Increasing the efficiency of antibody purification process by high throughput technology and intelligent design of experiment

A thesis submitted to University College London

for the degree of

DOCTOR OF ENGINEERING

Muazzam Ali Khan

2018

The Advanced Centre for Biochemical Engineering

Department of Biochemical Engineering

University College London

London UK

Declaration

I, Muazzam A. Khan 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.

Sign: ……

Date: ……

Acknowledgements

First I would like to be thankful to God for providing me with the provision to carry out this endeavour. I would like to thank my family who have stood by me and at times cajoled me to carry on when I had become preoccupied with life's challenges. I hope that the work presented is of use to the field of biochemical engineering and contributes to accelerating the delivery of life saving therapies to patients.

I am indebted to my primary supervisor Dr. Yuhong Zhou for her unwavering support and guidance throughout the duration of this EngD programme. I would also like to express my gratitude to my industrial supervisors, Dr. Dietmar Lang and Dr. Allen Lee for their intuition and hospitality during my time with them at Lonza. I also would like to thank all my colleagues in the department of Biochemical Engineering and the Purification development team at Lonza Biologics, Slough. The financial support provided by the Engineering and Physical Sciences Research Council (EPRSC) and sponsorship from Lonza Biologics are gratefully acknowledged without whom the project would not have been possible. I dedicate this thesis to my son Musa, and my late grandparents Sardar Mumtaz and Nazir Begum who are dearly missed.

Abstract

Design of experiments (DoE) is used in process development to optimise the operating conditions of unit operations in a cost-effective and time-saving manner. Along with high throughput technologies, the modern high throughput process development lab can turnover a tremendous amount of data with minimal feedstock. These benefits are most useful when applied to the purification bottleneck, which accounts for up to 80% of the total process operating costs. However due to complexities of biochemical reactions and the large number interacting factors in unit operations (which usually cross interact with each other), even carefully planned DoE experiments on high throughput platforms can become difficult to manage and/or not provide useful information.

This thesis examines the simplex search method and develops a set of protocols for use of the search method in combination with traditional DoE experimental design protocols. It is that is demonstrated in the developed in chapter 3 whilst also optimising a ammonium sulphate based precipitation step of an industrially relevant feedstock. Comparisons were drawn between a high resolution brute force study, a response surface DoE, the simplex method and then a combination of DoE and the simplex method. Various strategies were demonstrated that get the most out of the simplex method and mitigate against potential pitfalls. The precipitation step was optimised for yield and purity over the 3 factors, pH, ammonium sulphate concentration and initial MAb concentration and the results showed the simplex method was capable of rapidly identifying the optimum conditions in a very large 3 factor design space on an average of 18 experiments. The expansive study not only served as a testing ground for the methods comparison but demonstrated precipitation as a high throughput, low cost substitute for the expensive Protein A step.

The DoE –simplex search protocols are then refined in two complex case studies in chapter 4, a PEG precipitation primary capture step and an ammonium sulphate precipitation and

centrifugation sequence. The five factor precipitation and centrifugation sequence was especially complicated and utilised ultrascale down models provide accurate scale up data. This involved calibrating an acoustic device to provide shear treatment to the precipitate pre-centrifugation and using jet mixing equations to correlate precipitate conditioning between the TTecan robot's tips and an impeller in a stirred tank. The techniques developed were all applicable to microscale and high throughput. In both instances, the combined DoE-simplex approach retuned superior results both in terms of experimental savings and generating information-rich data from the final local regions DoE around the simplex located optimums.

A microscale chromatography protocol was developed on the Tecan liquid handling robot and demonstrated on screening work with different Protein A and cation exchange media. The caveats encountered when creating the running methods and the analytical methods supporting it for the Atoll robocolumns were highlighted and mitigation solutions implemented. The automated microscale Protein A method was successfully scaled up 50x from a 200 µL robocolumn to a conventional 10 mL labscale column. After selecting a cation exchange resin for developing an aggregate removal step, the DoE-simplex methodology was applied to an antibody product with an extremely high aggregate level and a comparison optimisation was made with a central composite design DoE. The difficult four factor design space overwhelmed the DoE and having used more experiment numbers than the DoE-simplex methodology, only went as far to show the high levels of curvature in the system and offer a poor prediction of the surface. The DoE-simplex methodology was able to provide a general model of the whole surface from the DoE, locate the optimum with the simplex in fewer experiment numbers. This subsequently allowed a local DoE to be applied to the optimum region to determine a robust operating range for the cation exchange step.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

Nomenclature and abbreviations

Abbreviations	Description
AS	Ammonium sulphate
AU	Absorbance units
BCA	Bicinchoninic acid
CCD	Central composite design
CCS	Cell culture supernatant
CEX	Cation exchange chromatography
CFD	Computational fluid dynamics
CHO	Chinese hamster ovary cell
COGs	Cost of goods
CPP	Critical process parameters
CQA	Critical quality attributes
CR	Combined response
CVs	Column volumes
DBC	Dynamic binding capacity
DF	Dilution factor
DoE	Design of Experiments
DSP	Downstream processing
ELISA	Enzyme-linked immunosorbent assay

Abbreviations	Description
FDA	Food and Drug Administration
GMP	Good Manufacturing Practice
HCPs	Host cell proteins
HIC	Hydrophobic interaction chromatography
HPLC	High pressure liquid chromatography
HTPD	High throughput process development
HTS	High throughput screening
IMAC	Immobilised metal affinity chromatography
LHR	Liquid handling robot
LPA	Leached protein A
MAb	Monoclonal antibody
OFAT	One-factor-at-a-time
PAT	Process analytical technology
PLW	Post load wash
QbD	Quality by Design
SEC	Size exclusion chromatography
USD	Ultra-scale down
USP	Upstream processing

1. Introduction

1.1 Perspective and motivation

In today's industry, pipelines bursting with drug candidates have pushed companies to invest in novel technologies and methods to find quicker and less expensive routes through process development. Developing and validating a good process requires time and finances that are often restricted to unproven candidates during the early phases of drug development when risk of failure is high. The delay to market of a would-be blockbuster not only gambles with first mover advantage but risks losing over $3 million per day of potential revenue (Subramaniam, 2003).

The typical purification development effort begins in the laboratory where experiments are designed around scarce upstream material and the unit operations and their ranges are defined to put together a purification process. As the candidate drug shows success in pivotal studies, the manufacturing process is further optimised and developed with manufacturing and final purity constraints in mind. Processes defined in the lab will be transferred to larger scales with pilot scale engineering runs validating all the lab work that has gone before it and the initiation of GMP grade manufacture. Due to the competitive nature of the industry, timelines are short for process development and representative material is scarce. To ensure an optimal manufacturing process, the process development work must be done quickly and accurately.

Process design and optimisation seeks to identify optimum and robust operating conditions for the unit operation by (i) gathering process insight, (ii) establishing critical operating ranges (iii) and satisfying all manufacturing constraints on time and resource. Process development has in recent years adopted approaches such as Design of Experiment (DoE) and quality by design (QbD) into their process optimisation efforts for the successful

validation of their products and manufacturing processes. Regulatory bodies like the FDA and EMA have set-up initiatives such as process analytical technology (PAT) that encourages the use of on-line process monitoring and modelling methods for a deeper understanding of the manufacturing process and improved product quality (FDA 2006, 2009; Mandenius and Brundin, 2008). Bioprocesses are often governed by complex functions, involving a large number of biological and engineering parameters. Encompassing every variable into the study can be experimentally taxing and of uncertain benefit without the proper experimental planning. Apart from managing sample numbers the solution should also provide the optimum operating conditions and good predictive capability of the bioprocess system. A logical and systematic development approach towards bioprocess optimisation holds great potential to address the high costs of downstream processing as well as accelerating process development so the drug can reach the patient as quickly as possible.

1.2 Biopharmaceutical downstream processing

By far the biggest group of biopharmaceutical products and key growth market is recombinant human antibody based therapy (Li et al., 2005). Of the antibody-based products, the monoclonal antibodies (MAbs) are expected to dominate market share in the pharmaceutical sector and maintain a healthy growth rate well into the next decade (Farid, 2007). The runaway success of first generation MAbs (safety and efficacy) has meant there are many MAb products in the development pipeline (Gagnon, 2006). Great strides in upstream technology have been made resulting in successful bioreactor scale up to 20,000 L with product titres of 5 – 15 g/L achievable (Challener, 2016). However the downstream purification side has not been able to keep up and is repeatedly highlighted as the critical limiting step in process development (Aldington and Bonnerjea, 2007; Birch and Racher, 2006; Rito-Palomares, 2008; Titchener-Hooker et al., 2001). Due to this productivity mismatch and the high cost of Protein A chromatography media, much of the manufacturing cost is attributed to the purification process. For some products, as much as

80% of total manufacturing cost has been reported (Lowe et al., 2001). In fact current trends in upstream process development have moved aware from maximising product titre (which is the primary factor that determines the cost of goods) and towards improving purity within the bioreactor (Challener, 2016). The higher level of impurities associated with high titre fermentations increase purification costs and risk campaign failure if the impurities are not cleared to specification (Cromwell et al., 2006).

1.3 The platform process

Developing and optimising a purification process is costly and time intensive due to the number of steps and variety of purification technologies involved. Time and investment are at a premium for candidate drugs of which only 1 in 5 make it to market (Steinmeyer and McCormick, 2008). The similarities between antibody based products however can be exploited to utilise platform processes. A platform process attempts to impose a 'one size fits all' approach that can drastically accelerate the process development effort and reduce the resources assigned to projects (Gagnon, 2006). The purification platform for Fc based products is centred around Protein A chromatography, which is unique in offering high purity and yield using virtually generic conditions. This is followed by a viral inactivation step, two further chromatography operations (usually anion exchange followed by cation exchange) before nanofiltration and formulation. Having two orthogonal chromatography and viral reduction steps are mandated by regulatory bodies such as the FDA. The latter steps (ion exchange chromatography, nanofiltration and formulation) all require some level of optimisation specific to the product and take up most of the development effort. The polishing steps receive the most attention as the chemistries employed by the step use subtle differences between product and the impurity to effect the separation. This sequence is shown in Figure 1.1 and has been succesful in being implemented for various antibody products. For the polishing steps there are multiple chromatography types available but the main ones being cation and anion exchange chromatography, hydrophobic interaction

chromatography (HIC), mixed mode and hydroxyapatite chromatography. Each type has a different capability in removing certain types of impurities.

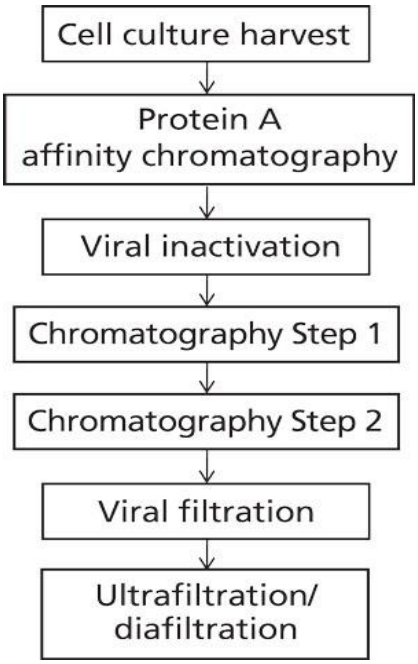


Figure 1.1: A generic downstream process for MAb purification

1.4 Main impurities

The three main impurities are aggregates, host cell proteins (HCPs) and host cell DNA whilst a 4th main impurity, leached protein A (LPA) is much less of a concern with modern protein A columns (Lintern et al., 2016). Aggregates are covalently bonded forms of the product (intact or denatured), usually dimers and trimers, which can lead to immunogenic responses in patients, essentially raising neutralising antibodies during therapy as well as increasing the occurrence of embolisms (Gamble, 1966; Gagnon, 1996; Fotiou et al., 2009). Cation exchange and HIC can remove these species quite well based on differences in protein surface properties. LPA is a documented immunotoxin and adjuvant protein with the potential to promote neutralising antibodies in the host, compromising the therapy

(Gagnon, 1996). It is one of the reasons why Protein A can only be used a limited number of times. HCPs are heterogeneous in biological properties making them difficult to remove so are removed over multiple unit operations although Protein A chromatography will do most of the work (Tarrant et al., 2012). DNA impurities are removed by anion exchange chromatography (binding the negatively charged nucleic acids). Impurity specification and acceptable levels are defined early on in the project and based on clinical data. Some high dose therapies and blood-brain barrier products will have very tight specifications on impurity levels (due to concentration effects), and the removal of trace impurities could decide which process to implement. Product aggregates are the main impurity and where most process development work is needed (Gagnon, 2006). They are formed during fermentation, storage, shipment, exposure to low pH buffers (such as in the viral inactivation step) however the majority are formed with the product during culture (Dintzis et al., 1989; Bachmann et al., 1993). Feed streams are typically 5-15 % for most antibodies and the aim is to typically reduce them to less than 1 %.

Table 1.1: General characteristics of the three main polishing step options

Chromatography	**Cation exchange**	**Hydrophobic interaction**	**Hydroxyapatite**
Capacity	+++++	+++	+++
Flow rates	+++++	+++	+++
Aggregate removal	++++	++++	+++++
LPA removal	++++	+	+++++
Optimisation work	Complex	Complex	Complex
Other	DNA/Endotoxins bind with product at loading conditions	High salt usage, costly disposal of salts, poor recovery	Unstable below pH 6.5, limited to phosphate buffers, special packing requirements

Aggregates are the key critical quality attribute (CQA) of a new drug product, which must be minimised or have data-supported justification for the set aggregate specification

(Cordoba-Rodriguez, 2008; FDA, 2013). Aggregate content is analysed by the tediously slow, size exclusion chromatography HPLC method, which is one of the few validated methods for lot release. The main chromatography options aggregate removal are cation exchange, hydrophobic interaction and hydroxyapatite chromatographies. Their most common features listed in Table 1.1. As well as aggregates, the polishing step is also useful in removing DNA, LPA and HCPs (Rosenburg and Worobec, 2005).

For some aggregate species and where the level is high, the task of removing aggregates is spread across two column steps and if antibody fragments require removal then this will be done in the polishing chromatography steps as well. Unlike affinity chromatography that is operated mostly under generic conditions, the polishing steps require considerable work. This is despite the harmonisation of the post-protein A feed that has very small stream variation across a range of MAb products. Process development at lab scale will utilise 'long and thin' columns (with volumes from 1mL – 50mL), which are usually packed by the scientist in the lab before use (Kelley et al., 2008). Lab chromatography systems will be used to identify the ideal chromatography media, and then optimise the operating conditions to achieve the best yield and purity. These experiments are very labour intensive, have long time constants and multiple factors to explore such as binding conditions (pH, salt concentration and flow rate) eluting conditions (pH, salt concentration and flow rate), loading capacity, media type, chromatography mode and stage length, buffer types etc. (Rege et al., 2006; Bergander et al., 2008; Coffman et al., 2008; Kelley et al., 2008).

The optimum conditions are variable even for molecularly similar MAb products. Where alternatives to Protein A are used prior to the step, it would require considerable rethinking of sequence selection as well as process conditions and operation (Ghose et al., 2008; Conley et al., 2011; Gagnon, 2005). Therefore a sub-optimal solution carries considerable loss as after affinity chromatography, the aggregate removal step usually has the highest operating cost (Chhatre and Titchener-Hooker, 2008).

Of the three intermediate column options listed in Table 1, cation exchange chromatography is the most widely used due to its high binding capacity (>40g/L) allowing whole batches of antibody to be purified in a single cycle of a reasonably sized column (Yigzaw et al., 2009). Hydrophobic-interaction chromatography (HIC) processing requires managing high salt concentration feed streams and requires salt disposal so it is not as popular however where a salt based operation such as precipitation precedes the step the decision for selecting HIC would certainly be easier to make. Hydroxyapatite chromatography is also very adept at removing aggregates, LPA and HCPs however it has operating process restrictions where the ligand-protein interaction is unstable under pH 6.5 and requires a special packing regime for its high density columns making it a costly option (Ghosh and Wang, 2006; Li et al., 2005).

For first generation MAbs products the platform performed very well with only subtle tweaks required to fit a wide range of products. However today's high titre and increasingly non-MAb products are stressing the platform by requiring extensive process development or very large columns and filters to deal with the greater material and impurity levels. Protein A chromatography has fast become the limiting step in the platform due to its high cost, limited binding capacity and slow flow rate (Ghose et al., 2007). Furthermore Protein A can only be used for Fc containing products (although a minority), and columns used for clinical trials are never used to full capacity as new columns must be used for every run making their use quite expensive (after clinical trials, for commercial supply columns and filters are re-used as the re-use number has been validated by then) . Consequently, alternatives to Protein A have been the subject of much industrial and academic interest of late with PEG precipitation, crystallisation two phase extraction and mixed mode chromatography (Przybycien et al., 2004; Thommes and Etzel, 2007; Knevelman et al., 2009). None of these however offer the high selectivity, yield and generic operation provided by Protein A chromatography, despite all being less expensive to run and turnover a higher throughput. Ion exchange chromatography for capture was once the norm but compared with affinity methods HCPs (host cell proteins) and DNA are not cleared

sufficiently. The operating conditions are also non-transferable between products despite the technique being higher throughput, inexpensive and robust (it is still used where affinity methods are unavailable, especially fragment antibodies purification). For most alternatives, the operating conditions will have to be defined and optimised virtually case by case and this will need to utilise the latest intensive, rapid downstream process development techniques to remain a competitive and realistic option (Nfor et al., 2008).

1.5 Precipitation as a capture step

Precipitation is a key process in the manufacture of chemicals and especially pharmaceuticals (Englard and Seifter, 1990). It involves the mixing of two fluid streams to form a relatively pure solid product. These solids may be removed from the bulk fluid by a solid-liquid separation step such as filtration or centrifugation. Precipitation is most effective after having removed large solids (i.e. cell debris) from the processing stream. As described in the previous section, the purification of proteins is dominated by chromatography. The target protein is captured in the equilibrated phase where impurities pass unhindered; and then eluted by a change in the mobile phase i.e. buffer concentration or pH (Jungbauer, 2005). The drawback to this operation is the high column inventory costs in terms of initial capex for packing and frequent replacements, especially with affinity columns (Richardson et al., 1990). Synthetic ligands as a replacement to affinity steps can be used to lower downstream costs, however further savings are limited (Farid, 2006). Although not as specific as high-resolution chromatography, precipitation allows low cost and high yield operation on much cruder, unfiltered feeds. As a primary capture step it can remove major impurities (and particulates) and reduce process volumes whilst retaining a high product yield (Stavrinides et al., 1993; Bonnerjea et al., 1986). Thus a refined process stream can be sent for column processing for a reduced protein load can greatly reduce inventory costs and extend resin lifetime (Janson and Ryden, 1997).

Protein precipitation uses the differences between the solubilities of individual proteins to separate them from a crude mixture. It is low cost, high yield, and simple to use and scale up (Stavrinides et al., 1993; Russo et al., 1983). Its applications in biotechnology is best known for its role in human plasma fractionation (Kistler and Friedli, 1980; Cohn et al., 1946). PEG precipitation has been demonstrated as a primary capture step for monoclonal antibodies and achieved high yields and good purification however it proved difficult to recover the solid phase product from the high viscosity of concentrated PEG (Knevelman et al., 2009). Salt based precipitation has long been used in purifying human blood plasma products so its industrial use is well established. Cohn's law and salting-out theory (see section below) is a simplified model of protein precipitation behaviour, however the difficulty in obtaining the necessary thermodynamic parameters has meant process development has always been empirical (Arakawa and Timasheff, 1985). Key precipitation factors include the type of precipitant and its concentration, buffer type and pH, temperature, feed characteristics such as conductivity and protein concentration, of which all affect the solubility of individual proteins (Stavrinides et al., 1993). It is historic and widely used for many applications in the food, chemicals, and pharmaceutical industries mainly due to ease of operation, scale up and high recoveries (Thommes and Etzel, 2007). Lyotropic salts (those that promote hydrophobic interactions between dissolved species) such as ammonium sulphate are the most popular precipitating agents as they can achieve high recoveries from complex mixtures without any danger of protein denaturation or irreversible aggregation (Cheng et al., 2006; Cohn et al., 1946; Foster et al., 1976).

Precipitation induces reversible protein aggregation until they become large enough to be insoluble;, it is fully reversible and safe for the product (Walsh and Headon, 1994). Solid precipitate is then separated by centrifugation or filtration. Fractional precipitation is the most common form of this operation due to its selective ability to precipitate contaminants and proteins according to their solubilities such as in double cut precipitation (Richardson et al., 1990). Precipitant is first added to a lower concentration to precipitate and separate less soluble impurities (larger particles often removed in first cut) followed by higher a concentration to recover the desired protein (second cut), leaving the more soluble

impurities in solution. Temperature and pH can also be adjusted to enhance yield and purity.

1.5.1 Types of precipitation

Protein precipitation works by interfering with solute-solvent interactions and facilitating solute-solute forces (hydrophobic interactions to promote temporary aggregation). This phase change is neatly explained in the Gibb's free energy equation as the solute-solvent system strives to maintain high entropy so rather than opening a cavity around hydrophobic solute particles it groups them together minimising entropic losses (Horvath et al., 1976). It is generally accepted that protein solubility is a property of its molecular surface groups and a surrounding hydration layer (colloid model). Electrostatic groups keep the protein soluble whereas hydrophobic patches (which are usually tucked up away from the solvent) are induced to form weak bonds between neighbouring molecules resulting in aggregation and precipitation (Stavrinides et al., 1993).

All forms of precipitation takes place in the liquid phase from the mixing of two streams, the dissolved protein and a precipitant reagent. The reaction creates a high level of supersaturation from which the protein undergoes phase change and is forced out of solution. Precipitants come in different classes characterised by their precipitation mechanisms as described in section 3.2.2. Salts such as ammonium sulphate initiate salting-out, metal ions act by charge neutralisation, acids work by isoelectric precipitation and solvents reduce the dielectric constant. Different precipitants also require different amounts relative to each other (see Figure 1.2).

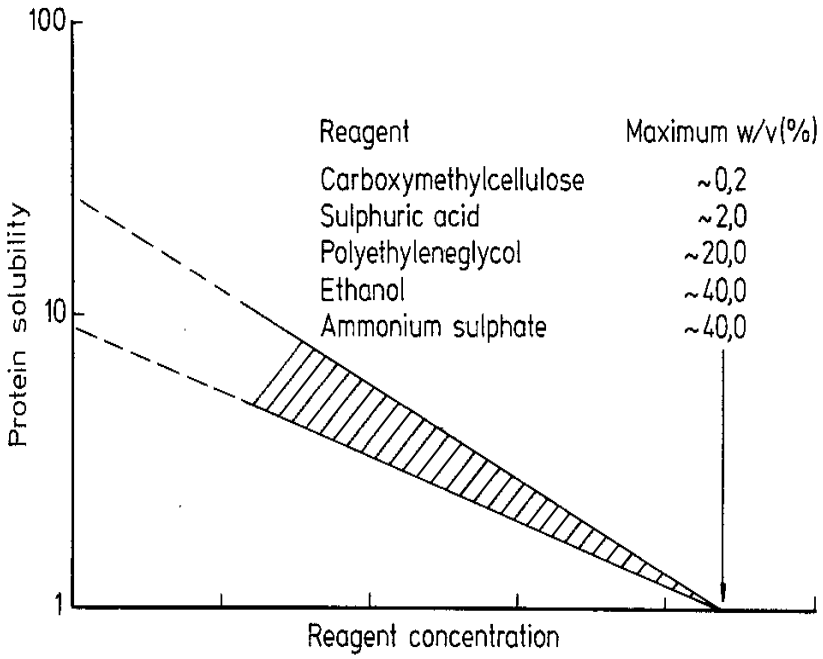


Figure 1.2: Comparative precipitant duties (equivalent quantities required) (Bell et al., 1983)

The Cohn expression is a widely used semi-empirical correlation for explaining salting-out (hydrophobic) and the effect of pH (isoelectric) on protein solubility:

$$\text{Log } S = \beta - K\, I, \qquad (1.1)$$

where S is the solubility ($g_{solute}/100g_{water}$); I, the ionic strength of the salt (M); β is a function of pH and temperature, essentially the solubility in a pure aqueous solution (extrapolated to zero ionic strength) and K, the salting-out constant, is a function of the salt cation valancy and the protein but it is independent of pH and temperature.

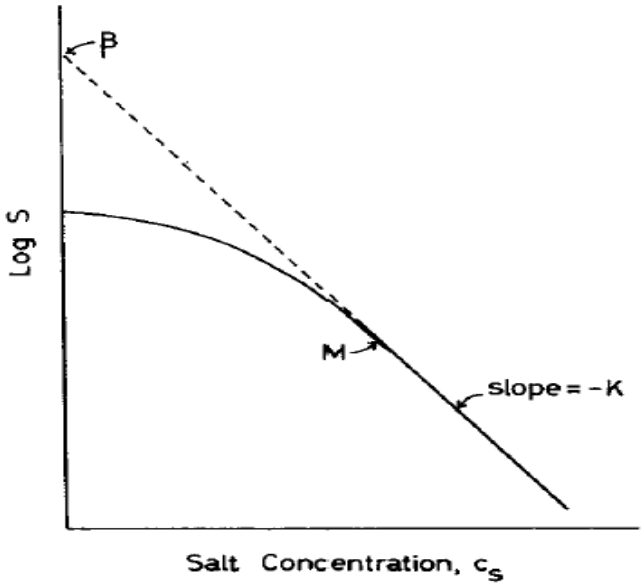


Figure 1.3: Salting-out curve (Foster et al., 1971), where S is solubility of gradient –k, B is the theoretical solubility in a pure aqueous solution, and M is the salt molarity.

Despite its widespread use, the Cohn equation is true for a limited range of salt concentration, deviating at both extremes of the salting out range. The curve represents actual solubility whereas the tangent is derived from the equation (Foster et al., 1971).

1.5.2 Salt based precipitation

The addition of salt to a protein solution results in electrostatic bonding between salt ions and water molecules to form, which at higher concentrations compete with the proteins that are being hydrated. The liberated protein molecules are forced by water molecules to aggregate (via their hydrophobic surface patches) in these conditions as the entropic cost of opening a cavity in the solvent to accommodate them is unfavourably high (as opposed to keeping polar salt ions solvated). As the system tends towards increasing entropy, protein molecules clump together and minimise their surface area to the solvent reducing the water molecules required to solvate them (or the area of the hydration layer, where

entropy is lower due to ordered molecular interactions). This loss in the water monolayer surrounding the protein reduces the barrier to aggregation that can be facilitated by mixing and more efficient solute collisions (the kinetic energy of the solutes should be high enough to overcome electrostatic repulsions and form hydrophobic bonds). The increase in free energy required to accommodate the protein in its dissolved state, i.e. opening a cavity in the solvent continuum, is unfavourably higher than keeping the proteins out of solution, in a high salt environment. However in salt concentrations, protein solubility actually increases (salting-in) above that of its pure aqueous solution as a result of solvating electrostatic interactions between salt ions and charged groups on the protein surface (Stavrinides et al., 1993).

The choice of salt is mainly determined by its solubility and valency, and acid by its potential for denaturation (Salt et al., 1982). Cations and anions of both acids and salts have been ranked in the Hofmeister series (also known as the lyotropic series), according to their lyotropic ability; additionally, this property is also inversely related to denaturation (see Figure 1.4). The Hofmeister series is especially useful to us as the salt order is the same for all proteins (Fruton, 1990) is useful in selecting suitable salts for protein precipitation (Zhang and Cremer, 2006) by their ability to reduce protein solubility. The classification is based on water specific lyotropic interactions for each salt (Hofmeister, 1888) which determines the formation of hydrophobic bonds between protein molecules. The salts increase solvent surface tension, consequently affecting the solubility of non-polar molecules (Melander & Horvath, 1977).

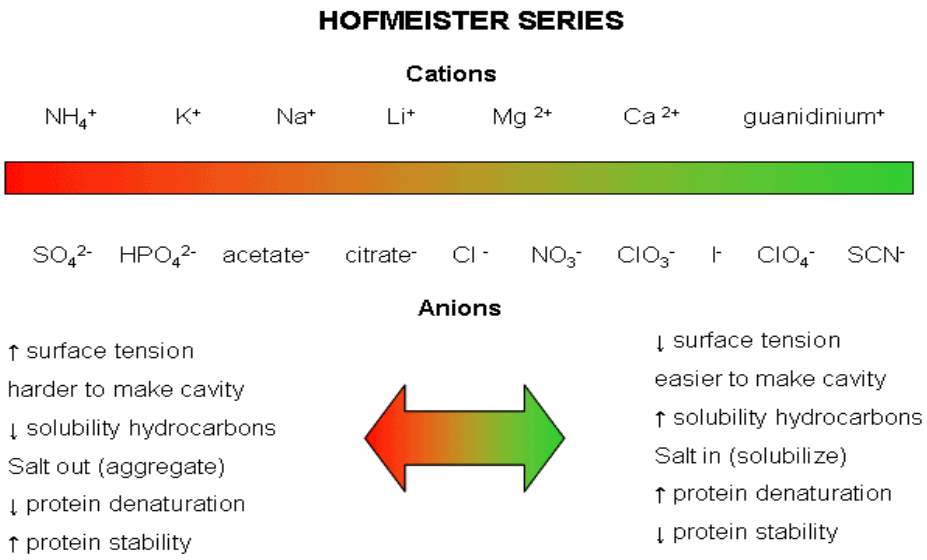


Figure 1.4: The Hofmeister series of salt ions

Ammonium sulphate is the most common protein precipitant used in industry, owing to its high yields and stabilising effect on proteins preventing denaturation (Walsh and Headon, 1994). Its density in solution at precipitating saturations (1.24 g/cm^3, 4.1 M, 20 ^{o}C) is sufficiently low that it does not interfere with processing (such as sedimentation during centrifugation) unlike other precipitants that often raise viscosity. However, precipitate sizes are small and large amounts of $(NH_3)_2SO_4$ must be used that can be corrosive to downstream ion-exchange resins (dialysis/diafiltration would be required to remove it) although hydrophobic interaction chromatography is its natural partner.

Melander and Horvath (1977) proposed that the observed change in protein solubility was directly linked to the electrostatic free energy available for solvation (the protein forming a hydrophobic cavity). They found that the increase in the solution's molar surface tension from adding salt ions was directly proportional to the hydrophobic interaction potential between protein particles (and this forms the basis of the Hofmeister series).

Sodium citrate is also a very capable precipitating agent as shown by the high position in the Hofmeister series but falls foul of limited solubility at room temperature. Sodium sulphate has been used for IgG_1 precipitation as the precipitated IgG from sodium sulphate is usually very stable. Sodium sulphate is especially suited to heat stable proteins due to its high solubility over 40 ^{o}C.

1.5.3 Isoelectric precipitation

Isoelectric precipitation is a technique that bring about precipitation by a change in pH. This mechanism is demonstrated in Figure 1.5 where the solubility of soya protein changes with pH. At the isoelectric point of the protein (pI) a zero net surface charge is exhibited (diminishing hydrophilic interactions) and the molecules will aggregate. This effect is most effective for proteins with many hydrophobic surface groups and low hydration constants (Shaw et al., 1966). Caution should be exercised with proteins with an acidic pI in case of denaturation (pH < 4); otherwise acids are mostly inexpensive and well tolerated by proteins (Daufin, 1997).

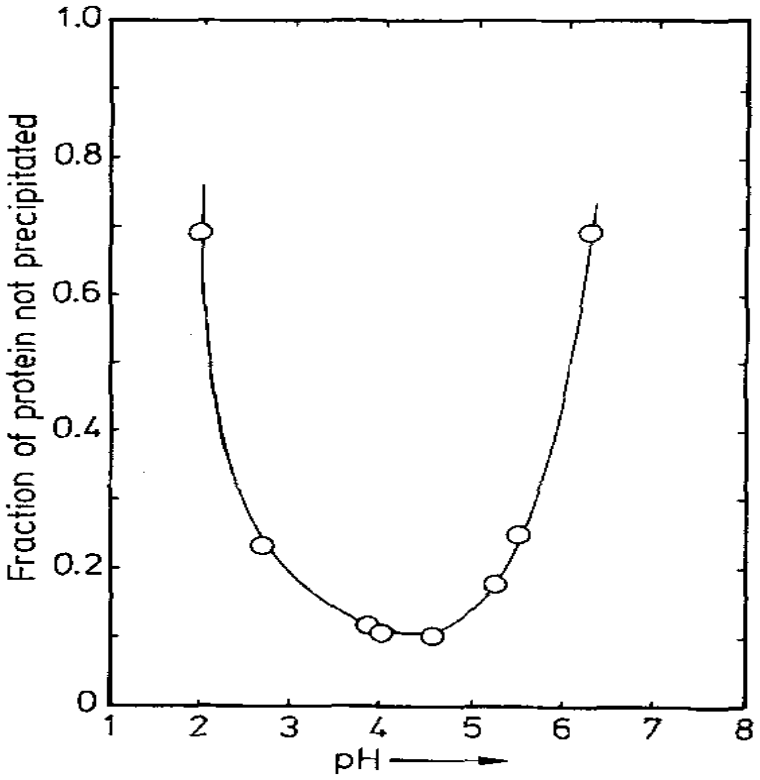


Figure 1.5: The effect of pH on soya protein precipitation (expressed as a fraction of initial concentration) (Virkar et al., 1982)

Acids are usually inexpensive, require low duties (with respect to salts) and are accepted in many protein food products. In fractional precipitation, it is sometimes possible to directly move to the next cut without removing the acid (contrary to many salts). Figure 1.6 shows the effect of pH on precipitation of Lysozyme with sodium chloride (Shih et al., 1992). Lysozyme's pI is at pH 10.5 where, as expected, its concentration is lowest but precipitation is also seen in the acidic data range.

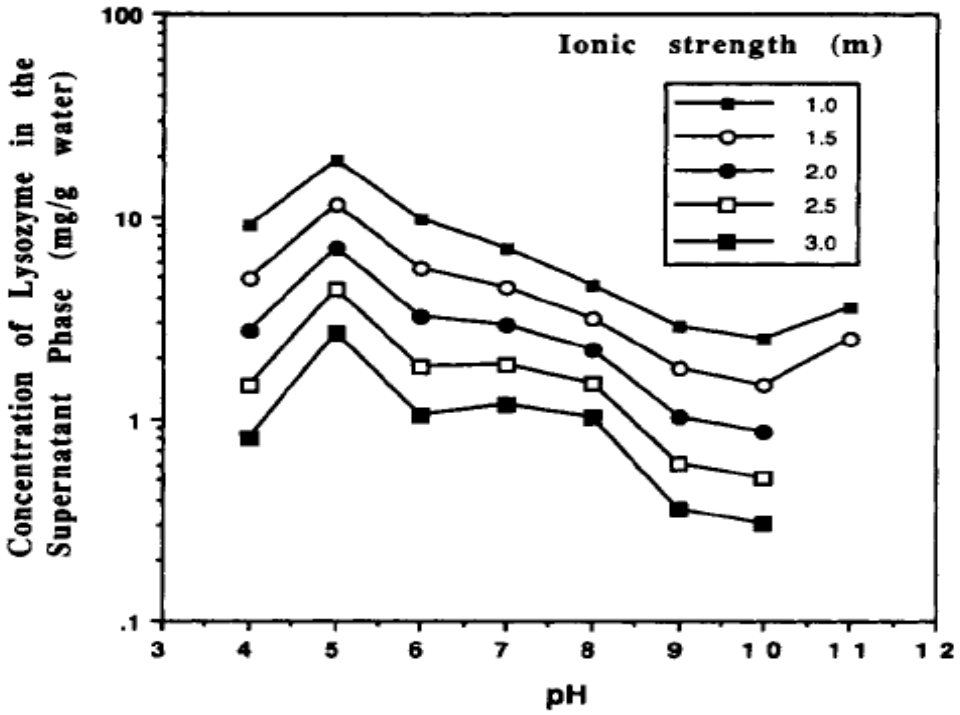


Figure 1.6: The effect of pH and NaCl on lysosyme solubility (Shih et al., 1992)

Other methods used to precipitate proteins utilise ionic polyelectrolytes such as in flocculation and polyvalent metal ions, which are known for a strong precipitating action. They can subsequently be removed by chelating or ion exchange chromatography.

When considering choice of reagent, the final process scale operation must be considered. The precipitant must meet processing limitations in a GMP environment and be safe for workers, i.e. hydrochloric acid is volatile and highly corrosive and their use in high pressure devices such as centrifuges and homogenisers is ill-advised. Also, salts precipitation can be integrated with a HIC column to complement a simpler process. PEG is useful with affinity techniques for it does not denature chromatographic ligands like salts may do at the concentrations required for precipitation (greater than 1 M). Protein integrity, ease of operation, the fractionation ability and the final use of the product are other factors to be

considered. Some precipitants such as metal ions and alcohols are associated with denaturation and safety concerns; ammonium sulphate is categorised as generally recognised as safe (GRAS) by the FDA and also used in vaccine formulation. Medicinal products are subject to strict purity requirements so any reagent residues in the final product need to be shown they are tolerated well in the patient.

1.6 Centrifugation

Centrifugation continues to play an important part in cell and debris removal from the product containing suspension. It is also used post precipitation to separate precipitated product and the impurity containing supernatant (or vice versa) usually with minimal yield losses. Centrifugation is most cost effective at larger scales where depth filtration is much lower throughput (non-continuous) and costly in terms of filter usage. However with larger centrifuges centrifugal performance becomes more dependent on what precipitation conditions were used as precipitate is a soft and loose biological solid, which is sheared easily by bioprocessing stresses found in pumps, air-liquid interphases and centrifuge feed zones (Manweiller and Hoare, 1990). Damage sustained to the precipitate results in its fragmentation, increasing the difficulty to separate and resulting in lower centrifugation recovery.

In the precipitation reaction, the formation of the precipitate takes place over four stages; nucleation, growth, aggregation of nuclei and ageing (Bell et al., 1983). Nucleation and growth is determined by diffusion controlled (Brownian motion), perikinetic growth resulting in a particles of roughly 0.2 μm (Hoare, 1982). Growth rate then becomes orthokinetic as large particles (> 1 μm) require additional energy provided by macro-mixing (via an impeller, shaker etc.) to overcome the energy barrier for aggregation (Smoluchowski, 1917). Upon arrival at a critical size, growth stops and precipitate ageing commences. Ageing requires a period in a low shear environment (such as from an impeller) for the erosion and re-arrangement of the precipitate, without this phase the

precipitate will not develop a tight and compact structure (Bell et al., 1983). Therefore this phase of precipitation is critical to good centrifugation recovery yield.

Studies have shown extensive ageing increases particle strength and consequently improve recovery (Bell and Dunnill, 1982a). Smaller finer particles are also 'mopped up' by larger particles in the ageing phase, 'filling in' gaps and increasing precipitate density (Smoluchowski theory). Figure 1.7 highlights the increased strength of extensively aged soya protein precipitate (by increasing mixing time) after they have been exposed to capillary shear treatment (Bell and Dunnill, 1982a).

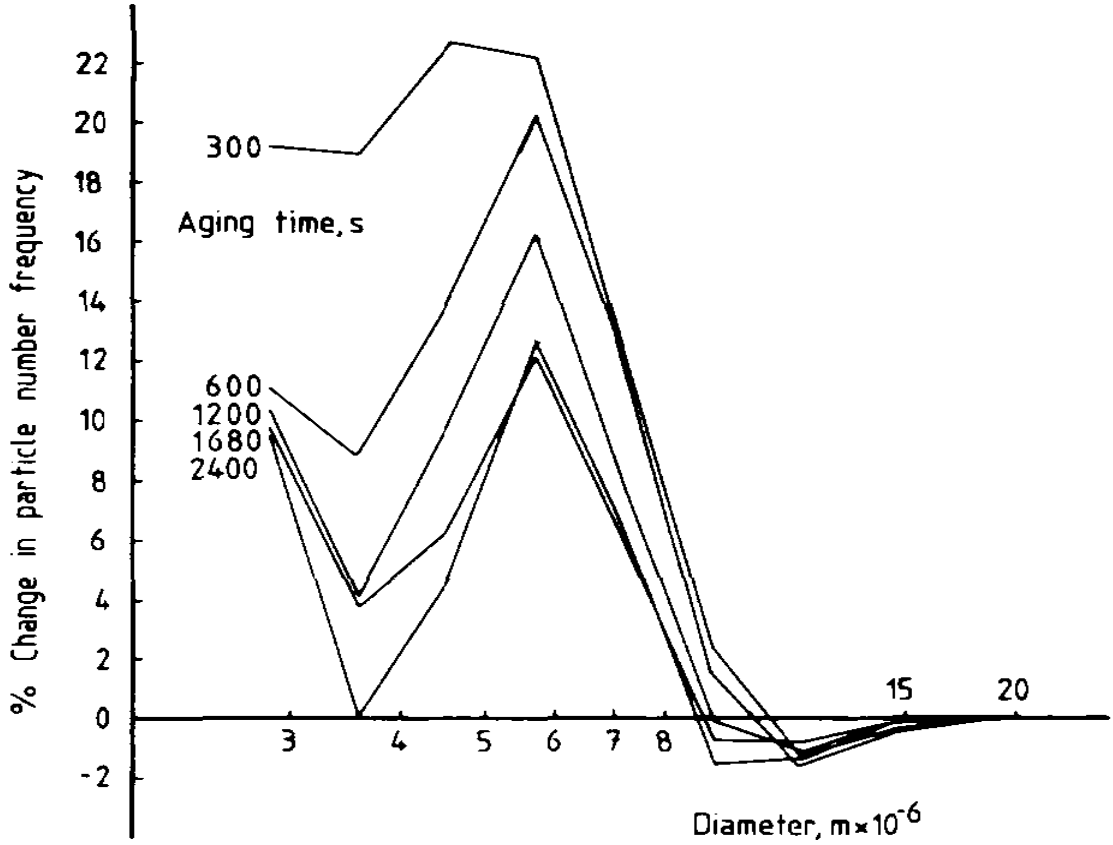


Figure 1.7: The effect of ageing on particle size post shear treatment. The frequency of 'broken' particles after capillary shear ageing (Bell and Dunnill, 1982a).

1.7 High throughput process development

High throughput process development (HTPD) facilitates accelerated routes through drug development for earlier milestone completion and commercial realisation. The increasing use of robotics and microscale platforms have considerably reduced the time, labour and resource costs of process development (Lye et al., 2003; Micheletti and Lye, 2006). However, most bioprocess experiments are still not high throughput friendly and must be done one at a time and require resource planning of scarce material. Analytics and operations that take a long time or require a lot of sample (e.g. protein A HPLC, cell cultures) are bottlenecks that are best negotiated using experimental design techniques. The double edged sword of automation and experimental design may allow us to derive the maximum useful process information from limited datasets (Lye et al., 2003; Islam et al., 2007).

1.8 Experimental design

The complexities of biochemical systems defined by multiple parameters and outputs, which often must be considered together to understand the factor interactions require a lot of experimental effort to identify robust operating spaces. Univariate methods (such as one factor at a time) offer simplified solutions though neglecting factor-factor interactions and assuming factors to be free from interactions can lead to inaccurate models (Eriksson et al., 2000). Most bioprocess systems are governed by complex functions and have interdependent factors that make the use of univariate methods extremely inefficient (Czitrom, 1999).

1.8.1 Design of Experiments

Design of experiments (DoE) methods build a model of the system using a list of generic points in the investigational design space. The model is refined by further rounds of experiments. The levels of the variables are varied simultaneously thereby estimating factor inter-dependency and keeping sample numbers manageable. There are many forms of DoE available for each stage of process development from screening to robustness testing. Depending on the design method, DoE generates a list of experiments from within the design space to create a model. It is then used to predict the optimum operating conditions and boundaries of failure that are experimentally verified. Recent bioprocess studies utilising experimental design have mainly focussed on upstream applications where the recent introduction of high throughput technology required a platform to efficiently and accurately process the data being generated (Nikerel et al., 2005; Ren et al., 2006; Zhang, 2006; Swalley et al., 2006; Islam et al., 2007). Consequently the use of experimental design and statistical data processing software (Design Expert 7, Stat-Ease Inc, MN, USA) and Minitab 15 (Coventry, UK)) has become commonplace to verify and optimise bioprocesses. The combination of multiple factors (>3) and complex factor-response relationships ('hilly' surfaces) can compromise the traditional DoE route by requiring intensive experimentation. Other options are: (i) to eliminate factors that have little effect on the overall function or (ii) to use extensive 'aliasing/confounding' factors though this gives a less useful model (Abu-Absi et al., 2010).

DoE is a valuable tool in bioprocess development that has led to more robust models, optimised operating conditions, and less batch failures in GMP manufacture. Its use also assists in new drug regulatory approval by the Quality by Design (QbD) principles. However, processes with many steps and factors are too experiment intensive for use with traditional DoE methods. DoE works well when the optimisation problem has a low number of factors and small ranges however this is rarely the case for early stage process development. Applying DoE to a large design surface with many unknowns will not

provide any clear or useful knowledge. Other issues with DoE are daunting software packages and a lack of statistical understanding by inexperienced users, which could potentially lead to a poorly controlled process. However, complex biochemical interactions inherent in bioprocess operations can still present a major obstable for DoE and process optimisation is still carried out empirically.

1.8.2 Simplex search method

When applied to bioprocessing, the main goals of DoE are to identify optimum and robust (in terms of process outputs) operating conditions in an efficient and resourceful manner. The simplex search method achieves these goals using a data-driven approach rather than simultaneously evaluating all variables for model building and predicting the optimum. Developed by Nelder and Mead (1965), the original simplex method is a popular search algorithm that optimises a function by comparing the function values at the vertices of a simplex. The simplex refers to the geometric shape formed by the coordinates of the initial conditions, which are one more in number (n+1) than the total number of variables in the experimental design (n). Beginning from the initial conditions the algorithm evaluates the system outputs to decide the direction to take the simplex. The simplex progresses by replacing its worst corner (where the output is least desirable) with a new point decided by the algorithm. The coordinates of the potentially better point are defined by ranking all simplex points and reflections away from worse points moving the simplex to a space where the process output is more favourable. These movements are repeated at each iteration until further improvement in the output is no longer possible and the simplex is terminated. Sample numbers are restricted to the points used in the trajectory of the simplex from the initial conditions to termination. This data-driven approach continues until the simplex has found the set of conditions in the design space where the optimum output is achieved. Such function-value based optimisation methods do not seek to build models of the design space but work in a retrospective fashion, constantly benefitting from data input to guide the search. The objective function requires careful definition for it is the driver

behind the simplex algorithm. The objective function can be a process or non-process output, a combination of outputs, and set to maximise or minimise (i.e. product yield, purification, process time, quantity of material used, etc). For exploration of higher variable design spaces ((n>5) the simplex method is expected to perform even more efficiently in terms of time and sample savings (Goldberg, 1989). Exploring such high variable spaces with conventional DoE methods is often compromised purely because of the vast number of experiments required that are simply not feasible to execute for bioprocessing problems.

The simplex method has many useful properties as an optimisation method. In particular, it is advantageous to the optimisation of complex processes since sequential methods do not require accurate initial models for a process. This is demonstrated by the many examples of the use of sequential methods for optimisation where the highly complicated biochemical interactions of bioprocessing are not clearly understood. Chhatre et al., successfully demonstrated the utility of the simplex method for early stage process design using FAb precipitation and chromatography case studies (Chhatre et al., 2011); Banerjee and Bhattacharyya (1993) applied simplex based evolutionary designs to maximise enzyme activity using inducers.

The adaptive nature of the simplex algorithm makes it very efficient at finding the optimum point in a high variable design space. The total number of experiments can be further minimised by strategic use of the simplex such as where in the design space it is initiated from, how large it should be as well as processing a rough idea of the response surface of the objective. However there are many variables that need to be defined when using the simplex search method, which can lead to variation in the solutions provided by it. There is very little in the literature to help select the initial conditions of the simplex search method, what termination criteria to use and what objective function to use. As the simplex method works by stepping in and around the design space it is also possible to get stuck at local optima in the response surface. The low experimental cost of the simplex method allows running it again from a different set of initial conditions though local optima

knowledge can still be useful for investigating alternative operating areas. Though the simplex search method as a bioprocess optimisation tool has been shown in recent studies (Chhatre et al., 2011, Konstantinidis et al., 2016 and Konstantinidis et al., 2017), the literature does not show how to use the method in a structured and clearly defined manner. The many possible initial conditions of the simplex, the objective function and the stop criteria are all customisable features of the method that can bring about a different solution and different partway to the solution. The work in this thesis will seek to standardise the wide variety of ways the simplex search method can be used so that the same solution is found every time regardless of user. This will be achieved by presenting in a framework with DoE so that each case study is processed in the same manner following one set of experimental design protocols.

