Development of aqueous two-phase separations by combining high-throughput screening and process modelling

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

by

Nehal Patel

Friday 21st July 2017

The Advanced Centre for Biochemical Engineering

Department of Biochemical Engineering

University College London

Gower Street, London, WC1E 6BT, UK

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

Abstract

Separation based on aqueous two-phase extraction (ATPE) is a promising downstream separation technology for the production of biological products. The advantages of using aqueous two-phase systems include but are not limited to easy scalability, ease of continuous operation and a favourable environment for biological compounds. One of the main challenges associated with aqueous two-phase systems is process development. This is in part due to the many factors which influence the separation of biological materials in such systems such as polymer and salt type, pH and charge. The large number of factors to consider makes the development of aqueous two-phase systems challenging due to the need to find a robust and efficient separation in a large experimental space. This work addresses this issue by considering the use of dynamic process models and high-throughput experimentation for the development of aqueous two-phase extraction processes for biological products.

The use of a dynamic equilibrium stage process model to simulate aqueous two-phase extraction is considered in Chapter 3. The process model is capable of simulating various modes of operation; and both multi-cycle batch and continuous counter-current modes of operation are considered. The capabilities of the model are demonstrated using a case study separation of enzyme α-amylase from impurities in a PEG 4,000-phosphate aqueous two-phase system containing NaCl. The dynamic model allowed investigation into the impact of upstream process variability on a continuous counter-current extraction process.

The development of aqueous two-phase systems requires detailed knowledge of the phase diagram. In Chapter 4, PEG 4,000-citrate aqueous two-phase system phase diagrams are determined using a combination of high-throughput screening and lab scale experiments. This involved the development of a systematic two-stage screening approach to determine the binodial curve location to a high accuracy using ~50% of the experimental resources that a single high-resolution screen would use. In addition, a novel method was developed to quantify uncertainty in the phase diagram due to the binodial curve location and tie-line fitting. The characterised phase diagrams were then used to estimate thermodynamic interaction parameters which are used in process models to describe phase equilibria.

In Chapter 5, the simulation and high-throughput screening methods of Chapter 3 and 4 are combined to develop an aqueous two-phase extraction separation process. The

approach is demonstrated by separating enzyme α-amylase from myoglobin in a PEG 2,000-phosphate aqueous two-phase system containing 6wt% NaCl. High-throughput experimentation is used to determine partitioning behaviour of α-amylase and myoglobin at different tie-line lengths and phase ratios. The experimental partitioning and phase diagram data was then used to simulate a counter-current extraction process. The insights gained using the process model allowed for better decisions to be made regarding selection, control and operation of aqueous two-phase separation equipment. Therefore, the combined approach of using process modelling and high-throughput experimentation allowed for greater amounts of process understanding to be gained for aqueous two-phase systems using limited resources where there is a large experimental space to be navigated.

Impact Statement

Currently the major expense in the production of new biopharmaceuticals is the cost of downstream separation processes (Azevedo et al., 2009). The work in this thesis is centred on the use of aqueous two-phase extraction, a potentially cheaper alternative to expensive chromatographic techniques currently used in biopharmaceutical manufacturing.

Specifically, the work presented in this thesis looks at the potential of using a combined high-throughput experimental and modelling approach for the development of aqueous two-phase extraction separations for the purification of biopharmaceuticals. The motivation for using high-throughput experimentation is that greater amounts of information can be extracted from limited experimental resources. This is particularly important in early stage development of new drugs where the availability of material to purify is limited. The use of appropriate process models allows for exploration of separation systems beyond the scope of the high-throughput experimentation without the need to perform additional experiments. Such an approach provides greater process insights to researchers so that robust separation processes can be developed more efficiently.

The methods presented in this work will help researchers develop commercially relevant aqueous two-phase extraction processes. This would benefit patients and the general public as the cost of a new biopharmaceutical drug could be cheaper for two reasons: 1) The combined approach using experiments and modelling allows for research and development efforts to be conducted more efficiently and 2) The use of aqueous two-phase extraction could save money as the inherent process may be cheaper than using conventional separation techniques.

Acknowledgement

I would like to thank my UCL supervisors, Dan Bracewell and Eva Sorensen for their continuous guidance and support throughout this project. Eva, I have never met someone who is as passionate as you are in the personal development of their students, it is truly inspiring. I feel very fortunate to have had your support during this project. Dan, throughout this project you have given me the freedom and flexibility to try new things. At the same time, you have always been very approachable when help was needed which is something I will always appreciate.

I would like to thank BioMarin for making this project possible. The input of Eric, Dan and John in project meetings was invaluable. In addition I would like to thank all the BioMarin employees I met for their hospitality during a visit to their facilities.

I would like to thank the many staff members at UCL for all their help in various forms. Dhushy and Elaine, thank you for your continuous support in all lab related matters.

I would like to thank the many contributors to online computing websites such as Stack Overflow and Stack Exchange for taking the time to share their invaluable knowledge.

I would like to thank the many friends I have made at UCL for their endless support and motivation. The EngD would not have been as fun without all of you.

I would like to thank Amira. You made the last few years of the EngD the best they could possibly be.

Finally, I would like to thank my family for their never ending support throughout all of my education and encouraging me to pursue what I enjoy.

Table of contents

Table of figures

List of tables

Nomenclature

Symbol	Definition
$C_{i,eq,q}$	Concentration of component i at equilibrium in location q (kg m^{-3})
$C_{i,q,n}$	Concentration of component i in location q in stage n (kg m^{-3})
F	Faraday's constant (C $mole^{-1}$)
R	Gas constant (J $mole^{-1}$ k^{-1})
$M^0_{i,overall,n}$	Initial overall mass holdup of i in stage n
F_{bottom}	Mass flow of bottom stream (kg s^{-1})
F_{in}	Mass flow of feed stream (kg s^{-1})
$F_{q,n}$	Mass flow of stream q in stage n (kg s^{-1}) where q is stream 1-4
F_{top}	Mass flow of top stream (kg s^{-1})
x_i	Mass fraction of component i in the bottom phase (-)
w_i	Mass fraction of component i in the feed stream (-)
y_i	Mass fraction of component i in the top phase (-)
$x_{i,q,n}$	Mass fraction of i in q in stage n where q is either top, bottom, overall or stream 1-4
$M_{i,q,n}$	Mass holdup of i in q in stage n where q is overall, top or bottom (kg)
M_i	Mass of i where i is either top, bottom or total (g)
m_i	Molality of component i (mol kg^{-1})
V_i	Molar volume of water where i = 3 (m^3 $mole^{-1}$)
Z_p	Net protein charge (-)
K_0	Partition coefficient in the absence of electrochemical interactions (-)
K_i	Partition coefficient of component i (-)
$K_{impurities,n}$	Partition coefficient of impurities in stage n
$K_{\alpha-amylase,n}$	Partition coefficient of α-amylase in stage n
M_{std}	Standard deviation in mass measurement used to calculate density (g)
V_{std}	Standard deviation in volume used to calculate density (mL)
T	Temperature (K)
V_j	Volume of j where j is either top, bottom or total (mL)
$w_{i,q,n}$	Weight percent of component i in q in stage n where q is overall, top or bottom (wt%)
z	Standard score from a normal probability distribution
Greek Letters	
$\beta_{i,i}$	Binary interaction parameter between component i and component i (kg $mole^{-1}$)
$\beta_{i,j}$	Binary interaction parameter between component i and component j where i ≠ j (kg $mole^{-1}$)
μ_i	Chemical potential of component i (J $mole^{-1}$)
$\rho_{q,n}$	Density in q in stage n (kg m^{-3}) where q is top or bottom
ρ_j	Density of j where j is either top or bottom (g mL^{-1})
φ_t or φ_b	Electrostatic potential of top and bottom phases (V)
α	Fraction of feed stream that ends up in the top phase (-)
$\emptyset_n$	Phase ratio in stage n
μ_i^0	Standard chemical potential of i (J $mole^{-1}$)
σ	Standard deviation
Subscripts and superscripts	
i	Component index
q	Location index i.e. top, bottom, overall or stream 1-4
n	Stage index

j	j corresponds top, bottom or total when calculating masses and volumes when determining phase ratios using density measurements

Acronyms

A	Binodial curve fitting parameter (-)
B	Binodial curve fitting parameter (-)
C	Binodial curve fitting parameter (-)
NC	Number of components
NP	Number of phases i.e. top and bottom
NS	Number of stages
NT	Number of tie-lines
STD	Standard deviation

Chapter 1: Introduction

1.1 Biopharmaceutical market

The biopharmaceuticals market is currently increasing at a strong pace; and is the fastest growing sector within the pharmaceutical industry. In 2009 and 2014, the market value for biopharmaceuticals alone was $99 billion and $140 billion, respectively (Walsh, 2010, Walsh, 2014). This can mainly be associated with the success of recombinant therapeutic proteins such as monoclonal antibodies (mAbs). Recombinant therapeutic proteins, often referred to as biopharmaceuticals, are molecules which are produced in a genetically modified host cell. The genetic modification of the host cell allows for the production/expression of large complicated human proteins which are used as therapeutic treatments once purified. The main differences between small molecules and biopharmaceuticals are described in Table 1 - 1. The main benefits of using proteins compared to small molecules are: 1) proteins are highly specific therefore are potent, 2) high specificity results in fewer unwanted adverse effects, 3) proteins are less likely to induce harmful immune responses, and 4) proteins therapeutics can provide a substitute to missing proteins such as insulin (Leader et al., 2008).

Table 1 - 1: Main differences between small molecules and biopharmaceuticals adapted from Lim and Suh (2015).

Major differences between small molecules and biopharmaceuticals		
	Small molecules	*Biopharmaceuticals*
Size	100-10,000 Da	15,000-150,000 Da
Structure	Well-defined and homogenous	Complex and variable protein structure
Production	Chemical synthesis from well-defined starting chemicals	Produced in recombinant cell
Process predictability	Predictable	Less predictable
Analytical characterisation	Easy	Challenging
Typical stability	High	Low
Example therapeutics	Ibuprofen, Advair, Revlimid	Humulin, Humira, Enbrel

Although biopharmaceuticals are highly effective therapies, the treatment cost associated with biopharmaceuticals cannot be overlooked. McCamish and Woollett (2011) state that the average cost of a branded small molecule drug is approximately $1 per day whilst the average cost of a branded biopharmaceutical is $22 per day. However, for orphan diseases such as Gaucher's diesease the annual cost of Cerdelga®

treatment can be more than $300,000 per patient per year (Weisman, 2014). Such high costs limits the access developing nations have to cutting edge healthcare where the healthcare spend per capita is significantly less than in US and European markets (The World Bank, 2016). Generic unbranded small molecules have made access to healthcare significantly cheaper by undercutting off-patent therapies by producing them at a lower cost. The introduction of biosimilars, the biopharmaceutical equivalent of generics, will help address the issue of the high costs. However, for a new biosimilar to be successful it needs to have a significantly cheaper manufacturing process. In addition, the speed at which it enters the market must also be rapid to ensure that benefits such as the first mover advantage can be fully capitalised (Chatterji et al., 2011, Bourgoin and Nuskey, 2013, Fuhr et al., 2015).

1.2 Challenges in biopharmaceutical production

In many cases therapeutic proteins such as monoclonal antibodies are usually administered in large dosages (>100 mg dose^{-1}) (Shire et al., 2004). This large dosage requirement, coupled with the high demand for the drugs, has resulted in an increased pressure on pharmaceutical companies to keep up with patient demand; especially if they want to supply developing countries at an affordable cost. As a result, there has been an increasing amount of research & development (R&D) aimed at improving upstream efficiency (Azevedo et al., 2009). This has resulted in cell titres (concentration of product in cell culture) increasing from milligrams to grams per litre of cell culture (Gottschalk, 2008). However, improvements in product titre and upstream efficiency have resulted in the productivity (and cost) bottleneck being shifted to downstream processing where separation equipment such as expensive chromatography columns must be scaled accordingly to cope with the increasing amounts of material generated (Balasundaram et al., 2009). For example, an industrial scale protein A chromatography operations can cost as much as $1.5 million a year (Walsh, 2010). For non-antibody therapeutic proteins such as enzyme replacement therapies, the purification process is far less straight forward as the initial powerful capture step provided by protein A chromatography cannot be used. In such cases purifications are typically conducted using a mixture of different types of chromatography such as ion-exchange, hydrophobic and size exclusion. According to Roque et al. (2004) and Walsh (2010), the total cost of downstream processing can cost as much as 80% of the total manufacturing cost. Such high costs highlight the importance of re-evaluating current strategies/methodologies in the design and implementation of downstream processing,

in particular the need to consider alternatives to chromatographic separations. Techniques such as aqueous two-phase extraction, precipitation and crystallisation are considered as alternative techniques by researchers (Przybycien et al., 2004).

Aqueous two-phase extraction is a promising alternative downstream processing technology; it is similar in operation to conventional liquid-liquid extraction using organic solvents, however, water is the only solvent therefore these aqueous based systems are much less likely to damage fragile biopharmaceuticals (Asenjo and Andrews, 2011). One of the benefits of using liquid-liquid extraction is the many different types of equipment already available which can facilitate operation of an extraction process in different ways such as simple batch extraction or more sophisticated continuous counter-current operations. Other technologies such as precipitation and crystallisation also do exist which should also be considered as viable alternatives, but will not be considered in this work.

One of the challenges to adopting new technologies is that they require additional research and development resources to be successfully developed before use when compared to mature technologies such as chromatography. This is particularly important in the production of biopharmaceuticals where time to market is critical. Therefore, there is a need to explore alternative separation techniques in a resource efficient way while still providing good overall process understanding and knowledge that can satisfy the Quality by Design (QbD) mandate (Rathore and Winkle, 2009a). Specifically, there is a need to develop a robust process that can deliver high-quality therapeutic material each and every time it is produced. This is a particularly challenging task to address given the inherent variability present when dealing with biological systems. To a certain extent, lack of robustness and understanding has prevented the adoption of alternative technologies such as aqueous two-phase extraction (Benavides and Rito-Palomares, 2008, Rosa et al., 2010). However, with good process development and understanding, process robustness can be achieved. The use of high-throughput screening and scale-down technologies has helped address this issue as they allow for greater amounts of information to be generated from limited resources (Nfor et al., 2009, Bhambure et al., 2011). For example, the use of automated liquid-handling has been used for aqueous two-phase system process development by Bensch et al. (2007). Additional process understanding can also be gained by implementing fundamental process models capable of simulations during process development, especially if modelling exercises are integrated with experimental efforts to provide a

greater overall process understanding (Rodrigues and Minceva, 2005, Rajendran et al., 2009).

1.3 Motivations and objectives

The aim of this research is to integrate high-throughput experimental techniques with process modelling for the development of aqueous two-phase extraction processes for the purification of non-antibody therapeutic proteins. In this work the enzyme α-amylase is considered as a model protein for acid-alpha glucosidase which is an enzyme replacement therapy what was being developed by BioMarin pharmaceuticals. The high-throughput experiments were conducted using automated liquid-handling capable of providing a greater amount of information through scaled-down experiments. The information generated using the high-throughput experimentation will be used to tailor a mechanistic process model capable of simulating aqueous two-phase extraction under a variety of conditions. The combined experimental and modelling approach will allow for greater process understanding and therefore allow for decisions to be made on whether or not a specific aqueous two-phase extraction process is feasible to deploy.

1.4 Outline of thesis

This thesis consists of five additional chapters. Chapter 2 consists of a detailed review of literature relating to aqueous two-phase systems. The literature review will start off broadly by covering the fundamentals of aqueous two-phase systems. This broad overview will then be followed by a detailed discussion of experimental and modelling techniques used in the development of aqueous two-phase systems.

Chapter 3 investigates the development of a process model capable of conducting simulations of aqueous two-phase systems under dynamic conditions using multi-cycle batch extraction and continuous count-current modes of operation. In addition, the power of dynamic modelling is shown by simulating aqueous two-phase extraction under variable operating conditions.

In Chapter 4, phase diagrams for aqueous two-phase systems are generated using a combination of high-throughput experimentation as well as standard bench-scale experiments. A unique method is determined to quantify the uncertainty in the phase diagram as a result of the experimental procedure utilised. The impact of the uncertainty in experimental phase diagrams on parameter estimation of interactions parameters is explored.

In Chapter 5, the modelling and experimental techniques developed in Chapters 3 and 4 are used to simulate an aqueous two-phase extraction process for a two-protein extraction system consisting of α-amylase and myoglobin. Simulations are conducted to explore the impact of various process parameters such as number of stages, tie-line length and phase ratio and the impact these have on equipment choice.

Chapter 6 summarises the work of the thesis and discusses opportunities for future work.

1.5 Contributions of this thesis

The following summarises the main contributions of this thesis:

- A process model based on first principles was developed to simulate aqueous two-phase extraction under dynamic conditions.
- The impact of process variability in continuous operation of counter-current aqueous two-phase extraction was investigated.
- Aqueous two-phase system phase diagrams were generated using a combination of high-throughput screening and lab scale experiments.
- An efficient two-stage screening approach to determine the location of the binodial curve (phase transition curve) to a high accuracy was developed.
- A unique method was developed to quantify the error in the phase diagrams of aqueous two-phase systems generated using high-throughput experiments.
- The impact of experimental uncertainty on the estimation of thermodynamic interaction parameters was investigated.
- A case study separation using two model proteins, α-amylase and myoglobin, was used to demonstrate the feasibility of using aqueous two-phase extraction using the combination of modelling and high-throughput experimental techniques developed in this work. The case study separation demonstrates an approach that could be utilised in the development of purification processes for non-antibody therapeutic processes.

Chapter 2: Literature Review

2.1 Overview of aqueous two-phase systems

2.1.1 What are two-phase systems?

Although two-phase systems exist between solid-liquid and liquid-vapour, in this work two-phase systems refer to those of the liquid-liquid type. Liquid-liquid two-phase systems are formed when two distinct layers of liquid are formed after two immiscible/partially immiscible liquids are mixed together. An everyday example of this is the two-phase system formed when oil and water are mixed together which is shown in Figure 2 - 1. Traditionally, two-phase systems using organic solvents are used for separation purposes.

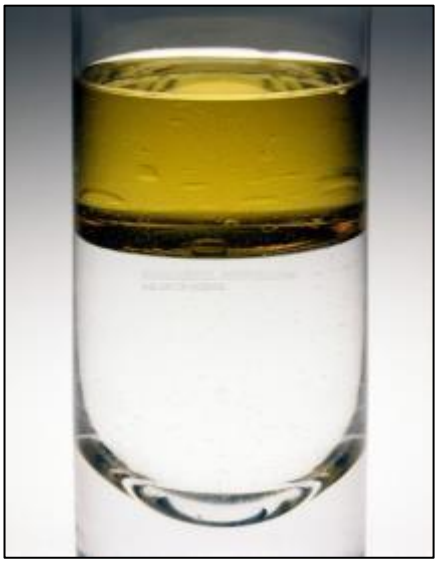


Figure 2 - 1: Oil-water two-phase system. Picture sourced from Peticolas (1986).

The density of the phases in two-phase systems determines which phase is the top or bottom phase, therefore the least dense phase is always the top phase.

Two-phase extraction works on the principle that a chemical species will have a different solubility in the two different phases. For example, a drug substance could be extracted from oil to water if its solubility in water was greater. The extent to which the drug substance is extracted into the water phase will depend on its relative solubility, this is characterised by the partition coefficient (K_i). The partition coefficient for species i is typically defined as follows:

$$K_i = \frac{C_{i,eq,top}}{C_{i,eq,bottom}} \qquad \text{Eq.(2.1)}$$

where $C_{i,eq,top}$ and $C_{i,eq,bottom}$ are the equilibrium concentrations of species i in the top and bottom phase, respectively. If K_i is equal to 1, there will be no extraction since both phases have equal solubility and hence an equal distribution of material. If the value of K_i is greater than 1, the extraction direction will be from the bottom phase to the top phase. If the K_i value is less than 1, the extraction direction will be from the top phase to the bottom phase. To enhance separation performance, multiple extraction stages can be utilised if the partition coefficient is weak.

The advantages of two-phase systems include (Asenjo and Andrews, 2011):

1) Low temperature of operation – this is important for biopharmaceuticals which are particularly sensitive to heat.
2) Simple set-up, especially if using a batch extraction system.
3) Potential for high selectivity.

Despite these advantages, traditional two-phase systems formed using organic solvents are rarely used in the production of biopharmaceuticals as organic solvents have poor interactions with proteins resulting in denaturation (Albertsson, 1958).

2.1.2 What are aqueous two-phase systems (ATPSs)?

Aqueous two-phase systems are a promising alternative to traditional liquid-liquid extraction systems. They operate in the same way as conventional liquid-liquid extraction systems; however, the process of establishing two-phases is different. In aqueous two-phase systems, water is the only solvent therefore the environment is more favourable for proteins when compared to organic two-phase systems. In addition, aqueous two-phase systems have a lower interfacial tension (1×10^{-7}-$1\times10^{-4}Nm^{-1}$) when compared to organic-aqueous systems ($1\text{-}20\times10^{-3}Nm^{-1}$) (Hatti-Kaul, 2000).

There are two main types of ATPSs, polymer-polymer and polymer-salt. The formation of the phases is associated with the incompatible interactions between the two polymers or the polymer and salt. The most common systems reported in literature are that of polyethylene glycol (PEG) and dextran (polymer-polymer) and PEG-phosphate (polymer-salt) (Rosa et al., 2010). In practice ATPSs are formed by mixing together, with water, two polymers or a polymer and salt above a certain critical concentration where the formation of two-phase is more favourable than a single homogenous phase. The practical application of this phenomenon was first realised by Albertsson (1958).

Hatti-Kaul (2000) points out that most hydrophilic polymer pairs are incompatible with each other, this should mean that there are many potential polymers combinations that could be used for the formation of two phases. PEG, however, is used due to the fact it is cheap, widely available and versatile. Figure 2 - 2 shows a simple illustration of an aqueous two-phase system. Depending on the type of system utilised, the composition of the two phases will be different.

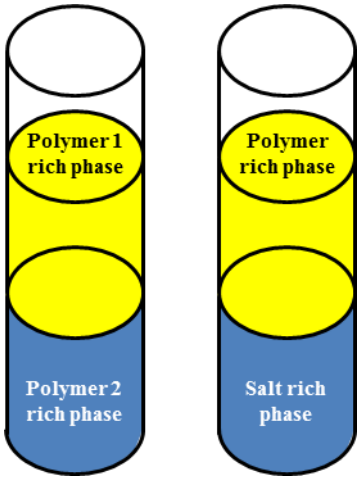


Figure 2 - 2: Illustration of polymer-polymer and polymer-salt aqueous two-phase systems. Salt rich phase is usually denser than polymer rich phase therefore forms bottom phase. In each phase, the major component present is water.

Industrial application of aqueous two-phase systems

To date there have been limited reports on the use of aqueous two-phase systems in industry. Akzo Nobel and Genentech have patent applications for the use of aqueous two-phase systems to purify human growth hormone (Kirk and Pascal 1999) and insulin like growth factor (Builder et al., 1995), respectively. Many researchers have also considered the use of aqueous two-phase systems to purify mAbs as highlighted in a review by Azevedo et al. (2009). The lack of widespread use of aqueous two-phase systems can be attributed to two main factors, firstly for proteins such as mAbs there already exists highly effective platform processes based on chromatographic techniques (Shukla et al., 2007). Secondly, to find the best operating conditions for an aqueous two-phase system requires navigation of a large experimental space as many different factors influence solute partitioning as well as phase formation (Hatti-Kaul, 2001, Rosa et al., 2010, Asenjo and Andrews, 2011).

Aqueous two-phase system phase diagrams

Aqueous two-phase systems are best described as partially immiscible. This is because in each phase there is a mixture of all components used to establish the two-phase system present. For example, in a polymer-salt aqueous ATPS, the top polymer rich phase would contain some salt and the bottom salt rich phase would contain some polymer. Therefore, it is useful to describe phase equilibria of aqueous two-phase systems with an appropriate phase diagram such as shown in Figure 2 - 3.

The phase diagram has two distinct areas separated by the binodial curve. The binodial curve represents the boundary between systems that will form two phases and those that will not. On the diagram the points M and M' corresponds to the total mixture/system composition. A composition of M will form two phases whereas a composition of M' will not. The point T and B corresponds to the composition of the top and bottom phases, respectively, of the mixtures with composition M.

The dashed tie lines allow one to calculate the composition of the top and bottom phases from the initial total mixture composition. Using the inverse lever-arm rule, the tie lines can also be used to calculate the relative amounts of top and bottom phase and hence the volumes of the phases. This is of particular importance if aqueous two-phase systems are used as a concentration step. The ratio of the amount of top phase to bottom phase is equal to the ratio of line MB to line TM. Generally speaking the tie lines for a certain type of system will all have the similar gradient, so once a few tie-lines have been determined, the positions of other tie lines can be interpolated or extrapolated (Hatti-Kaul, 2000, Amrhein et al., 2014). Often in literature, the tie line length (TLL) is varied to see the effect it has on the partition coefficient. The variation of the TLL is essentially a way of saying that the equilibrium concentrations of the top and bottom phase were varied. The equilibrium concentration of the major component in each phase increases as the TLL is increased. The TLL can easily be worked out using Pythagoras' theorem for a right angle triangle using the equilibrium compositions obtained from a phase diagram. The point C corresponds to the system critical composition, at this point the two phases will have the same composition and a partition coefficient value of 1 (Asenjo and Andrews, 2011). The critical point can also be found by determining where the tie-line length becomes zero.

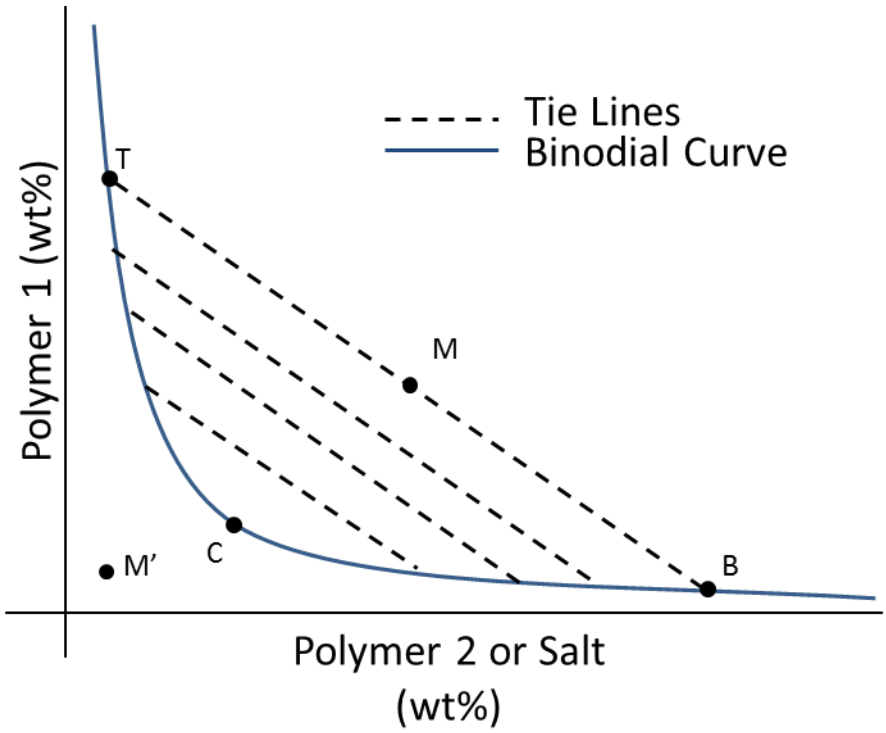


Figure 2 - 3: Illustration of a hypothetical aqueous two-phase system phase diagram.

Traditional phase diagrams for aqueous two-phase systems give information about how the components used to establish an aqueous two-phase system partition between the resultant phases. Unfortunately, the phase diagram is not capable of describing how additional solutes, such as therapeutic proteins, will partition between the two phases. The partition coefficient of solutes is dependent on both the two-phase system being used as well as the individual physical characteristics of the solute.

The partition coefficient

The partition coefficient, previously described by Eq.(2.1), is the most important parameter when considering aqueous two-phase extraction for separation purposes. It is important because it will determine whether or not using aqueous two-phase extraction as a separation step is feasible.

The partition coefficient of a specific solute is influenced by many different factors which can be split up into two main categories:

1) Intrinsic properties of the solute such as size, charge hydrophobicity etc.
2) Properties of the aqueous two-phase system such as tie-line length, polymer type etc.

Overview of solute properties which influence partitioning in aqueous two-phase systems

In a review conducted by Asenjo and Andrews (2011), it was reported that the major solute properties affecting partitioning behaviour were:

1) Hydrophobicity
2) Electrochemical properties such as charge and isoelectric point
3) Solute affinity
4) Solute size
5) Solute solubility

The overall partition coefficient has been written as a function of the above properties like so (Asenjo et al., 1994, Asenjo and Andrews, 2011, Tubio et al., 2004):

$$K = K_0 \times K_{hydrophobic} \times K_{electrical} \times K_{affinity} \times K_{size} \times K_{solubility} \quad \text{Eq.(2.2)}$$

where K_0, $K_{hydrophobic}$, $K_{electrical}$, $K_{affinity}$, K_{size} and $K_{solubility}$ are contributions to the partition coefficient due to no interactions, hydrophobic interactions, electrical interactions, affinity interactions, protein size/steric interactions and solubility interactions, respectively. Eq.(2.2) is very simplistic, and in reality aqueous two-phase systems are very complex and currently there exist no all-encompassing theory or models that can accurately predict the partition coefficient. It may therefore be better to write the partition coefficient in a more general way since the above mentioned properties definitely do have an effect on the partition coefficient where fn is generic function that takes into account all the different types of interactions:

$$K = fn(K_0 \times K_{hydrophobic} \times K_{electrical} \times K_{affinity} \times K_{size} \times K_{concentration}) \quad \text{Eq.(2.3)}$$

Hydrophobicity effects

Hydrophobicity has a significant impact on solute partitioning in aqueous two-phase systems (Andrews and Asenjo, 2010). This is interesting considering aqueous two-phase systems are formed using hydrophilic polymers and salts which readily mix with water. The result is that the established phases have different levels of hydrophobicity which can be used to influence solute partitioning. In PEG-dextran and PEG-salt systems the top PEG rich phase is usually more hydrophobic than the bottom dextran/salt rich phases.

Franco et al. (1996) investigated the sole effect of protein surface hydrophobicity on the partition coefficient in both PEG-dextran and PEG-phosphate aqueous two-phase systems. Franco et al. (1996) were able to chemically modify a protein to increase its hydrophobicity while maintaining its pI, size, protein conformation and solubility. They found that increases in protein surface hydrophobicity saw increases in the protein partition coefficient; this can be associated with the fact that the top PEG phase is the more hydrophobic phase. Furthermore, they found that the addition of NaCl salt led to increases in the partition coefficient for both PEG-phosphate and PEG-dextran systems. Figure 2 - 4 shows some of their findings. Although the partition coefficient is increased in the presence of NaCl, in multi-component separations typical of bioprocesses one must also consider the potential additional consequences such as increasing the partition coefficient of impurities.

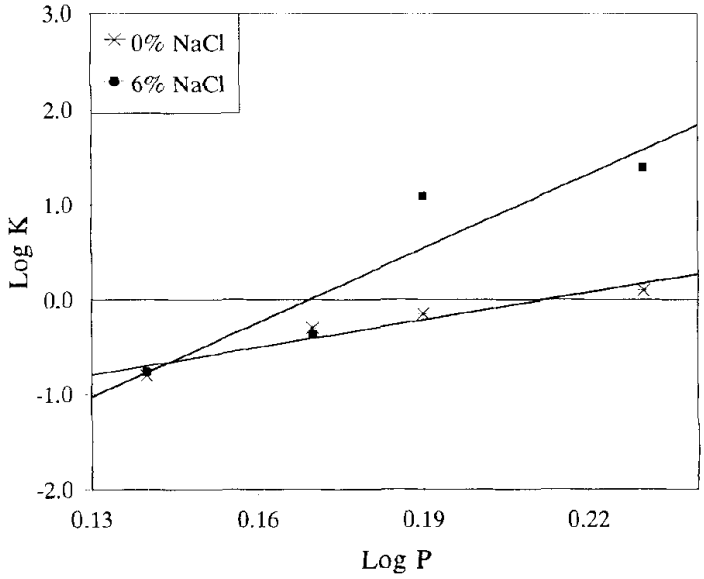


Figure 2 - 4: Partitioning of hydrophobically modified β-lactoglobulin as a function of protein surface hydrophobicity in the 15% PEG 1500-10% phosphate (pH 5.3) ATPS in the absence and presence of 6% NaCl by Franco et al. (1996).

Due to the significant effect of hydrophobicity on the partition coefficient, researchers have attempted to predict partition coefficients from hydrophobicity alone. For example, Salgado et al. (2008) have developed a way to predict partition coefficients based on average surface hydrophobicity (ASH) derived from protein amino-acid sequences. Their models were most successful in systems where hydrophobic interactions were the dominant. Furthermore, their best results obtained with PEG-dextran systems had an error of 15.6% when compared to experimental data. A method of predicting protein partitioning solely based on hydrophobicity is, however, somewhat limiting due to the

fact that many other interactions such as electrostatic and specific affinity influence solute partitioning.

Electrochemical/charge effects

The building blocks of proteins are amino acids which are amphoteric. An amphoteric species is capable of acting as both an acid and a base therefore they are capable of carrying both a negative and positive charge. The cumulative effect of the many amino acids is that proteins are capable of carrying different charge values which can be affected by pH. The pH that results in no net charge is known as the isoelectric point (pI). Protein charge is important and has many applications, for instance in chromatography. The application of protein charge in aqueous two-phase systems proves to be promising due to the presence of an interfacial potential difference (Andrews et al., 2005a).

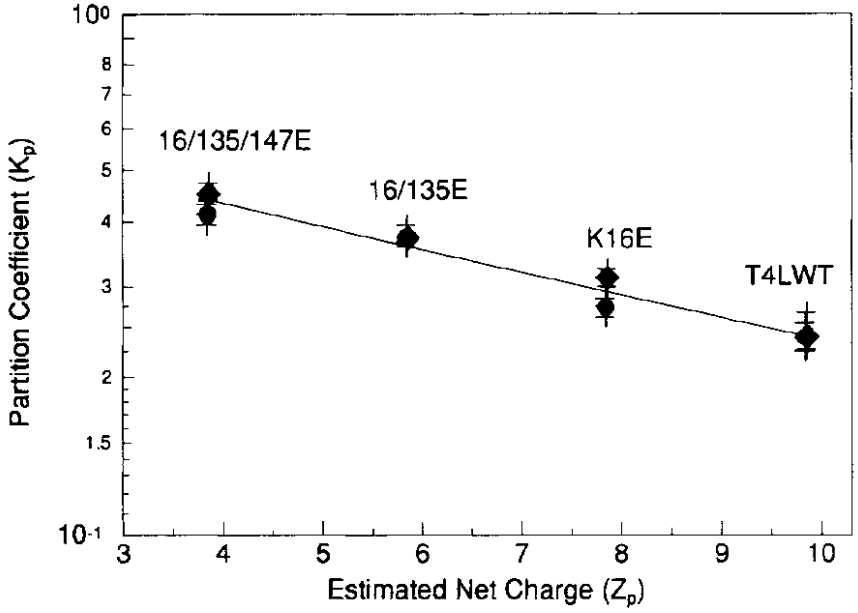


Figure 2 - 5: Different mutants of T4 lysozyme with different charge values (Luther and Glatz, 1994).

Luther and Glatz (1994) investigated the effect of protein charge on partitioning in PEG-dextran aqueous two-phase systems. They modified the protein T4 lysozyme by mutating it to replace positively charged lysine amino acids with negatively charged glutamic acid amino acids. Figure 2 - 5 shows how the partition coefficient varied with protein charge for the T4 lysozyme mutants. At higher net charge values, the partition coefficient is lower suggesting that the PEG rich top phase has a positive charge.

Andrews et al. (2005b) have reported similar behaviour in their PEG-phosphate systems in the absence of NaCl. However, when NaCl was added in their systems the opposite

trend was observed, i.e. partition coefficient increased with increasing net charge. This suggests that the Cl^- ions partition to the top PEG-rich phase resulting in a more electronegative phase. The addition of NaCl is likely to change the direction of interfacial potential difference. It has been proposed that the effects of interfacial potential difference and charge on the partition coefficient can be related by the following equation (Luther and Glatz, 1995, Luther and Glatz, 1994):

$$\ln K_i = \ln K_0 + \frac{FZ_p}{RT} \times (\varphi_B - \varphi_T) \qquad \text{Eq.(2.4)}$$

where K_0 is the partition coefficient in the absence of electrostatic interactions, F is the Faraday's constant, R is gas constant, T is absolute temperature, Z_p is net protein charge and φ is the electrostatic potential of the top and bottom phase. Eq.(2.4) provides a good qualitative description of how charge influences the partition coefficient, however, cannot solely be used for predictive purposes due to the presence of other interactions.

Affinity effects

Affinity effects are due to specific interactions between two species. For example protein A is capable of binding specifically to monoclonal antibodies (mAbs) and therefore has its use as a ligand in affinity chromatography resins for the highly selective separation of mAbs. Researchers have investigated the use of such interactions in aqueous two-phase systems as they are a promising way of increasing separation performance (Andrews et al., 1990, Barbosa et al., 2008, Maestro et al., 2008, Eriksson et al., 1976). There are two main ways affinity could be added to improve separation: 1) modification of the system, e.g. binding ligands to activated PEG, and 2) modification of product, e.g. tagged proteins.

The binding of ligands glutathione, trypsin inhibitor, protein A and anti-bovine serum albumin (BSA) to PEG was investigated by Andrews et al. (1990) as a means to improve separation performance in aqueous two-phase systems. They observed that the partition coefficient increased when there was an affinity ligand coupled to PEG in the top phase. Although there was an increase in the partition coefficient, the absolute value varied between the different proteins. For example, when using glutathione bound PEG which enhanced the recovery of thaumatin. A promising 17 fold increase in value of the partition coefficient, from 0.27 to 4.6, was observed. A 12.5 fold increase in the partition coefficient, from 0.04 to 0.5, was observed when Protein A bound PEG was used to enhance the recovery of Immunoglobulin G (IgG) antibody, however, the absolute value of the final partition coefficient is still very low. Maestro et al. (2008)

have looked at the affinity partitioning of proteins tagged with a choline binding group known as C-LytA. This choline binding group has an affinity for PEG and therefore partitions into the PEG rich phase. As a result C-LytA tagged protein also partitions into the PEG rich phase. An interesting aspect of this setup is that when choline is added to the two-phase system, it competes for the PEG polymer binding sites resulting in the partitioning of the C-LytA tagged protein into the lower dextran rich phase. This phenomenon was used as a tool to flip the partition behaviour (see Figure 2 - 6) of the material to be separated.

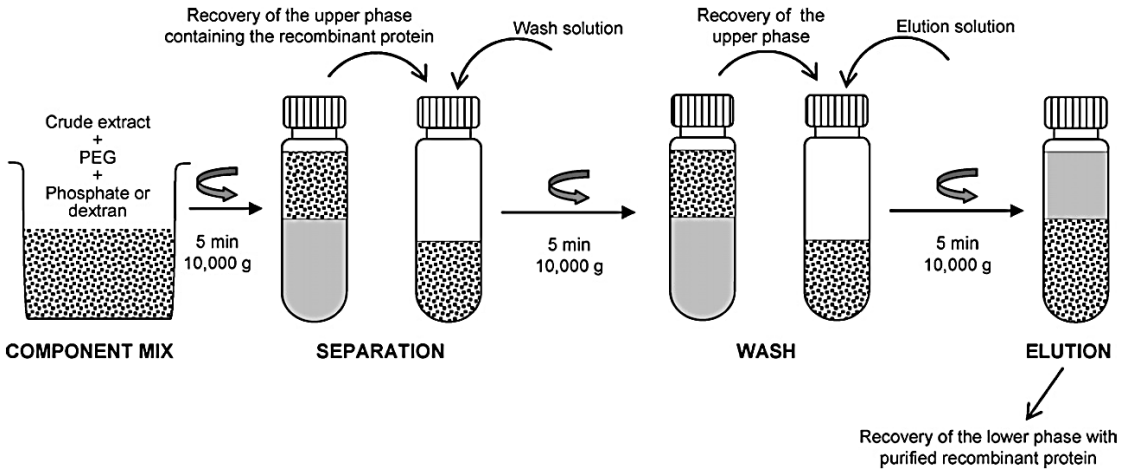


Figure 2 - 6: Concept of affinity separation in aqueous two-phase systems proposed by Maestro et al. (2008).

Protein size effects

Schmidt et al. (1996) discussed the effects of protein size on partitioning in PEG-phosphate aqueous two-phase systems. They stated that larger proteins have a lower maximum concentration in the top phase due to a steric exclusion effects caused by the presence of PEG. It is unclear how protein distribution is affected by steric exclusion in PEG-dextran systems since the dextrans used in aqueous two-phase systems are typically of a higher molecular weight than PEG, however, this was not covered in their study.

Protein solubility effects

The effect of total protein concentration on aqueous two-phase systems has been investigated by Schmidt et al. (1996) and Andrews and Asenjo (1996) . Their research shows that the partition coefficient does not necessarily remain constant with increasing total protein concentration. The concentration independent partition coefficient is observed at low protein concentrations; and the concentration threshold at which this is observed varies for each protein due to individual protein properties. Andrews and

Asenjo (1996) have also suggested that when the concentration of a protein is above the concentration independent threshold, there is likely to be protein precipitation at the interface of the aqueous two-phase systems.

Overview of system properties which influence protein partitioning

Aqueous two-phase systems can be modified in many different ways which will influence protein partitioning. Commonly investigated factors are:

1) Operational characteristics such as temperature and pH
2) Polymer characteristics such as molecular weight and concentration
3) Addition of chemicals such as salts to enhance/modify separations

Temperature and pH

Gautam and Simon (2006) investigated the influence of pH and temperature on the partitioning behaviour of β-glucosidase in a PEG-phosphate aqueous two-phase system. They used a pH and a temperature range of 6.0-7.5 and 25-55 ◦C, respectively. They found that for β-glucosidase, the partition coefficient increased with both pH and temperature.

The β-glucosidase is positively charged at the investigated pHs as it has a pI of ~8.7. The increase in partition coefficient due to increasing pH makes sense as β-glucosidase charge is likely to become less positive which would encourage partitioning into a positively charged PEG rich top phase.

The observed increase in partition coefficient with temperature does not seem to be a general trend, in fact Saravanan et al. (2008) observed the opposite when studying the partitioning of ovalbumin and myoglobin, however, they used a different system composed of PEG and poly(acrylic acid) (PAA) making the two studies more difficult to compare. Forciniti et al. (1991d) suggested that the influence that temperature has on protein partitioning is dependent on the protein characteristics, e.g. larger hydrophobic proteins are more susceptible to being influenced by temperature. What this suggests is that the effects of temperature on partitioning greatly depends on the individual characteristics of the system and solute being used. Given that most aqueous two-phase systems are different from each other, it is difficult to say that the effect of temperature observed in one study will be similar to the effect observed in another.

Polymer molecular weight, concentration and polydispersity

Two of the easiest polymer characteristics that can be changed in aqueous two-phase systems are the molecular weight and concentration, however, isolating the effects of molecular weight from changes in concentration is somewhat difficult. Albertsson et al. (1987) conducted an early study on the influence of polymer molecular weight on the partitioning of nine different proteins. The effect of changing dextran molecular weight can be seen in Figure 2 - 7.

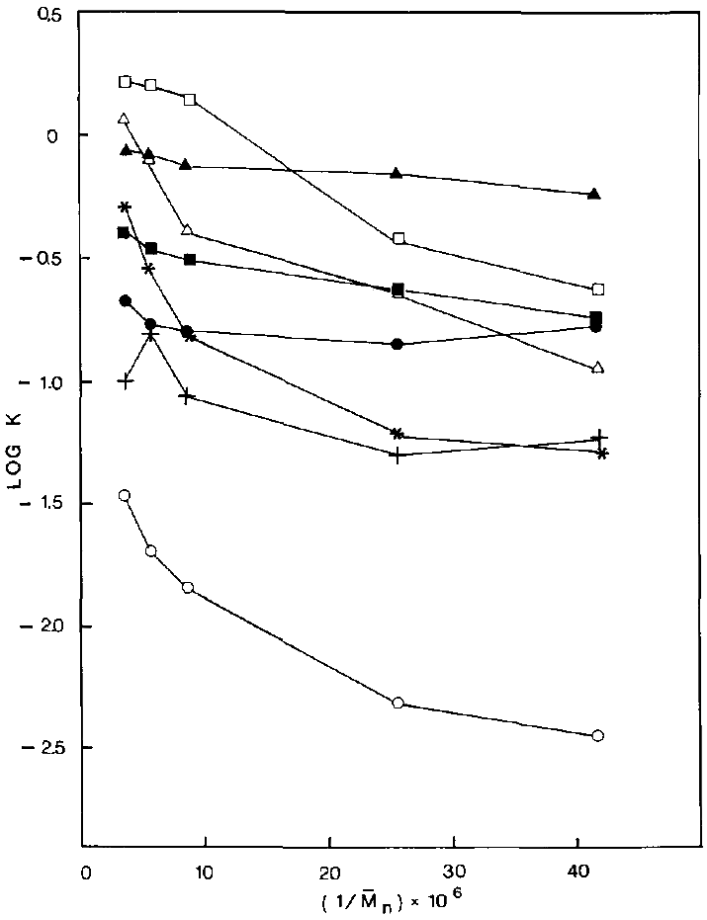


Figure 2 - 7: Effect of changing dextran MW in a PEG-dextran ATPS. (■) BSA, (Δ) catalse, (□) β-galactosidase, (○) phosphofructokinase, (*) ribulose diphosphate carboxylase, (●) cytochrome c, (▲) ovalbumin, (+) lactate dehydrogenase (Albertsson et al., 1987).

To better understand the results, Albertsson et al. (1987) used a theoretical model based on the Flory-Huggins theory which suggested that there is a linear relationship between $\log K$ and $\left(\frac{1}{M_n}\right)$ where M_n is average molecular weight. The experimental data in Figure 2 - 7 generally shows that increasing the molecular weight of dextran increases the partition coefficient. Increasing the PEG molecular weight results in a decrease in the partition coefficient (not shown here) (Albertsson et al., 1987). This relationship was expected according to their theory; however the experimental data suggests that a more

complicated non-linear relationship exists that is not explained by their model. This highlights that there is more to consider than just molecular weight. Another observation made by Forciniti et al. (1991c) was that the change in partition coefficient levelled off as the polymer molecular weight increased to large values, this is a similar to the data provided by Albertsson et al. (1987).

The theoretical model used in the above studies is very simple and does not take into account the fact that the concentration of the system also changes when the polymer molecular weight is modified. This was reported by Forciniti et al. (1991c) and is shown in Figure 2 - 8.

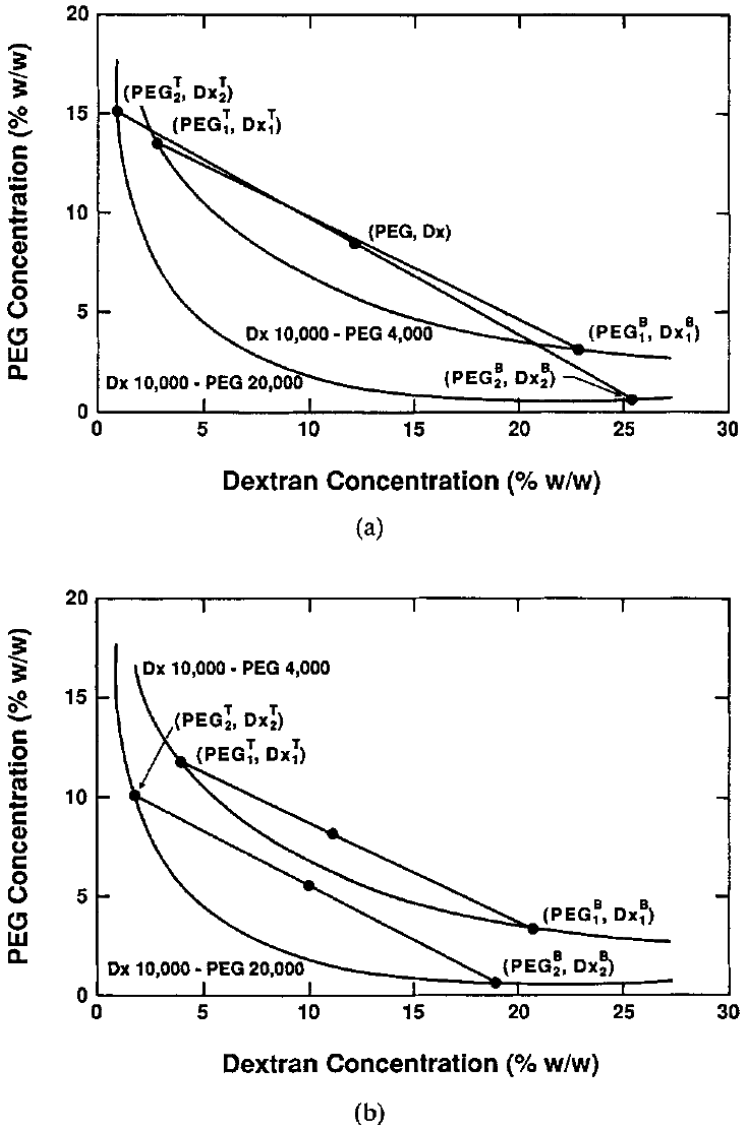


Figure 2 - 8: ATPS composition variation as a result of changing molecular weight. (a) Constant total concentration, (b) Constant tie line length, (Forciniti et al., 1991c).

Figure 2 - 8 shows how changing the polymer molecular weight whilst keeping the total polymer concentration constant results in different tie line lengths and hence a different

phase composition. Whereas keeping the tie line length constant results in the need to change the total polymer concentration which also results in a different phase composition. This means that it is not possible/extremely difficult to isolate molecular weight influences from concentration influences.

Another factor that is too simplistic in the model is the molecular weight; an average value is assumed which in reality is an oversimplification due to polydispersity. Polydispersity is a phenomenon associated with polymers that arises from the fact that polymer chains have different molecular weights as a result of differing degrees of polymerisation. This means that interpreting polymer molecular weight can be difficult as there is not one true value. Researchers have used models which have taken into account the impact of polydispersity for better prediction capability (Cabezas, 1996, Pazuki et al., 2010).

Forciniti et al. (1991a) actually investigated the molecular weight distribution of polymers in a PEG-dextran ATPS. They found that the dextran molecular weight in the top PEG rich phase and bottom dextran rich phase differed considerably. The molecular weight of dextran in the top phase was found to be much lower and more monodisperse than the molecular weight of dextran in the bottom phase. This also means that the molecular weight of the dextran in the two phases is different to the initial overall dextran molecular weight.

Polymer concentration is another important characteristic that needs to be considered, and its effects are usually investigated by changing the tie line length by modification of the total concentration of the polymer mixture. A larger tie line generally results in richer phase concentrations. Forciniti et al. (1991b) observed that if the value of the partition is greater than 1, then the partition coefficient value will increase when the tie line length is increased whereas the opposite was reported for cases where the partition coefficient value is less than 1, i.e. the direction of partitioning is enhanced.

Addition of salts

Rosa et al. (2007) conducted a design of experiment study on the purification of human IgG using a PEG-phosphate ATPS. Their study included the effects of ionic strength and salt. They found that increasing the NaCl concentration in the ATPS resulted in the partition coefficient for IgG increasing. Apparently the presence of NaCl increases the difference in hydrophobicity between the two phases therefore increasing the strength of separations based on hydrophobic interactions. A review by Huddleston et al. (1991),

however, suggests that the actual mechanism and interactions are likely to be more complicated. This is possibly due to the presence of different types of interactions such as hydrophobic and electrostatic.

