Model based process design for bioprocess optimisation: case studies on precipitation with its applications in antibody purification

A thesis submitted to the University College London

for the degree of

DOCTOR OF PHILOSOPHY

Yu Ji

2012

The Advanced Centre for Biochemical Engineering

Department of Biochemical Engineering

University College London

London UK

Declaration

I, Yu Ji 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:

To my loving parents

Acknoledgements

I would like to thank my supervisor Dr. Yuhong Zhou and my advisor Prof. Nigel Titchener-Hooker for their support and guidance throughout my three years PhD studies. Thanks to all colleagues in the department of Biochemical Engineering for supports and helps.

The financial support provided by UCL Overseas Research Student Scholarship (ORS) and UCL Graduate School Research Student Scholarship (GSRS) are gratefully acknowledged.

To my parents, thank you very much for unwavering support and love.

Abstract

Developing a bioprocess model can not only reduce cost and time in process development, but now also assist the routine manufacturing and guarantee the quality of the final products through Quality by Design (QbD) and Process Analytical Technology (PAT). However, these activities require a model based process design to efficiently direct, identify and execute optimal experiments for the best bioprocess understanding and optimisation. Thus an integrated model based process design methodology is desirable to significantly accelerate bioprocess development. This will help meet current urgent clinical demands and also lower the cost and time required. This thesis examines the feasibility of a model based process design for bioprocess optimisation. A new process design approach has been proposed to achieve such optimal design solutions quickly, and provide an accurate process model to speed up process understanding.

The model based process design approach includes bioprocess modelling, model based experimental design and high throughput microwell experimentation. The bioprocess design is based on experimental data and a computational framework with optimisation algorithm. Innovative model based experimental design is a core part in this approach. Directed by the design objectives, the method uses D-optimal design to identify the most information rich experiments. It also employs Random design and Simplex to identify extra experiments to increase the accuracy, and will iteratively improve the process design solutions.

The modelling and implementation method by high throughput experimentation was first achieved and applied to an antibody fragment (Fab') precipitation case study. A new precipitation model based on phase equilibrium has been developed using the data from microwell experimentation, which was further validated by statistical tests to provide high confidence. The precipitation model based on good data accurately

describes not only the Fab' solubility but also the solubility of impurities treated as a pseudo-single protein, whilst changing two critical process conditions: salt concentration and pH. The comparison study has shown the model was superior to other published models. The new precipitation model and the Fab' microwell data provided the basis to test the efficiency and robustness of the algorithms in model based process design approach. The optimal design solution with the maximum objective value was found by only 5 iterations (24 designed experimental points). Two parameterised models were obtained in the end of the optimisation, which gave a quantitative understanding of the processes involved. The benefit of this approach was well demonstrated by comparing it with the traditional design of experiments (DoE).

The whole model based process design methodology was then applied to the second case study: a monoclonal antibody (mAb) precipitation process. The precipitation model was modified according to experimental results following modelling procedures. The optimal precipitation conditions were successfully found through only 4 iterations, which led to an alternative process design to protein A chromatography in the general mAb purification platform. The optimal precipitation conditions were then investigated at lab scale by incorporating a depth filtration process. The final precipitation based separation process achieved 93.6% (w/w) mAb yield and 98.2 % (w/w) purity, which was comparable to protein A chromatography.

Polishing steps after precipitation were investigated in microwell chromatographic experimentation to rapidly select the following chromatography steps and facilitate the whole mAb purification process design. The data generated were also used to evaluate the process cost through process simulations. Both precipitation based and protein A chromatography based processes were analysed by the process model in the commercial software BioSolve under several relevant titre and scale assumptions. The results showed the designed precipitation based processes was superior in terms of process time and cost when facing future process challenges.

Table of Contents

List of Figures

List of Tables

Nomenclature and abbreviation

C_d	protein molar concentration in the dense phase ($\mathrm{mol \cdot L^{-1}}$)
C_i	other component molar concentration in the solution ($\mathrm{mol \cdot L^{-1}}$)
C_l	protein molar concentration in the light phase ($\mathrm{mol \cdot L^{-1}}$)
C_s	salt molar concentration ($\mathrm{mol \cdot L^{-1}}$)
$C_{l\,Fab'}$	Fab' concentration in the supernatant ($\mathrm{g \cdot L^{-1}}$)
$C_{l\,\mathrm{impu}}$	impurity concentration in the supernatant ($\mathrm{g \cdot L^{-1}}$)
$C_{l\,mAb}$	mAb concentration in the supernatant ($\mathrm{g \cdot L^{-1}}$)
C_T	maximum protein concentration in the solution ($\mathrm{g \cdot L^{-1}}$)
$C_{T\,Fab'}$	maximum Fab' concentration in the supernatant ($\mathrm{g \cdot L^{-1}}$)
$C_{T\,\mathrm{impu}}$	maximum impurity concentration in the supernatant ($\mathrm{g \cdot L^{-1}}$)
$C_{\mathrm{T}\,mAb}$	maximum mAb concentration in the supernatant ($\mathrm{g \cdot L^{-1}}$)
I	ionic strength ($\mathrm{mol \cdot L^{-1}}$)
k_i	component activity coefficient
k_s	salt activity coefficient
k	kinetic rate ($\mathrm{mol \cdot L^{-1} \cdot s^{-1}}$)
m_3	salt molar concentration ($\mathrm{mol \cdot L^{-1}}$)
R	ideal gas constant ($\mathrm{J \cdot mol^{-1} \cdot K^{-1}}$)
r_d	protein activity coefficient in the light phase
r_l	protein activity coefficient in the dense phase
T	temperature (K)

V_d	dense phase volume (L)
V_l	light liquid phase volume (L)
V_T	total solution volume (L)
w_1, w_2, w_3	constants (equation 4.11)
x	a $n \times m$ matrix of conditions (equation 3.1)
y	a $n \times 1$ matrix of observations (equation 3.1)
y_{exp}	experimental value (equation 2.3)
$\overline{y}_{\text{exp}}$	average of all experimental values (equation 2.4)
y_{model}	model value (equation 2.3)
Z	net charge of the protein
Z_i	charge number of that ion

Greek Symbols

β	k coefficients to be estimated (equation 3.1)
β_0	initial parameter (equation 3.2)
$\hat{\beta}$	the best estimated parameters (equation 3.7)
λ_L	real value at low limit (equation 2.1)
λ_{real}	real value (equation 2.1)
λ_{scaled}	scaled value (equation 2.1)
λ_U	real value at upper limit (equation 2.1)
μ_2°	protein standard chemical potential in the solution ($J \cdot mol^{-1}$)

$\mu_{2,w}^{\circ}$	protein standard chemical potential in the water ($J \cdot mol^{-1}$)
μ_d	chemical potential for dense phase ($J \cdot mol^{-1}$)
μ_d°	protein standard chemical potential in the dense phase ($J \cdot mol^{-1}$)
μ_l	chemical potential for liquid phase ($J \cdot mol^{-1}$)
μ_l°	protein standard chemical potential in the light phase ($J \cdot mol^{-1}$)
σ	standard deviation of the experimental error
$\gamma,\eta,\kappa,\phi,\xi,\rho,\nu,\varphi$	lumped coefficients

Abbreviation

AEX	Anion EXchange chromatography
ANN	Artificial Neural Network
CCD	Central Composite Design
CD	Circular Dichroism
CEX	Cation EXchange chromatography
CFD	Computational Fluid Dynamics
CHO	Chinese Hamster Ovary cell
COG	Cost of Goods
CPP	Critical Process Parameters
CQA	Critical Quality Attributes
DBC	Dynamic Binding Capacity
DoE	Design of Experiments
EMA	European Medicines Agency
Fab’	Fragment Antibody
FDA	Food and Drug Administration
GMP	Good Manufacturing Practice
HCP	Host Cell Protein

HIC	Hydrophobic Interaction Chromatography
HPTFF	High-Performance Tangential Flow Filtration
HTE	High Throughput Experimentation
HTS	High Throughput Screening
ICH	The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use
IEF	IsoElectric Focusing
IEX	Ion EXchange chromatography
mAb	monoclonal Antibody
PAT	Process Analytical Technology
QbD	Quality by Design
RSM	Response Surface Methodology
SD	Standard Deviation
UNIQUAC	UNIversal QUAsiChemical

Chapter 1. Introduction

1.1 Perspective and motivation

In the past three decades, the world has witnessed rapid growth and significant transformation in biotechnology, not only in biological areas involving genomics, cell and protein engineering, but also in the engineering field for bioprocess manufacturing, such as large-scale fermentation and downstream optimisation (Titchener-Hooker et al., 2008). A considerable number of candidate therapeutic proteins promising lucrative returns for biotech companies, increases the need for successful large-scale protein production. Meanwhile, the pressures for new low-cost and faster pharmaceutical production also arise from various other aspects, such as unsatisfied market needs, growing competition between companies and economic constrains of healthcare systems worldwide (Gottschalk, 2003). These needs inevitably place great pressure on typically expensive and time-consuming bioprocess development, mainly focusing on upstream fermentation and downstream purification.

With the successful scaling-up of bioreactor size up to 20,000L (Farid, 2007) and dramatical protein concentration improvement in cell culture broth from 5-50 mg/L to in excess of 5 g/L nowadays, downstream purification bioprocesses are being gradually identified as the crucial limiting step in biopharmaceutical development (Aldington and Bonnerjea, 2007; Birch and Racher, 2006; Rito-Palomares, 2008; Titchener-Hooker et al., 2001). Since the technological advances in downstream have failed to keep up with upstream productivity increase, 50-80% of the total manufacturing cost for one biotherapeutical product is from purification and polishing processes (Lowe et al., 2001). As a result, many manufacturers and researchers are exploring different technologies and platforms to reduce cost by such methods as decreasing the number of purification steps.

However, for a potential biotherapeutical product, the downstream bioprocess development is not only costly considering the capital investment and manpower, but also time limited. The process development is normally set up in pre-clinical trails and needs scaling-up, optimising and successfully transferring to multi-scale commercial manufacturing facilities under Good Manufacturing Practice (GMP) before clinical trial phase 3 and authority inspection can be implemented (Nfor et al., 2009). Once the manufacturing process has been reported to a regulatory body in a clinical trial dossier or market authorisation application, it is very difficult and complicated to change any operation and specification later unless there are abundant evidences to convince the regulatory authority that the safety, quality and efficacy of the product are equivalent, or better in the modified processes (Food and Drug Administration (FDA), 1995a, 1996). Moreover, sponsors take great risks with possible failure in clinical trials and marketing authorisation approval, with overall 18-29% historical success rate (Steinmeyer and McCormick, 2008), which is a large cost impediment.

The current downstream process development method normally initialises the bioprocess design by investigating several conditions in laboratory or pilot plant scale studies, and then scales up to large scale production in a general purification platform. This method invests much time, capital and efforts to design and optimise bioprocesses at both scales, but in the end may achieves little understanding and few improvements, because the process optimisation highly depends on experiences and lacks a proper design approach. Therefore, a systematic process design and development strategy at an early stage is extremely necessary to investigate bioprocesses, and so achieve manufacturing optimisation in a rapid and inexpensive way.

1.2 Downstream recombinant human antibody purification platform

Recombinant human antibodies are playing an important role in the pharmaceutical

industry with many successful applications including infectious diseases, cancer and autoimmune diseases, etc. In 2004, half of the new biopharmaceutics that were granted market approval in the USA and Europe were antibody-based products (Li et al., 2005). It is predicted that the monoclonal antibody products will have a large market share in pharmaceuticals and keep a significant growth rate in the near decade (Farid, 2007; Sommerfeld and Strube, 2005). Thus the downstream bioprocess development in biopharmaceutical industry mainly focuses on human antibody purification from mammalian cell culture.

Although the sequence of amino acids is different from one antibody to another, the structure of protein shares a basic biological framework and has some unique regions, e.g. Fc region. However, it is impossible to define a truly generic purification process for any antibody without knowledge of the operating conditions change (Kelley et al., 2009), but based on the similar characteristics of human antibodies, a purification platform with a similar sequence of bioprocesses can be achieved.

Currently, the general downstream purification platform for monoclonal antibody (mAb) consists of several chromatographic steps, starting with protein A affinity capture (Gottschalk, 2008; Lowe et al., 2001; Shukla and Thommes, 2010). Then the product is further polished by ion exchange chromatography and other processes such as ultrafiltration and virus inactivation (Sommerfeld and Strube, 2005). Using a platform approach has several advantages including greater harmonisation of process quality control and facility management in various manufacturing sites and better understanding for different products in the biopharmaceutical industry (Nfor et al., 2009). However, it is exactly the practical difficulties of bioprocess design and optimisation in this platform at the development stage that cause the cost and time problems mentioned above.

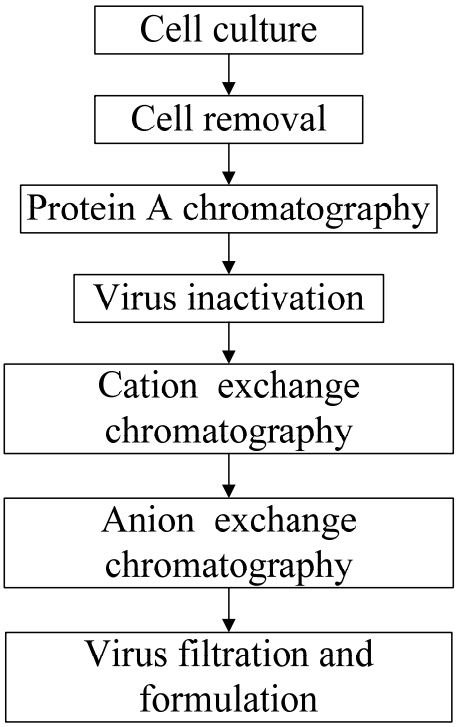


Figure 1.1 A general platform for downstream antibody purification.

1.2.1 Protein A affinity chromatography

As shown in Figure 1.1 (Shukla et al., 2007), the general platform starts with a clarified cell culture broth. A typical protein A chromatographic procedure needs the column equilibrated by neutral pH buffer first and then the cell free feedstock loaded under neutral pH conditions. The monoclonal antibody binds to the protein A ligand on the resin when the feedstock passes through the column. Most impurities flow through the column and will be washed out in the following washing steps. The antibody will be recovered by eluting in an acidic buffer, e.g. pH 3.5 sodium citrate buffer (Ghose et al., 2007). Since the solution is in an acidic buffer, it is relatively convenient to subsequently inactivate the virus (Phillips et al., 2007).

Protein A chromatography is the base of this platform, which is widely adopted in industrial commercial antibody production (Titchener-Hooker et al., 2008). However, there are several drawbacks to this technique. First of all, the cost of this process is very high due to low binding capacity, expensive resin and large volumes of buffer consumption (Ghose et al., 2007; Lowe et al., 2001). Considering that the requirement of purified antibody will significantly increase in the near future, the only solution is to increase the diameter of chromatographic column until the physical constraints, around 2m (Gottschalk, 2008) or invest in additional protein A chromatography plants, both of which will dramatically increase capital investment requirement (Kelley, 2007; Werner, 2004).

Due to the recent advances in upstream, high antibody titres at 5 g/L or more can be achieved in the future (Wurm, 2004). Meanwhile, the impurities, such as host cell protein (HCP), aggregates and DNA etc., also increase due to longer fermentation time, which may cause potential problems in protein A chromatography, e.g. easy blockage or pressure drop (Roque et al., 2007; Shukla et al., 2005). Secondly, the procedures of chromatography require several cycle phases in different buffer solutions. Long processing times including cleaning and regeneration, the

preparation and storage of large volumes of solution create a time bottleneck in manufacturing (Ghose et al., 2007; Shukla et al., 2007). Moreover, any leakage of protein A ligand from the purification step poses the potential risk to patients, and so requires extensive analysis and extra purification steps to remove it (Sommerfeld and Strube, 2005).

Although protein A chromatography has several drawbacks as mentioned above, it is still the most favoured process in the industry due to its high selectivity by affinity and relatively fixed, robust operation conditions (Roque et al., 2007). In the early development stage, with limited time and feedstock, protein A chromatography is currently the first choice for antibody purification because of its robustness and efficiency. However, once the process capacity requirement is more than protein A chromatography can provide in the future high titre, large scale mAb production, an extremely large number of protein A chromatographic purification cycles will be needed. It significantly increases either process time and/or cost per batch, which may make protein A chromatography too expensive to be considered in the platform. Thus alternative capturing processes will be preferred and investigated in future process development activity as early as possible rather than protein A chromatography.

1.2.2 Polishing chromatography and other purification processes

Several polishing steps (Figure 1.1), such as cation and anion exchange chromatography (IEX), follow after initial affinity capture in the platform to further remove impurities, e.g. host cell protein (HCP), genomic DNA etc, in order to achieve quality requirements set by regulatory authorities (Li et al., 2005; Shukla et al., 2007). In contrast to robust protein A chromatography, process development for IEX and other chromatographic processes is very complex. Although the binding of antibodies or other proteins to IEX is basically due to ionic interactions, the type of resins, ligands, buffer conditions and the property of the protein can influence the

process significantly (Gagnon, 2007; Yamamoto et al., 1983). The operation mode, e.g. binding-elution or flow through mode, step elution or gradient elution, and operation conditions, such as flowrate, also need careful investigations during process development (Ishihara and Yamamoto, 2005). Moreover, the scale-up development from lab or pilot scale to manufacture scale needs full examination and validation (Kaltenbrunner et al., 1997; GE Healthcare, 2009b). Sometimes, tradeoffs are needed as the optimal conditions found in lab scale can not be achieved in industrial scale, for instance, the optimal flowrate found in small scale exceeds critical velocity in large scale chromatography.

Besides ion exchange chromatography (IEX), many other separation techniques, such as hydrophobic interaction chromatography (HIC), absorption membrane filtration, gel filtration chromatography and ultrafiltration, are available and incorporated in polishing step (Nfor et al., 2008). Selecting proper candidate polishing processes from various available bioprocesses to form a reasonable sequence and subsequently screening operation conditions are complicated and time consuming. This part of development is normally achieved through large amounts of experiments and combining with researchers' experience and prior knowledge. Thus, a proper design method should be proposed to rationalise the bioprocess selection, the sequence of unit operations and operation conditions while minimise the time and materials requirement. Then whole polishing processes rather than a single process should be evaluated based on the overall performance.

1.2.3 Alternative processes to reduce chromatographic steps in platform

As already mentioned, downstream process development faces extra challenges from sophisticated upstream technology. The successful increases in cell culture titre make the downstream chromatographic process rate limiting, since all chromatography processes have limited upper capacity and can not be operated continuously (Shukla et al., 2007).

Currently, several non-chromatographic bioprocesses are being investigated and improved with the intention to replace one or two chromatographic steps, especially the low throughput protein A chromatography (Przybycien et al., 2004; Thommes and Etzel, 2007). Polyelectrolytes were successfully used by McDonald et al. (2009) for selective antibody precipitation. Ma and his co-workers (2010) then demonstrated polyamines precipitation as an alternative antibody separation process. Some researchers have evaluated the potential use of aqueous two phase systems (Andrews et al., 1996; Azevedo et al., 2009; Rosa et al., 2010) in antibody purification using PEG and salt solutions. High-performance tangential flow filtration (HPTFF) with absorption membranes also emerged as a new technology that enabled the antibody separation (Mohanty and Ghosh, 2007; Reis et al., 1999). All these innovative processes can achieve good antibody separation and the results were quite comparable to that of affinity separation.

However, these processes are relatively less generic for any antibody than protein A chromatography due to their non-affinity mechanism. If there is one possible alternative process to replace protein A chromatography in the platform for future large scale purification, the process and the operation conditions will have to be defined and optimised case by case in most circumstances. Therefore, the material specificity caused by different alternative capturing processes will require more intensive, rapid downstream process development and evaluate capturing and following polishing steps as a whole at the early stage concerning quite broad aspects, from the selection and sequence of processes to individual operation condition optimisation (Nfor et al., 2008).

1.2.4 Quality considerations in development and manufacture

For therapeutic antibody production, it is very important to address the quality and safety of the products, mandatorily required by regulatory bodies. Previously, the strict quality requirements and effort were mainly focused on manufacturing rather

than process development, quality controlled by specifications rather through understanding. Now both the regulatory authorities and industries desire the quality by design (QbD) concept which was initiated by the FDA and widely accepted in the industry (Yu, 2007). It is a comprehensive approach to product development and life cycle management with the following elements: design and developing processes, identifying critical quality attributes, process parameters and sources of variability (FDA, 2006). A thorough understanding of process parameters impacts and interactions on product quality should be obtained on a defined design space in the process development phase (Kelley et al., 2009).

It is an important change from the current regulatory view and efforts to guarantee the quality, efficacy and safety of the therapeutic products as early as possible (European Medicines Agency (EMA), 2000). In most cases, a mathematical equation or model, no matter whether derived from theory or purely empirical, is highly beneficial in the early bioprocess development in order to predict, control and analyse quality effects (Yu, 2007; Rathore, 2009). Nevertheless, currently there is no general condition or model which can be directly applied to different bioprocesses and products. Even under the same purification platform, the purification conditions and processes are still developed for one product and may vary significantly between each other. Therefore, a general systematic approach is urgently needed in order to develop and establish the relationship between the critical process parameters (CPP) and critical quality attributes (CQA) (ICH, 2008) while appreciating the differences of different products and their effects.

1.3 Existing solutions for bioprocess development

The issues mentioned above are widely recognised by researchers, so that various approaches have been designed and applied, with the aim to solve the problems or fulfil the requirements to accelerate downstream bioprocess development. These approaches include bioprocess modelling, experimental design and high throughput

technology (Carrier et al., 2010; FDA, 2004; Lee and Gilmore, 2006; Mandenius and Brundin, 2008; Milavec et al., 2002; Nfor et al., 2009).

1.3.1 Bioprocess modelling

1.3.1.1 The benefits of process modelling

Conventionally, process designs and improvements in industrial processes are achieved by trial and error methods, i.e. by empirical methods guided by intuition and experience. In this way, luck is an essential element of success and the method gives little insight into the processes involved (Simutis et al., 1997). Even a few years ago modelling has not been considered as an important element for industrial practice but was mainly used by academics. This is because it requires enormous efforts and resources to achieve an accurate bioprocess model.

Recently, modelling has attracted more attention from industry (Velayudhan and Menon, 2007). The main driving force behind this transition is the emerging requirements from Quality by Design (QbD) and Process Analytical Technology (PAT), highly suggested by regulatory bodies, such as US FDA and The International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH). These concepts and technologies aim to utilise mathematical modelling methods to enhance the understanding of complicated bioprocesses in pharmaceutical manufacturing, with the long term benefits for companies to develop and design processes more quickly, carry out online analysis with risk management and facilitate batch release (Mandenius and Brundin, 2008).

Several pharmaceutical companies have applied modelling to processes and gained benefits arising from their utilisation (Chhatre et al., 2011; Jimenez-Gonzalez and Woodley, 2010; Oatley et al., 2005; Rohner and Meyer, 1995). For instance, Lonza, a

Swiss company, found modelling very helpful to indentify the most cost-effective operation mode, and it also helped to develop and practice an adequate process control system (Vasic-Racki et al., 2003). Statistical or mathematical models, such as chromatography general rate model, now coupled with a powerful modern computer and special designed software, e.g. COMSOL (Fang, 2010), are far more effective in process design and productivity optimisation than conventional trial and error pathway (Banga, 2004; Van Impe, 1996; Schubert et al., 1994; Simutis et al., 1997).

The performance of a process sometimes, depends strongly on the operation of a few key unit operations (Gosling, 2005). If a mathematical model can describe the performance characteristics of one key operation with known input and control parameters, it will significantly reduce the number of experiments, which not only decrease the expense but also reduce the development time. That is exactly the classical and essential aim of the establishment of a bioprocess model. Now, besides its function in cost and time reduction, in the framework of QbD and PAT, the bioprocess model developed in the early research and development stage can go beyond the laboratory to benefit the routine manufacturing and guarantee the quality and safety of the final products.

1.3.1.2 Current modelling methods

Developing a bioprocess model is not an easy task, which itself is time-consuming and may involve complicated non-linearity computation due to the properties of biomaterials. An ideal process model should represent the process properties in a quantitative way and be able to predict the process results accurately and precisely by computational analysis and optimisation algorithms (Datta and Sabiani, 2007). However, an ideal model may not always exist and much effort is needed to develop in most cases due to the extremely complex biosystem. Moreover, a model may not be based on the fundamental science; in other words, it may be a totally empirical model based on quantitative conditional parameters.

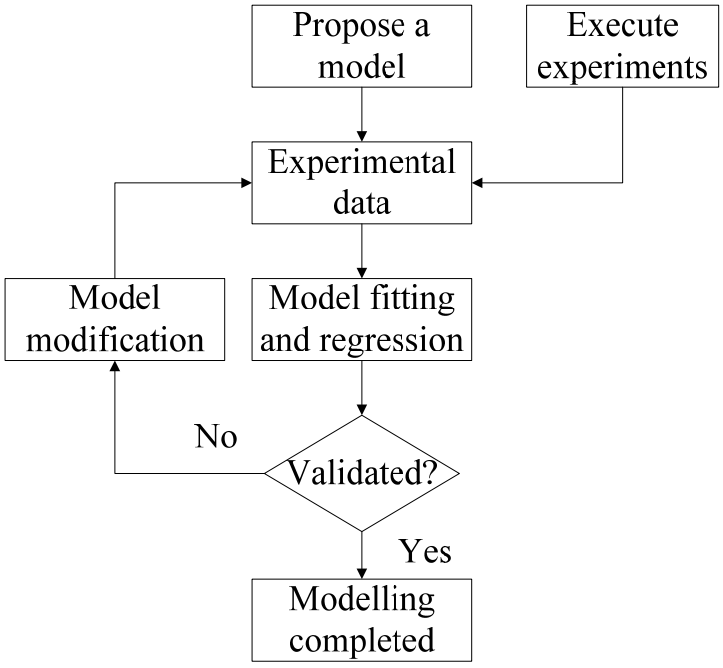


Figure 1.2 Process modelling flowchart.

1.3.1.2.1 Traditional modelling procedure

Figure 1.2 shows the traditional modelling flowchart. It starts by proposing a model structure according to the developer's knowledge and aims. The model can be proposed based on theories, when the mechanism of process is well understood and can be described by mathematical equations, such as an enzyme kinetic model (Michaelis and Menten, 1913). In engineering practice, it is quite common to build an empirical process model, when the process mechanism is unknown or too complicated to form a mathematical relationship. If there is a relatively simple mathematical equation which can then fit the data, it will be used to describe the process, for instance, Cohn's precipitation equation (Cohn, 1925). The third approach is to propose a hybrid model. A hybrid model is a mathematical equation combining the mechanistic knowledge with correlations of an empirical nature for aspects that are not yet fully understood (Galvanauskas et al., 2004). No matter which type of model is used, the model needs experimental data to fit and regress the parameters. The model fitting results will be evaluated and validated by statistical tests. Normally, it needs modifications or correlations to fine tune the differences between initial proposed structure and real experimental data often by few iterations.

Although the flowchart is quite simple and straightforward, the real difficulties of initial model proposal and the following model modifications can complicate the modelling activities and requires the model developer to have good mathematical background. Another drawback of this conventional modelling procedure is that the model does not interact readily with experiments. The experimental data may not be sufficient in the number or quality for modelling. It is also the concern of experimental design, which will be discussed later, that how to design experiments in order to obtain effective information and achieve maximum process understanding with minimum experimental number.

1.3.1.2.2 Innovative artificial neural network modelling

In order to simplify or solve the problems of model proposal and modification, artificial neural network (ANN) with hybrid process models can be used. The structure of the neural network based process model may be considered generic, in the sense that a specific initial model and its determination based on prior experience or mechanistic knowledge are not required (Massimo et al., 1991; Montague and Morris, 1994). ANN model is made up with highly interconnected layers of simple 'neuron-like' nodes, which act as non-linear processing elements within the network. An attractive property of ANN is that given the appropriate network topology, they are capable of learning and characterising non-linear functional relationships through experimental data.

The ANN based model is particularly useful in applications where the complexity of the data or task makes the design of such a function by hand impractical. It can be applied to modelling, especially time series prediction, sequential decision-making and data mining. Combining hybrid models with ANN is an advanced method to be used with either full mechanistic models or full empirical equations (Zorzetto et al., 2000). It works very well in many applications and gives more accurate modelling, optimisation and process control (Rivera et al., 2007; Schubert et al., 1994).

However, one significant weakness of ANN is that it requires a great amount of data to get the algorithm trained and it relies heavily on computers and software. This requirement creates a real dilemma especially for bioprocess development. In the early development stage, the design and optimisation method should reduce the time and material usage as much as possible. Moreover, when considering extra variables or when trying to apply the model to other similar biosystems, it needs training again, which is still time consuming and costly, without giving any valuable information from a previous model. The application area of this method is thus restricted to a system, which is able to provide thousands of data cheaply and rapidly, and only

some specialised personnel can understand the algorithms. Probably this is the reason why there is, as far as we know, not much applications of this technique in the real bioprocess industry yet (Galvanauskas et al., 1998).

1.3.1.3 Existing bioprocess models and research

There are several bioprocess models widely applied in downstream process development. Wong et al. (1996) applied a classic centrifuge model based on Stokes' Law to describe the centrifugal processing for inclusion bodies separation from *E. coli*. This model is also used to understand the tubular bowl centrifuge by Jain et al. (2005) with the help of CFD software. The general rate model proposed by Gu (1995) is frequently used in chromatography research for modelling and optimisation. However, the coefficients in the model require extra isotherm studies to provide, based on the regression of empirical isotherm models, e.g. Langmuir adsorption model. Not all models can be directly used for the bioprocesses, though they have been well developed and validated. Kuehner et al. (1996) proposed a precipitation model based on thermodynamic theory with several thermodynamic parameters. It successfully explained the precipitation in a scientific way, but hardly had applications in the engineering field due to a lack of accurate parameters for a real complex feedstock. In contrast, other two empirical models developed by Niktari et al. (1990) and Habib et al. (2000) respectively can be utilised in the precipitation process with only a few engineering based parameters.

If there is no process model structure available and meanwhile it is also impossible to create a new process model quickly, the current way is to make a temporary polynomial model based on statistical design of experiments (DoE) and thus achieve understanding and design through that polynomial equation. Although many published papers use DoE (Lee and Gilmore, 2006; Massimo et al., 1991) to model and optimise a bioprocess successfully, building a general pragmatic process model, which covers full range of all critical conditions in the process, still requires huge

efforts to collect experimental data, which is normally labour-intensive and costly. Therefore, the models available in the bioprocess industry are normally limited to certain design space or operations, e.g. Cohn's precipitation equation works well in a relatively high salt concentration range rather than low salting-in range. Especially, any polynomial models from DoE highly depend on the design space (typically a small space) and thus cannot be used to predict outside the original design range even under the same process with same materials. This is the reason why these polynomial equations are considered as temporary models rather than validated process models.

1.3.2 Experimental design

Bioprocesses are complex systems with various parameters and outputs, which need much effort to investigate the relationship between results and conditions. Initially, the experimental design and optimisation were achieved by trail and error methods or other empirical design based on intuition and previous experiences (Simutis et al., 1997). This empirical approach depends highly on experimental designer's knowledge and involves large uncertainty. Another disadvantage is that it hardly has an insight into the interactions between each parameter nor be able to achieve a satisfactory tradeoff while multiple objectives for one bioprocess exist (Sendin et al., 2006).

The aim of experimental design is to extract the most information possible from the minimum of elaborately designed experiments in the case of limited time and resources (Atkinson, 1996; Fedorov, 1971; Pukelsheim, 1993). It normally investigates the mathematical relationships between inputs and outputs combining with the purpose to optimise the system in most case (Karlin and Studden, 1966). Although the concept of this methodology and mathematical fundamentals have been well demonstrated and known for nearly eighty years (Box and Lucas, 1959; Fisher, 1932, 1935; Kiefer, 1959; Yates, 1959), it has not been widely used in

research until about thirty years ago with the advance of computers (Steinberg and Hunter, 1984). During these years, this methodology shows its benefits in nearly all science and engineering fields, such as pharmacology (Lewis et al., 2005), electricity engineering (DeVoe and Pisano, 1997), chemical engineering or known as chemometrics (Lazic, 2004), clinical trials (Fleiss, 1999), and even has applications in finance and marketing (Kuhfeld et al., 1994). However, cases in downstream purification processes are still very rare due to the complication of bioprocesses (Atkinson and Tobias, 2008).

1.3.2.1 Statistical DoE design

Among all well developed experimental design methods, statistical DoE is the most useful systematic approach to obtain and analyse information-rich data. DoE has been adopted and applied to sciences and engineering in quite broad fields, including bioprocesses (Mandenius and Brundin, 2008). It is more helpful for bioprocesses, which normally lacks an accurate mathematical model and has high noise levels in experimental data due to biological materials variation, interactions among variables and complex biochemical reactions (Lee and Gilmore, 2006).

Statistical DoE is a group of experimental design methods developed by statistical analysis. A design in which every setting of every factor appears with every setting of every other factor is a full factorial design and normally expressed using the notation I^k, where I is the number of factor level and k is the number of factors (Montgomery et al., 2005; Montgomery, 2008). However, even if the number of factors and its levels in a design is small, for example, 5 factors with only 2 levels each, it still requires $2^5 = 32$ runs. Therefore, a fractional factorial DoE was designed and adopted, which was a carefully chosen fraction of the complete factorial experiment runs (Box and Hunter, 1961). It reduces the number of experiment runs, particularly when both the number of factors and levels are high

(Gunst and Mason, 2009). Compared to full factorial experiments, the fractional factorial experiments may have some disadvantages such as less precision in the parameter estimates, and main effects and interaction effects are indistinguishable.

The response surface methodology (RSM) is also a derived statistical-based DoE with the power to predict the shape of the response surface for a system with several 3-D or contour figures (Anderson and Whitcomb, 2005). This method was first introduced by Box and Wilson in 1951 (Box and Wilson, 1951). RSM is very useful when the system response is non-linear. The traditional method involves a small design to fit the first-order model in the first stage. Then central composite designs (CCDs) with face-centred are utilised to fit second-order or even higher degree polynomial equations efficiently without challenging the estimation of the first-order linear effects (Gilmour, 2006). The DoE approach together with RSM provides an understanding of the bioprocess within a useful region in more depth (Chhatre et al., 2011) and helps bioprocess development from media optimisation (Gheshlaghi et al., 2005; Torres-Bacete et al., 2005) to antibody production (Garcia-Arrazola et al., 2005).

However, all statistical DoE generate empirical models only, in the most form of a polynomial equation without any physical understanding. They are based on statistical calculation with all parameters estimated. When there are many parameters and the design space is very large, statistical DoE will waste much effort in exploring uninteresting area and also produce equations which may not be appropriate, though empirical models are exceptionally useful in describing processes in which the mechanisms are extremely complex and incompletely understood as in most biological processes (Lee and Gilmore, 2006). Another disadvantage is that they can not predict the process performance outside the design space since it creates an empirical equation only in the defined space.

In the early development of bioprocess downstream purification, massive variables

or wide design space need evaluating while especially, time and feed are always limited. The normal timeline for the early stage process development is around 5 months (Jones et al., 2010) and the available material will be just 100 g to 1 kg from a small cell culture batch, for example, a 500 L scale bioreactor (Glynn et al., 2009; Coffman et al., 2008). Under such circumstances, a set of full design DoE with a large number of experimental runs is not feasible meanwhile a set of small DoE is not efficient to extract enough information from bioprocesses and achieve full understanding.

1.3.2.2 Optimal experimental design based on models

Optimal experimental design is a mathematical design methodology based on models. It was first introduced by Fisher (1935), then explored and extended by Kiefer (1959) and Box et al. (1959). The design methodology can be simplified when the model it is based on is a polynomial equation and gives the same design as above DoE design. However, it is more powerful than DoE design from a mathematical point of view. It allows to design experiments for any process containing discrete, continuous or both types of variables with a linear or nonlinear profile, no matter whether the model is mechanism based, hybrid or empiric (Ford et al., 1989). It also can be applied to discrete or irregular design space in some constrained or dynamic processes, where DoE cannot work at all (Atkinson, 1996).

The theory is initiated by choosing a set of experiments to accurately estimate the parameters in a model (Atkinson and Hunter, 1968). Thus the model is solely mathematically oriented rather than process performance oriented. The optimal experimental design itself is not considered as a direct optimisation technique but an important part of the modelling activity. However, because an accurate model can provide information for process optimisation, the optimal design which aims to make model more accurate actually promotes the process design and optimisation in an indirect way. Combining with optimisation algorithms, it can obtain useful

information from statistically or mathematically designed experiments to improve model, parameters and achieve optimisation at the same time (Franceschini and Macchietto, 2008). The methodology has broad applications in engineering (Box, 1966; Box and Draper, 1971; Galvanauskas et al., 1998), e.g. kinetic parameters estimation, to sciences (Harnett et al., 2008; Paterno, 1999), even in the social development science (Gregoire, 1998; Sarndal et al., 1978). The information extracted through model based optimal experimental design can be thus used for model parameter estimation (Watts, 1994), model validation (Franceschini and Macchietto, 2007) or process optimisation (Levisauskas et al., 2003).

The mathematical principle of model based optimal experimental design is to make the determinant of the parameters variance-covariance matrix as small as possible (Franceschini and Macchietto, 2008). There are several criteria and algorithms to minimise the variance-covariance matrix as follows (Montgomery, 2008; Pukelsheim, 1993):

D- optimality: minimise the generalised variance of the parameter estimators;

A- optimality: minimise the average variance of the parameter estimators;

G- optimality: minimise the maximum variance of the predicted values;

V- optimality: minimise the average variance of the predicted values.

Among these several design criteria, D-optimal design (Box, 1966; Box and Draper, 1971) is the most used algorithm to construct model based optimal experimental design (Mitchell, 1974; Cook and Bachtsheim, 1980; Papalambros and Wilde, 2000) due to its generality in parameter estimation and computational efficiency, which will be detailed in a later chapter.

Most downstream bioprocesses are developed based on known biochemical and biophysical interactions, which, in theory, can be mathematically modelled through an appreciation of the reaction mechanisms, e.g. mass balances, hybrid grey box or even black box (Papalambros and Wilde, 2000). These process models, which

contain a priori knowledge for the specific bioprocess, can be used by an elaborate model based optimal design algorithm to reduce not only the feedstock and time required but also facilitate the process improvement and understanding (Galvanauskas et al., 1998).

Galvanauskas et al. (1998) proposed a sequential design method shown in Figure 1.3. They changed the traditional batch mode in the optimal design to a continuous sequential mode. They started to design the first experiment with a process model and some priori knowledge. The results were analysed and regressed to the model. Then the design required iteration of model based experimental design, experiment execution and data analysis. When the preset objective was achieved, the iteration was terminated and design was completed. This method promoted the interplay between data and model, which further reduced the time and number of experiments. This design method was applied successfully in several bioprocess operations, such as *E. coli* fermentation and homogenisation (Chung et al., 2000; Galvanauskas et al., 1998; Middelberg et al., 1992; Souza et al., 2008). The method benefits from the prediction power of the model to obtain potential design candidates and on the other hand, the potential design candidates provide new experimental data for the model improvement. Therefore, the method and the final design are highly dependent on the correctness and robustness of the initial chosen model.

There are two obvious disadvantages. One is that it needs a correct model, which may not be available in some cases. Another disadvantage is that model based optimal experimental design is always computationally intensive due to the mathematical algorithm used and the complexity of model. Currently, for the applications in downstream bioprocesses, the difficulty in modelling a bioprocess and the complex computation required are the main limitations because most bioprocesses need nonlinear or even differential models. They may be too complicated to have a suitable and tractable mathematical model or computation algorithm.

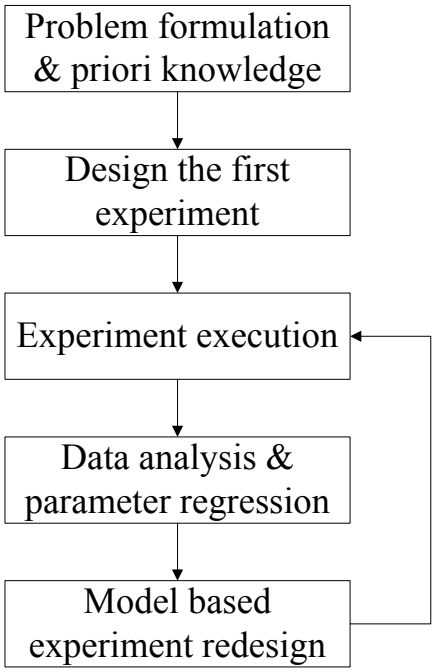


Figure 1.3 The flowchart of sequential model based optimal experimental design.

However, these limitations can be effectively overcome by developing advanced mathematical tools and more powerful computer technologies. Since both regulatory authorities and industries increasingly emphasise the benefits of bioprocess understanding and modelling, it is foreseen that optimal experimental design will play an important role in future bioprocess development.

1.3.2.3 Other experimental design

There are other non-mathematical experimental design techniques such as randomised designs, which are applied to one primary interest factor or variable through random assignments. Taguchi designs, the main idea of which is to test the system tolerance, are used to find the robust operation conditions (Rao et al., 2008). Simplex search is widely used when considering a mixture design subjected to a constraint (Morgan et al., 1990) and demonstrated its ability in optimising bioprocesses (Al-mhanna et al., 2010; Chhatre et al., 2011). Many papers have been published for the successful utilisation of these designs to improve processes and obtain very good results (Gohel et al., 2006; Weuster-Botz, 2000).

These methods are powerful because they can identify the optimal operation conditions without involving any model or regression equations but give poor insight into how the process performance changes in the vicinity of the solution due to their objective orientated design. Therefore, most of these methods are unable provide an understanding of the bioprocess even after successfully completing process optimisation. In the biological industries, regulations restrict the freedom of the engineer to alter a process design. It is thus quite crucial that the process engineer knows how the process responds in the vicinity of the optimal solution so as to determine how best to achieve the desired robust solution when subject to process uncertainties and changes (Zhou and Titchener-Hooker, 2003). Moreover, all industries and regulatory bodies now focus on Quality by Design (QbD) and so pay

significant attention to the importance of bioprocess understanding (Chhatre et al., 2011; Shah et al., 2010). These designs will not be sufficient to meet the future requirements alone. However, they can be used as supplemental designs in bioprocess design activities to provide extra information.

1.3.3 High throughput technology

High throughput technology had its breakthrough in the late 1990s (Bensch et al., 2005) due to the invention and development of fully automated liquid handling systems. It is a powerful platform for analytical and high-throughput screening experimentation especially in the fields of biology and chemistry (Lye et al., 2003; Micheletti and Lye, 2006). With the help of robotics and computer software, high throughput technology allows a researcher to conduct and analyse thousands of biological, chemical or pharmacological tests at a same time with minimum material requirement using 96, 384 or even 1536 microwell plates (Hertzberg and Pope, 2000; Titchener-Hooker et al., 2008).

This is now widely used in the biopharmaceutical industry to address many issues, from drug discovery to bioprocess development, and specifically valuable at the early stages of development and at low cost (Carrier et al., 2010; White, 2000). Merck applied this method in preclinical researches (Rodrigues, 1997) and Genetech used it in stable cell line generation (Torres et al., 2009). Technology providing companies also investigated and utilised this platform, for example, GE Healthcare and DSM developed a combinatorial protein-ligand library and evaluating affinity tags based on high throughput screening (Cawse, 2001; De Vries and Lefort, 2008).

Downstream protein purification development involves several separation steps with a significant number of variables to be tested. It should definitely benefit from the application of high throughput technology by providing variable screening and process understanding to speed up bioprocesses development. Using microwell

experiments also makes the minimal material consumption possible at the early phase (Nfor et al., 2009). Several researchers have already well demonstrated its power in developing downstream purification processes, such as chromatography screening (Coffman et al., 2008; Kelley et al., 2008; Kramarczyk et al., 2008), microfiltration (Jackson et al., 2006) and precipitation (Knevelman et al., 2010).

However, high throughput technology has it own challenges. The findings from the microwell experimentation should be able to predict the larger scale processes. Depends on the process investigated, the efforts of scaling up development may vary (Marques et al., 2010). Some bioprocess conditions may need correlation once scaling-up to pilot scale, for example, optimising mixing speed in cell culture reactor from microwell shaking speed (Micheletti et al., 2006). For some dynamic system such as chromatography, it is naturally difficult to scale down flowrate or pressure to test in a microwell system. Although there are some innovative devices created for handling this issue, such as PhyNexus tips (Chhatre et al., 2009) or Atoll columns (Lye et al., 2009; Wiendahl, 2008), it still needs extra tools, e.g. a correlation model, to scale up and predict large scale performance successfully (Bergander et al., 2008; Kaltenbrunner et al., 1997).

Another issue comes from the large amount of high throughput experimental data obtained with the consideration of data mining, information extraction and understanding (Diller and Hobbs, 2004). Due to its parallel capacity to test many variables, hundreds or even thousands of data points can be produced in a short time. It needs some mathematical tools, e.g. a mathematical model based algorithm, to design the experiments efficiently and then analyse the results rapidly and meaningfully after experimentation (Malo et al., 2006; Zhou and Titchener-Hooker, 1996). These tools are now encouraged by regulatory bodies, such as the FDA, through Process Analysis Technology (PAT) and Quality by Design (QbD) initiatives, to obtain the understanding of crucial process parameters and address their impact on the product quality attributes (Chhatre and Titchener-Hooker, 2009; Nfor, 2009).

1.4 A combination approach for bioprocess development and optimisation

All solutions mentioned above have the ability to accelerate a particular bioprocess development and achieve optimisation, if some optimisation algorithms are used. However, many of the researchers mentioned above applied only one of these methods alone to solve some specific problems in the process rather than a combined solution to facilitate process understanding and optimisation.

Now, the benefits of combination are being recognised by several researchers. Titchener-Hooker et al. (2008) proposed the following flowchart, Figure 1.4, to give a pioneer concept, which combined scale-down devices, e.g. microwell plates, with mathematical models to predict process scale performance (Willoughby, 2006). Salte et al. (2006) successfully used this approach to optimise the conditions for separating high density cell broth. Ma (2010) applied modelling and ultra scale-down shear filtration system to develop large scale diafiltration process.

However, the approach may need two potential improvements. First, it needs a proper feedback loop between modelling and scale-down experiments. A similar iteration as in Figure 1.3, between the model and experiments will be very helpful. The model based experimental design has the capability to not only produce a better experimental plan for scale-down experiments, but also enhances the modelling. Second, a scale-up method or process analysis is also required. Although this concept is a ‘top-down’ design rather than a ‘bottom-up’ design (Willoughby, 2006), it still needs a scale-up method or analysis to link and predict the performance of process at different scale.

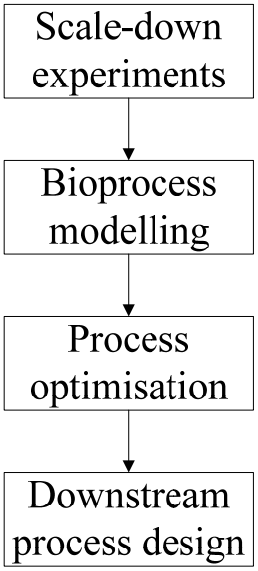


Figure 1.4 A process design and optimisation approach based on the combination of scale-down experiments and bioprocess modelling.

1.5 Research objectives

From the straightforward view of the biochemical engineering and based on previous research mentioned above, an efficient and cost-effective method for downstream purification processes design and optimisation is extremely important and desirable. This thesis examines these needs by focusing on the precipitation process, with the objectives as follows:

1. To present an integrated model based process design methodology for downstream bioprocess optimisation, combining process modelling, model based experimental design and high throughput experimentation to enable faster bioprocess development and optimal process identification.

2. To provide a bioprocess modelling procedure facilitated by high throughput experimentation. This will be carried out by applying the process modelling procedure in Fab' precipitation and also increase process understanding. The newly developed precipitation model will provide the mathematical equation for following studies.

3. To apply integrated model based process design in Fab' precipitation to find the optimal operation conditions based on the previous proposed precipitation model and microwell data. A comparison with the traditional DoE method will be used to evaluate the integrated model based process design methodology.

4. To transfer and test the generality of the integrated model based process design methodology to a different product process: a mAb precipitation system. Then the optimal conditions found at microwell scale will be scaled up at lab scale to examine the precipitation and following up solid-liquid separation.

5. To select and screen the potential polishing chromatography processes in microwell plates after the precipitation and use the whole process cost analysis to evaluate the effectiveness of precipitation based purification processes.

Chapter 2. Materials and Methods

2.1 Materials

2.1.1 Feed materials

2.1.1.1 Fab' feedstock

E. coli W3110 (ATCC 27325) was designed and donated by UCB-Celltech (Slough, UK). It had pTTOD A33 IGS2 plasmid encoded with tetracycline resistance and A33 antibody fragment (Fab') DNA sequence. The *E. coli* cells were produced with a fed-batch mode in a 20 Litre fermenter (Sartorius Stedim, UK). The Fab' was expressed and accumulated intracellularly. The fermentation protocol and medium composition were detail described by Bowering et al. (2002). Cells were then harvested by centrifugation in an Eppordorf 5810R centrifuge (Eppendorf GmbH, Germany) at 12,000 rpm for 2 hours. For pilot scale fermentation, *E. coli* cells were then harvested in CSA-1 disc-stack centrifuge (Westfalia Separator AG, Oelde, Germany). It was operated at 9800 rpm, room temperature with an inlet flow rate 50 L/h. The cell paste was collected in 500 g each lot and stored in a -70 ^{o}C freezer (New Brunswick Scientific, UK).

For one set of experiments, the same Fab' homogenate from one cell paste lot was used. Before each set, fresh Fab' homogenate was produced and stored in a +4 ^{o}C fridge for at most one week. Frozen *E. coli* cells were resuspended in 10 mM pH 7.0 phosphate buffer at 40% (w/w) and homogenised in APV Lab 40 Homogeniser (APV homogeniser GmbH, Lubeck, Germany) at 750 bar, three passes. Then the homogenised solution was centrifuged in an Eppendorf Centrifuge 5810R (Eppendorf GmbH, Germany) at 12,000 rpm for 2 hours. The clarified Fab' homogenate supernatant was collected and stored in the fridge as feedstock.

Pure Fab' solution was purified by MabSelect Hitrap 1ml column in AKTA Basic with UniCorn 5.1 software (GE Healthcare, Sweden). The column was equilibrated in 20 mM, pH 7.4 sodium phosphate loading buffer with 1 ml/min flowrate. The clarified Fab' homogenate was buffer exchanged to loading buffer and applied to column. 20 ml loading buffer and further 5 ml loading buffer with 10% (v/v) isopropanol was used to wash the column. The pure Fab' solution was eluted by 100 mM, pH 3.5 sodium citrate buffer at 1ml/min. The elution was stored in a +4 ^{o}C fridge after pH was neutralised by 50 µl per 1 ml elution 1 M pH 9.0 Tris buffer immediately.

2.1.1.2 mAb cell culture

The CHO *cyo1* cell line was developed and authorised by Lonza (Slough, UK) for culturing in the department of Biochemical Engineering, UCL. A recombinant humanised IgG_4 type monoclonal antibody (mAb), pI 7.4 to 7.8 determined by isoelectric focusing gel, was expressed extracellular by this CHO *cyo1* cell line.

The seed CHO cells were stored in the liquid nitrogen with 10% (v/v) DMSO. Before mammalian cell culture, the frozen CHO cells were revived and passaged in 50 ml flask, incubated at 5 % (v/v) CO_2, 37 ^{o}C for two generations. The CHO cells were then cultured in CD CHO Medium in a 20 Litre fermenter (Sartorius Stedim, UK) by a 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 antifoam were controlled by fermenter automatically according to preset value. Cell culture was harvested at mAb concentration around 1 g/L. It was then processed at 10,000 rpm by Eppendorf Centrifuge 5810R for 30 min and then filtered through a 0.22 mm filter (Millipore Limited, Dundee, UK). The detailed fermentation protocol can be referred to Galbraith et al. (2006).

The clarified cell culture was stored in a -70 ^{o}C freezer. For each set of experiments, the same batch of cell culture was used. During experiments, clarified cell culture was defrosted and then concentrated to around 2 g/L mAb using 5 kDa concentration tubes (Millipore Limited, Dundee, UK). The concentrated material was diluted by PBS to required mAb concentration in the following studies.

2.1.2 Chemicals

Sodium monobasic phosphate, sodium dibasic phosphate, sodium acetate, sodium glycine, PEG 6000 and ammonium sulphate et al. were purchased from Sigma Chemical Co. Ltd. (Dorset, UK). All chemicals were reagent grade. Ethanol and other HPLC solution were HPLC analysis grade. CHO cell culture medium: CD CHO Medium and CD CHO Medium AGT were brought from Invitrogen (Paisley, UK).

2.1.3 Antibody standard

Fab' and mAb standard were produced by two chromatographic steps. The first process was protein A chromatography. MabSelect Hitrap 5ml column (GE Healthcare, Sweden) was equilibrated in pH 7.4, 0.15 M sodium chloride, 20 mM sodium phosphate buffer with 1 ml/min flowrate on AKTA Basic (GE Healthcare, Sweden). The feedstock was buffer exchanged to pH 7.4 loading conditions in Millipore 5 kDa concentration tube (Millipore Limited, Dundee, UK) and loaded onto the column. Three column volume loading buffer was followed to wash the column and one column volume loading buffer with 10% (v/v) isopropanol was used to wash out hydrophobic impurities. Then, bound protein, Fab' or mAb, was eluted by 100 mM, pH 3.5 sodium citrate buffer at 1ml/min, with immediately adding 50 µl per 1 ml elution 1 M pH 9.0 Tris buffer to neutralise and protect antibody. Loading, washing, elution and fraction collection were executed and monitored by AKTA Basic with Unicorn software (GE Healthcare, Sweden). A gel filtration column was

used for the second step purification. 60 ml Superdex 200 (GE Healthcare, Sweden) was packed in a XK 16/40 column (GE Healthcare, Sweden). The protein A elution pool from above step was conditioned by 20 mM sodium phosphate buffer, 0.15 M sodium chloride pH 7.0 buffer. After the gel filtration column was connected to AKTA Basic and equilibrated for 120 minute at 1 ml/min flowrate, the conditioned post protein A solution was applied to column and the main antibody peak was collected as the standard.