1.8.3 Other experimental design methods

Evolutionary methods similar to the simplex have been used in biochemisty applications by Banerjee and Bhattacharyya (1993) where information on the process was insufficient, to maximize enzyme activity using induction. Tunga et al., (1999) used evolutionary operations (EVOP) to maximize the production of protease by optimizing the concentrations of vitamin, metal ion and plant hormone (Box, 1957; Box and Hunter 1957). The optimisation for product titre in fermentation studies has been achieved using a genetic algorithm (Saha et al., 2015).

Traditional DoE methods can be compromised by complicated non-linear multivariate systems but evolutionary methods, including the simplex method, have successfully been shown to be efficient optimisation methods in various applied science examples such as protein crystallisation, bioprocess media optimisation, multimodal chromatography, HPLC separation, two-phase partition and gas chromatography (Chhatre et al., 2011; Prater et al.,

1999; Wang et al., 1993; Karnka et al., 2002; Lancas et al., 1995; Backman and Shanbag, 1983; Bakeas and Siskos, 1996).

1.9 Research objectives

This thesis examines and develops the potential of the simplex method when applied to a microwell based bioprocess optimisation. A systematic approach will be developed that will tell the user how to divide the design space using DoE and where to initiate the simplex from, when to stop and how to select the objective function. This would be the first demonstration of combining DoE and the simplex method with high throughput automated microwell experimentation. Traditional DoE methods are used to compare the results of the simplex as well as conclude how they may be able to be used together so the maximum benefits of both methods may be achieved with the drawbacks of neither. The design strategy will benefit from the rapid optimum identification of the simplex method from minimal experiment numbers whilst the DoE will mitigate against the simplex method falling into local optima traps. The lack of surface knowledge revealed by the simplex will be covered by the DoE screening designs so surface insight is not lost. The novelty of the proposed experimental design strategy will cover both breadth and depth of design space in question. Furthermore, a precipitation step will be developed for primary recovery of a MAb purification process and assessed if it can be a substitute for Protein A chromatography. It was also serve to refine the use of the simplex based protocols that will then be demonstrated in optimising a precipitation and centrifugation sequence. Developing the two steps together will demonstrate the simplex based approach as an efficient tool for larger design spaces with multiple factors. The method will also be demonstrated on optimising a more challenging problem of aggregate removal by optimising a cation exchange chromatography step using the latest microwell based methods. The microscale chromatography methods will be a first demonstration of the methods used with DoE and the issues encountered with method development and experimental set up will be explored.

The DoE-simplex methodology should enable faster bioprocess development and an efficient identification of optimum processing conditions. To transfer and test the robustness of the combined methods approach to different product processes the optimisation method will then be used to optimise a complicated multiple variable precipitation and centrifugation operations sequence. Comparisons will be made with traditional optimisation routes. The developed optimisation method will also be applied to chromatography case studies using novel high throughput approaches to optimise antibody products provided by Lonza Biologics.

2. Materials & Methods

2.1 Materials

2.1.1 Chemicals

All chemicals used in this study including potassium phosphate (monobasic and dibasic), sodium phosphate (monobasic and dibasic), ammonium sulphate, sodium acetate, glycine, sodium citrate, sodium hydroxide, hydrochloric acid, sodium chloride, TRIS-EDTA, etc were obtained from Sigma Chemicals Co. Ltd. (Dorset, UK) and were of analytical grade quality, unless specified otherwise. HPLC solutions such as Ethanol were all HPLC analysis grade.

2.1.2 Clarified cell cultures

Monoclonal antibody - IgG_4

The product is derived from the CHO CY01 cell line developed by Lonza (Slough, UK) and licensed to department of Biochemical Engineering, UCL for research use. The recombinant human IgG_4 monoclonal MAb has a pI between 6.8 - 7.2 as determined by isoelectric focusing (personal communication with Dietmar Lang, Lonza Biologics), and is expressed extracellularly.

The seed CHO cells were stored in liquid nitrogen with 10% (v/v) DMSO. Before mammalian cell culture, the frozen CHO cells were revived and passaged in a 50 mL flask containing media, incubated at 5 % (v/v) CO_2, 37 ^oC for two generations. The CHO cells were then cultured in CD CHO Medium in a 20 L stirred tank fermenter (Sartorius Stedim,

UK) by fed-batch mode. CD CHO Medium AGT with 100 g/L glucose was used as the feed solution to keep glucose concentration at 2 g/L in the culture. Glucose concentration and cell viability was analysed every day and recorded to monitor the CHO cell culture. pH, oxygen and the addition of pluronic antifoam were controlled by fermenter automatically according to preset value. Cell culture was harvested at MAb concentration around 3.8 g/L. The broth was then centrifuged at 10000 rpm by Eppendorf Centrifuge 5810R for 30 min and then filtered through a 0.22 µm depth filter (Millipore Limited, Dundee, UK).

The detailed fermentation protocol can be referred to Galbraith et al. (2006). The clarified feedstock was stored in a -70 ^{O}C freezer. For each set of experiments, the same batch of cell culture was used. For the precipitation experiments requiring concentrated feed solutions, 5 kDa centrifugal concentrators were used (Millipore Limited, Dundee, UK).

Monoclonal antibody - IgG_1

The IgG_1 was derived from the CHO LB04 cell line developed by Lonza (Slough, UK). The pI of the MAb is between 7.4 - 7.8 as determined by isoelectric focusing. The culture process is identical to that described previously for IgG_4.

Anti-insulin antibody

The anti-insulin antibody is also derived from the cell line, CHO LB04 developed by Lonza (Slough, UK). The pI of the MAb is between 7.6 to 8 as determined by isoelectric focusing (personal communication, Lonza Biologics). The culture process is identical to that described previously.

MorAb

The MorAb product is derived from the CHO LB04 cell line Lonza (Slough, UK). The pI of the MAb is between 7.2 to 7.8 as determined by isoelectric focusing (personal communication with Dr Dietmar Lang, Lonza Biologics). The culture process is identical to that described previously.

2.2 Methods

2.2.1 Microscale precipitation

Microscale precipitation experiments were automated and performed on a four tip, Multiprobe II^{TM} EX (Packard Instrument Co., Meriden, Connecticut, US) liquid handling robot (LHR), which has been described elsewhere (Lye et al., 2003). The multiprobe is controlled by the winprep application software. Bio-robotix disposable tips with a size of $d_{internal}$ = 0.6 mm were used at injection flowrates of 400 μLmin^{-1} (in the turbulent range so they are well mixed). Well additions were made at the liquid surface, with liquid tracking to minimise droplet formation. During liquid transfers, 200 μl and 1 ml conductive disposable robotic tips (Tecan Group Ltd., Mannedorf, Switzerland) were used to eliminate cross contamination. Liquid aspiration would be tracked to liquid level and always be taken at just below the liquid surface (Revill, 1992). The microconducive tips used were both disposable and fixed; disposable tips are fabricated from virgin polypropylene impregnated with carbon and the robots own fixed tips are made from Teflon coated stainless steel. Experiments were cautiously done with disposable tips due to corrosion concerns from high molarity precipitant (saturated ammonium sulphate).

The performance files governing pipetting precision and accuracy had previously been optimised to within a 5% CV limit for low viscosity liquids (Nealon et al., 2005). For the

ammonium sulphate precipitation, the clarified broth was rapidly brought to 60% saturation with $(NH_4)_2SO_4$ (buffered with 0.1 M potassium phosphate). The precipitant was injected into the well containing the protein solution and then shaken in a Thermomixer® unit (Eppendorf thermomixer) at 1400 rpm, and ambient temperature for 2 hours.

Microplate

The tall aspect ratio microwell design (96-SRW Sarstedt) illustrated in the schematic (Figure 2.1), was chosen to minimise materials consumption and form a compact sediment (pellet) when centrifuged for ease of processing.

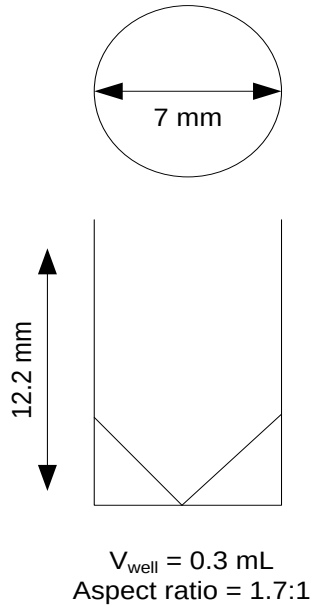


Figure 2.1: Microwell schematic used in the precipitation process

2.2.2 Microwell centrifugation

A J2-MI laboratory centrifuge (Beckman Instruments Ltd., High Wycombe, UK; 4000 rpm, 10 minutes, Σ_{lab} = 1.7 m^2, C_{lab} = 1) with a swing out microplate bucket was used to separate the precipitate (see Figure 2.2). Clear supernatant was transferred to Agilent 96 HPLC micro-well plate on Agilent HPLC 1200 series system (Agilent Technologies, Stockport, UK) and Tecan Safire plate reader (Tecan, Wisconsin, US) for analysis.

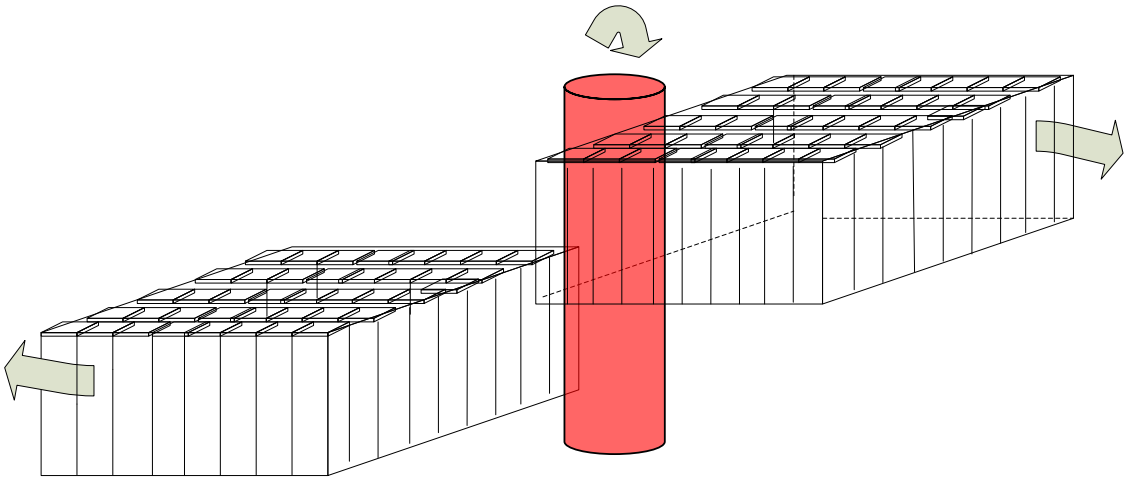


Figure 2.2: High throughput centrifugation. (As the centrifugal force increases the bucket holding the microplate will swing out, forming the precipitate pellet at the bottom of the wells).

2.2.3 Protein analysis

An Agilent HPLC 1200 series was used for protein of interest and total soluble protein assays of the supernatant. This method revealed the proteins that had not precipitated and

remained in solution. MAb absorbs strongly at 220 nm and other proteins at 280 nm. The clarified supernatant was transferred by the robotic arm to an Agilent 96-well sample loading plate (Agilent Technologies U.K.Limited, Cheshire, UK) for MAb HPLC analysis conducted on an Agilent 1200 system by loading 100 mL of sample on to a 1 mL Protein G HiTrap column (GE Healthcare, Buckinghamshire, UK) at 2 mL/min. 20 mM sodium phosphate, pH 7 was used for column equilibration and washing, and 20 mM glycine, pH 2.8 (pH adjusted with HCl) was used for elution. Both were filtered 0.22 mm sterile filter. The 220nm elution peaks were integrated and converted into concentrations by a calibration curve. For the purposes of the simplex algorithm, the goal was to identify conditions that maximised the amount of MAb in the solid phase precipitate. Hence, HPLC concentrations of MAb remaining in the liquid phase after precipitation were multiplied by -1 such that a large quantity of supernatant MAb resulted in a low objective function value (reflecting the small amount in the precipitate). Correspondingly, a low supernatant MAb value after multiplication resulted in a high objective function, indicating a large precipitated MAb concentration.

2.2.4 Overall precipitation method

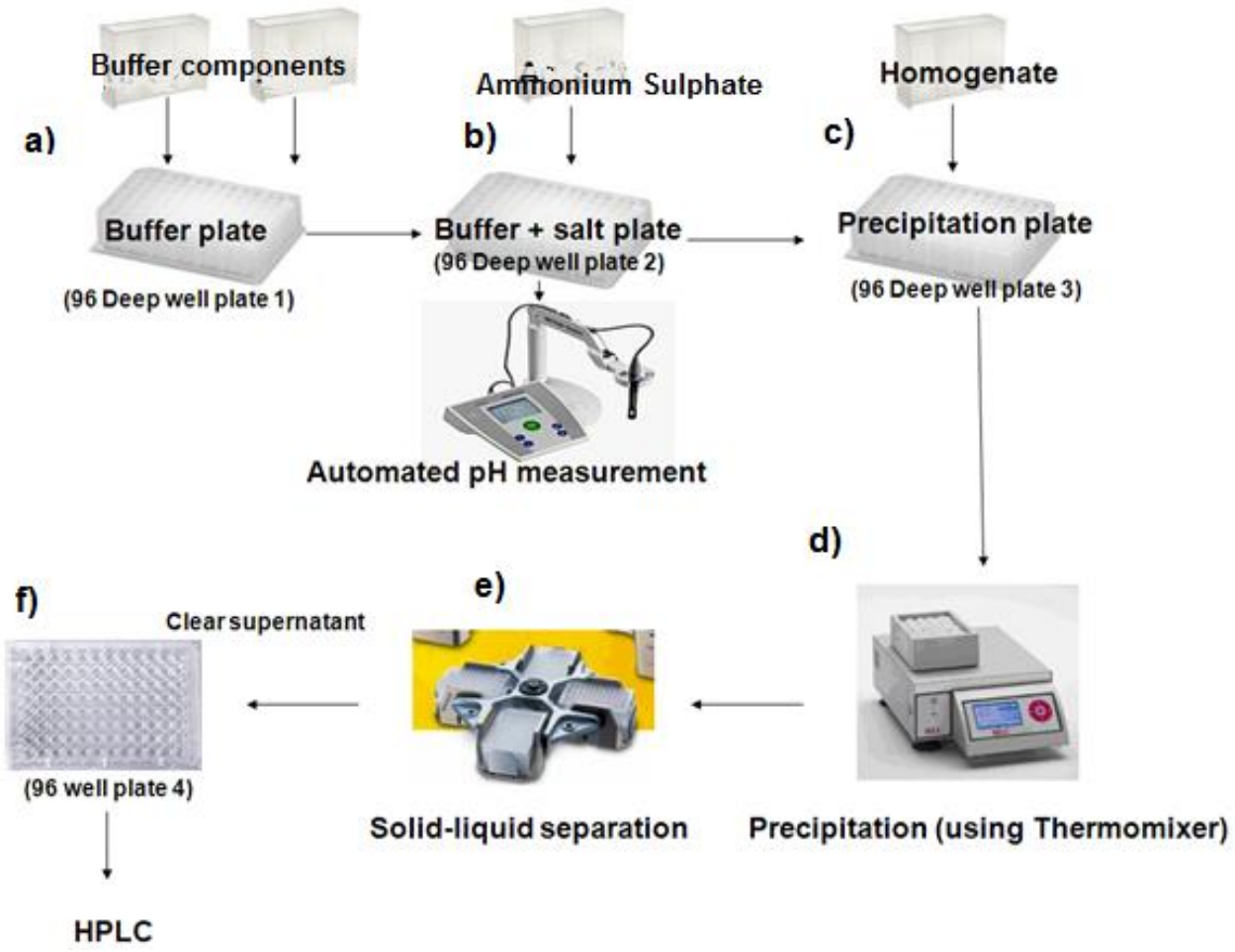


Figure 2.3: Microscale precipitation process flow chart

a) Buffer preparation (Plate 1) – Potassium Phosphate and Sodium Acetate buffers were made to ionic strengths, 1.5 M and 2.5 M respectively and then used to generate a pH range between pH 3 – pH 9. The desired pH was achieved by mixing the monobasic and dibasic forms to the required volumes according to the Hasselbalch-Henderson equation. The buffer concentrations were near the solubility limits of the salts and needed to be high as

later steps would dilute them and diminish their ability to maintain stable pH. The stock buffers were made manually whereas the preparation of plate 1 utilised the automated liquid handling (Multiprobe II, Hewlett Packard, California, USA). Plate 1 being a deep well plate with 2 mL working volume acted as a stock plate, which was to be used multiple times. Each column represented a single pH value.

b) Buffer + salt (Plate 2) – Ammonium Sulphate salt solution was made to 4.4 M (~ saturated at room temperature) before being mixed with buffer. 'Plate 2' represented a pH vs. salt concentration matrix as each well contained an equal volume of buffer (taken from the corresponding well in plate 1) and Ammonium Sulphate diluted as necessary to create a precipitant plate with a range of pH and salt concentrations. Plate 2 also acted as a stock plate as its volume capacity was much greater than the precipitation plate, plate 3.

c) Precipitation plate (plate 3) – The reaction took place in the smaller plate 3 (0.3 mL) allowing us to conserve material and proceed through the experiment quicker. Each well consisted of only 40% protein feed material, as the maximum volume was restricted by the solubility of the salt and the highest salt concentration we wished to investigate. Automated liquid handling dealt appropriate amounts of feed and salt dissolved in buffer from plate 2 so that our array of pH and salt concentration conditions were generated. The values in each well can be read off the axes of the microplate. Note, in the topmost 0M salt row, only buffer at 100 mM concentration was used to highlight the effect of pH on precipitation.

d) Agitation – Plate 3 was then covered and transferred onto a thermomixer (Eppendorf, UK) unit for controlled mixing and precipitate maturation. The shaking speed was set at 14000 rpm and the temperature at 21 ^{o}C. The shaking lasted for 2 hours (previous studies have suggested longer mixing times show little improvement in precipitate recovery).

e) Microwell centrifugation – The plate was then centrifuged in a Beckmann J2-MI laboratory centrifuge, with JS-13.1 swing-out rotor for 15 minutes at 4000 rpm (Beckman Instruments Ltd., High Wycombe, UK). The wells of the microplate were shaped with a conical base to facilitate the formation of a more compact pellet so the supernatant could be easily extracted without any disturbance (see Figure 2.1).

f) HPLC analysis – The supernatant was transferred to an analytical plate by the liquid handling robot, which was then assayed at 280nm and 220nm in the HPLC. Readings of total soluble protein and the protein of interest in in the supernatant were taken.

2.2.5 Lab scale precipitation

Lab scale precipitation used the IgG_4 feedstock and was executed in a 100 mL baffled beaker. The agitation was provided by a clamped drill with a 6 bladed mini impeller (2 cm diameter operated at 600 rpm). The precipitate provided material for the shear device mimicking studies so the acoustic device could be correlated to the rotating disc shear device (see section 4.3).

2.2.6 High throughput chromatography

The high throughput chromatography studies carried out in chapters 5 and 6 used an automated Tecan liquid handling robot (Tecan Group Ltd., Zurich, Switzerland) and miniature chromatography columns of 200 µL column volume (Robocolumns, Atoll GmbH, Weingarten, Germany). The robocolumns come prepacked in a 96 column layout however they can be used individually or up to 8 columns in parallel (limited only by the number of tips the robot has in this case it was 8). A method script was written on the Tecan Evoware software instructing the robot to carry out tasks such as loading the columns with buffers, moving 96 well plates in position under columns to collect fractions and transferring these plates to the plate reader for A280 absorbance reading and HPLC for further analysis. Multiple scripts were written including ones for bind and elute chromatography processes and for flow through chromatography. The script instructions were developed from generic lab scale protocols for Protein A and Cation exchange chromatography. The main difference being the phases are not continuously loaded onto the column but in a number of injections are the Tecan pipette is limited to 1 mL injections (which for a 200 µL robocolumn will allow a single injection of 5 column volumes).

2.2.7 Protein A chromatography

To prepare the feed for the cation exchange experiments, the XK16/20 column hardware (GE Healthcare, Uppsala, Sweden) was used to prepare the labscale protein A column (16 refers to the diameter in mm and 20 is the column height in cm). The column was packed using loose MabSelect Sure media, which was stored in tris buffer with azide. The packing protocol followed the procedures in GE Healthcare MabSelect manual (GE Healthcare, 2011). The labscale chromatography was operated on the AKTA 100 using a generic protein A bind and elute method written on unicorn 5.3 software and is as follows.

Five CVs (column volumes) of 25 mM sodium phosphate pH 7.4 equilibration buffer was run through the column, clarified cell culture supernatant is then loaded at 30 mg product per mL of column volume with a 4 minute residence time. This is followed by a 5 CV re-equilibration phase. A 1 CV wash of equilibration buffer with 10% (v/v) isopropanol is used to wash out hydrophobic impurities. The 25mM sodium citrate pH 3.5 elution buffer is then applied for 5 CVs. An absorbance watch command on the Akta method switches the outlet flow valve from waste to a collection vessel when the absorbance sensor reads above 100 mAu (absorbance units). This continues until the absorbance drops below 100 mAu and the valve switches again back to the waste stream. After the elution phase is complete, 3 CVs of 0.1 M citric acid strip buffer is run through the column. This is followed by 3 CVs of 0.1 M NaOH sanitisation buffer and finally the column is run with 2 CVs of storage buffer (25mM sodium phosphate pH 7.4 with 5 % azide).

2.3 Analytical methods

2.3.1 HPLC Protein A chromatography

The antibody concentration was also measured by a 1 mL protein G HiTrap column (GE Healthcare, Uppsala, Sweden) connected to an Agilent 1200 series HPLC system (Agilent Technologies, Stockport, UK).. 50 µL sample was injected into column with an autosampler. The sequence and methods were also pre-programmed in HPLC software Chemstation. UV 280 nm signal was recorded and used to measure the peak area. The MAb concentration was calculated based on a calibration curve, which was generated by several MAb concentration samples, ranging from 0 mg/mL to 1.5 mg/mL, diluted from MAb standard. The samples were transferred to 96 microwell filtration plates and centrifuged at 1000 rpm for 10 minutes before being loaded onto the HPLC. The flow rate of HPLC was kept at 1 mLmin with upper pressure limit of 85 bar. Total analysis time is 15 minutes with MAb eluting at around 6 minutes.

Analysis of the flow through peak from the HPLC protein A was made to calculate overall impurities included host cell protein, cell culture media protein and other impurities that absorb at 280 nm. HPLC standards of known protein concentration were used to make a A280 nm and protein concentration conversion equation curve. The original clarified feedstock was used as the standard. Several samples diluted from standard were made according to different dilution rates. The BCA total protein assay (see section 2.3.2) was used to measure the total protein concentration in each diluted sample. The impurities concentration for each sample was calculated by subtracting the corresponding antibody concentration from the total protein in that sample. The calibration curve was then regressed from the impurities concentration and HPLC peak area. The regression goodness of fit had R-square at 0.99 and random sample tests were validated by BCA total protein assay.

2.3.2 BCA total protein assay

BCA protein assay was used to analyse the total protein concentration in the samples. BSA standard and BCA protein assay kit were bought from Sigma-Aldrich (Dorset, UK). PBS was used to dilute samples and worked as blank. All samples were diluted to be in 0-40 μg/mL range, which was the standard and assay working range. 1 mL of standards, controls and samples were transferred to cuvettes and then 1 mL BCA reagent was added. Each cuvette was well mixed and covered from light for 10 minutes at room temperature. Both 2 mL cuvette and transparent 96 microwell plates, if samples were transferred from cuvettes, can be used to take samples and tests at UV 595 nm. The total protein concentration was then calculated based on corresponding calibration curves, depending on the method used.

2.3.3 HPLC size exclusion chromatography

The aggregates, monomer and half antibody were analysed by a TSKgel G3000 SWXL column (Tosoh Ltd, Tokyo, Japan) on Agilent 1100 HPLC (Agilent Technologies, Stockport, UK). The molecular weight separation range of column was 10 kDa to 500 kDa. The running buffer was pH 7.0, 20 mM sodium phosphate buffer with 0.15 M sodium chloride. All samples were filtered through 0.22 μm filter before loading to column. The loading concentration was around 1 mg/ml and flowrate at 1 ml/min with a total running time of 20 minutes. The three peaks came out in the sequence of aggregates (7.2 minutes), monomer (8.2 minute) and half antibody (10.3 minutes). The UV 220nm peak area was recorded and percentages of each component were calculated by Chemstation software (Agilent Technologies, Stockport, UK).

2.4 Adapted simplex algorithm

The simplex method used here is adapted from the Nelder - Mead simplex algorithm, which was developed for numerical function optimisation (1965). The method's original purpose is for the optimisation of numerical functions in a continuous and infinite design space but here it has been modified to work within an experimental design space of discrete and finite conditions. This involves carrying out the experiment and then inputting the objective function value to the algorithm for it to provide new condition to evaluate, this is repeated until the best conditions are found. Other typical experimental issues include:

(i) Establishing a grid every possible condition in the design space, this is usually limited by the selected ranges of the factors (e.g. pH 4- 8 and 0 – 2 M salt).

(ii) Smallest interval unit between points (e.g. 0.5 pH units and 0.2 M).

(iii) Where to start the simplex from (initial conditions).

For the case studies the experimental parameters (of n factors) and an objective function (i.e. combination of yield and purity) are used to form the design space, which also forms the limits for the simplex search. Each simplex point represents a set of process conditions for that the objective function value is experimentally established. The simplex method algorithm ranks the objective function values at each corner of the simplex and then replaces the least favourable corner with a new point it predicts to be in a more favourable region of the design space based on geometric manipulation of the simplex such as a reflection, expansion and contraction (for more background on the simplex movements please refer to section 1.5.

Each iteration updates the simplex to a new improved position in the response landscape until eventually it terminates upon an optimum. The size of the steps taken by the simplex is dependent on how favourable the direction is (Nelder, 1962). This is decided by the function value of the new point selected. If it is better than the other points in the simplex the length of the step expands enabling the simplex to move towards the optimum in fewer samples. Conversely, if the simplex strays into a low response area, the algorithm will make the simplex move away from it. Each new point is then experimentally verified. This carries on until the algorithm's termination criteria is met (this is triggered here by a minimization of the delta between the objective function values of successive points). When the simplex code suggests a point off the design space then those conditions will be rounded to the nearest feasible point. If the nearest feasible point is unavailable (because it already forms another point in the simplex or acceptance would degenerate the simplex to one lesser dimension) then the next nearest point will be accepted. A degenerated simplex is when during the iterations a point is selected that make the simplex lose a dimension (preventing it optimising in all dimensions/factors i.e. a 2 factor simplex needs to be of a triangular shape and not be a line of 3 points). This has been addressed by ensuring every simplex has a minimum area.

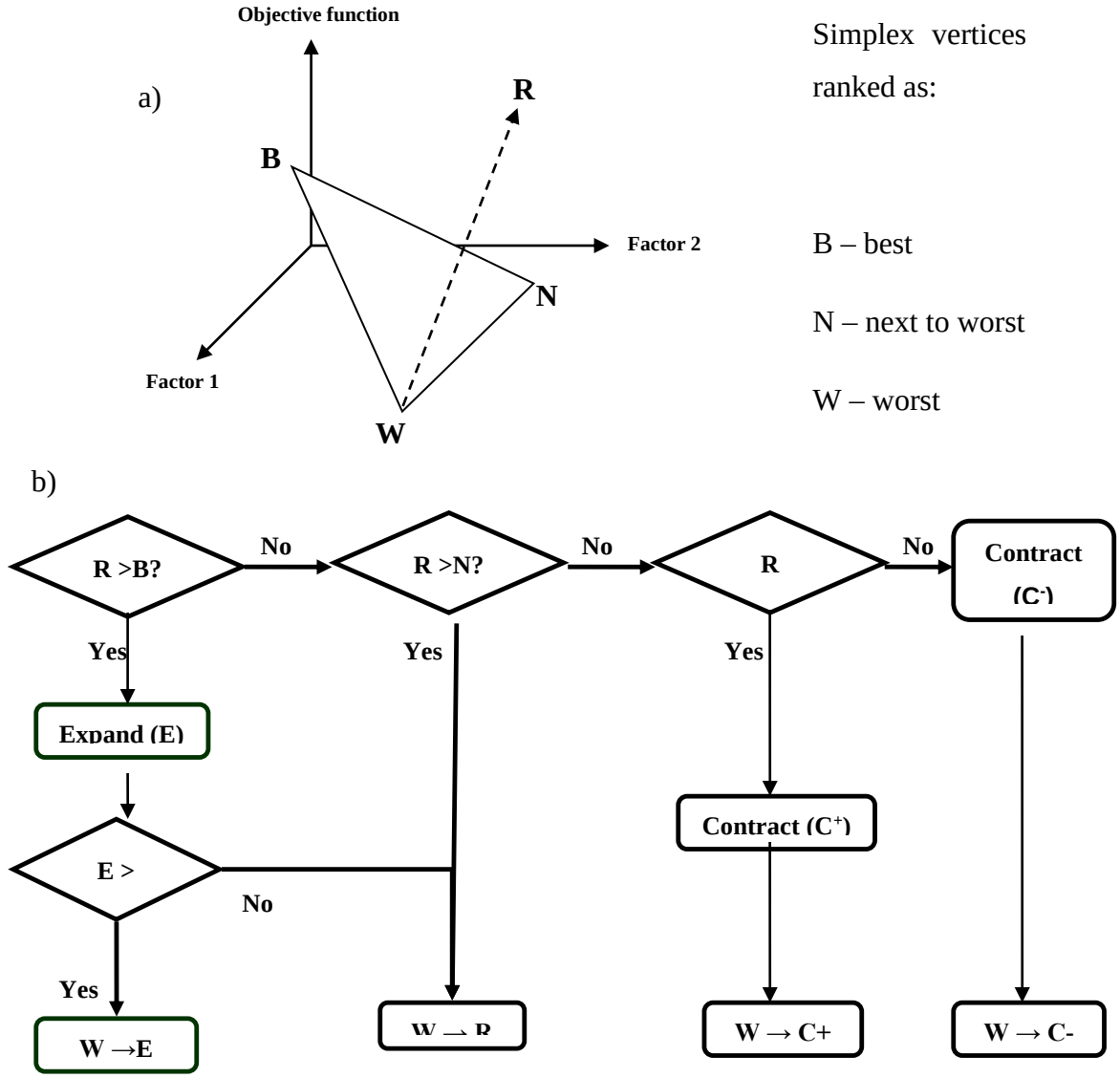


Figure 2.4: Simplex method logic rules. Example of the algorithm for a two variable optimisation. a) illustrates the simplex search scheme where the simplex vertices are ranked as B - best, W - worst, N - next to worst, and R - reflection. b) represents the decision tree the experimental data must go through to suggest the next experimental conditions.

3. Optimisation of MAb precipitation using the simplex method and comparison with traditional DoE optimisation

3.1 Introduction

Recombinant MAb technology dominates the biopharmaceutical industry having witnessed continuous growth since the licensing of the first MAb, muromonab-CD3 in 1986 (Birch, 2005). MAbs are indicated for a wide range of therapeutic targets and there are currently 30 approved therapies with many more in the pipeline (Labrijn et al., 2009). Despite their high effectivity and low rate of adventitious events, MAb commercialisation is hindered by the high cost of production, for example alemtuzumab indicated for Leukaemia, costs approximately £37,000 per year per patient (Shaughnessy, 2012). To reduce cost a lot of the process optimisation has focused on the upstream process, maximising product titre. This has however led to an enormous strain on the downstream process that has been playing catch up to deal with the higher productivity and accompanied increase in impurities (Sommerfeld and Strube, 2005). Protein A chromatography, which has long been the gold standard in MAb capture technology by offering unmatched yield and purity now finds itself quickly becoming the limiting factor in the chain (Przybycien et al., 2004). The high protein loads have increased inventory costs to what is already a costly operation especially in terms of limiting the resin lifecycle from excessive use of cleaning agents (Glynn, 2008). Furthermore, affinity chromatography lacks the throughput capacity of other resins to take advantage of the current high titres (Low et al., 2006, Natarajan and Zydney, 2013). As a result, a flurry of initiatives over the last ten years have investigated potential alternatives to Protein A chromatography (Przybycien et al., 2004; Kumar et al., 2003; Hilbrig and Freitag, 2003; Knevelman et al., 2009).

As introduced in section 1.5, protein precipitation is investigated here addressing two current areas of research, an alternative to protein A and a novel use of the simplex method to optimise precipitation process development. Protein precipitation is a well-established purification method that is used in various industrial processes (blood plasma production being the most common). To achieve high yields and sufficient impurity removal it requires considerable optimisation of several key factors such as pH, precipitant type and concentration, protein concentration, and temperature (Cheng et al., 2006; Cohn et al., 1946; Foster et al., 1976). Purification by precipitation is further discussed in chapter 1 and after a thorough review of the methods available, ammonium sulphate precipitation was chosen to be developed as a primary recovery step for MAb purification.

The LB01 MAb, indicated for acute myeloid Leukemia and Multiple Sclerosis, will be used as the test product with a precipitation mechanism expected to be representative of most MAbs (Labrijn and Buijsse, 2009). This chapter will use this case study to investigates the simplex method and compare its optimisation performance with traditional DoE methods. Main factors such as buffer pH, ammonium sulphate concentration and feed product titre will be optimised with respect to product yield and purification. Currently, it is very attractive if alternative processes can be used to replace affinity chromatography or even to reduce the number of chromatography steps (Ma et al., 2010). Therefore, in the early stage of the purification, a primary separation, such as protein precipitation, may be prudent and welcome to prepare a relatively clearer and less contaminated solution to lower the work burden for subsequent chromatography. In addition, the precipitation process is very simple and could be run in continuous mode to support perfusion reactor and continuous centrifuge operation. The study will also compare with simplex method with a brute force approach and central composite design. The simplex method itself can be run in different ways so this dataset generated by the brute force approach will allow different simplex strategies to be compared.

3.2 Objectives

To assess and compare the simplex method as an experimental design tool for process development challenges, it was applied to a MAb precipitation case study optimising MAb yield and purity (HCP clearance) as a function of pH, ammonium sulphate concentration and MAb concentration. Multiple simplex experiments were undertaken to examine and understand the method's performance in a microwell based experimentation to enable it to be used in the most efficient and useful manner from a bioprocess design viewpoint. Finally the simplex method was compared to a traditional DoE based optimisation. To carry out the above comparisons, an expansive dataset was collated using a high throughput brute force method describing MAb yield and purity as a function of pH, salt concentration and initial feed protein concentration (presented in the response surface diagrams in Figure 3.1).

3.3 Results

Preliminary range finding studies were carried out to assist with the brute force method. Data from the experiments to ascertain pH and salt concentration ranges is included in the appendix.

3.3.1 Selection of precipitation factors

The factors to be investigated by the experimental design methods were pH, ammonium sulphate and initial protein concentration. A wide pH range of 3 – 9 was selected that would be applied by using various buffer systems. Ammonium sulphate was picked as the precipitant due to its widespread use in industrial precipitation and well documented history as described previously. A range of 0 – 2.4 M ammonium sulphate was picked based on preliminary experiments. The feed used in the study was a 3.97 g/L MAb clarified

cell culture (Lonza, Slough, UK). A range from 1 g/L to 16 g/L MAb was generated representing typical titres currently being produced and predicative of future targets. The required concentrations in the study would be achieved by equivalent pH and conductivity buffer dilution or by concentration using ultrafiltration centrifuge tubes. The concentration method ensured the sample would retain the same pre-filtration purity and constant product/impurity ratio (despite the fact that in high producing cell lines, this ratio often decreases with increasing product titre).

As the MAb product would be collected in the solid phase the ageing phase of precipitation (low shear mixing) is vital for high centrifugation recovery. The shear resistant conditioning of the solid precipitate particles preventing breakage, and consequent losses during the solid/liquid separation process. For this precipitation study, the precipitate solids recovery was not being investigated therefore a constant time and mixing speed were used throughout the study of 2 hours at 800 rpm on an orbital shaker based on similar work on the same MAb feed by Knevelman et al., (2009) to ensure the precipitate was well formed. Lab scale centrifuges are also not representative of larger processing centrifuges in terms of shear treatment so a theoretical maximum recovery was assumed without re-solubilising the solid precipitate.

These results facilitated the primary goal of the process that was how the MAb concentration, salt concentration, and pH affected the experimental design space with respect to the yield and purity of the precipitated MAb. All precipitations were carried out in 96-microwell filter plates and assays were performed using high-throughput analytical methods. The precipitation process was conducted at microlitre scale and using the multiprobe automated high throughput platform.

Table 3.1 High throughput precipitation process time accountability

Process time per plate (96 samples)	Time (mins)	Time (% process)
Precipitant and Buffer preparation	20	1 %
Precipitation samples	45	2 %
Ageing	120	5 %
Centrifugation	20	1 %
Automated supernatant siphoning	10	0.5 %
BCA total protein assay	10	0.5 %
HPLC Protein G assay	2000	90 %

Table 3.1 shows the time spent on each task involved in the experiment. The HPLC Protein G assay accounted for 90 % of the time it took to process one plate, roughly 21 minutes per sample. Despite a fully high throughput process, analytical methods such as the HPLC protein G assay for product determination (as with various other examples of high resolution analytical techniques, such as HPLC gel filtration) seek to benefit the most from a reduction in sample numbers. The reduced time cost of analysis and interpretation of datasets is another area that would benefit from efficient experimental design methodologies.

The priority of traditional DoE methods is to try and understand the experimental relationships between factors and responses within the selected factor ranges. The experimental objective (i.e. find the conditions with optimum yield and purity in terms of

HCP clearance) does not play a part in selecting the points of the experimental design, rather they are a pattern suggested from the type of design and the selected factor ranges. The simplex method on the other hand, requires the experiment objective to be computed into it to drive the simplex towards maximizing this objective. Therefore the generation of a meaningful objective function satisfying the design objectives is even more important.

For the precipitation step, the two main objectives were maximizing yield and purification. The simplex method can go after either of these individually however a more efficient approach would be combining the yield and purity data into a single objective function and optimize within a single search. Therefore for the simplex method optimisation, a weighted approach was taken using minimum yield constraints.

Simplex Objective Function (OF):

For Y > 80 %,

$$OF = 0.3\ Y + 0.7\ P; \tag{3.1}$$

otherwise,

$$OF = (0.3\ Y + 0.7\ P)/10; \tag{3.2}$$

where Y is yield (%) and P is purity (%).

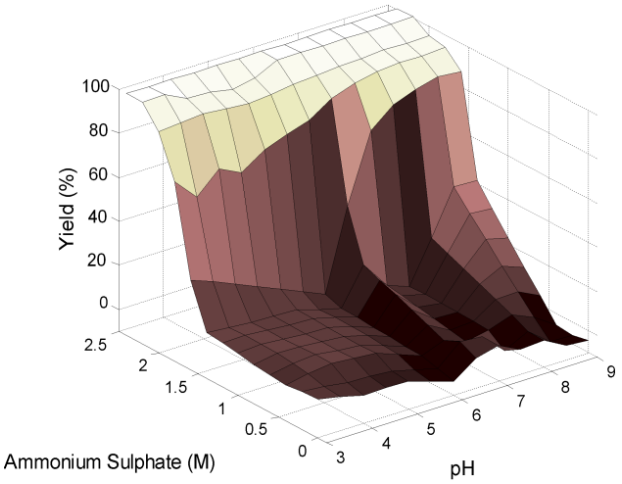
3.3.2 Brute force study

A brute expansive data set was generated using a series of high throughput platforms (liquid handling robot, HPLC, plate reader and microwell centrifugation) to facilitate the methods comparison. Having this high resolution data available meant the DoE and simplex method optimisation solutions could also be generated from it without conducting further experiments. The factors would be limited to the reliably resolvable intervals that are listed in table 3.2.

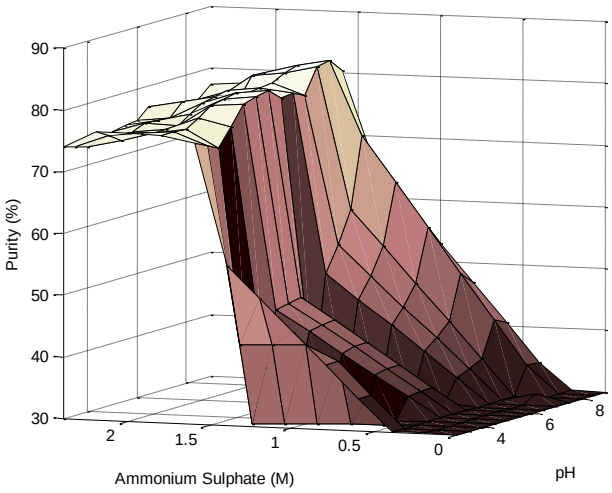
Table 3.2: Factors and ranges used in the precipitation brute force study.

Factor	Range	Minimum interval
pH	3 – 9	0.5
Ammonium Sulphate (M)	0 – 2.4	0.2
MAb concentration (mg/mL)	1 – 16	3

(a)



(b)



(c)

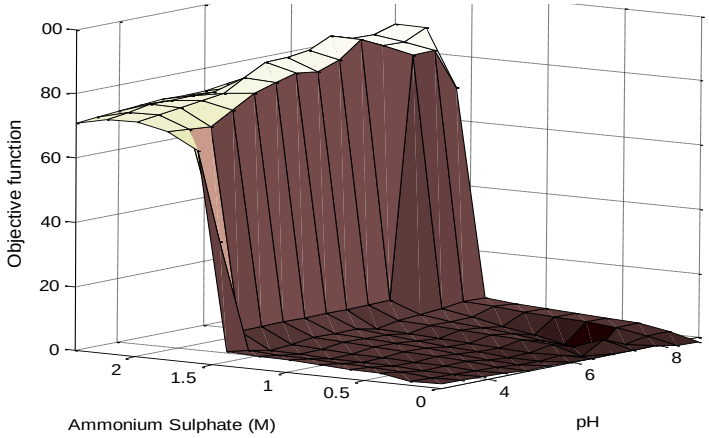


Figure 3.1: Response surfaces from brute force data of MAb precipitation. (a) MAb yield and (b) purity and (c) objective function that was defined by equations (3.1) and (3.2).

Figure 3.1 shows high resolution response surfaces from the brute force dataset for yield (a), purification (b), and objective function (c) (equations 2.1 and 2.2) for the simplex method optimisation. The objective function surface is useful regardless of the simplex method as it shows some of the compromises made in process design to isolate and present feasible operating spaces. All have the third variable, initial protein concentration fixed at 4g/L so the data can be presented in two dimensional plots. As mentioned earlier, the product is in the solid precipitate phase which would be resolubilised as the next step in the process. Hence the yield value in this study is worked out from titre analysis of the precipitation supernatant (i.e. if the titre is 0 we assume all the antibody has precipitated and is in the solid phase). Triplicate data sets were produced showing with good repeatability with the automated process. Variance for yield values was within 3 % of the mean value and for purity within 5 %. The brute force study revealed the ammonium sulphate concentration to be the dominant factor by increasing MAb solubility (salting-in) at low concentrations inducing precipitation (salting-out) at values over 1 M. Protein concentration and pH had minor effects on precipitation, both decreasing MAb solubility at the top end of the ranges. Co-precipitation of impurities does occur although the majority stays in solution. The optimum point of the combined response conditions was at pH 7.5, 2 M, and 10 g/L initial MAb concentration, achieving a yield of 95% at 82 % purity (from an initial feedstock purity of 51%). Figure 3.1 also reveals salting-in causes the yield to stay low from increasing salt concentration conditions from 0 – 0.8 M before salting-out takes over and yield begins to rise at 1.6 M and above. Impurities were less affected by the salting-in phase and consequently precipitated in this salt range.

A yield of 99% is achieved above the salt concentration of 2.2 M regardless of pH and initial MAb concentration clearing indicating the ammonium sulphate concentration is the most significant factor. Initial MAb concentration and pH show smaller effects on the yield and pH 7 – 9 seems to be where the MAb may be least soluble due its pI being in this range, shown by a high yield of 95% even when the ammonium sulphate is only 1.8 M. The optimum combined response is found in two spots, at (6.5, 1.9) and (7.5, 1.8). Neutral pHs, especially between 7 and 8, are most effective in reducing the MAb's solubility without

affecting impurity solubility. While isoelectric precipitation is used in some industries such as plasma fractionation, there is an increased risk of product denaturation therefore the process often targets impurities (Shuler and Kargi, 2002). Isoelectric precipitation employs a different mechanism to alter protein solubility exploiting protein surface charges to promote aggregation and precipitate formation (Virkar et al., 1982). This effect is demonstrated in Figure 3.1 (b) where pH precipitation enhances the precipitation specificity of the product protein. Impurity solubility showed resistance to salting-out in the range tested though it was more sensitive to pH change than the product, precipitating at low pH and salting-in at pH 8 and 9. Poor precipitation of CHO host cell proteins have been recorded previously in literature (Glynn, 2008; Kent, 1999). Hydrophobic residues found on protein surfaces are the means for aggregation when salting-out agents are added to solution, leaving protein precipitation the entropically favoured route to minimise free energy (Kita et al., 1994; Timasheff and Arakawa, 1997). The rate of salting-out is primarily affected by the extent of a protein's surface hydrophobicity and the size of the molecule. As a general rule, larger, hydrophobic proteins require less salt for effective precipitation (Wingfield, 1999). The relative solubilities of proteins is utilized in fractional precipitation processes to purify target proteins (Richardson et al., 1990).

The MAb product in our case is large (~150kDa) compared with the host cell proteins and the impurity solubility profile declines at a much more steady rate over the salt range. Mass spectrometry indicates the higher solubility of the impurities could be attributed to low molecular weight (or possibly few hydrophobic surface patches). Concentrating the feed achieved greater levels of purification as the impurities resisted precipitation whereas the MAb continued to follow its salting-out profile. It was interesting to note that impurity precipitation, did not have a pronounced 'salting-in' phase like the product, and were much more sensitive to pH, precipitating in low pH and low salt concentrations (pH 3 – 5, 0 – 1 M). The data suggests a two-cut precipitation starting with an impurity targeted precipitation and removal, may return a superior overall purification effect. Low salt and low temperature cuts have been shown to remove DNA, ribosomes, membrane fragments and even denatured proteins (Harrison and Roger, 1993). All the product is precipitated

above 2.2 M salt concentrations throughout the pH range, although between pH 7-9 a 100% precipitation is achievable at 2.0 M. The maximum combined response was found in two conditions, at (6.5, 2, 10) and (7.5-8.5, 2-2.2, 10). Neutral pHs, especially between 7 and 8, are most effective in reducing the MAb's solubility without affecting the solubility of impurities. The joint response applies weights of 30% to yield and 70% to purity to all samples above 80% yield, otherwise a penalty is applied. This represents a realistic trade-off that might be used in an industrial process development environment and as such, the response will be used in the comparison of the two methods.

3.3.3 Traditional DoE optimisation

Figure 3.2 shows the results of a commonly used DoE for early process design, the central composite design (CCD). The design used 20 samples including 6 centre point repeats and is represented by a quadratic model formed by least squares regression. The objective function model provides a very general overview of the system and indicates a wide band of feasible operating conditions in the high salt range. The optimum combined response conditions suggested by the DoE were at pH 3, 2.4 M, and 4 g/L with a short and wide robust region.