Summary of factors influencing partitioning in aqueous two-phase systems

Partitioning of solutes in aqueous two-phase systems will depend on two main factors: 1) properties of the solute itself, and 2) properties of the aqueous two-phase system. Properties of the solute include hydrophobicity, charge, specific affinity/binding, solute size and solubility. Solute properties can be exploited to enhance separation characteristics by changing aqueous two-phase system properties such as temperature, pH, polymer molecular weight and electrochemical behaviour through addition of salts. This wide range of interactions makes aqueous two-phase extraction potentially very powerful, however, challenging to develop due to the inherently large experimental space.

2.1.3 Two-phase extraction modes of operation modes

Aqueous two-phase extraction, like conventional liquid-liquid extraction systems, can be operated in many ways to suite different separation needs (Perry and Green, 2008). The three main set-ups used are:

1) Single stage operation using mixer settler for either batch or continuous operation.
2) Batch or continuous operation using a cross flow arrangement.
3) Batch or continuous operation using a counter-current operation.

An illustration of these modes of operation is shown in Figure 2 - 9. Single stage mixer-settler tank operations are the simplest form of two-phase extraction. For batch extraction systems a single tank may be sufficient for both mixing and settling. For continuous operations a stirred tank/mixer is linked in series to a settling stage which is designed to provide the residence time required for two phases to settle so that each phase can be continuously extracted. In cross-current operations fresh solvent is introduced at each stage which extracts product from the feed stream. The benefit of introducing fresh solvent in each stage is that additional product can be extracted with each additional stage. The main disadvantage of using cross-current operations is the large amount of solvent required (Kumar and Awasthi, 2009). Counter-current operations involve two process streams flowing in opposite direction to each other where one stream has an inlet at the first stage while the other stream has an inlet at the last stage. The benefit of counter-current operation is that the concentration differences which drive separations are maintained across all stages in a similar manner to counter-current heat exchangers, distillation columns and absorbers (Perry and Green, 2008). Counter-current operations can be implemented in a variety of ways as there exist many different types of equipment such as:

1. Continuous centrifugal separators.
2. Packed columns.
3. Agitated columns.
4. Multiple continuous mixer-settle tanks connected in counter-current mode.

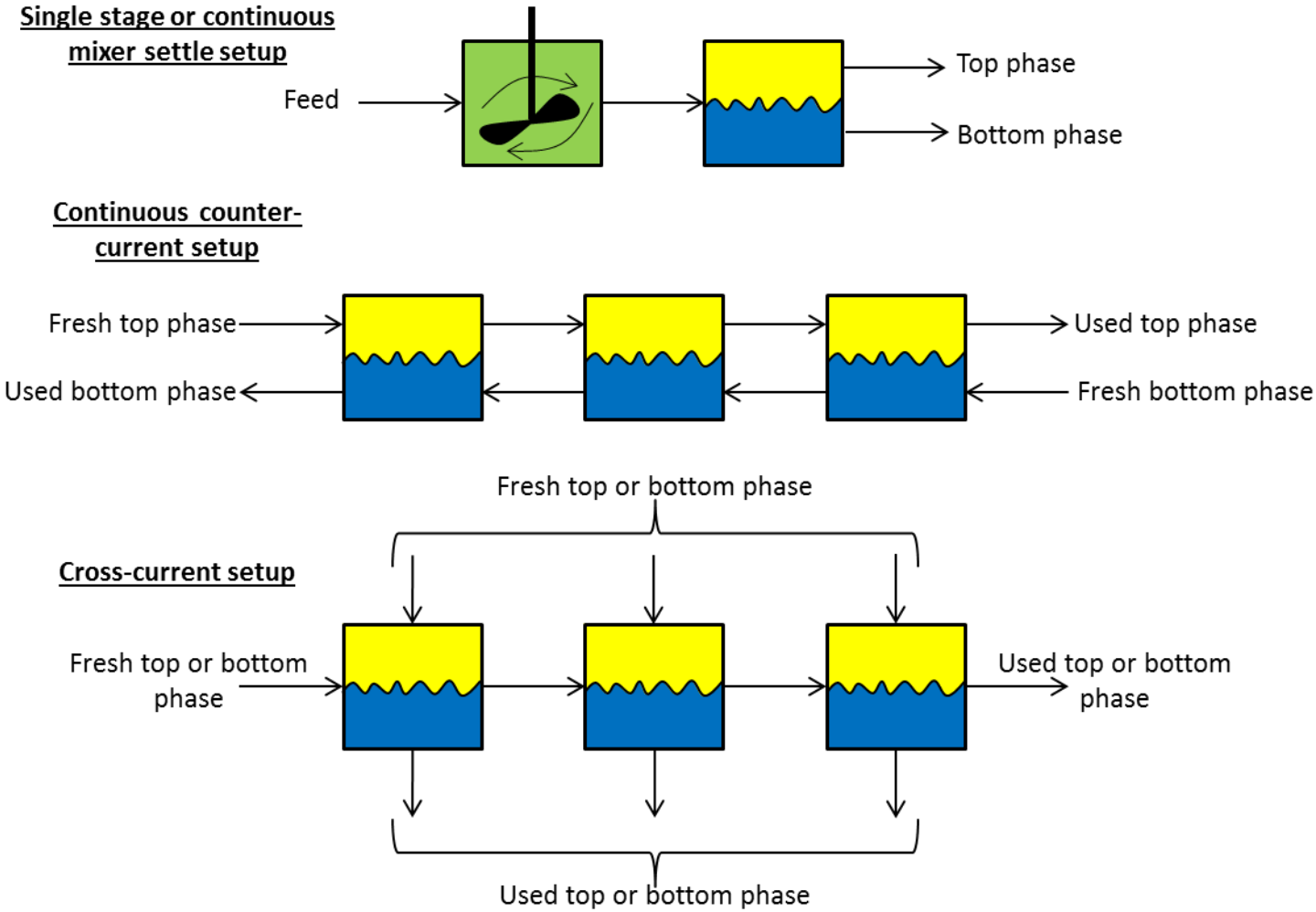


Figure 2 - 9: Schematic of different modes of two-phase extraction

2.2 Aqueous two-phase system experimentation

This section will cover the various experimental methods used when working with aqueous two-phase systems.

2.2.1 Conventional phase diagram determination techniques

When determining the phase diagrams of aqueous two-phase systems there are two main aspects that need to be determined.

1. The binodial curve.
2. Tie-lines relating top and bottom phase equilibrium compositions.

Experimental determination of the binodial curve

The binodial curve is traditionally determined experimentally by finding the location in which a phase change is observed. The simplest method for determining whether a two-phase system exists is by thoroughly mixing the system and testing for the presence of turbidity[1]. Typically when single phase systems are mixed they do not result in turbid solutions.

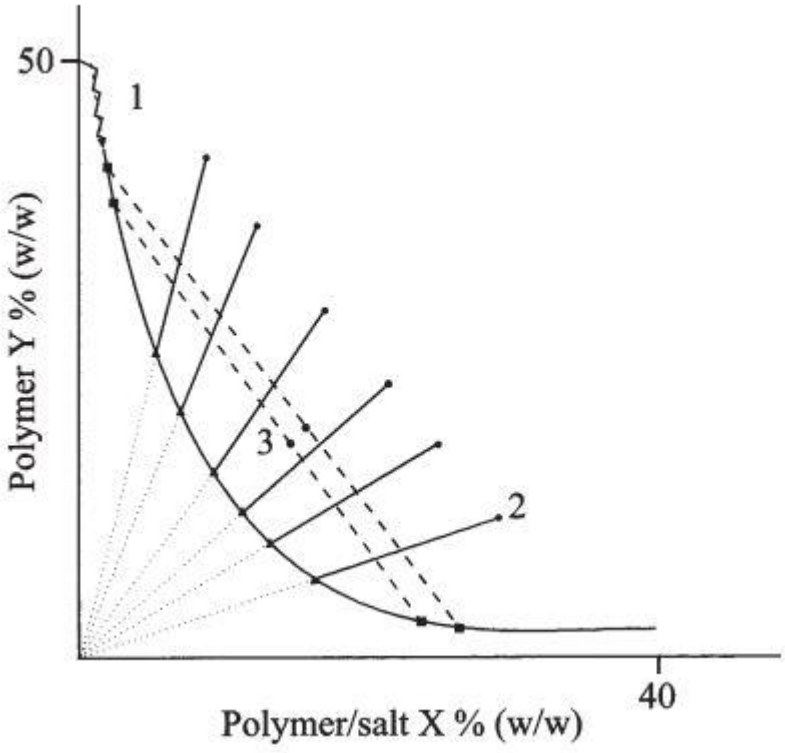


Figure 2 - 10: Illustration of two methods ("zig-zag" or dilution) to determine the binodial curve using turbidity. The "zig-zag" method starts at point 1 with pure polymer stock solution and involves "zig-zagging" above and below the binodial curve. The dilution method starts with many systems in the two-phase region, such as point 2, which are diluted until a phase transition is observed. Image obtained from Kaul (2000).

[1] Turbidity is the cloudiness or haziness observed during mixing of a two-phase system.

Kaul (2000) has described two different methods using turbidity to determine the binodial curve location which are illustrated in Figure 2 - 10.

The dilution method involves making a range of systems of known overall composition that are within the two-phase region. Each system is diluted until they form clear systems when mixed. The binodial curve composition can be calculated by taking into account the starting composition and mass of diluent added using a laboratory weighing scale. The benefit of this method is that it is simple. The main disadvantage of the dilution method is that a large number of aqueous two-phase systems must be prepared to obtain enough binodial curve points.

The "zig-zag" method involves starting with a pure stock solution which is a single phase solution. A polymer or salt is added to the pure solution until the mixture turns turbid. The mass of material added is recorded and used to determine the composition at which a phase transition occurred. The sample is diluted into the single phase region, the mass of diluent added is recorded. Further phase transition points can be determined by repeating the procedure. The benefit of the "zig-zag" method is that many binodial point locations can be determined in a single experiment. The main disadvantage is that any errors in mass measurement are propagated as the new compositions depend on all the previously weighed masses of added diluent and polymer/salt.

Experimental determination of tie-lines

Tie-lines are typically determined in two ways:

1. Direct measurement of compositions of top and bottom phases.
2. Using the inverse lever arm rule to fit tie-lines based on ratio of top phase mass to bottom phase mass (mass ratio).

Directly measuring the composition of the top and bottom phases requires the formation of a range of aqueous two-phase systems within the two-phase region. A sample of top and bottom phase is analysed to determine the composition. The advantage of this method is that it provides information on the location of the binodial curve if enough systems are analysed since the equilibrium points must also be on the binodial curve. The main disadvantage of this method is the need for specific analytics to determine polymer and salt compositions such as HPLC, flame photometry and refractive index measurements.

Using the inverse lever arm rule to fit tie-lines requires knowledge of both the mass ratio and the binodial curve location. The mass ratio is relatively simple to calculate experimentally using appropriate density and mass measurements (Amrhein et al., 2014, Atefi et al., 2016). The main advantage of this measurement is that the equipment used is relatively simple. The main disadvantage of this method is the need to determine the binodial curve which can be a cumbersome task (see previous section for details on methods to determine binodial curves). A detailed description of the tie-line fitting method is provided in Chapter 4.

2.2.2 High-throughput process development

The implementation of high-throughput techniques in research and development of biopharmaceuticals is continuously growing (Lacki, 2014). One of the main challenges when using automated liquid handling is that a large amount of initial work must be conducted to ensure that pipetting is accurate. This is particularly important when working with aqueous two-phase systems where polymers, salts, proteins all have very different liquid properties and therefore varying liquid handling characteristics (Xie et al., 2004, Bessemans et al., 2016). A further complication for two-phase systems is the need to extract top and bottom phases without cross-contamination, this is particularly challenging at extreme phase ratios where one of the two phases is very small. The benefit of using high-throughput experimentation is that a greater amount of information can be obtained from a limited amount of resources such as sample volume or time. The use of automated liquid handling robots as a means of performing high-throughput experiments for aqueous two-phase system process development was first reported by Bensch et al. (2007). In their approach binodial curves were determined by screening many systems in 96-well plates for the presence of two-phases using dyes which partition exclusively into one phase. Tie-lines were estimated using the inverse lever arm rule and volume ratios (also measured by quantification of coloured dye). Their approach allowed for the phase diagrams of aqueous two-phase systems to be determined reasonably quickly when compared to the conventional methods discussed in Section 2.2.1. A later study by Amrhein et al. (2014) found that using mass ratios resulted in tie-lines of a more consistent gradient, however, this required the measurement of phase densities which resulted in an experimental process which is slower than that reported by Bensch et al. (2007).

The use of microfluidic[2] devices has also been investigated for the development of aqueous two-phase systems. For example Silva et al. (2014) have developed a microfluidic chip (shown in Figure 2 - 11) which has three ports for polymer, salt and water solutions. The system composition is adjusted by altering the flow rate of each solution whilst the presence of an interface is detected using an optical microscope.

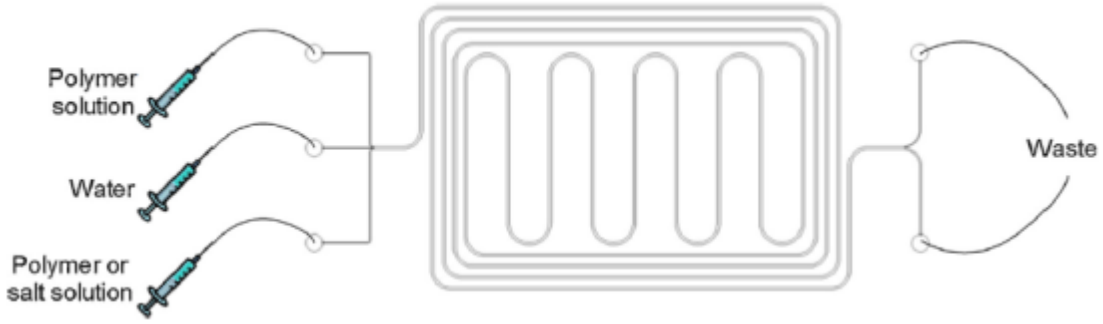


Figure 2 - 11: Schematic of microfluidic device used for aqueous two-phase systems by Silva et al. (2014).

A similar microfluidic device has been used as a pre-screening tool before conducting larger scale experiments by Espitia-Saloma et al. (2016) for the separation of human immunoglobulin. The separation of red and white blood cells in a microfluidic device using aqueous two-phase extraction was also investigated by SooHoo and Walker (2009).

2.2.3 Summary of aqueous two-phase system experimentation

Aqueous two-phase systems experimentation is used to determine phase diagrams as well as solute partitioning characteristics. Phase diagrams can be determined using established traditional methods such as titration for the binodial curve and phase analysis or inverse lever-arm rule for the phase compositions. More recently developed experimental techniques have drawn on automated liquid handling and microfluidics to scale-down aqueous two-phase systems. Advances in experimental techniques have been aimed at improving the speed and amount of information that can be obtained using a limited set of experimental resources.

[2] Microfluidic devices are capable of handling small volumes of liquid (<microliters) due to channels which are micrometres in diameter (Dittrich and Manz, 2006).

2.3 Modelling of liquid-liquid extraction

Modelling of liquid-liquid extraction is useful as it allows for process insights to be gained without conducting numerous experiments as conditions can be explored in-silico and then validated using targeted experiments. Models can be built using empirical relationships or from first principles understanding.

2.3.1 Empirical modelling

Empirical modelling involves generating relationships between independent and dependent variables. The relationships are determined using general mathematical relationships which correlate well between the independent and dependent variables, however, typically do not have any underlying physical meaning. The simplest form of empirical modelling is linear regression which uses the equation of a straight line. In situations where multiple independent variables influence the dependent variable multiple linear regression can be utilised. Many real systems cannot be described using linear relationships, therefore, in such situations linear regression can be extended to better suit such applications by introducing quadratic or higher order terms. If there are many independent variables the problem of multicollinearity and model over fitting may arise which can make estimation of fitting parameters difficult (Farrar, 1982, Graham, 2003). Techniques such as principle component analysis (PCA) have been used to address the problem of multicollinearity amongst many independent variables (Wold et al., 1987). The main drawback of PCA is that it focuses only on describing variability in independent variables and not dependent variables. To address this issue, the techniques of PCA have been extended to form partial least squares (PLS) analysis (Geladi and Kowalski, 1986). Neural networks are another form of empirical modelling which are inspired by neurons in the human brain. Neural networks are very good at describing non-linear problems where there are many independent variables (Tu, 1996).

The main disadvantage of using empirical models is that a large amount of upfront work must be completed to provide enough data to train/calibrate the models. Furthermore, empirical models typically have poor predictive capabilities when attempting to explore areas beyond the training/calibration dataset. Methods such as Design of Experiment (DoE) have helped researchers in many fields, including those working with aqueous two-phase systems, address this issue by providing a systematic framework for conducting experiments to build appropriate empirical models (Glyk et al., 2015, Rahimpour et al., 2016).

2.3.2 First principles models

First principles models describe a particular system based on fundamental physical laws such as the conservation of mass, energy and momentum. Such models are useful for conducting simulations where it is expensive to run an actual experiment for process development processes. Typically initial experiments are conducted to determine key model parameters, for example for chemical reactions, experiments are used to determine rate constants and activation energies. Once parameters are determined, the model can be used to design an appropriate process in-silico. Finally the in-silico process then has to be verified using a final set of validation experiments to ensure that the model was representative of a real process.

Equilibrium stages

It is often useful to think of liquid-liquid extraction operations in terms of equilibrium stages (as for distillation). An equilibrium stage is a theoretical section or area where the different phases (e.g. top and bottom phases) establish equilibrium with each other. Typically, using more equilibrium stages result in better separation performance. An illustration of a single and multiple equilibrium stages used for modelling purposes in this work is shown in Figure 3 - 2. For simple binary/ternary separations, the number of theoretical equilibrium stages can often be determined graphically using McCabe-Thiele style plots (Perry and Green, 2008). Rosa et al. (2009) have used such plots to describe the number of equilibrium stages required to purify an antibody in a PEG-phosphate aqueous two-phase system. The main limitation of using McCabe-Thiele style plots to determine the number of equilibrium stages is that they are difficult to use for systems with more than two components such as in the purification of biopharmaceuticals.

One of the disadvantages of using equilibrium stages is that they assume that equilibrium is reached instantly and therefore each stage is 100% efficient. In reality this is not the case as separations may take substantial time due to mass transfer resistances. A stage efficiency factor is usually applied to take into account such resistances. For example if a perfectly efficient extraction operation required 10 stages, in reality a physical system with 20 stages may be needed if the efficiency was 50%.

Non equilibrium rate based models

An alternative to using equilibrium stages of 100% efficiency is to use non-equilibrium models where mass transfer resistances are accounted for using forms of the Maxwell-Stefan equations (Krishna and Wesselingh, 1997). Such models are often used in the modelling of distillation. However, due to the increased complexity simulations can

take longer to calculate and converge to an appropriate numerical solution (Peng et al., 2002).

Thermodynamic description of aqueous two-phase systems

Thermodynamic equations are widely used in process modelling to describe phase equilibria, especially for vapour-liquid equilibria (Chen and Mathias, 2002). One of the simplest thermodynamic equations is the ideal gas law which can be used calculate the pressure of a gas at low pressure. To take into account non-ideal behaviour at higher pressures, cubic equations of state such as van der Waals and Redlich-Kwong have been used (Kontogeorgis and Folas, 2009). Cubic equations of state offer fast computational calculations; however, for complex systems with more than one component they offer poor performance which makes them unsuitable for description of liquid-liquid equilibria. Modelling of aqueous two-phase system is highly non-ideal as a result of the large amounts of polymer and salt present in solution.

The modelling of polymer solutions dates back to the work of Flory (1942) and Huggins (1942) who proposed a lattice type model which was later adapted by Scott (1949) to model ternary polymer-polymer-solvent mixtures. The UNIQUAC (universal quasichemical) activity coefficient equations were used by Kang and Sandler (1987) to model the phase equilibria of PEG-dextran aqueous two-phase systems. They found that the six adjustable UNIQUAC parameters could take on a range of values while still providing the same phase equilibria description as a result of parameter inter-correlation. To take into account the impact of polymer molecular weight without the need for estimating new parameters, Wu et al. (1998) developed a six parameter modified NRTL (non-random two-liquid) model which was used to describe PEG-dextran and PEG-salt aqueous two-phase systems. An alternative thermodynamic representation of ternary solutions of polymers has been presented by Edmond and Ogston (1968) using osmotic virial equations. King et al. (1988) and Zafarani-Moattar and Sadeghi (2001) later used the osmotic virial equations to describe PEG-dextran and PEG-phosphate aqueous two-phase systems, respectively. In both cases, the thermodynamic model calculations were fitted to experimental phase equilibria data by adjusting three binary interaction parameters.

An important factor that is not taken into account when using osmotic virial equations is the influence of electrolytes on phase equilibria. To take into account electrolyte interactions, the condition of electroneutrality must be met within each phase. Electrolyte interactions are usually taken into account by incorporating extensions or

modification of Debye-Hückel theory into a set of thermodynamic equations (Debye and Hückel, 1923). For example, Cameretti et al. (2005) extended SAFT (statistical associating fluid theory) equations to model vapour pressure and liquid densities of various electrolyte solutions. More recently, Naeem and Sadowski (2011) proposed a pePC-SAFT (polyelectrolyte perturbed chain-SAFT) model to describe phase equilibria of aqueous two-phase systems consisting of a polymer and polyelectrolyte which looks promising for the non-ideal aqueous two-phase systems.

A major hurdle for thermodynamics is the modelling of phase equilibria for biological materials such as proteins. This is because the size and complexity of proteins makes prediction of protein partitioning from first principles difficult. For example, Wu et al. (2017) have recently conducted a study where they attempted to describe the partitioning of ten proteins based on 57 unique protein features in 29 aqueous two-phase systems. They found that the unique features used to describe the partitioning varied between different types of aqueous two-phase systems therefore making it very difficult to predict how a protein will partition within a given system based on physical properties alone.

2.3.3 Simulation of aqueous two-phase systems

Simulations are useful for many tasks such as troubleshooting, process design, operation operator training etc. (Chen and Mathias, 2002, Biegler and Grossmann, 2004, Klatt and Marquardt, 2009). For liquid-liquid extraction, the majority of simulations are conducted using equilibrium stage models where phase equilibria are described using either empirical or thermodynamic equations. Black box models can also be used to simulate two-phase extraction, however, such models often require knowledge of process performance to be determined using more detailed models or appropriate experimentation. A black box model has been utilised by Torres-Acosta et al. (2016) for economic analysis of Uricase production using a PEG-ammonium sulphate aqueous two-phase systems.

To date the use of mechanistic models for simulation of aqueous two-phase systems has been fairly limited. Steady-state process simulations for aqueous two-phase systems have been first reported by Mistry et al. (1996) for the separation of α-amylase in a PEG-phosphate aqueous two-phase system. Similar modelling approaches have been used by other researchers to explore different aspects. For example, Huenupi et al. (1999) investigated the optimisation of a continuous ATPS process consisting of an initial extraction of α-amylase into the top phase followed by a back extraction into the

bottom phase. Simulation of a counter-current fractional extractor setup for monoclonal antibodies separation has been investigated by Ahmed Samatou et al. (2007). Simulation of a continuous multi-stage aqueous two-phase extraction for enzyme laccase and a monoclonal antibody have been presented by Prinz et al. (2014) and Mundges et al. (2015) respectively. Flowsheet optimisation was investigated for aqueous two-phase systems by Ahmad et al. (2010) to address the challenge of exploring the multiple different ways extraction stages can be arranged and connected together. Dynamic simulations are capable of providing greater information about a process than steady-state simulations as they are able to describe how a system change with time. Such information is important in understanding start-up and shut-down procedures as well as how to best control a given system. Simon and Gautam (2004) have investigated the use of dynamic models to better understand control of aqueous two-phase system processes, however, their approach resulted in more equations than variables resulting in an over specified system. To overcome this, they used a least squares minimisation approach to find an approximate numerical solution to the simulation problem. This approach allowed them to understand the likely dynamics of an aqueous two-phase system extraction process subject to process disturbances.

2.4 Concluding remarks

The literature review beforehand shows that aqueous two-systems are complicated systems for two main reasons. Firstly, partitioning of solutes such as proteins is influenced by many different factors resulting in a large experimental space. Secondly, the aqueous two-phase systems themselves are non-ideal mixtures composed of high concentration polymer and salt solutions. The former issue makes it difficult to develop aqueous two-phase systems experimentally, while the latter issue makes it difficult to model aqueous two-phase systems.

To address the large experimental space associated with aqueous two-phase systems, high-throughput techniques which include the use of microfluidic chips and automated liquid-handling robots have been developed (Bensch et al., 2007, Silva et al., 2014). The benefit of these techniques is that they allow more information to be obtained from limited experimental resources, this is particularly important in research and development of new drugs molecules. The main disadvantage of such techniques is that they are unlikely to fully replicate a scaled-up system which would be used in production, e.g. microfluidic chips operate almost exclusively in a laminar flow regime while mixing in a tank is likely to be turbulent. A potential way to address this issue is to derive fundamental properties which are unlikely to change and use them to build appropriate process models for simulations. To date an approach combining mechanistic modelling and high-throughput experimentation has not yet been reported.

The steady-state simulation of aqueous two-phase systems using mechanistic models has been reported by many authors (Mistry et al., 1996, Huenupi et al., 1999, Samatou et al., 2007, Prinz et al., 2014). Dynamic process models have also been reported using an over specified system of equations by Simon and Gautam (2004), which resulted in more equations than unknowns. To overcome this the authors used a least squares optimisation approach to approximate the solution to the system of equations. However, such an approach is not feasible when robust simulations need to be conducted on a regular basis. Therefore, there is a need to develop dynamic process models for aqueous two-phase systems which are capable of providing useful insights into process operation. In this work, this issue is addressed by developing a process model which is square i.e. the number of equations is equal to the number of unknowns.

2.5 Aim of this thesis by chapter

The aim of this thesis is to combine high-throughput experimental techniques with process simulations for the development of aqueous two-phase systems for the purification of non-antibody therapeutic proteins. Process simulations are conducted using mechanistic models that are capable of describing a variety of aqueous two-phase systems under dynamic operating conditions. High-throughput experimentation was conducted using automated liquid-handling robots to scale-down aqueous two-phase systems. The enzyme α-amylase was used as a model enzyme for acid alpha-glucosidase (GAA), a therapeutic protein for the treatment of Pompe disease that BioMarin pharmaceuticals was developing at the start of this project.

2.5.1 Chapter 3 will consider dynamic modelling of aqueous two-phase systems to quantify the impact of bioprocess design, operation and variability

The production of biopharmaceuticals is an inherently variable process due to the biological nature of upstream production. It is critical to understand the impact of this variability, especially as drug manufacturers are increasingly looking to take advantage of continuous technologies which will have to deal with dynamic process streams. To better understand the impact of this variability requires the development of a dynamic model which can take into account how process parameters change over time. In addition, aqueous two-phase systems can be deployed in a variety of modes of operation such as batch, continuous counter-current and multi-cycle batch extraction. To explore all of these configurations effectively requires a flexible process model. Chapter 3 addresses these issues by presenting a dynamic model to describe aqueous two-phase systems. The dynamic model presented in this work is a fully defined model consisting of the same number of equations as variables. The model is used to explore various modes of operation such as multi-cycle batch and continuous counter-current extraction, as well as the impact of process variability on bioprocess design, operation and performance. All of which provide useful information for researchers looking to use aqueous two-phase systems for batch or continuous manufacturing processes.

2.5.2 Chapter 4 will consider quantification of uncertainty in phase diagrams for estimation of interaction parameters

Effective process simulations for aqueous two-phase systems require parameters which represent the experimental system of interest. These parameters are often determined from appropriate experimental data. To use thermodynamic equations in process models for specific aqueous two-phase systems requires estimation of interaction parameters

from experimental phase equilibria data. The group of Hubbuch have presented experimental methods to determine phase diagrams for aqueous two-phase systems in a high-throughput manner using automated liquid handling robots (Bensch et al., 2007, Oelmeier et al., 2011, Diederich et al., 2013, Amrhein et al., 2014). Such methods involve first locating the location of the binodial curve which is then followed by determining tie-lines using the inverse lever arm rule. Therefore, uncertainty in the phase diagram is a combination of uncertainty in the actual binodial curve location as well as the uncertainty in the phase ratio measurements when using the inverse lever arm rule to determine tie-lines. To date, there have not been any methods to quantify such uncertainties. In Chapter 4, this issue is addressed by developing a systematic protocol for determining the uncertainty in the experimental phase diagrams generated using high-throughput experimentation. The developed protocol is used to calculate uncertainty due to experimental design as well as uncertainty in experimental measurements. The calculated levels of uncertainty were used in parameter estimation protocols to determine uncertainty in estimated thermodynamic interaction parameters which are used to describe phase equilibria of aqueous two-phase systems.

2.5.3 Chapter 5 will consider high-throughput screening and simulation of α-amylase and myoglobin separation

According to Soares et al. (2015), one of the main challenges in aqueous two-phase system process development is the lack of predictive design. To address this issue, the process modelling techniques developed in Chapter 3 are combined with the high-throughput experimental techniques developed in Chapter 4 to develop an aqueous two-phase extraction process. A case study separation using α-amylase and myoglobin is considered. The simulations are used to investigate the impact of various operating conditions, such as tie-line length and number of separation stages, have on process performance and the subsequent impact on equipment choice. The combined modelling and high-throughput experiment approach is shown to provide valuable information for researchers conducted early stage process development for aqueous two-phase systems.

Chapter 3: Dynamic modelling of aqueous two-phase systems

3.1 Introduction

To cope with increasing product and patient demands for biopharmaceuticals, drug developers are looking to take advantage of continuous manufacturing technologies for greater efficiencies. A continuous process must be able to consistently produce product of a high quality throughout its entire period of operation, and regardless of process disturbances and changes. This is, however, a very difficult task to achieve as the biological complexity of the cells used in upstream culture/fermentation is inherently variable. For instance, Valente et al. (2015) recently showed that as Chinese hamster ovary (CHO) cells age, the profile of hard to remove endogenous host cell proteins (e.g. contaminants with similar separation behaviour to the product) excreted into the media changes. In addition to such inherent variability, there are other sources of variability or disturbances that must also be evaluated, such as equipment failure, human error, contamination etc. Understanding the impact of such changes on whole bioprocess performance is extremely important as product quality could be compromised if the process is not sufficiently robust. Unfortunately, it is often costly and very time consuming to evaluate all sources of variability or disturbances experimentally.

Soares et al. (2015) recently conducted a SWOT (strengths, weaknesses, opportunities and threats) analysis of aqueous two-phase extraction. A key weakness identified was the lack of predictive design, as well as expertise, in operation of aqueous two-phase processes. Use of effective process models capable of capturing the dynamic behaviour of these systems can help address this issue. However, as highlighted before, the use of dynamic process models for aqueous two-phase system process development has been limited to the work of Simon and Gautam (2004), even though dynamic models offer greater insights than steady-state models as they are capable of simulating a wider variety of scenarios such as process start-up and shut-down.

In this chapter, these issues are tackled by the development of a dynamic stage-by-stage equilibrium model that can be used to simulate aqueous two-phase system (ATPS) processes under a variety of configurations and operating policies. Liquid-liquid equilibria data available from literature is used to estimate interaction parameters for the thermodynamic equations used in the process model. The approach permits a fast, yet

systematic, investigation of both process design and process operation for the separation of biomolecules using aqueous two-phase extraction. The approach is demonstrated by considering dynamic continuous counter-current, as well as multi-cycle batch, modes of operation for a case study involving the separation of enzyme α-amylase from impurities in a PEG 4,000-phosphate aqueous two-phase system (ATPS) in the presence of NaCl.

3.2 Mathematical methods

The modelling approach used in this work is illustrated in Figure 3 - 1. A general single stage dynamic equilibrium process model is used to describe both the multi-cycle batch and the counter-current mode of operation as the fundamental physical and chemical behaviour is the same in each mode. The system consists of a number of components:

1) Those characterising the two phases, either two hydrophilic polymers, or a polymer and a salt, together with water.
2) The desired protein and other biological material (in the following denoted impurities) from the upstream fermentation stage.

Chemical potential is used to describe the phase equilibria between the two aqueous phases and the thermodynamic parameters required to describe this are obtained from experimental data. Empirical correlations are used to represent the more complex behaviour of protein partitioning, as currently no thermodynamic prediction methods exist which can accurately describe this behaviour. The use of correlations to describe the behaviour of the protein and that of the impurity, reduces the complexity in describing the system when considering mixtures composed of a single desired protein product plus hundreds of impurities, as is often the case when dealing with primary recovery after cell culture/fermentation. The overall equation system is solved simultaneously using gPROMS ModelBuilder 4.1 (Process Systems Enterprise, 2017).

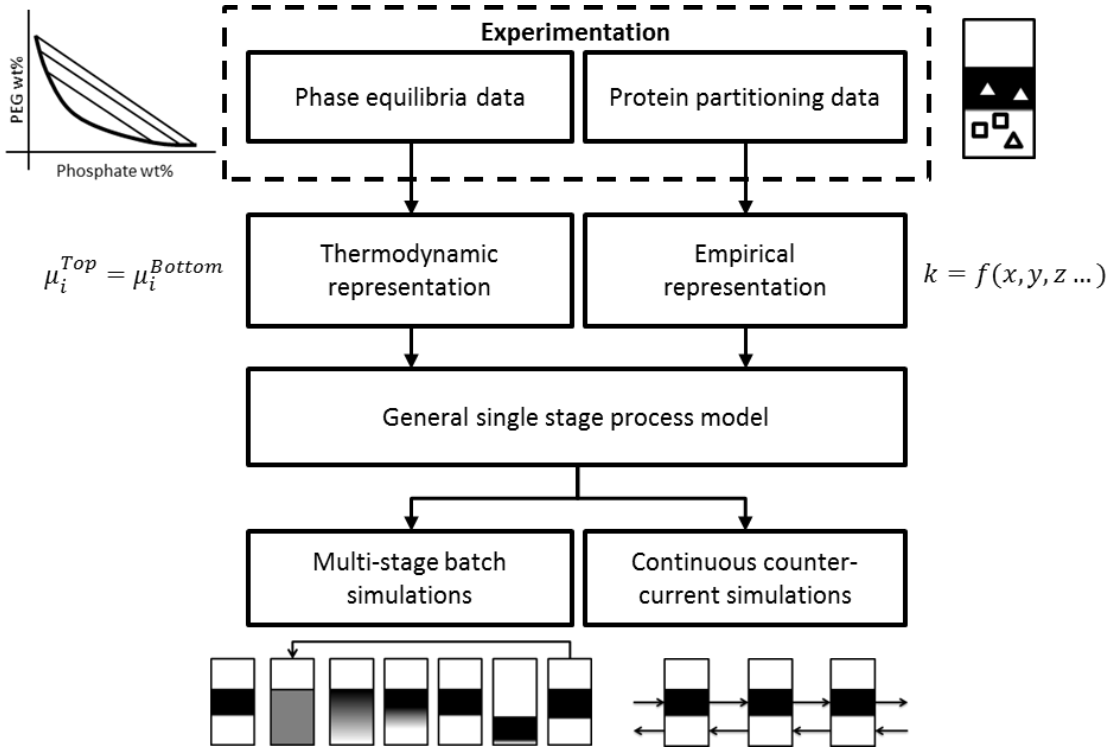


Figure 3 - 1: Overview of modelling framework. Experimental data is represented mathematically using appropriate equations.

3.2.1 Assumptions

In developing the process model, a number of assumptions had to be made. Most of these assumptions are similar to those made when considering other stage-wise separation methods such as distillation or absorption. The key assumptions are:

1. Major impurities were treated as a single component with a lumped partition coefficient. In reality, each impurity would have its own partition coefficient, however, it would be very difficult, and very time consuming, to determine all of these experimentally. In early stage separations such as after cell harvest, knowledge of detailed component behaviour is not necessary for most investigations, however, for later stage separations where the number of components is greatly reduced such information is very important.
2. An energy balance was not considered as the process is considered to be isothermal and adiabatic (Rosa et al., 2009, Mistry et al., 1996). This assumption is reasonable for biological separations as they are usually operated at ambient conditions where heat transfer is negligible.
3. It was assumed that equilibrium is reached instantly in each stage and for all components, i.e. each stage is 100% efficient. This assumption is reasonable for the multi-cycle batch model where in reality one would mix the phases thoroughly and then wait for them to settle. For the continuous counter-current model, the assumption of 100% stage efficiency is not realistic since equilibrium may not be reached in each stage due to the counter-current movement. However, as for other stage-wise processes such as distillation and absorption, the assumption of 100% efficiency can be made and combined with an appropriate stage efficiency. For instance, a 20 stage system with a 50% stage efficiency can be described as a 10 stage equilibrium system. Actual stage efficiency would have to be determined experimentally and would depend on many factors such as phase settling time, mass transfer resistance, strength of partitioning, equipment setup etc. The advantage of using a model which assumes 100% efficiency is that if separation performance is poor in an ideal system, then separation performance can be assumed to be worse in a real system therefore allowing potentially expensive experiments to be avoided so that resources can be directed to explore alternative separation strategies.
4. Perfect mixing was assumed, meaning that the composition of the phase outlet streams is assumed to be the same as the phase composition inside the stage.

Again, this is a common assumption when considering stage-wise separation models.

5. It was assumed that the presence of proteins and other biological materials do not significantly affect the phase equilibria or the density of the two aqueous phases as the relative amount of proteins and biological material is very small (~0.2wt%) compared to the other components.
6. It was assumed that the salt, if present, is distributed equally between top and bottom phases. In reality, the cations and anions of NaCl are likely to partition to different extent between the phases (Andrews et al., 2005b) therefore this is reasonably large assumption. However, a detailed model which takes into account partitioning of specific anions and cations was not considered in this work as there is unfortunately currently no experimental data in the literature to support such a model and determining such data was beyond the scope of this work.
7. For the continuous counter-current system, the mass holdup and phase ratio on each stage was assumed to be constant.

3.2.2 Experimental data

Experimental phase equilibria and protein partitioning data for the case study was obtained from Mistry et al. (1996). Their study examined the partitioning of α-amylase and impurities in a PEG 4,000-phosphate ATPS in the presence of NaCl. Binary interaction parameters ($\beta_{i,i}$ and $\beta_{i,j}$) required in this model were not considered by Mistry et al. (1996), hence were estimated in this work but based on their data. This study was chosen primarily for the completeness of experimental data which includes both partitioning data as well as phase diagrams including tie-lines, furthermore, this reduced the need to mix and match experimental data from different literature sources. The required data points were extracted from the experimental figures provided by Mistry et al. (1996) using the digitise tool in OriginPro 8.6 (OriginLab, 2012). The original data can be found in Appendix 7.11.

3.2.3 Process model

The process model described next is a dynamic equilibrium-stage process model which is simple, yet flexible, and is based on a single stage which can be used to describe both batch and continuous systems. To describe multi-stage continuous systems, a number of single stages are connected together to form a counter-current system equivalent to the approach used for distillation and absorption. More complicated scenarios, such as recycling of materials can also be considered, however, this is not covered in this work.

Mass balances

A simple process model with two inlets and two outlets was used to model both multi-cycle batch and counter-current modes of operation as outlined previously. A single stage as shown in Figure 3 - 2A was used for multi-cycle batch simulations. For continuous counter-current simulations, multiple single stages were connected together as shown in Figure 3 - 2B.

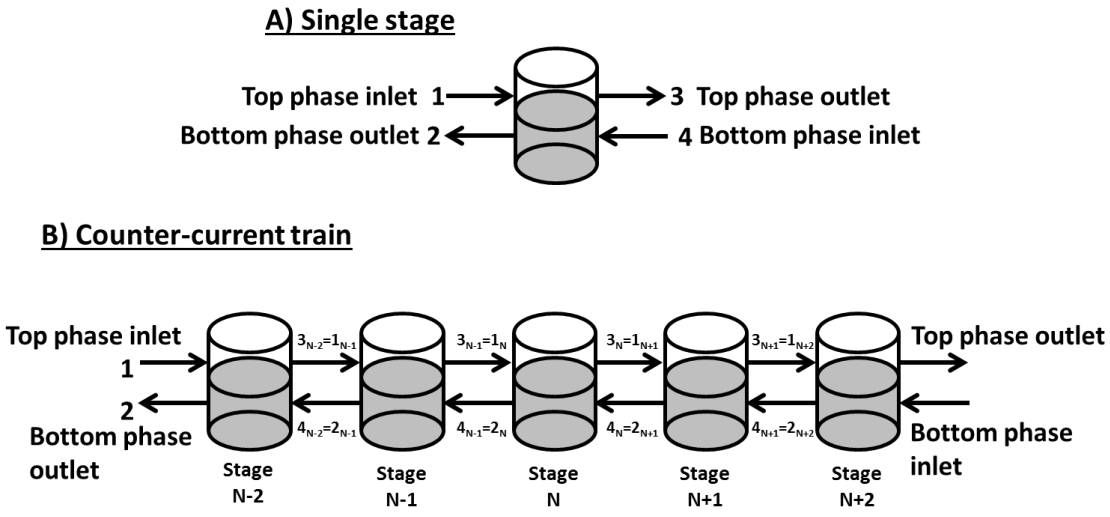


Figure 3 - 2: A) Set up of a single ATPS separation stage, B) a continuous counter-current ATPS separation system.

The following equations were considered for all components where *NC* is the number of components: Eq.(3.1) is an overall mass balance around the stage (stage *n*) with two inlet streams (streams 1 & 4) and two outlet streams (streams 2 & 3). Eq.(3.2) is the sum of the mass holdup of each component i in stage n in each phase which must add up to the total mass holdup of component i. Eq.(3.3) and Eq.(3.4) are a consequence of Assumption 4 which states that the outlet compositions are equal to the phase compositions inside the stage, i.e. perfect mixing, as well as the stream indexing

convention used in this work[3]. Eq.(3.5)-(3.7) are definitions of the mass fraction in terms of mass of components in the top, bottom and overall, respectively.

For a specific stage *n*:

$$\frac{dM_{i,overall,n}}{dt} = F_{1,n}x_{i,1,n} - F_{2,n}x_{i,2,n} - F_{3,n}x_{i,3,n} + F_{4,n}x_{i,4,n} \quad for\ i = 1 \dots NC\ for\ all\ n \qquad (3.1)$$

$$M_{i,overall,n} = M_{i,top,n} + M_{i,bottom,n} \quad for\ i = 1 \dots NC\ for\ all\ n \qquad (3.2)$$

$$x_{i,3,n} = x_{i,top,n} \quad for\ i = 1 \dots NC\ for\ all\ n \qquad (3.3)$$

$$x_{i,2,n} = x_{i,bottom,n} \quad for\ i = 1 \dots NC\ for\ all\ n \qquad (3.4)$$

$$x_{i,top,n} \times \sum_{i}^{NC} M_{i,top,n} = M_{i,top,n} \quad for\ i = 1 \dots NC\ for\ all\ n \qquad (3.5)$$

$$x_{i,bottom,n} \times \sum_{i}^{NC} M_{i,bottom,n} = M_{i,bottom,n} \quad for\ i = 1 \dots NC\ for\ all\ n \qquad (3.6)$$

$$x_{i,overall,n} \times \sum_{i}^{NC} M_{i,overall,n} = M_{i,overall,n} \quad for\ i = 1 \dots NC\ for\ all\ n \qquad (3.7)$$

Initial conditions

Since the model is dynamic, initial conditions must be specified. Different equations can be used to describe the initial state, and in this work, the initial holdup in stage *n* of each component is specified as:

$$M_{i,overall,n} = M^{0}_{i,overall,n} \quad for\ i = 1 \dots NC\ at\ t = 0\ for\ all\ n \qquad (3.8)$$

Additional equations for the continuous counter-current model

Additional equations are needed to link together the different stages in the counter-current model. Assuming the stages are numbered from left to right as shown in Figure 3 - 2B, then the top phase outlet of stage *n* is the inlet of stage *n+1*, whereas the bottom phase outlet of stage *n+1* is the inlet of stage *n* (*NS* is the total number of stages):

$$F_{3,n} = F_{1,n+1} \quad for\ n = 1 \dots NS - 1 \qquad (3.9)$$

[3] In this work streams 1 and 3 are inlets and outlets to the top phase respectively, whilst streams 4 and 2 are inlets and outlets to the bottom phase respectively. Other numbering conventions can be utilised as long as they are consistent in all equations.

$$x_{i,3,n} = x_{i,1,n+1} \qquad for\ n = 1 \dots NS-1,\ \ i = 1 \dots NC \qquad (3.10)$$

$$F_{4,n} = F_{2,n+1} \qquad for\ n = 1 \dots NS-1 \qquad (3.11)$$

$$x_{i,4,n} = x_{i,2,n+1} \qquad for\ n = 1 \dots NS-1,\ \ i = 1 \dots NC \qquad (3.12)$$

In the counter-current model, overall mass hold-up and phase ratio ($\emptyset_n$) in each stage are kept constant as stated by Assumption 7. The phase ratio is defined as the mass of the top phase to the mass of the bottom phase in each stage as described by Eq.(3.13) below. Eq.(3.14) ensures that the phase ratio is constant as it keeps the ratio of $F_{2,n}$ to $F_{3,n}$ equal to the phase ratio. Eq.(3.15) describes the assumption of constant stage holdup.

$$\emptyset_n \times \sum_{i}^{NC} M_{i,bottom,n} = \sum_{i}^{NC} M_{i,top,n} \qquad for\ n = 1 \dots NS \qquad (3.13)$$

$$F_{2,n} \times (\emptyset_n + 1) = F_{1,n} + F_{4,n} \qquad for\ n = 1 \dots NS \qquad (3.14)$$

$$F_{1,n} + F_{4,n} - F_{2,n} - F_{3,n} = 0 \qquad for\ n = 1 \dots NS \qquad (3.15)$$

Equilibrium descriptions for the case study

Equilibrium relationships are needed that describe how the components in each phase relate to each other.

Partition coefficients and density correlations

The system considered in this case study consists of six components:

1. PEG 4,000 (component 1)
2. phosphate (component 2)
3. water (component 3)
4. NaCl (component 4)
5. α-amylase (component 5 - protein)
6. bulk impurities (component 6)

PEG 4,000, phosphate, NaCl and water establish the ATPS into which the α-amylase, and the impurities, are distributed. The partition coefficients of the protein, α-amylase, and the impurities were empirically correlated to the NaCl mass fraction as this is the factor which affects partitioning characteristics the strongest. The correlations are based on the experimental data of Mistry et al. (1996). A polynomial function was used to correlate the α-amylase partition coefficient instead of the sigmoidal Boltzman function used by Mistry et al. (1996), as a polynomial was found to be a better fit when conducting curve fitting as the fit obtained by Mistry et al. (1996) could not be reproduced as their parameters were not provided:

$$\ln(K_{\alpha-amylase,n}) = 0.04874\left(100 \times x_{NaCl,overall,n}\right)^2 + 0.3043 \times 100 \times x_{NaCl,overall,n} - 2.731 \qquad for\ n = 1 \dots NS \quad (3.16)$$

$$ln(K_{impurities,n}) = 0.01412 \times 100 \times x_{NaCl,overall,n} - 0.01341 \qquad for\ n = 1 \dots NS \quad (3.17)$$

Experimental data used in fitting partition coefficient parameters can be found in Appendix 7.11 alongside a figure comparing fit of Eq.(3.16) to experimental data points. The equilibrium relationship for proteins is given by Eq. (3.18) below where $K_{i,n}$ is the partition coefficient for component *i* in stage *n*. The equilibrium relationship for NaCl is given by Eq.(3.19) and is a result of Assumption 6 which states that the salt (NaCl) is distributed equally between the two phases.

$$C_{i,top,n} = K_{i,n} C_{i,bottom,n} \qquad for\ i = 5 \dots 6, \qquad for\ n = 1 \dots NS \quad (3.18)$$

$$x_{i,top,n} = x_{i,bottom,n} \qquad for\ i = 4, \qquad for\ n = 1 \dots NS \quad (3.19)$$

The density (in kg m^{-3}) of each phase was correlated to the mass fractions of PEG 4,000, phosphate and NaCl as determined by Samatou (2012) who derived the correlation from experimental data which is shown in Appendix 7.12:

$$\rho_{q,n} = 1000 + 176 \times x_{PEG\ 4{,}000,q,n} + 888 \times x_{phosphate,q,n} + 808 x_{NaCl,q,n} \quad \begin{matrix} for\ q = top, bottom \\ for\ n = 1 \ldots NS \end{matrix} \tag{3.20}$$

Thermodynamic framework

In this work, the equations used to describe the phase equilibria are based on osmotic virial expansions. These equations are used to calculate the chemical potential (μ_i) of the polymer (component 1), phosphate (component 2) and water (component 3) in terms of molality (m_i) of polymer and salt, which allows the calculation of the phase equilibria compositions. The equations are based on those reported by Edmond and Ogston (1968) and Zafarani-Moattar and Sadeghi (2001) which they used to describe PEG-dextran and PEG-salt (citrate or phosphate) aqueous two-phase systems, respectively:

$$\mu_1 = \mu_1^0 + RT(\ln(m_1) + \beta_{1,1} m_1 + \beta_{1,2} m_2) \tag{3.21}$$

$$\mu_2 = \mu_2^0 + RT(\ln(m_2) + \beta_{2,2} m_2 + \beta_{1,2} m_1) \tag{3.22}$$

$$\mu_3 = \mu_3^0 - RTV_3(m_1 + m_2 + \frac{\beta_{1,1}}{2} m_1^2 + \frac{\beta_{2,2}}{2} m_2^2 + \beta_{1,2} m_1 m_2) \tag{3.23}$$

At equilibrium, the chemical potential of each component in each phase is the same:

$$\mu_i^{top} = \mu_i^{bottom} \qquad for\ i = 1 \ldots 3 \tag{3.24}$$

The μ_i^0 terms in Eq.(3.21)-(3.23) refer to the chemical potential in the standard states. For the polymer and salt this is the hypothetical solution of 1 unit molality, whereas for water it is as pure water. These terms cancel out due to Eq.(3.24). The interaction terms ($\beta_{i,i}$ and $\beta_{i,j}$) are constant terms that describes the interaction between the polymer and salt.

For the system considered in this work, the amount of PEG in the bottom phase was found to be almost negligible. Therefore, for numerical robustness in the numerical solution its mass fraction was assumed to be constant at a very small value in this phase ($x_{1,bottom} = x_{PEG,bottom} = 0.0001$). This assumption will not have any impact on the results, but is required to ensure stability of the mathematical solution.

Although more detailed thermodynamic models could have been used (e.g. those that take into account electrolyte interactions and more), a simple model (osmotic virial) was

chosen for ease of implementation and as the model will later be used for optimisation purposes where solution time becomes an issue. The advantages and disadvantages of this approach are discussed in the conclusions.

3.2.4 Simulation methodology

All the equations were solved simultaneously as one equation set using gPROMS (Process Systems Enterprise, 2017). Although the model may appear simple, the presence of the equilibrium conditions imposed by Eq.(3.24) means that initialisation is difficult as good initial estimates are needed, however, this was tackled using the initialisation protocol of gPROMS. Further details on initialisation protocol used is provided in Appendix 7.8. The DASOLV solver within gPROMS was used which is based on a variable time step/variable Backward Differentiation Formula (BDF).

A key focus of this work is the study of how upstream variability may have an impact on the performance of the separation section, in this case, the aqueous two-phase system (ATPS). In this work, this variability of process parameters such as NaCl wt% is described by an average value and a standard deviation such that x_{NaCl}~N(mean, standard deviation). The feed compositions of the input streams were therefore varied using the normal probability distribution assignment in gPROMS which picks a random value of a parameter or variable given an average value and the corresponding standard deviation for a specific process parameter.

