The native form of the protein is a 56 kDa covalent dimer. ApoA-I is the major protein component of high density lipoprotein, which serves to maintain cholesterol homoeostasis and eliminate excess cholesterol. The mutation results in an arginine residue at position 173 being substituted with cysteine (Vitello and Scanu,1976). Figure 1.17 & Figure 1.18 show the amino acid sequence of ApoA-IM and the structure of ApoA-I respectively. The cysteine substitution in the mutated form (ApoA-IM) allows a disulphide bond to be formed between two monomers to create an ApoA-IM dimer. This can be prevented during certain stages of production using a reducing agent such as DTT. A “simplified” process diagram for ApoA-IM dimer generation is shown in Figure 1.19.

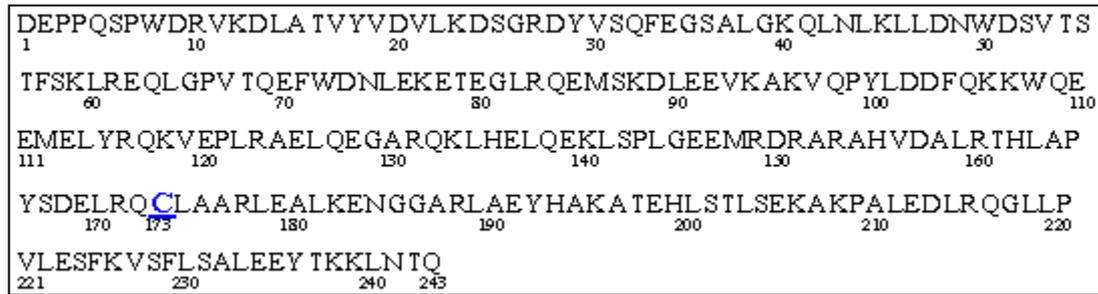


Figure 1.17 ApoA-IM amino acid sequence.

ApoA-IM is the naturally occurring variant of ApoA-I comprising of 243 amino acids with a cysteine substitution for arginine at position 173. The monomer & dimer have respective molecular weights of 28 & 56 kDas. The pI of ApoA-IM is between 5.1 and 5.3.

Image removed
due to copyright

Figure 1.18 The structure of truncated human apolipoproteinA-I (ApoA-I) (Borhani et al. 1997)

ApoA-IM is a major protein component of high density lipoprotein (HDL). The molecule consists of a pseudo-continuous alpha-helix with kinks at regularly spaced proline residues. It forms a horseshoe shape with dimensions (125 x 80 x 40 Å). Four molecules in an asymmetric unit associate via their hydrophobic faces to form an anti-parallel four helix bundle with an elliptical ring shape.

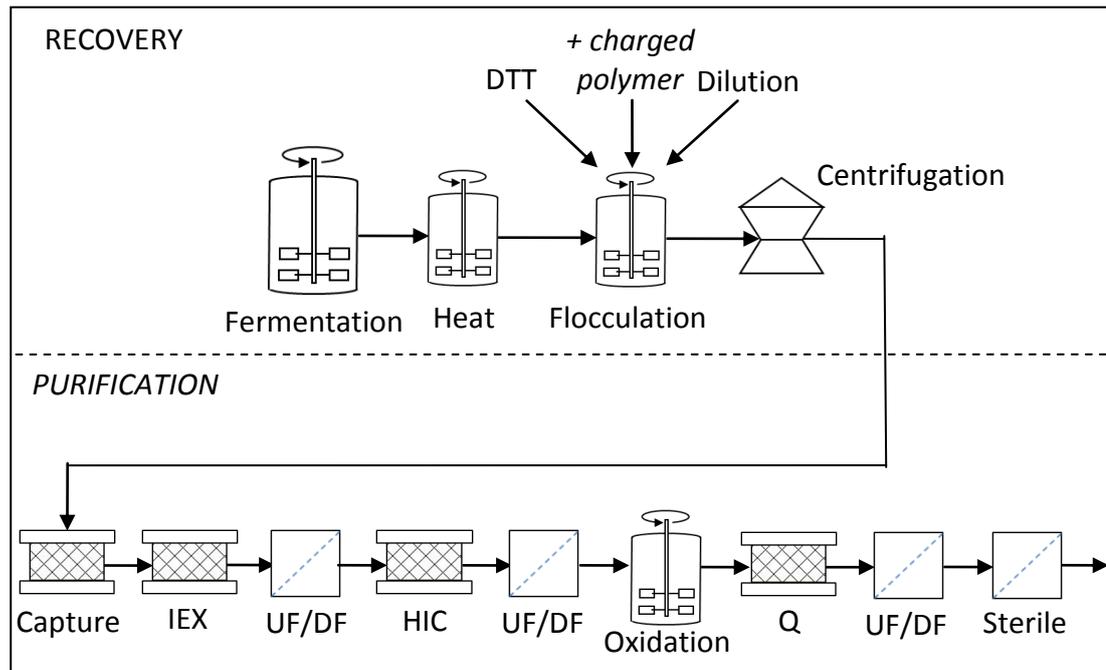


Figure 1.19 ApoA-IM dimer production process.

After the *E.coli* cells are ready for harvesting they are heat lysed to release ApoA-IM into solution. Reducing conditions are then introduced to keep ApoA-IM in its monomer form. The remaining cellular debris is flocculated ready to be removed by centrifugation. Diluting the material at this stage improves the yield of ApoA-IM during centrifugation. After centrifugation the ApoA-IM containing supernatant passes through a series of chromatography columns to reduce impurity levels and any non-reducible forms of ApoA-IM. The ordering of the columns is a common example of IEX being followed with a HIC column due to the IEX elution and HIC loading occurring with salt. After three columns ApoA-IM is oxidised to its dimer form and non-oxidisable ApoA-IM levels lowered with the Q column before final fill/finish steps.

1.6 Project objectives

In the demanding world of biopharmaceutical manufacture it is highly advantageous for a company to be able to rapidly develop well performing and cost effective processes with minimal resource requirements. This thesis investigates the use of ultra-scale down (USD) technologies for this purpose.

1.6.1 Objective 1: Develop the use of a mass spectrometry technology for bioprocess materials

In order to gain a greater insight into processes during and after their development improved analytical techniques are required. Protein mass spectrometry can now be carried out in a rapid and cost-effective manner and has varied application. Chapter two will apply a protein mass spectrometry technique to the bioprocess materials used in this thesis and investigate ways to improve their quality.

1.6.2 Objective 2: Application of the approach to an industrial process in conjunction with a USD device

To validate the use of the mass spectrometric technique for this project it will be used alongside traditional assays to analyse process streams in an existing process. Chapter three attempts to validate the mass spectrometry approach by applying it as an analytical technique for process monitoring of ApoA-IM dimer process development. Comparisons between the different approaches can then be drawn. The mass spectrometric technique will then be used to monitor the flocculation/centrifugation stage of the ApoA-IM process.

In chapter four another ultra-scale down technique for improving the flocculation/centrifugation in terms of cellular debris removal will also be applied.

1.6.3 Objective 3: Develop novel approaches for acquisition and analysis of mass spectra to expedite bioprocess development and monitoring

Further expansion on the mass spectrometry approach is necessary to gain further insights into proteins that exist in process streams. The aim is to develop a platform that can expedite data treatment while adding increased flexibility on data handling to improve bioprocess development.

Chapter five investigates the mass spectrometry protocols that will be used to characterise the relationship between product and contaminant proteins present in the ApoA-IM material after flocculation/centrifugation.

Chapter six describes the generation of the platform to handle the large volume of mass spectrometry data that was produced in chapter five. The Matlab code used is included as an appendix to this thesis.

Chapter seven introduces a novel method for using the mass spectrometry data produced in chapter five to elucidate some of the physicochemical properties of the proteins present in process materials and how the information can be used for process development.

The overall conclusions and future work for this research are summarized in chapter eight.

CHAPTER 2 CONFIGURATION OF A PROTEIN MASS SPECTROMETER FOR BIOPROCESS MATERIALS

2.1 Introduction

This engineering doctorate (EngD) was focussed at using surface enhanced laser desorption ionisation - time of flight - mass spectrometry (SELDI-TOF-MS) as a process monitoring tool and to aid purification decisions. Before this could be done, it was necessary to address the many issues inherent with this chosen experimental approach.

This chapter seeks to investigate how spectra resolution is affected by sample type, sample concentration, and laser strength. This is to give the SELDI-MS user the necessary awareness and knowledge to control these input variables in order to produce high resolution mass spectra peaks. This chapter will also evaluate the use of the mass spectrometry approach for later work in this EngD project as well as highlight some of the issues that will need to be addressed in later chapters.

Summary of chapter aims:

- Provide information about SELDI-MS spectra characteristics that can be used in future chapters.
- Provide guidelines for using the mass spectrometry technique for future chapters.

2.2 Materials & Methods

All chemicals, unless specified otherwise, were obtained from Sigma Chemical Co. Ltd. (St. Louis, MO, U.S.A.) and were of analytical grade.

2.2.1 Protein samples

Various protein samples were used in this EngD project. Single purified proteins were selected to get mass accuracy for the mass spectrometer and for initial case studies. Hen egg white (HEW) was chosen as a relatively simple mixture of a few abundant proteins making it ideal for deciding upon initial data treatment routines. A clarified *E.coli* heat lysate containing fab' produced at UCL and samples from a *E.coli* ApoA-IM dimer process used at Pfizer provided the industrially realistic feed materials to further develop the data treatment routines.

2.2.1.1 Single purified proteins

Purified cytochrome c horse, lysozyme, soy bean trypsin inhibitor, bovine serum albumin, and ovalbumin were used in this thesis. They were each dissolved in 10mM PBS, pH 7.4.

2.2.1.2 Hen egg white

HEW consists mainly of four proteins (Table 1.7) (Vachier et al. 1995). HEW was diluted with 2 volumes of 50mM Tris at pH 8 and stirred for 1 hour at 4°C. Any precipitate was removed by centrifugation at 27,000g for 15 minutes. The remaining supernatant aliquots were frozen at -20°C ready for use.

2.2.1.3 Fab' lysate

The humanised fab' lysate was produced by batch fermentation of *E.coli* W3310 transformed with the plasmid pAGP-4 (Bowering *et al.* 2002). The strain and plasmid were kindly provided by Celltech R&D Ltd., Slough, U.K.. For periplasmic

extraction, fermentation cell broth was centrifuged in a CARR Powerfuge and cell paste heat lysed to release periplasmic proteins into supernatant (Figure 1.16). Cells were removed using a Beckman centrifuge at 166.7rps for 1h 40 min at 4°C.

2.2.1.4 ApoA-IM dimer process samples

Samples were collected during different stages of Pfizer's (Chesterfield, MO, U.S.A.) ApoA-IM dimer process (Figure 3.1). After collection, the samples were immediately frozen at -20°C. When required for use, the samples were defrosted at room temperature and used immediately. All work was carried out at Pfizer (Chesterfield, MO, U.S.A.).

2.2.2 Mass spectrometry

All reagents and equipment specific to SELDI-MS except acetonitrile (ACN) and trifluoroacetic (TFA) acid were supplied by Bio-Rad Laboratories Inc. (Hercules, CA, U.S.A.).

2.2.2.1 Energy absorbing molecule preparation

The energy absorbing molecule (EAM) used was sinapinic acid (SPA). It was prepared by reconstitution with 200µl of 1% TFA in ultrapure water and 200µl of pure acetonitrile into a pre-supplied SPA vial. The vial was vortexed until dissolved for 10 mins and centrifuged at 220rps for 5 mins to pellet any undissolved SPA.

2.2.2.2 Non-Selective ProteinChip preparation

All samples were diluted ten-fold with 10mM PBS, pH 7.4 before deposition of 5µL onto the NP20 (normal phase) ProteinChip surface. After air drying for 20 minutes the spots were washed three times with 5µL of deionised water to eliminate non-adsorbed proteins. After 20 minutes from the last wash being removed, the spots were dry and 1µL of the SPA was applied. After 5 minutes the

EAM application was repeated before a final 5 minute dry step prior to ProteinChip analysis.

2.2.2.3 Selective ProteinChip preparation

The selective ProteinChips used in this chapter (CM10 & Q10) were prepared following identical protocols. Spots were equilibrated three times with 150 μ L of equilibration buffer for 5 minutes under shaking (8.3rps). The spots were then loaded with 150 μ L of the sample previously diluted in equilibration buffer. The sample was incubated with the spot surface for 30 minutes while being shaken (8.3rps). Each spot was then washed three times with 150 μ L of the equilibration buffer for 5 minutes under shaking (8.3rps) to eliminate non-adsorbed proteins. This was followed by a quick rinse of water to remove any salts and then the spots were left to dry at room temperature.

After 20 minutes the spots were dry and 1 μ L of the EAM is applied. After 5 minutes the EAM application was repeated before a final 5 minute dry step prior to ProteinChip analysis.

2.2.2.4 Data acquisition

All arrays were analyzed using a PCS4000 Reader (Figure 1.13) in a positive ion mode, with a source voltage of 25kV. The m/z range investigated was dependent on the sample. Focus mass was set to the m/z of the ApoA-IM monomer (28.1kDa/e). Each spot was divided into four partitions allowing the possibility of four different acquisition protocols to be carried out for each spot. Laser strength responsible for the desorption-ionization of proteins on the spot surface was adjusted based on the mass spectra intensity observed for each sample. Adsorption of sample proteins on arrays appeared as a signal at the appropriate molecular masses.

2.2.2.5 Data treatment

SELDI-MS data in this chapter was treated with Ciphergen Express 3.0 software using settings in Table 2.1.

| Analysis setting tab | Options | Description | Value set |
|---|--------------------------------|--|--------------------------|
| Baseline <i>Calculates electrical and chemical (EAM) noise.</i> | Smooth before fitting baseline | Can improve baseline accuracy by fitting data to moving average before baseline calculation | On |
| | Window | Specifies width of the moving average filter used to smooth baseline | 25 points |
| | Width | Selecting automatic assigns the automatic fitting width | Auto |
| Filtering | Filtering parameters | Removes high frequency noise to improve signal to noise ratio. | On Average |
| | Width | Excessive filter widths can distort peaks and reduce resolution | 0.2 times expected width |
| Noise <i>Calculates noise as variation in signal, not signal itself.</i> | Start measuring noise from: | Sets lower mass limit for noise calculation. Typically matrix m/z excluded. | Minimum m/z |
| | Measure noise to: | Sets upper mass limit for noise calculation. End at end is maximum mass in spectrum. | End at end |
| Spot Correction | On/Off | Corrects for slight systematic shifts in TOF data when spectra are collected from different spots on same array. | Off |

Table 2.1 Settings used when processing data with Ciphergen Express 3.0

2.3 Results

2.3.1 Sample analysis with mass spectrometry

Throughout this EngD different samples have been analyzed using SELDI-MS. One of the goals of this project was to create a framework that could be used for a variety of different samples including those realistic to bioprocess streams. It was therefore important to evaluate how the SELDI mass spectra would look for these different materials (Figure 2.1). From previous knowledge of our samples, small and large m/z regions were defined respectively as any peak less than or equal to 22.5kDa/e and any peak greater than 22.5kDa/e. The laser strength used was

higher for the large m/z region because these proteins required more energy to ionise.

The mass spectra of the samples presented in Figure 2.1 are clearly very different. They are shown in ascending order of expected complexity. Hen egg white (HEW) the least complex, exhibiting relatively few peaks compared to the fab' and ApoA-IM samples.

One of the novelties of using SELDI-MS was the ability to identify an array of proteins in the mixtures tested as opposed to total protein and specific protein assays for the product, which are the common analytical techniques used for microwell and chromatography experiments.

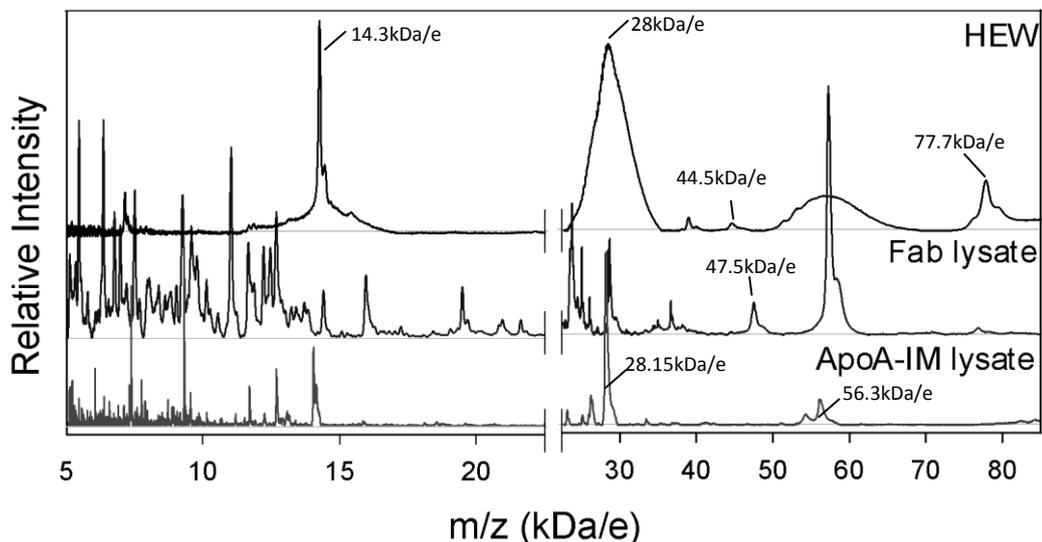


Figure 2.1 Representative mass spectra for adsorbed protein samples.

The samples investigated were hen egg white (HEW), and two clarified E.coli heat lysates; one containing an antibody fragment (fab') produced at UCL and the other containing ApoA-IM produced at Pfizer. The ProteinChips used in this study were normal phase (NP20) arrays which have a silicon oxide surface to bind hydrophilic and charged protein residues (Favre-Kontula et al. 2008).

Lysozyme, ovomucoid, ovalbumin, & conalbumin are four of the most abundant proteins in HEW and they have respective molecular weights (MW) of 14.3, 28, 44.5 and 77.7kDa (Vachier et al. 1995). In Figure 2.1, peaks corresponding to these MWs can be seen in the HEW mass spectra, although ovalbumin was

observed as a relatively small peak. These MWs were confirmed by using purified forms of the three proteins. The broadness of the peaks is due to the highly glycosylated nature of hen proteins compared to the unglycosylated proteins in *E.coli* (Mine,1995).

Figure 2.1 highlights the issues faced with using SELDI-MS to quantify proteins. It is noticeable in the HEW spectra that lysozyme (14.3kDa/e) was the largest peak, followed by conalbumin (77.7kDa/e) and lastly a very small peak for ovalbumin (44.5kDa/e). This is the opposite expected based on the naturally occurring v/v quantities of these proteins in HEW. Typically lysozyme, ovalbumin, & conalbumin constitute 3, 13 & 54% (v/v) of HEW respectively (Vachier et al. 1995).

Fab' has a MW of 47.5kDa and ApoA-IM exists as a monomer at 28.15kDa and a dimer at 56.3kDa, these can also be seen in their respective spectra (Figure 2.1). Some of the other peaks in these three spectra maybe multi-charged versions of the proteins present. CIPHERGEN Express 3.0 software (BioRad, Hercules, U.S.A.) has a peak wizard tool that may reduce these mistakes but since only a few samples will be used in this thesis it should be possible to manually select only true peaks.

2.3.2 Width of mass spectra peaks

As mentioned in the previous section, the peaks in the HEW spectra are significantly broader than in other samples and that is due to them being heavily glycosylated. It is harder to observe with this sample but as the other two samples show there is a drop in resolution at higher masses due to a proportional increase in peak width (Figure 2.1) (Dijkstra *et al.* 2006). This indicates that peak area rather than peak height maybe more accurate at measuring the amount of protein adsorbed.

At the time of this study it was reported by CIPHERGEN and the literature that peak accuracy was typically +/-0.3% of m/z (Brozkova et al. 2008;Panicker et al. 2009). This suggests that peak width typically increased by 0.6% of m/z. This value could be very useful for calculating peak area. ApoA-IM post centrifuge supernatant was chosen to validate this number due to the large number of detectable peaks.

Figure 2.2 shows the effect of m/z on peak width at half peak height using the ApoA-IM post centrifuge supernatant proteins (Figure 3.1) bound to a CM10 ProteinChip using 20mM sodium citrate, pH 3, 0M NaCl. This condition was chosen because it showed high levels of adsorption. Figure 2.2 shows a linear relationship from 10-30kDa/e. From this relationship we were able to derive Equation 2.1.

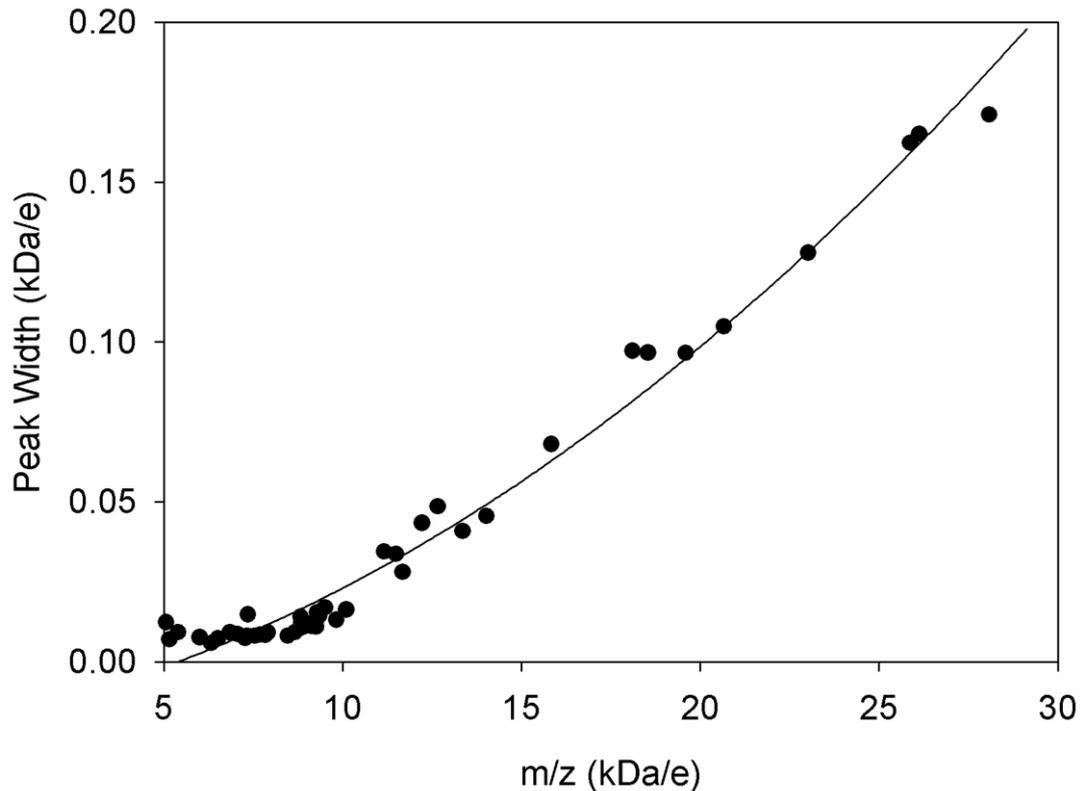


Figure 2.2 Relationship between m/z and peak width.

ApoA-IM post centrifuge supernatant was incubated on the CM10 ProteinChip with 20mM sodium citrate, pH 3, 0M NaCl. The total protein concentration incubated with the surface was 0.05mg/mL. Peaks were detected using the Ciphergen peak detection wizard at a setting of 3 times S/N ratio. Peak widths were recorded as peak width at half peak height. A laser strength of 1750nJ was used.

$$w_h \approx 0.006m/z$$

Equation 2.1 Width of mass spectra peaks.

Where w_h is peak width at half peak height; m is mass; and z is charge

2.3.3 Alignment of mass spectra peaks

The Ciphergen Express software comes with a number of tools to align corresponding peaks that can have percentage deviations of up to 0.3% due to different ProteinChip readers, aging of the readers or slight variations during acquisition (Jeffries,2005). Due to the need to selectively bind different proteins under different conditions it would be impractical to use these tools due to there being no single protein existing in all spectra to align others. It was therefore necessary to measure what effect alignment has on reproducibility. To do this, peaks were automatically selected from ApoA-IM post centrifuge supernatant proteins adsorbed to a Q10 ProteinChip at pH 8.0, 0M NaCl using the Ciphergen peak detection wizard. This adsorption condition was chosen because it bound a large number of proteins over a wide range of m/z values. Peaks were determined as those with a signal to noise (s/n) greater than 3. The coefficient of variation (%CV) was then calculated for the peaks across 3 out of 4 partitions on a spot before and after manual alignment.

Figure 2.3 shows that there was a slight difference before and after alignment which was magnified at the low masses. From this result it was believed that calculating an area under the peak rather than peak intensity may improve reproducibility between aligned and unaligned peaks. The benefit of using peak area to improve reproducibility will be shown later in Figure 6.9 & Figure 6.10.

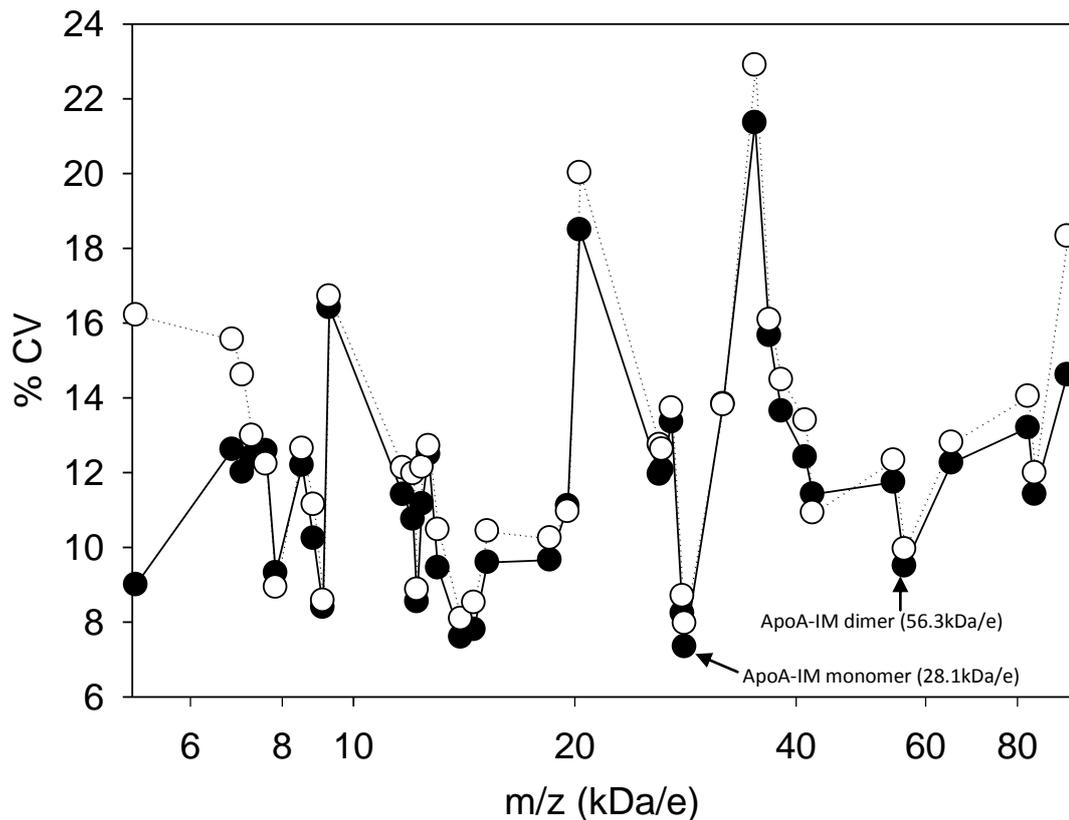


Figure 2.3 The effect of peak alignment on partition reproducibility.

(○) before manual peak alignment of automatically selected peaks, (●) after manual peak alignment of automatically selected peaks. The ApoA-IM post centrifuge supernatant proteins were adsorbed onto a Q10 ProteinChip with 50mM Tris, pH 8, 0M NaCl. Peaks were found using the CIPHERgen peak detection wizard at a setting of 3 times S/N ratio. A laser strength of 1750nJ was used on 3 out of 4 partitions.

2.3.4 The effect of protein concentration and laser strength on mass spectra peak resolution and intensity

From an experimental standpoint it was now of interest to find the sensitivity limits of the mass spectrometer and investigate what effect protein concentration and laser strength had on resolution.

2.3.4.1 Protein concentration

As shown previously with the HEW sample in Figure 2.1 it would be impossible to achieve absolute quantification using SELDI-MS without spiking samples with known standards. This would not be feasible in this thesis due to the complex nature of binding that will be experienced under the varying conditions used and the limited prior knowledge of which proteins are present in the samples examined. Even if it were possible to do this, it would no longer make this method high throughput which is one of the main advantages of using SELDI-MS over other techniques.

It was decided that to understand further the effect protein size has on peak intensity a predefined sample was required. The sample chosen for this was a mixture of five proteins; cytochrome c horse, lysozyme, soy bean trypsin inhibitor, bovine serum albumin, & conalbumin. This mixture was made by creating 1 mg/mL solutions of each protein in 10mM PBS, pH 7.4 and then combining all five in equal volume. This produced a clearly defined stock solution of proteins that covered a wide range of MWs from cytochrome c horse (12.43kDa) to conalbumin (77.7kDa). The stock solution was diluted to give eight different concentrations and these were then analyzed by spotting directly onto an NP20 ProteinChip. It was now possible to quantify the effect MW had on peak intensity.

Figure 2.4 shows that even when proteins are in equal proportions peak intensity is higher at lower masses. At higher masses intensity drops significantly. The biggest shift is observed between the two smallest proteins; cytochrome c (12.4kDa/e) and lysozyme (14.3kDa/e) and the remaining proteins. This suggests that different acquisition settings are required for smaller proteins.

The reason that smaller proteins have higher peak intensities is thought to be due to the relative ease at which smaller proteins ionize to form smaller ion clouds increasing their chance of hitting the ion detector (Dijkstra et al. 2007). This combined with a noticeable drop in resolution again supports the use of peak area rather than peak height for quantifying proteins. This observation will be used later in chapter six.

It can also be seen from Figure 2.4 that all the proteins have sigmoidal profiles with an approximate linear region between 5–100fM. This supports BioRads and others observations that the mass spectrometer is sensitive to femtomole (fM) quantities (Brenac, V et al. 2008;Weinberger et al. 2002a). All of these proteins are detectable above 1fM but to get strong signals the operator should use concentrations greater than 10fM.

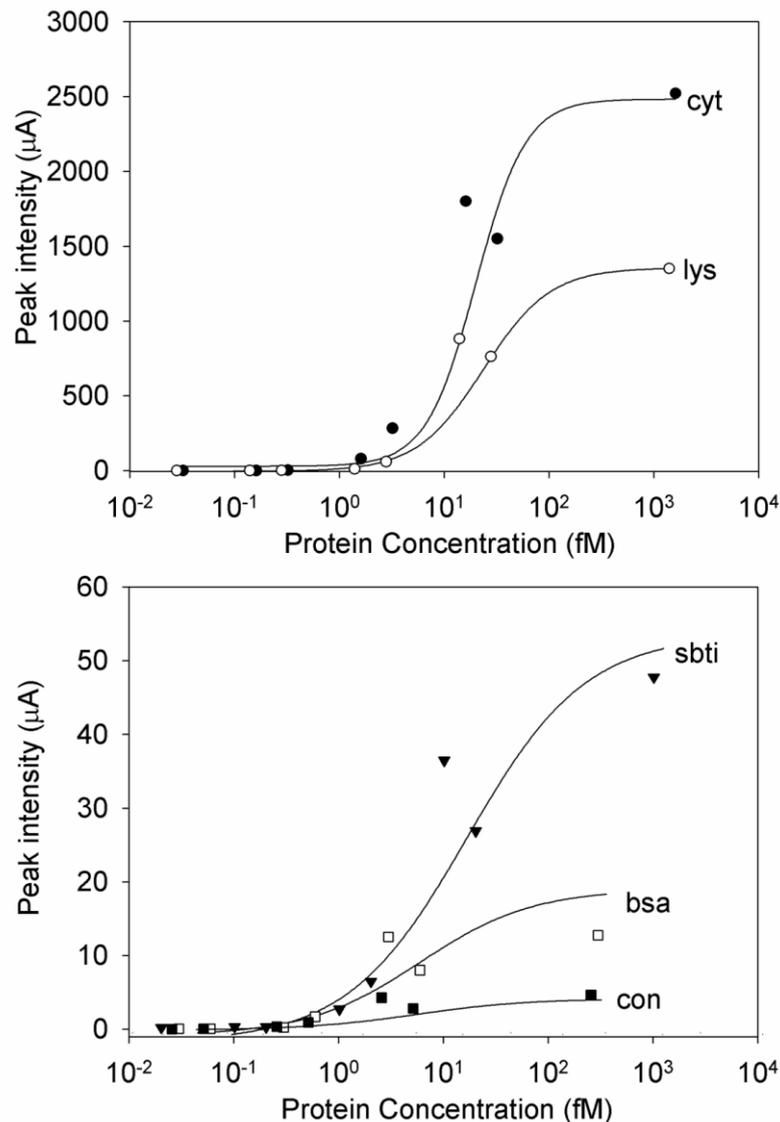


Figure 2.4 The effect of protein concentration on peak intensity.

A mixture of five proteins was used: (●) cytochrome *c* horse (*cyt*), 12.4kDa; (○) lysozyme (*lys*), 14.3kDa; (▼) soy bean trypsin inhibitor (*sbti*), 19.7kDa; (□) bovine serum albumin (*bsa*), 66.8kDa; (■) conalbumin (*con*), 77.7kDa. There is an

approximate linear range between 1-10² fM for all five proteins. Analysis was carried out directly on an NP20 ProteinChip with a laser strength of 2000nJ.

2.3.4.2 Laser strength

It was clear during analysis that the strength of the nitrogen laser used to ionize the proteins from the ProteinChip surface had a clear impact on peak intensity and resolution. Therefore the same five protein mixture was bound to one spot on an NP20 ProteinChip and the four partitions each analyzed with different laser strengths.

Figure 2.5 shows a hyperbolic relationship between laser strength and peak intensity. It would now be necessary to see what combinations of protein concentration and laser strength would give high peak resolutions.

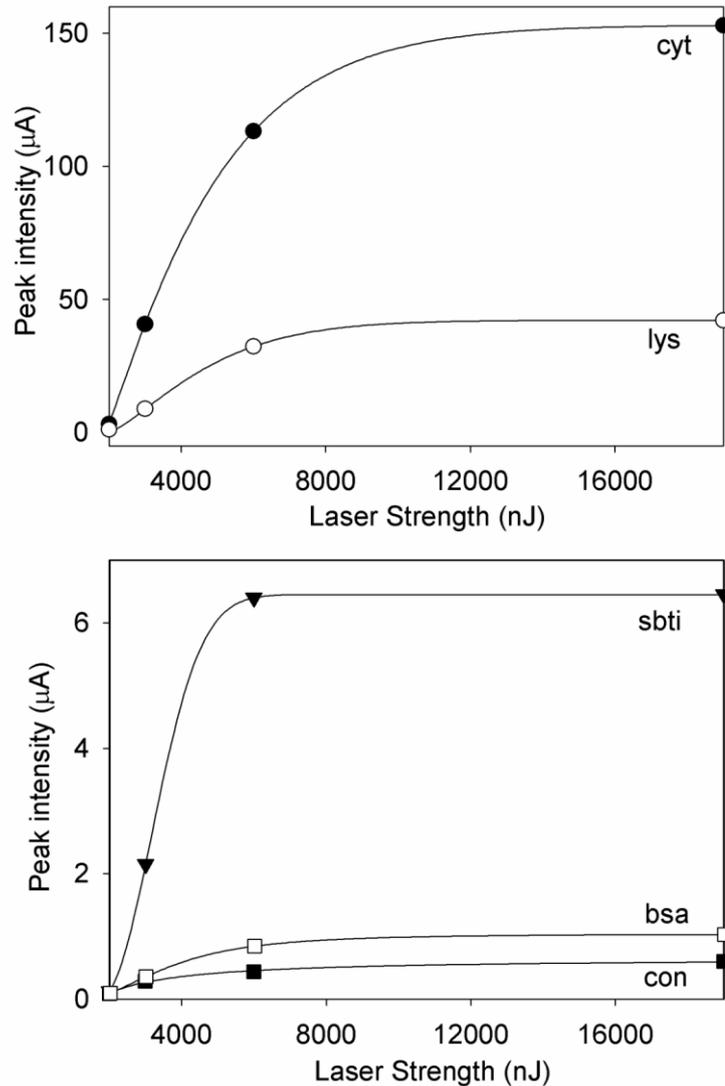


Figure 2.5 The effect of laser strength on peak intensity.

A mixture of five standard proteins was used: (●) cytochrome c horse (cyt), 12.4kDa; (○) lysozyme (lys), 14.3kDa; (▼) soy bean trypsin inhibitor (sbti), 19.7kDa; (□) bovine serum albumin (bsa), 66.8kDa; (■) conalbumin (con), 77.7kDa. The stock solution was diluted to 0.01mg/mL in 10mM PBS, pH 7.4 prior to NP20 ProteinChip application.

2.3.4.3 Optimising protein concentration and laser strength

Some proteins of interest in this thesis are fairly large (ApoA-IM dimer, 56.3kDa/e). These large proteins have been shown previously to have lower peak resolution (Figure 2.2). Effort was therefore made to reduce the impact that protein

concentration and laser strength would have on peak resolution particularly with these large proteins (Figure 2.6).