2.2 Methods

2.2.1 High throughput precipitation

The Fab' precipitation was carried out in ABgene 96 deep microwell plates (Epsom, UK) by Packard MultiPROBE II HT EX (Packard BioScience Company, Meriden, US). The liquid handling robot was controlled by WINPREP software with process programmed and volume information provided by preset Excel file. During liquid transfer, 200 µl and 1 ml conductive disposable robotic tips (Tecan Group Ltd., Switzerland) were used by arms of the multiprobe to avoid contamination.

Acetate buffer was used for pH from 4.5 to 5.5 and phosphate buffer was used for pH 6.0 to 8.0. The precipitation conditions chosen were pH from 4.5 to 8.0 with intervals of 0.5 and ammonium sulphate concentration from 0 mol/L to 3.0 mol/L with interval 0.2 mol/L for pure Fab' solution and interval 0.3 mol/L for clarified homogenised solution. 1.2 M sodium acetate and acetic acid were prepared as stock solution. Robot aspired different volume of stock solution to form 150 µl buffer with required pH. 1.2 M sodium monobasic phosphate and sodium dibasic phosphate were also prepared and used to form pH buffer. The final buffer concentration in each precipitation system was 100 mM due to 12 times dilution by salt solution, feedstock and water. 4 M ammonium sulphate solution was prepared as salt stock solution with pH adjusted to required condition by sulphuric acid and ammonia. The

corresponding volume of salt stock solution and water will be respectively taken to each microwell to give designed salt concentration. 300 µl feedstock was then dispensed into microwell to start precipitation. The total volume of small-scale precipitation was 1.8 ml, including salt solution, pH buffer, distilled water and Fab' solution. The precipitation plate was mixed on Eppendorf thermomixer (Eppendorf GmbH, Germany) at 600 rpm, room temperature for 2 hours with a lid on to prevent evaporation during incubation. The plate was then centrifuged in Eppendorf Centrifuge 5810R at 4000 rpm for 15 mins. Clear supernatant was transferred to Agilent 96 HPLC micro-well plate on Agilent HPLC 1200 series system (Agilent Technologies, Stockport, UK) for analysis.

The mAb precipitation was carried out in the 96 microwell filter plate with 0.45 µm Durapore membrane (Millipore Limited, Dundee, UK) by Tecan robot (Tecan Group Ltd., Switzerland). Since the mAb formed lipid-like precipitate, it required higher mixing speed and only filtration can successfully separate solids from liquid. The precipitation conditions chosen were pH from 5.0 to 8.5 with an interval of 0.5, ammonium sulphate concentration from 0 mol/L to 2.2 mol/L with an interval of 0.2 mol/L and PEG percentage from 0% to 22% (w/w) with an interval of 2%. Acetate buffer was used for pH 5.0 to 5.5 and phosphate buffer for pH 6.0 to 8.5. The total volume of small-scale precipitation was 250 µl, including salt solution or PEG solution, 20 mM pH buffer, distilled water and mAb solution. 25 µl pH buffer was first aspired from 200 mM stock solution and added into each microwell, followed by required volume of precipitant, either from 4 M ammonium sulphate stock solution or 40% (w/w) PEG stock solution, and water. The microwell plate was then mixed in the Eppendorf thermomixer for 15 mins to avoid unevenly distributed solution before dispensing 50 µl feedstock. The precipitation plate was incubated at 20 ^{o}C with a microwell lid on and mixed in Eppendorf thermomixer at 1000 rpm for 2 hours. Then microwell plate was centrifuged at 4000 rpm in the Eppendorf Centrifuge 5804R for 15 min with a 96 microwell receive plate underneath. Clear

permeate was transferred to Agilent 96 micro-well plate for HPLC and other analysis.

2.2.2 Lab scale precipitation

Lab scale precipitation used only mAb feedstock and was executed in 500 ml or 1 L Duran bottles. The one cut precipitation was carried out at pH 8.5, 1.6 M ammonium sulphate based on the microwell precipitation results. 250 ml precipitant solution, including 200 ml 4 M ammonium sulphate, 50 ml 200 mM pH 8.0 phosphate buffer, was pre-mixed and added slowly into 250 ml mAb feedstock. The initial mAb concentration was designed to have three different concentration, 1.2 mg/ml, 1.75 mg/ml and 2.3 mg/ml respectively, in order to investigate the effect of initial concentration. The solution was mixed on a magnetic stirrer (Bibby Scientific, Dunstable, UK) with a stir bar in the bottle at 1000 rpm for 1 hour. The solution was collected for later filtration.

The two cut precipitation was firstly carried out at pH 4.0, 1.2 M ammonium sulphate. The total volume and procedure were the same as above one cut precipitation, except pH 4.0, 200 mM sodium acetate buffer used instead. The initial mAb concentration was 1.2 mg/ml. After 1 hour, the solution was filtrated through CUNO Zeta plus EXP 90 sp filter (0.2 - 0.5 μm) (3M Purification Inc., Bracknell, UK) at 15 psi. Permeate solution was collected for second step precipitation. 450 ml permeate was conditioned to pH 8.5, 1.6 M ammonium sulphate by adding 45 ml pH 8.5, 4 M ammonium sulphate and 5 ml pH 8.5 200 mM phosphate buffer. It was then mixed on a magnetic stirrer with a stir bar at 1000 rpm for 1 hour again. The final solution was used for filtration later.

2.2.3 Membrane filtration

The membrane filtration was performed on various pore sizes of filter membranes: 0.2 μm, 0.45 μm, 1.2 μm and 5.0 μm on the XX15 047 filter holder (Millipore Limited, Dundee, UK). The mAb solution from previous lab scale one cut precipitation was loaded to filters by a vacuum pump. The loaded volume was recorded for capacity measurement when transmembrane pressure increased to 20 psi. Permeates were collected for analysis to calculate product recovery and loading capacity on different pore size membranes. Then two times volume of the recorded loading volume of pH 8.5, 1.6 M ammonium sulphate was used to washing the membrane and brought out remaining solution. After washing, pH 7.0 PBS solution, which had the same volume of loading volume, was used to resuspend the protein solids on the membrane and filtered through the membrane. The resuspension was recycled through the membrane for three times in order to dissolve mAb throughout. All solutions were collected and for further analysis.

2.2.4 Depth filtration

During depth filtration, mAb solution from lab scale precipitation was used for the studies. CUNO Zeta plus EXP 05 sp filter (3M Purification Inc., Bracknell, UK), which has a large pore size range from 1.8 μm to 10.0 μm, and CUNO Zeta plus EXP 30 sp filter (0.5 μm - 2.0 μm) were coupled to form a 0.5-10 μm depth filter system. The same loading capacity test was carried out as the above membrane filtration. 100 ml precipitation solution was filtered through this depth filter system by applying 20 psi pressure in nitrogen. In the washing step 100 ml same salt concentration and pH condition solution without mAb passed through and washed the depth filters. A second washing, same as the first one flushed the remaining first washing liquids outside the filter. Then 100 ml pH 7.0 PBS solution was used to resuspend and washed out the mAb solid in the depth filters by recycling through filter for three times. An extra 50ml PBS was followed to strip all remaining mAb and other molecules from the filters. All solutions were collected for further analysis.

The second depth filtration for two cut precipitation solution was the same as that of one cut precipitation except the solution had been already filtered through a CUNO Zeta plus EXP 90 sp filter in the first cut. Two processes were shown in Figure 2.1 flowchart.

2.2.5 Depth filtration improvement

In the improved experiments, the same solution and procedures were carried out until the end of washing step. Then gradient resuspension was adopted using ammonium sulphate solutions from 1.2 mol/L to 0.7 mol/L with an interval of 0.1 mol/L at pH 8.5. During resuspension, 100 ml ammonium sulphate solution from the highest concentration to the lowest concentration passed through depth filters sequentially. Finally, 100 ml PBS was used to strip all remaining proteins from the filters.

In fraction collection, the same procedures were carried out before resuspension. 100 ml 1.1 M ammonium sulphate at pH 8.5 was continuously passed through depth filters. Each 10 ml filter through resuspended solution was collected in Falcon tubes (Falcon Plastics, Los Angeles, US). The filter through solution, washing and resuspension solutions were all collected.

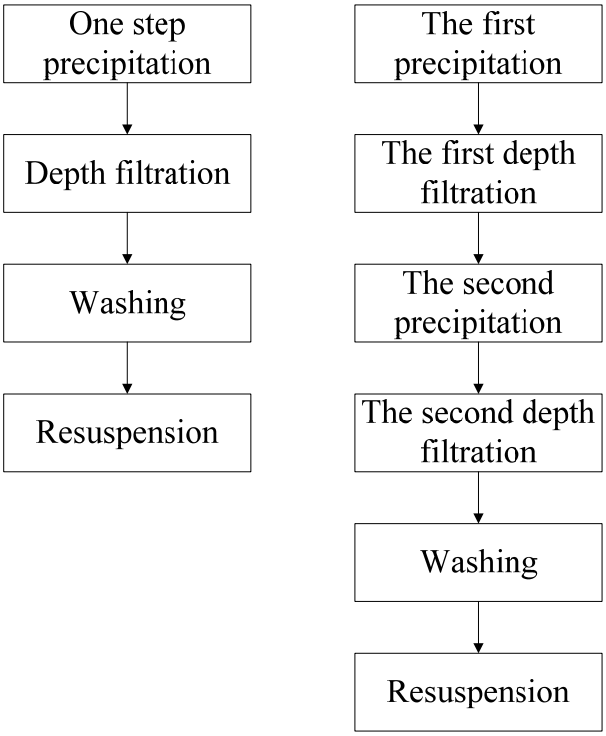


Figure 2.1 The flowcharts of two precipitation and filtration processes. Left: one cut precipitation; Right: two cut precipitation.

2.2.6 High throughput chromatography

The high throughput chromatography studies were carried out in the 96 microwell filter plate with 1.2 μm Durapore membrane (Millipore Limited, Dundee, UK) by Tecan robot (Tecan Group Ltd., Switzerland). After each incubation, mixing, washing and elution, the filterplates were centrifuged at 4000 rpm for 10 mins in an Eppendorf Centrifuge 5804R (Eppendorf, Cambridge, UK) to separate supernatant. The supernatant was collected by 96 microwell collection plates stacked beneath. The solution was analysed by HPLC and other analytical methods. All experiments were performed at room temperature.

2.2.6.1 CEX resin and binding screening

In the CEX resin screening, four CEX resins were evaluated to identify the optimal salt concentrations and pH for loading and elution. The four resins evaluated were Fractogel EMD SO^{-3} (M) (Merck KGaA, Germany), UNOsphere S (Bio-Rad, US), Capto S and SP Sepharose FF (GE Healthcare, Sweden). The Tecan system dispensed 100 μl different types of resin from 20% (v/v, in 20% (v/v) ethanol solution) stock slurries into designed microwells. For the following procedures, 250 μl solution was always added into each microwell.

Before the loading experiments, each resin was equilibrated with buffer at designed conditions. The conditions chosen were pH from 4.0 to 6.0 by 5 mM sodium citric buffer with appropriate sodium chloride concentration to form 1, 5, 10 ms/cm conductivity. The stock buffer solutions were made in pairs with one containing only 5 mM pH buffer and the other containing 1 M NaCl in the same pH buffer. The designed conditions were generated by mixing these stock solutions according to the salt concentration ratio into the microwell plates. The resin in the microwell was equilibrated by mixing the microwell plates on Eppendorf thermomixer for 15

minutes at 1200 rpm. Then supernatant was removed by centrifugation and equilibrium stage was repeated by three sequential cycles.

The mAb solution, after prepared by ammonium sulphate precipitation and resuspension, was concentrated to around 5 mg/ml and dialysed into 5 mM pH 7.0 sodium phosphate buffer by 5 kDa centrifugal concentration tubes (Millipore Limited, Dundee, UK). In loading phase, 200 µl loading buffer was first dispensed into microwell and then 50 µl mAb solution was added to obtain 1 mg/ml mAb final concentration. Then the microwell was incubated on an Eppendorf thermomixer for half an hour at 1200 rpm. The supernatant was separated by centrifuge and collected in 96-microwell collection plate.

A washing step was carried out by applying 250 µl loading buffer to each microwell with 15 minutes mixing and centrifuging supernatant for three times repeats. The elution step followed by dispensing 250 µl 1 M sodium chloride, with the same pH as loading conditions to contact resin. The plates were mixed half an hour on Eppendorf thermomixer at 1200 rpm. Supernatant was centrifuged and collected for further analysis.

2.2.6.2 HIC resin binding screening and gradient elution

In the HIC resin screening, four HIC resins were evaluated: Phenyl Sepharose 6 Fast Flow high sub, Butyl Sepharose 4 Fast Flow, Butyl-S Sepharose 6 Fast Flow and Butyl Sepharose High Performance, all from GE Healthcare, Sweden. The loading conditions screening procedures were same as CEX resins screening. The conditions chosen were pH from 6.0 to 8.0 by 20 mM phosphate buffer and ammonium sulphate at 1.0, 1.1 and 1.2 mol/L respectively. Concentrated mAb solution was conditioned to the required salt concentration at certain pH by salt stock solution and pH buffer with around 1 mg/ml mAb final concentration in loading stage.

After loading and three times washing, 300 µl elution solution was dispensed into each microwell and mixed for 15 minutes on Eppendorf thermomixer at 1200 rpm before centrifuging. The conditions of elution solutions had the decreasing ammonium sulphate concentration from 1.0 mol/L to 0.2 mol/L with an interval of 0.2 mol/L while had the same pH as the load pH. The gradient elution required five cycles with permeates collected each time. A final strip of 300 µl 10 mM phosphate buffer was used to wash out the residual molecules and complete studies. All filtered through solutions were collected and analysed.

2.2.6.3 AEX resin and flow through studies

Capto Q (GE Healthcare, Sweden) was used in anion exchange chromatography flow through mode. The conditions screened were pH from 4.0 to 7.0 by 5 mM sodium citric buffer and appropriate sodium chloride concentration to form 1, 5, 10 and 15 ms/cm conductivity. 1 mg/ml mAb solution was buffer exchanged to required loading conditions by 5 kDa centrifugal concentration tubes (Millipore Limited, Dundee, UK). After 20 µl resin in each microwell was equilibrated for 15 minutes in designed loading buffers with three cycles. 300 µl mAb solution was then dispensed into each microwell. The microwell plates were placed on Eppendorf thermomixer at 1200 rpm 15 minutes for solution-resin contact. The microwell plate was centrifuged and the supernatant was collected. The unbound mAb and impurities were analysed by HPLC and other analytical methods.

2.2.7 Lab scale chromatography

2.2.7.1 Protein A chromatography

Pre-packed 1 ml HiTrap MabSelect column (GE Healthcare, Sweden) and self-packed MabSelect column were used in protein A purification process for different scale. 20 ml MabSelect resin was packed in a XK16/20 column (GE

Healthcare, Sweden). The packing protocol followed the procedures in GE Healthcare MabSelect manual (GE Healthcare, 2011). AKTA basic was used to connect columns and control purification steps. 5 column volumes of loading buffer was applied to MabSelect column first. The conditions of loading buffer were pH 7.4, 20 mM sodium phosphate with 0.15 M sodium chloride. The antibody feedstock was buffer exchanged to pH 7.4 loading conditions and loaded to column at 1ml/min flowrate. 5 column volumes loading buffer was followed to wash the column and 1 column volume loading buffer with 10% (v/v) isopropanol was used to wash out hydrophobic contaminates. The antibody bound to resin was then eluted by by 100 mM, pH 3.5 sodium citrate buffer at 1ml/min, with immediately adding 50 µl 1 M pH 9.0 Tris buffer per 1 ml elution to neutralise and protect antibody. Fraction collection was used to collect elute with 1 ml each tube for 1 ml column and 5 ml each tube for 20 ml XK16 column. Collected elution at main peak was kept together in +4^{o}C fridge.

The column was then flushed with three column volumes of elution buffer until the UV 280 signal kept stable at UV base line. It was then equilibrated by five column volume equilibrium buffer for next sample running. For storage, 20 % (v/v) ethanol solution was used to sterilise the column and store the resin.

2.2.7.2 Gel filtration chromatography

60 ml Superdex 200 (GE Healthcare, Sweden) was packed in a XK 16/40 column (GE Healthcare, Sweden). The resin was poured into column and then distilled water was pumped through column at 1 ml/min for 2-3 hours until the levels of the bed was stable. Increased flow rate to keep pressure at 480±20 kPa for 45 minutes. Mounted the flow adaptor and adjusted to the gel bed surface (GE Healthcare, 2010). The running buffer was pH 7.0, 0.15 M sodium chloride, 20 mM sodium phosphate buffer. The gel filtration column was connected to AKTA Basic (GE Healthcare, Sweden) and equilibrated for 120 minute at 1 ml/min flowrate before each sample

run. The samples were exchanged to the same conditions as loading buffer. 0.6-1 ml each sample was injected into column by injection valve on the AKTA. The operational flowrate was 1 ml/min and total time for one sample was 90 minutes. The column regeneration and cleaning was the same as MabSelect column.

2.2.8 Analytical methods

2.2.8.1 Fab' HPLC analysis

The Fab' concentration in the samples were measured by a 1 ml protein G HiTrap column (GE Healthcare, Sweden) connected to an Agilent 1200 series HPLC system (Agilent Technologies, UK). The equilibrium and loading buffer was pH 7.4, 20 mM sodium phosphate. 100 μL sample was injected into column with an autosampler. The sequence and methods were pre-programmed in HPLC software Chemstation. The elution buffer was pH 2.5, 20 mM Glycine-HCl buffer. UV 220 nm signal was recorded and used to measure the peak area of Fab'. The Fab' concentration was calculated based on a calibration curve, which was generated by several Fab' concentration samples, range from 0 mg/ml to 1.2 mg/ml, diluted from Fab' standard. All samples were centrifuged at 1,000 rpm for 10 minutes and transferred to Agilent 96 microwell plates before loaded to HPLC. The flow rate of HPLC was kept at 2 ml/min with upper pressure 85 bar limit. Total analysis time is 16 minute with Fab' peaking at around 3 minute.

2.2.8.2 Bradford total protein assay

Bradford protein assay was used to analyse the total protein concentration in the samples. BSA standard and Bradford 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 Bradford reagent was added. Each cuvette was well mixed and covered from light for 10 minutes at room temperature (Bradford, 1976). 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 which method was used.

2.2.8.3 mAb HPLC analysis

The mAb concentration were also measured by a 1 ml protein G HiTrap column (GE Healthcare, Sweden) connected to an Agilent 1200 series HPLC system (Agilent Technologies, UK). The buffer system was the same as that of Fab'. 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, range from 0 mg/ml to 1.5 mg/ml, diluted from mAb standard. Before samples being loaded into HPLC, they were transferred to 96 microwell filtration plates and centrifuged at 1,000 rpm for 10 minutes. The permeates were transferred to Agilent 96 microwell plates for HPLC. The flow rate of HPLC was kept at 2 ml/min with upper pressure 85 bar limit. Total analysis time is 6 minute with mAb peaking at around 2.9 minute.

2.2.8.4 Impurities analysis

The overall impurities included host cell protein, cell culture media protein and other impurities. These impurities were measured by HPLC at UV 280 nm as the first flow through peak. The original feedstock, either Fab' or mAb, was used as the standard. Several samples diluted from standard were made according to different dilution rates. The Bradford total protein assay was used to measure the total protein

concentration in each diluted sample. The impurities concentration for each sample was calculated by subtracting the corresponding Fab' or mAb 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 Bradford assay.

2.2.8.5 DNA quantification

The Picogreen method was used for DNA quantification. The Picogreen kit, DNA standard and Molecular grade water were bought from Invitrogen (Invitrogen, UK). The standard was diluted into two concentration ranges: high range 1 ng/ml - 1000 ng/ml and low range 25 pg/ml - 25,000 pg/ml. Two calibration curves were made due to high or low DNA concentration range. The standards and samples were diluted by TE buffer within 100 pg/ml to 1000 ng/ml range to a final volume of 1.0 ml in tubes with all tips, tubes sterilised. 1.0 ml Picogreen reagent was added to each sample and incubated for 5 minutes at room temperature, protected from light. 200 µl of each sample was transferred to transparent microwell plate and the plate was read by Safire 2 microplate reader (Tecan Group Ltd., Switzerland) at standard fluorescein wavelenths (excitation ~480nm, emission ~520nm) (Invitrogen, 2008). DNA concentration in the samples was then calculated based on corresponding calibration curve.

2.2.8.6 Host cell protein Elisa

CHO HCP was analysed by Cygnus HCP ELISA kit 3G (Cygnus Technologies, US). Molecular grade water was bought from Invitrogen (Invitrogen, UK). All samples were diluted within 1 ng/ml to 100 ng/ml range. 100 µl anti-CHO:HRP antibody was transferred into each well and then pipetted 50µl of standards, controls and samples into each well. Covered and incubated the microwell plate at ~180rpm for 2 hours at room temperature. Washed each well with 350 µl wash solution for 4 washes.

Pipetted 100 µl TMB substrate into each well and incubated at room temperature for 30 minutes. Then pipetted 100 µl stop solution and read absorbance at 450/650nm by Safire 2 microplate reader (Tecan Group Ltd., Switzerland). Host cell protein concentration was then calculated from the CHO HCP calibration curve using standards provided in the Cygnus kit.

2.2.8.7 Aggregates, monomer and half antibody analysis

The aggregates, monomer and half antibody were analysed by a TSKgel G3000 SW_{XL} column (Tosoh Ltd, Japan) on Agilent 1100 HPLC (Agilent Technologies, UK). The molecule 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 total running 17 minute. The three peaks came out in the sequence of aggregates (7.2 minute), monomer (8.2 minute) and half antibody (10.3 minute). The UV 220nm peak area was recorded and percentages of each component were calculated by Chemstation software.

2.2.8.8 Protein structure analysis by Circular Dichroism (CD)

Protein structures were analysed by CD on AVIV CD spectrometer 400 (Biomedical Inc., USA). The machine was switched on after 15 minutes of liquid nitrogen cooling. After 1 hour of lamp warming up, software AVIV model 400 MxC was opened to set wave length and control analysis. The sample concentration was around 1 mg/ml and had at least 85 % (w/w) purity. All samples were buffer exchanged to pH 7.0, 20 mM phosphate buffer. Clean cuvettes were washed by ethanol, then ultrapure water and dried before being loaded by samples. The wavelength 200 nm to 280 nm was used to scan the secondary structure of proteins in the samples. Absorbance was recorded and plotted by software.

2.2.8.9 Protein electrophoresis by SDS-PAGE and IEF

Protein samples were also analysed and characterised by electrophoresis. Novex Tris-Glycine precast 8-16% SDS-PAGE gels, SDS-PAGE buffer kit and See Blue 2 marker were bought from Invitrogen (Invitrogen, UK). 20 µl each sample with loading buffer was loaded into gel well and run at 125 V for 90 minutes. Then the gel was stained for one hour by Coomassie blue G-250 (Bio-rad, US) after rinsing by ultrapure water three times. After stain, gel was kept in ultrapure water and a photo was taken when gel was clear.

Novex pH 3.0-10.0 IEF gels, buffer kit with IEF marker were bought from Invitrogen (Invitrogen, UK). 20 µl each sample was loaded to well and run at 100 V for 60 minutes, then 200 V for another 60 minutes (Invitrogen, 2011). The IEF gel was first fixed by 12 % (v/v) TCA for 30 minutes after eletrophorsis. Then the washing and stain step were the same as above SDS gel. A photo was taken after protein bands were shown on the gel clearly.

2.2.8.10 Measurement of conductivity

The conductivity of samples and buffers was measured by a conductivity meter (Jenway, UK) at room temperature (20°C).

2.3 Modelling and data processing methods

2.3.1 Data normalisation

In order to eliminate the errors in model parameter estimation caused by different dimensions and orders of variables, the experimental conditions of pH and salt concentration for protein precipitation, were scaled to 0-1 range by following scaling equation:

$$\lambda_{scaled} = \frac{\lambda_{real} - \lambda_L}{\lambda_U - \lambda_L}, \tag{2.1}$$

where λ_{scaled} is the scaled value, λ_{real} is the real value, λ_L is the real value at low limit, and λ_U is the real value at upper limit.

Fab' or mAb concentration in feedstock and impurities concentration were also normalised to the highest value of antibody concentration or impurities concentration. The initial concentration was not used in normalisation because there was a salting-in effect. The maximum concentration of target antibody and impurities during salting-in was regarded as the true protein concentration in the solution. In the reality, the true concentration was difficult to obtain, thus the maximum concentration in the whole set of data was used as the closest value to the upper limit concentration in equation (2.1).

2.3.2 Least squares regression

Least squares regression was used to determine the parameters and the objective of this method was to find a set of parameters to maximise R^2 in the following equations:

$$R^2 = 1 - \frac{SS_{err}}{SS_{tot}}, \tag{2.2}$$

$$SS_{err} = \sum_{i=1}^{n} (y_{\exp} - y_{\text{model}})^2, \tag{2.3}$$

$$SS_{tot} = \sum_{i=1}^{n} (y_{\exp} - \bar{y}_{\exp})^2, \tag{2.4}$$

where $y_{\exp}$ is the experimental value, y_{model} is the value calculated by model, $\bar{y}_{\exp}$ is the average of all experimental values.

Model parameters were estimated by 'lsqcurvefit' function with initial coefficient provided in matrix in MatLab (The MathWorks Inc., R2009b). The quality of parameter estimation was assessed by equation (2.2). The higher R^2, the better the parameter estimation was. When R^2 was larger than 0.90, the model regression would be acceptable and the model was considered as good quality.

2.3.3 Validation and statistic tests

9 DoE experiments under the same operation space were carried out to validate the model. When validating bioprocess models, it is not recommended to use error percentage to evaluate models because the range of bioprocess data may be very wide even after scaled or transformed, which will introduce mathematical error. Thus, statistical tests should be utilised to validate new model, no matter how good is the fitting of the data was in the regression step.

However, there are several unusual problems for bioprocess model validation. First of all, the number of samples used for validation is normally small e.g. 9 samples in this case, due to various reasons, such as high cost and long time of generating data. Secondly, the distribution of most bioprocess data is normally unknown or the data is hardly transformed to any known distribution, e.g. standard normal distribution (Goffaux and Wouwer, 2005). Statistically, the normal distribution can be assumed only when the number of samples is very large, normally more than 30 (Lamprecht, 2005). Therefore, for small validation group with unknown distribution, it is of great risk to use paired t-test due to high probability to fail.

There exist two solutions for this type of validation. One solution is to use Wilcoxon signed-rank test (Wilcoxon, 1945) for few samples in paired data. The other is to analyse validation samples together with previous regression data by paired t-test since the whole data set can be roughly considered as the normal distribution when sample number is large than 30. For Wilcoxon signed-rank test, 2-tailed significance

> 0.05 can be regarded as validation passed. For paired t-test, sig. > 0.05 can be considered as the null hypothesis that there is no difference between experimental data and model calculated value is accepted (Schiff and D'Agostino, 1996).

Chapter 3. Model based process design for bioprocess optimisation: methodology

3.1 Introduction

Biopharmaceutical development is becoming more time consuming and expensive due to the difficulties in bioprocess design and optimisation (Karlsson et al., 2004). Therefore, a systematic approach should be proposed to effectively tackle these difficulties and achieve optimised design efficiently. With the current techniques available, this promising approach should combine bioprocess modelling, model based experimental design incorporated with high throughput technology, under the framework of process optimisation to give a breakthrough for efficient bioprocess development.

The microwell-scale method can speed up the acquisition of process information, but alone it is rarely adequate to predict the overall production performance where a model will help (Titchener-Hooker et al., 2001). When bioprocess models exist but the material complexity precludes determination of key parameters, a model based experimental design with microwell experimentation can provide experimental data sufficiently and quickly covering the conditions needed (Nfor et al., 2009; Ziegel, 2003). Automation and the application of advanced computer software also make the iteration between experimental data and data analysis possible, and encourage the innovative model based experimental design by sequential feedback, which will significantly accelerate the experiments (Charaniya et al., 2008; Chernoff, 1959; Massimo et al., 1992). The interactions between these methods will bring mutual benefits and achieve better effects than the sum of effects of each method applied alone.

In this chapter, a systematic model based process design methodology for downstream bioprocess optimisation, integrating process modelling, model based experimental design and high throughput experimentation will be proposed and developed to enable faster bioprocess development and optimal process identification. The aims are to provide an efficient design methodology for bioprocess optimisation based on process model with minimum cost, to achieve a good understanding of process and to obtain detailed information to support QbD by generating accurate models in the end of design.

3.2 Integrated model based process design methodology development

3.2.1 Strategy and method overview

In this proposed integrated design, the interaction between three methods is the core strategy and decides the success of this methodology, shown in Figure 3.1. In this core strategy, each individual method receives information from other method while also gives out its results to the next method as a mutual benefit loop from the start of objectives input until the completion of optimisation. The information update and communication among different methods promotes information sharing and better process understanding.

Model based experimental design with an optimisation tool is a method that uses a theoretical or empirical model and guided by process performance objectives during design in order to achieve efficient and robust optimisation for an unexplored bioprocess unit operation (Galvanauskas et al., 1997). Thus this model based approach can be regarded as an optimisation procedure, which starts with restricted knowledge and aims to find the optimal process in a fast way.

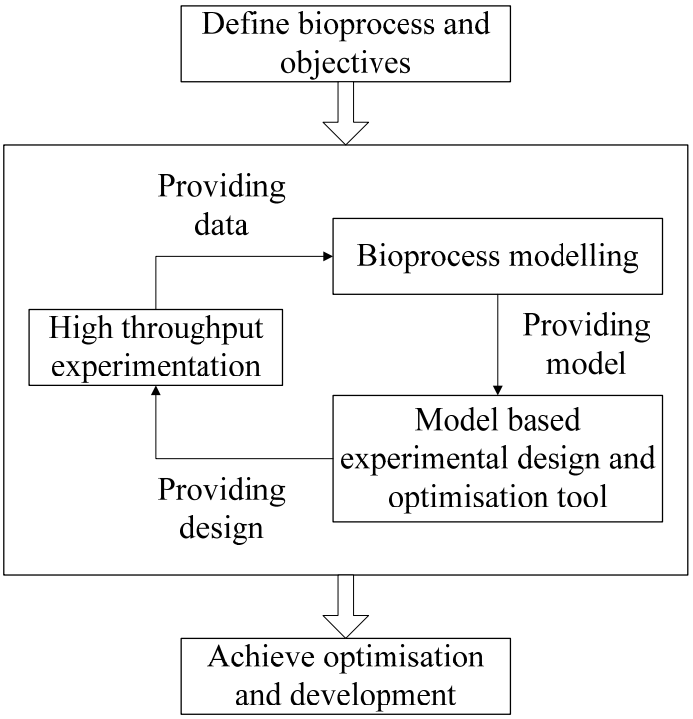


Figure 3.1 The core strategy of an integrated model based process design and optimisation approach.

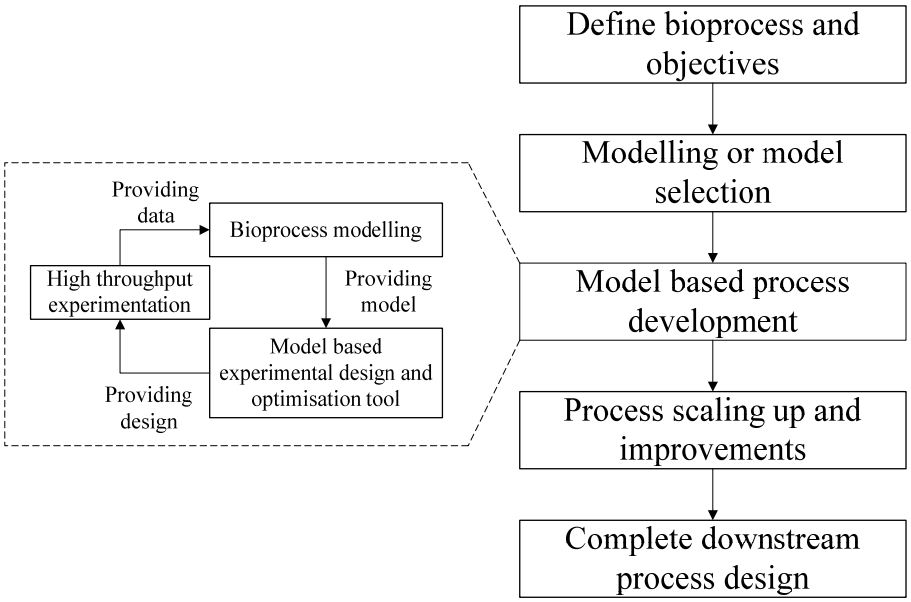


Figure 3.2 The whole flowchart for an integrated model based process design and optimisation approach.

Inspired by the flowchart in Figure 1.4, a whole process for this integrated model based process design methodology is developed and proposed, shown in Figure 3.2. It starts with choosing target bioprocess, through two crucial development stages in microwell plates utilising the core strategy and then scaling up to complete the process design. It can be regarded as a systematic approach for any bioprocess development starting with microwell scale and support scaling up by providing information. When polynomial equations are used, the design approach is the same as the commonly adopted DoE process design, while it can also be customised by using self defined models. The detailed steps are explained in following paragraphs.

3.2.2 Define bioprocess and objectives

First of all, the target bioprocess should be defined and ensure the following procedures in the design can be successfully executed. Each process development has one or more objectives. Therefore, it is quite important to define the objectives and tradeoffs between objectives considering some factors such as cost, model precision and time before the method starts. The objective may be described as a quantitative function in the most cases, but the other forms also exist. Since it is process performance oriented, this method should be very effective and efficiently.

3.2.3 Modelling or model selection

Select or develop a quantitative model based on a priori knowledge or previous literature data. The initial model can be a mechanistic or empirical model with some adjustable coefficients. For the modelling based on microwell experimentation, the procedures are nearly the same as those in Figure 1.2, except the experiments are carried out in microwell plates. High throughput experimentation is able to provide sufficient data for modelling, which is the most significant benefit of adopting this experimental platform. The detailed modelling procedure and flowchart will be explained in Chapter 4 with a modelling case study as a sample.

At this stage, if there are many established models available, it is not necessary to carry out modelling. However, the constraints and uses of all models should be assessed and project objectives will be the primary criteria. Moreover, the model structure may be modified according to current experiment facilities and industrial interests. Great attention must be paid to the statistical tests for model, because the quality of model selected in this step influences the following procedures.

3.2.4 Model based process development

In this step, it starts from a chosen model or a model developed, which contains some priori knowledge in its mathematical structure and initial parameters. However, they are definitely not accurate at the beginning and need updating all the time through optimisation. Thus a mathematical algorithm based on this model will be adopted to design the next set of experiments to collect new information and update the model with material specific experimental data. As the model becomes more accurate, the optimisation will be completed once the defined objective achieved, e.g. obtaining high model regression confidence. For the sake of time and effects, sequential design will be the most efficient method (DeGroot, 1962; Robbins, 1952). Therefore, a sequential model based experimental design is adopted and illustrated in Figure 3.3.

After the process design objectives are defined, the design and optimisation process starts with a chosen model and the best initial parameter guess. In most case, the best initial parameters are hardly available when facing a brand new process or a new biological feedstock. In order to obtain general information of the process, a set of initial DoE experiments covering all design space is recommended. The data from these preliminary experiments will quickly update the priori knowledge with accuracy to certain extent. Then the design procedure enters into a sequential design loop.

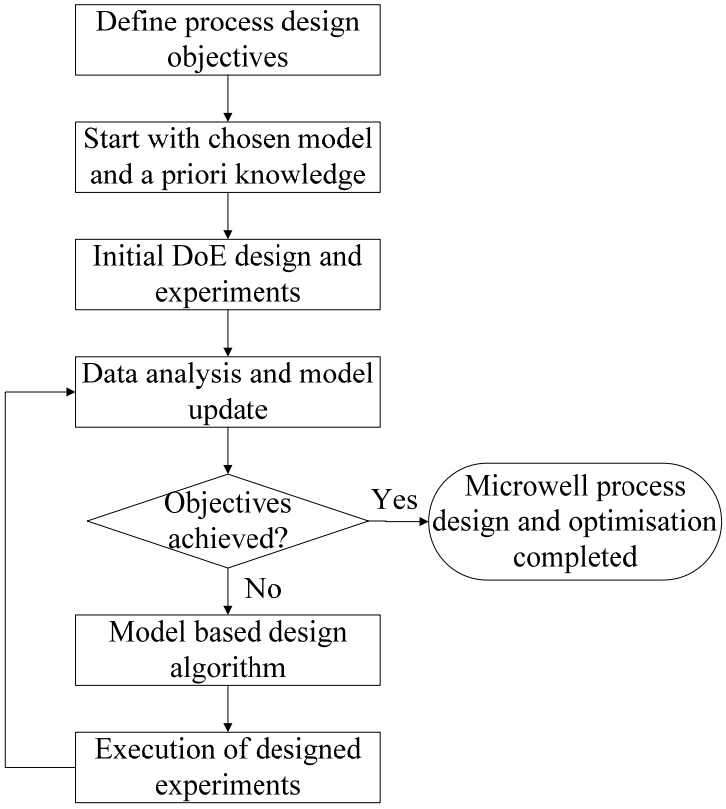


Figure 3.3 Sequential model based experimental design flowchart.

At the beginning of each loop, a set of experimental design is given by model based design algorithm based on currently best knowledge. Following experiments will be carried out according to design and data is feedback into the loop by analysis and model updating. Only a small number of information rich experiments selected by algorithm will be executed during each loop and once the objectives are achieved, the procedure terminates. Statistic tests will be used to judge whether model is accurate or not as one of the primary objectives in the algorithm. The effectiveness and efficiency of this approach are highly depends on the model based algorithm, which is illustrated in details as following.

3.2.4.1 D-optimal experimental design

In the model based methodology, the accuracy of model is the most important because it will give the insights of process and also facilitate optimisation. The accuracy highly depends on the model parameters, the generalised variance of which should be minimised. It is the reason why D-optimal experimental design is adopted as the core algorithm in this approach.

Suppose there is a mathematical model describing a bioprocess with following equation:

$$y = f(x, \beta) \tag{3.1}$$

where y is a $n \times 1$ matrix of observations, x is a $n \times m$ matrix of conditions, β is k coefficients to be estimated.

If equation (3.1) is expanded by a Taylor's series (Abramowitz and Stegun, 1970) at initial parameter β_0, it is then transformed to:

$$y = f(x, \beta_0) + \sum_{i=1}^{k} \left. \frac{\partial f(x, \beta)}{\partial \beta_i} \right|_{\beta=\beta_0} (\beta_i - \beta_{i0}). \tag{3.2}$$

Now let

$$Z = y - f(x, \beta_0), \tag{3.3}$$

$$X = \left\{ \begin{matrix} \frac{\partial f(x_1, \beta)}{\partial \beta_1} \cdots \frac{\partial f(x_1, \beta)}{\partial \beta_i} \\ \vdots \\ \frac{\partial f(x_n, \beta)}{\partial \beta_1} \cdots \frac{\partial f(x_n, \beta)}{\partial \beta_i} \end{matrix} \right\}_{\beta = \beta_0}, \tag{3.4}$$

$$B = \left\{ \begin{matrix} \beta_1 - \beta_{10} \\ \vdots \\ \beta_i - \beta_{i0} \end{matrix} \right\}, \tag{3.5}$$

equation (3.2) becomes

$$Z = XB. \tag{3.6}$$

It is well known that the best estimated parameters $\hat{\beta}$ will be obtained through the least squares estimation by minimising squares of Z (Atkinson and Hunter, 1968; Box and Draper, 1971; Mitchell, 1974; Papalambros, 2000).The variance-covariance matrix of $\hat{\beta}$ is expressed as:

$$\text{var}(\hat{\beta}) = (X'X)^{-1}\sigma^2 \tag{3.7}$$

where σ is the standard deviation of the experimental error and regarded as a normal distribution with zero mean value. It is inherited from the specific experimental system and usually is unknown but constant.

A correct and accurate model must have a very small value of equation (3.7), which means a big value of the determinant $|X'X|$ from mathematical calculation (Atkinson and Hunter, 1968; Box and Draper, 1971). Hence, D-optimal design was developed to choose some experimental conditions, which can maximise $|X'X|$. Since the invention of D-optimal design, many researchers had published several algorithms to select conditions (Cook and Bachtsheim, 1980). The first design algorithm (Fedorov, 1969; Kiefer, 1971) only managed to compute approximate D-optimal design due to

the poor computer ability at that time. Later, Fedorov (1972), Mitchell and Miller (1970) and Wynn (1972) improved the approximate design by exchange algorithms, which began with a set of *n* points non-singular design and exchange the one design point in the $|X'X|$ each time in order to achieve its maximum with fixed *n* point number. Mitchell (1974) further presented a multi-points exchange algorithm for constructing 'D-optimal' designs under the fixed number of experiments in 1974.

However, as the sequential experimental design will be adopted, the number of experiments is no longer fixed and increases in each loop. The D-optimal design utilised in the method is constructed based on that of Mitchell's and then modified to enable the number of design points to increase, as shown in Figure 3.4.

The algorithm starts the design in a sequential loop with *n* existing conditions. The first design to initialise the loop is from the DoE design in Figure 3.3. For a valid D-optimal design, the value of *n* should be at least one more than the degree of freedom, e.g. for a two variables model, $n \geq 3$. The following sub loop is used to search a set of experimental designs, which increases the new $|X'X|$ most. The value of *k* depends on the designer's decision on how many new experiments to be carried out in the next run. It should be an integer number at least equal one, and can be adjusted to different number in each loop according to detailed experimentation plan.

Nevertheless, the computation of $|X'X|$ is very complicated from a mathematical view because the model used is normally a multi-variable, nonlinear equation. Searching *k* points to maximise $|X'X|$ itself is a mathematical optimisation process, the complexity and computation time of which significantly increases with the number of *k* and the nonlinear extent of the model increase. Due to the practical consideration for experimentation, this computation should be as simple as possible to make waiting time short during each loop. Therefore, in this study, *k* is fixed and

equals one in each loop with the extra consideration of keeping the total experiment number as small as possible. In order to avoid matrix singularity, no repeat condition will be selected by algorithm. It is very reasonable even from the viewpoint of experimentation, that only new conditions will be carried out and analysed rather than repeating when facing limit material and time. Following Figure 3.4, at the end of each loop, *k* conditions will be output for the next experiments until the objectives in Figure 3.3 achieved.

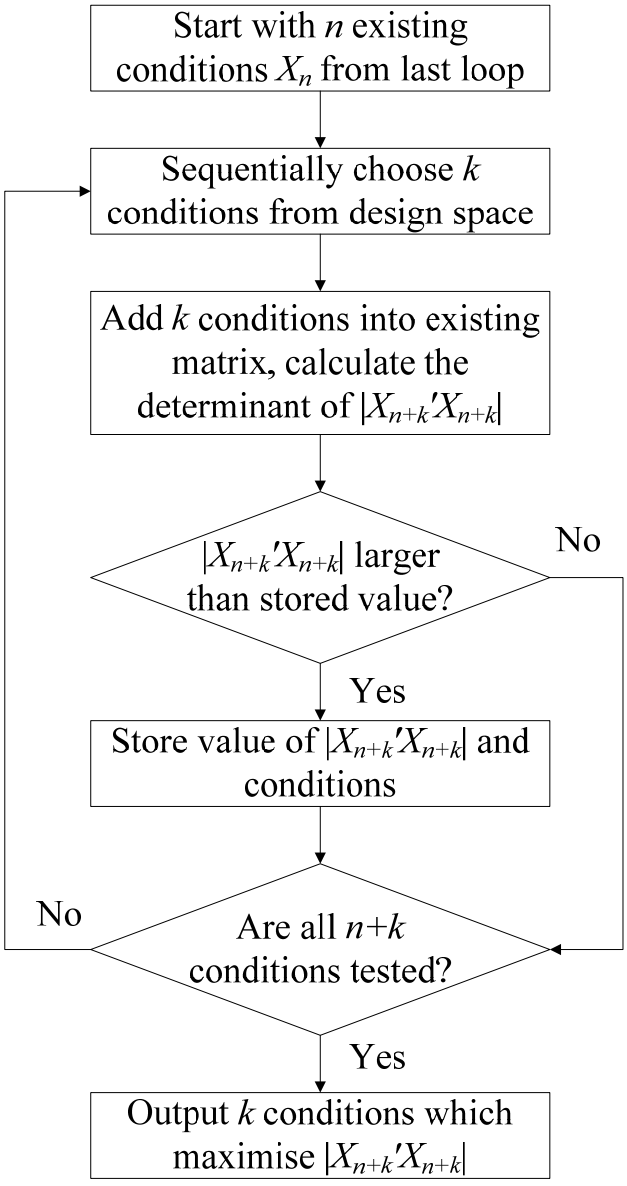


Figure 3.4 D-optimal experimental design flowchart for sequential design with k conditions chosen each time.

3.2.4.2 Optimisation and random design

The main purpose of D-optimal design is to correct the model parameters by elaborated experimental design and update the model with real results. For the purpose of process optimisation, it is a slow and indirect approach since the D-optimal algorithm is only responsible for finding informative rich points rather than the optimal point for process.

Therefore, mathematical optimisation based on the model, which is highly objective orientated, is adopted in a parallel design form alongside with D-optimal design. In each experimental loop, the experimental conditions, which give the maximum objective value according to the current updated model, will be selected as an extra point into the next set of experiments. For some complex models, predicted local maximum points can also be chosen for the next experiments together with the above global maximum point in order to avoid local optimisation (Mitchell, 1974).

In this case, each loop, only one point which has maximum model predicting objective value is output by algorithm. This optimisation is straight forward because the predicted maximum point can be quickly pointed out and easily validated in the next run. However, two difficulties exist: (1). There are various operation limits in the real bioprocess, for example, the sensitivity of pH meter is 0.1 pH value. Hence, the conditions based on mathematical solution need to be rounded to the nearest available experimental point. It will not like numerical solution that always gives different values each time. It may continuously give one repeat maximum point when the model parameters are approaching accuracy after only one or two runs of D-optimal design. (2). It has the same problem as many other optimisation algorithms that it may be trapped in a local optimisation point (Rodriguez-Fernandez et al., 2006; Sacks et al., 1989) and needs an effective mechanism to jump out.

Therefore, a random design is introduced to solve above problems. Once the maximum point predicted by model repeats within existing conditions, at which the experiments have already been carried out, or repeats with the point selected by other algorithms in the same loop, a random point is selected within unexplored design space instead of maximum condition point. The random point will be generated by a pseudo-random number generator in the MatLab software (Karnopp, 1963; Salcedo et al., 1990) and then be scaled and transformed to output corresponding conditions. It not only solves the repeating maximum condition point problem, which may cause the algorithm into a dead loop before achieving real optimisation, but also provides the extra information for global optimisation (Lee et al., 1999). Although the total number of experiments will slightly increase, it is a recommended tradeoff for better global optimisation in a nonlinear system.

3.2.4.3 Simplex

Simplex algorithm is not a model based method and only relies on real experimental results (Morgan et al., 1990) to iteratively drive the experimentation by the objective function to find the optimal point. The efficiency and accuracy of Simplex highly depend on the process it is investigating and the choice of the starting points (Chhatre et al., 2011). In this study, Simplex is modified and incorporated parallel with model based design based on following considerations: (1). The existing conditions pool can be reused by Simplex in parallel design form without increasing the workload in real experiments. It only slightly increases the total computation time by seconds, which can be completely ignored if compared to real time-cost experiments. (2). It is a direct search method and may extract information, which may complement the experiments from model based algorithm, e.g. direct comparison of real results or the trend of objective value. The information itself will also be a validation evidence for model based design. (3). Combining two different algorithms together using the same sample pool can achieve mutual benefits and accelerates each other. Inspired by mutation principle in genetic algorithm (Goldberg,

1989), the information exchange between two algorithms effectively avoid local maximum by preventing designed experiments becoming too similar or even repeating under one algorithm. If the information brought by other algorithm is richer than the next point designed by current algorithm, the design procedure will leap and speed up by using information rich point instead.

The sequential modified Simplex algorithm was used in this study (Morgan et al., 1990; Nelder and Mead, 1965). For a process with *n* variables, the Simplex needs *n+1* points to run and decide the next point according to its rules. Figure 3.5 shows the illustration of Simplex algorithm based on a two variables design, which needs three points to form a triangle (Morgan et al., 1990). The initial set of three points will be chosen from the DoE design. Three maximum points will be chosen so Simplex will be started at relatively high value level, reducing the time to optimal point as much as possible. After sorting points by value, the point will be either the best (B), worst (W), or next to worst (N) (Nelder and Mead, 1965). The next point (R) is defined as the reflection away from W across the BN line. After the value of R was evaluated, one of following routes can be taken, shown in Figure 3.5 (Chhatre et al., 2011; Morgan et al., 1990):

1. If R>B, an expansion occurs to point E, half length of WR away from R in the WR direction. If E>R, next three points are EBN. If not, RBN will be chosen.
2. If N<R<B, BRN.
3. If W<R<N, a contraction will take place to point C_R, one quarter length of WR away from R in the RW direction.
4. If R<W, a internal contraction will give point C_w, one quarter length of WR away from W in the WR direction.
5. If the new contraction point C_R is better than R or C_w is better than W, it will replace W in the new triangle.
6. If no point can be found to be better than W in any case, R' will be reflected away from N across the BW line instead.

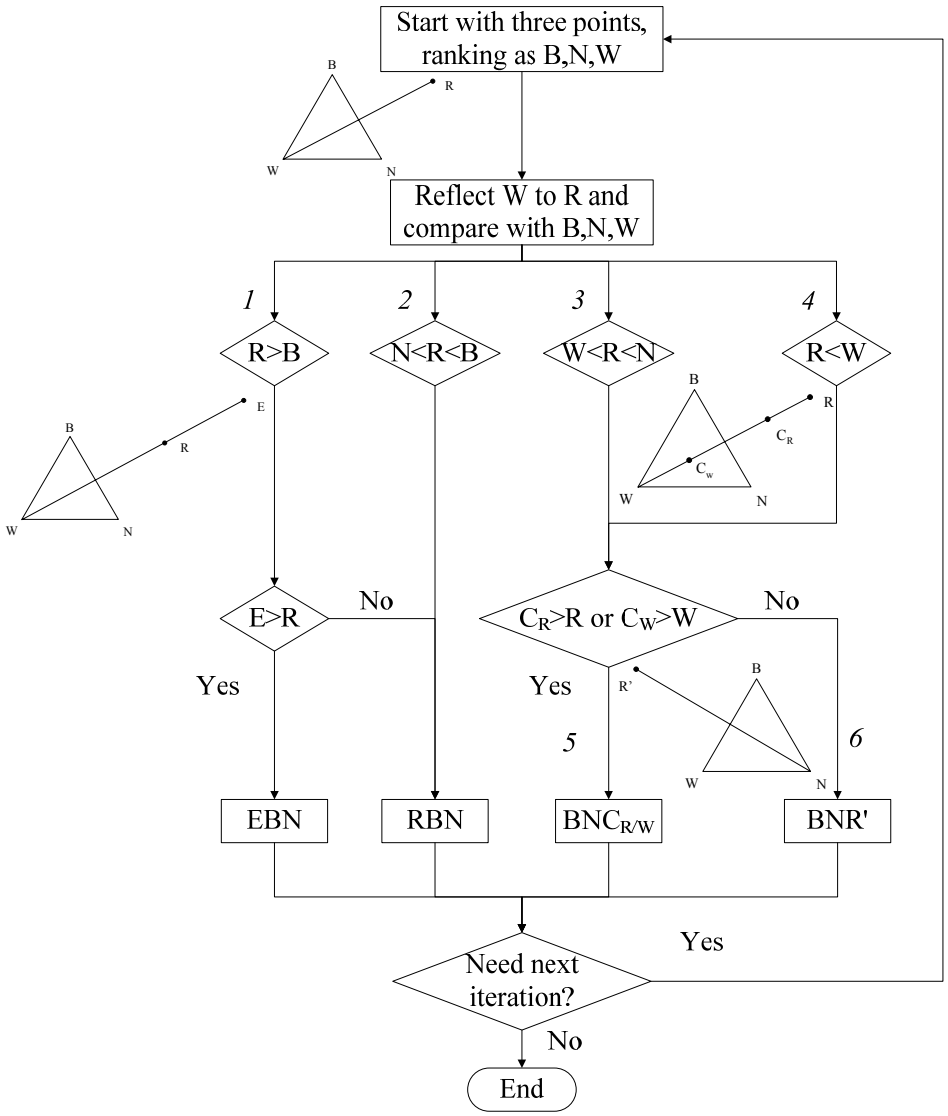


Figure 3.5 Flowchart of Simplex algorithm for 2 variables.

The new points selected will enter the next iteration of Simplex until optimisation is achieved. Since the algorithm is not case-specific and not model based, it can be used as the supplement design method to D-optimal algorithm to identify a good operating location.

3.2.4.4 Information exchange and algorithm integration

Since there are three different algorithms, how to share information and integrate them to form a core model based algorithm is the crucial step of this design methodology. For D-optimal design and optimisation/random design, incorporating with Simplex and exchanging information are relatively easy. Including the results of all points designed by Simplex when updating model will pass extra information to these design algorithms through updated model. Although the points selected by Simplex in most case are not 'D-optimal', the mathematical fundaments of D-optimal design and optimisation will not be influenced or changed by putting these points into $|X'X|$.

However, it is a huge challenge for Simplex modification. The normal sequential Simplex starts with *n+1* points (Morgan et al., 1990; Nelder and Mead, 1965), where *n* is number of variables, while in this case, it deals with a pool containing too many points. The original Simplex progresses by using points designed by itself only all the time and lacks a mechanism to exchange information with other non Simplex designed points. Therefore, the Simplex algorithm is modified to be able to exchange the points given by other algorithms, as shown in Figure 3.6.

The three points chosen for each Simplex loop are from an existing point pool, for example, in this study, the first three points are selected from the initial DoE set. It also avoids the point selection problem in the traditional Simplex, such as starting points are too close to each other or degenerated. During each loop, the criteria are to

select three points with the largest objective values. Choosing the largest value points helps to facilitate optimisation by reducing the iterations during hill climbing compared to a start with very low value points. If the starting points are linear, the fourth largest point in the pool will replace the third one to guarantee Simplex beginning with a triangle.