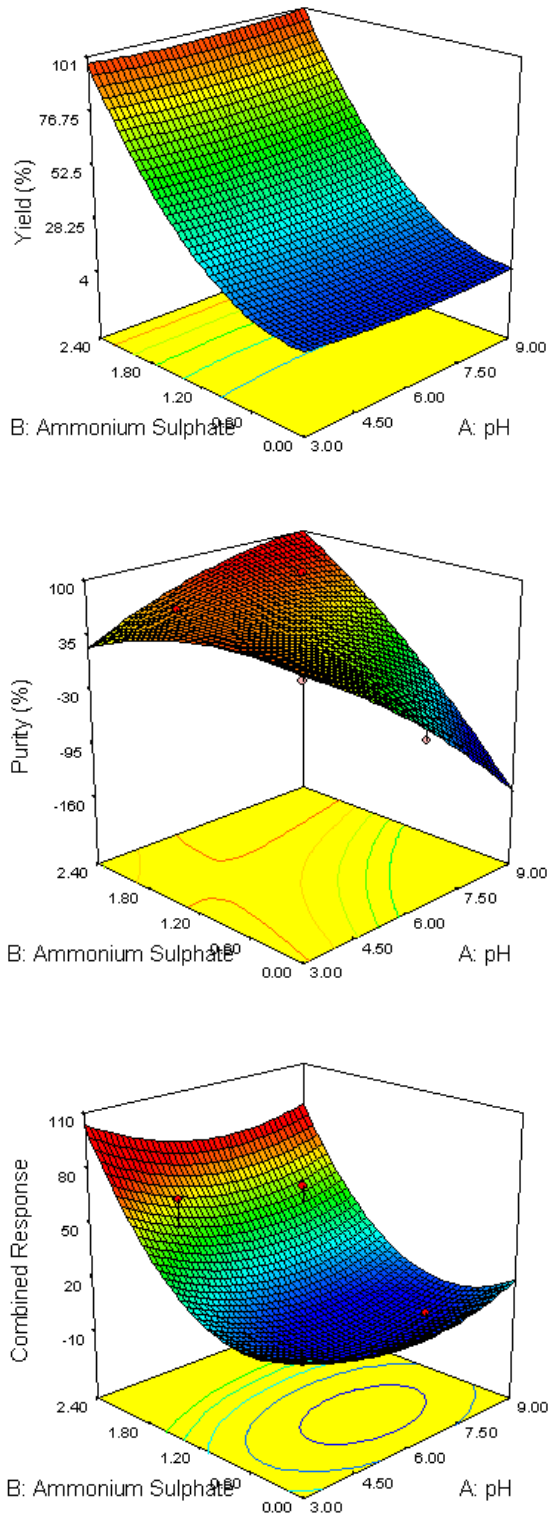


Figure 3.2: CCD DoE model response surfaces. MAb yield (a) and purity (b) and combined response (c) using a central composite design. Data shown for feed concentration 4 g_{MAb}/L with optimum located at pH 3, 2.4 M ammonium sulphate

The low pH would cause a stability concern for the product and its likely a second DoE would exclude the low pH range and it is being shown in the model pH is not a significant factor. The curvature of the system can be seen in Figure 3.2 where despite a quadratic model, the experimental conditions indicated by the dots are still some way off the plane of the response surface. The ANOVA values however (listed in Table 3.3) indicate a significant lack of fit. Whilst the R^2 value suggests the fit of the data to the model is satisfactory, the predicted R^2 is very low, which means the model is unfit for making predictions. The F value is not particularly large so the model is not very significant although the coefficient of variation is low and suggests good repeatability in the samples. Overall, a higher resolution DoE method would have been more appropriate given the large design space and number of factors.

Table 3.3: ANOVA statistics for DoE model

ANOVA data for Combined response quadratic model	
R^2	0.83
Adjusted R^2	0.68
Predicted R^2	-0.27
F value	5.52
Prob>F	0.0067
Lack of fit F value	39500
Coefficient of variation %	9.78

3.3.4 Simplex method optimisation

3.3.4.1 Initial simplex search

Optimisation with the simplex method requires nominating a set of initial conditions from where the algorithm can start searching. A 'good' starting point determines how efficiently the simplex converges on the optimum and helps in avoiding local optima. Just as traditional DoE requires defining the ranges of the search space and a good or bad definition affects the success of the solution, the simplex method lets us make further use of prior knowledge (or preliminary findings) by starting from an approximate trial solution. The initial simplex was deployed around the point pH 7.5, 2.4 M and 4 g_{MAb}/L, selected on the criteria that the MAb's pI is between 7.2 – 7.8, precipitation occurs at higher salt concentrations, and the feedstock was provided at 4g_{MAb}/L. The average of the triplicate responses at each condition was used to drive the simplex.

The stepping of the simplex was limited to within the factor ranges used in the DoE and the minimum variable intervals were set at 0.5, 0.2 and 3 for pH, ammonium sulphate and MAb concentration respectively. Since the DoE method used 21 conditions to construct the model, we limited the simplex method to 21 iterations. Figure 3.3 displays the initial simplex (tetrahedron) and its trail of new points in the three factor design space. In this case the simplex is a four sided tetrahedron and the response at each point is in the fourth dimension so cannot be shown on this graph.

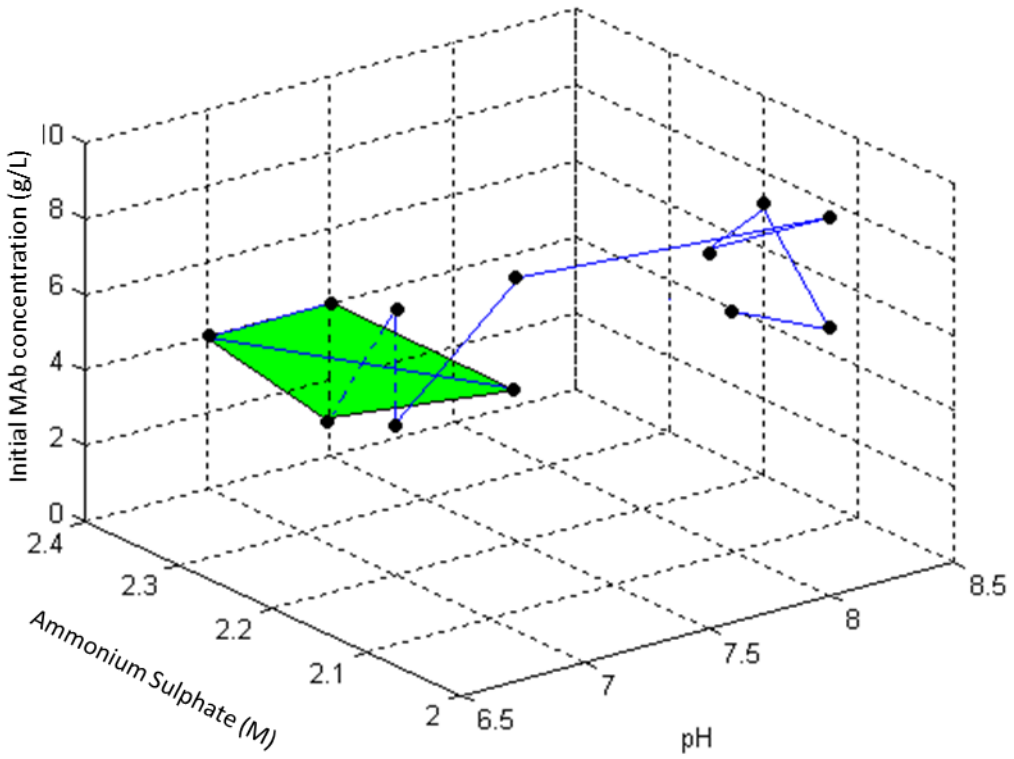


Figure 3.3: Simplex trail in a combined MAb yield and purity design space. The initial simplex is represented by the green tetrahedron.

Figure 3.3 shows the three factor simplex search initiating at (7.5, 2.4, 4), (7, 2.4, 4), (7.5, 2.2, 4), (7.5, 2.4, 1) and converging at an optimum at pH 7.5, 2 M, and 10 g/L, indicating a 95 % yield and 82 % purity. 16 points were evaluated to locate the peak, 12 are shown by the dots in figure 3.3 and 4 unsuccessful points are not shown. The peak was identified by the 9th iteration with subsequent iterations adding to the characterisation of the local region. Figure 3.4 shows the progression of the response values of MAb yield, purity and the combined response at each iteration in the optimisation.

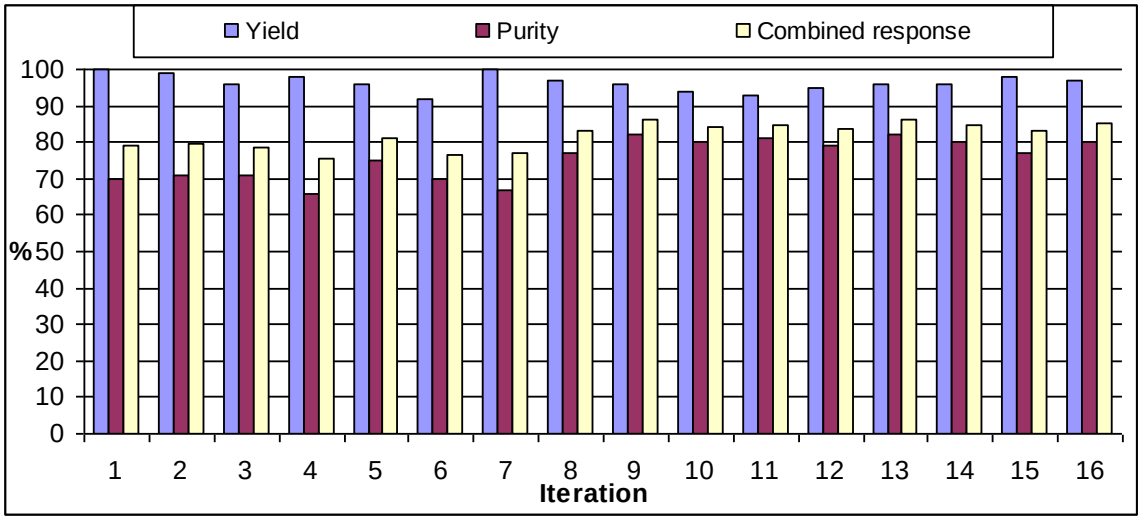


Figure 3.4: Change in response as the simplex search progresses. The combined response was used as the objective function to drive the search.

The differences from iteration to iteration in Figure 3.4 illustrate how the algorithm tries to find the right balance of yield and purity to maximise the combined response objective function. The decrease in the objective function of iterations 6, 7 and 15 represent the unsuccessful iterations where the simplex stepped in the wrong direction, prompting the algorithm to reject the point and try another direction. A peak in the design space was able to be located by iteration 9 with successive points converging around it and characterising some robustness to the simplex method solution. The identified peak is in fact one of the two global optima as illustrated in the combined response surface diagram of Figure 3.1 (c).

3.3.4.2 Initial conditions selection

Using the simplex method to optimize an experimental design space requires specifying the initial conditions, size of the design space and minimum increments between its factors. Selection of the simplex initial conditions can greatly affect how many samples are needed

to find the optimum. If intuition or prior experience can predict a fruitful area of the experiment and the simplex is initiated in this region it is naturally much more likely to converge upon the optimum in relatively few steps. Figure 3.5 demonstrates this principle using a two dimensional simplex method optimisation overlaid by a contour view from the brute force dataset. The initial simplex is shown by the green triangle and is formed at conditions (7, 2.2), (7.5, 2.4), (6.5, 2.4) with final simplex shown in grey.

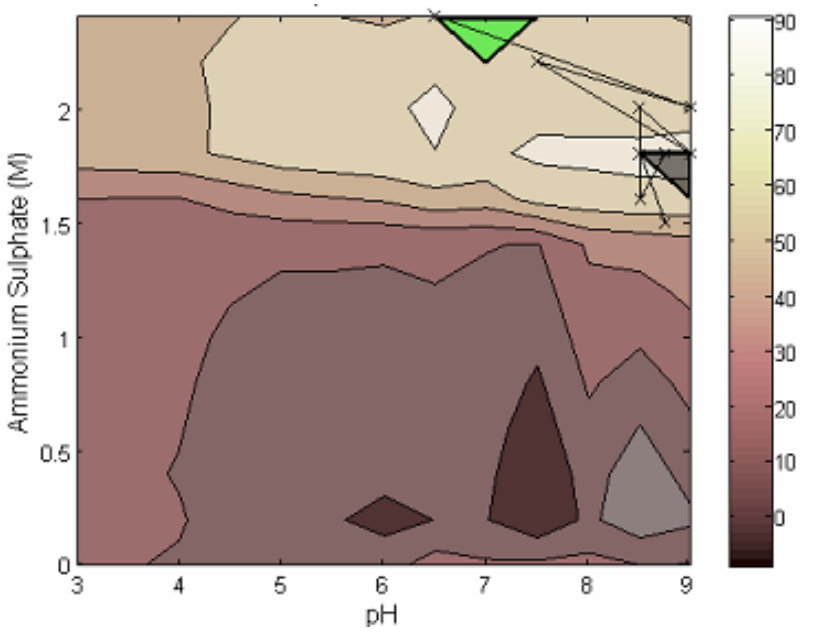


Figure 3.5: The simplex search overlaid on the objective function contour graph.

In this case the optimum is a little distance away and the first experiment results in an expansion. As the rising gradient is detected towards the peak where the simplex finally settles (pH 8.5-9, 1.7-1.8 M) in a total of 8 experiments obtaining a 91% objective function. Once the best point at the top of the peak is found the simplex shrinks as all the adjacent points do not lead it away. Data of these conditions local to the optimum are certainly handy for in depth characterisation studies if the conditions are accepted and taken further. The contour surface in Figure 3.5 has another peak at (6.5, 2) with an equally high objective function, however this peak is not as broad and may not be easily be discovered by the

simplex search. There is a chance it would be discovered by a simplex depending on where the initial conditions were selected so it may be beneficial to run multiple simplex searches from different areas. However a DoE based approach may be more robust in this smaller design space and more likely to discover and characterise this smaller subset of the design space.

3.3.4.3 Finding a local optima

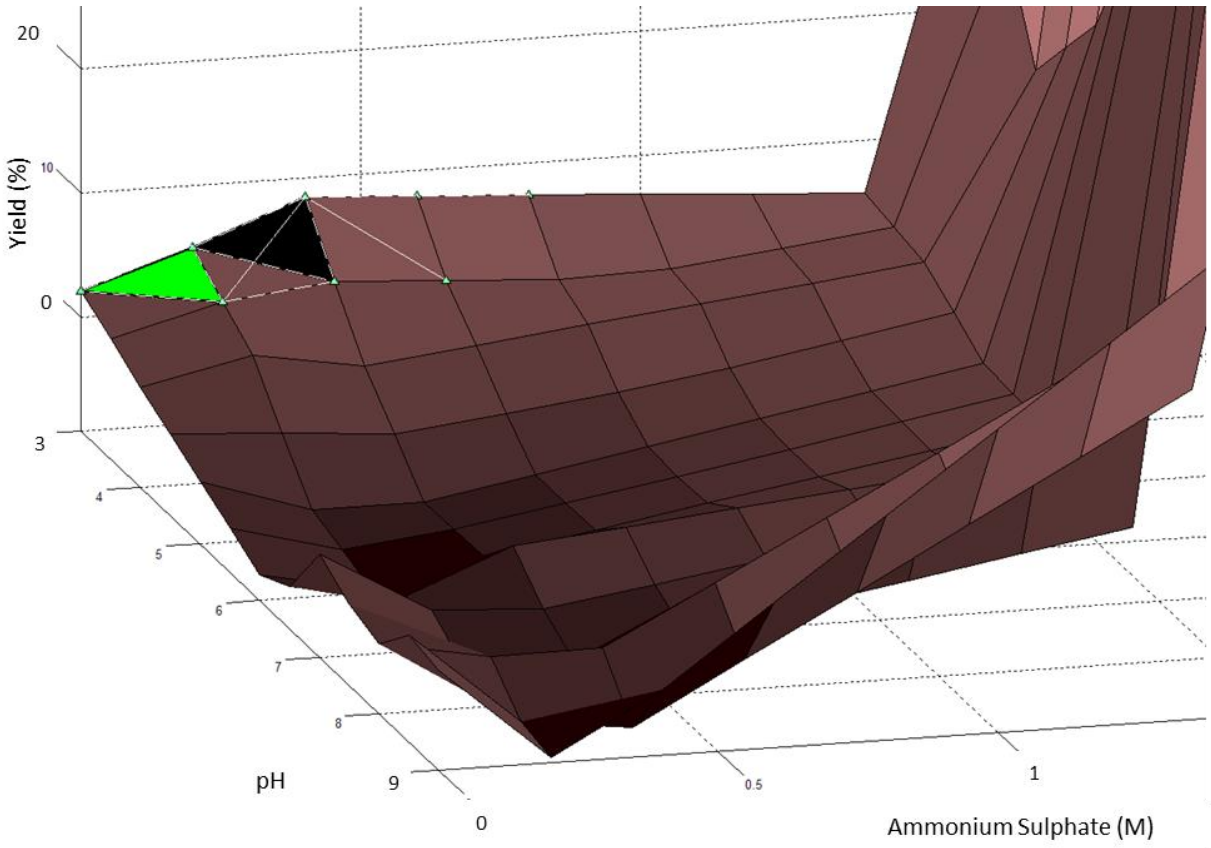


Figure 3.6: Example of a local optimum on the MAb yield response surface. Here the simplex fails to move away from a local optima at (3, 0.4).

One of the main criticisms with the simplex method is the issue of getting stuck at local optima because of its inability to distinguish between local and global optima (Brereton, 1993; Morgan et al., 1990; Aberg and Gustavsson, 1982; Morgan and Deming, 1974). Local optima is a general problem with any optimisation technique although DoE methods provide a more accurate overview of the whole response surface therefore they are more easily avoided. Figure 3.6 is a two dimensional simplex method optimisation of the MAb yield surface, initiating the simplex from a low pH and low salt corner of the design space. The low pH is causing precipitation despite the salting-in behaviour evident across the rest of the pH range. Above 0.4 M the delayed salting-in phase returns, resulting in a local peak at (3, 0.4) trapping the simplex.

The example highlights that the simplex will latch on to any incline in the surface and if there is a local peak with no surrounding superior points, it will not be able move away from it. It has been suggested to deploy a large simplex from a central location that may reduce convergence to local peaks (Morgan and Deming, 1974) however local optima can be a natural part of the response function as well as be caused by noise. Small simplices are more susceptible to noise than larger ones because variation in the response will be typically highest between points that are very close together (e.g. pH 6.0 and pH 6.1). A minimum sized simplex started from the wrong side of the valley (Figure 3.6) will be attracted to the local optimum. This is most easily addressed by using a large simplex or several small simplices intermittently spaced around the design space. The smallest size a simplex can be can also be limited in the algorithm to remain larger than the limits of detection of the assay. The simplex is at most risk to prematurely convergence in flat or zero-gradient regions so the difference between simplex points could be a good indicator of the risk of terminating and level of noise in that particular region. If the simplex terminates close to the optimum it is not so much of a problem as a local peak with a low response value. Avoiding noise induced termination can be achieved by de-sensitising the simplex by increasing its size and less sensitive to minor variations in the surface. By selecting a larger minimum simplex step size, a large simplex will be less likely to become trapped by noise because it is greatest between very small changes in factor conditions.

3.3.4.4 Multiple starting simplices

Multiple starting simplices initiated from spaced locations on the response surface can be used to identify multiple optima or increase the confidence in a global optimum if they all converge to the same point. The interpolated data from multiple simplices is also useful for achieving surface-wide information about the whole system (as would be done in simultaneous DoE techniques) and the more spaced the samples the greater the coverage will be achieved for the design space. In Figure 3.7 two simplices are started far away from each other yet they settle approximately in the same region. Convergence onto a common site does provide some verification of the solution as a global optimum however further experiments around the located site would be required for a characterizing a possible operating region.

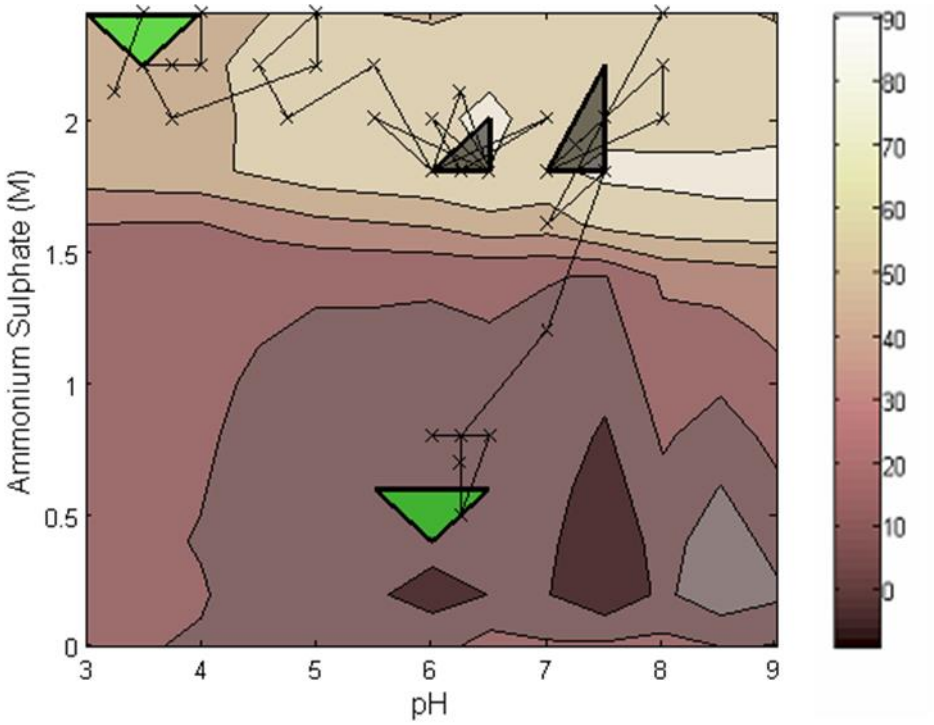


Figure 3.7: Use of multiple starting simplices with the response being purity. Two simplices started from (6, 0.4), (5.5, 0.6), (6.5, 0.6) and (3, 2.4), (4, 2.4), (3.5, 2.2). Each

simplex identified a separate optimum in the design space with the optimum located by the far left corner simplex proving to be greater.

In Figure 3.7 the simplex terminates on the same ridge as the solution in Figure 3.5 although the best point on this ridge is not found. A close-up look at the peak (Figure 3.5 (b)) reveals the level of uneven terrain around the optimum and highlights how noise in the system can prematurely terminate the simplex. However, the simplex is only likely to be affected by noise where the function gradient in the design space is relatively flat such as on a plateau. Using the diminishing change in the differences between the response values at the simplex points, a termination criteria has been set (Morgan et al., 1990). This saves from over-searching and also could be a point where the optimisation could be used as a trigger to switch to secondary experimental design, such as a DoE method as it would be much more informative and accurate in smaller design space around the optimum.

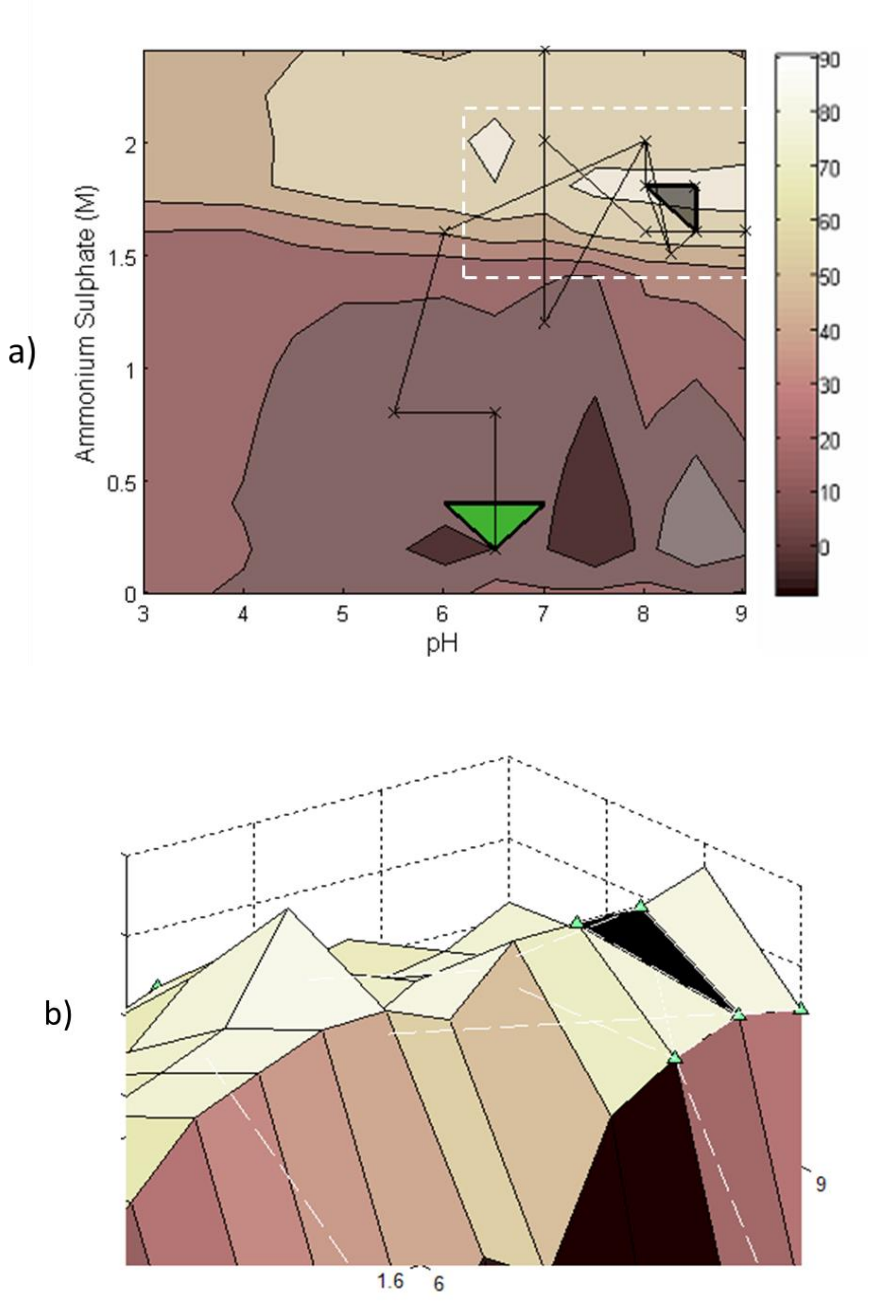


Figure 3.8: a) A simplex search of the same data as in Figure 3.5 but with the simplex initiated from different conditions (6.5, 0.2), (6, 0.4) and (7, 0.4). b) The area around the optimum at (8.5, 1.8) is enlarged and shown in a response surface.

3.3.4.5 Starting location

Varying the starting location of the simplex will not usually change the outcome of the optimisation however each simplex will traverse a different path to the optimum point. As seen previously in Figure 3.7, where two simplices were used as an optimum verification tool, the simplex closest to the optimum will often require fewer samples. This emphasizes the importance of even a basic understanding of the design space to benefit from starting from an ideal location. The distance of the initial simplex (centre of the simplex) to the optimum becomes important when the initial simplex is relatively small. Being close to the optimum the small initial simplex is more manoeuvrable to contract towards an optimum without overstepping. In Figure 3.9 the solutions of a large and small simplex, which approximately have the same simplex centre, are compared. Despite being initiated from about the same centre and settling on the same optimum the larger simplex takes almost 40% more experimental effort.

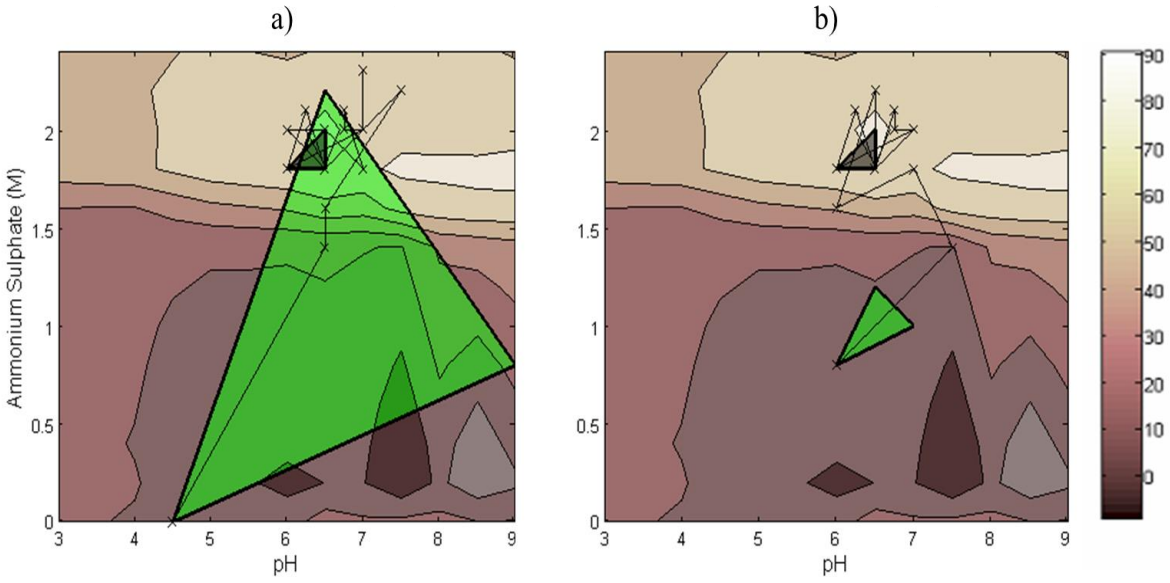


Figure 3.9: Effect of initial simplex size. Two simplices with very close centres requiring 13 (a) and 8 (b) experiments until circulation of an optima. Each cross denotes an experiment; the initial simplex is the lighter shaded triangle.

3.3.4.6 Orientation

Simplex orientation will be referred to as the relative position of the worst vertex of the simplex with respect to the location of the global optimum in the design space (see Figure 3.3). When the worst point of the simplex is facing directly away from the optimum, the simplex is in its best orientation to progress towards the optimum. Orientation will be measured using the angle formed between the lines optimum-worst and worst-simplex centre, as shown in Figure 3.9. In most circumstances minimising the angle formed between worst-optimum and simplex centre will bring about the most productive steps towards the optimum.

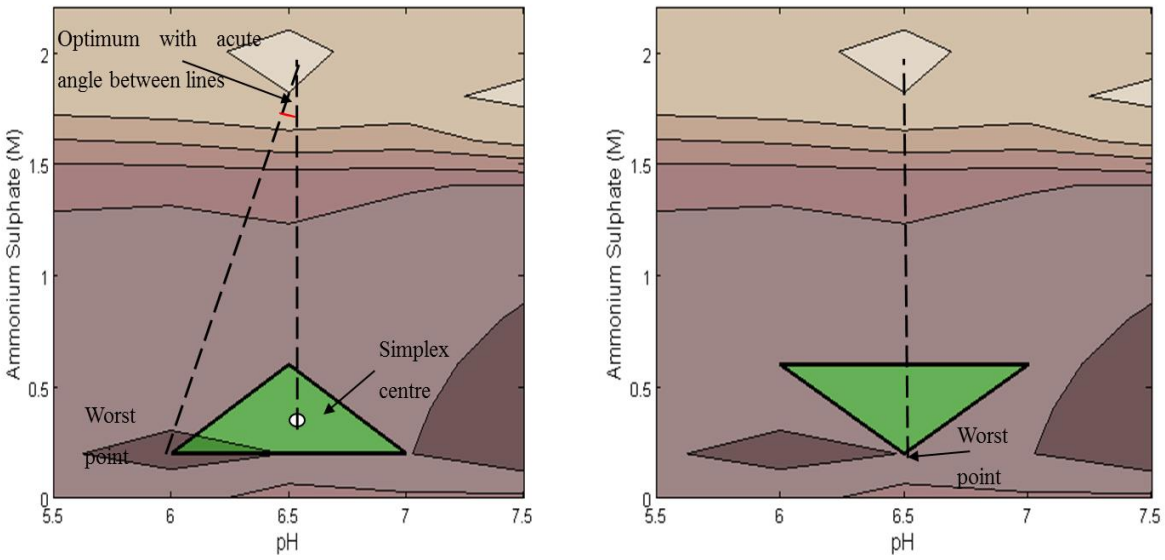


Figure 3.10: Defining the simplex orientation. The angle from the optimum to the worst point is used in the method.

When this angle is near zero then the orientation can be said to be ideal as the simplex will immediately expand towards the optimum. Thus getting an idea of the global optimum by using a crude DoE of the design space could help orientating the simplex in conserving

samples. In Figure 3.10, two simplices of identical size and centre but of different orientations (rotated 180^0 about the simplex centre) resulted in the same solution but at different search costs.

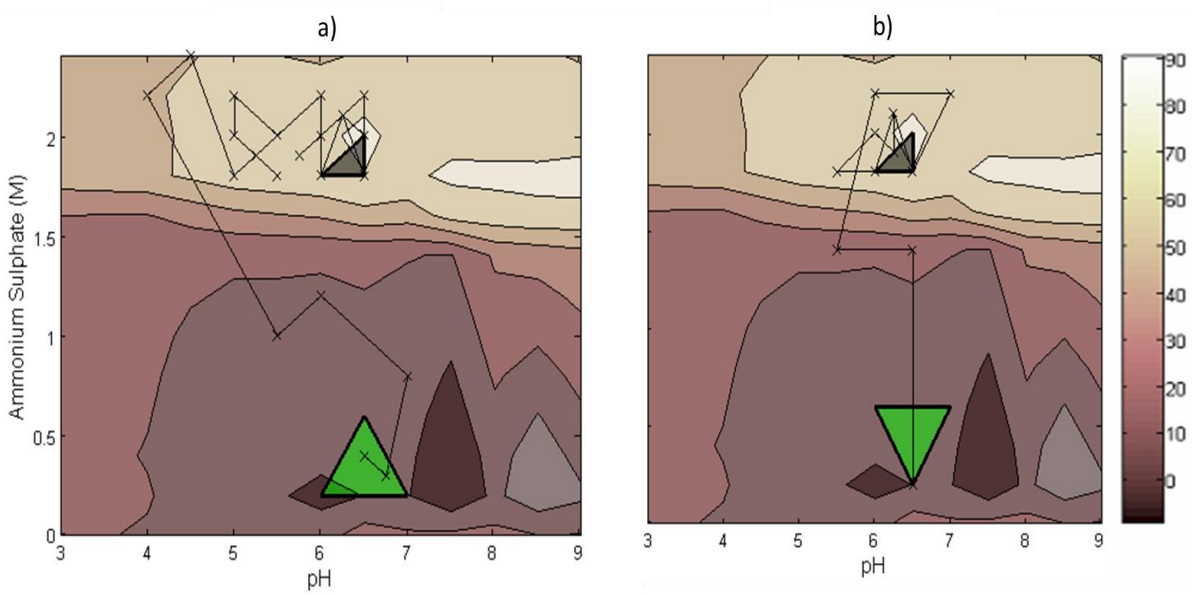


Figure 3.11: The effect of initial simplex orientation on search progress.

Figure 3.10 (a) begins ideally orientated with its lowest response vertex furthest or facing away from the optimum area. As the simplex algorithm works primarily by reflecting away from its worst point, this results in the simplex straight away moving up the slope in its first reflection that results in an expansion. The first reflection in simplex 3.10 (b) is not as productive leading to further steps, which re-orientate the simplex before it makes any decent progress. These examples of how initial simplex orientation can directly impact the algorithm's search cost are a source of variation from the simplex search. Both optimisations identified the same optimum but simplex (a) required only 15 samples as opposed to 36 samples of the poorly orientated, simplex (b). To make the search more robust and repeatable, it is imperative the search be conducted using a systematic set of protocols to eliminate the differences from using the simplex search in alternative ways.

3.3.4.7 Simplex size

A small initial simplex usually shows a rapid rise in response (due to simplex expansions) and then slowing down as an optimum approaches (contractions and shrinks). Small simplices are more sensitive to the minor variations in the surface so can quickly sense a good direction to move in and then expand. In our case the response surface is very steep around the area of the optimum and the simplex gains the greatest rise in response here due to the steep slope. Once a ridge is found the simplex will start to contract and shrink as the slope is much shallower here and finding the right orientation towards the optimum requires greater sensitivity. Once the optimum point is found the simplex will circulate it and repetitions of the surrounding points will begin until the algorithm is terminated. However the extra surface sensitivity makes the small simplex more prone to noise and consequently premature termination. Noise will not usually affect the progression of the simplex as the simplex moves by comparing points but when the points are adjacent to each other as in a small simplex and the gradient is generally flat the simplex is at greater risk on settling on a noise induced optimum. However when the small simplex is close to a genuine optimum, it will get there in fewer steps as well as providing a thorough characterisation of the region.

The simplex expansion takes a larger step in an attempt to arrive at the optimum in fewer samples. Upon approaching the maxima the expanded simplex will shrink before terminating around the peak. Expansions increase search efficiency especially when the simplex is initiated very far from the optimum point as the large simplex transverses very quickly through the design space. The process of expanding and shrinking the simplex requires the evaluation of extra points so in smaller design spaces it is often better to begin with a large simplex in the first place. Small initial simplices should be used when starting close to an estimated optimum whereas larger simplices are better for less informed design spaces.

3.3.5 Modelling with simplex data

Starting with a large simplex does provide rapid movements across the response surface and in a lot of cases is advantageous in quickly reaching the area of the optimum. Another possible use of a large simplex is using data from its wide-ranging steps to fit an empirical model. The assumed sparse data of a large simplex should be better in estimating the response surface than a smaller simplex with a more remote pathway. This feature is also useful for verifying the presence of other optima in the design space. The interpolation from the points used in a large simplex optimisation will indicate potential optima sites that can be further investigated. Figure 3.12 shows such an interpolation with the partial response surface created from only the points used in the simplex search. The surface fit to the simplex data uses triangle based cubic interpolation based on Delaunay triangulation (Barber et al., 1986) accounted for 40% of the total design space.

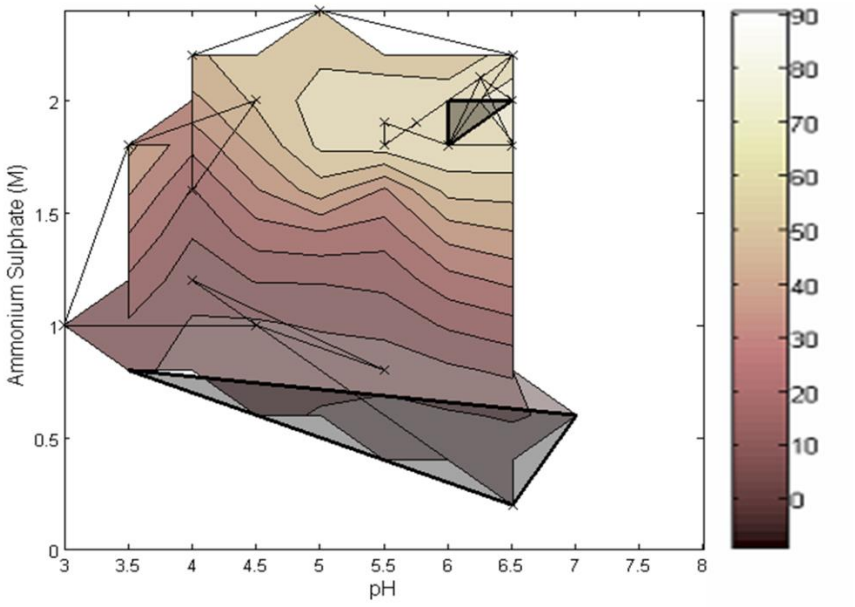


Figure 3.12: Interpolating between simplex points to create a partial response surface.

The accuracy of the interpolated regions of the map in Figure 3.12 are strongly dependent on the density of simplex points within that region. Although a simplistic linear model is used between points, a good prediction will exist where there are clusters of points. This is most apparent around any optima on the surface where the simplex has converged upon. A small starting simplex usually follow a narrow path to the optimum (expansions only take place in the direction of the optimum). Using its data to build a model of the response surface will lead to greater uncertainties where points are scarce (unexplored regions). The accuracy of the model in predicting the space should be used with caution as interpolated data near a cluster of points will be more accurate than areas where points are sparse. For optimisation purposes however it is usually the peaks that are most important anyway.

So if the experimental effort is to be used in contracting and expanding it makes sense to start big and let the simplex collapse around an optima. Simplex points can be used to create partial response surfaces that would provide sparse information in low response space and denser information along the path of the simplex. The points of a large simplex will provide models with greater coverage due to its sparse data points scattered over experimental space. The greater coverage should provide more robust models with better prediction capability. This approach is shown in Figure 3.13 where the initial conditions for the simplex search were selected to be wide apart. The search required 18 samples and using the delauney interpolation technique, provides a coverage of 63% of the total design space.

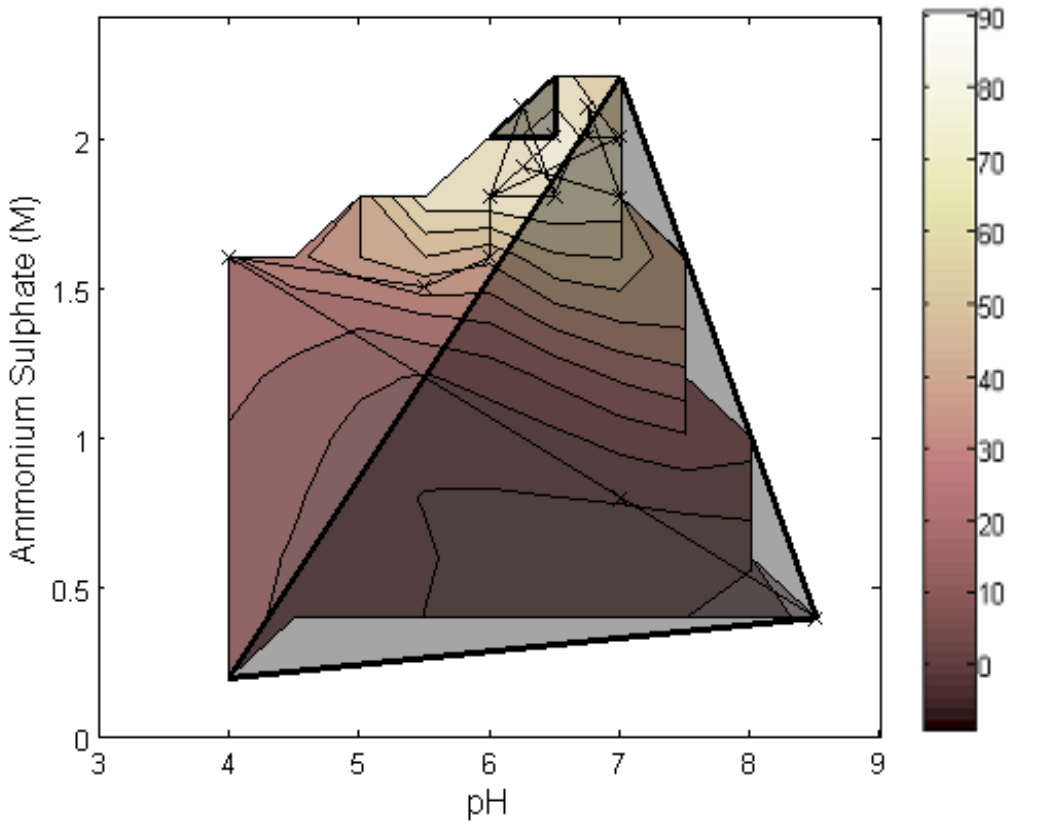


Figure 3.13: Interpolated data from a large initial simplex.

The large initial simplex started centrally was the most successful characteristics found in this study. Not in terms of reducing experimental workload but in being reliable enough to always find the optimum and provide greater information about experimental space. The most efficient simplex was the small simplex started close to the optimum but this requires prior system understanding or some luck. For early research and development work the large simplex was most dependable in focusing experimental effort into the region of the optimum. The use of multiple starting simplices from multiple corners of the design space can provide more coverage and predict with greater certainty.

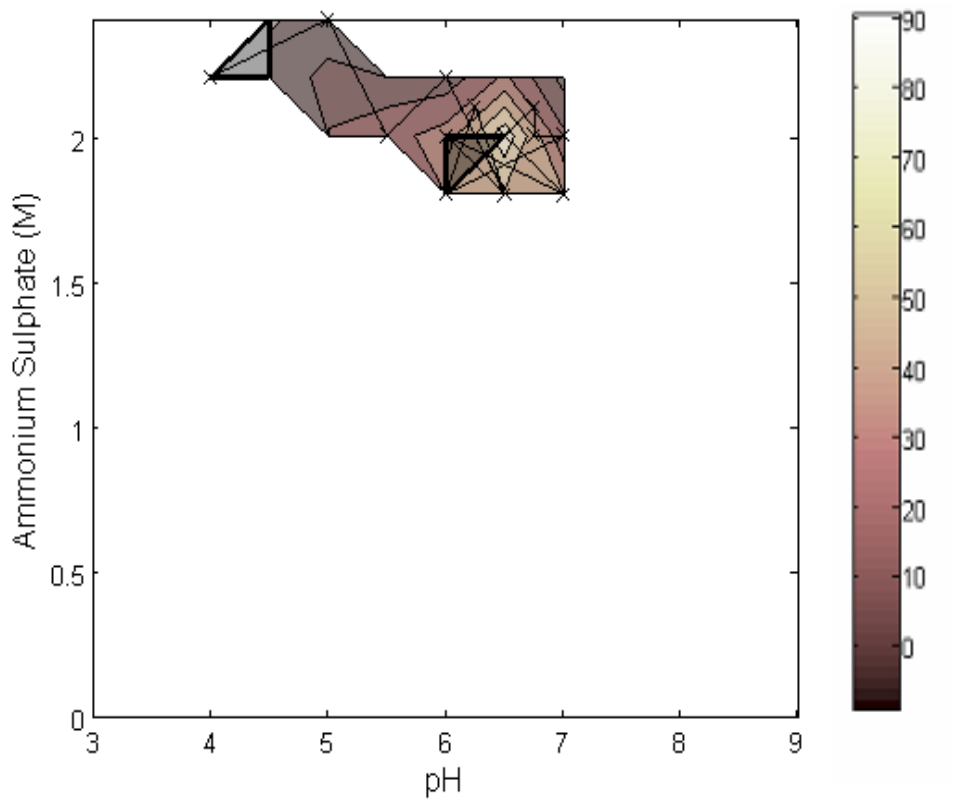


Figure 3.14: Interpolated data from a small initial simplex. Started from (4, 2.2), (4.5, 2.2) and (4.5, 2.4) the model is based on 18 samples but provided coverage for only 9.5% of the design space, although the interpolation is likely to be very accurate.

Figure 3.14 shows a very small simplex started much closer to the optimum than the larger simplex. The narrow pathway it creates is a very accurate match to the real experimental surface however the model covers less than 9.5 % of the total search space. The model required 18 samples to be created and the high density of points within the partial contour map provide very accurate predictive capability. The example in Figure 3.13 also used 18 samples but provided a 63% coverage. This additional mapping of the design space allows a deeper understanding and provides leads into areas worth investigating. Whilst both optimisations are successful in finding the optimum, the selection of the initial simplex in Figure 3.6 does not benefit from starting near the optimum, highlighting that the selection of initial conditions are less important to large simplices. Selecting the starting conditions

of a large simplex does not require much thought as it is expected to collapse onto the optimum, whereas a good selection of initial conditions for a small simplex greatly influences how many steps it will take to get to the optimum. In addition for a large simplex it is much more likely that at least one of its vertices will be close to the optimum to where it will shrink towards. When points are selected far away from the optimum the simplex will evaluate one extra point each step (iteration) to expand and then one extra point in each contraction and two extra points in shrinking. The experimental cost of expanding is avoided in Figure 3.6, which may be partially responsible for the similar total experiment number.