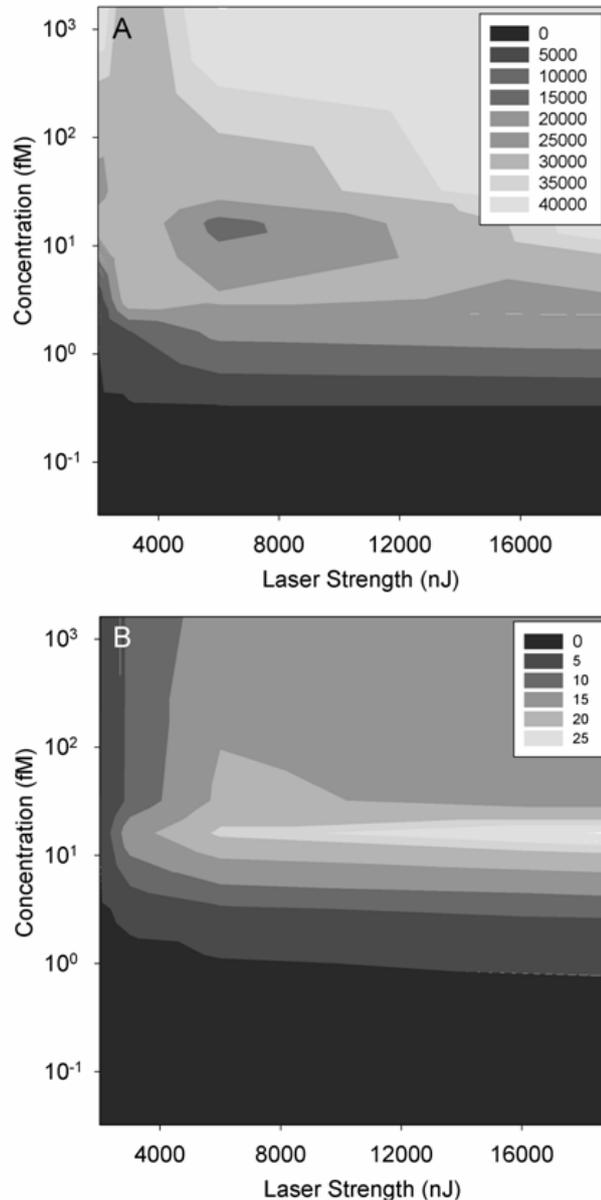


Figure 2.6 The effect of protein concentration and laser strength on resolution.

(A) cytochrome C horse and (B) conalbumin were adsorbed to an NP20 ProteinChip. Peak resolution values were taken manually from CIPHERGEN Express for each condition. Lighter areas indicate improved resolution.

For a small protein such as cytochrome c horse (12.4kDa/e) increasing concentration clearly improved peak resolution although laser strength had less of

an effect. For a large protein such as conalbumin (77.7kDa/e) less protein was required and more laser strength needed to get good peak resolution.

Figure 2.6 shows that for optimal resolution greater than 100fM should be loaded for small proteins and approximately 10fM should be loaded for large proteins. BioRad recommends loading between 0.1 and 0.3mg/mL of total protein to gain good reproducibility. The concentration of ApoA-IM monomer in the post centrifuge supernatant (Figure 3.1, sample E) was 1.6mg/mL. This meant a trade-off in future NP20 ProteinChip experiments as ApoA-IM post centrifuge supernatant would need to be diluted ten-fold to 0.16mg/mL (5.7fM of ApoA-IM monomer) in order to get reasonable resolution while diluting enough to allow adequate buffer exchange without the need for de-salting or dialysis. Based on Figure 2.6 a decision was also made to use a low laser strength (<6000nJ) for small proteins (≤ 22.5 kDa/e) and a high laser strength (>6000nJ) for large proteins (>22.5kDa/e) although this may change depending on the sample used.

2.4 Summary

This chapter has looked at the artefacts common with mass spectrometry particularly those associated with SELDI-MS. This should give the reader some guidelines on which conditions to use to get quality data along with examples of the data that is achievable using this approach.

Review of chapter aims:

- Provide information about SELDI-MS spectra characteristics that can be used in future chapters.
 - Investigated differences between the samples used in this thesis.
 - Peak width typically increases by 0.6% of m/z value.
 - Peak alignment shows some improvement in reproducibility but doesn't seem to be critical to this work.

- Provide guidelines for acquiring mass spectrometry data for future chapters.
 - ApoA-IM post centrifuge supernatant sample should be diluted ten-fold to 0.16mg/mL ApoA-IM monomer before application to an NP20 ProteinChip.
 - If sample concentration is unknown then a range of concentrations should be applied to the ProteinChip. The concentration chosen should be high enough to give a strong signal e.g. greater than 100μamps for masses around 5kDa/e at a laser strength of 1500nJ but low enough that it does not saturate the detector. Peaks that saturate the detector are marked by red ticks on CIPHERGEN Express 3.0.
 - Small and large proteins are defined respectively as any peak less than or equal to 22.5kDa/e and any peak greater than 22.5kDa/e.
 - A low laser strength (<6000nJ) should be used for small proteins and a high laser strength (>6000nJ) for large proteins although this may change depending on the sample used.

We are now able to proceed to the next chapter which aims to validate this technique by using it as a process monitoring tool.

CHAPTER 3 PROTEIN MASS SPECTROMETRY FOR MONITORING THE APOLIPOPROTEINA-IM PROCESS

3.1 Introduction

Currently a great deal of attention is being focused on quality by design (QbD) and process analytical technologies (PAT) in order to guide and accelerate robust bioprocess development for biologics such as therapeutic proteins (Garcia *et al.* 2008). Both are driven by a requirement to readily determine critical product attributes; such methods must not require large sample volumes and complicated sample preparation. This is especially important in early phase development when little is known about the target protein and process impurities, and well-developed quantitative analytical methods such as various forms of high pressure liquid chromatography (HPLC) and immunoassays are still to be established.

Information such as relative quantity & size (molecular weight) of the biological compounds in solution would be useful to gauge the effect of a process step with respect to yield and purity. Gel electrophoresis techniques have this capability but with limited mass range (5-100 kDa), require optimisation and have relatively low resolution and sensitivity (Caputo *et al.* 2003; Jenkins and Pennington, 2001). Surface enhanced laser/desorption ionization time of flight mass spectroscopy (SELDI-MS) could be an alternative method to address these issues (Woolley and Al Rubeai, 2009).

The ProteinChips used in this study are normal phase (NP20) arrays which have a silicon oxide surface for the binding of hydrophilic and charged protein residues (Favre-Kontula *et al.* 2008). Analytical results from more traditional assays such as various forms of HPLC were used as appropriate to compare with the SELDI-MS results.

A “simplified” process diagram for ApoA-IM dimer generation is shown in Figure 3.1 along with locations within the process where samples were taken for analysis. The product exists as a monomer under reducing conditions (samples E – J) and a dimer under oxidising conditions (samples K – N).

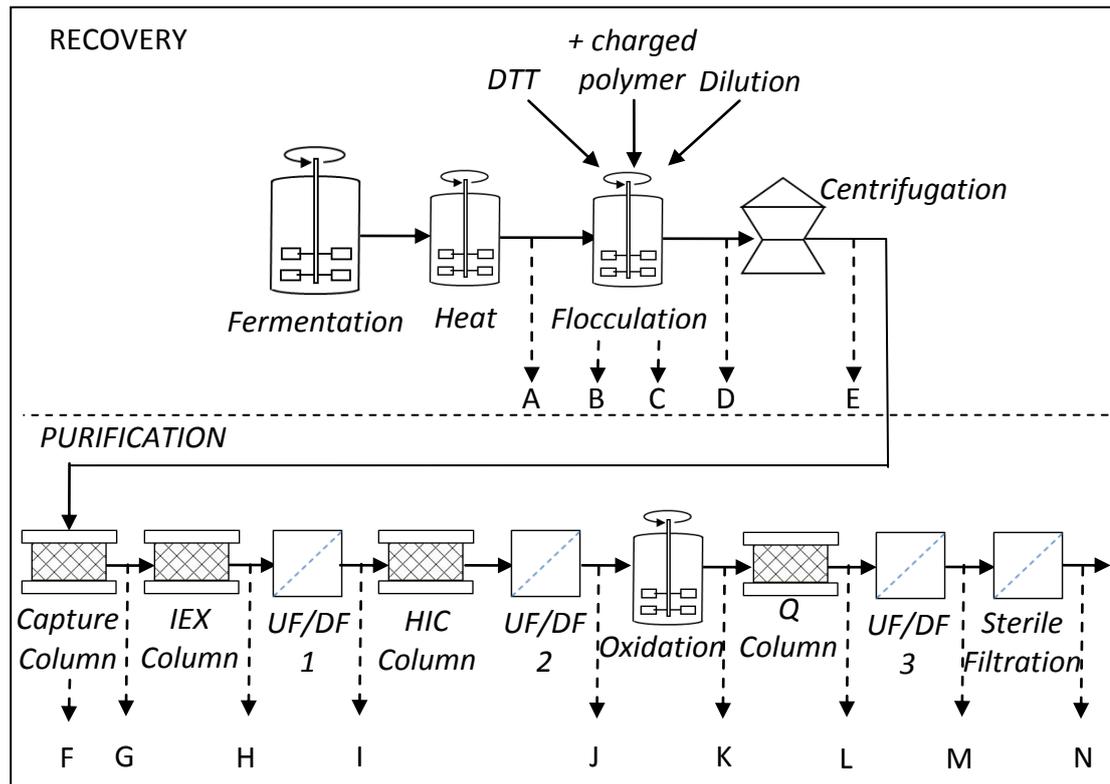


Figure 3.1 ApoA-IM dimer production process samples.

Solid and dashed arrows indicate bioprocess streams and sampling points respectively. Sampling points are denoted by letters for easier reference to analytical data.

Summary of chapter aims:

- Validate that the mass spectrometry technique is comparable to conventional analytical techniques such as HPLC and ELISA by using SELDI-MS to monitor in-process samples during the production of ApoA-IM Dimer Intermediate (ADI).
- Examine the protein profiles during the flocculation process in the hope that conditions can be changed to improve the yield and/or purity of ApoA-IM.

3.2 Materials & Methods

Unless specified otherwise the materials & methods are unchanged from section 2.2.

3.2.1 Protein samples

ApoA-IM process samples were collected during its operation and frozen at -20°C ready for use (Figure 3.1). These samples were diluted tenfold in 10mM PBS, pH 7.4 before their application to the NP20 ProteinChip surface.

3.2.2 Mass spectrometry

3.2.2.1 Data acquisition

The molecular weight range investigated was 1 to 90 kDa. Focus mass was set to 28.1kDa. Laser strength responsible for the desorption/ionization of proteins on the spot surface was set at 1750 nJ for 1-30 kDa range and 3000 nJ for 5-90 kDa range.

3.2.2.2 Data treatment

SELDI-MS data in this chapter was treated with Matlab using settings from (Coombes et al. 2005). This included a wavelet threshold of 10 and to isolate noise, and a threshold of 6 times the mean absolute deviation (MAD) of the wavelet coefficients.

Peaks (Table 3.1) were manually selected based on a signal/noise ratio > 5 and from prior SELDI-MS analysis of the samples. HCPs are representative of a number of unidentified peaks that were present between 5 - 13kDa/e.

| Peak | Mass/Charge (kDa/e) |
|---------------------|--|
| Host Cell Proteins | 6.86, 7.27, 7.85, 8.83, 9.28, 11.16, 12.22, 12.41, 12.66 |
| Possible Truncation | 26.2 |
| Monomer | 28.15 |
| Dimer | 56.3 |
| Possible Aggregate | 84.5 |

Table 3.1 Peaks manually detected in mass spectra.

From the data collected in chapter 2 it was known that peak width was proportional to 0.6% m/z. In order to reduce the bias of smaller proteins having increased peak height and narrower widths (Figure 2.4) summed peak intensities were therefore used instead of peak intensity using a Matlab function, assuming a peak width of 0.3% m/z. Figure 3.2 shows the method used to convert peak intensities into summed peak intensities based on this value.

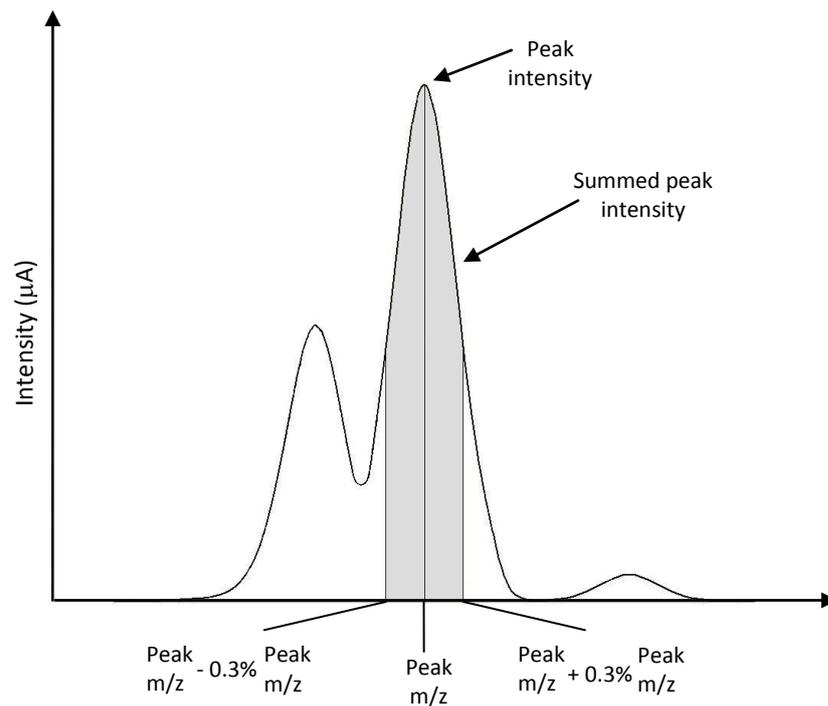


Figure 3.2 Calculation of summed peak intensity.

3.2.3 Reversed Phase HPLC

The Reverse Phase (RP) HPLC (Agilent Technologies Inc., Santa Clara, CA, U.S.A.) method was carried out by the analytical group at Pfizer Global Research & Development (St. Louis, MO, U.S.A). It was run under reducing conditions and utilized for in-process testing to determine the titre of ApoA-IM monomer species and several product-related impurities such as truncated species. An Agilent 1100 series chromatography system (Agilent Technologies Inc., Santa Clara, CA, U.S.A.) was used comprising solvent pump, degasser, injector, and fixed wavelength UV detector. The eluate was monitored for 280nm absorbance and peaks were integrated using the Agilent HPLC software. Component concentrations were based on the total integrated area of respective peaks.

3.2.4 Size Exclusion HPLC

The SDS Size Exclusion (SE) HPLC (Agilent Technologies Inc., Santa Clara, CA, U.S.A.) method was carried out by the analytical group at Pfizer Global Research & Development (St. Louis, MO, U.S.A). It was used to measure ApoA-IM monomer, dimer, high molecular weight, and monomer-related species in process samples. SDS was present in the mobile phase to dissociate non-covalent aggregates.

3.2.5 Host cell protein analysis

The Enzyme-linked immunosorbent assay (ELISA) method was carried out by the analytical group at Pfizer Global Research & Development (St. Louis, MO, U.S.A). It was used to determine the amount of host cell protein impurities in ADI and in-process samples. Samples to be assayed for host cell protein (HCP) impurities are applied to a micro-titre plate which has been coated with capture antibodies and blocked to prevent non-specific binding. The plate is then probed with secondary antibody (conjugated with biotin) and then with streptavidin horseradish peroxidase conjugate. The presence of HCP is monitored indirectly by the reaction of peroxidase with tetramethylbenzidine (TMB). The intensity of colour, measured

spectrophotometrically, is proportional to the amount of HCP present. Standards are analyzed in each assay and sample results calculated by interpolation from the standard curve. Assay range is 1 – 32 ng HCP/mg ApoA-IM.

3.3 Results

3.3.1 Initial capture chromatography

Spectra in Figure 3.3 show that the capture column performs as expected in retaining the ApoA-IM in the post centrifuge supernatant (sample E) and removes some lower molecular weight host cell proteins, which can be seen in the flow through (sample F). The capture pool (sample G) contains ApoA-IM mostly in the monomer form with a significant amount of host cell protein contaminants left. In Table 3.2, the RP-HPLC method shows ApoA-IM monomer at 1.6 mg/mL in the post centrifuge supernatant prior to column loading (sample E), increasing to 2.1 mg/mL in the capture pool (sample G). This corresponds to 7.6 and 18.3mA for the monomer peaks in the respective SELDI-MS spectra. By initial observation it seems that there is a discrepancy between the RP-HPLC and SELDI-MS results for these two samples but by comparing these two methods under reducing conditions (pre-oxidation samples, E-J) it seems that there is a linear relationship between the two techniques with an R^2 value of 0.93 (Figure 3.4). A linear relationship is also observed when measuring truncated species (Table 3.2) suggesting that both techniques could be measuring the same truncation at 26.2 kDa/e. The reason for SELDI-MS and RP-HPLC results not being comparable post-oxidation (samples K-N) is because SELDI-MS involves minimal sample pre-treatment whereas the RP-HPLC assay reduced all of the samples prior to analysis.

Due to the complexity of the capture pool solution, it was not possible to make SE-HPLC measurements to determine the relative amounts of monomer and dimer in that sample. SELDI-MS results shown in Table 3.3, however, show that there is significantly higher level of monomer relative to dimer for the capture pool (sample G), which is consistent with the reducing conditions present at this stage of the process. HCP levels measured with the ELISA method (Table 3.4) show HCP

reduction from over 10,000 ng/mg in the post centrifuge supernatant (sample E) to 1,500 ng/mg for the capture pool (sample G). A reduction in HCP levels is also shown by SELDI-MS but due to HCPs being recorded as a summation of different low molecular weight peak intensities (Table 3.1) that ionize by varying amounts the comparison appears to be more relative rather than absolute quantitation (Table 3.4).

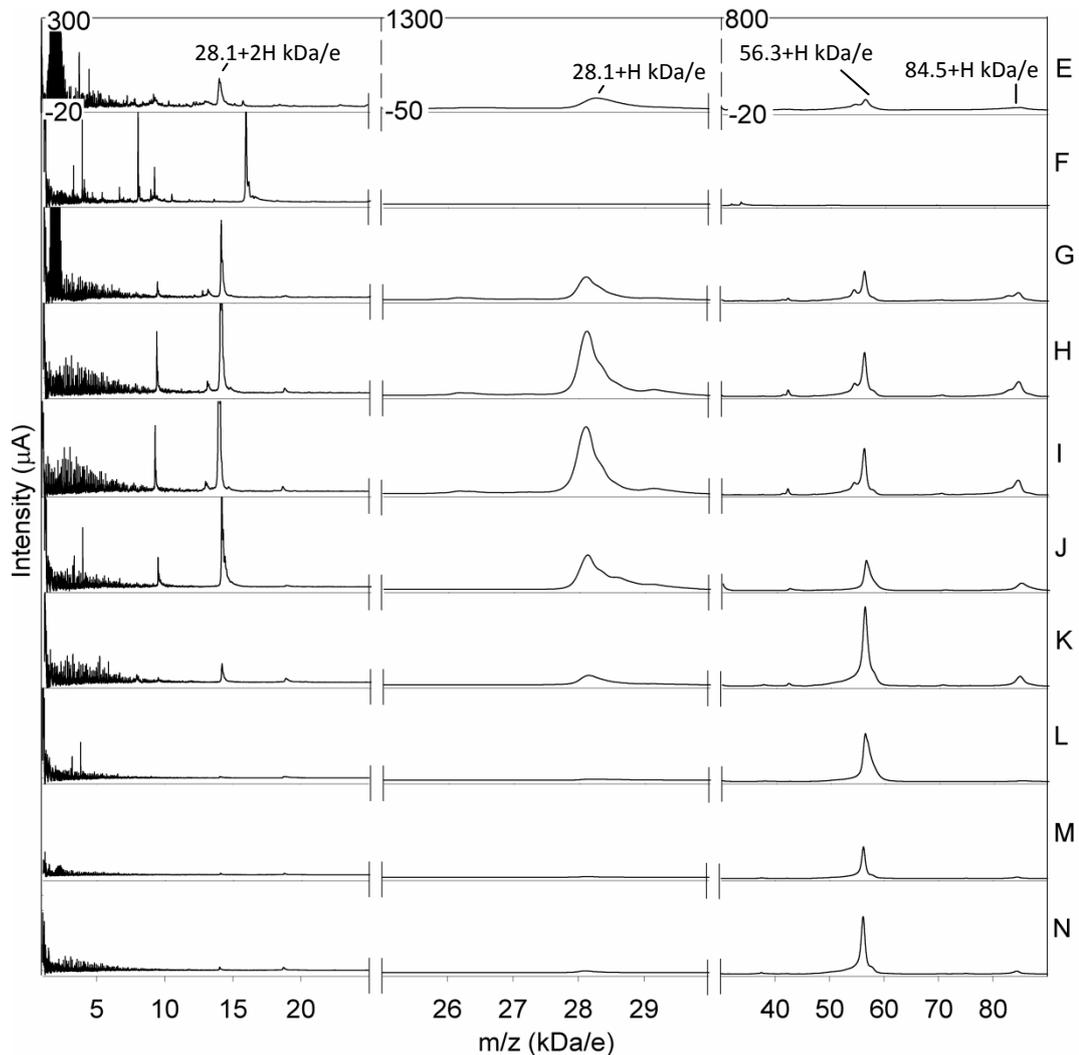


Figure 3.3 Mass spectra for ApoA-IM dimer production process samples.

(E) post centrifuge supernatant; (F) Capture flow-through; (G) Capture pool; (H) IEX pool; (I) UF/DF 1; (J) UF/DF 2; (K) Oxidation pool; (L) Q pool; (M) UF/DF 3; (N) ADI. The low mass/charge (1-30 kDa/e) and high mass/charge (5-90 kDa/e) regions were acquired at 1750 and 3000nJ respectively. The y-axis scale used for each mass to charge range is shown in the top spectra. The same scale is used for all samples.

| PROCESS STEP | | RP-HPLC | | SELDI-MS | |
|--------------|--------------------------------|---------------------------------|-----------------------------|--|-----------------------------|
| | | Reduced r-ApoA-IM (mg/mL) | Truncated Species (%) | Monomer Summed Peak Intensity (mA) | Truncated Species (%) |
| E | POST CENTRIFUGE SUPERNATANT | 1.6 | 14.7 | 7.6 | 7.2 |
| G | CAPTURE POOL | 2.1 | 14.3 | 18.3 | 7.7 |
| H | IEX POOL | 8.8 | 7.8 | 52.1 | 3.9 |
| I | UF/DF 1 | 9.5 | 7.1 | 53.1 | 3.5 |
| J | UF/DF 2 | 6.2 | 0 | 27.4 | 0.1 |
| K | OXIDATION POOL | 1.3 | n.d. | 7.8 | 0.2 |
| L | Q POOL | 5.7 | 0.2 | 1.0 | 0.5 |
| M | UF/DF 3 | 15.1 | 0.1 | 0.6 | 0.1 |
| N | ADI | 19.0 | n.d. | 1.4 | 7.2 |

Table 3.2 Quantification of monomer and truncated species using mass spectrometry and RP-HPLC.

Samples E-J are from the reduction stage of the process, whereas K-N are from the oxidation stage. All HPLC samples are analysed under reducing conditions. All SELDI-MS samples are analysed under similar conditions to that existing in the process. The SELDI-MS low mass (1-30 kDa) and high mass (5-90 kDa) regions were acquired at 1750 and 3000nJ respectively. Samples not analysed, are indicated as not determined (n.d.).

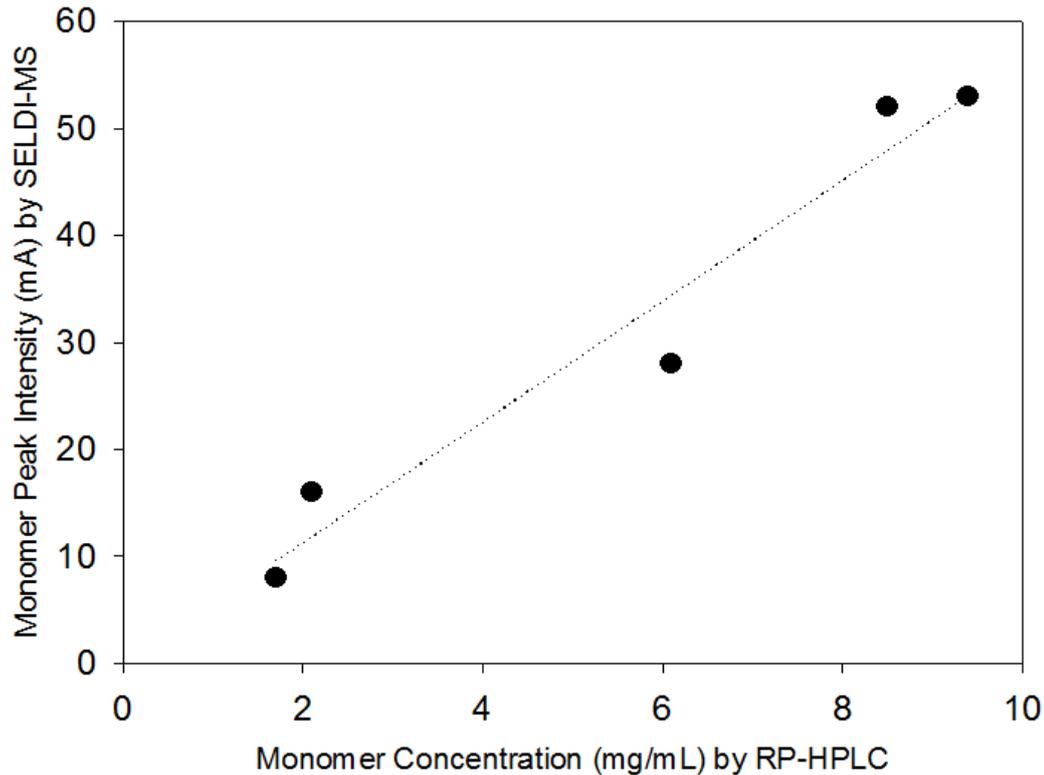


Figure 3.4 ApoA-IM monomer quantification using mass spectrometry and RP-HPLC.

Comparison between quantification using RP-HPLC and SELDI-MS for the ApoA-IM protein in samples prior to oxidation (samples E-J). The line of best fit shown has an R^2 value of 0.93

Due to the complexity of the capture pool solution, it was not possible to make SE-HPLC measurements to determine the relative amounts of monomer and dimer in that sample. SELDI-MS results shown in Table 3.3, however, show that there is significantly higher level of monomer relative to dimer for the capture pool (sample G), which is consistent with the reducing conditions present at this stage of the process. HCP levels measured with the ELISA method (Table 3.4) show HCP reduction from over 10,000 ng/mg in the post centrifuge supernatant (sample E) to 1,500 ng/mg for the capture pool (sample G). A reduction in HCP levels is also shown by SELDI-MS but due to HCPs being recorded as a summation of different low molecular weight peak intensities (Table 3.1) that ionize by varying amounts

the comparison appears to be more relative rather than absolutely quantitative (Table 3.4).

| PROCESS STEP | | SE-HPLC | | | SELDI-MS | | |
|--------------|-----------------------------|-------------|-----------|----------------|-------------|-----------|----------------|
| | | Monomer (%) | Dimer (%) | Aggregates (%) | Monomer (%) | Dimer (%) | Aggregates (%) |
| E | POST CENTRIFUGE SUPERNATANT | n.d. | n.d. | n.d. | 60.7 | 10.3 | 1.4 |
| G | CAPTURE POOL | n.d. | n.d. | n.d. | 50.9 | 32.7 | 12.1 |
| H | IEX POOL | n.d. | n.d. | n.d. | 66.1 | 22.7 | 9.3 |
| I | UF/DF 1 | n.d. | n.d. | n.d. | 63.7 | 23.9 | 10.6 |
| J | UF/DF 2 | >95 | <5 | 0 | 64.3 | 25.9 | 8.0 |
| K | OX POOL | 25 | 72 | 2.2 | 16.4 | 70.5 | 10.3 |
| L | Q POOL | <1 | 99 | 0 | 4.6 | 91.8 | 1.3 |
| M | UF/DF 3 | n.d. | n.d. | n.d. | 3.3 | 90.1 | 5.5 |
| N | ADI | 0.3 | 99 | 0.4 | 3.9 | 89.9 | 5.0 |

Table 3.3 Quantification of monomer, dimer and aggregates using mass spectrometry and SE-HPLC.

Samples E-J are from the reduction stage of the process, whereas K-N are from the oxidation stage. All samples are analysed under similar conditions to that existing in the process. The SELDI-MS low mass/charge (1-30 kDa/e) and high mass (5-90 kDa/e) regions were acquired at 1750 and 3000nJ respectively. Samples not analysed, are indicated as not determined (n.d.).

| PROCESS STEP | | ELISA | SELDI-MS |
|--------------|--------------------------------|------------------------|------------|
| | | HCP (ng/mg ApoA-IM) | HCP (%) |
| E | POST CENTRIFUGE SUPERNATANT | >10,000 | 27.7 |
| G | CAPTURE POOL | 1500 | 4.3 |
| H | IEX POOL | 100 to 200 | 1.9 |
| I | UF/DF 1 | n.d. | 1.8 |
| J | UF/DF 2 | ~10 | 1.8 |
| K | OX POOL | n.d. | 2.9 |
| L | Q POOL | 1 to 2 | 2.3 |
| M | UF/DF 3 | n.d. | 1.1 |
| N | ADI | <2 | 1.2 |

Table 3.4 Quantification of host cell proteins using mass spectrometry and ELISA.

ELISA measurement is based on all HCPs present whereas SELDI-MS based on a predetermined sample of HCPs in the range of 5 to 13 kDa/e. Peaks in the range of 13-15 kDa/e were ignored as possible multi-charged forms of monomer and its truncation. Samples E-J are from the reduction stage of the process, whereas K-N are from the oxidation stage. All samples are analysed under similar conditions to that existing in the process. Samples not analysed, are indicated as not determined (n.d.).

3.3.2 Ion-exchange chromatography

The RP-HPLC method shows that the concentration of monomer quadruples across the ion-exchange chromatography column (samples G & H), which is similarly shown with the SELDI-MS comparison (Table 3.2). There is also a concomitant two-fold reduction of truncated species and a reduction in the dimer and aggregate relative to the monomer that can be observed with both techniques (Table 3.2 & Table 3.3). The ion-exchange pool spectrum (sample H) in Figure 3.3 shows clearly that this column removed most of the host cell proteins (the group of peaks less than 2 kDa). Separate host cell protein analysis (Table 3.4) shows that HCP levels in the UF/DF 1 sample (sample I) are 100-200 ng/mg, dropping from over 1,500 ng/mg in the capture pool sample. HCPs are also seen to drop somewhat in the SELDI-MS results (Table 3.4) but do not reflect the one order of magnitude reduction found with the standard ELISA method. This represents a major

weakness of the SELDI-MS method used in this study for analyzing highly purified samples where accurate detection of trace impurities becomes critical in guiding process development effort. To overcome this, it may be necessary in future work to use ProteinChip conditions that select for these HCPs over the more abundant target protein in order to increase HCP detection.

3.3.3 Oxidation

The oxidation pool spectrum (sample K) shows that the size of the monomer peak drops significantly with dimer formation, as expected (Figure 3.3, sample K). For the oxidation pool and the samples that follow (samples K-N), it is no longer reasonable to compare the RP-HPLC with the SELDI-MS results due to the now oxidised proteins being reduced prior to RP-HPLC analysis, the expected mismatch for these samples can be observed (Table 3.2). This highlights a main advantage of SELDI-MS in that it requires limited sample pre-treatment and hence analyses material in the conformation present during process operation.

The comparison for monomer and dimer post oxidation between the SELDI-MS and SE-HPLC results is similar although not exact (Table 3.3). The reason for this could be due to an underestimation of the amount of dimer by SELDI-MS. The dimer being twice as large has a much larger time of flight, allowing its ion cloud to have more time to expand from same charge repulsions and hence more of the protein missing the detector (Dijkstra et al. 2007). This results in broader, shorter peaks. For more quantitative assessment with the SELDI-MS method, monomer and dimer should therefore be calibrated separately.

3.3.4 Anion-exchange chromatography & final step

Figure 3.3 shows the spectra for samples prior to and post the Q column (samples K & L respectively) along with the final purified ApoA-IM dimer (ADI) sample (sample N). Qualitatively, the dimer content increases substantially across the Q column, consistent with HPLC analysis shown in Table 3.3. However, only 91.8% is dimer in the Q pool and 89.9% in the ADI sample based on SELDI-MS

intensity values (Table 3.3), lower than the 99% dimer measured by the HPLC method.

3.3.5 Host cell protein removal

It is of interest to investigate the ability of SELDI-MS analysis to monitor HCP reduction across each step in the manufacturing process. The biggest change in impurities levels during the process was between 5 and 13 kDa/e and so it was assumed that this was the location for many of the HCPs. There was also typically more noise within this range perhaps due to multi-charged ionisation or proteolysis of large molecular weight proteins. Efforts were made to limit potential proteolysis by using samples directly after defrosting. Where possible, information from literature and knowledge obtained from earlier research in Chapter 2 was used to improve spectra quality including the reduction in multi-charged ionisations (Cordingley et al. 2003).

By assigning the peaks in the range of 5 to 13 kDa as host cell proteins (Table 3.1, peaks between 13 and 15 kDa could be multi-charged forms of monomer and its truncated species and so were not included), the total intensities of those peaks at different points in the process could reflect the relative amounts of HCP (Table 3.4). As expected, there is a general trend downward as the protein solution becomes purer. Table 3.4 shows the actual HCP values measured for some of the samples by the ELISA method as measured relative to the product. Qualitatively, there is order-of-magnitude agreement between HCPs measured with ELISA and SELDI-MS, especially in the early stages of the process. As the solution becomes cleaner, SELDI-MS analysis appears to overestimate the level of HCPs left. This could be due to smaller proteins including the HCPs shown in this work having much smaller times of flight relative to larger molecules such as the monomer. This results in the HCPs having less time for their ion clouds to expand, increasing the likelihood of their ions colliding with the ion detector (Dijkstra et al. 2007). This would lead to a higher than expected HCP signals for the SELDI-MS data. This would be further exaggerated as impurities including larger proteins (>13kDa/e) are

removed. Further development of the SELDI-MS method may give scope to circumvent such shortcomings.

There was a desire to identify HCPs found within this dataset with existing *E.coli* protein databases. It is known that many proteins share similar native masses so although matches were found based on native mass, the level of certainty was limited. To increase confidence each of these HCPs would have needed to be isolated and proteolysed to produce the unique peptide maps that could then be compared with relevant databases. Unfortunately, with other work that needed to be completed, this was too large a task to be carried out within the timeframe of this study and hence this has been left to future work (section 8.3).

3.3.6 Mass spectrometry analysis during flocculation process

Figure 3.5 shows a series of SELDI-MS spectra obtained for samples during an ultra scale down mimic of the flocculation process (Berrill *et al.* 2008); the top spectrum is for the heat extracted lysate (sample A), followed by the addition of DTT to the lysate to reduce covalently linked ApoA-IM molecules (sample B), then with addition of the flocculating reagent to precipitate impurities and cell debris (sample C), and finally with water dilution prior to centrifugation (sample D). These samples represent the sequential treatment of the cell broth for extracting ApoA-IM from the cell periplasm then clarifying the resultant cell lysate. On the vertical axis is the relative intensity of each peak, an indication of the amount of that particular species in the solution. The horizontal axis shows the mass over number of charges for each peak; the value shown in the plot is equivalent to kDa over number of electronic charges. For most cases, e is equal to 1 however peaks at ~ 14.1 kDa in these spectra are likely monomer with 2 electronic charges. The peak at about 28.2 kDa is the ApoA-IM monomer and the small peak at 56.3 kDa its dimer form. Interestingly, the addition of DTT to the heat extracted cell lysate (sample B) appears to release into solution ApoA-IM monomer, which is likely associated with the cell membrane. The release of ApoA-IM into solution seems to further increase with the addition of the flocculating reagent (sample C).

Also shown in Figure 3.5 is the large number of protein contaminants in the molecular weight range from 1 to 10 kDa, some of them are seen to be reduced post flocculation & centrifugation (Figure 3.3, sample E). This contaminant profile information obtained with these crude samples demonstrates a particular power of the SELDI-MS method. For these crude samples SE-HPLC does not work effectively and the more quantitative RP-HPLC analysis is limited to the quantification of ApoA-IM amount and relative amounts of some product related impurities such as truncated species (Table 3.2).

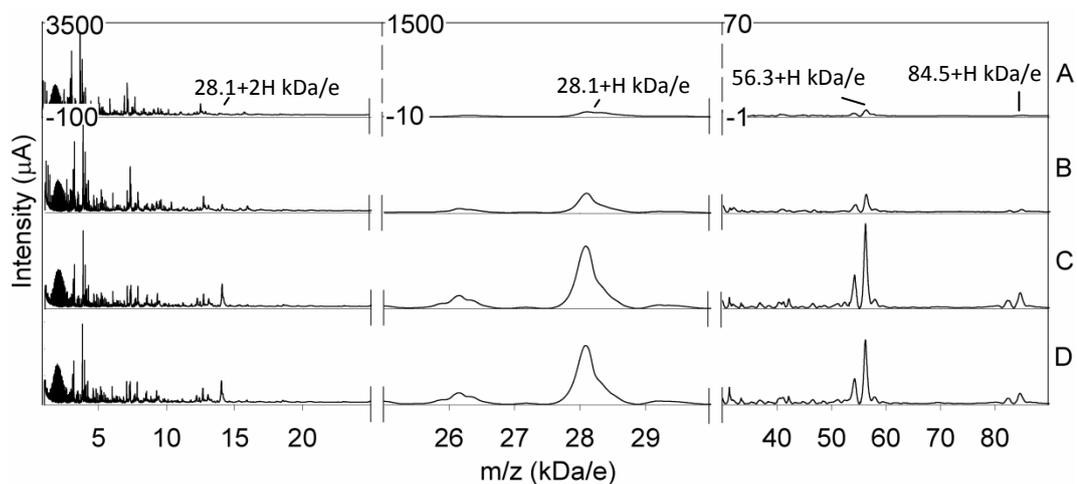


Figure 3.5 Mass spectra of USD flocculation samples.