In each loop, the information exchange happens by deciding which three points will be used in current Simplex design from the existing point pool or last Simplex loop. The main selection method created (In the flowchart, system variable number $n=2$) is that at the beginning of each loop, the three points $\{B, N, W\}$ from last Simplex run will compare with three largest points $\{X_1, X_2, X_3\}$ from existing point pool. There are three pathways as following:

1. If they are same, which means the points from other algorithms are not large enough to change Simplex pathway, therefore, Simplex will run with $\{B, N, W\}$.
2. If they are not same and at least one point in $\{X\}$ has larger value than the second best point N in the original simplex points $\{S\}$, the information from other algorithm is thought to be valuable enough to terminate current Simplex loop and to start a new Simplex search by using these new points. A new set of three maximum points will be selected from a combining points group $\{S\}U\{X\}$.
3. If they are not same but no point in $\{X\}$ is larger than the value of point N, the three points will not be replaced and the procedure still follows pathway 1.

After the points are exchanged in above pathways, these selected points will still follow the conventional Simplex in Figure 3.5 (Morgan et al., 1990; Nelder and Mead, 1965). However, one thing worth noting is that the new design point Y from this Simplex loop may already existed in the point pool, which means the conditions were already explored by other algorithms before. In this case, the value will automatically evaluated by software and feedback to Simplex algorithm to form iterations until a non repeat point designed, shown in the bottom part of Figure 3.6.

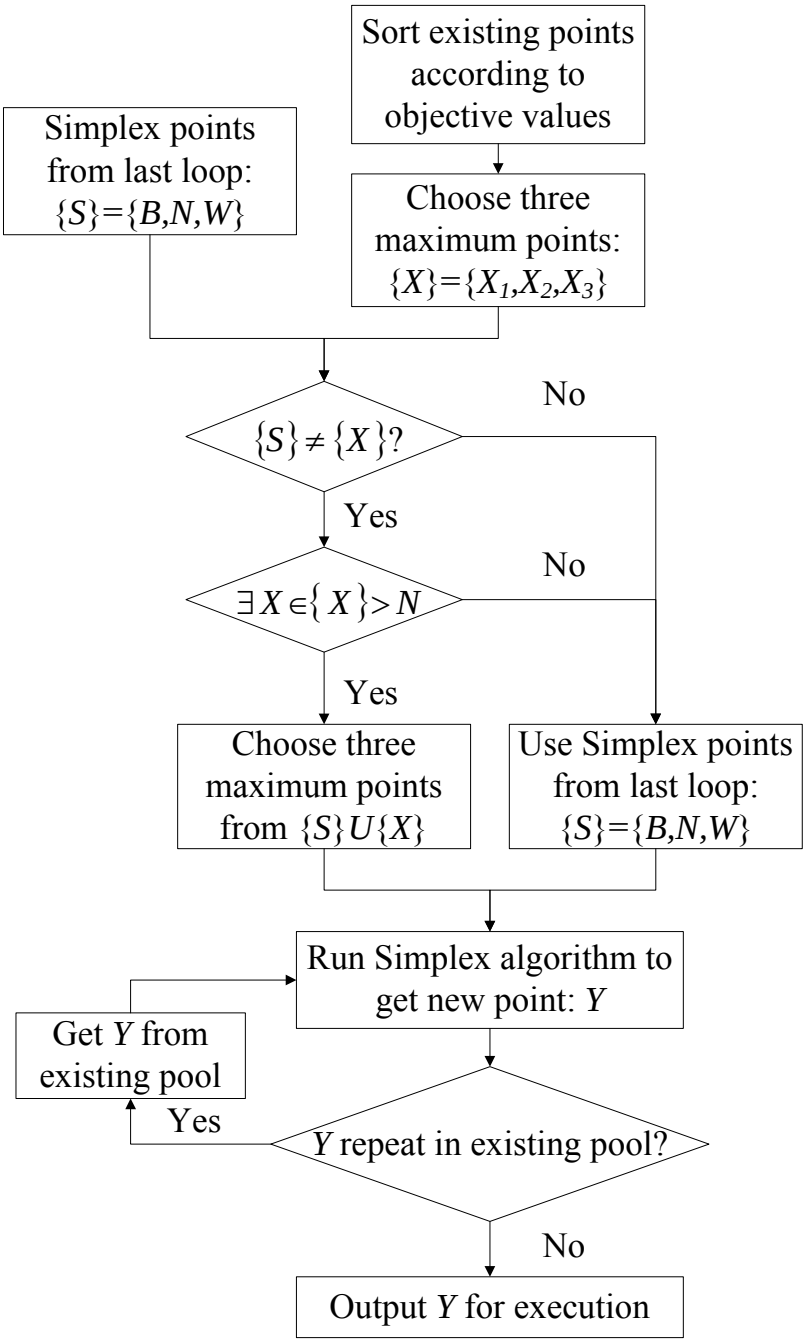


Figure 3.6 Modified Simplex experimental design flowchart with points exchange at each loop.

By this selection method, the points from previous two algorithms: D-optimal and optimisation/random design are able to enter into Simplex loop and externally change the direction of Simplex search. If the objective values of those points are smaller than current points used in Simplex, the Simplex will not be interrupted and continues as the results of Simplex are currently superior to other parallel algorithms. Once the point found by D-optimal design, optimisation or random design having larger value than the second best point in current Simplex, the worst point in the Simplex will be abandoned and replaced by this better point, which speeds up the approach and avoid local maximum.

3.2.4.5 Step summary

If linking all algorithms together, the full flowchart is illustrated in Figure 3.7. The initial DoE provides the first design for the following algorithms. Based on the DoE experimental data, the algorithm is divided into two different pathways. One is based on the direct search Simplex, the other is based on the mathematical model. The model based branch also utilised two different design methods. The D-optimal is used to make the model parameters more accurate and the optimisation is used to optimise and validate the model within each loop. When the algorithm is not working due to some repeat points, the alternative random design is used to minimise the risk of sticking to the local maximum point due to model or the improper initial guess. On the other hand, the modified simplex will select points from a pool of experimental results offered by D-optimal design, optimisation / random design and Simplex itself. This makes Simplex more efficient by jumping from low value direct to high objective value point, which eliminates several iterations in the normal Simplex. Three design methods are based on different theories but using the same dataset. This kind of parallel designs will extract and utilise the most information from experimental data while still work in an overall sequential framework.

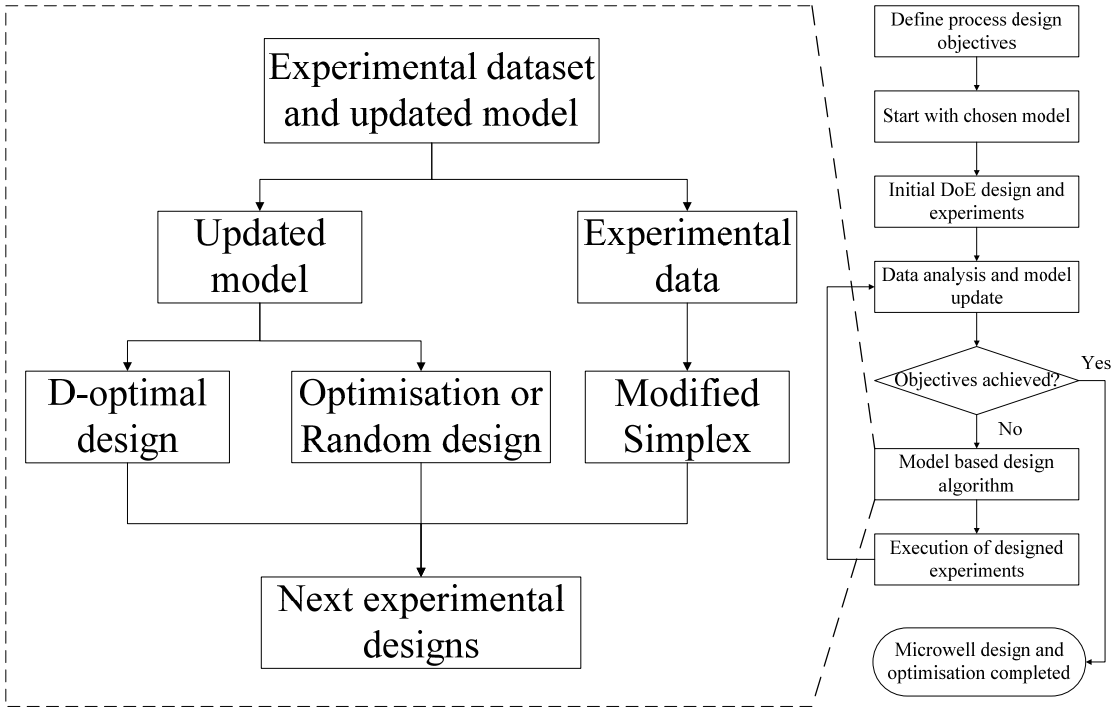


Figure 3.7 Model based experimental design flowchart with all three algorithms.

3.2.5 Larger scale process and improvements

After the optimal operation conditions found in microwell experiments, it should be tested and improved at larger scale. In this step, the differences between microwell scale and larger scale are mainly focused on, e.g. technical feasibility. If it is possible, the whole purification processes should be evaluated by current data, models and process design software based on the cost and process time as soon as possible. These issues will be explained in details in Chapter 6 and Chapter 7 respectively with flowcharts and applications in a mAb purification process development.

3.2.6 Results assessed by objectives

If the objectives have been achieved by the main algorithm, this process design is completed. The design will give both the improved model and the best experimental results, or maybe even an improved experimental system with new equipments and materials if an optional loop for equipment improvement was executed.

3.3 MatLab programming

The algorithms were then programmed in MatLab (The MathWorks Inc., R2009b) by one main design algorithm and 11 sub-function files, as shown in Appendix 1. The initial coefficients, conditions and results of existing points are required to input into the program to initialise the algorithm. The program will automatically display the conditions of points for the next experimental set with other information, e.g. current maximum objective value and SD^2 in each loop after computation and wait until new experimental data are input to start the next loop.

The current interface requires the response from researchers. However, the MatLab codes can be modified to output conditions to Tecan software in order to implement high throughput experimentation automatically and receive results in excel files

directly from analytical instruments, such as HPLC. Therefore, full automatic process development according to preset algorithms without researchers on site can be achieved in the future if hardware and software can be integrated successfully.

3.4 Conclusions

This model based process design approach has been developed to reduce both the cost and time during process development. The core strategy is to integrate bioprocess modelling, model based experimental design and high throughput experimentation together to achieve fast process development that directly addresses the data analysis challenges on microwell experimentation. It iteratively designs the experiments, carries out process modelling with updated data to improve the accuracy of process model and promotes process optimisation till the final solution is found.

The model based experimental design and optimisation was the main step in the process design approach. The conventional D-optimal design was modified by considering the practical experimental constraints. A feasible sequential D-optimal design method was developed to make the complicated matrix determinant optimisation manageable. Optimisation was introduced into the algorithm to find optimal solution as well as provide extra useful experimental points. Random design was also incorporated to prevent dead loop and local optimisation. Simplex was modified to select the most useful points from a point pool before running into the next design directly. Most importantly, a complicated information sharing method integrated these different algorithms together. The method is able to find information rich and optimal points based on model and steer away from the local optimal solution through non model based algorithms.

The model based process design method thus provided a framework to link high throughput experimentation with effective data analysis in an integrated fashion. It

has the potential to achieve fast development for a complex bioprocess with fewer experiments but better process understanding.

Chapter 4. Bioprocess modelling: a case study in Fab' precipitation facilitated by high throughput microwell scale experimentation

4.1 Introduction

From Chapter 3, process modelling is one of the key elements in the model based process design method. Establishing or selecting a model structure is required before the design algorithm and optimisation can be applied. The accuracy of the model structure will influence the effectiveness of the method. However, bioprocess modelling faces several issues such as biological complexity, the scarcity of data etc, which impose great challenges to traditional modelling (Afima, 1982; Riel, 2006). A systematic modelling approach is needed to achieve system level understanding of complex bioprocesses based on mathematical models (Bailey, 1998; Marquardt, 1996). Therefore, in this chapter, a bioprocess modelling and implementation method by high throughput experimentation will be demonstrated by using Fab' precipitation as an example. 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. The developed model will also be used to examine the effectiveness of model based design algorithm in the next chapter.

4.2 Bioprocess modelling in protein precipitation

4.2.1 Precipitation background

Protein separation using differences of protein solubility to precipitate components is a technique widely used in the biotechnology industry due to ease of operation and

clean products in the later refining stages (Thommes and Etzel, 2007). As ammonium sulphate does not denature protein and has very high salting out effect, it is extensively used to separate protein from complex solutions (Cheng et al., 2006; Cohn et al., 1946; Foster et al., 1976). Currently, it is very attractive that alternative processes can be used to replace 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 further expensive and complicated chromatography purification processes.

In the aqueous solution, protein molecules are considered as charged poly-ions. During precipitation, the solubility of protein depends primarily on process conditions including pH, salt concentration and temperature (Knevelman et al., 2010). In order to optimise the precipitation process operation, a good understanding of the impact of these conditions on the behaviour of protein is needed. For industrial scale process engineering and design purpose, a protein precipitation model that directly links protein solubility with critical operating conditions is beneficial. This can support industrial process development and further quality control, e.g. scale-up, predict process optimal conditions and provide information for online process control (Kell and Sonnleitner, 1995).

4.2.2 Previous protein precipitation modelling

The first attempt to model the protein solubility was by Cohn (1925). His log-linear equation gives a very simple but general relationship between the soluble protein concentration and ionic strength in the solution, but Cohn's equation only accurately describes the protein behaviour over a high salt concentration range. Melander and Horvath (1977) then improved Cohn's empirical equation by linking hydrophobic effect with thermodynamical parameters such as hydrophobic surface and tension.

However, their work focused on thermodynamic theory and the improved model shed little light on the bioprocess operation and design. Agena et al. (2000) proposed a thermodynamic approach based on the UNIQUAC model to describe protein solubility by protein activity coefficients. Ruppert et al. (2008) went further in modelling protein behaviour and used the polynomial relationship proposed by Winzor et al. (2001) between protein activity coefficients and osmotic second virial coefficients. Both models required protein activity data which had to be gained through extra experiments. Chiew et al. (1995) and Kuehner et al. (1996) proposed theoretical thermodynamic equations to predict protein solubility with molecule radius and surface parameters. These models work quite well in a defined simple system with all physical properties known such as lysozyme-salt solution.

However, problems arise when applying these above thermodynamical-based models to real bioprocess. For example, fermentation broth is complex and contains not only target protein, but also various impurity components such as DNA, host cell proteins, whose thermodynamical information is not available. The models based on thermodynamics are thus limited in the use of process design and control as the parameters representing thermodynamical properties for complex processing material are unknown.

Niktari et al. (1990) and later, Habib et al. (2000) proposed modified empirical exponential models to describe the traditional sigmoid shape of the precipitation curve. The empirical equations link predictions with process conditions directly. Nevertheless, these models have a disadvantage that they lack the fundamental understanding. In addition, although it has been reported that pH can make a strong impact on protein precipitation, such an impact has not been considered in the models. Temperature is another variable that influences protein precipitation. However, most proteins are sensitive to temperature, so a fixed temperature will be applied during industrial precipitation process, typically at low temperature to prevent protein denature.

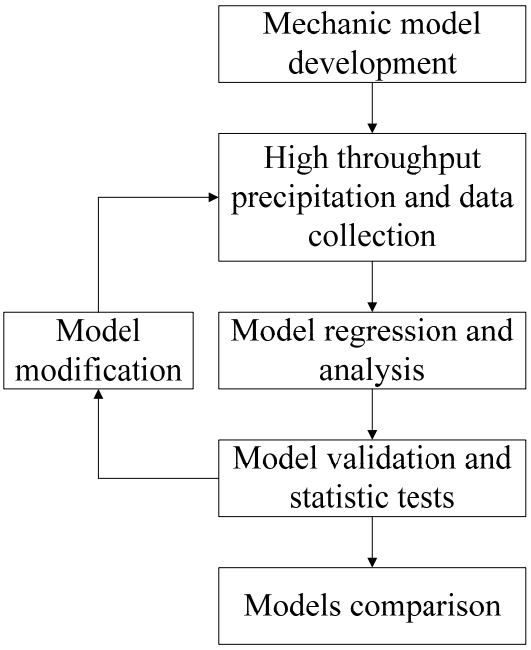


Figure 4.1 The flowchart of the modelling approach by high throughput experimentation.

4.3 Precipitation modelling objectives and flowchart

This modelling research is to propose and validate a protein precipitation model that uses bioprocess conditions as inputs and predicts the protein solubility for complex multi-components materials such as a clarified protein solution after harvest step. Figure 4.1 shows the flowchart of the modelling approach in this case study. The model will be developed based on theoretical phase equilibrium and use two process conditions, pH and salt concentration, as model variables in order to achieve an improved process understanding for process operation and design. The structure of the model will be examined by two different feedstock, a pure fragment of antibody (Fab') solution and a clarified Fab' solution from *E. coli* homogenate to determine how well the model can describe the Fab' and impurity solubility in both feedstock. The massive data required to regress will be provided by high throughput microwell precipitation. The model will be validated by experimental data and statistical tests will be used to evaluate the quality of the model. The model regression and evaluation will be supported by MatLab programs. If the statistic tests fail, the model needs modification and iterates with experimental data through modelling loop. The predictions of the model will also be compared with four empirical models where pH will be introduced as an extra variable (Cohn, 1925; Habib et al., 2000; Niktari et al., 1990).

4.4 Phase equilibrium based protein precipitation model

Protein precipitate has been thermodynamically regarded as a pure crystal since the solution has only protein and salt (Chiew et al., 1995; Kuehner et al., 1996; Winzor et al., 2001). However, for proteins in the fermentation broth or other complex biological materials, the precipitation will not form pure crystal but an amorphous mixture (Arakawa and Timasheff, 1985; Shih and Prausnitz, 1992). Arakawa and Timasheff (1985), Shih and his co-workers (1992) treated the precipitation as a distribution between a light liquid phase (supernatant) and a dense liquid phase

(precipitation). Therefore, the proposed model in this thesis will be developed based on the phase equilibrium for the target protein in a multi-component solution:

$$Protein^{light\ phase} \overset{k}{\longleftrightarrow} Protein^{dense\ phase}, \tag{4.1}$$

where k is the kinetic rate. When two phases are in equilibrium, the chemical potentials of protein must be equal:

$$\mu_l = \mu_d , \tag{4.2}$$

where μ_l is the chemical potential for liquid phase and μ_d is the chemical potential for dense phase. Based on equation (4.1) and (4.2), it became

$$\mu_l^\circ + RT \ln C_l \cdot r_l = \mu_d^\circ + RT \ln C_d \cdot r_d , \tag{4.3}$$

where C_l is the protein molar concentration in the light phase, C_d the protein molar concentration in the dense phase, r_l the protein activity coefficient in the light phase, r_d the protein activity coefficient in the dense phase, R the ideal gas constant, T the temperature, μ_l° the protein standard chemical potential in the light phase, μ_d° the protein standard chemical potential in the dense phase.

Equation (4.3) can be rearranged as

$$\frac{C_l}{C_d} = \frac{r_d}{r_l} \cdot e^{-\frac{\mu_l^\circ - \mu_d^\circ}{RT}} . \tag{4.4}$$

Suppose V_l is the light liquid phase volume, V_d the dense phase volume, C_T the maximum protein concentration in the solution and V_T the total solution volume with the assumption that there is no volume change during precipitation. Then,

$$V_T = V_l + V_d , \tag{4.5}$$

$$C_l \cdot V_l + C_d \cdot V_d = C_T \cdot V_T . \tag{4.6}$$

Introduce equation (4.5) and (4.6) into equation (4.4) and transform it to

$$\frac{C_l}{C_T} = \frac{1}{\frac{V_l}{V_T} + \frac{V_d}{V_T} \cdot \frac{r_l}{r_d} \cdot e^{\frac{\mu_l^\circ - \mu_d^\circ}{RT}}} . \tag{4.7}$$

In the processing, dense phase volume is very small compared to the total volume due to relatively low protein concentration, so it is reasonable to assume that $\frac{V_l}{V_T} \approx 1$.

However, $\frac{V_d}{V_T}$ depends on the protein properties in the materials and also the process conditions. According to prior knowledge, the dense phase volume in most precipitation cases will increase with salt concentration and reach a nearly constant level at high salt concentration. It will also change with pH and depend on all components in the material, probably without an apparent isoelectric point (pI). In this case, pH was assumed to have only a simple linear effect on the multi-components precipitation. Considering the salt effects, the trend of $\frac{V_d}{V_T}$ in real experiments was very similar to the shape of Michaelis-Menten enzyme kinetic model (Michaelis and Menten, 1913). Thus, it is proposed that $\frac{V_d}{V_T}$ is represented by an empirical equation of salt concentration and pH with the same structure as above model as follows:

$$\frac{V_d}{V_T} = a_1 + \frac{b_1 \cdot (pH - c_1) \cdot C_s}{d_1 + C_s}, \tag{4.8}$$

where a_1,b_1,c_1,d_1 are coefficients to represent the effect of pH and salt concentration on $\frac{V_d}{V_T}$. By arranging equation (4.8),

$$\frac{V_d}{V_T} = \frac{a_1 \cdot d_1 + a_1 \cdot C_s + b_1 \cdot pH \cdot C_s - b_1 \cdot c_1 \cdot C_s}{d_1 + C_s} = \frac{\alpha + \beta \cdot C_s + \chi \cdot pH \cdot C_s}{\delta + C_s}, \tag{4.9}$$

where C_s is the salt molar concentration, $\alpha, \beta, \chi, \delta$ are the lumped parameters and will be estimated by real experimental data.

In 1943, Kirkwood (1943) found protein activity coefficient in a multi-component solution is a simple function of the concentrations of all solute species. Long and

Mcdevit (1952) further justified protein activity coefficient can be assumed by a log-linear function:

$$\log r_p = k_s C_s + \sum_{i=1}^{n} k_i C_i \,, \tag{4.10}$$

where k_s is the salt activity coefficient, k_i the components activity coefficient, C_i the other component concentration in the solution.

In a multi-component solution containing biomolecules and salt, the protein activity coefficient is dominantly affected by salt concentration. The other effects caused by biomolecules can be regarded as constant due to their very low molar concentrations. Therefore in the liquid phase, the second part of equation (4.10) can be represented by a constant. In the dense phase, the concentration of salt is considered not changing much while other molecules still have no or little effects, hence the overall protein activity coefficient can be regarded as a constant at all time. From the above assumptions equation (4.10) provides,

$$\frac{r_l}{r_d} = \frac{e^{k_s \cdot C_s + w_1}}{e^{w_2}} = e^{k_s \cdot C_s + w_3} \,, \tag{4.11}$$

where w_1, w_2 and w_3 are the constants, k_s is the coefficient.

Equation (4.11) was a modification of Kirkwood's equation (4.10). However in some case, the protein property and its main interaction with salt will depend on the type of salt so equation (4.11) may need the second order or even higher order of salt concentration term to describe the strong effect of salt concentration on the change of protein activity coefficient (Kirkwood, 1943). In this study, only first order term was used in the model. The higher order model will be considered only if the first order one fails to work.

In 1985, Arakawa and Timasheff (1985) published a theoretical protein precipitation model for a single component, which gave following theoretical chemical potential equation,

$$\frac{\mu_2^\circ - \mu_{2,w}^\circ}{RT} = AZ^2\sqrt{I}/(1+B\sqrt{I}) - (1/2.303RT)\int_0^m (\partial\mu_2/\partial m_3)_{T,m_2}\, dm_3 , \quad (4.12)$$

where μ_2° is the protein standard chemical potential in the solution, $\mu_{2,w}^\circ$ the protein standard chemical potential in the water, Z the net charge of the protein, I the ionic strength, m_3 the salt molar concentration, R the gas constant, T the temperature, A, B the coefficients.

The second differential term can be approximated by a first order term of salt concentration $\phi \cdot C_s$, as $(\partial\mu_2/\partial m_3)_{T,m_2}$ is an empirical constant over a wide range of salt concentrations, especially in salting-out range (Arakawa and Timasheff, 1985). As proteins are sensitive to thermal damage, so the process usually is operated at a low fixed temperature. Although increasing temperature will facilitate precipitation, it will also denature targeted protein and require large energy input in large scale precipitation. Therefore, the temperature was not considered as a variable in this study. All the experiments were carried out at room temperature and RT thus can be regarded as a constant in the equation (4.12).

Since ionic strength is defined as:

$$I = \frac{1}{2}\sum C_i \cdot Z_i^2 , \quad (4.13)$$

where C_i is the molar concentration of ion i, and Z_i is the charge number of that ion. For a neutral salt, such as ammonium sulphate, $\sqrt{I}$ is linear proportional to square root of salt concentration $\sqrt{C_s}$, according to equation (4.13). As protein surface net charge is a function of pH without a general mathematical model, Z^2 in equation (4.13) was assumed to be described and approximated by a second order pH

polynomial equation. Therefore, equation (4.12) can be simplified into a function of salt concentration and pH with the similar structure:

$$\frac{\mu_2^\circ - \mu_{2,w}^\circ}{RT} = \varphi(pH - \gamma)^2\sqrt{C_s}\,/(\eta + \sqrt{C_s}) + \phi \cdot C_s \tag{4.14}$$

where $\phi, \varphi, \gamma, \eta$ are lumped coefficients.

Under the assumptions that the salt concentration in the dense phase is very small and does not change much, which means the value of equation (4.14) for dense phase protein will be considered as a coefficient without bulk salt concentration and pH influence, above assumptions giving

$$\frac{\mu_l^\circ - \mu_d^\circ}{RT} = \frac{(\mu_l^\circ - \mu_w^\circ) - (\mu_d^\circ - \mu_w^\circ)}{RT} = \phi \cdot C_s + \frac{\varphi \cdot (pH - \gamma)^2 \cdot \sqrt{C_s}}{\eta + \sqrt{C_s}} + \kappa\,, \tag{4.15}$$

where $\phi, \varphi, \gamma, \eta, \kappa$ are lumped coefficients.

The second term mainly describes the protein salting-in effect at low salt concentration. At high salt concentration, this phenomena does not exist or the effect is very small compared to the first term in equation (4.15) (Arakawa and Timasheff, 1985). To simplify the calculation, the second term was rearranged to be expressed by pH effect, which is dominated described by a simplified second order polynomial function of pH effect only, and the salt effect at low concentration range, which is separated from this term and lumped into the first term to give:

$$\frac{\varphi \cdot (pH - \gamma)^2 \cdot \sqrt{C_s}}{\eta + \sqrt{C_s}} \approx \sigma \cdot pH^2 + \nu \cdot pH + \rho + \xi C_s\,, \tag{4.16}$$

where σ, ν, ρ, ξ are coefficients.

Combine equation (4.9), (4.11), (4.15) and (4.16) together, the model in equation (4.7) becomes:

$$\frac{C_l}{C_T} = \frac{1}{1 + \dfrac{f + g \cdot C_s + h \cdot pH \cdot C_s}{i + C_s} \cdot \exp^{(a \cdot C_s + b \cdot pH + c \cdot pH^2 + d)}}, \tag{4.17}$$

where *a, b, c, d, f, g, h, i* are lumped coefficients.

This model is able to describe the strong nonlinearity of the precipitation surface due to its sigmoid structure. All parameters in the equation (4.17) are lumped and thus it is difficult to predict their values or limit their ranges. However, according to prior knowledge and modelling assumption, parameters *a* and *i* should have positive values and $f + g \cdot C_s + h \cdot pH \cdot C_s$ term also be positive. At low salt concentration, the exponential expression in the model is not a dominate effect and thus the decrease of dense phase volume caused by salting-in effect, which makes the value of $\frac{V_d}{V_T}$ smaller, explains the protein concentration increase. At high salt concentration, the second term which contains an exponential expression will have a much larger value than one, ignore value 1 in the denominator and thus this model is similar to the exponential structure of Cohn's equation.

4.5 Model comparison

In order to evaluate the capability of this new model, it is useful to compare the model with three published models, Cohn's (1925), Niktari's (1990) and Habib's (2000) models plus a polynomial model. For the process design purpose, all selected models were modified to contain pH factor by introducing a second order polynomial expression of pH to substitute model coefficients without changing the model structure, in order to link protein solubility directly to operating variables.

Expansion on Cohn's equation is

$$\ln \frac{S}{S_0} = a_2 + b_2 \cdot pH + c_2 \cdot pH^2 - d_2 \cdot C_s . \tag{4.18}$$

Expansion on Niktari's equation is

$$y = \frac{1}{1 + (\frac{C_s}{a_3 + b_3 \cdot pH + c_3 \cdot pH^2})^{d_3 + e_3 \cdot pH + f_3 \cdot pH^2}} . \quad (4.19)$$

Expansion on Habib's sigmoid model is

$$y = 1 - (a_4 + b_4 \cdot pH + c_4 \cdot pH^2) \cdot (1 - \exp[-(\frac{C_s}{d_4 + e_4 \cdot pH + f_4 \cdot pH^2})^{g_4 + h_4 \cdot pH + i_4 \cdot pH^2}]) . \quad (4.20)$$

A second order polynomial equation with interaction terms is used to represent the conventional two factors polynomial model in DoE:

$$y = a_5 + b_5 \cdot C_s + c_5 \cdot pH + d_5 \cdot C_s^2 + e_5 \cdot pH^2 + f_5 \cdot C_s \cdot pH , \quad (4.21)$$

where y is Fab' concentration in the supernatant, S/S_0 is the percentage of Fab' concentration in the supernatant to the Fab' concentration in the feedstock, and others are parameters.

4.6 Results and discussion

The phase equilibrium based model structure, equation (4.17), was used to describe the Fab' solubility in pure Fab' solution and clarified homogenate with regarding to pH and ammonium sulphate concentration. As the focus of this study was on investigation of accurate precipitation model, a relevantly large data set was needed for model accuracy and thus the brute-force design method for experimental design was adapted. Then a nonlinear least square method was used to estimate the parameters in the model. The accuracy of the model, equation (4.17), and the other models, equation (4.18), (4.19), (4.20) and (4.21) was measured by R^2. In addition, statistical tests such as F-test and t-test were used to establish and validate the new model and parameters.

4.6.1 Model parameters for pure Fab', Fab' in clarified homogenate and impurities

In total, 119 experiments were carried out for pure Fab' solution and 79 experiments for clarified homogenate. The normalised pure Fab' concentration shown in Figure 4.2.(a) slightly increased at salting-in phase and then gradually decrease with salt concentration, showing the same pattern as many previous pure protein precipitation curves (Arakawa and Timasheff, 1985; Cohn, 1925). The main difference between the two feedstocks was the salting-in behaviour. For the clarified homogenate shown in Figure 4.2.(b), when salt concentration was low, Fab' concentration was significantly affected by other components in the solution and its concentration was dramatically altered compared to that of pure Fab' solution. Under the unfavourable conditions, e.g. low pH and low salt concentration, the Fab' concentration was only 40% of the highest concentration. Compared with the impurity solubility, shown in Figure 4.2.(c), the same phenomena existed. It may be explained that the low Fab' concentration at certain conditions, while it was not pure solution, was probably caused by the co-precipitation between Fab' molecule and other impurity proteins, the solubility of which were dramatically changed by pH at low ionic strength (Akita and Nakai, 1993; Bramaud et al., 1997). When the salt concentration increased, the salting out effects dominated, and thus the solubility was the same as that of pure Fab' solution.

These experimental data sets were then used to develop the pure Fab' precipitation model based on equation (4.17). The estimated parameters were shown in Table 4.1 and R-square value was 0.975, which indicated a very good match between the model and the experimental data. The F-test value of the model fitting was 624.81, indicating that 95% confidence of model accuracy was achieved. Figure 4.2 also showed the model prediction surfaces. The model can describe flexibly both the salting-in and salting-out features of the concentration surface without the cost of losing accuracy at any phase, except for a few stray points probably caused by experimental errors.

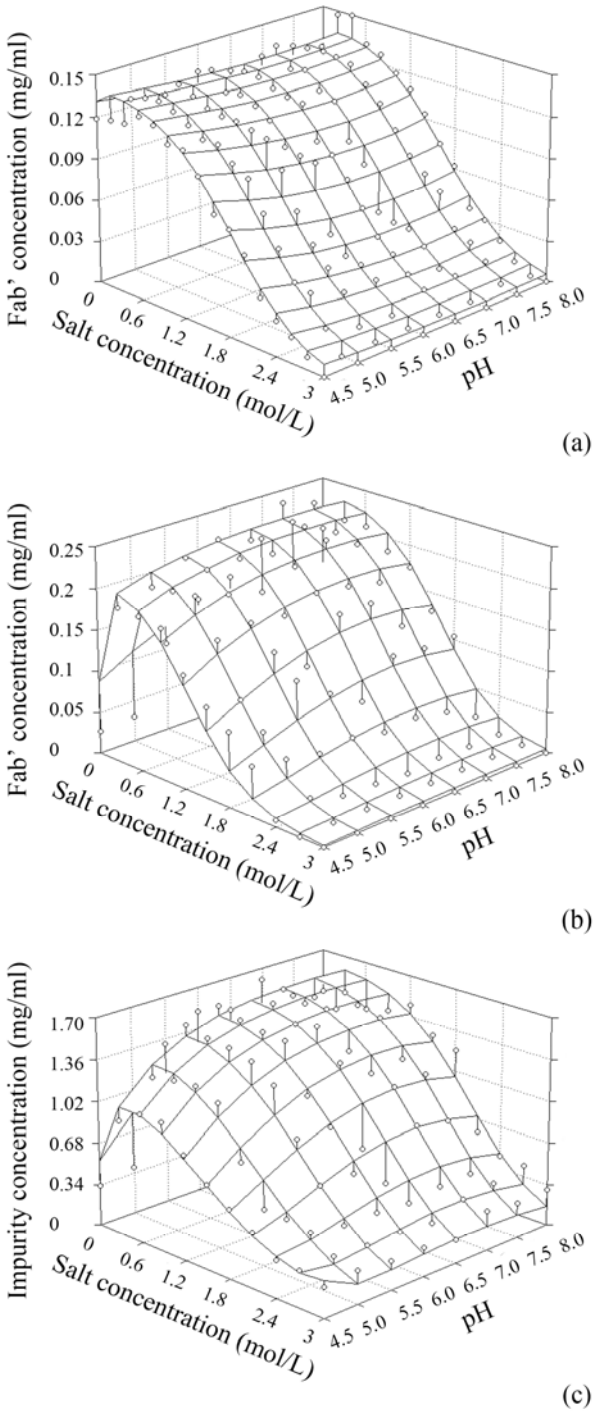


Figure 4.2 The predicted surfaces provided by simplified model equation (4.17), with real experimental results (dots): (a) pure Fab' solution; (b) Fab' in clarified homogenate; (c) impurities in clarified homogenate.

	Parameters								R^2	F-test value
	a	*b*	*c*	*d*	*f*	*g*	*h*	*i*		
Pure Fab'	7.97	1.62	0.53	-5.30	1.15	-1.05	0.16	0.03	0.975	624.81
Fab' in clarified homogenate	7.61	-6.60	4.34	-4.27	1.04	-0.75	21.12	0.002	0.972	320.36
Impurities	5.51	-7.88	5.03	-3.43	1.32	-0.41	20.60	0.01	0.945	172.68

Table 4.1 Parameters, F-test value and R^2 value of developed model using equation (4.17) for pure Fab' precipitation, Fab' precipitation in clarified homogenate and impurities precipitation.

To examine the generality of the model structure, the model was then assessed by applying to Fab' precipitation in clarified homogenate, where multi-components exist. The parameters of pure Fab' model were used as the initial guess for this new fitting. The estimated parameters were also shown in Table 4.1 and R-square value was 0.972, which was very similar to that of pure Fab' model. The F-test value of model fitting was 320.36, which was smaller than that in pure Fab' model. However, these measures indicated that the model was also accurate with 95% confidence. The predicted Fab' concentration surface was shown in Figure 4.2.(b). It can be seen that Fab' solubility in the clarified homogenate was predicted well by the model in general. Nevertheless, there was a slight discrepancy between predicted solubility and experimental data at low pH range as well as very low salt concentration. As the exposure to these conditions can cause the formation of soluble high molecular weight aggregates and/or insoluble precipitate (Shukla et al., 2007), it is very difficult to be predicted with no theoretical model available.

To further explore the quality of the model, the model was assessed by applying to describe impurities precipitation, where a mixture of proteins was treated as an assumed pseudo-single molecule with average characters of all proteins in the solution, e.g. average electronic charges and hydrophobic behaviour (Mahadevan and Hall, 1992). 79 experiment data from impurities precipitation in clarified homogenate were used for parameters estimation. The results were given in Table

4.1 and the R-square value was 0.945, which was lower than that of Fab' models. The F-test value of the model was 172.68. These measures showed the model was accurate with 95% confidence. The predicted impurity concentration surface was shown in Figure 4.2.(c). The geometrical pattern in real data points and model predicted surface were slightly different, especially at high salt concentration region. The difference was caused by the simplification of a mixture of proteins into a pseudo-single protein. Although the impurities were regarded as a pseudo-single molecule with average values of all mixture, in reality, different protein will have different sensitivity to salt and pH (Hoskins et al., 1996). Conditions may dramatically affect one protein with no great effect on other molecules (Mahadevan and Hall, 1992). Thus, the assumed average properties of a pseudo-single molecule were not constant under all conditions, especially for extreme conditions, e.g. low pH or high salt concentration.

4.6.2 Model modification

It has been demonstrated that the new model can represent the experimental data with good quality. However, there were nine parameters in the equation (4.17) and the model exhibits a high level of nonlinearity. Generally, a simple model is more useful than a complicated one for application in processes operation and design (King et al., 2007). To simplify the model is also beneficial for computation, since the higher number of parameters, the longer computing time and the less accuracy. The t-test for individual parameters in the model can be used to evaluate and decide if a parameter is necessary. If a parameter fails the t-test, it is neither accurate nor needed (Freedman, 2005). After carrying out t-test for each parameter, the parameters *f*, *g* and *h* in the upper part of $\frac{V_d}{V_T}$ have failed t-test in all three models. Therefore, let the upper part equal to one for model simplification by considering $\frac{V_d}{V_T}$ as a ratio and thus the model can be simplified to

$$\frac{C_l}{C_T} = \frac{1}{1 + \frac{1}{f_1 + C_s} \cdot e^{(a_1 \cdot C_s + b_1 \cdot pH + c_1 \cdot pH^2 + d_1)}}. \qquad (4.22)$$

The simplified model, equation (4.22), was developed for pure Fab', Fab' in clarified homogenate and impurities in clarified homogenate by using the experimental data sets again and the parameters were shown in Table 4.2. The R-square values were evaluated and were a little lower than the previous ones, but all tests still showed the models had excellent statistic confidence. All parameters passed t-tests with 95% confidence.

	Parameters					R^2	F-test value
	a_1	b_1	c_1	d_1	f_1		
Pure Fab'	8.21	1.49	-1.08	-5.34	0.03	0.973	1037.82
Fab' in clarified homogenate	9.46	-3.16	2.19	-4.60	0.005	0.937	308.28
Impurities	7.02	-5.01	3.70	-3.51	0.02	0.914	195.48

Table 4.2 Parameters, F-test value and R^2 value of modified model using equation (4.22) for pure Fab' precipitation, Fab' precipitation in clarified homogenate and impurities precipitation.

Replacing the parameters in equation (4.22) with the corresponding values in the Table 4.2, all three models showed the similar trends. When salt concentration was low (<0.2 mol/L), the value of exponential term was small and changed little while the $\frac{1}{f_1 + C_s}$ term dominated and changed rapidly. It described the salting in phenomenon at low salt concentration (Arakawa and Timasheff, 1984). When salt concentration was higher, e.g. in salting out range, the value of exponential term dominated due to large value of parameter a_1, while the effect of $\frac{1}{f_1 + C_s}$ term was small. It explained the protein salting out with the similar mathematical structure as Cohn's equation at high salt concentration (Cohn, 1925). The values of parameters

linking with pH varied relatively little. The parameter of second order pH term for impurities, c_1, was the largest with the fact that impurity concentration was influenced the most by the pH in all three materials. The effect of pH at neutral pH range was very small due to the small parameter value and its second order structure. However, extreme pH conditions would have large effects according to the models, describing the protein concentration changes in the experiments at extreme pH conditions (Petrucceli and Anon, 1996).

4.6.3 Model validation

9 DoE experiments under the same operation space were carried out to validate the model, as shown in Table 4.3. When validating bioprocess models, it was not recommended to use error percentage to evaluate models because the range of bioprocess data may be very wide even after scaled or transformed, which will introduce mathematical error (Hancock et al., 1988). Thus, statistical tests should be utilised to validate new model, no matter how good the fitting of the data was in the regression step (Hills and Leslie, 2003; Mayer and Butler, 1993).

Conditions		Pure Fab' (mg/ml)		Fab' in clarified homogenate (mg/ml)		Impurities in homogenate (mg/ml)	
Salt	pH	Real value	Model value	Real value	Model value	Real value	Model value
0.6	5.0	0.122	0.128	0.181	0.187	1.229	1.336
1.8	5.0	0.075	0.064	0.067	0.056	0.464	0.646
3.0	5.0	0.000	0.007	0.000	0.003	0.272	0.095
0.6	6.5	0.117	0.125	0.188	0.197	1.601	1.481
1.8	6.5	0.054	0.057	0.059	0.074	0.658	0.899
3.0	6.5	0.000	0.006	0.000	0.004	0.197	0.168
0.6	8.0	0.127	0.122	0.185	0.205	1.307	1.573
1.8	8.0	0.060	0.051	0.094	0.094	1.193	1.144
3.0	8.0	0.000	0.005	0.000	0.006	0.301	0.284

Table 4.3 Validation DoE with real experimental value and model predicting value for pure Fab', Fab' in homogenate and impurities.

However, there are several unusual problems for bioprocess model validation. First of all, the number of samples used for validation is normally small, e.g. 9 samples in this case, due to various reasons, such as high cost of materials and long time of experimentation. Secondly, the distribution of most bioprocess data is normally unknown or the data is hardly transformed to any known distribution, e.g. standard normal distribution (Goffaux and Wouwer, 2005). Statistically, the normal distribution can be assumed only when the number of samples is very large, normally more than 30 (Lamprecht, 2005). Therefore, for small validation group with unknown distribution, it is of great risk to use paired t-test due to high probability to fail.

There exist two solutions for this type of validation. One solution is to use Wilcoxon signed-rank test (Wilcoxon, 1945) for few samples. The other is to analyse validation samples together with previous regression data by paired t-test since the whole data set can be roughly considered as the normal distribution when sample number is large than 30. For Wilcoxon test, 2-tailed significance > 0.05 can be regarded as validation passed. For paired t-test, sig. > 0.05 can be considered as the null hypothesis that there is no difference between experimental data and model calculated value is accepted (Schiff and D'Agostino, 1996). Table 4.4 shows the test results for equation (4.22) with t-test sig. > 0.05 and Wilcoxon 2-tailed significance > 0.05 (from SPSS calculation) in all three materials. It demonstrates the simplified model passed validation with strong statistical confidence (95 %).

	Pure Fab'	Fab' in clarified homogenate	Impurities
t-test value	1.201	1.220	-0.198
sig. (2-tailed)	0.232	0.227	0.843
Wilcoxon 2-tailed sig.	0.767	0.086	0.515

Table 4.4 Results with 9 samples Wilcoxon signed-rank test and all samples paired t-test results for modified model using equation (4.22).

4.6.4 Model comparison

The experimental data sets were also used to estimate the parameters in the four models described by equations (4.18), (4.19), (4.20) and (4.21). Table 4.5 presented the R-square and the F-test values for all four compared models in different feedstock. The predicted Fab' and impurity concentration surfaces by the four models were shown in Figure 4.3 to Figure 4.6 respectively. A large difference between predicted surface and experimental data was observed for all three models based on Cohn's equation in Figure 4.3.(a), (b), (c). Due to its linear model structure, Cohn's equation can not describe protein salting-in effect at low salt concentration in multi-components solution (Curtis et al., 2002) as all statistical tests failed with all R-square values below 0.80.

Nitark's expansion model and Habib's expansion model can fit quite well in pure Fab' precipitation, shown in Figure 4.4.(a) and Figure 4.5.(a), with R-square value 0.956 and 0.971, F-test value 402.72 and 400.30 respectively. However, these models for Fab' precipitation in multi-components solution and impurities precipitation were poor, shown in Figure 4.4.(b), (c) and Figure 4.5.(b), (c). The R-square values and F-test values of these two models shows less good quality of model and less predicting capability, particularly for impurity precipitation. Besides, both models did not consider pH effect originally and a second-order polynomial expression for pH expansion may not effectively describe the real complicated impacts.

	Cohn's equation		Niktari's model		Habib's model		Polynomial model	
	R^2	F-test	R^2	F-test	R^2	F-test	R^2	F-test
Pure Fab'	0.795	148.53	0.956	402.72	0.971	400.30	0.950	433.74
Fab' in clarified homogenate	0.621	40.88	0.795	46.55	0.864	48.78	0.825	68.90
Impurities	0.721	64.50	0.859	73.03	0.877	54.48	0.858	88.00

Table 4.5 R^2 and F-test values for all four compared models.

Theoretically, the polynomial model has the most flexibility to fit data. However, the same results were obtained as previous two models, shown in Figure 4.6. The model in pure Fab' precipitation was quite good with R-square value 0.95, but the models for Fab' precipitation in multi-components were less good as R-square was less than 0.90 in non-ideal solution. Moreover, at high salt concentration, the predicted concentration surface of the polynomial model was below zero, which also occurred in Habib's model. Therefore, both models were quite misleading because the predicted value below zero had no physical meaning. In order to solve this problem, the parameters in these models can be adjusted at the cost of losing accuracy at other conditions. The polynomial model was considered as not adequate to give good fitting in clarified homogenate precipitation. The higher order polynomial model may be more accurate but will inevitably introduce more parameters, complicate computation and have little physical meanings.

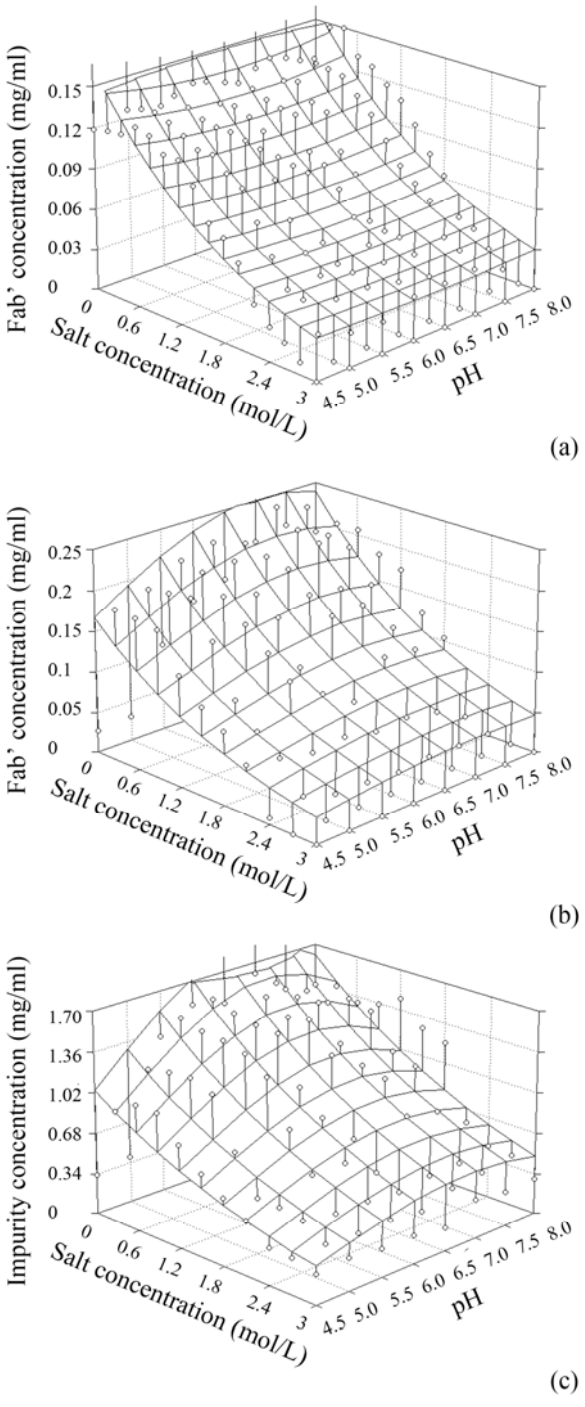


Figure 4.3 The predicted surfaces provided by Cohn's model equation (4.18) with real experimental results (dots): (a) pure Fab' solution; (b) Fab' in clarified homogenate; (c) impurities in clarified homogenate.

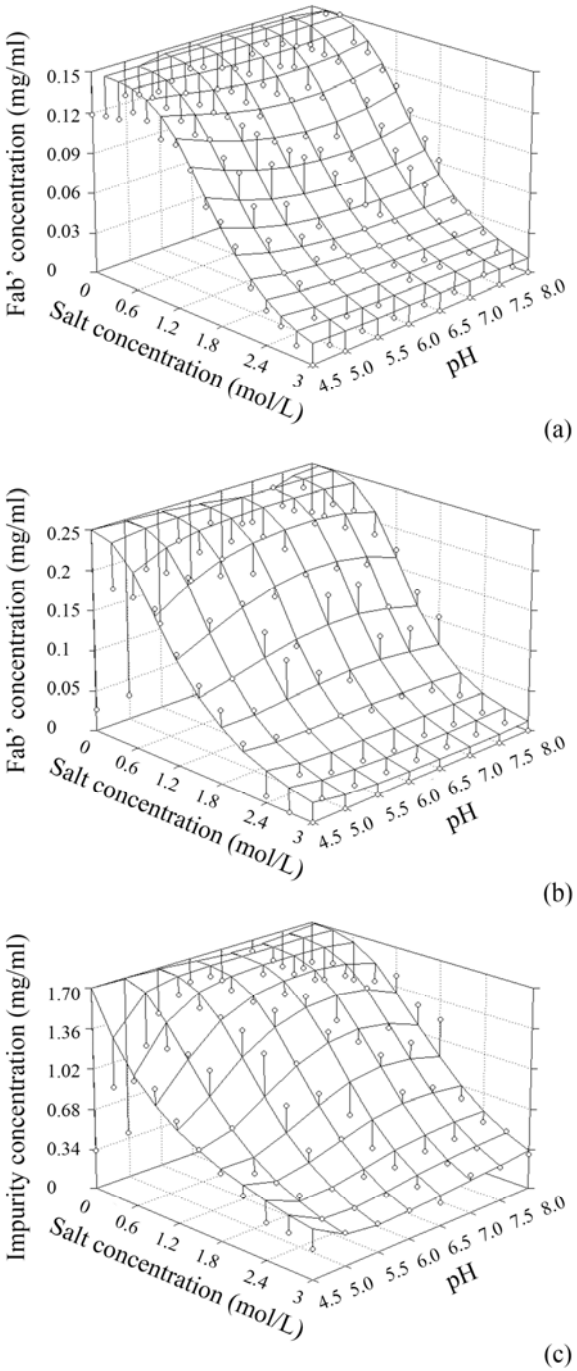


Figure 4.4 The predicted surfaces provided by Nitark's model equation (4.19) with real experimental results (dots): (a) pure Fab' solution; (b) Fab' in clarified homogenate; (c) impurities in clarified homogenate.

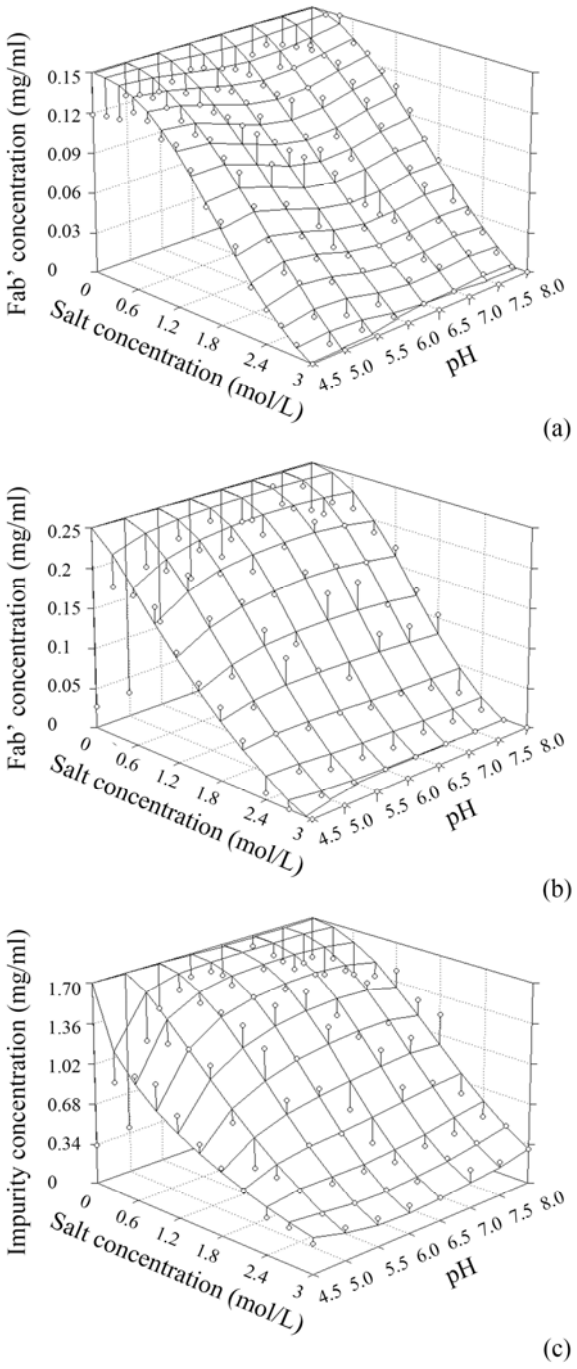


Figure 4.5 The predicted surfaces provided by Habib's model equation (4.20) with real experimental results (dots): (a) pure Fab' solution; (b) Fab' in clarified homogenate; (c) impurities in clarified homogenate.