3.2.5 Parameter estimation of interaction parameters

Although the description of the PEG 4,000-phosphate phase equilibria data was based on the work by Mistry et al. (1996), their empirical correlations were not considered because firstly, they required too many fitted parameters which reduces the statistical confidence in each parameter, and secondly, the use of thermodynamics to describe phase equilibria allows for a model based on greater fundamental principles, hence Eq.(3.21)-(3.23) were used instead. The interaction parameters of these equations, ($\beta_{i,i}$ and $\beta_{i,j}$), therefore needed to be estimated using the experimental tie-line data from Mistry et al. (1996). The estimation of the interaction parameters, $\beta_{i,i}$ and $\beta_{i,j}$, was conducted using the parameter estimation entity in gPROMS. This entity estimates parameters based on Maximum Likelihood which takes into account experimental variation. Interaction parameter values were scaled according to the initial guesses within gPROMS. Since an experimental variation was not provided by Mistry et al. (1996), a standard deviation of 2wt% for phase equilibria compositions was assumed in

order to illustrate the methodology. This is a relatively conservative error as PEG and salts can be quantified accurately using techniques such as refractive index and flame photometry (Planas et al., 1997, Zafarani-Moattar et al., 2004). In addition, the amount of PEG in the bottom phase and salt in the top phase reported by Mistry et al (1996) is less than 2wt% which suggests that assuming a standard deviation of 2wt% is likely larger than the true experimental error. The NLPSQP (nonlinear programming sequential quadratic programming) solver in gPROMS was used for the parameter estimation. Further details of this solver can be found in Appendix 7.5.

3.3 Results and discussion

This work presents an approach based on a dynamic stage-by-stage equilibrium model that can be used to simulate aqueous two-phase extraction (ATPE) processes under a variety of configurations and operating policies. The approach permits a fast, yet systematic, investigation of both process design and process operation for the separation of biomolecules using aqueous two-phase extraction. The approach presented is demonstrated using a case study involving the separation of enzyme α-amylase from impurities in a PEG 4,000-phosphate aqueous two-phase system (ATPS) in the presence of NaCl.

In addition, the quality of the phase equilibria descriptions determined using the estimated interaction parameters for Eq.(3.21)-(3.23) is also examined by taking into account potential experimental error. This is an important first step of any model development as it is not possible to model ATPS processes, or any separation process for that matter, without accurate representation of the phase equilibria.

Two different types of aqueous two-phase extraction operations are considered to show how the model can be used to describe the dynamic behaviour of each process. First a batch extraction process operated over multiple cycles is investigated to show the capabilities of dynamic modelling for even a simple process. Next, the impact of upstream variability is investigated for a continuous counter-current extraction process. This is especially important given the demand for quality by design in manufacturing processes which requires the specification of a robust design space.

3.3.1 Estimation of $\beta_{i,j}$ parameters

The top and bottom phase compositions of PEG 4,000 (component 1) and phosphate (component 2) in their respective phases at different NaCl concentrations were obtained from the experimental tie-lines reported by Mistry et al. (1996). The feed composition was taken to be the mid-point of the tie-line of each respective tie-line since this information was not provided. The process model described previously was simplified to a single steady-state extraction stage consisting of one inlet and two outlet streams using Eqs.(3.1)-(3.7), Eq.(3.19) and Eqs.(3.21)-(3.24) for parameter estimation purposes. This simplified extraction stage model was used to determine equilibrium compositions in the top and bottom phase given inlet stream feed compositions. The parameter estimation entity in gPROMS (Process Systems Enterprise, 2017) is capable of determining the uncertainty in the fitted parameters (standard deviation) given information regarding the experimental uncertainty in the data used for the fitting. This accumulated uncertainty is important to consider (as a small uncertainty may in some cases still lead to a large uncertainty in the fitted parameter values) but is unfortunately often ignored. Since the actual experimental uncertainty was unknown, a conservative standard deviation of 2wt% in equilibrium phase compositions was assumed as described before.

The resulting estimated values of the interaction parameters are reported in Table 3 - 1. The PEG 4,000-PEG 4,000 interaction parameter ($\beta_{1,1}$) has a standard deviation (5.8) which is in the same order of magnitude as the parameter value (-10.5). This suggests that there is insufficient experimental data, or that the assumed uncertainty in the data is too high, in order to estimate this parameter accurately. The PEG 4,000-phosphate and phosphate-phosphate interaction parameters, $\beta_{1,2}$ and $\beta_{2,2}$, respectively, have much lower standard deviations, which means that the uncertainty associated with these parameters is lower. A more thorough investigation into the impact of uncertainty in estimated interaction parameters is conducted in Chapter 4.

Parity plots were used to compare model predictions of tie-line compositions (y-axis) to experimental tie-line compositions (x-axis) for the different NaCl concentrations (Figure 3 - 3). A perfect model fit would have all points lie on the line y = x. Parity plots A and B in Figure 3 - 3 for 0wt% NaCl show a poor model fit. This is likely due to the fact that experimental equilibrium compositions in systems containing NaCl are similar (see Appendix 7.11 for data from Mistry et al. (1996)) and therefore the parameter estimation solver is minimising the maximum likelihood objective function

more by fitting to the experimental phase diagrams containing more than 0wt% NaCl. In addition, the model used in this work does not take into account electrolyte interactions (as a result of Assumption 6) which is likely to result in an under fitting model across a wide range of NaCl concentrations. In the model, NaCl does not explicitly interact with PEG 4,000 and phosphate, and the presence of NaCl only reduces the overall mass fraction of water, which in turn increases the molality of PEG 4,000 and phosphate. Nevertheless, the changes in PEG 4,000 and phosphate molality due to NaCl presence are still not significant enough to describe a PEG 4,000-phosphate aqueous two-phase system from 0wt% to 10wt% NaCl. However, for systems containing NaCl concentrations above 2wt%, the parity plots show that the phase equilibria predictions are sufficiently accurate for process simulations as the experimental tie-lines reported for these systems by Mistry et al. (1996) are very similar. If a good model fit is required at 0wt% NaCl, then parameter estimation could be conducted using tie-line data of only the 0wt% NaCl phase diagram, however, this is undesirable since it would result in two sets of parameters.

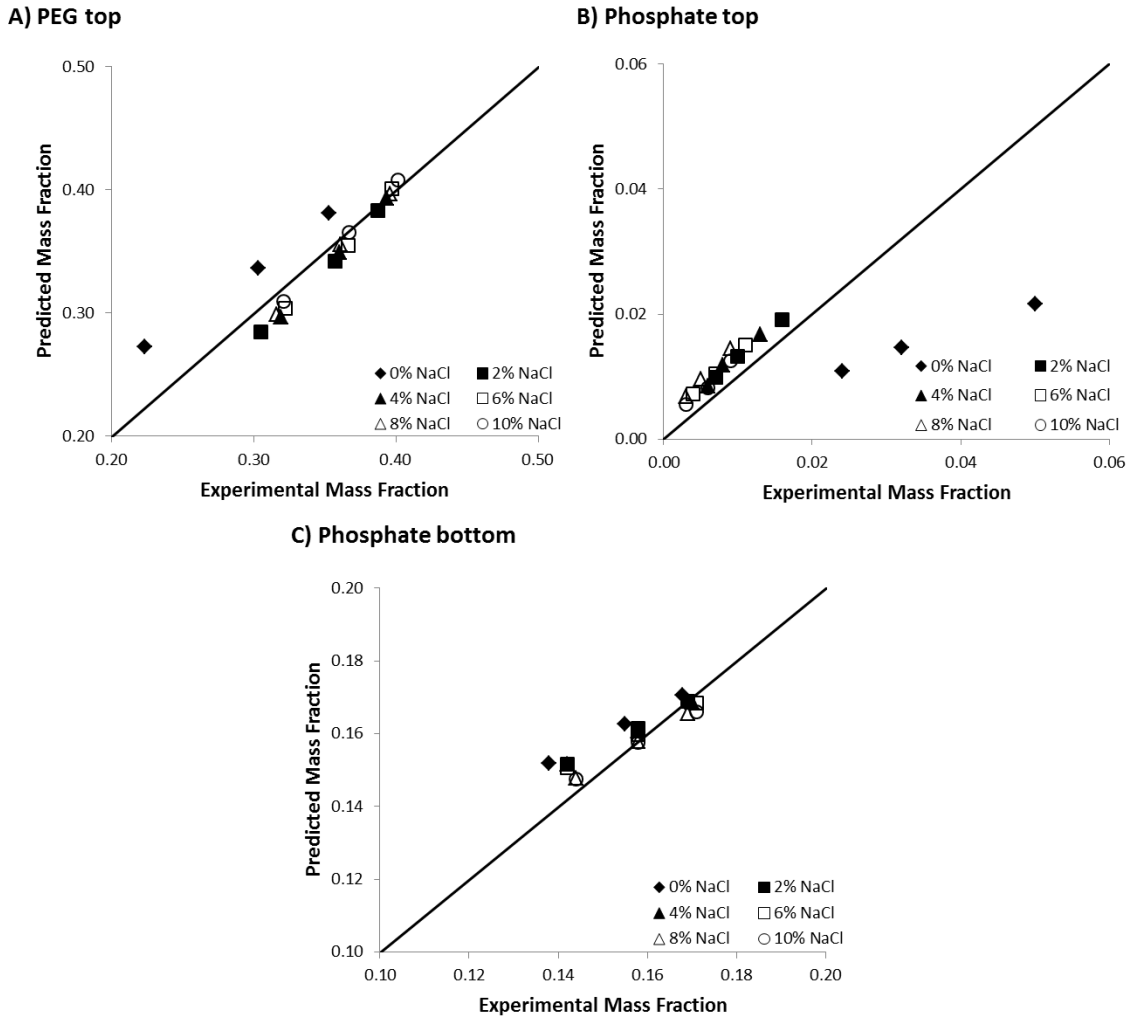


Figure 3 - 3: Parity plots for a PEG 4,000-phosphate aqueous two-phase system from 0wt% to 10wt% NaCl. Experimental data were obtained from Mistry et al. (1996). A) PEG top phase mass fraction. B) Phosphate top phase mass fraction. C) Phosphate bottom phase mass fraction.

Table 3 - 1: Estimated interaction parameters ($\beta_{i,j}$) for Eq.(3.21)-(3.23) in a PEG 4,000 (component 1)-phosphate (component 2) aqueous two-phase system.

$\beta_{i,j}$ $(kg\ mole^{-3})$	PEG 4,000		Phosphate	
	Value	Standard Deviation	Value	Standard Deviation
PEG 4,000	-10.5	5.8	6.1	1.5
Phosphate	6.1	1.5	-1.1	0.1

The PEG 4,000 mass fraction in the bottom phase is usually very small, therefore a constant mass fraction of 0.0001 was assumed in the simulations to increase the numerical robustness as explained earlier, hence no parity plot is shown for this concentration. The presence of NaCl can be taken into account by extending the chemical potential equations (Eq.(3.21)-Eq.(3.23)) to a 4 component system; however, this would require more interaction parameters to be estimated, and sufficient

experimental data would need to be available to ensure statistically significant parameters values are found. As for all parameter estimations, there needs to be a balance between the number of parameters that are to be estimated and the amount of useful experimental data available for parameter estimation. This is because, as the number of parameters is increased, the ability to estimate parameters with low uncertainty decreases.

The uncertainty in the estimated parameter values can be reduced in two ways:

1. Reducing the uncertainty in experimental measurement values, in this case 2wt% error was assumed. This can be achieved using sufficiently accurate and precise analytical techniques such as HPLC.
2. Increasing the number of experimental data points by using Design of Experiments, or model-based experimental design, to determine which experiments should be conducted next so that the most information can be obtained using the fewest possible resources.

Often, reducing the uncertainty in measurements is difficult as it may require development of new experimental techniques which could be costly. Model based experimental design is useful if a model is known to be a good representation of experimental data, however, often one still needs an initial set of experimental data to estimate initial parameters (Dechambre et al., 2014, Tulsyan et al., 2012). Once an initial set of parameters are estimated, the model is then used to design new experiments which target areas which will provide the greatest information about the parameters to be re-estimated. For example, the sampling points during a chemical reaction could be optimised to take many samples during the times the model predicts there will be the greatest change in reaction rates. The result is that the subsequent experimental data yields more useful information which allows parameters to be estimated to a lower degree of uncertainty provided that the model is an accurate representation of the system being studied.

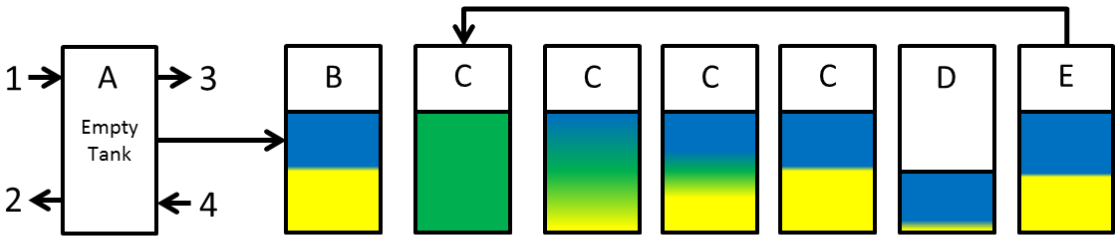
3.3.2 Multi-cycle batch extraction simulations

The process model developed in this work can be used to describe the operation of a batch extraction process for an ATPS. The process is simple to operate, and if the purity required is not obtained directly in the first batch, the process can be repeated over another cycle by removing the solvent (bottom) phase, reloading the tank with pure

solvent, and repeating the separation as shown in Figure 3 - 4. The separation achieved will thus depend not only on the salt concentration and the phase ratio, but also on the number of cycles considered.

The operation of a batch operated ATPS over a number of extraction cycles is considered first, as most biopharmaceutical operations are currently operated either batch-wise or semi-batchwise. The streams profiles are shown in Table 3 - 2 and are based on Figure 3 - 2A. Streams 1 and 4 feed the PEG rich top phase and phosphate rich bottom phase, respectively. The phase compositions chosen in Table 3 - 2 were chosen as they are close to the shortest tie-line in the experimental data presented by Mistry et al. (1996). Typically for high molecular weight species operating at low tie-line lengths is recommended (Benavides and Rito-Palomares, 2008). A feed concentration of 1kg m^{-3} was chosen as most upstream production processes are capable of expressing therapeutic proteins at concentrations greater than 1kg m^{-3}. NaCl wt% was not explored below 6wt% as the partition coefficient of α-amylase is less than 1 at NaCl concentrations less than 5wt%.

The influence of the phase ratio (mass of top phase to mass of bottom phase) on yield and purity was investigated by varying the mass flow rate of stream 4 during initial loading. A detailed schematic of the operating protocol used can be found in Figure 3 - 4. The responses in Figure 3 - 5 and Figure 3 - 6 at 30 minute intervals correspond to the extraction cycles outlined in the operating protocol. The flat sections in Figure 3 - 5 and Figure 3 - 6 consisting of ~ 7 data points each correspond to the time associated with mixing (10 minutes) and settling (15 minutes). The steps in Figure 3 - 5 and Figure 3 - 6 which look like discrete steps correspond to when the fresh bottom phase is being fed into the multi-cycle batch extraction process. Due to the way the model was programmed into gPROMS there are not very many points during this transition step. This issue can be solved using a while loop rather than a continue until loop within gPROMS to find the condition when the total hold up within the tank exceeds 300kg.



A. Initially the tank is empty.
B. Streams 1 & 4 are turned on until the mass in the tank ($M_{overall}$) is greater than 300 kg. (Stream 1 contains the protein).
C. Allow 10 minutes for mixing (t_{mixing}) and 15 minutes for settling ($t_{settling}$). (Mixing and settling is not modelled).
D. Drain the bottom phase (containing protein) via stream 2 (1kg s^{-1}) until the mass in the bottom phase (M_{bottom}) is less than 5 kg.
E. Fill with fresh phosphate rich bottom phase via stream 4 until the mass in the tank ($M_{overall}$) is greater than 300 kg. Go to step C.

Repeat steps C to E for multiple extraction cycles.

Figure 3 - 4: Schematic of operating protocol for multi-cycle batch extraction for case study.

Table 3 - 2: Operating conditions considered for multi-cycle batch extraction of α-amylase (component 5) with impurities (component 6) in a PEG 4,000 (component 1)-phosphate (component 2) aqueous two-phase system.

Stream	Operating flow rate ($kg\ s^{-1}$)	α-amylase conc ($kg\ m^{-3}$)	Impurity conc ($kg\ m^{-3}$)	PEG (wt%)	Phosphate (wt%)	NaCl (wt%)
1	1.00	1.00	1.00	27.38	1.77	6 to 10
4	1 to 4.00	N/A	N/A	0.01	14.46	6 to 10

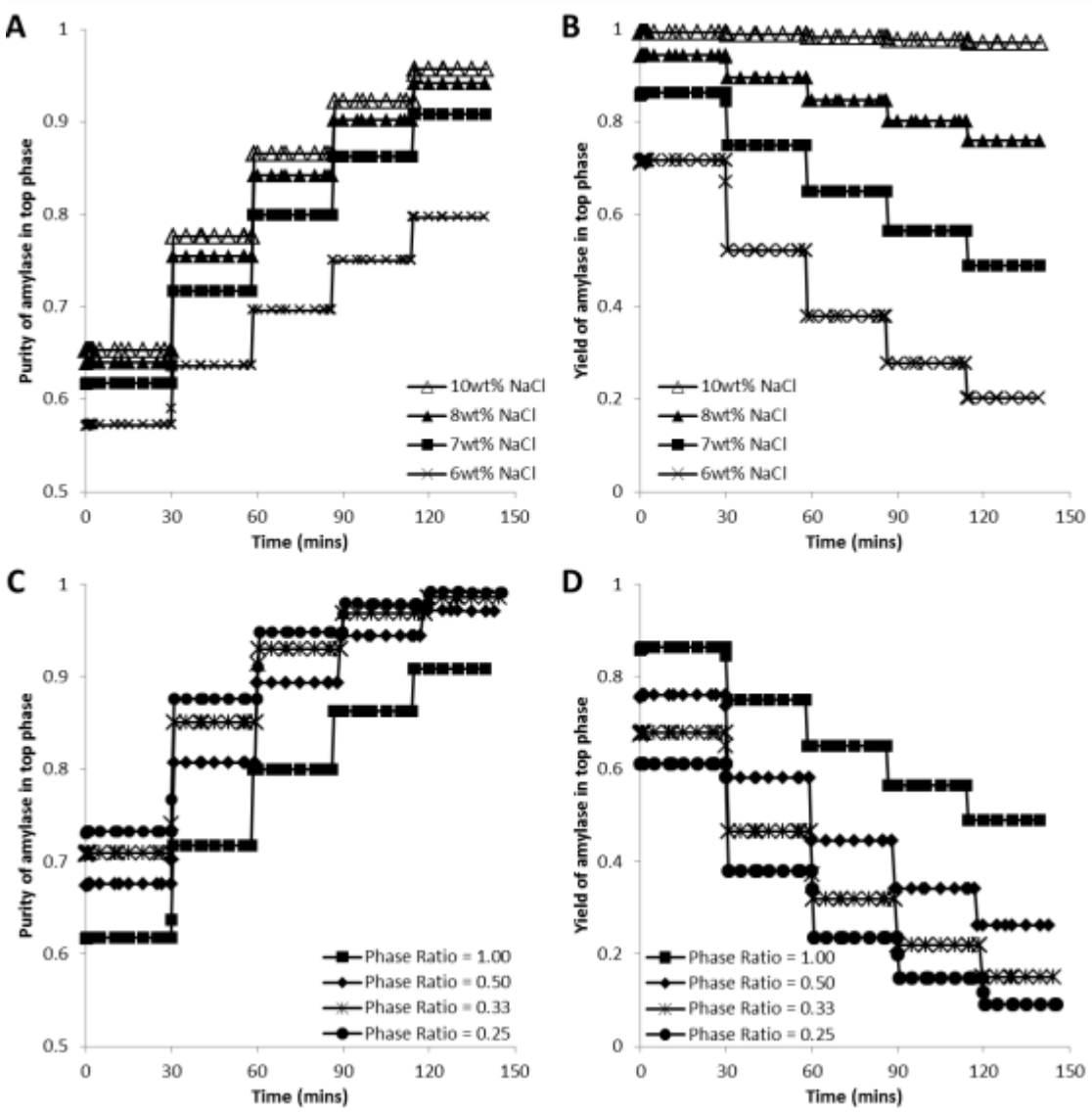


Figure 3 - 5: Simulated protein (α-amylase) yield and purity in the top phase of a PEG 4,000-phosphate aqueous two-phase system multi-cycle batch extraction process. The steps of 30 minute duration correspond to time associated with mixing, settling and draining as shown in points C-E in Figure 3 - 4. The changes in step height correspond to refilling the tank with fresh bottom phase as shown in point E in Figure 3 - 4.
A & B) Influence of NaCl concentration (phase ratio ~ 1). C & D) Influence of phase ratio (7wt% NaCl).

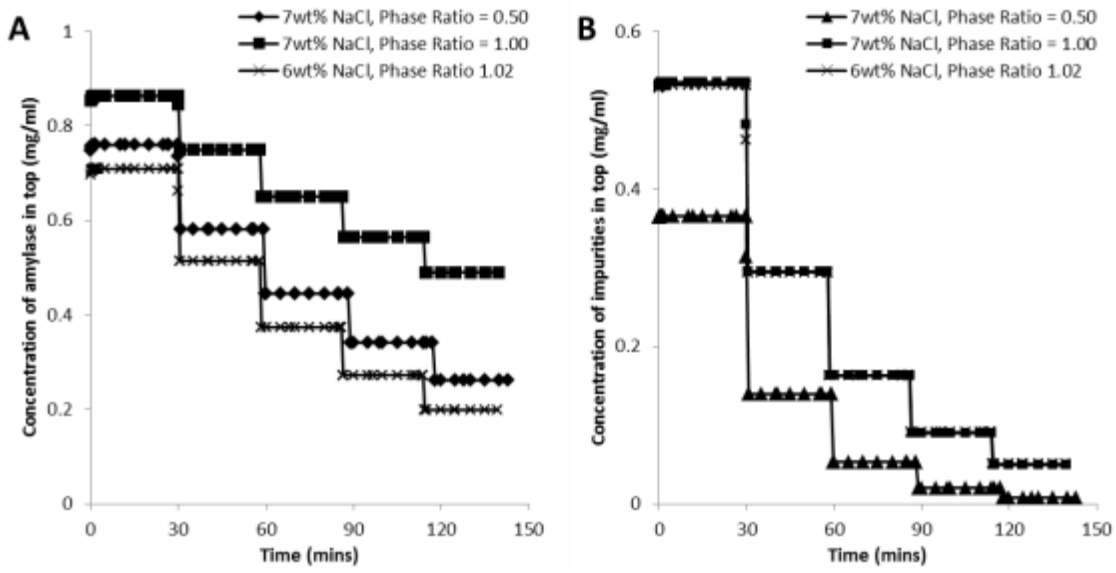


Figure 3 - 6: Concentration of protein (α-amylase) (A) and impurities (B) changes as a result of NaCl, phase ratio and extraction cycles in a PEG 4,000-phosphate aqueous two-phase system multi-cycle batch extraction process. The steps of 30 minute duration correspond to time associated with mixing, settling and draining as shown in points C-E in Figure 3 - 4. The changes in step height correspond to refilling the tank with fresh bottom phase as shown in point E in Figure 3 - 4.

Extraction cycles

The purity, i.e. the ratio of α-amylase concentration to total concentration of α-amylase and impurities, can be improved by repeating the extraction process over a number of cycles as shown in Figure 3 - 5. Increasing the number of extraction cycles is akin to increasing the number of stages in a continuous counter-current process. The advantage of a multi-cycle batch setup is that this operation is simple to execute as the remaining product is just recycled back to the tank and the process runs again for each cycle.

As expected, increasing the number of extraction cycles results in a higher purity of protein (α-amylase) in the product (top) phase (Figure 3 - 5A), however, this comes at the cost of a reduced yield (see Figure 3 - 5B). The yield is reduced because for each new extraction cycle, more α-amylase is being lost to the fresh bottom phase (contains no α-amylase or impurities) as a new equilibrium is established. In addition, one must also consider whether the reduction in concentration as a result of increasing the number of extraction cycles is worthwhile since, as shown in Figure 3 - 6, each cycle results in more material being lost, therefore the maximum achievable concentration is reduced.

Amount of NaCl

Figure 3 - 5 shows that increasing the NaCl concentration from 6wt% to 10wt% will cause both the purity and the yield to increase. This is because the partition coefficient of α-amylase is increased from ~2 to ~170 as the salt concentration is increased, whilst the partition coefficient of the impurities remains relatively constant. This can be explained by the differences in the correlations for the partition coefficient described by Eq.(3.16)-(3.17). The addition of salts often results in the partition coefficient of proteins changing, however, the magnitude and direction of this change differs between proteins (Tubio et al., 2007, Gunduz and Korkmaz, 2000). This effect is most noticeable in plot B in Figure 3 - 6 where the concentration of impurities in the top phase remains almost identical at both 6wt% and 7wt% NaCl. Operating using 10wt% NaCl would see a purity and yield of 0.86 and 0.98 after three extraction cycles, respectively. At NaCl concentrations less than 5wt%, the α-amylase partition coefficient is less than 1, meaning that the operating protocol could be adapted to replace the top phase instead of the bottom phase, thus allowing for the product to be loaded into the bottom phase. This could be useful for situations where a high salt concentration feed is required for the next purification step such as in hydrophobic interaction chromatography. However, these simulations were not conducted in this work as the aim was to purify α-amylase into the top phase.

Phase ratio

Figure 3 - 5 also shows how the phase ratio can be manipulated to alter purity and yield. A smaller phase ratio (Plots C & D) results in higher purity because, relative to α-amylase, more impurities partition into the bottom phase. Unfortunately, the amount of α-amylase in the bottom phase is also increased when the phase ratio is decreased, hence resulting in a lower yield. Using a higher amount of NaCl resulting in higher partition coefficients for α-amylase would provide the ability to operate at lower phase ratios while still maintaining a high concentration. The optimum conditions would have to be determined via taking into account costs of additional salt as well as potential impact on product quality such as via aggregation or precipitation.

3.3.3 Continuous counter-current simulations

Although batch operation is currently the preferred mode of operation for biopharmaceuticals, there is a drive towards continuous operation where possible, due to the increased efficiency associated with these processes (Jungbauer, 2013). To consider continuous operation of the ATPS system, a base case study of a perfusion bioreactor setup was considered. Perfusion bioreactors are well suited to a continuous counter-current setup because they can be set up so that there is a continuous flow of cell culture material containing the product of interest to subsequent downstream separations.

In the case study, a perfusion bioreactor of 300L was assumed with a perfusion rate of 300L per day (i.e. 1 reactor volume). Pollock et al. (2013) state that product titre in a perfusion setup can range from 20% to 45% of a fed-batch setup. Assuming a fed batch titre of 2 g/L, a perfusion titre of 0.65 mg ml^{-1} was used in the following for both product and impurities. The flow rate of the feed is based on a perfusion rate of 300L per day assuming a density of ~1000 kg m^{-3}. The feed composition of protein (α-amylase) and impurities was set to half of the perfusion titre, i.e. a 1x dilution factor, because in reality one would have to dilute the perfusion culture feed with the polymers, salt and water. This is because PEG above a certain molecular weight is only available as a powder and must be diluted from a concentrated stock solution prepared in water. PEG 4,000 has a maximum solubility in water of ~64wt%, however, typically stock solutions are made in the range of 50-60wt%. Therefore, to make a 27wt% PEG 4,000 feed stream without diluting the product too much would require directly mixing the PEG stock solution with the feed stream containing product in an approximately 1 to 1 ratio resulting in a 1x dilution factor.

Specific details of stream compositions and the initial compositions in each stage are shown in Table 3 - 3. In this work the impurities are deemed to make up 50% of the total proteins. In reality the concentration of impurities such as host cell proteins number would vary considerably based on the performance of different cell lines and growing conditions in upstream bioreactor processes, however, is still likely to be within the same order of magnitude as the concentration of the therapeutic protein being expressed (Tait et al., 2012).

The impact of feed location on separation performance is considered as it is important to decide if the product should be processed into the top or bottom phase inlets. The impact of process variability due to poor NaCl control and drifting titre is also considered. This is important as it helps identify potential process robustness issues as well as critical process parameters. In addition, the impact of stage size is considered as a method of designing in inherent process robustness to potential disturbances.

Feed location

So far, it has been assumed that the proteins (α-amylase and impurities) are introduced to the system in the top PEG 4,000 rich phase and that the phosphate rich bottom phase is used as the solvent to extract impurities. The model developed in this work can also be used to explore how changing the feed location of the proteins from the top (PEG 4,000) to the bottom (phosphate) phase may influence the accumulated protein purity. Figure 3 - 7 shows this scenario for a 3 stage system where the feed location of protein is changed from stage 1 (top) to stage 3 (bottom). As can be seen, the switch from the top to the bottom phase results in a reduction in purity from ~0.75 to ~ 0.57 for the protein (α-amylase). This is because when α-amylase is fed in the top phase (stage 1), there is more efficient utilisation of each extraction stage as a result of the system maintaining a larger concentration difference between the phases across all stages due to there being a α-amylase partition coefficient value greater than 1. This can be clearly seen in Plot C in Figure 3 - 7 where the concentration difference of α-amylase between the phases remains higher than 0.20 mg ml^{-1} for all three stages when fed via the top phase. Plot D in Figure 3 - 7 shows how the concentration difference drops dramatically from stage 3 to 1, therefore resulting in most of the purification of α-amylase happening in stage 3 for the bottom phase fed scenario, and hence a less efficient utilisation of the 3 stages. These findings will change depending on the partitioning characteristics of the components. If, for example, the desired product had a partition coefficient value smaller than 1, then it would make more sense to feed via the bottom phase. In addition,

if one wanted to concentrate a product into the top phase, then feeding via the bottom phase is also an option (this is demonstrated in Chapter 5).

Table 3 - 3: Operating conditions considered for continuous counter-current extraction of α-amylase (component 5) with impurities (component 6) in a PEG 4,000 (component 1)-phosphate (component 2) aqueous two-phase system (phase ratio ~ 1).

Stream #	Operating flow rate (kg s^{-1})	α-amylase concentration (kg m^{-3})	Impurity concentration (kg m^{-3})	PEG (wt%)	Phosphate (wt%)	NaCl (wt%)
1 in stage 1	0.00347	0 or 0.325	0 or 0.325	27.38	1.77	6 to 10
4 in last stage	0.00347	0 or 0.325	0 or 0.325	0.01	14.46	6 to 10

Table 3 - 4: Initial composition of each stage for continuous counter-current extraction of α-amylase in a PEG 4,000 (component 1)-phosphate (component 2) aqueous two-phase system (phase ratio ~ 1).

Variable	Symbol	Value
Mass hold up (kg)	-	2 or 10
Composition of PEG 4,000	$w_{1,overall,n}$ (wt%)	13.7
Composition of phosphate	$w_{2,\ overall,n}$ (wt%)	8.2
Composition of NaCl	$w_{4,\ overall,n}$ (wt%)	6 to 10

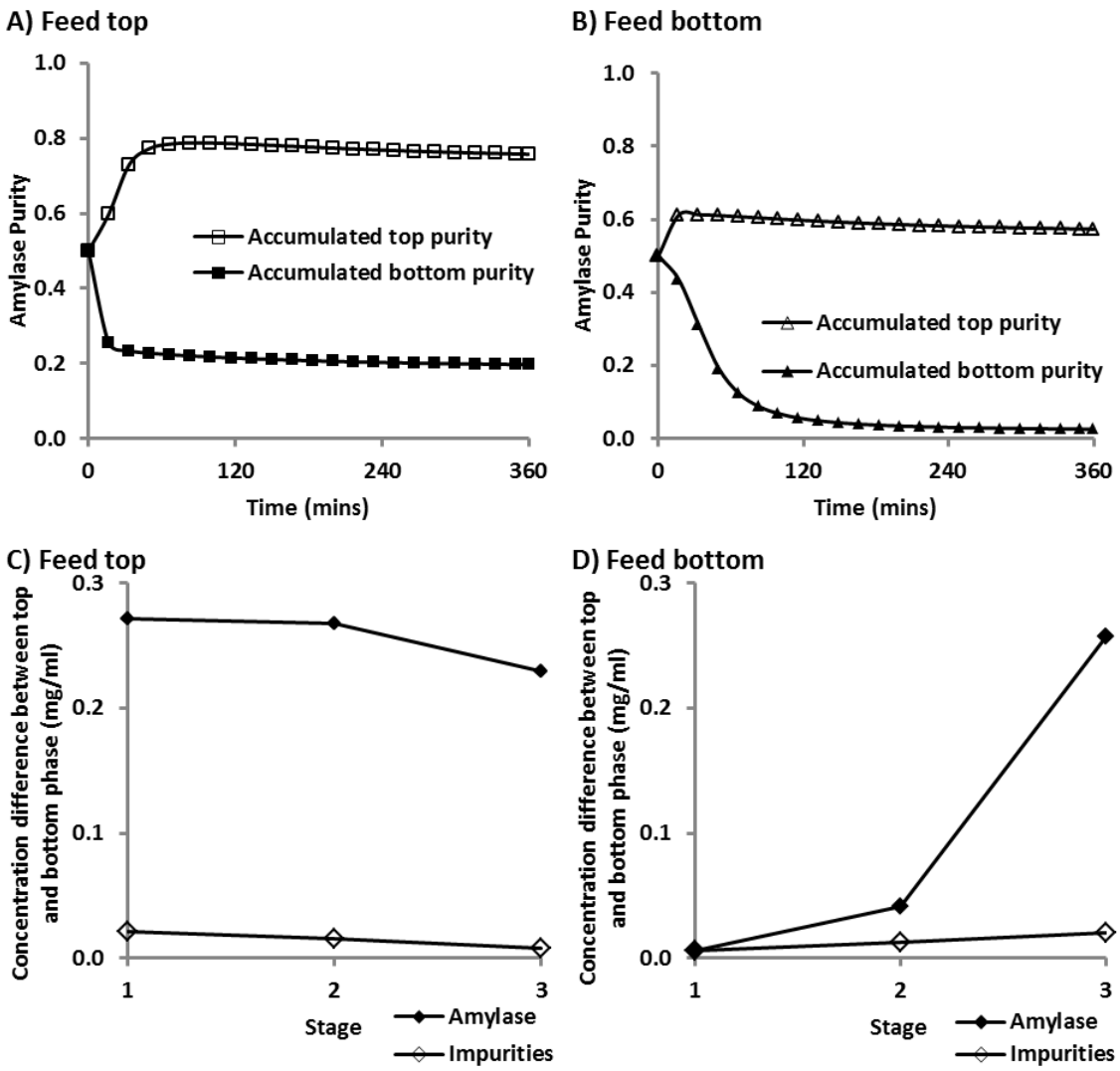


Figure 3 - 7: Influence of feed location on process design and purification performance of a 3 stage counter-current operation of a PEG 4,000-phosphate ATPS (7wt% NaCl, phase ratio ~1). Simulated product feed stream is from a perfusion reactor as described in Section 3.3.3. Plots A & C: feeding protein in stage 1 stream 1 (top phase). Plots B & D feeding protein in stage 3 stream 4 (bottom phase). Plots C & D are concentration differences between top and bottom phase in each stage at 360 minutes.

Influence of NaCl variability

The partitioning of α-amylase and impurities between the two phases is also influenced by the amount of NaCl in the aqueous two-phase system, as well as by the scale of the operation. To understand the importance of NaCl, the sensitivity of process performance due to disturbances in NaCl in the process feeds is investigated. NaCl is present throughout the whole separation system; this can only be achieved if NaCl is fed via streams in both first and final stages. Therefore it was important to consider poor control of NaCl in both these feed streams. As outlined below, the variability was considered by randomly assigning the NaCl wt% value based on a normal probability distribution with a mean of 7wt% and a standard deviation of 2wt%. A composition of 7wt% of NaCl was chosen as the partition coefficient for α-amylase is ~6 whilst a standard deviation of 2wt% was chosen because at 5wt% the partition coefficient for α-amylase is ~1. A random value is chosen every 60 seconds after ~8.5 hours of continuous operation.

NaCl concentration in stream 1 of stage 1 is normally distributed:

$$x_{NaCl,1,1} \sim N(\,7wt\%\,,2wt\%\,) \tag{3.25}$$

NaCl concentration in stream 4 of stage NS is normally distributed:

$$x_{\,NaCl,4,NS} \sim N(\,7wt\%\,,2wt\%\,) \tag{3.26}$$

Variations were started after ~8.5 hours because this is when the simulated concentration of α-amylase in the top phase of the final stage was at steady-state. Figure 3 - 8A shows how when the variability is introduced, the concentration of α-amylase fluctuates as a result. The difference between the maximum and minimum concentrations (i.e. the range) of α-amylase after start of variations is 0.11 mg ml^{-1} which is 38% of the mean α-amylase concentration value (0.28 mg ml^{-1}). Such product concentration variations due to inherent process variability may be very important to control because they could have serious implications on subsequent separation stages where the tolerance of the feed material being processed is quite stringent. The model has shown the impact of this particular variation and can also be used to investigate others process conditions by defining them in a similar way.

Influence of stage size on variability

One way to control variability is to have a system which is inherently less susceptible to disturbances; this can be achieved by increasing the size of the system. Plots A and B in

Figure 3 - 8 shows how simply increasing the size of the system, i.e. changing the total amount of holdup in each stage from 2 kg to 10 kg, results in a dampening effect when the NaCl wt% of the feed streams starts to randomly vary based on the same NaCl wt% mean and standard deviation value as before. For the 10kg stage holdup system, the difference between the maximum and minimum α-amylase concentrations after start of variations is 0.04 mg ml^{-1} which is 14.8% of the mean α-amylase concentration value (0.28mg ml^{-1}). This is less than 40% of the range observed in the 2kg holdup system. In addition, the resulting relative standard deviation for the downstream α-amylase concentration in the top phase of the final stage of the ATPE, for both 10kg and 2kg stage holdups, is 7.08% and 2.50%, respectively. However, Figure 3 - 8C, shows that if the size of the tank is increased, there is a lower yield of α-amylase as the system takes longer to get to steady-state suggesting that a trade-off would have be made between system robustness to disturbances and process yield.

The model has here been used to show one example of how a modelling approach can be used to investigate process behaviour following uncertainties for different designs. The severity of the impact of these uncertainties will entirely depend on the requirements and robustness of subsequent downstream separation processes. It should be noted that one disadvantage of increasing the holdup mass of each stage is that the length of time taken to reach steady-state of the yield is increased (see Figure 3 - 8C); this may, however, be insignificant if one is running a 60 day perfusion culture.

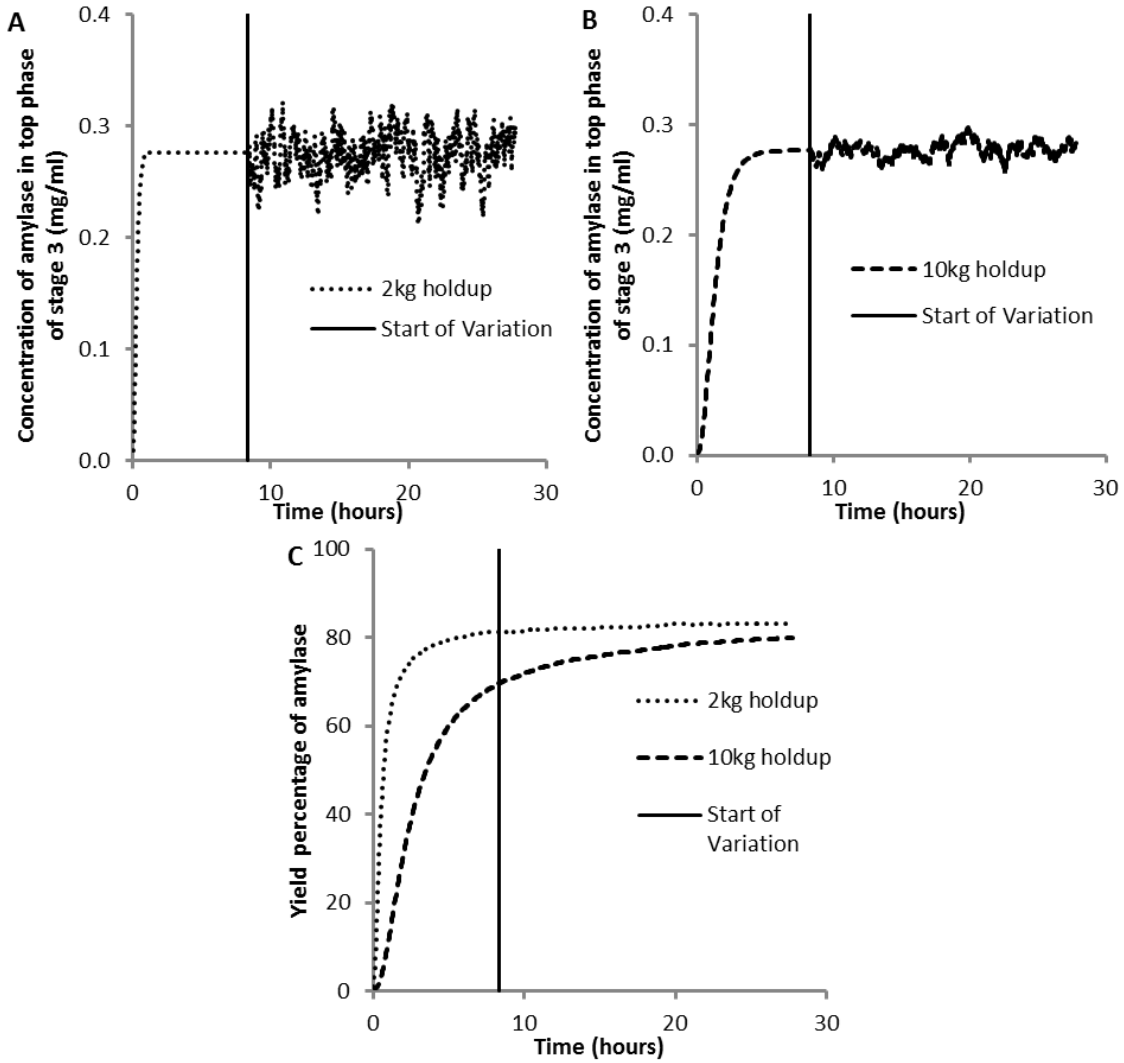


Figure 3 - 8: A&B) Impact of NaCl (mean 7wt%, standard deviation 2wt%) variability on α-amylase concentration in the top phase of the final stage in a 3 stage continuous counter-current extraction process with 2kg and 10kg stage holdup. Product feed stream is from a perfusion reactor as described in Section 3.3.3. Variation of NaCl is implemented once start-up is completed (i.e. when there is a steady-state concentration of α-amylase).
C) Total yield percentage of α-amylase with 2kg and 10kg stage holdup.

Influence of drifting product titre and number of stages

The model can also be used to investigate the impact of a drifting product titre over the course of a 65 day perfusion. Product titre is likely to change due to the inherent variability associated with producing therapeutic proteins using genetically modified cells. Figure 3 - 9 shows a case where the product and impurity titre were kept constant for the first 5 days of operation and was then allowed to drift by ~8% over the course of the 60 days, this is shown in Figure 3 - 9. During the simulations the new decayed/lower concentration of α-amylase is calculated as 0.3% less than the α-amylase concentration in the previous 24 hours while the new non-decaying impurity concentration is calculated as 0.3% higher than the impurity concentration in the previous 24 hours. This was chosen to show a worst case scenario where an unstable cell line was shifting from producing product to impurities as the product gene is not fully integrated into the genome of all cells. The equations used to describe this are:

$$C_{1,new,\alpha-amylase} = 0.997 \times C_{1,old,alpha-amylase} \tag{3.27}$$

$$C_{1,new,Impurities} = 1.003 \times C_{1,old,Impurities} \tag{3.28}$$

where the old concentrations are randomly chosen from a normal distribution of the concentration in the past 24 hrs and with a standard deviation of 0.001:

$$C_{1,old,alpha-amylase} \sim N(Concentration\ of\ \alpha - amylase\ in\ previous\ 24\ hours, 0.001) \tag{3.29}$$

$$C_{1,old,Impurities} \sim N(Concentration\ of\ impurities\ in\ previous\ 24\ hours, 0.001) \tag{3.30}$$

Plot B in Figure 3 - 9 shows how the purity of α-amylase out of each stage decreases as the titre drifts down. Although the product purity in the feed (stream 1, stage 1) drops from 50% to ~42% over the simulated 65 days, the absolute drop in α-amylase purity in stream 4 of stage 3 (purity before titre drift minus purity at end of simulation) is ~6.5%. Such a reduction in purity could potentially cause subsequent purification steps to fail as the final product purity required for pharmaceuticals is more than >99%. In this simulation, 7wt% NaCl was assumed which results in an α-amylase partition coefficient of ~6.0. If the same simulation is run but with 12 counter-current stages instead of 3 stages, an absolute purity drop of ~4.7% is observed.

Figure 3 - 10B shows that increasing the number of stages results in a smaller drop in purity; however, it seems that the drop in purity is not linear with number of stages

suggesting that after a certain number of stages are added, the benefit gained in terms of purity drop is negligible. In addition to reducing the drop in purity, increasing the number of stages 3 to 24 increases the total α-amylase purity from 67% to 83% respectively; this is similar to increasing the number of cycles in the multi-cycle batch setup. Increasing the number of stages will reduce the yield, however, due to the high partition coefficient of α-amylase at 7wt% NaCl the yield remained relatively constant at ~84%.

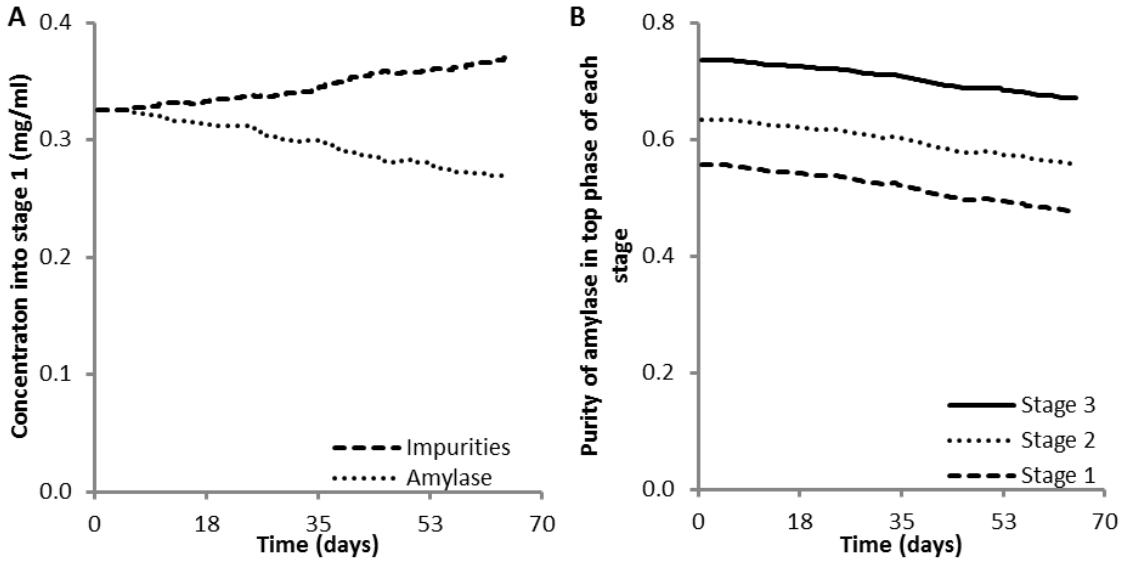


Figure 3 - 9: Impact of drifting titre (increasing impurities & decreasing product) on purity in a 3 stage continuous counter-current extraction process. (NaCl kept constant at 7wt%.) A) Titre of α-amylase and impurities into stage 1. B) in-line purity of α-amylase in the top phase of each stage of the counter-current train.

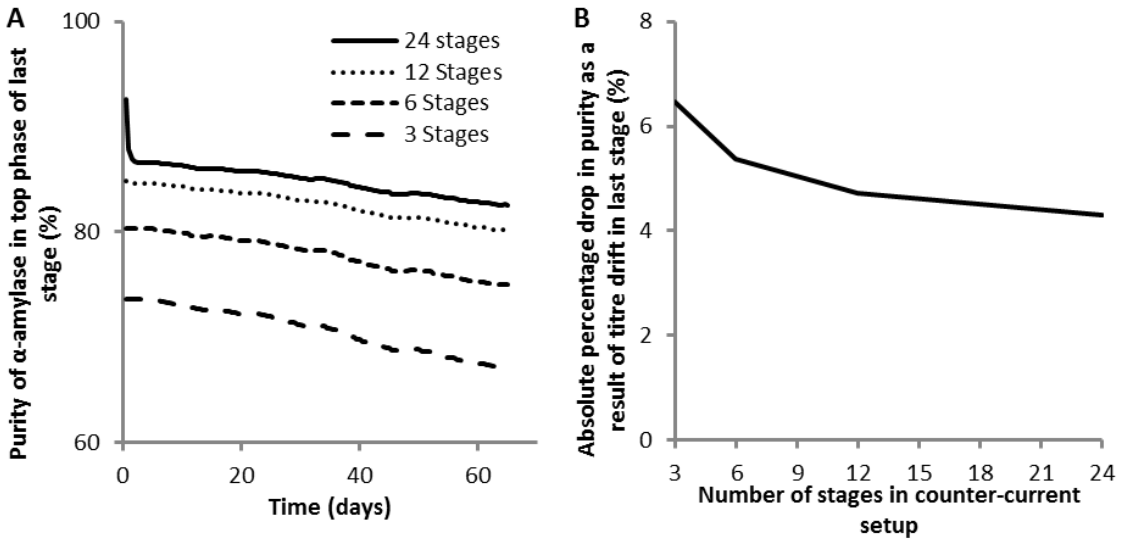


Figure 3 - 10: Impact of process design on product purity in the presence of drifting titre. A) Influence of number of stages on in-line purity of α-amylase in final stages, and B) Influence of number of stages on absolute

3.4 Advantages, limitations and challenges of this modelling approach

The main advantages of the approach presented in this work for the modelling and simulation of aqueous two-phase systems are:

1. Flexibility to use one model to investigate different modes of operation, such as multi-cycle batch and continuous counter-current modes of operation, as well as the flexibility to adapt the model to different experimental systems using appropriate parameter estimation and description of component partitioning.
2. The ability to conduct dynamic simulations which are important in understanding impact of bioprocess variability.
3. The inclusion of thermodynamic equations to describe phase equilibria which is something current literature has not considered for aqueous two-phase system process simulations. The use of thermodynamics will help alleviate experiments which require generation of phase equilibria is parameters in thermodynamic models for different two-phase systems are well documented.

Important limitations to the modelling approach presented include the need to validate model simulations with experimental data, for example stage efficiency which is unlikely to be 100%. In addition, the approach presented in this work considered only two components, α-amylase and impurities, as all other components generated during upstream fermentation/cell culture are lumped together as impurities. These components could be considered individually. To be able to model these types of multi-component systems will require a careful balance of model detail and experimental effort as the experimental effort would be phenomenal. One solution to this is to use high-throughput analytical techniques to quantify only the major impurity types. Finally, the thermodynamic equations used in this work are relatively simple as they do not consider factors such as ionic interactions and polymer molecular weight. The implementation of a more detailed thermodynamic equation set would make the modelling approach presented in this work more comprehensive.

3.5 Conclusion

The main aim of the work in this chapter was to present a dynamic equilibrium stage process model, and a methodology for prediction of key process parameters from limited experiments, capable of describing aqueous two-phase extraction (ATPE) separations under both multi-cycle batch and continuous counter-current modes of

operation. The model can be used to predict the separation performance of the process, as well as for the investigation of suitable design and operating conditions.

The capabilities of the methodology were demonstrated using a case study of a PEG 4,000-phosphate-NaCl aqueous two-phase extraction (ATPE) process for the purification of α-amylase using both multi-cycle batch or continuous counter-current extraction modes. Thermodynamic interaction parameters were estimated from experimental phase equilibria data from literature, and were found to be sufficiently accurate for phase equilibria predictions for aqueous two-phase systems (ATPSs) containing more than 2wt% NaCl.

This work has demonstrated how a relatively simple dynamic process model can be used to better understand the behaviour of the aqueous two-phase extraction system, and to predict process behaviour of downstream processing as a result of upstream product variability which is often difficult or impossible to gauge when using steady-state process models. The development of the model also involved estimation of model parameters from a limited number of experiments, and an analysis of the impact of experimental uncertainty on the accuracy of the parameters obtained. The model developed in this chapter is used for further simulations in Chapter 5.