SELDI-MS analyses of heat extract samples before and during an ultra scale down mimic of the flocculation process. Samples were centrifuged at 216.7 rps for 5 minutes prior to analysis. (A) heat extract; (B) sample A with DTT addition; (C) sample B with flocculating agent addition (D) sample C with water addition, also known as centrifuge feed supernatant. The low mass/charge (1-30 kDa/e) and high mass/charge (5-90 kDa/e) regions were acquired at 1750 and 3000nJ respectively. The y-axis scale used for each mass to charge range is shown in the top spectra. The same scale is used for all samples.

| PROCESS STEP | | RP-HPLC | |
|--------------|--------------------|------------------------------|--------------------------|
| | | Reduced r-ApoA-IM (mg/mL) | Truncated Species (%) |
| A | HEAT EXTRACT | 1.6 | 15.6 |
| D | CENTRIFUGE FEED | 1.7 | 15.5 |

Table 3.5 RP-HPLC analysis before and after the pilot scale flocculation.

It was observed that flocculation conditions had less of an impact on protein profile than on cellular debris removal. For this reason the next chapter will focus on using an ultra scale-down model to improve cellular debris clearance rather than protein contaminant removal.

3.4 Summary

Review of chapter aims:

- Validate that mass spectrometry technique is comparable to conventional analytical techniques such as HPLC and ELISA by using SELDI-MS to monitor in-process samples in the production of ApoA-IM Dimer Intermediate (ADI).
 - The SELDI-MS method offers an effective analytical tool for monitoring in-process samples during early-stage process development when little is known about the protein and impurities, and well developed quantitative analytical methods are not yet in place.
 - The normal phase ProteinChip (NP20) provides a broad spectrum adsorption surface that enables detection of product and contaminant proteins not readily measured with any other single analytical technique. For these crude samples SE-HPLC does not work effectively and the more quantitative RP-HPLC analysis is limited to the quantification of ApoA-IM and relative amounts of some product related impurities such as truncated species.
 - Intensity values of the observed peaks along with their molecular weights provide rapid assessment with respect to the relative abundance and

functional “identity” of those peaks, i.e., ApoA-IM monomer, dimer, and host cell proteins.

- Near the end of the process, when the protein solution becomes very pure, the SELDI-MS method as implemented has reduced value since the obtained spectra are dominated by the target protein whereas interest at that stage focuses more on minor impurities such as host cell proteins and product-related species. A possible solution to this problem could be the use of surfaces with increased selectivity for contaminant capture while reduced binding of the product protein.
- Examine the protein profiles during the flocculation process in the hope that conditions can be changed to improve the yield and/or purity of ApoA-IM.
 - The addition of DTT to ApoA-IM heat extract appears to release into solution substantial amount ApoA-IM monomer, which is likely associated with the cell membrane.
 - The release of ApoA-IM into solution seems to further increase with the addition of the flocculating agent.
 - Some protein contaminants in the molecular weight range from 1 to 10kDa, seem to be reduced post flocculation.
 - There is a much greater change in cellular debris removal than protein profiles at this stage so the decision was made to concentrate on improving cellular debris removal.

The next chapter will now address cellular debris removal during the flocculation/centrifugation stage of the ApoA-IM process using another USD device. ApoA-IM post centrifuge supernatant material produced from this stage will be used for purification development in chapters five through to seven.

CHAPTER 4 DEFINING & IMPROVING THE FLOCCULATION & CENTRIFUGATION STAGE OF THE APOLIPOPROTEINA-IM PROCESS USING A USD DEVICE

4.1 Introduction

Due to increased titres and fermentation volumes, the size of chromatography columns and the volumes of buffers to be used with them are being pushed ever higher (Thommes and Etzel,2007). For this reason there is a desire to find methods to lower the burden on the chromatographic steps in a process.

Flocculation combined with a preceding centrifugation step is one method to reduce the contaminants that can have disastrous effects on subsequent steps that usually include chromatography. The flocculating agent causes these contaminants, usually cellular debris to form larger aggregates which are then easier to remove by centrifugation.

The required floc characteristics depend on the subsequent separation operation. In a disc stack centrifuge there are typically relatively high levels of shear in the feed zone so there is a demand for floc particles to be strong and dense enough to prevent breakup. The more resistant the particles are to shear the higher the speed the centrifuge can be operated at. This in turn allows higher flow rates to be used through the centrifuge without sacrificing clarification or yield capabilities.

If the floc particles are shear sensitive and the targets for clarification are not achieved then it is necessary at laboratory scale to investigate what factors during the flocculation process enable the production of stronger flocs.

The USD method to measure the shear sensitivity of floc particles expands on previous work carried out at University College London. The concept is to mimic the pilot scale centrifuge in the laboratory, by finding the shear that is similar to that experienced in the feed zone of a pilot centrifuge. This is carried out by exposing the feed material to different levels of shear in the laboratory and then measuring the ease of clarification in a laboratory scale centrifuge compared to pilot scale. The theory and method of mimicking entry shear has been thoroughly

described in the literature (Boychyn et al. 2001;Boychyn et al. 2004;Boychyn et al. 2000).

Summary of chapter aims:

- Demonstrate how a mimic for centrifugal clarification can be applied to a flocculated *E.coli* extract and give insight into the interaction between flocculation and centrifugation.

4.2 Materials & Methods

Unless specified otherwise the materials & methods are unchanged from section 2.2.

The process material used in this study was a flocculated ApoA-IM heat extract (Figure 3.1, sample A) provided by Pfizer (Chesterfield, MO, U.S.A.). Material was collected and used directly from the process within an hour of it being made. The initial solid concentration of the feed stream was determined by centrifuging the feed in 50 mL graduated centrifuge tubes in an Sorvall RC6 centrifuge (Thermo Electron Corp., Asheville, NC, U.S.A.), rotor SLA-600TC (N=166.7rps, t=0.5 h). The feed was found to vary slightly between flocculation batches but averaged at ~20% (w/w) solids, packed volume basis.

Multiple centrifugation runs were performed and each time duplicate measurements were taken with standard deviation used to estimate error. This error never exceeded +/- 5% in the laboratory or pilot centrifuge. Each run showed similar trends with figures showing data from a representative run.

4.2.1 Bench-top studies

4.2.1.1 Flocculation

E.coli cells were heat lysed before exposure to a 1% w/v solution of flocculating agent in ultrapure water. The flocculant used is proprietary and cannot be identified other than it being a cationic polymer but the same flocculant grade

and quality was used for both bench-top and pilot studies. A R1312 Turbine stirrer (IKA Works Inc., Wilmington, NC, U.S.A.) was used to mix 400 mL of *E.coli* extract at 6.66 s^{-1} . Flocculating agent was added at 3 g/kg, 5 g/kg, and 7 g/kg relative to *E.coli* extract. The material was mixed for 15 minutes before water was added prior to centrifugation. Water was added to make final weight ratio of 1.0 w/w (total additions including flocculating agent solution/heat extract) to improve product yield and centrifuge performance.

4.2.1.2 Centrifugation

E.coli flocs were diluted ten-fold to ~2.0% (w/w, packed volume basis) in 50mM PBS, pH 7.4, to make a 1 L suspension. This was used to produce well spun supernatant, low shear (100 rps) exposed material, and high shear (300 rps) exposed material. Shear exposure was carried out for 20 s in a high-speed rotating-disc device. The construction of this device has been described previously (Hutchinson et al. 2006). These two shear conditions were used prior to centrifugation to mimic the impact of high-shear regions in the inlet of centrifuges, leading to potential cell disruption and carry-over of solids contaminants. The remaining suspension was kept as a no shear control condition. The optical density at 600nm was measured for both the sheared and the non-sheared feed suspensions (OD_f) to generate data to calculate % clarification from Equation 4.1.

$$\%C = \frac{OD_f - OD_s}{OD_f - OD_r} \times 100$$

Equation 4.1 % clarification for centrifugation.

Where OD_f , OD_s and OD_r are the optical density of the feed, the sample supernatant and the supernatant of the reference sample, respectively (Boychyn et al. 2004).

Clarification was mimicked by taking 1.5 mL samples of non-sheared and sheared material and placing these in 2.2 mL Eppendorf tubes. The tubes were spun in an Eppendorf 5810R centrifuge, rotor T-60-11 (Eppendorf North America Inc.,

New York; $N = 25$ to 208.3 rps, $\Sigma_{lab} = 0.009$ to 0.617 m², $t_{lab} = 0.2$ h). The supernatant was removed by pipetting and the OD_{600} of the supernatant was measured (OD_s). The well-spun material ($N=208.3$ rps, $t_{ref}=0.5$ h) was prepared for the purpose of a clarification reference sample with OD_{600} measured (OD_r). Where necessary this material was also used to dilute samples in order to operate in the linear optical range of the spectrophotometer. 50mM PBS, pH 7.4 was used as a blank.

The hindered settling correction factor was calculated on the % solids of the material just prior to laboratory centrifugation.

4.2.2 Pilot plant studies

4.2.2.1 Flocculation

E.coli cells were heat lysed before exposure to a 1% w/v solution of flocculating agent in water for injection (WFI). The amount of flocculating agent added was changed from 3 g/kg to 4.5 g/kg as well as the addition rate of the flocculating agent from 9 L/min to 4.5 L/min. The material was mixed for 15 minutes before water was added prior to centrifugation. Water was added to make final weight ratio of 1.0 w/w (total additions including flocculating agent solution/heat extract) to improve product yield and centrifuge performance.

4.2.2.2 Centrifugation

Flocculated material was clarified with a CSA-8 disk-stack centrifuge (Westfalia Separator AG, Oelde, Germany) at a flowrate of 120 L/hr. For one run additional flowrates of 180 and 270 L/hr were used to verify the pilot clarification with the laboratory scale. The optical density of the supernatant (OD_s) was measured at 600nm. Temperature was maintained at $\sim 10^{\circ}\text{C}$ using glycol cooling circulating at -4°C . Solids were discharged immediately upon the bowl reaching capacity (indicated by breakthrough of solids to supernatant). Equation 4.1 was also used to calculate measurements of % clarification for pilot scale. The well-spun

material ($N=208.3$ rps, $t_{ref}=0.5$ h) was prepared for the purpose of a clarification reference sample with OD_{600} measured (OD_r).

4.2.3 Microscope images

Microscope images were taken on an Axioscope 2 plus microscope (Carl Zeiss MicroImaging Inc., Thornwood, NY, U.S.A.) at 10x and 100x magnification.

4.3 Theoretical considerations

4.3.1 Sigma theory

Sigma theory, is a concept used to describe the solid-liquid separation performance of a centrifuge and can be used to determine the clarification time. The Sigma factor, also called equivalent settling area, is defined as the surface area of a gravity settling tank with equivalent settling performance. The derivation of the Sigma concept is described elsewhere (Ambler,1961).

Each type of centrifuge design has a unique correlation for calculating a Sigma value. Equation 4.2 & 4.3 show the formulas for calculating the Sigma value for a laboratory and disc-stack centrifuge respectively.

$$\Sigma_{lab} = \frac{V_{lab}\omega^2}{2g \ln\left(\frac{2R_o}{R_o + R_i}\right)}$$

Equation 4.2 Sigma for a swing-out laboratory centrifuge.

Where Σ_{lab} is the equivalent settling area of the laboratory centrifuge (m^2), V_{lab} is the volume of process material in the centrifuge tube (m^3), $\omega = 2\pi N$ ($rad.s^{-1}$), N is the rotational speed (rps), g is acceleration due to gravity (ms^{-2}), R_i is the inner radius (m) (the distance between the centre of rotation and the top of the liquid), R_o is the outer radius (m) (the distance between the centre of rotation and the bottom of the tube) (Maybury et al. 1998).

$$\Sigma_{ds} = \frac{2\pi n_d \omega^2 (r_2^3 - r_1^3)}{3g \tan \theta}$$

Equation 4.3 Sigma for a disc-stack centrifuge.

Where Σ_{ds} is the equivalent settling area of a disk stack centrifuge (m^2), n_d is the number of active discs, r_1 and r_2 are the inner and outer disc radius (m), θ is the half disc angle (degrees).

The ease of broth clarification for any particular centrifuge is given by the ratio of the volumetric flow rate, Q , and settling area, Σ . Therefore comparisons between different laboratory centrifuges, or the same centrifuge using different settings, can be performed using Equation 4.4.

$$\frac{Q_{lab}}{c_{lab} \Sigma_{lab}} = \frac{Q_{ds}}{c_{ds} \Sigma_{ds}}$$

$$\frac{V_{lab}}{c_{lab} t_{lab} \Sigma_{lab}} = \frac{Q_{ds}}{c_{ds} \Sigma_{ds}}$$

Equation 4.4 Comparing two different centrifuges.

Where Q is the flow rate ($m^3 s^{-1}$), t is the residence time (s), and c is a correction factor to account for non-ideal fluid flow patterns, the value of which depends on centrifuge design. Subscripts *lab* and *ds* represent the laboratory and disk stack centrifuge respectively. The laboratory centrifuge is usually considered a reference, hence $c_{lab} = 1.0$ for this centrifuge.

For high solids feed material there is a need to overcome the phenomenon of hindered settling in order to get a good comparison between the clarification in laboratory scale to that observed at pilot scale.

4.3.2 Hindered settling

Hindered settling starts to occur at high solid concentrations. As the solids concentration increases, the settling rate of particles decreases due to the increase

in interparticle forces. Particles settling slowest will be hindered by those underneath them. This is particularly noticeable in the confined spaced of a laboratory centrifuge. At very high solids what is known as “blanket” sedimentation can occur. This is when particles all settle at the same rate and form a large floc that sweeps up the fine particles in the liquid. This leads to an over prediction in the laboratory scale to that seen in the pilot scale continuous flow centrifuge.

Due to the high shear in the feed zone and short residence time in the centrifuge bowl it is believed that flocs do not have enough time to reform and therefore hindered settling is not a main concern at pilot scale. This is not the case at laboratory scale where the length of time between shearing material and clarification may be enough to allow flocs to reform and cause hindered settling which consequently increases clarification performance in the laboratory scale compared to the pilot scale at an equivalent separation condition. In order to allow a more accurate comparison between the USD and pilot centrifuges, we dilute the floc for the USD experiments to such an extent that floc reformation is minimized. A correction to the USD data to allow for hindered settling at high solids concentration in the pilot centrifuge must then be applied in Equation 4.5.

It has been shown in previous USD studies (Boychnyn et al. 2001; Boychnyn et al. 2004; Maybury et al. 2000) that diluting material to approximately 2% w/w may help reduce hindered settling, without greatly affecting the liquid viscosity.

In 1954, Richardson & Zaki developed a mathematical relationship for hindered settling, shown in Equation 4.5 that can be used to predict the clarification of high solid density feeds in an industrial centrifuge from laboratory clarification data (Richardson and Zaki, 1997).

$$\frac{\left(\frac{Q}{c\Sigma}\right)_{ds}}{\left(\frac{Q}{c\Sigma}\right)_{lab}} = \frac{(1 - \varphi_{ds})^{4.8}}{(1 - \varphi_{lab})^{4.8}} = h$$

Equation 4.5 Hindered settling correction factor

Where φ is the cellular solids fraction and h is the hindered settling correction factor that equals to 1 for no hindered settling.

The derivation of the hindered settling correction factor concept is described elsewhere (Salte et al. 2005).

4.4 Results

4.4.1 Evaluation of shear

To understand the size and morphology of the flocculated solids and how shear affects these parameters, microscope images were taken (Figure 4.1). It can be seen that particle size decreases with increasing shear represented here by the speed of the rotating disc in the shear device. Particle size was approximately 75 μm for the no and low shear condition and 25 μm for the high shear condition suggesting that as shear was increased, particles were being broken down. Data shown in Figure 4.2 supports these results of shear sensitivity by giving a quantitative measure of shear effects. In this figure and other figures like this centrifuge speed was changed to alter the equivalent settling area. An increase in the equivalent flow rate to equivalent settling area ratio corresponded to a decrease in centrifuge speed from 208.3 to 25 rps while maintaining the time for centrifugation constant at 12 minutes. There is clearly a difference in % clarification between the material before and after exposure to high shear at any particular equivalent flow rate to equivalent settling area ratio (Figure 4.2). Both Figure 4.1 & Figure 4.2 results determined that the flocculated particles were sensitive to the shear conditions they were exposed to, which were selected to be relevant to process scale centrifugation (Neal et al. 2003).

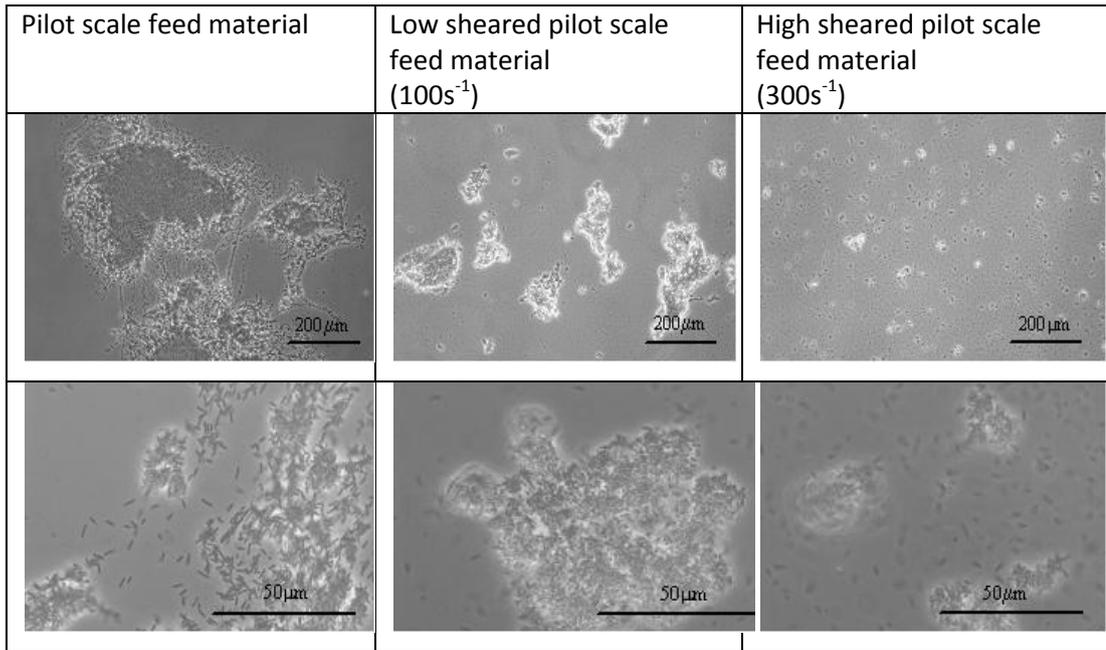


Figure 4.1 Microscope images of flocculated ApoA-IM material exposed to shear.

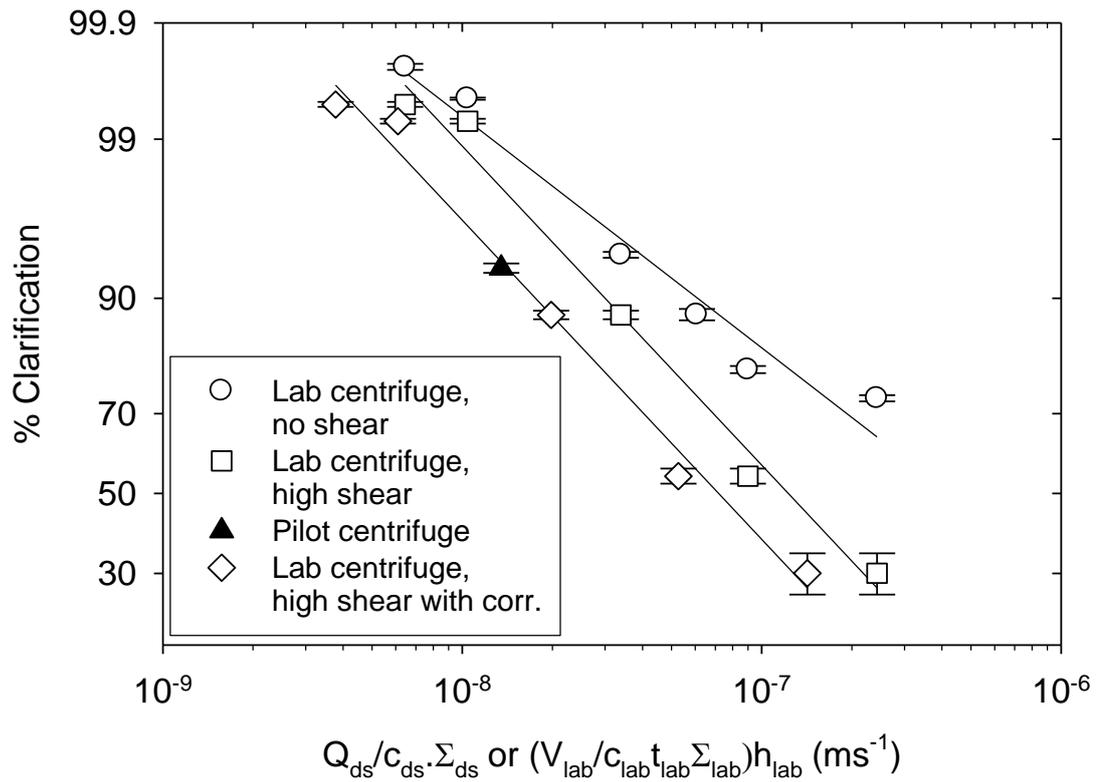


Figure 4.2 Developing a mimic of the ApoA-IM dimer recovery process

Flocs were exposed to the following shear conditions in the shear device: (\circ) non-sheared; (\square) high sheared, $300s^{-1}$ (relates maximum energy dissipation rate of 2×10^6)

W/kg) (\diamond) high sheared with hindered settling correction factor applied; (\blacktriangle) pilot centrifuge. The x-axis is the ratio of equivalent flow rate to equivalent settling area. All laboratory samples were sheared for 20s. Data points are plotted on probability-logarithmic axes indicate mean values with errors from duplicate runs. Solid lines represent the line of best fit for each data set. The hindered settling correction used (h_{lab}) was calculated from Equation 4.5 as 0.57. The values of c_{ds} and c_{lab} were 0.4 and 1 respectively.

4.4.2 Adoption of a hindered settling correction factor

The solids percentage in the pilot centrifuge feed of ~20% was far greater than the 2% threshold where this type of USD work has previously been carried out successfully without the hindered settling correction factor (Boychyn et al. 2001;Boychyn et al. 2004;Maybury et al. 2000). For this reason, the pilot and USD data (USD feed material is diluted to 2% solids before centrifugation) consistently didn't match one another by similar degrees each time. After a hindered settling correction factor of 0.57 was taken into account, the USD data was found to be in much greater agreement with the pilot data (Figure 4.2). The phenomenon for hindered settling is described in section 4.3.2.

4.4.3 Comparing % clarification between pilot scale and USD mimic

For one of the pilot plant runs, it was possible to operate the disc-stack centrifuge at three different flowrates. The objective was to investigate the effect of flowrate on % clarification and to get a better comparison between lab and pilot scale (Figure 4.3). A generally good agreement in trends can be seen between scales.

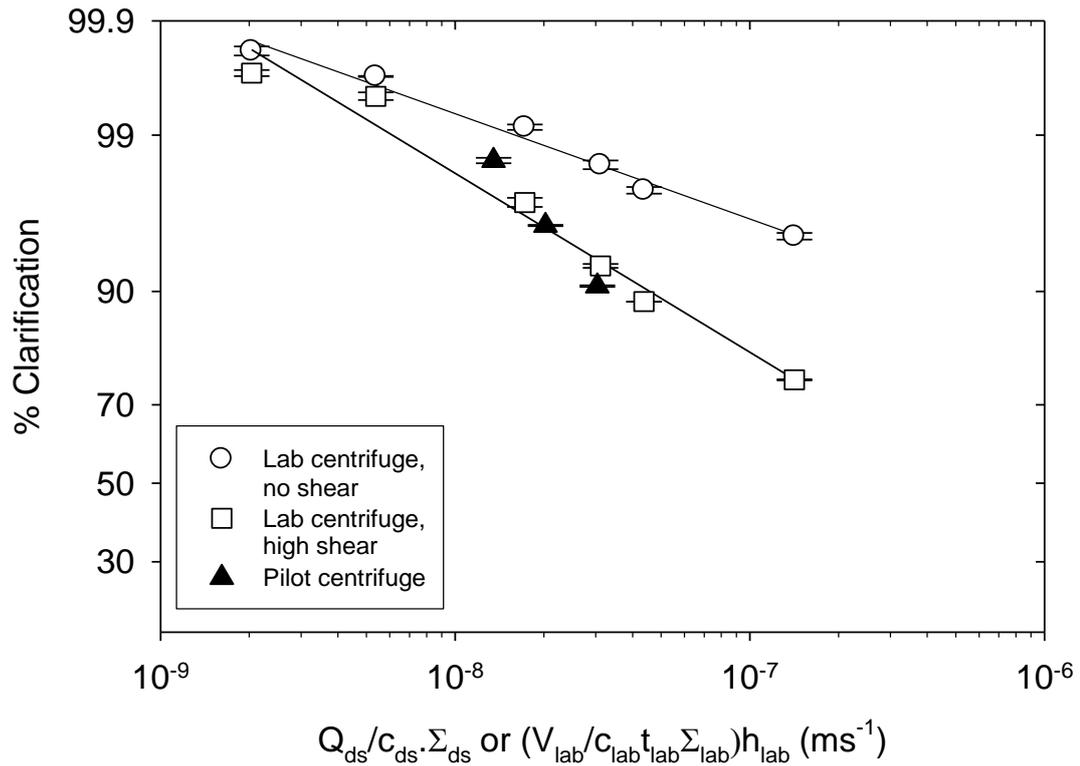


Figure 4.3 Comparing the USD % clarification with pilot scale

Floc suspension was exposed to the following shear conditions in the shear device: (○) non-sheared; (□) high sheared, 300s⁻¹; (▲) pilot centrifuge. The x-axis is the ratio of equivalent flow rate to equivalent settling area. All laboratory samples were sheared for 20s. Data points are plotted on probability-logarithmic axes and indicate mean values with errors from duplicate runs. Solid lines represent the line of best fit for each data set. The hindered settling correction used (h_{lab}) was calculated as 0.57. The values of c_{ds} and c_{lab} were 0.4 and 1 respectively.

4.4.4 Varying flocculating agent addition rate for both scales

Two different addition rates were used in generating pilot scale flocculated material (Figure 4.4). The rates used were 4.5 L/min and 9 L/min. It was observed that the higher addition rate produced a higher variability in flocculated particle size by visual observation (results not shown). This high addition rate may reduce the chance of an even distribution of flocculating agent which would result in a more heterogeneous population of particle sizes.

It can be seen that for both addition rates at the low equivalent flow rate to equivalent settling area ratio (high centrifuge speed, 208.3rps) a greater than 99% clarification is achieved for both the no and high shear conditions (Figure 4.4). It seems that shear doesn't seem to have a big affect at such high centrifugation speeds, probably due to even smaller particles being removed at such high speeds. Unfortunately we cannot feasible reach these conditions in the pilot centrifuge as the data point shown is already at the highest centrifugation speed possible. Such high solids concentration also requires the flowrate to be set in a very narrow range. Low flowrates result in clarifications taking too long, while high flowrates can cause the centrifuge to become overloaded with material, this can cause some flocs to overspill into the supernatant, resulting in inadequate clarification.

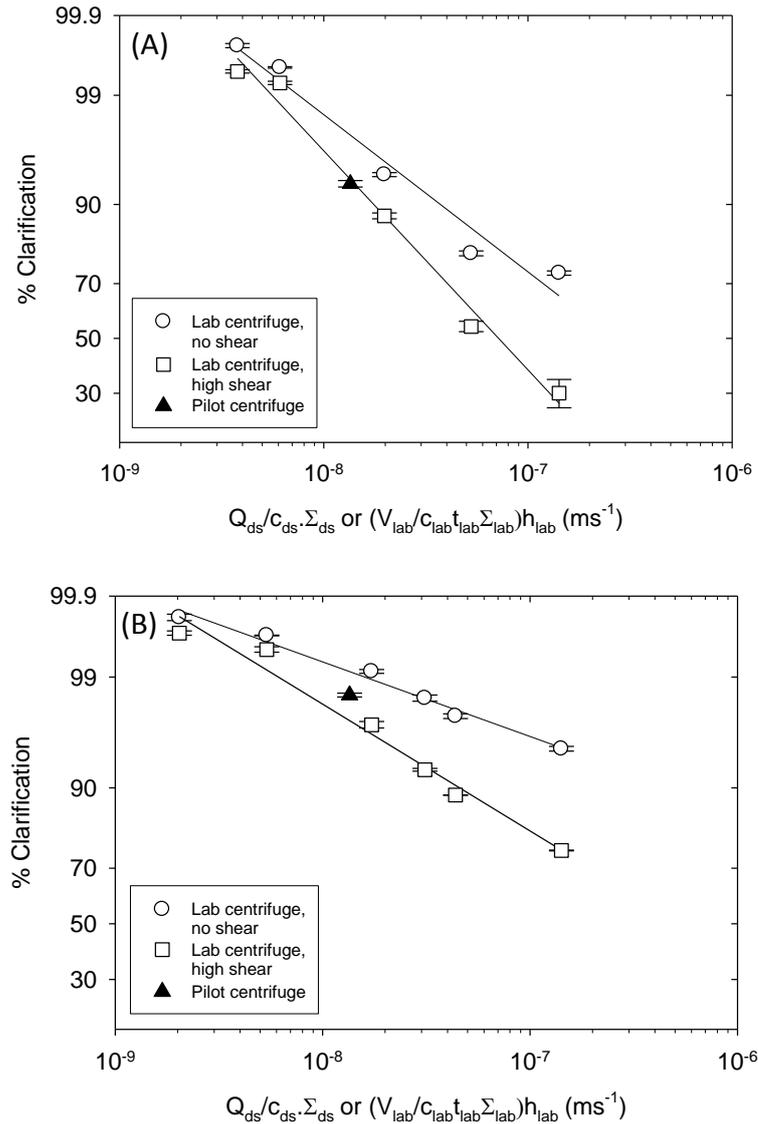


Figure 4.4 Varying flocculating agent addition rate at both scales.

(A) Flocs prepared with a flocculating agent addition rate of 9 L/min (B) Flocs prepared with a flocculating agent addition rate of 4.5 L/min. Floc suspension exposed to the following shear conditions in the shear device: (○) non-sheared; (□) high sheared, 300s⁻¹; (▲) pilot centrifuge. The x-axis is the ratio of equivalent flow rate to equivalent settling area. All laboratory samples were sheared for 20s. Data is corrected with a hindered settling correction factor. Data points are plotted on probability-logarithmic axes and indicate mean values with errors from duplicate runs. Solid lines represent the line of best fit for each data set. The hindered settling correction used (h_{lab}) was calculated as 0.57. The values of c_{ds} and c_{lab} were 0.4 and 1 respectively.

For the high addition rate, clarification in both the no and high shear conditions drops more rapidly as the ratio of equivalent flow rate to equivalent settling area is increased (reducing centrifuge speed). This may be explained by the previous observation of increased particle size heterogeneity at the higher addition rate which would lead to the presence of smaller floc particles causing reduced clarification when compared to the lower addition rate.

Another observation is that at the high addition rate, the difference in clarification between the no and high shear condition is greater than that for the low addition rate. At the 9 L/min addition rate and the highest equivalent settling area, the % clarification is 73 and 30%, more than two-fold drop, for the no and high shear rates respectively. Whereas for the 4.5 L/min addition rate the % clarification is 95 and 76% for the no and high shear rates respectively. This suggests that the particles at the high addition rate are more shear sensitive which could also be due to inadequate mixing causing the particles to be larger, more fragile and more heterogeneous than they were at the lower addition. The time for mixing and the flocculant addition rate are not independent from one another, using a lower flow rate not only results in a steadier and controlled flocculation but also gives the particles more time to age and hence increases their mechanical strength. Once the lower addition rate was introduced there was also a notable increase in the reproducibility of these USD experiments (results not shown). Note that the pilot plant data points more closely align with high shear results in both runs.

4.4.5 Varying flocculating agent addition ratio with USD mimic

An experiment was carried out at laboratory scale to study the effect of flocculant levels on clarification. Three different flocculant amounts were used; 3, 5 and 7 g/kg of heat extract. Figure 4.5 shows that the amount of recoverable solids increased as the amount of flocculating agent added was increased. The biggest shift observed was from 3 to 5 g/kg, which may indicate that the original amount of

flocculant added in the pilot plant of 3 g/kg may not be sufficiently flocculating all of the smaller cell debris.

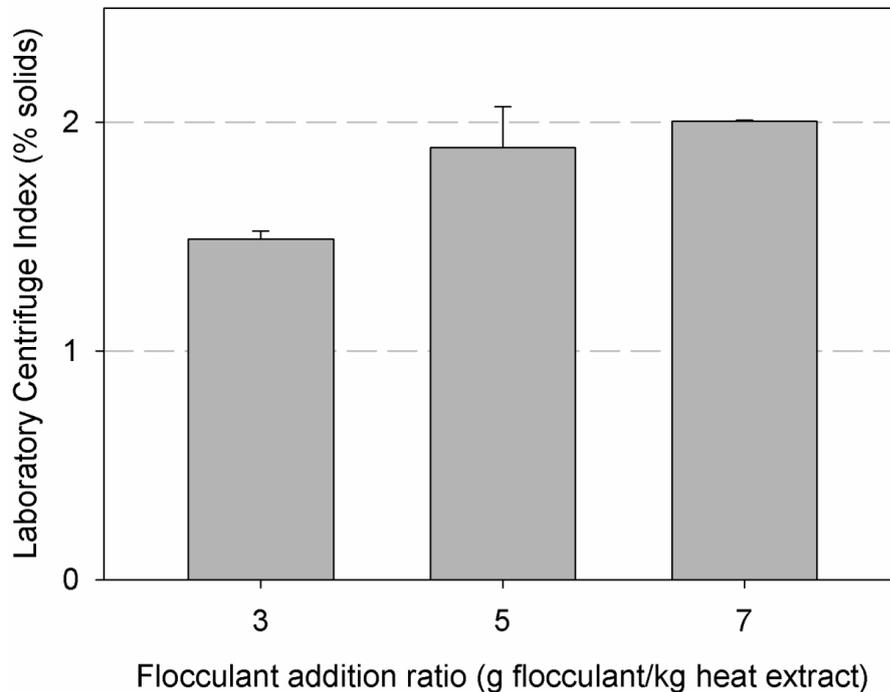


Figure 4.5 Using the USD mimic to study the effect of varying the ratio of flocculating agent addition on the laboratory centrifuge index.

The laboratory centrifuge index was calculated as the % solids (w/w) in the floc after varying amounts of flocculating agent had been added. The solid pellet was created at 166.7 rps for 30 minutes in a Sorvall RC6 centrifuge. All laboratory samples were sheared under the high shear condition for 20s. Bars indicate mean values with errors from duplicate runs.

At these three different flocculating agent addition levels, representative samples underwent two shear conditions, low shear (100s^{-1}) and high shear (300s^{-1}). From the % solids the hindered settling correction factor was determined using Equation 4.5 for each of these six samples. Figure 4.6 shows that little change in clarification was observed for the material that was exposed to the low shear. However, Figure 4.2 shows that the high shear condition was the closest match to the shear experienced at pilot scale and Figure 4.6 shows that under this condition

there was an increase in % clarification from 88 to 99% as the level of flocculating agent was increased. This indicated that as the level of flocculating agent was increased, the floc particles produced were stronger and hence easier to clarify.

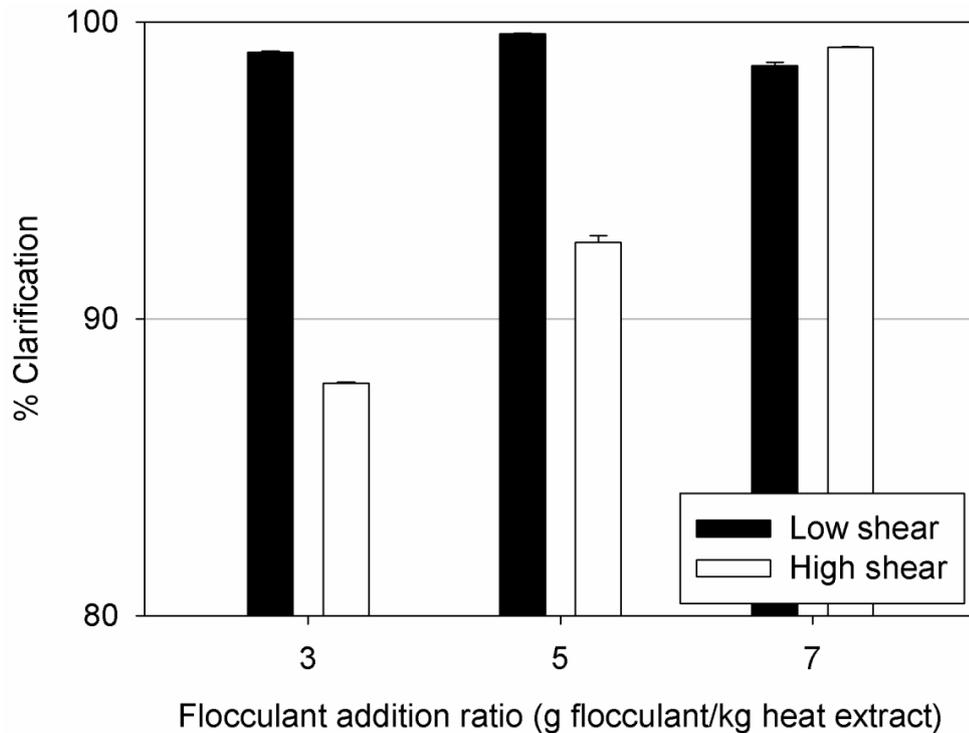


Figure 4.6 Using the USD mimic to study the effect of varying the ratio of flocculating agent addition on the % clarification.