3.3.6 Monte Carlo simulation

Knowledge of multiple peaks allows further operation constraints to be evaluated such as robustness of operation. A good operating area should have a good response and be 'spread out' so any operational deviances are tolerated by product/process quality. A statistically accepted way of assessing such scenarios is using Monte Carlo simulations. Monte Carlo analysis uses simulations of simplices of a random size and from random starting points run multiple times to view the statistical performance of the system. From the Monte Carlo analysis (Figure 3.15) the global optimum was achieved in 90% of the simulations with the mean sample number of 19. These numbers are very much dependent on the shape of the response surface and will differ accordingly. Along with the actual function that determines the relationship between parameters and responses, the error from system noise and the resolution of the experimental grid (feasible factor intervals) can also affect how the simplex progresses. However the latter two are only ever likely to lead the simplex astray where the response surface is relatively flat (response gradient and difference in vertices close to zero) and the simplex size is small. This is easily addressed by limiting simplex size according to noise levels. In poorly understood research areas it is sometimes necessary to run the simplex from randomly selected initial conditions. In Figure 3.15 the results of a Monte Carlo analysis of 1000 random simplex simulations are presented.

The Monte Carlo method used to calculate the probable number of samples needed for the simplex method to isolate the global optimum in the brute force purification dataset. By running many simulations starting the simplex from completely random conditions (1000 simplex method searches) the Monte Carlo analysis shows the probability distribution of how many samples were needed to find the optimum. Almost 900 of the simulations managed to find the optimum. The average number of experiments used in a simplex search was 19 experiments although a quarter of all simulations was able to locate the optimum in 9 experiments or under. The initial conditions of all the simulations were selected randomly.

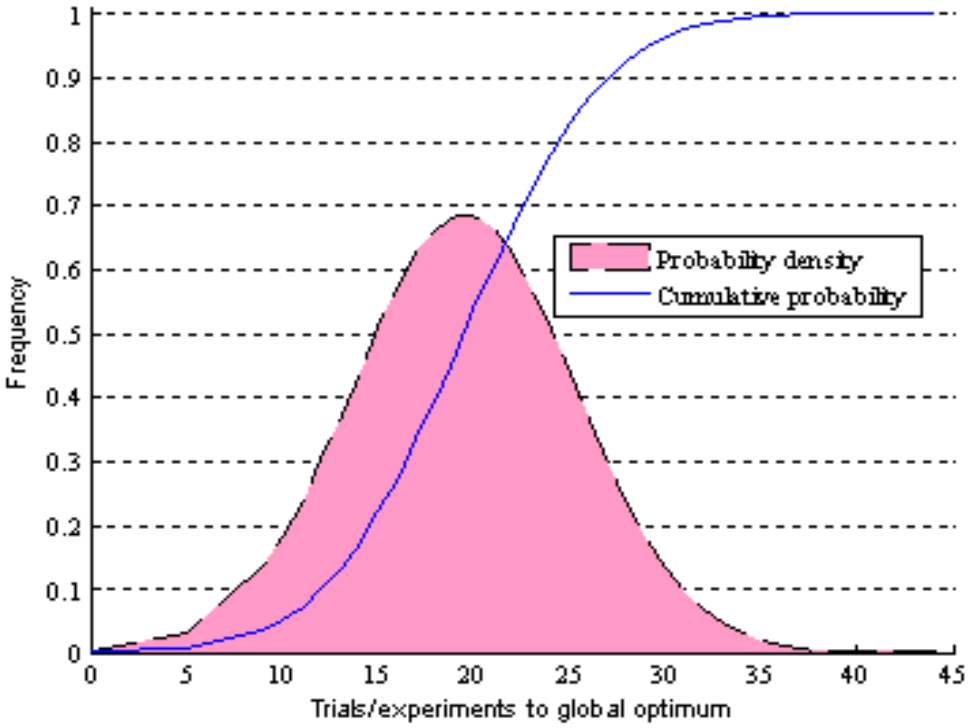


Figure 3.15: Probability distribution graph of Monte Carlo simulations. A 1000 simplex searches were simulated on the MATLAB software with randomly selected start conditions.

Of the 1000 samples 124 were unable to locate the global optimum and settled on local peaks in the system. The robustness of the technique is demonstrated as almost 90% of

simplex searches started from random conditions were successful in a design space with known local optima.

The simplex method was designed to optimise functions in a continuous, infinite space unlike experimental optimisation that is bound by a finite set of conditions making up the design space. Therefore to keep the selection of new simplex points within our chosen factor ranges, as well as choosing conditions of an practical resolution (e.g. selecting pH conditions to the nearest 0.5 units rather than 0.05), some modifications are needed to the simplex algorithm.

The successful Monte Carlo simulations shared certain simplex features that suggest the initial conditions could be used to influence optimisation efficiency. These include the location of the initial conditions, the size or internal area of the initial simplex and its orientation with respect to the boundaries and the global optimum.

The simplex method will provide the exact processing conditions where the best response is obtained usually at a low experimental cost. Noise is most prevalent where the gradient of the response surface function is flattest such as on ‘plateau-like optima in Figure 3.6 and can slow down the simplex progress. Premature termination of the simplex is unlikely unless local peaks are stumbled upon.

However, since much of the surface remains unexplored it would be wise to run another, larger simplex to verify the provided solution. Larger simplices have shown to be more successful in uncertain search spaces as its larger steps cover much more of the surface and have a higher probability of coming across multiple optima (Deming and Morgan, 1973). Using more than one simplex started from the different corners of the design space can also be used in combination to greater coverage of the design space, increasing the confidence

in the solution. The simplex method does not predict an optimum like DoE but searches for it and provides its exact location as well as data for characterisation of the surrounding area.

Experimental noise is less detrimental to search progress as the simplex method only compares the responses at the simplex vertices and the magnitude is unimportant. Determining the true optimum with traditional DoE methods requires multiple rounds of experiments depending on model complexity. The simplex method has shown to cope well with complex response functions and in high dimension search spaces whereas the sample requirements of traditional DoE techniques become overwhelming (Plakett and Burman, 1946).

3.3.7 Combining DoE and the simplex method

Comparing experimental design techniques and sequential optimisation methods is not straightforward as the former provides the solution using computational curve fitting prior to any experimentation while the latter arrives at it directly through experiments. If our goal was to find the optimum in the least number of samples the simplex may prove the efficient option however the solution from a single simplex will lack the confidence that DoE techniques impart. The average samples used by the simplex method in the Monte Carlo simulations was 18, which represents a 33% saving over the central composite design and 70 % over the D-optimal design. However in 10% of simulations the global optimum was not found. The points evaluated in the path of the simplex can be used to estimate the entire surface using delauney triangulation (see Figure 3.12-3.14) although it only provides good prediction in areas where there are a high density of points.

Table 3.4: Comparison of DoE model and simplex data. Comparing the 2 alternative DoE methods with the simplex method average from the Monte Carlo analysis

	CCD	D-optimal	Simplex method
Samples	27	39	18 (mean)
F-value	18.46	14.7	
R^2	0.959	0.93	
Predicted R^2	0.605	0.646	
Signal/Noise ratio	10	11.2	
Coverage (%)	100	100	~

Simplex features such as initial simplex volume, its position and orientation affect the success of the optimisation however all these depend on a little surface knowledge to use wisely. For this reason a low resolution screening study of all the variables can be used to set the ideal simplex conditions. Conditions to use would be a starting position close to the optimum, a favourable orientation facing a predicted optimum area and a size dependent on how local the optimum area is and the noise levels in the system. If the DoE shows multiple optima a simplex can be started next to each one. In Figure 3.16 the objective function design space is divided into four equal segments and a low level DoE was run for each one. Each DoE segment can be evaluated for high response and the decision taken if a simplex should be deployed. In the top two segments the deployed simplices reveal the location of the true optima that the DoE model has missed. For the lower two segments it was decided not to deploy the simplices as the DoE model did not indicate any favourable response areas.

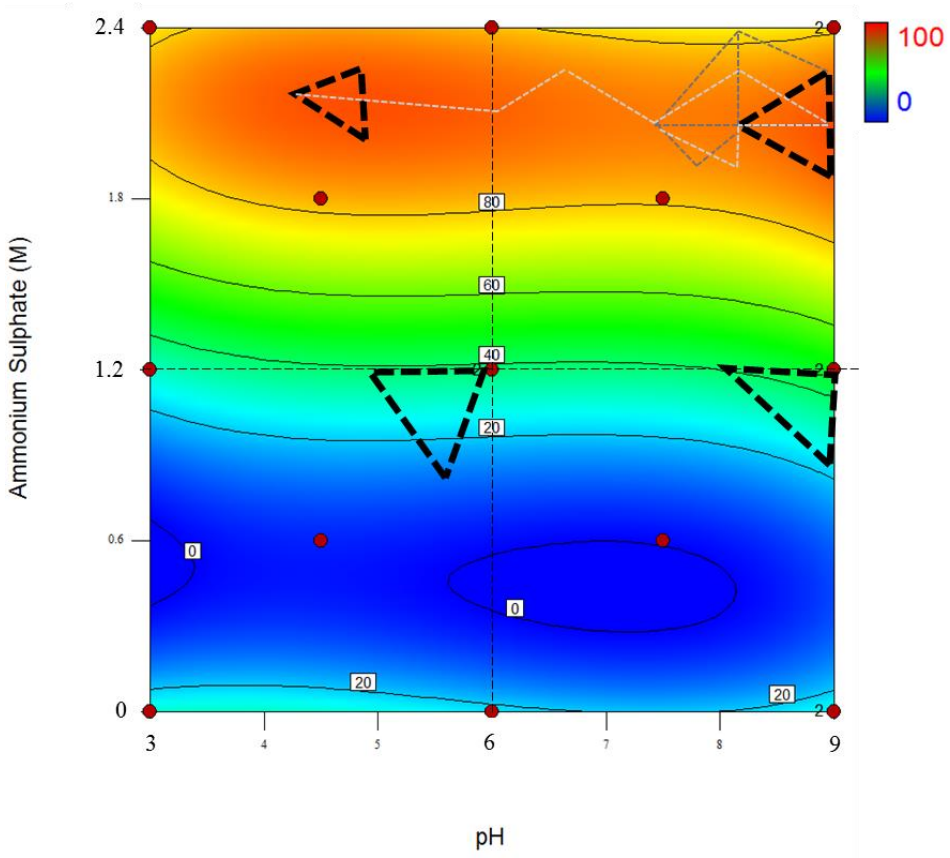


Figure 3.16: Design space segmentation. Use of DoE and the simplex method to provide comprehensive information about a response surface. The simplex provides detailed location of the optimums and surrounding area while the DoE finds suitable starting places for the initial simplex.

The combined approach gains from the strengths of both methods as well as covering each other's weaknesses. The DoE provides enough surface information to prevent the simplex getting stuck on local optimums whereas the DoE on its own cannot determine the exact location of the optimum. The design space is divided into 4 equal parts and from the most promising area in each subspace a simplex may be started depending on whether the researcher judges any part of the DoE response is worth investigating. The DoE data will also provide extra information on how to orientate and select simplex size.

Alternatively we can use the simplex method alone by beginning with a large simplex and letting it collapse onto an optimum or by using multiple smaller simplices spaced about the grid. Both these approaches will most likely require less experimental cost however provide less comprehensive data and carry a risk of locating on local optimums (of course using multiple simplices lowers this risk).

It was also noted that in cases where the simplex started near the boundaries of the design space it tended to spend a lot of time there, evaluating a lot of points. This is of course a product of the function of the response surface however it is suggested to avoid starting simplices near the boundaries to conserve experimental cost. Boundary violation rules in the program tell simplex to reflect in the 2^{nd} most favourable direction (as best direction is off the map). This way the simplex crawls slowly until it is away from the boundary. Movements defined by boundary violations rules are inefficient compared to the standard algorithm rules in open space and are for keeping the simplex moving and not degenerating (e.g. a two dimensional simplex becoming one dimensional).

Where the simplex method should really excel in bioprocess optimisation is when the number of variables is raised to 3 and above. Simplex allows researchers to continue investigating in a multivariate design space without resorting to eliminating less influential system parameters (out of the equation as is commonly done in routine DoE optimisations). The scalability of the simplex method should allow a reasonable experimental cost when handling many variables, whereas traditional DoE would be applied after reducing the number of variables.

3.4 Conclusions

We found that the size of the initial simplex played a pivotal role in noise affected datasets as larger initial simplices were a good choice for noise prone data and estimating factor interactions by creating a model for the increased coverage.

Traditional DoE and the simplex method have been successfully applied to the optimisation of MAb precipitation. The automated microscale methods and assays involved required microlitre quantities of the MAb feed material making the optimisation rapid and economical. The simplex method demonstrated it can quickly locate the optimum in terms of sample numbers. Further assessment of the simplex method revealed the need of useful surface information can be addressed using a suitable initial simplex strategy. The experimental noise present in microwell platforms can hinder experimental design but the simplex method copes admirably due to only requiring a comparison of samples (ranking them best, next to best and worst) and not their numerical value. This feature of the simplex method makes it appropriate for use with experimental factors lacking absolute responses (such as visual assessment or highly subjective quality scoring methods).

The DoE-simplex-DoE approach with a structured use of the simplex search method can be generically applied to bioprocess optimisation to thoroughly characterise robust operating ranges from minimal prior system knowledge. The combination of designs focuses experimentation at the optimum points in the design space. The initial DoE data provides the simplex method with the right starting features so it can locate suitable optima in just a few steps. The final DoE experiments can characterise the located optimum in sufficient resolution. It is possible that some of the simplex search conditions are identical to the DoE conditions allowing re-use of these response values or at least providing a replicate of those conditions.

The simplex method was able to successfully provide potential optimum operating conditions as well mapping the surrounding conditions within the design space. The optimum area with respect to a weighted response of MAb yield and purity were found by the brute force study at pH 7.5 – 8.5, 2 - 2.2 M ammonium sulphate, and a feed concentration of 10 g_{MAb}/L. This robustly achieved at least a yield of 96 % and 82 % purity within the specified ranges. On the other hand, the DoE approach predicted the optimum at the conditions pH 3, 2.4 M, and 4 g/L feed concentration. The simplex method also located the correct optimum in half the number of samples used by in the central composite design. The results clearly indicate a potential time and resource benefit of using the simplex method over the traditional DoE approach for process development. Having assessed the key factors of the simplex method such as initial simplex size, orientation and location, a combined DoE-simplex strategy seems to be most efficient option for the precise exploration of experimental design spaces.

Experimental design in bioprocess development is not straight forward due to the biological complexity and high specificity from product to product. Therefore it is naive to rely on modelling only and much of the work undertaken in research labs is wholly empirical. The key elements of modelling: how to derive the model structure from theory, how to simplify the model based on biological property assumptions and how to validate the model, will be presented.

4. Optimisation of PEG precipitation and a precipitation–centrifugation sequence using the DoE-simplex methodology

4.1 Introduction

As we discovered in chapter three, it is possible to reduce the number of experiments required for system optimisation by careful selection of the simplex conditions. We implemented the findings into a package combining it with a traditional screening DoE method. The DoE-simplex protocol proposed here is applied here to a more complicated study combining two unit operations, precipitation and centrifugation to create a five factor design space. The case study will also serve to optimise the primary recovery process for the MAb we have been using.

A three factor PEG precipitation step is investigated as an alternative precipitation mechanism using the combined DoE-simplex protocols. The case study serves as an introduction to the issues faced when optimising problems in higher dimensions and the difficulty in relaying the results. We then move on to a difficult five factor precipitation and centrifugation sequence. The proposed methodology is used to find and characterise the large space composed from variables from both unit operations. A high throughput experimental approach is maintained, which also uses ultrascale down principles so the microscale work is representative of process scale operation.

4.1.1 Optimisation strategy

The methodology proposed here uses DoE and the simplex method so that the overall optimisation benefits from the advantages of both techniques with none of the drawbacks. The combination of the two methods covers up each other's weaknesses and provides an end solution that neither could achieve alone. The strategy framework is presented in Figure 4.1.

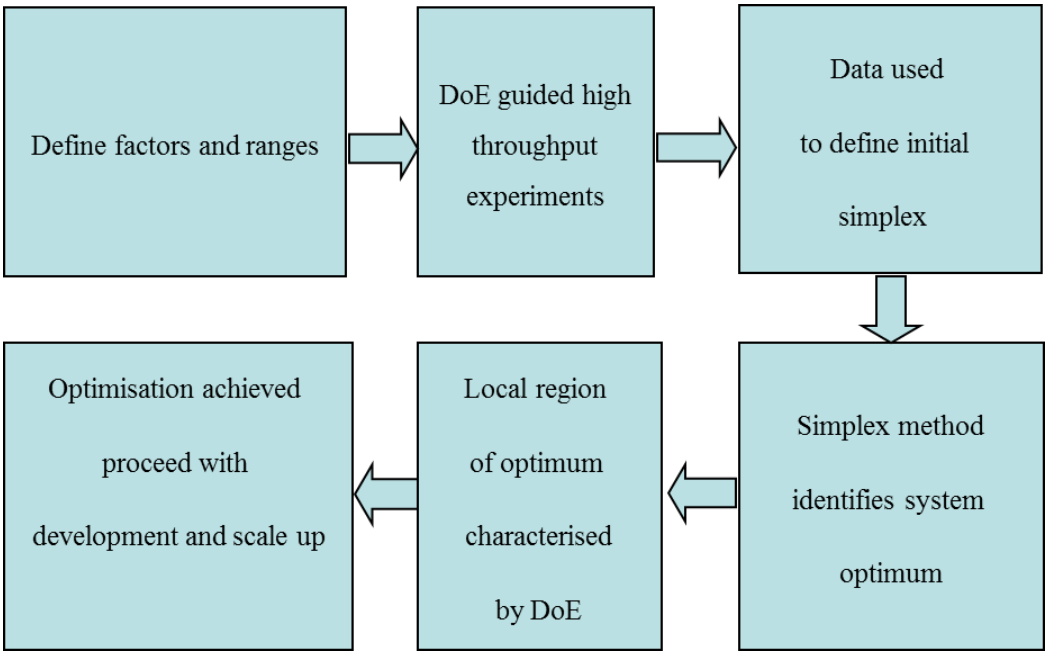


Figure 4.1: DoE – Simplex methodology process flow chart

4.1.1.1 Initial DoE

The approach begins by selecting the variables and range setting. This will require having prior knowledge of the study or preliminary experiments will be needed to ascertain the variables. If there are a lot of variables, screening DoE methods can be used to identify the most important. Biological systems often have many complex interactions and factors can be pH, concentration of salts, amino acids, enzymes, gases, mixing and flow characteristics etc. The responses are equally varied and range from product concentration, DNA clearance, HCP clearance, total protein concentration to overall yield and purity. Models can present a combination of these to suggest an overall optimum operating conditions and likewise the simplex method can optimise according to the weighting in the combination set so it is very important that the weighting between responses is carefully selected. The simplex method can also use for complex weighting and penalty systems as demonstrated in chapter three where a poor yield for precipitation resulted in a greater penalty being applied to the objective function.

Regular 2 level fractional factorial design provides a good basic understanding between the factors and responses, as well as to locate a potential region of the design space to evaluate further in search of the optimum conditions for the objective. Varying the factor only over two levels is unlikely to provide sufficient data for building a useful model or clearly identify, which variables are the most important however the aim here is to identify ideal conditions to deploy the simplex. However, we can use this crude model for the next stage of the optimisation strategy.

4.1.1.2 Design space segmentation

The factorial model of objective function response is then used to find its estimation of the global optimum conditions. This point and its surrounding region in the design space are then 'cordoned off' into their quadrant, bounded by 2^n half-axes. This subset of the design space will be used in the selection of the vertices of the initial simplex.

The conditions at each corner of this n-dimensional hyper quadrant are then numerically evaluated using the factorial model formed in the initial DoE experiment. The number of samples is dependent on the number of variables forming the design space (see Table 4.1).

Table 4.1: The size of the hyper-quadrant and number of vertices

Number of dimensions/factors	% Design space occupied by hyper-quadrant	Vertices (corners) of hyper-quadrant
2	25%	4
3	12.5%	8
4	6.25%	16
5	3.125%	32

From the values of the numerically calculated quadrant corners the vertices of the initial simplex are selected (see Figure 4.2). To select a simplex with favourable orientation, the simplex must face the (to be discovered) optimum by having its worst value point furthest away from it. Since the DoE optimum is only an estimation of the conditions, our method

appoints the quadrant corner closest to this point as the favourable region the initial simplex must be orientated towards. And so is then barred from being selected as a potential simplex point. Rather, the quadrant corner directly opposite this 'optimum' quadrant corner (perceived to be the worst point in the first simplex iteration) is selected as the first choice for a simplex vertex. The remaining simplex vertices are then selected from the *n* highest value corners of the quadrant (apart from the barred 'optimum' corner). With protocol for initial simplex point selection we take account for simplex orientation, size, and location in the design space so the simplex is forced to search towards the region where the optimum has been suggested.

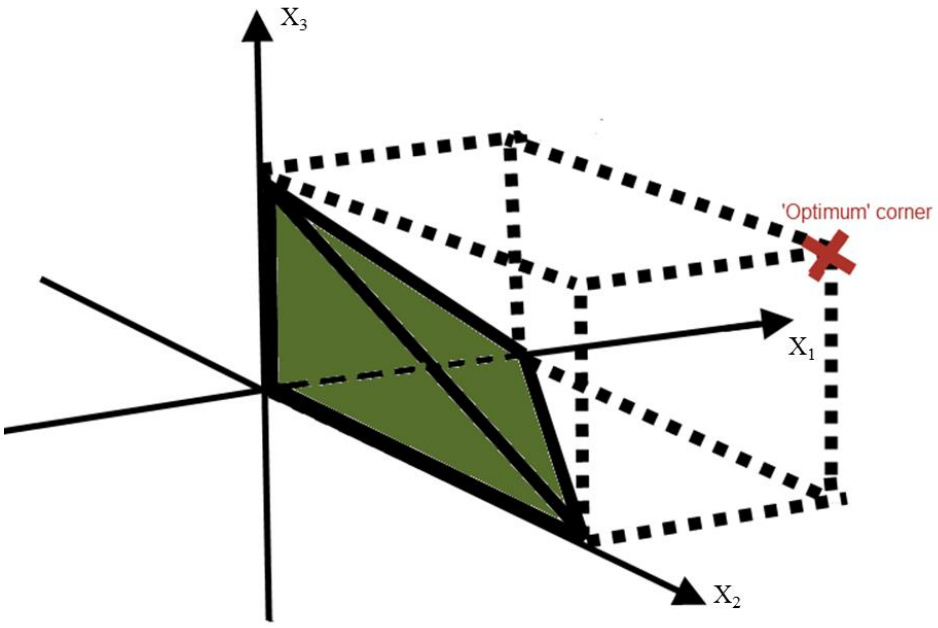


Figure 4.2: Selection of initial conditions using the protocol. For a three factor design space the proposed points selected at the vertices within an octant are shown.

4.1.1.3 Terminating the simplex search

The decision to stop the simplex method search is made after considering a number of features of the current simplex. As the simplex converges upon the optimum, shrinking in the process the difference between the values at each vertex usually become very small.

This small delta can be used that the optimisation is near the end. The size of the simplex will also shrink as it converges so the minimum possible size simplex is another criterion. If the best point of the simplex remains unchanged for a set number of iterations this is another strong indicator the search is at its conclusion. The possibility of previously evaluated points being recycled in new iterations can also be used as a sign to end the search. For this algorithm, an unchanged optimum for 3 iterations and a repeat of 2 previous conditions was used as the strop criteria. This reduced the overall number of experiments. It is preferred the area around the optimum are explored systematically using DoE instead of the simplex search. The main role of the simplex method is to locate the optimum area in the design space.

4.1.1.4 Local DoE modelling

After the simplex search has ended a secondary DoE experiment can be based around the optimum condition to gain process insight from an operational view. The ranges of the design will be now much more defined therefore the model will be a much more accurate fit. The data can be used for process validation to establish the critical process parameters, proofing range and the boundaries of failure.

4.1.1.5 Visualisation of multivariate data

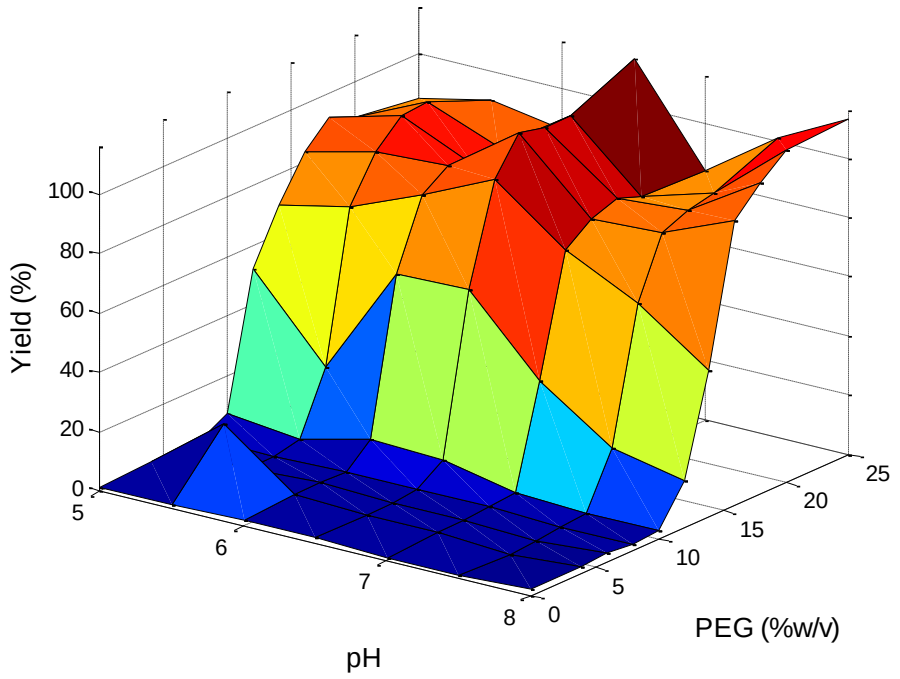
As the number of variables in the experiments increase it becomes increasingly difficult to present the data using two dimensional graphs and even three dimensional response surfaces. One of the most common techniques used to visualise and analyse high dimensional data is using the parallel coordinate plot (Inselberg, 1985). Data for each condition (a point in an n-dimensional space) is represented by a polyline passing through an n number of vertical parallel axes with each axis defining the range of one variable/response. The positions on each axes where the polyline crosses, corresponds to the coordinate of the point in the n-dimensional space and the value of the response(s).

To visualize patterns in the data, all the variables/responses are normalized to a common scale. Inselberg (1997) makes the general observations regarding interpreting the plot for understanding the data; a positive relationship is suggested when most lines are somewhat parallel to each other; when the lines cross similar to X-shapes, a negative relationship is insinuated; when lines cross randomly then no relationship is obvious. Adding more variables or responses simply involves adding more axes. For our high dimensional experiments this tool will be used to present the data.

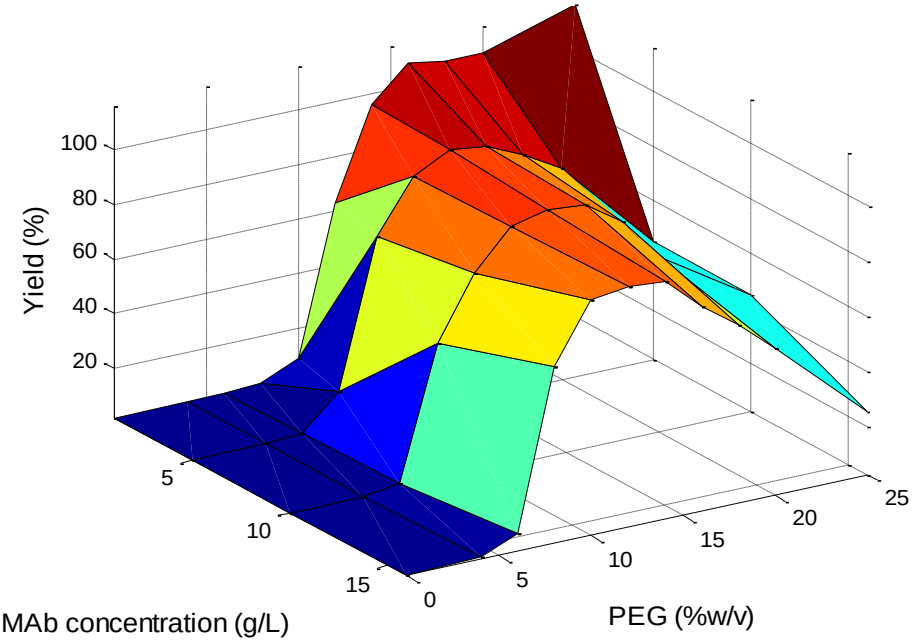
4.2 Case study 1: Optimising PEG Precipitation with the DoE-Simplex methodology

4.2.1 PEG precipitation

Polyethylene glycol (PEG) is one of the few organic solvents that can be used for protein precipitation without causing denaturation (Janson and Ryden, 1997). It is thought to cause precipitation using an excluded volume mechanism, reducing the potential of water-protein interactions inducing protein aggregation until precipitation occurs (Bloomfield, 1996; Marquet, 1995). PEG has been previously demonstrated for antibody precipitation (Knevelman et al., 2009; Giese, 2009; Li et al., 2013). Its main advantage over salts such as ammonium sulphate is its ability to tolerate surfactants (e.g. Pluronic F68®) typically found in cell culture media, which can destabilise the salt precipitation mechanism by dissociating aggregates. PEG also requires less than half the amount of ammonium sulphate (see section 3.5) to achieve comparable precipitation (although it is more expensive). Method development of the technique is most empirical due to the large number of factors requiring optimisation and the uniqueness of protein surface chemistry. To demonstrate the efficiency of using a combined DoE and simplex approach to process optimisation, a previously completed study of the PEG precipitation of MAb is used as an example (Knevelman, 2009). The study encompassed a high resolution response surface dataset for PEG precipitation yield as a function of pH, PEG % (w/v) and MAb concentration (using the same Lonza sourced antibody as in the chapter 3 study).



a)



b)

Figure 4.3: Response surfaces of PEG precipitation. (a) (pH vs PEG, MAb concentration = 1g/L), reveals the global optimum at pH 6.5, 25 % PEG. (b) (PEG vs MAb concentration, pH = 6.5) highlights the curvature in the system showing the same optimum in the right back corner.

The design space consisted of the ranges pH 5 - 9, PEG 0 – 25 % (w/v) and initial MAb concentration, 1 – 15.3 g/L. 385 conditions were used in the study. The yield response surfaces in Figure 4.3 are markedly different to the ammonium sulphate yield response surface in the previous chapter. There is significant curvature and yield drops at PEG concentrations above 13 % w/v for higher titre loads (> 5 g/L). The pH also has more significant effect on yield at high precipitant concentrations, even reducing yield for some conditions. With Ammonium sulphate the precipitation yield would not fall at high precipitant concentration regardless of initial protein concentration or pH (within the ranges tested).

It is suggested that this recovery loss at high PEG concentrations was due to high viscosity that caused premature blocking of the filtration plates. This could be overcome by using larger surface area filter plates or centrifugation to clarify the precipitate. The strength of the precipitate pellet is also effected by the precipitant type (see section 1.5). The optimum of the surface is visible at pH 6.5, 25 % PEG and 1 g/L. For the lower MAb concentrations of 1.0 and 3.8 mg/mL, the highest recovery was attainable at 25 % PEG concentration range. Upon increasing the MAb concentration, the MAb recovery shifted to 12 % PEG and is also lower overall. pH also plays a lesser role in the precipitation, returning the highest yields at pH 6.5 and then pH 8. The loss of yield at higher MAb concentrations and acidic pH is thought to be caused by increased viscosity causing losses over the filtration and resolubilising steps (Knevelman, 2009).

4.2.2 DoE – Simplex method optimisation

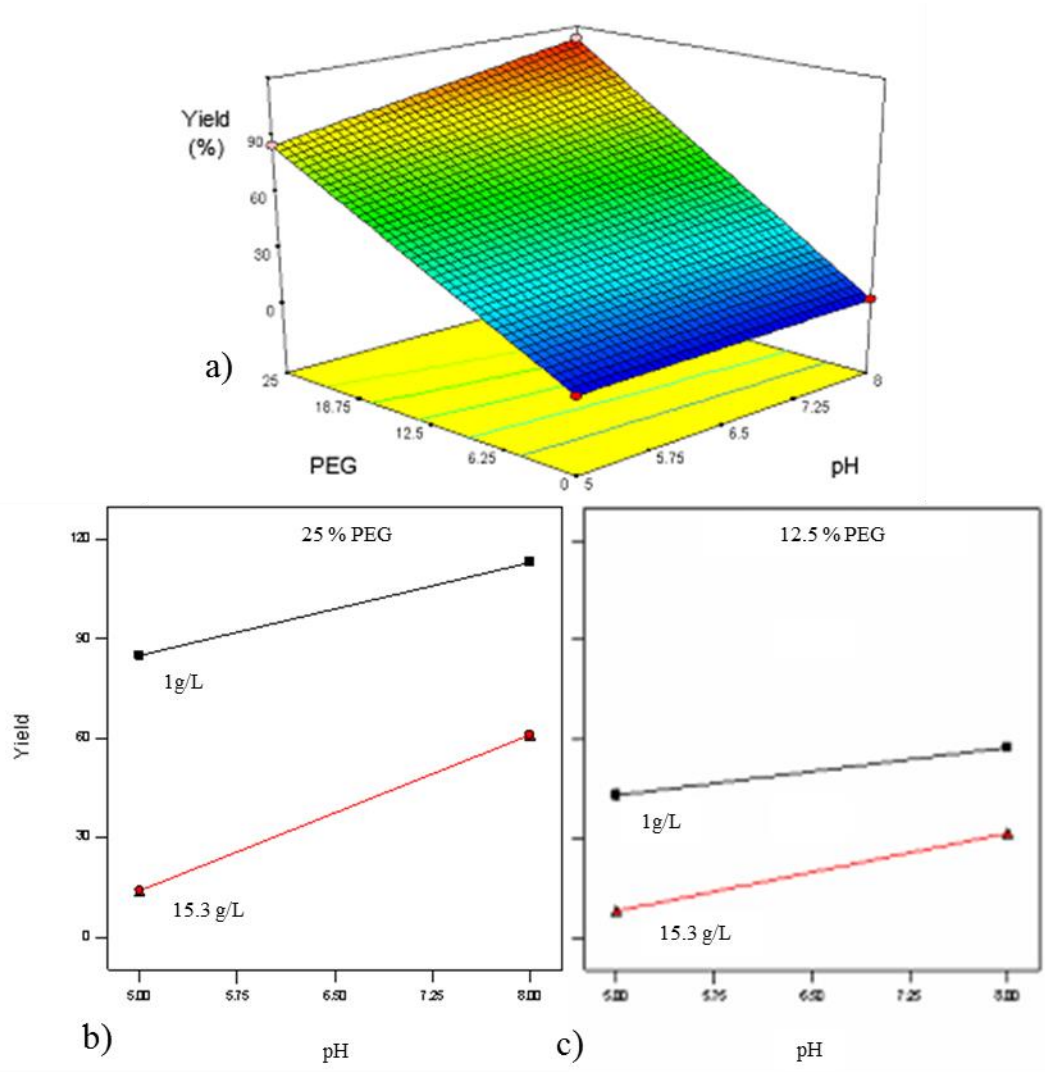


Figure 4.4: Results from the initial DoE of PEG precipitation case study. a) Shows the response surface for yield as a function of pH and PEG concentration at 1 g/L MAb concentration, b) and c) show how pH affects yield at 25 and 12.5 % PEG, and at 1 g/L and 15.3 g/L MAb concentrations.

A two level full factorial DoE was used to provide preliminary data about the factors and response for the PEG precipitation of the antibody from crude supernatant. This generated nine conditions from where promising areas would be used to deploy the simplex from. The DoE results are shown in Figure 4.4 where the data was fitted to a linear model. The

data shows there is 100 % recovery at pH 9 and 25 % PEG, when at 1 g/L (Figure 4.4 (a)) however at 15.3 g/L MAb concentrations the optimimum yield drops to 60 %. As this discussed earlier, the more viscous samples were difficult to separate by filtration and this had an adverse effect on yield. At high PEG and MAb concentrations, the viscosity can be assumed to be highest.

The findings from the DoE data and the model indicate MAb yield is predominantly dependent upon PEG concentration and the initial MAb concentration whereas pH has a minor significance. Figure 4.4 presents some snapshots of the 3 dimensional design space capturing the location of a potential optimum area. The data was fitted to a linear polynomial equation using the Design Expert software (Design Expert, Minneapolis, MN):

$$\text{Yield} = -22 + 6.33\ \text{pH} + 2.66\ C_{PEG} - 2.132\ C_{MAb}, \qquad (4.1)$$

where pH, C_{PEG}, and C_{MAb} refer to the actual values for pH, PEG concentration (% w/v) and initial MAb concentration (g/L).

Due to only using 9 samples in a 3 variable design space the model cannot offer much confidence in the optimum. An R^2 value of 0.66 and a low curvature F value of 3.8 highlight that the model fails to explain the variation about the mean well and is blind to the curvature that is evident in the high resolution dataset in Figure 4.3 despite using a mid-point in the design. Figure 4.4 c) shows the lack of fit between the model's estimation for the mid-point (green X) and the experimental value (green circle), although it actually achieved 89% and was the second highest yield sample in the study.

4.2.2.1 Defining the initial simplex

The simplex experiment consisted of forming the design space, which was carried over from before (pH 5 - 8, PEG weight 0-25 % w/v and initial MAb concentration 1 – 15.3 g/L) and yield would be the driver for the algorithm. The location of the optimum yield was estimated by the DoE model to be in the high pH, high PEG and low MAb concentration corner of the design space at pH 8, 25 % w/v PEG and 1g/L MAb

concentration. As per the proposed DoE methodology described in section 4.1.3 the 3 factor design space was split into 8 segments with the octant containing the optimum would be used by the simplex method. The values at the corners of the octant are predicted using equation 1 and are shown in Figure 4.5.

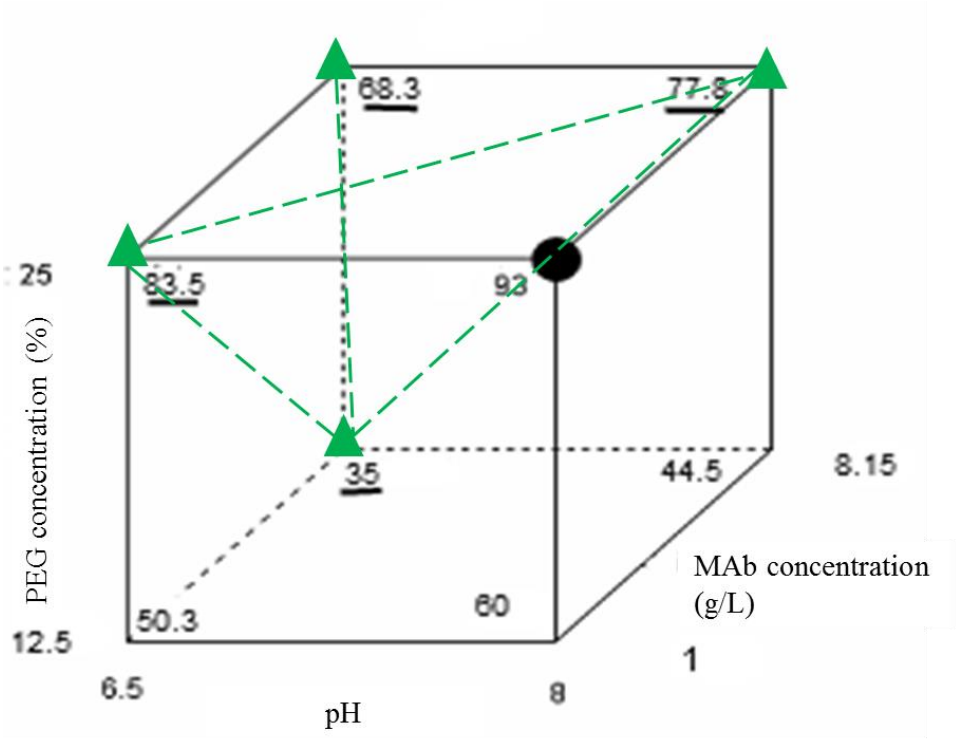


Figure 4.5: Octant of yield design space containing the model optimum (black circle). Model estimated yield values at each vertex of the octant. Vertices with underlined response estimations will form the 4 initial simplex conditions

Four conditions were needed for the initial simplex, which meant its shape was a tetrahedron. According to the methodology the conditions for the initial simplex were selected and are listed in Table 4.2. The corner of the octant with the highest yield estimation was avoided, instead selecting the corner opposite this (6.5, 12.5, 8.15) as well as the next 3 highest response value corners.

Table 4.2: Conditions of the initial simplex

Vertex	pH	PEG (%)	MAb (g/L)
1	6.5	12.5	8.15
2	6.5	25	8.15
3	6.5	25	1
4	8	25	8.15

The simplex search is shown in Figure 4.6 with the first set of points for the initial simplex represented by the green tetrahedron. It is shown in the whole design space where its coverage is 6 % (by volume) and its orientation has been selected so the corner with the worst conditions is exactly opposite to the DoE predicted optimum at pH 8, 25 %, 1 g/L. The model was a poor fit with a low 0.66 R^2 value however the real optimum is still expected to be in the general vicinity of the predicted conditions. By following the simplex trail in Figure 4.6 The worst vertex in the initial simplex was in fact at pH 8, 25 %, 8.15 g/L therefore the simplex was led into the low pH range before settling around pH 6.5, 25 %, 1 g/L. This was the global system optimum yielding 116% which inadvertently had also been selected as an initial simplex condition by virtue of the criteria described in the methodology. As the simplex converged around this point, the second highest condition (113%) was discovered at pH 6.5, 22.5 % PEG and 1 g/L MAb concentration. The simplex search used a total of 13 conditions including the conditions adjacent to the optimum were also evaluated after discovering the optimum as the simplex terminated.

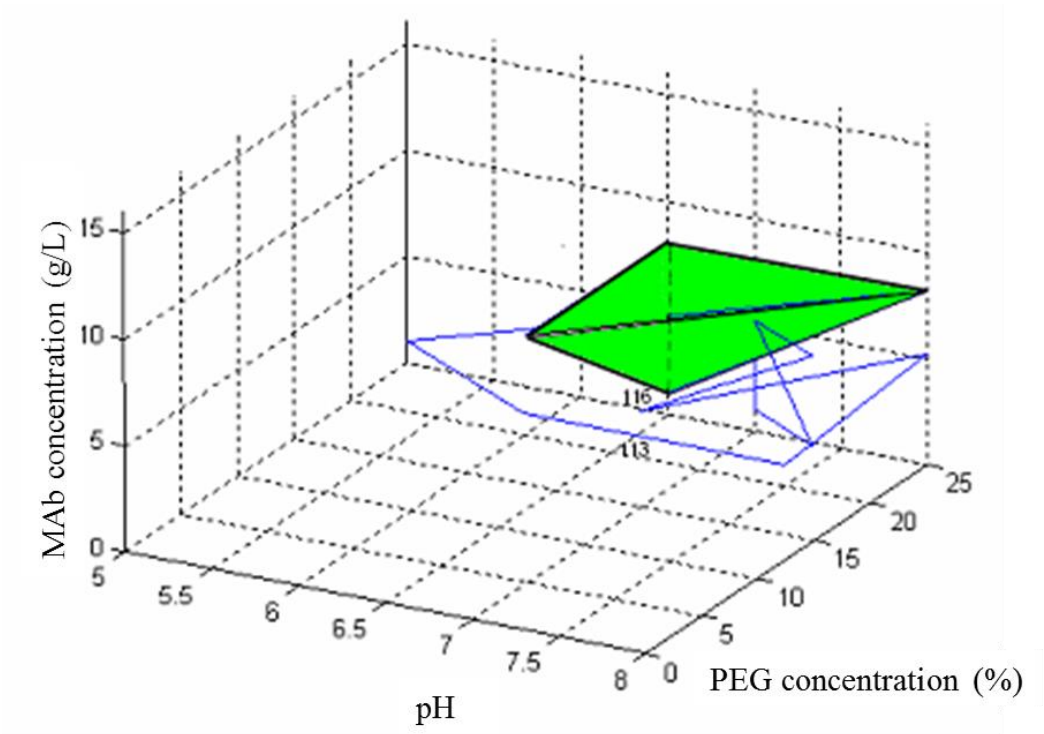


Figure 4.6: Progression of the simplex from the chosen octant. The second part of the proposed optimisation protocol involves initiating the simplex search from the chosen octant of the design space. The figure shows the trail of the simplex in the 3 factor space. The objective function in this case is solely yield.

The difference in the two systems is stark as DoE is able to provide good oversight from a position of knowing very little but not much in terms of pinpointing exact optimum locations. The simplex search is more localized and finds the exact spot of the optimum but reveals much less about the total design space as a whole. In this particular example there are possible 385 conditions in the design space (determined the factor ranges and accepted intervals between the units) and using DoE alone might have taken several rounds to locate the exact spot. The number of experiments needed were only 3.4 % of the design space. The extra points of the simplex can be re-used in a more localised DoE to establish the limits of failure and begin characterisation of this point as potential operating conditions.

4.2.2.2 Local region characterisation

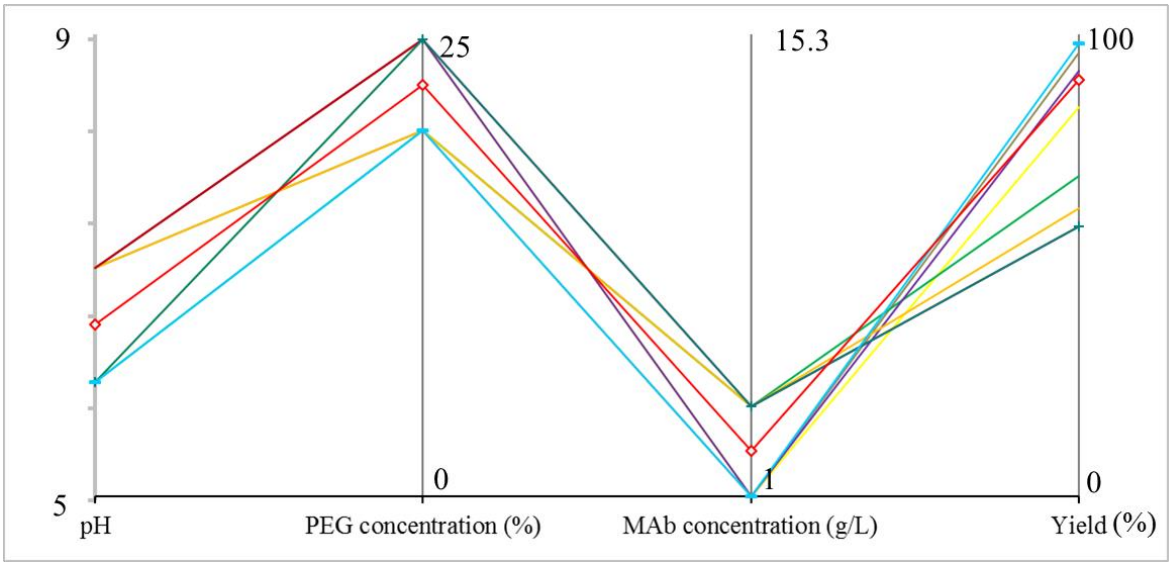


Figure 4.7: Results of the secondary DoE. Sensitivity of the response to % variation in the conditions is also shown, calculated using the model fitted to the data.

After the simplex search, a secondary DoE experiment was initiated around the optimum conditions identified by the simplex. The results are shown in Figure 4.7, a 2-level 3 factor DoE was used with the ranges selected as pH 6 – 7, PEG concentration 20 – 25 %, and MAb concentration 1- 4 g/L. As the local region was much smaller the model created from this data would have a much better fit and it could provide confidence in the conditions were they to be used as operating conditions for this step. Some of the previously used points in the simplex search were also used in the design therefore these conditions did not need to be repeated. It was found there is no other condition with a better yield than the simplex optimum, although the extra experiments have characterised and placed further confidence in the solution to be used as the operating condition. The data from the study was fitted to a cubic equation using regression modeling, which returned a very good fit having a high R^2 value and very low lack of fit F-value (0.992 and 3.4 respectively; see Table 4.3).