Chapter 4: Quantification of uncertainty in phase diagrams for estimation of interaction parameters

4.1 Introduction

Chapter 3 looked at the use of dynamic models to simulate an aqueous two-phase extraction process. Estimation of interaction parameters was conducted based on literature phase diagram data with an unknown experimental certainty. For models to be used in the production of biopharmaceuticals, the impact of experimental uncertainty must be reflected in the description of the model and the appropriate estimated parameters. This is of particular importance within the implementation of the Quality by Design (QbD) methodology advocated by the US Food and Drug Administration (FDA), which calls for greater understanding of all processes so that the quality of the drug substances being manufactured is guaranteed using rigorous scientific methods (Rathore and Winkle, 2009b). This need for greater understanding has resulted in more sophisticated technologies, such as high-throughput screening and process modelling, in the development of novel biopharmaceuticals.

The use of high-throughput screening for the development of aqueous two-phase systems has been well documented by the group of Hubbuch (Bensch et al., 2007, Oelmeier et al., 2011, Diederich et al., 2013, Amrhein et al., 2014). Their approach uses automated liquid-handling robots to determine phase diagrams and protein partitioning characteristics in scaled down aqueous two-phase systems. But to date, there have not been any reports in the literature of methods to determine the uncertainty in phase diagrams generated using these high-throughput screening methods, and the resulting impact that this uncertainty may have on the estimation of thermodynamic interaction parameters and ultimately the accuracy of model predictions.

In this work, this issue is considered based on a formalised method to quantify the uncertainty in phase diagrams generated using a combination of high-throughput screening and lab scale experiments. In addition, we investigate the impact of this experimental uncertainty on the estimation of thermodynamic interaction parameters, something which has not been considered so far in the literature for aqueous two-phase system phase diagrams. The proposed approach is illustrated in Figure 4 - 1 and described overleaf.

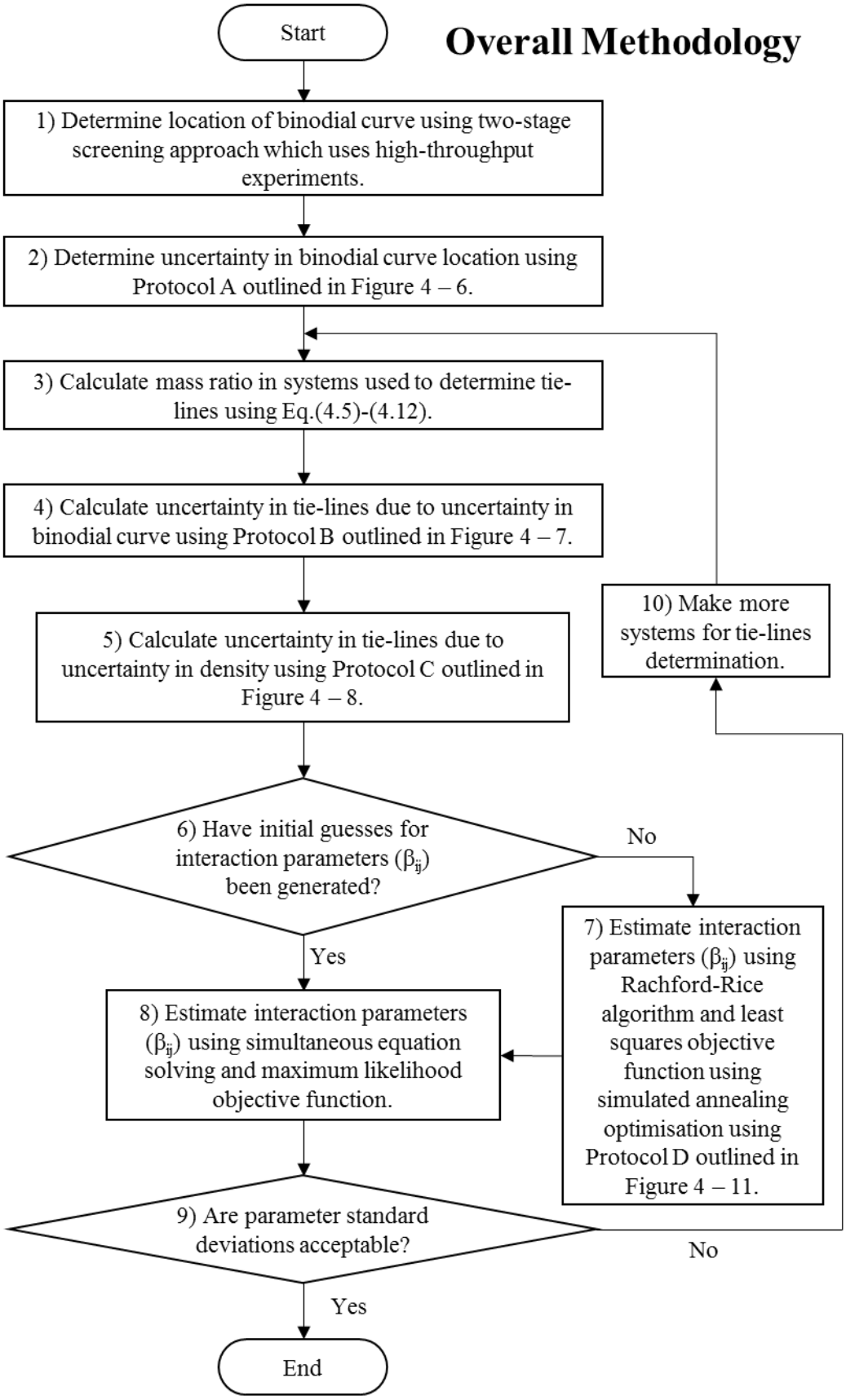


Figure 4 - 1: Proposed overall methodology to estimate uncertainty in phase diagrams for aqueous two-phase systems and the impact of this uncertainty on estimated interaction parameters.

The approach presented in Figure 4 - 1 was developed because there is a need for detailed knowledge of the uncertainty associated with phase equilibrium composition when conducting maximum likelihood parameter estimation within gPROMS to determine interaction parameters. This is because in this work phase equilibrium compositions are determined by firstly inferring the location of the binodial curve and then fitting tie-lines to the binodial curve using mass ratios. Therefore, the uncertainty in the equilibrium compositions is directly linked the fundamental experimental design during binodial curve screening and the accuracy of experimental techniques when determining mass ratios in two-phase systems. Once the uncertainty was quantified, parameter estimation techniques were used to estimated interaction parameters. The gPROMS parameter estimation tool required good initial values for the solvers to initialise. To address this issue a separate parameter estimation protocol based on using the Rachford-Rice algorithm was developed to generate good initial guesses for maximum likelihood parameter estimation within gPROMs.

Step 1) Determine location of binodial curve using two-stage screening approach

To determine a complete phase diagram for a specific aqueous two-phase system, the location of the binodial curve is required. As explained previously, the binodial curve is the curve which lies at the boundary between a single-phase and two-phase system. The high-throughput method used in this work to determine the location of the binodial curve involves the formation of multiple systems of different composition in 96 well plates which are then tested for the presence of two-phases.

A two-stage screening approach was developed to locate the binodial curve (see Figure 4 - 2). An initial low resolution screening experiment is designed to find an approximate location. The benefit of a low resolution screen is that a large composition range can be covered using fewer experimental systems, however, the resultant uncertainty in the binodial curve location is greater. To reduce the uncertainty in the location of the binodial curve, a second targeted screening experiment of a higher resolution is then conducted. The two-stage approach uses less overall experimental resources when compared to a single high-resolution screen across the entire feasible composition range. The compositions of experimental systems used in the targeted screen are based on both the approximate location of the binodial curve and the size of the initial screening resolution utilised. Further information on this approach is given in Section 4.1.7.

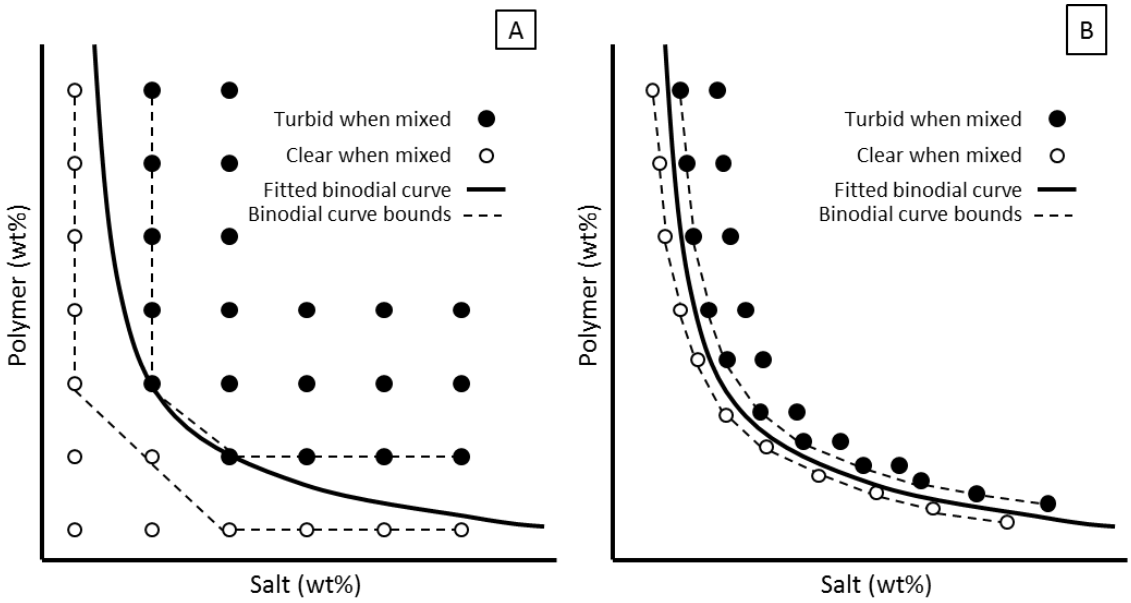


Figure 4 - 2: Illustration of low (A) and targeted higher resolution (B) binodial curve screen.

Step 2) Determine uncertainty in binodial curve location

Screening resolution and design of experiments have a direct impact on the uncertainty in the location of the binodial curve. To quantify the uncertainty in the location a method was developed which is illustrated in Protocol A in section 4.2.4. Note that uncertainty in the binodial curve location cannot be reduced by repeating the same screening experiments multiple times since the fundamental uncertainty in the location of the binodial curve is due to the design of the experiment.

Step 3) Calculate mass ratio in systems used to determine tie-lines

A range of aqueous two-phase systems are made and the mass ratio (the mass of top phase/the mass of bottom phase) is calculated so that it can be used to fit tie-lines. Tie-lines relate the equilibrium compositions of components used to establish the two-phase system in the top and bottom phase to their overall composition.

Step 4) Calculate uncertainty in tie-lines due to the uncertainty in the binodial curve

Since tie-lines are determined by fitting them to the binodial curve, any uncertainty in the binodial curve will result in uncertainty in the tie-lines and the corresponding equilibrium compositions. A method was developed to quantify the uncertainty in tie-lines due to the uncertainty in the binodial curve which is illustrated in Figure 4 - 7Protocol B in section 4.2.5.

Step 5) Calculate uncertainty in tie-lines due to uncertainty in density measurements

Tie-lines are determined by fitting them to the binodial curve such that the inverse lever arm rule is equal to the calculated mass ratio of each system used to determine the tie-lines as shown in Figure 4 - 3B. Any uncertainty in the mass ratio will therefore also result in uncertainty in the tie-line and the resultant equilibrium compositions. To quantify the uncertainty in the mass ratio due to density, Protocol C was developed which is outlined in section 4.2.6.

Step 6, 7 and 8) Estimate interaction parameters

Interaction parameters (β_{ij}) in Eq.(3.21)-(3.23) are used to fit model predicted phase equilibria compositions to experimentally calculated phase equilibria compositions. In this work, interaction parameters are first estimated using the Rachford-Rice algorithm[4] and a least squares objective function (Eq.(4.18)) , as shown in Protocol D in section 4.2.8. The implementation of the Rachford-Rice algorithm is useful as it is far more robust than using simultaneous equation solving to conduct phase equilibria calculations. Minimisation of the least squares objective function was conducted using a simulated annealing optimisation algorithm to reduce the likelihood of becoming stuck in a local minimum.

The estimates of interaction parameters from Step 7 are then used as initial guesses in Step 8 when estimating the parameters using a maximum likelihood objective function. The gPROMS software used in this work solves equations simultaneously rather than sequentially therefore there is a need to have good initial guesses for both interaction parameters and phase compositions. The maximum likelihood objective function is minimised in gPROMS using a local deterministic optimisation algorithm.

Step 9 and 10) Evaluate parameter values and uncertainty

The parameter estimation entity in gPROMS is used to evaluate the statistical significance of estimated parameters, which includes parameter standard deviations and confidence ellipsoids. If uncertainty in estimated parameters is deemed unacceptable, then more experimental tie-lines need to be determined.

[4] The Rachford-Rice algorithm was used in this work for its simplicity. There exists many more modifications and alternatives to the Rachford-Rice algorithm which can be used (Nichita and Leibovici, 2013).

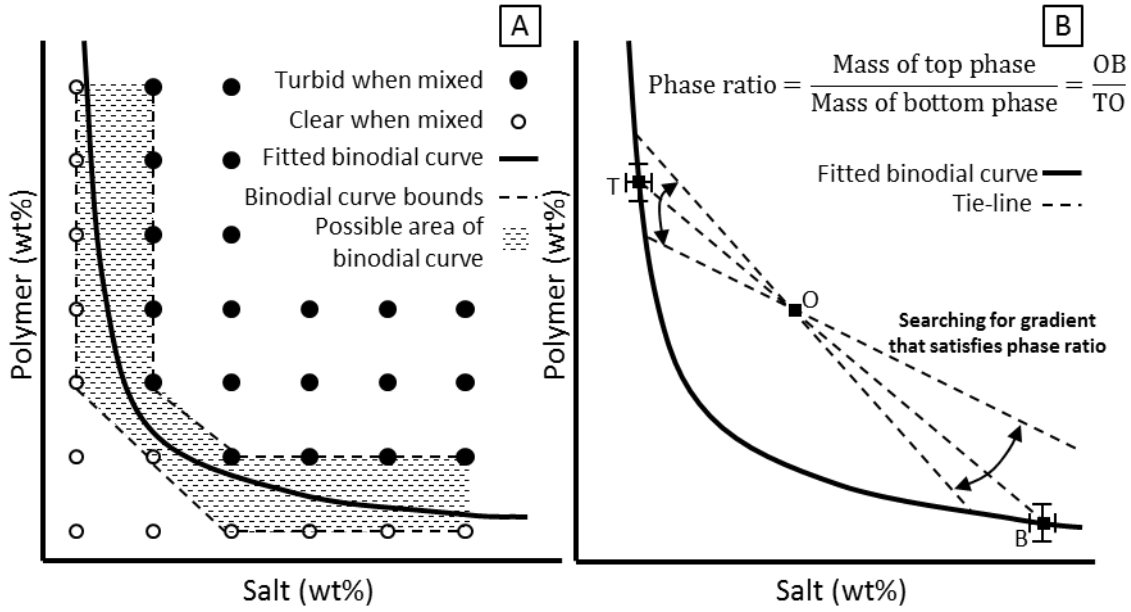


Figure 4 - 3: Phase diagram generation. A) Method to screen for binodial curve. Multiple systems of different composition formed using automated liquid handling. Systems mixed and checked for turbidity which indicates a two-phase system. Binodial curve fitted in gaps where transition from clear to turbid system is observed. B) Method for determining tie-lines using mass based phase ratio. T, O and B represent top, overall and bottom phase compositions, respectively. Gradient of tie-line is adjusted such that the ratio of lines OB and TO, which intersect with the binodial curve, is equal to the phase ratio.

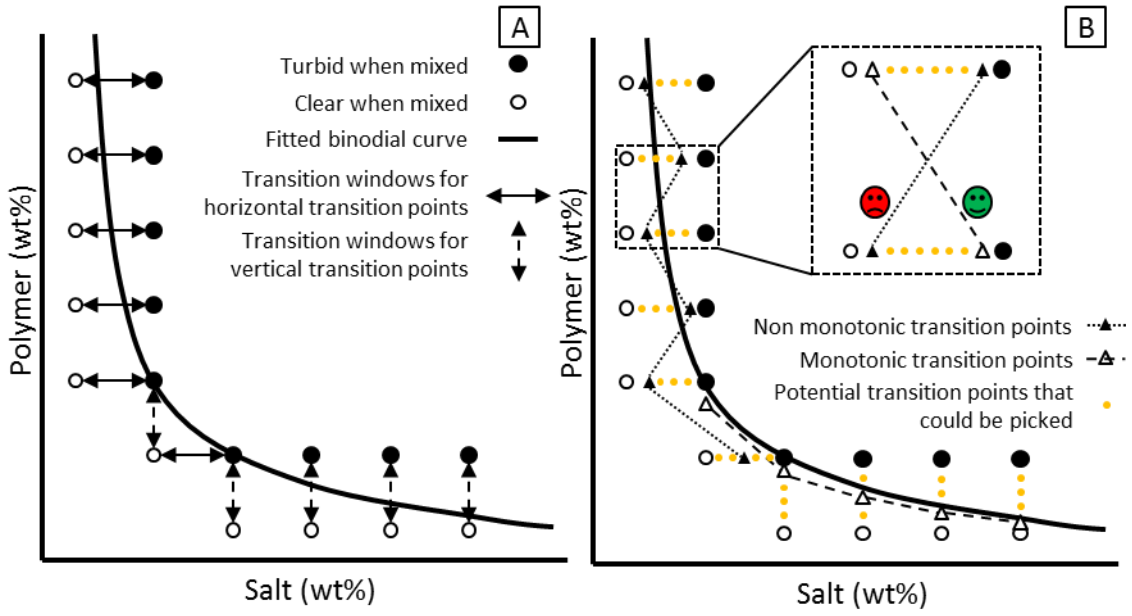


Figure 4 - 4: Method to determine uncertainty in binodial curve. A) Transition windows for generation of binodial curve. The transition windows are assumed to be uniformly distributed. Random points are picked from each transition window. B) Random points are accepted if they are monotonic[5] i.e. in size order. In the above example, the filled black triangles would be rejected, the clear triangles would be accepted and used to estimate the binodial curve.

[5] A monotonic function is a function which is either entirely nonincreasing or nondecreasing i.e. if the first derivative of a function does not change sign.

4.1 Experimental materials and methods

The experimental materials and methods described here are used for the work in this chapter as well as in Chapter 5 where process simulations are conducted.

4.1.1 Chemicals and stock solutions

All chemicals were purchased from Sigma-Aldrich (UK) and are more than 99% pure, where possible Sigma-Aldrich's BioUltra grade reagents were used. 50wt% PEG 4,000, 50wt% PEG 2,000, 25wt% NaCl, 40wt% NaH_2PO_4, 40wt% K_2HPO_4 and 30wt% sodium citrate tribasic stock solutions were prepared by mixing together with appropriate amounts of each chemical with MilliQ water. pH 7 and pH 7.5 40wt% phosphate stock solutions were prepared by titrating the 40wt% solutions of NaH_2PO_4 and K_2HPO_4 until the desired pH was obtained. pH was measured using a Mettler Toledo InLab Versatile Pro (Mettler Toledo UK, UK) electrode. It is important to use a probe with a liquid electrolyte (e.g. KCl) to ensure a fast and stable pH response. A two component protein stock solution of ~8mg mL^{-1} α-amylase and ~15mg mL^{-1} myoglobin was prepared using α-amylase from *Bacillus subtilis* (Sigma-Aldrich, UK) and myoglobin from equine heart (Sigma-Aldrich, UK), in 10mM pH 7 phosphate buffer.

4.1.2 Concentration determination

Concentration of single protein stock solutions were determined using absorbance at 280nm using theoretical extinction coefficients obtained from the ExPASy ProtParam tool (Gasteiger et al., 2003). Extinction coefficients were based on amino acid sequence of α-amylase and myoglobin obtained from the UniProt protein database using entry names AMY_BACSU and MYG_HORSE respectively. For α-amylase and myoglobin extinction coefficients[6] of 1.640 and 0.818, respectively, were used.

Concentration in mixtures

Myoglobin has a strong absorbance at 409nm. Concentration of α-amylase was inferred from total protein concentration and myoglobin concentration. Total protein concentration was determined using either absorbance at 280nm or using the BCA (bicinchoninic acid) assay with absorbance at 562nm. Two sets of equations were then solved simultaneously to determine protein concentration:

$$Total\ absorbance_{280\ or\ BCA} = Absorbance_{280\ or\ BCA}^{Myoglobin}(c_{Myoglobin}) + Absorbance_{280\ or\ BCA}^{\alpha-amylase}(c_{\alpha-amylase}) \quad (4.1)$$

$$Total\ absorbance_{409} = Absorbance_{409}^{Myoglobin}(c_{Myoglobin}) + Absorbance_{409}^{\alpha-amylase}(c_{\alpha-amylase}) \quad (4.2)$$

The absorbance of α-amylase at 409nm is almost negligible when compared to myoglobin, however, it was taken into account to provide a more general approach. Single component standard curves for α-amylase and myoglobin were determined for absorbance using 409nm, 280nm and 562nm (BCA assay). The single component standard curves were used to determine single protein absorbance contribution to total absorbance measured in Eq.(4.1) and Eq.(4.2) above. This allows Eq.(4.1)-(4.2) to be solved simultaneously as the right hand side can be written in terms of myoglobin and α-amylase concentration. Standard curves are available in Appendix 7.1. Blanks from control aqueous two-phase systems with no added protein were used when determining protein concentration in samples from aqueous two-phase systems.

4.1.3 Density measurements

[6] Extinction coefficient values are for a 1cm path length where concentration is measured in units of mg mL^{-1}.

Density of solutions was measured by pipetting 250µL of sample using a Gilson Microman 250 positive displacement pipette into a 1.5mL reaction tube. The Gilson Microman 250 has a calibration error of 2.5µL when pipetting 250µL (Gilson, 2016). The sample was then weighed using a Mettler Toledo AB204-S analytical balance which has a calibration error of 1mg (Mettler-Toledo, 2005). Density was calculated by dividing the volume and the measured mass. All density measurements were taken at a lab temperature of 22°C.

4.1.4 Lab scale aqueous two-phase systems

Initial screening of partitioning in aqueous two-phase systems was conducted using a total system size of 4g in 15mL Eppendorf tubes. Systems were mixed thoroughly using a vortexer for at least 20 seconds each. After mixing, all systems were centrifuged at 2000g for 5 minutes.

Samples of top and bottom phases were extracted manually using a syringe and needle. Bottom phase samples were extracted using a positive pressure to minimise the risk of cross-contamination between the top phase and bottom phase samples. Specifically, as the needle was withdrawn through the top phase, a slight pressure was applied to ensure outward flow of bottom phase sample thereby stopping any top-phase entering the syringe and needle.

4.1.5 Automated liquid handling

Tecan Freedom Evo 150 and 200 automated liquid handling robots were used for binodial curve screening (Tecan Group Ltd, Switzerland). The robots are equipped with disposable tips (DiTis) of 1000µL and 200µL. Tips were not reused in any experiments. Aspirating and dispensing of samples is governed by liquid classes, a collection of different settings which can be changed to reach the desired level of liquid handling performance.

Liquid-class calibration

Liquid classes were calibrated such that all volumes greater than 200µL were pipetted using 1000µL, pipette tips while all volumes less than 200µL were pipetted using 200µL pipette tips. This ensured that the air gap between the system fluid in the liquid handling robot and the sample is minimised. Preliminary liquid class calibration was achieved using clear tips which allowed visual inspection of sample aspirating and dispensing. Final liquid class calibration was then carried out using conductive tips which are capable of sensing liquid level height, however, are opaque. Liquid classes

were calibrated in two brackets, 800µL-200.01µL and 200.01µL-10µL. The bracket boundary was set to 200.01µL instead of 200µL to ensure that this boundary is inclusive of 200µL. Within each bracket, volumes were measured in triplicates at the maximum, middle and minimum bracket volumes. Sample liquid class calibration curves and calibration settings can be found in Appendix 7.3.

Viscous solutions

50wt% PEG 4,000 is viscous compared to water. To achieve consistent pipetting, slow aspirating (25 µL s^{-1}) and dispensing (45 µL s^{-1}) speeds were used. During aspiration, the tips were held within the sample for an additional 10 seconds to allow any lagging liquid movement to equilibrate with the suction pressure created due to aspirating. Contact dispensing onto the walls of the deep well plates was used to minimise variability due to drop formation on the end of the disposable tips.

Experiment execution

Scripts for experiments on the Tecan automated liquid handling platforms were prepared in the Tecan Evoware software (Tecan, 2017). Tecan worklist files were used to instruct how much liquid should be dispensed to each well.

4.1.6 High-throughput screening of protein partitioning

A Tecan Evo 150 automated liquid handling platform (Tecan Group Ltd, Switzerland) was used to dispense 50wt% PEG 2,000, 40wt% pH 7 phosphate, 25wt% NaCl and 10mM pH 7 phosphate stock solutions using disposable tips. Volumes less than 200µL were pipetted using 200µL tips while volumes greater than 200µL were pipetted using 1000µL tips.

System formation

A total system size of 1g was used when preparing systems using the automated liquid handling platform. Systems were prepared in 2mL Eppendorf tubes as these provided superior mixing characteristics to the 1.5mL Eppendorf tubes. 125µL of protein solution was pipetted into each system manually using a positive displacement pipette.

Mixing, settling and sample extraction

Systems were mixed at 3000RPM for 8 minutes using an IKA MS3 basic mixing device (IKA England Ltd, UK). After mixing, the systems were centrifuged at 2000g for 10 minutes. Samples of top and bottom phase were manually extracted using a pipette. Again, a positive pressure was maintained when extracting bottom phase samples.

Automated liquid handling was not used for phase extraction as it is difficult to minimise cross-contamination between samples of top and bottom phase.

4.1.7 Binodial curve two-stage screening

The binodial curve was screened by forming many systems of different compositions in deep well plates using automated liquid-handling. Total system mass per well was 800mg. The formation of a two-phase system was verified visually by eye for the presence of turbidity upon mixing. Mixing was achieved by aspirating and dispensing 450µL using a hand operated multichannel pipette with clear tips. The clear tips allowed for easy determination of turbidity.

Experiment design

Initial binodial curve screen

An initial screening resolution of 4wt% was chosen in the PEG 4,000 and sodium citrate tribasic compositions. Screening compositions were generated using a custom written MATLAB script that takes into account the compositions of stock solutions, total system mass and the desired screening resolution. Figure 4 - 5 shows an illustration of square markers corresponding to the overall composition of PEG 4,000 and sodium citrate tribasic in the formed systems. Black markers represent those systems which form two-phases whereas clear markers represent those which form a single-phase.

High resolution second stage screen

Once an approximate location of the binodial curve was found using the 4wt% screening resolution, a second high-resolution screening experiment was designed. The compositions of systems were generated using a custom written MATLAB script which targets experiments around the initial binodial curve generated. Figure 4 - 5 illustrates the high-resolution screen generated using this approach. Again the square markers corresponding to the composition of PEG 4,000 and sodium citrate tribasic in the formed systems, however, in this targeted screen the screening area covered follows the curvature of the initial binodial curve location (see Figure 4 - 14A). The script also takes into account the initial screening resolution in the sodium citrate tribasic wt% axis. For example, if the initial screening resolution was 4wt% in the sodium citrate tribasic axis, then to cover this range of uncertainty with a 1wt% screen, 5 systems would need to be formed, an illustration of this is shown in Figure 4 - 5. To reduce the number of total experiments to fit onto a single 96 well plate, the screening resolution in the PEG wt% axis was set to a constant value of 2wt%.

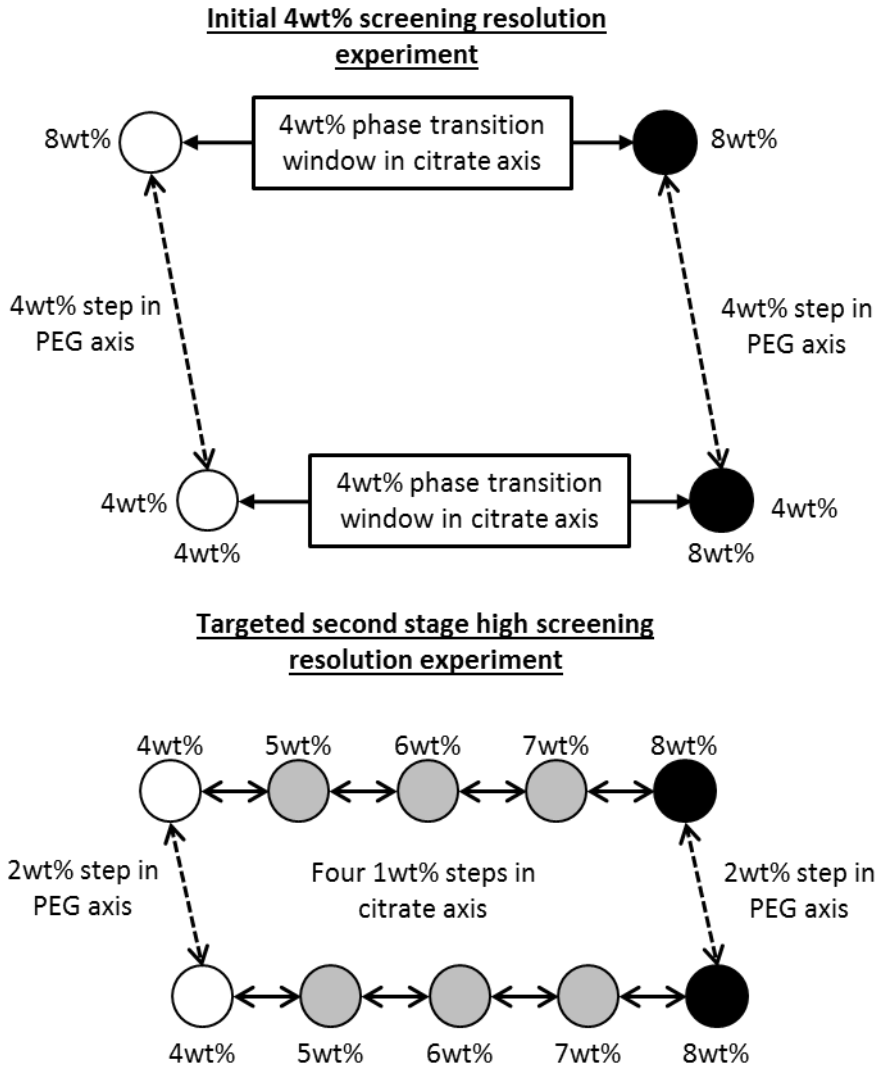


Figure 4 - 5: Targeted second stage screen covers initial range of uncertainty. The white circles correspond to systems which do not form a two-phase system. The black circles correspond to systems which do form a two-phase system. The grey circles correspond to systems which may form a two-phase system based on the initial low resolution screen.

4.2 Mathematical methods

4.2.1 Density correlation for process simulations

The density of phases was calculated using the following correlation which has been determined from densities used to determine mass ratios for tie-line determination:

$$\rho = 1017 + 231 \times x_{PEG} + 954 \times x_{phosphate} \quad (4.3)$$

x_{PEG} and $x_{phosphate}$ are mass fractions of PEG 2,000 and phosphate. A parity plot of the experimental and predicted density using Eq.(4.3) can be found in Appendix 7.2.

4.2.2 Confidence ellipsoids for estimated parameters

Confidence ellipsoids for estimated parameters were generated using the Parameter Estimation entity within gPROMS Model Builder 4.1 (Process Systems Enterprise, 2017). Confidence ellipsoids represent how the value of a single estimated parameter changes as a result of changes to another estimated parameter. Confidence ellipsoids are generated by gPROMS using the joint confidence region which also utilised the covariance values of parameters.

4.2.3 Phase ratio optimisation for process simulations

The phase ratio was optimised to achieve a yield of 90% using the Optimisation entity in gPROMS Model Builder 4.1 (Process Systems Enterprise, 2017). An end-point inequality constraint was used which ensures that the yield is between 89.9 and 90.1% at the end of the simulation.

4.2.4 Binodial curve fitting

To take into account the impact of screening resolution in the determination of the binodial curve (step 2 in the overall methodology), a range of single random points were picked from a uniform distribution where a transition from a single-phase to a two-phase system was observed. The random points were then tested to see if they were monotonic, if they were not, then they were rejected. It is important to reject non-monotonic points because the binodial curve must be a smooth monotonic function as each point on the binodial curve must have a unique composition (e.g. there cannot be multiple polymer compositions associated with a single salt composition). An illustration of monotonic and non-monotonic points was shown in Figure 4 - 4B.

To mathematically represent the binodial curve an exponential function was used which is described below by Eq.(4.4).

$$Polymer\ wt\% = A \times \exp(B \times Salt\ wt\%^{0.5} - C \times Salt\ wt\%^{3}) \quad (4.4)$$

Where A, B and C are parameters which are adjusted to fit the data and polymer and salt are the two components used to form an aqueous two-phase system such as PEG and phosphate. Eq.(4.4) has been used successfully numerous times to correlate binodial curves for aqueous two-phase systems (Merchuk et al., 1998, Bensch et al., 2007, Atefi et al., 2016, Alvarez-Guerra et al., 2016). A step by step method is presented next which corresponds to the steps of Protocol A in Figure 4 - 6:

Step 1) Set the number of desired fits, $N_{binodial}$, which corresponds to the number of times the binodial curve is repeatedly fitted

The first step is to set the number of times Eq.(4.4) will be fitted to the binodial curve transition window. The number, $N_{binodial}$, needs to be chosen such that there is negligible variation in the estimated values of A, B and C if the method in Figure 4 - 6 is repeated. This can easily be checked by calculating standard deviation for A, B and C as $N_{binodial}$ is increased. For the purpose of illustration, an $N_{binodial}$ value of 2000 will be used.

Steps 2,3 and 4) Pick random points within observed phase transition windows

A random point is picked in each observed transition window from a uniform distribution to create a set of transition points which represent the binodial curve. The point has to be randomly chosen as it is unclear where in the transition window the true binodial curve lies. The set of randomly picked transition points are tested to see if they are monotonic as shown in Figure 4 - 4B. If they are, then they are stored. However, if the picked points were not monotonic, they are discarded and new points are picked.

Step 5) Checking the number of stored sets of monotonic transition points

If the number of stored sets of monotonic transition points representing the binodial curve is less than the number of desired fits ($N_{binodial}$), Steps 2, 3 and 4 are repeated.

Step 6) Fit the binodial curve equation (Eq.(4.4)) to each set of monotonic transition points.

Eq.(4.4) is fitted to each set of randomly generated monotonic transition points by adjusting parameters A, B and C. The fitting process generates combinations of A, B and C parameters which are stored. Each unique combination of A,B and C describes a different binodial curve which could lie within the observed phase transition windows.

Step 7) Calculate mean and standard deviation of A, B and C parameters used in Eq.(4.4)

The mean and standard deviation of the $N_{binodial}$ stored values of A, B and C is calculated.

Steps 8 and 9) Calculate binodial curve confidence intervals

To calculate binodial curve confidence intervals each combination of estimated A, B and C parameters is used to predict polymer wt% using Eq.(4.4) at a range of salt wt% values. The predicted values are stored resulting in $N_{binodial}$ sets of predicted polymer wt% values. Lower and upper confidence intervals are determined by calculating the appropriate percentiles for the $N_{binodial}$ predicted polymer wt% values. Percentiles are cut-points which correspond to a cumulative probability distribution, e.g. the 2.5^{th} percentile would be the value at which the first 2.5% of the data is below. In this work, 95% confidence intervals were used, however, any confidence intervals can be generated using appropriate percentiles. For the 95% confidence intervals, 2.5 and 97.5 percentiles were used.

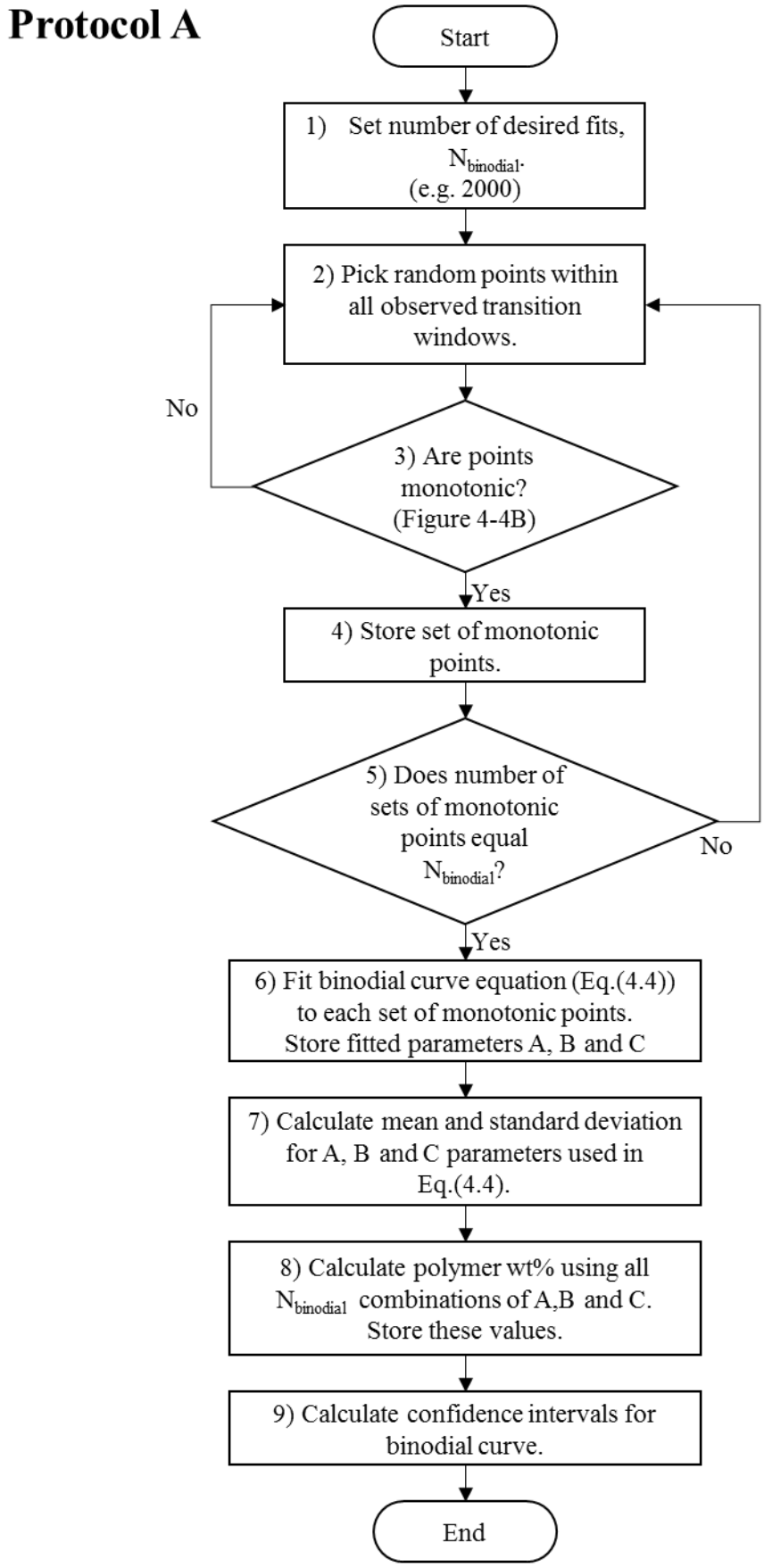


Figure 4 - 6: Protocol A for determining uncertainty in binodial curve due to screening resolution.

4.2.5 Determine equilibrium compositions by fitting tie-lines

To determine tie-lines for aqueous two-phase system phase diagrams, multiple systems were made in 15mL centrifuge tubes. The total mass of each prepared system was 5g. Systems were thoroughly mixed using a vortexer for 10 seconds at 2000RPM. They were than centrifuged for 10 minutes at 2000g to settle the phases. 250µL of each top and bottom phase are needed to determine the density of each phase. A minimum of 500µL of top and bottom phases were obtained using 19G and 21G, needles respectively. Bottom phase samples were extracted by passing the 21G needle through the top phase with a positive pressure, when the needle was a sufficient distance into the bottom phase the sample was aspirated. When the needle was withdrawn from the system through the top phase, a slight positive pressure was maintained which ensured that any top phase sample could not contaminate the bottom phase sample.

Mass ratios were determined using the method described by Amrhein et al. (2014) which is described using the following equations. Experimentally determined variables in Eq.(4.5)-(4.12) are highlighted in **bold.** Density in this work was measured by accurately dispensing solutions using a positive displacement pipette onto an analytical balance. Masses of stock solutions were measured using an analytical balance.

$$M_{top} = \boldsymbol{\rho_{top}} \times V_{top} \tag{4.5}$$
$$M_{bottom} = \boldsymbol{\rho_{bottom}} \times V_{bottom} \tag{4.6}$$

Eq.(4.5)-(4.6) are relations between density, mass and volume for top and bottom phases.

$$M_{total} = M_{top} + M_{bottom} \tag{4.7}$$
$$V_{total} = V_{top} + V_{bottom} \tag{4.8}$$

Eq.(4.7)-(4.8) are total system mass and volume balances.

$$M_{total} = \sum_{i=stock\ solution}^{Number\ of\ stock\ solutions} \boldsymbol{M_i} \tag{4.9}$$

Total mass is calculated using Eq.(4.9) where mass of individual stock solutions added is measured experimentally.

$$V_{total} = \sum_{i=stock\ solution}^{Number\ of\ stock\ solutions} V_i = \sum_{i=stock\ solution}^{Number\ of\ stock\ solutions} \frac{\boldsymbol{M_i}}{\boldsymbol{\rho_i}} \tag{4.10}$$

The sum of stock solution volumes is used to determine total volume using Eq.(4.10).

$$\emptyset = \frac{M_{top}}{M_{bottom}} \tag{4.11}$$

The phase ratio, the ratio between the mass of the top and bottom phase is calculated using Eq.(4.11).

$$M_{top} = \frac{(M_{total} - V_{total} \times \rho_{bottom}) \times \rho_{top}}{\rho_{top} - \rho_{bottom}} \tag{4.12}$$

Eq.(4.12) is a result of substituting Eq.(4.5)-(4.6) and Eq.(4.8) into Eq.(4.7). The volume of individual stock solutions added is calculated using experimentally determined density measurements.

A schematic of how the tie-lines were fitted to the binodial curve was shown in Figure 4 - 3B. The tie-line fitting process was automated using a MATLAB script which uses the intersection function created by Scwarz (2010). Example MATLAB script for this can be found in Appendix 7.6.

4.2.6 Calculation of uncertainty in tie-lines

There are two main sources of uncertainty in equilibrium compositions when fitting tie-lines using the method described above:

1. Uncertainty in equilibrium compositions due to the screening resolution used to determine the binodial curve.
2. Uncertainty in equilibrium compositions due to uncertainty in density measurements propagating into calculations of mass ratios through Eq.(4.5)-(4.12).

Uncertainty due to binodial

The procedure to determine the uncertainty in tie-line equilibrium compositions due to the binodial curve is outlined in Protocol B in Figure 4 - 7. The steps of the protocol B are explained below in more detail:

Step 1) Load stored estimated binodial curve parameter sets generated from when determining uncertainty in the binodial curve location

The first step involves loading all combinations of A, B and C parameters used in Eq.(4.4) which are generated using Protocol A in Figure 4 - 6.

Step 2) Load calculated mass ratios

The mass ratios for all systems used to determine tie-lines are loaded from experimental data.

Step 3) Fit tie-lines for each combination of A, B and C parameters

Tie-lines are fitted for each system using each combination of A, B and C parameters generated using Protocol A. For example if $N_{binodial}$ values of A, B and C parameters were estimated using Protocol A, a single tie-line would be fitted $N_{binodial}$ times to a binodial curve described by each of the $N_{binodial}$ combinations of A,B and C parameters resulting in $N_{binodial}$ different tie-lines.

Step 4) Remove infeasible binodial curve parameters

Some combinations of estimated A, B and C parameters result in predicted binodial curves which cannot satisfy the inverse lever arm rule when fitting tie-lines. The reason they cannot satisfy the inverse lever arm rule is because the tie-lines do not intersect the binodial curve twice when using these combinations of A, B and C parameters. This means that equilibrium compositions cannot be calculated. These parameters are

therefore removed from the initial set of estimated parameters generated using Protocol A.

Step 5) Re-calculate mean and standard deviation of estimated A,B and C parameters
The mean and standard deviation of estimated A, B and C parameter values is re-calculated since infeasible combinations have been removed.

Step 6) Calculate mean and standard deviation of tie-line equilibrium compositions
The mean and standard deviation of all tie-line equilibrium compositions is calculated based on tie-lines fitted to all feasible binodial curves.

Step 7) Determine confidence intervals based on standard deviation
Confidence intervals are determined for equilibrium compositions by multiplying standard deviation values of equilibrium compositions by the appropriate z score for a normal distribution. For example 95% confidence intervals for equilibrium compositions are determined by multiplying standard deviation values of equilibrium compositions by 1.96 as in a normal distribution 95% of the population is present within plus/minus 1.96 standard deviations.

Uncertainty due to density
The procedure to determine the uncertainty in tie-line equilibrium compositions due to the density is outlined in Protocol C in Figure 4 - 8. The mass ratio is calculated using various density measurements using Eq.(4.5)-(4.12) described above in section 4.2.5. An explanation of Protocol C is given below:

Step 1) Set the number of desired fits, $N_{density}$.
The first step is the set the number of times ($N_{density}$) a tie-line is fitted to a randomly generated mass ratio. $N_{density}$, must be set such that if Protocol C were repeated, there would be a negligible difference in mean and standard deviations of calculated equilibrium compositions. Again, this can be achieved by increasing the value of $N_{density}$ and calculating the mean and standard deviations of equilibrium compositions.

Step 2) Load mean binodial curve parameters (A, B and C) from Protocol B.
The mean values of A, B and C parameter values from Protocol B are used to describe the binodial curve when investigating uncertainty due to density measurements.

Step 3) Determine standard deviation of experimental volume and mass measurements

The standard deviation of experimental mass and volume measurements can be obtained by dividing the error in the mass and volume measurements by z score of 1.96. This assumes the errors correspond to a 95% confidence interval.

Step 4) Load experimental mass data

Experimental data for mass of stock solutions is loaded into MATLAB so that total system mass can be calculated. Experimental data for mass of 250µL of stock solutions is loaded so that stock solution densities can be calculated. Experimental data for mass of 250µL of top and bottom phase samples is loaded so that densities of top and bottom phase can be calculated.

Step 5) Generate $N_{density}$ random values of M_{total}, V_{total}, ρ_{top} and ρ_{bottom}

In this step, $N_{density}$ random values of total mass, total volume and phase densities are generated based on the experimental measurement and the corresponding measurement error.

For example, if a density of 1.250g mL^{-1} was calculated using an experimental mass of 1.250g and a volume of 1mL which have standard deviations of 0.001g and 0.01mL, respectively. Then a random density measurement could be generated by picking a random mass from a normal distribution with a mean of 1.250g and standard deviation of 0.001g and dividing it by a randomly picked volume from a normal distribution with a mean of 1mL and standard deviation of 0.01mL. A similar approach is used for total mass and volume.

Step 6) Calculate mass ratio

Using the $N_{density}$ random values of total mass, total volume and phase densities from Step 5, $N_{density}$ mass ratios can be calculated using Eq.(4.5)-(4.12).

Step 7) Determine tie-line equilibrium compositions

An individual tie-line is fitted to the binodial curve described by A, B and C parameters from Step 2 using each randomly generated mass ratio from Step 6. The result is a set of $N_{density}$ tie-line equilibrium compositions which have random variations due to propagation of the uncertainty in experimental mass and volume measurements through Eq.(4.5)-(4.12).

Step 8 and 9) Calculate mean, standard deviation and confidence intervals of tie-line equilibrium compositions

In this step, the mean and standard deviation of the $N_{density}$ tie-line equilibrium composition is calculated. 95% confidence intervals are calculated by multiplying equilibrium compositions standard deviations by a z score of 1.96.

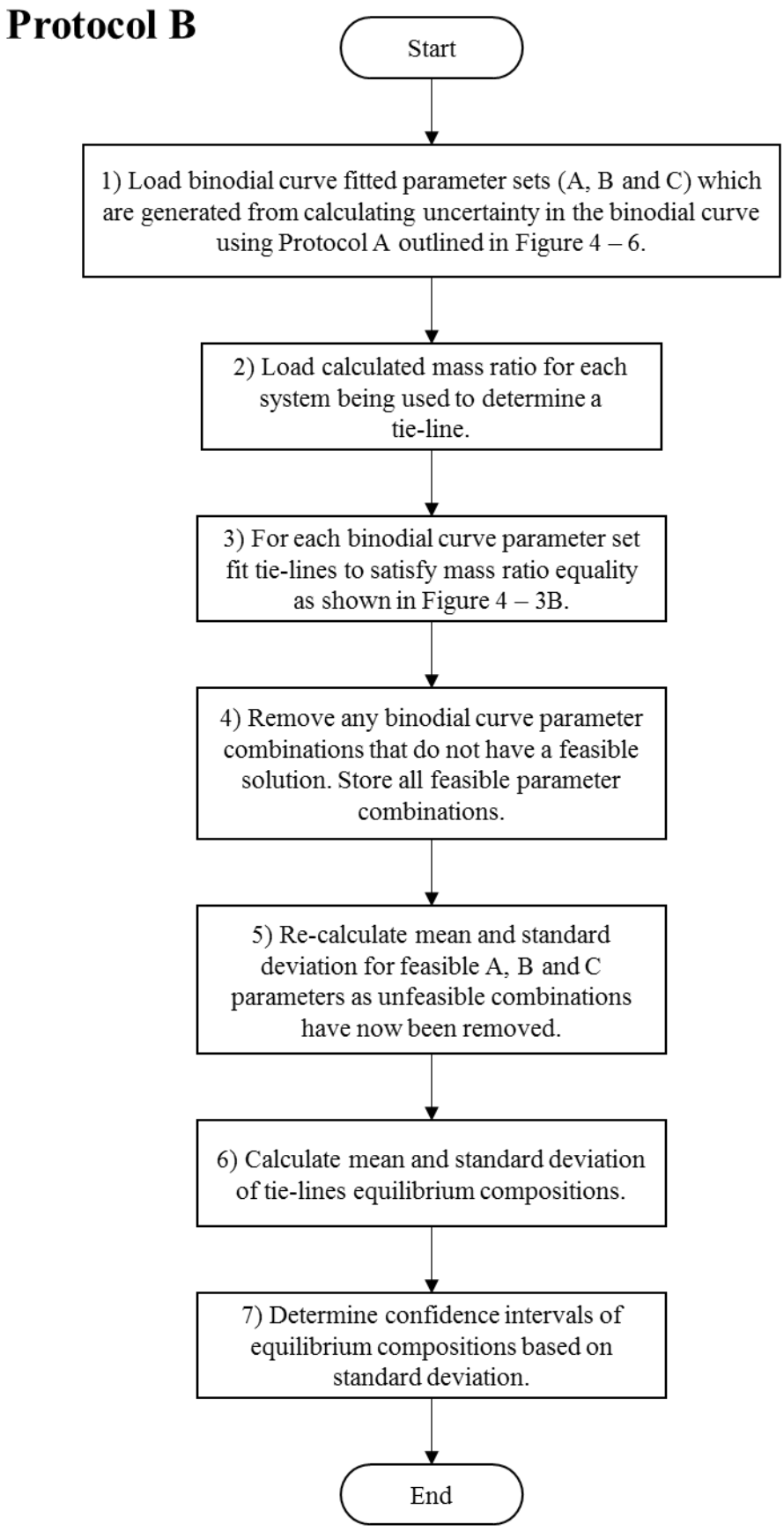


Figure 4 - 7: Protocol B for determining uncertainty in tie-lines due to binodial curve.