USD conditions were chosen that matched the pilot plant and each condition was corrected for hindered settling using the laboratory centrifuge index (Equation 4.5). Flocculation suspension exposed to the following shear conditions in the shear device: (dotted bars) low shear ($100s^{-1}$); (dashed bars) high shear ($300s^{-1}$). The high shear condition was used as earlier results indicated that this was the closest match to the shear experienced at pilot scale. All laboratory samples were sheared for 20s. Bars indicate mean values with errors from duplicate runs.

4.4.6 Validation of USD flocculating agent addition ratio at pilot scale

Based on laboratory data shown previously as well as Pfizer supporting data it was decided that the amount of flocculating agent added at pilot scale should be increased to 4.5 g flocculating agent/kg heat extract.

Material produced in the pilot plant by flocculation with 4.5 g flocculant/kg heat extract was used in laboratory studies to look at how sensitive this new material was to shear and hence if the % clarification was improved over a range of centrifugation conditions. The USD data in Figure 4.7 shows that clarification was improved overall compared to the experiments carried out with material made with 3 g/kg and the same addition rate.

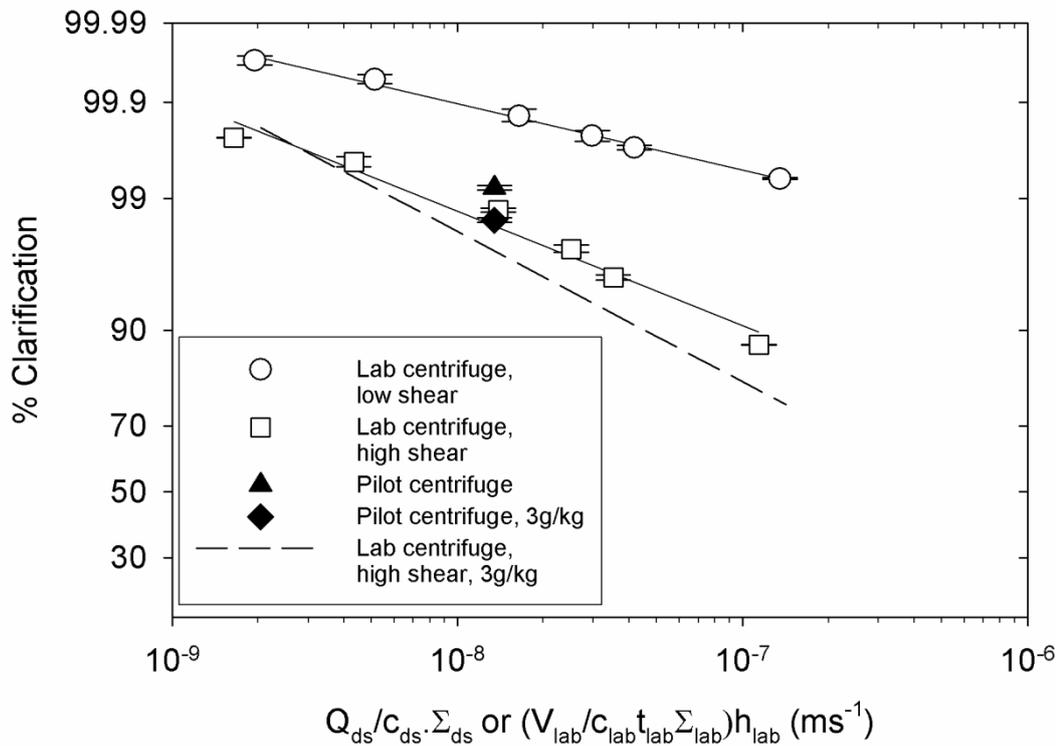


Figure 4.7 Validating at pilot scale the USD predictions on the effect of flocculating agent addition ratio.

A new flocculation addition ratio of 4.5 g flocculating agent/kg heat extract was used unless stated otherwise while addition rate was kept constant at 4.5L/min. Floc suspension exposed to the following shear conditions in the shear device: (\circ) non-sheared; (\square) high sheared, 300s^{-1} ; (\blacktriangle) pilot centrifuge, (\blacklozenge) pilot centrifuge with

0.3kg flocculating agent/kg heat extract, (– line) high sheared, $300s^{-1}$, with 3 g flocculating agent/kg heat extract. The x-axis is the ratio of equivalent flow rate to equivalent settling area. All laboratory samples were sheared for 20s. Data points are plotted on probability-logarithmic axes and indicate mean values with errors from duplicate runs. Solid lines represent the line of best fit for each data set.

This improvement was more pronounced at the lowest centrifugation speed where, under high shear with 3g/kg, the % clarification obtained was 76% compared to the 88% at the new 4.5g/kg addition. Possibly, higher flocculant levels could lead to stronger particles. There could also be a greater number of smaller debris being flocculated, which improves the clarification. Laboratory % solids results support the latter assessment (Figure 4.5).

4.5 Summary

This chapter described the use of an ultra scale-down methodology for solid-liquid separation in industrial centrifuges for a flocculated feed material.

Review of chapter aims:

- Demonstrate how a mimic for centrifugal clarification can be applied to a flocculated *E.coli* extract and give an insight into the interaction between flocculation and centrifugation.
 - The feedstock is a high solids (20% solids) shear sensitive flocs.
 - The high solid content necessitated the inclusion of a dilution step and a hindered settling correction factor for the USD centrifugation.
 - Using this approach it was possible with small volumes (approx. 100 ml for full characterization) to rapidly investigate the interaction between flocculation and centrifugation conditions for this material.
 - The developed method was utilized to optimize the flocculation step at large-scale, which was shown to improve the overall clarification.

The next three chapters of this thesis will now investigate how to improve the purification of the material produced from this flocculation/centrifugation stage.

CHAPTER 5 DEVELOPING AN EXPERIMENTAL METHODOLOGY TO DEFINE PRODUCT AND CONTAMINANT RELATIONSHIPS

5.1 Introduction

Cordingley *et al.*, investigated some of the factors involved in SELDI sample preparation and application (Table 5.1). The response variables chosen were based on the number of peaks, mass range of detectable peaks and resolution parameters.

| Factor | Type | Low level (-) | Midpoint | High level (+) |
|------------------------------------|------------|-----------------------|------------------------|----------------------|
| Concentration and volume of sample | Continuous | 100 μ L at 1mg/mL | 40 μ L at 2.5mg/mL | 25 μ L at 4mg/mL |
| Length of washes | Continuous | 0 min | 5 min | 10 min |
| Number of washes | Continuous | 2 | 3 | 4 |

Table 5.1 Factors in SELDI ProteinChip preparation.

This group found that the best settings from those investigated were; minimal wash times and fewest number of washes. Volume and protein concentration of sample could be set as convenient.

This work along with information supplied by BioRad gives a good background on some of the factors that have an effect on SELDI-MS. This chapter however investigates a couple of specific factors involved in the optimisation of the protocols for the selective binding of proteins to the ProteinChip surface, including decisions on which solutions to use.

These adapted protocols for the selective ProteinChips will be used for later purification development work in chapter seven.

Chapter aims:

- Finalise SELDI protocols for later purification development study, including the selection of solutions A & B.

- Use protocols to generate adsorption data for the ApoA-IM material produced in the previous chapter for use in the remaining sections of this thesis.

5.2 Materials & Methods

Unless specified otherwise the materials & methods are unchanged from section 2.2.

5.2.1 Reducing pH below the pI of the ApoA-IM monomer

Due to ApoA-IM having a pI of approximately 5.4, it was necessary to add 1% polysorbate 20 to the ApoA-IM post centrifuge supernatant (Figure 3.1, sample E) before dropping the pH to 3 to avoid protein precipitation. Once the polysorbate was added and the pH was reduced, the polysorbate 20 was removed using a 3mL slide-alyser cassette (Thermo Fisher Scientific Inc., IL, U.S.A.) with a 2.5kDa cut-off.

5.2.2 Mass spectrometry

5.2.2.1 Selective ProteinChip preparation

Before ProteinChip equilibration, the H50 ProteinChips were activated by incubating twice with 50µL of 50% Acetonitrile for 5 minutes and the IMAC10 ProteinChips were activated by incubating twice with 50µL of 100mM nickel sulphate for 5 minutes.

After this stage the selective ProteinChips (CM10, Q10, H50 & IMAC10) used in this chapter were prepared with identical protocols (Figure 5.1) but with buffers specific to their surface chemistry. Spots were equilibrated three times with 150µL of solution A for 5 minutes under shaking (8.3 rps). The spots were then loaded with 150µL of the sample previously diluted in solution A based on previous load optimisation results. The sample was incubated with the spot surface for 30 minutes while being shaken (8.3 rps). Each spot was then washed three times with 150µL of solution A for 5 minutes under shaking (8.3 rps) to eliminate non-

adsorbed proteins. This was followed by a quick rinse with 150 μ L of solution B to reduce salt levels and then the spots were left to dry.

After 20 minutes the spots were dry and 1 μ L of the EAM was applied. After 5 minutes the EAM application was repeated before a final 5 minute dry step prior to ProteinChip analysis.

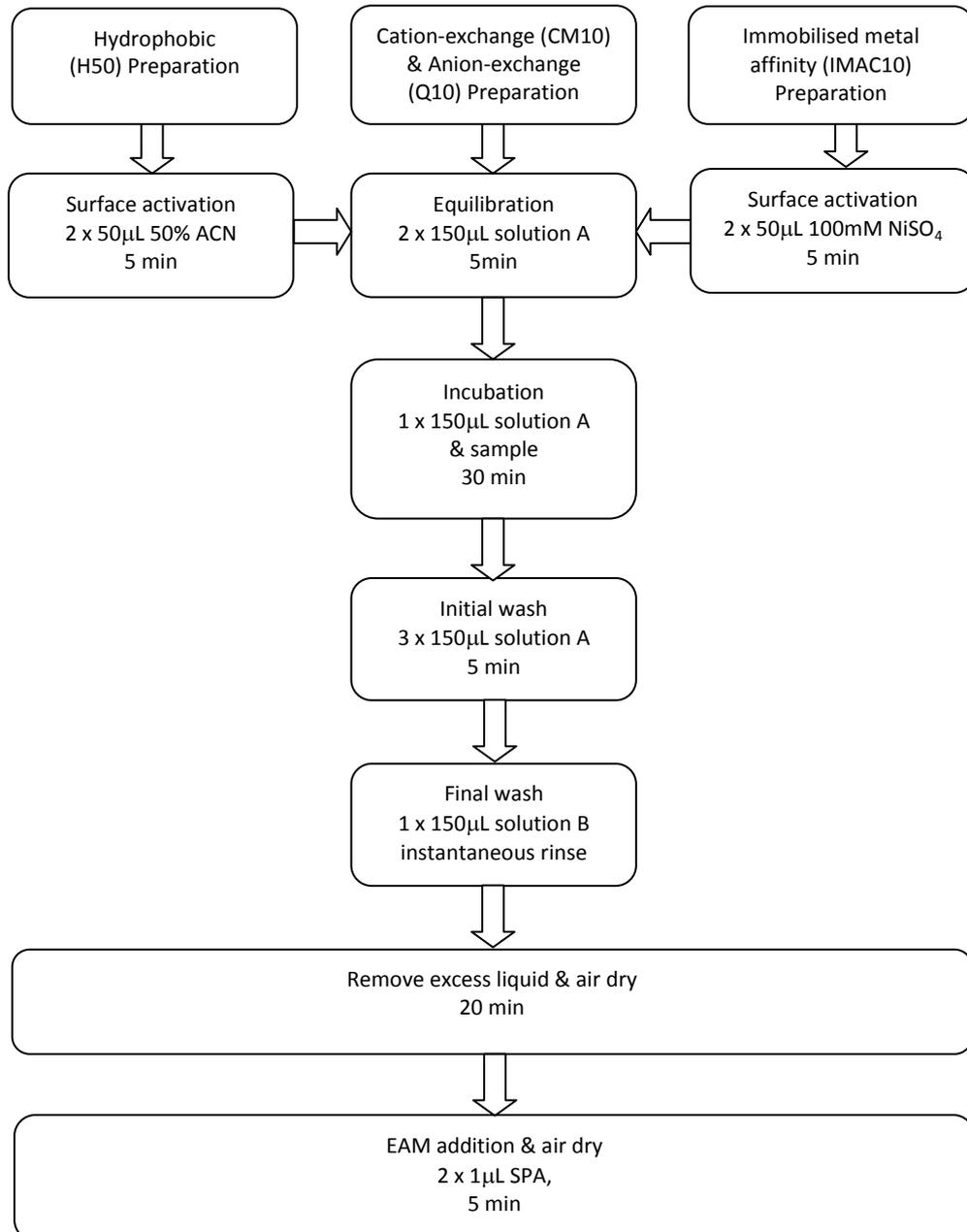


Figure 5.1 Steps for selective ProteinChip preparation.

Solution A & B will vary depending on the experiments in later sections.

5.2.2.2 Data treatment

SELDI-MS data in this chapter was treated with Matlab as described in section 3.2.

5.3 Results

Typical protocols for the preparation of the various ProteinChips are shown in Figure 5.1. This chapter will look at the selection of solution A and solution B.

5.3.1 Selection of wash step solutions for hydrophobic ProteinChips

The protocols supplied by BioRad (and previously CIPHERGEN) suggest using the H50 ProteinChips in a reverse phase mode, with acetonitrile (ACN) and trifluoroacetic acid (TFA). These solvents have an advantage because of the absence of salt, removing the need for a final water wash and because they have been used previously with HPLC for elucidating the hydrophobic characteristics of proteins (Wall et al. 2002). The down side of using these solvents are their minimal use at larger scales and the possibility of unknown damaging effects on the proteins themselves. This damage is one of the reasons why aqueous biphasic systems are sometimes used to measure hydrophobicity.

Along with investigating the use of these reverse phase solvents further we therefore also wanted to investigate ammonium sulphate buffers which are more commonly used at industrial scale and maybe less damaging to the proteins.

5.3.1.1 Single Protein solutions

In Figure 5.2, polar (1% TFA in ultrapure water) and non-polar (50% ACN in ultrapure water) solvents were used for binding BSA & Lysozyme on an H50 ProteinChip following the protocol in Figure 5.1. As can be seen in Figure 5.2 there is no binding of lysozyme with the 1% TFA condition on the hydrophobic ProteinChip but considerable binding of BSA. However the opposite is observed with the use of 50% ACN. A possible reason for this is that hydrophobic proteins

have increased stability adsorbed to a hydrophobic surface when exposed to increased solvent polarity. A hydrophilic protein displays the opposite, being increasingly stable in solution with increased solvent polarity. The higher adsorption of BSA relative to lysozyme in the polar solvent suggests that BSA is a more hydrophobic protein. This observation is supported by conventional RP-HPLC where lysozyme elutes before BSA with an increasing percentage of non-polar solvent (Parris and Baginski,1991).

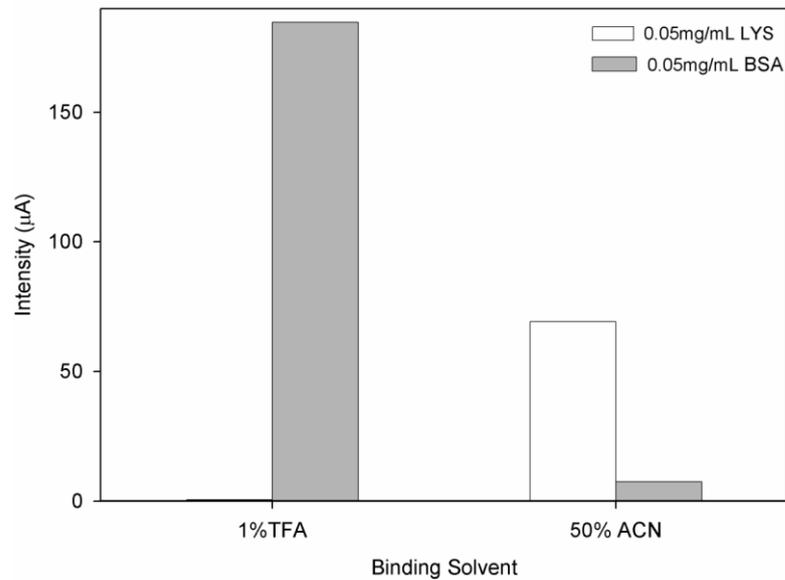


Figure 5.2 Competitive adsorption of BSA and lysozyme to a hydrophobic ProteinChip using reverse phase solvents.

Equal quantities of BSA and lysozyme were mixed in 10mM PBS, pH 7.4 to make 0.1mg/mL total protein. The TFA solution was 1% TFA diluted in ultrapure water and the ACN solution was 50% ACN diluted in ultrapure water. The solutions were used for all steps in the ProteinChip preparation. The laser strength used was 2500 & 1750nJ for BSA & lysozyme respectively.

Due to the concerns raised with reverse phase solvents it was decided to compare TFA with ammonium sulphate solutions. Since BSA exhibited hydrophobic properties it was used in this study. Figure 5.3 shows that the two solvent systems show similar trends, with the binding of BSA dropping as the percentage of TFA or ammonium sulphate concentration was lowered.

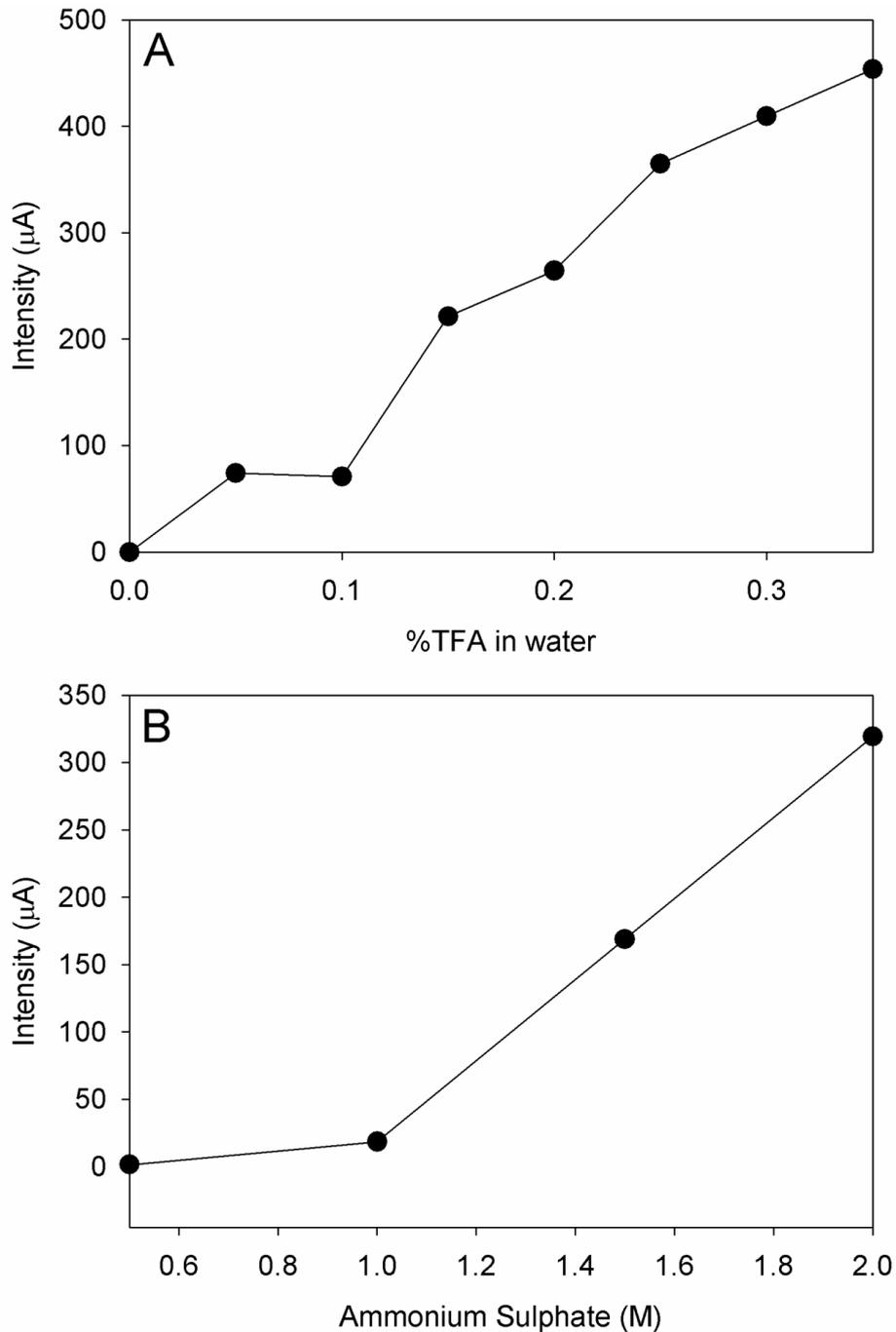


Figure 5.3 Non-competitive adsorption of BSA to a hydrophobic ProteinChip using TFA or ammonium sulphate.

A solution of 0.1mg/mL of BSA was made with 10mM PBS, pH 7.4. This solution was incubated with either (A) varying TFA concentration (used as solution A & B) in ultrapure water or (B) 50mM Tris, pH 7 with varying ammonium sulphate concentration (solution A) and a final wash with 50mM Tris, pH 7 with no ammonium sulphate (solution B). A laser strength of 6000nJ was used.

5.3.1.2 Complex mixtures

To further aid the decision on whether to use ammonium sulphate or reverse phase solvents, a comparison of the two methods was carried out using a complex process material. The material chosen was the ApoA-IM post centrifuge supernatant produced from the previous chapter.

Figure 5.4 shows that ammonium sulphate improved the adsorption of ApoA-IM post centrifuge supernatant proteins to the H50 ProteinChip in greater magnitude than ACN even though the ammonium sulphate spots had a final water wash which could have removed proteins.

Precipitation may explain why the spectrum without ammonium sulphate has a higher signal intensity than the spectra with ammonium sulphate even though visually there was no apparent visual turbidity. The “salting-in” effect by ammonium sulphate may increase the surface charge of the proteins, reducing their adsorption to the non-polar surface (Arakawa and Timasheff,1984).

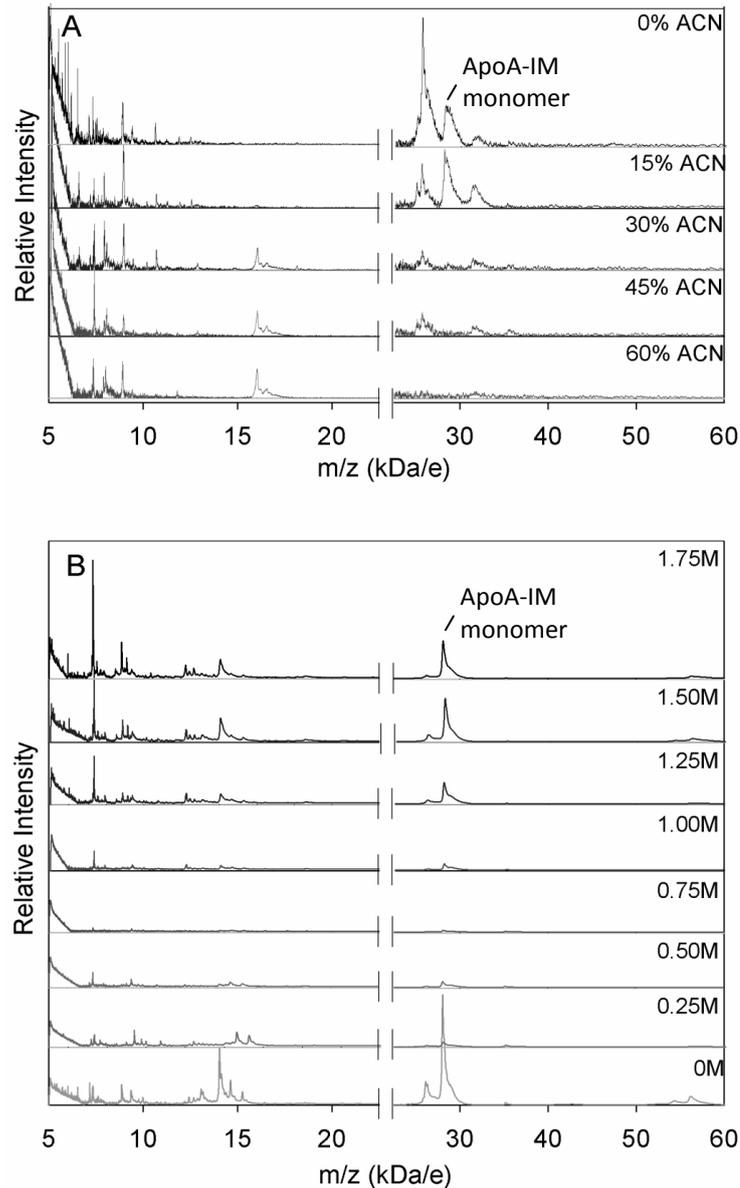


Figure 5.4 Competitive adsorption of ApoA-IM post centrifuge supernatant proteins to a hydrophobic ProteinChip with ACN or ammonium sulphate.

A solution of 0.05mg/mL ApoA-IM post centrifuge supernatant was incubated with; (A) varying ACN concentration in ultrapure water (used for solutions A & B) or (B) 50mM Tris, pH 7 with varying ammonium sulphate concentration (solution A) and a final wash with 50mM Tris, pH 7 with no ammonium sulphate (solution B). A laser strength of 2000 & 6000nJ was used for the low and high m/z regions respectively (regions separated by breaks on x-axis). For figure A the y-axis scale is 0 – 60 μ A and 0 – 20 μ A for low and high m/z respectively. For figure B the y-axis scale is 0 – 1000 μ A and 0 - 1200 μ A for low and high m/z respectively.

5.3.2 Selection of final wash step solutions for all of the selective ProteinChips

The final wash step to prepare the selective ProteinChips (CM10, Q10, H50 & IMAC10) involves an instantaneous ultrapure water wash of the ProteinChips to remove salts that may interfere with laser desorption/ionisation (LDI) (Figure 5.1., solution B). Since later work required using the ProteinChips with an analogous method to that used with a chromatography column, it would be beneficial to remove this water wash step. It was also a concern that washing with water may remove or irreversibly affect the proteins bound. This was a particular concern for the hydrophobic (H50) ProteinChips since proteins typically elute from hydrophobic surfaces in low salt conditions.

The aim was to replace the water with a low molarity solution at the same pH as the buffer used in the preceding steps (Figure 5.1, solution A). It was hypothesized that if the molarity was low enough there would be negligible adverse effects on ionization particularly as the proteins already contribute their own ion suppression (Annesley,2003).

5.3.2.1 Hydrophobic ProteinChips

As shown previously, hydrophobic ProteinChips satisfactorily adsorbed proteins in the presence of ammonium sulphate. The only remaining foreseeable downside of using salt instead of reverse phase solvents was that the low salt concentrations during the final wash step would be very different to the high salt concentrations used during adsorption and hence proteins may desorb from the surface.

To test this hypothesis, a hydrophobic ProteinChip was pre-incubated with BSA using 50mM Tris, pH 7 with either 1 or 2M ammonium sulphate (Figure 5.5). By varying the final water wash conditions for this ProteinChip it was possible to witness desorption. The result demonstrated that once BSA was adsorbed its desorption within a time frame of 9.5 minutes was gradual. This result was the same for both the 1 & 2M ammonium sulphate conditions. The desorption even

seemed to be less than a control spot that used 0.35% TFA in ultrapure water as a final wash condition, which was known to be a good adsorption condition from Figure 5.3 (A).

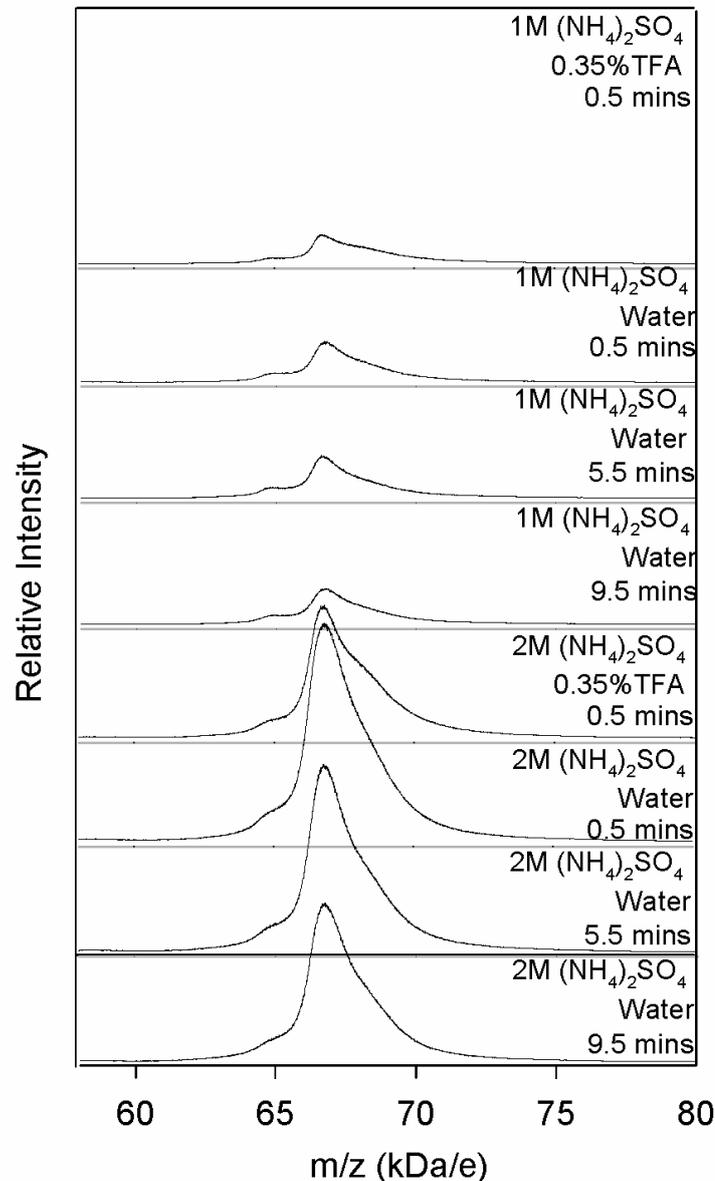


Figure 5.5 Non-competitive adsorption of BSA to a hydrophobic ProteinChip by varying ammonium sulphate concentration and incubation time.

A solution of 0.1mg/mL BSA was incubated with 50mM Tris, pH 7. In the legend for each spectrum, the top line is the ammonium sulphate concentration in 50mM Tris, pH 7 used for equilibration, sample incubation and wash; the middle line is the final wash solution and the bottom line is the time for the final wash step. A laser strength of 3000nJ was used.

5.3.2.2 Ion Exchange ProteinChips

It now seems unlikely that adsorbed proteins are removed by water within the time frames of the protocols used. A change in pH from the equilibration, binding and wash solution to the final wash with water could still however affect the adsorption/desorption of the proteins. ApoA-IM post centrifuge supernatant was adsorbed to two spots on a anion exchange (Q10) ProteinChip using 50mM Tris at pH 8.0 as solution A. One of the spots had a final wash with water and the other had no final wash and was left to dry.

Interestingly, as can be observed from Figure 5.6, a better response was observed for the ApoA-IM monomer (main 28.1kDa/e peak) without a water wash than with one although there is minimal effect on the contaminant levels. Overall, having 50mM Tris present didn't seem to have a negative impact on the response observed. This shows that a lower concentration of Tris such as 10mM should be an adequate replacement for water while still possessing the buffering capabilities desired.

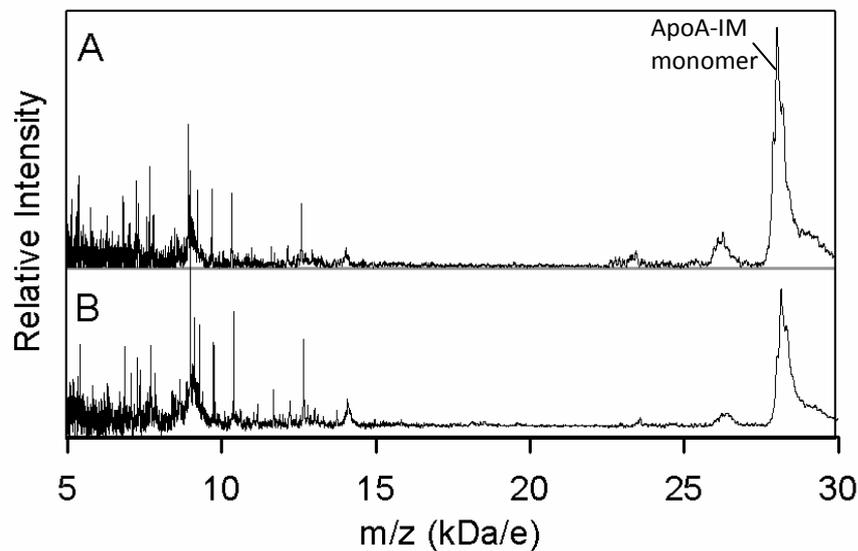


Figure 5.6 Varying the final wash step for adsorbed ApoA-IM post centrifuge supernatant proteins.

(A) 50mM Tris, pH 8.0 with no final water wash; (B) 50mM Tris, pH 8.0 with final water wash. Identical scales are used for axes with the y axis from 0 – 80 μ amps. Laser strength used was 900 & 1250nJ for low & high m/z regions respectively.

5.3.3 Final decision on which buffers to use for selective ProteinChips

After adapting the protocols in Figure 5.1 it was now necessary to decide which buffer species to use in later protein adsorption experiments. Proteins in the various samples have highly variable pI values so buffer selection was highly dependent on the sample used. Solutions were chosen so that they covered a pH range that was 2 pH units from the pI of the protein of interest depending on the surface being used. For instance, on a CM10 ProteinChip, buffers were selected that provided buffering capabilities covering at least 2 pH units below the pI and on a Q10 ProteinChip at least 2 pH units above the pI. The main aim during solution choice was to use the fewest number of buffer species to cover the largest pH range. This reduced the impact that varying solution species may have on adsorption. This led to the selection of the buffers shown in Table 5.2.

| ProteinChip | Buffer | | |
|---------------|---------|---|---|
| | pH | Equilibration species (solution A) | Final wash species (solution B) |
| CM10 & Q10 | 3.0-6.0 | 20mM Citrate, 0-1M NaCl | 5mM Citrate, 0M NaCl |
| | 6.1-6.9 | 50mM Bis-Tris, 0-1M NaCl | 10mM Bis-Tris, 0M NaCl |
| | 7.0-9.0 | 50mM Tris, 0-1M NaCl | 10mM Tris, 0M NaCl |
| H50 | 3.0-6.0 | 20mM Citrate, 0-2M (NH ₄) ₂ SO ₄ | 5mM Citrate, 0M (NH ₄) ₂ SO ₄ |
| | 6.1-6.9 | 50mM Bis-Tris, 0-2M (NH ₄) ₂ SO ₄ | 10mM Bis-Tris, 0M (NH ₄) ₂ SO ₄ |
| | 7.0-9.0 | 50mM Tris, 0-2M (NH ₄) ₂ SO ₄ | 10mM Tris, 0M (NH ₄) ₂ SO ₄ |
| IMAC10 | 7.4 | 10mM PBS, 0-0.2M imidazole | 5mM PBS, 0M imidazole |

Table 5.2 Equilibration and final wash buffers for the selective ProteinChips.

5.3.4 Selective adsorption profiles for the ApoA-IM post centrifuge supernatant proteins

The ApoA-IM post centrifuge supernatant proteins were bound to the four different ProteinChips under varying pH and salt conditions (Figure 5.7) to explore the adsorption characteristics of the proteins present.

Solution A was used for equilibration, sample incubation, and wash in a single experiment using a single bioprocessor arranged as shown in Figure 5.7. The corresponding lower molarity buffer was used for the final wash buffer (solution B, Table 5.2). For example, when 20mM sodium citrate, 0M NaCl at pH 3 was used for

equilibration, sample incubation and wash, 5mM sodium citrate, 0M NaCl at pH 3 was used as the final wash.

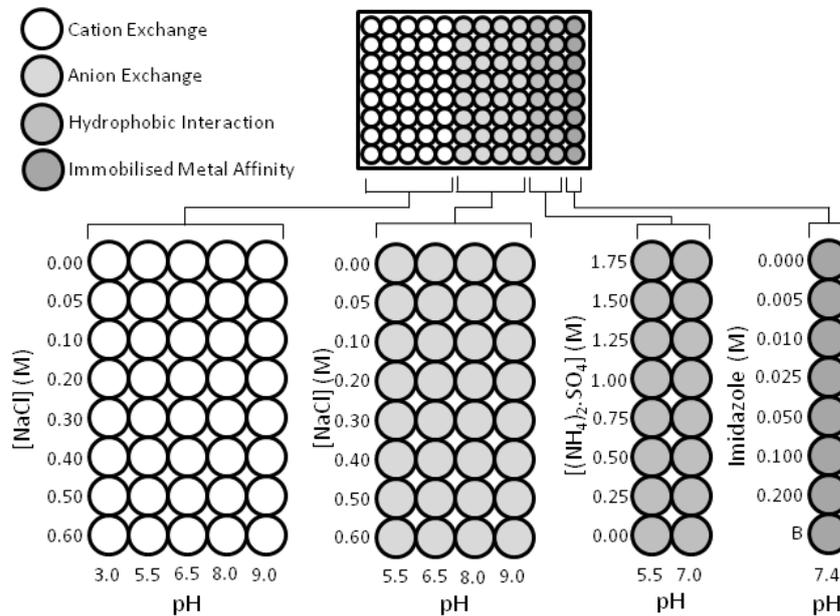


Figure 5.7 Condition array for profiling post centrifuge supernatant

Four different ProteinChip arrays, each with distinct affinity capture properties are aligned side by side. Each array has eight specific spots with defined surface chemistry allowing eight different experimental conditions. A specialised bottomless plastic 96 well reservoir is placed over these ProteinChips to isolate each of these spots. This allows separate exposure of each spot to 200 μ L of buffer. Proteins are selectively retained due to the surface chemistry and buffer used. Each condition produces a single mass spectrum that displays the proteins retained on the surface under that condition. Well B represents an unused spot.