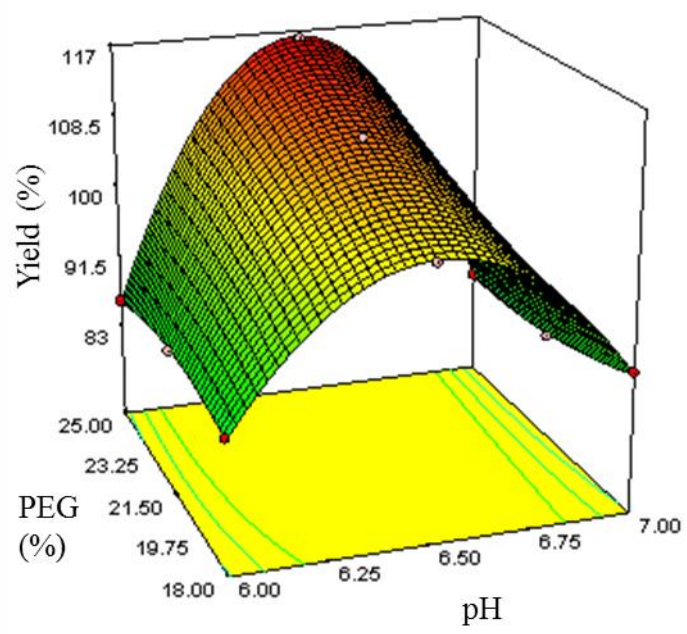
$$Y = 2854 - 886pH - 361C_{PEG} + 1185C_{MAb} + 114pH.C_{PEG} - 343pH.C_{MAb} - 8.2C_{PEG}\,C_{MAb} + 70pH^2 + pH\,C_{PEG}\,C_{MAb} - 9pH^2\,C_{PEG} + 25pH^2\,C_{MAb}\ ; \quad (4.2)$$

where Y is yield, C_{PEG} is the PEG % (w/v), and C_{MAb} is the MAb concentration.

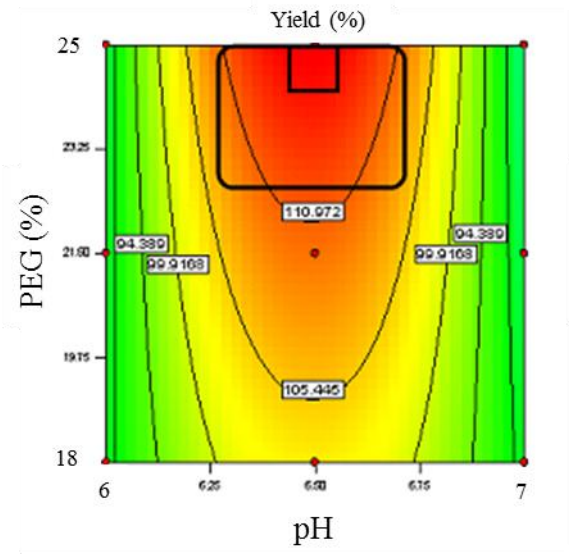
Table 4.3: ANOVA data for the final DoE model (local region only)

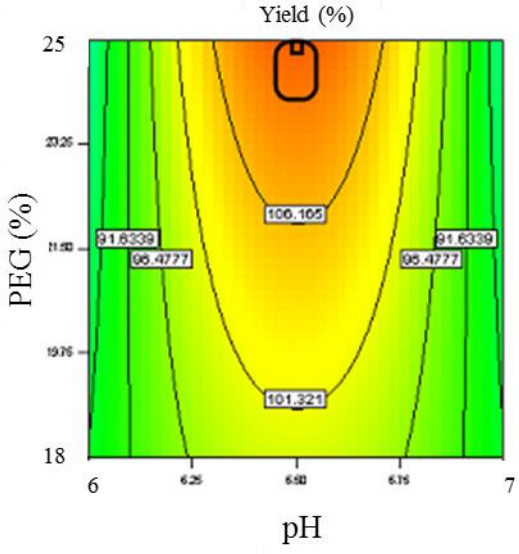
Statistical ANOVA data for quadratic model (calculated by Design Expert 7)	
R-Squared	0.993
Lack of fit F value	3.44
Model F value	194.4
Prob>F	0.0001
Lack of fit F value	3.44
Coefficient of variation %	1.32 %

The design used 10 samples including 2 replicates of the centre-point and the model suggests the region is very sensitive to variation in pH and MAb concentration but can tolerate a drop in PEG without affecting the yield. The R^2 is high suggesting the model is good fit and can be used to predict the optimum. The coefficient of variation is small at 1.32 % so suggesting there is good repeatability of the data. The predicted optimum is at pH 6.5, 25 % PEG and 1 g/L MAb concentration. The response surface graphs of the data are presented in Figure 4.8.



a)





b) 1g/L c) 1.5g/L

Figure 4.8: Secondary DoE heat maps with potential operating range. a) proposed optimum region for MAb concentration of 1 g/L; b) and c) show the same yield function at 1 and 1.5 g/L MAb concentrations highlighting a potential operating range.

It can be predicted with confidence that a yield within 10 % of the optimum can be achieved by keeping the operating conditions within pH 6.25 – 6.75, 22.5 – 25 PEG % (w/v), and 1 – 1.5 g/L MAb concentration. This narrow ridge of a potential operating range is illustrated

in Figure 4.8. Potential operating regions and boundaries of failure are also indicated in the response surfaces. The model was used to calculate the response sensitivity from parameter variation.

4.2.3 Standalone simplex method optimisation

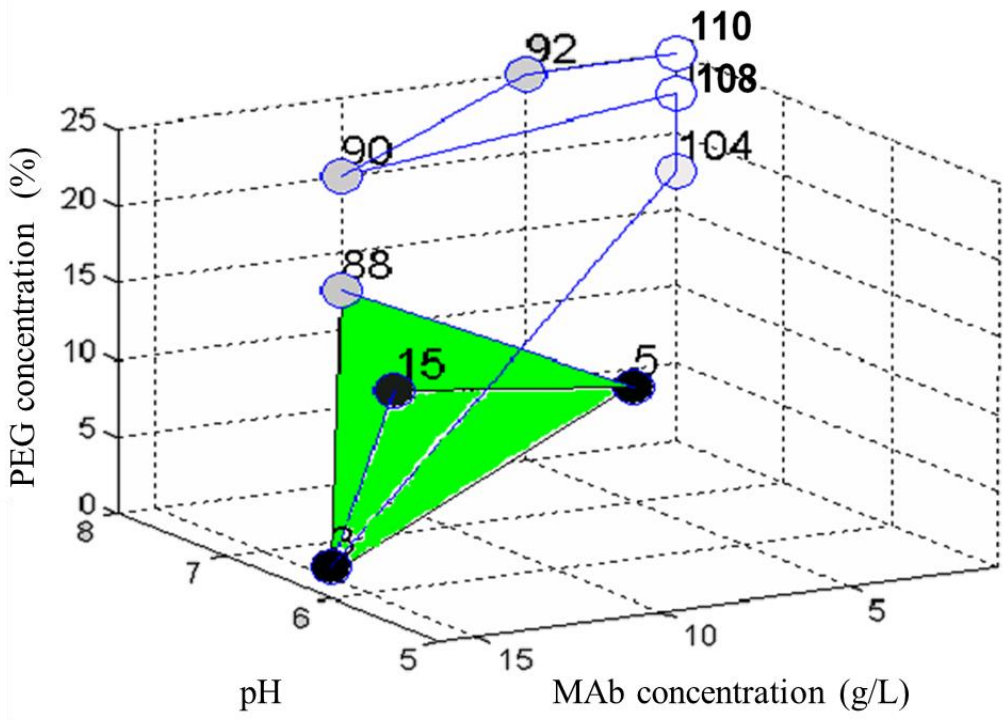


Figure 4.9: Standalone simplex optimisation of PEG precipitation. A local optimum found at pH 8, 1 g/L, 25 % PEG using only 9 samples. Yield values are indicated for each vertex and also represented by the colour intensity of each point (black to white circles – low to high yield)

To compare the results of the Simplex-DoE methodology with just using the simplex method alone a search was initiated from a randomly selected starting location in the design space. Figure 4.9 shows the initial simplex (tetrahedron shaded green) starting from an area of high MAb concentration, low pH and PEG % and progressing its way to an optimum. Three of the simplex points were very low in response that meant it led to an expansion making use of its adaptive 'stepping length' to generate points in the corner of the design space where the optimum happens to be. When the simplex is started close to the optimum as in Figure 4.6, it is less likely to take large steps. The search did not find the global

optimum (at pH 6.5, 25 % PEG, 1 g/L) but did identify another peak of slightly lower response value (110%) at pH 8, 25 PEG % and 1 g/L MAb concentration, which coincidently was also found by the initial DoE. Although not the global optimum, this example again demonstrates the ability of the simplex method in rapidly locating a solution despite no criteria being used in selecting the starting conditions (which actually were disadvantageous as the simplex was in an area of low response). Despite this the search took just 9 samples, which is only 2.3 % of the design space and the same amount of experiments required in the initial screening DoE.

This remarkable optimum locating efficiency does suggest the use of alternative strategy with the simplex method where an initial simplex search be followed by a local DoE although at a small cost of losing some surface-wide process information. The proposed methodology though manages to capture the efficiency and the confidence building features of the simplex and DoE methods.

4.3 Case study 2: Optimisation of a five variable precipitation – centrifugation sequence using the DoE–Simplex methodology

The DoE-simplex methodology was applied to a complicated 5 factor centrifugation study. As discussed in chapter 1, centrifugation is the primary method at scale to separate product containing supernatant from the cell culture. It is also the most common technique used to separate precipitated product from impurities, which remain in solution. Centrifugation success depends on solids that are large and dense to be separated effectively and that can handle the shear stresses encountered in the process. Precipitate ageing is what makes the solids resistant to shear breakage and replicating this at lab scale a scale down model was used using the mean constant velocity gradient (Ğ). The mixing conditions for ageing can be described using the mean velocity gradient, which is a common engineering parameter for describing the shear rate within a stirred vessel and is commonly used in bioreactor scale down (Camp and Stein, 1943):

$$\check{G} = [P / V\mu]^{0.5}; \quad (4.3)$$

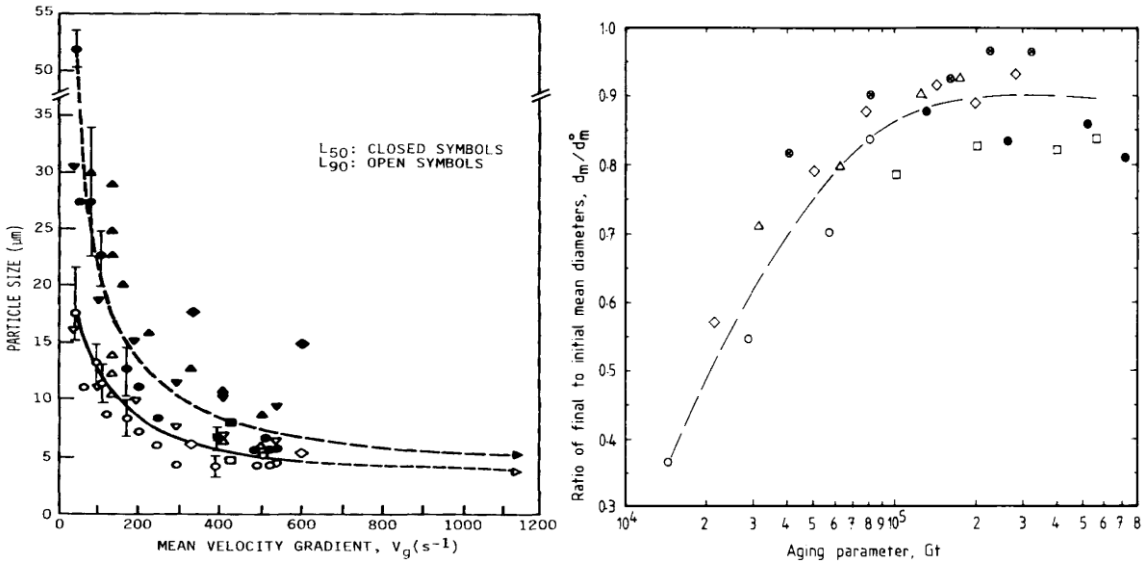
where μ is the dynamic viscosity (Pa.s); V is the volume (m^3) and P is the power dissipation (N/m^3).

The size of precipitate is determined by a dynamic equilibrium between the rates of aggregate growth and impeller induced breakage (Ayazi Shamlou and Titchener-Hooker, 1993). Precipitate particle strength has been correlated with Ğ and the time spent within the shear regime by the dimensionless Camp number:

$$Ca = \check{G}\, t, \qquad (4.4)$$

where t is the mixing time (s) and Ğ is the mean velocity gradient, (s^{-1}).

Ğ has long been used to characterise turbulent velocity gradients in a variety of stirred vessels based on power dissipation (P) per unit volume (V) for turbulent flow (Rushton et al., 1950). In studies with soya protein it was seen that low values (<200 s^{-1}) produced large precipitates while higher values (>400 s^{-1}) gave rise to smaller aggregates (Bell and Dunnill, 1982a). It has been suggested that there is a direct correlation between the final particle size and Ğ (Tambo and Hozumi, 1979; Ayazi Shamlou et al., 1996a).



(a) (b)

Figure 4.10: The effect of shear and ageing on precipitate stability. a) The effect of the mean velocity gradient on particle size for soya protein precipitates (Glatz et al., 1986). (b) Camp number versus precipitate susceptibility to capillary shear (Bell and Dunnill, 1982a).

Ca gives us an indication of precipitate strength, which is difficult to determine otherwise and has also been shown to accurately characterise precipitate size (Fisher and Glataz, 1987). The approach has long been used in the water treatment industry for flocculation processing (Bell et al., 1982). The strength and density of a floc have been shown to be directly related (Tambo and Hozumi, 1979) and to particle strength it is recommended that Ca is $1 \times 10^4 - 10^5$ with Ğ ranging from $10 - 100 s^{-1}$ (Camp, 1955; Bell and Dunnill, 1982).

4.3.1 Estimating shear in the tip micro-environment

Ageing in microwells can be accomplished in a number of ways, shaking in an orbital mixer, micro-magnetic stirrers or jet-mixing using tips (Nealon et al., 2006). Magnetic stirrers and shaking systems are ideal for applying a single mixing condition to many conditions on a microplate however when varying mixing conditions, tips afford a higher level of mixing control and the ability to be programmed into a run on a liquid handling

robot. The shear levels of tip generated jet mixing is also greater than plate shakers and was correlated with impeller shear. The only drawback was the mixers were limited by the number of tips on the robot eight in this case). Equations from literature can be used to estimate power per unit volume imparted by jet-mixing into the well to draw a comparison with stirred vessels.

The Hagen-Poiseuille law (Bennett and Myers, 1962) was used to calculate the pressure drop across the tip orifice of the jet while the tip itself was assumed to be a cylindrical capillary of the same diameter as the jet. The pressure drop lets us calculate power imparted into the microwell and then it is possible to use the established precipitation scale down criteria.

Hagen-Poiseuille equation:

$$Q = \pi r^4 \Delta P / 8 \mu L, \quad (4.5)$$

Reynolds number:

$$Re_{pipe} = (\rho u d) / \mu, \quad (4.6)$$

Mean velocity gradient:

$$\check{G} = [P / V \mu]^{½}, \quad (4.7)$$

Pressure drop:

$$\Delta P = P / Q, \quad (4.8)$$

where Q is the flowrate, L is length of capillary, μ is viscosity, ΔP is pressure drop across an orifice, P is power input (W), ρ is density, u is velocity, and P_0 is a dimensionless power number (6 for turbulent flow in a 4 baffled vessels, approximated for a square well microplate).

The ageing mix was carried out on the Tecan work station using repetitive tip aspirations and dispenses of the precipitate suspension at various speeds. Together the number of aspiration and dispense cycles and the speed of dispense was used to define the camp number, estimated for tip mixing using equations 4-7. The extreme values for the ageing range are shown in Table 4.4 and correspond to a Ca range of 10^4 – 10^5. A linear

relationship is used to increase the number of cycles through the tip and dispense speed (and hence the shear from passing through the tip), so both G and t of the Ca are increased at the same time. For the variable range to be used in the study the number of aspiration and dispense cycles (3 – 12) will be used where each step up will also carry an increase in dispense speed.

Table 4.4: The range of the ageing factor used in the study. The precipitate suspension would be cycled through a 0.06 mm diameter tip between the indicated speeds and passes.

Aspirate and dispense cycles	Time (s)	Dispense speed (ms^{-1})	G (s^{-1})	Ca
3	10	0.71	11	4260
12	40	6.36	77	34090

4.3.2 Centrifugation

Centrifugation is a mature industrial technology capable of high recoveries offering low-cost continuous operation, lower contamination risks, brief preparation times and a small footprint (Foster, 1994). The process uses accelerated sedimentation forces to clarify solids from liquids according to the density differences between the two phases. Precipitation using ammonium sulphate is an advantage here due to the salt's low density even at saturation whilst the challenge remains to optimize the precipitate particle size and strength to create larger, robust precipitate solids. An efficient clarification and good level of dewatering ensures high product yields and a reduced drying load for subsequent protein purification.

A scale down microplate centrifugation technique was used to vary the large scale equivalent centrifuge flowrate and capacity (Boychyn et al., 2004; Hutchinson et al., 2006;

Levy et al., 1999 and Tait et al., 2009). The samples were given shear treatment before centrifugation replicating the particle destabilizing shear forces encountered in centrifuge feed zones that can reduce clarification.

4.3.3 High throughput ultrascale down shear treatment

This shear is not present in lab scale centrifuges and its effects were not accounted for in chapter 3. To recreate the effects of centrifuge shear stress an ultrascale down device known as the rotating disc shear device has been used (Boychn, 2001). Together with the device a much more accurate scale up is achieved, which has been shown to be comparable with pilot and process scale centrifuge performance (Hutchinson et al., 2006). However this device however requires lab scale volumes (~20 mL) and is not suited for high throughput use. In this study a Covaris sonicator disruption device (see Figure 4.11) was used to selectively apply shear to the samples in the microplate.

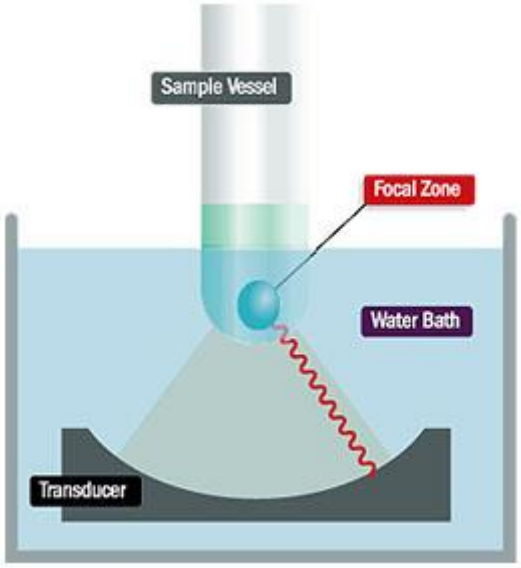


Figure 4.11: The Covaris sonicator acoustic shear device. It uses acoustic energy to apply shear to the precipitate

The Covaris uses acoustic energy packets as the source of the disruptive shear. To select the right conditions preliminary experiments were carried out at lab scale using MAb precipitate (without any ageing) and then the material was sheared in the rotating disc shear

device at various rotational speeds. Some material was also sheared using the Covaris device over a range of intensities. The sheared samples were then centrifuged for 5 minutes at 4000 rpm in a bench top unit before analysis of their supernatants revealed the clarifications.

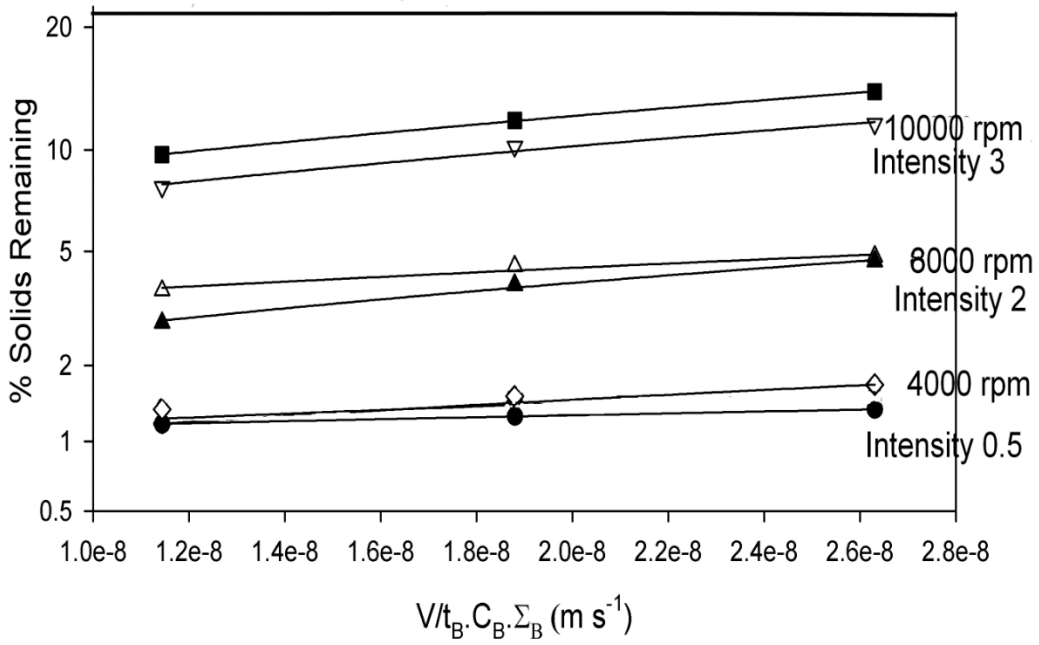


Figure 4.12: Mimicking the labscale shear device with the acoustic shear device. The rotating disc shear device was operated at speeds between 4000 – 10000 rpm for 20 s shear treatment periods. The Covaris intensity was varied to find the clarification matches for the shear device (Covaris at duty cycle 2 %, cycles/burst 100, time 20 s).

Figure 4.12 summarises the shear conditions of the two devices that produced similar levels of clarification. Shear was applied to samples for 20 seconds in both devices and only the disc's speed and the Covaris's intensity were varied to find a match. There was good agreement between the precipitate dewatering levels between 3 sets of conditions from the 2 devices. For the centrifugation study, the Covaris would be operated using intensity 3 for 20 s to mimic the shear from the rotating disc shear device at 10000 rpm for 20 s.

4.3.4 Microscale Centrifugation

High throughput centrifugation using microplates and a labscale centrifuge is achieved by varying the sample volume in the microwell using the method detailed by Tait et al., 2009. Centrifugation time and force are kept constant and in one operation a multitude of conditions may be processed (see Figure 4.13).

The different volumes in the microwells correspond to varying particle settling distances (see Table 4.5). A low volume sample mimics the centrifuge performance using a low flowrate achieving high clarification. When the volume is increased the settling distance is more therefore the results match the performance of a high flowrate operation.Figure 4.13 describes the distances used in Stoke's Law to calculate the sigma factor for the centrifuge are derived using the trigonometric relationship between a theoretical well at the center of the plate and the well of interest (Tait et al., 2009). The range of centrifugation conditions as the fifth variable was set based on the comparative relative centrifugal force (RCF) used in the pilot scale CSA-1 disc stack centrifuge (Westfalia, Berlin, Germany), which would be used for scale up studies (see Table 4.5). By varying volume in the microplate and the position of the well on the plate, the equivalent flowrate could be calculated using Sigma theory (Maybury et al., 2000).

$$[V/(ct\,\Sigma)]_{lab} = [Q/\Sigma]_{pilot}; \qquad (4.9)$$

where V is volume for the laboratory scale, c is the correction factor, t is time, Q is the equivalent flowrate in an continuously operating centrifuge and Σ is sigma.

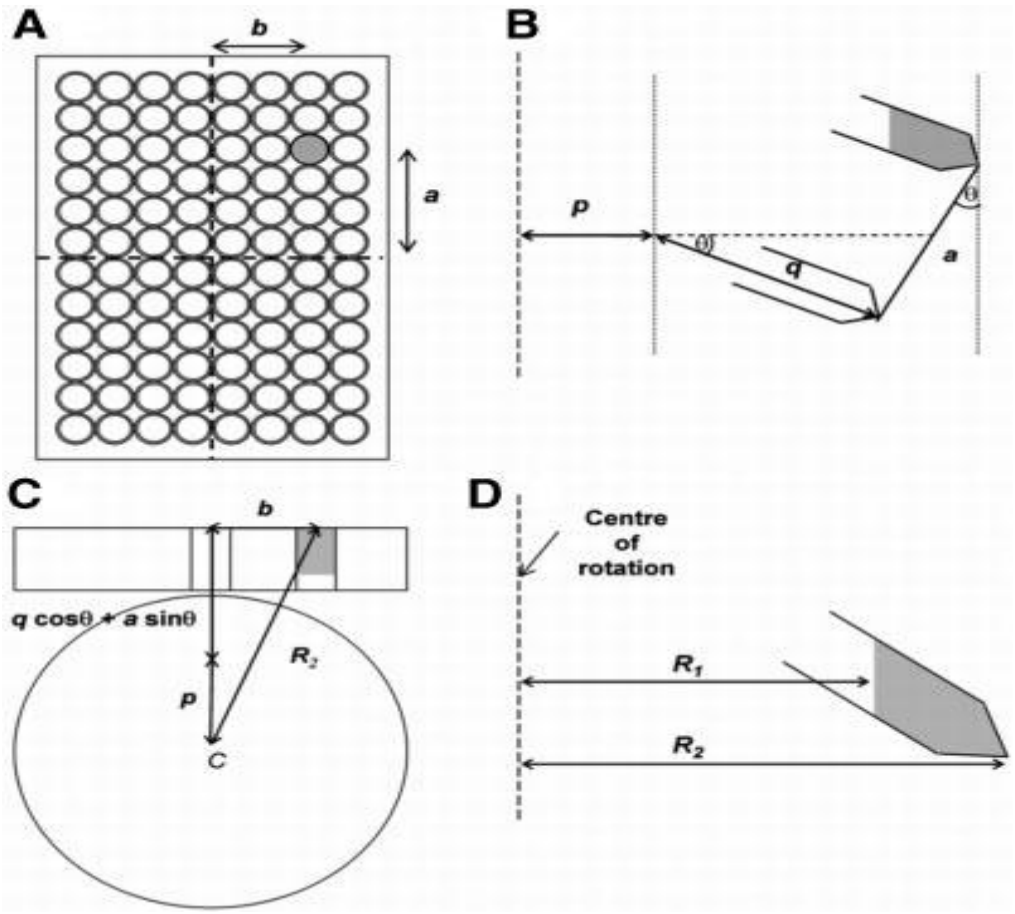


Figure 4.13: Microplate centrifugation. High throughput centrifugation condition screening by varying the well volume (for more information please refer to Tait et al. (2009).

Table 4.5: Equivalent pilot scale centrifugation flowrates. The range for the well volumes used in DoE

CSA-1 flowrate (L/hr)	Vol (µL)	Relative centrifugal force (g)	Time (mins)
65	500	2025	5
85	800	2025	5
125	1100	2025	5
145	1500	2025	5
205	2000	2025	5

After the process was completed, the supernatant was then removed using the Tecan workstation and then analysed for optical density, total protein and HPLC protein G protein quantification.

4.4 DoE study

Experiments with multiple factors are much more difficult to optimise using DoE techniques due to the exponentially rising sampling required with each additional factor. Therefore process development usually begins with crude screening DoEs with the aim of identifying significant factors and ranges in relation to the responses. After reducing the number of factors to ones deemed statistically significant, factorial designs are utilized to optimize the process and generate response surface models from the data (Kalil et al., 2010). The predicted optimums are then experimentally evaluated. Determining statistical significance of process factors can be highly subjective and affected by the chosen ranges defined by the scientist. These are often decided by comparison with previously optimized and similar experiments so it is possible a suboptimal choice carried forward. The option of conducting preliminary range-finding experiments for each factor would make the experiment extensively sample intensive

For this investigation a two level 5 factorial DoE design was generated to model the overall design space in assistance for selecting favourable initial conditions for the simplex. The experiment consisted of 33 samples including 1 centre point and using MAb yield and purity as the responses and basis of the objective function. The combined response function will also be needed for the objective function to drive the simplex search:

For $Y \geq 80\%$

$$CR = 0.3Y + 0.7P; \quad (4.10)$$

otherwise,

$$CR = (0.3Y + 0.7\,P)/10; \quad (4.11)$$

where CR is combined response, Y is yield and P is purity. The penalty to low yield samples produced a more desireable response as the possibility of high purity at low recovery could distort the simplex search. Recovery is arguably more important than purity however, the purpose of the step is to purify so therefore trade-off in recovery for purity only becomes meaningful after a comfortable yield is achieved. The salting-in phase that affected the product more than impurities also caused negative purity values therefore setting a minimum yield constraint was a good compromise for excluding such phenomena.

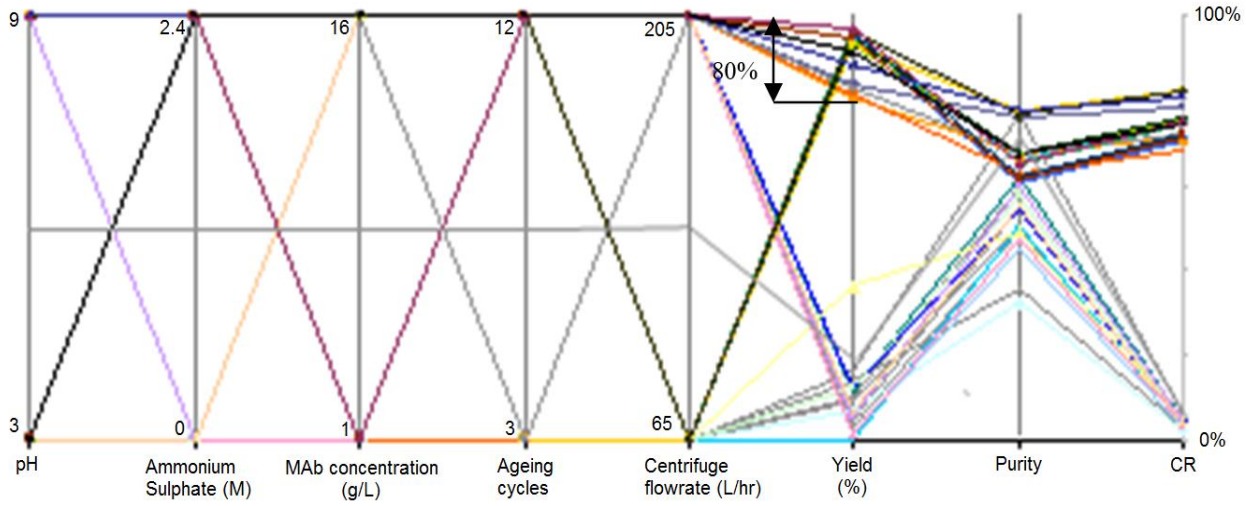


Figure 4.14: Parallel coordinate plot of precipitation and centrifugation DoE. The first 5 vertical axes on the left as the variables and the 3 on the right being the responses. The variables are scaled from 0–1 but the responses are the actual values.

Each line in the parallel coordinate plot in Figure 4.14 represents one of the 32 experiments conducted in the full factorial DoE. Samples with less than 80 % yield were penalised in the combined response function that has caused two 'bundles of lines on the combined response axis. The crisscrossing pattern of the lines over the variable axis is caused by the two level (high and low) DoE design used in the study. The data for the combined response was then fitted to a factorial model defined by:

$$CR = 0.046 - 0.257\, C_{AS} + 0.001pH + 0.001C_{MAb} - 5x10^{-5}A + 0.004pH.C_{AS} + 0.001C_{AS}.C_{MAb} + 0.001C_{AS}.A - 7.15 \times 10^{-5}\, Q_{Cent}\ ; \qquad (4.12)$$

where C_{AS}, Ammonium Sulphate concentration (M), C_{MAb}, MAb concentration (g/L), A is number of ageing cycles, and Q_{Cent}, the equivalent centrifuge flowrate (L/Hr).

The DoE model suggested Ammonium sulphate concentration, MAb concentration and ageing were the most significant factors. Table 4.6 suggests the model created with the DoE software is not a great fit of the data with only an R^2 value of 0.64. Also the discrepancy between the model's R^2 value of 0.63 and predicted R^2 of 0.48, the accuracy of the predicted response values is poor. Model fit will inevitably suffer when a low resolution DoE is applied to an experiment with many factors.

Table 4.6: Statistical ANOVA data from the combined response model.

R-Squared	0.63
Predicted R-Squared	0.48
Model F-value	9.1
Prob > F	< 0.0001
Coefficient of variation	62 %

From the model the combined response has a multi peak surface with the best two optima listed in Table 4.7. Both of the conditions use 2.4 M ammonium sulphate, 16 g/L MAb concentration and a low flowrate however the variables pH and centrifugation flowrate are at opposite ends of their respective ranges. To increase confidence in the DoE results the usual route to proceed would be to undertake further DoE studies localised to the indicated optimal conditions. This would increase the experimental cost to 99 conditions using the same DoE method.

Table 4.7: The DoE models predictions of the optimum conditions. Two best optimal conditions produced by the DoE factorial model with their predicted response values.

pH	Ammonium sulphate concentration (M)	MAb concentration (g/L)	Ageing (pipetting cycles)	Equivalent centrifuge flowrate (L/hr)	Yield (%)	Purity (%)	Combined Response
9	2.4	16	11	75	96	76	0.81
3	2.4	16	3	65	98	70	0.76

Figure 4.15 is a response surface from the CR model predicting where the best conditions (pH 9, 2.4 M, 16 g/L, 11 ageing cycles, 75 L/Hr) are located for the combined response with a value of 0.81 (a yield of 96 % and a purity of 76 %). The linear model is unable to account for the penalty applied to sub-80 % yield samples so is a poor estimate of the surface in the middle of the design space. The penalty function also causes an overestimation of the effect of ammonium sulphate concentration while masking the effect of other parameters that is why the model appears independent of them. Purity values seem to be independent of ageing and centrifugation flowrate variables and mainly affected by pH, salt and MAb concentrations.

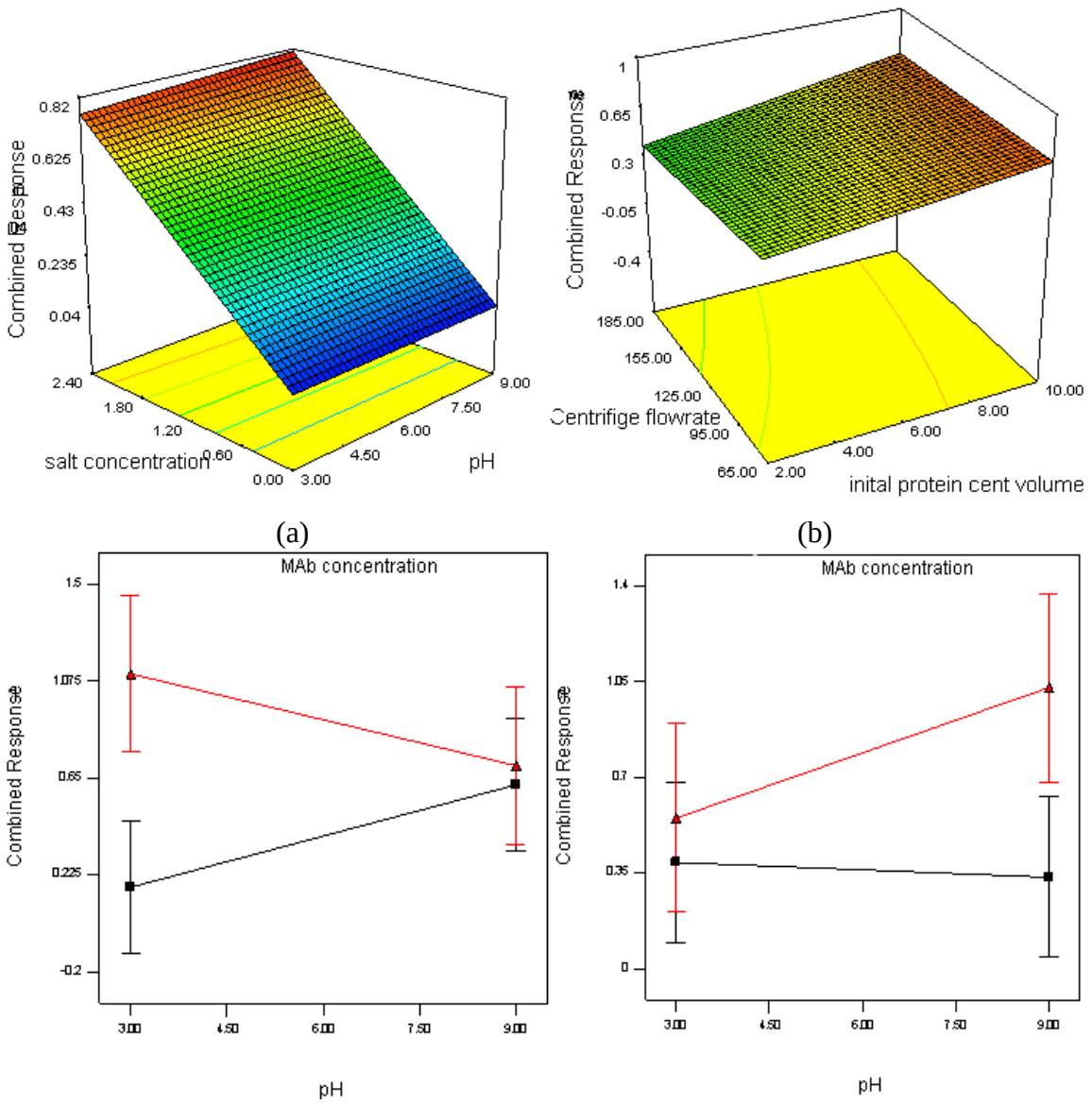


Figure 4.15: Initial DoE data for the precipitation and centrifugation study. a) CR, pH vs ammonium sulphate and other factors are 16 g/L MAb, 11 ageing cycles, 75 L/h flowrate. b) The model suggests flowrate and MAb concentration have little effect on the combined response, other factors are pH 7.5, 2.4 M, and 4.5 ageing cycles. The factor interaction plots show the combined response at 16g/L (red) and 1g/L (black) MAb concentrations for (c) 65L/Hr and (d) 205 L/Hr flowrates. The other variables are fixed at pH 9, 2.4 M, and 7.75 ageing cycles.

Figure 4.15 indicates high MAb concentration feeds provide a higher combined response, however when the centrifugation flowrate is low, a low pH is more beneficial whilst a high pH is better at higher flowrates. In the absence of salt the precipitate does not form and therefore centrifugation recovery is low throughout. Samples with higher MAb concentrations achieved higher clarification possibly due to the larger precipitate being better at 'mopping' up finer particles that are still in solution. As with previous experiments, the DoE suggests salt and pH strong interactions an ideal optimum area at pH 9, 2.4 M ammonium sulphate. Low centrifuge flowrate and high initial MAb concentration suggest a higher response value. Samples with higher MAb concentrations will be denser, which may affect the precipitate particle size distribution. However there is no clear effect on the combined response although a higher concentration achieves a slightly higher response.

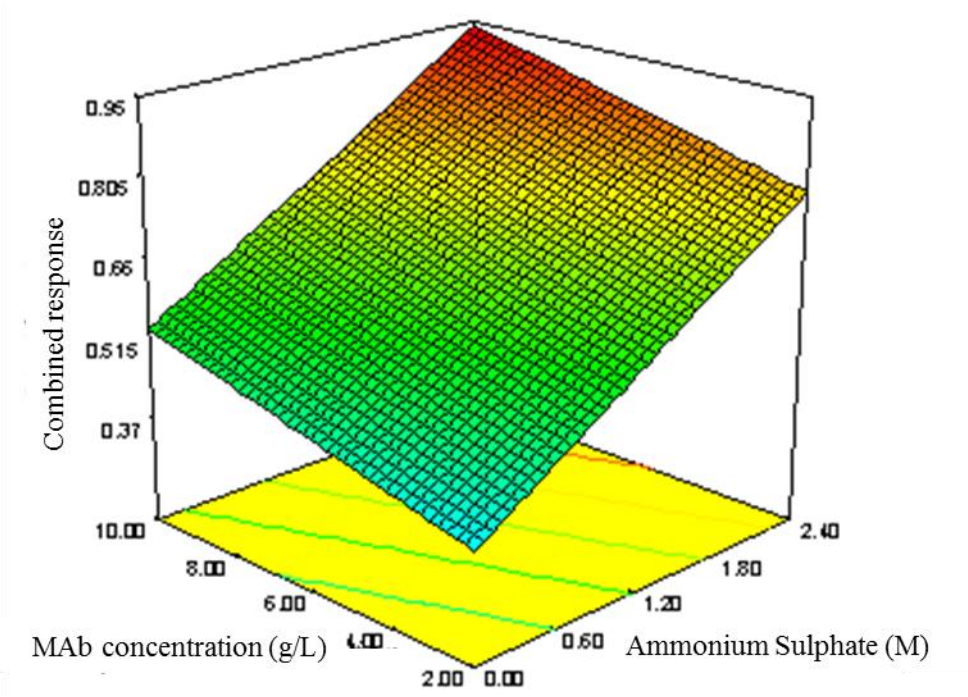


Figure 4.16: Centrifugation DoE data. Clarification as a function of ammonium sulphate concentration and initial MAb concentration, at constant pH (6), ageing (7.5) and centrifuge volume (125 L/Hr).

Recovery is dependent upon a good clarification, which itself requires the formation of a strong precipitate that has had sufficient ageing and time in the centrifuge. Figure 4.16 highlights the positive effect of initial MAb concentration upon clarification, the more concentrated the solution the better the separation, possibly due to the higher densities of the precipitate when ample protein is available (Shih et al., 1992). The precipitate is larger when the initial protein concentration and pH is high, resulting in easier centrifugation (higher clarification). MAb concentration and pH and are also effective in improving the purification factor whereas the other factors have little effect for this response.

4.5 Design space segmentation and simplex method optimisation

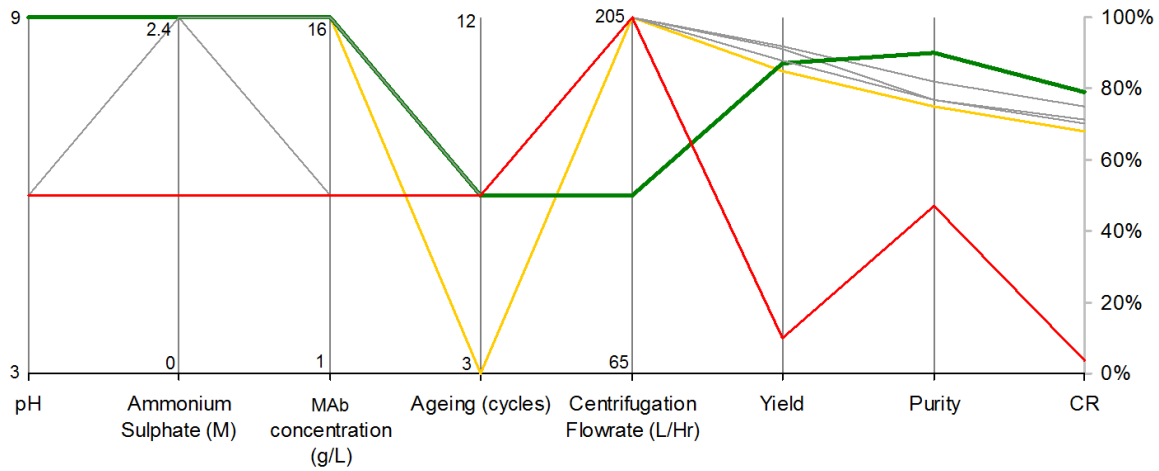


Figure 4.17: Six conditions of the initial simplex. Vertices used in the first iteration are indicated by the coloured lines - green, yellow and red (best, next to worst and worst). All variables are scaled from 0 – 1.

The initial simplex was chosen using the principles of segmenting the design space into the hyper-quadrant of the design space containing the model's predicted optimum conditions. The simplex is difficult to visualise therefore its vertices are shown in Figure 4.17 using the parallel coordinate plot including the ranking of the first simplex iteration. The objective function driving the simplex method was the combined response as discussed earlier. Some of the vertex locations had to be adapted to bring them onto our experimental grid, such as for ageing and centrifuge flowrate, the model suggested conditions of 7.5 tip

mixing cycles and 135 L/Hr were rounded to 8 and 145, respectively. The worst point of the simplex was correctly predicted opposite the model's optimum location at pH 6, 1.2 M, 8 g/L, 8 cycles, 145 L/Hr, which resulted in a combined response value of 0.04 when experimentally evaluated.

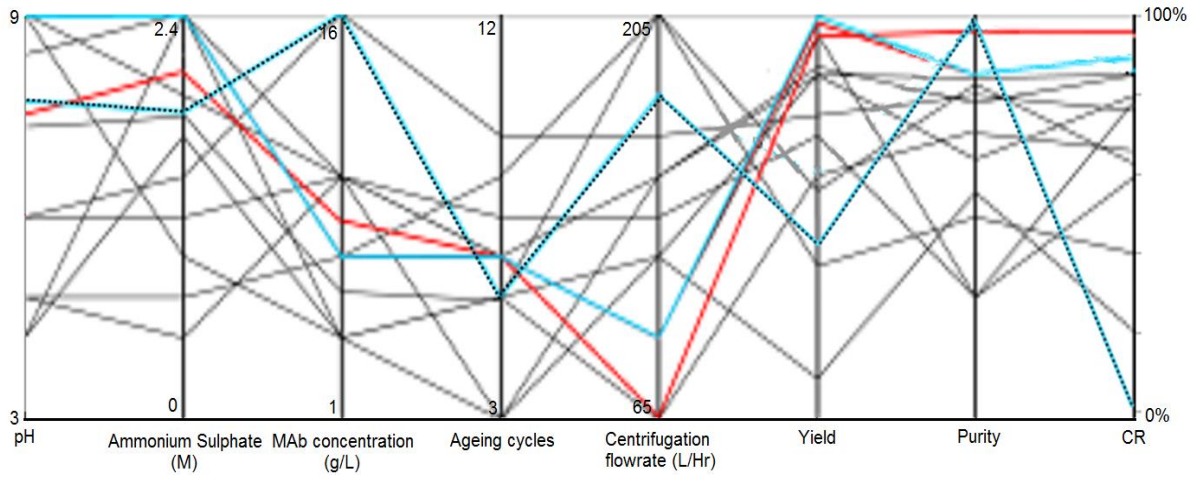


Figure 4.18: Conditions used by the simplex search. 13 lines representing the 13 points used in the simplex iterations following the initial simplex. Red line is the highest combined response (CR) value, while the blue and dotted blue lines represent vertices used in the simplex iterations that gave the highest yield and purity values.

Table 4.8: The best condition for the CR identified by the simplex search.

pH	Salt (M)	MAb concentration (g/L)	Ageing (cycles)	Centrifugation (L/Hr)	Yield	Purity	CR
7.5	2	8	7	65	89%	92%	0.91

Figure 4.18 and Table 4.8 reveal the results of the simplex search. In total 19 conditions used including the initial simplex conditions to identify the optimum at pH 7.5, 2 M salt concentration, 8 g/L MAb concentration, 7 cycles and 65 L/Hr. This point would then be the focus of the next DoE characterisation study.

4.6 DoE characterisation experiment for optimum local region

A two level DoE design was applied to the optimum local region to characterise the design space surrounding the optimum. Small variable ranges were used (see Table 4.9) requiring 33 samples although most of these already been evaluated in the simplex search therefore only the new conditions were experimentally evaluated. The results for the local DoE and the model that was fitted to the data indicate the optimum conditions remain at pH 7.5, 2 M, 8 g/L, 7 cycles, 65 L/Hr. Figure 4.19 highlights this and the flexibility in parameter variance that maintain a minimum combined response value of 0.82. All the conditions with the proofing range have a minimum yield of 88% while purity will be at least 81%. The model used to verify the range has high R^2 of 0.99 and F- value of 28000.