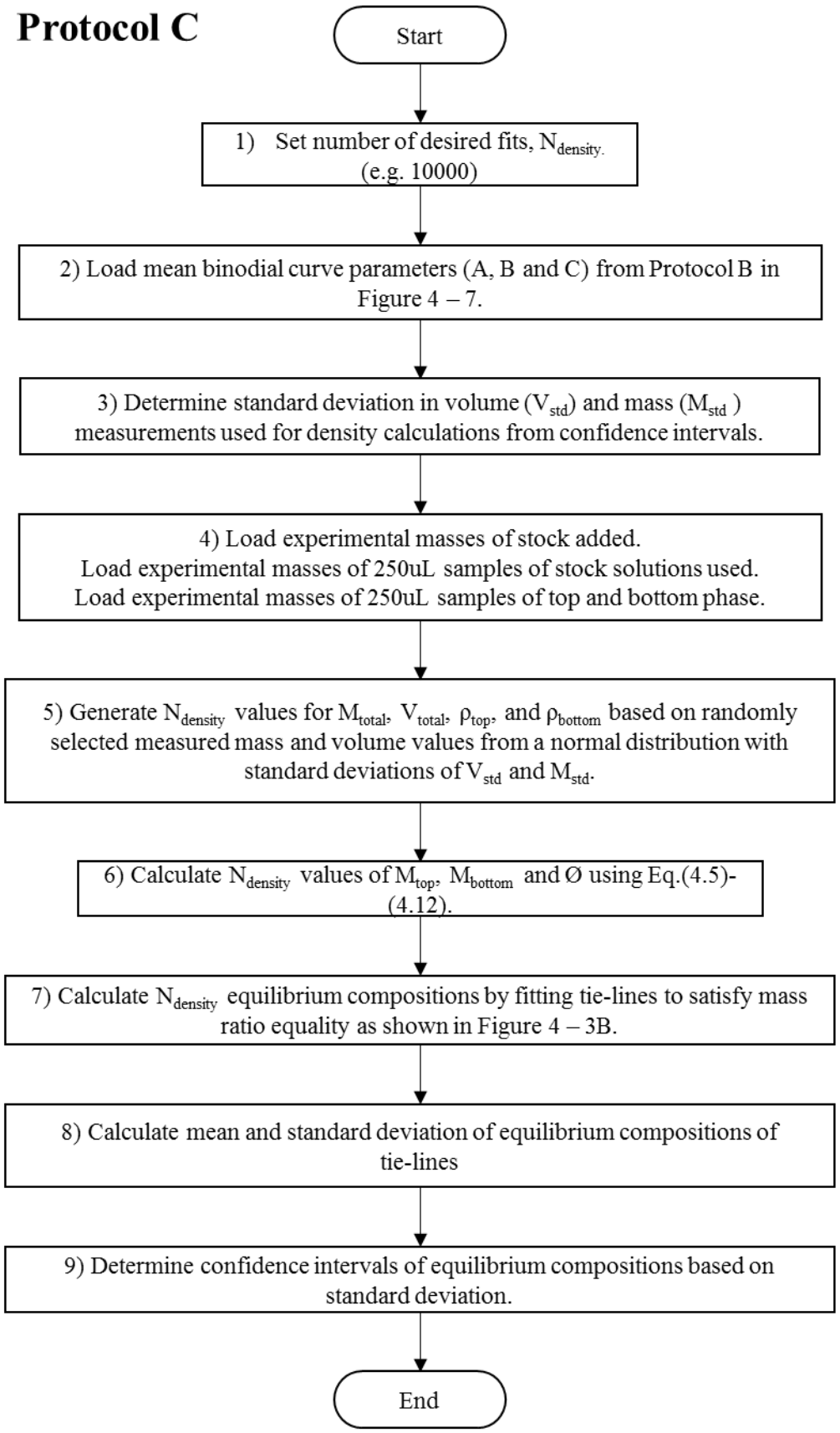


Figure 4 - 8: Protocol C for determining uncertainty in tie-lines due to the density measurements.

4.2.7 Liquid-liquid equilibria calculations

The liquid-liquid equilibria compositions needs to be calculated using given interaction parameters to determine how well the calculated liquid-liquid equilibria compositions fit the experimental data as shown in Steps 7 and 8 in Figure 4 - 1. In this, work liquid-liquid equilibria calculations were conducted using two methods:

1. Calculation of liquid-liquid equilibria compositions using the Rachford-Rice algorithm as shown in Figure 4 - 10. Such an approach is referred to as an equation-decoupling or substitution method (Joulia et al., 1986, Teh and Rangaiah, 2002).
2. Calculation of liquid-liquid equilibria compositions by solving all mass balance and thermodynamic equations simultaneously. Such an approach is referred to as a simultaneous equation solving or Newton type method (Joulia et al., 1986, Teh and Rangaiah, 2002, Adewumi, 2017).

The Rachford-Rice algorithm was used when generating initial estimates of interaction parameters using Protocol D shown in Figure 4 - 11. This decision was made because equation-decoupling methods are more robust at converging to a solution (compared to Newton type methods) when initial guesses are not well known (Joulia et al., 1986, Teh and Rangaiah, 2002, Adewumi, 2017). Teh and Rangaiah (2002) have shown that methods based on the Rachford-Rice algorithm are reliable for conducting phase equilibria calculations in vapour-liquid and liquid-liquid equilibrium systems.

gPROMS ModelBuilder 4.1.0 (Process Systems Enterprise, 2017) on the other hand, finds numerical solutions to sets of equations by solving all equations simultaneously. The main disadvantage of such an approach is that, in the context of this work, the numerical solvers need good initial guesses to initialise and find a solution when conducting liquid-liquid equilibria calculations. When conducting process simulations, the need for excellent initial guesses can be addressed in part by using appropriate initialisation protocols (see Appendix 7.8 for more information on initialisation protocols) (Pantelides et al., 2015). However, for parameter estimation this is more difficult for two main reasons:

1) Estimates of initial parameters are often not known therefore need to be guessed.
2) If parameters are guessed, initial values of equilibrium compositions that will allow for successful initialisation of the numerical solvers are unknown.

The main advantage of using the Parameter Estimation entity within gPROMS ModelBuilder 4.1.0 is that it can take into account the impact of experimental uncertainty on estimation of parameters and provide useful statistical information. Therefore, the Parameter Estimation entity within gPROMS was used once good estimates of interaction parameters (β_{ij}) were obtained using Protocol D shown in Figure 4 - 11.

Rachford-Rice algorithm for liquid-liquid equilibria calculations

The Rachford-Rice objective function (Eq.(4.22)) was used for calculation of liquid-liquid equilibria as it is a more robust objective function for iterative liquid-liquid equilibria calculations (Prausnitz, 1980). This is particularly important when estimating interaction parameters for thermodynamic equations as explained at the start of section 4.2.7. A derivation of the Rachford-Rice objective function is given below:

Figure 4 - 9: Liquid-liquid equilibrium stage used to derive Rachford-Rice objective function

Assumptions in derivation of Rachford-Rice objective function

1. The liquid-liquid equilibrium stage is assumed to be operating at steady-state.
2. The phase separation is 100% efficient and occurs instantly.
3. The liquid-liquid extraction operation is isothermal and isobaric.

$$F_{in}w_i = F_{top}y_i + F_{bottom}x_i \qquad for\ i = 1 \ldots NC \qquad (4.13)$$

Eq.(4.13) is a component mass balance around a single liquid-liquid extraction stage.

$$\sum_{i}^{NC} x_i = 1 \qquad (4.14)$$

$$\sum_{i}^{NC} y_i = 1 \qquad (4.15)$$

Eq.(4.14) and Eq.(4.15) represent the condition that mass fractions must add up to1.

$$F_{in} = F_{top} + F_{bottom} \qquad (4.16)$$

Eq.(4.16) is a total mass balance around a single liquid-liquid extraction stage.

$$\alpha \times F_{in} = F_{top} \tag{4.17}$$

Eq.(4.17) is a definition of α, the fraction of feed that ends up in the top phase, in Eq.(4.22).

$$K_i \times x_i = y_i \qquad for\ i = 1 \ldots NC \tag{4.18}$$

Eq.(4.18) is a definition of the component partition coefficient K_i.

Substituting Eq.(4.18) and Eq.(4.16) into Eq.(4.13) results in Eq.(4.19):

$$F_{in}\left(\frac{w_i}{x_i} - 1\right) = F_{top}(K_i - 1) \qquad for\ i = 1 \ldots NC \tag{4.19}$$

Then re-arranging Eq.(4.19) results in Eq.(4.20):

$$x_i = \frac{w_i}{1 + \alpha(K_i - 1)} \qquad for\ i = 1 \ldots NC \tag{4.20}$$

Eq.(4.14) minus Eq.(4.15) results in Eq.(4.21):

$$\sum_{i=1}^{NC} x_i - y_i = 0 \tag{4.21}$$

Substituting Eq.(4.20) and Eq.(4.18) into Eq.(4.21) results in Eq.(4.22) which is the Rachford-Rice objective function:

$$\sum_{i}^{NC} \frac{(K_i - 1) \times w_i}{(K_i - 1) \times \alpha - 1} = 0 \tag{4.22}$$

The method for conducting iterative flash calculations using Eq.(4.18), Eq.(4.20) and Eq.(4.22) is outlined in Figure 4 - 10 with a detailed explanation of each step described below.

Step 1) Assign interaction parameters

First interaction parameters (β_{ij}) used in the osmotic virial equations (Eq.(3.21)-(3.23)) are assigned. The value of these parameters will dictate the calculated equilibrium compositions for a given feed composition.

Step 2) Assign feed stream compositions

The mass fraction of components in the feed stream (w_i) is assigned. Typically the components are a polymer, salt and water.

Step 3) Guess a value for K_i

A value for K_i is guessed as the final equilibrium K_i is unknown. If experimental tie-lines are available then they can be used generate a reasonable initial guess for K_i by dividing mass fractions of components in the top phase by mass fractions of components in the bottom phase.

Step 4) Calculate fraction of feed in top phase (α)

The fraction of feed flow that ends up in the top phase (α) can now be determine because the feed composition (w_i) and component partition coefficients (K_i) have been defined. The value of α is determined by solving Eq.(4.22) such that the left-hand side equals zero.

Step 5) Calculate mass fraction of components in top and bottom phase

The top and bottom phase mass fractions are calculated using Eq.(4.18) and Eq.(4.20).

Step 6) Calculate component chemical potential values

Chemical potential values of components in the top and bottom phase are calculated using the mass fractions calculated in Step 5 using osmotic virial equations (Eq.(3.21)-(3.23)). The osmotic virial equations are written in terms of molality therefore an appropriate conversions between units must be made.

Steps 7,8 and 9) Test component chemical potentials are at equilibrium

At equilibrium, the chemical potentials for a specific component in both the top and bottom phases is equal. Therefore, to test for thermodynamic equilibrium the difference between top and bottom phase chemical potentials for each component is calculated.

If there is a difference in the chemical potential, then new K_i value needs to be determined as shown in Step 8. If, however, the difference in chemical potential values is zero, then equilibrium mass fraction values have been found.

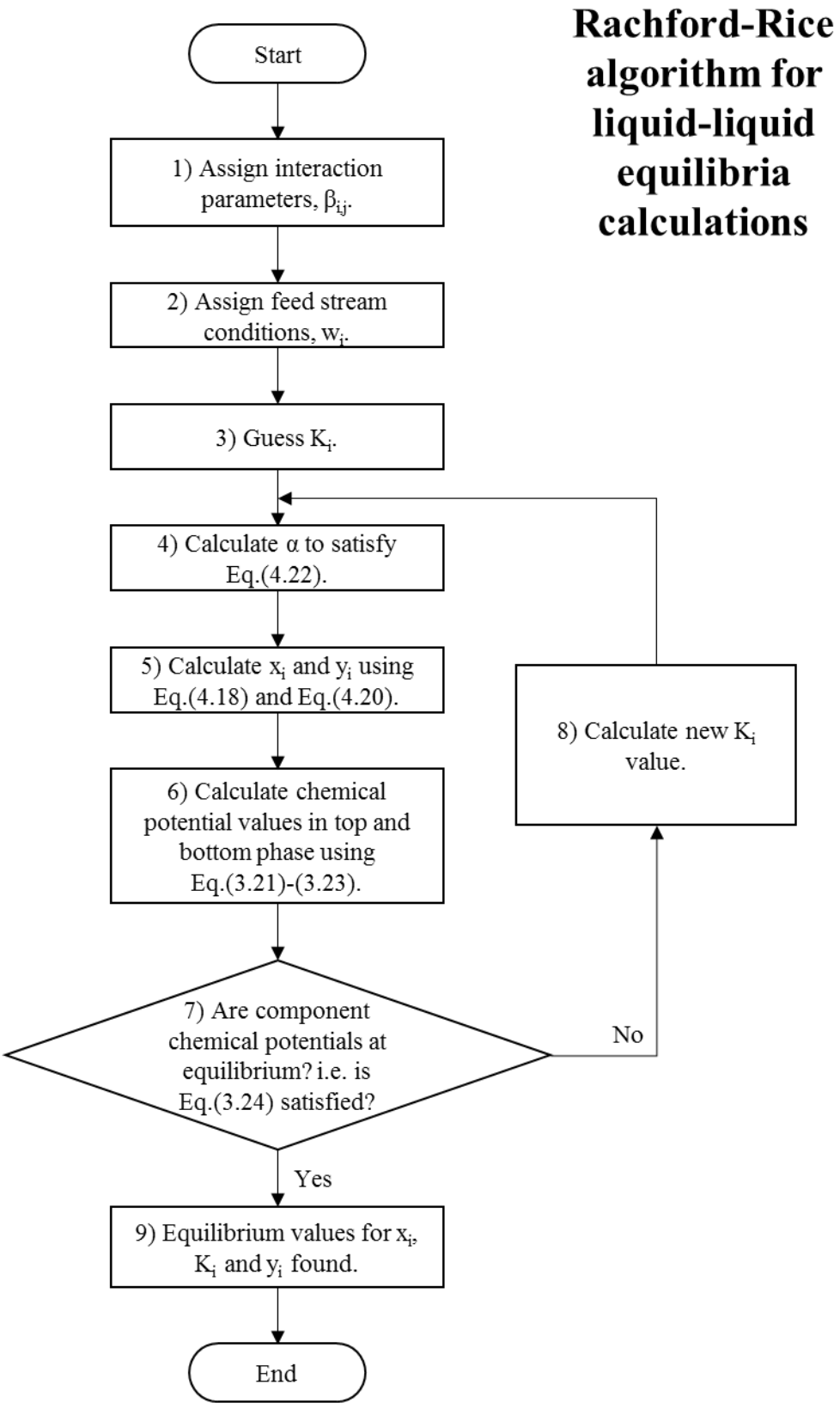


Figure 4 - 10: Rachford-Rice algorithm for liquid-liquid equilibria calculations.

4.2.8 Estimation of interaction parameters

Interaction parameters in this work were first estimated using the Rachford-Rice algorithm combined with a simulated annealing algorithm. The estimates from this approach were then used as initial guesses using Maximum Likelihood Parameter estimation within gPROMS. Details of each approach are discussed next.

Rachford-Rice algorithm combined with simulated annealing

Interaction parameters in Eq.(3.21)-(3.23) were initially estimated in MATLAB using the simulated annealing algorithm within the Global Optimisation Toolbox (MathWorks®, 2017). Simulated annealing optimisation algorithms are heuristic/probabilistic optimisation algorithms which are good at avoiding local minima (Eglese, 1990). More information on simulated annealing algorithms can be found in Appendix 7.7.

$$Objective\ function = 1 \times 10^5 \times \sum_{i}^{NT} \sum_{j}^{NC} \sum_{k}^{NP} \left(x_{i,j,k}^{model} - x_{i,j,k}^{experimental} \right)^2 \quad (4.23)$$

Eq.(4.23) is a least squares objective function which was used when estimating parameters using the simulated annealing optimisation algorithm. The objective function is a sum of the squared differences between the model predicted and experimental phase equilibria compositions. The objective function was multiplied by 1×10^5 to scale it to a value closer to 1 for better numerical performance. A better way to have scaled the objective function would have been to divide the difference terms in Eq.(4.23) by $x^{experimental}$. The aim of the optimisation algorithm is to minimise the value of Eq.(4.23) so that model predictions closely match experimental data. Protocol D in Figure 4 - 11 outlines the steps taken to estimate initial sets of interaction parameters. An explanation of each step is given below.

Step 1) Assign feed stream conditions

The overall composition of the feed stream is assigned in terms of mass fractions for each component.

Step 2) Guess initial interaction parameter (β_{ij}) values

Initial values of interaction parameters are guessed which will be used as the starting values when running the simulated annealing optimisation algorithm.

Step 3) Calculate model predicted liquid-liquid equilibria compositions

In this step, the liquid-liquid compositions are calculated using the Rachford-Rice algorithm described previously in Figure 4 - 10.

Step 4) Calculate least squares objective function for simulated annealing optimisation

In this step experimental phase equilibria data and model predicted phase equilibria data is used to calculated the least squares objective function outlined in Eq.(4.23). The value of this objective function is used by the simulated annealing algorithm to determine new interaction parameter values.

Steps 5,6 and 7) Check termination criteria in the simulated annealing optimisation algorithm is satisfied

In this step, various termination criteria are checked to see if the simulated annealing optimisation algorithm should terminate. Termination of the algorithm can occur due to:

1. Maximum number of iterations is reached.
2. Maximum number of function evaluations is reached.
3. Time limit is reached.
4. Function tolerance is reached.
5. Objective function limit is reached.

However, in this work the simulated annealing algorithm will terminate if the average change in Eq.(4.23) is less than the function tolerance (10^{-6}) or if the number of objective function evaluations exceeds 9,000. If the termination criteria are satisfied, then the interaction parameters are found. If, however, termination criteria are not satisfied, new interaction parameter values are generated in Step 7. The termination criteria were based on the default values within MATLABs global optimisation toolbox. Making the function tolerance more stringent did not improve estimated parameter values as the change in estimated parameter values was deemed to be negligible. Typically the optimisation would end as a result of meeting the function tolerance criteria being satisfied rather than the maximum number of function evaluations being reached as the objective function value was not improved any further by new estimated parameter values.

Maximum likelihood parameter estimation

The gPROMS Parameter Estimation entity was used to estimate parameters with confidence intervals based on experimental error in the form of standard deviations as shown previously in Step 8 in Figure 4 - 1. A good initial guess for the fitted parameters

was obtained using the Protocol D. The parameter estimation entity in gPROMS uses the maximum likelihood objective function:

$$Objective\ function = \frac{N}{2}\ln(2\pi) + \frac{1}{2}\sum_{i}^{NT}\sum_{j}^{NC}\sum_{k}^{NP}\left(\ln\left(\sigma_{ijk}^{2}\right) + \frac{\left(x_{i,j,k}^{model} - x_{i,j,k}^{experimental}\right)^{2}}{\sigma_{ijk}^{2}}\right) \quad (4.24)$$

This objective function takes into account the standard deviation of measured variables and weights the fit of the model to the areas of highest certainty.

The standard deviations of equilibrium compositions were obtained from those calculated when determining uncertainty in the phase diagram using Protocol B and C. A combined variance which takes into account all calculated uncertainty is obtained using Eq.(4.25)

$$\sigma_{total}^{2} = \frac{\sigma_{binodial}^{2} + \sigma_{density}^{2}}{2} + \left(\frac{\mu_{binodial} - \mu_{density}}{2}\right)^{2} \quad (4.25)$$

where $\sigma_{binodial}$ is the standard deviation in equilibrium compositions due to the uncertainty in the binodial curve and $\sigma_{density}$ is the standard deviation in equilibrium compositions due to the uncertainty in the density measurements when determining the mass ratio. $\mu_{binodial}$ is the mean equilibrium composition when taking into account uncertainty due to the binodial curve and $\mu_{density}$ is the mean equilibrium compositions when taking into account uncertainty due to density. Eq.(4.25) is based on the combination of two variances from equal sized large populations. A full derivation of Eq.(4.25) is given by Burke (2014) and can be found in Appendix 7.9.

Finally, for tie-lines where more than 99.99% of polymer was in the top phase, the composition of polymer in the bottom phase was assumed to be 0.01wt% with a 99.7% confidence interval of 0.01wt% which corresponds to a standard deviation of 0.003wt%. This was to ensure numerical robustness during parameter estimation activities.

Software and numerical solvers

MATLAB parameter estimation

The simulated annealing algorithm from the Global Optimisation Toolbox within MATLAB R2015b was used. Specific algorithm settings which were used are reported in Appendix 7.7.

gPROMS parameter estimation

The NLPSQP (nonlinear programming sequential quadratic programming) solver in gPROMS was used for the parameter estimation. A description of the NLPSQP solver can be found in Appendix 7.5.

Protocol D

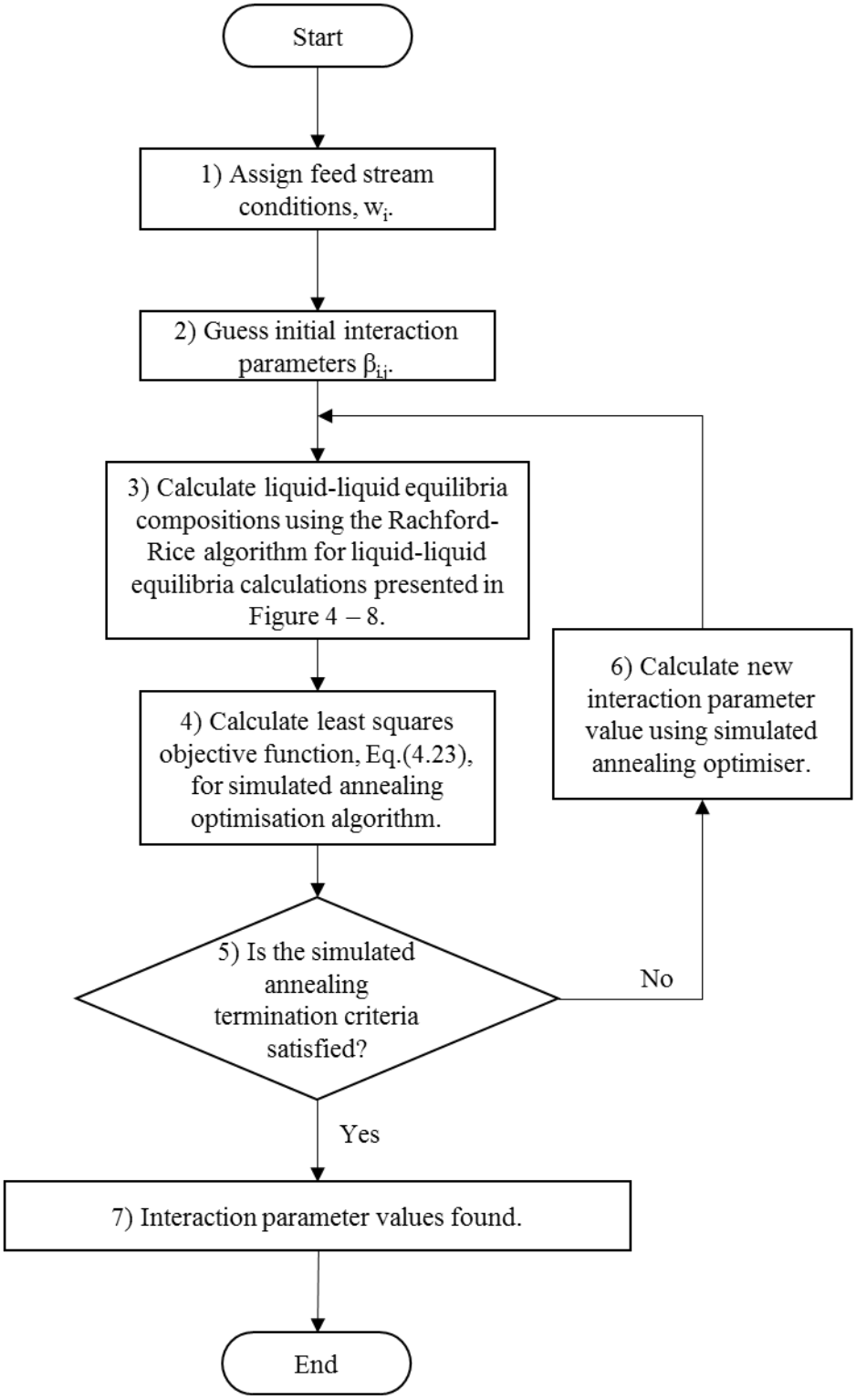


Figure 4 - 11: Protocol D for estimation of interaction parameters using simulated annealing and Rachford-Rice algorithms.

4.2.9 Tie-line interpolation

In this work four experimental tie-lines were determined. Often estimates of parameters can be improved by increasing the quantity of experimental data used for parameter estimation. Therefore, additional tie-lines were interpolated to quantify the impact of additional experimental data on uncertainty of estimated interaction parameters. Note that the interpolation of tie-lines is not included in the overall methodology presented in Figure 4 - 1. The new tie-lines were fitted by interpolating an average gradient between two tie-lines and using this gradient to determine the equilibrium compositions of the new tie-line using the fitting method illustrated in Figure 4 - 3B. The standard deviation of equilibrium compositions was also interpolated as the average between the two tie-lines used for interpolation. An example illustration of this is shown in Figure 4 - 12. In this work, the four experimental tie-lines were used to interpolate 3 additional tie-lines resulting in 7 tie-lines and the resulting 7 tie-lines were used to interpolate 6 additional tie-lines. This procedure was repeated until there was a total of 49 tie-lines. This process could be repeated indefinitely, however, was stopped at 49 tie-lines as this was deemed far in excess of the number of tie-lines that would practically be determined experimentally.

4.2.10 Summary of mathematical methods

This section covered the details of the mathematical methods used in the overall methodology presented in Figure 4 - 1. This involved the development of:

1. A method to calculate uncertainty in phase diagrams due to the screening design used to determine the binodial curve.
2. A method to calculate uncertainty in phase diagrams due to errors in density measurements.
3. The fully characterised phase diagrams were then used to estimate interaction parameters using a combination of the Rachford-Rice and simulated annealing algorithms.
4. Estimates of interaction parameters were then used as initial guesses when performing Maximum Likelihood Parameter estimation within gPROMS which was used to determine the statistical significance of estimated parameters.

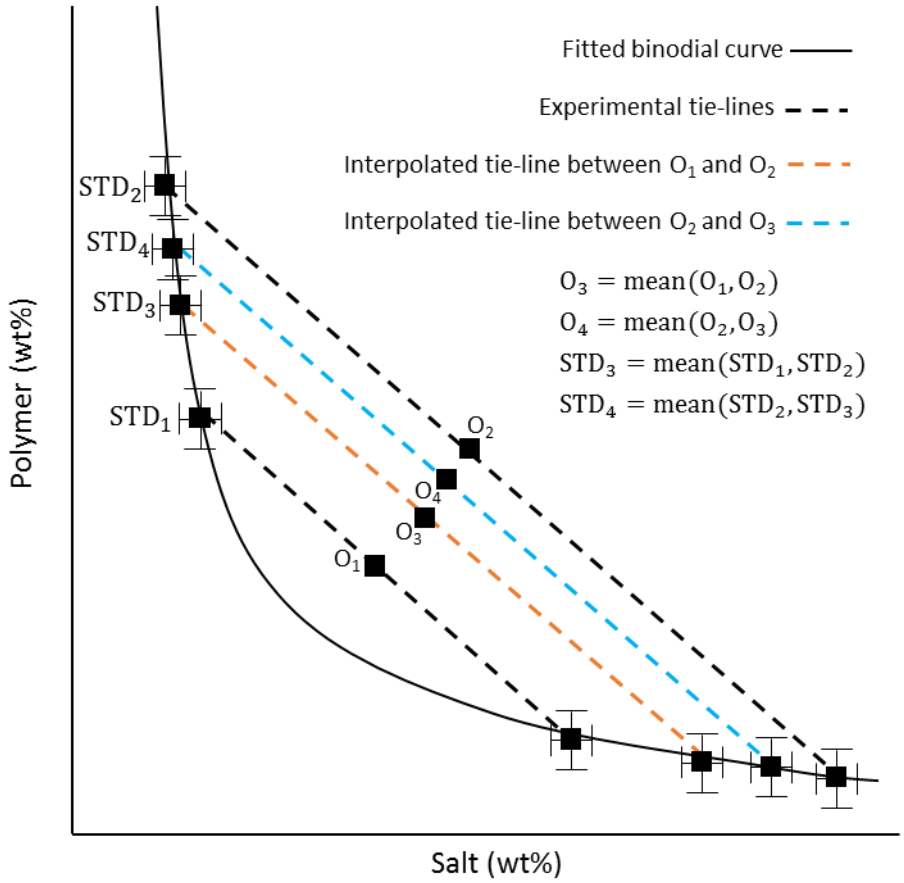


Figure 4 - 12: Illustration to show how tie-lines were interpolated.

4.3 Case study: PEG 4,000-citrate phase diagram

A case study is conducted to illustrate the application of the overall methodology presented in Figure 4 - 1 which involves the determination of uncertainty in experimentally determined phase diagrams for aqueous two-phase systems and the impact this experimental uncertainty has on accuracy/uncertainty of estimated of interaction parameters. Often parameter estimation is conducted without taking into account experimental limitations, however, this work shows the importance of taking into account experimental limitations encountered when determining the binodial curve and fitting tie-lines. The aqueous two-phase system considered for the case study is made up of PEG 4,000 and citrate tribasic in the presence of water.

Thermodynamic framework

The thermodynamic framework from Chapter 3 was used which uses the osmotic virial equations to describe chemical potentials of polymer, salt and water. The only difference is the components considered in this work are PEG 4,000, sodium citrate tribasic and water.

Components

The system considered in this chapter consists of three components:

1. PEG 4,000 (component 1)
2. Sodium citrate tribasic (component 2)
3. Water (component 3)

The molecular weight of PEG 4,000, Sodium citrate tribasic and water are 4000g mol^{-1}, 258.07g mol^{-1} and 18g mol^{-1}, respectively.

4.3.1 Two-stage binodial curve screening

A two-stage binodial curve screening process was developed so that the binodial curve could be determined to a high level of certainty. This process is the first step in the proposed overall methodology presented earlier in Figure 4 - 1. A two-stage approach is beneficial because it makes more efficient use of experimental resources while still obtaining high quality experimental data. This is because it reduces the number of experimental systems which would provide no additional information on the location of the binodial curve.

Figure 4 - 13 shows an initial screen, using a 4wt% screening resolution, used to find an approximate location of the binodial curve for a PEG 4,000-sodium citrate tribasic aqueous two-phase system.

The binodial curve equation (Eq.(4.4)) was fitted to this phase transition location using the method described in Protocol A in Figure 4 - 6. Using this approximate location of the binodial curve allowed for a detailed screen, using a 1wt% and 2wt% screening resolution in the sodium citrate tribasic and PEG 4,000 axes, respectively, to be targeted to the area of the greatest experimental information. This is shown in Figure 4 - 14 where the phase transitions all occur within the area screened. This approach used a total of 142 aqueous two-phase systems, 49 using the 4wt% screen and 93 using the targeted screen. If a detailed screen was conducted, without approximating using a two-stage approach, a total of 350 systems would be needed.

The time taken to prepare all 142 systems using automated liquid handling was less than 2 hours. To obtain the binodial curve using traditional titration/dilution (e.g. zig-zag method previously described in Section 2.2.1) methods would take considerably longer as at least two titrations would need to be performed starting from opposite ends of the binodial curve to cover the entire phase boundary. Furthermore, the masses of diluent, polymer and salt stock solutions added during turbidometric titration methods needs to be recorded meticulously as any errors will results in incorrect calculation of subsequent phase transition areas, this is especially the case for the zig-zag method. Another benefit of the multi-stage screening method using automated liquid handling is that the desired level of accuracy in the binodial can be set by the researcher.

The benefit of reducing the uncertainty in the binodial curve can be visually seen by the reduction in the 95% confidence intervals shown in Figure 4 - 13B and Figure 4 - 14B. In addition, Table 4 - 1 shows that standard deviations of the A, B and C fitting parameters used in the Eq.(4.4) for the binodial curve are reduced by more than a third.

Table 4 - 1: Parameters for fitted binodial curve equation (Eq.(4.4)) using 4wt% initial screen and targeted screen using a 1wt% and 2wt% screening resolution in the sodium citrate tribasic and PEG 4,000 axes respectively.

Parameters in Eq.(4.4)	Binodial curve screening resolution			
	4wt% initial screen		Targeted screen	
	Value	Standard Deviation	Value	Standard Deviation
A	73.5	26.2	78.8	8.0
B	-0.481	0.165	-0.475	0.050
C	5×10^{-4}	2.9×10^{-4}	5×10^{-4}	6.3×10^{-5}

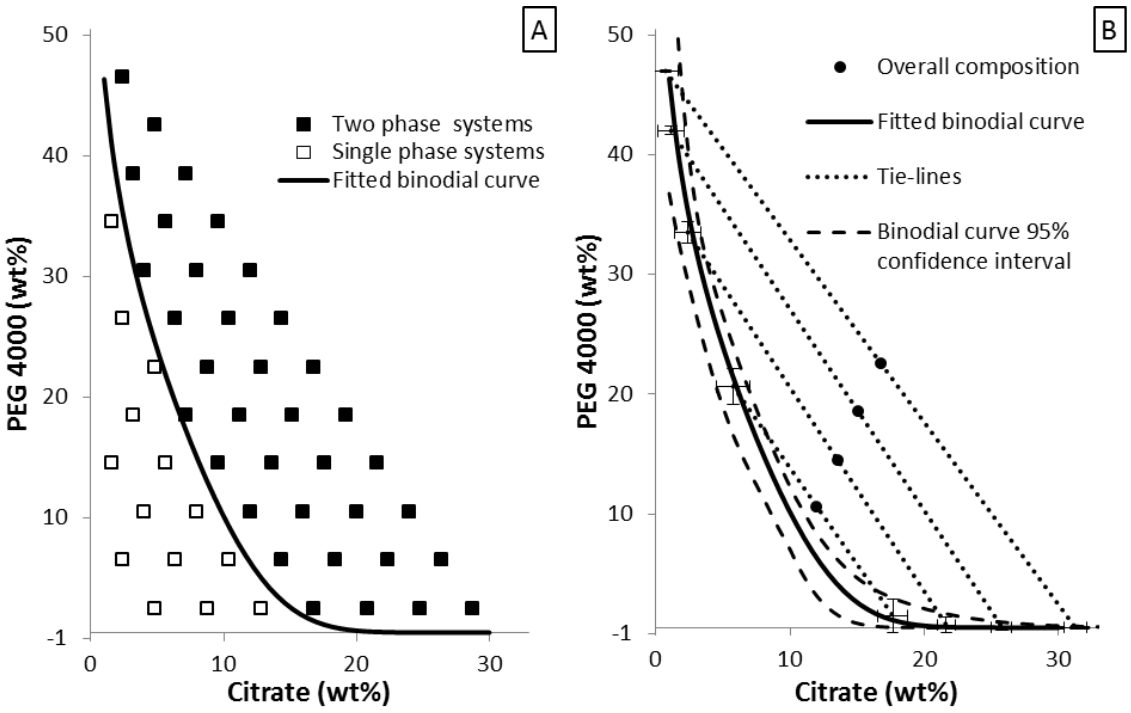


Figure 4 - 13: A) Binodial curve generated using a screening resolution size of 4wt% for a PEG 4,000-sodium citrate tribasic aqueous two-phase system. B) Phase diagram generated by fitting tie-lines to binodial curve obtained using 4wt% screening resolution using Protocol A in Figure 4 - 1. Error bars in tie-lines are 95% confidence intervals based on standard deviations of equilibrium compositions generated using Protocol B in Figure 4 - 7.

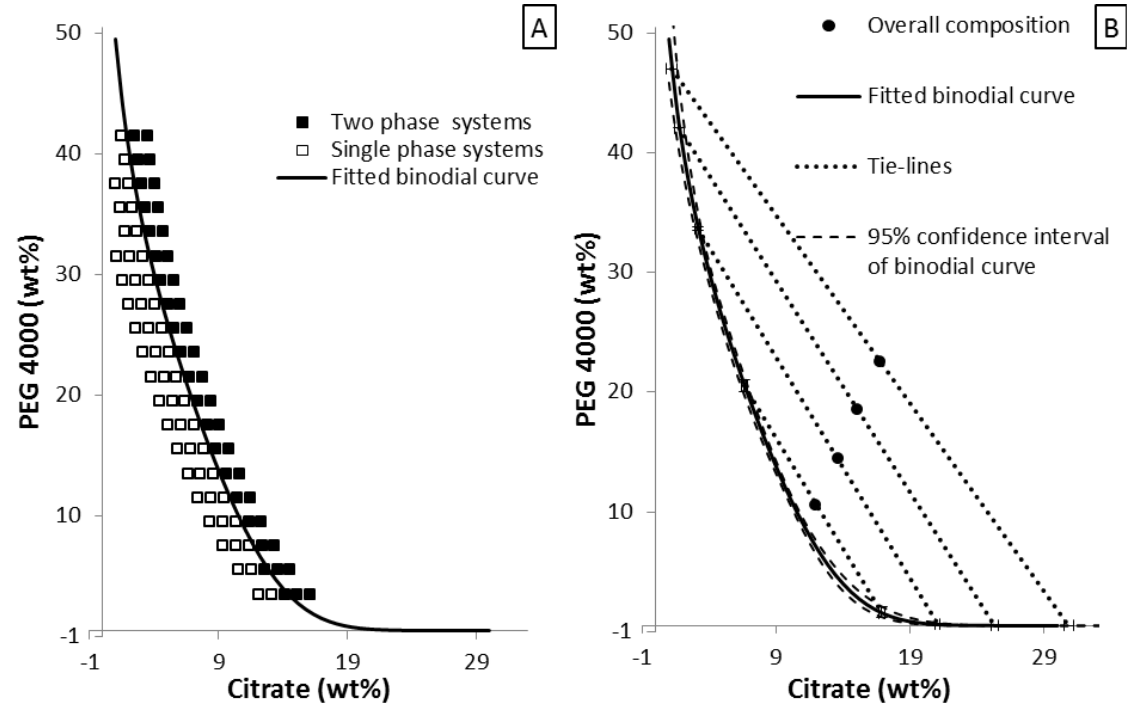


Figure 4 - 14: A) Binodial curve generated using a screening resolution size of 1wt% for a PEG 4,000-sodium citrate tribasic aqueous two-phase system. Experiment design based on initial screen shown in Figure 4 - 13A. B) Phase diagram generated by fitting tie-lines to binodial curve obtained using targeted screen using Protocol A. Error bars in tie-lines are 95% confidence intervals based on standard deviations of equilibrium compositions generated using Protocol B in Figure 4 - 7.

4.3.2 Uncertainty in equilibrium compositions

In this work, equilibrium compositions are determined by fitting tie-lines to the binodial curve. To fit tie-lines to the binodial curve requires knowledge of the mass ratio of a specific aqueous two-phase system. Therefore, the uncertainty in the equilibrium compositions will depend on the uncertainty in the binodial curve location as well as the uncertainty in the value of the experimentally determined mass ratio.

Uncertainty due to the binodial curve

Uncertainty in the binodial curve location is directly dependent on the screening resolution used. This uncertainty will then be propagated to uncertainty in the fitted tie-lines and hence the resultant equilibrium compositions, as was shown in Figure 4 - 13B and Figure 4 - 14B. It can clearly be seen in Figure 4 - 13 and Figure 4 - 14 that increasing the resolution of the binodial curve reduces the error bars of the equilibrium compositions significantly for all tie-lines. This will have an impact on the total uncertainty in equilibrium compositions which is calculated using Eq.(4.25).

Table 4 - 2 shows the mean equilibrium compositions and corresponding standard deviations for the four experimental tie-lines using the initial 4wt% binodial curve screen and the targeted screen. For the longest tie-line, the standard deviations of PEG 4,000 equilibrium compositions is reduced by more than 99% when using the targeted binodial curve screen, whereas, for the shortest tie-line there is only a 67% reduction in standard deviations for PEG 4,000 equilibrium compositions. Note that the reduction in standard deviation for PEG 4,000 in the bottom phase must be taken with caution as the amount of PEG 4,000 in the bottom phase is very low. The opposite trend is observed for the standard deviations of sodium citrate tribasic equilibrium compositions where the shortest tie-line has an 81% reduction in standard deviations of equilibrium compositions whereas the longest only has a 58% reduction in standard deviations of equilibrium compositions. It is unclear why this trend is observed but one suggestion could be due to the asymmetrical curvature of the binodial curve and the impact this has on the sensitivity of fitting tie-lines.

Table 4 - 2: Equilibrium compositions and their corresponding standard deviations when using a 4wt% initial screen and targeted screen using a 1wt% and 2wt% screening resolution in the sodium citrate tribasic and PEG 4,000 axes, respectively. This data does not take into account uncertainty due to mass ratio as a result of density measurements. Experimental equilibrium compositions are listed in order of descending tie-line length.

		Equilibrium compositions (wt%)			
Tie-line #		Top		Bottom	
		PEG	**Citrate**	**PEG**	**Citrate**
1	4 wt% initial screen	46.49 ± 0.06	0.72 ± 0.48	0.01 ± 0.05	31.24 ± 0.43
	Targeted screen	$46.50 \pm 7.72\times10^{-5}$	1.2 ± 0.2	$1.96\times10^{-5} \pm 6.93\times10^{-5}$	30.81 ± 0.18
	% Reduction in standard deviations	99.9	58.3	99.9	58.3
2	4 wt% initial screen	41.49 ± 0.18	1.14 ± 0.51	0.05 ± 0.14	25.77 ± 0.39
	Targeted screen	41.55 ± 0.01	1.76 ± 0.18	$4.24\times10^{-3} \pm 0.01$	25.29 ± 0.14
	% Reduction in standard deviations	96.3	64.3	96.3	64.3
3	4 wt% initial screen	33.01 ± 0.44	2.38 ± 0.48	0.22 ± 0.31	21.63 ± 0.34
	Targeted screen	33.17 ± 0.08	3.16 ± 0.13	0.1 ± 0.05	21.08 ± 0.09
	% Reduction in standard deviations	82.6	73.4	82.6	73.4
4	4 wt% initial screen	20.14 ± 0.78	5.77 ± 0.62	0.99 ± 0.7	17.65 ± 0.56
	Targeted screen	20.04 ± 0.25	6.62 ± 0.12	1.09 ± 0.23	16.88 ± 0.11
	% Reduction in standard deviations	67.4	81.0	67.4	81.0

Uncertainty due to mass ratio

The major source of uncertainty in this work was found to be due to the mass ratio. This is because various density measurements are used to calculate the mass ratio and the uncertainty associated with these are propagated through Eq.(4.5)-(4.12) which are used to calculate the masses of phases. Protocol C in Figure 4 - 8 was used to quantify the uncertainty in equilibrium compositions due to the calculated mass ratio.

Table 4 - 3 shows how the uncertainty in the mass ratio is influenced by the difference in density of the top and bottom phase. It was found that the standard deviation of the mass ratio increases from 0.05 (longest tie-line length) to 0.30 (shortest tie-line length) even though the standard deviations of the density is constant. This can be attributed to the fact that the density difference between phases is smaller at shorter tie-line lengths.

Table 4 - 3: Experimental densities and calculated mass ratio values of samples of top and bottom phases for PEG 4,000-sodium citrate tribasic aqueous two-phase system. (System 1-4 are in order of decreasing tie-line length.)

System	Top Density (g ml^{-1})		Bottom Density (g ml^{-1})		Mass Ratio (-)	
#	Value	Standard Deviation	Value	Standard Deviation	Value	Standard Deviation
1	1.090	0.006	1.226	0.007	0.89	0.05
2	1.080	0.006	1.182	0.006	0.75	0.06
3	1.074	0.006	1.148	0.006	0.71	0.10
4	1.077	0.006	1.115	0.006	0.92	0.30

Due to the way the mass ratio is used to fit tie-lines, the greatest uncertainty in equilibrium compositions is observed in the top PEG and bottom sodium citrate tribasic compositions. This is because the mass ratio influences the gradient of the tie-lines and equilibrium compositions are determined by finding the intersection of the binodial curve with the tie-lines. This is can be seen in Figure 4 - 15A which shows a phase diagram for a PEG 4,000-sodium citrate tribasic aqueous two-phase system when the uncertainty in equilibrium compositions due to the binodial curve is not taken into account.

Figure 4 - 15B shows the phase diagram obtained when using a targeted binodial curve screen whilst taking into account uncertainty due the mass ratio using Eq.(4.25). It can be seen clearly by comparing Figure 4 - 14B and Figure 4 - 15B that taking into account the additional effect of uncertainty due to the mass ratio results in a larger error bars for all equilibrium compositions. Figure 4 - 15C shows the same phase diagram when using the binodial curve obtained using the initial 4wt% screen. In this situation it is more

difficult to see the impact of mass ratio uncertainty as the uncertainty in the binodial curve is also relatively high.

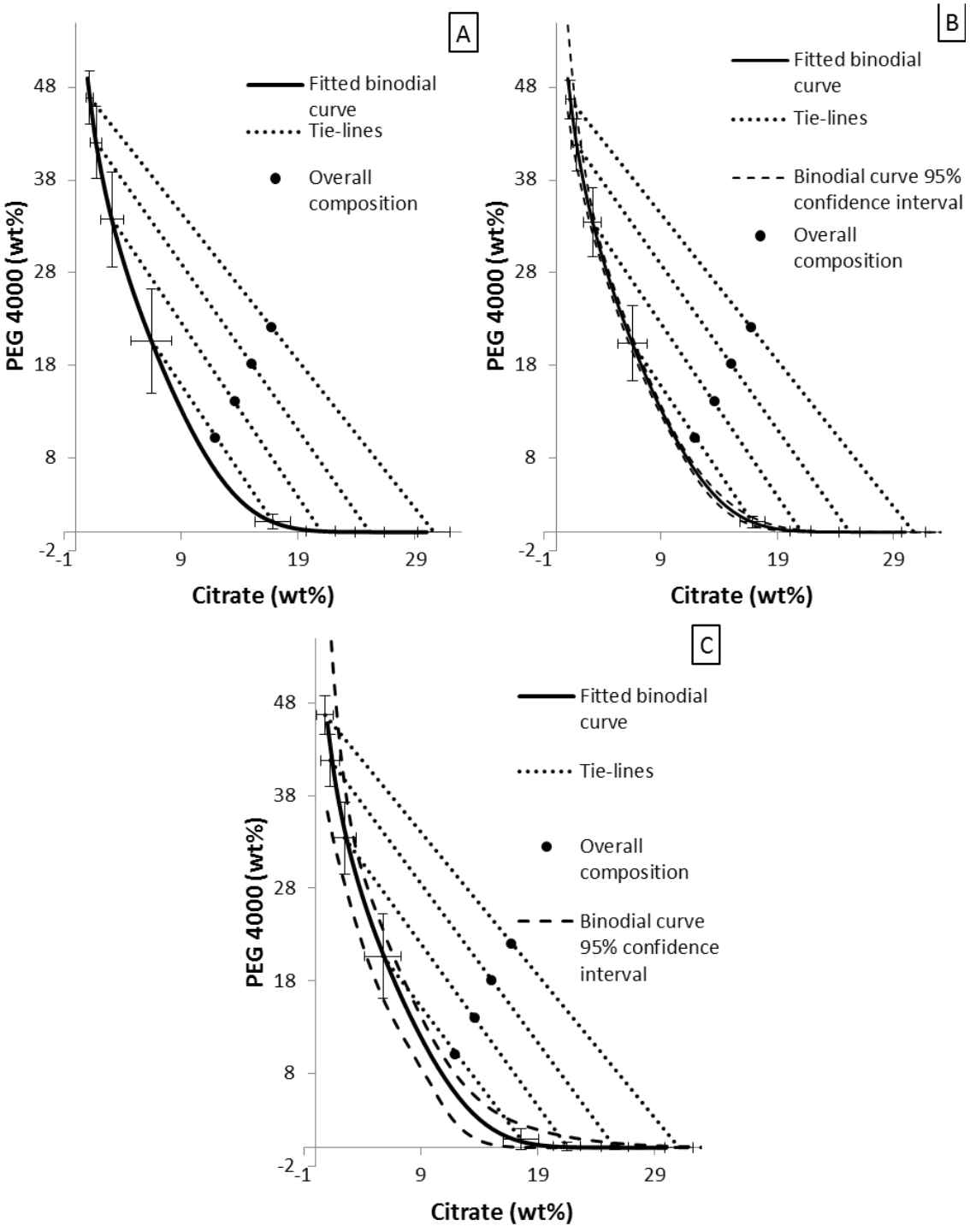


Figure 4 - 15: Major source of uncertainty is mass ratio. Phase diagrams with tie-lines taking into account uncertainty due to A) mass ratio using binodial curve from 1wt% screening resolution (error due to uncertainty in binodial curve is ignored) B) a 1wt% screening resolution for the binodial curve and uncertainty due to the mass ratio. C) a 4wt% screening resolution for the binodial curve and uncertainty due to the mass ratio.

4.3.3 Estimation of interaction parameters

Once the uncertainty in equilibrium compositions due to the binodial curve and mass ratio were calculated, estimation of the interaction parameters used in Eq.(3.21)-(3.23) was conducted taking into account both sources of uncertainty. Interaction parameters were estimated using both Protocol D outlined in Figure 4 - 11 and the Parameter Estimation entity within gPROMS. Estimation of interaction parameters was conducted using 3 different scenarios to see whether quantity of experimental information significantly impacted parameter estimates:

1. Using equilibrium composition data generated using the 4wt% initial screen for the binodial curve.
2. Using equilibrium composition data generated using the targeted screen for the binodial curve.
3. Using equilibrium composition data generated using the 4wt% initial screen and the targeted screen for the binodial curve.

Least squares parameter estimation vs. maximum likelihood parameter estimation

Table 4 - 4 shows the estimated interaction parameters obtained using the least squares (Protocol D) and the maximum likelihood gPROMS parameter estimation entity objective functions. The initial guesses of interaction parameters for maximum likelihood parameter estimation were obtained using Protocol D as the gPROMS parameter estimation entity required good initial guesses. The interaction parameters estimated using the least squares objective function are similar in value to those obtained with the maximum likelihood objective function (Table 4 - 4), however, the model prediction of equilibrium compositions for the top phase are different, this is can be seen in Figure 4 - 16. In addition, interactions parameters estimated using Protocol D do not have calculated standard deviation values as the optimisation method is a gradient free method. The model prediction using interaction parameters obtained using the least squares objective function fit the mean equilibrium compositions well. Whereas when using the maximum likelihood objective function, the model predictions are weighted to the experimental data of highest certainty which are equilibrium compositions at higher tie-line lengths. This is in line with the previous findings which state that the uncertainty in equilibrium compositions due to the mass ratio is lowest at higher tie-line lengths.

In addition to weighting the model fit to areas of highest experimental certainty, the Parameter Estimation entity within gPROMS calculates standard deviations for estimated parameters based on covariance; these are reported in Table 4 - 4. When using equilibrium compositions generated using the 4wt% initial screen of the binodial curve, the PEG-PEG and Citrate-Citrate interaction parameters have standard deviation values which are ~62% and ~40%, respectively, of the estimated parameter value. The standard deviation of the PEG-Citrate interaction parameter (9) is greater than the estimated parameter value (-6). When conducting parameter estimation using equilibrium compositions generated from the targeted binodial curve screen, the uncertainty in the PEG-PEG and PEG-Citrate interaction parameters is reduced slightly to ~50% and ~31%, respectively, of the estimated parameter value. The PEG-Citrate interaction parameter still has a standard deviation value (7) which is greater than the estimated parameter value (-5). The uncertainty in parameters can be reduced a little further by using equilibrium compositions from both the 4wt% screen and the targeted screen, however, PEG-Citrate interaction parameter standard deviation is still large. One possible reason why the uncertainty in the PEG-Citrate parameter is high is that the experimental tie-lines determined in this work may not be in the area of maximum information for estimation of this parameter. One possible way to overcome this would be to conduct further experiments where the model fit is particularly sensitive to the value of the PEG-Citrate interaction parameter.

The findings suggest that the experimental phase diagram generated in this work is not known to a sufficiently high accuracy in order to estimate interaction parameters with a low level of uncertainty. This is primarily due to the fact that the density measurements used in this work result in large uncertainties in the equilibrium compositions.

Potential ways to address this are to:

1) Increase the number of experimental tie-lines determined.
2) Reduce the uncertainty in density measurements using dedicated density meters.