During ProteinChip data acquisition, spots were divided into four partitions and two of these partitions were read using different laser strengths. A low laser strength (1750nJ) for the low m/z region (5-22.5kDa/e) and a high laser strength (3000nJ) for the high m/z region (22.5-60kDa/e). This resulted in two spectra for the same adsorption condition. The other two remaining partitions were not used in case acquisition needed to be repeated. The whole experiment took less than a day to complete and generated a vast array of data.

Figure 5.8 shows some of the mass spectra from this experiment at the pH that exhibited the best binding of our target protein for each surface.

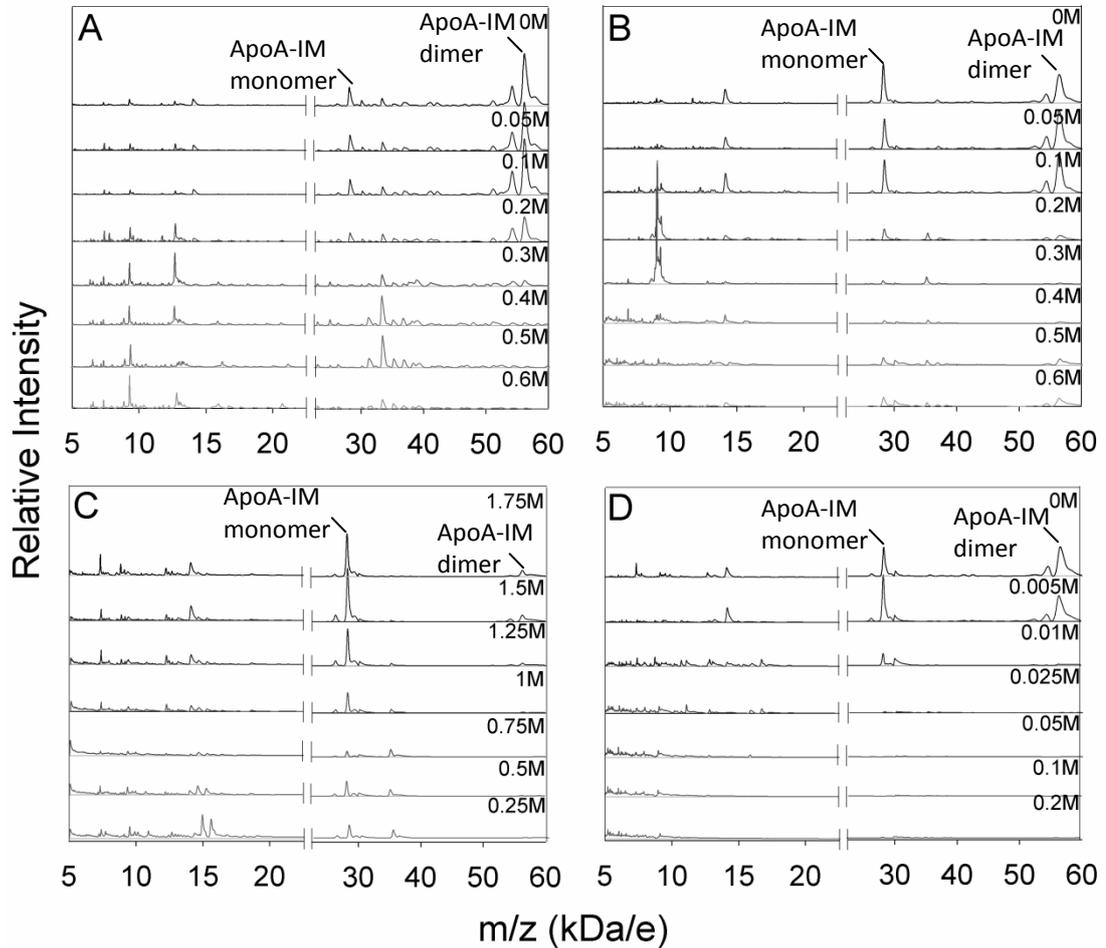


Figure 5.8 Mass spectra for adsorbed ApoA-IM post centrifuge supernatant proteins using conditions in Figure 5.7.

(A) CM10 ProteinChip at pH 3 under varying NaCl concentration; (B) Q10 ProteinChip at pH 8 under varying NaCl concentration; (C) H50 ProteinChip at pH 5.5 under varying $(\text{NH}_4)_2\text{SO}_4$ concentration; (D) IMAC10 ProteinChip at pH 7.4 under varying imidazole concentration. Laser strength of 1750 & 3000nJ was used for low and high mass respectively.

It was observed that the CM10 ProteinChip at pH 3 bound more proteins in the ApoA-IM post centrifuge supernatant than any of the other conditions used and interestingly many impurities were bound when the ApoA-IM no longer adsorbed at

0.3M NaCl. It could be argued therefore that binding at 0.1M NaCl and eluting at 0.3M NaCl could be a very good capture step on a cation exchange column at pH 3. The major downside to this is the extra sample preparation required in order to reduce the pH of solution past 4.8, the pI of ApoA-IM (section 5.2). The Q10 ProteinChip also bound many of the proteins.

The best adsorption conditions on the H50 ProteinChip were at pH 5.5 which may be due to it being close to the pI of ApoA-IM. The ApoA-IM adsorption lessens as you lower ammonium sulphate but then binds the most at 0M, probably due to “salting in” as described previously (Arakawa and Timasheff,1984). It would have been very interesting to see if a good separation was achievable on the hydrophobic ProteinChip at pH 3 which may have validated further scouting of a cation exchange and a hydrophobic interaction chromatography sequence. Of course, this would be based on assuming that the hydrophobic ProteinChip could be related to real chromatography, which is still to be determined.

5.4 Summary

Review of chapter aims:

- Finalise SELDI protocols for purification development
 - Hydrophobic ProteinChips can produce adsorption data with ammonium sulphate buffers so the use of reverse phase solvents and the consequent damage they may cause to the proteins can be avoided.
 - Water in the final wash step of ProteinChip preparation can be replaced with a low molarity buffer to maintain the same pH used during equilibration, sample incubation and washing.
 - Buffers selected that used minimal solution species while covering a large pH range in order to keep any impact from varying solution species to a minimum.
 - Buffer solutions chosen that cover at least plus and minus two pH units from the pI of the protein of interest for the Q10 and CM10 respectively.

- Use protocols to generate adsorption data for the remaining sections of this thesis.
 - Ninety three conditions were used to generate a large array of adsorption data for ApoA-IM post centrifuge supernatant proteins to be used in the following chapters.

Due to the large volume of mass spectra data, the next chapter will address the creation of a flexible platform to expedite data handling with the ultimate aim of creating unique ways to aid purification development in chapter seven.

CHAPTER 6 DATABASE FOR HANDLING PROTEIN MASS SPECTRA

6.1 Introduction

Ciphergen Express is preinstalled in a computer linked to the Bio-Rad ProteinChip Reader. The platform has various functions for improving data quality. Although excellent software with an easy to use graphical user interface, extra programming flexibility would be beneficial to organize, treat and visualize mass spectra data within a single platform for purification development.

Matlab (The Mathworks Inc., Natick, U.S.A.) was the platform of choice due to its common use in engineering research and because other groups had previously written programming code for various SELDI mass spectrometry treatment algorithms (Coombes et al. 2005).

This chapter will discuss the elements required and the resultant functions created to produce a program within the Matlab environment to handle SELDI-MS data. Once the Matlab functions were in place they could then be compared with the Ciphergen platform based on the reproducibility of data sets processed using the two routines (Figure 6.1). All Matlab code created for this chapter is located in the appendix to this thesis.

After elucidating issues associated with SELDI-MS data (chapter two) and modifying selective ProteinChip preparation to generate adsorption mass spectra (chapter five) it is now necessary to find a method to automate the organisation and treatment of the data for use in chapter seven.

Chapter aims:

- Create Matlab functions to organise data based on adsorption conditions used.
- Apply Matlab functions to carry out typical mass spectra data treatment.
- Compare Matlab with Ciphergen Express data treatment.

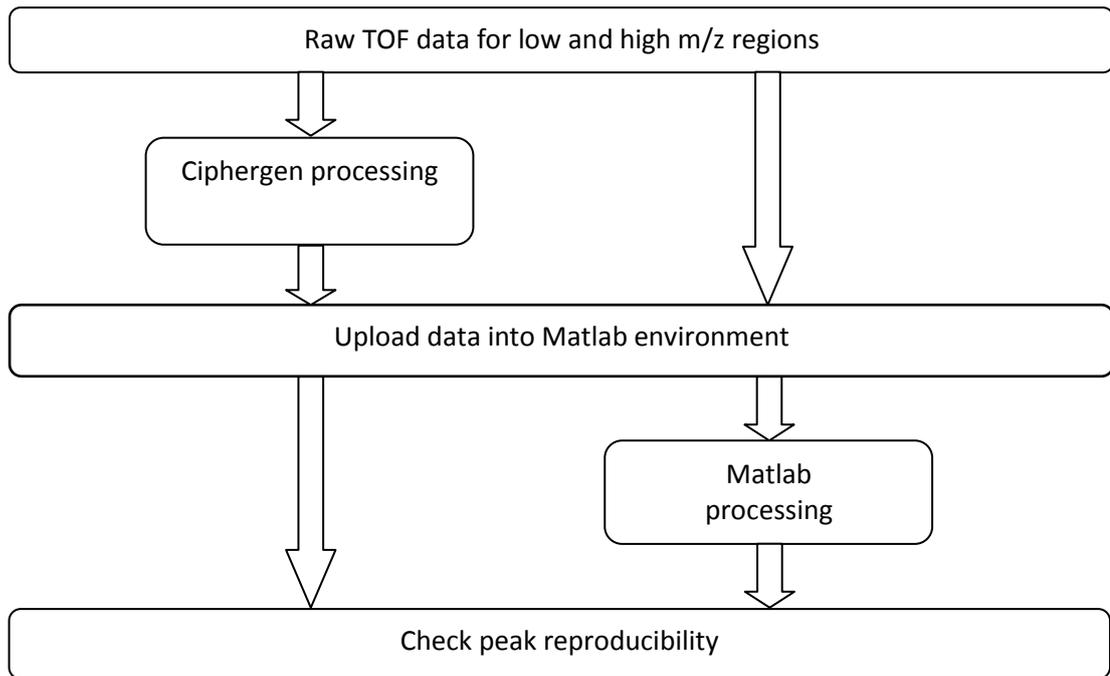


Figure 6.1 Options for mass spectra data treatment.

6.2 Materials & Methods

Unless specified otherwise the materials & methods are unchanged from section 2.2.

6.2.1 Mass spectrometry

6.2.1.1 Selective ProteinChip preparation

Two Q10 ProteinChips (16 spots in total) were equilibrated three times with 150 μ L of 50mM Tris, pH 8, 0M NaCl (solution A) for 5 minutes under shaking (8.3 rps). The spots were then loaded with 150 μ L of ApoA-IM post centrifuge supernatant previously diluted ten-fold in solution A. The sample was incubated with the spot surface for 30 minutes while being shaken (8.3 rps). Each spot was then washed three times with 150 μ L of the solution A for 5 minutes under shaking (8.3 rps) to eliminate non-adsorbed proteins. Each spot then finally had instantaneous exposure to 10mM Tris, pH 8, 0M NaCl (solution B) to reduce salts while maintaining pH. After 20 minutes the spots were dry and 1 μ L of SPA was

applied. After 5 minutes the SPA application was repeated before a final 5 minute dry step prior to ProteinChip analysis.

6.2.1.2 Data acquisition

All arrays were analyzed using a PCS4000 Reader (Figure 1.13) in a positive ion mode, with a source voltage of 25kV. The m/z range investigated was from 5 to 60kDa/e for all samples. All of the proteins of interest were within this range. Focus mass was set to the m/z of the target protein (28.1kDa) being investigated. All spots were divided into four partitions and one partition read with the same laser strength for all spots. One spot was chosen to have all four partitions read. Laser strength responsible for the desorption-ionization of proteins was 1850nJ for low and high m/z regions. Adsorption of sample proteins on arrays appeared as a signal at the appropriate molecular masses.

6.2.2 Data treatment

Ciphergen Express 3.0 & Matlab 7.5 were the two platforms used in this study. Ciphergen Express 3.0 and Matlab 7.5 were used as described in section 2.2 and 3.2 respectively.

6.3 Results

To create a Matlab programming space for use in chapter 7 the following was required; a way of easily uploading and sorting the large volume of data from the Ciphergen Express software and then carrying out batch treatment of the data.

6.3.1 Organisation & storage of mass spectrometry data

Before any data treatment could be carried out on Matlab, it was necessary to import the data into the Matlab environment while maintaining information about its preparation. Data from the ProteinChip reader can be exported from

Ciphergen Express as comma separated value (.csv) files. Ciphergen Express names the files with the name given to the spectrum by the user during data acquisition. In order to keep a record of preparation parameters in Matlab, a short code was created to name each spectrum within Ciphergen Express before exporting them (Figure 6.2). A script could then be written in Matlab to de-code how the ProteinChip was prepared from the filename of each spectrum. The code created for each spectrum is based on the minimum number of characters necessary to describe the level used for each of the factors chosen. The four factors used were ProteinChip surface, pH, salt species, and salt concentration. Since there were two sets of data for each adsorption condition (low and high m/z data) Ciphergen Express automatically appends “_1” to the filename of one of the files to make the filenames unique.

Raw data was exported and stored in a “Raw” folder in the Matlab environment (Figure 6.5). The same data processed with Ciphergen functions was exported into a “Processed” folder. Metadata was also exported with the processed data to keep a record of the acquisition parameters used.

The process of organising and storing mass spectra data is summarised in Figure 6.3.

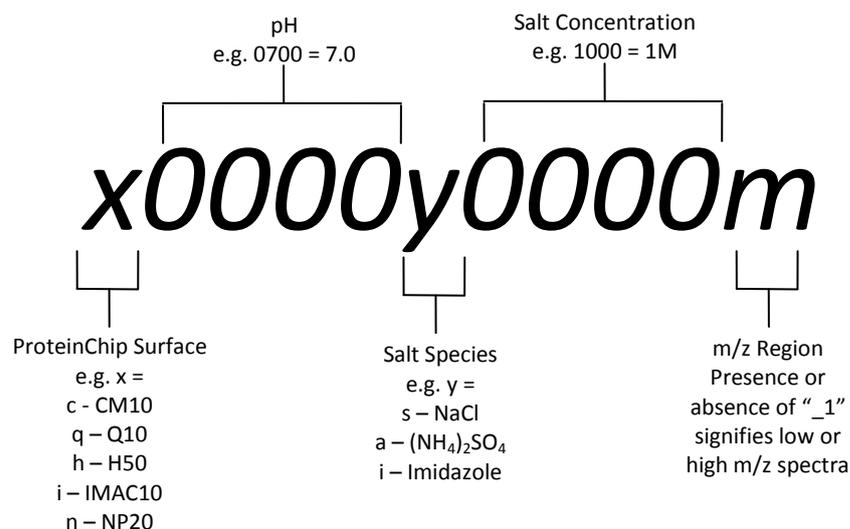


Figure 6.2 Identification code for mass spectra files.

Code created so each spectra in Matlab can be linked to the adsorption conditions used to generate them.

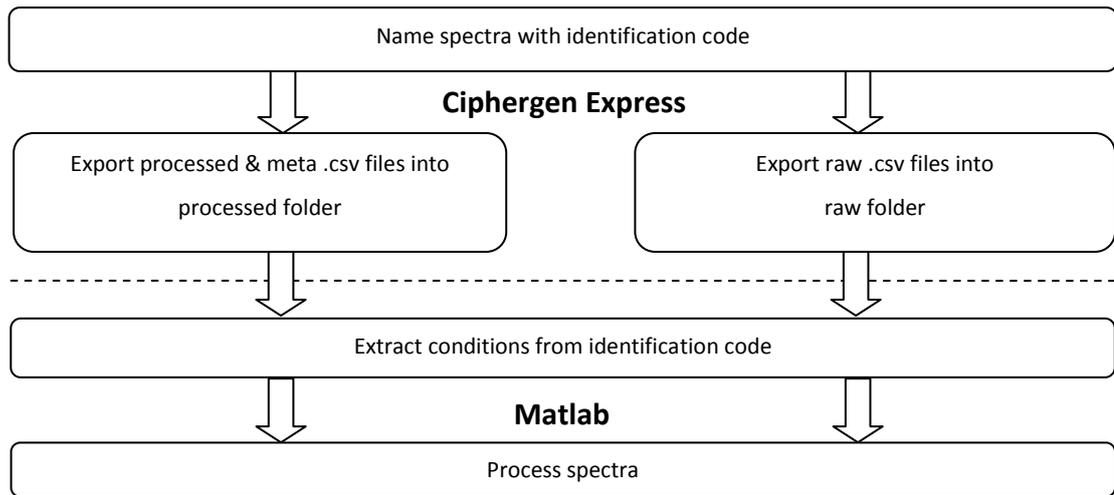


Figure 6.3 Uploading mass spectra data into the Matlab environment.

6.3.2 Program execution

A "ReadRoot" program was created that when executed, opened the previously created "Processed" folder and read each of the .csv files individually to extract the m/z , intensity and metadata information. It then collated all this data into m/z and intensity matrices plus matrices for metadata and filenames. The program also clipped the "_1" from the ends of any filenames in the filenames matrix so that the low and high m/z data now had identical names. The spectra files in the "Raw" folder were then read based on the filenames matrix to make sure all the processed data had supporting raw data for later data treatment comparisons.

6.3.3 Treatment of raw mass spectrometry data

There were various options for the treatment of SELDI-MS data, many of which have been published (Coombes et al. 2005). Typically the initial data treatment involves noise reduction and baseline correction (Figure 6.4). Different treatment routines will later be compared based on peak reproducibility.

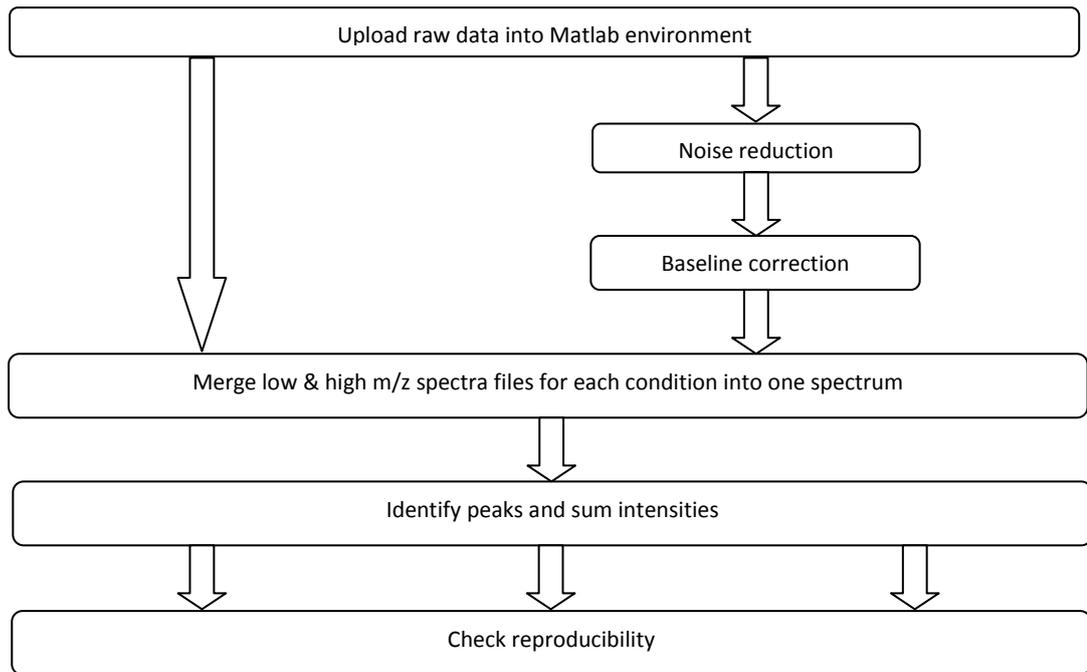


Figure 6.4 Processing mass spectra data within the Matlab environment.

6.3.3.1 Noise reduction

Usually the first processing step used on raw mass spectrometry data (Figure 6.5) is denoising. The noise is believed to be caused by fluctuations in current output (Coombes et al. 2005).

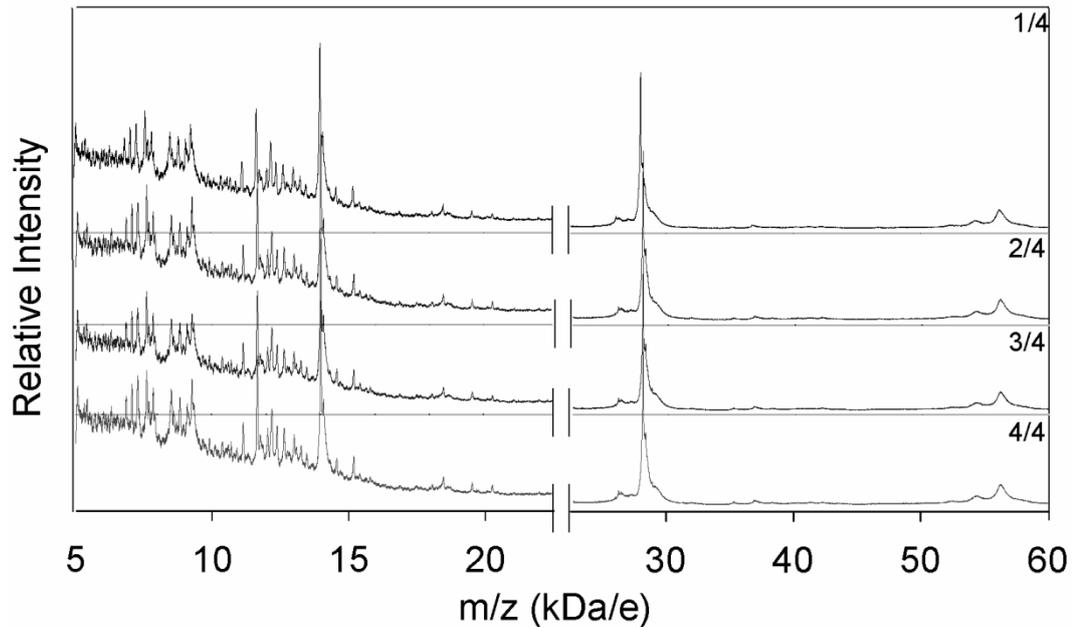


Figure 6.5 Raw mass spectra for adsorbed ApoA-IM post centrifuge supernatant proteins.

A solution of 0.05mg/mL ApoA-IM post centrifuge supernatant was bound to a Q10 ProteinChip at pH 8.0, 0M NaCl. Each spectrum is from 1 of 4 partitions. The laser strength used was 1850nJ for the low and high m/z regions.

After researching different denoising functions it was decided to use one publicized in a journal by a group at the University of Texas (Coombes et al. 2005). In their work they used an Undecimated Discrete Wavelet Transform (UDWT) that was shown to be very effective at denoising spectra. This transform uses small waves located at different times as the basis function. The transform is achieved by scaling and translation of a scaling function and wavelet function, meaning that the wavelet transform unlike other transforms (e.g. Fourier) is localised in both time and frequency.

In this work the method involved transformation of the data from the initial time domain to the wavelet domain. The variability of the wavelet coefficients is then computed based on their median absolute deviations (MAD). Any coefficients below a chosen threshold ($0.67 \times \text{MAD}$) are then set to zero. The data is then transformed back to the original time domain. Figure 6.6 & Figure 6.7 respectively

show the estimated noise and its subtraction from the raw spectra shown in Figure 6.5.

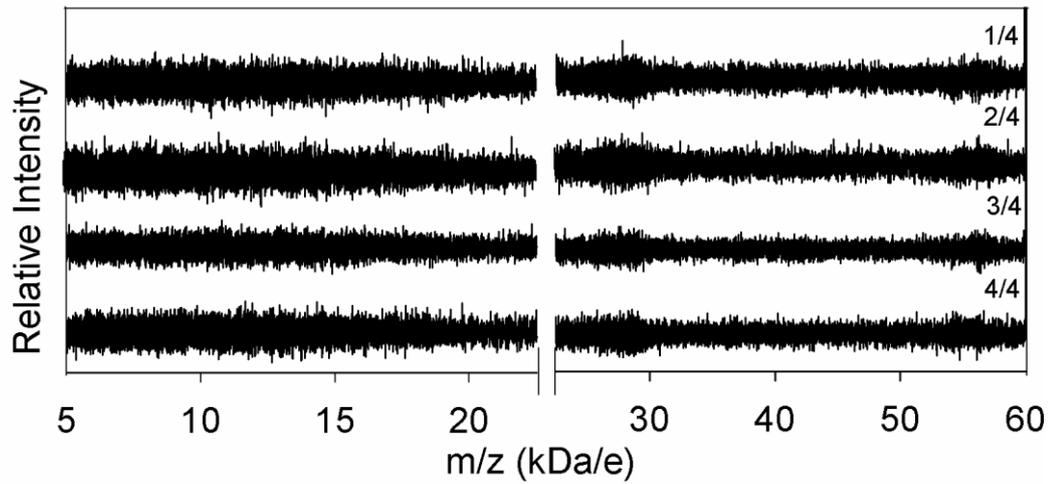


Figure 6.6 Mass spectra signal noise for adsorbed ApoA-IM post centrifuge supernatant proteins.

A solution of 0.05mg/mL ApoA-IM post centrifuge supernatant proteins were bound to a Q10 ProteinChip at pH 8.0, 0M NaCl. Each spectrum is from 1 of 4 partitions. The laser strength used was 1850nJ for the low and high m/z regions. Signal noise was calculated using an undecimated discrete wavelet transform.

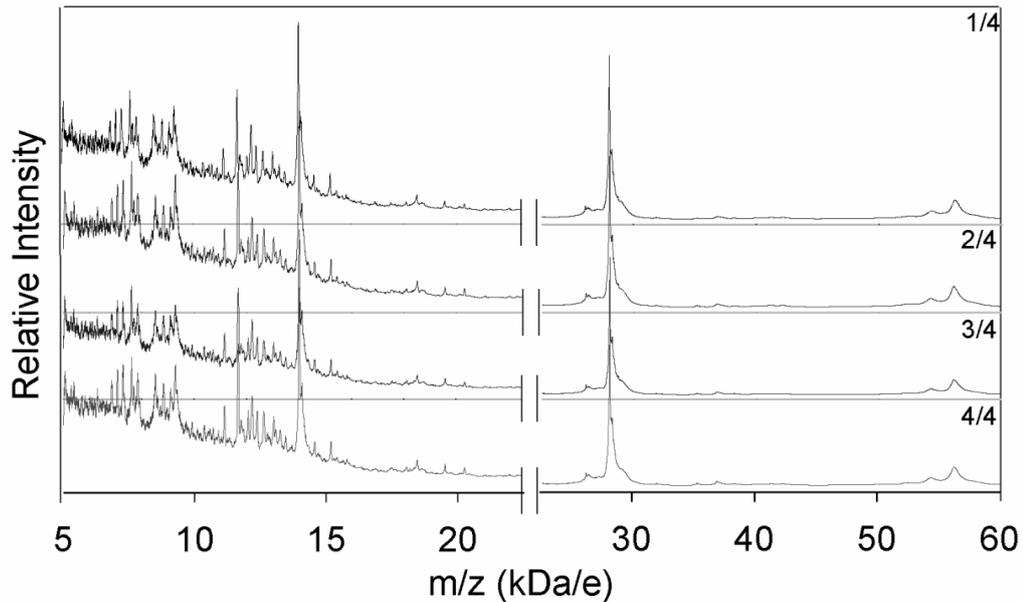


Figure 6.7 Denoised mass spectra for adsorbed ApoA-IM post centrifuge supernatant proteins.

A solution of 0.05mg/mL ApoA-IM post centrifuge supernatant was bound to a Q10 ProteinChip at pH 8.0, 0M NaCl. Each spectrum is from 1 of 4 partitions. The laser strength used was 1850nJ for the low and high m/z regions.

6.3.3.2 Baseline correction

After denoising, baseline correction is the next common step for treatment of mass spectra. The function used by Coombes *et al.* was a monotone minimum estimation. This served the purpose for their work but they did not recommend it for others to use. In this thesis, a baseline correction function (`msbackadj`) was used with default settings from the Matlab bioinformatics toolbox. This function estimates the baseline within multiple shifted windows of width 200 m/z and then regresses the varying baseline to the window points using a spline approximation. This function allowed increased flexibility due to the various methods of baseline adjustment that could be used. Figure 6.8 shows the removal of the baseline from the spectra in Figure 6.7 using the default settings for the Matlab function.

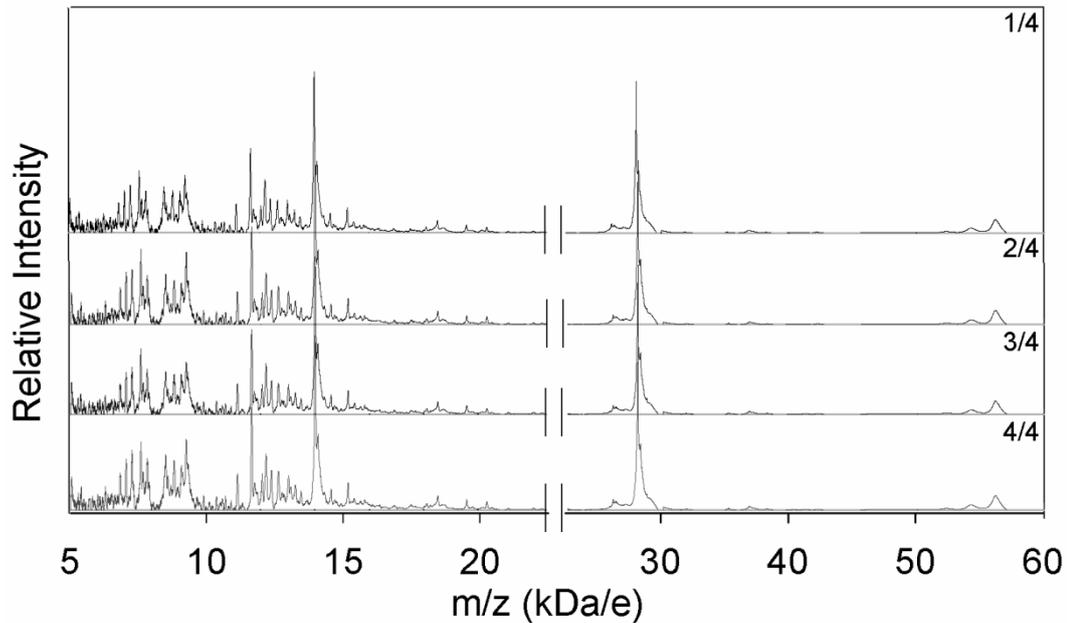


Figure 6.8 Denoised & baselined mass spectra for adsorbed ApoA-IM post centrifuge supernatant proteins.

A solution of 0.05mg/mL ApoA-IM post centrifuge supernatant was bound to a Q10 ProteinChip at pH 8.0, 0M NaCl. Each spectrum is from 1 of 4 partitions. The laser strength used was 1850nJ for the low and high m/z regions.

6.3.3.3 Splicing low and high m/z spectra data

After the raw spectra had been processed for both low and high m/z regions it would now be beneficial to splice the low and high m/z data together to get one spectrum for each condition instead of the original two. Since ‘_1’ had been removed from the ends of half of the filenames, the filenames matrix now had identical names for the low and high m/z data. This allowed the use of Matlabs “unique” function to find the pairs of filenames that were identical. Based on this the corresponding mass spectra data was then also paired. The low and high m/z data from each pair were then identified by comparing the ratio of their start and end m/z values. The data was then spliced together correctly and clipped at the extremes of m/z. As can be seen from Figure 6.8 the values chosen for these spectra were; minimum m/z of 5kDa/e, intersect between low and high m/z data of

22.5 kDa/e and maximum m/z of 60kDa/e. Now each adsorption condition only had one spectrum created from the original low and high m/z spectra.

6.3.3.4 Identifying peaks

One of the biggest hurdles in biomarker discovery is finding an algorithm that can successfully and consistently locate real peaks. In biomarker discovery many samples are investigated with relatively few different adsorption conditions, in purification development the opposite is true. The advantage of this is that peak selection only needs to be done once on conditions where many peaks bind.

After trialling many different peak selection algorithms it was decided that manual peak selection would be preferred in this work due to individual users being more inclined to select real peaks rather than artefacts of mass spectrometry. The next step in the data treatment routine involved summing peak intensities as described in section 3.2.

6.3.4 Reproducibility

6.3.4.1 Partitions within a spot

To compare the Matlab and Ciphergen Express platforms, the coefficient of variation (%CV) was calculated for thirty manually selected peaks (also used in the next chapter) in ApoA-IM post centrifuge supernatant across four out of four partitions on a Q10 ProteinChip spot. Interestingly Figure 6.9 shows that de-noising raw spectra marginally improves peak reproducibility across partitions but baseline correction carried out with Ciphergen Express and Matlab actually reduces reproducibility compared to the raw spectra. Summed peak intensity also marginally improves peak reproducibility compared to peak intensity for all treatments including the raw data.

Before final conclusions could be drawn on these different treatment methods, it was decided to see if similar observations were seen when investigating reproducibility across spots on different ProteinChips.

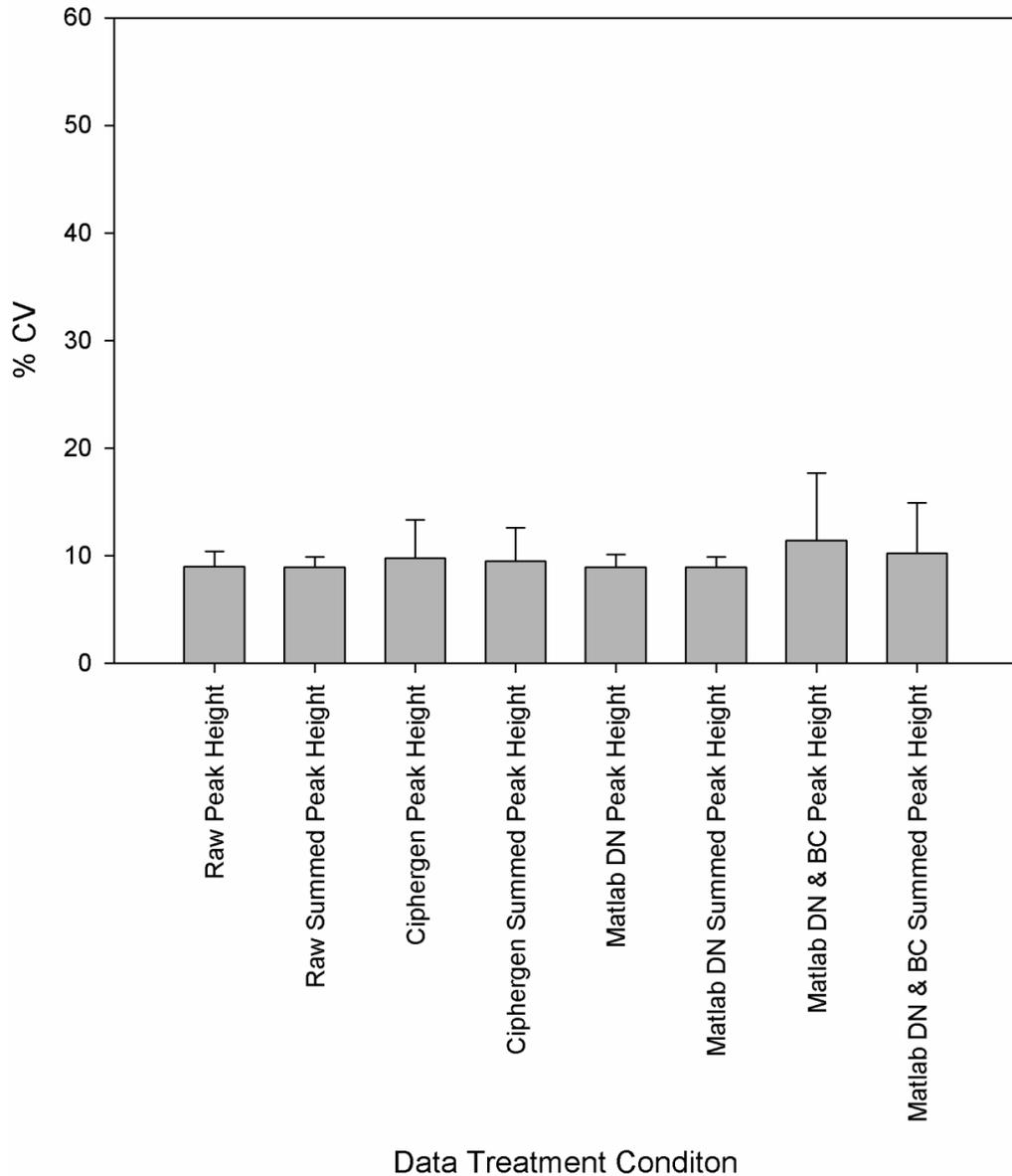


Figure 6.9 Peak reproducibility between spot partitions for adsorbed ApoA-IM post centrifuge supernatant proteins.