Table 4.9: Factor ranges used in the secondary local DoE experiment

	Low	Mid-point	High
pH	7	7.5	8
Ammonium Sulphate (M)	1.8	2	2.2
Mab concentration (g/L)	6	8	10
Ageing (cycles)	5	7	9
Centrifugation flowrate (L/Hr)	65	95	125

Had the optimisation used the traditional route of applying a higher resolution design method after the initial DoE, it is likely pH and flowrate would be have been fixed at 9 and 75 L/Hr due to their lower statistical significance and to make the design space more manageable (i.e. for a face centred central composite design, a 3 variable design would need $3^3 + 3 = 30$ experiments and 5 variables $3^5 + 5 = 248$ experiments). Optimising in only ammonium sulphate, MAb concentration and ageing factors would have excluded the optimum condition identified by the simplex method at pH 7.5 and 65 L/hr. Another advantage for the simplex search method is it doesn't need to compromise and exclude factors that the DoE model deems less significant.

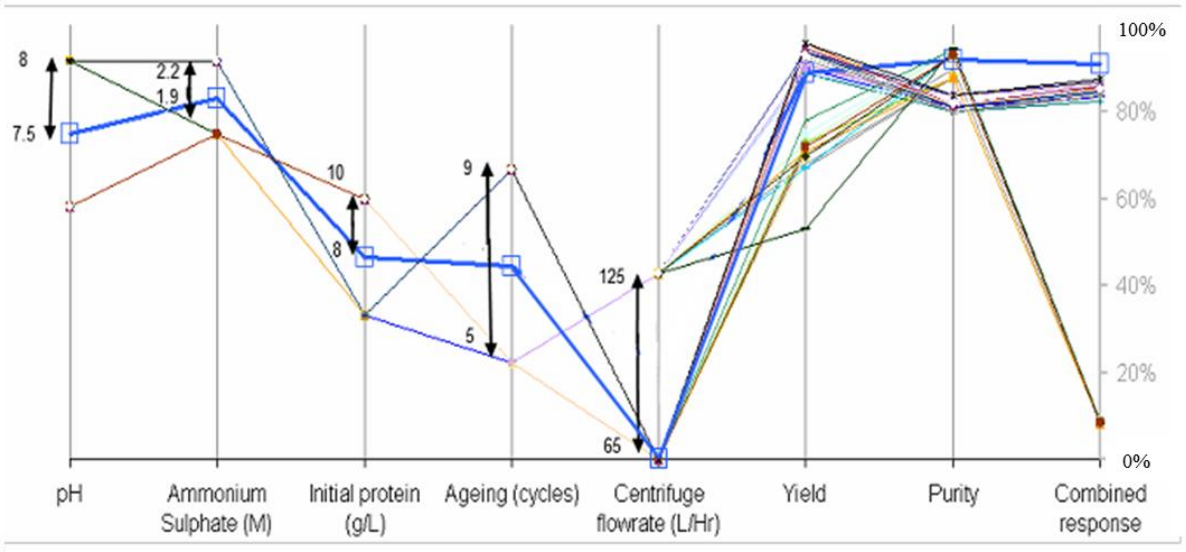


Figure 4.19: Local DoE results. The highest combined response value at (7.5, 2, 8, 7, 65) is highlighted by light blue line and squares. The overall ranges are used to scale the variable axes and not the local DoE ranges.

4.7 Conclusions

As seen in chapter 3, bioprocess design spaces are often complex and multi-peak systems that can cause the simplex search method to converge to local peaks when used in isolation. This chapter proposes a standardised set of protocols of using the simplex search based on design space segmentation so local optima are not an issue. The other ambiguities surrounding the use of the simplex search method such as initial point selection, termination and objective function setting are also defined so the optimisation process itself is made robust. The protocol is explained in Section 4.1.1 and executed on a three factor PEG precipitation step and a 5 factor precipitation and centrifugation sequence. For the PEG precipitation case study the simplex search discovered in just 13 experiments the global optimum at 6.5, 25 %, 1 g/L The secondary local DoE confirmed the yield at this point was the maximum in the design space and provided an accurate model with a R^2 of 0.993. The proposed experimental design identified the best operating conditions for precipitation and centrifugation (which were optimised at the same time) at pH 7.5, 2 M ammonium sulphate, 8 g/L MAb concentration and 7 ageing cycles for the precipitation step and 65 L/Hr flow rate for centrifugation. This achieved an overall recovery of 8 9%

and 92 % HCP clearance. The secondary DoE allowed an assessment of robustness in the system to be made and a far more accurate model of the local region around the optimum found by the simplex search.

The combination of DoE and simplex methods provided comprehensive and focused experiment optimisation using considerable less time and sample numbers then DoE would have alone. The two case studies present examples of three and five factor design spaces that are complex and would take considerable experimental effort to optimise. The approach takes the study from an initial screening DoE experiment to in-depth characterisation of potential operating space using the simplex methods rapid search ability. The studies point to the DoE-simplex method approach would be of greater benefit for high variable problems and the option to optimise whole process sequences, where the multiple factor-factor and unit-unit interactions would make conventional DoE optimisations very difficult to comprehend and implement. Large design spaces also benefit the most as seen in the precipitation/centrifugation case study, as their no need to remove factors deemed insignificant to the statistical model ensuring no factor interaction will be missed or escape the simplex. DoE methods require compromises such as shedding less significant variables from the study and aliasing variables into pseudo terms in high dimension design spaces. The combined DoE-simplex method removes the compromise to reduce the design space and risk losing a superior condition whilst still maintaining the confidence assuring insight of DoE.

5. Development of an automated microscale chromatography process

5.1 Introduction

Efforts to improve process efficiencies and practice lean manufacture within the biopharmaceutical industry have led to the innovation of miniaturised unit operations for process development that consume only microlitre quantities of material (Titchener-Hooker et al., 2008). 'Walk away automation' has been demonstrated at this scale for entire sequences fully integrated with DoE based process optimisation (Micheletti and Lye, 2006; Islam et al., 2007). Microscale chromatographic offerings have been few and far between up until recent years, despite the process often proclaimed as the workhorse of the pharmaceutical industry (Pryzbycien et al., 2004). Microscale chromatography techniques would be especially valuable due to the high manufacturing cost of the unit operation and the difficulties involved in the lab scale optimisation procedures. Several methods employing microlitre volumes of sample and resin have been explored recently, which are aimed at the high-throughput and cost-effective exploration of the design space for chromatographic separations. The technology would facilitate the identification of suitable operating conditions using micro-quantities of feed and buffer materials in much shorter timeframes making the most of the scarce time and resources typically on hand at the start of a campaign.

5.1.1 Chromatography process development

The traditional approach to column development utilises self-packed laboratory columns of 1-5 mL scale and is run on lab based purification systems (e.g. Akta purifiers) for resin and buffer screening, evaluating binding capacities and eluate volume, flow rates etc. Pre-packed columns of similar sizes can also be used, (e.g. Hi-trap, Hi-Prep, GE Healthcare (Amersham, Buckinghamshire, UK), PRC columns, Pall Life Sciences (Ann Arbor, Michigan, USA) but are usually more costly, increase waste and increase the dependence on supply chain (e.g. ordering and shipping time).

Methods for scale down will keep parameters such as column heights, linear flow rates, sample mass: resin volume ratio, residence the same between the scales (Sofer and Hagel, 2011). Consequently even lab scale columns require relatively high volume feed material and lengthy operating times, which often restricts the number of experiments that can be carried out. The purification system used to run the experiment itself may consume extra buffer and material for its hold up volumes and piping. DoE methods are used to optimise experiments but due to the mentioned difficulties design spaces are rarely explored in a vigorous manner. The use of previously built up knowledge and experience is a major influence in chromatography development so the final solution may often be sub optimal with novel resins, processes and products. Such approaches make it hard to take advantage of the regulatory concessions one might receive with QbD, which rewards experimental design led process optimisation with a smoother route through process validation (Wechsler, 2008).

5.2 Microscale chromatography methods

The mainstream adoption of microscale chromatography methods promises to ease the financial and time costs of optimising chromatographic techniques. There is a large gap between lab scale chromatography development costs and the recently available microscale options. Material and product use is comparatively very low, parallel operation offers high throughput processing allowing multiple operating strategies to be evaluated side by side (Chhatre and Titchener-Hooker, 2009). Automated liquid handling will also remove worker error and provides accurate and reproducible results. The relaxation of time and experimental pressures by microscale chromatography offers to build a better, optimised process and shorten the time to market (Farid et al., 2000; Lakhdar et al., 2006). And if a product fails in clinical trials while process development is ongoing, the cost of failure is lower although the technology should facilitate a rise in new drug applications from the faster turnaround times (Titchener-Hooker et al., 2008).

Table 5.1: general features of the three main microscale chromatography formats

	Micropipette tips	Miniature column	Batch plates
Costs	*	***	*****
Sizes	5µL - 50 µL	50µL - 600 µL	5µL - 320 µL
Flexilibity	*****	**	***
Ease of use	*	**	*****
Adsorption	Y	Y	Y
Size Exclusion	N	Y	N
Ease of method development	****	*****	**
Throughput	*****	**	*

High demand for microscale chromatography in biopharmaceutical screening phase and elucidation of process parameters. Experiments to be done in parallel and at high throughput processing of many small volume samples. Previously, gravity driven columns that were handled manually have been used and columns of the 1 mL scale have shown small scale chromatography is very accurate for predicting large scale performance taking into account the extra column volume (Kaltenbrunner et al., 1997). However, it was the improvements in liquid handling robots that has facilitated the introduction of novel, high throughput microscale chromatography systems.

5.2.1 Phynexus chromatography tips

Phynexus chromatography microtips (San Hose, California, USA) are available in 5-50 µL column volume sizes that can be ordered with most commercially available resins. The resin is packed into the tip and held in place by frites and filters. As such they offer the smallest size options of the microscale devices, which can be useful for screening extremely limited feed material. A similar method was demonstrated by Wiliams and Tomer (2004) where they packed standard 10 µL tips with slurry and mobile phase while customised the inlet and outlet with frits. Shukla et al., (2007) also use a similar resin packed tips although with sealed bottoms. Phynexus tips are reusable a limited number of

times until the sealed inlet and become compromised. The resin is fixed into the tip during manufacture using screens to create a packed bed. Operation is a little unconventional in that sample and buffers are loaded by aspirating and dispensing a set number of times, representing the column operations steps such as loading, equilibrium elution etc. Due to this adaption, developing an accurate scale down model has been reported to be a challenging task however there correlation has been shown between equivalent residence time calculations (Chatre et al., 2011). Parallel operation is possible using liquid handling robots (usually up to 8 tips at once) and the phases are processed sequentially just like packed bed operation. Due to the miniscule volumes involved screening very scarce material is possible with samples and buffers being prepared and stored on microplates and reservoir troughs. Caveats may include surface evaporation of buffers/sample, air bubbles, unknown effects the flow of buffer/sample in the reverse direction, and the tips also have the potential to dry out from air aspiration and the frequency of aspirations and dispenses required during operation make this matter worse (from air bubbles or exhausting buffer/sample). Buffers and samples should normally be stocked in excess in the microplate wells as a preventative measure to guard against introducing air into the tips. Certain well geometry plates (e.g. conical bottom) are also favoured over flat bottom wells.

Although residence time has been used as an equivalent scale down parameter it has been shown the first few aspiration/dispense cycles are more critical than latter ones, which suggests the process could be shortened to increase sample throughput. Understanding the difference in flow dynamics between the tips and lab scale columns will be critical to refining the scale up model. Furthermore, the extent of reducing tip cycles in the interest of throughput will be limited to product binding properties and kinetic uptake profiles, which will determine recovery (Wenger et al., 2007). Variable flow restrictions and resistance from different matrix types, buffers, samples, flowrate, creates a lag time during aspiration/dispense, which must also be factored into automated aspirations and dispense steps.

There have been technical issues noted compatibility between the tips and robot pipettor arm. It has been recommended to use fresh buffer in each aspiration-dispense cycle stage

(e.g. washing) as re-aspirating will become contaminated with washed out buffer (from preceding dispense). Its operation does require a lot of microwell usage as each well could be 1 tip volume of used buffer, which makes the process a lot longer and risks a higher frequency of robot errors. Flow rates of 5-20 µL/s have been suggested to ensure adequate interaction of ligant-solute and mass transfer. These flow values are very close to limit of most liquid handling robots range of accurate dispense volumes (Wenger et al., 2007).

Wenger investigated pre packed ion exchange chromatography tip method for VLP (virus like particles) purification from yeast. A 1000 fold comparison was made between 10 microlitre tips and 10 millilitre scale columns. The phynexus tips had simpler automation so were chosen over the other available format at the time, chromatography batch plates, which had a more complicated plate manipulations involved (vacuum/centrifugation). The experiments over the 2 scales had equivalent load capacity, flow, mobile phase conditions and gel electrophoresis of the results were highly similar.

5.2.2 Microlitre batch incubation plates

Using microplates with filters in the base, small volumes of resin slurry are loaded onto the well. Feed is incubated with the slurry and agitated via a shaking device. The plate is then filtered with a plate vacuum manifold or centrifuged and the filtrate collected. The process of loading sample/buffer. Holding for a set period and then filtering with the vacuum is repeated for each chromatography phase. The unique aspect about the batch plate system is that they can be self-prepared with loose resin and are much higher throughput as the whole plate can be processed at one time so are more flexible than the other options. They are also lower in cost, have a lower lead time to acquire (just the resin slurry is needed) and are very amenable to high throughput screening. The drawbacks are they are more labour intensive and each filtering step introduces air into the slurry. The process can be automated with a liquid handling robot to reduce the workload although it has been noted the resin slurry tends to settle over time, which can lead to heterogeneity in the results. Charleton et al. (2006) demonstrated high throughput evaluation of ligands and buffer types and successively used the batch plates for scale down indicator of separation performance.

Coffman et al. (2008) demonstrated a 50-100 µL binding system to quantify binding behaviour of multiple media to MAb feed streams. Comparisons with lab and pilot scale studies were arguably favourable and differences were attributed to remaining liquid hold up volume in wells in between successive stages (washing, loading, elution) and that it is a single equilibrium phase at a time whereas a column operation has multiple phases at any one time.

Most literature sources recommend the batch chromatography plates for performing media and buffer screening studies and studying trends main interactions only. The process takes much longer than the others due to hold times in between each chromatography however it is much higher throughput. Coffman et al., (2008) suggests scale up to be based on the incubation times needed as a minimum for it to be comparable to a column.

The self-dispensing of slurry for batch plates does make the process more sensitive to variation and robotic liquid handling is advised. The procedure for dispensing matrix can be a potentially error-prone e.g. resin may settle in the trough and disposable tips and not be fully dispensed, there may be variability in the volume of resin slurry pipetted or resin may adhere to the outside of the tip. Mixing will also be affected by the large dense matrix beads increasing the viscosity so must be factored in as well as mixing difficulties at lower volumes. The pipette dispensing height and speed also requires some optimisation work. As the wells are not sealed and the plate is vacuum filtered using a plate manifold, the risk of carry over risk from well to well is high and residue remaining in wells can transfer from phase to phase. Evaporation is also a major issue, more so on wells on the plate edge and low aspect ratio well designs.

5.2.3 Atoll robocolumn

Miniature columns are the most similar in form and operation to conventional packed bed columns. Atoll GmbH (Weingarten, Germany) provide pre-packed robocolumns of 50 µL

– 1 mL sizes that can be specified with most commercially available resins. The similarity to a labscale columns means the results are more indicative of larger scale performance (same flow direction and mass transfer dynamics). The robocolumns use dynamic flow that makes dynamic binding capacity and column lifetime studies possible. Such studies are much more difficult to translate with the other options. Chromatograms can be created from analysis of the fractions collected in plates. Additional Tecan hardware is needed that moves the collection plates on a conveyor belt unit underneath the columns in sync with the stages. Plate handling with the robots arms allow transfer of the plates to storage or analytical devices (plate reader or HPLC) to quantify product or impurity levels within each fraction. Evaporation and the chance of cross contamination of droplets falling into the wrong well on the fraction plate are again likely to impact data.

Of the three main options commercially available the robocolumns are the most similar to larger scale columns and as all the equipment was in place to use them it made sense to use them to develop the microscale chromatography method. The results here investigate their feasibility application on a range of resins that were included in the starter pack from atoll and conclude with lab scale experiment to verify the finding.

5.3 Results

A microscale chromatography method was created using the Tecan liquid handling robot and two versions of the pre-packed robocolumns supplied by Atoll (Weingarten, Germany) were evaluated (in 50 µL and 200 µL column volume sizes). Protein A and cation exchange chromotography methods were developed to using the robocolumns on the Tecan liquid handing robot followed by scale up to 10 mL scale using atoll's media scout columns on an Akta system. The following results show good agreement between laboratory scale and microscale purification of the test antibody. The process flowchart of chromatography methods as executed on the Tecan liquid handling robot is shown in Figure 5.1.

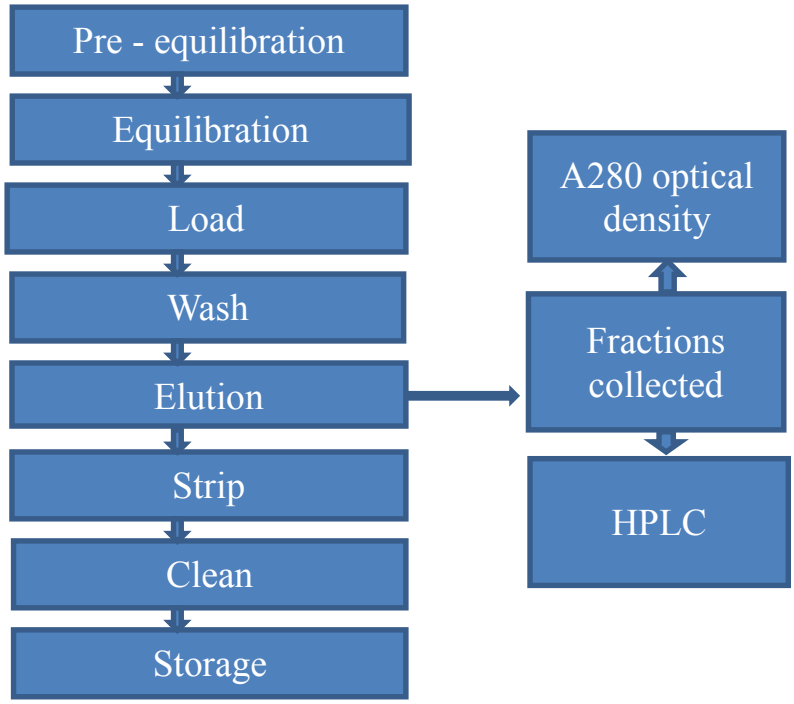


Figure 5.1: Chromatography method flow chart. The generic chromatography steps to be executed on the automated high throughput Atoll chromatography system

5.3.1 Preliminary study using Atoll 200 µL Columns

A preliminary study was conducted using pre-packed Atoll UNOsphere S miniature columns to assess the fraction volume accuracy. Formats in 50 µL and 200 µL colunm volumes were trialled and it was later decided to use only the 200 µL format as the 50 µL data had a lot variation. The variation mainly came from the Tecan's inaccurate pippeting at very low flowrates that the 50 µL columns required. A bind and elute Tecan script was written based on the established platform ion exchange conditions for this particular product using the pH 5, 25 mM Sodium Acetate, and pH 5, 25 mM sodium acetate, 250 mM Sodium Chloride equilibration and elution buffers at a linear flowrate of 120 cm/hr. The linear flowrate translated to a very slow volumetric flowrate on the Tecan, which was just within the bounds of accurate Tecan pipetting for the 200 µL format. The columns are mounted upon a special 96 column holder plate. As liquid is passed through the column, droplets are collected in a 96 well fraction plate positioned below the columns. The Te-

Chrom Shuttle device moves the collection plate one well row for each column operation phase (eluate, strip etc.) dispensed so a separate fraction is collected in each new well. This allows the fractions to be collected as phases (unbound fraction (UBF), Post Load Wash (PLW), Product Eluate (PE) and Strip fractions in Figure 5.3).

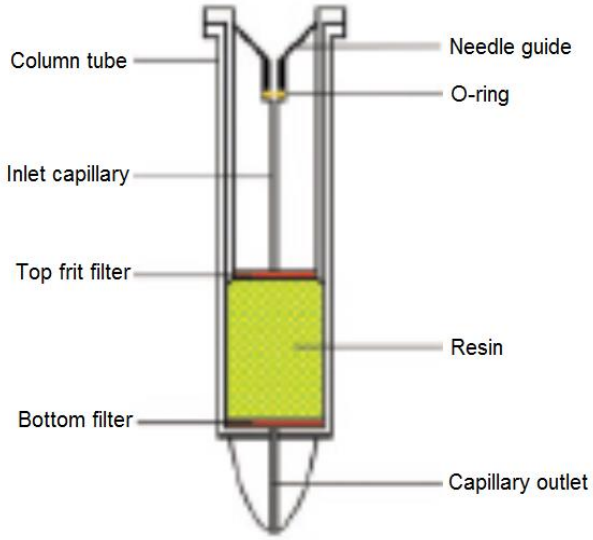


Figure 5.2: Robocolumn cross-section. MediaScout® RoboColumn® (Atoll)

The columns have a pressure tight inlet that seals to the fixed tips of the Tecan liquid handling workstation (see Figure 5.2). This entrance port allows a flexible connection between the automated robot's fixed tips and the packed bed inside. A rubber O-ring maintains the pressure of the flow and prevents back flow /leakage during injection. The inlet capillary guides the fixed tip to the suitable height for dispensing. Care must be taken the tip does not penetrate the top filter so tip insertion depths are specific to the column heights.

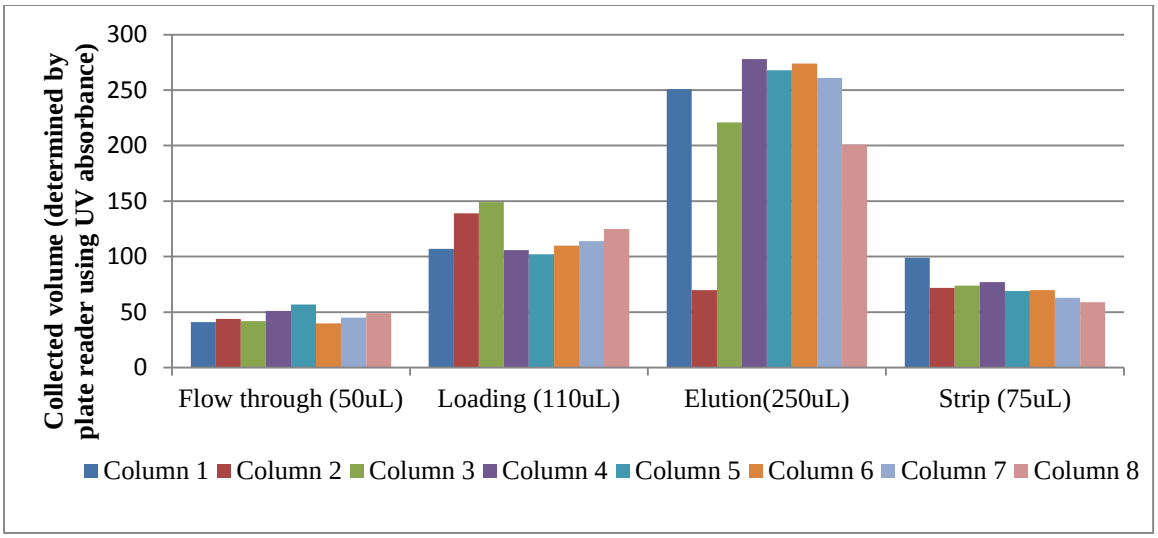


Figure 5.3: Tecan volume dispense discrepancies over eight replicates. Observed variation in between volumes collected in each fraction columns in earlier experiments

The dispense volumes were generally in good agreement but all under-predicted the dispensed values. There is also some carry over of the previous fraction into new fractions by droplets still adhering to the end of the column tip which cannot be avoided. Fraction sample volumes were also checked manually using a 0 – 20 µL manual pipette and all product eluate fractions were found to be in the range 230 – 240 µL. Post centrifugation the computed sample volumes for the product eluate fractions appeared to more closely reflect the measured sample volumes.

Any errors in the computed sample path length and volume will affect the % Yield calculation. The discrepancies noted with computed sample volumes may also be due to the variability in absorbance data at 900 nm and 977 nm, and subsequent computations used for path length correction and the calculation of protein concentration. Variability in path length corrected 280 nm absorbance data is shown in Figure 5.4 highlighting the plate reader is another source of variation.

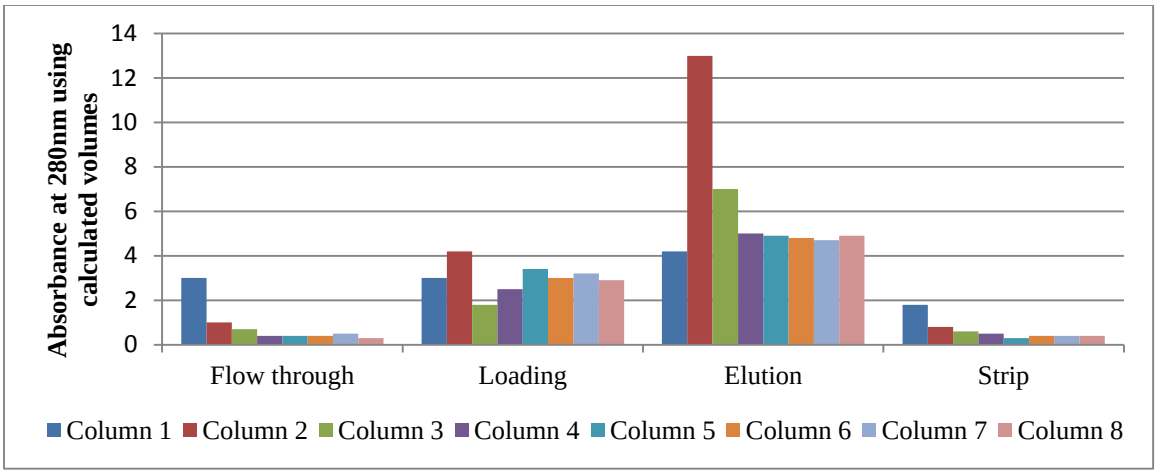


Figure 5.4: Bubble impact on A280. Bubbles in the fractions affected the plate reader's path length calculations thereby distorting the readings.

The samples collected in the fraction plate were visually frothy with lots of small bubbles, especially the flow through and elution samples where the protein was concentrated. To calculate the protein concentration and the liquid volume, absorbance readings are taken by the plate reader at multiple wavelengths (280 nm, 320 nm, 900 nm and 977 nm) of which it is likely the bubbles would have caused light scattering and distorted the readings.

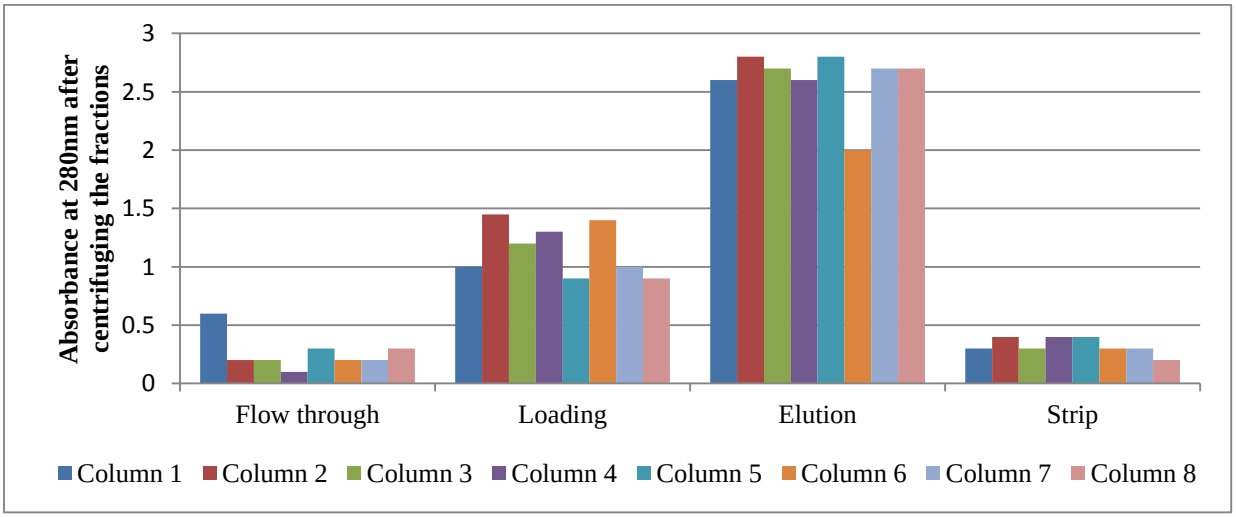


Figure 5.5: Centrifuging to remove bubbles. Same samples as Figure 5.4 after centrifuging for 2 minutes at 3000 rpm which removed bubbles within the samples

In an effort to remove the bubbles, the fraction plate was centrifuged at 3000 rpm for 2 minutes, despite this step adding further time and effort to the high throughput experiment. The samples were read by the plate reader immediately afterwards to reduce evaporation losses. Figure 5.5 illustrates the reduction in bubbles does improve the consistency between data, indicating the plate reader is a potential source of error.

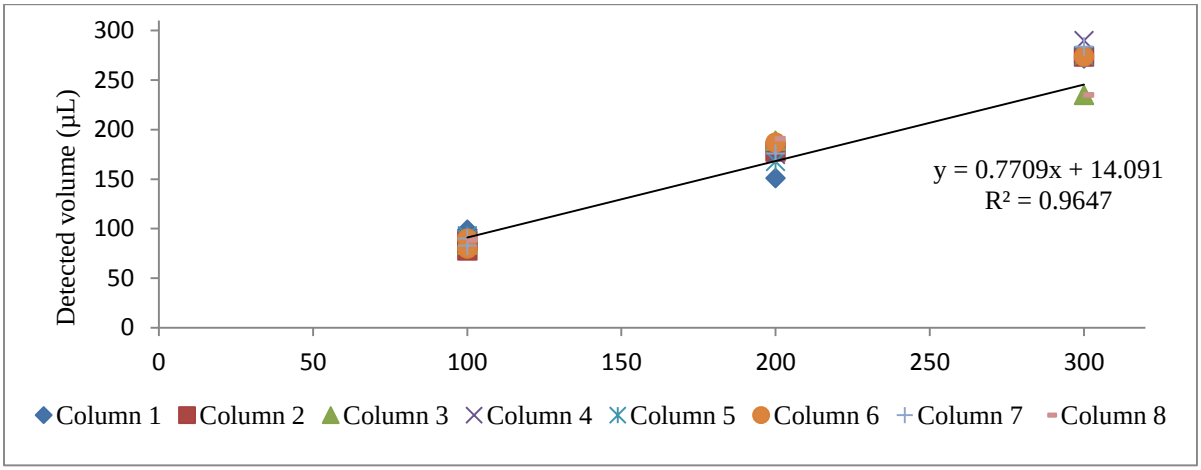


Figure 5.6: Detected volume to actual volume conversion. A calibration curve from the data allows improved calculations of volume and yield to be made

$$vol_{dt} = 0.77vol_{d} + 14.1; \quad (5.1)$$

where vol_{dt} and vol_{d} refer to the detected and dispensed volumes.

The linear relationship between dispense and detected volumes was used to establish a correction factor for sample volume. The sample volume was set as a variable, using equation 1. This computation was used for 2 confirmation runs that demonstrate a good correlation between dispensed and measured sample volume. It was stipulated the difference in dispensed volume and the volume measured by the plate reader was down to losses in the fraction collection phase (stray droplets collecting in the wrong well, sample not dispensing (i.e. the final drop adhering to the end of the tip and not being dispensed in within the allocated phase time), and liquid properties which affect the plate readers

volume calculations to be under reading (e.g. protein concentrations affecting the shape of the meniscus and causing indifferent light scattering).

5.3.2 Cation exchange chromatography starter kit

Included in the starter pack was the S-kit which is 8 commonly available CEX resins which were used to test the developed bind and elute chromatography method on the Lonza test IgG_4 antibody. The feed material had previously been purified by protein A at labscale and a stock was used to supply these experiments. Each column was loaded to a dynamic binding capacity of 15 mg/mL and the Tecan carried out the purification in parallel. Usually the number of available tips are 8 so up to 8 columns can be operated at once which is not as high throughput at the chromatography filter plates however tip heads with 96 tips are also available on the Tecan. The antibody's specific buffers and conditions were used although they did not differ much from the generic MAb platforms use of sodium acetate to buffer and sodium chloride to elute. The method shown in Figure 5.1 shows the operations in the process. The feed had a conductivity of 5.2 mS/cm and a step elution performed with the elution buffer having a conductivity of 13.8 mS/cm. The yield for each CEX format was determined by pooling the total amount of product collected in the wells of the 96 well plate corresponding to the elution fractions. The 8 different CEX options and the results of the Tecan screen are shown in Figure 5.7.

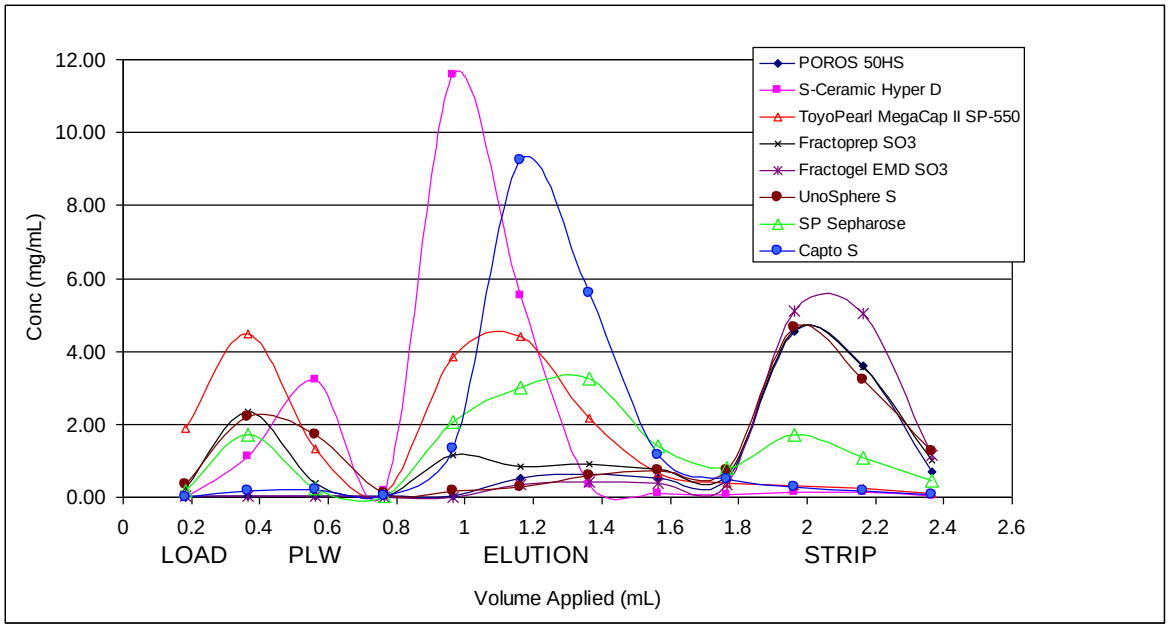


Figure 5.7: Chromatograms from the CEX screening study. Chromatograms were created from the fractions protein concentration (elution with 25 mM Sodium Acetate / 125 mM NaCl pH 5.0)

All the fractions for every phase were collected in a 96 well plate and the eluate fractions (Rows 5 to 9, corresponding to a 5 CV elution block) added up to calculate yield. . Figure 5.7 represents an offline chromatogram for all the CEX formats tested which is comparable to A280 traces that are found on labscale chromatography machines such as the Akta series (GE, Uppsala, Sweden).

The load and post load wash (PLW) fractions indicate some of the CEX options had better binding capacity for the IgG_4 than others. POROS 50HS, Fractogel EMD SO_3 and Capto S showed no breakthrough (product in the flow through) whereas Toyopearl, Unosphere and S-ceramic have notable product loss during the load and PLW stages. Elution was with 25 mM sodium acetate / 125 mM NaCl pH 5.0, and product peaks were only observed with S-Ceramic, Capto S, Toyopearl and SP Sepharose resins. Higher salt elution buffers are likely to be needed for Fractogel, Unosphere, Fractoprep and SP sepharose which all had little to no protein in the elution and significant product in the high salt strip. The highest

yielding CEX was Capto S followed by S-Ceramic. Both show low amounts of protein in the load and wash flow through fractions (although S-Ceramic was a little overloaded) and high concentrated eluate fractions with a sharp peak, which usually is indicative of good selectivity. Both of these options would be ideal to scale up and conduct lab scale optimisation for aggregate removal. Aggregates tend to bind more strongly to CEX columns due to their multiplicity of binding sites over the monomeric product. Elution buffer optimisation therefore seeks to find the conditions which elute the monomer but not aggregates which are later eluted in the strip and sanitisation phases.

Most of the process development effort in downstream processing is focused on the IEX steps as the other steps in the MAb platform are operated under mostly generic conditions. Each antibody product responds differently to the various chemistries available in IEX chromatography and removing aggregates is a key CQA due to their immunogenicity. An unwanted consequence of high titre cell cultures is the greater amount of aggregate content and other impurities which puts a strain on the downstream process. Product quality analysis of the pooled eluate was performed using SDS PAGE which revealed additional bands, indicating the product IgG_4 had degraded to high levels of fragmentation, most likely due to sample age. Antibody fragments, while still being able to bind to CEX ligands, have a much weaker bond and may come out in the flow though and the wash. This may be the case for most of the CEX options tested as most showed protein in the flow-through fractions which was mostly likely enriched with antibody fragments.

5.3.3 Protein A starter kit

Similar to the S-kit experiment, the Protein A capture kit was also evaluated using the same test antibody. . The pack consisted of 8 different Protein A resins in the 200 µL column format and the buffers used were generic protein A buffers, equilibrating and washing with 50 mM TRIS pH 7.4, eluting with 50 mM sodium acetate pH 3.6 and strip with 0.1 M acetic acid. The feed material was cell culture supernatant containing IgG_4 which was

loaded at 25mg/mL for all Protein A columns. This capacity was chosen based on the developed downstream process for this particular product.

The results of the 8 different Protein As are shown in a chromatogram in Figure 5.8. The loading and wash phases are omitted as these fractions give a high A280 signal as they contain HCP, DNA and other non-Fc region containing impurities which are not bound to the matrix. These fractions were directed to waste during the Tecan run and not collected on the 96-well plate, only the eluate and strip fractions were collected for analysis by absorbance at 280 nm.

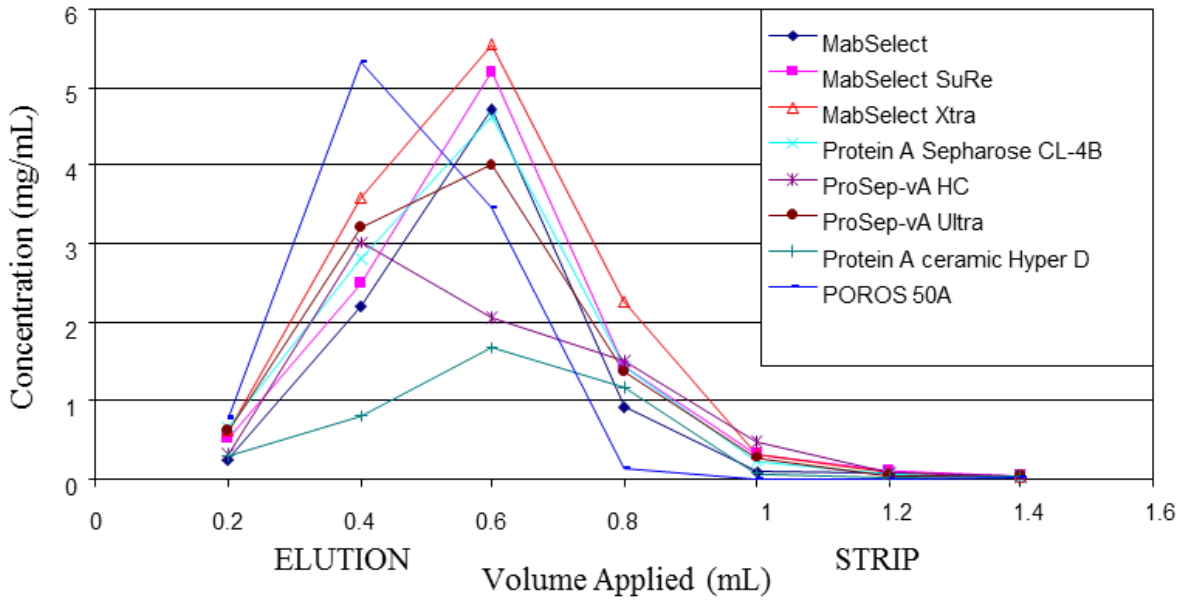


Figure 5.8: Protein A capture using atoll starter kit

The elution profiles do appear to vary slightly across the 8 different resins especially Ceramic Hyper D and Prosep HC which have much less product in the eluate than the other protein As. This shows despite all the resins having the protein A ligand attached to agarose beads, there are many other features to the media (such as ligand density, bead and pore size) which can effect process performance or benefit from screening or some optimisation. Protein A chromatography is the centre piece of the MAb platform process representing its

one size fits all proposition however screening studies like this show the current generation of antibodies sourced from high density cell cultures are displaying features not historically seen before. This is where high throughput capacity allows process development to remain on top of optimal processes.

The yields for the 8 Protein A resins varied from 38 – 114 % with Poros 50A proving the best protein A in terms of recovery. Results of over 100 % were calculated most likely from the error which is inherent to such a small scale system such as hold up volume in the column and drops which don't dispense in the right phase. Also as the material is expelled as droplets from the column outlet to a 96 well plate below, there is a small chance these can drop into the wrong well. There is some evaporation occurring between processing and the final read in the plate reader. The analytical method assumes each fraction volume per well is 200 μL although they have previously been measured by pipette to be slightly less. However for high throughput screening experiments at such small scales and with small volumes this is to be expected and there is no easy means of determining the exact well volume.

The elution profiles indicate the majority of the resins have eluate volume of 2-3 column volumes, which is typical for this step. A low eluate volume is preferred as it concentrates the product making subsequent processing easier and also as the product is less diluted by low pH elution buffer, the eluate pH is generally higher which is beneficial for product stability. The Applied Biosystems POROS 50A resin gave an elution volume of approximately 2 CV's and a corresponding yield of 114 %. Both POROS 50A and the two GE resins, Mabselect sure and Mabselect Xtra gave good results for yield and elution volume and would be ideal candidates to begin buffer optimisation with. As Mabselect Sure has been the protein A of choice throughout the industry, alternative protein A resins are often overlooked due to MabSelect's long history of the being the choice protein A and GE's excellent security of supply. High throughput screening kits such as this can efficiently isolate the protein As best for the product and accelerate process development.

5.3.4 Scale up comparison with lab scale protein A

To compare the robocolumn results with labscale chromatography, a pre-packed 10 mL protein A MabSelect SuRe column was run on an akta purifier chromatography system (GE Healthcare, Buckinghamshire, UK). The same starting material was used, loading the column to 25 mg product to mL column volume). The parameters of the column are listed in Table 5.2. As the bed height is different for each format, traditional scale down by e linear flow rate would be inaccurate to use therefore the residence time was kept the same.

Table 5.2: Microscale and labscale Protein A column comparison. The column dimensions and operating conditions of the MediaScout® MiniChrom® and Atoll Robocolumn.

Column format	Labscale	Robocolumn
CV (mL)	10	0.2
Loading capacity (mg/mL)	25	25
material loaded (mg)	250	5
Column diameter (cm)	1.1	0.5
Column height (cm)	10	1
Flowrate (mL/min)	2.5	0.05
Residence time (mins)	4	4

Going from 200 µL robocolumn to a 10 mL lab column is a scale up by a factor of 50. A single cycle was run using 50 mL of clarified cell culture supernatant, using the same set of buffers used in the Protein A resin screening study. Note, loading to an equivalent capacity only required a 1 mL of the feed, tremendously reducing the sample requirements. No adjustments of the supernatant pH or conductivity were made and it was at pH 7.53 and 8 mS/cm.

The chromatogram for the run is shown in Figure 5.9 starting from the point of sample injection (red vertical dashed line), there is a short re-equilibration and a high salt wash

followed by elution. The impurities are removed mainly in the flow through signaled by the high A280 absorbance during the load phase and any impurities sticking to the bound product or resin by non-specific interactions are disrupted in the high salt wash. The elution peak is sharp and symmetrical, concentrating the material 2.5-fold. The cycle finishes with a 0.1 M sodium hydroxide strip to strip off any anything remaining on the resin and then it is put into storage buffer.

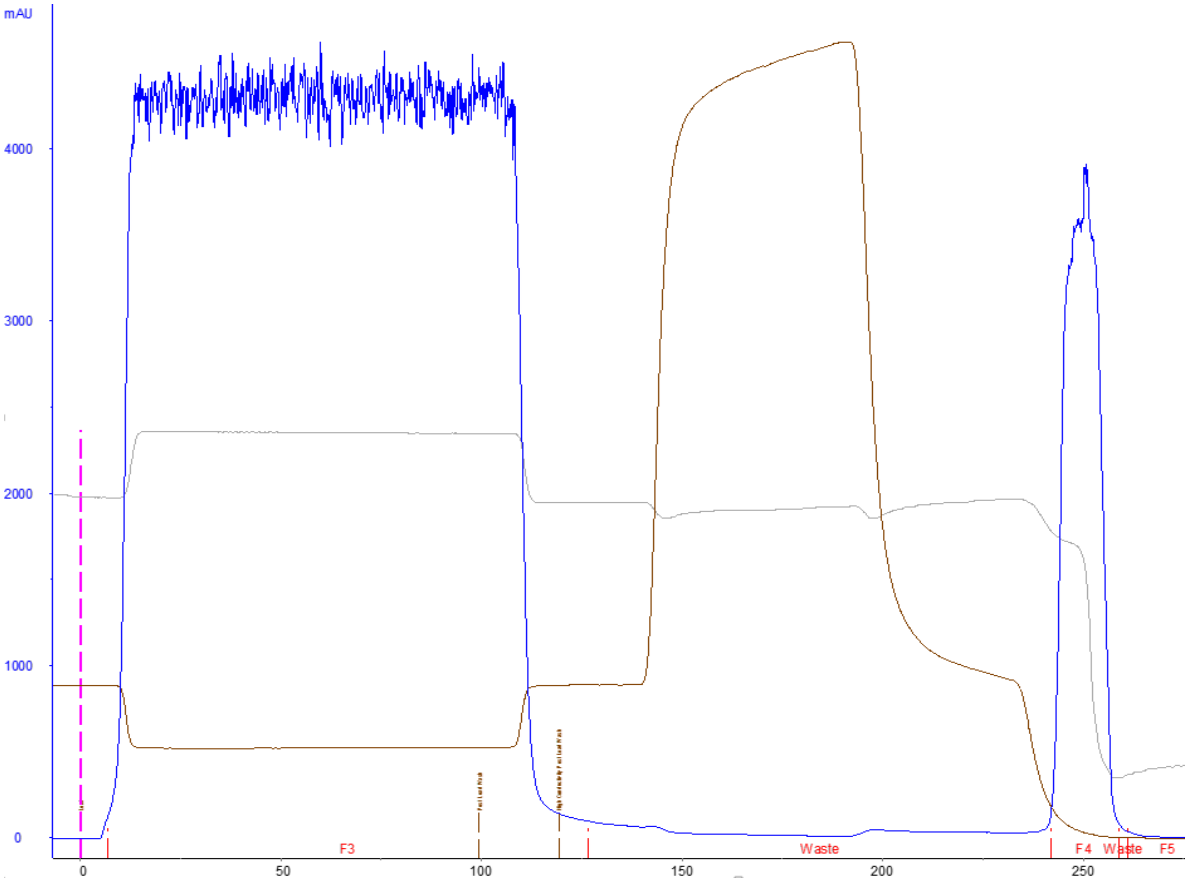


Figure 5.9: Labscale Protein A chromatogram. The blue trace is the A280 (protein signal), grey is pH, brown is conductivity and the pink dashed line is the start of load injection

The lab scale Protein A step had a recovery of 83 % whereas the robocolumn achieved slightly higher at 93 %. The scale down system seems to be over predictive in terms of yield, however on consultation with my industrial supervisor the recovery for the lab scale protein A is typically more than 90 %. It is postulated the age and cycle history of the

labscale column was to blame for its slightly lower recovery. The atoll column was new out of the pack and the lab scale column had been run almost a hundred cycles since packing with fresh resin and it is expected with this level of use the binding capacity of protein A is reduced (mainly due to the repeated exposure of the protein A ligand to 0.1 M sodium hydroxide for sanitisation in every cycle). Table 5.3 sums up the results of the two scale protein A experiments. The perceived difference in eluate CVs (5 vs 1.7) and concentration is result of the collection criteria, where the Akta a uses A280 absorbance sensor to isolate the eluate the same method cannot be applied on the Tecan (due to the small volumes and droplet elution) so the whole elution phase is collected (having a diluting effect). Individual analysis of the robocolumn fractions do suggest the actual eluate A280 peak is very similar with an eluate CV between 2-2.5 CVs (comparing Figures 5.8 and 5.9).