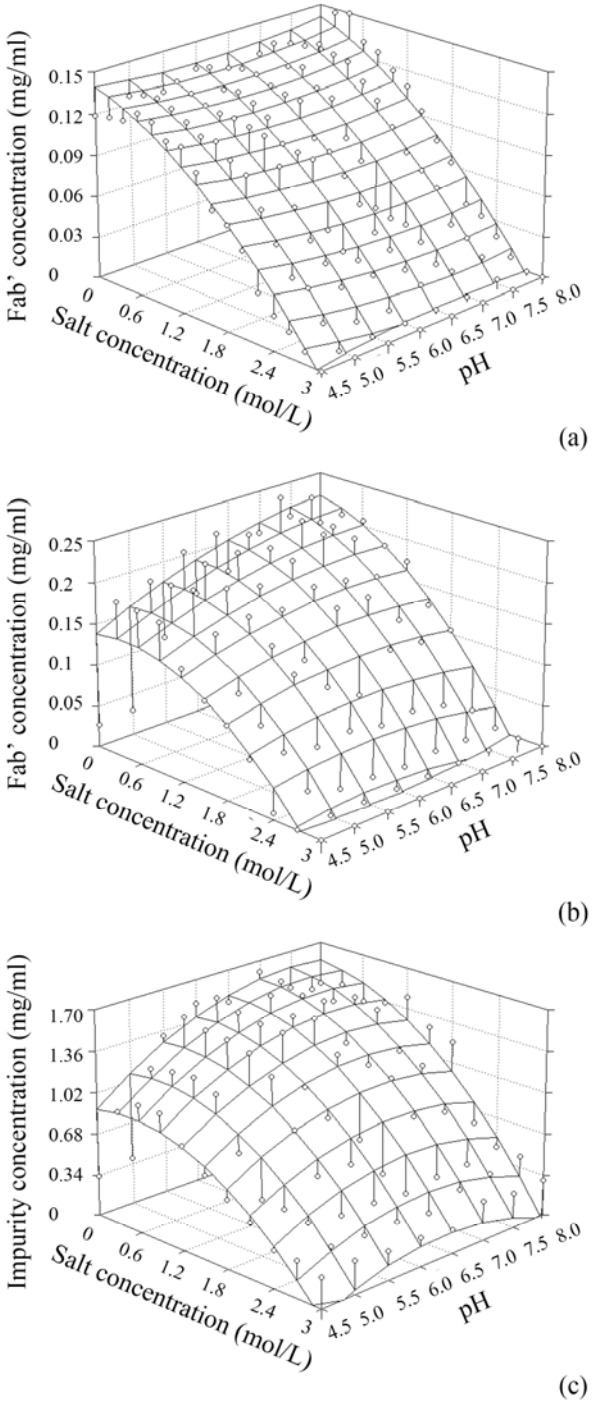


Figure 4.6 The predicted surfaces provided by polynomial model equation (4.21) with real experimental results (dots): (a) pure Fab' solution; (b) Fab' in clarified homogenate; (c) impurities in clarified homogenate.

4.7 Modelling procedure and limitation

The results well demonstrated this modelling procedure incorporating with high throughput experimentation was better than traditional modelling procedure, as a large amount of data for modelling can be produced in a short time and few materials required (Nfor et al., 2009). With more data output, it made the accuracy of bioprocess model much more possible even in a complex system. The most difficult thing in modelling was still the mechanic model development starting from theories, which really required sufficient understanding of the process. The large amount of data collected created a challenge for mathematical and statistical algorithm (Katare et al., 2004). A proper algorithm should be able to analyse data, validate model and help to modify model through modelling and high throughput experimentation loop shown in Figure 4.1.

High throughput experimentation platform collaborating with mathematical algorithm also has a potential application in empirical model setting up but requires more advanced software to execute model selection and regression (Pfeifer et al., 2010). The limitation of this method is that the model developed by microwell technique needs scaling up and validating in large scale process. In order to scale up, extra large scale experimentations are needed to provide data support and the model may need correlation in some case (Micheletti et al., 2006). However, this method is still faster and requires less time and cost than traditional modelling procedure without the help of high throughput technology.

4.8 Conclusions

Phase equilibrium based models have been developed and modified using parameter t-tests. The model equation (4.22) can precisely describe the precipitation surface of pure Fab' and Fab' in clarified homogenate with two variables, salt concentration and pH. The model structure based on single protein can be applied to a protein in

multi-components as well as a pseudo-single molecule such as impurities with slightly reduced quality.

The results showed the modified model equation (4.22) can predict Fab' precipitation in pure Fab' solution very well without failing any regression statistical tests. When applied to clarified homogenate with multi-components present, the model demonstrated excellent robustness and passed all of the regression statistical tests. Even when applied further to impurities in clarified homogenate, the accuracy of the model prediction was only slightly reduced but was still within the acceptable statistical confidence of 95%. In addition, the model developed was superior to the four existing models published, further indicates the suitability of the model.

The model structure was generic to any protein solution precipitation and allowed the researchers to modify the model structure according to the first principle based on. The fact that it worked well for Fab' and impurities in clarified homogenate showed this model can be used to predict precipitation outputs and conditions in order to obtain a low cost precipitation step as an alternative to protein A based chromatography process (Przybycien et al., 2004). It could also be used to define a pre-purification step for the chromatographic routine by reducing the burden of high impurity concentration to avoid plugging of affinity columns (Peram et al., 2010).

Chapter 5. Model based process development and experimental design at microwell scale: a case study in Fab' precipitation

5.1 Introduction

When facing a complex nonlinear bioprocess with limited time and cost, there are two efficient strategies based on different concepts in order to achieve good understanding and promote process development. One is to reduce the volume of each experiment and thus increase the number of experiments, e.g. high throughput experimentation. The other available approach is, in the contrast, to reduce the number of experiments by a well-planned experimental design, such as model based experimental design. It seems there is a conflict between two methods and it is impossible to apply them together in a process simultaneously. However, the fundamental idea behind them is exactly same, that is to squeeze abundant information by using limited materials in a constrained situation. Through this concept, optimal conditions can thus be found because an optimisation procedure for bioprocesses normally starts with some limited priori knowledge with a defined objective e.g. maximum yield, and aims to achieve the final process optimality rapidly with increasing understanding (Galvanauskas et al., 1998).

Therefore, in this chapter, the third step, model based process development, in previous proposed methodology (Figure 3.2) will be tested with the combination of above two strategies. The process development procedure will follow the flowchart in Chapter 3, Figure 3.3, in order to illustrate the capability of this model based design and optimisation method. It will be applied to a case study: downstream Fab' precipitation processes using the precipitation model developed and validated in previous Chapter 4. Then it will be compared with traditional DoE design and optimisation methodology to understand its powers and limitations.

5.2 Case study: Fab' precipitation

In order to investigate the quality and efficacy of the proposed model based experimental design and optimisation for bioprocesses, a Fab' precipitation by ammonium sulphate at different pH value was used as a case study to provide an evaluation based on real experimental data. The following precipitation model, which was developed in Chapter 4, will be used both for Fab' concentration and impurity concentration calculation:

$$\frac{C_l}{C_T} = \frac{1}{1 + \frac{1}{f + C_s} \cdot e^{(a \cdot C_s + b \cdot pH + c \cdot pH^2 + d)}} \tag{5.1}$$

where C_l / C_T is the predicted normalised protein concentration in the supernatant, C_s is normalised salt concentration and *a, b, c, d, f* are parameters.

The system inputs are salt concentration and pH value with Fab' value and impurity value calculated by equation (5.1) in normalised values respectively, while the objective of the process design is to find an optimal process condition for Fab' primary capture based on an equation of purity ratio and yield as following:

$$Objective\ \ value = 0.4 \cdot Fab'\ yield + 0.6 \cdot purity\ ratio - 0.1 \cdot salt\ concentration\ , \tag{5.2}$$

where

$$Fab'\ yield = \begin{cases} \dfrac{C_{l\,Fab'}}{C_{T\,Fab'}} & \dfrac{C_{l\,Fab'}}{C_{T\,Fab'}} > 30\% \\ 1 - \dfrac{C_{l\,Fab'}}{C_{T\,Fab'}} & \dfrac{C_{l\,Fab'}}{C_{T\,Fab'}} < 30\% \end{cases}, \tag{5.3}$$

$$Purity\ ratio = \begin{cases} \dfrac{C_{l\,Fab'}}{C_{T\,Fab'}} \Big/ \dfrac{C_{l\,impu}}{C_{T\,impu}} & \dfrac{C_{l\,Fab'}}{C_{T\,Fab'}} > 30\% \\ \left(1 - \dfrac{C_{l\,Fab'}}{C_{T\,Fab'}}\right) \Big/ \left(1 - \dfrac{C_{l\,impu}}{C_{T\,impu}}\right) & \dfrac{C_{l\,Fab'}}{C_{T\,Fab'}} < 30\% \end{cases}, \tag{5.4}$$

$\dfrac{C_{l\,Fab'}}{C_{T\,Fab'}}$ and $\dfrac{C_{l\,impu}}{C_{T\,impu}}$ are the normalised concentration in the supernatant for Fab' and impurities respectively, salt concentration in equation (5.2) is in normalised value.

In the precipitation process, the product can either be recovered from the soluble phase or the solid phase. As the precipitation is normally used as the first capturing step to collect solids, the product recovery yield from solids should be larger than 70%, which is $\frac{C_{l\,Fab'}}{C_{T\,Fab'}} < 30\%$ in equation (5.3) and (5.4). Otherwise, the soluble phase supernatant will be collected, when $\frac{C_{l\,Fab'}}{C_{T\,Fab'}} > 30\%$. So the yield and purity ratio equations have two expressions according to the criteria whether the Fab' remaining in the solution is more than 30% of overall Fab' concentration or not. The objective function will have a tradeoff between yield, purity ratio and conditions. The target yield was defined to be more than 70% and the weight of yield in objective function is 0.4. The purity ratio has a weight of 0.6 in equation (5.2) because it aims to achieve both high primary recovery and certain extent purification during this bioprocess. The negative value of 0.1 for salt concentration in the equation (5.2) is designed as the penalty term, which attempts to use as little salt as possible in the process.

Experimental data of Fab' and impurities precipitation was provided from previous Chapter 4 in the same normalised form. The clarified homogenate experiments were done by brute-force design and thus the whole dataset was available to give comparison during methodology tests. However, the experimental design algorithm will only acquire the data at designed point from above dataset and the final result will be compared with overall real surface. Because the real experiments have minimum operation space, e.g. 0.5 interval for pH, while the algorithm based on model is numerical method, which will give floating number, the points given will be automatically rounded to the nearest available experiments points in the MatLab code. This may slightly slow down the computation time of the model based process design but will not cause serious problems.

In this case, a set of DoE with nine points was used as start points, shown in Table 5.1. Both Fab' and impurities data will be fitted to the precipitation model (5.1), which will give one model for Fab' and another for impurities. Therefore, after the DoE initialisation in the algorithm, the following designs, such as D-optimal design, will give two points respectively during each loop due to two sets of parameters, one for Fab', the other for impurities.

Experiment No.	Initial DoE		Validation DoE	
	Salt (M)	pH	Salt (M)	pH
1	0.6	5.0	0.0	4.5
2	1.8	5.0	1.5	4.5
3	3.0	5.0	3.0	4.5
4	0.6	6.5	0.0	6.0
5	1.8	6.5	1.5	6.0
6	3.0	6.5	3.0	6.0
7	0.6	8.0	0.0	7.5
8	1.8	8.0	1.5	7.5
9	3.0	8.0	3.0	7.5

Table 5.1 The conditions of initial DoE design and validation DoE used in model based algorithm.

Another set of validation DoE from whole dataset is used to calculate variance (SD^2) between real data and the model predicted value, also shown in Table 5.1. The termination criteria of design algorithm is set as the SD to be lower than 0.05 (5%) in the optimisation procedure for normalised dataset. The real maximum objective point will be given by full dataset plotting and compared with the final maximum point found by algorithm when it terminates. When all three SDs for Fab', impurities and objectives are less than pre-set accepted value, the algorithm terminates and the procedure of design and optimisation is completed. The initial parameters are from the pure Fab' regression in Chapter 4, shown in Table 5.2.

	Parameters				
	a	*b*	*c*	*d*	*f*
Pure Fab'	8.21	1.49	-1.08	-5.34	0.03

Table 5.2 Parameters from pure Fab' model regression, which was used as initial parameters for model based design.

5.3 Results and discussion

5.3.1 The real experimental surfaces

The real surfaces of pure Fab' solution, Fab' in clarified homogenate and impurities in clarified homogenate were plotted in Figure 5.1 (a), (b) and (c) respectively according to the experimental results from Chapter 4. These real surfaces can be compared with the predicted surface in each loop of following model based design and optimisation algorithm, though not in a quantitative way.

The real objective surface was calculated based on equation (5.2) using clarified homogenate results from Chapter 4 and plot in Figure 5.1.(d). It also offered the ground to compare and evaluate the design and optimisation algorithm. The objective surface was quite fluctuated and the optimal area was identified around salt 0.7, pH 0.7-1.0 in normalised value, the real value of which was salt 2.1M, pH 7.0-8.0 with maximum objective value 1.219. The objective surface showed high nonlinearity because it was the mathematical calculation from equation (5.2) based on Fab' and impurities results. No simple model was able to describe this surface directly and thus its optimisation highly depended on two precipitation models and algorithms adopted.

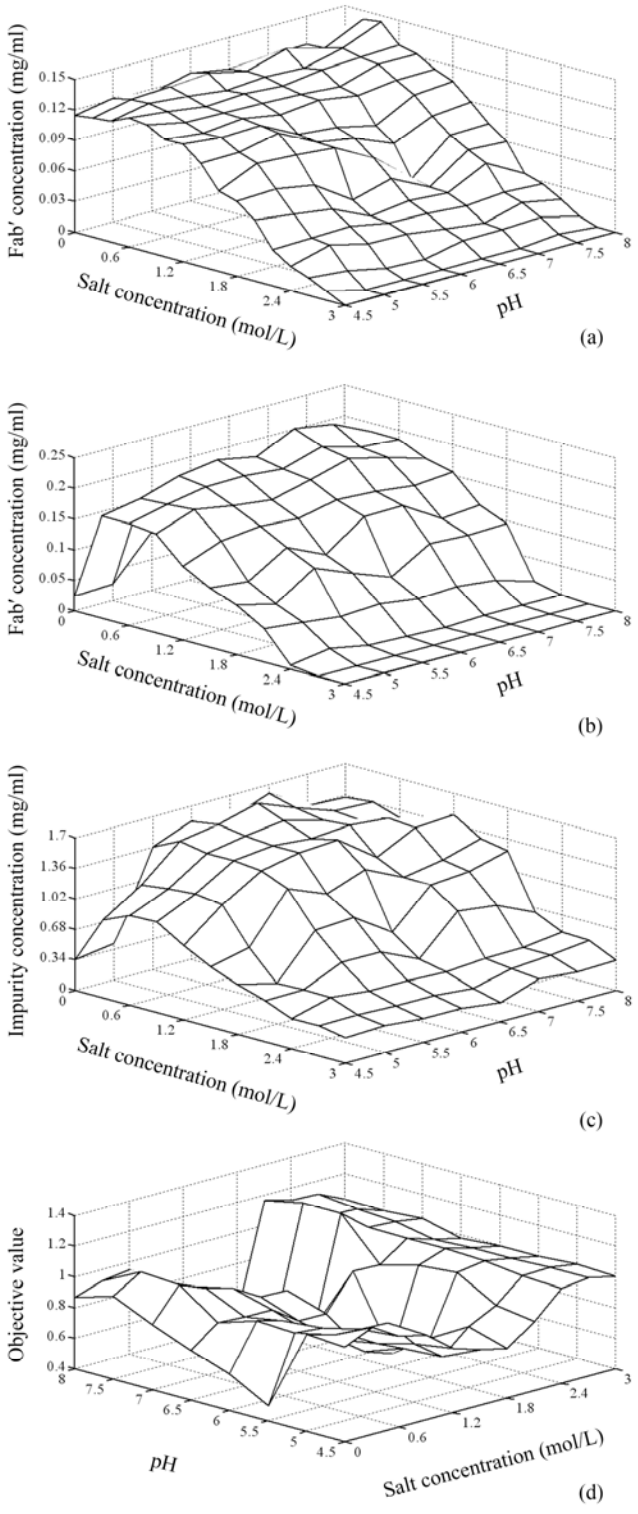


Figure 5.1 The protein solubility surfaces from experimental results: (a) pure Fab' solution; (b) Fab' in clarified homogenate; (c) impurities in clarified homogenate; (d) the objective surface based on clarified homogenate.

5.3.2 Model based design and optimisation

5.3.2.1 The first run

The surfaces predicted by model were really inaccurate due to lack of information in the first run, shown in Figure 5.2. The predicted Fab' solubility surface in Figure 5.2 (a) was not accurate with a sharp drop at normalised salt concentration 0.3, while the real solubility surface started to decrease at 0.3 but with a less steep slop until 0.8, shown in Figure 5.1 (b). Comparing Figure 5.2 (b) with Figure 5.1 (c), the surfaces of impurity solubility had the similar trend but lost the details. These surfaces were the predicted results from the regression of the initial DoE and parameters. The large gap between predicted value and real experimental results showed that a simple set of DoE was not accurate enough to investigate a complex bioprocess, especially in a non linear system (raw data of each validation run shown in Appendix 3). The SD^2 for Fab' and impurities models were 0.1799 and 0.0537 respectively, shown in Table 5.3. Although the values of SD^2 were larger than 5%, the models were initialised by the general information from DoE and then the parameters in the models will be adjusted by D-optimal designs.

Comparing the predicted objective surface in Figure 5.2.(c) with the real objective surface in Figure 5.1 (d), the objective surface calculated from the first run algorithm was completely wrong. It showed the optimal area located at low salt concentration range. Mathematically, the SD^2 for objective surface was very large at 0.4624 and the value of the optimal point predicted was 4.949, nearly 4 times of the real one, shown in the Table 5.3. It was mainly caused by the inaccurate Fab' precipitation model obtained above. Thus, the algorithm continued and four extra points were designed by algorithm for next experiments, also shown in the Table 5.3.

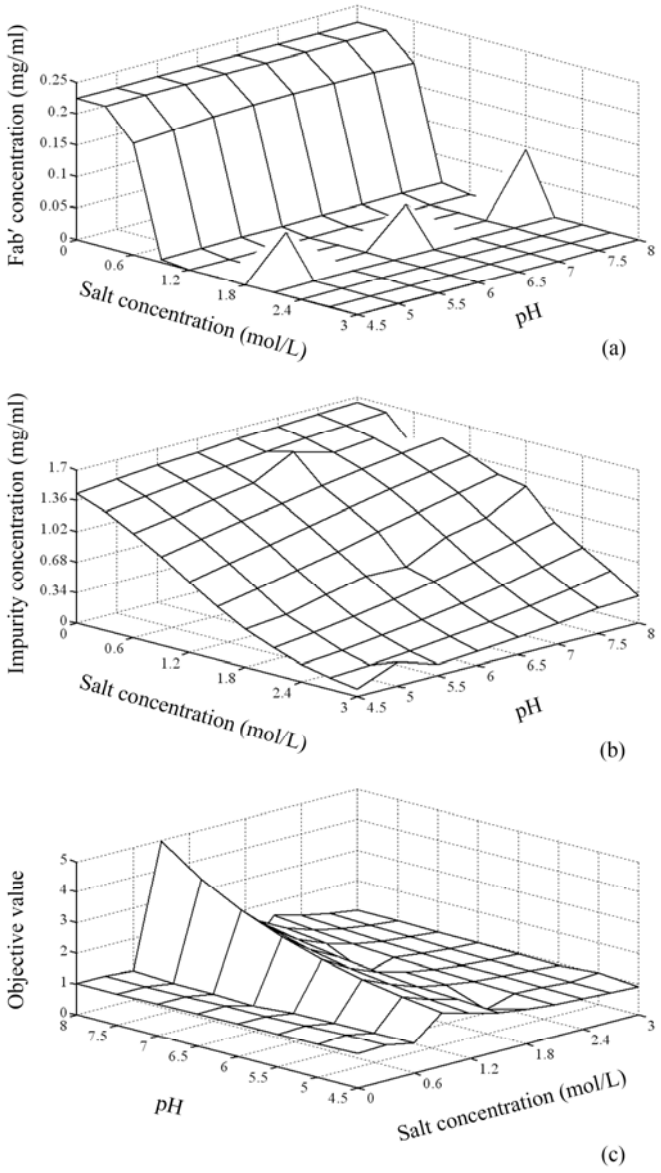


Figure 5.2 The predicted protein solubility surfaces from the 1st run: (a) Fab' in clarified homogenate; (b) impurities in clarified homogenate; (c) the objective surface.

5.3.2.2 The second run

The results from the four experiments designed in the first run (Table 5.3) were input into the algorithm and started the second run of model based design. The information offered by these extra four experiments was quite rich and changed the surface dramatically, shown in Figure 5.3. The objective surface (Figure 5.3.(c)) had the same pattern as the real surface with only 13 sample points (9 DoE points plus 4 new points), although the detail information was still not accurate enough. The predicted precipitation surfaces (Figure 5.3.(a) and Figure 5.3.(b)) started to show the conventional sigmoid curve with certain pH effects on the slops of surfaces. However, due to the lack of the information at the low salt concentration, the model can only predict the solubility value decreasing at the beginning and can not demonstrate the salting in phase, if compared to the real surfaces in Figure 5.1. The objective SD^2 value decreased significantly to 0.0171 with predicted maximum value at 1.176 (Table 5.3), which meant the models were greatly improved by just one run of D-optimal design.

5.3.2.3 The third run

Another four points were designed based on the current knowledge following the second run and repeated in a same point given by both Fab' D-optimal design and impurities D-optimal design (Table 5.3). In this case, the next set of experiments had only three new experiments. The data were feedback to the algorithm and surfaces were plotted in Figure 5.4. In this run, the information extracted from last run design brought in small changes on the models particularly in the low salt concentration area and the SD^2 for Fab' solubility model slightly decreased. Thus the design and optimisation continued.

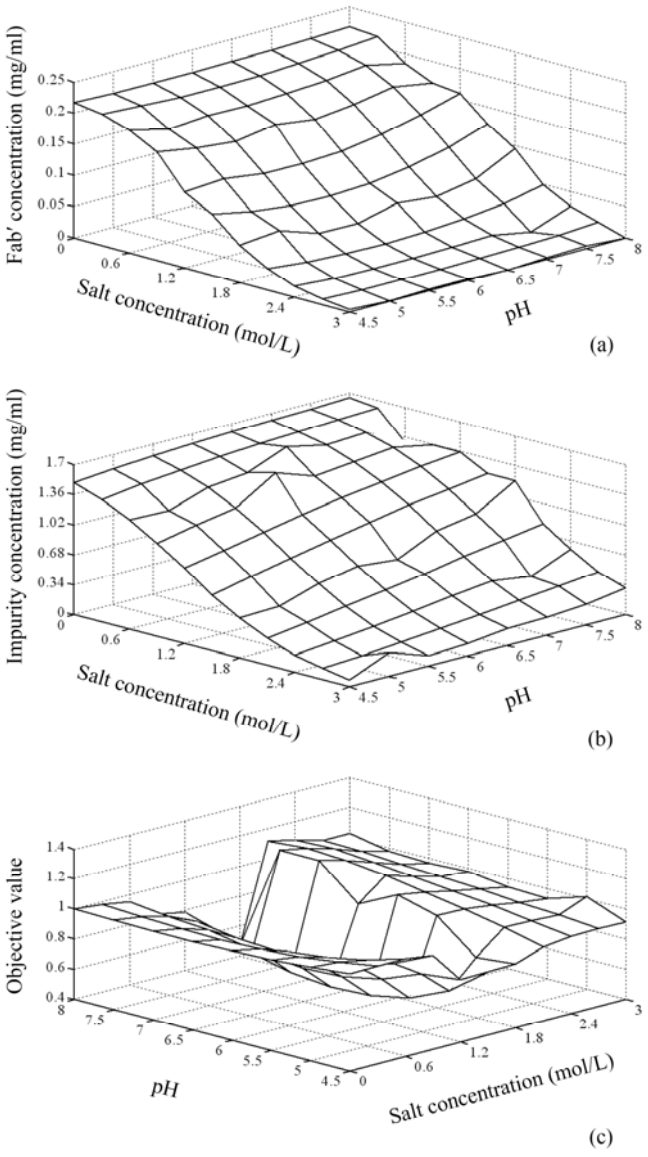


Figure 5.3 The predicted protein solubility surfaces from the 2nd run: (a) Fab' in clarified homogenate; (b) impurities in clarified homogenate; (c) the objective surface.

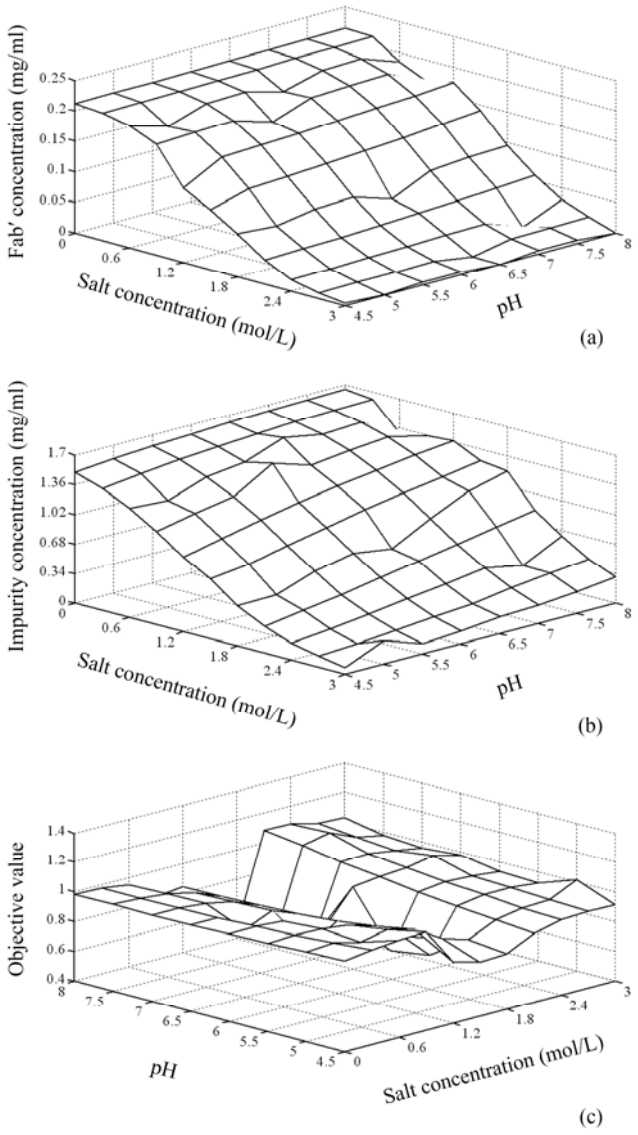


Figure 5.4 The predicted protein solubility surfaces from the 3rd run: (a) Fab' in clarified homogenate; (b) impurities in clarified homogenate; (c) the objective surface.

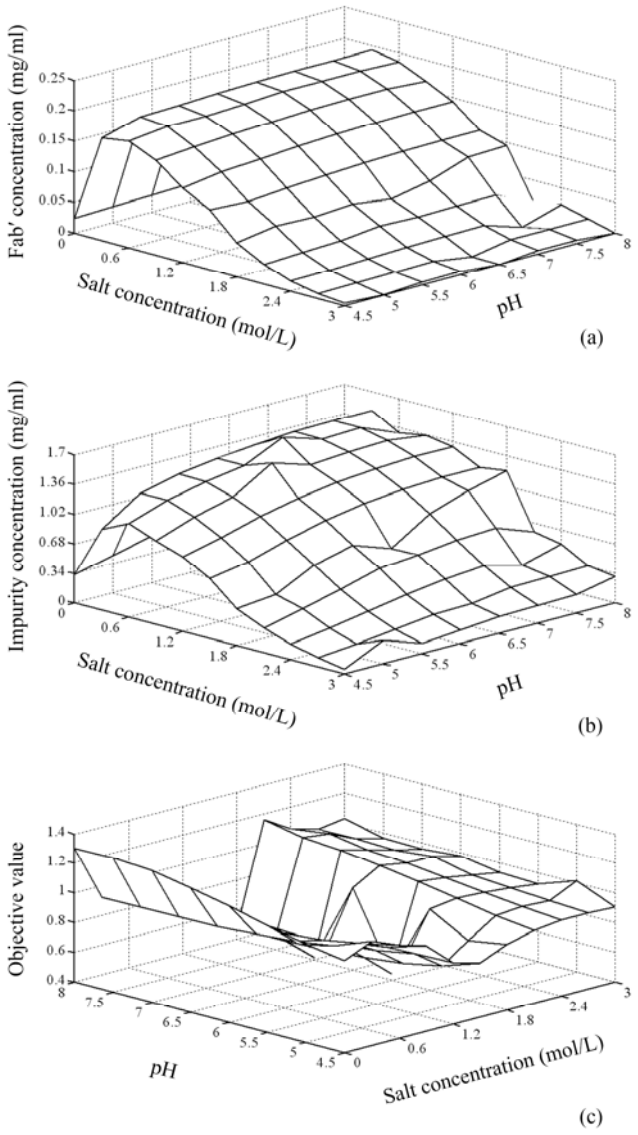


Figure 5.5 The predicted protein solubility surfaces from the 4th run: (a) Fab' in clarified homogenate; (b) impurities in clarified homogenate; (c) the objective surface.

5.3.2.4 The fourth run

The experimental points designed from the third run started to locate at the low salt concentration area. It was mainly driven by D-optimal design, which tried to make the parameters more accurate. The optimisation algorithm gave the point around the real maximum point (salt concentration 2.1 and pH 8.0). Simplex still iterated around the maximum area due to the relatively small operation windows, in which required very few points to leap to the possible maximum region.

The predicted surfaces were shown in Figure 5.5. Because the points at low salt concentration were introduced into the data, both precipitation surfaces began to have the same salting in phase, but the information was not accurate enough by only two points thus the objective surface gave the inaccurate pattern at low salt concentration range, which demonstrated high pH and low salt concentration would give the best results. The SD^2 values for precipitation models decreased due to more accurate parameters regressed except SD^2 value for the objective surface, but none of them hit the criteria set before (Table 5.3). Therefore, a new set of four points was given and all points produced by model algorithm were focusing on the low salt concentration area while Simplex was not influenced by inaccurate model parameters and still jumping around the real optimal area.

5.3.2.5 The fifth run

After inputting the latest four point results, the predicted surfaces were shown in the Figure 5.6. Compared with the surfaces in Figure 5.1 correspondingly, the predicted surfaces were very close to the real brute force experimental surfaces. Both precipitation surfaces showed the salting in effects and salting out effects perfectly and also demonstrated pH factor on the solubility. The objective surface (Figure 5.6.(c)) had the correct pattern and the surface around low salt concentration also gave the right trend.

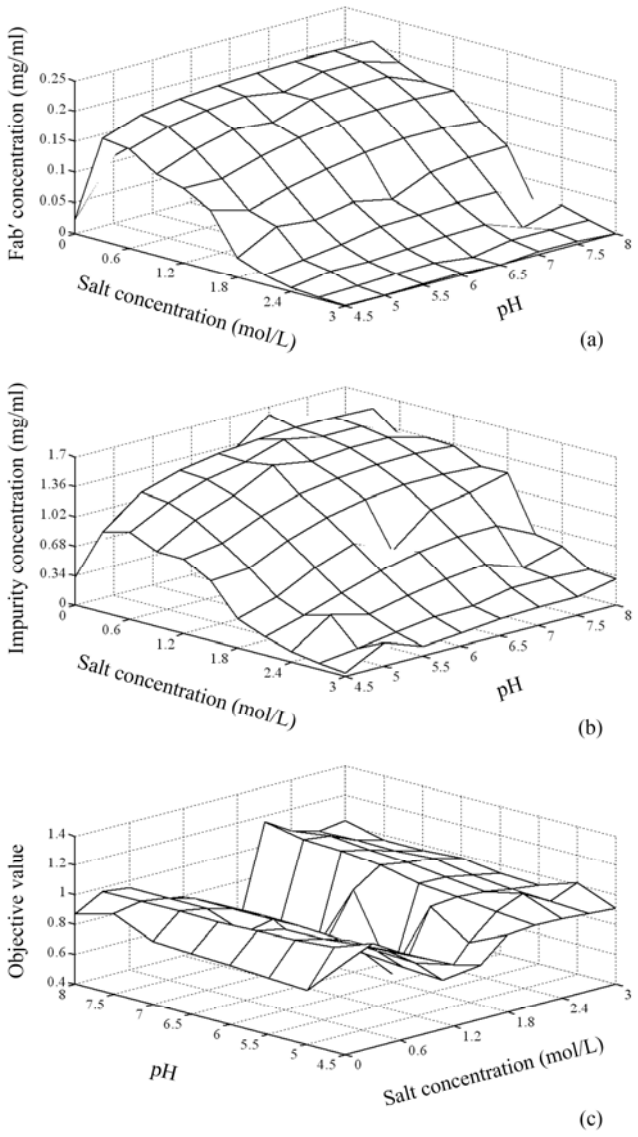


Figure 5.6 The predicted protein solubility surfaces from the 5th run: (a) Fab' in clarified homogenate; (b) impurities in clarified homogenate; (c) the objective surface.

The SD^2 shown in Table 5.3 had already achieved the termination criteria with all values lower than 0.05. Since the information of protein behaviour at low salt concentration added into model, the SD^2 of two models significantly decreased to less than 0.01, which also lowered the SD^2 of objective surface to 0.0063. The maximum point was found at pH 7.5, salt concentration 2.1M, which was the real optimal operation condition. Predicted maximum objective value was 1.109, while the real value was 1.219, less than 10 % variation. Although the algorithm still gave four next experimental points, the optimisation was achieved and the model produced through this procedure had the best accurate parameters to describe both precipitation processes with only totally 24 experimental points (9 DoE points plus 15 points designed by four runs of model based algorithm).

Run No.	Fab' d-optimal		Impurities d-optimal		Optimal point / random		Simplex		SD^2 Fab'	SD^2 Impurities	SD^2 Objective	Maximum value
	Salt (M)	pH	Salt (M)	pH	Salt (M)	pH	Salt (M)	pH				
1	0.9	6.0	1.2	4.5	0.9	8.0	2.4	7.5	0.1799	0.0537	0.4624	4.949
2	1.5	4.5	1.5	4.5	1.8	7.5	2.7	6.5	0.1006	0.0584	0.0171	1.176
3	0.0	4.5	0.3	4.5	2.1	8.0	2.1	6.0	0.0961	0.0588	0.0226	1.088
4	0.0	8.0	0.0	7.0	0.0	7.5	2.4	5.0	0.0759	0.0442	0.0497	1.307
5	0.0	5.5	0.0	5.0	2.1	7.5	1.8	4.5	0.0016	0.0061	0.0063	1.109

Table 5.3 The conditions designed by model based algorithm in each run and the SD^2 values with the maximum value at predicted optimal point.

5.3.3 Comparison with traditional DoE design

The 9 points DoE (Table 5.1) which used as the initial design in above algorithm was used again for the traditional DoE method comparison. The data were regressed to a second order polynomial equation and response surfaces were plotted in Figure 5.7. The traditional polynomial model used in DoE design was not sufficient to predict the dramatically change of the real surfaces. The predicted objective surface, shown in Figure 5.7 (c), had no detailed information as being obtained through

above algorithm. Both the value of optimal point and the trend towards the optimal area were far away from the correct one. Even the precipitation surfaces, which were relatively continuous and smooth in the real situation, had poor prediction from the model. Form the predicted objective surface, the next round DoE will design 9 points to explore a small design space at low pH and high salt concentration, which will further lose the chance to find the global maximum point. Therefore, the widely used sequential DoE with second order polynomial equation was of quite risk to be utilised to investigate the optimal point with a complex objective function and only few points providing information (Box, 1970).

Apparently, it is not fair enough to use only 9 or 18 points DoE to compare with a method using 24 points. Then, the same 24 points, which were used to extract information and support optimisation in above model based algorithm, were utilised subsequently in a new polynomial fitting in order to compare the efficacy of two methods fairly.

All predicted surfaces were shown in Figure 5.8. The results were much better than 9 points polynomial fitting in Figure 5.7. Surfaces had the same trends and pattern as the real surfaces. However, the detailed information was less provided than model based algorithm using the same set of experiments. As shown in Figure 5.8.(a) and Figure 5.8.(b), the predicted precipitation surfaces deceased continuously without limits, which caused negative concentration in predicted Fab' surfaces. The misleading negative value had no physical meaning in real process and casted doubts on whether a traditional DoE design and following methods have the capacity to describe a bioprocess correctly. The predicted objective surface in Figure 5.8.(c) suggested a direction to the optimal area outside of the operation windows, which gave a misleading judgement that the optimal process should be carried out at salt concentration and pH as high as possible. It definitely missed the real optimal area and may require more following studies at a new operation area.

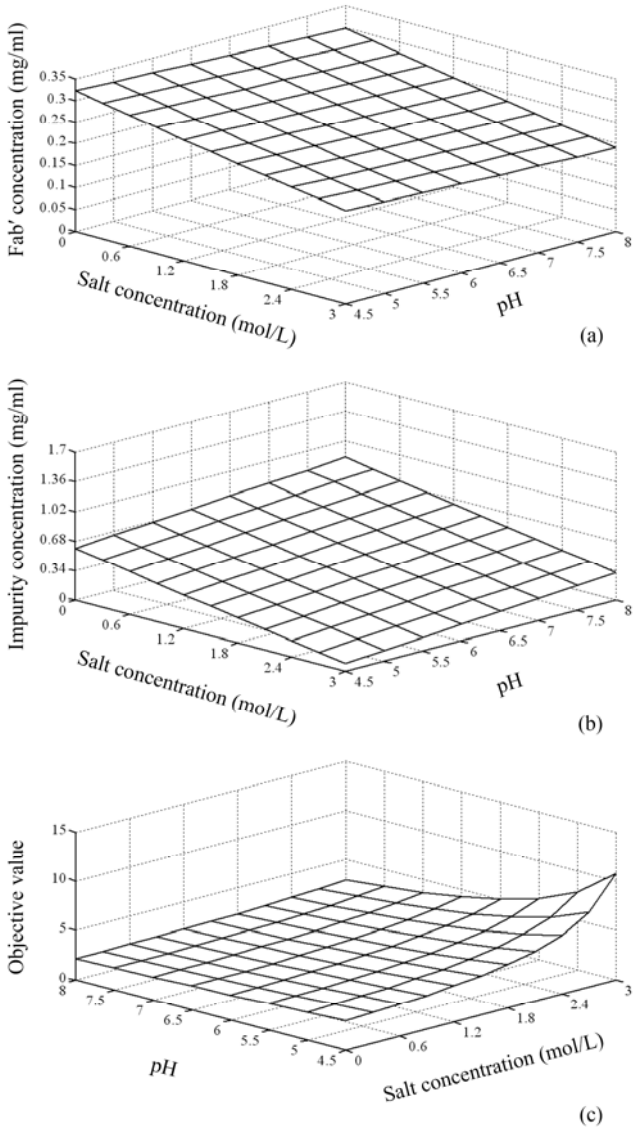


Figure 5.7 The predicted protein solubility surfaces from polynomial DoE (9 points): (a) Fab' in clarified homogenate; (b) impurities in clarified homogenate; (c) the objective surface.

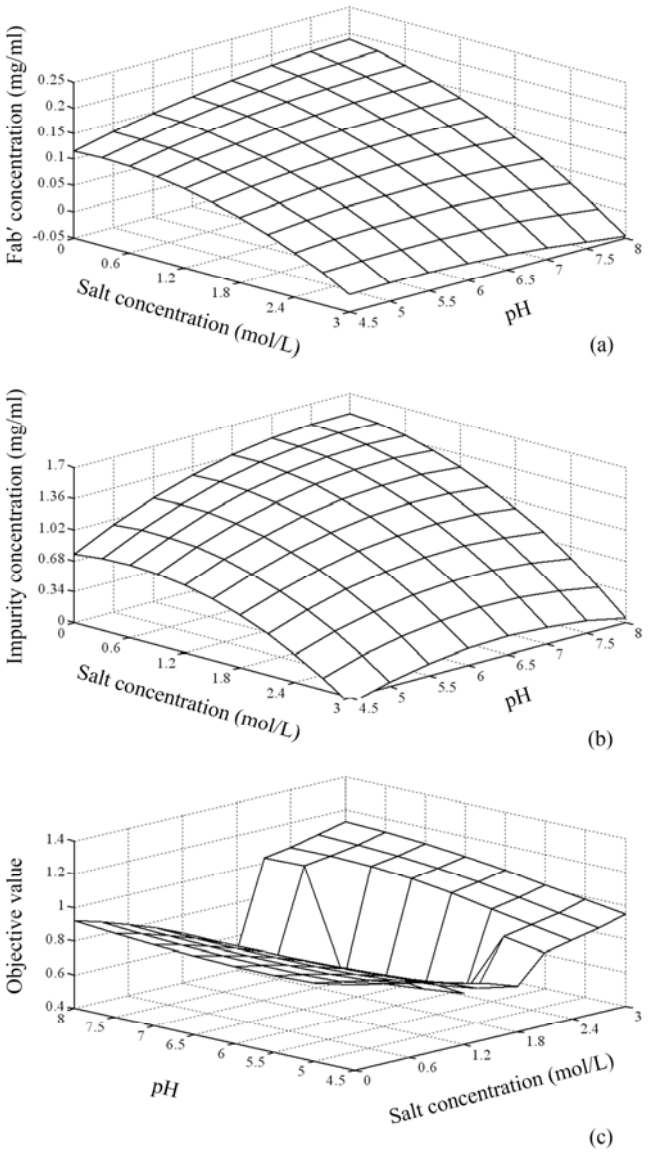


Figure 5.8 The predicted protein solubility surfaces from polynomial DoE (24 points): (a) Fab' in clarified homogenate; (b) impurities in clarified homogenate; (c) the objective surface.

5.4 Conclusions

In this chapter, it was well demonstrated how to start with a process model structure to successfully find the optimal point in a Fab' precipitation process. The new model based design and optimisation method only requires 24 experimental points in 5 runs to locate the maximum point according to the objective value. The conventional DoE, which was widely used in the experimental design and optimal search in quite broad fields, was not powerful to describe and optimise this nonlinear bioprocess accurately. It missed the optimal area using the same set of experimental points. At the end of the algorithm, accurate mathematical models were also given by the algorithm to characterise the bioprocess with SD^2 values less than 5 %.

The new design methodology showed better effects to speed up optimal conditions identification and bioprocess characterisation in the early process development. Not only the optimisation can be achieved successfully with few experiments, validated process models were also provided after the completion of optimisation, which can be utilised to further support some regulatory requirements or recommends, such as QbD. The only drawback of this algorithm was that it required at least one model to describe the bioprocess. However, considering more and more importance was emphasised on bioprocess modelling, this algorithm was very useful for the future optimisation of various bioprocesses, which can greatly reduce the cost, labour and material requirement in the development phase.

Chapter 6. Model based process design for bioprocess optimisation: a case study in alternative mAb purification process development

6.1 Introduction

The advances in mammalian cell culture development have made large scale monoclonal antibody production with 5 g/L or higher titres possible (Low et al., 2007). This efficiency in cell culture increasingly presents challenges to achieve economic downstream purification processes of antibodies considering the strict safety and quality requirement set by the regulator. The biological material containing monoclonal antibody produced from mammalian cell culture has complex impurities including impurity proteins, host genomic DNA, endotoxin and other potential hazards such as allergy molecules (Wurm, 2004). For therapeutic antibodies, these impurities must be removed and the residual impurities in the product must be controlled to an extremely low level to ensure the safety of the drug (FDA, 2001). Current practice relies on a general mAb purification platform, which is typically of three chromatographic steps starting with protein A chromatography, to produce high quality drugs (Shukla et al., 2007). Affinity protein A chromatography is the most effective step in downstream antibody purification and may achieve up to 95% purity in one single step, depending on the compositions of feedstock (Shukla et al., 2007; Sommerfeld and Strube, 2005; Roque et al., 2007).

However, when handling large scale higher titers feedstock, protein A chromatography has its own disadvantages (Gottschalk, 2008). The significant increase of mAb titre in the feedstock requires large volume of expensive protein A resin to capture antibody due to the relatively low binding capacity of resin. Nevertheless, the volume of the column is practically limited due to the reduced wall support in the large diameter column and the physical properties of the beads, both of which will slow the flow rate by increasing pressure drop (Stickel and Fotopoulos,

2001). As protein A chromatography is operated in the batch mode, the only solution is to increase the number of column and purify feedstock in parallel, which will demand a large initial capital investment and higher operation cost. Currently, although many efforts focus on protein A chromatography improvements, such as increasing the resin binding capacity or beads rigidity (Hober et al., 2007), or on membrane chromatography, the manufacture cost will still be high due to a large number of processing cycles, high concentrations of column-fouling materials such as cell debris or lipids, and more frequent use of harsh cleaning agents (Tugcu et al., 2008). Therefore, an economical process which is able to reach nearly the same quality results of protein A chromatography or able to significantly reduce the bioburden of feedstock before applied to protein A chromatography, as well as easy to scale up and fit into currently manufacturing platform, is extremely desirable (Shukla and Thommes, 2010).

In this chapter, the model based process design method will be used to find optimal mAb precipitation conditions at microwell scale experimentation. The model structure developed for Fab' precipitation in Chapter 4 will be used to examine how well the model structure can be transferred to a different protein and test the generality of the process design method in another precipitation process. In order to further examine the precipitation at process scale, the lab scale precipitation followed by solid-liquid filtration separation will be developed and key process performance, e.g. yield, purity, will be studied.

6.2 Methodology

Protein precipitation has been widely used in biotechnology for example, in blood product fractionations (Cohn et al., 1940; Cohn, 1941; Edsall, 1947). Recently, polyelectrolytes e.g. PEG and short-chain fatty acids also became useful precipitants for purifying protein or removing impurities, especially in mAb purification (Atha and Ingham, 1981; Ingham, 1990; Ma et al., 2010; McDonald et al., 2008). The

precipitation process can be continuously operated at very large scale e.g. thousands litre, with relatively low cost and thus is well suited to high titre biological feedstock (Glatz et al., 1986; Knevelman et al., 2009; Starvrinides et al., 1993). The operation time will be short due to the fast reaction with continuous operation. The facility requires only simple cleaning and easy maintenance. Another advantage of precipitation is the effect of concentration, which need fewer buffers and cleaning reagents, making the manufacturing process more economic (Glatz et al., 1986; Starvrinides et al., 1993). In this study, the potential of protein precipitation as an alternative approach for mAb recovery with the intention to eliminate expensive protein A chromatography in the primary recovery step will also be investigated.

In order to develop this process, the systematic model based process design approach proposed in Chapter 3 was adopted, shown in Figure 6.1 (same as Figure 3.2), which was well demonstrated in Chapter 4 and Chapter 5 by a Fab' precipitation case study at microwell scale. The mAb precipitation process was first developed and optimised through modelling and model based design in microwell format. Then the designed process was investigated at lab scale. In order to be applied to lab scale, process modification and improvements were necessary during these procedures, for example, finding a proper lab scale solid-liquid separation process. The possible improvements were tested and evaluated based on real results since no model was available for technical feasibility tests.

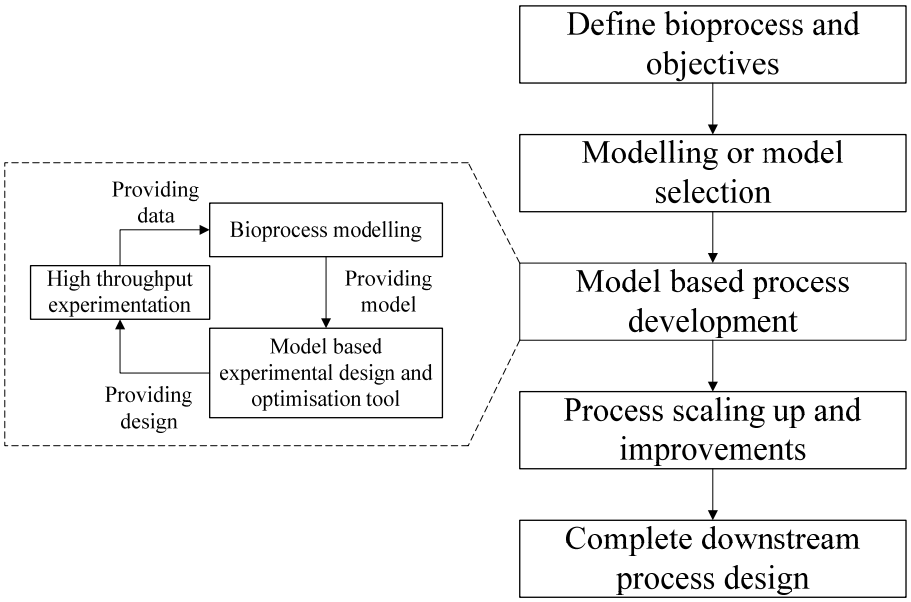


Figure 6.1 The whole process development and optimisation flowchart for mAb process design.

6.2.1 Process objectives

The objective of this study was to design and optimise a precipitation based mAb purification process. The process was selected from two precipitation systems with ammonium sulphate and PEG 6000 as precipitant respectively. Then it was further evaluated as an alternative mAb capturing step in the general purification platform. Therefore, working as the initial primary recovery process, the mAb yield was emphasised as the main target while the purity should be as high as possible. Similar to the objective equation (5.2) in Chapter 5, the mAb yield was set to be more than 70% recovered from solids, otherwise collect mAb from supernant. The corresponding objective value, yield and purity were calculated as following:

$$Objective\ value = 0.4 \cdot mAb\ yield + 0.6 \cdot purity - 0.1 \cdot precipitant\ concentration \quad (6.1)$$

$$mAb\ yield = \begin{cases} \dfrac{C_{l\,mAb}}{C_{T\,mAb}} & \dfrac{C_{l\,mAb}}{C_{T\,mAb}} > 30\% \\ 1 - \dfrac{C_{l\,mAb}}{C_{T\,mAb}} & \dfrac{C_{l\,mAb}}{C_{T\,mAb}} < 30\% \end{cases}, \quad (6.2)$$

$$purity = \begin{cases} \dfrac{C_{l\,mAb}}{C_{l\,mAb} + C_{l\,impu}} & \dfrac{C_{l\,mAb}}{C_{T\,mAb}} > 30\% \\ \dfrac{C_{T\,mAb} - C_{l\,mAb}}{(C_{T\,mAb} - C_{l\,mAb}) + (C_{T\,impu} - C_{l\,impu})} & \dfrac{C_{l\,mAb}}{C_{T\,mAb}} < 30\% \end{cases} \quad (6.3)$$

$C_{l\,mAb}/C_{T\,mAb}$ is the normalised concentration in the supernatant for mAb, $C_{l\,mAb}$ and $C_{l\,\mathrm{impu}}$ are the real concentration in the supernatant without normalisation for mAb and impurities respectively, $C_{\mathrm{T}\,mAb}$ and $C_{T\,\mathrm{impu}}$ are the maximum concentration for mAb and impurities in the solution respectively, precipitant concentration is in normalised value for both ammonium sulphate and PEG.

Since the yield has already a constraint of more than 70% in equation (6.1), the weight of yield was designed as 0.4, with the weight of purity at 0.6. The third term in equation (6.1) was the penalty term for the precipitant used in order to minimise the reagents used in the process. This objective function was used to optimise the

mAb precipitation process in the following microwell experiments by using model based experimental design algorithm, which has been developed in Chapter 3, as shown in Figure 3.7.

The mAb yield and final product purity in lab scale studies were investigated in more detail according to FDA quality requirements. The contaminants, such as impurity proteins including CHO host cell protein (HCP) and residual DNA were tested during lab scale precipitation and filtration in order to further optimise and evaluate this process. The optimised results were also compared with a standard post protein A chromatography solution based on the mAb yield, purity, HCP and DNA removal. The results and comparison elucidated whether the precipitation would be effective or not as an alternative approach in monoclonal antibody purification.

6.2.2 Microwell modelling

The next step in this case study was the mAb precipitation evaluation and precipitants selection. However, precipitation was not a highly selective approach. Its performance was influenced by many factors, particularly the feed material composition, pH and precipitant concentration (Kuczewski et al., 2010). The optimal conditions may vary substantially for different antibodies. When facing this type of investigation, high throughput experimentation was very useful due to its rapid process, simultaneously multifactor screening and little material requirement. Knevelman et al. (2010) had already demonstrated mAb precipitation in microwell plates by high throughput screening for pH, PEG and initial mAb concentration conditions. Previous Fab' case study in Chapter 4 and Chapter 5 also proved the feasibility of microwell precipitation. Therefore, the microwell scale mAb precipitation can be used as an applicable rapid platform in this model based process design to facilitate the development in this study.

The precipitation model developed for Fab' in Chapter 4 was transferred and applied

to mAb precipitation system. The precipitation model was based on the equilibrium phase and was generic to any precipitation. However, as mentioned before, the coefficients of the model depended on the specific material composition and should be estimated again by experimental data. Then the model and parameter should be judged and validated by statistic tests, as the flowchart in Figure 4.1. Therefore, if a new protein was used, the specificity of the material may change the significance of the coefficients in the model and thus may change the structure of the model, due to protein properties, such as the sensitivity to the salt or pH. A preliminary experiment using pure material was an easy and prudent step to collect the information of model, whether the model needed modification or simplification due to the new material, and also to give the best initial coefficient guess of the model for complex feedstock. Because the precipitation model developed in Chapter 4 was based on the first principle, the model structure can be modified according to the precipitation theory and the observed real experimental data. A set of pure mAb precipitation in microwell was used as the preliminary experimental study for the initial model structure evaluation and initial parameter guess for the following model based experimental design in mAb cell culture.

6.2.3 Model based mAb precipitation development and optimisation

In order to select the best precipitant and operating conditions, two sets of microwell precipitation were carried out using a clarified cell culture containing mAb with ammonium sulphate and PEG respectively at various concentration/percentage and pH, designed by model based experimental design algorithm. The mAb precipitation model with the initial coefficients, which was found from the study in 6.2.2, was used in the model based experimental design. Same as the design method in Chapter 5, one DoE set was used as the initial input and another set of independent DoE was used as validation dataset, shown in Table 6.1 and Table 6.2, for ammonium sulphate and PEG respectively. The criterion for either initial DoE or validation DoE selection was to cover all design space as much as possible or cover the most interested area.

Therefore, all initial DoE designs adopted in this study used the points located around the whole design space due to the design space was relatively small. The validation DoE sets used different pH conditions and relatively small salt concentration range in ammonium sulphate precipitation as the preliminary study showed high salt concentration range was more promising in process development. For PEG precipitation, both DoE set covered the whole design space as no prior knowledge were available about which range was more useful.