These two solutions are discussed next.

Table 4 - 4: Estimated interaction parameters using a least squares objective function in MATLAB and maximum likelihood objective function using the Parameter Estimation Entity within gPROMS.

β_{ij}	MATLAB value	gPROMS parameter estimation					
		4wt% initial screen		Targeted screen		4wt% and targeted screen	
		Value	Std	Value	Std	Value	Std
PEG-PEG	-114	-110	68	-102	52	-106	42
PEG-Citrate	-10	-6	9	-5	7	-6	6
Citrate-Citrate	-3.4	-3.0	1.2	-2.9	0.9	-3.0	0.7

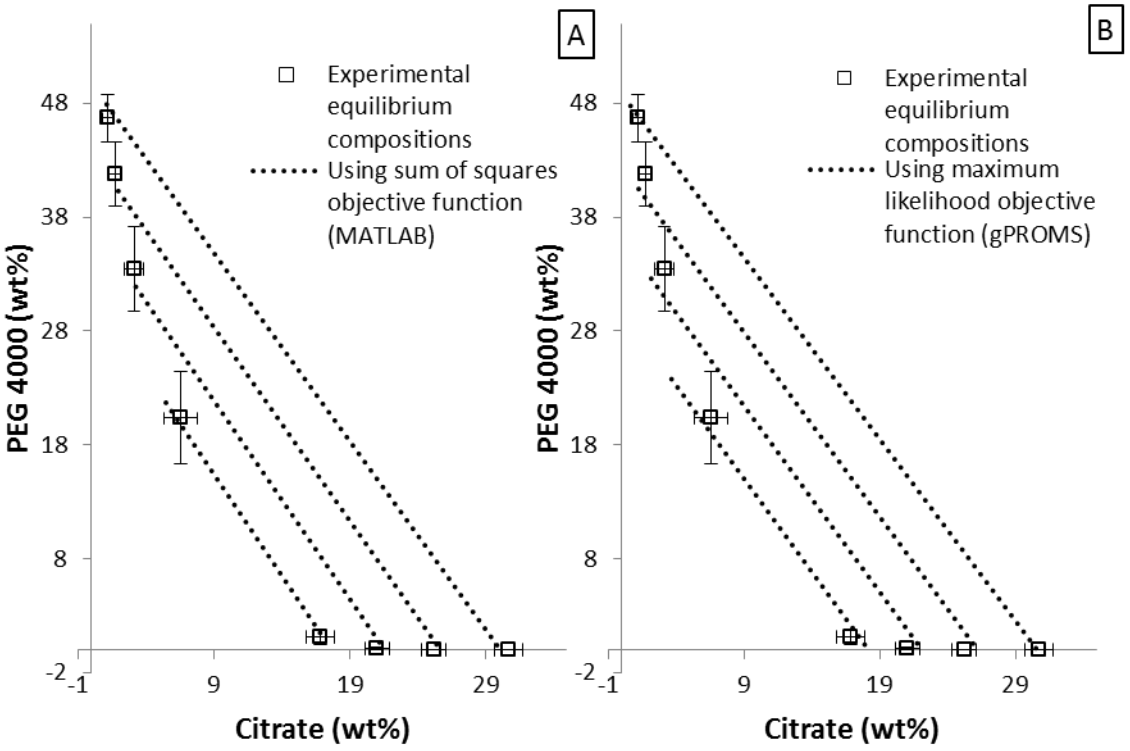


Figure 4 - 16: Model predictions for a PEG 4,000-sodium citrate tribasic aqueous two-phase system using parameters generated using A) sum of squares objective function and B) maximum likelihood objective function. Experimental equilibrium points are from tie-lines shown in Figure 4 - 15B.

Impact of number of tie-lines and lower density error

Additional tie-lines were interpolated to investigate the quantity of experimental data required to obtain interaction parameter estimates with a low level of uncertainty. The analysis was conducted using 4, 7, 13, 25 and 49 tie-lines. In addition, the impact of using density measurement values with a lower error (0.001g ml^{-1}) was also investigated. This data is reported in Table 4 - 5 and Table 4 - 6, respectively.

Generally speaking, increasing the number of tie-lines reduces the uncertainty of the estimated parameters. For example, when using a 4wt% screening resolution for the binodial curve the relative standard deviation percentage[7] (Rel Std%) of the PEG-PEG interaction parameter changes from 62% to 13% when increasing the number of tie-lines from 4 to 49. In reality, it would not be practically feasible to determine 49 tie-lines. Interestingly, the Rel Std% of the PEG-PEG interaction parameter when using the targeted screen is almost identical at 13, 25 and 49 tie-lines. The Rel Std% of the PEG-PEG interaction parameter is consistently lower when using data from both the 4wt% and targeted screen. The same types of trends are also observed for the Citrate-Citrate interaction parameter.

If the density could be measured more accurately for example by using an analytical density meter with an error of 0.001g ml^{-1} then the interaction parameters can be determined to a much lower uncertainty. This is because the standard deviation of density measurements in this work is ~0.006g ml^{-1} whereas the standard deviation of an analytical density meter would be ~10 times less at 0.00051g ml^{-1} assuming that the error of the analytical density meter is normally distributed. For example, the Rel Std% of PEG-PEG and Citrate-Citrate interaction parameters using 4 tie-lines and both the 4wt% and targeted binodial curve screen is 11% and 6%, respectively as can be seen in Table 4 - 6. This is equivalent to conducting 25 tie-line experiments with the density measurement techniques used in this work. What these findings suggest is that provided the binodial curve is known to a high accuracy, the best way to improve the estimates of PEG-PEG and Citrate-Citrate interaction parameters is by increasing the quality of the mass ratio values. One way to do this would be to use an analytical grade density meter, however, as these meters generally require sample sizes greater than 250µL therefore the size of two-phase systems made would also have to be increased.

[7] Relative standard deviation percentage is the parameter standard deviation value divided by the parameter value multiplied by 100.

Unfortunately the PEG-Citrate interaction parameter Rel Std% increases from 150% to 1201% when increasing tie-lines from 4 to 7 when using the 4wt% screening resolution data (Table 4 - 5). It is unclear why this is the case, however, after 7 tie-lines the Rel Std% does decrease. When considering the targeted screen equilibrium composition data, the PEG-Citrate Rel Std% initially decreases as the number of tie-lines is increased from 4 to 7, however after that there is an increase to around ~200%. Interestingly, it also seems to be difficult to obtain an accurate estimate of the PEG-Citrate interaction parameter even when using a hypothetical scenario where the error in density was assumed to be lower at 0.001g ml^{-1} (Table 4 - 6).

To investigate the challenge in determination of the PEG-Citrate interaction parameter, a sensitivity analysis was conducted where the value of a single interaction parameter was adjusted while the remaining two interaction parameters were kept constant. Figure 4 - 17 shows how the predicted phase equilibrium compositions change as a result of changes in the value of PEG-PEG, PEG-Citrate and Citrate-Citrate interaction parameters. When the PEG-PEG and Citrate-Citrate interaction parameters are more positive, the two-phase region decreases in size whereas when the PEG-Citrate interaction parameter is more positive the two-phase region increases in size. This relationship may explain why it is difficult to obtain a good estimate of the PEG-Citrate interaction parameter since changes in model fit due to PEG-PEG and Citrate-Citrate interaction parameters can be compensated for by appropriate adjustments of the PEG-Citrate interaction parameter by the optimisation algorithms. This is not a surprising finding since the parameter estimation report generated when using gPROMS also suggests that the interaction parameters are highly correlated as shown in the correlation matrix presented in Table 4 - 7.

Figure 4 - 17 shows that the shifts in the two-phase region due to changes in interaction parameters are not linear. For example, at PEG concentrations of ~40wt%, the two-phase region boundary does not shift as much as when the PEG concentration is ~10wt%. This suggests that one way to obtain better estimates of interactions parameters would be to obtain experimental tie-line data where the change in model predictions due to small changes in parameters is the greatest. In this case, it would specifically involve looking at systems of a low tie-line length. These additional experiments could be designed using a formalised approach using model based design of experiment as described by Dechambre et al. (2014) for liquid-liquid extraction. However, in this work the experimental protocols used have enabled a good

determination of the binodial curve and hence the two-phase region. Therefore there is potential to develop a new method to estimate interaction parameters which takes into account both knowledge of the tie-lines as well as the complete binodial curve location. One possible way this could be achieved is if an estimation protocol determined a model predicted binodial curve as well as tie-line equilibrium compositions for a given set of interaction parameters.

Table 4 - 5: Impact of number of tie-lines on estimated interaction parameter (β_{ij}) values and their corresponding relative standard deviations.

PEG-PEG						
Number of tie-lines		**4**	**7**	**13**	**25**	**49**
4wt% initial screen	Value	-110	-71	-65	-62	-60
	Rel Std %	62	32	24	18	13
Targeted screen	Value	-102	-93	-70	-69	-65
	Rel Std %	51	41	25	19	13
4wt% and targeted screen	Value	-106	-67	-63	-61	-60
	Rel Std %	40	21	17	12	9
PEG-Citrate						
Number of tie-lines		**4**	**7**	**13**	**25**	**49**
4wt% initial screen	Value	-5.9	0.2	0.8	1.0	1.1
	Rel Std %	150	1201	257	139	89
Targeted screen	Value	-5	-4	-1	-1	0
	Rel Std %	129	116	220	167	244
4wt% and targeted screen	Value	-5.7	0.2	0.5	0.6	0.7
	Rel Std %	97	1085	293	171	107
Citrate-Citrate						
Number of tie-lines		**4**	**7**	**13**	**25**	**49**
4wt% initial screen	Value	-3.0	-2.3	-2.2	-2.2	-2.2
	Rel Std %	39	17	12	9	6
Targeted screen	Value	-2.9	-2.8	-2.4	-2.4	-2.3
	Rel Std %	32	25	13	10	7
4wt% and targeted screen	Value	-3.0	-2.3	-2.2	-2.2	-2.2
	Rel Std %	25	11	8	6	4

Table 4 - 6: Impact of number of tie-lines on estimated interaction parameter (β_{ij}) values and their corresponding relative standard deviations. A hypothetical density error of 0.001g ml^{-1} was used for these calculations.

PEG-PEG						
Number of tie-lines		**4**	**7**	**13**	**25**	**49**
4wt% initial screen	Value	-109	-79	-76	-75	-74
	Rel Std %	26	15	11	9	6
Targeted screen	Value	-65	-67	-57	-59	-57
	Rel Std %	13	14	10	8	5
4wt% and targeted screen	Value	-73	-58	-59	-60	-60
	Rel Std %	11	9	7	6	4
PEG-Citrate						
Number of tie-lines		**4**	**7**	**13**	**25**	**49**
4wt% initial screen	Value	-6.0	-1.0	-0.9	-0.8	-0.8
	Rel Std %	61	150	137	107	78
Targeted screen	Value	-1.5	-2.0	-0.4	-0.7	-0.4
	Rel Std %	97	76	277	119	152
4wt% and targeted screen	Value	-2.3	0.4	0.1	-0.1	-0.2
	Rel Std %	52	262	544	672	273
Citrate-Citrate						
Number of tie-lines		**4**	**7**	**13**	**25**	**49**
4wt% initial screen	Value	-3.0	-2.5	-2.4	-2.4	-2.4
	Rel Std %	16	8	6	5	3
Targeted screen	Value	-2.3	-2.4	-2.2	-2.2	-2.2
	Rel Std %	7	8	5	4	3
4wt% and targeted screen	Value	-2.4	-2.2	-2.2	-2.2	-2.2
	Rel Std %	6	5	4	3	2

Table 4 - 7: Parameter correlation matrix for interaction parameters estimated using a hypothetical density error of 0.001g ml^{-1} and both 4wt% and targeted binodial curve screens.

	Correlation matrix		
	PEG-PEG	**PEG-Citrate**	**Citrate-Citrate**
PEG-PEG	1		
PEG-Citrate	0.992	1	
Citrate-Citrate	0.993	0.981	1

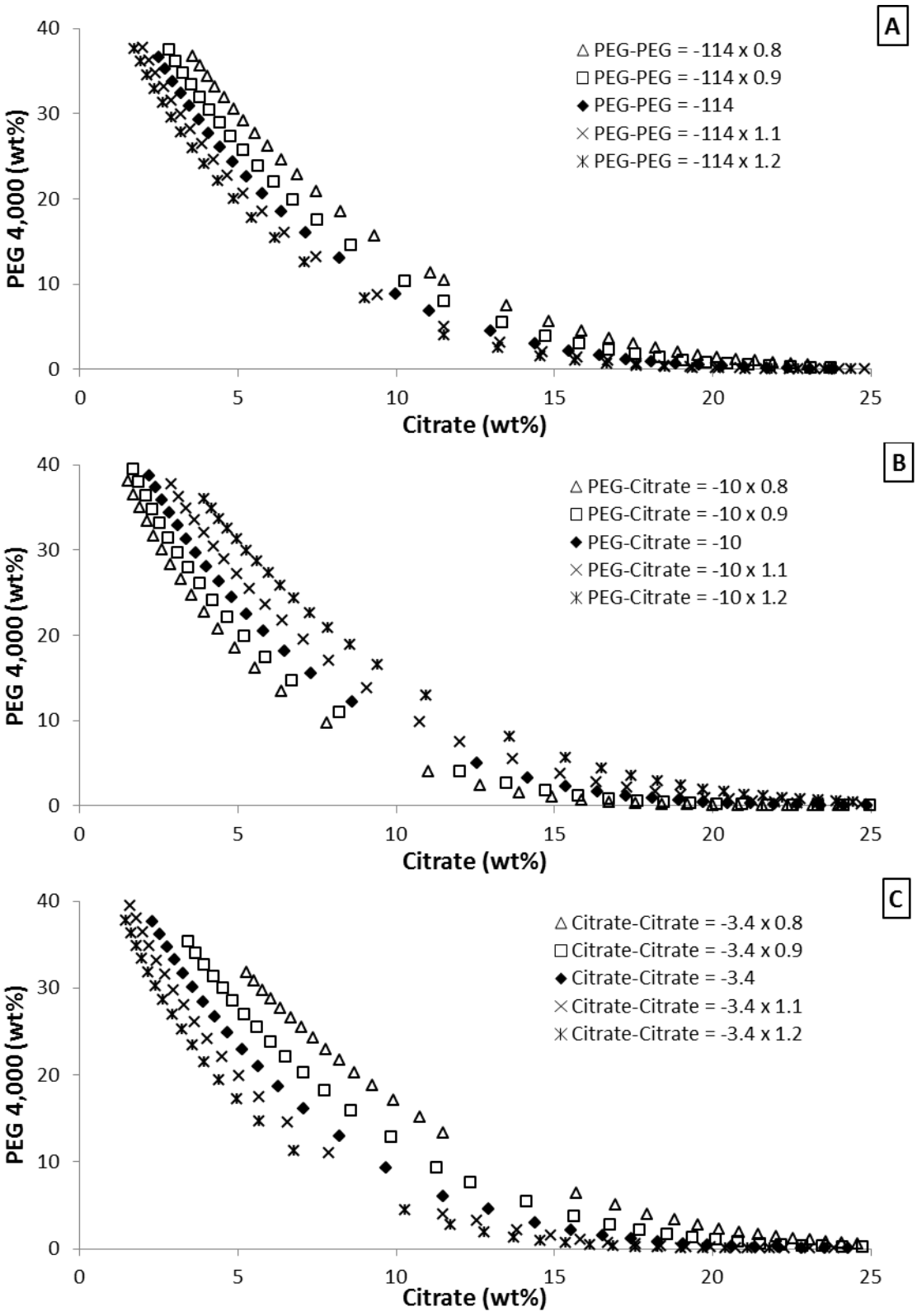


Figure 4 - 17: Sensitivity of phase equilibria compositions to value of $\beta_{PEG\text{-}PEG}$ (A), $\beta_{PEG\text{-}Citrate}$ (B) and $\beta_{Citrate\text{-}Citrate}$ (C) interaction parameters. Base values of $\beta_{PEG\text{-}PEG}$, $\beta_{PEG\text{-}Citrate}$ and $\beta_{Citrate\text{-}Citrate}$ are -114, -10 and -3.4, respectively.

4.4 Conclusion

The main aim of this work was to understand how uncertainty in phase diagrams generated using high-throughput screening of aqueous two-phase systems impacts on the estimation of interaction parameters used in thermodynamic models which are used to calculate phase equilibria compositions in process models. This has not been considered in the literature so far, but is particularly important considering the need to understand the impacts of process uncertainty in the development of robust bioprocesses.

Two main sources of uncertainty in equilibrium compositions were identified and quantified: uncertainty due to the binodial curve and uncertainty due to the mass ratio used to fit tie-lines. A two-stage screening approach, demonstrated using a PEG 4,000-sodium citrate tribasic aqueous two-phase system, was developed to minimise the uncertainty in equilibrium compositions due to the binodial curve. The developed two-stage screening approach involves conducting an initial screen of a low resolution which is used to identify an approximate location of the binodial curve. The low resolution screen is followed by a high resolution screen which is targeted to the approximate location of the binodial curve therefore resulting in a binodial curve location which is known to a high accuracy. The combined approach uses less overall experimental resources (142 systems) than if a high resolution screen was conducted across all compositions (350 systems). In addition, the automated two-stage approach took less than 2 hours to prepare all 142 systems. Conventional dilution methods would take considerably longer and are more susceptible to human error as multiple interlinked mass measurements need to be taken. Tie-lines were fitted using the inverse lever arm rule and experimentally determined mass ratios for various aqueous two-phase systems. Uncertainty in the mass ratio was determined by propagating experimental errors in density measurements. It was found that the major source of uncertainty in determining equilibrium compositions was due to the mass ratio.

Next estimation of thermodynamic interaction parameters was also conducted using a two-stage approach. Firstly, interaction parameters were estimated by minimising a least squares objective function using a simulated annealing optimisation algorithm where phase equilibria calculations were conducted using the Rachford-Rice algorithm. These estimated parameters were then used as initial guesses within the gPROMS Parameter Estimation entity which estimated parameters by minimising a maximum likelihood objective function. The benefit of using a maximum likelihood objective

function over a least squares objective function is the ability to take into account experimental uncertainty during parameter estimation. A two-stage approach was beneficial in this work as the Parameter Estimation entity within gPROMS requires very good initial guesses to run the optimisation and equation solving algorithms robustly. Analysis of the estimated parameters suggests that the determined experimental equilibrium compositions in this work may not be of a sufficient accuracy to determine interaction parameters to a low level of uncertainty where the interaction parameter standard deviation is significantly less than the parameter value.

For the case study, it was found that uncertainty in the estimated PEG-PEG and Citate-Citrate interaction parameters could be significantly reduced to a relative standard deviation of less than ~11% by increasing the amount of experimental data used in parameter estimation or by alternatively increasing the accuracy of density measurements. The findings suggest that increasing accuracy of density measurements results in a greater positive impact on estimated parameters than increasing the number of tie-lines. Therefore, future efforts to improve experimental techniques for parameter estimation purposes should focus on improving accuracy of mass ratio measurements or density measurements.

Chapter 5: High-throughput screening and simulation of α-amylase and myoglobin separation

5.1 Introduction

Chapter 3 looked at the use of dynamic modelling techniques to simulate aqueous two-phase extraction, while Chapter 4 looked at the determination of phase diagrams and estimation of interaction parameters for aqueous two-phase systems. The work in this chapter aims to combine these two methods. High-throughput screening experiments were used to generate phase diagrams as well as partitioning data. The experimental data was then be used to build a suitable process model so that an aqueous two-phase extraction process can be developed.

The development of aqueous two-phase systems is challenging as discussed earlier. This is in part due to the many factors which can influence protein partitioning such as polymer molecular weight, polymer type, pH, charge etc. (Asenjo and Andrews, 2011). To better tackle this large experimental space, researchers have resorted to the use of design of experiment and high-throughput screening techniques for process development (Bensch et al., 2007, Azevedo et al., 2007). Such techniques are useful in initial process development; however, two-phase extraction processes are operated in a multitude of setups such as batch, continuous counter-current, continuous counter-current with recycle etc. An experimental exploration of all such setups is extremely resource intensive due to the number of operating factors which must be adjusted, such as number of stages, amount of holdup, flow rates etc. The use of process models to perform simulations of aqueous two-phase extraction can help reduce the experimental effort. This, however, requires appropriate experimental information to determine the model parameters required to ensure the model is an accurate description of the specific system.

To date, the use of a combined experimental and modelling approach for aqueous two-phase system development has been limited to the works of Mistry et al. (1996) and Mündges et al. (2015). The work in this chapter improves on their approaches by using a combination of high-throughput experimentation and process modelling for the development of an extraction process to purify enzyme α-amylase from myoglobin. The use of high-throughput experimentation is advantageous as it allows for greater information to be obtained from fewer experimental resources while the use of process

models allows for exploration of potential modes of operation without the need to perform additional experiments. Specifically, high-throughput experiments are used to determine phase diagrams and partition coefficients. The phase diagrams and partition coefficients are then used in a mechanistic process model to simulate aqueous two-phase extraction operation at a larger scale in continuous counter-current mode. The process model allows the exploration of additional conditions which are difficult to investigate experimentally such as number of extraction stages, phase ratio and residence time; all of which are critical in the choice of equipment used.

In this work, the separation of a two-component mixture of α-amylase and myoglobin is considered in a PEG-phosphate aqueous two-phase system. α-amylase was chosen as a model protein and product as it has similar characteristics to the therapeutic protein acid-alpha glucosidase (GAA). Myoglobin was chosen as a contaminant protein as it is easy to detect experimentally due to its strong spectral absorbance at 409nm.

5.2 Results and discussion

5.2.1 α-amylase and myoglobin initial screening

Prior to conducting experiments in high-throughput using automated liquid handling, a 3 factor screening study was conducted to determine which factors influence α-amylase and myoglobin partitioning the most. This approach was chosen as it reduces the number of liquid-class calibrations required when using the automated liquid handling platform. An overall composition of 15wt% PEG and phosphate was used for each system formed. The factors considered were PEG molecular weight, pH of phosphate stock solution and overall NaCl wt%. The low and high limits of these factors are shown in Table 5 - 1. The concentration of α-amylase and myoglobin in the top and bottom phases is presented in Table 5 - 2 for all screened systems. The molecular weight of PEG was screened between 2000 and 4000 because for high molecular weight species such as proteins these are favourable conditions to promote partitioning into the top phase (Benavides and Rito-Palomares, 2008). A pH range of 7.0 to 7.5 was chosen as these two pHs are either side of the reported pI of myglobin (7.36). NaCl composition was chosen based off the initial studies reported by Mistry et al. (1996) for partitioning of partitioning of partitioning of α-amylase in PEG-phosphate aqueous two-phase systems.

Table 5 - 1: Factors considered for initial α-amylase and myoglobin screen. Overall composition of PEG and phosphate is 15wt%.

Factors	Low	Medium	High
PEG molecular weight	2,000	N/A	4,000
pH of phosphate stock	7.0	N/A	7.5
NaCl wt%	0	3	6

PEG molecular weight

The initial screening results in Figure 5 - 1A and Figure 5 - 1B show that the molecular weight of PEG used has a significant impact on the concentration of α-amylase and myoglobin in the top and bottom phase, respectively. The difference in concentration is most likely due to increasing amounts of protein precipitation at the higher PEG molecular weight. Table 5 - 2 shows the total recovery[8]of α-amylase and myoglobin in in the PEG 4,000 system is ~50-60% whilst in the PEG 2,000 system recoveries are close to 100%. In some cases the calculated recovery is greater than 100% which is not possible; a possible reason for this may be due to assay error resulting in an over

[8] Recovery is calculated as the total amount of protein in both phases compared to what was put in.

calculation of α-amylase concentration. Figure 5 - 2 shows the interface in the PEG-phosphate aqueous two-phase systems using a molecular weight of 2,000 and 4,000. Precipitation is observed in both systems, however, the amount of precipitation visually observed in the PEG 4,000 system is more as can be seen from the greater amounts of precipitate on the walls of the tube. The formation of precipitation at the interface was linked to the solubility of proteins in the aqueous two-phase systems (Asenjo and Andrews, 2011). A study by Atha and Ingham (1981) found that higher molecular weight PEGs reduce the solubility of proteins. It is interesting that there is a loss of myoglobin as it primarily partitions into the bottom phase, which has very little PEG, which suggests that the precipitation is occurring during mixing where myoglobin is exposed to both phases. A PEG molecular weight of 2,000 was therefore chosen for high-throughput screening experiments.

pH of phosphate stock

Systems were formed using a pH 7.0 and pH 7.5 phosphate stock solution. Figure 5 - 1C and Figure 5 - 1D show that there was not any significant effect of pH on top phase α-amylase and bottom phase myoglobin concentration. The predicted pI of α-amylase (AMY_BACSU) using the ExPASy ProtParam tool (Gasteiger et al., 2003) is 5.59, therefore the fact that no major difference is observed makes sense as protein charge strength is likely to not have changed considerably. Using the same tool, a myoglobin (MYG_HORSE) pI of 7.36 was calculated which suggests that changing from a pH of 7.0 to 7.5 should change myoglobin protein charge sign. A change in charge sign could potentially alter partitioning behaviour, however, this was not observed. A reason for this could be that myoglobin is still relatively neutrally charged at the pHs investigated. A phosphate stock solution pH of 7.0 was arbitrarily chosen for high-throughput screening as the investigated pHs did not seem to significantly impact the concentrations.

NaCl wt%

Figure 5 - 1E and Figure 5 - 1F show the impact of NaCl wt% on top phase α-amylase and bottom phase myoglobin concentration. Adding more NaCl to the aqueous two-phase system had no observable impact on concentration of myoglobin, however, increasing overall NaCl amount from 0wt% to 6wt% saw a slight increase in α-amylase concentration (Figure 5 - 1E). The partitioning of α-amylase in the aqueous two-phase systems investigated in this work does not seem to be impacted as much by NaCl wt% as reported in literature where the partition coefficient changes from less than 1 to more

than 1 when NaCl wt% is increased from 0wt% to 10wt% (Mistry et al., 1996) . It is unclear why this is the case, however, one reason for this could be due to differences in the type of α-amylase used which may result in different protein characteristics and therefore partitioning. A 6wt% overall composition of NaCl was used for high-throughput screening as this resulted in the highest α-amylase concentration in this work.

Table 5 - 2: Concentration of α-amylase and myoglobin in top and bottom in different aqueous two-phase systems. Abbreviations: Amy is α-amylase and Myo is myoglobin.

Factors			Concentration (mg/mL)				Recovery (%)	
			Top		Bottom			
PEG molecular weight	pH	NaCl wt%	Amy	Myo	Amy	Myo	Amy	Myo
4,000	7.5	6.0 (~1.2M)	0.93	0.03	0.09	0.85	69	62
4,000	7.0	3.0 (~0.6M)	0.77	0.02	0.06	0.80	60	55
4,000	7.5	0.0	0.64	0.01	0.09	0.84	57	49
4,000	7.0	6.0 (~1.2M)	0.89	0.02	0.09	0.87	67	59
4,000	7.0	0.0	0.65	0.02	0.04	0.94	53	50
4,000	7.5	3.0 (~0.6M)	0.79	0.02	-0.06	0.96	54	60
2,000	7.0	0.0	1.34	0.01	0.14	1.19	119	98
2,000	7.5	6.0 (~1.2M)	1.46	0.03	0.07	1.24	112	103
2,000	7.5	0.0	1.29	0.02	0.06	1.21	113	100
2,000	7.5	3.0 (~0.6M)	1.31	0.02	0.05	1.23	107	102
2,000	7.0	3.0 (~0.6M)	1.29	0.02	0.04	1.12	98	93
2,000	7.0	6.0 (~1.2M)	1.47	0.03	0.02	1.22	108	101

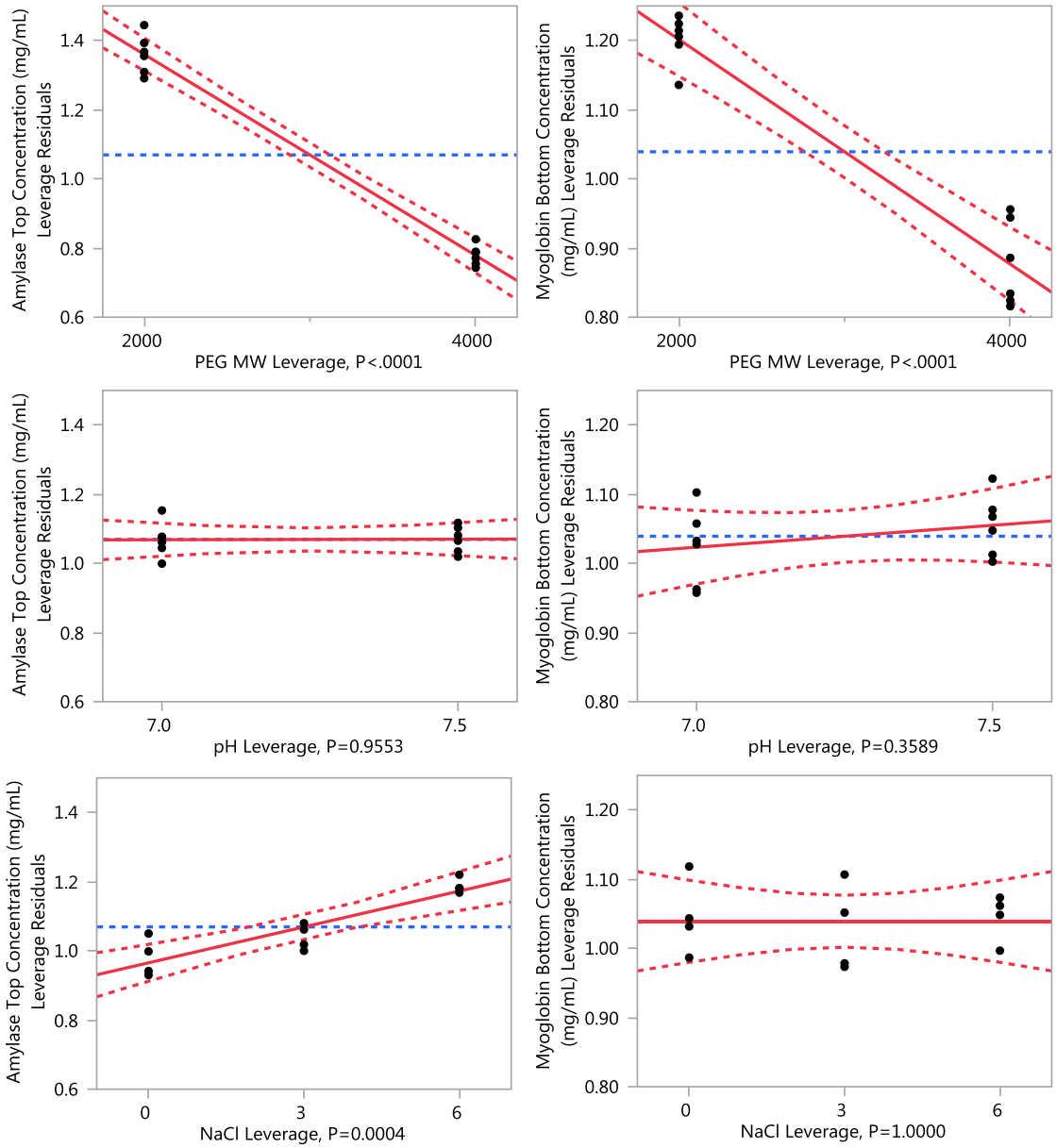


Figure 5 - 1: Concentration of α-amylase and myoglobin in the top and bottom phases respectively fitted to an empirical least squares model in JMP with PEG molecular weight, pH and NaCl as factors. Dashed blue line is mean concentration value across all conditions. Solid red line is line of fit taking into account specific interaction. Dashed red lines are 95% confidence intervals. A, C and E) α-amylase top concentration. B, D and F) myoglobin bottom concentration.

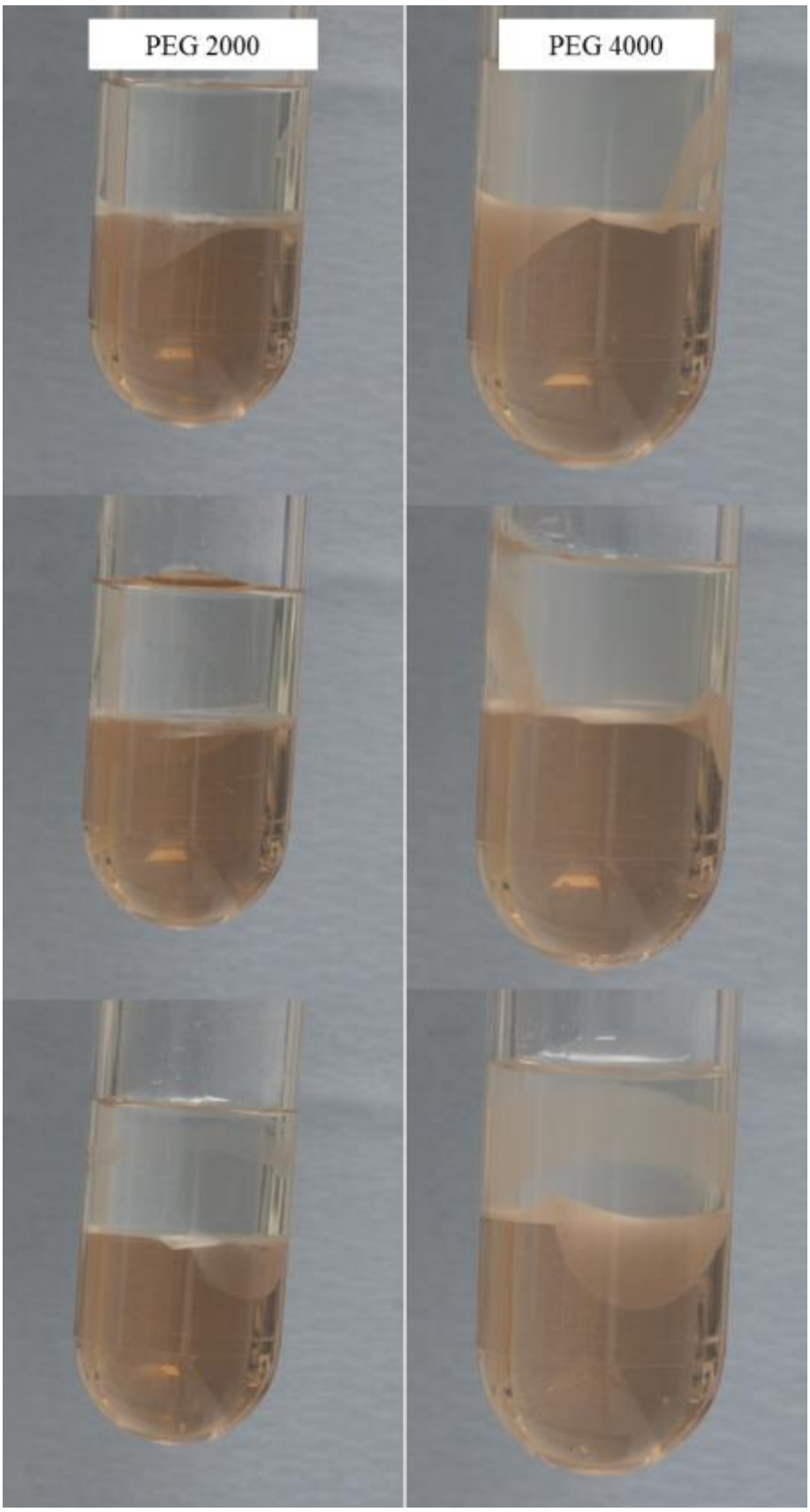


Figure 5 - 2: Precipitation at the interface in PEG 2,000 (left) and PEG 4,000 (right) aqueous two-phase systems with 0wt% NaCl. Precipitation observed at the interface in both systems. In PEG 4,000 systems, the interface is more saturated and excess amounts of precipitate is deposited onto walls during centrifugation.

5.2.2 PEG 2,000-phosphate-6wt% NaCl phase diagram

Based on the findings of the initial screening study a PEG 2,000-phosphate (pH 7) aqueous two-phase system with an overall NaCl concentration 6wt% was used in high-throughput screening. Since the type of aqueous two-phase system was now determined, the next step was to determine the phase diagram for estimation of interaction parameters used in the process model. This involved determination of the binodial curve and appropriate tie-lines using the methods established in Chapter 4 (see Figure 4 - 1 for overall methodology for phase diagram determination).

Binodial curve screening

The methods developed in Chapter 4 were used to successfully screen for the location of the PEG 2,000-phosphate (pH 7)-6wt% NaCl aqueous two-phase system binodial curve. Figure 5 - 3A shows the binodial obtained using a 4wt% screening resolution in PEG 2,000 and phosphate composition. Using the approximate location of the binodial curve a second screen was conducted, as shown in Figure 5 - 3B, to reduce the size of the confidence intervals. The majority of transition points are found within the left side of the targeted screening area i.e. the first two systems in the PEG composition row. This highlights the importance of taking into account the size of the initial screen as if the targeted screen examined a smaller area then there is a possibility of not finding any transition points. The parameters for the binodial curve equation are shown in Table 5 - 3.

Table 5 - 3: Parameters for fitted binodial curve equation (Eq.(4.4)) using 4wt% initial screen and targeted screen using a 1wt% and 2wt% screening resolution in the phosphate and PEG 2,000 axes, respectively. Binodial curve is for a PEG 2,000-phosphate (pH 7)-6wt% NaCl aqueous two-phase system.

	Binodial curve screening resolution			
	4wt% screen		Targeted screen	
Parameter	Value	Standard deviation	Value	Standard deviation
A	53.7	8.2	51.3	4.6
B	-0.423	0.082	-0.442	0.052
C	0.0004	1.43×10^{-4}	0.0006	9.11×10^{-5}

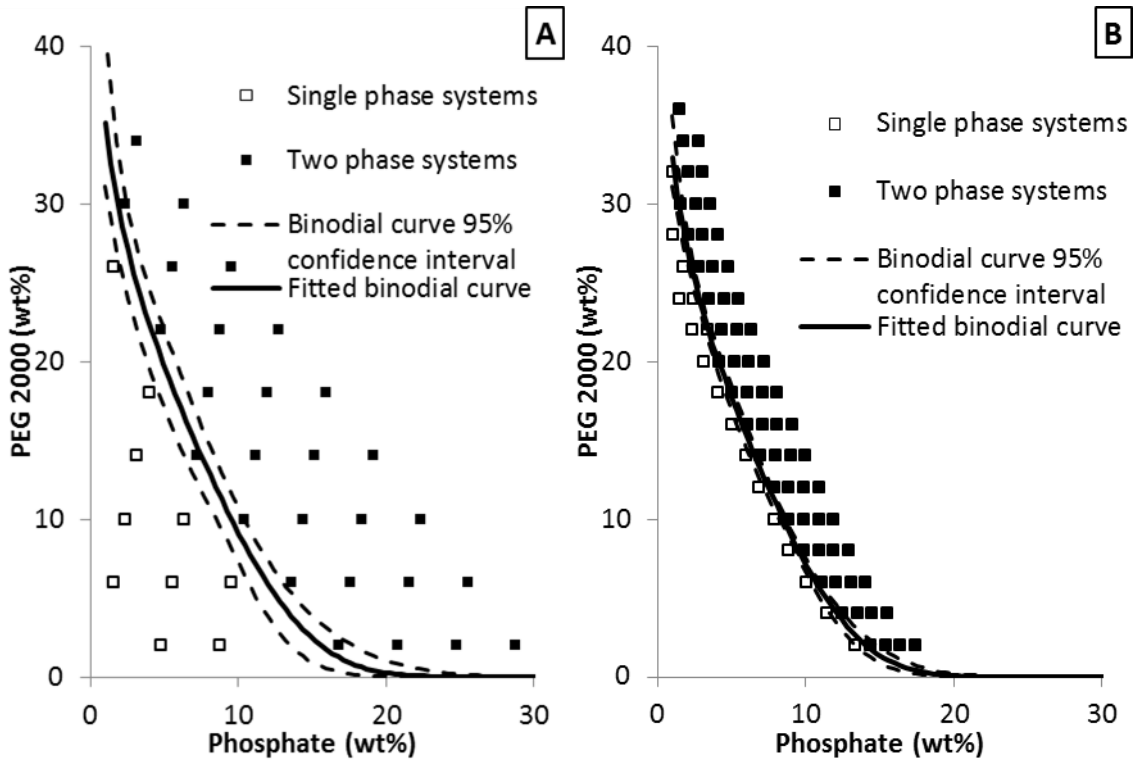


Figure 5 - 3: Binodial curve for a PEG 2,000-phosphate (pH 7)-6wt% NaCl aqueous two-phase system. A) Binodial curve and confidence intervals obtained using 4wt% screening resolution for PEG 2,000 and phosphate. B) Binodial curve and confidence intervals obtained using 1wt% screening resolution in phosphate and 2wt% screening resolution in PEG 2,000.

5.2.3 Estimation of interaction parameters

Using the methods of Chapter 4, five tie-lines were determined for the PEG 2,000-phosphate (pH 7)-6wt% NaCl aqueous two-phase system phase diagram. Ten sets of equilibrium compositions were obtained by fitting the five tie-lines to the binodial curves generated using both the 4wt% screening resolution and the targeted screen. The resultant equilibrium compositions were then used to estimate interaction parameters for the osmotic virial expansion equations (Eq.(3.21)-(3.23)) as outlined in Figure 4 - 1. Table 5 - 4 shows the estimated interaction parameters along with the corresponding parameter standard deviations. In addition high and low values for parameters are reported which are based on confidence intervals.

The standard deviation of the interaction parameter values is almost equal to the parameter values. This is not surprising considering the findings of Chapter 4 which highlighted the impact of density error on phase diagram uncertainty. To quantify the impact of this uncertainty on model prediction, low and high parameter combinations were generated from confidence ellipsoids which are shown in Figure 5 - 4. Confidence ellipsoids relate how the value of an estimated parameter relates to the value of another estimated parameter. They are useful in understanding how correlated estimated parameters are. Highly correlated parameters are often difficult to estimate independently of each other and result in confidence ellipsoids which are narrow and long. In this work confidence ellipsoids were used because the interaction parameters are highly correlated. The problem of parameter correlation and the resultant impact on parameter estimation is well documented in the literature, especially for the Arrhenius equation (Schwaab and Pinto, 2007). High and low combinations of interaction parameters were obtained by taking values at the extremes of the confidence ellipsoids.

Figure 5 - 5 shows model predictions using the optimal, low and high interaction parameter combinations with the corresponding experimental equilibrium point compositions. Although the uncertainty in the interaction parameters is high, the extreme values of each parameter combination still result in similar model prediction values, which are also within the calculated experimental error. This means that the estimated interaction parameters can still be used for preliminary process development simulations without the need to conduct further experiments to reduce the uncertainty in the parameters. The uncertainty in the parameters is likely due to the high-correlation between interaction parameters. Similar findings have also been reported by Kang and Sandler (1987) for estimated parameters in a UNIQUAC model. However, for the

osmotic virial equations used in this work, previous authors have failed to highlight this issue (Zafarani-Moattar and Sadeghi, 2001, Zafarani-Moattar and Gasemi, 2002, Zafarani-Moattar et al., 2004). To address this issue, further work would need to be conducted which aims at breaking the correlation between interaction parameters. If the correlation between interaction parameters cannot be broken, then the thermodynamic equations used would have to be re-evaluated to see if they are fit for purpose. If the aim of the model is to capture and describe the system well and there is not such an importance on uncertainty in parameter values then further work aimed at reducing parameter uncertainty may not be so beneficial. However, if the model is to be used in critical processes such as during manufacturing, then all aspects of the model need to be fully understood, which includes uncertainty in model parameter values.

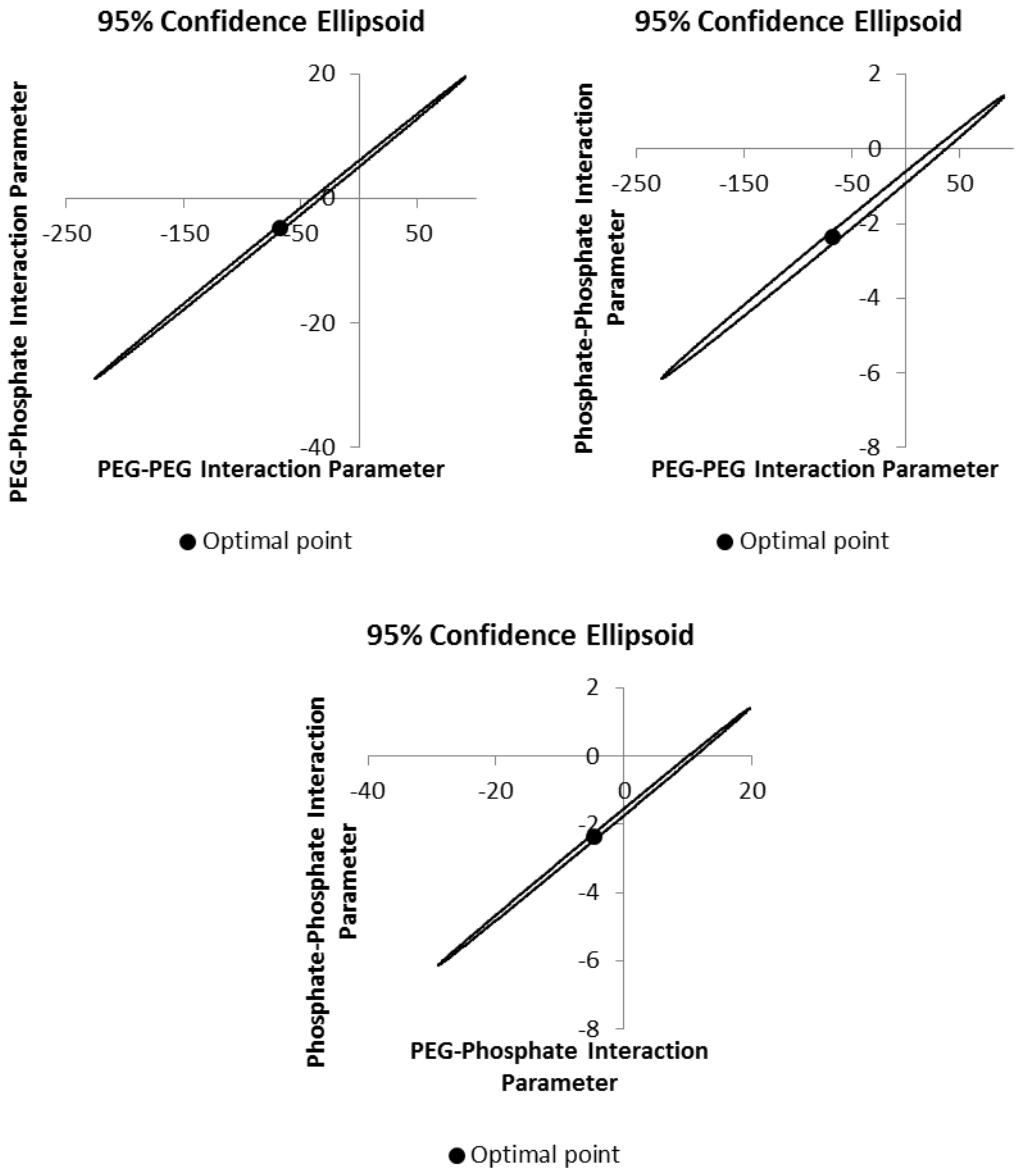


Figure 5 - 4: Confidence ellipsoids for estimated interaction parameters for PEG 2,000-phosphate (pH 7)-6wt% NaCl ATPS phase diagram.

Table 5 - 4: Estimated interaction parameters using gPROMS for PEG 2,000-phosphate (pH 7)-6wt% NaCl aqueous two-phase system. Highest and lowest parameter values obtained from parameter confidence ellipsoids.

βij	gPROMS parameter estimation			
	Optimal		Low	High
	Value	STD	Value	Value
PEG-PEG	-67	42	-225.15	90.59
PEG-Phosphate	-4.7	6	-28.96	19.57
Phosphate-Phosphate	-2.4	1.0	-6.14	1.41

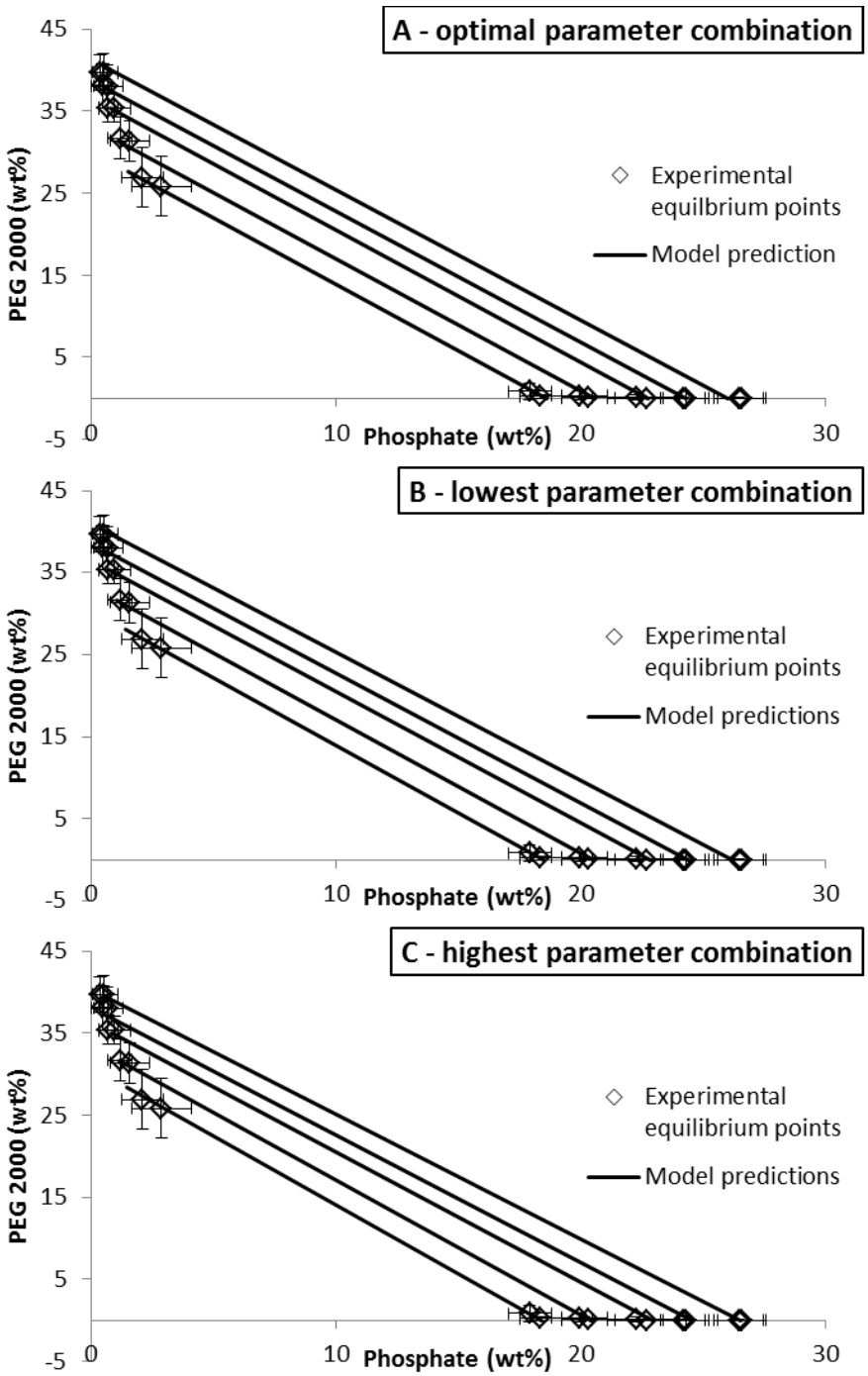


Figure 5 - 5: Model prediction and experimental phase equilibria compositions. Error bars in experimental equilibria points corresponds to +/- 2 standard deviations Tie-lines generated using 4wt% and 1wt% binodial curve screen used for interaction parameter estimation. A) Model prediction using optimal estimated parameter set. B) Model prediction using lowest possible combination of parameters. C) Model prediction using highest possible combination of parameters.