Table 5.3: Comparison of labscale and microscale protein A chromatography

Column volume (mL)	**Eluate Volume (CVs)**	**Eluate pH**	**Eluate concentration (mg/mL)**	**Yield (%)**
10	1.7	4.5	12.6	83
0.2	5	4.5	6.6	93

5.4 Conclusion

After a thorough review of the literature for scale down chromatography formats work was initiated using Atoll robocolumns. Out of the 3 major formats, they provide the best small scale mimic and use the conventional mode of operation. The major downside being the much greater upfront costs in having a robotic workstation such as the Tecan platform to operate the columns. However the station itself has many other uses in process development and when combined with analytical devices such as a plate reader and HPLC, the high throughput potential greatly surpasses other devices. The robotic handler-analytics integration also allows in theory a feedback driven experimental strategy (such as the simplex method, or successive rounds of DoE) to run itself.

A proof of concept study for automation of high throughput chromatography screening was performed using Atoll's robocolumn units operated on the Tecan EVO 200 liquid handling robot. Clarified cell culture containing IgG_4 served as the feed for the protein A experiment and whilst a protein A purified IgG_4 stock (by labscale protein A) fed the cation exchange chromatography screening. Chromatography methods mimicking the operations at labscale were written on the Tecan evoware software as scripts trying to make the scale down model as representative as possible. Due to the much smaller volumes and heights of the robocolumns, the usual packed column scale down method of maintaining an identical linear flowrate between sizes would not be accurate therefore the flowrates were adjusted for equivalent residence time.

Protein A and CEX starter packs were used to screen preliminary performance of a selection of current resins on the test feed. Out of the 8 protein A resins Poros 50A and Mabselect Sure Xtra performed the best on yield and eluate volume. The options available in protein A chromatography are a recent occurrence due to the expiration of the GE's patent therefore it is expected many alternatives will be available to screen for antibody capture. Of the cation exchangers, Ceramic Hyper D and Capto S gave the highest recoveries at the preliminary screen. The next stage for cation exchange development would be to screen loading buffer conditions for increasing binding capacity and to screen elution buffers to remove aggregates. An experimental design approach would be the most efficient way to conduct this.

During the processing it was noted that bubbles in the fraction samples could severely affect the absorbance readings at 900 nm and 977 nm, used to compute fraction path length and sample volume. This subsequently led to inaccurate calculation of product eluate sample concentrations and process yield values. Centrifuging the fraction plates prior to scanning in the UV reader appears to eliminate the bubbles and the variability observed at 900 nm and 977 nm. However, the use of centrifugation would require the introduction of a manual step to an automated process. It was also noted the reproducibility of the robocolumn results seemed to deteriorate after 4 – 5 cycles. This is unlikely to be caused

by resin deterioration as the labscale columns use the same resin and can provide reproducible results for up to a hundred cycles. From close observation of the Tecan chromatography process it was seen the robocolumns begin to leak during injection from the top insertion point. This suggests the inlet seal (rubber O-ring) becomes leaky over time from the repeated injections of the metal Tecan tips. Or possibly the rubber is being compromised by chemicals in the buffers. It is recommended the re-use of the Atoll columns should be limited to 3 cycles until the cause of the leaking inlets is determined.

6. Cation exchange chromatography optimisation with the DoE-simplex methodology

6.1 Introduction

Aggregate level is one of the key critical quality attributes (CQA) of an antibody therapeutic due to the potential of them causing an immunogenic response (Rosenburg, 2006). Cation exchange chromatography (CEX) uses the chemistry of negatively charged ligands attached to media to selectively bind positively charged species. Antibodies, which are zwitterionic, display a net positive surface charge in mildly acidic conditions and this feature is used to purify them in CEX. Whereas the affinity chromatography step is excellent at isolating the antibody from HCPs and DNA, aggregated and fragmented antibodies (which also require removal) carry the same affinity mechanism as the antibody and will co-elute with the product. Fortunately, ion exchange chromatography is excellent at removing aggregates as the product antibody is usually monomeric, and multimeric aggregate species display a much stronger ionic charge. CEX has been shown as a robust method at removing or reducing aggregates to acceptable levels (Faherner et al., 2001; Stein and Kiesewetter, 2006; Suda et al., 2009; Rea et al., 2011). Of all the steps in bioprocessing CEX is perhaps most tedious to optimise due to the variation in antibody surface chemistry and many processing parameters in a typical bind and elute operation. To improve column utilisation high binding capacity is desired which is a function of loading conditions (buffer pH, salt concentration, flow rate), if fragment removal is required the wash phase conditions are critical and elution phase will target product elution whilst trying to keep the aggregates bound (as they are most positively charged of the antibody species). Strict purity target of less than 1% aggregate are often the norm though this is determined in safety testing in phase 1 trials (FDA guidance, 2001).

This chapter investigates CEX as a polishing step for removing aggregates from a range of antibody products. CEX will also remove HCPs and DNA to some extent however these impurities can be removed in other downstream process steps, whereas CEX is usually the only aggregate removing step (McDonald et al., 2016). The DoE-simplex methodology will be used to optimise a range of factors and all experiments will take place using the atoll robocolumns and the experimental methods developed on the Tecan liquid handling robot in the previous chapter. Every feature of the experiment has been developed to be high throughput friendly and integrated with analytics, namely the plate reader for determining protein concentration and the HPLC for aggregate quantification (using high pressure size exclusion chromatography). Four different products with differing aggregate levels are used in this study and each will most likely have different optimum buffer conditions for aggregate removal. The CEX step is typically where a lot of the process development optimisation work is focused in a mostly generic platform purification process so it will be a realistic development scenario for the DoE-simplex methodology to demonstrate its uses.

6.2 Results

For this investigation four different antibody based products were used for CEX optimisation. An IgG_4 with high aggregates is initially used for resin selection. The DoE simplex methodology is then applied to optimise an anti-insulin antibody and this is compared with a response surface DoE optimisation. An IgG_1 with a high aggregate clearance requirement is then optimised with the protocol. And finally the simplex method is used in two standalone experiments looking at improving productivity over CEX for a MorAb product and removing aggregates from an IgG_4 feed using flow through mode. All products are owned by Lonza Biologics Plc and were produced in a CHO mammalian cell line.

6.2.1 Cation exchange resin selection

Initially a screen of potential CEX media was conducted on the liquid handling robot to select the resin to use for the experimental design studies. All the CEX options fall in the SO_3 strong cationic chemistry which are a new generation of high flow rate and high binding capacity resins. A test feed of protein A purified IgG_4 was used, which had artificially been made aggregate rich. The feed was subjected to a hold step at 50 ^oC so product degradation would be induced, taking the aggregate content from 4% to 15%. Although not strictly necessary, it was postulated the high aggregate feed would magnify the differences between resin aggregate removal allowing a clear selection (this method also mitigates against sources of error and variation which are more prevalent at this scale as shown in chapter 5).

The CEX resins were operated in parallel using the Tecan handling robot. Equilibration and loading conditions were identical, using 25 mM sodium acetate pH 5 buffer and load adjusted to pH 5. The product was loaded to 30g/L on the 200 µL columns and washed with the equilibration buffer before the elution buffer was applied. As it was not known what salt concentration in the elution buffer to use, 4 different buffers were tested ranging from 150 mM - 300 mM NaCl, and buffered with 25 mM NaAc pH 5. This formed 32 individual chromatography experiments for the resin and buffer screen and the overview is presented in Figure 6.1.

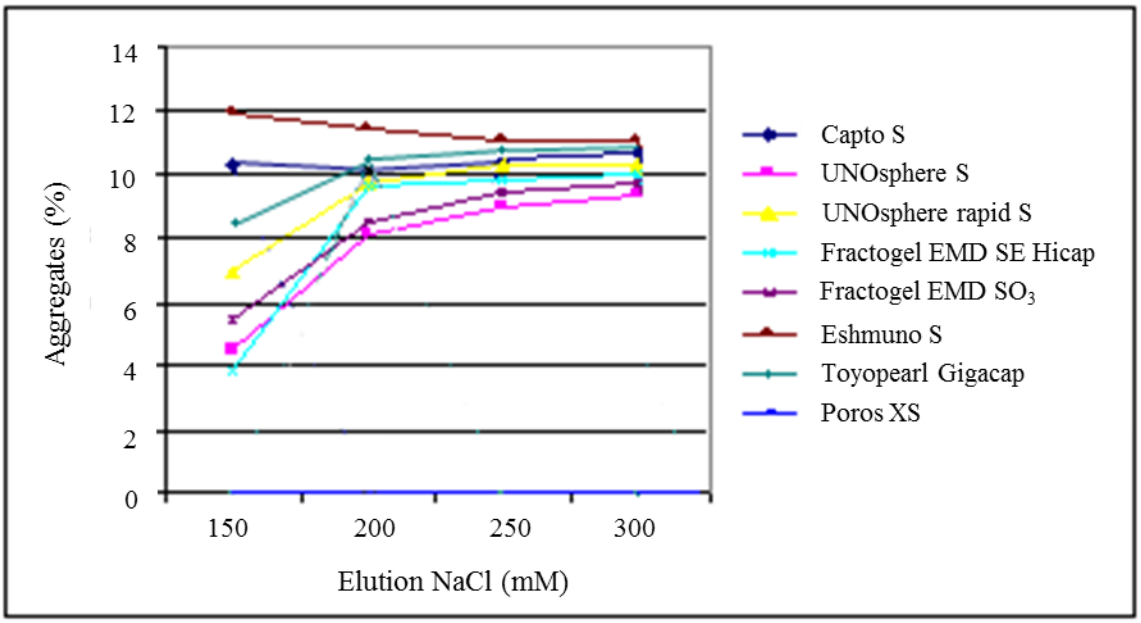


Figure 6.1: CEX options and elution buffer screen. CEX options and elution buffer screen. 8 different CEX media were assessed over 4 elution buffers for aggregate removal from a feed containing 15 % aggregate. UNOSphere S and Fractogel SO_3 displayed greater removal over all salt conditions and Fractogel SE Hicap reduced to 4 % at 150 mM NaCl although it is not the best for the other salt conditions.

The loading conditions were not investigated here however 25 mM NaAc pH 5 is a stable and commonly used buffer for the CEX loading phase. A lower pH would increase binding capacity however at the risk of product stability. Wide intervals of 50mM were used to cover a wider range of elution conditions, and the eluates were analysed by HPLC size exclusion column (Tosoh, Tokyo, Japan) to determine the aggregate level and compared with the feed level to determine the reduction.

The lowest elution buffer salt concentration of 150 mM gave the best aggregate clearance for all the CEX options. Fractogel SE Hicap (Merck-Millipore, Watford, UK) was able to reduce the aggregate content down to 4 % which is a remarkable 11 % removal. UNOSphere S (Biorad, Watford, UK) and Fractogel SO_3 (Merck-Millipore, Watford, UK)

were also good, clearing to 4.2 % and 5.5 % aggregate at 150 mM but exceeded Fractogel SE Hicap in all the other elution salt concentrations tested. Eshmuno S (Merck-Millipore, Watford, UK) and Poros XS (Thermofisher, Paisley, UK) were the worst performing, providing no better than 12% final aggregate level. Analysis of the flow through fractions for protein concentration revealed UNOSphere S and Fractogel OSO_3 to have the least protein, therefore most of the product had bound indicating better dynamic binding capacities. UNOsphere S also had best aggregate clearance at 200, 250 and 300 mM NaCl elutions therefore due to this consistency and the higher binding capacity it was selected to be developed an optimised CEX bind and elute step with the antibody products.

6.2.2 DoE-simplex methodology applied to optimising anti-insulin CEX chromatography

Optimising the mid-process polishing step has a number of desired goals apart from the removal of aggregates (although it is the primary aim). Generic processing goals such as a high yield, maintaining product stability (less salt in the eluate), process productivity (fast flow rate and high binding capacity) are all part of the optimisation.

The first product to be investigated was an anti-insulin antibody which already has low level of aggregate in the starting material at 3.5%. As per the methodology, an initial screening DoE was used to map out product yield and purity (aggregate clearance) by varying loading pH, loading NaCl concentration, elution pH and elution NaCl concentration (the equilibration buffer matched the loading pH and salt concentration to keep the loading phase uniform in pH and conductivity) . The ranges for the conditions were chosen after conducting preliminary experiments in which the residence time and binding capacity were also fixed at 4 minutes and 25 mg/mL (see Chapter 5). The columns used were atoll 200 µµL robocolumns, pre-packed with UNOsphere resin. The mode of chromatography was a bind and elute operation therefore the product would be collected

in the eluate. The factor ranges determined from preliminary work were pH 5 – 6.5 for both equilibration and elution buffers, and the NaCl concentrations were 25 mM – 100 mM, and 150 mM – 250 mM for the equilibration and elution buffers, respectively. For the simplex method, a weighted response combining yield and purity would be used representing the best compromise. This response is represented in equations (6.1), (6.2) and (6.3).

$$P = 1- Agg\ \%; \tag{6.1}$$

for P > 98 %

$$CR = 0.5Y + 0.5P; \tag{6.2}$$

else,

$$CR = (0.5Y + 0.5P)/10; \tag{6.3}$$

where P is purity, Agg is aggregate level in the eluate, CR is the combined response objective function, and Y is yield.

An equal weighting between the two responses is used only if the condition has a purity greater than 98%. If the purity is less than 98% then a penalty applied to result essentially making it very undesirable to the simplex (equation 6.3). This ensured that yield would only be considered in the optimisation decision after the conditions returning the purest eluates had been screened by the objective function.

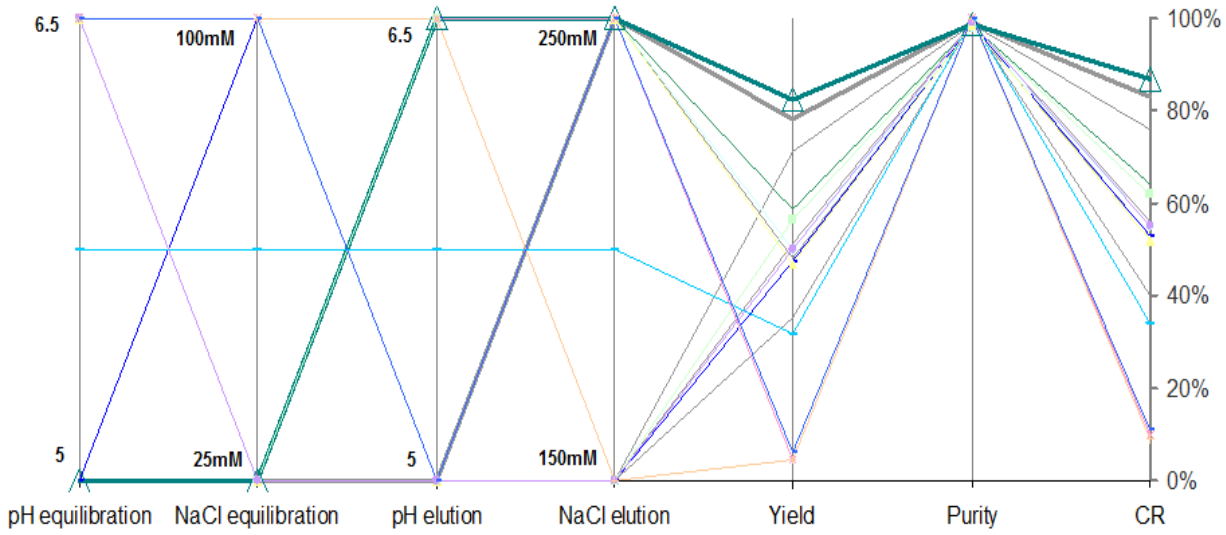


Figure 6.2: Initial DoE data on UNOSphere S for aggregate removal

The initial DoE was a 2 level factorial design requiring 17 samples (including a centrepoint) of which the design space and results are shown in the parallel coordinate plot in Figure 6.2. The parallel coordinate plot shows a complete overview of all the conditions and results in one graph and is especially useful in experiments carrying more than 3 factors (a heat map or contour surface can only show one response and 2 factors at a time). Heat maps of the data are also shown in Figure 6.3 where some factors are held constant. The conditions were evaluated using a strip of 8 UNOsphere 200 µL robocolumns. Care was taken that each robocolumn was not used for more than 3 cycles before a new strip was opened. The data was fit to the factorial model defined by:

$$CR = 156 - 20\ pH_{eq} - 0.46\ C_{eq} + 4.67\ pH_{el} + 0.068\ C_{el}; \qquad (6.4)$$

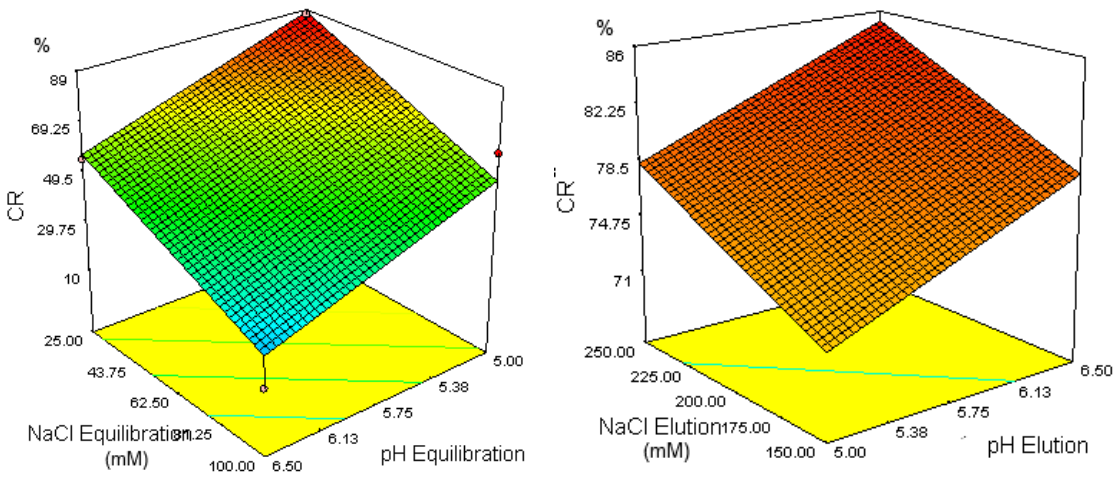
where pH_{eq} is equilibration phase pH, C_{eq} is equilibration phase NaCl concentration, pH_{el} is elution phase pH and C_{el} is elution phase NaCl concentration.

The models had a R^2 value of 0.65 and the equilibration pH and NaCl concentration were the most significant factors. The predictive ability is not very good due to significant curvature for the linear fit model although it is in a very large design. The optimum point suggested by the model was at pH_{eq} 5, C_{eq} 25 mM, pH_{el} 6.5, C_{el} 250 mM. This was a condition evaluated in the experimental design and achieved a CR value of 0.87, which corresponds to 82.5 % anti-insulin yield and 1.4 % aggregates remaining (see Figure 6.3). The model was also used to predict conditions for the highest yield where aggregates were under 1% and the result was pH_{eq} 5, C_{eq} 70 mM, pH_{el} 5, C_{el} 150 mM, at a yield of 49.1 %.

Table 6.1: DoE model predicted optimums. Best conditions and responses. Responses for (a) highest combined response (CR) value and (b), best yield at 1% or under aggregate level.

	Loading pH	Loading NaCl (mM)	Elution pH	Elution NaCl (mM)	Yield (%)	Aggregate (%)	CR
(a)	5	25	6.5	250	82	1.4	0.87
(b)	5	70	5	150	49	1.0	0.74

Generally a little to no amount of salt in the equilibration and loading phases allows maximum available binding sites on the CEX beads to bind the antibody. Aggregates are typically dimers, trimers and tetramers of the monomer antibody so will exhibit a stronger charge due to multiple potential binding sites. Aggregates bind more strongly than the monomer antibody so can only be removed by finding the right elution conditions which dislodge the monomer but leave the aggregate bound to the column. Anything left bound is removed in the high salt strip (1 M NaCl) and sanitisation phases (0.1 M Sodium hydroxide).



(a) (b)

Figure 6.3: Combined response surfaces from DoE. Where a) the elution variables were fixed at pH 6.5 and 250 mM; b) the loading conditions were fixed at pH 5 and 25 mM.

The fact that the model has a low R^2 value and that it is trying to model 4 factors with a combined response which includes a penalty to sub 98 % purity conditions makes for a very complicated system. Figure 6.3 displays heat maps of the linear model for (a) the loading conditions and (b) the elution conditions. Neither are surprising as it suggests the best CR function would be low pH and low salt for loading and high pH and high salt for elution which is in line with CEX theory. The model equation suggests equilibration pH as the most significant factor while the other terms are appointed lower coefficient values in equation 3.

Loading pH on a CEX step primarily determines how strongly the antibody (and aggregate) binds. This is possibly due to the low level of aggregate in the first place hence elution conditions are not so critical. As most conditions cleared the aggregate to under 2 % the penalty was not applied so the model is very much dependent on yield as the driving force. The effect of 98 % purity as opposed to 99 % has little bearing on the CR whereas yield has much more variation The DoE data suggests the CR having equal weighting for yield

and purity may not be as useful. Based on the best conditions in Table 6.1, a simplex of 5 conditions was deployed. The same design space was used and the intervals between units were set at 0.25 pH units and 25mM, which allowed the simplex to make very small steps. The results of the simplex and its conditions are shown in the parallel coordinate plot in Figure 6.4. The design space was segmentation so the initial simplex contains the corners of the segment in which the predicted optimum with the highest CR (condition (a) from Table 6.1) was found.

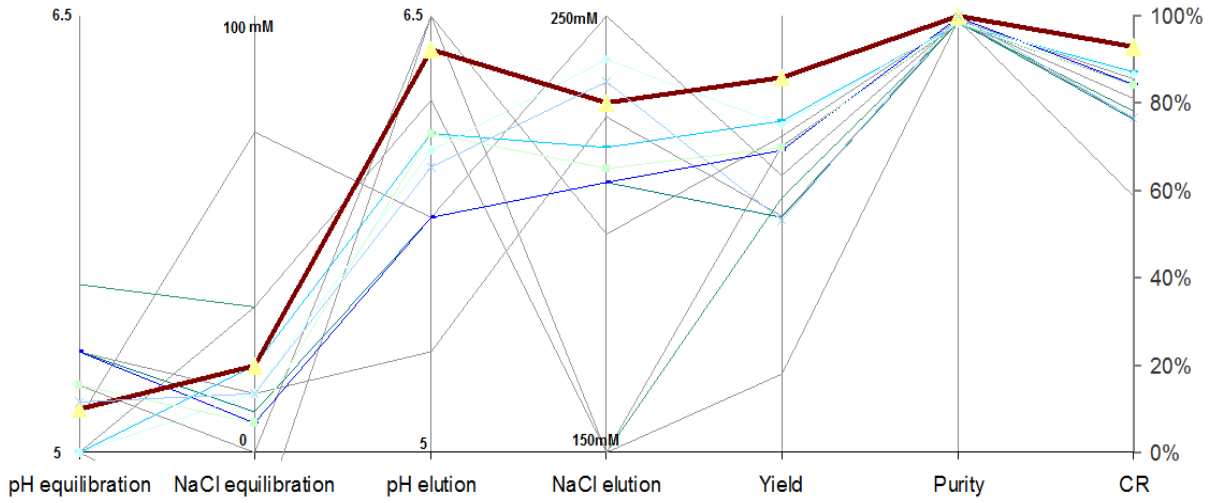
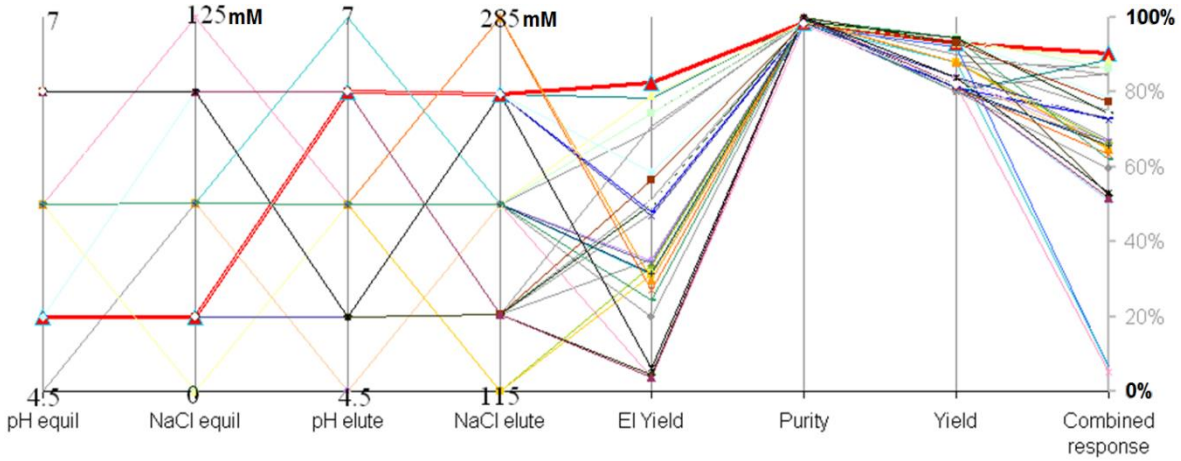


Figure 6.4: Simplex optimisation of aggregate removal with UNOSphere S. The second part of the optimisation uses the simplex method. 13 samples were required to locate the optimum (conditions on brown line), achieving a yield of 87 % and purity of 99.0 %

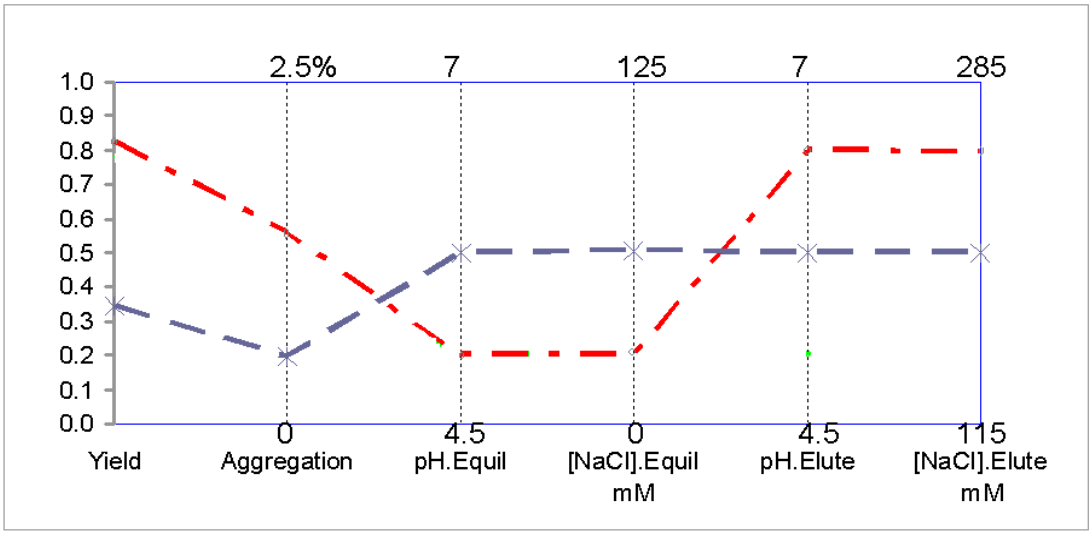
The simplex used 13 conditions (which are shaded in Figure 6.4) to locate the optimum CR at pH_{eq} 5.25, C_{eq} 25mM, pH_{el} 6.25 and C_{el} 225mM. This corresponds to a combined response of 0.93, the yield was 87 % and purity, 99 % (marked by the thick brown line/yellow triangles in Figure 6.2). The DoE-simplex methodology required a total of 30 conditions to be evaluated to find a set of conditions which provide good yield and purity (87 and 90 % respectively). As part of the investigation the DoE showed a local area of the design space where the global optimum might be and mitigated against the chances the

simplex method would converge on a local optima. The 2 rounds of experimental design provide a robust characterisation of the design space on a local and whole design space levels which provides great confidence in the solution and robustness around the global optimum.

A comparison experimental design to the DoE-simplex methodology a higher resolution DoE was selected to optimise the same design space. A central composite design (CCD) was selected which is a response surface method (RSM) which requires a higher number of samples but can account much better for the curvature of complicated functions with good predictive capability. Using the same UNOSphere S columns, anti-insulin starting material and the four factors a design was created and the necessary buffers. A CCD varies the factors in a pattern of axial, midpoint and star points (outside design space) so there are 5 levels in each plane. It is almost a supplemented full factorial design. For the DoE a total of 36 samples were evaluated in the study and included conditions which were outside the factor boundaries used in the previous work due to the star points. The larger dataset allowed the results to be fitted to a quadratic model. The design and results from the CCD experiment are shown in a parallel coordinate plot in Figure 6.5 (a) and as heat maps in Figure 6.6. The two best yield and aggregate conditions are highlighted in Figure 6.5 (b).



(a)



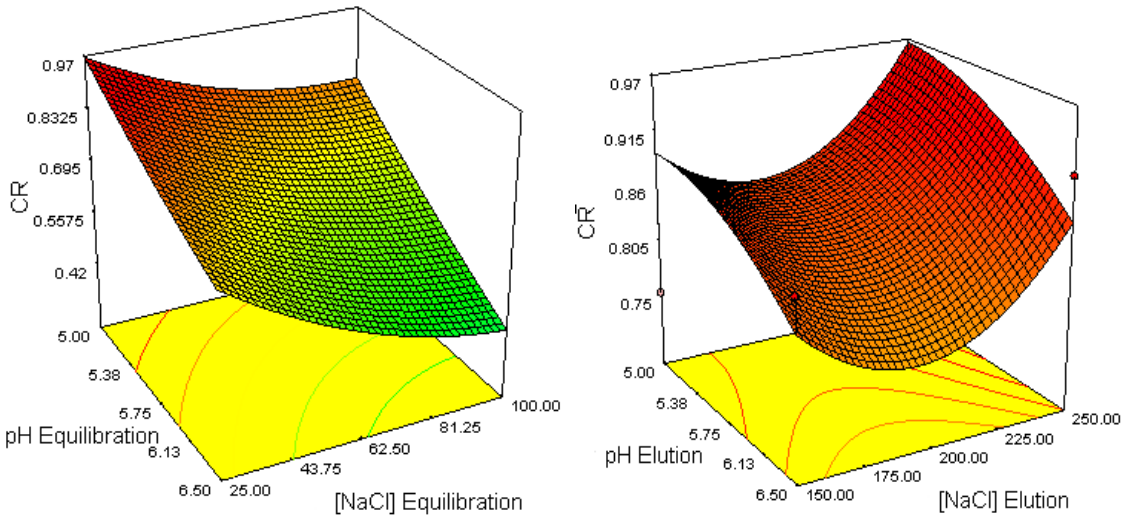
(b)

Figure 6.5: CCD results for aggregate removal. (a) DoE results and (b) the optimum conditions for CR (red dotted line) and yield with 1 % aggregate (blue dotted line).

The CCD model has a low R^2 value of 0.34 and a low model F value of 7.7 indicating a significant lack of fit. This is interesting compared to the factorial model used in the simplex pre-screen which had a better fit. The more samples in this dataset reveal just how much more curvature there is over this system of 4 factors which the simpler model did not reveal. The factorial model had half the number of data points with only 1 centre point (for all 4 factors) therefore the model is unaware of how inaccurate it is, although its R^2 is still

higher at 0.64 (though still poor). The more levels per factor of the CCD revealed the complexity of CR response surface which shows in the statistics. The predicted optimum of the CCD model is identical to that suggested by the factorial DoE (pH_{eq} 5, C_{eq} 25 mM, pH_{el} 6.5, C_{el} 250 mM).

The heat maps in Figure 6.6 are split for the loading and elution conditions. The loading response surface is similar to the one created earlier with the factorial DoE with the optimum at low pH and low salt. The surface is mostly linear. The Elution factors in fugire 6 (b) display more curvature. The optimum this time is at low pH and high salt, although the CR rises again to create a secondary optimum at 150mM salt and low pH. Elution pH seems to have less of an effect in the quadratic model.



(a) (b)

Figure 6.6: CCD model response surfaces. Response surfaces showing the optimum for loading (a) and elution (b) conditions (where the elution variables are fixed at pH 6.5 and 250 mM for (a) and the loading conditions are fixed at pH 5 and 25 mM for (b).

The calculated response values of the quadratic model and the linear model are quite different however and neither is close to the real experimentally determined value. The quadratic model predicts the best CR of 0.96 (75 % yield and 0.85 % aggregates remaining) whereas the experimental value was 0.87 (82 % yield and 1.4 % aggregates remaining). The model also predicts the highest yield of the sample with a <1% aggregate level of 36% at pH_{eq} 5.75, C_{eq} 63 mM, pH_{el} 5.75, C_{el} 200 mM. The two experiments show that modeling of high dimension and large design spaces is very difficult with DoE alone and can require many experiments. The models generated are often poor at predicting where the optimum is and the mapping of the local region can be very inaccurate. A technique like the simplex method which does not use models to locate optimum conditions has the advantage in such scenarios.

The case study is typical in Chromatography optimisation where many variables have strong interactions that can greatly affect the objective function. Using the factorial design and the simplex method required a total 38 conditions to be evaluated and the optimum condition found at pH_{eq} 4.5, C_{eq} 35 mM, pH_{el} 6.25 and C_{el} 225 mM provided a CR of 0.93 which had superior yield and aggregate clearance (87 % yield and 1 % aggregates remaining) than the CCD located optimum. The CCD required a slightly more 36 samples and provided further data for the complexity of the response surface, yet its estimation of the optimum was no better than the 17 sample factorial design used before the simplex method. To develop the process from the CCD data would certainly require another DoE but this time with narrower ranges centred on the predicted optimum, which would double the sample count. The data from the simplex method could be used to create a local region model using advanced modelling techniques such as delauney triangulation which would achieve a similar result but with little to no extra experiments.

6.2.3 DoE-simplex method applied to IgG_1 aggregate removal with UNOsphere S

The DoE-simplex methodology was used to optimise aggregate removal from a protein A purified IgG_1 feed using the sample binding and elution buffer factors in the previous work. The range for $pH_{equilibration}$ was tweaked slightly to pH 4.5 – 6.5 due to the optimum found for anti-insulin to be at the lowest loading pH setting. The other factors and ranges remained the same ($pH_{elution}$ was pH 5 – 6.5, $NaCl_{equilibration}$ was between 25 – 100 mM and $NaCl_{elution}$ between 150 mM – 250 mM). The experiment was done using a Tecan liquid handling robot with 200 µL UNOSphere S robocolumns. A 2 level factorial design DoE was applied to generate a list of conditions to evaluate. The initial aggregate level in the feed was 5 % and the target level was a high 99 % purity. The results from the factorial design experiment are presented in Figure 6.7. A combined response was used again to aid the simplex method using the same weighting and penalty as before:

For Purity (P) > 98 % $\quad CR = 0.5Y + 0.5P; \quad (6.5)$

or else, $\quad CR = (0.5Y + 0.5P)/10; \quad (6.6)$

The samples with the penalty applied can be seen in Figure 6.7 with the very low CR values. The optimum conditions for the best CR are predicted to be at pH_{eq} 4.5, C_{eq} 25 mM, pH_{el} 6.5 and C_{el} 250 mM which achieves CR of 0.93, or a yield 87 % and purity of 98.3 %.

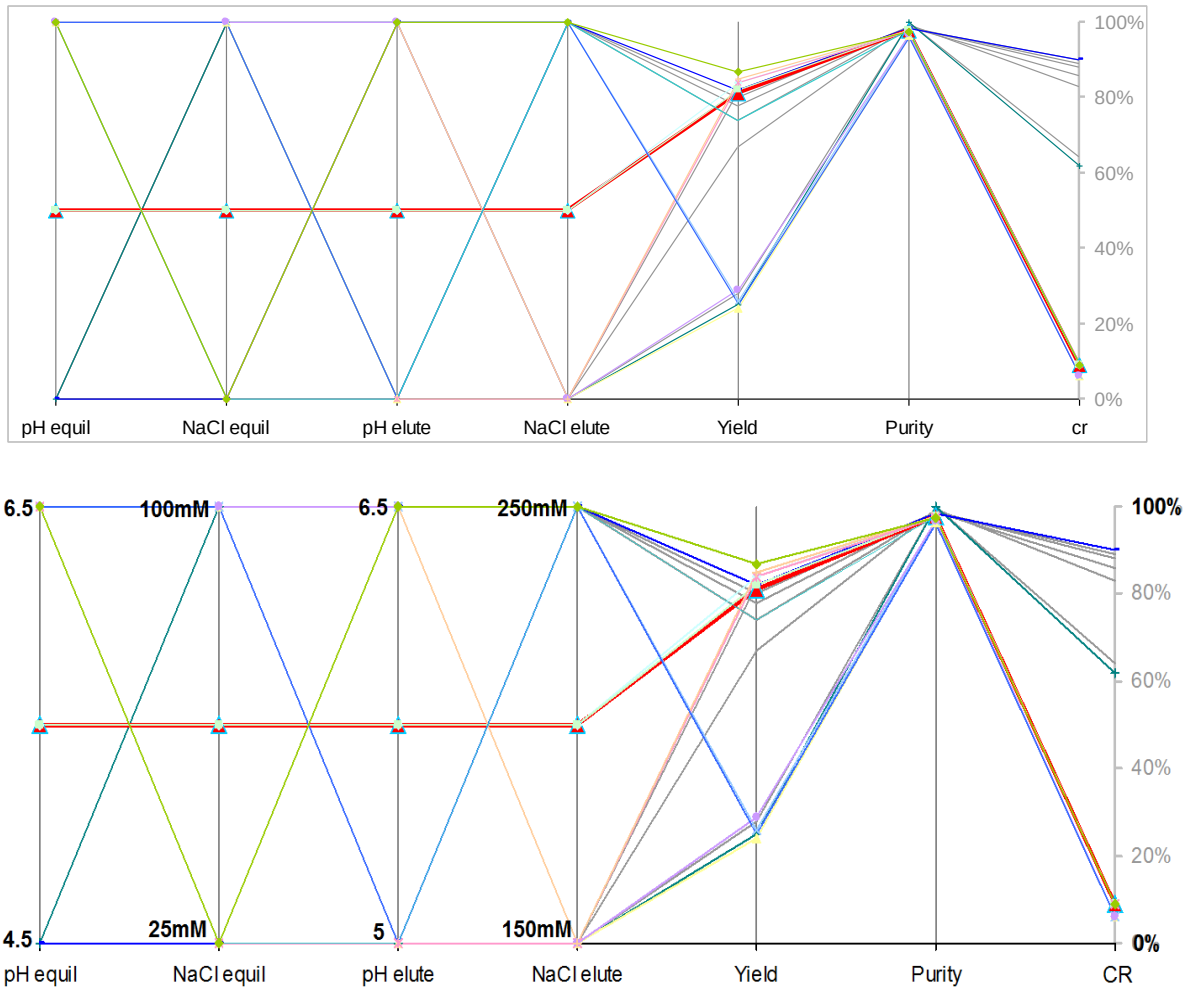
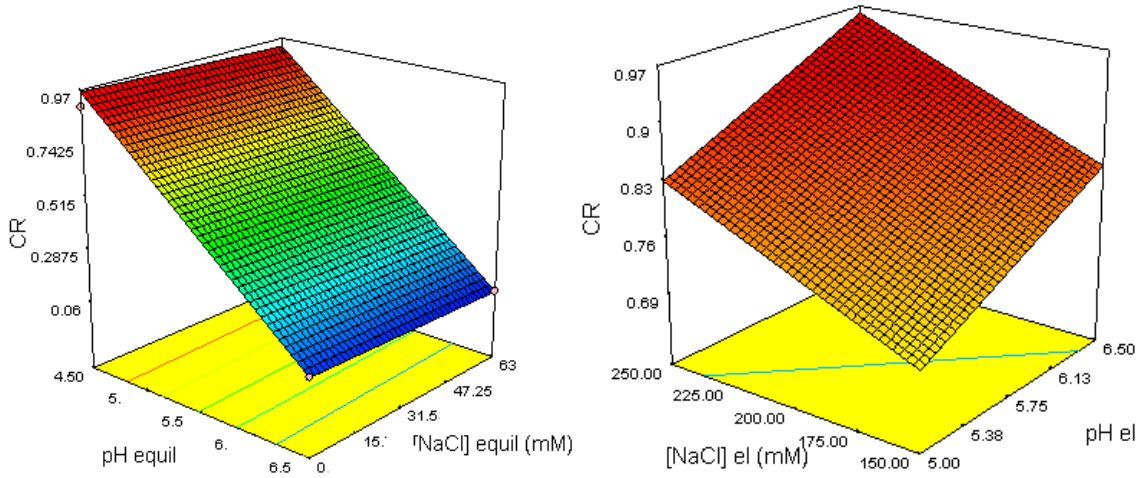


Figure 6.7: Factorial design DoE result for the IgG_1 feed

The dataset of 17 samples (including 1 centre-point) was fit to a linear polynomial model and response surfaces from the resulting model are shown in Figure 6.8. As the purity was still lower than required for the target of the study, the CR penalty was changed from 2 % to 1 % and this dramatically altered the model moving the optimum to pH_{eq} 4.5, C_{eq} 50 mM, pH_{el} 5.5 and C_{el} 250 mM with the purity at 1 % and yield dropping to 73 %. The response surfaces in Figure 6.8 show the loading salt concentration has less of an effect on CR than the loading pH and elution conditions. A low pH loading is the most critical factor to a high CR as shown by the high CR values across the surface in Figure 6.8 (b). This is likely due to the penalty not being applied in a low loading pH samples as they are purer,

the aggregates bind more strongly in an more acidic environment and are less are eluted with the product. The drawback of using such low pH for the loading phase is the product also binds more strongly and this is reflected in the lower yield. Product stability may also be a concern for the product will remain at low pH for most of the chromatography cycle.



(a) (b)

Figure 6.8: Response surface of IgG_1 factorial design DoE. The effect of loading (a) and elution (b) conditions on the CR (for (a) the elution buffer is fixed at pH 6.5 and 250 mM, NaCl and for (b) loading is fixed at pH 4.5 and 25 mM) NaCl).

For selecting the conditions for the initial simplex the design space was segmented and the corners of the segment that contained the DoE predicted optimum was used. The CR which applied a penalty to anything below 99% pure was used. The simplex method used an extra 8 experiments which were evaluated sequentially on the Tecan using a strip of UNOSphere robocolumns (shown in Figure 6.9). The limited uses of the robocolumns do add another element of variation as the same column is not used throughout the experiment. So far this has not been an issue as a new column is used after 3 cycles, and a column will only begin to leak from the inlet after 4 or 5 cycles.

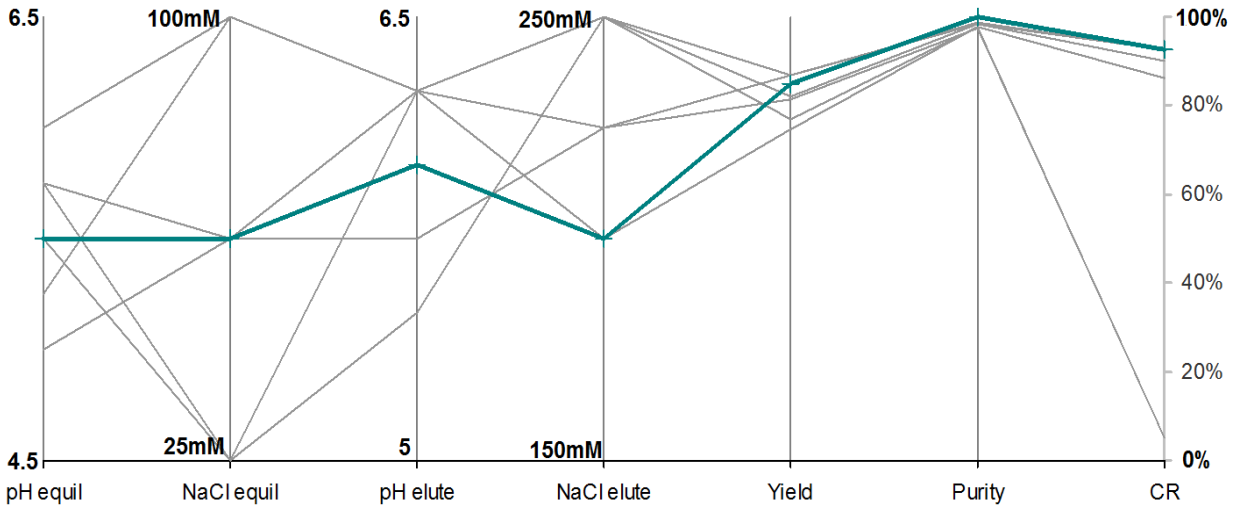


Figure 6.9: Simplex search for optimum IgG_1 yield and aggregate clearance

An optimum was found by the simplex at pH_{eq} 5.5, C_{eq} 63 mM, pH_{el} 6 and C_{el} 200 mM (shown by the bold dark blue line in Figure 6.9). This corresponded to a CR of 0.92 or a 99.3 % purity and 92 % yield. Only one condition of the simplex search had a sub 99 % purity and all the conditions were above 78 %. The conditions which achieved more than 99 % purity had quite a spread of factor values and therefore present a number of potential conditions which be used as mid-points for further for a robustness DoE and the mid-point showing the least sensitivity to the responses used for scaling up and developing the step.

6.2.4 Improving CEX productivity for MorAb using the simplex method

During method development for the Tecan based atoll CEX chromatography system the lack of literature initially gave poor binding of the antibodies when using equivalent linear flow rates with lab scale column processes. The little background data of which conditions are best suited for binding for the different antibodies used (which was largely unknown) by UNOSphere S compounded matters. Here the simplex method was then used to investigate how the step productivity is affected by how much product is loaded (binding

capacity) and the flow rate it loaded at. Product throughput calculated from yield and loaded protein was used as the objective function for the simplex method.