Experiment No.	Initial DoE		Validation DoE	
	Salt (M)	pH	Salt (M)	pH
1	0.0	5.5	0.6	5.0
2	1.0	5.5	1.4	5.0
3	2.0	5.5	2.2	5.0
4	0.0	7.0	0.6	6.5
5	1.0	7.0	1.4	6.5
6	2.0	7.0	2.2	6.5
7	0.0	8.5	0.6	8.0
8	1.0	8.5	1.4	8.0
9	2.0	8.5	2.2	8.0

Table 6.1 The conditions of initial DoE and validation DoE used in model based algorithm for mAb precipitation by ammonium sulphate.

Experiment No.	Initial DoE		Validation DoE	
	PEG percentage (% w/w)	pH	PEG percentage (% w/w)	pH
1	0	5.5	0	5.0
2	10	5.5	10	5.0
3	20	5.5	20	5.0
4	0	7.0	0	6.5
5	10	7.0	10	6.5
6	20	7.0	20	6.5
7	0	8.5	0	8.0
8	10	8.5	10	8.0
9	20	8.5	20	8.0

Table 6.2 The conditions of initial DoE and validation DoE used in model based algorithm for mAb precipitation by PEG.

The algorithm developed in Chapter 3, same as the Fab' case study in Chapter 5, gave four designed conditions for the next experiments in each run. PEG and ammonium sulphate precipitation had its own separate optimisation routine. The next set of experiments was then executed by Tecan in microwell plates and results were feedback to algorithm. The primary termination criterion was same as that in Chapter 5, which required variance (SD^2) for any models and objective values were all lower than 0.05 (5%). An extra requirement was added once the primary termination criterion was reached. The new requirement was that if the real maximum point was not changed by consequent three runs, the algorithm will be terminated in order to avoid repeated design because in this case, unlike that in previous Fab' case study, there was no real objective surface based on a whole 96 microwell plate to compare with the predicted value. The algorithm was objective value orientated based on equation (6.1), (6.2) and (6.3) with the aim of process optimisation. The objective values were compared between two precipitation systems and only the precipitant with optimal conditions which achieved the maximum objective value was selected for later process study at lab scale.

6.2.4 Process development at lab scale and improvements

Subsequently, lab scale precipitation was executed based on the results of microwell precipitation optimisation. The following strategy was to select a feasible solid-liquid separation to retain the mAb in solids from the liquid or separate the impurity solids from the solution containing mAb. Solid-liquid separation by filtration was investigated in this study. Membrane pore size and process capacity were firstly screened due to lack of precipitation particle size information. A depth filtration system was chosen and designed to intercept the mAb solids on the proper pore size membrane. Thus the mAb solids can be washed by buffer to remove residual impurties and then resuspended by PBS buffer to recover mAb. The yield and purity of this process were evaluated and compared with those in protein A chromatography process.

Since more analytical results were available in lab scale precipitation, such as DNA and HCP assay, several innovative procedures, inspired by chromatography process e.g. intermediate washing, were adopted in lab scale precipitation and filtration to achieve higher purity in the final products. In order to evaluate these factors which can not be investigated in microwell studies, the further lab scale optimisation and improvements were mainly based on the experimental results without any model involved.

6.3 Results and discussion

6.3.1 Microwell modelling and design

In this preliminary experiment, pure mAb was precipitated in microwell plates by ammonium sulphate in pH 5.0-8.0. The mAb solubility surface was shown in Figure 6.2. The simplified Fab' precipitation model in Chapter 4, equation (4.22)

$$\frac{C_l}{C_T} = \frac{1}{1 + \frac{1}{f_1 + C_s} \cdot e^{(a_1 \cdot C_s + b_1 \cdot pH + c_1 \cdot pH^2 + d_1)}}. \qquad (4.22)$$

with the Fab' coefficients as the initial coefficient guess for mAb was applied to mAb precipitation results to regress the model. However, the R-square was only 0.83 with two coefficients failed parameter t-test. This meant the mAb had a different sensitivity to salt and pH. The parameter d_1 in the equation (4.22) failed the t-test, which means it was insignificant. Therefore, it can be the first parameter removed from model (4.22) to modify the precipitation model.

According to the t-test, the coefficient for the second order pH effect also had poor t-test value, suggesting the second order pH term was not needed. It can also be observed from Figure 6.2 that the pH did not have the same strong effect in mAb precipitation as in Fab' precipitation, so a linear pH term for the exponential function

was adopted to describe the trend. As mentioned in Chapter 4, the protein activity coefficient function (4.11), was simplified with only the first order of salt concentration in Fab' case. While the mAb precipitation showed the protein had a stronger reaction to the salt concentration than Fab', a second order salt concentration term was necessary to be added into model to describe the strong salt effect. Therefore, equation (4.22) was modified as following for mAb precipitation:

$$\frac{C_l}{C_T} = \frac{1}{1 + \frac{1}{f_1 + C_s} \cdot e^{(a_1 \cdot C_s + b_1 \cdot pH + c_1 \cdot C_s^2)}}. \tag{6.4}$$

Compared with equation (4.22), the second order pH term was replaced by the second order salt concentration term to show the differences between Fab' and mAb sensitivity to salt and pH.

The coefficient values and test value were validated and shown in Table 6.3 and achieved both above 95% confidence in t-test and F-test.

	Parameters				R^2	F-test value
	a_1	b_1	c_1	f_1		
Pure mAb	-4.97	0.58	10.00	3.09	0.916	145.1
t-test value (confide.)	-4.42 (99%)	2.47 (95%)	7.04 (99%)	4.15 (99%)	P>F	0.000 (99%)

Table 6.3 Parameters, F-test value and R^2 value of mAb precipitation model using equation (6.4).

6.3.2 High throughput mAb precipitation development and optimisation

6.3.2.1 Optimisation for ammonium sulphate precipitation

The model equation (6.4) and the coefficients from pure mAb precipitation were used to initialise the model based design algorithm. Figure 6.3 shows the predicted surface from the new precipitation model in the first run. The subsequent experimental conditions and SD^2 were listed in Table 6.4. The SD^2 values were all

smaller than 0.05, which proved the models for mAb precipitation and impurities were quite accurate after only one run in model based design method.

As in this study, there was no brute-force design dataset to compare the real surface with predicted surface, the algorithm continued until the second termination criteria achieved (raw data of each validation run shown in Appendix 3). Figure 6.4 and Figure 6.5 show the next two runs results with the designed points and SD^2 values in Table 6.4. There were only slightly improvements in SD^2 values, and the predicted surfaces changed little. The maximum points found by algorithm were listed in Table 6.5. Since the 2nd run, salt concentration 1.6 M and pH 8.5 were selected as the conditions for maximum point three times. Hence the algorithm terminated at the end of the 4th run. The predicted maximum objective value by model was 0.698, while the real objective value was 0.721, a little higher than the predicted value. Total 21 points were evaluated in this optimisation procedure.

Run No.	mAb d-optimal		Impurities d-optimal		Optimal point / random		Simplex		SD^2 mAb	SD^2 Impurities	SD^2 Objective	Maximum value
	Salt (M)	pH	Salt (M)	pH	Salt (M)	pH	Salt (M)	pH				
1	1.6	8.5	2.2	8.5	1.8	6.5	1.6	6.5	0.0360	0.0012	0.0119	0.669
2	1.2	5.0	2.2	5.0	1.8	8.5	2.2	6.0	0.0567	0.0011	0.0119	0.695
3	1.4	8.5	0.0	5.0	0.6	7.5	1.6	7.0	0.0303	0.0010	0.0097	0.694
4	1.4	5.0	2.2	8.0	2.0	7.5	1.8	5.0	0.0311	0.0008	0.0099	0.698

Table 6.4 The conditions from model based experimental design algorithm in each run and the SD^2 values at predicted optimal point (ammonium sulphate).

Maximum point from algorithm	Salt (M)	pH	Objective value
1st run	2.0	5.5	0.669
2nd run	1.6	8.5	0.721
3rd run	1.6	8.5	0.721
4th run	1.6	8.5	0.721

Table 6.5 The conditions at maximum point objective values (ammonium sulphate).

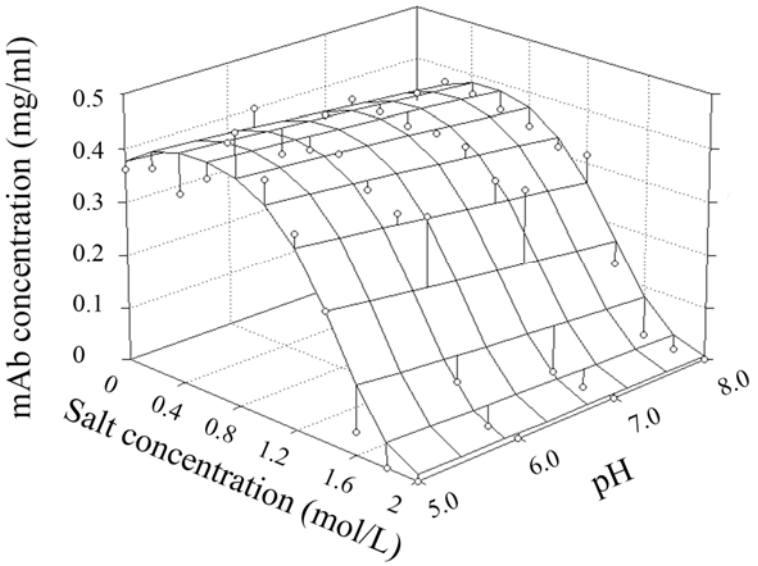


Figure 6.2 The predicted surfaces provided by modified precipitation model (6.4) with real experimental data shown in dots.

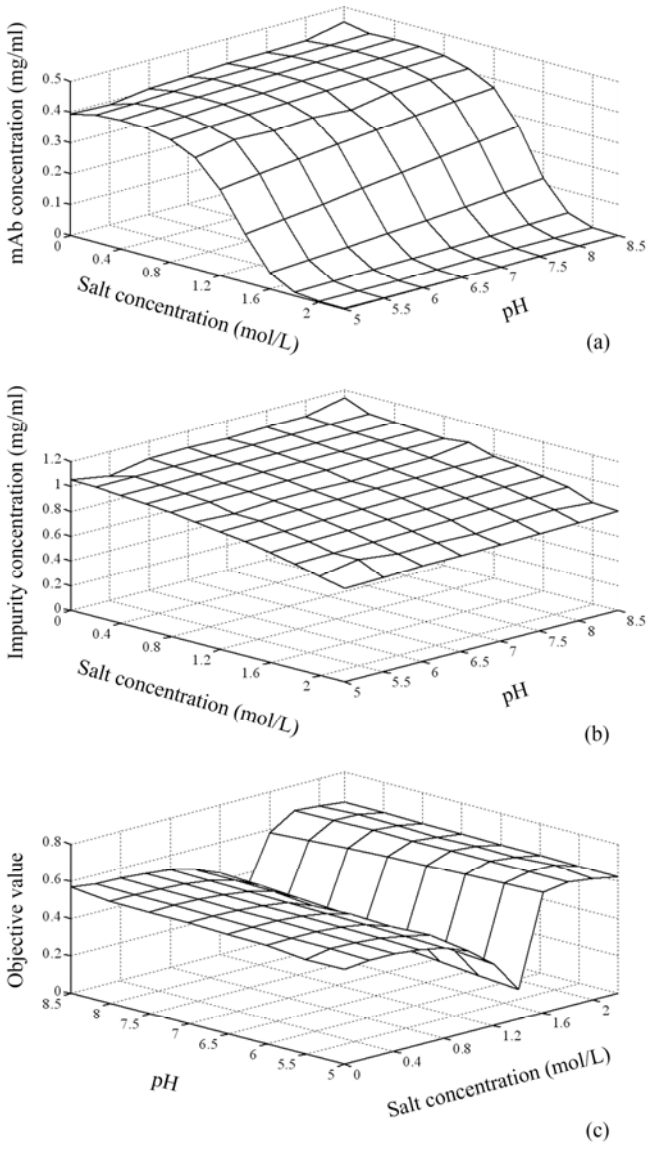


Figure 6.3 The predicted protein solubility surfaces from the 1st run (ammonium sulphate): (a) mAb; (b) impurities; (c) the objective surface.

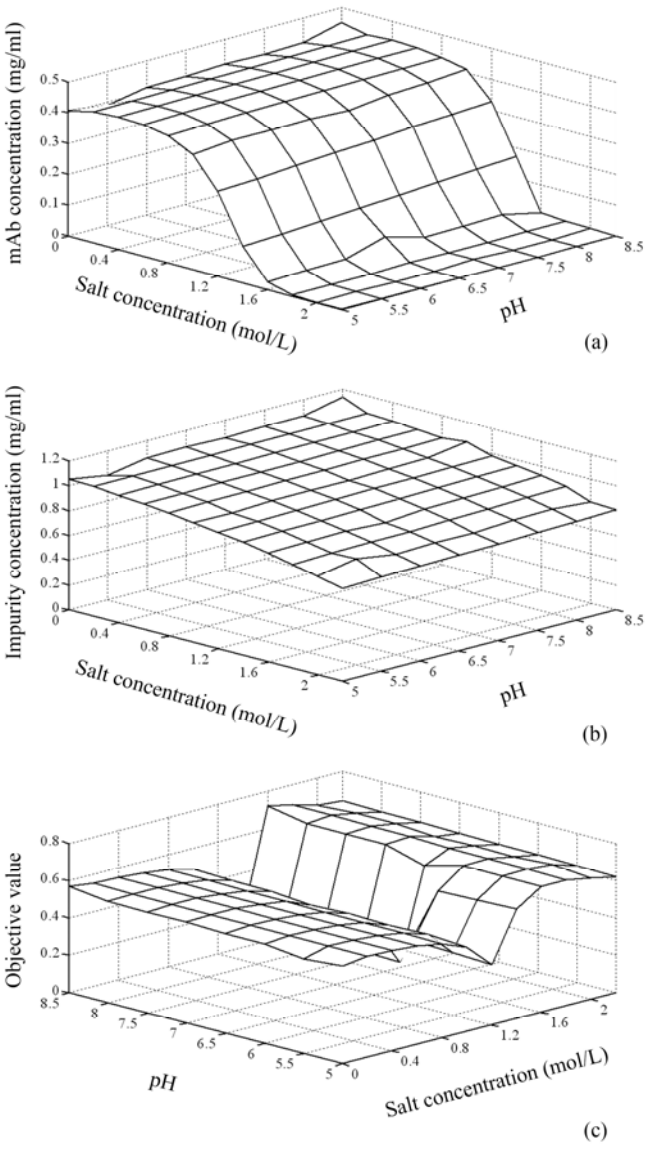


Figure 6.4 The predicted protein solubility surfaces from the 2nd run (ammonium sulphate): (a) mAb; (b) impurities; (c) the objective surface.

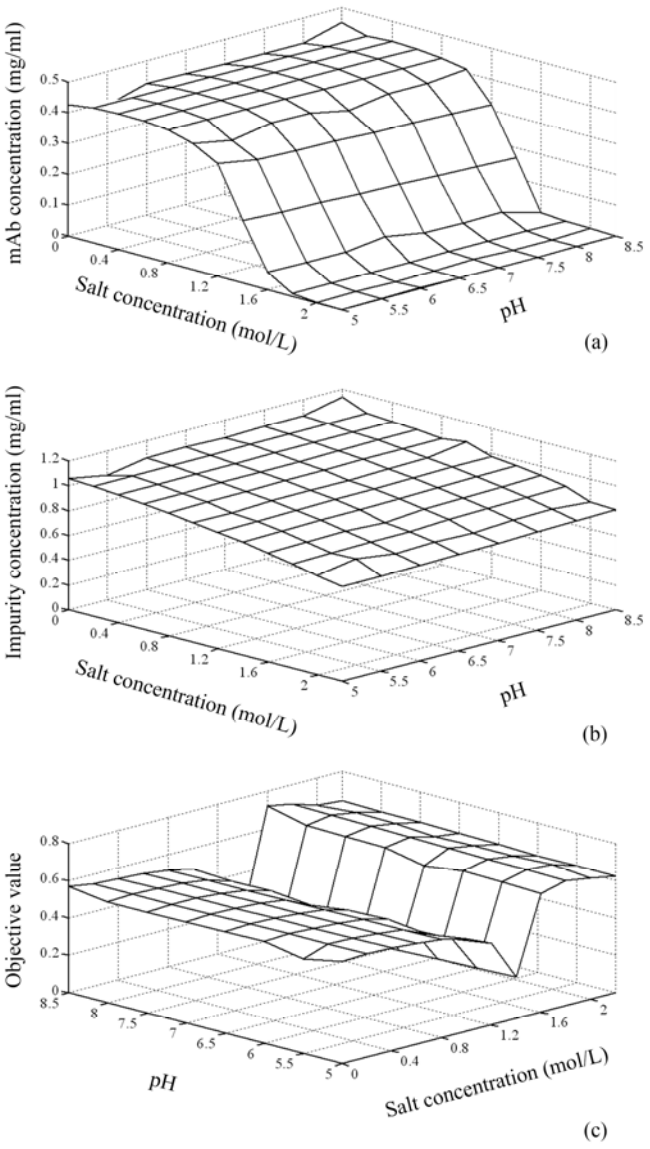


Figure 6.5 The predicted protein solubility surfaces from the 3rd run (ammonium sulphate): (a) mAb; (b) impurities; (c) the objective surface.

6.3.2.2 Optimisation for PEG precipitation

The same procedure was applied to PEG precipitation, which aimed to compare and select the best precipitant (raw data of each validation run shown in Appendix 3). The model (6.4) and initial guess from preliminary ammonium sulphate precipitation were used and the results showed the algorithm was quite flexible and the model fit well in PEG precipitation system. The SD^2 values after the 1st run were all lower than 0.01, as shown in Table 6.6. The designed experimental conditions in each run were also all listed in Table 6.6. In the 1st run and 4th run, both D-optimal designs gave repeat points, which reduced the total experiments points used. However, it took five runs to terminate the algorithm as the conditions of maximum point at PEG 14% (w/w) and pH 8.5, shown in Table 6.7 was found at 3rd run and kept for three times. Therefore, the total experiment point number was 23 in this case. The model predicted objective value at the optimal condition was 0.602, while the real objective value was 0.655.

Run No.	mAb d-optimal		Impurities d-optimal		Optimal point / random		Simplex		SD^2 mAb	SD^2 Impurities	SD^2 Objective	Maximum value
	PEG (%)	pH	PEG (%)	pH	PEG (%)	pH	PEG (%)	pH				
1	14	5.0	14	5.0	16	8.5	10	8.0	0.0047	0.0041	0.0014	0.593
2	8	5.0	12	5.0	14	8.5	10	7.5	0.0046	0.0046	0.0014	0.646
3	10	5.0	20	5.0	14	7.0	12	7.5	0.0038	0.0041	0.0011	0.620
4	0	5.0	0	5.0	8	8.5	2	8.5	0.0018	0.0036	0.0004	0.607

Table 6.6 The conditions from model based experimental design algorithm in each run and the SD^2 values at predicted optimal point (PEG).

Maximum point from algorithm	PEG (%)	pH	Objective value
1st run	0.0	8.5	0.590
2nd run	14	5.0	0.614
3rd run	14	8.5	0.655
4th run	14	8.5	0.655
5th run	14	8.5	0.655

Table 6.7 The conditions at maximum point objective values (PEG).

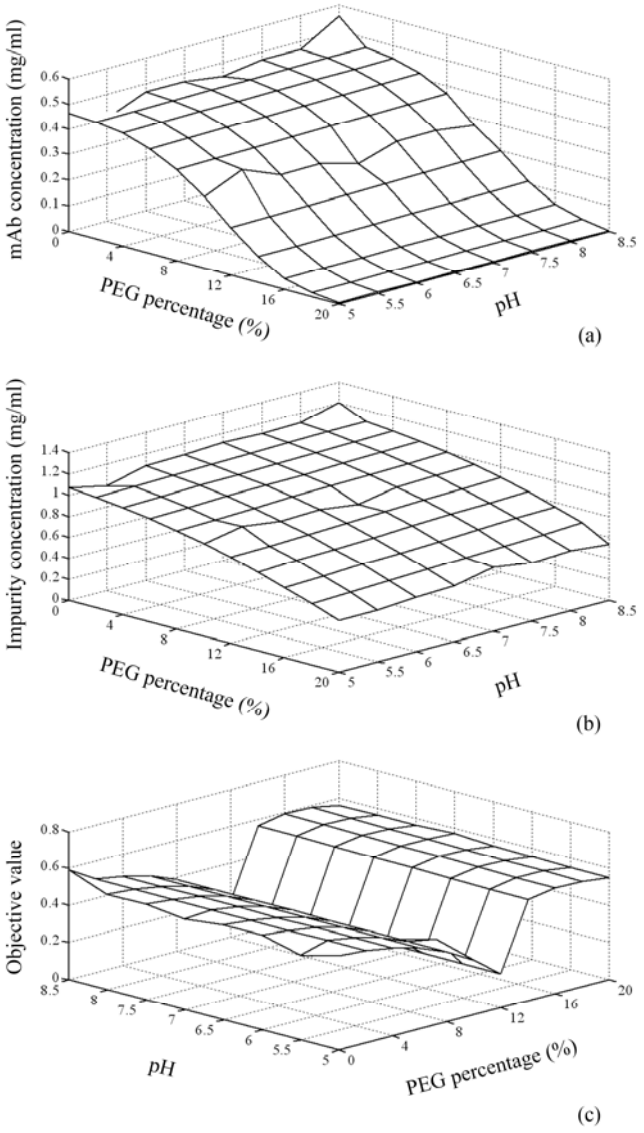


Figure 6.6 The predicted protein solubility surfaces from the 1st run (PEG): (a) mAb; (b) impurities; (c) the objective surface

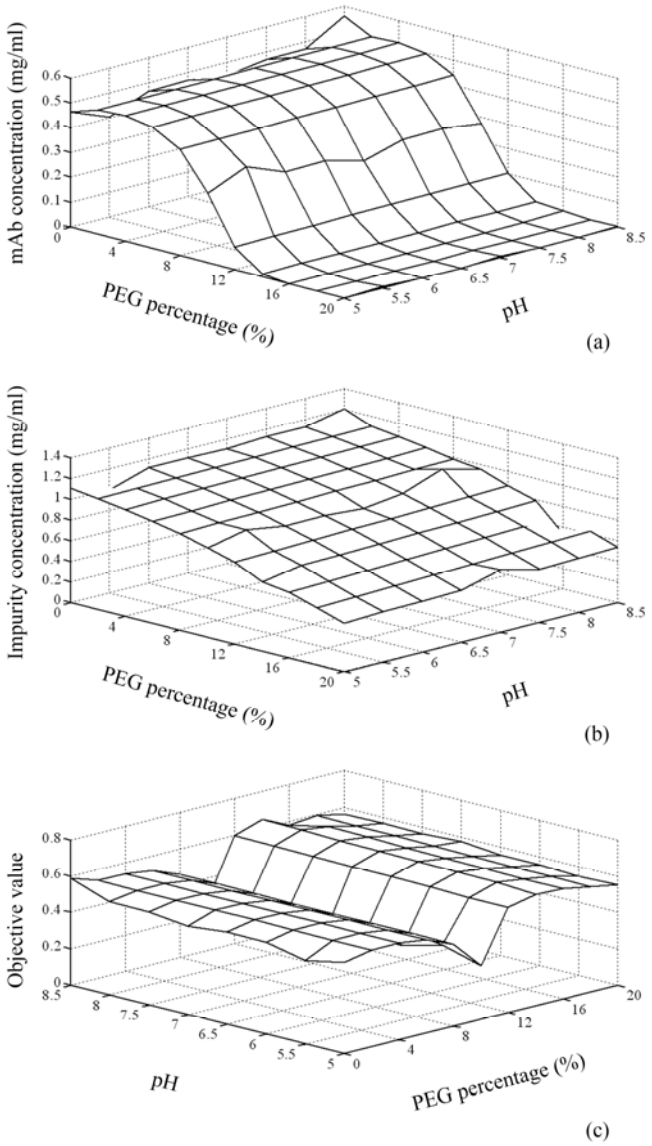


Figure 6.7 The predicted protein solubility surfaces from the 2nd run (PEG): (a) mAb; (b) impurities; (c) the objective surface.

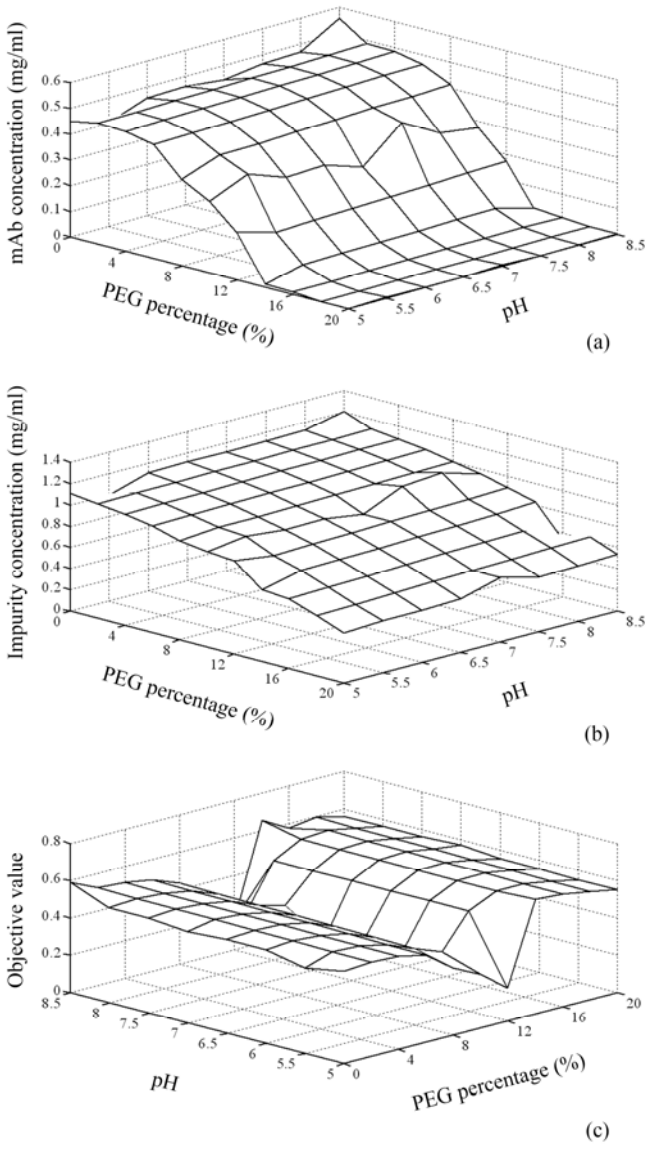


Figure 6.8 The predicted protein solubility surfaces from the 3rd run (PEG): (a) mAb; (b) impurities; (c) the objective surface.

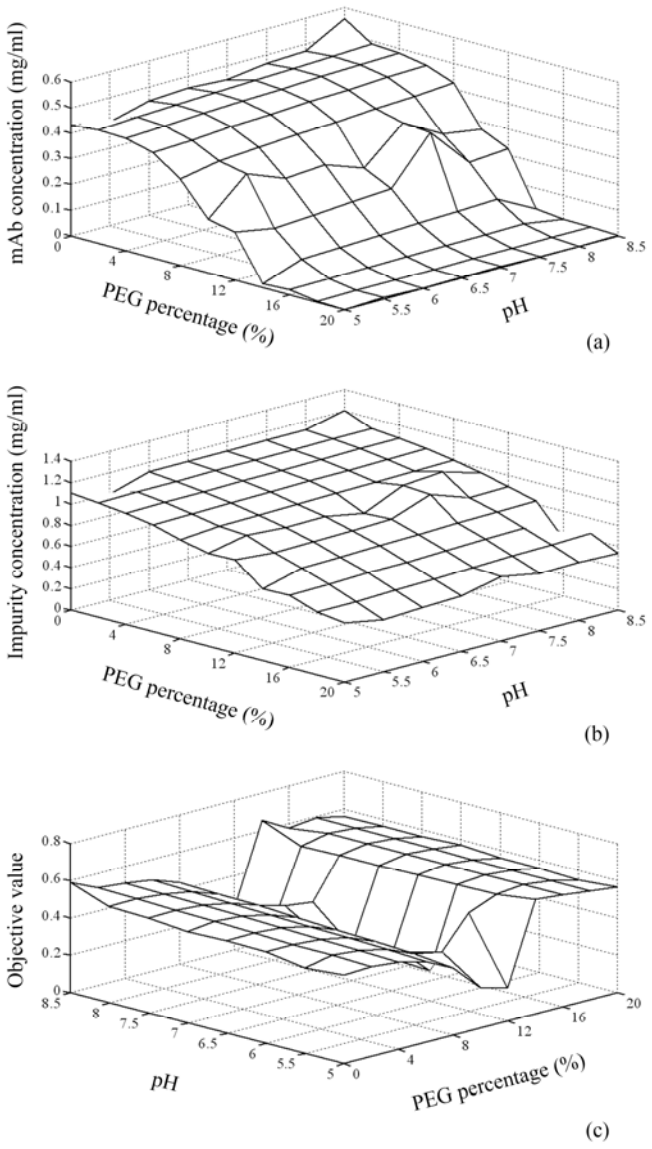


Figure 6.9 The predicted protein solubility surfaces from the 4th run (PEG): (a) mAb; (b) impurities; (c) the objective surface.

Figure 6.6 to Figure 6.9 show the predicted surfaces in PEG precipitation at each run. The surfaces changed slightly in each run but the SD^2 kept decreasing due to more information added to make model accurate.

The objective value at maximum point in PEG system was smaller than that in ammonium sulphate precipitation system. If compared the impurities surfaces, the results demonstrated that ammonium sulphate has more selective effects than PEG in this feedstock. When mAb was nearly precipitated in solid phase, the concentration of impurities in ammonium sulphate precipitation, shown in Figure 6.5 (b), only decreases 30%, while it decreased around 60% in PEG precipitation, shown in Figure 6.9 (b). It showed that ammonium sulphate precipitation would have higher purity in final product under the same mAb yield percentage, which was also proved by the objective values. Therefore, the optimal conditions 1.6 mol/L ammonium sulphate at pH 8.5 was preferred with 97.6 % yield and 67.3 % purity at microwell scale, and applied in later lab scale experiments.

6.3.3 Lab scale experiments of mAb precipitation with filtration

Since the conditions found in above study precipitated mAb out into solids, a solid-liquid separation process, which can capture solids and then recover the mAb solids, has been investigated in lab scale process development. In this step, dead-end membrane filtrations were first carried out with different pore size to test the capacities and screen the proper pore size. In order to recovery all mAb, the smallest pore size filter was required to keep all solids on the membrane, while in reality, the particle size of mAb precipitant was probably a normal distribution (Aucamp et al., 2005; Shih et al., 1992). In filtration, the relatively large particles will easily block the membranes and the low capacity of the process will become the bottleneck of downstream purification, shown in Table 6.8. Moreover, it was observed that the particle size of protein colloids was apparently affected by initial mAb concentration, the profile of other impurties in the feedstock, precipitation time, mixing speed and

ways of mixing (George and Wilson, 1994; Grabenbauer and Glatz, 1981; Tsao et al., 1951). Currently, there is no clear and general explanation or model between particle size and these variables. Therefore, in order to achieve higher capacity and a robust process, depth filters were adopted. The CUNO Zeta plus EXP 05 sp filter, which has a large pore size, coupled with CUNO Zeta plus EXP 30 sp filter intercepted the mAb solids in a broad size range (0.5-10 μm) based on the membrane pore size studies.

mAb concentration in feedstock (g/L)	Membrane filter pore size (μm)	mAb capacity on membrane ($g \cdot m^{-2}$)	Transmembrane flux ($L \cdot min^{-1} \cdot m^{-2}$)
	0.2	4.27	1.015
1.20	0.45	5.52	1.321
	1.2	9.43	1.915
1.75	5.0	9.41	1.194
2.30	5.0	9.32	1.234
1.20	0.5 - 10.0*	86.83	39.40

Table 6.8 Pore size sieving results. * depth filters coupled to form 0.5 - 10.0 μm range.

Table 6.8 shows the results of dead-end membrane filtrations with different pore size. For the feedstock with initial mAb concentration at 1.2 mg/ml, the filter with pore size of 5.0 μm did not successfully separate the solids and liquid (protein solids passing through membrance were observed). Some particles were much smaller than 5.0 μm and passed through the membrane easily. With the pore size of 1.2 μm, the membrane was efficient to hold all mAb solids with the capacity of 9.43 g/m^2 while smaller pore size membrane, 0.2 μm and 0.45 μm, had lower processing capacity.

For higher mAb concentration feedstock, mAb molecules had more chance to aggregate with each other and probably form large particles (McPherson et al., 1999). Both 1.75 mg/ml and 2.30 mg/ml mAb feedstock were tested and the separation of both feedstock after precipitation can be operated on a 5.0 μm membrane filter without mAb penetrating through the membrane.

However, the capacity of the 5.0 μm membrane for two high concentration feedstock did not increase very much as anticipated. This probably was caused by the maximum capacity of dead-end filtration itself as it was always around 9.4 g mAb /m^2 using any different concentration feedstock and different membrane pore size. All the dead-end filtration had low capacity and the transmembrane flux in the experiments was very low, shown in the Table 6.8. It was impossible to scale up and design a bioprocess based on such a low flux. In depth filtration with pore size from 0.5 to 10.0 μm, both the capacity and flux significantly increased. The maximum capacity was 86.83 g mAb /m^2 with 39.40 $L \cdot min^{-1} \cdot m^{-2}$ operational flux (~2400 $L \cdot m^{-2} \cdot h^{-1}$), which was feasible for scale up and further bioprocess design.

mAb concentration in feedstock (mg/mL)	Membrane filter pore size (μm)	mAb loaded (mg)	mAb recovered (mg)	Yield (% w/w)	Total protein in resuspension (mg including mAb)	Purity (% w/w)
1.20	1.2	2.93	2.47	84.3	4.56	54.2
1.75	5.0	2.80	2.67	95.4	3.04	88.0
2.30	5.0	9.37	8.16	87.1	8.75	93.3
1.20	0.5 - 10.0*	59.82	58.01	97.0	68.56	84.6

Table 6.9 mAb yield and purity in filtration studies. * depth filters coupled to form 0.5 - 10.0 μm range.

The mass balance data from above studies were also summarised in Table 6.9. For 1.2 g/L feedstock, 1.2 μm filter had the lowest yield and purity due to several reasons. Small pore size filter may have better effect to hold protein solids than 5 μm filter but in the washing step, fewer impurities were washed away due to the same reason. The low yield was attributed to the resuspension step. It was less efficient to dissolve mAb solids completely in the buffer on small pore size membrane in a short time and brought them through filter, while in 5.0 μm filter, the mAb can be quickly recovered and pass through the filter even in the form of fine particles, e.g. 1.0 μm diameter particles.

From Table 6.9, the best result for membrane filtration was 95.4 % yield and 88 % purity. The results from depth filter were more promising, achieving 97 % yield and around 85 % purity. Although the same condition in microwell achieved 97.6 % yield, which were better than any lab scale results, the lab scale results were still comparable to those of micro scale considering the protein lost and the complexity in lab scale process. The purity at lab scale was higher than that in microwell studies, which was 67.2 %. It may be casued by the same reason as explained in previous 1.2 μm membrane filtration results since the pore size in microwell filter plate was only 0.45μm and the mircowell precipitation did not have a washing step. Due to the structure of depth filter, depth filtration can capture mAb solids more efficiently than membrane as demonstrated by above results. High flow rate made washing step more throughout with less impurities remaining. It also made resuspension step easier and provided high recovery yield (Low et al., 2007). Therefore, the depth filtration was selected as the separation process at lab scale.

6.3.4 Process improvement

In order to further improve the purity, some innovative steps were introduced after precipitates captured in the depth filters. In resuspension step, gradient resuspension and fraction collection were utilised to further enhance the purity. Figure 6.10.(a) shows the results of gradient resuspension from ammonium sulphate concentration 1.2 M to 0.7 M. It was obvious that salt concentration at 1.1-1.2 mol/L was enough to dissolve mAb from filter and kept some impurity proteins in solids. The resuspenion buffer with its salt concentration lower than 1.1 M only increased the impurities and lowered the final purity. It was better than PBS resuspension since PBS has a much lower salt concentration than 1.0 mol/L which dissolved and introduced more impurties into final solution.

Fraction collection data was shown in Figure 6.10.(b). As the salt precipitation was not a highly selective process, the resolution between impurty peak and mAb peak

was not as good as expected. With mAb concentration peaking at 30 ml fraction volume, impurity concentration was also increasing. After 7th fraction collection (70ml), there was nearly no mAb in the solution while impurty concentration was quite stable in the fractions. Therefore, the first six fractions of 1.1 M resuspension were collected and the result was improved to 97.3% yield and 88.2% purity.

6.3.5 Host cell protein (HCP), DNA removal with a two cut precipitation process

6.3.5.1 HCP removal

Although the yield and overall purity by above precipitation process were reasonably good, the results from HCP analysis demonstrated simple one cut precipitation can not compete with protein A chromatography. The HCP in the feedstock was 242, 465 ng/mg mAb (ppm), post precipitation solution 90, 407 ng/mg mAb (ppm) and protein A elution pool 403 ng/mg mAb (ppm), as shown in Table 6.10. Previous researches on CHO HCP by several researchers (Arunakumari and Wang, 2008; Glynn, 2008) proved that HCP levels decreased with salt concentration increasing and lower pH precipitated more HCP (Mao et al., 2010). The one cut precipitation precipitated HCP together with mAb and thus increased the HCP levels in the final solution.

	mAb yield (w/w)	Purity (w/w)	HCP (ng/mg mAb, ppm)	DNA (ng/ml)
Feedstock	100%	28.4%	242, 465	6, 592
Post protein A	92.4%	95.7%	403	11
One cut precipitation	97.3%	88.2%	90, 407	256
Two cut precipitation	93.6%	98.2%	8, 737	15

Table 6.10 The comparison between protein A chromatography, one cut precipitation and two cut precipitation.

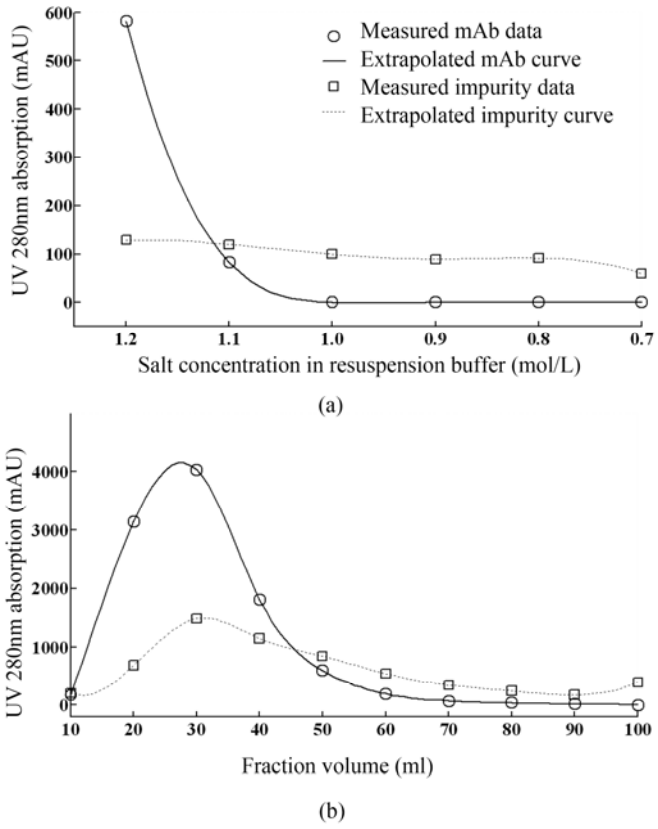


Figure 6.10 (a) The UV absorption of mAb and impurities in gradient resuspension procedure; (b) The UV absorption of mAb and impurities in 1.1 M ammonium sulphate fraction collection procedure.

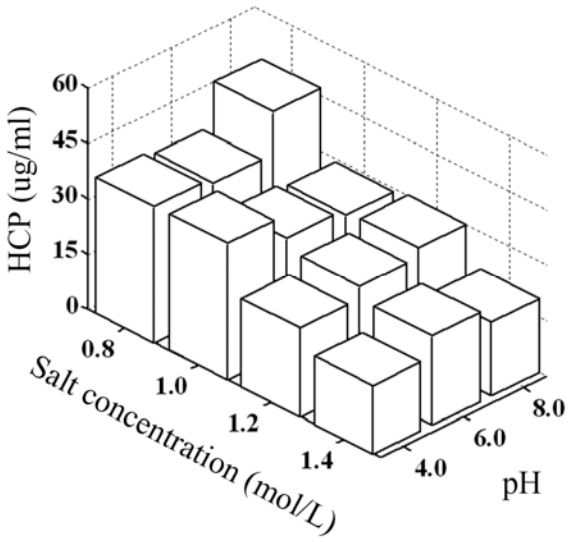


Figure 6.11 The HCP concentrations in mAb precipitation system under different salt concentration and pH conditions.

In order to achieve better purity and particularly focus on HCP, a set of preliminary studies were carried out to investigate the possibility of HCP reduction, shown in Figure 6.11. It showed that pH 4.0 and 1.2 mol/L ammonium sulphate precipitated nearly half of HCP and some other impurities, while kept mAb in the feedstock. Although 1.4 mol/L ammonium sulphate at any pH conditions would precipitate more HCP, the mAb would also be precipitated and thus significantly lower the yield. Therefore, a two cut precipitation process was preferred to eliminate some HCP first. The first precipitation conditions were pH 4.0 and ammonium sulphate 1.2 mol/L to reduce the HCP amount in the feedstock. After CUNO Zeta plus EXP 90 sp filter, the HCP precipitates and other impurity solids were retained in the filter. The positive charge of the membrane also helped to absorb soluble negative charged HCP molecules and further reduced the HCP level as a bonus (Prashad and Tarrach, 2006; Shukla et al., 2007; Yigzaw et al., 2008).

	HCP (ng/mg mAb, ppm)	DNA (ng/ml)	mAb yield (%)
Feedstock	242,465	6,592	100
After 1st depth filtration	105,891	4,354	98.1

Table 6.11 The HCP, DNA concentration and mAb yield after the first filtration in a two cut precipitation process.

Through this simple flow through procedure, the remaining HCP level in the solution decreased to 105,891 ng/mg mAb (ppm) with only less than 2% mAb loss, shown in Table 6.11. Then the same processes were carried out as those in the above studies (6.3.3 and 6.3.4). The final overall yield of mAb was 93.6% with purity dramatically increasing to 98.2%, very competitive to standard one step protein A chromatography, which achieved 92.4% yield and 95.7% purity using the same feedstock (Table 6.10). HCP level was 8, 737 ng/mg mAb (ppm), which was still higher than that in the protein A elution, but ten times less than previous result in the one cut precipitation procedure.

6.3.5.2 DNA removal

The DNA in the initial precipitation system was 6, 592 ng/ml (Table 6.10). After one cut precipitation, it was 256 ng/ml in the resuspension solution, nearly 25 times more than the 11 ng/ml DNA remaining in the protein A elution (Table 6.10). This was caused by the material of the depth filters, which was positive charged cellulose. DNA in the feedstock was normally negative charged and bound to the membrane. In resuspension step, the bound DNA was released into the final solution. While in the two cut precipitation, the first filter worked as the anion membrane chromatography, bound and reduced the DNA amount in the solution (Prashad and Tarrach, 2006; Yigzaw et al., 2008). Therefore, the DNA concentration reduced around 35% to 4,354 ng/ml after the first filtration (Table 6.11) and the final DNA concentration in the two steps processes decreased to 15 ng/ml (Table 6.10), which matched the quality of post protein A chromatography solution.

6.3.6 Protein property analysis and comparison

Although the precipitation process was successfully developed and scaled up with comparable yield and purity to standard protein A chromatography process, the protein property should be analysed and kept in nature active configuration. Figure 6.12 shows the CD results of three final mAb products. The salt precipitation is generally considered as a reversible process, which can be reversed by low ionic strength buffer without denaturing the protein (Manning et al., 1989). The protein secondary structure of mAb can be considered as in natural form by salt precipitation compared to protein A pH 3.5 elution (mAb standard) in Figure 6.12. The difference between precipitation resuspension and protein A pH 3.5 elution may come from the purity, as the resuspension used was from one cut precipitation, which had purity only around 89%. The impurities in the solution will interfere the CD results (Kelly et al., 2005). However, protein A chromatography has a potential risk to denature the mAb at low pH elution, e.g. pH 2.5, well demonstrated in Figure 6.12.

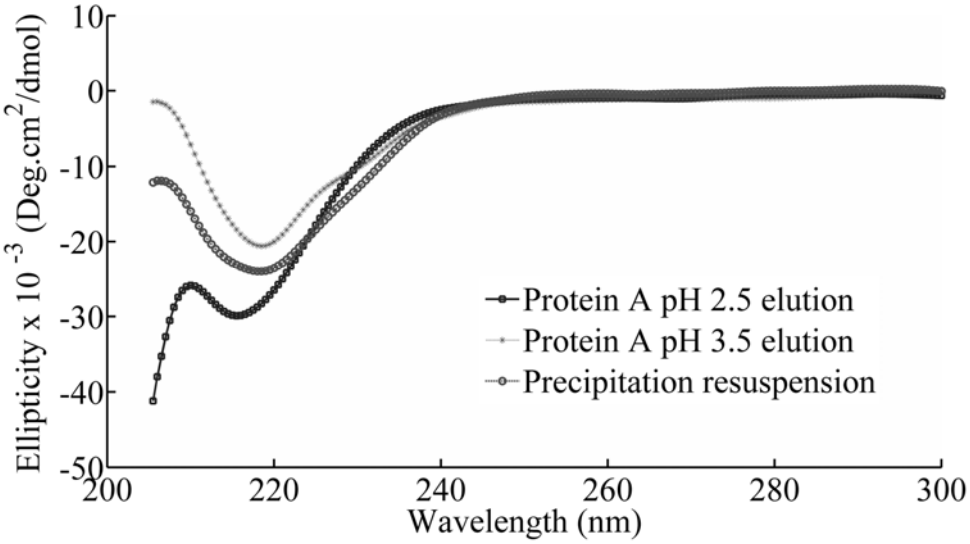


Figure 6.12 The CD curves of precipitation resuspension, protein A pH 2.5 elution and protein A pH 3.5 elution respectively.

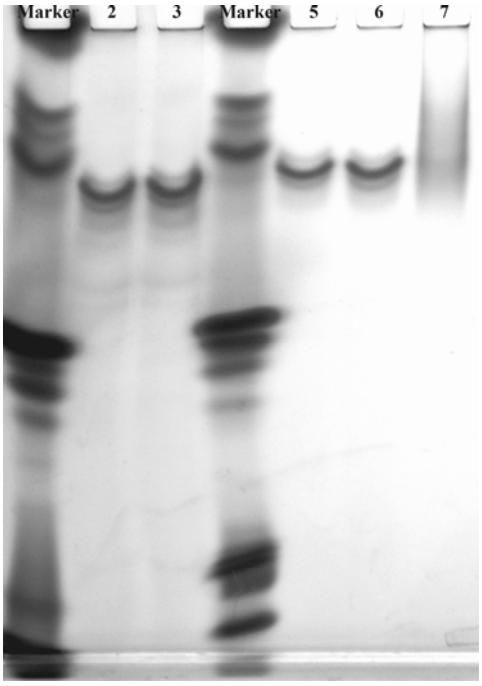


Figure 6.13 IEF gel image. Lane 1 and Lane 4: Marker; Lane 2 and Lane 3: feedstock; Lane 5: precipitation resuspension; Lane 6: protein A pH 3.5 elution; Lane 7: protein A pH 2.5 elution.

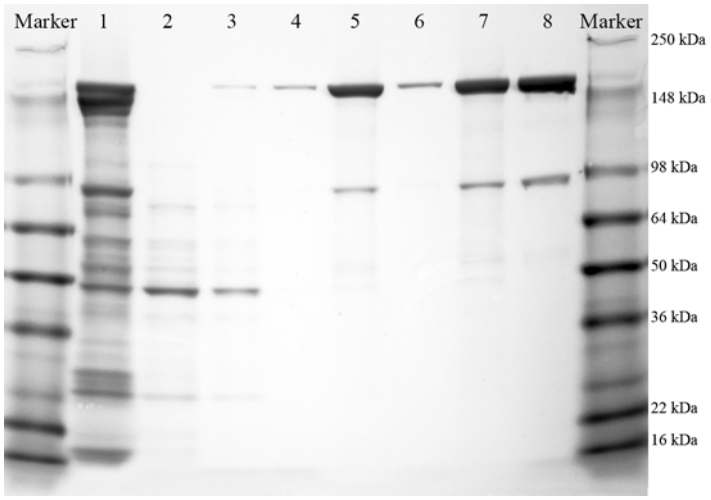


Figure 6.14 8-16% SDS-PAGE image. Lane 1: feedstock; Lane 2: filter through after desalting; Lane 3: 1st wash; Lane 4: 2nd wash; Lane 5: resuspension from one cut precipitation, mAb at 155 kDa band; Lane 6: strip; Lane 7: concentrated resuspension; Lane 8: protein A pH 3.5 elution.

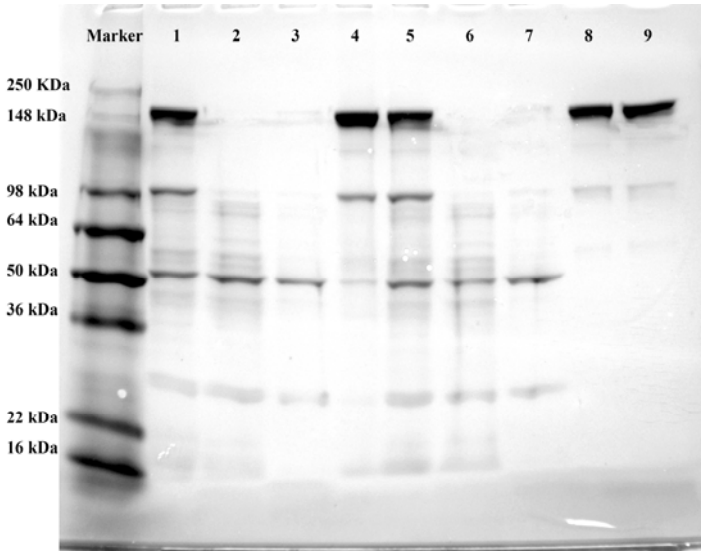


Figure 6.15 8-16% SDS-PAGE image. Lane 1: feedstock; Lane 2: one cut precipitation: filter through; Lane 3: one cut precipitation: wash; Lane 4: one cut precipitation: resuspension; Lane 5: two cut precipitation: solution after 1st filtration; Lane 6: two cut precipitation: filter through; Lane 7: two cut precipitation: wash; Lane 8: two cut precipitation: resuspension; Lane 9: protein A pH 3.5 elution.

The same phenomenon can also be observed in IEF gel, as shown in Figure 6.13. The precipitation resuspension and the protein A pH 3.5 elution shown the same pI mAb band as that in the feedstock, while the protein A pH 2.5 elution apparently showed that the mAb structure was denatured with different surface charges so that the molecules can not be focused to a fixed band (Prats, 2010).

Figure 6.14 shows the SDS-PAGE gel of one cut mAb precipitation compared with post protein A solution. The precipitation process really separated intact mAb efficiently from contaminating proteins compared with feedstock. Because precipitation was not a selective process, post precipitation solution still demonstrated some impurities background in the SDS gel, while post protein A solution lane had cleaner and distinct bands. The lower band at around 98 kDa, which existed in both processes, was considered as a mAb relative fragment or half mAb from SDS-PAGE sample preparation as the following size exclusion chromatography did not show the peak at this molecule weight.

Figure 6.15 shows the results of both one cut precipitation and two cut precipitation. The initial precipitation and the 1st filtration, shown in Figure 6.15 Lane 5, did not change the components very much compared to feedstock, shown in Lane 1. Although the HCP Elisa results showed that nearly half of HCP (Table 6.11) is removed from this step, it can not be observed from SDS-PAGE gel due to their relatively low concentration. The resuspension from the two cut precipitation (Lane 8) has cleaner background than that from one cut precipitation (Lane 4), proved by the purity in Table 6.10 with 98.2% to 88.2%. Lane 8 is quite similar to Lane 9, protein A pH 3.5 elution (mAb standard), which can also be expected from the purity data in Table 6.10.

6.3.7 Size exclusion chromatography and aggregation

The results from gel filtration column also proved the efficiency of protein

precipitation, which was shown in the Figure 6.16. The feedstock result, Figure 6.16.(a), shows several peaks besides the main mAb peak at around 30ml. The one cut precipitation was not capable to remove the first high molecule weight peak (Figure 6.16.(c)) and that was probably the reason why its purity was only around 88%. Two cut precipitations was able to eliminate that impurities peak, shown in Figure 6.16.(d), and matched the result of protein A chromatography (Figure 6.16.(b)).

	Aggregates (%)	Monomer (%)	Half antibody (%)
Protein A 3.5 elution	0.85	94.53	4.61
One cut precipitation resuspension	1.34	94.93	3.72
Two cut precipitation resuspension	0.32	96.31	3.37

Table 6.12 The aggregates, monomer and half antibody percentage in three processes.

Table 6.12 shows the results of aggregates, monomer and half antibody in three solutions analysed by Tosch TSKgel G3000 SW_{XL} column. Both precipitation processes had less half antibody percentage in the final solution than protein A chromatography. As the Figure 6.16.(c) shows that one cut precipitation resuspension had an aggregate peak at 20 ml, this solution really had a high aggregate percentage (1.34%) than other two solutions. In the two cut precipitation process, the initial precipitation and filtration also reduced aggregates to 0.32% in the final solution. Normally, the aggregates, which are high molecule weight components, are much easier to be precipitated out in relatively low salt concentration (Arakawa and Timasheff, 1985) and thus separated by the following 1^{st} depth filtration. That is probably the reason why two cut precipitation solution has low aggregates percentage.