5.2.4 High-throughput screening of α-amylase and myoglobin separation

After establishing the phase diagram for the aqueous two-phase system of choice (PEG 2,000-phosphate (pH 7)-6wt% NaCl), experiments were conducted using high-throughput techniques to determine partitioning of α-amylase and myoglobin at different tie-line length and phase ratios. The partition coefficients were then used in process models to conduct simulations of aqueous two-phase extraction. The different tie-line lengths correspond to phases higher in polymer and salt concentration. The phase ratio, as described previously, is the mass of top phase divided by mass of bottom phase.

Mixing time

The high-throughput experiments involved scaling down aqueous two-phase systems into 2.0mL Eppendorf tubes using a total system size of 1 gram. Mixing on such small scales can often be difficult, especially when handling viscous polymer solutions. Therefore, to gauge whether mixing was adequate and equilibrium had been reached in the 2.0mL Eppendorf tubes, a simple mixing time study was conducted prior to conducting high-throughput screens. Figure 5 - 6 shows the concentration of α-amylase and myoglobin in the top and bottom phase when using different mixing times. It can be seen that the concentrations of both α-amylase and myoglobin in the top and bottom phase are reasonably constant at all the times tested which suggest that equilibrium is reached quickly in the 2.0ml Eppendorf tubes used. There is a slight downward trend in the top-phase myoglobin concentration (~0.2mg/mL) with mixing time, however, the bottom phase myoglobin concentration (~2mg/mL) at the 4 and 8 minutes is relatively constant which suggests equilibrium is likely reached. Note that the concentrations of α-amylase and myoglobin used in this study are different to those reported in due to different protein stock solution concentrations. However, to be certain of this conclusion, a more comprehensive mixing time study with many more time points would have to be conducted.

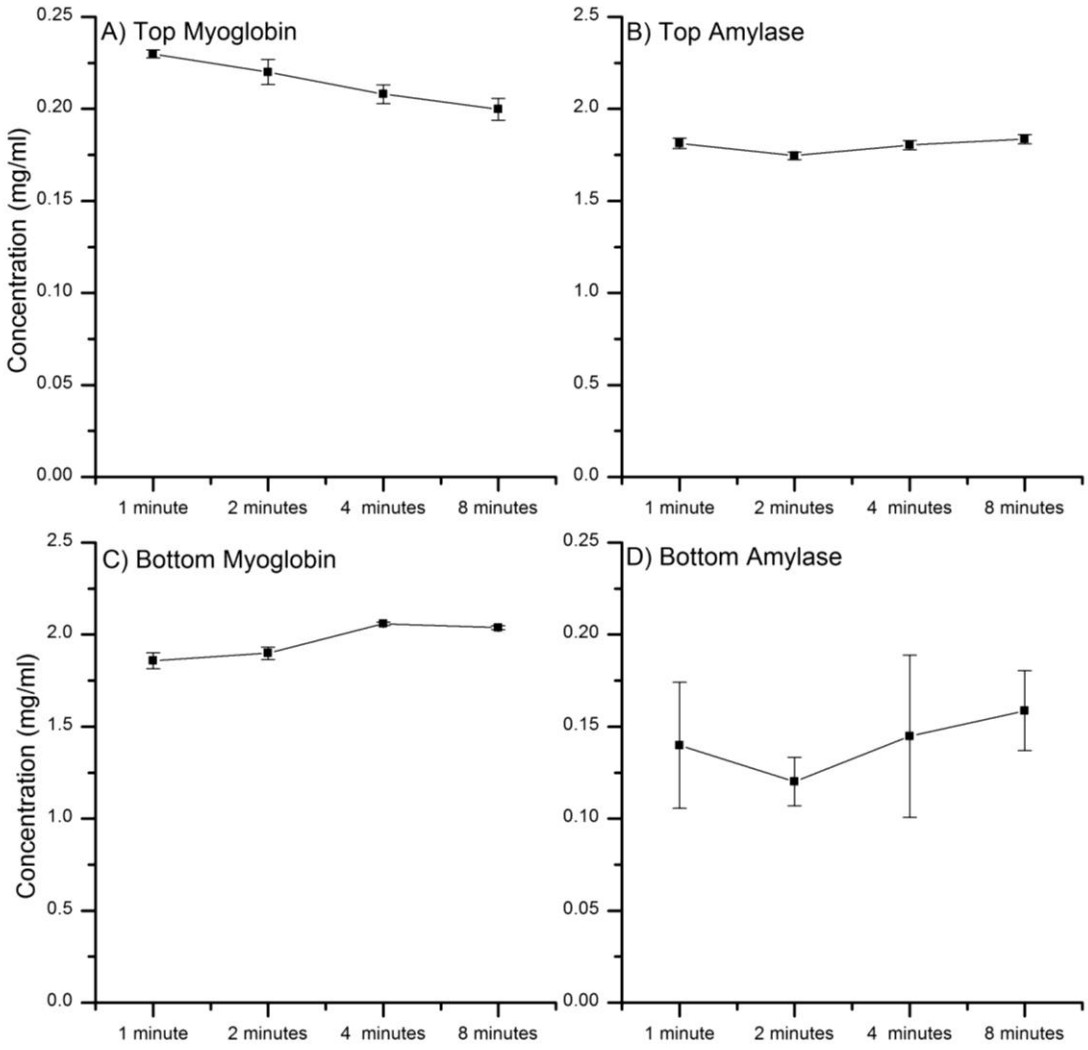


Figure 5 - 6: Impact of mixing time on protein partitioning in 2.0mL Eppendorf reaction tubes. Markers are mean value of triplicate results with error bars representing +/- 1 standard deviation.

Screening using automated liquid handling

Automated liquid handling was used to screen 21 aqueous two-phase systems. A total of 42 systems were produced when taking into account systems used as blanks[9] for absorbance measurements. The systems produced are illustrated in Figure 5 - 7. Five systems were produced on each tie-line with the middle system being produced in triplicates. The concentrations and partition coefficients for α-amylase and myoglobin are reported in Table 5 - 5.

The average myoglobin concentration in the top phase is 0.19mg mL^{-1}. At a lower phase ratio, the volume of the top-phase is reduced therefore there is potential to increase the concentration of myoglobin, however, this is not observed suggesting that the solubility limit of myoglobin has been reached. The concentration of myoglobin in the bottom phase, however, changes significantly in the systems investigated due to the change in phase size. For example, at a predicted volume ratio of 0.52 in system 1 the concentration of myoglobin in the bottom phase is 2.49mg mL^{-1} whereas the concentration is 5.23mg mL^{-1} in system 21 where the predicted volume ratio is 2.27. Similar observations are made for α-amylase where the average bottom phase concentration average is 0.07mg mL^{-1}. As with myoglobin, the concentration of α-amylase in the top phase changes significantly with the volume of the top-phase. This means that the aqueous two-phase systems can be used for concentration as well as purification.

The partition coefficients of α-amylase and myoglobin reported in Table 5 - 5 are very high across all 21 systems. For α-amylase, when changing from the low to high tie-line, the median partition coefficient increases from 25 to 46. For myoglobin, the median partition coefficient decreases from 0.072 to 0.050 when changing from low to high tie-line length. These findings are in agreement with literature which suggests that increasing the tie-line length typically amplifies the partitioning characteristics (Johansson et al., 1998, Rito-Palomares and Hernandez, 1998, Zhi et al., 2004).

To test whether calculated volume ratios were reasonable, the concentration of α-amylase and myoglobin in the top and bottom phase, respectively, were calculated. The concentrations were calculated using the median partition coefficients in Table 5 - 5 and the total mass of α-amylase (0.880mg) and myoglobin (1.55mg) added to each system. This represents a system where the feed stream is ~37% pure based on protein

[9] Blank systems containing no protein were used to correct for any potential absorbance due to the components used to make the aqueous two-phase system (PEG, phosphate, NaCl and water).

concentration. In reality the ratio of product to impurity will vary based on characteristics of cell lines used in upstream bioreactors. Figure 5 - 8 is a parity plot comparing calculated concentrations of α-amylase in the top phase and myoglobin in the bottom phase to experimentally determined concentrations. The majority of the points are close to the line y=x suggesting that the model predicted volume ratios are reasonably accurate.

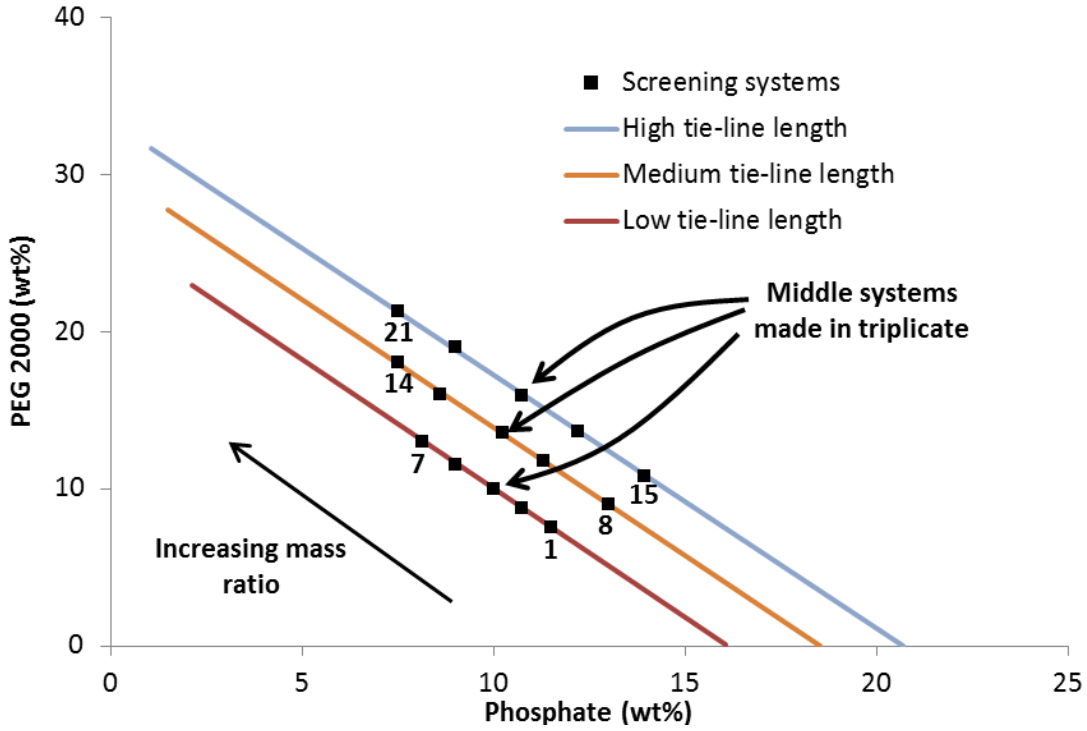


Figure 5 - 7: Screening space design for high-throughput experiments. Five systems screened for low, medium and high tie-line lengths. Middle systems formed in triplicates. Tie-lines determined from model predictions. Numbers correspond to system number.

Table 5 - 5: Concentration and partition coefficients of α-amylase (Amy) and myoglobin (Myo) in the PEG 2,000-phosphate (pH 7)-6wt% NaCl aqueous two-phase system at different locations in the phase diagram. Systems within dashed lines correspond to middle systems which were formed in triplicates.

System #	Predicted mass ratio (-)	Predicted volume ratio (-)	Concentration (mg/mL) Top Amy	Top Myo	Bottom Amy	Bottom Myo	Partition Coefficient, Ki Amy	Partition Coefficient, Ki Myo	Median tie-line Ki Amy	Median tie-line Ki Myo
1	0.48	0.52	3.23	0.20	0.09	2.49	35	0.080		
2	0.61	0.66	2.75	0.20	0.11	2.72	26	0.074		
3	0.77	0.83	2.42	0.21	0.09	3.02	27	0.071		
4	0.77	0.83	2.41	0.21	0.10	2.93	24	0.073	25	0.072
5	0.77	0.83	2.39	0.21	0.10	2.90	25	0.072		
6	1.02	1.09	2.01	0.23	0.10	3.31	20	0.069		
7	1.31	1.41	1.13	0.17	0.10	3.81	11	0.045		
8	0.48	0.52	3.23	0.17	0.05	2.57	68	0.068		
9	0.74	0.80	2.36	0.18	0.06	3.09	42	0.058		
10	0.95	1.04	2.03	0.18	0.06	3.47	34	0.052		
11	0.95	1.04	2.04	0.18	0.07	3.45	30	0.053	31	0.053
12	0.95	1.04	2.04	0.18	0.07	3.43	31	0.053		
13	1.38	1.51	1.69	0.19	0.07	4.13	26	0.045		
14	1.83	2.00	1.42	0.18	0.06	4.92	24	0.037		
15	0.52	0.57	2.93	0.20	0.05	2.79	56	0.071		
16	0.76	0.84	2.29	0.18	0.03	3.07	66	0.059		
17	1.03	1.13	1.94	0.18	0.04	3.70	46	0.048		
18	1.03	1.13	1.98	0.18	0.06	3.67	33	0.050	46	0.050
19	1.03	1.13	1.97	0.18	0.10	3.76	21	0.048		
20	1.48	1.64	1.63	0.17	0.08	4.49	22	0.038		
21	2.05	2.27	1.43	0.27	0.01	5.23	96	0.051		
Average concentration			**2.16**	**0.19**	**0.07**	**3.47**				

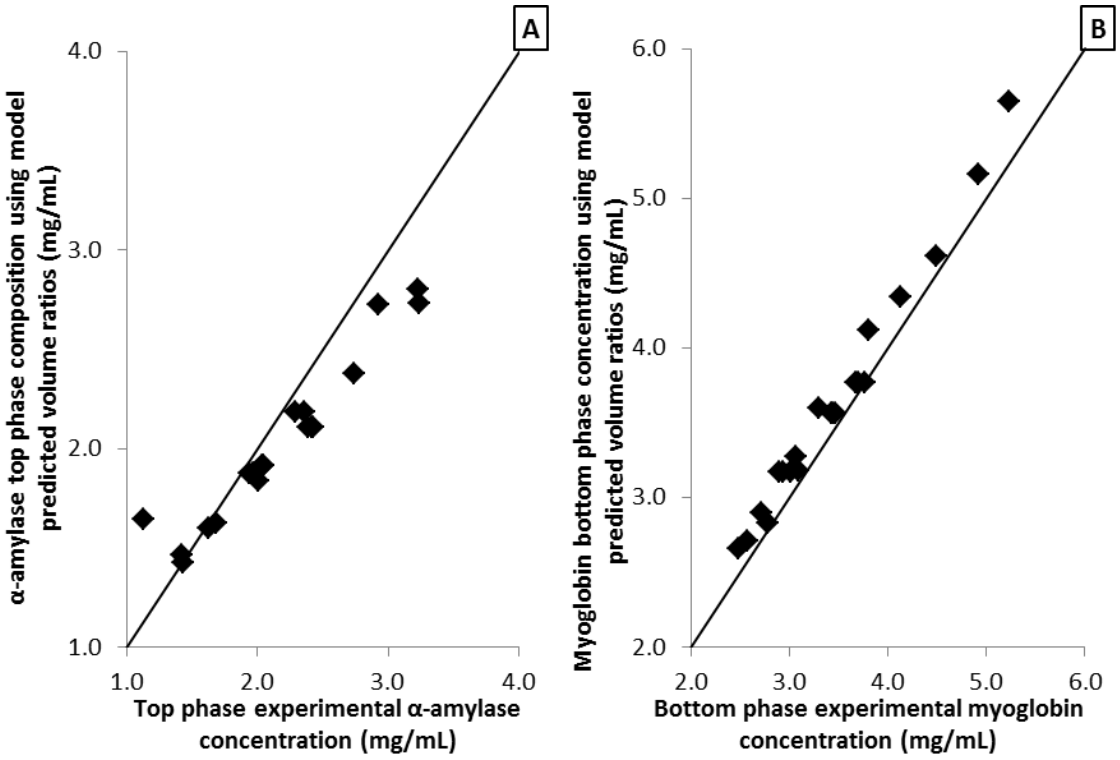


Figure 5 - 8: Parity plot of calculated concentration using predicted volume ratios against experimental concentrations. Experimental concentrations are from high-throughput screening experiments (see Table 5 - 5 for concentration data). A) Predicted α-amylase top phase concentration compared to experimental α-amylase top phase concentration. B) Predicted myoglobin bottom phase concentration compared to experimental myoglobin bottom phase concentration.

5.2.5 α-amylase and myoglobin process simulation

Estimation of interaction parameters using the experimental phase diagram and determination of partition coefficients for α-amylase and myoglobin have now provided enough information to conduct process simulations (using the methods established in Chapter 3) to explore further the extraction of α-amylase from myoglobin using a PEG 2,000-phosphate (pH 7)-6wt% NaCl aqueous two-phase system further.

The aim of the simulations was to develop a process to separate α-amylase (the product) from myoglobin (the impurity). The following specifications were made regarding the simulated process:

1. The feed (stream 4 in stage N) to the aqueous two-phase extraction process is from a 300L perfusion bioreactor with a perfusion rate of 300L day^{-1} (Gorenflo et al., 2002, Pollock et al., 2013).
2. Composition out of the bioreactor is 50% α-amylase and 50% myoglobin each with a concentration of 0.650mg mL^{-1}. Typically host cell proteins are created in a similar order of magnitude to that of the protein for example 1mg host cell protein for 1mg of therapeutic protein (Zhu-Shimoni et al., 2014).
3. The feed from bioreactor needs to be mixed with phosphate stock solution before entering the aqueous two-phase extraction process therefore the initial 0.650mg mL^{-1} protein concentration is diluted.

A volume of the perfusion bioreactor and perfusion rate were chosen based off assumptions presented by Pollock et al. (2013) for perfusion processes. A titre of 0.650mg mL^{-1} was used based on a fed batch culture titre of 2mg mL^{-1} as perfusion titres typically range from 20-40% of fed batch cultures.

A process diagram of the counter-current simulation setup is shown in Figure 5 - 9. Simulations were conducted to investigate the impact of tie-line length, phase ratio and partition coefficient on separation performance. The partition coefficients used are presented Table 5 - 6. There are three groups of partition coefficients which are named low, medium and experimental. The experimental partition coefficients correspond to the median partition coefficients reported in Table 5 - 5 obtained using high-throughput screening. The low and medium partition coefficients correspond to hypothetical partition coefficients used in process simulations to investigate the impact of poorer separation. Further parameters used in process simulations for the counter-current extraction process are listed in Table 5 - 7.

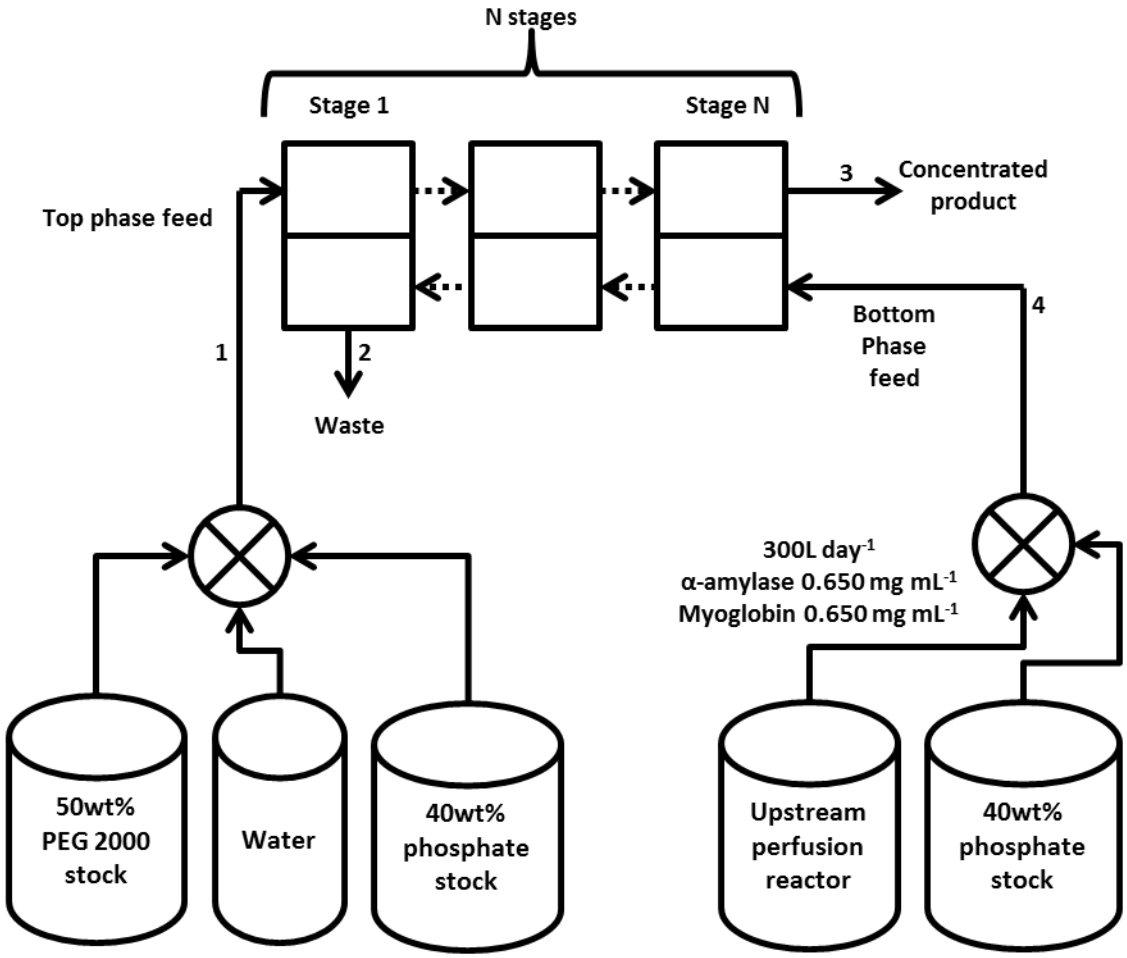


Figure 5 - 9: Schematic of counter-current operation used in simulations.

Table 5 - 6: Partition coefficients used in process simulations for separation of α-amylase (Amy) and myoglobin (Myo). Experimental partition coefficients correspond to those obtained from high-throughput screening. Low and medium correspond to hypothetical partition coefficients resulting in a poorer separation. The low and medium partition coefficients were deliberately chosen to be different to the experimental partition coefficient values to determine the impact K_i has on process performance.

Tie-line length	Partition coefficient (K_i)					
	Low		Medium		Experimental	
	Amy	Myo	Amy	Myo	Amy	Myo
Low	3	0.528	15	0.106	25	0.072
Medium	3	0.528	15	0.106	31	0.053
High	3	0.528	15	0.106	46	0.050

Table 5 - 7: Model input parameters/variables used for process simulations of aqueous two-phase extraction of α-amylase and myoglobin.

Model parameter	Value
Number of stages (N)	1 to 6
Molecular weight (g/mole):	
Water	18
PEG	2000
Phosphate	157.01
NaCl	58.44
α-amylase	72377
myoglobin	17082
Interaction parameter (kg/mole)	
PEG-PEG	-67
PEG-Phosphate	-4.7
Phosphate-Phosphate	-2.4
Stage 1 feed	
Flow rate (kg/s)	$(5.56\times10^{-3}$ to $7.31\times10^{-3})$ × desired phase ratio
α-amylase and myoglobin concentration (mg/mL)	0.001
Mass fraction PEG 2,000	0.23 to 0.33
Mass fraction phosphate	0.01 to 0.02
Mass fraction NaCl	0.06
Stage N feed	
Flow rate (kg/s)	5.56×10^{-3} to 7.31×10^{-3}
α-amylase and myoglobin concentration (mg/mL)	0.36 to 0.45
Mass fraction PEG 2,000	2.8×10^{-5}
Mass fraction phosphate	0.16 to 0.21
Mass fraction NaCl	0.06
Stage hold up (kg)	40
Desired phase ratio (-)	0.125 to 1.500

Impact of phase composition (tie-line length)

Tie-line length is linked to the overall and equilibrium compositions of components used to form an aqueous two-phase system. At higher overall compositions (and equilibrium compositions) the length of the tie-line is greater. This makes the tie-line length a useful metric to describe the effects of overall and equilibrium compositions. It is important to consider the tie-line length because they influence aqueous two-phase system performance in many ways. Firstly, operating at a higher tie-line length increases the strength of partitioning as shown in the experimental partition coefficients reported in Table 5 - 5. Secondly, operating at a higher tie-line results in greater density differences between the phases (Amrhein et al., 2014). The difference in density will influence the settling time of phases and therefore the residence time of an extraction operation. Thirdly, operating at higher tie-lines requires higher concentrations of polymers and salts which will influence process costs. All of these considerations will greatly influence the choice of equipment utilised.

To investigate the impact of operating at different tie-line lengths, a single continuous extraction stage was simulated using the median experimental partition coefficients reported in Table 5 - 6. Figure 5 - 10 shows how operating at different tie-line lengths influences concentration factor, process yield and purity at different phase ratios for a continuous single stage extraction process.

Concentration factor was defined as:

$$Concentration\ factor = \frac{Concentration\ of\ protein\ out\ of\ two-phase\ system}{Concentration\ of\ protein\ in\ reactor\ outlet\ i.e.\ 0.650 mg\ ml^{-1}} \quad (5.1)$$

Operating at higher tie-lines resulted in lower concentration factors, this is an interesting result because the partition coefficient of α-amylase at the low tie-line length is 25 whereas at the high tie-line length it is 46; typically increases in partition coefficient result in higher concentrations. The reduction in concentration factor (Figure 5 - 10A) can therefore be explained by the fact that operating at a higher tie-line length requires the need to dilute the reactor outlet stream with more phosphate stock solution to reach the desired bottom phase phosphate composition. The reduction in protein concentration as a result of greater dilution cannot be overcome by the higher α-amylase partition coefficient of 46.

One way to increase the concentration factor is to reduce the phase ratio, however, this comes at a cost of lower yield as shown in Figure 5 - 10B. For example, for the high tie-line length, decreasing phase ratio from 0.250 to 0.125 increases concentration factor from 1.9 to 3.4 but decreases yield from 91% to 85%. As with the multi-cycle batch extraction system examined in Chapter 4, operating at a lower phase ratio results in higher purities.

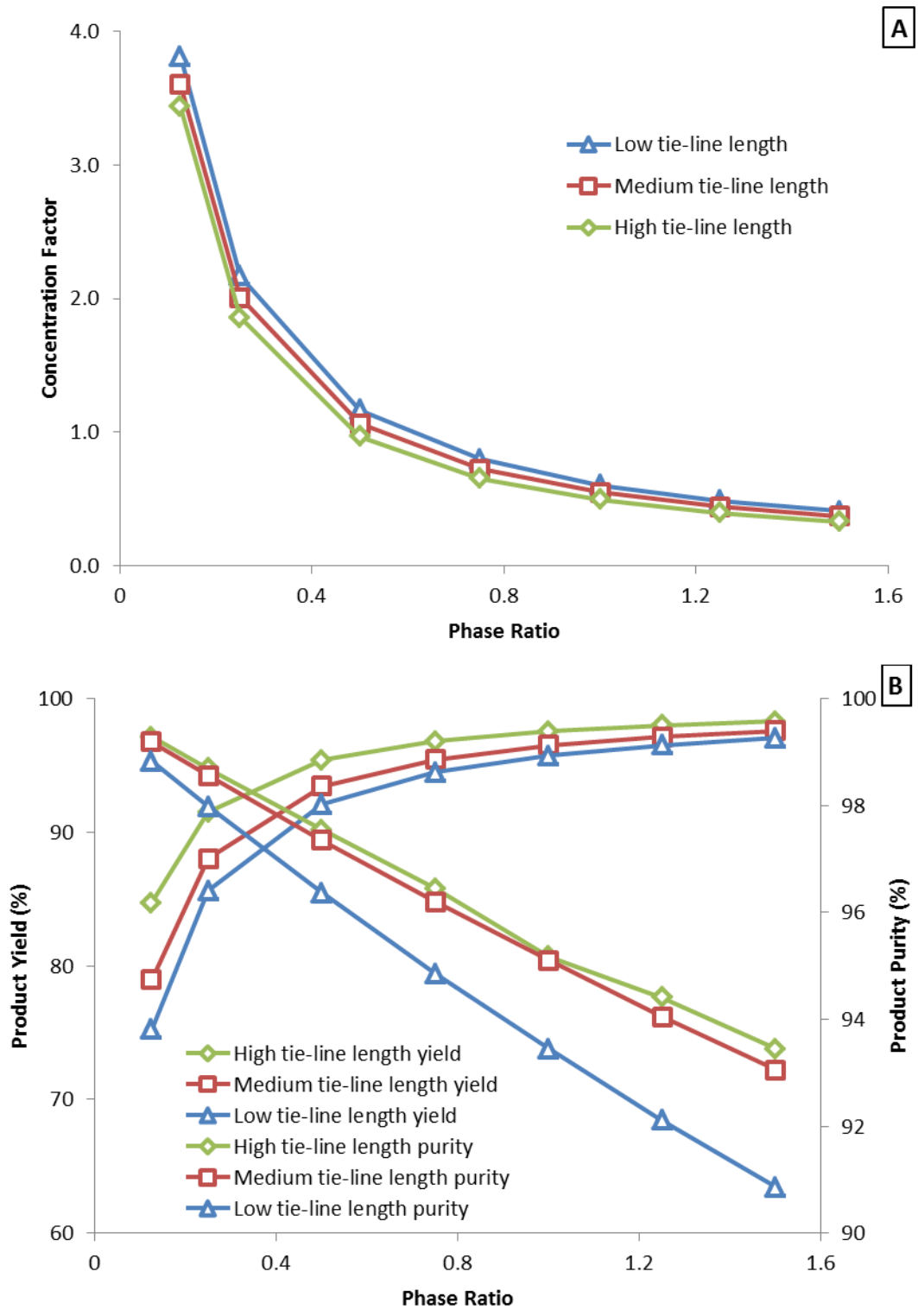


Figure 5 - 10: Impact of operating at different tie-lines and phase ratios using experimental partition coefficient data using a single extraction stage. The shorter tie-line lengths result in greater product concentration factors, however, lower yield and purity. A) Impact of phase ratio on α-amylase concentration factor. B) Impact of phase ratio on product yield and purity.

Impact of partition coefficient (K_i) and number of stages

The experimental partition coefficients reported in this work are reasonably high, however, it is useful to understand how robust a process would be if the operating under poorer performing conditions. Such insights are useful in designing appropriate control strategies which can be used to mitigate the effects of key process parameters such as the partition coefficient changing. This would aid in the development of a robust design space such as is often discussed in Quality by Design. Therefore, additional "low" and "medium" partition coefficient values were considered to investigate the importance of partition coefficient magnitude on process design and performance. These additional partition coefficients are reported in Table 5 - 6.

Figure 5 - 11 shows that changing from the "medium" to "low" partition coefficient values results in an absolute yield drop of 38% at a phase ratio of 0.125 and 14% at a phase ratio of 1.5. The absolute drop in purity is 17% at a phase ratio of 0.125 and 23% at a phase ratio of 1.5. These findings suggest that changes in partition coefficient have a nonlinear influence on process performance where yield is more sensitive to partition coefficient at low phase ratios and process purity is more sensitive to partition coefficients at high phase ratios. The higher yield drop at low phase ratios is likely due to the larger volume of the bottom phase which results in greater amounts of product being retained in the bottom phase at weaker partition coefficients. The higher purity drop at high phase ratios is likely due to the larger volume of the top phase which results in greater amounts of impurities being extracted from the bottom phase at weaker partition coefficients. It is important to identify these sensitivities early on in process development so that a robust experimental design space can be specified which takes into account potential deviations from optimal feed conditions.

Yield constraints on continuous counter-current extraction

Often in process development, a certain performance specification is required for each unit operation, for example a yield greater than 90% or purity greater than 99.99%. To test how such constraints influence process design of aqueous two-phase systems, a yield constraint of 90% was set for aqueous two-phase systems, consisting of 1 to 6 stages.

Figure 5 - 12 shows how concentration factor, phase ratio and purity are affected when the number of stages is increased and yield is kept constant at 90%. Figure 5 - 12A shows that increasing the number of stages results in greater concentration factors for

the low, medium and experimental partition coefficient values. This is because increasing the number of stages allows for lower phase ratios to be utilised while still maintaining a constant yield, this is shown in Figure 5 - 12B. However, the simulations also show that the benefit of adding additional extraction stages becomes limiting as more stages are added. For example, the percentage increase in concentration factor when increasing from 1 extraction stage to 2 extraction stages is ~250% whereas the percentage increase in concentration factor when increasing from 5 extraction stages to 6 extraction stages is ~7%. Another benefit of increasing the number of stages is greater purities, this is particularly important if the partitioning is weak as is the case when using the low magnitude partition coefficients where purity can be increased from ~60% to ~80% by using 6 counter-current extraction stages.

However, in reality the number of available stages is entirely dependent on the type of liquid-liquid extraction equipment available. If the process simulations show that performance requirements cannot be met using the required equipment, then there is a need to re-evaluate separation techniques or search for ways to increase the strength of partitioning. The benefit of using process models in such a situation is that model validation experiments which are likely to not meet performance specifications can be avoided. To show how process simulations can be used to guide equipment choice, the types of equipment available are discussed next.

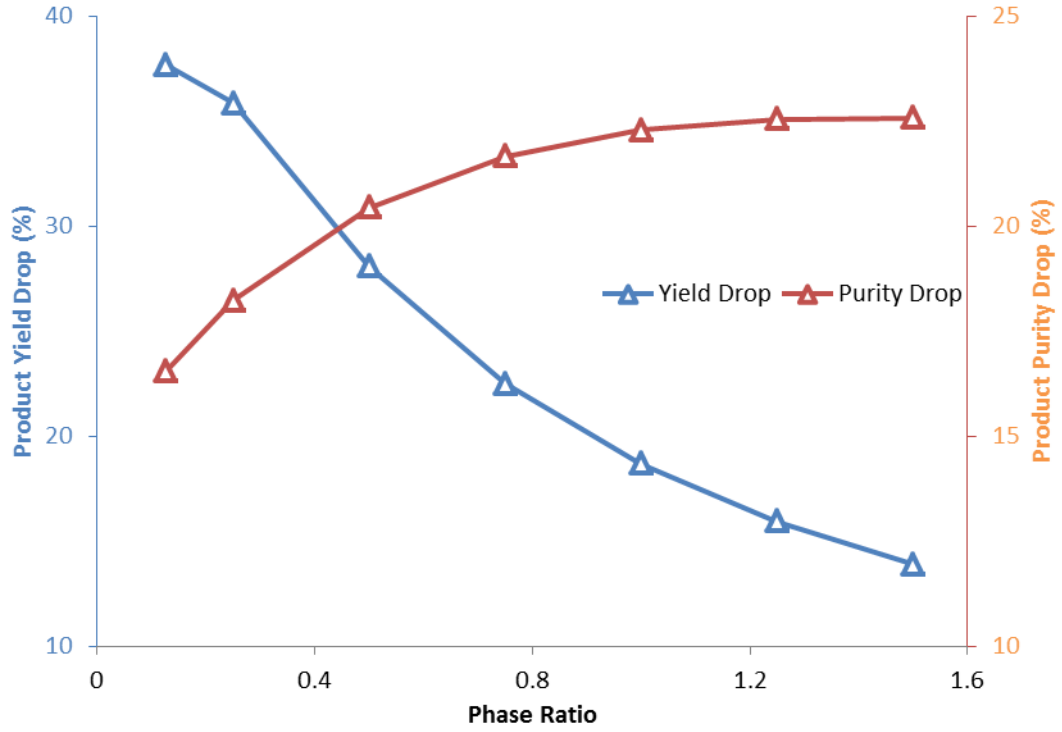


Figure 5 - 11: Absolute drop in yield and purity at different phase ratios when switching from medium magnitude partition coefficient values ($K_{amylase}$=15 and $K_{myoglobin}$=0.106) to low magnitude partition coefficient values ($K_{amylase}$=3 and $K_{myoglobin}$=0.528) in a single stage extraction process operating at the low tie-line length. For example if the yield using K_1 was 80% and K_2 was 60% at a phase ratio of 0.1 then the absolute yield drop would be 20%. See Table 5 - 8 for original yield and purity data. The medium and low partition coefficients used are reported in Table 5 - 6.

Table 5 - 8: Yield and purities corresponding to absolute drops which are plotted in Figure 5 - 11. The medium and low partition coefficients used are reported in Table 5 - 6.

	Yield of α-amylase						
Phase Ratio	**0.125**	**0.25**	**0.5**	**0.75**	**1.0**	**1.25**	**1.5**
Medium Ki	65	78	88	91	93	95	95
Low Ki	27	43	60	69	75	79	82
	Purity of α-amylase						
Phase Ratio	**0.125**	**0.25**	**0.5**	**0.75**	**1.0**	**1.25**	**1.5**
Medium Ki	98	97	95	93	91	89	88
Low Ki	81	79	74	71	68	67	65

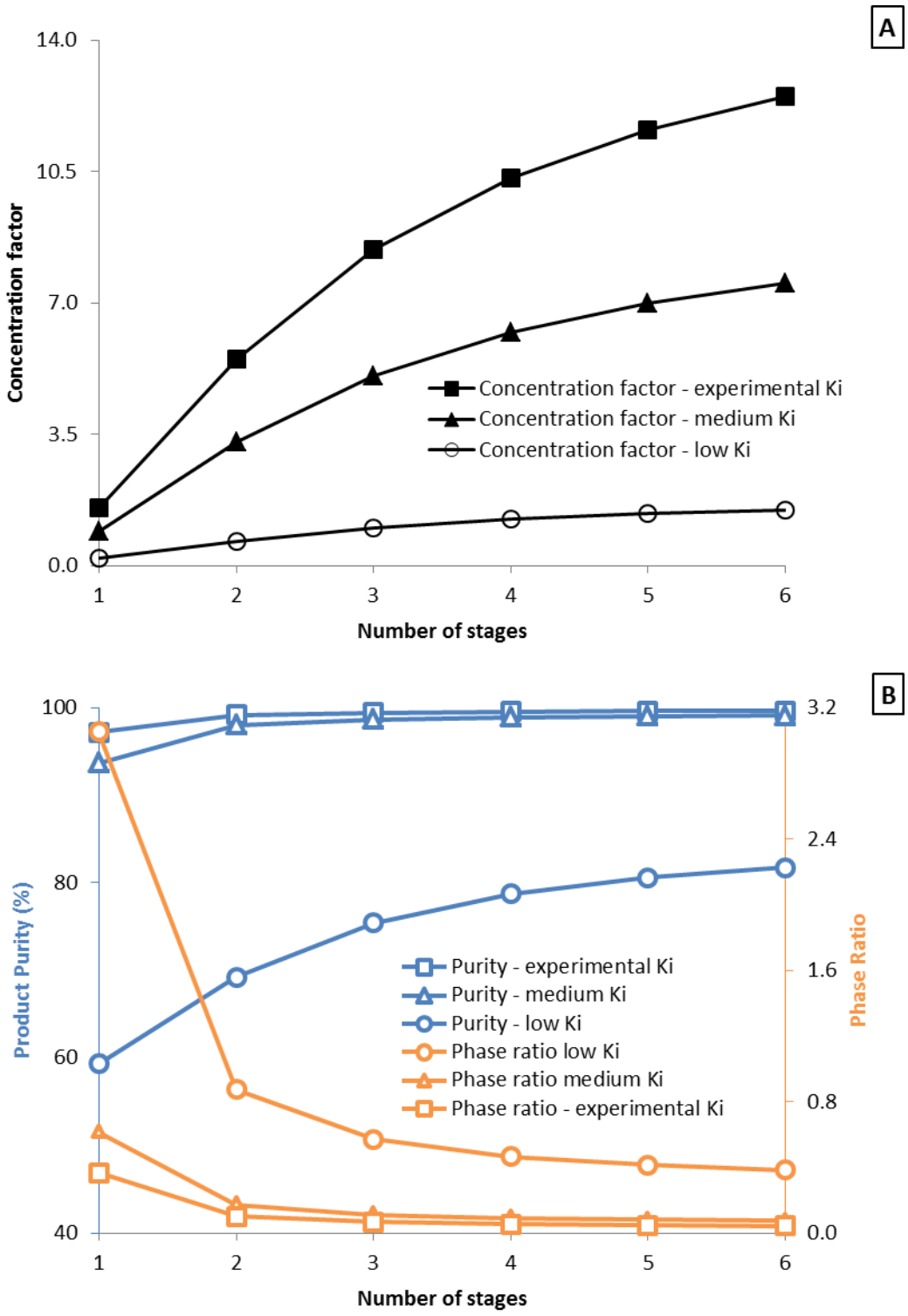


Figure 5 - 12: Impact of number of stages on separation performance when a 90% yield constraint is set using low, medium and experimental K_i values. Process is simulated at the low tie-line length. A) Impact of number of stages on concentration factor. B) Impact of number of stages on product purity and phase ratio.

Equipment choice

There are three main types of liquid-liquid extraction equipment (Law and Todd, 2008):

1) Mixer-settler tanks
2) Counter-current columns
3) Centrifugal contactors

The choice of equipment will depend primarily on strength of partitioning which will govern process design and operation as discussed above. The characteristics of the different types of equipment are displayed in Table 5 - 9.

For example, using a mixer-settler setup to extract α-amylase from myoglobin seems like a reasonable choice due to the high partition coefficients, however, if a process requirement was to concentrate the product, then using a mixer-settler setup may not provide the ability to operate at extreme phase ratios required to increase concentrations as shown in the previous simulations.

If large numbers of extraction stages are required such as when there is weak partitioning, then again the use of a mixer-settler set-up should be avoided as it is cumbersome to operate such equipment in counter-current format; in such cases counter-current columns and centrifugal contactors may be more appropriate.

In addition, further considerations must be made such as the available plant floor space. Mixer-settler tanks take up a large amount of floor space whereas columns take up a lot of vertical space, both restrictions would impact on the maximum number of possible stages. Further considerations include the cost of the equipment which is not covered here.

With such restrictions on design, process simulations can provide useful insights into whether or not a given extraction process is worth pursuing further experimentally. For example, if using the very simplified model in this work suggests many extraction stages are needed to achieve the desired level of separation, then a separation may not be worth further developing as real systems are far from ideal.

Table 5 - 9: Characteristics of liquid-liquid extraction technologies compiled from Law and Todd (2008) and Edison et al. (2016).

	Equipment Choice		
	Mixer-settler	**Counter-current column**	**Centrifugal-contactor**
Residence time	High	High	Low
Height	Short	Tall	Short
Floor space	High	Low	Medium
Scale-up difficulty	Average	Average	Easy
Hold-up volume	High	Medium	Low
Steady-state speed	Slow	Medium	High
Ability to use extreme phase ratios	Average	High	High
Potential to use many stages	Low	High	Medium

5.3 Conclusions

The main aim of this work was to use a combination of high-throughput experimentation and process modelling to develop a suitable aqueous two-phase extraction process to separate α-amylase from myoglobin. The combination of process modelling and experimentation was chosen so that the greater information could be gained from limited experimental resources.

Prior to high-throughput screening, a three factor lab scale screening study (4g system size) was conducted which investigated the impact of PEG molecular weight, pH and NaCl amount of partitioning of α-amylase and myoglobin in the corresponding aqueous two-phase systems. It was found that reducing PEG molecular weight from 4,000 to 2,000 resulted in higher concentrations of both α-amylase and myoglobin which is likely due to less precipitation at the interface. The recovery in the PEG 2,000 aqueous two-phase systems was ~40-50% higher than in the PEG 4,000 aqueous two-phase systems for both α-amylase and myoglobin. Increasing NaCl wt% from 0wt% to 6wt% resulted in higher α-amylase concentrations in the top phase. Changing pH from 7.0 to 7.5 had no observable effect on partitioning of α-amylase and myoglobin. α-amylase partitioned into the top phase while myoglobin partitioned into the bottom phase. Based on this initial screening study, an aqueous two-phase system composed of PEG 2,000-phosphate (pH 7.0)-6wt% NaCl was characterised in more detail using high-throughput screening.

To describe experimental phase equilibria PEG 2,000-phosphate (pH 7.0)-6wt% NaCl aqueous two-phase system interaction parameters for the osmotic virial equations were estimated using the parameter estimation and experimental techniques presented in Chapter 5. Interaction parameters were found to have high standard deviation values, however, model predicted phase equilibria compositions when using minimum and maximum parameter combinations were still within the calculated experimental error therefore were suitable for process simulations.

Partitioning of α-amylase and myoglobin was investigated at three different tie-line lengths and multiple phase ratios using high-throughput experimentation (1g system size) using automated liquid handling. Partition coefficients of α-amylase and myoglobin were amplified at increasing tie-line lengths while concentration of each protein could be altered by changing phase ratio. Model predictions of phase ratio were

found to be reasonable as predicted concentration of α-amylase in the top phase and myoglobin in the bottom phase matched experimentally determined concentrations.

Simulations were conducted to investigate the performance of a continuous counter-current extraction process downstream of a continuous 300L perfusion bioreactor. Operating at longer tie-line lengths resulted in greater yields and purities but lower concentration factors due to increased dilution. Sensitivity of process performance to partition coefficient value was investigated, and as expected weaker partition coefficients resulted in poorer purities and yields. Interestingly, simulations showed the resultant drop in yield and purity due to poor partition coefficients is greatest at low and high phase ratios, respectively. Depending on what is deemed more important, a separation system can therefore be designed to be more robust to yield or purity changes using appropriate phase ratios. When a yield constraint of 90% was set, increasing the number of stages allowed for greater concentrations and purities to be obtained. This is because extraction systems can operate at lower phase ratios when compared to operating a single stage.

The benefit of the combined modelling and experimental approach presented in this chapter is that resource intensive experiments such as continuous counter-current extractions can be instead simulated to see whether a separation is feasible based on process performance requirements and available equipment. In addition, the approach allows the exploration of potential robustness issues, based on fundamental understanding of the separation system, which is useful when meeting the requirements of Quality by Design principles. The main limitation of the approach presented in this work is that the process modelling is simple and does not take into account more complicated phenomena such as thermodynamics of electrolytes or mechanics of phase settling. These issues could be addressed in future work by making use of more detailed thermodynamic models and extending the equilibrium stage model to take into account process inefficiencies.

Chapter 6: Conclusions and future work

6.1 Review of project objectives

The aim of this thesis was to combine high-throughput experimental techniques with process modelling/simulation techniques to better understand aqueous two-phase systems for their use in the separation of biopharmaceuticals. To conduct simulations a mechanistic process model describing an aqueous two-phase system extraction stage was developed. To generate the data required to customise the mechanistic process model to specific experimental systems a series of high-throughput experimental techniques were employed which used automated liquid-handling. The enzyme α-amylase was chosen in simulation and experimental studies as a mimic to acid alpha-glucosidase (GAA) which was a therapeutic protein that was being developed by BioMarin Pharmaceuticals to treat Pompe disease. The motivation for using a combined approach was to address the issue of resource intensive process development for aqueous two-phase systems. This is because the performance of an aqueous two-phase extraction separation process is influenced by many factors (e.g. tie-line length and phase ratio) which are challenging to completely explore experimentally. The use of process models can help gain insights into the impact of such factors and help develop a robust process whilst minimising the use of experimental resources.

6.1.1 Dynamic modelling of aqueous two-phase systems to quantify the impact of bioprocess design, operation and variability

In Chapter 3, the modelling framework for conducting dynamic simulations of aqueous two-phase extraction was formulated and executed. Specifically an equilibrium stage model was developed to simulate dynamic aqueous two-phase extraction using multi-cycle batch extraction and continuous counter-current modes of operation. The motivation for developing a dynamic model was based on the fact that most separation processes have dynamic regimes of operation (e.g. start-up/shutdown, response to disturbances or set-point changes etc.) which are important to understand and investigate. The model is particularly useful to understand the dynamics of continuous processes which are being increasingly considered by pharmaceutical companies to improve processing efficiencies. Thermodynamic interaction parameters, which were used to describe phase equilibria in the process model, were estimated from experimental phase equilibria data reported in literature. The developed process model was used to simulate the purification of α-amylase from impurities. The impact of

operating conditions, such as number of extraction cycles and phase ratio, were investigated. The benefits of dynamic modelling were demonstrated by exploring the performance of aqueous two-phase extraction performance under variable operating conditions. Two scenarios were considered: (1) impact of drifting product titre from upstream production and (2) impact of poor control of NaCl concentration in the feeds of the extraction process. The work of Chapter 3 showed how a relatively simple dynamic process model can be used to gain powerful insights into the operation of an aqueous two-phase extraction process beyond the scope of conventional lab scale experimental techniques. The simplicity of the model is also its main limitation. This is because the model uses a relatively simple thermodynamic model which does not take into account factors such as electrolyte interactions and polymer molecular weight which in reality are present and have a significant impact on phase equilibria calculations. In addition, the dynamics of phase settling is not considered which makes it challenging to understand how long an actual phase separation will take. Finally, each equilibrium stage is assumed to be 100% efficient, in reality one would likely need more equilibrium stages than the model predicts due to inefficiencies such as mass transfer resistances.

6.1.2 Quantification of uncertainty in phase diagrams for estimation of interaction parameters

Chapter 4 looked at the understanding of experimental phase diagrams for aqueous two-phase systems generated using a combination of high-throughput and lab scale experiments. This included the development of a two-stage high-throughput binodial curve screening strategy to minimise the uncertainty in the binodial curve location. It was found that using a two-stage screening approach resulted in an overall reduction of required experimental resources when compared to doing a single high-resolution screen. The resultant binodial curve was used to determine tie-lines based on phase ratios measured from aqueous two-phase systems prepared at lab scale.

A major aim of Chapter 4 was to quantify the uncertainty in phase diagrams generated using high-throughput screening as to date there have not been any reported methods to do this. This involved the development of a novel protocol to quantify the uncertainty in the corresponding experimental phase diagrams as a result of experimental design and experimental measurement errors. The major source of uncertainty in the phase equilibria compositions was found to be due to errors in density measurement. Once the uncertainty in phase equilibria was quantified, binary interaction parameters used in the

process model described in Chapter 3 to describe liquid-liquid phase equilibria were estimated. In this work, improving the quality of density measurements was found to be the best way to improve uncertainty in estimated thermodynamic interaction parameters. In addition, increasing the number of experimentally determined tie-lines was also shown to reduce the uncertainty of estimated interaction parameters. Using these improvements resulted in interaction parameters with a relative standard deviation of less than 11%. The main assumptions in Chapter 4 are similar to those of Chapter 3 where it was assumed that electrolyte interactions, do not exist as the thermodynamic equations used did not take into account effect of charge. Future work should look at applying the methodology to a more comprehensive thermodynamic set of equations. The methods developed in Chapter 4 were utilised in Chapter 5 where the separation of α-amylase from myoglobin is investigated using aqueous two-phase extraction.

6.1.3 High-throughput screening and simulation of α-amylase and myoglobin separation

Chapter 5 looked at using aqueous two-phase extraction to separate α-amylase from myoglobin. The aim of Chapter 5 was to integrate the methods developed in Chapters 3 and 4 to demonstrate the possibility of integrating high-throughput experimental screening techniques with process models used for predictive simulations of aqueous two-phase extraction. The information obtained from the process simulations can then be used to better understand aqueous two-phase system process performance.

A preliminary lab scale screening study was first conducted to narrow down the aqueous two-phase system to further characterise using high-throughput screening. This helped reduce the number of calibrations required when using the automated liquid handling robots. A PEG 2,000-phosphate (pH 7)-6wt% NaCl aqueous two-phase system was chosen. The phase diagram and interaction parameters were determined using the methods of Chapter 4. In addition, high-throughput screening of α-amylase and myoglobin partitioning was conducted to investigate the impact of different phase ratios and tie-line lengths. In this work myoglobin was chosen as a contaminant protein as it is easy to quantify experimentally. In reality there would hundreds of other contaminants which would need to be appropriately quantified.