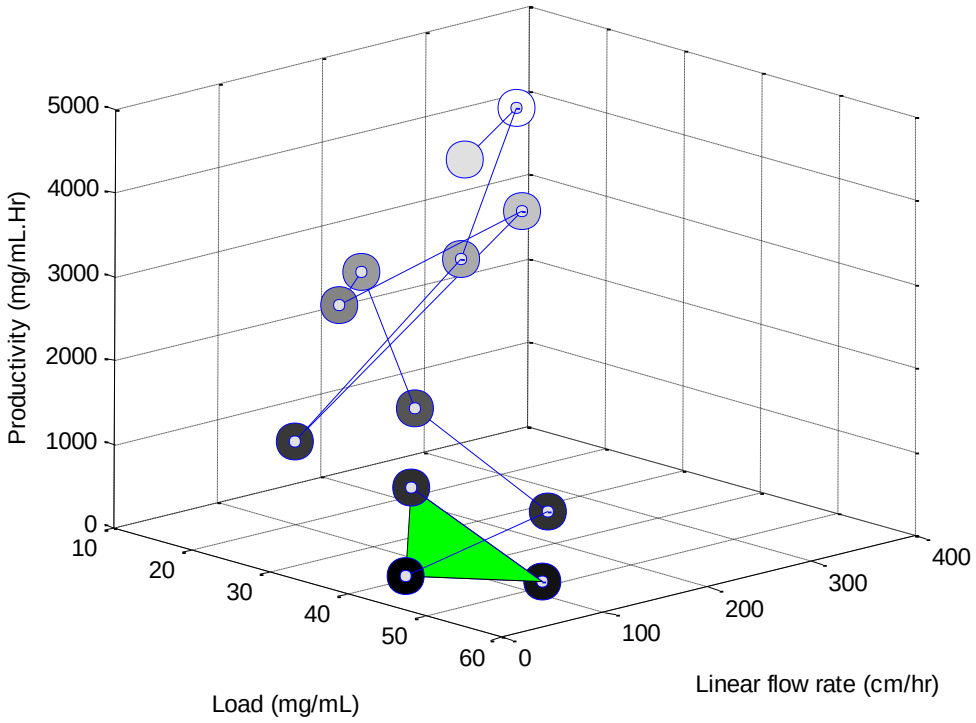


Figure 6.10: Optimising UNOSphere S productivity with the simplex method

6.2.5 Flow through mode optimisation of IgG_4 using the simplex method

Operating a CEX step in flow through mode as opposed to bind and elute has many benefits. The process is simpler and high throughput in that the impurities are bound in the loading phase and the product comes out in the flow through. For highly aggregated products conventional bind and elute mode chromatography does struggle to reduce to acceptable levels and extra steps may be required. In flow through mode it is possible to clear high levels aggregate (although it is less effective at clearing other impurities such as HCPs and DNA). As aggregates bind more strongly to CEX ligands, they will preferentially bind to the matrix compared to the monomer. When the column becomes

saturated, product will breakthrough and emerge in the flow through. Flow through chromatography will generally specify smaller size columns and larger feed volumes to saturate the column quickly and keep yield high. Once the column is saturated and monomer product emerges the theory is further aggregates loaded onto the column will displace monomers and bind to the column. This should result in an aggregate-free flow through up until the aggregates start breaking through. After the product is collected the column is stripped and the aggregate rich fraction is removed.

The high aggregate IgG_4 product used previously was used for this work (FT). The simplex method was used to optimise loading pH and NaCl concentration (of equilibration buffer and feed) for a combined yield and purity objective function. The experiment was conducted using 200 µL UNOSphere S robocolumns using the Tecan liquid handling robot. Figure 6.11 shows the results of the simplex search, with the initial simplex shown as the red triangle and the final 3 points displayed as the small green simplex triangle.

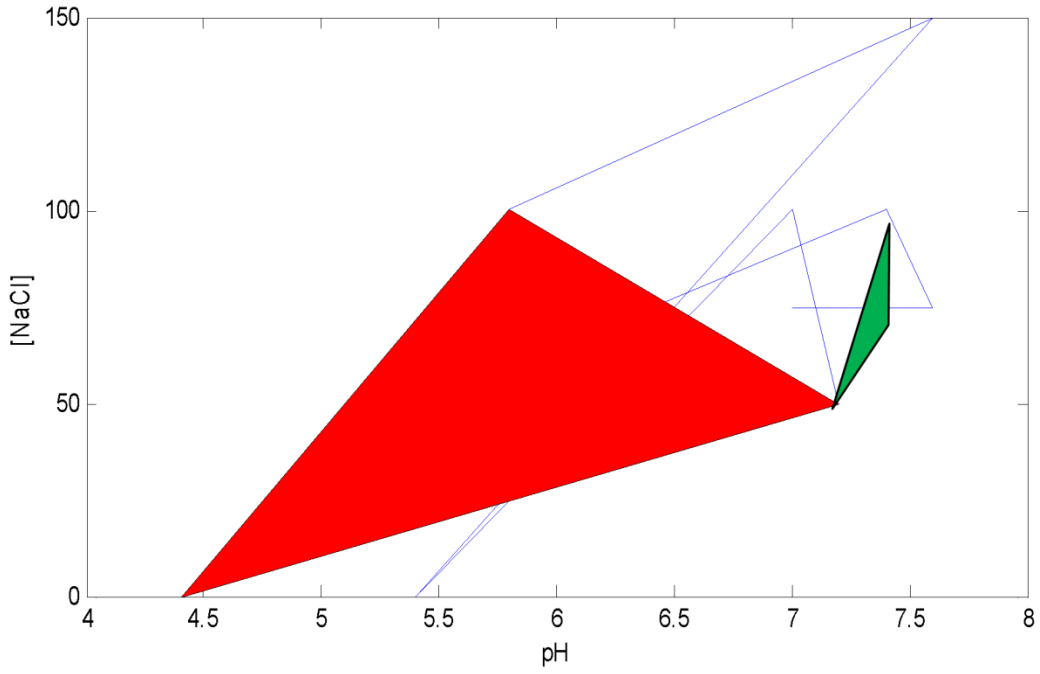


Figure 6.11: Optimising IgG_4 aggregate clearance in flow through mode. The simplex search was used to optimise the flow through conditions for maximum aggregate clearance. The red and green triangles represent the initial and final simplices.

The design space was quite large with the loading NaCl concentration ranging from 0 – 150 mM and the pH from 4 – 8. A fixed loading capacity of 100 mg product per mL column volume was added for all samples. The starting aggregate level was 15% which had been artificially increased by heat pre-treatment of the feed. Once again the 200 µL format of the UNOSphere S robocolumn was used which required 20 mg of product to be loaded in each injection. A large simplex was chosen from random conditions shown by the 3 corners of the red triangle in Figure 6.11 and the simplex method required 9 further conditions to be evaluated before converging around pH 7.5 and 75 mM NaCl, which gave a yield of 61 % yield and 98.7 % purity. The aggregate clearance is outstanding compared with the bind and elute experiment with UNOSphere S with the same feed in section 6.2.2 achieved a 95.8 % purity. The yield is quite low considering a high loading pH is being used which in theory should only bind aggregates which have the stronger charge of the antibody species. It is likely a smaller column size or loading to a higher capacity will improve the yield although the change in aggregate level will have to be monitored. Some forms are aggregate can be extremely difficult to remove and may require different chemistries and chromatography operations to remove. A 61 % over one step is equivalent to two column operations with 80 % yields so it sometimes can be the preferred option especially as one step is less costly and simpler to operate.

6. 3 Conclusions

The CEX case studies presented here demonstrates how the DoE-simplex methodology can break down a complicated process development problem and optimise for the most suitable conditions. Using DoE and the simplex method sequentially provides process knowledge in breadth as well as depth. Examples of 4 factor design spaces with significant curvature were given which a comparative RSM technique was shown to struggle in modelling the design space both in model quality and in the high numbers of experiments required.

Process modelling is recommended for validation of operating spaces however RSM methods are wasted when used earlier on in process design when there are a lot of factors and large ranges. The models they generate will typically be poor predictors and a poor fit to the data. For screening, lower resolution designs such as 2 level factorials and half fraction factorials are arguably more efficient in that they will require much less experiments but offer most of the information gained by RSM methods. Higher resolution designs are most effective in the later stages of process definition where a subset of the design space has been isolated and the resulting model of the RSM will provide a very good fit of the smaller design space. The simplex method is affected much less by large spaces and will home in on the optimum efficiently. Using some DoE to provide the simplex method with a smaller area mitigates against the chances of it falling into a local optimum as shown in chapter 3. The main drawback experienced here with the simplex method was the experiments must be done sequentially, due to the feedback loop. Whilst this isn't an issue when material is very scarce it only uses a fraction of the high throughput capabilities of modern lab systems (such as the Tecan which can do 8 chromatography experiments at once). There is still benefit in saving time and effort in assays such as HPLC SEC which for aggregate determination required 15 minutes per sample. Of course where high throughput capability is unavailable and the chromatography experiments (taking 2-3 hours each time) have to be done on a one at a time system such as an Akta, the simplex method will save time.

Apart from aggregate removal, HCPs, DNA and leached protein A (LPA) molecules are also CQAs and need to be reduced to acceptable safe levels. They are often removed with IEX chromatography although they was not looked at in this chapter. Throwing these responses into the process development challenge will only make modelling of the system even more difficult as so many responses are considered simultaneously. This should improve the results gained with simplex method which does not rely on models to optimise like DoE. High throughput HCP, DNA, and LPA assays were not readily available during the investigations but is something that would be necessary in a real process development campaign. HCPs and DNA would require sandwich ELIZA assays which do take a long

time and would benefit from the simplex methods reduction in experiments. The simplex method, guided by a small factorial design DoE is a much more potent combination.

7. Challenges to validation and commercialisation

7.1 Introduction

The development of the purification steps in this thesis have shown encouraging results as alternative technologies to protein A chromatography and as polishing steps to remove aggregates. The methods were developed at microscale using automated high throughput techniques wherever possible. The development of the steps using the simplex and Doe modelling techniques created a large volume of data for the operations being studied. The DoE-simplex is especially efficient at selectively concentrating experiments (and hence predictive capability) in the more fruitful regions of the design space. DoE helps nominate the CPPs and establishes process understanding of the design space. This approach will satisfy QbD targets by offering greater understanding of how CQAs vary within the potential operating ranges of the CPPs.

This will assure the manufacturing process quality is consistent and costly batch failures are avoided. Monte carlo simulations could have been run to assess the failure rate of CQAs being out of spec for random variation of the CPPs in the proposed operating ranges. The DoE simplex methodology presented here is an efficient way to conduct process development using elements of multiple optimisation methods to achieve a superior product. It is readily scaleable to multiple factors and can help save time and investment.

*This chapter is included in partial requirement for the award of the UCL Engineering Doctorate in Biochemical Engineering and Bioprocess Leadership

7.2 Technical challenges

The key challenge remains in integrating the optimisation method with high throughput screening platforms such the Tecan and its ancillary components. Steps to commercialise the presented simplex code and experiment planning framework can be achieved after it is demonstrated to work in an industry lab. The protocol and methods have shown good scaleable performance on multi-variable experimental systems. Key challenges which remain are to convert the protocol to a robust package compatible with commonly used bioprocessing laboratory softwares. The technical review of the package would have to be made after it is trialled by process developers ideally in an industrial purification development department for a number of case studies. The key commercial benefits of the package would be savings from reduced experiments, more focused design space search. This would translate to more optimum manufacturing processes and the time created to increase high throughput screening for new drugs. Finding the optimum conditions would save on the extensive pilot scale development and/or revalidation work required if sub optimal operating conditions are taken forward. Further work is required to make the package accessible in a software package. The software should have seamless integration with major automated liquid handling platforms as well as key analytical devices software. This would the not only make the robotic platform carry out all the experiments but the software would take over decision making for identifying in which direction of the search space to proceed and evaluate.

The DoE-simplex methodology was used to take an unknown product and optimise the conditions in a large design space for a high throughput microscale mimic of a unit operation. It needs to be shown the well mapped out potential operating regions that were established for each product and step are representative at larger scales. Due to a lack of material and time the scale up work to demonstrate the conditions translated well to the lab scale work was not completed for every step. It would be ideal to carry out a number of conditions at lab scale to verify the applicability of the local region DoE models for design

space verification studies. As background and some ultrascale down work was carried out (for example with the ageing and shear treatment at the centrifugation stage), any discrepancies seen due to scale should be minor. The inter-scale transferability of the experimental methods developed here will prove the microscale models are wholly representative of the manufacturing process. The method falls complete into the FDA's PAT framework, as it sets processing design limits from the measurement of CPPs (such as pH and NaCl concentration for CEX loading and ammonium sulphate concentration for precipitation) which affect CQAs like purity and aggregate level. These CQAs were used in creating the combined responses and setting penalties for exceeding limits.

8. Conclusions and future work

In chapter 1 it was set out to examine and develop the nelder-mead simplex method so it could be used successfully in the process development of bioprocessing unit operations. The method originating from computer science and for the optimisation of algebraic functions needed to be adapted to be utilised in the lab as an experimental design method. A literature search revealed very little had been publicised on the use of the simplex method in optimising a biological or experiment let alone bioprocessing examples (not until the work done by Chhatre et al., and Konstantinidis et al., in 2011 and 2012, which was done at the same time as the work presented here). DoE methods are compared with the simplex method and are investigated how they are fully harnessing the capability afforded by high throughput platforms and microscale process development. Protein precipitation is proposed as a scaleable, low-cost capture step for antibody purification and together with centrifugation and cation exchange chromatography serve as case studies on a number of antibody products optimised with a combined DoE-simplex method approach.

The experimental simplex method was developed and tweaked using the high resolution dataset of IgG_4 precipitation in chapter 3. It was found that simplex size, location and orientation with respect to the optimum greatly affect the number of steps it takes to reach the optimum. The minimum size allowed for the simplex also affected how susceptible it was to noise and getting stuck at local optima. Comparisons with a CCD experimental design showed the simplex method was much more efficient, locating the optimum in half the number of experiments. Screening DoE designs were also shown to help the simplex method in mitigating against falling into local optima and make the search even more efficient. The global optimum found by the proposed combined use of the DoE and simplex method was found at pH 7.5 – 8.5, 2 – 2.2 M ammonium sulphate, and a feed concentration of 10g/L IgG_4 (achieving a yield of > 96 % and 82 % purity). The precipitation was done in 300 µL microwells and for future scale up for verification of the identified conditions, a scale of 100 mL – 500 mL should be used with an impeller stirred vessel. Some further work may be required to correlate impeller speeds with plate shaking speed however correlations are widely available in literature (Micheletti et al., 2006; Kensy et al., 2009).

In chapter 5 the proposed DoE –simplex method approach was applied to a complicated 5 factor precipitation and centrifugation sequence. Seeking a robust operating region for IgG_4 capture and purification from clarified cell culture, the method located ideal precipitation and centrifugation conditions at pH 7.5, 2 M ammonium sulphate, a feed IgG_4 concentration of 8 g/L, and 7 cycles of precipitate ageing by well mixing with an equivalent centrifuge flowrate of 65 L/Hr (giving a minimum recovery of 88 % and purity of 81 % within the operating range). Scale up was not carried out to verify this at lab scale and would be item for future work. The conditions would be easy to replicate in an impeller driven vessel for the precipitation and a centrifuge step including a shear step using ultrascale down principles (Boychyn et al., 2004). Some work was already carried out to correlate said shear device with the acoustic shear device used here based on shear conditions and equivalent particle size. For the ageing conditioning, equations from literature were used to correlate the cycles of jet mixing to shear values which could serve as the starting point of defining impeller mixing conditions in a reactor.

Where Protein A chromatography has usually been free from process development efforts there are cases which are becoming more frequent with newer biopharmaceuticals (more complicated possibly due to modern cell line development and upstream practices) where the generic Protein A protocol has not been adequate. While some have required tweaks to elution pH, others have required the development of bespoke washes to remove higher levels of HCPs and DNA (a negative effect of high cell density and high titre fermentation) which are co-eluting with the product (Shukla and Hinckley, 2008).

A review of the microscale chromatography methods available in chapter 5 led to the development of high throughput Protein A and CEX processes using atoll robocolumns and the Tecan platform. After method development it was decided robocolumn use would be limited to 3 cycles before results became unreliable, the absorbance reading methods on the plate reader were investigated and centrifugation of the plates was introduced to remove sources of error. The microscale methods used product residence time on the column to mimic the labscale process which was demonstrated for Protein A to be accurate in terms

of eluate yield and CVs. A screen of strong CEX options on a test IgG_4 product revealed just how different the results can be despite all the options being based on a strong cation SO_3 ligand. With the buffer conditions used, Ceramic Hyper D and Capto S were the best in terms of yield whilst the elution buffer was not strong enough to elute product from Fractogel or UNOSphere S. As each resin responds so differently to buffer conditions a potential future work would be to run a stepwise gradient elution of a range of elution conditions so screening experiments like these can indicate better which buffer is best for a step elution.

In chapter 6 the problem of removing antibody aggregates was investigated using UNOSphere S CEX chromatography on a number of different products. The proposed DoE-simplex methodology was applied to a 4 factor experiment and optimised loading and eluting buffer conditions whereas a comparison response surface method was unable to do so in the same number of samples. As the analytical step for aggregate determination was very slow (HPLC size exclusion chromatography), as well as the chromatography step itself taking, the time saving of the reduced samples numbers was evident with the simplex method. The combined approach allowed the design space to be viewed in breadth and in depth using a 2 level factorial DoE for scouting the surface and the simplex method to rapidly locate the global optimum for potential operating conditions. For the IgG_1 with the artificially high aggregates the simplex used a weighted response of yield and purity to find the optimum at pH_{eq} 5.25, C_{eq} 25mM, pH_{el} 6.25 and C_{el} 225mM (achieving a yield of 87 % and purity of 99.0 %). 3 other Lonza products were also optimised over the CEX step using the simplex method including an example looking at how the step's productivity is affected by loading capacity and flowrate and also aggregate removal by operating the step in flow through mode (targeting impurity/aggregate capture and allowing the product monomer to come out in the flow through).

The result chapters 3, 4 and 6 demonstrate how using the combined simplex method - DoE strategy is complementary of both optimisation techniques help in getting the most out of the simplex method in terms of selecting favourable initial conditions, simplex size, orientation. The highly opportunistic nature of the simplex method does mean a fortunately

placed simplex may result in a rapid identification of the global optimisation however the extra DoE screening goes a long way in getting the best out of it and replacing the element of luck for robustness, at the expense of a few extra samples.

The combined DoE-simplex methodology approach provides the best compromise between rapidly identifying the optimum operating conditions for an unknown bioprocess operation(s), and providing highly relevant process understanding. The least possible information is used at each stage and no account of past positions is kept. No assumptions are made about the surface except that it is continuous. A general problem shared by all optimisation techniques is that missing the global optimum and converging no a local optimum. DoE and the simplex is used here to avoid this issue. Some work to fully integrate the DoE-simplex method approach with automated high throughput platforms is still required to reap the greatest benefits for intelligent process design. For the methodology, it is also possible to create a final model from the simplex method data using methods such as delauney triangulation (as explored in two dimensions in chapter 3). This would provide considerable savings in time and experimental cost and as shown in Chapter 3 is quite accurate in describing local regions of the design space. The choice of initial simplex conditions could also be factored into setting up this final model. For the final DoE in the methodology we were already re-using conditions which the simplex method had evaluated.

High throughput microwell platforms make it feasible to perform high resolution study (such as the 385 condition precipitation example in chapter 3) and using under 1g of material for the PEG precipitation example in chapter 4 (Knevelman et al., 2009). The response surfaces provide in-depth understanding of the interactions between the factors and clearly illustrate the regions of high yield which may be taken for further characterisation. The newfound capabilities afforded by automated high throughput process design platforms do come at considerable capital expenditure however, and in an industry which demands reliable scale down models the presence of lower throughput lab systems will carry on for some time yet. And even if a high throughput platform is introduced in process development, it is unlikely that every element in the experiment is

just as fast and material-conservative such as most biochemical analytic methods. HPLC Protein A and SEC techniques which were used in chapters 3,4 and 6 each take 20 minutes per sample, consuming large amounts of buffer and exhaust expensive analytical columns. Assays for quantifying most impurities which as DNA and CHO HCP kits are also relatively labour intensive and time consuming, all of which would see a benefit in reducing experiment numbers. The time and cost of retrieving and analysing the data is also an unaccounted feature of processing large dataset.

To compute the mean velocity gradient and camp number with tip mixing is not so straightforward and as with other scale sown systems such as orbital mixing, a computational fluid dynamics (CFD) analysis would arguably be a better option. During process development for new biopharmaceuticals there are two economical drivers which counteract each other and result in a compromise to progress. The financial pressures to enter the market as quickly as the drug is proven in clinical trials and the development of the process to the extent that an optimum manufacturing process is used.
By this time, strong pressures to move into the marketplace may force companies to choose between either entering the market with an economically under-performing process or delaying product launch whilst improvements are carried out, thus risking loss of revenue. Conversely, by using only small amounts of feed, microscale methods facilitate the early stage evaluation of many process strategies in parallel, thus reducing development costs and allowing later pilot work to be more highly focused upon the most feasible option.

The overall objective of this research was achieved by the successful development of a DoE-simplex method process design approach and optimisation methodology for downstream bioprocess. This can significantly reduce the development time and cost in the early stage. The method which has been proposed uses three main components in an integrated optimisation framework: the initial DoE model, simplex method and then a local space DoE in the optimum area found by the simplex. These components worked with each other iteratively in a loop of receiving, analysing and passing useful process information. This combination approach was more effective to reduce material consumption in process design than the application of traditional DoE design and response surface methods such

as CCD. The difference was much more apparent in the more complicated nonlinear bioprocesses. This approach was faster and produced a more accurate process model due to the focused data points provided in the design.

References

Aldington S, Bonnerjea J. 2007. Scale-up of monoclonal antibody purification processes. J Chromatogra B. 848:64-78.

Andrews BA, Nielsen S, Asenjo JA. 1996. Partitioning and purification of monoclonal antibodies in aqueous two-phase systems. Bioseparation. 6:303-313.

Anderson M, Whitcomb P. 2005. RSM simplified-optimizing processes using response surface method for design of experiments. New York: Productivity Press. 293 p.

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.

Arakawa T, Timasheff SN. 1985. Theory of protein solubility. Meth Enzymol. 114:49-77.

Arunakumari A, Wang J. 2008. Purification of human monoclonal antibodies: non-protein A strategies. In: Gottschalk U, editor. Process scale purification of antibodies. Hoboken, NJ: John Wiley & Sons. 103-124.

Atha DH, Ingham KC. 1981. Mechanism of precipitation of proteins by polyethylene glycols. J Biol Chem. 256:12108-12117.

Atkinson AC, Hunter WG. 1968. The design of experiments for parameter estimation. Technometrics. 10:271-289.

Atkinson AC. 1996. The usefulness of optimum experimental designs. J Roy Stat Soc B. 58:59-76.

Atkinson AC, Tobias RD. 2008. Optimal experimental design in chromatography. J Chromatogra A. 1177:1-11.

Aucamp JP, Cosme AM, Lye GJ, Dalby PA. 2005. High-throughput measurement of protein stability in microtiter plates. Biotechnol Bioeng. 89:599-607.

Banerjee R, Bhattacharyya BC. 1993. Evolutionary operation (EVOP) to optimize three-dimensional biological experiments. Biotechnology and bioengineering. 41:67-71.

Bensch M, Wierling PS, Lieres E, Hubbuch J. 2005. High throughput screening of chromatographic phases for rapid process development. Chem Eng Technol. 28:1274-1284.

Bergander T, Nilsson-Valimaa K, Oberg K, Lacki KM. 2008. High-throughput process development: determination of dynamic binding capacity using microtiter filter plates filled with chromatography resin. Biotechnol Progr. 24:632-639.

Biorad Laboratories. 2000. UNOsphere™ S ion exchange media instruction manual. Document No.: 4110109 Rev B.

Birch JR, Racher AJ. 2006. Antibody production. Adv Drug Deliver Rev. 58:671-685.

Box GEP, Lucas HL. 1959. Design of experiments in non-linear situations. Biometrika. 46:77-90.

Box GEP, Wilson KB. 1951. On the experimental attainment of optimum conditions. J Roy Stat Soc. 13:1-45.

Box MJ, Draper NR. 1971. Factorial designs, the |XX'| criterion and some related matters. Technometrics. 13:731-742.

Bradford M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of dye-binding. Anal Biochem. 72: 248-254.

Carrier T, Heldin E, Ahnfelt M, Brekkan E, Hassett R, Peppers S, Rodrigo G, Slyke G, Zhao D. 2010. High-throughput technologies in bioprocess development. In: Flickinger MC, editor. Encyclopedia of industrial biotechnology: bioprocess, bioseparation and cell technology. New York: John Wiley & Sons. 1-75.

Carta G, Jungbauer A. 2010. Protein chromatography: process development and scale-up. Weinheim: Wiley-VCH. 364.

Cawse JN. 2001. Experimental strategies for combinatorial and high-throughput materials development. Acc Chem Res. 34:213-221.

Challener CA. 2016. Modeling Bioreactor Performance. BIOPHARM INTERNATIONAL. Sep 1;29(9):52-4.

Chernoff H. 1959. Sequential design of experiments. Ann Math Stat. 30:755-770.

Chhatre S, Bracewell DG, Titchener-Hooker NJ. 2009. A microscale approach for predicting the performance of chromatography columns used to recover therapeutic polyclonal antibodies. J Chromatogra A. 1216:7806-7815.

Chhatre S, Titchener-Hooker NJ. 2009. Review: Microscale methods for high-throughput chromatography development in the pharmaceutical industry. J Chem Technol Biot. 84: 927-940.

Chhatre S, Farid SS, Coffman J, Bird P, Newcombe AR, Titchener-Hooker NJ. 2011. How implementation of Quality by Design and advances in biochemical engineering are enabling efficient bioprocess development and manufacture. J Chem Technol Biot. 86:1125-1129.

Chhatre S, Konstantinidis S, Ji Y, Edwards-Parton S, Zhou Y, Titchener-Hooker NJ. 2011. The Simplex Algorithm for the Rapid Identification of Operating Conditions During Early Bioprocess Development: Case Studies in FAb' Precipitation and Multimodal Chromatography. Biotechnol Bioeng. 108:2162-2170.

Chiew YC, Kuehner DE, Blanch HW, Prausnitz JM. 1995. Molecular thermodynamics for salt-induce protein precipitation. AIChE J. 41:2150-2159.

Coffman JL, Kramarczyk JF, Kelley B. 2008. High-throughput screening of chromatographic separations: I. Method development and column modelling. Biotechnol Bioeng. 100:605-618.

Cohn EJ. 1925. The physical chemistry of the proteins. Physiol Rev. 5:349-437.

Cohn EJ, McMeekin TL, Oncley JL, Newell JM, Hughes WL. 1940. Preparation and properties of serum and plasma proteins. I. size and charge of proteins separating upon equilibration across membranes with ammonium sulphate solutions of controlled pH, ionic strength and temperature. J Am Chem Soc. 62:3386-3393.

Cook RD, Bachtsheim CJ. 1980. A comparison of algorithms for construction exact D-optimal designs. Technometrics. 22:315-324.

EMA. 2000. Development pharmaceutics for biotechnological and biological products - annex to note for guidance on development pharmaceutics.

Faherner RL, Knudsen HL, Basey CD, Galan W, Feuerhelm D, Vanderlaan M. 2001. Industrial purification of pharmaceutical antibodies: Development, operation, and validation of chromatography processes. Biotechnol Genet Eng Rev 18:301–327

Fang, Z. 2010. Large-scale chromatography columns, modeling flow distribution. Encyclopedia of Industrial Biotechnology: Bioprocess, Bioseparation and Cell Technology. 1-18.

Farid SS, Washbrook J, Titchener-Hooker NJ. 2006. Modelling biopharmaceutical manufacture: design and implementation of SimBiopharma. Comput Chem Eng. 31:1141-1158.

Farid SS. 2007. Process economics of industrial monoclonal antibody manufacture. Journal of Chromatography B. 848:8-18.

Farid SS. 2008. Process economic drivers in industrial monoclonal antibody manufacture. In: Gottschalk U, editor. Process scale purification of antibodies. Hoboken, NJ: John Wiley & Sons. 239-262.

FDA. 1995a. Guidance: changes to be reported for product and establishment license applications.

FDA. 1995b. Guidance document concerning use of pilot manufacturing facilities for the development and manufacture of biological products.

FDA. 1996. Guidance: demonstration of comparability of human biological products, including therapeutic biotechnology-derived products.

FDA. 2001. Guidance for industry: monoclonal antibodies used as reagents in drug manufacturing.

FDA. 2004. Guidance for industry: process analytical technology, a framework for innovative pharmaceutical development, manufacturing and quality assurance.

FDA. 2006. Guidance for industry: Q8 pharmaceutical development.

FDA. 2013. Biologics Guidance 21 CFR 10.115(b).

Follman DK, Fahrner RL. 2004. Factorial screening of antibody purification processes using three chromatography steps without protein A. J Chromatogr A. 1024:79-85.

Foster PR, Dunnill P, Lill MD. 1976. The kinetics of protein salting-out: precipitation of yeast enzymes by ammonium sulphate. Biotechnol Bioeng. 18:545-580.

85. Gagnon P. 2007. Polishing Methods for Monoclonal IgG Purification. In: Shukla AA, editor. Process scale bioseparations for the biopharmaceutical industry. Boca Raton, FL: Taylor & Francis Group. 491-506.

GE Healthcare. 2006a. Capto S cation exchanger for post-Protein A purification of monoclonal antibodies. Application Note: 28-4078-17 AA.

GE Healthcare. 2006b. Process-scale purification of monoclonal antibodies - polishing using Capto Q. Application Note: 28-9037-16 AA.

GE Healthcare. 2006c. Sreening and optimization of the loading conditions on Capto S. Application Note: 28-4078-16 AA.

GE Healthcare. 2009a. Rapid screening of a scalable, intermediate purification step for recombinant EGFP with HiTrap HIC Selection Kit. Application note: 28-9192-12 AB.

GE Healthcare. 2009b. Scale-up of a downstream monoclonal antibody purification process using HiScreen™ and AxiChrom™ columns. Application note: 28-9403-49 AA.

GE Healthcare. 2011a. MabSelect. Instructions: 71-5020-91 AE.

GE Healthcare. 2011b. Superdex 200 10/300 GL. Instructions: 71-5017-96 AG.

George A, Wilson WW. 1994. Predicting protein crystallization from a dilute solution property. Acta Cryst. 50:361-365.

Ghosh R, Wang L. 2006 Purification of humanized monoclonal antibody by hydrophobic interaction membrane chromatography. J Chromatogr A. 1107:104-109.

Ghose S, McNerney T, Hubbard B. 2007. Protein A affinity chromatography for capture and purification of monoclonal antibodies and Fc-Fusion proteins: practical considerations for process development. In: Shukla AA, editor. Process scale bioseparations for the biopharmaceutical industry. Boca Raton, FL: Taylor & Francis Group. 463-490.

Ghose S, Jin M, Liu J, Hickey J. 2008. Integrated polishing steps for monoclonal antibody purification. In: Gottschalk U, editor. Process scale purification of antibodies. Hoboken, NJ: John Wiley & Sons. p 145-168.

Giese G. Polyethylene glycol precipitation of monoclonal antibodies and the impact on column chromatography. In: BioProcess Int. Raleigh, NC; 2009.

Glatz CE, Hoare M, Landa-Vertiz J. 1986. The formation and growth of protein precipitates in a continuous stirred-tank reactor. AIChE J. 32:1197-1204.

Glynn J. 2008. Process-scale precipitation of impurities in mammalian cell culture broth. In: Gottschalk U, editor. Process scale purification of antibodies. Hoboken, NJ: John Wiley & Sons. 309-323

Glynn J, Hagerty T, Pabst T, Annathur G, Thomas K, Johnson P, Ramasubramanyan N, Mensah P. 2009. The development and application of a monoclonal antibody purification platform. BioPharm Intl Suppl. Oct.

Goldberg DE. 1989. Genetic algorithms in search, optimization and machine learning. Boston: Addison-Wesley. 432 p.

Gottschalk U. 2003. Biotech manufacturing is coming of age. Bioprocess Int. Apr, 1-7.

Gottschalk U. 2008. Bioseparation in antibody manufacturing: the good, the bad and the ugly. Biotechnol Progr. 24:496-503.

Grabenbauer GC, Glatz CE. 1981. Protein precipitation analysis of particle size distribution and kinetics. Chem Eng Commun. 12:203-219

Hertzberg RP, Pope AJ. 2000. Review: High-throughput screening: new technology for the 21st century. Curr Opin Chem Biol. 4:445-451.

Hober S, Nord K and Linhult M. 2007. Protein A chromatography for antibody purification. J Chromatogr B. 848:40-47.

125.ICH. 2008. Guideline Q10: pharmaceutical quality system.

Ingham KC. 1990. Precipitation of proteins with polyethylene glycol. Method Enzymol. 182:301-306.

Ishihara T, Yamamoto S. 2005. Optimization of monoclonal antibody purification by ion-exchange chromatography: Application of simple methods with linear gradient elution experimental data. J Chromatogra A. 1069:99-106.

Kelley B, Switzer M, Bastek P, Kramarczyk JF, Molnar K, Yu T. 2008. High-throughput screening of chromatographic separations: IV. Ion-exchange. Biotechnol Bioeng. 100:950-963.

Kelley B, Blank G, Lee A. 2009. Downstream processing of monoclonal antibodies: current practices and future opportunities. In: Gottschalk U, editor. Process scale purification of antibodies. New York: John Wiley & Sons. 1-24.

Kensy, Frank, Engelbrecht, Christoph and Büchs, Jochen (2009) Scale-up from microtiter plate to laboratory fermenter: evaluation by online monitoring techniques of growth and protein expression in Escherichia coli and Hansenula polymorpha fermentations. Microbial Cell Factories 2009 8:68

Knevelman C, Davies J, Allen L, Titchener-Hooker NJ. 2010. High-throughput screening techniques for rapid PEG-based precipitation of IgG4 mAb from clarified cell culture supernatant. Biotechnol Progr. 26:679-705.

Konstantinidis S, Chhatre S, Velayudhan A, Heldin E, Titchener-Hooker NJ, 2012. The hybrid experimental simplex algorithm – An alternative method for ‘sweet spot’ identification in early bioprocess development: Case studies in ion exchange chromatography. Analytica Chimica Acta , 743 19 - 32

Konstantinidis S, Welsh JP, Roush DJ, Velayudhan A. 2016. Application of simplex-based experimental optimization to challenging bioprocess development problems: Case studies in downstream processing. Biotechnol Progress, 32: 404–419.

Konstantinidis S, Titchener-Hooker N, Velayudhan A. 2017. Simplex-based optimization of numerical and categorical inputs in early bioprocess development: Case studies in HT chromatography. Biotechnol. J., 12: n/a, 1700174.

Kramarczyk JF, Kelley BD, Coffman JL. 2008. High-throughput screening of chromatographic separations: II. Hydrophobic interaction. Biotechnol Bioeng. 100:707-720.

Kuczewski M, Schirmer E, Lain B, Zarbis-Papastoitsis G. 2010. PEG precipitation: a powerful tool for monoclonal antibody purification. Biopharm Intl Suppl. Mar, 11-18.

Li F, Zhou JX, Yang X, Tressel T, Lee B. 2005. Current Therapeutic Antibody Production and Process Optimization. BioProcessing J. 5:16-25.

Lintern K, Pathak M, Smales CM, Howland K, Rathore A, Bracewell DG. 2016. Residual on column host cell protein analysis during lifetime studies of protein A chromatography. Journal of Chromatography A. 1461: 70-77.

Long FA, McDevit WF. 1952. Activity coefficients of nonelectrolyte solutes in aqueous salt solutions. Chem Rev. 51:119-157.

Low D, O'Leary R, Pujar NS. 2007. Future of antibody purification. J. Chromatogr B. 848:48-63.

Lowe CR, Lowe AR, Gupta G. 2001. New developments in affinity chromatography with potential application in the production of biopharmaceuticals. J Biochem Biophys Methods. 49:561-574.

Lye GJ, Ayazi-shamlou P, Baganz F, Dalby PA, Woodley JM. 2003. Accelerated design of bioconversion processes using automated microscale processing techniques. Trends Biotechnol. 21:29-37.

Lye GJ, Hubbuch J, Schroeder T, Willimann E. 2009. Shrinking the costs of bioprocess development. Bioprocess Int Suppl. 10:18-22.

Mao LN, Rogers JK, Westoby M, Conley L, Pieracci J. 2010. Downstream antibody purification using aqueous two-phase extraction. Biotechnol Progr. 26:1662-1670.

Massimo CD, Lant PA, Saunders A, Montague GA, Tham MT, Morris AJ. Bioprocess applications of model-based estimation techniques. 1992. J Chem Tech Biotechnol. 53:265-277.

McDonald P, Tran B, Williams CR, Wong M, Zhao T, Kelley BD, Lester P. 2016. The rapid identification of elution conditions for therapeutic antibodies from cation-exchange chromatography resins using high-throughput screening. Journal of Chromatography A. 1433:66-74.

Melander W, Horvath C. 1977. Salt effects on hydrophobic interactions in precipitation and chromatography of proteins: an interpretation of the lyotropic series. Arch Biochem Biophys. 183:200-215.

Micheletti M, Barrett T, Doig SD, Baganz F, Levy MS, Woodley JM, Lye GJ. 2006. Fluid mixing in shaken bioreactors: implications for scale-up predictions from microliter-scale microbial and mammalian cell cultures. Chem Eng Sci. 61:2939-2949.

Micheletti M, Lye GJ. 2006. Microscale bioprocess optimization. Curr Opin Biotech. 17:611-618.

Morgan E, Burton KW, Nickless G. 1990. Optimization using the modified simplex method. Chemometr Intell Lab. 7:209-222.

Natarajan, V. and Zydney, A. L. (2013), Protein a chromatography at high titers. Biotechnol. Bioeng., 110: 2445–2451. doi:10.1002/bit.24902

Nelder JA, Mead R. 1965. A simplex method for function minimization. Comput J 7:308-313.

Nfor BK, Ahamed T, Dedem G, Wielen L, Sandt E, Eppink M, Ottens M. 2008. Design strategies for integrated protein purification processes: challenges, progress and outlook. J Chem Technol Biot. 83:124-132.

Nfor BK, Verhaert PD, Wielen LA, Hubbuch J, Ottens M. 2009. Rational and systematic protein purification process development: the next generation. Trends Biotechnol. 27:673-679.

Niederauer MQ, Glatz CE. 1992. Selective precipitation. Adv Biochem Eng Biot. 47:159-188.

Peram T, McDonald P, Carter-Franklin J, Fahrner R. 2010. Monoclonal antibody purification using cationic polyelectrolytes: an alternative to column chromatography. Biotechnol Progr. 26:1322-1331.

Przybycien TM, Pujar NS, Steele LM. 2004. Alternative bioseparation operations: life beyond packed-bed chromatography. Current Opinion in Biotechnology. 15:469-478.

Queiroz JA, Tomaz CT, Cabral JMS. 2001. Hydrophobic interaction chromatography of proteins. J Biotechnol. 87:143-159.

Rao SR, Kumar CG, Prakasham RS, Hobbs PJ. 2008. The Taguchi methodology as a statistical tool for biotechnological applications: a critical appraisal.
Biotechnol J. 3:510-523.

Rathore AS. 2009. Roadmap for implementation of quality by design (QbD) for biotechnology products. Trends Biotechnol. 27:546-553.

Rea, J., Moreno, G., Lou, Y., Farnan, D. (2011), Validation of a pH gradient-based ion-exchange chromatography method for high-resolution monoclonal antibody charge variant separations. Journal of Pharmaceutical and Biomedical Analysis, 54 (2): 317-323

Rege K, Pepsin M, Falcon B, Steele L, Heng M. 2006. High-throughput process development for recombinant protein purification. Biotech Bioeng. 93:618-630.

Rito-Palomares M. 2008. Bioseparation: the limiting step in bioprocess development. J Chem Technol Biot. 83:115-116.

Roque A, Silva C, Taipa MA. 2007. Affinity-based methodologies and ligands for antibody purification: advances and perspectives. J Chromatogra A. 1160:44-55.

Rosa P, Azevedo AM, Aires-Barros MR. 2007. Application of central composite design to the optimisation of aqueous two-phase extraction of human antibodies. J Chromatogr A. 1141:50-60.

Suda EJ, Thomas KE, Pabst TM, Mensah P, Ramasubramanyan N, Gustafson ME. 2009. Comparison of agarose and dextran-grafted agarose strong ion exchangers for the separation of protein aggregates. J Chromatogr A 1216(27):5256–5264

Saha G, Barua A and Sinha S. Dynamic optimization of a bioprocess using genetic algorithm.2015 Bioreactors, 31 -52

Shah RB, Park JT, Read EK, Khan MA, Brorson K. 2010. Quality by Design (QbD), Biopharmaceutical manufacture. In: Flickinger MC, editor. Encyclopedia of industrial biotechnology: bioprocess, bioseparation and cell Technology. New York: Wiley.1-36.

Shaughnessy A.F. Monoclonal antibodies: magic bullets with a hefty price tag. BMJ. 2012;345:8346.

Shih YC, Prausnitz JM. 1992. Some characteristics of protein precipitation by salts. Biotechnol Bioeng. 40:1155-1164.

Shukla AA, Hinckley PJ, Gupta P, Yigzaw Y, Hubbard B. 2005. Strategies to address aggregation during Protein-A chromatography. BioProcess Int. 3:36-45.

Shukla AA, Hubbard B, Tressel T, Guhan S, Low D. 2007. Downstream processing of monoclonal antibodies-Application of platform approaches. J Chromatogra B. 848:28-39.

Shukla AA and Hinckley P. 2008. Host cell protein clearance during protein a chromatography: Development of an improved column wash step. Biotechnol Progress, 24: 1115–1121

Shukla AA, Thommes J. 2010. Recent advances in large-scale production of monoclonal antibodies and related proteins. Trends Biotechnol. 28:253-261.

Smith TS, Wilson E, Scott RG, Misczak JW, Bodek JM, Zabriskie DW. 1998. Establishment of operating ranges in a purification process for a monoclonal antibody. In: Kelley B, Ramelmerier RA, editor. Validation of biopharmaceutical manufacturing processes. Milwaukee, WI: Quality press. 80-92.

Sommerfeld S, Strube J. 2005. Challenges in biotechnology production – generic processes and process optimization for monoclonal antibodies. Chem Eng Process. 44:1123-1137.

Stavrinides S, Ayazi-shamlou P, Hoare M. 1993. Effects of engineering parameters on the precipitation recovery and purification of proteins. In: Ayazi-Shamlou P, editor. Processing of solid-liquid suspensions. London: Butterworth Heinemann. 118-158.

Stein A, Kiesewetter A. 2007. Cation exchange chromatography in antibody purification: pH screening for optimised binding and HCP removal. J Chromatogr B. 848:151-158.

Steinmeyer DE, McCormick EL. 2008. The art of antibody process development. Drug Discov Today. 13:613-618.

Tarrant RDR, Velez-Suberbie ML, Tait A. 2012. Host Cell Protein Adsorption Characteristics During Protein A Chromatography. Biotechnol. Progr. 2:1037–1044.

Titchener-Hooker NJ, Harri S, Davies E, Zhou Y, Hoare M, Dunnill P. 2001. Biopharmaceutical process development Part 2, Methods of reducing development time. Biopharm Europe. September, 68-74.

Titchener-Hooker NJ, Dunnill P, Hoare M. 2008. Micro biochemical engineering to accelerate the design of industrial-scale downstream processes for biopharmaceutical proteins. Biotechnol Bioeng. 100:473-487.

Thommes J, Etzel M. 2007. Alternatives to chromatographic separations. Biotechnol Prog. 23:42-45.

Torres-Bacete J, Arroyo M, Torres-Guzman R, De La Mata I, Acebal C, Castillon MP. 2005. Optimization of culture medium and conditions for penicillin acylase production by streptomyces lavendulae ATCC 13664. Appl

Biochem Biotech. 126:119-131.

Tugcu N, Roush DJ, Goklen KE. 2008. Maximizing productivity of chromatography steps for purification of monoclonal antibodies. Biotechnol

Bioeng. 99:599-613.

Wurm FM. 2004. Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol. 22:1393-1398.

Yu LX. 2007. Pharmaceutical quality by design: product and process development, understanding and control. Pharmacol Res. 25:781-791.

Ziegel ER. 2003. Experimental design for combinatorial and high throughput materials development. Technometrics. 45:365-365.

Zhou YH, Titchener-Hooker NJ. 1996. The use of windows of operation as a bioprocess design tool. Bioprocess Eng.14:263-268.

Appendix 1: Simplex code

The simplex code used in Matlab:

```
% Initialize parameters
rho = 1; chi = 2; psi = 0.5; sigma = 0.5;
onesn = ones(1,n);
two2np1 = 2:n+1;
one2n = 1:n;

% Set up a simplex near the initial guess.
xin = x(:); % Force xin to be a column vector
v = zeros(n,n+1); fv = zeros(1,n+1);
v(:,1) = xin;    % Place input guess in the simplex! (credit L.Pfeffer
at Stanford)
x(:) = xin;    % Change x to the form expected by funfcn
fv(:,1) = funfcn(x,varargin{:});
func_evals = 1;
itercount = 0;
how = '';
% Initial simplex setup continues later

% sort so v(1,:) has the lowest function value
[fv,j] = sort(fv);
v = v(:,j);

exitflag = 1;

% Main algorithm
% Iterate until the diameter of the simplex is less than tolx
%   AND the function values differ from the min by less than tolf,
%   or the max function evaluations are exceeded. (Cannot use OR
instead of
%   AND.)
while func_evals < maxfun && itercount < maxiter
    if max(abs(fv(1)-fv(two2np1))) <= tolf && ...
            max(max(abs(v(:,two2np1)-v(:,onesn)))) <= tolx
        break
    end

    % Compute the reflection point

    % xbar = average of the n (NOT n+1) best points
    xbar = sum(v(:,one2n), 2)/n;
    xr = (1 + rho)*xbar - rho*v(:,end);
    x(:) = xr; fxr = funfcn(x,varargin{:});
    func_evals = func_evals+1;

    if fxr < fv(:,1)
        % Calculate the expansion point
        xe = (1 + rho*chi)*xbar - rho*chi*v(:,end);
        x(:) = xe; fxe = funfcn(x,varargin{:});
```

```
        func_evals = func_evals+1;
        if fxe < fxr
            v(:,end) = xe;
            fv(:,end) = fxe;
            how = 'expand';
        else
            v(:,end) = xr;
            fv(:,end) = fxr;
            how = 'reflect';
        end
    else % fv(:,1) <= fxr
        if fxr < fv(:,n)
            v(:,end) = xr;
            fv(:,end) = fxr;
            how = 'reflect';
        else % fxr >= fv(:,n)
            % Perform contraction
            if fxr < fv(:,end)
                % Perform an outside contraction
                xc = (1 + psi*rho)*xbar - psi*rho*v(:,end);
                x(:) = xc; fxc = funfcn(x,varargin{:});
                func_evals = func_evals+1;

                if fxc <= fxr
                    v(:,end) = xc;
                    fv(:,end) = fxc;
                    how = 'contract outside';
                else
                    % perform a shrink
                    how = 'shrink';
                end
            else
                % Perform an inside contraction
                xcc = (1-psi)*xbar + psi*v(:,end);
                x(:) = xcc; fxcc = funfcn(x,varargin{:});
                func_evals = func_evals+1;

                if fxcc < fv(:,end)
                    v(:,end) = xcc;
                    fv(:,end) = fxcc;
                    how = 'contract inside';
                else
                    % perform a shrink
                    how = 'shrink';
                end
            end
            if strcmp(how,'shrink')
                for j=two2np1
                    v(:,j)=v(:,1)+sigma*(v(:,j) - v(:,1));
                    x(:) = v(:,j); fv(:,j) = funfcn(x,varargin{:});
                end
                func_evals = func_evals + n;
            end
        end
    end
    [fv,j] = sort(fv);
```

```
    v = v(:,j);
    itercount = itercount + 1;
    if prnt == 3
        disp(sprintf(' %5.0f        %5.0f     %12.6g         %s',
itercount, func_evals, fv(1), how))
    elseif prnt == 4
        disp(' ')
        disp(how)
        v
        fv
        func_evals
    end
    % OutputFcn call
    if haveoutputfcn
        [xOutputfcn, optimValues, stop] =
callOutputFcn(outputfcn,x,xOutputfcn,'iter',itercount, ...
            func_evals, how, fv(:,1),varargin{:});
        if stop  % Stop per user request.
            [x,fval,exitflag,output] =
cleanUpInterrupt(xOutputfcn,optimValues);
            if  prnt > 0
                disp(output.message)
            end
            return;
        end
    end
end   % while
```

Appendix 2: Yield results of robocolumn screens

The results of the resin screens conducted in chapter 5 for method development on the Tecan LHR.

Protein A resin	Yield (%)
MabSelect	77
MabSelect SuRe	94
MabSelect Xtra	115
Protein A Sepharose	92
ProSep-vA HC	70
ProSep-vA Ultra	89
Ceramic Hyper D	38
POROS 50A	91

Yield data from Protein A robocolumns screen

CEX Resin	Yield (%)
POROS 50HS	15
S-Ceramic Hyper D	117
ToyoPearl MegaCap II SP-550	76
Fractoprep SO3	29
Fractogel EMD SO3	10
UnoSphere S	17
SP Sepharose	70
Capto S	119

Yield results from CEX screen using the Atoll S-kit robocolumns