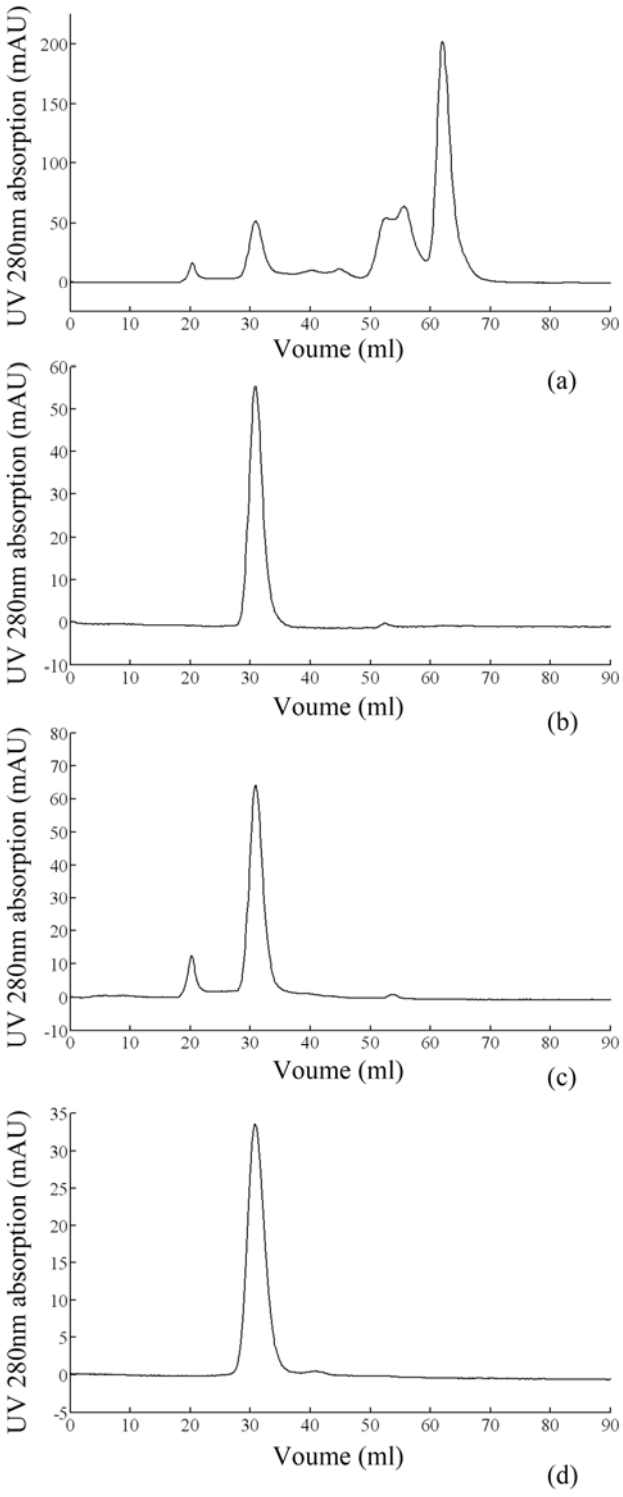


Figure 6.16 Size exclusion chromatography results: (a) feedstock; (b) protein A pH 3.5 elution; (c) one cut precipitation resuspension; (d) two cut precipitation resuspension. mAb (MW 155 KDa) peaks at 30 ml.

6.4 Conclusions

The model based process design method was successfully applied in this mAb precipitation case study. The Fab' precipitation model was first evaluated by preliminary mAb precipitation data, modified based on the first principles and then updated by microwell experimental data. Two sets of precipitations by ammonium sulphate and PEG respectively were optimised through the model based experimental design and optimisation method proposed in Chapter 3. The optimisation was completed after only a few iterations, requiring very few experiments. Strong statistic evaluations in the algorithms ensured the accuracy of the model and supplied more key conclusions on the bioprocess. The best precipitant was thus selected and the optimal conditions were found by comparing the objective values predicted by the updated models. The results provided the basis for following lab scale depth filtration studies. The study further proved the effectiveness of a model based procedure that integrated bioprocess modelling, model based experimental design and high throughput experimentation. This approach had the capacity to rapidly design and develop a precipitation based mAb purification process starting at microwell scale.

This study also provided more detailed insights into the mAb precipitation and demonstrated the possibility of precipitation to replace protein A chromatography in the initial capture step after cell culture harvest. With further improved procedures adopted in lab scale, such as two cut precipitation, high mAb yield with acceptable purity process can be achieved, which were comparable to protein A chromatography process. The normal platform for mAb production will carry out IEX after the initial protein A chromatography capture step (Shukla et al., 2007). Considering the robustness and separate capacity of IEX, the remaining impurities in resuspension such as HCP, probably will be further separated (GE Healthcare, 2006, 2008; Smith et al., 1998). The final product utilising new precipitation based platform has the potential to achieve the same quality of the product following traditional protein A

chromatography platform. Under the same quality requirement from FDA, the high process capacity of precipitation and the low cost (Honig and Kula, 1976; Knevelman et al., 2010; Niederauer and Glatz, 1992) will make it a very promising alternative monoclonal antibody purification approach to protein A chromatography in the future.

Chapter 7. Bioprocess scale up assessment based on high throughput chromatographic experiments and whole process analysis: a case study in precipitation based mAb purification process development

7.1 Introduction

The potentials of precipitation process as an alternative to protein A chromatography has been demonstrated in Chapter 6 in terms of yield and impurities removal. To truly understand whether precipitation can be used at manufacturing scale, the whole process evaluation for both processes including capacity and cost analysis should be carried out and compared. The conclusion will be drawn from the results of whole process assessment. Therefore, in this chapter, a rapid whole process evaluation approach combining high throughput chromatographic screening and cost of goods (COG) analysis will be proposed. The effectiveness of precipitation will be further assessed by this approach and compared with protein A chromatography from a whole process view.

7.2 Polishing process development and whole process analysis

Currently, chromatography processes are the main workhorses in the general purification platform for monoclonal antibody. It normally starts with affinity protein A chromatography, followed by two or more polishing chromatographic steps (Shukla et al., 2007). Ion-exchange chromatography (IEX), both anion and cation, hydrophobic-interaction chromatography (HIC) and other chromatographic processes, e.g. Ceramic Hydroxyapatite or anion membrane, are commonly utilised techniques in downstream polishing steps (Fahrner, 2001; Ghosh and Wang, 2006; Knudsen et al., 2001; Li et al., 2005). The aims of polishing steps are to reduce the remaining impurity molecules, such as host cell protein (HCP), aggregates, DNA, and leached protein A if using protein A chromatography in the primary capture step,

to a safe level, which can be accepted and approved by regulatory bodies (FDA, 2001; Gagnon, 2007; Gottschalk, 2008).

However, the development of a polishing chromatographic process is very challenging due to limited overall development time but very long experimental time for an individual test and need to explore great amount of variables and large parameter space (Bensch et al., 2005; Coffman et al., 2008). The conventional methods for a chromatography step development utilise scale down columns, pre-packed or self packed, to optimise process conditions by dozen runs and then keep the parameter values, e.g. column height, or ratio to scale up (Keener et al., 2008; Kelley et al., 2008). Therefore, even the development in a small to lab scale makes heavy demands for feed material and long timescale studies. This restriction inhibits the fully understanding of chromatography process with still relatively large space unexplored, and thus results in the final process is actually not fully optimised from technical or economic aspects (Chhatre and Titchener-Hooker, 2008).

In the current purification platform, the affinity protein A chromatography can easily achieve up to 95% purity in one step and also very high yield during the initial capture step (Roque et al., 2007; Shukla et al., 2007; Sommerfeld and Strube, 2005). The following polishing step development is relatively robust with very small stream variation from the initial capture step and thus can achieve process harmonisation in a board range of products under the same platform. Nevertheless, as mentioned in previous chapters, the low capacity and high cost limit its further application in large scale and high titre antibody purification (Gottschalk, 2008; Przybycien et al., 2004; Shukla and Thommes, 2010). Several alternative processes, e.g. aqueous two-phase extraction, precipitation and HPTFF, are recently explored by many researchers (Andrews et al., 1996; Etzel, 2008; Knevelman et al., 2010; Ma et al., 2010; McDonald et al., 2008; Rosa et al., 2007) to replace affinity chromatography and form a non protein A chromatography platform (Przybycien et al., 2004).

These new processes have certain capacity to separate antibody from other impurities, but the final materials are normally not as pure as post protein A chromatography solution due to less selectivity (Conley et al., 2011; Low et al., 2007). Moreover, the conditions of final materials after these innovative processes, such as pH, ionic strength, and the purity of the solution may be various for each product. It means that in the most cases, the following polishing step needs to be individually developed, not only the conditions of the bioprocess but also the proper operation selection and sequence (Follman and Fahrner, 2004; Ghose et al., 2008; Conley et al., 2011). The disadvantage of any alternative approach is that it hardly has harmonised techniques or knowledge sharing between different product lines unless a general rapid developing and evaluation solution is designed to tackle this complex task with the ability to accommodate various different starting materials.

Besides the technical feasibility analysis, the whole process economics should also be analysed based on the experimental results as early as possible, even before the development of full scale operation (Mustafa et al., 2004). The purification platform may not be changed, of course, if the whole process containing the alternative process needs more polishing steps or time than protein A based platform to achieve the same quality standard, even though the new unit operation is very successful and a great cost saver (Farid, 2008; Kelley, 2007). The new alternative platform will be adopted only if it is superior to the traditional protein A chromatography platform overall when facing the future upstream development and quality challenges, either in saving cost or reducing time or both (Farid et al., 2006; Farid, 2008).

Currently, high throughput screening is a reasonable and powerful technique to handle such a complicated task, considering the huge efforts needed for the selection of chromatography types, resins and conditions screening, which was well demonstrated in many researchers' work (Bergander et al., 2008; Coffman et al., 2008; Kelley et al., 2008; Rege et al., 2006).

In this chapter, the effectiveness of precipitation developed in Chapter 6 as an alternative process to protein A chromatography in current platform was assessed from a whole process view by the application of high throughput screening (HTS) in rapid chromatographic process development of polishing steps and a cost of good (COG) analysis by a commercial software.

7.3 Methodology

The aim of this study is to propose and assess a rapid technical and cost analysis method for evaluating an alternative platform design in the early stage. Under this circumstance, the needs to evaluate whether there are any polishing steps available to follow the new alternative process and whether they are able to form a reasonable and cost-friendly platform are urgent. Because the downstream purification process is not the results of only one unit operation but the combination of several bioprocesses, it is very worthwhile to evaluate the impact of new alternative process replacement from a whole process prospective before any further development, such as scale up to very large scale. It will reduce the risk of the process development and minimise the cost in case that the new developed bioprocess can not be well incorporated into whole purification processes or not economical at all for whole processes operation in the term of time, cost or labour.

However, this kind of risk management concept creates a paradox. The information needed to support the judgement normally requires a large number of experiments results from all processes in the new platform, while in the reality, it is almost impossible to provide enough detail information from the second or third process when the first one is still under developing. Therefore, a rapid high throughput screening method for early stage whole process analysis with very few material requirements was proposed as shown in Figure 7.1.

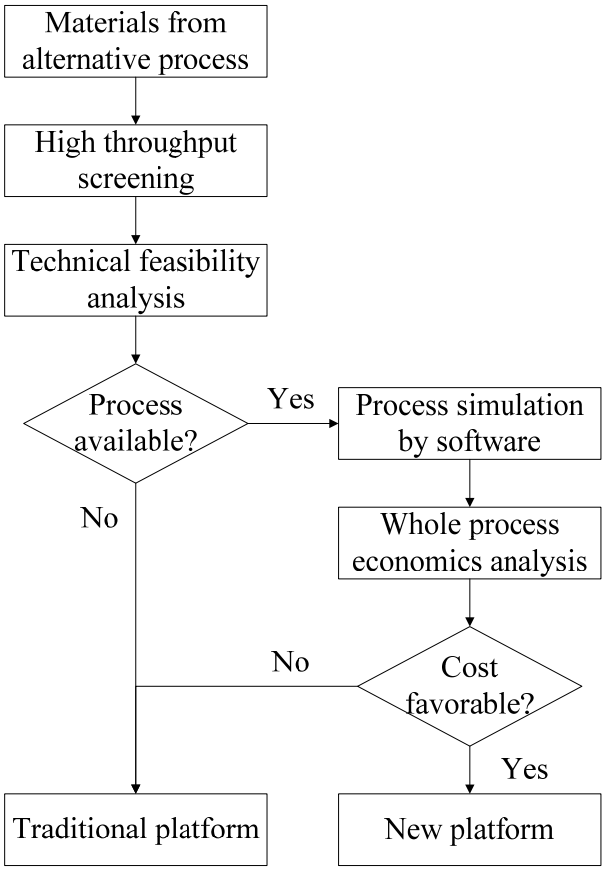


Figure 7.1 The flowchart for early stage rapid whole process analysis based on high throughput screening.

It starts with the product stream from alternative process, the resuspension from depth filters in this case. The high throughput screening experiments are used to collect the data for technical feasibility analysis. If there is no feasible process, the benefit of alternative process can not be realised. If the following up processes are feasible, the process simulation by using process software, such as BioSolve, is performed to carry out whole process economics analysis for both protein A based processes and the alternative processes. The cost analysis will provide quantitative data whether the alternative process is favourable or not. The HTS, technical feasibility analysis, process simulation and whole process economics analysis will be described in details and the results will provide the data to assess the benefits of alternative process in real term to give the insights.

7.3.1 HTS and technical feasibility analysis

With extremely limited material available in the early stage, high throughput screening is the best technology to provide evaluation data. HTS requires only several millilitres processing material to run the essential experiments for the technical feasibility analysis. It is not intended to conduct full process investigation. In this study, ion exchange chromatography and HIC were the candidate polishing processes for selection. In order to simplify the analysis of Host Cell Protein (HCP) and other impurities, the resuspension solution from one cut precipitation was used. This preparation had relatively spiked impurities compared to two cut resuspension in Chapter 6, which can be easily and rapidly analysed by HPLC instead using HCP Elisa. The resins types, binding conditions and capacity were rapidly evaluated on 96-microwell filter plates packed with resins by DoE designs. Binding-elution, flow through modes and gradient elution will also be tested in microwell filter plates with the technique developed by Coffman and his co-workers (Coffman et al., 2008; Kelley et al., 2008).

The technical analysis was based on the real experimental results from above studies

according to selectivity, capacity, yield and purity. Although the concentration of material was not optimised, it at least represented one typical stream, which was sufficient to carry out feasibility studies. The yield or purity in the microwell studies may be further optimised in the column based process (Kato et al., 2005), but the HTS data provided practical information on selectivity and capacity (Kelley et al., 2008) as well as the possible operation mode and the sequence of steps (GE Healthcare, 2006; Rege et al., 2006). The main target is to evaluate the separation of mAb and HCP, since the results from Chapter 6 showed the HCP in precipitation process was the main concerns compared to protein A chromatography. The potential processes will be selected as candidates of the next stage for further whole process economics analysis. If there is no process able to complete the polishing steps, the traditional protein A chromatography based platform will still be used, no matter how successful the precipitation purification process was.

7.3.2 Whole process simulation and economics analysis

Once there are several possible polishing processes, the combination of alternative process and the selected candidate polishing steps will be assessed in the commercial process simulation software BioSolve with integrated whole process models and a cost data base, developed by Biopharm Services UK. Because the information of large scale process was not available at this stage, most purification process parameters would be assumed based on the literatures and existing data base in the software, which was collected from real industrial processes and updated by Biopharm Services. Part of the assumptions were made based on above high throughput screening data, e.g. binding capacity, as the best estimator for large scale process (Bergander et al., 2008; GE Healthcare, 2006), if necessary, and used in simulation to give cost of goods analysis for this product.

The possible alternative platform would be compared with traditional protein A platform by cost and time evaluation. The software calculated not only the cost and

time for individual process but also for the whole process under cGMP conditions based on the input initial mAb concentration and scale. The proper size and number of vessels, bioreactors and columns were automatically decided by the software once the process flow sheet and process parameters were input. Different product titres and manufacturing scales were tested in the software to provide a full assessment according to various production capacity and future trends. If the platform based on the new alternative process did not have the advances in cost and time saving, the new bioprocess development would be terminated as the protein A platform was still superior. Otherwise, the new process was worth to be scaled up and optimised to replace the current platform.

7.4 Results and discussion

7.4.1 High throughput screening and technical feasibility studies

7.4.1.1 CEX resins and binding capacity screening

The first set of HTS experimentation evaluated a CEX chromatography process for the binding conditions of mAb in a post precipitation solution. As the pI for mAb is 7.4-7.8 (Figure 6.13), the protein was predicted to bind to the CEX resins at low pH and the operation model would be binding-elution (Ghose et al., 2008). The screening results for four CEX resins, which were designed to identify the optimal salt concentration and pH for binding, were shown in Figure 7.2.

Capto S resin, shown in Figure 7.2.(a), had the lowest binding capacity among all four resins. The maximum binding capacity at pH 4.0 was only 1 mg/ml, and nearly no binding in other conditions. Figure 7.2.(b), (c), (d) demonstrated the binding capacity of UNOsphere S, Fractogel EMD SO^{-3} (M) and SP Sepharose FF respectively. The trends of all three resins were similar, that the binding capacity decreased when pH and conductivity increased, which was the typical results of

CEX screening (Stein and Kiesewetter, 2007; Gagnon, 2007). However, the mAb binding capacity was very low in all resins. The maximum capacity achieved was only around 8 mg/ml at pH 4.0, 1 $mS \cdot cm^{-1}$ by UNOsphere S. As the mAb was loaded into each microwell for far more than 20 mg per ml resin challenge and the normal capacity for CEX resins will be around 100 mg/ml (Bio-Rad, 2000; GE Healthcare, 2006; Kelley et al., 2008), the above results demonstrated this mAb was not suitable to process by CEX binding-elution mode due to its low throughput.

It was considered as a quite special case result as the pure mAb solution, which was purified by standard protein A chromatography, was also tested with the nearly same poor results (data was not shown). This phenomenon was probably caused by the unusual properties of mAb, such as hydrophobicity or structures (Burgess, 1987; GE Healthcare, 2006). Since the binding mode failed due to low mAb binding in all resins, the flow through mode was considered.

The impurities binding results were shown in Figure 7.3 and these results limited the application of CEX flow through mode. The results for Capto S were not as expected, as shown in Figure 7.3.(a), that neither impurities nor mAb bound to resin very well. The results from other three resins were very close (Figure 7.3.(b) to Figure 7.3.(d)), that nearly half impurities were flowed away in loading phase while the remaining bound to the resin, around 40% to 50% in all conditions. These results predicted the flow through mode using the later three resins was able to lower the impurities by 2 fold, but it was not enough for a polishing step. Therefore, cation exchange chromatography failed the feasibility analysis.

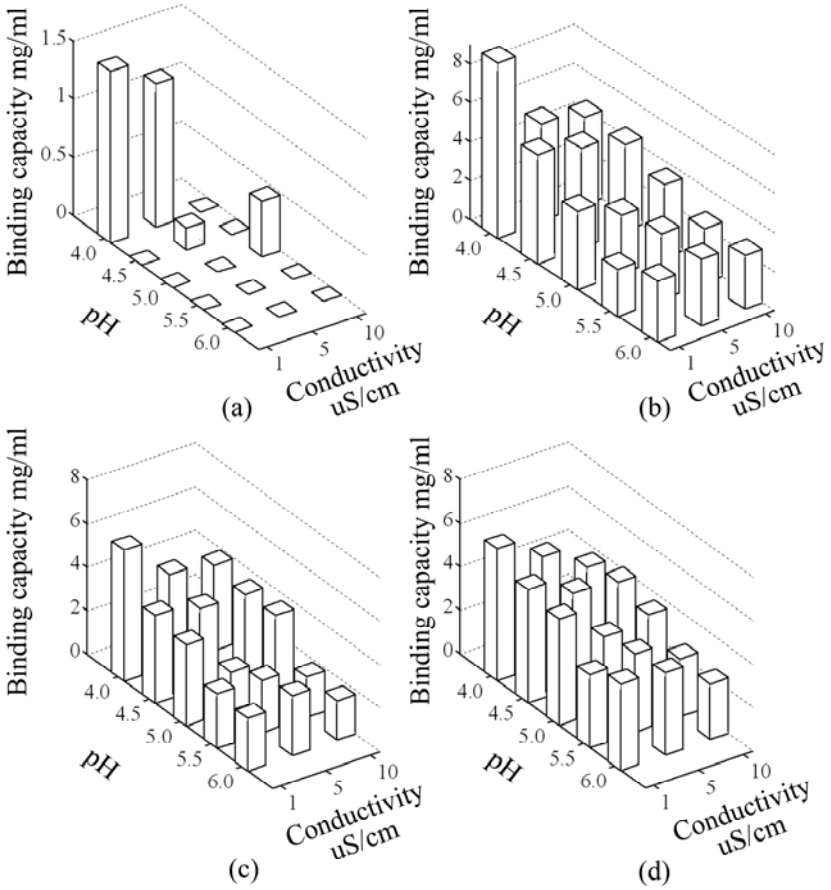


Figure 7.2 The screening studies of mAb binding capacity for four CEX resins: (a). Capto S; (b). UNOsphere S; (c). Fractogel EMD SO^{-3} (M); (d). SP Sepharose FF.

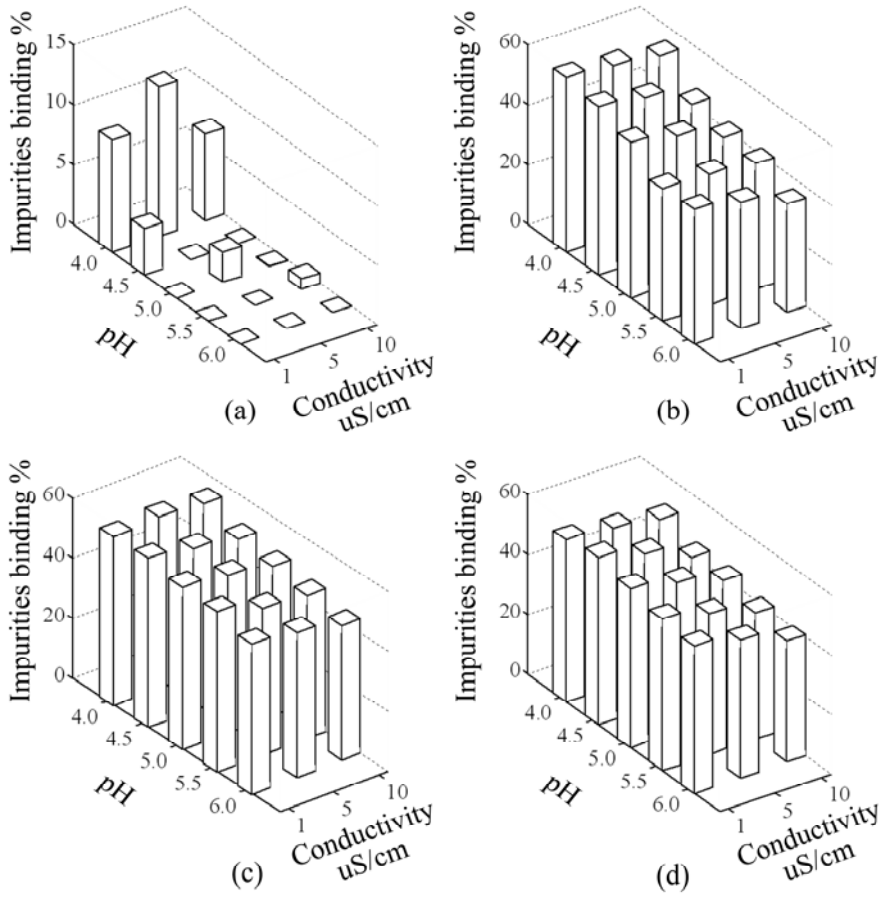


Figure 7.3 The impurity binding results for four CEX resins: (a). Capto S; (b). UNOsphere S; (c). Fractogel EMD SO^{-3} (M); (d). SP Sepharose FF.

7.4.1.2 HIC resins and binding capacity screening with gradient elution

In the HIC screening, it was designed to find the conditions for mAb binding and elution in four resins. Since HIC normally has low binding capacity, typically around 30 mg/ml for mAb, and the elution conditions are the main factors to separate components (GE Healthcare, 2009; Machold et al., 2002; Queiroz et al., 2001), the focus of this screening was to find the optimal elution conditions within four resins rather than the optimal binding conditions to achieve maximum binding capacity.

In the challenge of 29 mg mAb to 1 ml HIC resin tests, mAb was completely bound to the resin in all conditions for Phenyl 6, Butyl 4 and Butyl high performance. Only Butyl-S 6 had relatively low binding capacity, shown in Figure 7.4. Low pH at 6.0 and low salt concentration 1.0 M ammonium sulphate was not enough to allow strong hydrophobic interaction between mAb molecules and resin ligands. By increasing pH and salt concentration in the loading phase, the binding capacity for Butyl-S 6 also increased to the same levels as the capacity of other three resins, which was expected as Butyl-S 6 was the weakest hydrophobic interaction resin among these four resins (From GE Healthcare technical information).

The gradient elution was then carried out and all results from four resins in nine loading conditions each were demonstrated in Figure 7.5. In Figure 7.5.(a), the strong hydrophobic binding property of Phenyl 6 resin caused the mAb peaked at relatively low salt concentration, from 0.6 to 0 M. Even the 0 M strip can not completely elute all mAb from the strong bound resin. The yield of mAb through binding mode using Phenyl 6 was not satisfied due to large amount of mAb lost. The impurity normally had strong binding ability and eluted later than monomer mAb (Jiang et al., 2010). In Figure 7.5.(b), it was shown that impurity peak started at 0.2 M and the same as mAb, most of them still bound to the resin. Therefore, the resolution between mAb peak and impurity peak was not good enough. Considering the probably poor mAb yield, Phenyl 6 resin was not suitable for polishing step.

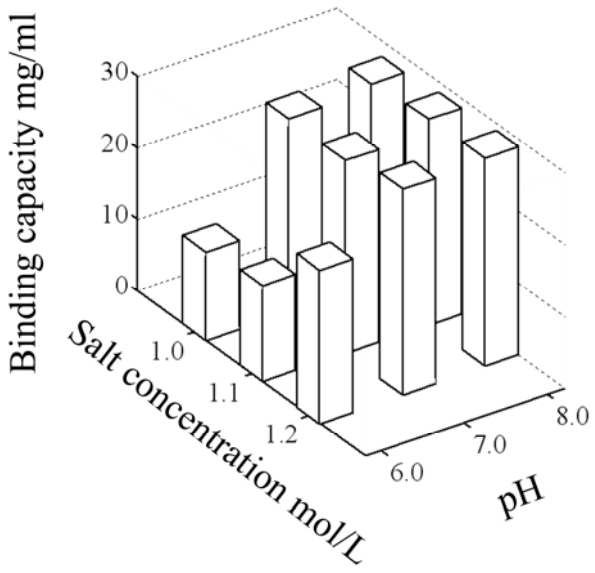


Figure 7.4 Binding capacity of Butyl-S 6.

Figure 7.5.(c) showed the mAb elution peak in Butyl 4 resin. As Butyl 4 was one of the weak HIC resins, the mAb elution peak came earlier than Phenyl 6 between salt concentrations 1.0 to 0.4 M. The impurity elution peak shown in Figure 7.5.(d) appeared in higher salt concentration than that in Phenyl 6. It started at 0.6 M and peaked at 0.2 M. Although there were still some impurities remained in the resin, if the elution condition was well chosen, e.g. salt concentration 0.4 M, pH 7.0, nearly 95% mAb yield and a relatively good separation can be achieved.

Figure 7.5.(e) and Figure 7.5.(f) showed the results from Butyl high performance, which gave high resolution and was recommended by GE healthcare as the ideal resin for intermediate and final purification steps. The mAb elution peak was in the middle range of salt gradient, with mAb started to elute at 0.8 M and completed at 0.2 M. The impurity peak was narrower and sharper than that in Butyl 4 elution and peaked at 0.2 M. Compare the results from Butyl 4 and Butyl high performance, latter one really had better resolution to separate mAb peak from impurity. If 0.3 M salt concentration was chosen as one step elution condition, Butyl high performance resin will have the most promising separation ability with higher mAb yield and better resolution than Butyl 4 resin, shown in Figure 7.6.

The elution results from Butyl-S 6 (Figure 7.5.(g) and (h)) seemed to prove again that this resin has the weakest hydrophobic interaction between mAb molecules and ligand. As shown in Figure 7.4, some conditions had relatively low binding capacity. Therefore, in Figure 7.5.(g), the mAb elution curves at these loading conditions did not show a complete peak even at high salt concentration elution. Other elution curves shown the mAb eluted at around 0.8 M with sharp peaks. However, the impurity elution curves demonstrated Butyl-S 6 was not able to separate impurity and mAb at all. The impurity also peaked at high salt concentration, around 0.8 M and had a quite wide salt concentration range.

In these screening studies, initial salt concentration in the loading conditions did not

show large difference in binding capacity and following elution peaks. Although pH normally was not considered as a big factor in HIC, the mAb elution curves in Figure 7.5 had an obvious common phenomenon for all four resins, that pH 8.0 binding conditions had strong hydrophobic effects than low pH as the elution peaks at these loading conditions came later than other conditions. This may be caused by the pI of mAb, which is around 7.4, since the protein will show more hydrophobic when approaching pI (Carta and Jungbauer, 2010; Queiroz et al., 2001). The impurity elution curves did not demonstrate any pH or salt concentration effects from loading conditions. Therefore, HIC separation in this case highly depended on how the interaction between mAb and resins changed with conditions and how much difference from impurity-resin interaction during elution phase. From the screening results, the Butyl high performance resins with proper elution conditions would be selected as the potential polishing process.

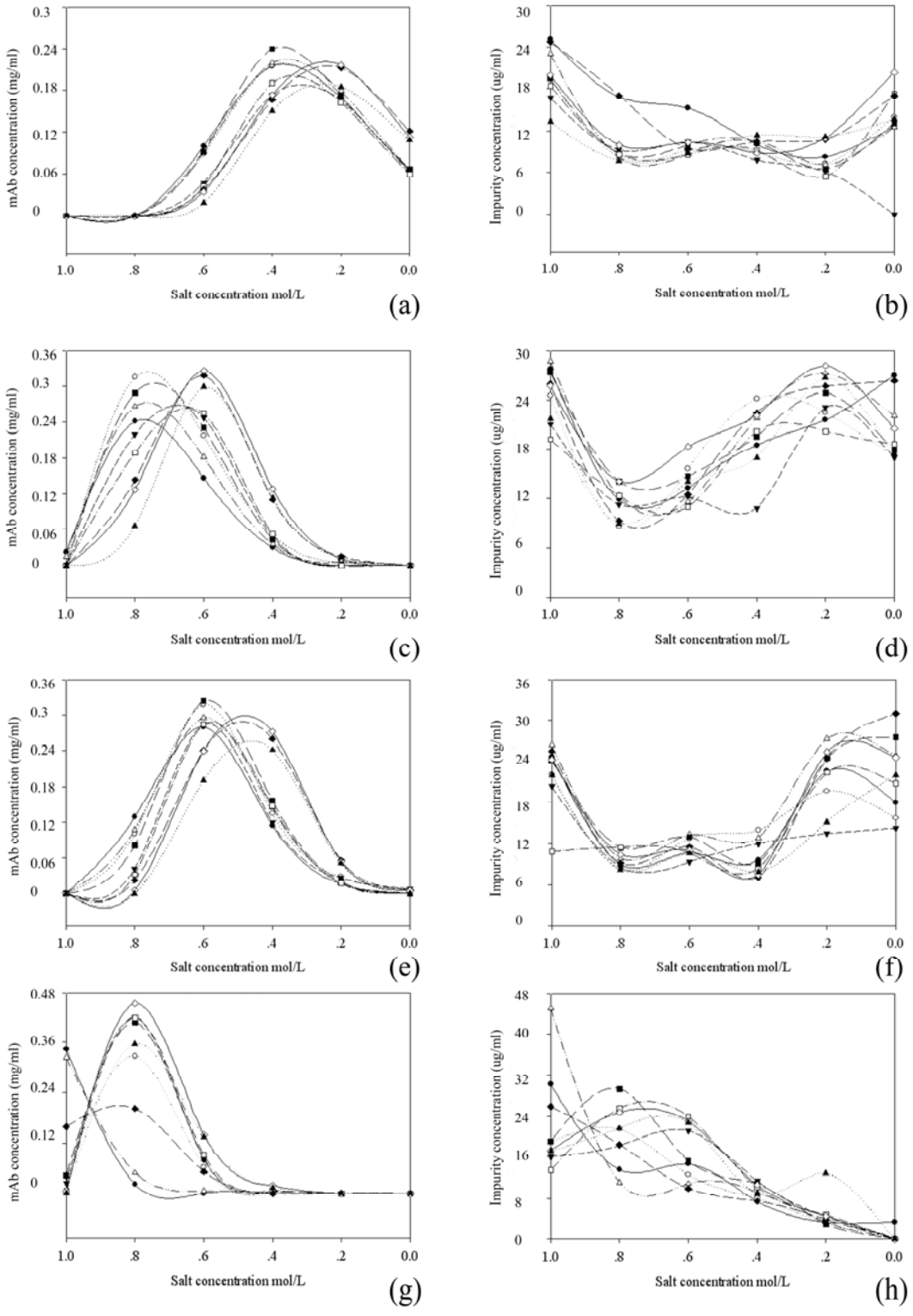


Figure 7.5 Gradient elution screening of Phenyl 6 (a, b), Butyl 4 (c, d), Butyl high performance (e, f) and Butyl-S 6 (g, h). Symbol represents loading conditions: solid triangle (up): pH 8.0, salt 1.2M; diamond: pH 8.0, salt 1.1M; solid diamond: pH 8.0,

salt 1.0M; square: pH 7.0, salt 1.2M; solid square: pH 7.0, salt 1.1M; triangle (up): pH 7.0, salt 1.0M; solid triangle (down): pH 6.0, salt 1.2M; star: pH 6.0, salt 1.1M; solid star: pH 6.0, salt 1.0M. Dots are data points and curves are extrapolated by MatLab.

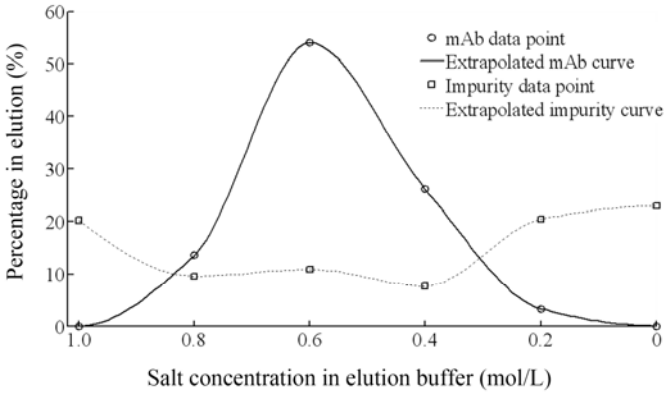


Figure 7.6 The gradient elution graph for Butyl high performance with loading conditions pH 7.0, salt 1.1 M.

7.4.1.3 AEX resin and flow through screening

Anion exchange chromatography in a flow through mode was frequently used to remove impurities in the downstream polishing steps (Shukla et al., 2007). The impurity proteins, such as CHO HCP, were generally more acidic than mAb (Jin et al., 2010; Stein and Kiesewetter, 2007). Therefore, these impurities may be removed by binding to AEX resin in a pH condition lower than mAb pI. AEX process was also utilised to remove residual DNA and leached protein A in affinity chromatography platform (Low et al., 2007; Shukla and Thommes, 2010).

Shown in the Figure 7.7.(a), as the pH in the loading conditions were all lower than mAb pI, the yield of mAb were all above 90%. Yield increased towards pH 7.0 because when it is close to mAb pI, fewer charges on the surface of mAb decreased the binding ability. Conductivity did not show very strong effects on the mAb binding to resins. The impurity remove surface was illustrated in Figure 7.7.(b). Contrary to mAb yield, conductivity had a significant effect on impurity remove. Low conductivity promoted the binding between Capto Q resins and impurities with 80% removal rate at high pH, while high conductivity lowered the rate nearly 25% in all conditions to around 50%. Higher pH had a slight positive influence on binding due to the impurities have more negative charges at higher pH range.

This initial screening of the AEX flow through conditions showed a Capto Q chromatography step after salt precipitation was also very promising to be a polishing process with high mAb recovery yield and impurities remove percentage. The possible operation conditions would be at pH 7.0 and low conductivity range.

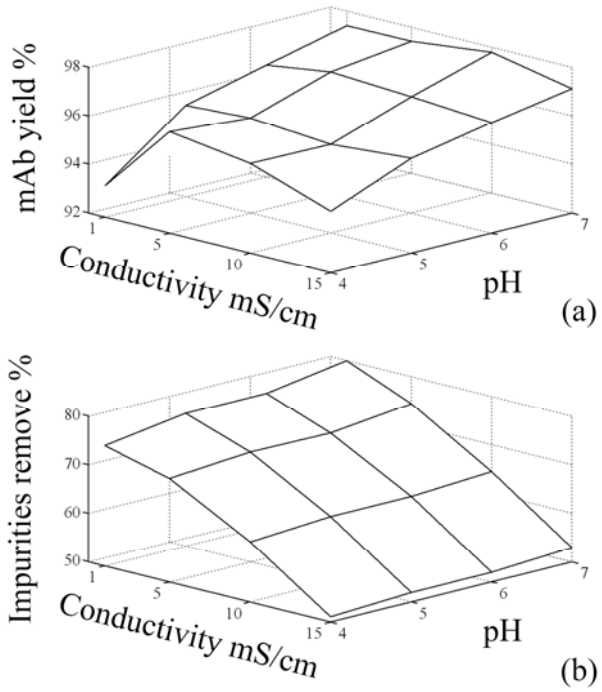


Figure 7.7 The condition screening studies of Capto Q resin in flow through mode: (a) mAb yield percentage; (b) impurities remove rate.

7.4.2 Whole process simulation and cost of goods analysis

7.4.2.1 Process flow sheet and operating conditions

According to the previous technical feasibility studies, two downstream purification processes were designed, shown in Figure 7.8. One was the protein A chromatography based processes and the other was the precipitation based processes. Two purification processes would be compared based on the time and cost.

In platform (a), it started with the conventional protein A chromatography as the first capture step. Since the cation exchange chromatography failed in the previous microwell studies, the second main purification step would be a binding-elution mode hydrophobic interaction chromatography (HIC). Between the first and the second chromatography steps, a buffer exchange process was needed to condition the feedstock to the required operating conditions (mainly salt concentration). The final step was an AEX flow through chromatography.

In platform (b), the previous developed two cut precipitation will replace the protein A chromatography. It started with the first solubilisation and the first depth filtration to remove the impurity protein and HCP, described in Chapter 6. Then the second precipitation and depth filtration would be carried out. The second depth filtration was different from the normal depth filtration process, since it had a mAb capacity limitation and needed washing, resuspension steps. Therefore, the depth filtration model in the software was modified to reflect these innovative changes. The following two steps were the same as those in platform (a).

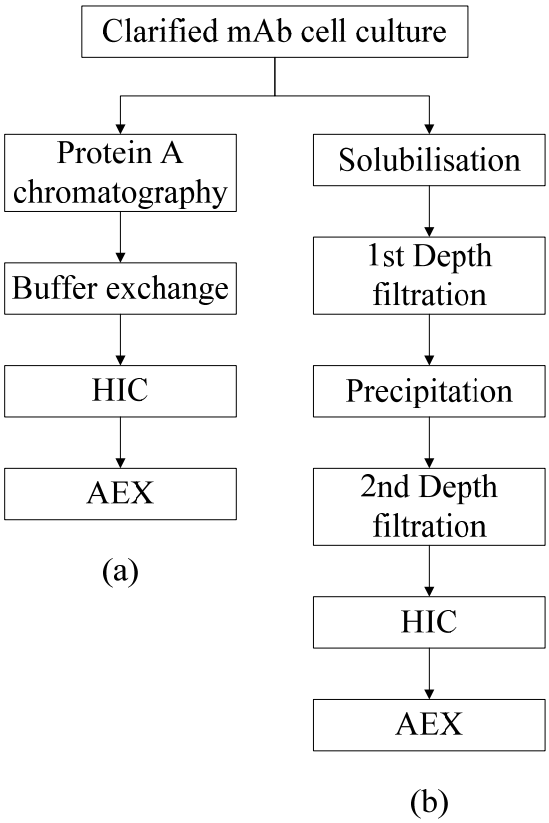


Figure 7.8 Two downstream purification processes: (a) protein A chromatography platform; (b) precipitation based platform.

The detailed information of operating conditions was listed in Table 7.1. Two levels of mAb concentration: 5 g/L and 10 g/L and three manufacturing scales: 1,000 L, 10,000 L and 100,000 L were assumed as the initial conditions to investigate the impact of concentration and scales. In the small scale, the small chromatography column was selected to minimise the capital cost with the target one chromatography cycle. For the large scale, the biggest column was chosen to maximise the capacity of each cycle in order to reduce operating time. The values for linear velocity, binding capacity and other parameters came from manufacture manuals or previous experiments. The salt stock solution added in solubilisation or buffer exchange steps and precipitation was 4.0 M ammonium sulphate. All materials were conditioned to required pH by buffer exchange before each process, which was not shown in both process flow sheets. It was assumed that these processes caused little process and cost difference in both platforms and thus can be ignored in comparisons.

In precipitation platform, depth filtration was the crucial process. In order to reduce process time, three filtration racks with a maximum 30 filters capacity each rack were used, according to the information of current available equipments and filters, provided by main suppliers (3M Cuno, Millipore and Pall). The largest filter with 2.8 m^2 is applied and the maximum reuse number for filter was assumed as 5. From the lab scale results in Chapter 6, the filter capacity in the second depth filtration was assumed as 85 g mAb/m^2, which meant each depth filtration rack has a maximum capacity. Once the maximum capacity was achieved, the filters should be washed and then resuspended to release the capacity. Therefore, the second depth filtration worked like a chromatography process, requiring several cycles to process one batch materials. From the flux results in Chapter 6, the maximum flux can be around 2,400 $L \cdot m^{-2} \cdot h^{-1}$, if the linear scale up was adopted. Low flux resulted in greater capacity and longer filter life span (3M Cuno, 2007), but longer process time. While in the software calculation, the flux used in both depth filtrations was set at 1600 $L \cdot m^{-2} \cdot h^{-1}$, which was the upper limit of optimal flux recommended by the depth filter supplier (3M Cuno, 2008).

Protein A chromatography		Solubilisation in precipitation platform	
column diameter	200 cm*	salt volume add (%)	30
bed height	20 cm		
bed volume	627 L*	1st depth filtration	
linear velocity	300 cm/hr	flux	1600 LMH
dynamic binding capacity	30 g mAb/L	filter area	2.8 m^2
elution volume	2.5 CV	number of filter racks	3
resin max reuses	200	number of filters	30
Buffer exchange in protein A chromatography platform		Precipitation	
salt volume add (%)	27.5	salt volume add (%)	16.7
Hydrophobic interaction chromatography		2nd depth filtration	
column diameter	240 cm*	flux	1600 LMH
bed height	20 cm	filter area	2.8 m^2
bed volume	905 L*	number of filter racks	3
linear velocity	200 cm/hr	number of filters	30
dynamic binding capacity	30 g mAb/L	filter capacity	85 g mAb/m^2
elution volume	2.5 CV		
resin max reuses	50		
Anion exchange chromatography		* The maximum size of one chromatography column to handle large scale materials. For the small scale, the size will be adjusted by the software to the most economical size.	
column diameter	240 cm*		
bed height	20 cm		
bed volume	905 L*		
linear velocity	150 cm/hr	** In the flow through mode, very few mAb will bind to the resin and the value here indicates the maximum mAb loaded to column in each cycle.	
dynamic binding capacity	75 g mAb/L**		
resin max reuses	50		

Table 7.1 Details of two mAb purification processes operation conditions.

7.4.2.2 Cost of goods and process time comparisons

Table 7.2 shows the cost and time analysis of protein A chromatography purification platform based on two mAb titres and three scales. It has been already pointed out by several researchers (Farid, 2007; Werner, 2004) that low titre and small production scale caused high cost per gram product. It can be noticed that the cost decreased from 23.1 US $/g to 11.3 US $/g when mAb mass increased 200 times from 5,000 g

to 1,000,000 g. However, processing time increased from 20 hours to 270 hours, which was nearly 11 days. It was impossible to process a batch for 11 days in a real production, which meant the manufacturer needed invest probably three more sets of purification facility to reduce one purification shift to less than 3 days at very large scale. Calculated by software, the cost and throughput did not increase when the processes were operated at very large scale, e.g. 100,000 L, while the process time and capital still increased with mAb titre. It was caused by the physical volume limitation of a protein A chromatography column, which can not be scaled up any more once its maximum capacity was exceeded by the process scale. The only solution was to build more chromatographic facilities, which inevitably increased the cost and capital.

MAb concentration (g/L)	Volume (L)	Protein A chromatography platform				
		Cost (US $/g)	Batch time (hr)	Batch per year	Throughput (Kg/Year)	Capital (US $M)
5	1,000	23.1	20	376	1,527	37.1
10	1,000	19.3	20	376	3,054	52.1
5	10,000	14.3	26	376	15,270	78.1
10	10,000	13.5	41	376	30,540	81.2
5	100,000	11.3	147	94	38,175	103.9
10	100,000	11.3	270	47	38,175	118.6

Table 7.2 Protein A chromatography purification processes costs and time analysis.

Table 7.3 shows the cost and time analysis of precipitation based purification platform, in order to compare the results to protein A chromatography platform in Table 7.2. The trends of cost and process time against the titre and process scale were observed as same as those in Table 7.2. The unit operation cost of a precipitation process normally depended on the scale, which was more volume, more cost as it required more salt to precipitate mAb. That was why the cost of 10 g/L, 10,000 L process was slightly lower than that of 5 g/L, 100,000 L process in the overall cost calculation. At low titre and small scale, the precipitation based platform did not have any advantages in cost and time, if compared with protein A

chromatography based platform. While the titre and scale increased, precipitation showed its cost and time saving abilities. At very large scale, the cost and time were reduced by 22% (11.3 US $/g compare with 8.8 US $/g) and 26% (270 hours compare with 199 hours) respectively. Since precipitation process had no physical limitation, the throughput would increase when more capital was invested, demonstrated in Table 7.3.

MAb concentration (g/L)	Volume (L)	Precipitation based platform				
		Cost (US $/g)	Batch time (hr)	Batch per year	Throughput (Kg/Year)	Capital (US $M)
5	1,000	23.2	23	376	1,579	31
10	1,000	17.5	23	376	3,159	45.1
5	10,000	12.8	26	376	15,796	67.1
10	10,000	10.3	34	376	31,592	81.5
5	100,000	11.1	113	94	39,490	106.2
10	100,000	8.8	199	62	52,094	114.4

Table 7.3 Precipitation based purification processes costs and time analysis.

The results sufficiently showed that precipitation based purification platform was quite comparable to protein A chromatography based platform. When the upstream produces high titre at large scale cell culture, precipitation based platform is more preferable than the traditional protein A chromatography platform, due to its low cost and large throughput. However, in this designed processes, the second depth filtration had a batch mode operation, which was the bottleneck in whole processes. The nearly 200 hours for processing 10 g/L, 100,000 L cell culture was largely contributed by the second depth filtration process. It highly depended on the filter capacity and flux. If higher filter capacity and flux can be achieved, the operating time will be much shorter.

7.5 Conclusions

In this chapter, a HTS development and analysis technique was demonstrated in

order to select and design polishing chromatographic steps quickly based on a post precipitation mAb solution. Four CEX resins with both binding and flow through mode were evaluated. It showed the CEX was not able to purify the solution further and thus saved the cost and time for large scale CEX studies. Four HIC resins were also investigated by binding and gradient elution mode. The separation effects can be clearly demonstrated in elution profiles. The best resin and conditions would be selected to be further investigated in lab scale studies. Capto Q AEX was also tested in a flow through mode, which showed good polishing results.

The automatically HTS can rapidly assess the suitability of a downstream purification process with regards to binding capacity and impurity removal. The experiments carried out in microwell plates required small amount of feedstock and resins with huge data produced in one set, far more efficiently than typical chromatographic development. However, they are not the complete replacement of a laboratory scale column study. The lab scale tests with the resins and conditions chosen from HTS experimentations still need quite a lot conditions to be investigated in column study because HTS can not predict the optimal pressure, flowrate and DBC etc. The HTS studies can provide useful information on processes selection and conditions screening, which can be used as the starting point for later scale up and column studies.

Combining the HTS technical feasibility study results with the process simulation software based on whole process models will give a quick judgment and analysis on which bioprocess is needed and how to form a purification platform to purify current feedstock by little cost and research work. In this study, whole process analysis showed that at relatively small scale (1,000 L), both processes had similar unit production cost. However, with the increase of the scale, the precipitation based process was cheaper and the unit production cost reduced 20-30 %. It is also noticed that the operation time at large scale for protein A chromatography based process was 35 % longer than precipitation based process, which was a major concern in

future large scale production. The whole process analysis showed the new precipitation process developed in Chapter 6 was able to be incorporated into a feasible platform providing a low cost and faster approach for future large scale mAb production than current protein A chromatography platform.

Chapter 8. Conclusions and future work

8.1 Conclusions

The overall objective of this research was achieved by the successful development of a model based process design and optimisation methodology for downstream bioprocess. This can significantly reduce the development time and cost in the early stage. The new model based process design method has been proposed with three main components in an integrated optimisation framework: bioprocess modelling, model based experimental design and high throughput experimentation. 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 in a nonlinear bioprocess system. The integrated design methodology has been applied to two protein precipitation case studies and proved its effectiveness for process design and optimisation.

The model based process design started with the bioprocess modelling steps. The modelling step not only helped to understand the bioprocess through a quantitative equation but also provided the basis for the following model based experimental design and optimisation. The general approach for bioprocess modelling was designed to develop a reliable model from the process mechanism or empirical data, regressed by high throughput microwell experimental results, which was then to be validated by statistical tests. Compared to the conventional modelling procedure, this generic approach was faster and produced a more accurate process model due to the massive number of data points provided from the microwell studies. This also accelerated the loop of model modification.

A Fab' precipitation system with ammonium sulphate concentration and pH as

operation conditions was then used as the first case study to test this modelling procedure in microwell plates. A precipitation model based on phase equilibrium was proposed to describe this Fab' precipitation. Microwell precipitation provided a full set of data for modelling development and enabled statistical validation after modification. The newly developed model can accurately describe not only the solubility of pure Fab' and Fab' in the clarified homogenate, but also can be applied to impurities, considered as a pseudo component. The model showed superiority over other existing models, including the widely used polynomial equation from DoE method, judged by statistical tests values. Since the model was based on the theory of precipitation, the model can be transferred to other precipitation systems after the training and parameterisaton by real data. However, the significances of the parameters in the model should be fully evaluated by statistical tests based on the new data and modified if necessary in any new precipitation environment or system. The validated precipitation model can also be used to test the following model based experimental design algorithm.

In the core step of the whole process design methodology, model based experimental design adopted three different algorithms to achieve mutual benefits and accelerate model accuracy and process optimisation. D-optimal design was the driving force of model accuracy due to its power to choose the information rich points to parameterise model and minimise regression errors. A key contribution of this D-optimal design work was that the new sequential design reduced the computation efforts. Optimisation was carried out by predicting the maximum or optimal points in each design loop for the next run of experiments. This optimisation was achieved by simulation method based on experimental conditions. Random and a modified Simplex design provided extra points to work as an error preventing function, in case the algorithm was trapped in a local optimum or became deadlocked. This innovative experimental design algorithm avoided the disadvantages of each design method when used alone in the traditional experimental design. Combining them together and working in a parallel mode in one design loop can facilitate optimisation without

increasing time and cost. The overall sequential design structure and high throughput experimentation further reduced the development time significantly and promoted the interplay between algorithms and experiments.

The same Fab' precipitation system was used as the test base for the model based experimental design algorithm evaluation. It was initialised by a DoE design in the clarified homogenate feedstock precipitation. The real results were continuously feedback to the algorithm to make the model more accurate and design the next experimental runs. Accurate models and the optimal point were found efficiently after only 5 iterations. The error between predicted value and real experimental data was in the acceptable range. This approach was compared with current widely used DoE with a second order polynomial regression method. The results showed that the new model based experimental design and optimisation method required fewer experiments but gave better predictions and without missing the real optimal area.

The whole methodology was then applied and tested in a second case study: a mAb precipitation process. The pure mAb precipitation preliminary studies provided sufficient data for precipitation model modification and the initial parameterisation, which is the necessary procedure for any existing model transfer from one reaction to another. After the proper model has been selected, the model based experimental design utilised the modified mAb precipitation model with the best initial guess from a pure mAb system to design and search the optimal point for the mAb precipitation process. Two precipitation reagents: ammonium sulphate and PEG, were used and both systems were simultaneously optimised by the algorithm in microwell plates. The optimal points in the two reaction systems were rapidly located with only a few iterations. The ammonium sulphate precipitation system with the optimal operational conditions was selected by comparing two precipitation systems based on the maximum objective value at the optimal conditions. The mAb precipitation models were also evolved to accurate models during this design procedure with very small SD^2 values when the iteration were completed, and providing sufficient information

for the process understanding.

The next step was to test the process at larger scale based on the results of above two steps in the design methodology. The optimised microwell results were then investigated at lab scale in order to design a bioprocess based on precipitation for an alternative mAb purification platform to replace the conventional protein A chromatographic route. Depth filtration with innovative separation procedures was adopted following precipitation to efficiently separate solids containing mAb from supernatant with most impurities remaining in solution. Filter capacity and flux proved the engineering feasibility for precipitation at large scale production and the results were comparable to those at microwell scale. The further study for some specific impurities, HCP and DNA, required process modification. Therefore, a two cut precipitation was introduced to eliminate more HCP and DNA to match the protein A chromatographic process. The first precipitation and depth filtration reduced the concentration of HCP, aggregates and DNA in the mAb feedstock. mAb was precipitated out in the second step precipitation and then retained, washed and resuspended in the second depth filtration process. The yield and purity were very comparable to protein A chromatographic process without denaturing mAb.

A rapid evaluation approach for whole purification processes was further proposed with the help of results from microwell and lab scale processes to carry out the whole process technical feasibility tests for precipitation based mAb purification processes and cost of goods analysis as early as possible. Microwell chromatographic experimentation provided the platform for a quick technical feasibility test. The commercial software BioSolve gave a whole process cost analysis based on the results from microwell screening and the process model in the software. Hydrophobic interaction chromatography (HIC) in binding-elution mode and anion exchange chromatography (AEX) in flow-through mode were selected as two potential candidate processes for polishing steps according to the high throughput screening results. Then a precipitation based purification platform was

compared with the traditional protein A chromatography based platform using the same mAb feedstock, assuming that the same properties were as those tested in the microwell studies. Various titre and scales of feedstock were set in the software to assess the cost and time of the two different purification platforms. The results showed that the alternative purification platform based on the precipitation had the ability to lower cost and reduce process time in high titre and large scale production, which was also the trend of the upstream cell culture. Therefore, in the near future, the precipitation based purification platform has the potential and advantages to replace the conventional protein A chromatography platform with large manufacturing capacity.