A solution of 0.05mg/mL ApoA-IM post centrifuge supernatant was bound to a Q10 ProteinChip at pH 8.0, 0M NaCl. Reproducibility is from 4 out of 4 partitions on one spot. The %CV was calculated for each manually selected peak across partitions. The average and standard deviation of the %CV for the thirty peaks was then plotted for each treatment method. DN and BC refer to de-noised and baseline corrected spectra respectively. The laser strength used was 1850nJ for the low and high m/z regions.

6.3.4.2 Partitions between spots

Figure 6.10 shows that the % CV is higher when examined across spots and ProteinChips than within partitions, ranging from 20 to 40% depending on the treatment method used. Other groups have also reported coefficients of variation above 30 percent due to the multi-variable nature of the SELDI-MS technique (Panicker et al. 2009; Semmes et al. 2005).

Apart from the % CV, the patterns seen between Figure 6.9 and Figure 6.10 are similar. Figure 6.10 shows that denoising raw spectra marginally improves peak reproducibility across spots & ProteinChips but baseline correction carried out with Ciphergen Express and Matlab show a proportionally large deterioration in reproducibility compared to the raw and de-noised spectra. Summed peak intensity also marginally improves peak reproducibility compared to peak intensity for all treatments including the raw data.

In the next chapter the method used for purification development will therefore de-noise the raw data but no baseline correction will be carried out. Also, summed peak intensity will be used. When deciding on data visualisation techniques in the next chapter, the approach chosen will incorporate a method to improve reproducibility further.

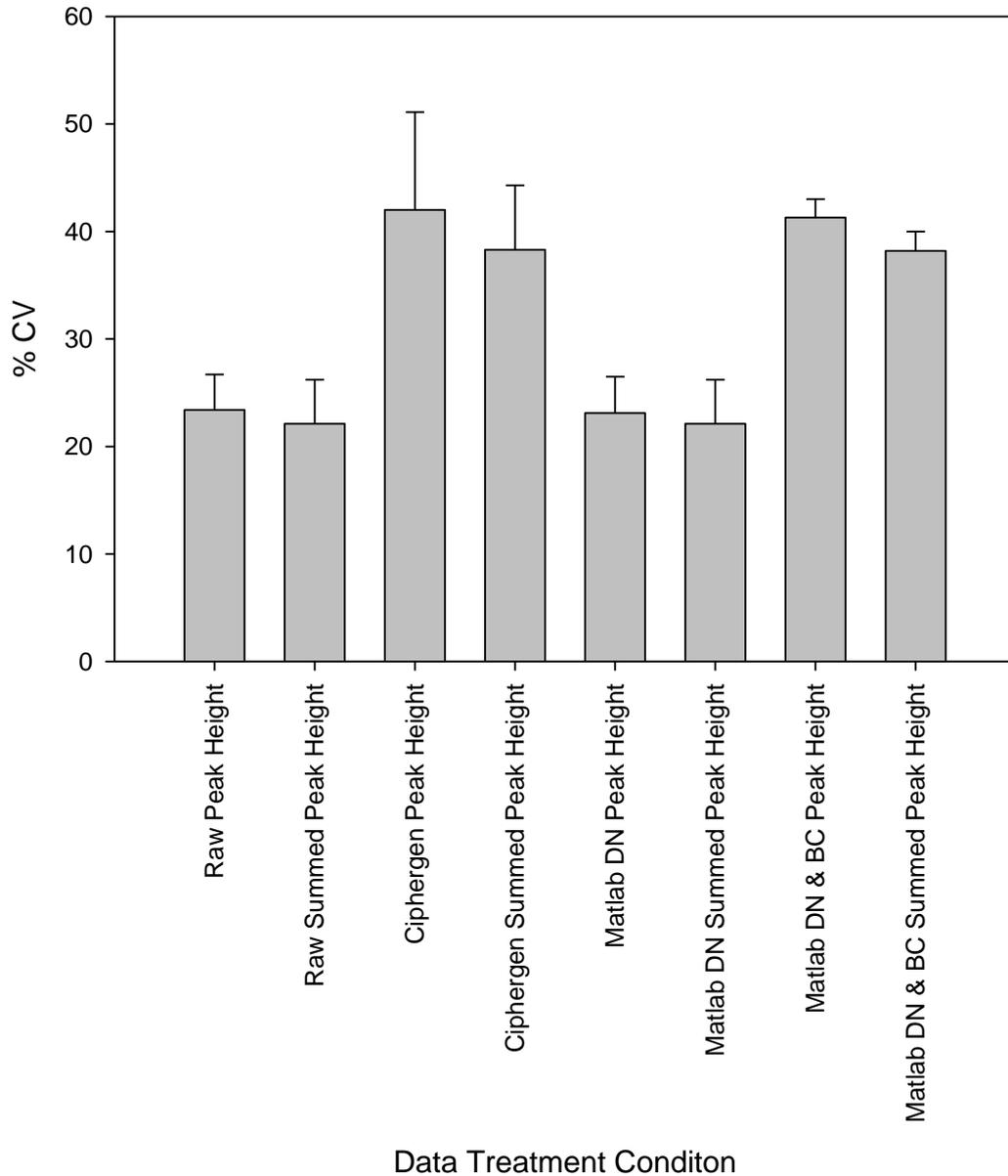


Figure 6.10 Peak reproducibility between ProteinChips for adsorbed ApoA-IM post centrifuge supernatant proteins.

A solution of 0.05mg/mL ApoA-IM post centrifuge supernatant was bound to a Q10 ProteinChip at pH 8.0, 0M NaCl. Reproducibility is from 1 out of 4 partitions on sixteen spots (two ProteinChips). The %CV was calculated for each manually selected peak across partitions. The average and standard deviation of the %CV for all thirty peaks was then plotted for each treatment method. DN and BC refer to de-noised and baseline corrected spectra respectively. The laser strength used was 1850nJ for the low and high m/z regions.

6.4 Summary

Review of chapter aims:

- Create Matlab functions to organise data based on adsorption conditions used.
 - Matlab functions created that can upload and interpret spectra files in order to sort the spectra data based on adsorption conditions used.
 - Functions allow the retention of meta data.
 - Spectra files for low and high m/z are spliced together to produce one spectrum for each adsorption condition.

- Apply Matlab functions to carry out typical mass spectra data treatment.
 - Functions from the literature and the Matlab bioinformatics toolbox have been used to treat the data within the Matlab framework and those successful for improving reproducibility retained for use in the next chapter.
 - Baseline correction showed a drop in reproducibility hence Matlab will now only be used to treat the data using denoising and summed peak intensity calculations.
 - To reduce the variability of the mass spectrometry data further, the approach chosen in the next chapter for purification development will incorporate a method that uses peak normalisation.

- Compare Matlab with Ciphergen Express data treatment.
 - Baseline corrected data from Matlab and Ciphergen Express give similarly large coefficient of variations.
 - The Ciphergen software was found to be good for quick, initial screening of data so that decisions could be made on which datasets to export to Matlab.
 - Matlab will be used in the next chapter because it provides the freedom to try many different data visualisation techniques.

CHAPTER 7 MASS SPECTROMETRY TO DESCRIBE PRODUCT AND CONTAMINANT RELATIONSHIPS FOR BIOPROCESS DEVELOPMENT

7.1 Introduction

Traditionally, biopharmaceutical process development starts without full characterization of the material produced from fermentation. It would be particularly advantageous to determine the proteins that have similar physicochemical properties to the product prior to purification. In upstream operation, cell lines and the conditions they are grown under could be changed or the host's genetic information modified to alter the properties of protein contaminants similar to the product, relieving some of the burden on later purification stages (Humphreys et al. 2004). For the downstream processing sequence, the physicochemical information of near neighbour contaminants could be used to select the best purification strategy.

Current methods for full physicochemical property characterisation are very time consuming, labour intensive and carry with them many unfavourable assumptions and could fall on the process development critical path, something which will always be avoided. This lack of a rapid method for full characterisation leads to reduced flexibility in the process due to laborious, time consuming experimentation with every process change.

Conditions for the removal of cellular debris have been found previously in chapter four to avoid fouling of the chromatography column. The focus of this chapter is to use SELDI-MS adsorption data generated during chapter five and treated in chapter six for detecting retained compounds with shared physicochemical properties in order to rapidly characterize feed materials (Merchant and Weinberger,2000). Patterns in the differential adsorption of proteins to ProteinChip surfaces have been investigated as a route to address this. Similarities between proteins are then deduced and hence impurity proteins that are hardest to separate from the product identified.

Summary of chapter aims:

- Incorporate peak normalisation, where peaks are compared against one another within a spectrum to reduce variability.
- Develop a physicochemical map to compare proteins within a given process material.
- Determine nearest neighbours to target protein.
- Commence investigation into the conditions that may be promising for separating the target protein from nearest neighbours.

7.2 Materials & Methods

Unless specified otherwise the materials & methods are unchanged from section 5.2.

7.2.1 Protein samples

A stock solution of 2mg/mL ovalbumin in 50mM Tris, pH 7 was created. A hen egg white (HEW) solution was made following protocol in section 2.2. The process materials used in this study came from an *E. coli* process generating ApoA-IM (Figure 3.1) and was provided by the Pfizer Global Biologics group (Chesterfield, MO, U.S.A.).

7.2.2 Mass spectrometry

7.2.2.1 Data treatment

SELDI-MS spectra were analysed using functions written within Matlab using settings as defined in section 3.2 (Coombes et al. 2005). Peaks were manually selected based on a signal/noise ratio > 5 and from prior SELDI-MS analysis of the samples. SELDI-MS results were recorded by summing signal intensities between +/- 0.3% of peak m/z. The value of 0.3% was decided from previous work, discussions with Bio-Rad Inc. (Hercules, CA, U.S.A.) and its frequent use in the

literature as a parameter for SELDI-MS peak detection (Brozkova et al. 2008; Panicker et al. 2009). Baseline correction was not used for star plots, dendrograms or adsorption maps due to the treatment incurring decreased peak signal reproducibility as shown in the previous chapter.

7.2.3 Chromatography

Chromatography was performed with a HiTrap column (7mm i.d. x 250mm height, 1mL) on an AKTA Explorer 100 system. These were supplied by GE Healthcare (Uppsala, Sweden).

The column was delivered pre-packed with a Q Sepharose fast flow sorbent. The buffer condition used was; equilibration, wash with 50mM Tris, pH 8, 100mM NaCl and elution with 50mM Tris, pH 8, 400mM NaCl.

Chromatography separations were accomplished at a flowrate of 1mL/min equivalent to a linear velocity of 0.04cm/s. Absorbance of column effluents was recorded at 280nm.

The column was equilibrated with 10 CV (column volumes) of equilibration buffer. The sample was diluted to 1 mg/mL and adjusted to column equilibration conditions before direct loading of 10mL onto the column. The non-adsorbed proteins were washed out with 5 CV of equilibration buffer. The column was eluted with 10 CV of elution buffer before being cleaned with 10 CV of equilibration buffer with 2M NaCl. The column was stored in 1M sodium hydroxide.

Fractions of 1mL each were collected throughout and diluted 1in100 with 50mM Tris, pH 8.0, 0mM NaCl before being analysed directly on an anion exchange (Q10) ProteinChip as described in section 6.2. This buffer was chosen because it showed favourable binding of many proteins in the ApoA-IM post centrifuge supernatant including ApoA-IM and its nearest neighbour impurities. Fractions were also analysed using RP-HPLC following the protocol in section 3.2 to confirm that Q10 ProteinChips are able to produce quantitative data.

7.2.4 Data visualisation techniques

7.2.4.1 Star plots

Star plots (Figure 7.1) consist of equiangular spokes around an origin. Each spoke represents a variable and its length represents a normalised response. Normalisation is based on the maximum response observed for that variable across all the data. The data values for each spoke are connected by a line.

Image removed
due to copyright

Figure 7.1 A star plot example from The Mathworks Inc. (Natick, MA, U.S.A.)

This example shows how different cars can be compared, with each spoke representing a different property e.g. engine size.

7.2.4.2 Dendrograms

There are two main steps before dendrogram creation; calculating a distance between observations and then linking these observations by clustering. There are many distance metrics available such as Mahalanobis, Manhattan and Chebyshev but the most commonly used distance metric and the one used for this work was Euclidean distance (Bonet et al. 2008). This was used to measure differences in peak intensities over the range of adsorption conditions used. For three identified peaks, A, B & C we have (Equation 7.1);

$$\overline{AB} = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$$

$$\overline{AC} = \sqrt{\sum_{i=1}^n (a_i - c_i)^2}$$

$$\overline{BC} = \sqrt{\sum_{i=1}^n (b_i - c_i)^2}$$

Equation 7.1 Euclidean distance

where a , b and c are peak intensities for peak masses A , B and C respectively. Subscript i refers to individual adsorption conditions; and the number of conditions studied in this case was 93. \overline{AB} , \overline{AC} and \overline{BC} are the Euclidean distances between peak masses A , B and C .

For the average linkage method, otherwise known as the unweighted pair group method with arithmetic mean (UPGMA), the two proteins with the smallest value from Equation 7.1, for instance A & B , are joined at a node connected by two branches to start the formation of a dendrogram. The combined height of both branches represents the differences in peak intensities between them. The difference in peak intensities between the new cluster (AB) and another peak mass (C) is then an average of the difference in peak intensities of A from C and of B from C (Equation 7.2).

$$(\overline{AB})C = \frac{1}{2}(\overline{AC} + \overline{BC})$$

Equation 7.2 Calculating the new distance from cluster AB to peak mass C using UPGMA.

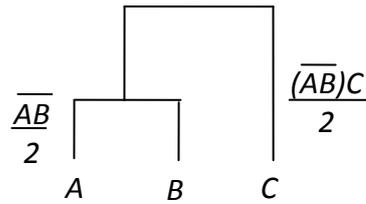


Figure 7.2 Dendrogram formation

Once the new distances are calculated the next nearest observations would be joined (Figure 7.2). In this simplified example, this could only be C since it is the only peak mass left but in reality it could be any of the peak masses chosen at the start of the algorithm. The algorithm would continue until a dendrogram is formed.

7.3 Results

7.3.1 Measurement of contaminants during purification

Figure 7.3 shows chromatography samples analysed from the ApoA-IM process. It reveals that two chromatography steps are necessary to remove the 26.2 and 27.2kDa/e proteins. All the other product related impurities (deduced from their m/z values) were not removed even after three chromatography steps although it could be argued that dimers at 54.6 and 57.5kDa/e may no longer be present after two steps. The most distantly related impurities in mass terms are removed after one chromatography step, particularly proteins less than 14.6kDa/e. The hardest proteins to separate from the product are the 28.6 and 29.1kDa/e proteins which are shown to remain with our product even after three chromatography steps. The peak at ~ 14 kDa/e could be a double charged form of the product. The analysis reflects what is seen in many other separations e.g. monoclonal antibodies, where product related forms are the most challenging to separate.

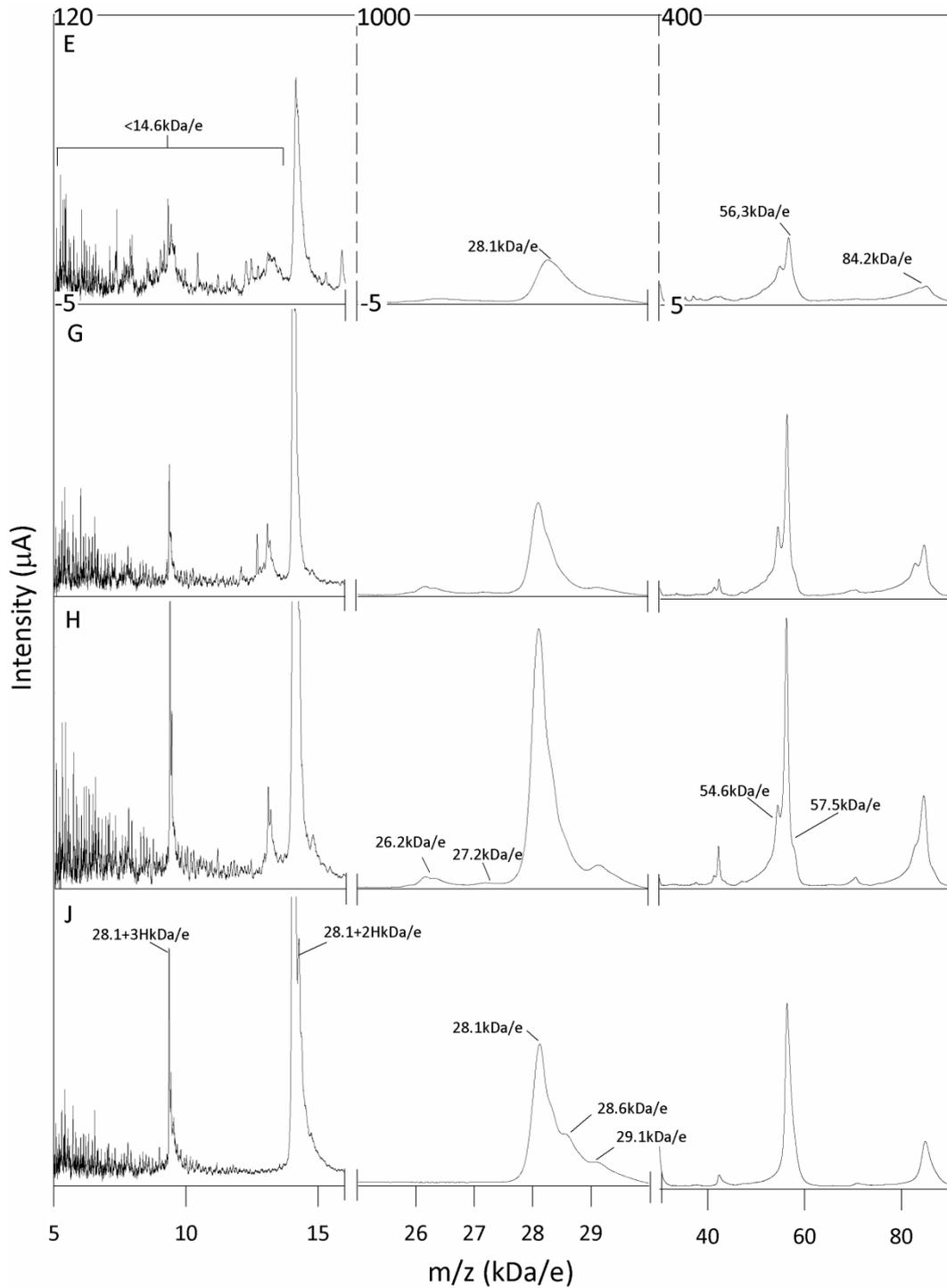


Figure 7.3 Mass spectra for process chromatography samples.

As in Figure 3.3; (E) post centrifuge supernatant; (G) Capture pool; (H) IEX pool; (J) UF/DF 2, after HIC. Analysis was carried out on an NP20 ProteinChip. The low m/z (1-30 kDa/e) and high m/z (5-90 kDa/e) regions were acquired at 1750 and 3000nJ respectively. The y-axis scale used for each mass to charge range is shown in the top spectra. The same scale is used for all samples.

The hypothesis was to investigate the possibility of predicting the observed separation difficulty of the feed material components based upon adsorption behaviour (physicochemical properties) using SELDI-MS. The ability to determine the extent of the separation challenge presented by near neighbour proteins for any given process feed materials could be used to determine manufacturability e.g. based on the cell line or upstream processing strategy.

7.3.2 Generating basic physicochemical properties using mass spectrometry

A physicochemical property of proteins is the isoelectric point (pI). The theoretical pI of a protein is the pH at which the net electrical charge of all the constituent amino acids is zero. The product protein sequence will most likely be known but this is not necessarily the case for contaminants. For proteins with unknown amino acid sequences an empirical method such as two-dimensional electrophoresis is necessary to find their pI values. A downside to this approach however is the amount of time necessary for optimising the method for each new sample.

Using the isoelectric point to find conditions to separate a protein of interest from contaminants also has its drawbacks. It is usually assumed that when the net charge of a protein is zero it will not adsorb to an ion-exchange column. However, the isoelectric point gives no information on charge distributions and there may be localised patches of charge on its surface allowing adsorption to occur.

A direct method is needed to find conditions where proteins adsorb to a sorbent. It was considered that SELDI-MS could be used to get this information by using cation and anion exchange ProteinChips. Figure 7.4 and Figure 7.5 respectively show how this can be done using a single protein and a complex mixture as examples. As can be seen, there is some correlation between the neutral adsorption point and the isoelectric points of the proteins in HEW, with the rank order of the adsorption point pH following the rank order of the isoelectric points for the proteins (Table 1.7).

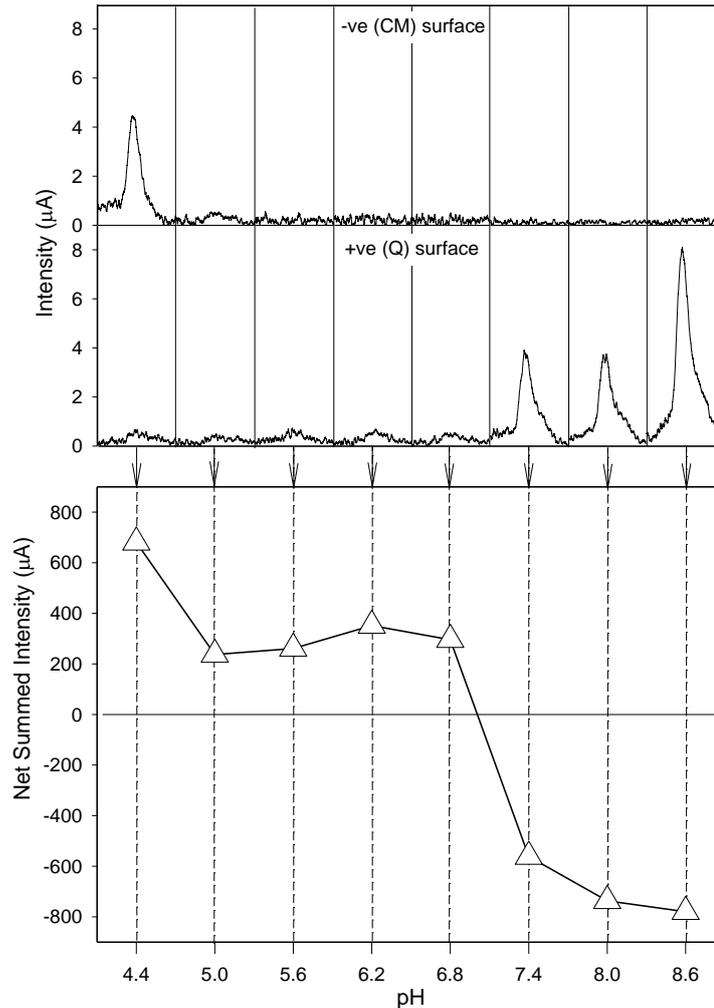


Figure 7.4 Non-competitive adsorption of ovalbumin to electrostatically charged surfaces.

A solution of 2mg/mL ovalbumin in 50mM Tris, pH 7 was diluted ten-fold in buffers with pH from 4.4 to 8.6 pH (Table 5.2) before incubation with: a cation exchange (CM10) ProteinChip (top spectrum) & an anion exchange (Q10) ProteinChip (bottom spectrum). Subtracting the summed peak intensity from the two spectra gives an approximation of the neutral adsorption point of a protein. This can be carried out for any protein present. The pH at which the neutral adsorption point is equal to zero can sometimes be similar to the pI but as this figure shows this is not always the case. A laser strength of 1000nJ was used during acquisition and the spectra were processed using Matlab as described in section 7.2. For ease of visualisation, baseline correction was applied to the spectra in the top half of the figure but this was not used when calculating net summed intensities.

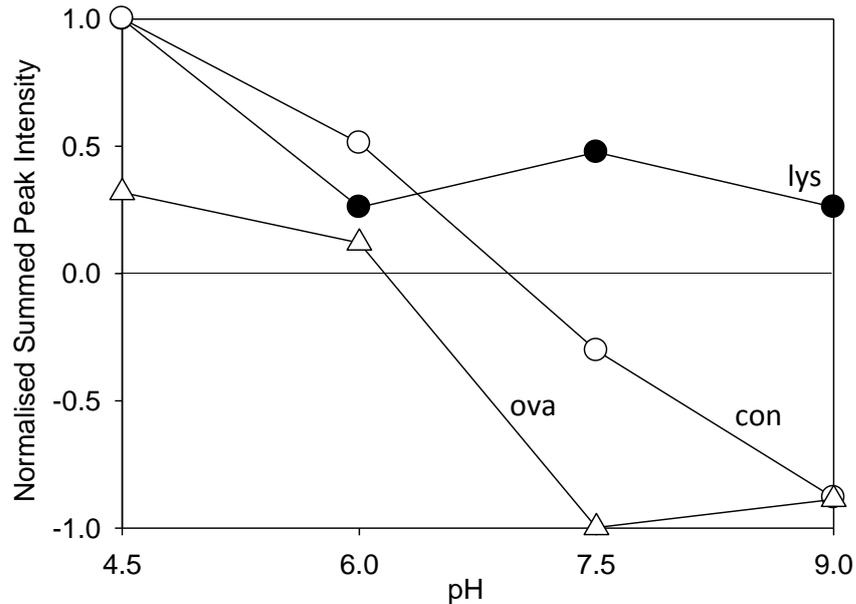


Figure 7.5 Competitive adsorption of hen egg white proteins to electrostatically charged surfaces.

Adsorption of lysozyme, lys (●), ovalbumin, ova (Δ) and conalbumin, con (○) onto cation- exchange (CM10) & anion-exchange (Q10) ProteinChips at pH 4.5 to 9. A laser strength of 1000nJ was used during acquisition and the spectra were processed using Matlab as described in section 7.2.

7.3.3 Visualising nearest neighbour contaminants using mass spectrometry data

To find more physicochemical properties, the next step involved incubating the feed material to four surface chemistries representative of those applied in process chromatography using an array of buffers of varying pH and salt concentration. The level of adsorption of each protein for each of these conditions is then detected by laser desorption/ionisation mass spectrometry.

Star plots were used to represent the differences between proteins based on their binding to these surfaces under the various conditions investigated. The star plots were created by observing the binding of each protein under all the conditions chosen (Figure 5.7) and normalising against the maximum response observed for that protein. As well as producing star plots that are readily

comparable to one another, this normalisation also improved reproducibility. Plotting the data equiangular around a single origin produced a star plot for each protein. These plots were created in Matlab using the “glyphplot” function. There were 93 conditions analysed (Figure 5.7), which meant peak intensities plotted on the glyphplot every 3.87 degrees (360 degrees/93).

Figure 7.6 shows how the star plots were plotted including how the conditions were placed around the origin. Star plots were created for thirty proteins selected in the post centrifuge supernatant (Figure 7.6). The star plots highlight that proteins are shown to be distinct while a significant portion are very similar to one another.

The next step was to create a method to quantify and visualise these relationships. This was achieved by creating dendrograms. Dendrograms are commonly used to visualise large amounts of multi-dimensional data and are commonly used to cluster genes in computational biology. This technique allows visual comparison between the data represented in Figure 7.6. Star plots were compared to produce a dendrogram showing the relationship between their adsorption characteristics. Those proteins linked by small branches are closely related and those with long branches between them are most likely distinct proteins with little if any similarities.

Three distinct clusters are formed (Figure 7.7). Cluster 1, are proteins distantly related to the product (28.1kDa/e). Based on their m/z values this group of proteins doesn't seem to include any product related species and with the exception of the 35.2kDa/e protein are all fairly small (<14.6kDa/e). Cluster 2, are a group of closely related proteins. This group includes the product. All of these proteins have m/z values that suggest they may be product related. These proteins could possibly be truncations (27.2, 28.6 and 29.1kDa/e), dimers (54.6, 56.3, 57.5kDa/e), and a trimer form (84.2kDa/e). The final major group is cluster 3. Apart from the 26.2kDa/e protein none of these proteins were believed to be product related. Within this group are a sub-group of proteins that have m/z values from 40.1 to 48.6kDa/e. These proteins have near identical physicochemical properties.

The method identifies the near neighbour contaminants hardest to separate from the product. In this case these impurities are shown to be product related proteins, particularly the 29.1 and 28.6kDa/e proteins. This correlates well with the observations of the pilot scale downstream processing (Figure 7.3).

This technique also allows the possibility of weighting impurities prior to dendrogram formation to bias the removal of some impurities over others. For example, this could be particularly useful if there was a desire to remove particular proteases early that maybe known to be damaging to the product protein.

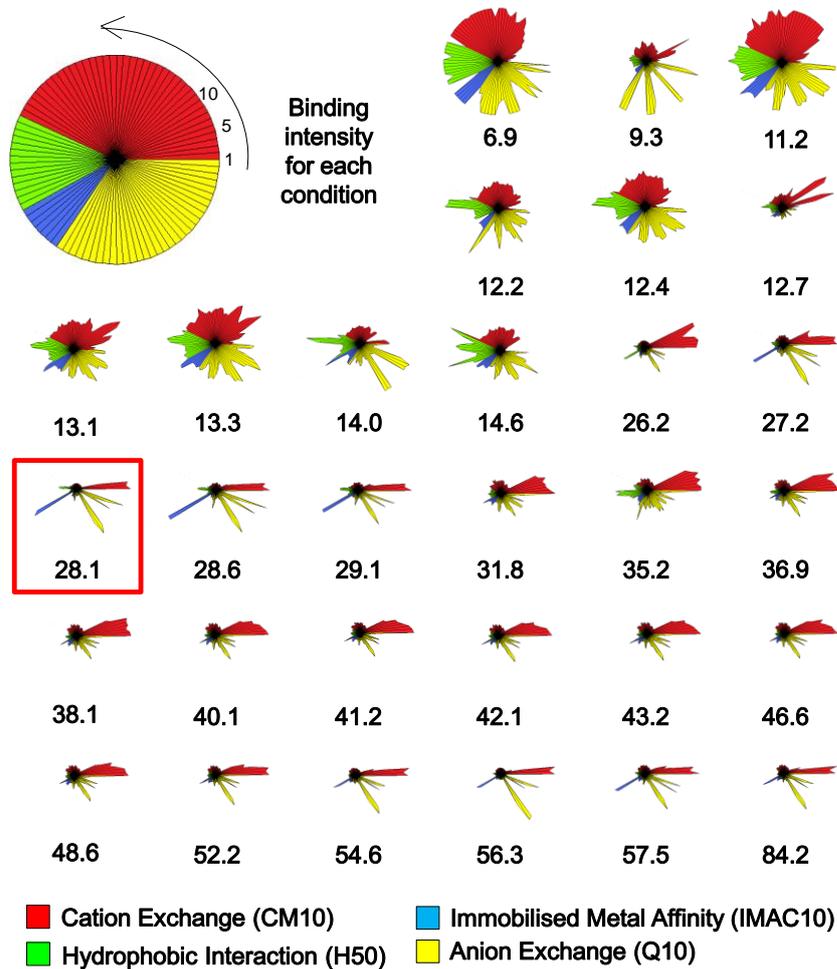


Figure 7.6 Star plots for thirty proteins contained within post centrifuge supernatant.

Each spoke of a star represents one of ninety three different binding conditions shown in Figure 5.7. Proteins are identified based on their m/z ratio shown

underneath each star. The star plot highlighted with a red box is the product at 28.1 kDa/e. Stars give a qualitative indication of how proteins are related in terms of physicochemical properties.

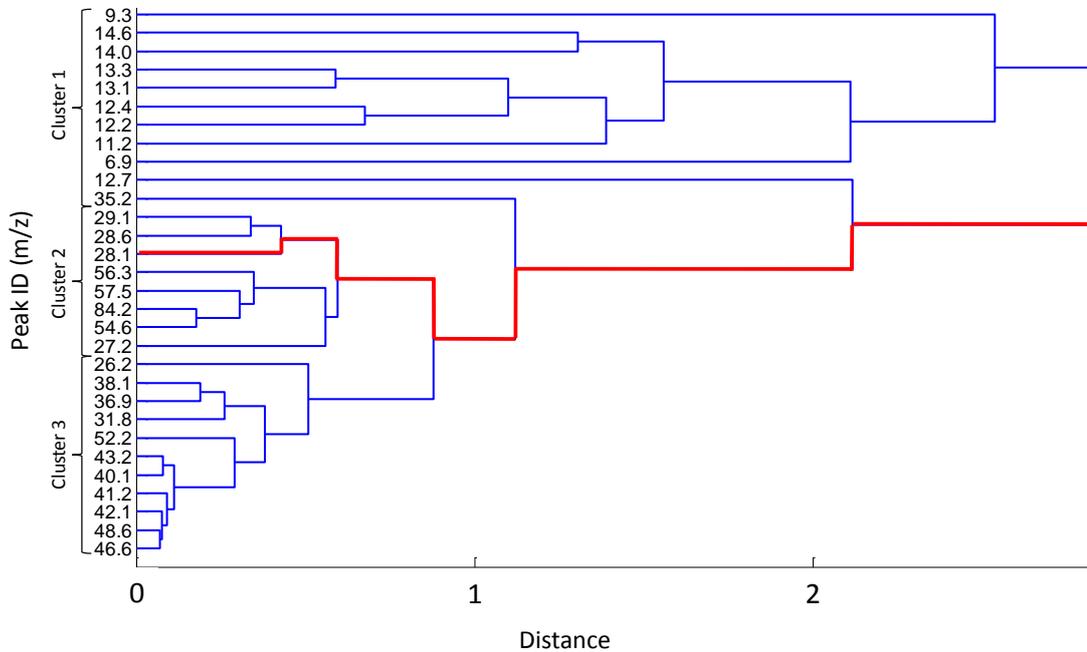


Figure 7.7 Dendrogram to show relationships between post centrifuge supernatant proteins.

The dendrogram is based on adsorption characteristics from SELDI-MS and shows how major proteins in the post centrifuge supernatant are related. Proteins with relatively short branch lengths between them are closely related and those with relatively long branch lengths are distantly related. The product (28.1kDa/e protein) is indicated by red branches. The dendrogram was created using Euclidean distance and the average linkage method (UPGMA) for clustering.

7.3.4 Identifying conditions for separating nearest neighbour contaminants using mass spectrometry

After predicting which proteins are problematic to remove (product related species) it is necessary to find ways to separate them from the product. As mentioned previously, it is unlikely that ProteinChips fully represent the subtleties of surface interactions for the wide range of chromatography resin available but

they should however give insight into where to start scouting experiments in order to find conditions to separate the nearest neighbour contaminant from the product.

Figure 7.8 shows adsorption maps for the product, a closely related protein (29.1kDa/e) and distantly related protein (12.7kDa/e) on the different ProteinChips under varying pH (x-axis) and salt concentration (y-axis). Red regions indicate high protein adsorption and blue regions indicate low protein adsorption. As expected, Figure 7.8 shows that the product has very different adsorption characteristics to the distantly related contaminants and displays very similar behaviour on both cation exchange and hydrophobic surfaces.

It was found that there were some slight differences between adsorption on the anion exchange ProteinChip particularly near the edges of the binding region for the product and nearest neighbours while being in the low adsorption region for the distantly related proteins (Figure 7.8). These were conditions at high pH (≥ 8) and with the addition of sodium chloride ($\geq 0.1M$).

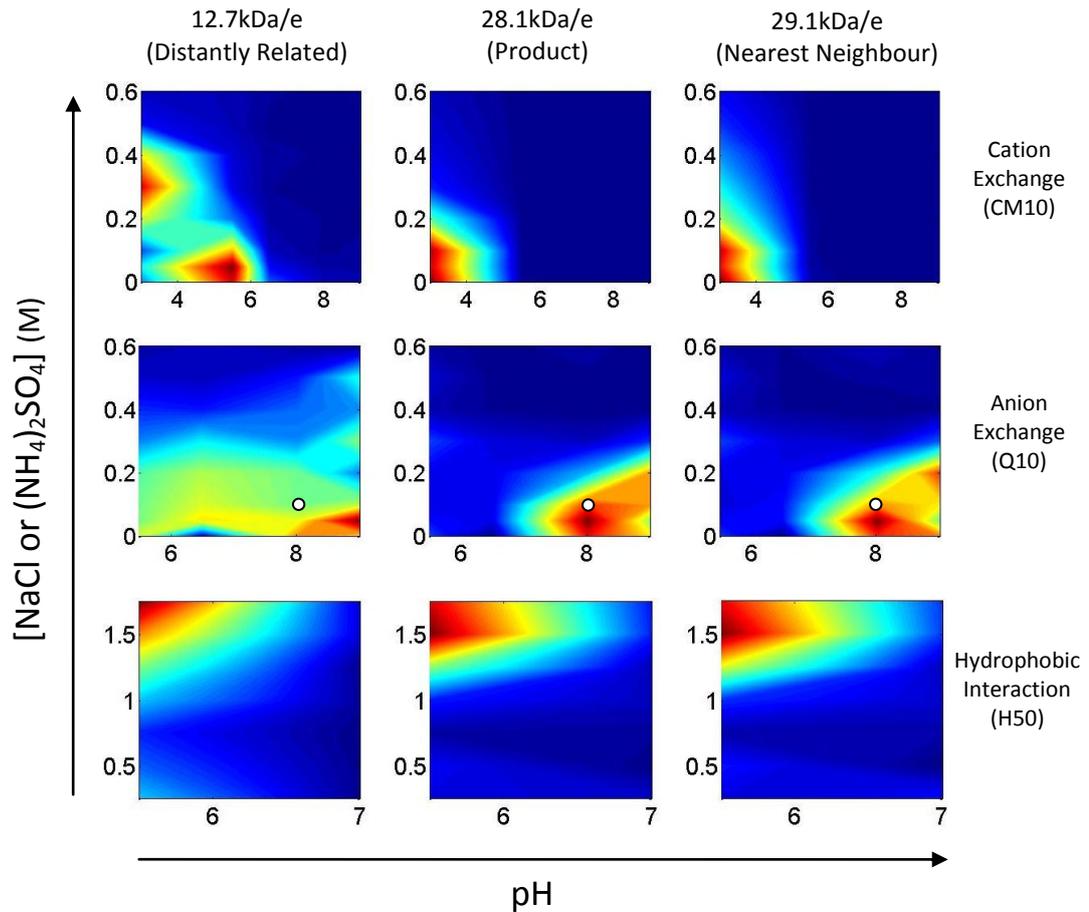


Figure 7.8 Selectivity maps for product and nearest neighbour contaminants.

Post centrifuge supernatant was incubated with a cation (CM10), anion (Q10) & hydrophobic (H50) ProteinChip using the conditions Figure 5.7. Red regions correspond to high protein adsorption while blue corresponds to low protein adsorption. A laser strength of 1750nJ & 3000nJ was used respectively for the low and high m/z regions. The white circle indicates the selected pH 8, 100mM NaCl condition used for the equilibrium, load and wash for the Q Sepharose FF chromatogram in Figure 7.9.

7.3.5 Nearest neighbour removal analysis

The chromatography condition chosen to test the separation on a column was a Q Sepharose FF resin with 50mM Tris, pH 8, 100mM NaCl buffer for equilibration, loading and washing (indicated by white circle in Figure 7.8). Figure 7.9 shows the chromatogram generated for the Q column using 50mM Tris, pH 8,

100mM NaCl buffer for equilibration, loading and washing. Elution fractions from this chromatogram were pooled and analysed for product using Q10 ProteinChips with 50mM Tris, pH 8.0 and RP-HPLC to confirm results. The anion exchange condition was selected to provide good adsorption of product related species to improve analytical sensitivity. The representative impurities selected were: (a) very distantly related to the product (9.3 and 12.7kDa/e proteins), (b) a protein that was moderately related to the product (35.2kDa/e) and (c) some nearest neighbours to the product (28.6, 29.1, and 56.3).

Figure 7.9 shows that conditions chosen seemed to separate the product from distantly related proteins (9.3 and 12.7kDa/e) but not the 35.2kDa/e proteins or the nearest neighbour contaminants (28.6, 29.1, and 56.3). This illustrates the difficulty of separating such closely related proteins in a single step, a fact borne out in the previously developed process (Figure 7.3) which requires four chromatography steps (Berrill et al. 2009).

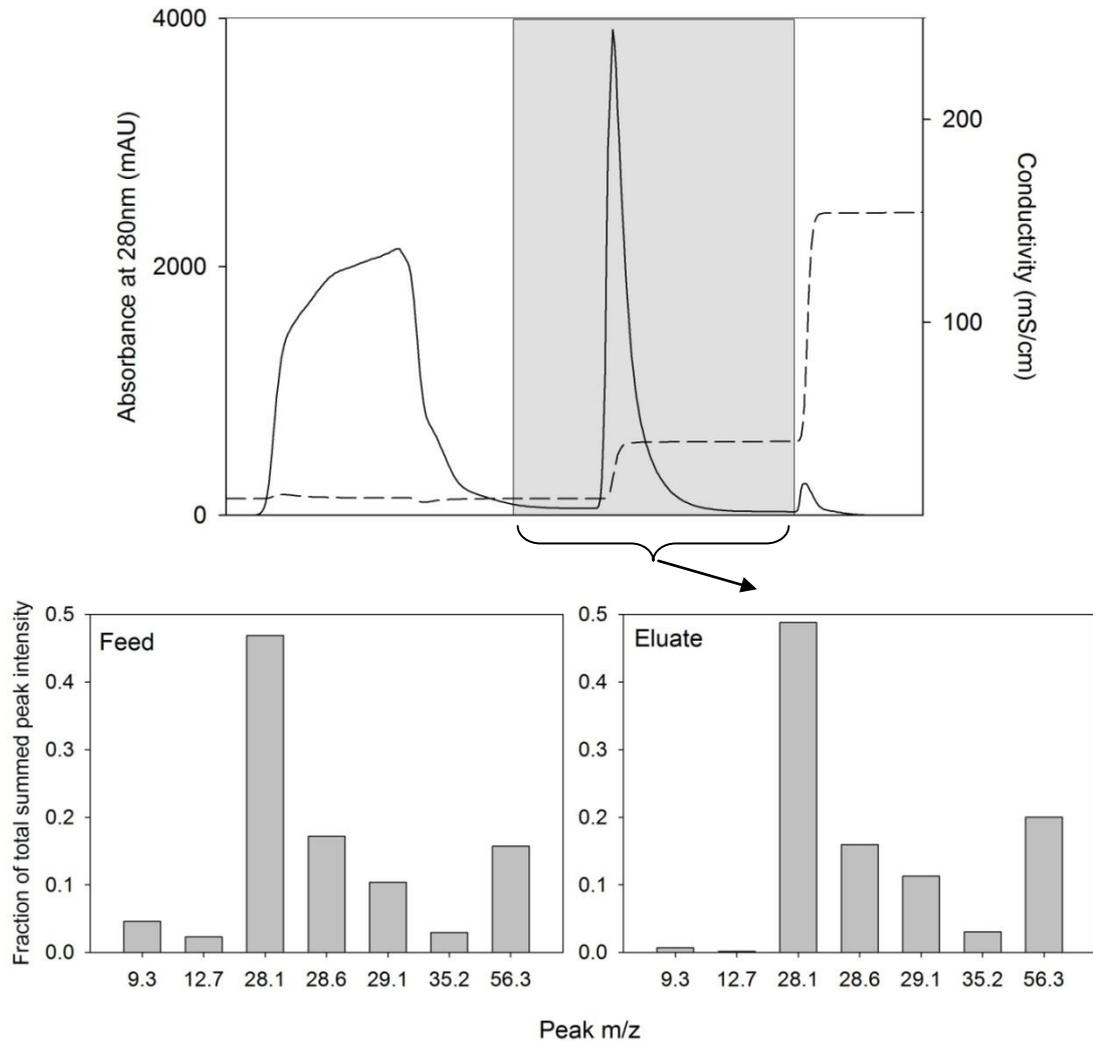


Figure 7.9 Impurity removal analysis from chromatography elution fractions.

The column used was a 1mL Q Sepharose FF HiTrap column using 50mM Tris, pH 8, 100mM NaCl for equilibration, load and washing. For elution 50mM Tris, pH 8, 400mM NaCl was used. Pooled elution fractions are shown by shaded region. Analysis was carried out on a Q10 ProteinChip with 50mM Tris, pH 8.0, 0mM NaCl. Proteins chosen for study were either very distantly related to the product (9.3 and 12.7kDa/e proteins), moderately related to the product (35.2kDa/e) or nearest neighbours to the product (28.6, 29.1, and 56.3). Bar charts are normalised based on the total summed peak intensity for all seven peaks for each sample.

7.4 Summary

Review of chapter aims:

- Incorporate peak normalisation, where peaks are compared against one another within a spectrum to reduce variability.
 - Star plots and dendrograms incorporate normalisation by comparing responses between peaks.

- Develop a physicochemical map to compare proteins within a given process material.
 - Using dendrograms, a visualisation tool usually used in taxonomy for grouping species or genes it was possible to compare proteins based on their physicochemical properties for bioprocess development.

- Determine nearest neighbours to target protein.
 - SELDI-MS was used to gather the data required for this analysis in a matter of days enabling identification of impurities that are hard to separate from the product.
 - Impurities that were observed to be difficult to remove in a previously developed bioprocess were predicted correctly by this approach.
 - Impurities that were predicted to be hard to remove were most likely product related.

- Commence investigation into the conditions that may be promising for separating the target protein from nearest neighbours.
 - Conditions for the separation of the product from these nearest neighbours were found to be challenging, on the basis of simple hydrophobic and electrostatic adsorption chemistries.
 - Although conditions for separation of ApoA-IM monomer from its nearest neighbours were not found, separation from very distantly related proteins was observed.

- The ProteinChips could give some clues on the conditions to focus on in future sorbent scouting experiments.
- The method gives a rapid indication of the separation challenge for a particular product and contaminant profile using minimal material. This is an important metric to determine the resources required for process development.

CHAPTER 8 CONCLUSIONS & RECOMMENDATIONS FOR FUTURE WORK

8.1 Review of project objectives

8.1.1 Objective 1: Develop the use of a mass spectrometry technology for bioprocess materials

This EngD project has generated Matlab code (appendix) for the easy upload and treatment of SELDI-MS data for further applications. This significantly benefited this project but could also be beneficial for researchers who do not want to be confined to a platform limited to a graphical user interface.

8.1.2 Objective 2: Application of the approach to an industrial process in conjunction with a USD device

During this project I was fortunate enough to observe the ApoA-IM process at Pfizer. This gave me an excellent opportunity to test the capabilities of SELDI-MS as a process monitoring tool against other more conventional approaches. Although SELDI-MS would require a large amount of optimisation to allow absolute quantification, it did demonstrate some key advantages over other techniques. This included its ability to handle very crude feed streams, even those near the start of a process. This meant that unlike other approaches, SELDI-MS could be used at every stage of protein production. SELDI-MS could also provide information that would usually require all three conventional analytical approaches.

It was also possible to incorporate an ultra scale-down shear device study with the aim of improving the flocculation/centrifugation stage of the ApoA-IM process.

This work which was an extension on previous work at UCL showed that the flocculation/centrifugation could be improved by adding more flocculating agent at a slower rate. This also showed the benefit at pilot scale that the ultra-scale down predicted.

This improvement could have been due to the slower flow rate allowing an increased mixing time for the flocculating agent and therefore increased aging of the flocs. If flocs had a longer time to form their mechanical strength could have greatly increased. Adding more flocculant would also ensure complete flocculation of the material which would also improve clarification.

The material generated from the flocculation/centrifugation stage was used for later purification development work.

8.1.3 Objective 3: Develop novel approaches for acquisition and analysis of mass spectra to expedite bioprocess development and monitoring

Aware of the limitations of using SELDI-MS for chromatography sorbent screening meant the need to apply a different approach. SELDI-MS could be used to examine the relationship between proteins present in process streams and reveal which impurities maybe hardest to remove down-stream. This may then suggest which impurities should be knocked out of the process during fermentation by changing or modifying the host or the conditions it is grown under. This technique could also propose initial purification conditions for impurity removal.

8.2 Publications

The above objectives provide a summary of the original aspects of the work presented in this thesis, of which a significant portion has been published, or accepted for publication in a peer reviewed journal. Selections of this work have also been presented in poster and oral form.

Peer reviewed publications:

Berrill A, Ho SV, Bracewell DG. Ultra scale-down to define and improve the relationship between Flocculation and disc-stack centrifugation. *Biotechnology Progress* 2008; 24:426-431.

Berrill A, Ho SV, Bracewell DG. Product and Contaminant Measurement in Bioprocess Development by SELDI-MS. *Biotechnology Progress*. In press.

Conference presentations:

Berrill A, Ho SV, Bracewell DG. Ultra scale-down to define and improve the relationship between flocculation and disc-stack centrifugation. American Institute of Chemical Engineers. 1st Annual Meeting for the Society of Biological Engineering, San Diego, CA, US, March 19-23, 2007. (Poster)

Berrill A, Ho SV, Bracewell DG. Ultra scale down to improve the flocculation disc stack centrifuge interaction. American Institute of Chemical Engineers. Annual Meeting, Salt Lake City, UT, US, Nov 4-9, 2007. (Oral)

Berrill A, Ho SV, Bracewell DG. Accelerated physicochemical characterization of protein mixtures to aid purification sequence selection. American Institute of Chemical Engineers. Annual Meeting, Salt Lake City, UT, US, Nov 4-9, 2007. (Oral)

8.3 Final comments and future work

This section will evaluate the three objectives of this thesis, in terms of their future potential, highlighting those ones of greatest importance.

8.3.1 Related to objective 1: Develop the use of a mass spectrometry technology for bioprocess materials

Key developments to this objective:

- 1) Analysis of data treatment routines, particularly baseline correction and absolute quantification.

It is widely known that obtaining quality SELDI-MS data is difficult and more research in treating this data is required. A surprising result in this thesis was that

baseline correction actually reduced reproducibility. Further investigation is required to identify where these sources of variation occur. Once this is clearer better suited data treatment algorithms can be chosen to reduce this variation. One of the major downsides of the TOF-MS approach is the increased signal intensities for peaks at lower m/z . This makes absolute quantification without standards for each protein currently impossible. A data treatment technique needs to be developed to correct for the expansion in the ion clouds that are known to cause this phenomenon. One way to do this may involve bridging the proteins of interest with protein standards. If the protein standards are quantifiable then it may also be possible to relate changes in the peak signals between the standards to known quantities.

8.3.2 Related to objective 2: Application of the approach to an industrial process in conjunction with a USD device

Key developments to this objective:

1) Varying additional variables for the flocculation and centrifugation.

Only one type of flocculating agent was used in the ApoA-IM process but it would be of interest to investigate the use of other flocculating agents. Increasing the number of flocculating agent concentrations and the levels added would also be beneficial.

In terms of analysis, further particle size measurements could be implemented to examine further how particle morphology changes with different flocculation conditions. In this work, microscope images were taken but further analysis using equipment such as light scattering would also provide useful information.

8.3.3 Related to objective 3: Develop novel approaches for acquisition and analysis of mass spectra to expedite bioprocess development and monitoring

Key developments to this objective:

1) Investigate ability of SELDI-MS to give further information on HCPs and other impurities such as nucleic acids and endotoxins.

To gather detailed information on HCPs present in feed materials is to isolate them prior to their peptide mapping to confirm their identities with existing protein databases. A far less labour intensive and timely method maybe possible if HCP standards are available such as those used for ELISA. These standards could be run either individually or in a mixture on a non-selective ProteinChip to confirm their m/z values. A mixture of these HCPs could then be profiled under the same array of adsorption conditions as those used for the feed material. Provided there is minimal competition between proteins, an HCP that is present in the HCP standard as well as the feed material should have a similar adsorption pattern. The closer the masses of the two proteins and the more similar the adsorption pattern the greater the likelihood they are indeed the same protein.

At present there is very little literature on the use of SELDI-MS to detect nucleic acids and endotoxins but this is also an area where it may be possible to expand the SELDI-MS technique.

2) Identify nearest neighbours in other feed materials and confirm the difficulty of their removal and define how close an impurity has to be to the product in order for it to become problematic to remove.

By analysing other feed materials with SELDI-MS and using existing knowledge of the difficulties of purifying certain components it should be possible to build up a database of how close impurities have to be on a dendrogram to become difficult to separate from one another. This could result in defining a cut-off (annotated by C in Figure 8.1) for any new material examined. The cut-off would group those impurities that co-elute with the target protein under the majority of separation conditions (e.g. cluster 2 in Figure 8.1). It may then mean that more novel separation techniques such as mixed-mode chromatography would need to be adopted to separate these nearest neighbours.

This work may also allow a cost for removal to be assigned to each impurity which is linked to the difficulty of their separation from the target protein. The

closer the nearest neighbour, the more costly purification will be. If there are many near neighbours, column capacities may also be lower for the target protein due to nearest neighbours competing for adsorption sites.

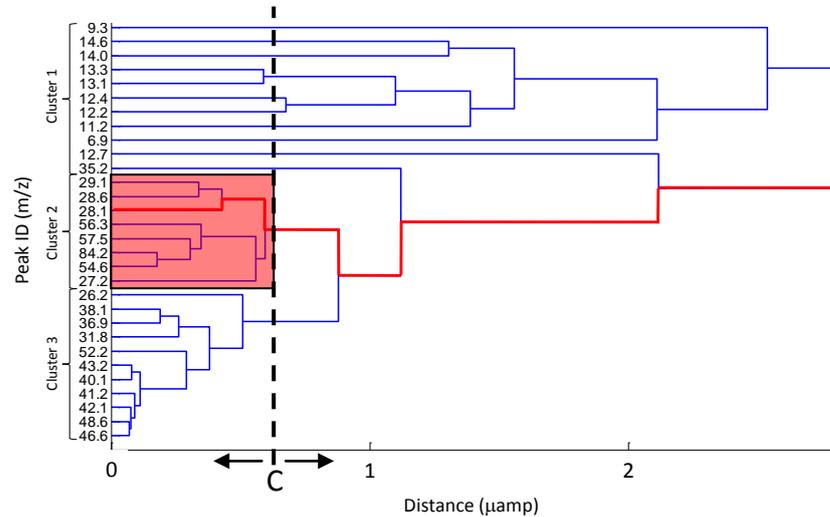


Figure 8.1 Defining cut-off to cluster hard to remove proteins

The cut-off is defined by C and could move depending on further analysis of how hard these impurities are to remove using conventional process purification techniques.

3) Inform cell biologists of ways of cell engineering to help enhance recovery/purification performance.

If impurities are found that are close enough to the target protein that the cost of separating them downstream is greater than the cost of engineering the host so they are no longer expressed, then perhaps cell engineering is the preferred option. For cell engineering these impurities would need to be identified using an advanced LC-MS system combined with a database search. The cell biologist would then be able to decipher which genes produced them and how they should be knocked out of the hosts DNA to prevent their expression.

Without these nearest neighbours being produced, downstream processing would be greatly simplified and less resource intensive.

- 4) Investigation the adsorption characteristics of the ProteinChips relate to chromatography column formats.

In order to develop chromatography operations based on ProteinChip results further understanding of the correlation between the two is required. This thesis shows it is possible to define which proteins are likely to be problematic to separate. Further dendrograms can be created to show the relationship between nearest neighbours and product on the individual SELDI surface chemistries available. What remains to be defined is what this establishes in terms of the chromatographic operation considering the variety of materials and surface chemistries available as adsorbents.

- 5) Design of chromatographic sequences.

As further correlations between adsorption on ProteinChips and modes of chromatographic resins are made there is increased likelihood that SELDI-MS data could be used to design chromatography sequences. This could involve selecting the first chromatography chemistry in a chromatography sequence to be the one that shows the greatest distance between the target protein and the overall nearest neighbour when all chemistries are considered. There are many other elements that could be considered in the complex algorithms that would need to be generated for this work. For instance, a ranking system could be introduced based on the removal of some proteins early on. For example, if a protease is present this may need to be weighted heavily so that the distances in the dendrogram for this impurity are reduced making it more of a concern to remove this impurity early on.

CHAPTER 9 VALIDATION OF BIOPROCESSES: IMPLICATION OF SCALE-DOWN METHODOLOGIES

9.1 Process validation

Biopharmaceutical products need to be efficacious and economically viable to produce while satisfying regulatory requirements on their safety. Since they are biologically derived these medicines usually have to be purified from very complex mixtures of DNA, lipids and other proteins including those with similar properties to the protein of interest.

Since the protein produced is defined by the process that produces it, it is the process needs to be validated. Throughout production, data must be collected that gives assurances that the process implies product quality. The process must consistently meet these specifications. This includes a robust and well characterised process that can withstand inevitable variance in operating conditions including operating error.

9.2 Scale-down models

At pilot scale and larger it is rarely economically feasible to test a wide range of operating conditions. This is due in part to the large amount of resources that would need to be employed. Also, this stage is likely to be on the critical path of process development and time spent here will ultimately delay product launch and hence the recuperation of the capital invested along with any expected profit.

Technologies that can generate the data required with minimal resource expenditure while running in parallel with large scale development are in high demand. These technologies include scale down mimics of the large scale operations. An ideal scale down device will use minimal material while having high throughput and minimal burden on timely analytical assays.

The scale down mimic must match the larger scale in terms of key process parameters with an equivalent set of operating parameters. Any change in the operating parameters at either scale should result in comparative outputs. If this is

achievable then this scale-down device would be suitable to be used for process validation of the larger scale operation.

Scale-down models have been used for a number of large scale unit operations including; centrifugation, filtration, hydrophobic interaction chromatography, expanded bed chromatography, and bioreactors (Boychyn et al. 2004;Boychyn et al. 2000;Neal et al. 2003;Reynolds et al. 2003;Willoughby et al. 2004).

9.2.1 Process validation scenario: Effect of contaminant properties on purification

The initial concept behind the ProteinChip experiments used in this project were to simulate the adsorption at ultra scale-down to that experienced during large scale chromatography. Due to the complexity of the mechanisms and different operation factors of the two scales it is very hard to precisely mimic the chromatography using ProteinChip technology.

After continued efforts the decision was made to concentrate on using SELDI-MS as an analytical technique to provide wide arrays of information about our process material in a minimal amount of time.

Using SELDI-MS in this way allows the user to identify impurities that maybe particularly difficult to remove and adjust their host organism or fermentation conditions to reduce these or concentrate on the unit operations that will be the most effective in removing these problem proteins. Once an initial process is developed around this framework SELDI-MS can then be used to validate the process by analysing different feed streams shown in chapter three.

9.2.2 Process validation scenario: A new protein candidate for a process development team

Imagining the scenario of a biopharmaceutical company discovering a protein or wanting to acquire or buy the license for a protein that displays clinical benefit. How does the company evaluate early on if producing this protein will be economically feasible to produce at industrial scale? The methodology outlined

could be a useful way to do this before many resources have committed that may have been better spent on more viable alternatives.

Figure 9.1 gives an outline of the types of steps that may occur when a new protein candidate is presented to process development team.

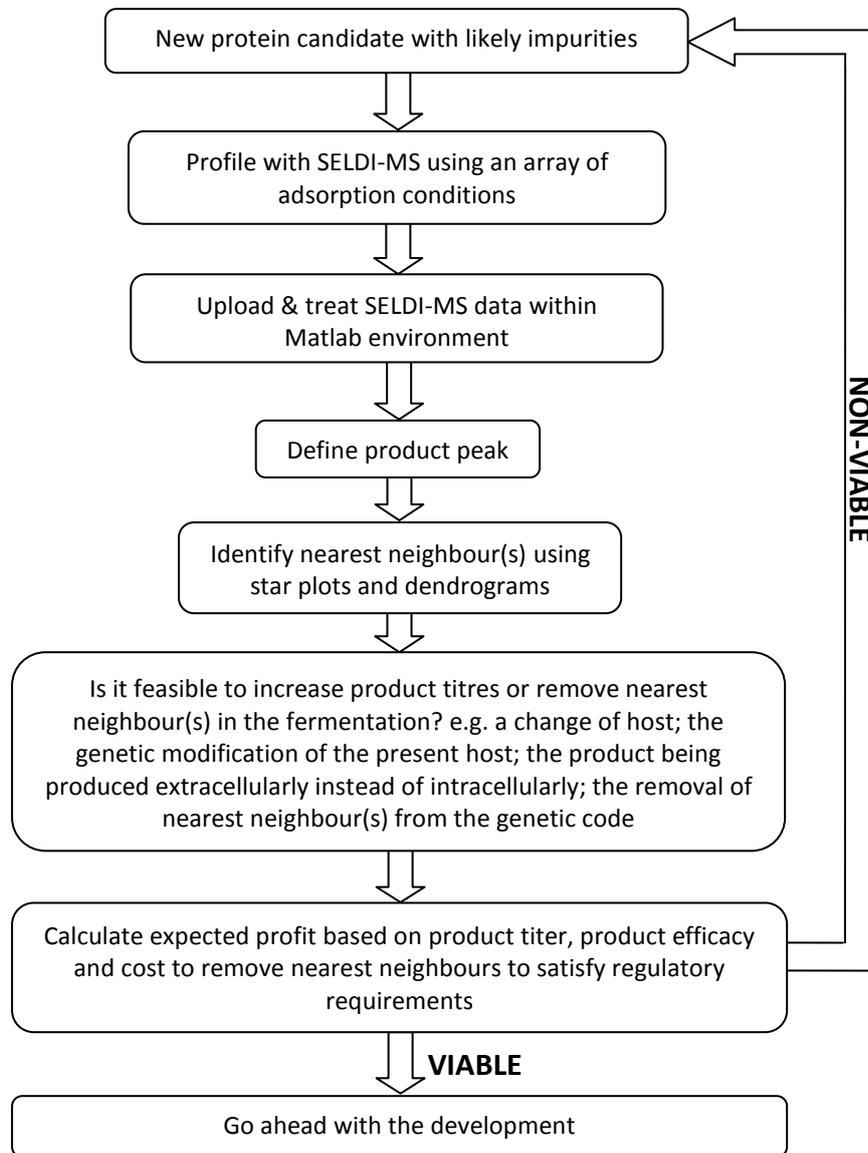


Figure 9.1 Theoretical framework for the rapid evaluation of protein candidates at early stage process development.

The steps are based on product titres, product efficacy and expected cost of removing nearest neighbour impurities to satisfy regulatory requirements.

ABBREVIATIONS

| | |
|-----------------|--|
| 2D-PAGE: | Two Dimensional-PolyAcrylamide Gel Electrophoresis |
| ABS: | Aqueous Biphasic Systems |
| ACN: | Acetonitrile |
| ADH: | Alcohol Dehydrogenase |
| ADI: | ApoA-IM Dimer Intermediate |
| ApoA-IM: | ApolipoproteinA-IM |
| BSA: | Bovine Serum Albumin |
| CM: | Carboxy Methyl |
| CV: | Coefficient of Variation |
| DEAE: | Diethylaminoethane |
| DRT: | Dimensionless Retention Time |
| DTT: | Dithiothreitol |
| EAM: | Energy Absorbing Molecule |
| EBA: | Expanded Bed Adsorption |
| <i>E.coli</i> : | <i>Escherichia coli</i> |
| ELISA: | Enzyme-linked immuno sorbent assay |
| ESI-TOF-MS: | ElectroSpray Ionisation-Time of Flight-Mass Spectroscopy |
| Fab: | Antibody Fragment |
| FDA: | Food and Drug Administration |
| HCP: | Host Cell Protein |
| HEW: | Hen Egg White |
| HIC: | Hydrophobic Interaction Chromatography |
| HPLC: | High Performance Liquid Chromatography |
| IEX: | Ion-Exchange |
| IMAC: | Immobilised Metal Affinity Chromatography |
| kDa: | kiloDalton |
| MALDI-TOF-MS: | Matrix Assisted Laser Desorption/Ionisation-Time of Flight- Mass Spectrometry |
| MS: | Mass Spectrometry |

| | |
|---------------|--|
| MW: | Molecular Weight |
| NP: | Normal Phase |
| PAT: | Process Analytical Technology |
| PBS: | Phosphate Buffered Saline |
| pI: | Isoelectric point |
| Q: | Quaternary amine |
| QbD: | Quality by Design |
| QSPR: | Quantitative Structure Property Relationships |
| RP: | Reverse Phase |
| SDS-PAGE: | Sodium Dodecyl Sulphate-PolyAcrylamide Gel Electrophoresis |
| SE: | Size Exclusion |
| SEAC: | Surface Enhance Affinity Capture |
| SELDI-TOF-MS: | Surface Enhanced Laser Desorption/Ionisation-Time of Flight - Mass Spectroscopy |
| SP: | Sulfopropyl |
| SPA: | Sinapinic Acid |
| TFA: | Trifluoroacetic acid |
| TMB: | Tetramethylbenzidine |
| UDWT: | Undecimated Discrete Wavelet Transform |
| UF/DF: | Ultrafiltration/Diafiltration |
| UPGMA: | Unweighted Pair Group Method with Arithmetic mean |
| USD: | Ultra-Scale Down |
| UV: | Ultra Violet |

NOMENCLATURE

| Symbol | Description | Unit |
|-------------------------|---|--------------|
| g | Acceleration due to gravity | $m.s^{-2}$ |
| Φ_{surface} | Average surface hydrophobicity value | - |
| ϕ | Cellular solids fraction | - |
| z | Charge | e |
| C_i | Concentration of protein i | $mol.L^{-1}$ |
| c | Correction factor to account for non-ideal fluid flow patterns, the value of which depends on centrifuge design | - |
| DRT | Dimensionless retention time | - |
| Σ_{ds} | Equivalent settling area of the disc-stack centrifuge | m^2 |
| Σ_{lab} | Equivalent settling area of the laboratory centrifuge | m^2 |
| Q | Flow rate | $m^3.s^{-1}$ |
| θ | Half disc angle in disk stack centrifuge | degrees |
| h | Hindered settling correction factor | - |
| ϕ_{aai} | Hydrophobicity value for amino acid 'i' | - |
| r_1 | Inner disc radius in disk stack centrifuge | m |
| R_i | Inner radius, the distance between the centre of rotation and the top of the liquid in USD centrifuge | m |
| M | Mass of protein | kDa |
| OD_f | Optical density of the feed | nm |
| OD_r | Optical density of the reference | nm |
| OD_s | Optical density of the supernatant | nm |
| r_2 | Outer disc radius in disk stack centrifuge | m |
| R_o | Outer radius, the distance between the centre of rotation and the bottom of the liquid in USD centrifuge | m |
| n_d | Number of active discs in industrial scale centrifuge | - |
| K_i | Partition coefficient | - |
| %C | Percentage Clarification | - |
| T_{lab} | Residence time in the laboratory centrifuge | s |
| ω | Rotational speed | $rad.s^{-1}$ |

| | | |
|-----------|--|----------|
| N | Rotational speed, rps | s^{-1} |
| S_{aai} | Solvent accessible area occupied by amino acid 'i' | - |
| t_f | Time corresponding to the end of the elution gradient | s |
| t_r | Time corresponding to the retention time of the target protein | s |
| t_0 | Time corresponding to the start of the elution gradient | s |
| t_t | Time of flight | s |
| S_p | Total solvent accessible area of the protein | - |
| V_{lab} | Volume of process material in laboratory centrifuge tube | m^3 |

APPENDIX - MATLAB CODE

Reading spectra data

```
%Enter directory where the data is located

clear all

path=cd;

directory=find(path=='\');

currentfile=path(directory(end-1)+1:end);

currentfile(currentfile=='.')='_';
currentfile(currentfile=='\')='_';

%Read Data

cd([cd '\ 'Processed']);

[MZ SELDIProcessed,Metapropnames,Metapropvalues,filenames]=readProcessed;

cd ..

cd([cd '\ 'Raw']);

[Raw]=readRaw(filenames);

%Remove redundant data from Processed & Raw
[u numcols]=size(MZ);

for i=1:numcols;
    b(:,i)=min(find(isnan(MZ(:,i))==1));
end

clearInd=max(b);

MZ(clearInd:end,:)=[];
SELDIProcessed(clearInd:end,:)=[];
Raw(clearInd:end,:)=[];

% Process Raw Data
[Raw,Smoothed]=basecorr2(Raw, MZ, 6, 10, 6);
```

```
%Mass=MZ;

clear b

AverageMass1=[];
AverageMass2=[];

clear e

for i=1: numel(filenamees);

    %remove csv
    a=strmatch(filenamees{i}{:,end-3:end},'.csv');

    if a==1;
        filenamees{i}=filenamees{i}{:,1:end-4};
    end

    %remove _1
    b=strmatch(filenamees{i}{:,end-1:end},'_1');

    if b==1;
        filenamees{i}=filenamees{i}{:,1:end-2};
    end

end

for i=1: numel(filenamees);

    f=numel(strmatch(filenamees{i},filenamees));

    e(i,1:f)=strmatch(filenamees{i},filenamees)';

end

[l m n]=unique(e,'rows');

variablenames=[];
variables=[];
variablelevels=repmat(NaN,numel(filenamees),1);

%Sorting Variables
if numel(filenamees(Karpievitch et al. 2007))==10 & numel(filenamees{end})==10;
    [variablenames variables variablelevels]=datasorting2(filenamees,MZ);
end
```

```
MinCutOff=5000;
MidCutOff=22500;
MaxCutOff=90000;

lowStepSize=200;
highStepSize=200;

lowWindowSize=900;
highWindowSize=1800;

%if numel(f)>=2;

newMass= repmat(NaN,25000,numel(m));
newSELDIProcessed=repmat(NaN,25000,numel(m));
newRaw=repmat(NaN,25000,numel(m));
newBaselineCorrected=repmat(NaN,25000,numel(m));
newSmoothed=repmat(NaN,25000,numel(m));

%if numel(m)<numel(l);

%Merging of low & high mass spectra
if m(1)==2;

    for i=1:numel(m);

        o=min(MZ(:,l(i,1)));
        p=max(MZ(:,l(i,1)));
        AverageMass1=(o+p)/2;

        x=min(MZ(:,l(i,2)));
        y=max(MZ(:,l(i,2)));
        AverageMass2=(x+y)/2;

        if AverageMass1<AverageMass2;

            %Process low and high mass data
            lowMass=MZ(:,l(i,1));
            highMass=MZ(:,l(i,2));
            lowSmoothed=Smoothed(:,l(i,1));
            highSmoothed=Smoothed(:,l(i,2));

            lowMass(isnan(lowMass))=[];
            highMass(isnan(highMass))=[];
            lowSmoothed(isnan(lowSmoothed))=[];
            highSmoothed(isnan(highSmoothed))=[];
```

```

baselineCorrectedlow=msbackadj(lowMass,lowSmoothed,'WindowSize',low
WindowSize,'StepSize',lowStepSize,'QuantileValue',0.1);

baselineCorrectedhigh=msbackadj(highMass,highSmoothed,'WindowSize',hi
ghWindowSize,'StepSize',highStepSize,'QuantileValue',0.1);

%Splice treated data together
a=max(find(MZ(:,l(i,1))<MinCutOff));
b=max(find(MZ(:,l(i,1))<MidCutOff));

c=max(find(MZ(:,l(i,2))<MidCutOff));
d=max(find(MZ(:,l(i,2))<MaxCutOff));

if isempty(a);
    [u a]=min(MZ(:,l(i,1)));
end

if isempty(b) || isempty(c) ;
    [u b]=max(MZ(:,l(i,1))); %some NaN so can not have numel, need max
    c=max(find(MZ(:,l(i,2))<=max(MZ(:,l(i,1)))));
end

if isempty(d);
    [u d]=max(MZ(:,l(i,2)));
end

k=length([MZ(a:b,l(i,1));MZ(c:d,l(i,2))]);

newMass(1:k,i)=[MZ(a:b,l(i,1));MZ(c:d,l(i,2))];

newSELDIProcessed(1:k,i)=[SELDIProcessed(a:b,l(i,1));SELDIProcessed(c:d,l(i,
2))];
newMetapropvalues(:,i)=[Metapropvalues(:,l(i,1));Metapropvalues(:,l(i,2))];
newRaw(1:k,i)=[Raw(a:b,l(i,1));Raw(c:d,l(i,2))];

newBaselineCorrected(1:k,i)=[baselineCorrectedlow(a:b,:);baselineCorrecte
dhigh(c:d,:)];
newSmoothed(1:k,i)=[Smoothed(a:b,l(i,1));Smoothed(c:d,l(i,2))];
V(i,:)=variablelevels(l(i,1),:);

end

if AverageMass1>AverageMass2;

% Process low and high mass data
lowMass=MZ(:,l(i,2));

```

```

highMass=MZ(:,l(i,1));
lowSmoothed=Smoothed(:,l(i,2));
highSmoothed=Smoothed(:,l(i,1));

lowMass(isnan(lowMass))=[];
highMass(isnan(highMass))=[];
lowSmoothed(isnan(lowSmoothed))=[];
highSmoothed(isnan(highSmoothed))=[];

baselineCorrectedlow=msbackadj(lowMass,lowSmoothed,'WindowSize',low
WindowSize,'StepSize',lowStepSize,'QuantileValue',0.1);

baselineCorrectedhigh=msbackadj(highMass,highSmoothed,'WindowSize',hi
ghWindowSize,'StepSize',highStepSize,'QuantileValue',0.1);

%Splice treated data together
a=max(find(MZ(:,l(i,2))<=MinCutOff));
b=max(find(MZ(:,l(i,2))<=MidCutOff));

c=max(find(MZ(:,l(i,1))<=MidCutOff));
d=max(find(MZ(:,l(i,1))<=MaxCutOff));

if isempty(a);
    [u a]=min(MZ(:,l(i,2)));
end

if isempty(b) || isempty(c) ;
    [u b]=max(MZ(:,l(i,2))); %some NaN so can not have numel, need max
    c=max(find(MZ(:,l(i,1))<=max(MZ(:,l(i,2)))));
end

if isempty(d);
    [u d]=max(MZ(:,l(i,1)));
end

k=length([MZ(a:b,l(i,2));MZ(c:d,l(i,1))]);

newMass(1:k,i)=[MZ(a:b,l(i,2));MZ(c:d,l(i,1))];

newSELDIProcessed(1:k,i)=[SELDIProcessed(a:b,l(i,2));SELDIProcessed(c:d,l(i,
1))];
newMetapropvalues(:,i)=[Metapropvalues(:,l(i,2));Metapropvalues(:,l(i,1))];
newRaw(1:k,i)=[Raw(a:b,l(i,2));Raw(c:d,l(i,1))];

newBaselineCorrected(1:k,i)=[baselineCorrectedlow(a:b,:);baselineCorrecte
dhigh(c:d,:)];
    
```

```

        newSmoothed(1:k,i)=[Smoothed(a:b,l(i,2));Smoothed(c:d,l(i,1))];
        V(i,:)=variablelevels(l(i,1),:);

    end
end

% if one spectrum per condition no splicing necessary
else
    for i=1:numel(variablelevels(:,1));

        a=max(find(MZ(:,i)<=MinCutOff));
        b=max(find(MZ(:,i)<=MidCutOff));
        d=max(find(MZ(:,i)<=MaxCutOff));

        if isempty(a);
            [u a]=min(MZ(:,i));
        end

        if isempty(b);
            b=(a+d)/2;
        end

        if isempty(d);
            [u d]=max(MZ(:,i));
        end

        k=length([MZ(a:d,i)]);

        newSELDIProcessed(1:k,i)=[SELDIProcessed(a:d,i)];
        newMetapropvalues(:,i)=[Metapropvalues(:,i)];

        lowMass=MZ(a:b,i);
        highMass=MZ(b+1:d,i);
        lowSmoothed=Smoothed(a:b,i);
        highSmoothed=Smoothed(b+1:d,i);

        lowMass(isnan(lowMass))=[];
        highMass(isnan(highMass))=[];
        lowSmoothed(isnan(lowSmoothed))=[];
        highSmoothed(isnan(highSmoothed))=[];

        baselineCorrectedlow=msbackadj(lowMass,lowSmoothed,'WindowSize',low
        WindowSize,'StepSize',lowStepSize,'QuantileValue',0.1);
    end
end

```

```

baselineCorrectedhigh=msbackadj(highMass,highSmoothed,'WindowSize',hi
ghWindowSize,'StepSize',highStepSize,'QuantileValue',0.1);

newMass(1:k,i)=[lowMass;highMass];
newRaw(1:k,i)=[Raw(a:b,i);Raw(b+1:d,i)];
newBaselineCorrected(1:k,i)=[baselineCorrectedlow;baselineCorrectedhigh];
newSmoothed(1:k,i)=[lowSmoothed;highSmoothed];
V(i,:)=variablelevels(i,:);
end
end

%End of Sorting Masses function

%clear a
clear variablelevels;
clear Metapropvalues;
clear Mass;
clear SELDIProcessed;
clear Raw;
clear BaselineCorrected;
clear Smoothed;

variablelevels=V;
Metapropvalues=newMetapropvalues;
Mass=newMass;
SELDIProcessed=newSELDIProcessed;
Raw=newRaw;
BaselineCorrected=newBaselineCorrected;
Smoothed=newSmoothed;

%crop data at max mass
[z clearInd]=max(Mass(:,1));

Mass(clearInd+1:end,:)=[];
SELDIProcessed(clearInd+1:end,:)=[];
Raw(clearInd+1:end,:)=[];
BaselineCorrected(clearInd+1:end,:)=[];
Smoothed(clearInd+1:end,:)=[];

Mass=Mass/1000;
%Mass(Mass==0)=NaN;

Noise=Raw-Smoothed;
Baseline=Smoothed-BaselineCorrected;

cd ..