8.2 Future work

8.2.1 Applications of model based process design in dynamic systems

The model based process design methodology has the potential to be used in any bioprocess when a model exists or modelling can be carried out. In this thesis, two precipitations have a static nonlinear model due to the relatively simple mechanism behind these precipitations. The method has proved its power and capability to design and optimise this type of process. However, some bioprocesses, such as chromatography, have a more complicated dynamic model, e.g. general rate model, to predict the behaviour and results of the process with certain feedstock. An extension of the model based process design method in this thesis to dynamic systems is needed to find the solution from dynamic equations. As currently there is no well established method to design and optimise a dynamic process, it is very worthwhile to apply this integrated methodology to investigate its performance in the dynamic bioprocess.

8.2.2 Whole process design and optimisation

The model based process design method developed in this thesis has demonstrated its ability in a single unit operation design and optimisation. However, the downstream purification platform is normally composed of several bioprocesses. In most cases, tradeoffs between different processes and objectives are required when expanding to whole process optimisation. The ultimate aim is to achieve whole process design and optimisation, which cannot be simply achieved by adding together all designs based on the single unit operation.

However, if each unit operation in the whole purification process is able to be described by one or several mathematical models, the whole process can be described by a group of process models with multiple design space and constraints. The ability of the methodology in whole process design and optimisation based on a group of process models should be assessed and evaluated in future work.

8.2.3 Scale up modelling and prediction

In this thesis, all model based design and optimisation was based on the high throughput experimentation. However, it is quite difficult to predict the process performance at large scale due to lack of scale up model to link the high throughput experimentation results and large scale results. Knowledge including fluid dynamics, is heavily involved in this type of modelling activity.

Some bioprocesses, such as chromatography, works significantly differently at small scale and large scale, which is mainly limited by the instrument, e.g. no flowrate conditions to be analysed in mircrowell plates. Therefore, carrying out scaling up modelling for those bioprocesses requires new experimental designs or instruments to test these conditions throughout. The benefits of scale-up modelling is that once the proper models for process operation and scale up are available, the model based process design methodology can directly carry out process design and optimisation for large scale process based on these models with the small scale experimentation.

It will further accelerate the development of downstream purification processes.

8.2.4 Improve precipitation and depth filtration

Some other precipitants, such as PEI, or the combination of different precipitants, e.g. PEG with NaCl, should be investigated to understand more about the relevant process of precipitation and the possibility of better selective precipitation. The conditions, e.g. pH, salt types, of the washing buffer in the second depth filtration should also be screened to further reduce HCP in the solution. From the engineering aspects, various depth filters should be screened to study whether there are any improvements in filter capacity and flux. Increasing capacity and flux is the crucial step in the second depth filtration steps. It is anticipated that this will reduce the operation times. Moreover, other separation processes, especial continuous operations, should be evaluated to find whether they are able to improve the current batch mode mAb recovering depth filtration to a continuous mode separation process, which is more suitable in large scale processing.

References

1. Abramowitz M, Stegun IA. 1970. Handbook of mathematical functions with formulas, graphs and mathematical tables. New York: Dover Publications. 1046 p.
2. Afima RRC. 1982. The development of methodologies of mathematical modelling. Teach Math Appl. 1:125-131.
3. Agena SM, Bogle D, Pessoa F. 2000. An activity coefficient model for proteins. Biotechnol Bioeng. 55:65-71.
4. Akita EM, Nakai S. 1993. Comparison of four purification methods for the production of immunoglobulins from eggs laid by hens immunized with an enterotoxigenic *E. coli* strain. J Immunol Methods. 160:207-214.
5. Aldington S, Bonnerjea J. 2007. Scale-up of monoclonal antibody purification processes. J Chromatogra B. 848:64-78.
6. Al-mhanna NMM, Huebner H, Buchholz R. 2010. Optimization of parameters growth conditions of yeast biomass during single cell protein production by using simplex method. Chem Eng Trans. 21:475-480.
7. Andrews BA, Nielsen S, Asenjo JA. 1996. Partitioning and purification of monoclonal antibodies in aqueous two-phase systems. Bioseparation. 6:303-313.
8. Anderson M, Whitcomb P. 2005. RSM simplified-optimizing processes using response surface method for design of experiments. New York: Productivity Press. 293 p.
9. 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.
10. Arakawa T, Timasheff SN. 1985. Theory of protein solubility. Meth Enzymol. 114:49-77.

11. 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. p 103-124.
12. Atha DH, Ingham KC. 1981. Mechanism of precipitation of proteins by polyethylene glycols. J Biol Chem. 256:12108-12117.
13. Atkinson AC, Hunter WG. 1968. The design of experiments for parameter estimation. Technometrics. 10:271-289.
14. Atkinson AC. 1996. The usefulness of optimum experimental designs. J Roy Stat Soc B. 58:59-76.
15. Atkinson AC, Tobias RD. 2008. Optimal experimental design in chromatography. J Chromatogra A. 1177:1-11.
16. Aucamp JP, Cosme AM, Lye GJ, Dalby PA. 2005. High-throughput measurement of protein stability in microtiter plates. Biotechnol Bioeng. 89:599-607.
17. Azevedo AM, Rosa PA, Ferreira I, Aires-Barros MR. 2009. Chromatography-free recovery of biopharmaceuticals through aqueous two-phase processing. Trends Biotechnol. 27:240-247.
18. Bailey JE. 1998. Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. Biotechnol Progr. 14:8-10.
19. Banga JR. 2004. Optimization in computational systems biology. BMC Syst Biol. 2:47-52.
20. 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.
21. 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.

22. Bio-Rad Laboratories. 2000. UNOsphere™ Q & S ion exchange media instruction manual. Document No.: 4110109 Rev B.
23. Birch JR, Racher AJ. 2006. Antibody production. Adv Drug Deliver Rev. 58:671-685.
24. Bowering L, Bracewell D, Keshavarz-Moore E, Hoare M, Weir A. 2002. Comparison of techniques for monitoring antibody fragment production in E. coli fermentation cultures. Biotechnol Progr. 18: 1431-1438.
25. Box GEP, Hunter JS. 1961. The 2^{k-p} fractional factorial designs part I. Technometrics. 3:311-351.
26. Box GEP, Lucas HL. 1959. Design of experiments in non-linear situations. Biometrika. 46:77-90.
27. Box GEP, Wilson KB. 1951. On the experimental attainment of optimum conditions. J Roy Stat Soc. 13:1-45.
28. Box MJ. 1966. A comparison of several current optimization methods and the use of transformations in constrained problems. Comp J. 9:67-77.
29. Box MJ. 1970. Some Experiences with a Nonlinear Experimental Design Criterion. Technometrics. 12:569-589.
30. Box MJ, Draper NR. 1971. Factorial designs, the |XX'| criterion and some related matters. Technometrics. 13:731-742.
31. 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.
32. Bramaud C, Aimar P, Daufin G. 1997. Whey protein fractionation: Isoelectric precipitation of α-lactalbumin under gentle heat treatment. Biotechnol Bioeng. 56:391-397.
33. Burgess RR. 1987. Protein purification. In: Oxender D, Fox CF, editor. Protein engineering. New York: Alan R. Liss. p 71-82.
34. 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. p 1-75.

35. Carta G, Jungbauer A. 2010. Protein chromatography: process development and scale-up. Weinheim: Wiley-VCH. 364 p.
36. Cawse JN. 2001. Experimental strategies for combinatorial and high-throughput materials development. Acc Chem Res. 34:213-221.
37. Charaniya S, Hu WS, Karypis G. 2008. Mining bioprocess data: opportunities and challenges. Trends Biotechnol. 26:690-699.
38. Cheng YC, Lobo RF, Sandler SI, Lenhoff AM. 2006. Kinetics and equilibria of lysozyme precipitation and crystallization in concentrated ammonium sulphate solutions. Biotechnol Bioeng. 94:177-188.
39. Chernoff H. 1959. Sequential design of experiments. Ann Math Stat. 30:755-770.
40. 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.
41. 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.
42. 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.
43. 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.
44. Chiew YC, Kuehner DE, Blanch HW, Prausnitz JM. 1995. Molecular thermodynamics for salt-induce protein precipitation. AIChE J. 41:2150-2159.

45. Chung SH, Ma DL, Braatz RD. 2000. Optimal model-based experimental design in batch crystallization. Chemometr Intell Lab. 50:83-90.
46. Coffman JL, Kramarczyk JF, Kelley B. 2008. High-throughput screening of chromatographic separations: I. Method development and column modelling. Biotechnol Bioeng. 100:605-618.
47. Cohn EJ. 1925. The physical chemistry of the proteins. Physiol Rev. 5:349-437.
48. 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.
49. Cohn EJ. 1941. The properties and functions of the plasma proteins, with a consideration of the methods for their separation and purification. Chem Rev. 28:395-417.
50. Cohn EJ, Strong LE, Hughes WL Jr, Mulford DJ, Ashworth JN, Melin M, Taylor HL. 1946. Preparation and properties of serum and plasma proteins. IV. A system for the separation into fractions of the protein and lipoprotein components of biological tissues and fluids. J Am Chem Soc. 68:459-475.
51. Conley GP, Viswanathan M, Hou Y, Rank DL, Lindberg AP, Cramer SM, Ladner R C, Nixon AE, Chen J. 2011. Evaluation of protein engineering and process optimization approaches to enhance antibody drug manufacturability. Biotechnol Bioeng. 108:2634-2644.
52. Cook RD, Bachtsheim CJ. 1980. A comparison of algorithms for construction exact D-optimal designs. Technometrics. 22:315-324.
53. Curtis RA, Ulrich J, Montaser A, Prausnitz JM, Blanch HW. 2002. Protein-protein interactions in concentrated electrolyte solutions. Biotechnol Bioeng. 79:367-380.
54. Datta AK, Sabiani SS. 2007. Mathematical modelling techniques in food and bioprocesses: an overiew. In: Sabiani SS, editor. Handbook of food and

bioprocess modelling techniques. Boca Raton, FL: Taylor & Francis Group. p 2-11.

55. DeGroot MH. 1962. Uncertainty, information and sequential experiments. Ann Math Stat. 33:404-419.
56. DeVoe DL, Pisano AP. 1997. Modeling and optimal design of piezoelectric cantilever microactuators. J Microelectomech S. 6:266-270.
57. De Vries JG, Lefort L. 2008. High-throughput experimentation and ligand libraries. In: De Vries JG, Elsevier CJ, editor. The Handbook of Homogeneous Hydrogenation. Weinheim: Wiley-VCH. p 1245-1278.
58. Diller DJ, Hobbs DW. 2004. Deriving knowledge through data mining high throughput screening data. J Med Chem. 47:6373-6383.
59. Edsall JT. 1947. The plasma proteins and their fractionation. Adv Protein Chem. 3:383-479.
60. EMA. 2000. Development pharmaceutics for biotechnological and biological products - annex to note for guidance on development pharmaceutics.
61. Etzel MR. 2008. Charged ultrafiltration and microfiltration membranes in antibody purification. In: Gottschalk U, editor. Process scale purification of antibodies. Hoboken, NJ: John Wiley & Sons. p 325-348.
62. Fahrner RL. 2001. Industrial purification of pharmaceutical antibodies: development, operation and validation of chromatography process. Biotechnol Gen Eng Rev. 18:301-327.
63. Fang, Z. 2010. Large-scale chromatography columns, modeling flow distribution. Encyclopedia of Industrial Biotechnology: Bioprocess, Bioseparation and Cell Technology. 1-18.
64. Farid SS, Washbrook J, Titchener-Hooker NJ. 2006. Modelling biopharmaceutical manufacture: design and implementation of SimBiopharma. Comput Chem Eng. 31:1141-1158.
65. Farid SS. 2007. Process economics of industrial monoclonal antibody manufacture. J Chromatogra B. 848:8-18.

66. 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. p 239-262.
67. Fasth M, Heijbel A, Karlsson A, Kurt-Fuentes R, Lindqvist S, Stenklo K. 2008. A new prepacked column format with Capto ion exchange media for process development. GE Healthcare. Application note: 28-9351-74 AA.
68. FDA. 1995a. Guidance: changes to be reported for product and establishment license applications.
69. FDA. 1995b. Guidance document concerning use of pilot manufacturing facilities for the development and manufacture of biological products.
70. FDA. 1996. Guidance: demonstration of comparability of human biological products, including therapeutic biotechnology-derived products.
71. FDA. 2001. Guidance for industry: monoclonal antibodies used as reagents in drug manufacturing.
72. FDA. 2004. Guidance for industry: process analytical technology, a framework for innovative pharmaceutical development, manufacturing and quality assurance.
73. FDA. 2006. Guidance for industry: Q8 pharmaceutical development.
74. Fedorov VV. 1969, 1971, 1972. Theory of optimal experiments. New York: Academic press. 292 p.
75. Fisher, RA. 1932. Statistical method for research workers. Can Med Assoc J. 27:460.
76. Fisher, RA. 1935. The design of experiments. Edinburgh: Oliver & Boyd. 251 p.
77. Fleiss JL. 1999. The design and analysis of clinical experiments. New York: John Wiley & Sons. 432 p.
78. Follman DK, Fahrner RL. 2004. Factorial screening of antibody purification processes using three chromatography steps without protein A. J Chromatogr A. 1024:79-85.
79. Ford I, Titterington DM, Kitsos CP. 1989. Recent advances in nonlinear experimental. Technometrics. 31:49-60.

80. 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.
81. Franceschini G, Macchietto S. 2007. Validation of a model for biodiesel production through model-based experiment design. Ind Eng Chem Res. 46:220-232.
82. Franceschini G, Macchietto S. 2008. Novel anticorrelation criteria for model-based experiment design: theory and formulations. AIChE J. 54:1009-1024.
83. Franceschini G, Macchietto S. 2008. Model-based design of experiments for parameter precision: State of the art. Chem Eng Sci. 63:4846-4872.
84. Freedman DA. 2005. Statistical models: theory and practice. New York: Cambridge University Press. 414 p.
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.
86. Galbraith DJ, Tait AS, Racher AJ, Birch JR, James DC. 2006. Control of culture environment for improved polyethylenimine-mediated transient production of recombinant monoclonal antibodies by CHO cells. Biotechnol Progr. 22:753-762.
87. Galvanauskas V, Simutis R, Lubbert A. 1997. Model-based design of biochemical processes: simulation studies and experimental tests. Biotechnol Lett. 19:1043-1047.
88. Galvanauskas V, Simutis R, Volk N, Lubbert A. 1998. Model based design of a biochemical cultivation process. Bioproc Eng. 18:227-234.
89. Galvanauskas V, Simutis R, Lubbert A. 2004. Hybrid process models for process optimisation, monitoring and control. Bioproc Biosyst Eng. 26:393-400.
90. Garcia-Arrazola R, Dawson P, Buchanan I, Doyle B, Feam Y, Titchener-Hooker NJ, Baganz F. 2005. Evaluation of the effects and interactions of mixing and

oxygen transfer on the production of Fab' antibody fragments in *Escherichia Coli* fermentation with gas blending. Bioproc Biosyst Eng. 27:365-374.

91. GE Healthcare. 2006a. Capto S cation exchanger for post-Protein A purification of monoclonal antibodies. Application Note: 28-4078-17 AA.
92. GE Healthcare. 2006b. Process-scale purification of monoclonal antibodies - polishing using Capto Q. Application Note: 28-9037-16 AA.
93. GE Healthcare. 2006c. Sreening and optimization of the loading conditions on Capto S. Application Note: 28-4078-16 AA.
94. 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.
95. GE Healthcare. 2009b. Scale-up of a downstream monoclonal antibody purification process using HiScreen™ and AxiChrom™ columns. Application note: 28-9403-49 AA.
96. GE Healthcare. 2011a. MabSelect. Instructions: 71-5020-91 AE.
97. GE Healthcare. 2011b. Superdex 200 10/300 GL. Instructions: 71-5017-96 AG.
98. George A, Wilson WW. 1994. Predicting protein crystallization from a dilute solution property. Acta Cryst. 50:361-365.
99. Gheshlaghi R, Scharer JM, Moo-Young M, Douglas PL. 2005. Medium optimization for hen egg white lysozyme production by recombinant *Aspergillus niger* using statistical methods. Biotechnol Bioeng. 90:754-760.
100. Ghosh R, Wang L. 2006 Purification of humanized monoclonal antibody by hydrophobic interaction membrane chromatography. J Chromatogr A. 1107:104-109.
101. 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. p 463-490.

102. 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.

103. Gilmour SG. 2006. Response surface designs for experiments in bioprocessing. Biometrics. 62:323-331.

104. 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.

105. 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. p 309-323

106. 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.

107. Goffaux G, Wouwer A. 2005. Bioprocess state estimation: some classical and less classical approaches. Control and observer design for nonlinear finite and infinite dimensional systems. Lect Notes Contr Inf. 322:111-128.

108. Gohel V, Chaudhary T, Vyas P, Chhatpar HS. 2006. Statistical screenings of medium components for the production of chitinase by the marine isolate *Pantoea dispersa*. Biochem Eng J. 28:50-56.

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

110. Gosling I. 2005. Process simulation and modelling for industrial bioprocessing: tools and techniques. Ind Biotechnol. 1:106-109.

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

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

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

114. Gregoire TG. 1998. Design-based and model-based inference in survey sampling: appreciating the difference. Can J Forest Res. 28:1429-1447.

115. Gu TY. 1995. Mathematical modeling and scale-up of liquid chromatography. London: Springer. 123 p.

116. Gunst RF, Mason RL. 2009. Fractional factorial design. Wiley Interdiscip Rev Comput Stat. 1:234-244.

117. Habib G, Zhou Y, Hoare M. 2000. Rapid monitoring for the enhanced definition and control of a selective cell homogenate purification by a batch-flocculation process. Biotechnol Bioeng. 70:131-142.

118. Hancock AA, Bush EN, Stanisic D, Kyncl JJ, Lin CT. 1988. Data normalization before statistical analysis: keeping the horse before the cart. Trends Pharmacol Sci. 9:29-32.

119. Harnett CK, Templeton J, Dunphy-Guzman KA, Senousy YM, Kanouff MP. 2008. Model based design of a microfluidic mixer driven by induced charge electroosmosis. Lab Chip. 8:565-572.

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

121. Hills RG, Leslie IH. 2003. Statistical validation of engineering and scientific models: validation experiments to application. Sandia National Laboratories. SAND99-1256.

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

123. Honig W, Kula MR. 1976. Selectivity of protein precipitation with polyethylene glycol fractions of various molecular weights. Anal Biochem. 72:502-512.

124. Hoskins R, Robb ID, Williams PA, Warren P. 1996. Phase separation in mixtures of polysaccharides and proteins. J Chem Soc. 92:4515-4520.

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

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

127. 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.
128. Invitrogen. 2008. Quant-iT PicoGreen dsDNA reagent and kits. MP 07581.
129. Invitrogen. 2011. Novex IEF gels. MAN 0003907.
130. Jackson NB, Liddell JM, Lye GJ. 2006. An automated microscale technique for the quantitative and parallel analysis of microfiltration operations. J Membrane Sci. 276:31-41.
131. Jain M, Paranandi M, Roush D, Goklen K, Kelly WJ. 2005. Using CFD to understand how flow patterns affect retention of cell-sized particles in a tubular bowl centrifuge. Ind Eng Chem Res. 44:7876-7884.
132. Jiang C, Liub J, Rubachaa M, Shukla AA. 2009. A mechanistic study of Protein A chromatography resin lifetime. J Chromatogra A. 31:5849-5855.
133. Jiang C, Flansburg L, Ghose S, Jorjorian P, Shukla AA. 2010. Defining process design space for a hydrophobic interaction chromatography (HIC) purification step: application of quality by design (QbD) principles. Biotechnol Bioeng. 107:985-997.
134. Jimenez-Gonzalez C, Woodley JM. 2010. Review bioprocesses: modeling needs for process evaluation and sustainability assessment. Comput Chem Eng. 34:1009-1017.
135. Jin M, Szapiel N, Zhang J, Hickey J, Ghose S. 2010. Profiling of host cell proteins by two-dimensional difference gel electrophoresis (2D-DIGE): implications for downstream process development. Biotechnol Bioeng. 105:306-316.
136. Jones SD, Seymour P, Levine HL. 2010. CMC activities for development of mAbs. Contract pharma. Apr, 60-64.
137. Kaltenbrunner O, Jungbauer A, Yamamoto S. 1997. Prediction of the preparative chromatography performance with a very small column. J Chromatogra A. 760:41-53.

138. Karlin S, Studden WJ. 1966. Optimal Experimental Designs. Ann Math Stat. 37:783-815.

139. Karlsson D, Jakobsson N, Axelsson A, Nilsson B. 2004. Model-based optimization of a preparative ion-exchange step for antibody purification. J Chromatogra A. 1055:29-39.

140. Karnopp DC. 1963. Random search techniques for optimization problems. Automatica. 1:111-121.

141. Katare S, Caruthers JM, Delgass WN, Venkatasubramanian V. 2004. An intelligent system for reaction kinetic modelling and catalyst design. Ind Eng Chem Res. 43:3484-3512.

142. Kato R, Nakano H, Konishi H, Kato K, Koga Y, Yamane T, Kobayashi T, Honda H. 2005. Novel strategy for protein exploration: high-throughput screening assisted with Fuzzy Neural Network. J Mol Biol. 351:683-692.

143. Keener RN, Fernandez EJ, Maneval JE, Hart RA. 2008. Advancement in the modeling of pressure-flow for the guidance of development and scale-up of commercial-scale biopharmaceutical chromatography. J Chromatogr A. 1190:127-140.

144. Kell DB, Sonnleitner B. 1995. GMP - Good Modelling Practice: an essential component of Good Manufacturing Practice. Tibtech. 13:481-492.

145. Kelley B. 2007. Very large scale monoclonal antibody purification: the case for conventional unit operations. Biotechnol Prog. 23:995-1008.

146. 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.

147. 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. p 1-24.

148. Kelly SM, Jess TJ, Price NC. 2005. How to study proteins by circular dichroism. BBA-Mol Cell Biol L. 1751:119-139.

149. Kiefer J. 1959. Optimum experimental designs, J Roy Stat Soc B. 21:273-319.

150. Kiefer J. 1971. The role of symmetry and approximation in exact design optimality. In: Gupta SS, Yachel J, editor. Statistical decision theory and related topics. New York: Academic Press. p 109-118.

151. King JMP, Titchener-Hooker NJ and Zhou Y. 2007. Ranking bioprocess variables using global sensitivity analysis: a case study in centrifugation. Bioproc Biosyst Eng. 30:123-134.

152. Kirkwood JG. 1943. The theoretical interpretation of the properties of solutions of dipolar ions. In: Cohn EJ, Edsall JT, editor. Protein, amino acids and peptides as ions and dipolar ions. New York: Reinhold Publishing. p 276-303.

153. 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.

154. Knudsen H, Fahrner RL, Xu Y, Norling LA, Blank GS. 2001. Membrane ion-exchange chromatography for process-scale antibody purification. J Chromatogr A. 907:145-154.

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

156. 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.

157. Kuehner DE, Blanch HW, Prausnitz JM. 1996. Salt-induced protein precipitation: Phase equilibria from an equation of state. Fluid Phase Equilibr. 116:140-147.

158. Kuhfeld WF, Tobias RD, Garratt M. 1994. Efficient experimental design with marketing research applications. J Mar Res. 31:545-557.

159. Lamprecht J. 2005. Applied data analysis for process improvement: a practical guide to Six Sigma black belt statistics. Milwaukee, WI: Quality press. 283 p.

160. Lazic ZR. 2004. Design of experiments in chemical engineering: a practical guide. Weinheim: Wiley-VCH. 610 p.

161. Lee KM, Gilmore DF. 2006. Statistical experimental design for bioprocess modeling and optimization analysis. Appl Biochem Biotech. 135:101-115.

162. Lee YP, Rangaiah GP, Luus R. 1999. Phase and chemical equilibrium calculations by direct search optimization. Comput Chem Eng. 23:1183-1191.

163. Levisauskas D, Galvanauskas V, Henrich S, Wilhelm K, Volk N, Lubbert A. 2003. Model-based optimization of viral capsid protein production in fed-batch culture of recombinant *Escherichia coli*. Bioproc Biosyst Eng. 25:255-262.

164. Lewis GA, Mathieu D, Phan-Tan-Luu R. 2005. Pharmaceutical experimental design. New York: Marcel Dekker. 512 p.

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

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

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

168. 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.

169. 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.

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

171. Ma G. 2010. Development of ultra scale-down shear filtration system and modelling of large scale diafiltration system. Doctoral thesis, UCL (University College London).

172. Ma JF, Hoang H, Myint T, Peram T, Fahrner R, Chou JH. 2010. Using precipitation by polyamines as an alternative to chromatographic separation in antibody purification processes. J Chromatogr B. 878:798-806.

173. Machold C, Deinhofer K, Hahn R, Jungbauer A. 2002. Hydrophobic interaction chromatography of proteins: I. comparison of selectivity. J Chromatogr A. 972:3-19.

174. Mahadevan H, Hall CK. 1992. Theory of precipitation of protein mixtures by nonionic polymer. AIChE J. 38:573-591.

175. Malo N, Hanley JA, Cerquozzi1 S, Pelletier J, Nadon R. 2006. Statistical practice in high-throughput screening data analysis. Nat Biotechnol. 24:167-175.

176. Mandenius CF, Brundin A. 2008. Bioprocess optimization using design of experiments methodology. Biotechnol Prog. 24:1191-1203.

177. Manning MC, Patel K, Borchardt RT. 1989. Stability of protein pharmaceuticals. Pharmacol Res. 6:903-918.

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

179. Marquardt W. 1996. Trends in computer-aided process modelling. Comput Chem Eng. 20:591-609.

180. Marques MPC, Cabral JMS, Fernandes P. 2010. Bioprocess scale-up: quest for the parameters to be used as criterion to move from microreactors to lab-scale. J Chem Technol Biot. 85:1184-1198.

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

182. Massimo CD, Willis MJ, Montague GA, Tham MT, Morris AJ. 1991. Bioprocess model building using artificial neural networks. Bioprocess Eng. 7:77-82.

183. Mayer DG, Butler DG. 1993. Statistical validation. Ecol model. 68:21-32.

184. McDonald P, Victa C, Carter-Franklin JN, Fahrner R. 2009. Selective antibody precipitation using polyelectrolytes: A novel approach to the purification of monoclonal antibodies. Biotechnol Bioeng. 102:1141-1151.

185. McPherson A, Malkin AJ, Kuznetsov YG, Koszelak S, Wells M, Jenkins G, Howard J, Lawson G. 1999. The effects of microgravity on protein crystallization: evidence for concentration gradients around growing crystals. J Cryst Growth. 196:572-586.

186. 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.

187. Michaelis L, Menten ML. 1913. Kinetk der Invertinwirkung. Biochem Z. 49:333-369.

188. 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.

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

190. Middelberg APJ, O'Neill BK, Bogle IDL. 1992. Modelling bioprocess interactions for optimal design and operating strategies (homogenisation and centrifugation of insoluble inclusion bodies). Food Bioprod Process. 70:8-12.

191. Milavec P, Podgornik A, Stravs R, Koloini T. 2002. Effect of experimental error on the efficiency of different optimization methods for bioprocess media optimization. Bioproc Biosyst Eng. 25:69-78.

192. Mitchell TJ, Miller FL. 1970. Use of design repair to construct designs for special linear models. Math Div Ann Progr Rept. Oak Ridge, TN: Oak Ridge National Laboratory. p 130-131.

193. Mitchell TJ. 1974. An algorithm for the construction of "D-optimal" experimental designs. Technometrics. 16:203-210.

194. Mohanty K, Ghosh R. 2007. Novel tangential-flow countercurrent cascade ultrafiltration configuration for continuous purification of humanized monoclonal antibody. J Membrane Sci. 307:117-125.

195. Montague G and Morris J. 1994. Neural-network contributions in biotechnology. Trends Biotechnol. 12:312-324.

196. Montgomery DC. 2008. Design and analysis of experiments. 7th edition. Hoboken NJ: John Wiley & Sons. 696 p.

197. Montgomery DC, Myers RH, Carter WH, Vining GG. 2005. The hierarchy principle in fesigned industrial experiments. Quality and Reliability Engineering International, 21: 197-201.

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

199. Mustafa MA, Washbrook J, Lim AC, Zhou Y, Titchener-Hooker NJ, Morton P, Berezenko S, Farid SS. 2004. A software tool to assist business-process decision-making in the biopharmaceutical industry. Biotechnol Progr. 20:1096-1102.

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

201. 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.

202. 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.

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

204. Niktari M, Chard S, Richardson P, Hoare M. 1990. The monitoring and control of protein purification and recovery processes. In: Pyle DL, editor. Separations for biotechnology. London: Elsevier Applied Science. p 622-634.

205. Oatley DL, Cassey B, Jones P, Bowen WR. 2005. Modelling the performance of membrane nanofiltration - recovery of a high-value product from a process waste stream. Chem Eng Sci. 60:1953-1964.

206. Papalambros PY, Wilde DJ. 2000. Principles of optimal design: modeling and computation. 2nd edition. Cambridge: Cambridge university press. 416 p.

207. Paterno F. 1999. Model-based design and evaluation of interactive applications. London: Springer-Verlag. 192 p.

208. 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.

209. Petrucceli S, Anon MC. 1996. pH-Induced modifications in the thermal dtability of doybean protein isolates. J Agric Food Chem. 44:3005-3009.

210. Pfeifer AC, Kaschek D, Bachmann J, Klingmuller U, Timmer J. 2010. Model-based extension of high-throughput to high-content data. BMC Syst Biol. 4:106-118.

211. Phillips MW, Bolton G, Krishnan M, Lewnard JJ, Raghunath B. 2007. Virus filtration process design and implementation. In: Shukla AA, editor. Process scale bioseparations for the biopharmaceutical industry. Boca Raton, FL: Taylor & Francis Group. p 333-367.

212. Prashad M, Tarrach K. 2006. Depth filtration: cell clarification of bioreactor offloads. Filtr Separat. 43:28-30.

213. Prats M. 2010. Minireview: Isoelectric focusing. Biochem Edu. 20:109-111.

214. Przybycien TM, Pujar NS, Steele LM. 2004. Alternative bioseparation operations: life beyond packed-bed chromatography. Curr Opin Biotech. 15:469-478.

215. Pukelsheim F. 1993. Optimal design of experiments. New York: John Wiley & Sons. 454 p.

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

217. 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.

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

219. Reis R, Brake JM, Charkoudian J, Burns DB, Zydney AL. 1999. High-performance tangential flow filtration using charged membranes. J Membrane Sci. 159:133-142.

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

221. Riel NAW. 2006. Dynamic modelling and analysis of biochemical networks: mechanism-basedmodels and model-based experiments. Brief Bioinform. 7:364-374.

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

223. Rivera EC, Costa AC, Andrade RR, Atala DP, Maugeri F, Filho RM. 2007. Development of adaptive modeling techniques to describe the temperature dependent kinetics of biotechnological processes. Biochem Eng J. 36:157-166.

224. Robbins H. 1952. Some aspects of the sequential design of experiments. Bullet Am Math Soc. 58:527-536.

225. Rodrigues, AD. 1997. Preclinical drug metabolism in the age of high throughput screening: an industrial perspective. Pharmacol Res. 14:1504-1510.

226. Rodriguez-Fernandez M, Mendes P, Banga JR. 2006. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. Biosystems. 83:248-265.

227. Rohner M, Meyer HP. 1995. Applications of modelling for bioprocess design and control in industrial production. Bioproc Biosyst Eng. 13:69-78.

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

229. 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.

230. Rosa P, Ferreira I, Azevedo AM, Aires-Barros MR. 2010. Aqueous two-phase systems: a viable platform in the manufacturing of biopharmaceuticals. J Chromatogr A. 1217:2296-2305.

231. Ruppert S, Sandler S, Lenhoff A. 2008. Correlation between the osmotic second virial coefficient and the solubility of proteins. Biotechnol Progr. 17:182-187.

232. Sacks J, Welch WJ, Mitchell TJ, Wynn HP. 1989. Design and analysis of computer experiments. Stat Sci. 4:409-435.

233. Salcedo R, Goncalves MJ, Azevedo SF. 1990. An improved random-search algorithm for nonlinear optimization. Comput Chem Eng. 14:1111-1126.

234. Salte H, King JMP, Baganz F, Hoare M, Titchener-Hooker NJ. 2006. A methodology for centrifuge selection for the separation of high solids density cell broths by visualisation of performance using windows of operation. Biotechnol Bioeng. 95:1218-1227.

235. Sarndal CE, Thomsen I, Hoem JM, Lindley DV, Barndorff-Nielsen O, Dlenius T. 1978. Design-Based and Model-Based Inference in Survey Sampling. Scand J Stat. 5:27-52.

236. Schiff D, D'Agostino RB. 1996. Practical engineering statistics. New York: John Wiley & Sons. 309 p.

237. Schubert J, Simutis R, Dors M, Havlik I, Lubber A. 1994. Bioprocess optimization and control: application of hybrid modelling. J Biotechnol. 35:51-68.

238. Sendin OH, Vera J, Torres NV, Banga JR. 2006. Model based optimization of biochemical systems using multiple objectives: a comparison of several solution strategies. Math Comput Model Dyn. 12:469-487.

239. 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. p 1-36.

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

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

242. 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.

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

244. Simutis R, Oliveira R, Manikowski M, Azevedo SFD, Lubbert A. 1997. How to increase the performance of models for process optimization and control. J Biotechnol. 59:73-89.

245. Singer AB, Taylor JW, Barton PI, Green WH. 2006. Global dynamic optimization for parameter estimation in chemical kinetics. J Phys Chem A. 110:971-976.

246. 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. p 80-92.

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

248. Souza SM, Araujo OQF, Coelho MAZ. 2008. Model-based optimization of a sequencing batch reactor for biological nitrogen removal. Bioresource Technol. 99:3213-3223.

249. 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. p 118-158.

250. 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.

251. Steinberg DM, Hunter WG. 1984. Experimental design: review and comment. Technometrics. 26:71-97.

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

253. Stickel JJ, Fotopoulos A. 2001. Pressure-flow relationships for packed beds of compressible chromatography media at laboratory and production scale. Biotechnol Progr. 17:744-751.

254. 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.

255. 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.

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

257. 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.

258. Torres JZ, Miller JJ, Jackson PK. 2009. High-throughput generation of tagged stable cell lines for proteomic analysis. Proteomics. 9: 2888-2891.

259. Tsao TC, Bailey K, Adair GS. 1951. The size, shape and aggregation of tropomyosin particles. Biochem J. 49:27-36.

260. Tugcu N, Roush DJ, Goklen KE. 2008. Maximizing productivity of chromatography steps for purification of monoclonal antibodies. Biotechnol Bioeng. 99:599-613.

261. Van Impe. 1996. Power and limitations of model based bioprocess optimization. Math Comput Simulat. 42:159-169.

262. Vasic-Racki D, Bongs J, Schorken U, Sprenger GA, Liese A. 2003. Modeling of reaction kinetics for reactor selection in the case of L-erythrulose synthesis. Bioprocess Biosyst Eng. 25:285-290.

263. Velayudhan A, Menon MK. 2007. Modeling of purification operations in biotechnology: enabling process development, optimization and scale-up. Biotechnol Progr. 23:68-73.

264. Watts DG. 1994. Estimating rate parameters in nonlinear rate equations. Can J Chem Eng. 72:701-710.

265. Werner RG. 2004. Economic aspects of commercial manufacture of biopharmaceuticals. J Biotechnol. 113:171-182.

266. Weuster-Botz. 2000. Experimental design for fermentation media development: statistical design or global random search? J Biosci Bioeng. 5:473-483.

267. Wiendahl M, Wierling SP, Nielsen J, Christensen FD, Krarup J, Staby A, Hubbuch J. 2008. High throughput screening for the design and optimization of chromatographic processes - miniaturization, automation and parallelization of breakthrough and elution studies. Chem Eng Technol. 31:893-903.

268. Wilcoxon F. 1945. Individual comparisons by ranking methods. Biometrics. 1:80-83.

269. Willoughby N. 2006. Perspective scaling up by thinking small: a perspective on the use of scale-down techniques in process design. J Chem Technol Biotechnol. 81:1849-1851.

270. Winzor D, Carrington L, Harding S. 2001. Analysis of thermodynamic non-ideality in terms of protein salvation. Biophys Chem. 93:231-240.

271. White RE. 2000. High-throughput screening in drug metabolism and pharmacokinetic support of drug discovery. Annu Rev Pharmacol. 40: 133-157.

272. Wong HH, O'neil BK, Middelberg APJ. 1996. Centrifugal processing of cell debris and inclusion bodies from recombinant *Escherichia coli*. Bioseparation. 6:361-372.

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

274. Wynn HP. 1972. Results in the theory and construction of D-optimum experimental designs. J Roy Stat Soc B. 34:133-147.

275. Yamamoto S, Nakanishi K, Matsuno R, Kamikubo T. 1983. Ion exchange chromatography of proteins - prediction of elution curves and operating conditions. I. Theoretical considerations. Biotechnol Bioeng. 25:1465-1483.

276. Yates F. 1959. Discussion on optimum experimental designs. J Roy Stat Soc B. 21:308-309.

277. Yigzaw Y, Piper R, Tran M, Shukla AA. 2008. Exploitation of the adsorptive properties of depth filters for host cell protein removal during monoclonal antibody purification. Biotechnol Progr. 22:288-296.

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

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

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

281. Zhou YH, Titchener-Hooker NJ. 2003. The application of a Pareto optimisation method in the design of an integrated bioprocess. Bioproc Biosyst Eng. 25:349-355.

282. Zorzetto LFM, Filho RM, Wolf-Maciel MR. 2000. Processing modelling development through artificial neural networks and hybrid models. Comput Chem Eng. 24: 1355-1360.

283. 3M Cuno. 2007. Zeta Plus™ Maximizer EXT series: featuring SP filter media. Document No.: LITZPMEXTSP.E0208.

284. 3M Cuno. 2008. Installation and operation instructions for Zeta Plus™ BC25 disposable capsules. Document No.: LITOPSBC25ML.0808.

Appendix 1. Matlab codes for Model-based process design algorithm

1. Seachalgorithm.m

```
% This will be the overall algorithm for model-based,greedy and simplex
% methods (This version 1.0 only for model-based, greedy/random test.
% simplex will be corporated in later version due to its too complicated code)

fprintf('This algorithm is based on D-optimal design, greedy/random search and modified simplex\n')
fprintf('Specific for precipitation model with five parameters\n')
fprintf('Designed by Yu Ji, Biochemical Engineering Department, UCL\n')
fprintf('**********************************************************************\n')
% title complete
fprintf('This algorithm is for Fab AS precipitation with fixed design space\n')
fprintf('Salt from 0M to 3M with step 0.3, pH from 4.5 to 8 with step 0.5. All value normalized to 0-1\n')
fprintf('Designspace is written in the code, if want to use new space, change value or use ''input'' phrase replace this ''fprintf'' phrase\n')
fprintf('Input require initial coefficients for both Fab and impurities model in 1x5 matrix format\n')
fprintf('Existing conditions in [salt concentration, pH] format\n') % real data used in following algorithm are index, need convert
fprintf('Corresponding results at that conditions in [Fab, impurities] format\n') % data need convert due to following algorithm
fprintf('**********************************************************************\n')
% initial input
coeff1=input('Initial Fab coefficient in 1x5 matrix form [x,x,x,x,x...]\n');
coeff2=input('Initial impurities coefficient in 1x5 matrix form [x,x,x,x,x...]\n');
excondition=input('Existing conditions in a matrix form [x,y;x,y;...], real pH and salt concentration, not normalized, at least six points\n'); % real value
exdata=input('Results corresponding to exconditions in matrix form [x1,x2;x1,x2;x1,x2;..] fab data first and impurities data second\n'); % input data
% designspace is a normalized matrix for reaction conditions with first
% column salt and second pH
ph=(4.5-4.5)/3.5:0.5/3.5:(8-4.5)/3.5;
salt=(0-0)/3.0:0.3/3.0:(3-0)/3.0; % give normalized value in matrix form to give designspace matrix
```

```
[x,y]=meshgrid(salt,ph); % for surface function 8x11 matrix
condition1=[0,0.3,0.6,0.9,1.2,1.5,1.8,2.1,2.4,2.7,3];
condition2=[4.5,5,5.5,6,6.5,7,7.5,8];
for i1=1:1:8
    for j1=1:1:11
        designspace((i1-1)*11+j1,:)=[condition1(j1),condition2(i1)];
    end
end
designspace(:,1)=designspace(:,1)./3;
designspace(:,2)=(designspace(:,2)-4.5)./3.5; % covert to normalized value
% convert excondition to conditions by index number
excondition(:,1)=excondition(:,1)./3;
excondition(:,2)=(excondition(:,2)-4.5)./3.5;
[a1,b1]=size(excondition); % a1=n, the number of initial experiments which also
equals to the number of exdata
for i2=1:1:a1
    for j2=1:1:88
        if excondition(i2,:)==designspace(j2,:) % find the corresponding condition
            conditions(i2)=j2; % give index
        end
    end
end
realresults=zeros(88,2); % initial results for fab and impurities data
objresults=zeros(88,1); % initial results for objective function value
for i3=1:1:a1
    n=conditions(i3); % find the index of that condition
    realresults(n,:)=exdata(i3,:); % store data at realresults matrix in corresponding
index
    objresults(n)=objectivefun(realresults(n,1),realresults(n,2),designspace(n,:)); %
store objresults according to real data and index in the designspace
    sortmatrix(i3,:)=[n,objresults(n)]; % store in sort matrix for following sort in
initial simplex points choose step
end
% all data convert to required format in the following algorithm
ctrlnum=1;
reflection=0;
reflectindex=[0,0];
spenum=0; % initial value for simplex
exva=[0,0]; % initial value for sorting, format as sortmatrix one row
% choose points from the inital set without linearity
for i4=1:1:a1-1
    for j4=1:1:a1-1
        if sortmatrix(j4,2)>sortmatrix(j4+1,2)
            continue;
```

```
            else exva=sortmatrix(j4,:);
                    sortmatrix(j4,:)=sortmatrix(j4+1,:);
                    sortmatrix(j4+1,:)=exva;
            end
        end
end % famous bubble sorting algorithm, the largest value on the top of matrix, with
first column is index number and second is objvalue
% take two largest points out and one lowest point (the steepest increase) and store
other index in one matrix
candidatepoints=[sortmatrix(1,1);sortmatrix(2,1);sortmatrix(a1,1)]; % the largest
first
for i5=1:1:a1-3
        remindex(i5)=sortmatrix(i5+2,1); % store remaining sorted index for random
choose in following case
end
% following check linearity using loop, if linear, change the third with
% the second lowest point to get the faster and so on repeat until no
% longer linear
lopnu=0; % initial value for linear loop in case repeat need to give index
linnum=1; % control value for loop
while linnum==1
        linnum=linearcheck(candidatepoints,designspace); % ******need
linearcheck.m, return 1 means linear, 0 means no linear
        if linnum==0
            break; % no linear
        elseif linnum==1
            lopnu=lopnu+1; % index count increases one
            if lopnu==a1-2 % means all points checked and all in the same line
                fprintf('All points on the same line, close and restart program, use
another set of data\n')
                break; % break, but still need a stop sentence after while loop(do not
coded due to very rare case), in this algorithm, once happen, need mannually close
            else candidatepoints(3)=sortmatrix(a1-lopnu,1); % the next lowest index
            end
        end
end
% all parameters needed in the following algorithm completed
% model based algorithm requires two coefficients, conditions in index
% matrix,designspace with normalized data in matrix
% coefficients regression will happen before model based algorithm but
% still remains in the loop, it requires excondition and exdata (the final
% of the loop will store new conditions and data to this two matrix)
%***************main algorithm modules starts here to three different
%function, which can be replaced by other module function if possible, or
```

```
%only use one or two modules.
loopctronum=1; % initial loopctronum
while loopctronum==1 % loopctronum is used to control loop, when 0 stop, The stop
criteria:
    % 1. all points searched, 2. simplex stops due to linear or small area,
    % 3. achieve the required level, in this case, mannually stop by input
    % loopctronum=0, 4. some break in the unpredictable case.
    % the first will be regress function, which should be specificed by the
    % model and range of parameter, in this model, the fifth parameter
    % should be limited positive
    coeff1=coreg(coeff1,excondition,exdata(:,1)); % **********program in
coreg.m
    coeff2=coreg(coeff2,excondition,exdata(:,2)); % **********program in
coreg.m

[nexdopt1,dmatrixvalue1,nexrandom1]=doptimal(coeff1,conditions,designspace); %
nextpoint from Fab

[nexdopt2,dmatrixvalue2,nexrandom2]=doptimal(coeff2,conditions,designspace); %
nextpoint from impurities

[nexpoigreedy,maxvalue,nexrandom3,objvalue,modelvalue1,modelvalue2]=greedyse
arch(coeff1,coeff2,conditions,designspace); % nextpoint from greedy/random search
    % the following loop was used for surface plot but in reality, the
    % existing poits value should replace the model calculated value to
    % show half reality and remaining model based value, otherwise, the
    % model based surface will totally show a different pattern
    [ncon1,ncon2]=size(conditions); % get the size of matrix
    for replanum=1:1:ncon2
        indexforreal=uint8(conditions(replanum)); % give index

objvalue(indexforreal)=objectivefun(realresults(indexforreal,1),realresults(indexforr
eal,2),designspace(indexforreal,:)); % give real objvalue according to real data
        modelvalue1(indexforreal)=realresults(indexforreal,1);
        modelvalue2(indexforreal)=realresults(indexforreal,2);
    end % after this loop model based data at existing points are replaced by real
value according to experiments, this will cause the following surface plot not smooth
    for ynum=1:1:8
        for xnum=1:1:11
            objsurface(ynum,xnum)=objvalue((ynum-1)*11+xnum); % make into
surface matrix with 8x11 specific for objective value
            fab(ynum,xnum)=modelvalue1((ynum-1)*11+xnum); % make into
surface matrix with 8x11 specific for fab value
            tot(ynum,xnum)=modelvalue2((ynum-1)*11+xnum); % make into
```

```
surface matrix with 8x11 specific for impurities value
            end
      end

[nextsimpoint,reflection,reflectindex,ctrlnum,spenum,candidatepoints]=nextsimplex(
candidatepoints,conditions,designspace,ctrlnum,spenum,reflection,reflectindex,objva
lue); % simplex, due to it requires real
      % objvalue replacing model calculated value.
      % ctrlnum,spenum,reflection,reflectindex initialized before loop,and
      % keep updated in internal loop or points select algorithm in the end
      % of this algorithm before a new loop starts. designspace is a fixed
      % matrix. candidatepoints will be updated in points select algorithm,
      % and conditions keep all carried out experiments index in it when loop
      % goes on. reflectindex,ctrlnum,spenum,nextsimpoint,objvalue will be
      % required in points select algorithm with real data inputed at
      % nextsimpoint condition
      % ***above algorithm gives nextsimpoint from simplex
      % objvalue updated in line 124,
      i6=1; % for skip and store nextpoint
      nexpoint=[0]; addpoint=[0]; % initial nexpoint&addpoint as a zero matrix to
avoid remaining value from last run

%1111111111111111111111111**********************************start
doptimal module by Fab and impurities, with two new
      % points
      if nexdopt1==0
            if nexrandom1==-1
                  fprintf('No d-optimal points for Fab, all points used\n'); % end pharse
                  loopctronum=0;
            else
                  nexpoint(i6)=nexrandom1; % give nexpoint(1) index
                  i6=i6+1;
            end
      else      nexpoint(i6)=nexdopt1; % give nexpoint(1) index
                  i6=i6+1;
      end % finish Fab doptimal point
      if nexdopt2==0
            if nexrandom2==-1
                  fprintf('No d-optimal points for impurities, all points used\n'); % end
pharse
                  loopctronum=0;
            else
                  nexpoint(i6)=nexrandom2; % give nexpoint(2) index, if nexdopt does
not exists, i6=1 here. in this case it will not happen, so always i6=2
```

```
                i6=i6+1;
            end
        else        nexpoint(i6)=nexdopt2; % give nexpoint(2) index
                i6=i6+1;
        end % finish impurities doptimal point
        %
2222222222222222222**********************************************
**start greedy/random module by objvalue, with one
        % points from model max or random
        if nexrandom3==-1
            fprintf('No max points for objective value, all points used\n'); % end
pharse
            loopctronum=0;
        elseif nexrandom3==0 % max point available without repeat
                nexpoint(i6)=nexpoigreedy; % give index to nexpoint(3)
                i6=i6+1;
        else nexpoint(i6)=nexrandom3; % give random index to nexpoint(3)
             i6=i6+1;
        end
        % finish greedy/random point
        %
333333333333333333**********************************************
*
        % start simplex module by objvalue and candidatepoints, nexpoint input
        if nextsimpoint==-1 % small area or repeat with all existing points
            fprintf('No simplex point due to small area,ignore following next point from
simplex search\n'); % end pharse
            loopctronum=0;
            nexpoint(i6)=nexpoint(1);
        else nexpoint(i6)=nextsimpoint;
        end % finish simplex point
        % check repeat and then output
        [b2,a2]=size(nexpoint); % get the row number of nexpoint
        nexcon=[0,0]; % initial nexcon for loop
        for i7=1:1:a2
            nexcon(i7,:)=designspace(nexpoint(i7),:); %conditions in normalized value
        end
        nexrealcon=zeros(a2,2); % initial value
        nexrealcon(:,1)=nexcon(:,1).*3; % convert to real salt concentration
        nexrealcon(:,2)=nexcon(:,2).*3.5+4.5; % real ph value
        % find the real max from existing conditions, the max value existing,
        % not from model but from excondition,exdata,
        [exi,exj]=size(exdata); % find the total number of existing points
        maxobjreal=0;maxrealind=0;
```

```
    for nreali=1:1:exi
tobjvalue(nreali)=objectivefun(exdata(nreali,1),exdata(nreali,2),excondition(nreali));
        if tobjvalue(nreali)>maxobjreal
            maxobjreal=tobjvalue(nreali);
            maxrealind=nreali;
        end
    end % find the max index and objvalue
    predimax(1,:)=designspace(nexpoigreedy,:); % model predicted max conditions
    predimaxcon(1,1)=predimax(1,1).*3;
    predimaxcon(1,2)=predimax(1,2).*3.5+4.5;
    curmax(1,:)=excondition(maxrealind,:); % current max conditions
    curmaxcond(1,1)=curmax(1,1).*3;
    curmaxcond(1,2)=curmax(1,2).*3.5+4.5;
    cmi=uint8(curmaxcond(1,1)/0.3+1);
    cmj=uint8((curmaxcond(1,2)-4.5)/0.5+1);
    fprintf('The next point from Fab d-optimal is                salt %f pH
%f\n',nexrealcon(1,1),nexrealcon(1,2));
    fprintf('The next point from impurities d-optimal is    salt %f pH
%f\n',nexrealcon(2,1),nexrealcon(2,2));
    fprintf('The next point from greedy/random search is        salt %f pH
%f\n',nexrealcon(3,1),nexrealcon(3,2));
    fprintf('The next point from simplex search is               salt %f pH
%f\n',nexrealcon(4,1),nexrealcon(4,2));
fprintf('The max predicted objective value from model is                %f at
conditions salt %f pH %f\n',maxvalue,predimaxcon(1,1),predimaxcon(1,2));
    fprintf('The max current objective value    is                %f at conditions
salt %f pH %f\n',objsurface(cmj,cmi), curmaxcond(1,1),curmaxcond(1,2));
    figure (1)
    surface(x,y,fab);
    figure (2)
    surface(x,y,tot);
    figure (3)
    surface(x,y,objsurface);
    [sfab,stot,sobj]=validcalculate(fab,tot,objsurface); % give value of derivation
    fprintf('The derivation value for Fab, impurities and objective value is \n Fab:
%f   impurities: %f   Objvalue: %f\n',sfab,stot,sobj);
    for i8=1:1:a2-1
        for i9=i8+1:1:a2
            if nexpoint(i9)==nexpoint(i8) % repeat
                nexpoint(i9)=0; % 0 value repeat point
            end
        end
    end
    n2=1; % initial value for remove repeat
```

```
    for i10=1:1:a2
        if uint8(nexpoint(i10))==0
        else addpoint(n2)=nexpoint(i10);
             n2=n2+1;
        end
    end % addpoint was used for adding to conditions and get results
    [b3,a3]=size(addpoint); % get the row number of all point without repeat
    addcon=[0,0];    % inital value for loop
    for i11=1:1:a3
        addcon(i11,:)=designspace(addpoint(i11),:); %conditions in normalized
value
    end
    addrealcon=zeros(a3,2); % initial matrix
    addrealcon(:,1)=addcon(:,1).*3; % convert to real salt concentration
    addrealcon(:,2)=addcon(:,2).*3.5+4.5; % real ph value
    inputsure=0; addresults=[0,0]; % initial value for following loop
    while inputsure==0 % input wrong number, reinput
        for i12=1:1:a3
            fprintf('Next point salt %f pH
%f\n',addrealcon(i12,1),addrealcon(i12,2));
            addresults(i12,:)=input('Input Fab and impurities results at that point
in [Fab, impurities] format\n');
        end
        fprintf('Is the data correct? correct=1, no=0\n');
        inputsure=input('Input correct value    ');
    end
    [a4,b4]=size(excondition); % find how many runs have done
    % add new points to excondition,exdata and conditions, coeff1 & 2 are
    % updated by loop. designspace does not change
    for addi=1:1:a3
        realresults(uint8(addpoint(addi)),:)=addresults(addi,:); % give to
realresults for loop plot purpose
        excondition(a4+addi,:)=addcon(addi,:); % in normalized value adding
        exdata(a4+addi,:)=addresults(addi,:); % results adding
        conditions(a4+addi)=addpoint(addi); % in index form adding

objvalue(uint8(addpoint(addi)))=objectivefun(realresults(uint8(addpoint(addi)),1),re
alresults(uint8(addpoint(addi)),2),designspace(uint8(addpoint(addi)),:));
    end % updated by new results
    %
**********************************************************************
**start simplex points select algorithm

[candidatepoints,ctrlnum,spenum,reflection,reflectindex,loopctronum]=simpointsele
```

```
ct(candidatepoints,conditions,addpoint,realresults,designspace,ctrlnum,spenum,refle
ction,reflectindex,nextsimpoint,objvalue,loopctronum);
      % for full version here need a evaluation program according to results
      % and change loopctronum value to stop the loop********************
end
```