```

```

save([currentfile
'Spectra'],'Mass','SELDIProcessed','Raw','Noise','Smoothed','Baseline','BaselineCorr
ected')%,'AUCMass','AUCSELDIProcessed','AUCRaw','AUCSmoothed','AUCBaselineC
orrected');
save([currentfile
'MetaData'],'Metapropnames','Metapropvalues','filenames','variablenames','variabl
es','variablelevels');

[nummasses numcols]=size(Mass);

%stop it crashing from too much data
if numcols<=32;

    load([currentfile 'Spectra']);

    xlsxwrite([currentfile 'Raw'],[Mass Raw]);
    xlsxwrite([currentfile 'SELDIProcessed'],[Mass SELDIProcessed]);
    xlsxwrite([currentfile 'Noise'],[Mass Noise]);
    xlsxwrite([currentfile 'Smoothed'],[Mass Smoothed]);
    xlsxwrite([currentfile 'Baseline'],[Mass Baseline]);
    xlsxwrite([currentfile 'BaselineCorrected'],[Mass BaselineCorrected]);

    load([currentfile 'MetaData']);

    if AverageMass1~=AverageMass2;
        xlsxwrite([currentfile 'MetaData'],[repmat(Metapropnames,2,1)
Metapropvalues]);
    else
        xlsxwrite([currentfile 'MetaData'],[Metapropnames Metapropvalues]);
    end

end

```

Reading processed data

```

%Read data

%Enter directory files are in and run

function [MZ,Processed,propertyname,propval,csvfiles]=readProcessed

%Upload data and get conditions

%function
[operatingparameters,MZ,Int,datafilenames]=uploadSELDInp20data(fullpath);

```

```

%fullpath=['c:\My Documents\Matlab\Functions & Data\Data\' sample '\ 'SELDI' '\
experiment '\ intensity '\ datatreatment '\ 'Raw' '\'];
Int=[];
%cd(fullpath);

[stat,mess]=fileattrib('*');

[r numfiles]=size(mess);

for i=1:numfiles;
datafullpath{i}=mess(i).Name;
end

[pathrows pathcols]=size(cd);

for i=1:numfiles;
datafilenames{i}=datafullpath{i}{:,pathcols+2:end};
end

datafilenames=datafilenames';

chardatafilenames=char(datafilenames);

for i=1:numfiles;
a=char(datafilenames(i,:));
[numletters]=length(deblank(a));
x=strmatch(a(:,numletters-2:numletters),'csv','exact');
if x==1;
j(i)=i;
end
end

csvfiles=datafilenames(nonzeros(j),:);

[numfiles]=length(csvfiles);

%csvfiles

[MZ, Processed, propertyname, propval]=GetSpectraData(csvfiles);

```

Reading raw data

```

%Read data

%Enter directory files are in and run

```

```

function [Raw]=readRaw(csvfiles)

%path=['c:\My Documents\Matlab\Data\' sample '\' samplenummer '\' equipment
 '\' experiment '\' intensity '\' datatreatment '\'];

%Upload data and get conditions

%function
[operatingparameters,MZ,Int,datafilenames]=uploadSELDInp20data(fullpath);

[numfiles]=length(csvfiles);
Raw= repmat(NaN,150000,numfiles);

for i=1:numfiles;
[propertyname propertyvalue]=textread(csvfiles{i},'%s
%s','delimiter','=', 'emptyvalue',NaN);

%propertyname

%Take mass spec data from propertyname
[a]=size(str2num(char(propertyvalue)));

TOFmultiplier=str2num(char(propertyvalue(a(1)-1)));
TOFdata=str2num(char(propertyname(a(1)+2:end,:)));

size(TOFmultiplier);
lenTOFdata=length(TOFdata);

Raw(1:lenTOFdata,i)=TOFdata*TOFmultiplier;
end

```

Retrieve spectra data

```

function [lowcatMZ, lowcatInt, propertyname, propval]=GetSpectraData(csvfiles);

numcsvfiles=numel(csvfiles);

lowcatMZ=repmat(NaN,150000,numcsvfiles);
lowcatInt=repmat(NaN,150000,numcsvfiles);
propertyname=[];

propval=cell(130,numcsvfiles);

for i=1:numcsvfiles;
[propertyname propertyvalue]=textread(csvfiles{i},'%s %s','delimiter','=');

```

```

a=strmatch(char('M/Z, Intensity'),char(propertyname));

%a
%propertyname

%Take mass spec data from propertyname
[massspecdata]=str2num(char(propertyname(a+1:end)));

MZ=massspecdata(:,1);
Int=massspecdata(:,2);

lowcatMZ(1:numel(MZ),i)=MZ;
lowcatInt(1:numel(Int),i)=Int;

%Clear mass spec data at end

propertyvalue(a:end,:)=[];

%size(propval)
t(:,i)=a;

propval(1:a-1,i)=propertyvalue;
end

maxt=max(t);

propval(maxt:end,:)=[];
propertyname(maxt:end,:)=[];
    
```

Sort data

```

function [variablenames variables variablelevels]=datasorting2(filenamees,Mass);

variablelevels=[];
variables=[];

charcsvfiles=char(filenamees);

[Chip]=charcsvfiles(:,1);
ChipInd= repmat(NaN,numel(Chip),1);
pHInd= repmat(NaN,numel(Chip),1);
SaltInd= repmat(NaN,numel(Chip),1);
SaltConcInd= repmat(NaN,numel(Chip),1);

%ChipType
cm=strmatch('c',Chip);
    
```

```
qu=strmatch('q',Chip);
hyd=strmatch('h',Chip);
nor=strmatch('n',Chip);
ima=strmatch('i',Chip);
dea=strmatch('d',Chip);

ChipInd(cm,:)=1;
ChipInd(qu,:)=2;
ChipInd(hyd,:)=3;
ChipInd(nor,:)=4;
ChipInd(ima,:)=5;
ChipInd(dea,:)=6;

%Mass
[Mass b MassInd]=unique(min(Mass));

%pH
[pH b pHInd]=unique(str2num(charcsvfiles(:,2:5)));
pH=pH/100;

%SaltType
[Salt]=charcsvfiles(:,6);

sod=strmatch('s',Salt);
amm=strmatch('a',Salt);
imi=strmatch('i',Salt);

SaltInd(sod,:)=1;
SaltInd(amm,:)=2;
SaltInd(imi,:)=3;

%Salt Conc
[SaltConc b SaltConcInd]=unique(str2num(charcsvfiles(:,7:10)));

variables(Karpievitch et al. 2007)={'c';'q';'h';'n';'i';'d'};
variables(Karpievitch et al. 2007;Merchant and Weinberger,2000)=Mass;
variables(Coombes et al. 2005;Karpievitch et al. 2007)=pH;
variables(Karpievitch et al. 2007;Xu and Lenhoff,2008)={'s';'a';'i'};
variables(Chen et al. 2008;Karpievitch et al. 2007)=SaltConc;
size(ChipInd)
size(MassInd)

variablelevels(:,1)=ChipInd;
variablelevels(:,2)=MassInd;
variablelevels(:,3)=pHInd;
variablelevels(:,4)=SaltInd;
```

```
variablelevels(:,5)=SaltConcInd;  
%FractionInd  
%variablelevels(:,6)=FractionInd;  
%variablelevels(:,7)=SpotInd;  
%variablelevels(:,8)=PartitionInd;
```

Raw data treatment

The code in this section is from the UT M.D. Anderson Cancer Center and follows these conditions:

CONDITIONS FOR USE:

Copyright (c) 2003, 2004 UT M.D. Anderson Cancer Center. All rights reserved.

This software is distributed and licensed to you on a non-exclusive basis, free-of-charge. Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistribution of source code must retain the above copyright notice, this list of conditions and the following disclaimer.
2. Redistribution in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.
3. All advertising materials mentioning features or use of this software must display the following acknowledgment: This product includes software developed by the University of Texas M.D. Anderson Cancer Center, Houston, Texas and its contributors.
4. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY UNIVERSITY OF TEXAS M.D. ANDERSON CANCER

CENTER, HOUSTON, TEXAS, AND CONTRIBUTORS AS IS AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL M.D. ANDERSON OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTIONS) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE), PRODUCT LIABILITY, OR OTHERWISE ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

Applicable code:

```
function [RawEnd, sm]= basecorr2(raw, rawMass, M, L, thld)

w = size(raw);           % how much data are we handling?
n = floor(w(1)/2^L);     % number of length 2^L pieces available

echo on
status = 'start baseline correction'
tic
echo off

midmass=22500;

bc = repmat(NaN, [w(1) w(2)]);
sm = repmat(NaN, [w(1) w(2)]);
RawEnd = repmat(NaN, [w(1) w(2)]);
for i = 1:w(2),

    Mass=rawMass(:,i);
    RawInt=raw(:,i);

    Mass(isnan(Mass))=[];
    RawInt(isnan(RawInt))=[];

    [s] = waveletSmoothAndBaselineCorrect(Mass,RawInt, thld, M, L, midmass);
```

```

c=length(Mass);

RawEnd(1:c,i)=RawInt;
sm(1:c,i)=s;

end
echo on
toc
status = 'end baseline correction'
echo off

Peak selection

%Enter directory where the data is located
clear all

path=cd;

directory=find(path=='\');

currentfile=path(directory(end-1)+1:end);

currentfile(currentfile=='.')='_';
currentfile(currentfile=='\')='_';

load([currentfile 'Spectra']);
load([currentfile 'MetaData']);

PeakMass=[6.855;9.277;11.16;12.22;12.41;12.66;13.06;13.28;14.03;14.6;26.2;27.18
;28.15;28.62;29.05;31.75;35.19;36.88;38.09;40.13;41.18;42.06;43.18;46.57;48.59;5
2.18;54.62;55.95;57.48;84.23];

[a b]=size(PeakMass);

[peakindexrange peakindexmid]=mspeakfindman(Mass(:,1),PeakMass);

RawPeakHeight=Raw(peakindexmid,:);
BcPeakHeight=BaselineCorrected(peakindexmid,:);
CIPHERGENPeakHeight=SELDIProcessed(peakindexmid,:);

for i=1:a;
    %SELDIInt(i,:)=sum(NormSELDIAUC(peakindexrange(i,1):peakindexrange(i,2),:),1);

    BcPeakArea(i,:)=sum(BaselineCorrected(peakindexrange(i,1):peakindexrange(i,2),:),
    1);
  
```

```
CiphergenPeakArea(i,:)=sum(SELDIProcessed(peakindexrange(i,1):peakindexrange(i,2),:),1);
```

```
[a b]=size(BcPeakArea);
```

```
maxBcPeakArea=max(BcPeakArea,[],2);
maxBcPeakArea= repmat(maxBcPeakArea,[1 b]);
normBcPeakArea=BcPeakArea./maxBcPeakArea;
```

```
xlswrite([currentfile 'RawPeakHeight'],[PeakMass RawPeakHeight]);
xlswrite([currentfile 'CiphergenPeakHeight'],[PeakMass CiphergenPeakHeight]);
xlswrite([currentfile 'CiphergenPeakArea'],[PeakMass CiphergenPeakArea]);
xlswrite([currentfile 'BcPeakHeight'],[PeakMass BcPeakHeight]);
xlswrite([currentfile 'BcPeakArea'],[PeakMass BcPeakArea]);
xlswrite([currentfile 'normBcPeakArea'],[PeakMass normBcPeakArea]);
```

```
save([currentfile
'Peaks'],'Mass','PeakMass','RawPeakHeight','CiphergenPeakHeight','BcPeakHeight','
CiphergenPeakArea','BcPeakArea','normBcPeakArea');
```

Nearest neighbour predictor

```
%Enter directory where the data is located
```

```
path=cd;
```

```
directory=find(path=='\');
```

```
currentfile=path(directory(end-1)+1:end);
```

```
currentfile(currentfile=='.')='_';
currentfile(currentfile=='\')='_';
```

```
load([currentfile 'Peaks']);
load([currentfile 'MetaData']);
```

```
MassString=char(num2str(PeakMass,'%1f'));
figure
h=glyphplot(normBcPeakArea,'Obslabels',MassString);
set(h(:,3),'FontSize',16);
```

```
Y = pdist(normBcPeakArea);
Z = linkage(Y);
```

```
figure
[H,T] = dendrogram(Z,'colorthreshold', 0,'labels',MassString,'orientation','right');
```

```

set(gca,'FontSize',14);
set(H,'LineWidth',2);
xlabel('Relationship (\muA)','FontSize',16)
ylabel('Protein ID (kDa/e)','FontSize',16)

TargetInd=13;

SquareDist=squareform(Y);
TargetDist=SquareDist(TargetInd,:);
TargetDist(TargetDist==0)=NaN;

[a NearestNeighbourInd]=min(TargetDist);
NearestNeighbour=PeakMass(NearestNeighbourInd)

ph=variables(Coombes et al. 2005)(variablelevels(:,3));
conc=variables(Chen et al. 2008)(variablelevels(:,5))/1000;

[a]=find(variablelevels(:,1)==1);
[b]=find(variablelevels(:,1)==2);
[c]=find(variablelevels(:,1)==3);

%Cation Exchange
if size(a)>0;

    figure
    subplot(1,2,1);
    [XI,YI] = meshgrid(unique(ph(a)),unique(conc(a)))
    [catph,catconc,catpeakarea] =
    griddata(ph(a),conc(a),normBcPeakArea(TargetInd,a),XI,YI,'v4');

    contourf(catph,catconc,catpeakarea,500,'EdgeColor','none')

    view(0,90)
    axis tight

    subplot(1,2,2);
    [XI,YI] = meshgrid(unique(ph(a)),unique(conc(a)))
    [catph,catconc,catpeakarea] =
    griddata(ph(a),conc(a),normBcPeakArea(NearestNeighbourInd,a),XI,YI,'v4');

    contourf(catph,catconc,catpeakarea,500,'EdgeColor','none')

    view(0,90)
    axis tight
end

```

```
if size(b)>0;

    figure
    subplot(1,2,1);
    [XI,YI] = meshgrid(unique(ph(b)),unique(conc(b)))
    [catph,catconc,catpeakarea] =
griddata(ph(b),conc(b),normBcPeakArea(TargetInd,b),XI,YI,'v4');

    contourf(catph,catconc,catpeakarea,500,'EdgeColor','none')

    view(0,90)
    axis tight

    subplot(1,2,2);
    [XI,YI] = meshgrid(unique(ph(b)),unique(conc(b)))
    [catph,catconc,catpeakarea] =
griddata(ph(b),conc(b),normBcPeakArea(NearestNeighbourInd,b),XI,YI,'v4');

    %colormap('default')
    contourf(catph,catconc,catpeakarea,500,'EdgeColor','none')

    view(0,90)
    axis tight
end

if size(c)>0;

    figure
    subplot(1,2,1);
    [XI,YI] = meshgrid(unique(ph(c)),unique(conc(c)))
    [catph,catconc,catpeakarea] =
griddata(ph(c),conc(c),normBcPeakArea(TargetInd,c),XI,YI,'v4');

    contourf(catph,catconc,catpeakarea,500,'EdgeColor','none')

    view(0,90)
    axis tight
    subplot(1,2,2);
    [XI,YI] = meshgrid(unique(ph(c)),unique(conc(c)))
    [catph,catconc,catpeakarea] =
griddata(ph(c),conc(c),normBcPeakArea(NearestNeighbourInd,c),XI,YI,'v4');

    contourf(catph,catconc,catpeakarea,500,'EdgeColor','none')

    view(0,90)
```

axis tight
end

REFERENCES

Ambler CM. 1961. Theory - the Fundamentals of Separation, Including Sharples Sigma Value for Predicting Equipment Performance. *Industrial and Engineering Chemistry* 53:430-433.

Annesley TM. 2003. Ion suppression in mass spectrometry. *Clin Chem* 49:1041-1044.

Arakawa T, Timasheff SN. 1984. Mechanism of Protein Salting in and Salting Out by Divalent-Cation Salts - Balance Between Hydration and Salt Binding. *Biochemistry* 23:5912-5923.

Asenjo JA, Andrews BA. 2004. Is there a rational method to purify proteins? From expert systems to proteomics. *J Mol Recognit* 17:236-247.

Barzaghi D, Isbister JD, Lauer KP, Born TL. 2004. Use of surface-enhanced laser desorption/ionization -time of flight to explore bacterial proteomes. *Proteomics* 4:2624-2628.

Berggren K, Wolf A, Asenjo JA, Andrews BA, Tjerneld F. 2002. The surface exposed amino acid residues of monomeric proteins determine the partitioning in aqueous two-phase systems. *Biochim Biophys Acta* 1596:253-268.

Berrill A, Ho SV, Bracewell DG. 2008. Ultra scale-down to define and improve the relationship between Flocculation and disc-stack centrifugation. *Biotechnology Progress* 24:426-431.

Berrill A, Ho SV, Bracewell DG. 2009. Product and contaminant measurement in bioprocess development by SELDI-MS. *Biotechnol Prog*.

Biddlecombe JG, Craig AV, Zhang H, Uddin S, Mulot S, Fish BC, Bracewell DG. 2007. Determining antibody stability: creation of solid-liquid interfacial effects within a high shear environment. *Biotechnol Prog* 23:1218-1222.

Bonet I, Rodríguez A, Grau R, García MM, Saez Y, Nowé A. 2008. Comparing Distance Measures with Visual Methods. *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg. p 90-99.

Bonnerjea J, Oh S, Hoare M, Dunnill P. 1986. Protein purification: The right step at the right time. *Biotechnology* 4:954-958.

Borhani DW, Rogers DP, Engler JA, Brouillette CG. 1997. Crystal structure of truncated human apolipoprotein A-I suggests a lipid-bound conformation. *Proc Natl Acad Sci U S A* 94:12291-12296.

Bowering LC, Bracewell DG, Kesharvarz-Moore E, Hoare M, Weir AN. 2002. Comparison of techniques for monitoring antibody fragment production in *E. coli* fermentation cultures. *Biotechnol Prog* 18:1431-1438.

Boychyn M, Doyle W, Bulmer M, More J, Hoare M. 2000. Laboratory scaledown of protein purification processes involving fractional precipitation and centrifugal recovery. *Biotechnol Bioeng* 69:1-10.

Boychyn M, Yim SS, Bulmer M, More J, Bracewell DG, Hoare M. 2004. Performance prediction of industrial centrifuges using scale-down models. *Bioprocess Biosyst Eng* 26:385-391.

Boychyn M, Yim SSS, Shamlou PA, Bulmer M, More J, Hoare A. 2001. Characterization of flow intensity in continuous centrifuges for the development of laboratory mimics. *Chemical Engineering Science* 56:4759-4770.

Brenac B, V, Schapman A, Santambien P, Britsch L. 2008. Fast purification process optimization using mixed-mode chromatography sorbents in pre-packed mini-columns. *J Chromatogr A* 1177:226-233.

Brenac V, Mouz N, Schapman A, Ravault V. 2006. Expression optimization and purification process development of an engineered soluble recombinant mouse linker of activation of T cells using surface enhanced laser desorption/ionization-mass spectrometry. *Protein Expr Purif* 47:533-541.

Brozkova K, Budinska E, Bouchal P, Hernychova L, Knoflickova D, Valik D, Vyzula R, Vojtesek B, Nenutil R. 2008. Surface-enhanced laser desorption/ionization time-of-flight proteomic profiling of breast carcinomas identifies clinicopathologically relevant groups of patients similar to previously defined clusters from cDNA expression. *Breast Cancer Res* 10:R48.

Caputo E, Moharram R, Martin BM. 2003. Methods for on-chip protein analysis. *Anal Biochem* 321:116-124.

Chen J, Cramer SM. 2007. Protein adsorption isotherm behavior in hydrophobic interaction chromatography. *J Chromatogr A* 1165:67-77.

Chen J, Luo Q, Breneman CM, Cramer SM. 2007. Classification of protein adsorption and recovery at low salt conditions in hydrophobic interaction chromatographic systems. *J Chromatogr A* 1139:236-246.

Chen J, Yang T, Cramer SM. 2008. Prediction of protein retention times in gradient hydrophobic interaction chromatographic systems. *J Chromatogr A* 1177:207-214.

Coombes KR, Tsavachidis S, Morris JS, Baggerly KA, Hung MC, Kuerer HM. 2005. Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform. *Proteomics* 5:4107-4117.

Cordingley HC, Roberts SL, Tooke P, Armitage JR, Lane PW, Wu W, Wildsmith SE. 2003. Multifactorial screening design and analysis of SELDI-TOF ProteinChip array optimization experiments. *Biotechniques* 34:364-373.

Cramer SM, Jayaraman G. 1993. Preparative chromatography in biotechnology. *Curr Opin Biotechnol* 4:217-225.

Cromwell ME, Hilario E, Jacobson F. 2006. Protein aggregation and bioprocessing. *AAPS J* 8:E572-E579.

Dijkstra M, Roelofsen H, Vonk RJ, Jansen RC. 2006. Peak quantification in surface-enhanced laser desorption/ionization by using mixture models. *Proteomics* 6:5106-5116.

Dijkstra M, Vonk RJ, Jansen RC. 2007. SELDI-TOF mass spectra: a view on sources of variation. *J Chromatogr B Analyt Technol Biomed Life Sci* 847:12-23.

Dyr JE, Suttner J. 1997. Separation used for purification of recombinant proteins. *J Chromatogr* 699:383-401.

Favre-Kontula L, Sattonnet-Roche P, Magnenat E, Proudfoot AEI, Boschert U, Xenarios L, Vilbois F, Antonsson B. 2008. Detection and identification of plasma proteins that bind GlialCAM using ProteinChip (TM) arrays, SELDI-TOF MS, and nano-LC MS/MS. *Proteomics* 8:378-388.

Fortis F, Guerrier L, Righetti PG, Antonioli P, Boschetti E. 2006. A new approach for the removal of protein impurities from purified biologicals using combinatorial solid-phase ligand libraries. *Electrophoresis* 27:3018-3027.

Garcia T, Cook G, Nosal R. 2008. PQLI key topics - Criticality, design space, and control strategy. *J Pharm Innov* 3:60-68.

Gasner LL, Wang DIC. 1970. Microbial Cell Recovery Enhancement Through Flocculation. *Biotechnology and Bioengineering* 12:873-&.

GE gel filtration principles and methods. 2003. GE gel filtration principles and methods.

GE HIC principle and methods. 2003. GE HIC principle and methods.

GE protein purification. 2003. GE protein purification.

Groep ME, Gregory ME, Kershenbaum LS, Bogle IDL. 1997. Performance Modelling and Simulation of Biochemical Process Sequences with Interacting Unit Operations. *Biotechnol Bioeng* 67:300-311.

Gu Z, Glatz CE. 2007. A method for three-dimensional protein characterization and its application to a complex plant (corn) extract. *Biotechnol Bioeng* 97:1158-1169.

Guerrier L, Lomas L, Boschetti E. 2005. A simplified monobuffer multidimensional chromatography for high-throughput proteome fractionation. *J Chromatogr A* 1073:25-33.

Guerrier L, Lomas L, Boschetti E. 2007. A new general approach to purify proteins from complex mixtures. *J Chromatogr A* 1156:188-195.

Humphreys DP, Heywood SP, King LM, Bowering LC, Turner JP, Lane SE. 2004. Engineering of *Escherichia coli* to improve the purification of periplasmic Fab' fragments: changing the pI of the chromosomally encoded PhoS/PstS protein. *Protein Expr Purif* 37:109-118.

Hunter AK, Wang X, Suda EJ, Herberg JT, Shell RE, Thomas KE, Dufield RL, Gustafson ME, Mozier NM, Ho SV. 2009. Separation of Product Associating *E. coli* Host Cell Proteins OppA and DppA from Recombinant Apolipoprotein A-I-milano in an Industrial HIC Unit Operation. *Biotechnology Progress* 25:446-453.

Hutchens TW, Yip TT. 1993. New Desorption Strategies for the Mass-Spectrometric Analysis of Macromolecules. *Rapid Communications in Mass Spectrometry* 7:576-580.

Hutchinson N, Bingham N, Murrell N, Farid S, Hoare M. 2006. Shear stress analysis of mammalian cell suspensions for prediction of industrial centrifugation and its verification. *Biotechnology and Bioengineering* 95:483-491.

Jeffries N. 2005. Algorithms for alignment of mass spectrometry proteomic data. *Bioinformatics* 21:3066-3073.

Jenkins RE, Pennington SR. 2001. Arrays for protein expression profiling: towards a viable alternative to two-dimensional gel electrophoresis? *Proteomics* 1:13-29.

Karpievitch YV, Hill EG, Smolka AJ, Morris JS, Coombes KR, Baggerly KA, Almeida JS. 2007. PrepMS: TOF MS data graphical preprocessing tool. *Bioinformatics* 23:264-265.

Kim JS, Akeprathumchai S, Wickramasinghe SR. 2001. Flocculation to enhance microfiltration. *Journal of Membrane Science* 182:161-172.

Kumar N, Maurya P, Gammell P, Dowling P, Clynes M, Meleady P. 2008. Proteomic profiling of secreted proteins from CHO cells using Surface-Enhanced Laser desorption ionization time-of-flight mass spectrometry. *Biotechnol Prog* 24:273-278.

Ladiwala A, Rege K, Breneman CM, Cramer SM. 2005. A priori prediction of adsorption isotherm parameters and chromatographic behavior in ion-exchange systems. *Proc Natl Acad Sci U S A* 102:11710-11715.

Ladiwala A, Xia F, Luo Q, Breneman CM, Cramer SM. 2006. Investigation of protein retention and selectivity in HIC systems using quantitative structure retention relationship models. *Biotechnol Bioeng* 93:836-850.

Lienqueo ME, Asenjo JA. 2000. Use of expert systems for the synthesis of downstream protein processes. *Computers & Chemical Engineering* 24:2339-2350.

Lienqueo ME, Mahn A, Asenjo JA. 2002. Mathematical correlations for predicting protein retention times in hydrophobic interaction chromatography. *J Chromatogr A* 978:71-79.

Lienqueo ME, Mahn A, Navarro G, Salgado JC, Perez-Acle T, Rapaport I, Asenjo JA. 2006. New approaches for predicting protein retention time in hydrophobic interaction chromatography. *J Mol Recognit* 19:260-269.

Lienqueo ME, Mahn A, Salgado JC, Asenjo JA. 2007. Current insights on protein behaviour in hydrophobic interaction chromatography. *J Chromatogr B Analyt Technol Biomed Life Sci* 849:53-68.

Lienqueo ME, Mahn A, Vasquez L, Asenjo JA. 2003. Methodology for predicting the separation of proteins by hydrophobic interaction chromatography and its application to a cell extract. *J Chromatogr A* 1009:189-196.

Lodish H, Berk A, Zipursky ZL, Matsudaira P, Baltimore D, Darnell JE. 1999. *Molecular Cell Biology*. W.H. Freeman & Company.

Mahn A, Asenjo JA. 2005. Prediction of protein retention in hydrophobic interaction chromatography. *Biotechnol Adv* 23:359-368.

Mahn A, Lienqueo ME, Asenjo JA. 2007. Optimal operation conditions for protein separation in hydrophobic interaction chromatography. *J Chromatogr B Analyt Technol Biomed Life Sci* 849:236-242.

Mahn A, Lienqueo ME, Asenjo JA. 2004. Effect of surface hydrophobicity distribution on retention of ribonucleases in hydrophobic interaction chromatography. *J Chromatogr A* 1043:47-55.

Mahn A, Zapata-Torres G, Asenjo JA. 2005. A theory of protein-resin interaction in hydrophobic interaction chromatography. *J Chromatogr A* 1066:81-88.

Malmquist G, Nilsson UH, Norrman M, Skarp U, Stromgren M, Carredano E. 2006. Electrostatic calculations and quantitative protein retention models for ion exchange chromatography. *J Chromatogr A* 1115:164-186.

Maybury JP, Hoare M, Dunnill P. 2000. The use of laboratory centrifugation studies to predict performance of industrial machines: Studies of shear-insensitive and shear-sensitive materials. *Biotechnology and Bioengineering* 67:265-273.

Maybury JP, Mannweiler K, Titchener-Hooker NJ, Hoare M, Dunnill P. 1998. The performance of a scaled down industrial disc stack centrifuge with a reduced feed material requirement. *Bioprocess Engineering* 18:191-199.

McGregor WC, Finn RK. 1969. Factors affecting the flocculation of bacteria by chemical additives. *Biotechnol Bioeng* 11:127-138.

Merchant M, Weinberger SR. 2000. Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. *Electrophoresis* 21:1164-1177.

Mine Y. 1995. Recent Advances in the Understanding of Egg-White Protein Functionality. *Trends in Food Science & Technology* 6:225-232.

Neal G, Christie J, Keshavarz-Moore E, Shamlou PA. 2003. Ultra scale-down approach for the prediction of full-scale recovery of ovine polyclonal immunoglobulins used in the manufacture of snake venom-specific Fab fragment. *Biotechnol Bioeng* 81:149-157.

Ou K, Seow TK, Liang RC, Ong SE, Chung MC. 2001. Proteome analysis of a human hepatocellular carcinoma cell line, HCC-M: an update. *Electrophoresis* 22:2804-2811.

Panicker G, Lee DR, Unger ER. 2009. Optimization of SELDI-TOF protein profiling for analysis of cervical mucous. *J Proteomics* 71:637-646.

Park JT, Bradbury L, Kragl FJ, Lukens DC, Valdes JJ. 2006. Rapid optimization of antitoxin antibody fragment production by an integral approach utilizing RC-SELDI mass spectrometry and statistical design. *Biotechnol Prog* 22:233-240.

Parris N, Baginski MA. 1991. A Rapid Method for the Determination of Whey-Protein Denaturation. *Journal of Dairy Science* 74:58-64.

Rege K, Pepsin M, Falcon B, Steele L, Heng M. 2006. High-throughput process development for recombinant protein purification. *Biotechnol Bioeng* 93:618-630.

Reynolds T, Boychyn M, Sanderson T, Bulmer M, More J, Hoare M. 2003. Scale-down of continuous filtration for rapid bioprocess design: Recovery and dewatering of protein precipitate suspensions. *Biotechnol Bioeng* 83:454-464.

Richardson JF, Zaki WN. 1997. Sedimentation and fluidisation: part 1. *Chemical Engineering Research & Design* 75:S82-S99.

Richardson P, Hoare M, Dunnill P. 1990. A new biochemical engineering approach to the fractional precipitation of proteins. *Biotechnol Bioeng* 36:354-366.

Salgado JC, Andrews BA, Ortuzar MF, Asenjo JA. 2008. Prediction of the partitioning behaviour of proteins in aqueous two-phase systems using only their amino acid composition. *J Chromatogr A* 1178:134-144.

Salgado JC, Rapaport I, Asenjo JA. 2005a. Is it possible to predict the average surface hydrophobicity of a protein using only its amino acid composition? *J Chromatogr A* 1075:133-143.

Salgado JC, Rapaport I, Asenjo JA. 2005b. Prediction of retention times of proteins in hydrophobic interaction chromatography using only their amino acid composition. *J Chromatogr A* 1098:44-54.

Salgado JC, Rapaport I, Asenjo JA. 2006a. Predicting the behaviour of proteins in hydrophobic interaction chromatography. 1: Using the hydrophobic imbalance (HI) to describe their surface amino acid distribution. *J Chromatogr A* 1107:110-119.

Salgado JC, Rapaport I, Asenjo JA. 2006b. Predicting the behaviour of proteins in hydrophobic interaction chromatography. 2. Using a statistical description of their surface amino acid distribution. *J Chromatogr A* 1107:120-129.

Salte H, King JP, Titchener-Hooker N. 2005. Centrifuge selection for recovery of high solids density cell broths and visualization of separation performance via windows of operation. *Abstracts of Papers of the American Chemical Society* 229:U179.

Semmes OJ, Feng Z, Adam BL, Banez LL, Bigbee WL, Campos D, Cazares LH, Chan DW, Grizzle WE, Izbicka E, Kagan J, Malik G, McLerran D, Moul JW, Partin A, Prasanna P, Rosenzweig J, Sokoll LJ, Srivastava S, Srivastava S, Thompson I, Welsh MJ, White N, Winget M, Yasui Y, Zhang Z, Zhu L. 2005. Evaluation of serum protein profiling by surface-enhanced laser desorption/ionization time-of-flight mass spectrometry for the detection of prostate cancer: I. Assessment of platform reproducibility. *Clin Chem* 51:102-112.

Shan JG, Xia J, Guo YX, Zhang XQ. 1996. Flocculation of cell, cell debris and soluble protein with methacryloyloxyethyl trimethylammonium chloride-acrylonitrile copolymer. *Journal of Biotechnology* 49:173-178.

Suurkuusk M, Hallen D. 1999. Apolipoprotein A-I(Milano) unfolds via an intermediate state as studied by differential scanning calorimetry and circular dichroism. *Eur J Biochem* 264:183-190.

Tang N, Tornatore P, Weinberger SR. 2004. Current developments in SELDI affinity technology. *Mass Spectrom Rev* 23:34-44.

Thommes J, Etzel M. 2007. Alternatives to chromatographic separations. *Biotechnol Prog* 23:42-45.

Tustian AD, Salte H, Willoughby NA, Hassan I, Rose MH, Baganz F, Hoare M, Titchener-Hooker NJ. 2007. Adapted ultra scale-down approach for predicting the centrifugal separation behavior of high cell density cultures. *Biotechnology Progress* 23:1404-1410.

Vachier MC, Piot M, Awade AC. 1995. Isolation of hen egg white lysozyme, ovotransferrin and ovalbumin, using a quaternary ammonium bound to a highly crosslinked agarose matrix. *J Chromatogr B Biomed Appl* 664:201-210.

Vitello LB, Scanu AM. 1976. Studies on human serum high-density lipoproteins. Self-association of human serum apolipoprotein A-II in aqueous solutions. *Biochemistry* 15:1161-1165.

Wall DB, Parus SJ, Lubman DM. 2002. Three-dimensional protein map according to pI, hydrophobicity and molecular mass. *J Chromatogr B Analyt Technol Biomed Life Sci* 774:53-58.

Weinberger SR, Boschetti E, Santambien P, Brenac V. 2002a. Surface-enhanced laser desorption-ionization retentate chromatography mass spectrometry (SELDI-RC-MS): a new method for rapid development of process chromatography conditions. *J Chromatogr B Analyt Technol Biomed Life Sci* 782:307-316.

Weinberger SR, Dalmasso EA, Fung ET. 2002b. Current achievements using ProteinChip Array technology. *Curr Opin Chem Biol* 6:86-91.

Weinberger SR, Viner RI, Ho P. 2002c. Tagless extraction-retentate chromatography: a new global protein digestion strategy for monitoring differential protein expression. *Electrophoresis* 23:3182-3192.

Wenger MD, Dephillips P, Price CE, Bracewell DG. 2007. An automated microscale chromatographic purification of virus-like particles as a strategy for process development. *Biotechnol Appl Biochem* 47:131-139.

Wierling PS, Bogumil R, Knieps-Grunhagen E, Hubbuch J. 2007. High-throughput screening of packed-bed chromatography coupled with SELDI-TOF MS analysis: monoclonal antibodies versus host cell protein. *Biotechnol Bioeng* 98:440-450.

Willoughby N, Martin P, Titchener-Hooker N. 2004. Extreme scale-down of expanded bed adsorption: Purification of an antibody fragment directly from recombinant *E. coli* culture. *Biotechnol Bioeng* 87:641-647.

Woolley JF, Al Rubeai M. 2009. The application of SELDI-TOF mass spectrometry to mammalian cell culture. *Biotechnol Adv* 27:177-184.

Xu L, Glatz CE. 2009. Predicting protein retention time in ion-exchange chromatography based on three-dimensional protein characterization. *J Chromatogr A* 1216:274-280.

Xu X, Lenhoff AM. 2008. A predictive approach to correlating protein adsorption isotherms on ion-exchange media. *J Phys Chem B* 112:1028-1040.

Yang T, Breneman CM, Cramer SM. 2007a. Investigation of multi-modal high-salt binding ion-exchange chromatography using quantitative structure-property relationship modeling. *J Chromatogr A* 1175:96-105.

Yang T, Sundling MC, Freed AS, Breneman CM, Cramer SM. 2007b. Prediction of pH-dependent chromatographic behavior in ion-exchange systems. *Anal Chem* 79:8927-8939.