2. Objectivefun.m

```
% This code is used to calculate the objective function in greedysearch for
% optimization, modelvalue1 is fab, modelvalue2 is impurities. The
% criteria will be specified
% if yield(i,j)<0.7
% obj(i,j)=0.4.*yield(i,j)+0.6.*pf(i,j)-0.1.*(j-1)*0.1;

function objvalue=objectivefun(modelvalue1,modelvalue2,designspace) %
designspace will not be matrix in this function
fab=modelvalue1;
tot=modelvalue2;
if fab>0.3
      objvalue=0.4.*fab+0.6.*fab./tot-0.1.*designspace(1);
else objvalue=0.4.*(1-fab)+0.6.*(1-fab)./(1-tot)-0.1.*designspace(1);
end
end
```

3. Linearcheck.m

```
% This code is used to check the linearity of candidatepoints or in any
% other case, if linear return 1, if not, return 0

function linnum=linearcheck(candidatepoints,designspace)
bpoint=designspace(candidatepoints(1),:);
npoint=designspace(candidatepoints(2),:);
wpoint=designspace(candidatepoints(3),:);
a=npoint-bpoint;
b=wpoint-bpoint;
matxdetermintant=det([a;b]);
if matxdetermintant==0
      linnum=1;
else linnum=0;
end
end
```

4. coreg.m

```
% This code is used for parameters regression for precipitation model with
% the fifth parameter limited positive

function coeff=coreg(coeff,excondition,exdata) % excondition nx2 matrix in
normalized value, exdata nx1 matrix, different from the main algorithm
% coeff will be updated in this algorithm
n=0; % number for stop regress in case that the coeff(5) can not be positive
coeff=lsqcurvefit(@ppmodel,coeff,excondition,exdata); % lsq regress, better
algorithm can be used here
while coeff(5)<0
      coeff(5)=rand(1); % random give value between [0,1]
      n=n+1; % account 1 more
      if n==20 % 20 loop to stop
            break
      else coeff=lsqcurvefit(@ppmodel,coeff,excondition,exdata); % repeat regress
until the right coeff comes out
      end
end
end
```

5. doptimal.m

```
% This code is used as a part of new combining methods for (n+1) d-optimal
% design choose from candidate points in the designspace
% The input will be 1. model coefficients, 2. design space in matrix format
% according to the experiments and interval steps, 3. the model functions
% which depends on the model used, which is changable, 4. data alreday used
% including conditions and results
% symbol: coeff, designspace(defined by model and experiments), interstep
% (information contained in design space), modelname(for model function,
% in the first case study two variables precipitation model), conditions
% it can be change to other two factor model such as polynomial equation
% and if upgrade the code to three or four variables it will therotical works
% with the number of variables also changed in corresponding code (simplex)
% The output should be conditions where d-optimal gives max, d-efficiey
% maybe ok to give out as a optional judge value
% In some case, there is no d-optimal max point (any extra point added
% into d-matrix will decrease dmatrix value, matrix size n), use random
% search
% in optional methods, delete the min. conditions to (n-1) matrix and then
% exchange two conditions to (n+1) matrix. thus this code will give two
% point extra, then the following loop need change the existing conditions
% and simplex point choice criteria, which is more like mitchell sequential
% exchange d-optimal method
```

```
% symbol: nexdopt, dmatrixvalue,deffic(optional, not programmed in this case)
% nexrandom(same as greedysearch.m store random information)
% function file used in this function dmatrix.m

function [nexdopt,dmatrixvalue,nexrandom]=doptimal(coeff,conditions,designspace)
%%% when nexrandom=0, means d-optimal design is availabel, otherwise, random
search. the only exception is nexrandom=-1, which means, all points are used
% conditions are index number, in order to use in dmatrix calculation, need
% convert to real normalized conditions in designspace
[a,b]=size(conditions);
[a1,b1]=size(designspace);
realcon=[0,0]; % initial value
for i=1:1:b
      j=uint8(conditions(i)); % let j=index number, which show the conditions in
designspace
      realcon(i,:)=designspace(j,:);
      designspace(j,:)=2; % let corresponding space have one special value, which
can be recongized as existing points in later algorithm,since normalized to 0-1, 2 is
ok to use
end % this loop convert realcon matrix with real conditions
inidmat=dmatrix(coeff,realcon); % give d-optimal matrix according to realcon
matrix & coefficients &****** dmatrix.m file need programed according to model
chosen******
tempdmatvalue=det(inidmat'*inidmat); % give the template d-matrix the initial
value used for the next (n+1) search
num=0; % give the num initial value zero for index number, it will store the max
dmatrix conditions index in the following code
for ii=1:1:a1
      if uint8(designspace(ii,1))==2    % 2 means that points existing in the realcon,
thus skip this ii to next
            onepoidmatrix(ii)=0; % matrix in order to store dmatrix at that point is
zero, can not be used for sorting since maybe the real matrix is less than zero
            continue;
      else
            tempcon=realcon; % let tempcon has the realcon value (transfer existing
conditions to tempcon matrix ax2, in precipitation case)
            tempcon(a+1,:)=designspace(ii,:); % let tempcon has one more conditions
from candidate points (n+1) matrix
            tempdmatrix=dmatrix(coeff,tempcon); % give (n+1) matrix form to
tempdmatrix by adding iith designspace into the new d-matrix
            onepoidmatrix(ii)=det(tempdmatrix'*tempdmatrix); % store new dmatrix
value to corresponding index number matrix onepoidmatrix, if not the new point,
that index store 0 in above 'if' pharse
            if onepoidmatrix(ii)>tempdmatvalue
```

```
                tempdmatvalue=onepoidmatrix(ii); % let tempdmatvalue always store
the max (n+1) dmatrix value
                num=ii; % store the index in the num
            end
        end
end % after whole loop, onepoidmatrix store all det value at that point adding, num
store the max index, if no point adding can give larger dmatrix value, num=0
if num==0 % start random search, in some cases, it can be change to other d-optimal
methods or even g-, a- optimal methods, in this case for simple, random search
        nexdopt=0; % no option for d-optimal
        dmatrixvalue=det(inidmat'*inidmat); % existing dmatrix value
        repn=1; %initial value for random search
        jj=0;    % if all points used, jj=a1;let nexrandom=-1
        while repn==1 %loop for random search until no repeat point comes out or no
candidate points (very rare situation), in this case, give nexrandom value -1
            if jj==a1      % which means all points used
                nexrandom=-1;break; % while loop break with three value returned
            else
            nexrandom=randi(a1,1); % random choose an integer from designspace
conditions
            repn=reptest(conditions,nexrandom); % chech whether new point is also
repeated
            jj=jj+1; % add one for each loop
            end
        end % after this while loop, one random point will be choose or give
nexrandom=-1
else nexdopt=num; % let num value to nexdopt output
         dmatrixvalue=onepoidmatrix(num); % return the new dmatrix value
         nexrandom=0; % 0 means no random search.
end
end
```

6. greedysearch.m

```
% This code is used as a part of new combining methods for greedy/max
% search with random search following when repeat happens
% The input will be 1. model coefficients, coeff1 fab, coeff2, impurities
% 2. design space in matrix format according to the experiments and interval
% steps, 3. the model functions which depends on the model used, which is
% changable. 4. Data already used
% symbol: coeff, designspace(defined by model and experiments), interstep
% (information contained in design space), modelname(for model function,
% in the first case study two variables precipitation model), conditions
% it can be change to other two factor model such as polynomial equation
```

```
% and if upgrade the code to three or four variables it will therotical works
% with the number of variables also changed in corresponding code (simplex)
% The output should be the max point conditions with value
% symbol: nexpoigreedy, maxvalue,nexrandom(when max point repeat in the loop)
% function file used in this function ppmodel.m and reptest.m,
% objectivefun.m

function
[nexpoigreedy,maxvalue,nexrandom,objvalue,modelvalue1,modelvalue2]=greedysea
rch(coeff1,coeff2,conditions,designspace) %%% when nexrandom=0, means new
max point design is availabel, otherwise, random search. the only exception is
nexrandom=-1, which means, all points are used
% designspace always use ixj matrix, i=n experiments, j=number of
% variables, it is not easy for surface plotting, but good for model value
% calculation and easy for upgrade to higher order
[a,b]=size(designspace); % a,b get value of i and j
tempmax=-1; % tempmax used to store template max value produce in the following
loop. Since the ppmodel range from 0 to 1, so initial =-1. for other models change
value.
num=0; % num used to store template max condition index number in the following
loop.
for i=1:1:a % this loop used to find the greedy/max point in all designspace
      modelvalue1(i)=ppmodel(coeff1,designspace(i,:)); % calculate model1 value at
certain coeff and i row conditions,fab
      modelvalue2(i)=ppmodel(coeff2,designspace(i,:)); % calculate model2 value at
certain coeff and i row conditions,tot
      % ppmodel will be function file contains required model structure, in
      % ****** this model file need program in ppmodel.m file ***********
      % this case, precipitation model with five parameters, for other use,
      % it should be changed, i.e. polynomial equation
      objvalue(i)=objectivefun(modelvalue1(i),modelvalue2(i),designspace(i,:)); %
**** objectivefun.m should be programed ****
      if objvalue(i)>tempmax % max algorithm to compare and store template max
value and index number
            tempmax=objvalue(i);
            num=i;
      end % after all i=1:1:a loop, tempmax will be the max value in all design space
and num will be the corresponding condition index
end % since all model value were calculated, it can be used for further study, i.e. find
all local and global points as candidate points for design, which requires rigor
      % changes in simplex algorithm (the basic choice will be the same, but
      % need consider the code for filter all local points, maybe by second
      % order differential?)
% after the max points found out, need check the repeat, whether it is
```

```
% already in the existing design points contained in conditions
% conditions had better to be index matrix (ix1), i=n experiments (it is easy to deal
with, the real conditions can be reversed from designspace)
% if max point repeat, which happens frequently in the small design space
% with large loop number, change to random search. if no repeat, output num
% to nexpoigreedy with maxvalue.
repn=reptest(conditions,num); % repeat check function, if repeat return 1, if not,
return 0    % ****** this need program in reptest.m file ***********
nexrandom=0; % initial value
if repn==0
      nexpoigreedy=num; % give index value for conditions
      maxvalue=tempmax; % give model value for evalutation. i.e. model accuracy,
error=realvalue-maxvalue
      nexrandom=0; % 0 means no random search.
else
      nexpoigreedy=num; % still give index value for conditions
      maxvalue=tempmax; % still give model value for evalutation. i.e. model
accuracy, error=realvalue-maxvalue
      j=0; % used in following while loop in the case of all points are used.
      while repn==1 % loop for random search until no repeat point comes out or no
candidate points (very rare situation), in this case, give nexrandom value -1
            if j==a      % which means all points used
                nexrandom=-1;break; % while loop break with three value returned
            else
            nexrandom=randi(a,1); % random choose an integer from designspace
conditions
            repn=reptest(conditions,uint8(nexrandom)); % chech whether new point is
also repeated
            j=j+1; % add one for each loop
            end
      end
end
end
```

7. nextsimplex.m

```
% This code used for simplex algorithm incorporating in the new combining
% algorithm. Dealing with internal loop and external loop together. The
% main loop is the same as v2 version
% input will be changed from v2 to include all the information in the case
% that in the following code requires full data assess except model
% structure and coefficients(that is not necessary in simplex). The input
% will be all existing conditions in index matrix, with all results in 88x2
% matrix, then can be transformed to simplex using objective surface matrix
```

```
% 88x1 with no existing conditions at value zero. ctrlnum and spenum were
% used to control the switch in the loop at different cases. The search
% design space was also required. extra informatoin including last run of
% three points simplex and reflection points in the comparsion, these
% information need updated each run both in internal loop or external
% loop. the return value will be the index number of next point, when break
% happen, the index value will be negative for main algorithm to
% recognize,the last run of three point with reflection
% index,ctrlnum,spenum (if applicable),

function
[nextsimpoint,reflection,reflectindex,ctrlnum,spenum,candidatepoints]=nextsimplex(
candidatepoints,conditions,designspace,ctrlnum,spenum,reflection,reflectindex,objva
lue)
% candidatepoints will be three candidate points for this run simplex in
% index format, data sorted; conditions will be all existing conditions in index
format;
% excondition existing conditions in normalized nx2 matrix; exdata existing
% results in nx2 matrix corresponding to excondition and conditions;
% designspace 88x2 matrix in normalized value; ctrlnum control value for
% loop; spenum special case control number, lastrun will store previous one
% run simplex three points from candidatepoints, reflection will be store
% the reflect point when the ctrlnum=1 exists,objvalue will be 88x2
% matrix contain all existing conditions objective value
repn=1; % initial value for loop, =0 means loop close and output value or end whole
simplex search due to e.g. linearity
repnc=1; % initial value for NW exchange to prevent dead loop, when repnc>7
break;
expanobj=0;
reflectobj=0; % two initial value for the loop following
for i=1:1:3
    indexnum=candidatepoints(i); % point index, first always the best point, 2nd
the next to best, the third worst point
    candidobj(i)=objvalue(indexnum); % get relevant obj value for later
comparsion
    point(i,:)=designspace(indexnum,:); % get real conditions from corresponding
designspace
end % after this loop point matrix store real conditions for three candidate points
while repn==1 % when repn==0, means no repeat or break by following break
sentence
% according to ctrlnum do simplex algorithm
outrange=1; % inital value for outrange loop, =0 means not out range
    while outrange==1
        if ctrlnum==1 % which means all points are new and start normal
```

```
reflective point search
                temppoi=point(1,:)+point(2,:)-point(3,:); % give tempoi value since
need check range and steps, after that convert to index and transfer value to
nextsimpoint
            elseif ctrlnum==2 % which means expansion step
                temppoi=(3.*point(1,:)+3.*point(2,:)-4.*point(3,:))./2;
            elseif ctrlnum==3 % external contraction
                temppoi=(3.*point(1,:)+3.*point(2,:)-2.*point(3,:))./4;
            elseif ctrlnum==4 % internal contraction
                temppoi=(point(1,:)+point(2,:)+2.*point(3,:))./4;
            elseif ctrlnum==5 % special case for repeat and all 1,2,3,4 fail
                temppoi=point(1,:)+point(2,:)-point(3,:); % in this case, the
point(3,:) will be N
                ctrlnum=1; % use it as new start but with spenum be the mark
            end % after this if-end, the simplex point for next given, following
check three things 1. out of range? 2. choose the nearest integer step point. e.g. pH
4.15 to 4.0 3.check the repeat in conditions
            % first convert to real pH and salt condition
            % specifical case pH 4.5-8.0, step 0.5; salt 0-3 step 0.3
            realx=3*temppoi(1); % x is salt
            realy=3.5*temppoi(2)+4.5; % y is pH
        if realx>3|realx<0|realy>8|realy<4.5
            if ctrlnum==1
                ctrlnum=3; % change to contraction
                reflectobj=-1; % when out of range, objective value of reflection
point will be a very small number
            elseif ctrlnum==3
                    ctrlnum=4; % change to internal contraction
            elseif ctrlnum==2
                    temppoi=reflection; % reflection will be in 1x2 matrix same as
one row in designspace
                    realx=3*temppoi(1); % x is salt
                    realy=3.5*temppoi(2)+4.5; % y is pH
                    realconx=round(realx*10)/10;
                    realcony=round(realy*2)/2; % round to the nearest floating
number for salt is to 0.1,for pH is to 0.5
                    num=(realcony-4.5)/0.5*11+(realconx-0)/0.3+1; % floating
number for chozen point
                    num=int8(num); % change to integer for matrix index use
                    expanobj=-1; % give -1 means out of range
                    ctrlnum=1; % once expansion out of range use R as new
points
                    outrange=0;
            end
```

```
            else
                % ****all following for round to the nearest point and convert to
index**********
                realconx=round(realx*10)/10;
                realcony=round(realy*2)/2; % round to the nearest floating number for
salt is to 0.1,for pH is to 0.5
                num=(realcony-4.5)/0.5*11+(realconx-0)/0.3+1; % floating number for
chozen point
                num=int8(num); % change to integer for matrix index use
                if ctrlnum==1;
                    reflection=designspace(num,:); % if ctrlnum=2,3,4, reflection and
reflectindex will be the same as input, while ctrlnum=1,no reflection input, except
ctrlnu=2 and outrange or special case
                    reflectindex=num;
                end % store reflection when new reflection given
                outrange=0;
                % **** above finish index change
            end
        end % above finish outrange loop with a index point in the boundary given
    % following code for repeat check,first should check whether repeat with
    % candidatepoints, which means simplex is too small or maybe at the end
    % of iteration due to gather at one local or global point
    repn=reptest(candidatepoints,num); % check whether repeat with candidate points
    if repn==1 % repeat
        BN=point(1,:)-point(2,:); % vector BN
        BW=point(1,:)-point(3,:); % vector BW
        angNBW=acos(dot(BN,BW)/norm(BN)/norm(BW)); % calculate angle
        area=0.5*sin(angNBW)*norm(BN)*norm(BW); % triangle NBW area
        if area<0.5/7*0.1
            nextsimpoint=-1;
            break; % -1 means too small area, either simplex go to end or need new
point from outsider
        else
            if repnc>7 % in case dead loop repeat the NW exchange 7 times, break
                nextsimpoint=-1; % same as small area
                break;
            else exchangepoint=point(2,:); % exchange the NW
                point(2,:)=point(3,:);
                point(3,:)=exchangepoint;
                ctrlnum=1; % new loop with BWN points
                repnc=repnc+1; % each time count one more
            end
        end
    else repn=reptest(conditions,num); % chech whether repeat with existing points
```

```
        if repn==1
            % divert to five consequence
            previousdata=objvalue(num); % get the repeat objvalue for comparsion
            if spenum==1 % spenum=1 means special case
                ctrlnum=5; % special case, no matter what ctrlnum go to 5
                spenum=0; % del spenum mark, if no repeat, mark remains to
output
            end
            if ctrlnum==1 % repeat when ctrlnum=1 without special case, but may
coma from 2 with expanobj=-1
                if expanobj==-1 % deal with expanobj=-1 case
                    point(3,:)=point(2,:);
                    point(2,:)=point(1,:);
                    point(1,:)=designspace(num,:);
                    expanobj=0; % reset value once done
                    ctrlnum=1; % new run with RBN
                else % real ctrlnum=1 case, four following consequences
                    if previousdata>candidobj(1)
                        ctrlnum=2; % into expansion mode,same points, same
reflectindex
                    elseif
previousdata<candidobj(1)&&previousdata>candidobj(2)
                            point(2,:)=designspace(num,:);
                            ctrlnum=1; % into new loop with BRN
                    elseif
previousdata<candidobj(2)&&previousdata>candidobj(3)
                            ctrlnum=3; % contraction, same points, same
reflectindex
                    else ctrlnum=4; % internal contraction, same points
                    end
                end
            elseif ctrlnum==2 % repeat when ctrlnum=2,outrange problem solved in
1 special case
                if previousdata>objvalue(reflectindex)
                    ctrlnum=1;
                    point(3,:)=point(2,:);
                    point(2,:)=point(1,:);
                    point(1,:)=designspace(num,:); % new loop EBN
                else ctrlnum=1;
                    point(3,:)=point(2,:);
                    point(2,:)=point(1,:);
                    point(1,:)=designspace(reflectindex,:); % new loop RBN
                end
            elseif ctrlnum==3 % repeat when ctrlnum=3,special case when
```

```
reflectobj=-1, R point out of range
                    if reflectobj==-1 % defined by line 59
                        reflectobj=0; % reset once done
                        ctrlnum=1; % new BNCr loop sort by next line

point=sortpoint(candidatepoints(1),candidatepoints(2),num,objvalue,designspace);
                    else % real ctrlnum=3 case with R in the range
                        if previousdata>objvalue(reflectindex)
                            ctrlnum=1; % new loop with BNCr sorted by next line

point=sortpoint(candidatepoints(1),candidatepoints(2),num,objvalue,designspace); %
**********code for sort three points
                        else ctrlnum=5; % entry into special case
                                point(3,:)=point(2,:);
                                point(2,:)=designspace(num,:); % kick out W point, N
become the lowest
                                spenum=1; % mark
                        end
                    end
                elseif ctrlnum==4 % repeat when ctrlnum=4, no special case
                    if previousdata>candidobj(3)
                        ctrlnum=1; % new loop with BNCw sort by next line

point=sortpoint(candidatepoints(1),candidatepoints(2),num,objvalue,designspace);
                    else ctrlnum=5; % entry into speical case
                        point(3,:)=point(2,:);
                        point(2,:)=designspace(num,:); % kick out W point, N become
the lowest
                        spenum=1; % mark
                    end
                elseif ctrlnum==5 % when mark exists, all 1,2,3,4 become ctrlnum 5,
with new loop BCr or BCw with N reflective point sorted
                    ctrlnum=1;

point=sortpoint(candidatepoints(1),candidatepoints(2),num,objvalue,designspace);
                end
            else repnc=1; % no repeat, repnc get initial value
                 nextsimpoint=num; % when repn=0,num the next simplex point index
            end
    end
end
% the initial candidatepoints may be change due to existing points, the final step is to
find the current candidatepoints and output
for ii=1:1:3
```

```
        for jj=1:1:88
            if designspace(jj,1)==point(ii,1)&&designspace(jj,2)==point(ii,2)
                candidatepoints(ii)=jj;
            end
        end
end
end
```

8. validcalculate.m

```
% this code is used to calculate the derivation between model based value
% and real value in a 9 DoE validation mode (points fixed as modelling validation
parts)

function [sfab,stot,sobj]=validcalculate(fab,tot,objvalue)
% for 8X11 matrix, corresponding position is (1,1) (1,6) (1,11)
%                                            (4,1) (4,6) (4,11)
%                                            (7,1) (7,6) (7,11)
validmatrix=zeros(8,11);
validmatrix(1,1)=1.021160116;
validmatrix(1,6)=0.764229739;
validmatrix(1,11)=1.00447783;
validmatrix(4,1)=0.640376318;
validmatrix(4,6)=0.546911361;
validmatrix(4,11)=0.997322858;
validmatrix(7,1)=0.942834051;
validmatrix(7,6)=0.575906734;
validmatrix(7,11)=1.023508504;
validfab=zeros(8,11);
validfab(1,1)=0.107116053;
validfab(1,6)=0.433502454;
validfab(1,11)=0;
validfab(4,1)=0.593506604;
validfab(4,6)=0.452633809;
validfab(4,11)=0;
validfab(7,1)=0.831231493;
validfab(7,6)=0.485283396;
validfab(7,11)=0;
validtot=zeros(8,11);
validtot(1,1)=0.1931851;
validtot(1,6)=0.405882959;
validtot(1,11)=0.148305348;
validtot(4,1)=0.883690382;
validtot(4,6)=0.653060399;
```

```
validtot(4,11)=0.139566425;
validtot(7,1)=0.817147341;
validtot(7,6)=0.674327245;
validtot(7,11)=0.170707744;
sfab=0;
stot=0;
sobj=0;
for i=1:1:8
      for j=1:1:11
            if validmatrix(i,j)==0
            else sfab=sfab+(validfab(i,j)-fab(i,j))^2;
                  stot=stot+(validtot(i,j)-tot(i,j))^2;
                  sobj=sobj+(validmatrix(i,j)-objvalue(i,j))^2;
            end
      end
end
sfab=sfab/9;
stot=stot/9;
sobj=sobj/9;
end
```

9. simpointselect.m

```
% This code is used to select the next run of candidatepoints for simplex
 % in the main algorithm.

 function
[candidatepoints,ctrlnum,spenum,reflection,reflectindex,loopctronum]=simpointsele
ct(candidatepoints,conditions,addpoint,realresults,designspace,ctrlnum,spenum,refle
ction,reflectindex,nextsimpoint,objvalue,loopctronum)
 ctn=1; % initial value for following loop
      if nextsimpoint==-1
            fprintf('Simplex loop terminated due to small area, probably objective
achieved\n');
            fprintf('Do you want to continue a new run simplex? Y=1,N=0\n');
            ctn=input('Input Y/N value\n');
            if ctn==0
                loopctronum=0;
            elseif ctn==1 % start a new set of simplex, in this case use random choose
from existing points in conditions matrix
                  loopctronum=1; % loop continue
                  [a5,b5]=size(conditions);
                  rechnum=1; % in case random choose all repeat, rechnum=0 means
successfully choose the new set, so loop stop
```

```
                while rechnum==1
                newcandid=randi(b5,9,3); % random choose 9 group of three points

[candidatepoints,rechnum]=randomstart(newcandid,designspace,objvalue,conditions
); % ******the file select three max area points sorted to start new set simplex from
existing points
                end
                ctrlnum=1;
                reflection=0;
                reflectindex=[0,0];
                spenum=0; % initial value for simplex
            end
        else bobj=objvalue(candidatepoints(1));
            nobj=objvalue(candidatepoints(2));
            wobj=objvalue(candidatepoints(3));
            if reflectindex==0
                reflpoiobj=0; % when reflectindex does not exists, give its objvalue
zero;
            else reflpoiobj=objvalue(reflectindex);
            end % get all the value for the following code
            % the real objective value at simplex point
            % first crossover then if no crossover simplex orginal loop
            simindex=nextsimpoint;

resuvalue=objectivefun(realresults(simindex,1),realresults(simindex,2),designspace(
simindex,:));
            % get other crossover points with objvalue and index.
            % a3 is the number of points added, in the case of nextsimpoint
            % does not equals to -1,theoretically there are always four points, except
all points run out, in this situation, algorithm will wrong
            % notice, when repeat, a3 will be less than four but always no
            % less than one. special case: all four points are the same one
            % including one given by simplex
            [a6,b6]=size(addpoint); % get the number of points add
            for i=1:1:b6
                addobj(i)=objvalue(addpoint(i));
            end
            cronum=1; % initial value for points crossover, 1 means use old run
simplex points, 0 means new simplex with extra points introduced by other methods
            if b6==1
                cronum=1;
            elseif b6==2
                if addobj(1)>nobj
                    cronum=0;
```

```
            end
        elseif b6==3
            if addobj(1)>nobj|addobj(2)>nobj
                cronum=0;
            end
        elseif b6==4
            if addobj(1)>nobj|addobj(2)>nobj|addobj(3)>nobj
                cronum=0;
            end
        end % total four situations % in this case, once the extra one point
objvalue is larger than the second simplex point, new simplex loop starts, otherwise,
continue old simplex loop
        if cronum==1
            if spenum==1
                spenum=0;
                ctrlnum=1;
                reflectindex=0;
                reflection=[0,0];

newchanmatrix=[candidatepoints(1),objvalue(candidatepoints(1));candidatepoints(2)
,objvalue(candidatepoints(2));simindex,resuvalue];
                exchange1=[0,0];
                for i=1:1:2
                    for j=1:1:2
                        if newchanmatrix(j,2)>newchanmatrix(j+1,2)
                        else exchange1=newchanmatrix(j+1,:);
                            newchanmatrix(j+1,:)=newchanmatrix(j,:);
                            newchanmatrix(j,:)=exchange1;
                        end
                    end
                end % sort by buble algorithem with index also changed
according to the objvalue

candidatepoints=[newchanmatrix(1,1),newchanmatrix(2,1),newchanmatrix(3,1)]; %
give new candidatepoints by sort BR'Cr or Cw
            else
                if ctrlnum==1
                    if resuvalue>bobj
                        ctrlnum=2;
                    elseif resuvalue<bobj&&resuvalue>nobj
                        candidatepoints(3)=candidatepoints(2);
                        candidatepoints(2)=simindex;
                        ctrlnum=1;
                    elseif resuvalue<nobj&&resuvalue>wobj
```

```
                    ctrlnum=3;
                elseif resuvalue<wobj
                    ctrlnum=4;
                end
            elseif ctrlnum==2
                if resuvalue>reflpoiobj
                    candidatepoints(3)=candidatepoints(2);
                    candidatepoints(2)=candidatepoints(1);
                    candidatepoints(1)=simindex;
                    ctrlnum=1;
                else candidatepoints(3)=candidatepoints(2);
                        candidatepoints(2)=candidatepoints(1);
                        candidatepoints(1)=reflectindex;
                        ctrlnum=1;
                end
            elseif ctrlnum==3
                if resuvalue>reflpoiobj

candidatepoints=sortpoint2(candidatepoints(1),candidatepoints(2),simindex,objvalue
); %********new function file a little change from sortpoint
                    ctrlnum=1;
                else candidatepoints(3)=candidatepoints(2);
                        candidatepoints(2)=simindex;
                        ctrlnum=5;
                        spenum=1;
                end
            elseif ctrlnum==4
                if resuvalue>wobj

candidatepoints=sortpoint2(candidatepoints(1),candidatepoints(2),simindex,objvalue
); %********new function file a little change from sortpoint
                    ctrlnum=1;
                else candidatepoints(3)=candidatepoints(2);
                        candidatepoints(2)=simindex;
                        ctrlnum=5;
                        spenum=1;
                end
            end % ctrlnum=5, spenum must be 1, so do not need this
conditions in this if phrase
        end
    else % start crossover

newsortmatrix=[candidatepoints(1),objvalue(candidatepoints(1));candidatepoints(2),
objvalue(candidatepoints(2));candidatepoints(3),objvalue(candidatepoints(3))]; %
```

```
matrix for sorting with first index, second objvalue
                for i9=1:1:b6
                    newsortmatrix(3+b6,:)=[addpoint(b6),objvalue(addpoint(b6))];
                end % add all points together
                exchange2=[0,0]; % initial value for sorting
                for i=1:1:2+b6
                    for j=1:1:2+b6
                        if newsortmatrix(j,2)>newsortmatrix(j+1,2)
                        else exchange2=newsortmatrix(j+1,:);
                                newsortmatrix(j+1,:)=newsortmatrix(j,:);
                                newsortmatrix(j,:)=exchange2;
                        end
                    end
                end % with index sorted, largest one in the first row

candidatepoints=[newsortmatrix(1,1),newsortmatrix(2,1),newsortmatrix(3,1)]; % get
index for the first three largest points
                reflection=[0,0];
                reflectindex=0;
                spenum=0;
                ctrlnum=1; % reset to initial parameters for crossover simplex loop
            end
        end
    end
```

10. dmatrix.m

```
% This code was used to make dmatrix according to coeff and realcon based
% on specific model. In this case, the model used is the developed
% precipitation model with five parameters.
% Exchangable algorithm due to the model, so this is specific for
% precipitation model

function dmatx=dmatrix(coeff,realcon) % dvalue is a matrix nx5, n is the same as the
rows of realcon, and 5 comes from the model paremeters
x=realcon(:,1); % separate realcon to x matrix, in this case, salt concentration, in
normalized value 0-1
y=realcon(:,2); % separate to y matrix, in this case, pH
[a,b]=size(realcon); % get the size of realcon, a equals to n
for i=1:1:a

minielement=1./(x(i)+coeff(5)).*exp(coeff(1).*x(i)+coeff(2).*y(i)+coeff(3).*y(i).*y(
i)+coeff(4)); % calculate minielement for dmatrix
    melement=-1.*((1+minielement).^(-2)).*minielement; % 2nd minielement for
```

```
dmatrix
        dmatx(i,1)=melement.*x(i); % column one
        dmatx(i,2)=melement.*y(i); % column two
        dmatx(i,3)=melement.*y(i).*y(i); % column three
        dmatx(i,4)=melement; % column four
        dmatx(i,5)=melement.*(-1./(coeff(5)+x(i))); % column five
end
end
```

11. reptest.m

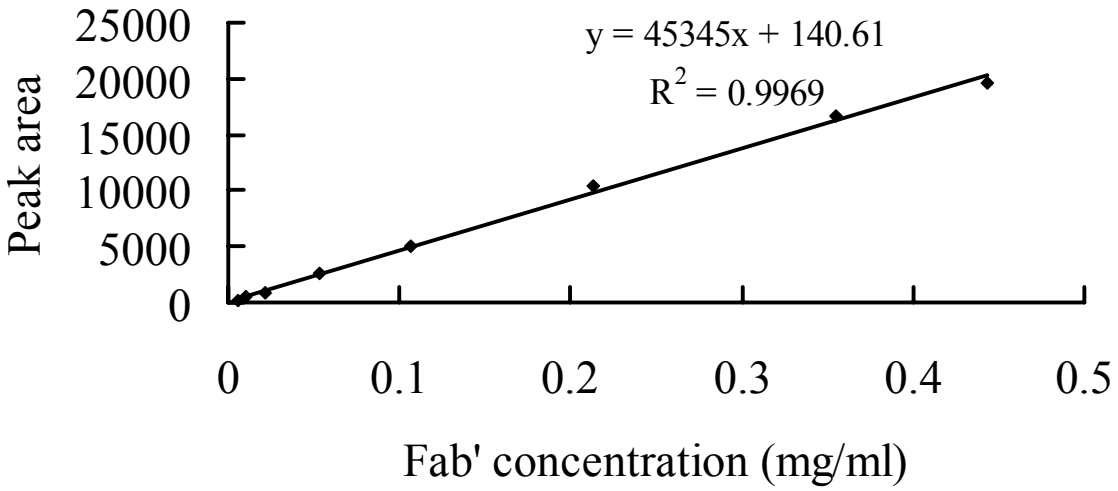
```
% this function used for judge whether the point is repeated or not in the
% existing conditons, return value 1 means repeat, 0 means no repeat

function repn=reptest(conditions,num)
% conditions will be ix1 matrix, num will be the index num
[n,m]=size(conditions); % get i value
for j=1:1:m
        if uint8(conditions(j))==uint8(num)
                repn=1;break;
        else repn=0;
        end
end
end
```

12. ppmodel.m

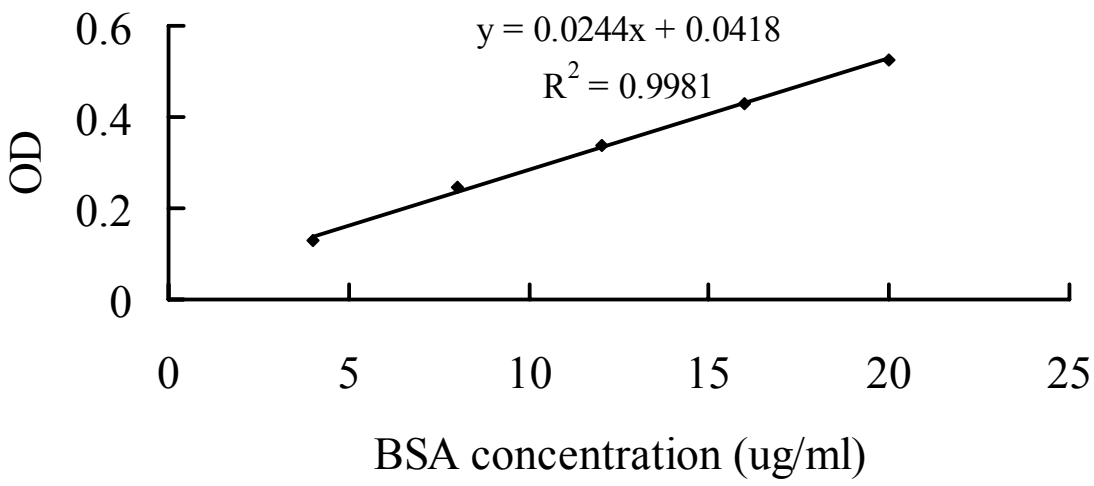
```
% this is the model file which contains model structure
% this file is used to calculate model value and also for parameter
% regression

function modelvalue=ppmodel(coeff,conditions) % five coefficients in coeff, 1x5
matrix
x=conditions(:,1); % x is salt concentration in normalized value
y=conditions(:,2); % y is pH in normalized value
modelvalue=1./(1+(1./(coeff(5)+x)).*exp(coeff(1).*x+coeff(2).*y+coeff(3).*y.*y+co
eff(4)));
end
```

Appendix 2. Calibration curves

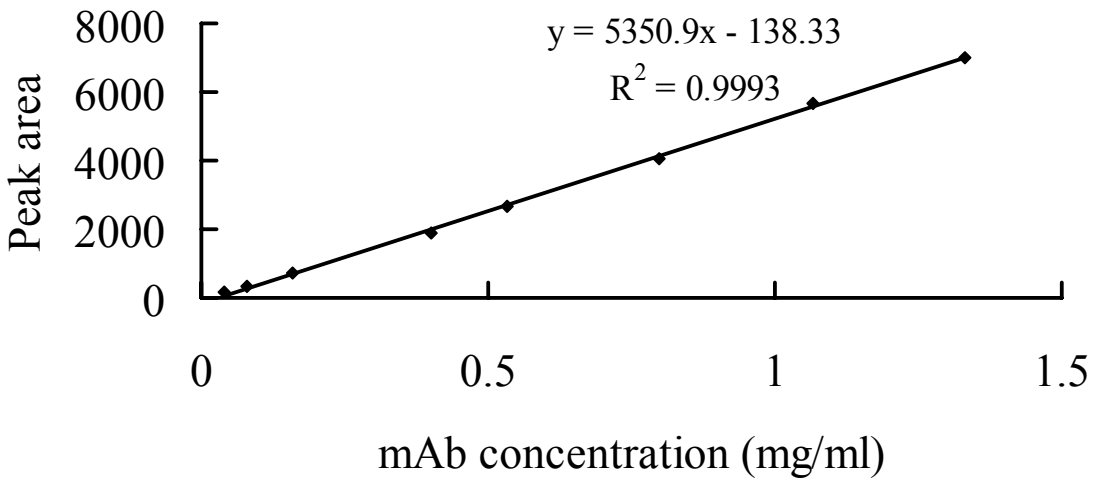
1. Fab' HPLC calibration curve



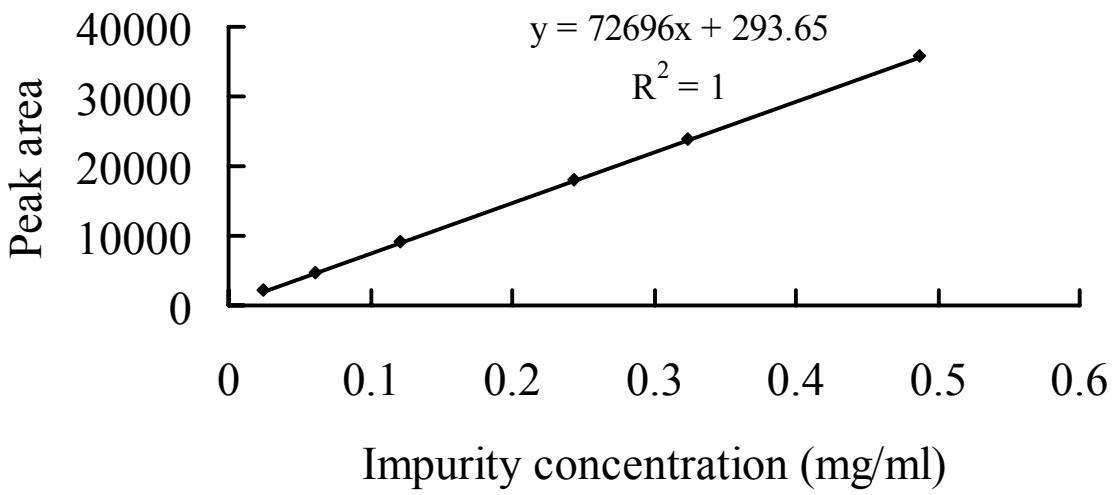
2. Bradford calibration curve



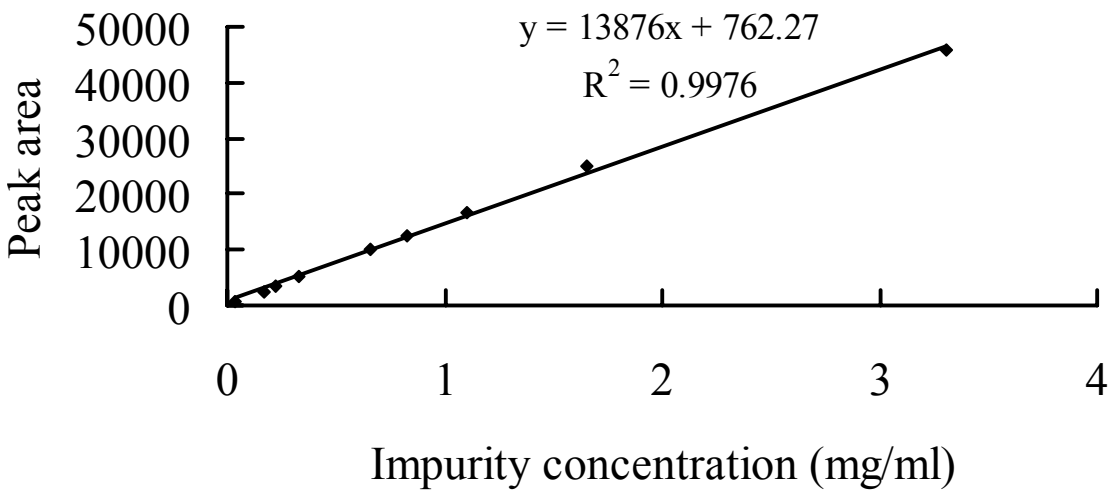
3. mAb HPLC calibration curve



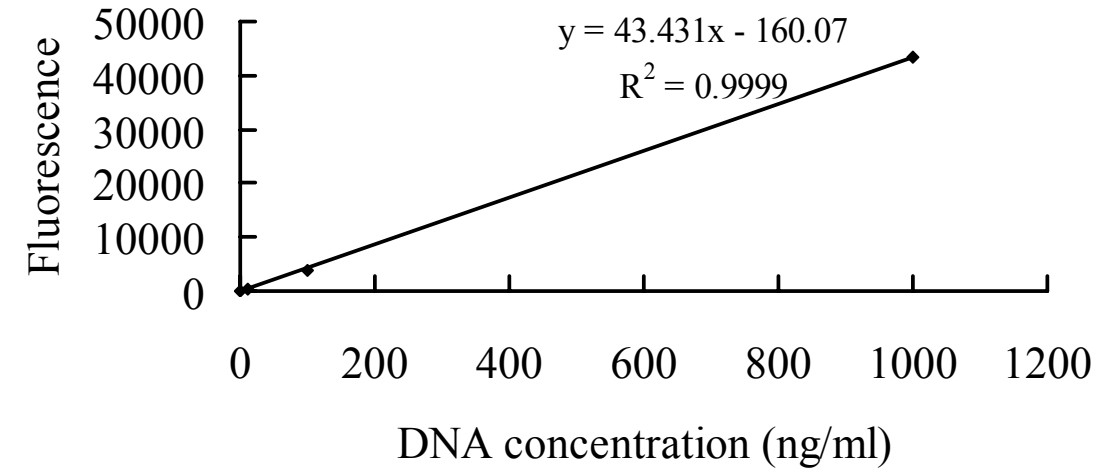
4. Fab’ solution impurity HPLC calibration curve



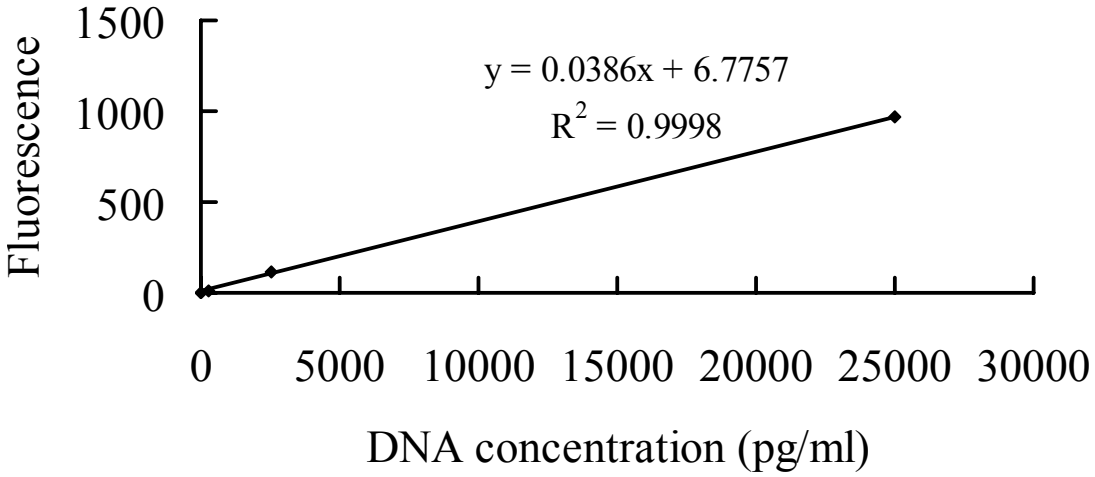
5. mAb solution impurity HPLC calibration curve



6. Picogreen DNA calibration curve

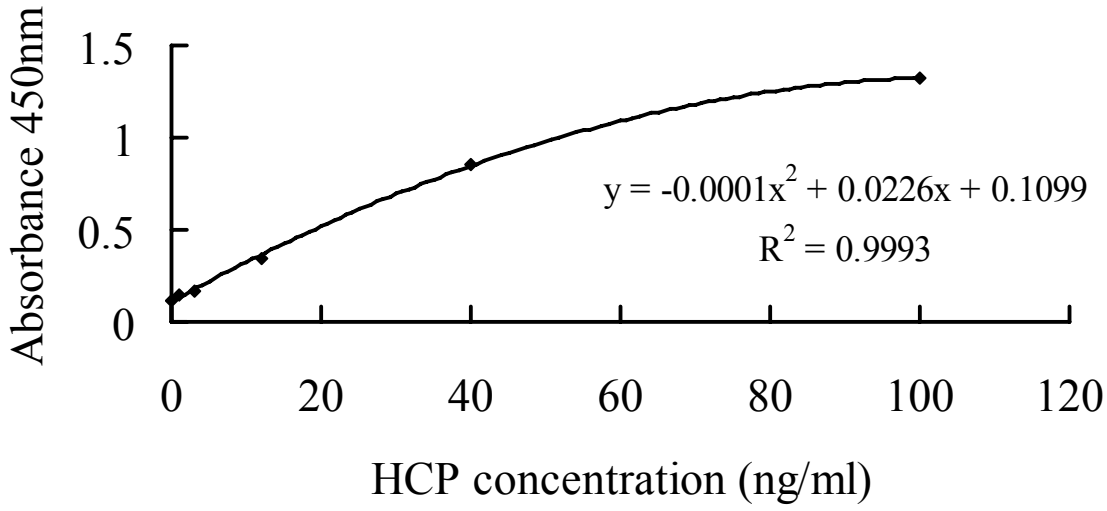


High range calibration curve



Low range calibration curve

7. CHO host cell protein calibration curve



The CHO HCP Elisa protocol provided by Cygnus Technology recommended the calibration curve was regressed by polynomial fitting rather than linear regression.

Appendix 3. Raw data for Fab' and mAb validation

1. Fab' validation data

Validation DoE		Real value		
Salt (M)	pH	Fab' (mg/ml)	Impurity (mg/ml)	Objective value
0	4.5	0.181	1.228	1.021
1.5	4.5	0.067	0.465	0.764
3	4.5	0.000	0.272	1.004
0	6	0.188	1.601	0.640
1.5	6	0.059	0.658	0.547
3	6	0.000	0.197	0.997
0	7.5	0.185	1.307	0.943
1.5	7.5	0.094	1.193	0.576
3	7.5	0.000	0.301	1.024
		1st run		
Salt (M)	pH	Fab' (mg/ml)	Impurity (mg/ml)	Objective value
0	4.5	0.224	1.438	1.109
1.5	4.5	0.000	0.592	1.271
3	4.5	0.000	0.084	0.931
0	6	0.224	1.553	1.057
1.5	6	0.000	0.861	1.566
3	6	0.000	0.154	0.960
0	7.5	0.224	1.625	1.028
1.5	7.5	0.000	1.152	2.211
3	7.5	0.000	0.288	1.023
		2nd run		
Salt (M)	pH	Fab' (mg/ml)	Impurity (mg/ml)	Objective value
0	4.5	0.217	1.497	1.047
1.5	4.5	0.097	0.633	0.824
3	4.5	0.004	0.078	0.909
0	6	0.214	1.589	0.992
1.5	6	0.075	0.909	0.460
3	6	0.003	0.144	0.942
0	7.5	0.218	1.635	0.994
1.5	7.5	0.102	1.136	0.540
3	7.5	0.005	0.237	0.975
		3rd run		
Salt (M)	pH	Fab' (mg/ml)	Impurity (mg/ml)	Objective value
0	4.5	0.211	1.505	1.015
1.5	4.5	0.097	0.690	0.764
3	4.5	0.004	0.078	0.910

0	6	0.213	1.596	0.988
1.5	6	0.121	0.928	0.762
3	6	0.005	0.147	0.935
0	7.5	0.216	1.641	0.982
1.5	7.5	0.135	1.165	0.717
3	7.5	0.006	0.249	0.973
			4th run	
Salt (M)	pH	Fab' (mg/ml)	Impurity (mg/ml)	Objective value
0	4.5	0.024	0.328	1.021
1.5	4.5	0.097	0.690	0.764
3	4.5	0.004	0.062	0.905
0	6	0.030	0.643	1.182
1.5	6	0.103	1.008	0.600
3	6	0.005	0.153	0.935
0	7.5	0.033	0.797	1.307
1.5	7.5	0.109	1.154	0.573
3	7.5	0.006	0.214	0.959
			5th run	
Salt (M)	pH	Fab' (mg/ml)	Impurity (mg/ml)	Objective value
0	4.5	0.024	0.328	1.021
1.5	4.5	0.097	0.690	0.764
3	4.5	0.001	0.039	0.908
0	6	0.143	1.194	0.799
1.5	6	0.114	1.067	0.638
3	6	0.004	0.151	0.940
0	7.5	0.186	1.389	0.943
1.5	7.5	0.129	1.218	0.663
3	7.5	0.005	0.216	0.963

2. mAb validation data by ammonium sulphate precipitation

Validation DoE			Real value	
Salt (M)	pH	mAb (mg/ml)	Impurity (mg/ml)	Objective value
0.6	5	0.378	0.943	0.478
1.4	5	0.358	0.945	0.417
2.2	5	0.004	0.805	0.627
0.6	6.5	0.392	0.984	0.490
1.4	6.5	0.340	0.897	0.402
2.2	6.5	0.005	0.761	0.620
0.6	8	0.396	0.997	0.494
1.4	8	0.203	0.927	0.224

2.2	8	0.000	0.807	0.645
			1st run	
Salt (M)	pH	mAb (mg/ml)	Impurity (mg/ml)	Objective value
0.6	5	0.413	1.014	0.511
1.4	5	0.155	0.919	0.160
2.2	5	0.000	0.750	0.621
0.6	6.5	0.426	1.024	0.526
1.4	6.5	0.202	0.934	0.222
2.2	6.5	0.000	0.772	0.630
0.6	8	0.435	1.033	0.535
1.4	8	0.251	0.948	0.284
2.2	8	0.000	0.793	0.639
			2nd run	
Salt (M)	pH	mAb (mg/ml)	Impurity (mg/ml)	Objective value
0.6	5	0.427	1.010	0.528
1.4	5	0.119	0.909	0.585
2.2	5	0.000	0.746	0.620
0.6	6.5	0.432	1.021	0.533
1.4	6.5	0.140	0.928	0.139
2.2	6.5	0.000	0.772	0.630
0.6	8	0.436	1.031	0.536
1.4	8	0.163	0.945	0.169
2.2	8	0.000	0.797	0.640
			3rd run	
Salt (M)	pH	mAb (mg/ml)	Impurity (mg/ml)	Objective value
0.6	5	0.438	1.010	0.542
1.4	5	0.202	0.910	0.224
2.2	5	0.000	0.763	0.627
0.6	6.5	0.437	1.021	0.538
1.4	6.5	0.187	0.927	0.202
2.2	6.5	0.000	0.777	0.632
0.6	8	0.434	1.030	0.535
1.4	8	0.172	0.943	0.181
2.2	8	0.000	0.800	0.642

3. mAb validation data by PEG precipitation

Validation DoE			Real value	
PEG (%)	pH	mAb	Impurity (mg/ml)	Objective value

		(mg/ml)		
0	5	0.429	1.038	0.481
10	5	0.210	0.867	0.217
20	5	0.000	0.550	0.569
0	6.5	0.426	1.017	0.481
10	6.5	0.350	1.055	0.349
20	6.5	0.003	0.445	0.545
0	8	0.435	1.094	0.481
10	8	0.295	1.060	0.291
20	8	0.000	0.638	0.589
			1st run	
PEG (%)	pH	mAb (mg/ml)	Impurity (mg/ml)	Objective value
0	5	0.463	1.071	0.512
10	5	0.279	0.875	0.295
20	5	0.005	0.482	0.550
0	6.5	0.467	1.090	0.513
10	6.5	0.286	0.911	0.297
20	6.5	0.005	0.524	0.558
0	8	0.470	1.108	0.514
10	8	0.292	0.944	0.300
20	8	0.006	0.567	0.567
			2nd run	
PEG (%)	pH	mAb (mg/ml)	Impurity (mg/ml)	Objective value
0	5	0.463	1.106	0.507
10	5	0.279	0.869	0.295
20	5	0.000	0.466	0.552
0	6.5	0.467	1.118	0.510
10	6.5	0.286	0.895	0.299
20	6.5	0.000	0.496	0.557
0	8	0.470	1.129	0.512
10	8	0.295	1.060	0.291
20	8	0.000	0.527	0.564
			3rd run	
PEG (%)	pH	mAb (mg/ml)	Impurity (mg/ml)	Objective value
0	5	0.451	1.103	0.496
10	5	0.277	0.894	0.290
20	5	0.000	0.464	0.551
0	6.5	0.461	1.115	0.504
10	6.5	0.293	0.919	0.304
20	6.5	0.000	0.495	0.557
0	8	0.470	1.127	0.512

10	8	0.295	1.060	0.291
20	8	0.000	0.526	0.563
			4th run	
PEG (%)	pH	mAb (mg/ml)	Impurity (mg/ml)	Objective value
0	5	0.430	1.100	0.476
10	5	0.210	0.867	0.217
20	5	0.000	0.550	0.569
0	6.5	0.454	1.112	0.497
10	6.5	0.298	0.927	0.309
20	6.5	0.001	0.516	0.560
0	8	0.474	1.123	0.516
10	8	0.295	1.060	0.291
20	8	0.001	0.546	0.566