The experimental partition coefficients were used to simulate the continuous counter-current extraction of α-amylase from myoglobin using the modelling techniques presented in Chapter 3. The impact of process operating conditions such as number of stages, tie-line length and phase ratio on yield, purity and concentration factor were

investigated. The resultant information is very useful as it helps determine the type of equipment required to make an aqueous two-phase system extraction feasible. For example, in this work, operating at a low phase ratio results in higher concentration factors, especially for multi-stage systems, however, this would require equipment capable of operating at extreme phase ratios such as a counter-current column. One of the main limitations of the work conducted in Chapter 5 is that the simulations were not validated against a real counter-current system. This would be useful to do as it would enable researchers to determine key information such as stage efficiency and realistic settling times.

6.2 Recommendations for future work

The work in this thesis shows that there is great promise in using a combination of high-throughput experimentation and process modelling in the development of aqueous two-phase systems; however, there are many improvements that can be made to improve the work of this thesis. Below is a list of proposals for future work which are grouped into short and long term time scales.

6.2.1 Short term proposals

Short term proposal 1: Implementation of complimentary high-throughput analytical techniques to explore real protein systems

The work presented in this thesis looked at the separation of two-proteins, α-amylase and myoglobin, as it was relatively simple to quantify the concentration of each protein experimentally. In reality, there would be many more components to consider. Therefore to fully test the capabilities of the approach presented in this work would require investigations using more realistic systems containing more impurities. Quantification of these numerous impurities requires analytical techniques which are also high-throughput. A few technologies that could address this include ultra-performance liquid chromatography (UPLC), microfluidic based electrophoresis and use of soft sensors which combine spectral data with multivariate data analysis such as PLS modelling (Rathore et al., 2008, Glassey et al., 2011, LeSaout et al., 2016, Rüdt et al., 2017). After implementation of an appropriate analytical technique and determination of appropriate partition coefficients, the methods presented in this thesis would be used to explore the separation of complex mixtures.

Short term proposal 2: Model validation of continuous counter-current operations

The simulations in this thesis explored the continuous counter-current extraction of α-amylase from myoglobin. Unfortunately, there was not the opportunity to validate these models against experimental data to test the extent of assumptions such as 100% stage efficiency. Therefore, another proposal for short term future work would be to conduct lab-scale continuous counter-current experiments so that model validation could be conducted.

Short term proposal 3: Increasing the detail of the modelling approach presented

The modelling approach presented in this work is very simple as it uses equilibrium stages which assume 100% stage efficiency. As mentioned previously, this is not realistic therefore a proposal for future work would include extending the model to take

into account real process inefficiencies. In addition, the thermodynamic equations used in this work are relatively simple compared to more detailed approaches such as the SAFT equations of state which can take into account additional properties such as molecular weight and electrolyte interactions (Reschke et al., 2014). The benefit of using a more detailed model is that ability to use one set of parameters for a wider range of process conditions. Therefore future work should also include the consideration of a more detailed description of thermodynamics, however, care must be taken balance the level of modelling detail vs. the amount of experimental effort required.

6.2.2 Long term proposals

Long term proposal 1: Development of lab equipment specifically for the high-throughput development of aqueous two-phase extraction operations

A big challenge in using liquid handling robots to develop aqueous two-phase extraction is the sampling of phases without cross-contamination. This is particularly challenging when attempting to sample the bottom phase where the pipette tip has to travel through the top phase as well as potential precipitation at the interface. A proposal to combat this is to develop suitable labware which can be used to sample both phases without cross contamination. One way to achieve this would be to design a robot compatible 96 well-plate with mini valves at the bottom which could be opened and closed to drain the required amount of sample. In addition, such a device would also need to have good mixing characteristics to ensure that viscous polymer solutions mix well. Such devices would be useful as it would eliminate the need to perform pipetting calibrations for each type of aqueous two-phase system considered.

Long term proposal 2: Real time optimisation of aqueous two-phase extraction using model based high-throughput experimentation

In this proposal initially a pre-determined set of aqueous two-phase system conditions are screened for a given product using a liquid handling robot. The experimental partitioning data for each system considered is automatically passed to a simple mechanistic model (like presented in this thesis) of the desired extraction process. Process performance parameters such as purity and yield are used to calculate an objective function which is used by an optimiser to design new experiments given appropriate experimental constraints. The aim of the approach is to fully integrate and automate initial process development of aqueous two-phase extraction. The benefit of such an approach would be the ability to rapidly assess whether aqueous two-phase extraction is a feasible separation strategy for a given product. The ideas of this

proposal could also be used to with other separation techniques such as chromatography where there already exists good process models and high-throughput technologies.

References

ADEWUMI, M. 2017. *Solution Algorithms for VLE Problems* [Online]. Available: https://www.e-education.psu.edu/png520/m17_p4.html [Accessed 17-6 2017].

AHMAD, M. M., HAUAN, S. & PRZYBYCIEN, T. M. 2010. Flowsheet simulation of aqueous two-phase extraction systems for protein purification. *Journal of Chemical Technology and Biotechnology,* 85**,** 1575-1587.

AHMED SAMATOU, J., ENGBERT WENTINK, A., ALEXANDRA J ROSA, P., MARGARIDA AZEVEDO, A., RAQUEL AIRES-BARROS, M., BÄCKER, W. & GÓRAK, A. 2007. Modeling of counter current monoclonal antibody extraction using aqueous two-phase systems. *Computer Aided Chemical Engineering,* 24**,** 935-940.

ALBERTSSON, P. A. 1958. Partition of Proteins in Liquid Polymer-Polymer 2-Phase Systems. *Nature,* 182**,** 709-711.

ALBERTSSON, P. A., CAJARVILLE, A., BROOKS, D. E. & TJERNELD, F. 1987. Partition of Proteins in Aqueous Polymer 2-Phase Systems and the Effect of Molecular-Weight of the Polymer. *Biochimica Et Biophysica Acta,* 926**,** 87-93.

ALVAREZ-GUERRA, E., VENTURA, S. P. M., ALVAREZ-GUERRA, M., COUTINHO, J. A. P. & IRABIEN, A. 2016. Modeling of the binodal curve of ionic liquid/salt aqueous systems. *Fluid Phase Equilibria,* 426**,** 10-16.

AMRHEIN, S., SCHWAB, M. L., HOFFMANN, M. & HUBBUCH, J. 2014. Characterization of aqueous two phase systems by combining lab-on-a-chip technology with robotic liquid handling stations. *Journal of Chromatography A,* 1367**,** 68-77.

ANDREWS, B., HEAD, D., DUNTHORNE, P. & ASENJO, J. 1990. PEG activation and ligand binding for the affinity partitioning of proteins in aqueous two-phase systems. *Biotechnology Techniques,* 4**,** 49-54.

ANDREWS, B., SCHMIDT, A. & ASENJO, J. 2005a. Correlation for the partition behavior of proteins in aqueous two-phase systems: Effect of surface hydrophobicity and charge. *Biotechnology and Bioengineering,* 90**,** 380-390.

ANDREWS, B. A. & ASENJO, J. A. 1996. Protein partitioning equilibrium between the aqueous poly(ethylene glycol) and salt phases and the solid protein phase in poly(ethylene glycol) salt two-phase systems. *Journal of Chromatography B-Biomedical Applications,* 685**,** 15-20.

ANDREWS, B. A. & ASENJO, J. A. 2010. Theoretical and Experimental Evaluation of Hydrophobicity of Proteins to Predict their Partitioning Behavior in Aqueous Two Phase Systems: A Review. *Separation Science and Technology,* 45**,** 2165-2170.

ANDREWS, B. A., SCHMIDT, A. S. & ASENJO, J. A. 2005b. Correlation for the partition behavior of proteins in aqueous two-phase systems: Effect of surface hydrophobicity and charge. *Biotechnology and Bioengineering,* 90**,** 380-390.

ASENJO, J. A. & ANDREWS, B. A. 2011. Aqueous two-phase systems for protein separation: A perspective. *Journal of Chromatography A,* 1218**,** 8826-8835.

ASENJO, J. A., SCHMIDT, A. S., HACHEM, F. & ANDREWS, B. A. 1994. Model for Predicting the Partition Behavior of Proteins in Aqueous 2-Phase Systems. *Journal of Chromatography A,* 668**,** 47-54.

ATEFI, E., FYFFE, D., KAYLAN, K. B. & TAVANA, H. 2016. Characterization of Aqueous Two-Phase Systems from Volume and Density Measurements. *Journal of Chemical and Engineering Data,* 61**,** 1531-1539.

ATHA, D. H. & INGHAM, K. C. 1981. Mechanism of precipitation of proteins by polyethylene glycols. Analysis in terms of excluded volume. *Journal of Biological Chemistry,* 256**,** 12108-12117.

AZEVEDO, A. M., ROSA, P. A. J., FERREIRA, I. F. & AIRES-BARROS, M. R. 2007. Optimisation of aqueous two-phase extraction of human antibodies. *Journal of Biotechnology,* 132**,** 209-217.

AZEVEDO, A. M., ROSA, P. A. J., FERREIRA, I. F. & AIRES-BARROS, M. R. 2009. Chromatography-free recovery of biopharmaceuticals through aqueous two-phase processing. *Trends in Biotechnology,* 27**,** 240-247.

BALASUNDARAM, B., HARRISON, S. & BRACEWELL, D. G. 2009. Advances in product release strategies and impact on bioprocess design. *Trends in Biotechnology,* 27**,** 477-485.

BARBOSA, H., HINE, A. V., BROCCHINI, S., SLATER, N. K. H. & MARCOS, J. C. 2008. Affinity partitioning of plasmid DNA with a zinc finger protein. *Journal of Chromatography A,* 1206**,** 105-112.

BENAVIDES, J. & RITO-PALOMARES, M. 2008. Practical experiences from the development of aqueous two-phase processes for the recovery of high value biological products. *Journal of Chemical Technology and Biotechnology,* 83**,** 133-142.

BENSCH, M., SELBACH, B. & HUBBUCH, J. 2007. High throughput screening techniques in downstream processing: Preparation, characterization and optimization of aqueous two-phase systems. *Chemical Engineering Science,* 62**,** 2011-2021.

BESSEMANS, L., JULLY, V., DE RAIKEM, C., ALBANESE, M., MONIOTTE, N., SILVERSMET, P. & LEMOINE, D. 2016. Automated Gravimetric Calibration to Optimize the Accuracy and Precision of TECAN Freedom EVO Liquid Handler. *Jala,* 21**,** 693-705.

BHAMBURE, R., KUMAR, K. & RATHORE, A. S. 2011. High-throughput process development for biopharmaceutical drug substances. *Trends in Biotechnology,* 29**,** 127-135.

BIEGLER, L. T. & GROSSMANN, I. E. 2004. Retrospective on optimization. *Computers & Chemical Engineering,* 28**,** 1169-1192.

BOURGOIN, A. F. & NUSKEY, B. 2013. An outlook on US biosimilar competition. *New York: Thomson Reuters*.

BUILDER, S., HART, R., LESTER, P., OGEZ, J. & REIFSNYDER, D. 1995. *Aqueous multiple-phase isolation of polypeptide - US 5723310 A.* USA patent application.

BURKE, S. E. 2014. Combined Variance of Two Groups with Equal Numbers of Observations. Available: http://saraemilyburke.com/stats/CombinedVarianceEqualNs.pdf [Accessed 21/09/16].

CABEZAS, H. 1996. Theory of phase formation in aqueous two-phase systems. *Journal of Chromatography B-Biomedical Applications,* 680**,** 3-30.

CAMERETTI, L. F., SADOWSKI, G. & MOLLERUP, J. M. 2005. Modeling of aqueous electrolyte solutions with perturbed-chain statistical associated fluid theory. *Industrial & Engineering Chemistry Research,* 44**,** 3355-3362.

CARDOSO, M., SALCEDO, R., DE AZEVEDO, S. F. & BARBOSA, D. 2000. Optimization of reactive distillation processes with simulated annealing. *Chemical Engineering Science,* 55**,** 5059-5078.

CAVAZZUTI, M. 2013. Deterministic Optimization. *Optimization Methods: From Theory to Design Scientific and Technological Aspects in Mechanics.* Berlin, Heidelberg: Springer Berlin Heidelberg.

CHATTERJI, A., HISEY, R. T., JACOBY, R. & HOFFMAN, T. 2011. The follow-on biologics market: enter at your own risk.

CHEN, C. C. & MATHIAS, P. M. 2002. Applied thermodynamics for process modeling. *AIChE Journal,* 48**,** 194-200.

DEBYE, P. & HÜCKEL, E. 1923. ON THE THEORY OF ELECTROLYTES. I. FREEZING POINT DEPRESSION AND RELATED PHENOMENA. *Physikalische Zeitschrift,* 24**,** 185-206.

DECHAMBRE, D., WOLFF, L., PAULS, C. & BARDOW, A. 2014. Optimal Experimental Design for the Characterization of Liquid-Liquid Equilibria. *Industrial & Engineering Chemistry Research,* 53**,** 19620-19627.

DIEDERICH, P., AMRHEIN, S., HAMMERLING, F. & HUBBUCH, J. 2013. Evaluation of PEG/phosphate aqueous two-phase systems for the purification of the chicken egg white protein avidin by using high-throughput techniques. *Chemical Engineering Science,* 104**,** 945-956.

DITTRICH, P. S. & MANZ, A. 2006. Lab-on-a-chip: microfluidics in drug discovery. *Nature Reviews Drug Discovery,* 5**,** 210-218.

EDISON, M., WESORICK, S., ROBERTSON, M., PALAZZOLO, J. & MINBIOLE, K. 2016. *ENCYCLOPEDIA OF CHEMICAL ENGINEERING EQUIPMENT - EXTRACTORS* [Online]. Available: http://encyclopedia.che.engin.umich.edu/Pages/SeparationsChemical/Extractors/Extractors.html [Accessed 26/08 2016].

EDMOND, E. & OGSTON, A. 1968. An approach to the study of phase separation in ternary aqueous systems. *Biochem. J,* 109**,** 569-576.

EGLESE, R. W. 1990. Simulated Annealing - a Tool for Operational-Research. *European Journal of Operational Research,* 46**,** 271-281.

ERIKSSON, E., ALBERTSSON, P. A. & JOHANSSON, G. 1976. Hydrophobic Surface Properties of Erythrocytes Studied by Affinity Partition in Aqueous 2-Phase Systems. *Molecular and Cellular Biochemistry,* 10**,** 123-128.

ESPITIA-SALOMA, E., VAZQUEZ-VILLEGAS, P., RITO-PALOMARES, M. & AGUILAR, O. 2016. An integrated practical implementation of continuous aqueous two-phase systems for the recovery of human IgG: From the microdevice to a multistage bench-scale mixer-settler device. *Biotechnology Journal,* 11**,** 708-716.

FARRAR, D. E. 1982. Citation Classic - Multicollinearity in Regression-Analysis - the Problem Revisited. *Current Contents/Social & Behavioral Sciences***,** 22-22.

FERRARI, J. C., NAGATANI, G., CORAZZA, F. C., OLIVEIRA, J. V. & CORAZZA, M. L. 2009. Application of stochastic algorithms for parameter estimation in the liquid–liquid phase equilibrium modeling. *Fluid Phase Equilibria,* 280**,** 110-119.

FLORY, P. J. 1942. Thermodynamics of high polymer solutions. *The Journal of chemical physics,* 10**,** 51-61.

FORCINITI, D., HALL, C. K. & KULA, M. R. 1991a. Analysis of Polymer Molecular-Weight Distributions in Aqueous 2-Phase Systems. *Journal of Biotechnology,* 20**,** 151-161.

FORCINITI, D., HALL, C. K. & KULA, M. R. 1991b. Influence of Polymer Molecular-Weight and Temperature on Phase-Composition in Aqueous 2-Phase Systems. *Fluid Phase Equilibria,* 61**,** 243-262.

FORCINITI, D., HALL, C. K. & KULA, M. R. 1991c. Protein Partitioning at the Isoelectric Point - Influence of Polymer Molecular-Weight and Concentration and Protein Size. *Biotechnology and Bioengineering,* 38**,** 986-994.

FORCINITI, D., HALL, C. K. & KULA, M. R. 1991d. Temperature-Dependence of the Partition-Coefficient of Proteins in Aqueous 2-Phase Systems. *Bioseparation,* 2**,** 115-128.

FRANCO, T. T., ANDREWS, A. T. & ASENJO, J. A. 1996. Use of chemically modified proteins to study the effect of a single protein property on partitioning in aqueous two-phase systems: Effect of surface hydrophobicity. *Biotechnology and Bioengineering,* 49**,** 300-308.

FUHR, J. P., CHANDRA, A., ROMLEY, J., SHIH, T., MAY, S. G. & VANDERPUYE-ORGLE, J. 2015. Product naming, pricing, and market uptake of biosimilars. *Generics Biosimilars Initiat J,* 4**,** 64-71.

GASTEIGER, E., GATTIKER, A., HOOGLAND, C., IVANYI, I., APPEL, R. D. & BAIROCH, A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research,* 31**,** 3784-3788.

GAUTAM, S. & SIMON, L. 2006. Partitioning of beta-glucosidase from Trichoderma reesei in poly(ethylene glycol) and potassium phosphate aqueous two-phase systems: Influence of pH and temperature. *Biochemical Engineering Journal,* 30**,** 104-108.

GEAR, B. 2007. Backward differentiation formulas. *Scholarpedia,* 2**,** 3162.

GELADI, P. & KOWALSKI, B. R. 1986. Partial Least-Squares Regression - a Tutorial. *Analytica Chimica Acta,* 185**,** 1-17.

GILSON, I. 2016. *Microman - Specifications* [Online]. Available: http://www.gilson.com/en/Pipette/Products/44.224/Default.aspx?d=335 [Accessed 09/09 2016].

GLASSEY, J., GERNAEY, K. V., CLEMENS, C., SCHULZ, T. W., OLIVEIRA, R., STRIEDNER, G. & MANDENIUS, C. F. 2011. Process analytical technology (PAT) for biopharmaceuticals. *Biotechnology Journal,* 6**,** 369-377.

GLYK, A., SOLLE, D., SCHEPER, T. & BEUTEL, S. 2015. Optimization of PEG-salt aqueous two-phase systems by design of experiments. *Chemometrics and Intelligent Laboratory Systems,* 149**,** 12-21.

GOCKENBACH, M. S. 2003. Introduction to sequential quadratic programming. *Course material, Michigan Technological University,* 106.

GORENFLO, V. M., SMITH, L., DEDINSKY, B., PERSSON, B. & PIRET, J. M. 2002. Scale-up and optimization of an acoustic filter for 200 L/day perfusion of a CHO cell culture. *Biotechnology and Bioengineering,* 80**,** 438-444.

GOTTSCHALK, U. 2008. Bioseparation in antibody manufacturing: The good, the bad and the ugly. *Biotechnology Progress,* 24**,** 496-503.

GRAHAM, M. H. 2003. Confronting multicollinearity in ecological multiple regression. *Ecology,* 84**,** 2809-2815.

GUNDUZ, U. & KORKMAZ, K. 2000. Bovine serum albumin partitioning in an aqueous two-phase system: Effect of pH and sodium chloride concentration. *Journal of Chromatography B,* 743**,** 255-258.

HATTI-KAUL, R. 2000. *Aqueous two-phase systems : methods and protocols,* Totowa, N.J., Humana Press.

HATTI-KAUL, R. 2001. Aqueous two-phase systems - A general overview. *Molecular Biotechnology,* 19**,** 269-277.

HUDDLESTON, J., VEIDE, A., KOHLER, K., FLANAGAN, J., ENFORS, S. O. & LYDDIATT, A. 1991. The Molecular-Basis of Partitioning in Aqueous 2-Phase Systems. *Trends in Biotechnology,* 9**,** 381-388.

HUENUPI, E., GOMEZ, A., ANDREWS, B. A. & ASENJO, J. A. 1999. Optimization and design considerations of two-phase continuous protein separation. *Journal of Chemical Technology and Biotechnology,* 74**,** 256-263.

HUGGINS, M. L. 1942. Some Properties of Solutions of Long-chain Compounds. *The Journal of Physical Chemistry,* 46**,** 151-158.

JEŻOWSKI, J., BOCHENEK, R. & POPLEWSKI, G. 2007. On application of stochastic optimization techniques to designing heat exchanger-and water

networks. *Chemical Engineering and Processing: Process Intensification,* 46**,** 1160-1174.
JOHANSSON, H. O., KARLSTROM, G., TJERNELD, F. & HAYNES, C. A. 1998. Driving forces for phase separation and partitioning in aqueous two-phase systems. *Journal of Chromatography B,* 711**,** 3-17.
JOULIA, X., MAGGIOCHI, P., KOEHRET, B., PARADOWSKI, H. & BARTUEL, J. 1986. Hybrid method for phase equilibrium flash calculations. *Fluid Phase Equilibria,* 26**,** 15-36.
JUNGBAUER, A. 2013. Continuous downstream processing of biopharmaceuticals. *Trends in Biotechnology,* 31**,** 479-492.
KANG, C. H. & SANDLER, S. I. 1987. Phase-Behavior of Aqueous 2-Polymer Systems. *Fluid Phase Equilibria,* 38**,** 245-272.
KAUL, A. 2000. The Phase Diagram. *In:* HATTI-KAUL, R. (ed.) *Aqueous Two-Phase Systems: Methods and Protocols: Methods and Protocols.* Totowa, NJ: Humana Press.
KING, R. S., BLANCH, H. W. & PRAUSNITZ, J. M. 1988. Molecular Thermodynamics of Aqueous 2-Phase Systems for Bioseparations. *AIChE Journal,* 34**,** 1585-1594.
KIRK , J. H. & PASCAL , P. V. 1999. *Methods for protein purification using aqueous two-phase extraction - US 6437101 B1*. USA patent application.
KLATT, K. U. & MARQUARDT, W. 2009. Perspectives for process systems engineering-Personal views from academia and industry. *Computers & Chemical Engineering,* 33**,** 536-550.
KONTOGEORGIS, G. M. & FOLAS, G. K. 2009. Cubic Equations of State: The Classical Mixing Rules. *Thermodynamic Models for Industrial Applications: From Classical and Advanced Mixing Rules to Association Theories***,** 39-77.
KRISHNA, R. & WESSELINGH, J. A. 1997. Review article number 50 - The Maxwell-Stefan approach to mass transfer. *Chemical Engineering Science,* 52**,** 861-911.
KUMAR, A. & AWASTHI, A. 2009. *Bioseparation Engineering,* I. K. International Publishing House.
LACKI, K. M. 2014. High throughput process development in biomanufacturing. *Current Opinion in Chemical Engineering,* 6**,** 25-32.
LAW, J. D. & TODD, T. A. 2008. Liquid-Liquid Extraction Equipment. *Introduction to Nuclear Chemistry and Fuel Cycle Separations Course, Consortium for Risk Evaluation With Stakeholder Participation, http://www. cresp. org/education/courses/shortcourse*.
LEADER, B., BACA, Q. J. & GOLAN, D. E. 2008. Protein therapeutics: A summary and pharmacological classification. *Nature Reviews Drug Discovery,* 7**,** 21-39.
LESAOUT, X., COSTIOLI, M., JORDAN, L., LAMBERT, J., BEIGHLEY, R., PROVENCHER, L., GERWE, B., MCGUIRE, K., VERLINDEN, N. & BARRY, A. 2016. Automated small-scale protein purification and analysis for accelerated development of protein therapeutics. *Engineering in Life Sciences,* 16**,** 179-184.
LIM, S. & SUH, M. 2015. Product and Process Innovation in the Development Cycle of Biopharmaceuticals. *Journal of Pharmaceutical Innovation,* 10**,** 156-165.
LUTHER, J. R. & GLATZ, C. E. 1994. Genetically-Engineered Charge Modifications to Enhance Protein Separation in Aqueous 2-Phase Systems - Electrochemical Partitioning. *Biotechnology and Bioengineering,* 44**,** 147-153.
LUTHER, J. R. & GLATZ, C. E. 1995. Genetically-Engineered Charge Modifications to Enhance Protein Separation in Aqueous 2-Phase Systems - Charge Directed Partitioning. *Biotechnology and Bioengineering,* 46**,** 62-68.

MAESTRO, B., VELASCO, I., CASTILLEJO, I., AREVALO-RODRIGUEZ, M., CEBOLLA, A. & SANZ, J. M. 2008. Affinity partitioning of proteins tagged with choline-binding modules in aqueous two-phase systems. *Journal of Chromatography A,* 1208**,** 189-196.

MATHSWORKS®. 2016a. *How Simulated Annealing Works* [Online]. Available: http://uk.mathworks.com/help/gads/how-simulated-annealing-works.html [Accessed 14/09 2016].

MATHSWORKS®. 2016b. *Simulated Annealing Options* [Online]. Available: http://uk.mathworks.com/help/gads/simulated-annealing-options.html [Accessed 14/09 2016].

MATHWORKS®. 2017. *Global Optimization Toolbox* [Online]. Available: https://uk.mathworks.com/products/global-optimization.html [Accessed 17-06 2017].

MCCAMISH, M. & WOOLLETT, G. 2011. Worldwide experience with biosimilar development. *Mabs,* 3**,** 209-217.

MERCHUK, J. C., ANDREWS, B. A. & ASENJO, J. A. 1998. Aqueous two-phase systems for protein separation: Studies on phase inversion. *Journal of Chromatography B: Biomedical Sciences and Applications,* 711**,** 285-293.

METAL SUPERMARKETS. 2016. *DIFFERENCE BETWEEN ANNEALING AND TEMPERING* [Online]. Available: https://www.metalsupermarkets.com/difference-annealing-tempering/ [Accessed 14/09 2016].

METTLER-TOLEDO 2005. Operating Instructions - Mettler Toledo B-S line of balances: AB-S and PB-S.

MISTRY, S. L., KAUL, A., MERCHUK, J. C. & ASENJO, J. A. 1996. Mathematical modelling and computer simulation of aqueous two-phase continuous protein extraction. *Journal of Chromatography A,* 741**,** 151-163.

MUNDGES, J., ZIEROW, J. & ZEINER, T. 2015. Experiment and simulation of an aqueous two-phase extraction process for the purification of a monoclonal antibody. *Chemical Engineering and Processing,* 95**,** 31-42.

MÜNDGES, J., ZIEROW, J. & ZEINER, T. 2015. Experiment and simulation of an aqueous two-phase extraction process for the purification of a monoclonal antibody. *Chemical Engineering and Processing: Process Intensification,* 95**,** 31-42.

NAEEM, S. & SADOWSKI, G. 2011. pePC-SAFT: Modeling of polyelectrolyte systems 2. Aqueous two-phase systems. *Fluid Phase Equilibria,* 306**,** 67-75.

NFOR, B. K., VERHAERT, P. D. E. M., VAN DER WIELEN, L. A. M., HUBBUCH, J. & OTTENS, M. 2009. Rational and systematic protein purification process development: the next generation. *Trends in Biotechnology,* 27**,** 673-679.

NICHITA, D. V. & LEIBOVICI, C. F. 2013. A rapid and robust method for solving the Rachford–Rice equation using convex transformations. *Fluid Phase Equilibria,* 353**,** 38-49.

OELMEIER, S. A., DISMER, F. & HUBBUCH, J. 2011. Application of an Aqueous Two-Phase Systems High-Throughput Screening Method to Evaluate mAb HCP Separation. *Biotechnology and Bioengineering,* 108**,** 69-81.

ORIGINLAB 2012. OriginLab, Northampton, MA. 8.6 ed.

PANTELIDES, C. C., NAUTA, M., MATZOPOULOS, M. & GROVE, H. Equation-oriented process modelling technology: recent advances & current perspectives. 5th Annual TRC-Idemitsu Workshop, 2015.

PAZUKI, G. R., TAGHIKHANI, V. & VOSSOUGHI, M. 2010. Prediction of the Partition Coefficients of Biomolecules in Polymer-Polymer Aqueous Two-Phase

Systems Using the Artificial Neural Network Model. *Particulate Science and Technology,* 28**,** 67-73.
PENG, J. J., LEXTRAIT, S., EDGAR, T. F. & ELDRIDGE, R. B. 2002. A comparison of steady-state equilibrium and rate-based models for packed reactive distillation columns. *Industrial & Engineering Chemistry Research,* 41**,** 2735-2744.
PERRY, R. H. & GREEN, D. W. 2008. *Perry's chemical engineers' handbook,* New York, McGraw-Hill.
PETICOLAS, K. 1986. *IMMISCIBILITY: OIL AND WATER SEPARATION* [Online]. Available: http://fphoto.photoshelter.com/image/I0000yZ9mGuSItSU [Accessed 19/11 2012].
PHILIP, E. & ELIZABETH, W. 2010. Sequential quadratic programming methods. Technical Report, UCSD Department of Mathematics, NA-10-03.
PLANAS, J., LEFEBVRE, D., TJERNELD, F. & HAHN-HÄGERDAL, B. 1997. Analysis of phase composition in aqueous two-phase systems using a two-column chromatographic method: Application to lactic acid production by extractive fermentation. *Biotechnology and Bioengineering,* 54**,** 303-311.
POLLOCK, J., HO, S. V. & FARID, S. S. 2013. Fed-batch and perfusion culture processes: Economic, environmental, and operational feasibility under uncertainty. *Biotechnology and Bioengineering,* 110**,** 206-219.
PRAUSNITZ, J. M. 1980. *Computer calculations for multicomponent vapor-liquid and liquid-liquid equilibria,* Englewood Cliffs, N.J., Prentice-Hall.
PRINZ, A., KOCH, K., GORAK, A. & ZEINER, T. 2014. Multi-stage laccase extraction and separation using aqueous two-phase systems: Experiment and model. *Process Biochemistry,* 49**,** 1020-1031.
PROCESS SYSTEMS ENTERPRISE 2017. gPROMS.
PRZYBYCIEN, T. M., PUJAR, N. S. & STEELE, L. M. 2004. Alternative bioseparation operations: life beyond packed-bed chromatography. *Current Opinion in Biotechnology,* 15**,** 469-478.
RAHIMPOUR, F., HATTI-KAUL, R. & MAMO, G. 2016. Response surface methodology and artificial neural network modelling of an aqueous two-phase system for purification of a recombinant alkaline active xylanase. *Process Biochemistry,* 51**,** 452-462.
RAJENDRAN, A., PAREDES, G. & MAZZOTTI, M. 2009. Simulated moving bed chromatography for the separation of enantiomers. *Journal of Chromatography A,* 1216**,** 709-738.
RATHORE, A. S. & WINKLE, H. 2009a. Quality by design for biopharmaceuticals. *Nature biotechnology,* 27**,** 26.
RATHORE, A. S. & WINKLE, H. 2009b. Quality by design for biopharmaceuticals. *Nat Biotech,* 27**,** 26-34.
RATHORE, A. S., WOOD, R., SHARMA, A. & DERMAWAN, S. 2008. Case study and application of process analytical technology (PAT) towards bioprocessing: II. Use of ultra-performance liquid chromatography (UPLC) for making real-time pooling decisions for process chromatography. *Biotechnology and Bioengineering,* 101**,** 1366-1374.
RESCHKE, T., BRANDENBUSCH, C. & SADOWSKI, G. 2014. Modeling aqueous two-phase systems: I. Polyethylene glycol and inorganic salts as ATPS former. *Fluid Phase Equilibria,* 368**,** 91-103.
RITO-PALOMARES, M. & HERNANDEZ, M. 1998. Influence of system and process parameters on partitioning of cheese whey proteins in aqueous two-phase systems. *Journal of Chromatography B: Biomedical Sciences and Applications,* 711**,** 81-90.

RODRIGUES, A. E. & MINCEVA, M. 2005. Modelling and simulation in chemical engineering: Tools for process innovation. *Computers & Chemical Engineering,* 29**,** 1167-1183.

ROQUE, A. C. A., LOWE, C. R. & TAIPA, M. A. 2004. Antibodies and genetically engineered related molecules: Production and purification. *Biotechnology Progress,* 20**,** 639-654.

ROSA, P. A. J., AZEVEDO, A. M. & AIRES-BARROS, M. R. 2007. Application of central composite design to the optimisation of aqueous two-phase extraction of human antibodies. *Journal of Chromatography A,* 1141**,** 50-60.

ROSA, P. A. J., AZEVEDO, A. M., SOMMERFELD, S., MUTTER, M., AIRES-BARROS, M. R. & BACKER, W. 2009. Application of aqueous two-phase systems to antibody purification: A multi-stage approach. *Journal of Biotechnology,* 139**,** 306-313.

ROSA, P. A. J., FERREIRA, I. F., AZEVEDO, A. M. & AIRES-BARROS, M. R. 2010. Aqueous two-phase systems: A viable platform in the manufacturing of biopharmaceuticals. *Journal of Chromatography A,* 1217**,** 2296-2305.

RÜDT, M., BRISKOT, T. & HUBBUCH, J. 2017. Advances in downstream processing of biologics–Spectroscopy: An emerging process analytical technology. *Journal of Chromatography A,* 1490**,** 2-9.

SALGADO, J. C., ANDREWS, B. A., ORTUZAR, M. F. & ASENJO, J. A. 2008. Prediction of the partitioning behaviour of proteins in aqueous two-phase systems using only their amino acid composition. *Journal of Chromatography A,* 1178**,** 134-144.

SAMATOU, J. A. 2012. *Modelling and Simulation of Antibody Purification by Aqueous Two-Phase Extraction.* Technischen Universität Dortmund.

SAMATOU, J. A., WENTINK, A. E., ROSA, P. A. J., AZEVEDO, A. M., AIRES-BARROS, M. R., BACKER, W. & GORAK, A. 2007. Modeling of Counter Current Monoclonal Antibody Extraction using Aqueous Two-Phase Systems. *17th European Symposium on Computer Aided Process Engineering,* 24**,** 935-940.

SARAVANAN, S., RAO, J. R., NAIR, B. U. & RAMASAMI, T. 2008. Aqueous two-phase poly(ethylene glycol)-poly(acrylic acid) system for protein partitioning: Influence of molecular weight, pH and temperature. *Process Biochemistry,* 43**,** 905-911.

SAS INSTITUTE 2017. JMP Pro. 12.0.1 ed.: SAS Institute.

SCHMIDT, A. S., ANDREWS, B. A. & ASENJO, J. A. 1996. Correlations for the partition behavior of proteins in aqueous two-phase systems: Effect of overall protein concentration. *Biotechnology and Bioengineering,* 50**,** 617-626.

SCHWAAB, M. & PINTO, J. C. 2007. Optimum reference temperature for reparameterization of the Arrhenius equation. Part 1: Problems involving one kinetic constant. *Chemical Engineering Science,* 62**,** 2750-2764.

SCOTT, R. L. 1949. The thermodynamics of high polymer solutions. V. Phase equilibria in the ternary system: polymer 1—polymer 2—solvent. *The Journal of chemical physics,* 17**,** 279-284.

SCWARZ, D. M. 2010. *Fast and Robust Curve Intersections* [Online]. Available: http://uk.mathworks.com/matlabcentral/fileexchange/11837-fast-and-robust-curve-intersections.

SHIRE, S. J., SHAHROKH, Z. & LIU, J. 2004. Challenges in the development of high protein concentration formulations. *Journal of Pharmaceutical Sciences,* 93**,** 1390-1402.

SHUKLA, A. A., HUBBARD, B., TRESSEL, T., GUHAN, S. & LOW, D. 2007. Downstream processing of monoclonal antibodies—application of platform approaches. *Journal of Chromatography B,* 848**,** 28-39.

SILVA, D. F. C., AZEVEDO, A. M., FERNANDES, P., CHU, V., CONDE, J. P. & AIRES-BARROS, M. R. 2014. Determination of aqueous two phase system binodal curves using a microfluidic device. *Journal of Chromatography A,* 1370**,** 115-120.

SIMON, L. & GAUTAM, S. 2004. Modeling continuous aqueous two-phase systems for control purposes. *Journal of Chromatography A,* 1043**,** 135-147.

SOARES, R. R., AZEVEDO, A. M., VAN ALSTINE, J. M. & AIRES-BARROS, M. R. 2015. Partitioning in aqueous two-phase systems: Analysis of strengths, weaknesses, opportunities and threats. *Biotechnology Journal,* 10**,** 1158-1169.

SOOHOO, J. R. & WALKER, G. M. 2009. Microfluidic aqueous two phase system for leukocyte concentration from whole blood. *Biomedical Microdevices,* 11**,** 323-329.

TAIT, A. S., HOGWOOD, C. E., SMALES, C. M. & BRACEWELL, D. G. 2012. Host cell protein dynamics in the supernatant of a mAb producing CHO cell line. *Biotechnology and Bioengineering,* 109**,** 971-982.

TECAN. 2017. *Freedom EVOware®* [Online]. Available: http://diagnostics.tecan.com/products/software/freedom_evoware [Accessed 13/06/ 2017].

TEH, Y. S. & RANGAIAH, G. P. 2002. A study of equation-solving and Gibbs free energy minimization methods for phase equilibrium calculations. *Chemical Engineering Research & Design,* 80**,** 745-759.

THE WORLD BANK 2016. 2.15 World Development Indicators: Health systems. http://wdi.worldbank.org/table/2.15#: The World Bank.

TORRES-ACOSTA, M. A., AGUILAR-YANEZ, J. M., RITO-PALOMARES, M. & TITCHENER-HOOKER, N. J. 2016. Economic analysis of uricase production under uncertainty: Contrast of chromatographic purification and aqueous two-phase extraction (with and without PEG recycle). *Biotechnology Progress,* 32**,** 126-133.

TU, J. V. 1996. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *Journal of Clinical Epidemiology,* 49**,** 1225-1231.

TUBIO, G., NERLI, B. & PICO, G. 2004. Relationship between the protein surface hydrophobicity and its partitioning behaviour in aqueous two-phase systems of polyethyleneglycol-dextran. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 799**,** 293-301.

TUBIO, G., NERLI, B. & PICO, G. 2007. Partitioning features of bovine trypsin and alpha-chymotrypsin in polyethyleneglycol- sodium citrate aqueous two-phase systems. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences,* 852**,** 244-249.

TULSYAN, A., FORBES, J. F. & HUANG, B. A. 2012. Designing priors for robust Bayesian optimal experimental design. *Journal of Process Control,* 22**,** 450-462.

VALENTE, K. N., LENHOFF, A. M. & LEE, K. H. 2015. Expression of Difficult-to-Remove Host Cell Protein Impurities During Extended Chinese Hamster Ovary Cell Culture and Their Impact on Continuous Bioprocessing. *Biotechnology and Bioengineering,* 112**,** 1232-1242.

WALSH, G. 2010. Biopharmaceutical benchmarks 2010. *Nature biotechnology,* 28**,** 917.

WALSH, G. 2014. Biopharmaceutical benchmarks 2014. *Nature biotechnology,* 32**,** 992-1000.

WEISMAN, R. 2014. New Genzyme pill will cost patients $310,250 a year. *The Boston Globe* [Online]. Available: https://www.bostonglobe.com/business/2014/09/02/new-genzyme-pill-treat-rare-gaucher-disease-will-cost-patients-year/5thkIb587nKi7zRAb9GgxM/story.html [Accessed 28/10/2016].

WOLD, S., ESBENSEN, K. & GELADI, P. 1987. Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems,* 2**,** 37-52.

WU, Y. T., LIN, D. Q. & ZHU, Z. Q. 1998. Thermodynamics of aqueous two-phase systems - the effect of polymer molecular weight on liquid-liquid equilibrium phase diagrams by the modified NRTL model. *Fluid Phase Equilibria,* 147**,** 25-43.

WU, Z., HU, G., WANG, K., ZASLAVSKY, B. Y., KURGAN, L. & UVERSKY, V. N. 2017. What are the structural features that drive partitioning of proteins in aqueous two-phase systems? *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics,* 1865**,** 113-120.

XIE, I. H., WANG, M. H., CARPENTER, R. & WU, H. Y. 2004. Automated calibration of TECAN Genesis liquid handling workstation utilizing an online balance and density meter. *Assay and Drug Development Technologies,* 2**,** 71-80.

ZAFARANI-MOATTAR, M. T. & GASEMI, J. 2002. Liquid-liquid equilibria of aqueous two-phase systems containing polyethylene glycol and ammonium dihydrogen phosphate or diammonium hydrogen phosphate. Experiment and correlation. *Fluid Phase Equilibria,* 198**,** 281-291.

ZAFARANI-MOATTAR, M. T. & SADEGHI, R. 2001. Liquid-liquid equilibria of aqueous two-phase systems containing polyethylene glycol and sodium dihydrogen phosphate or disodium hydrogen phosphate - Experiment and correlation. *Fluid Phase Equilibria,* 181**,** 95-112.

ZAFARANI-MOATTAR, M. T., SADEGHI, R. & HAMIDI, A. A. 2004. Liquid-liquid equilibria of an aqueous two-phase system containing polyethylene glycol and sodium citrate: experiment and correlation. *Fluid Phase Equilibria,* 219**,** 149-155.

ZHI, W., SONG, J., BI, J. & OUYANG, F. 2004. Partial purification of α-amylase from culture supernatant of Bacillus subtilis in aqueous two-phase systems. *Bioprocess Biosyst Eng,* 27**,** 3-7.

ZHU-SHIMONI, J., YU, C., NISHIHARA, J., WONG, R. M., GUNAWAN, F., LIN, M., KRAWITZ, D., LIU, P., SANDOVAL, W. & VANDERLAAN, M. 2014. Host cell protein testing by ELISAs and the use of orthogonal methods. *Biotechnology and Bioengineering,* 111**,** 2367-2379.

Chapter 7: Appendices

7.1 Protein absorbance standard curves

Below are standard curves used for protein quantification using absorbance at 562nm using the BCA assay, 409nm and 280nm. For 280nm and 409nm curves are generated using 200uL of sample in low UV absorbance 96 well plates with an approximate path length of 0.6cm. For the BCA assay standard curves, absorbance is measured using 25µL of sample and 200µL of BCA assay reagent.

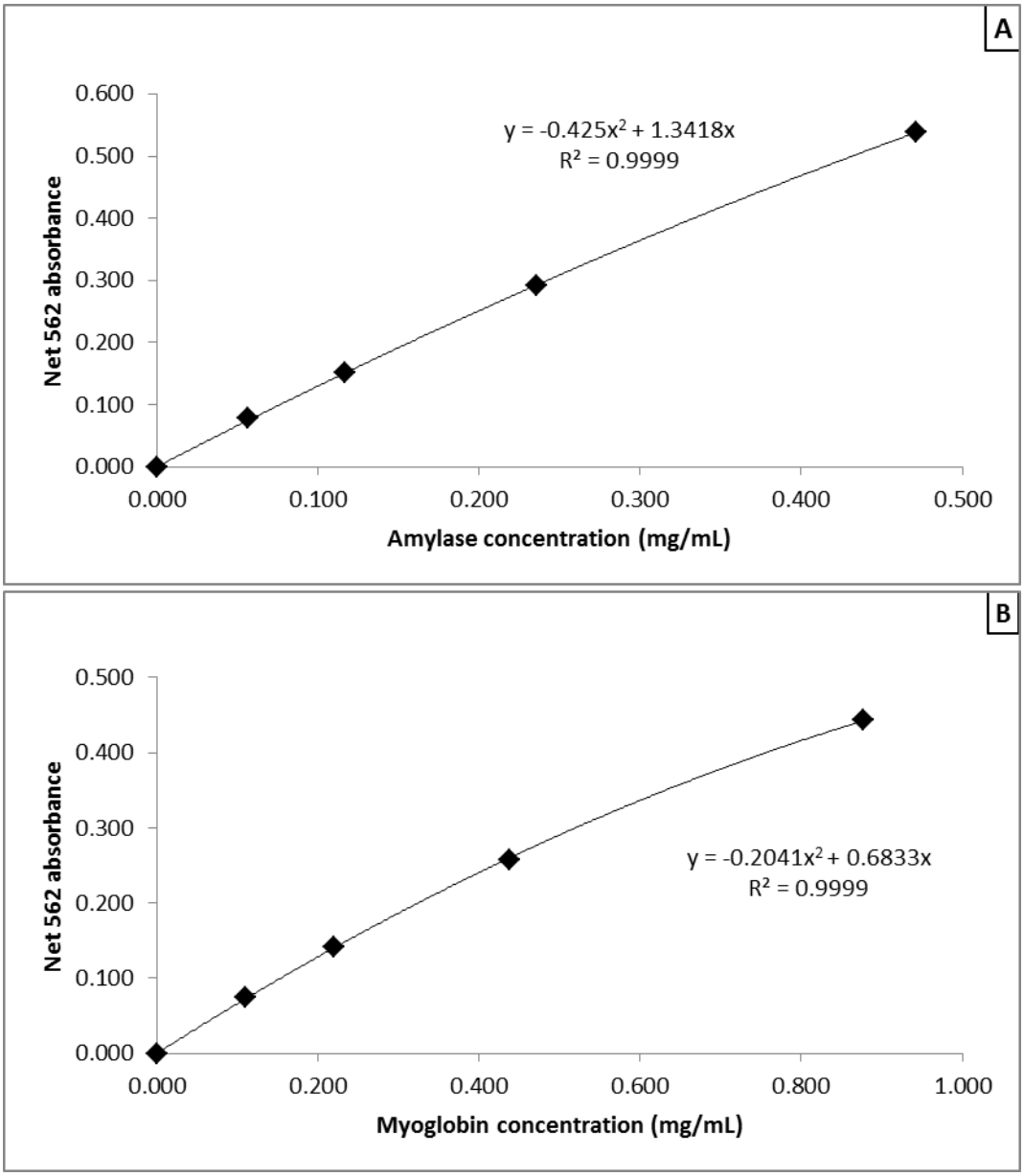


Figure 7 - 1: Absorbance at 562nm for amylase (A) and myoglobin (B) using BCA assay.

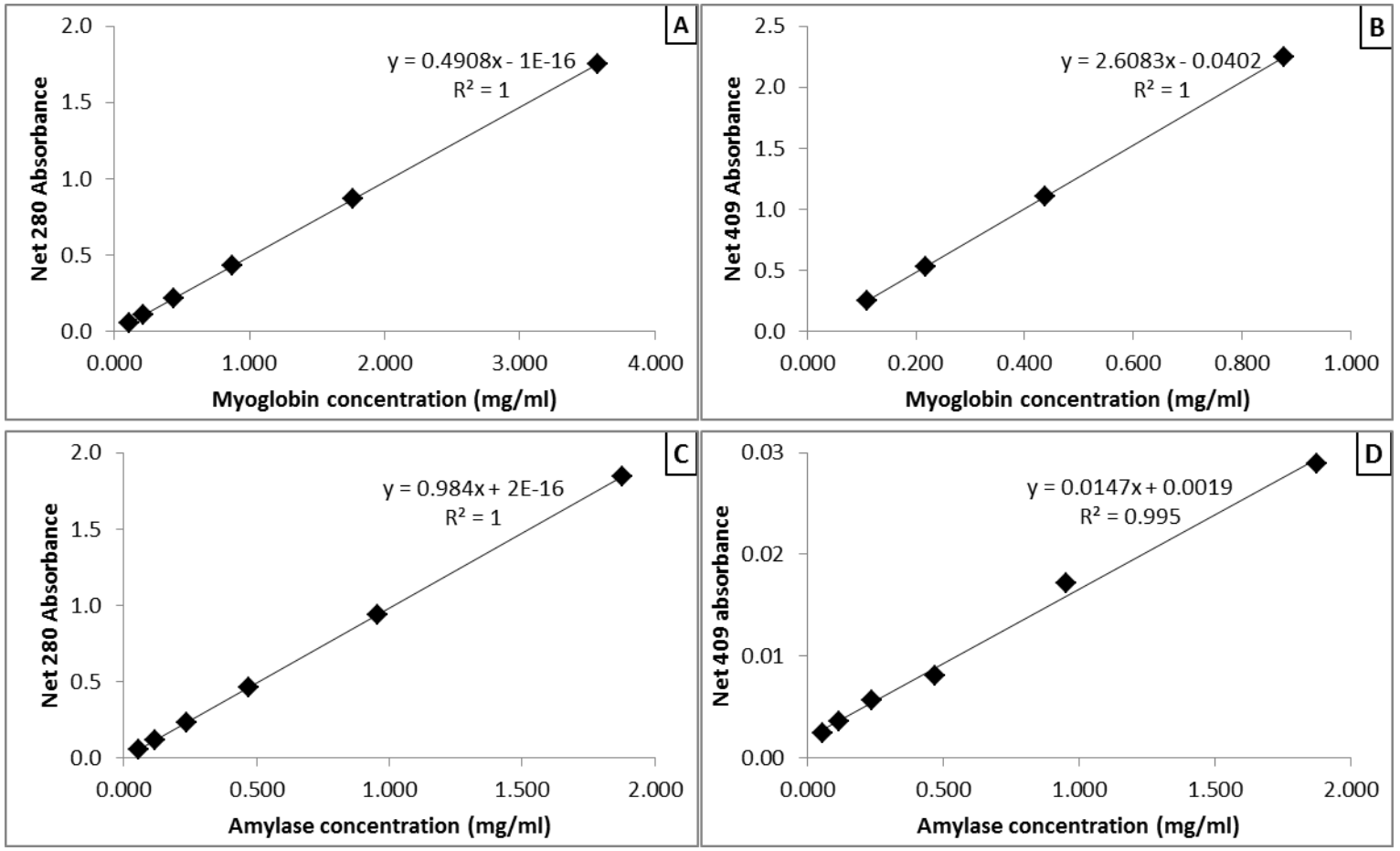


Figure 7 - 2: Absorbance standard curves for myoglobin and α-amylase at 280nm and 409nm.

7.2 Density correlation

The following is a parity plot of experimentally calculated density (see section 4.1.3) vs. predicted density using the empirical Eq.(4.3). The equilibria compositions are determined from the tie-lines fitted to the binodial curve generated using a targeted screen.

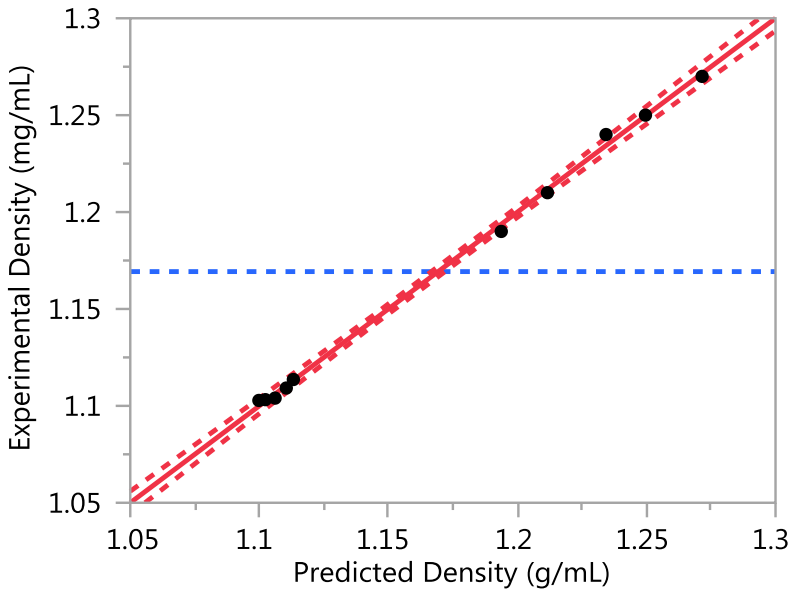


Figure 7 - 3: Parity plot of experimental density vs. density predicted using Eq.(4.3).

The correlation was obtained using the JMP Pro (SAS Institute, 2017) software which fitted a linear least-squares model with PEG and phosphate equilibria composition as effects to a response of experimental density. Parameter details can be found below in Table 7 - 1.

Table 7 - 1: Parameters used in Eq.(4.3) in units of g/mL with corresponding standard error determined in JMP Pro (SAS Institute, 2017).

Correlation parameter	Value	Standard error
Intercept	1.018	0.009
PEG coefficient	0.231	0.024
Phosphate coefficient	0.955	0.038

7.3 Liquid class calibration curves

Liquid classes are settings within the automated liquid handling platform which govern how much material is pipetted. Many factors can be changed such as air gaps, dispense and aspirate speeds etc. Two sample liquid-class calibration curves are presented below volumes above 200µL and below 200µL. Liquid class calibration curves are determined by weighing the amount of liquid that was pipetted and comparing it to the volume of diluter movement based on the Tecan calibration settings.

Table 7 - 2 shows the liquid class offset and intercept values for each solution pipetted using the Tecan Evo 150 automated liquid handling platform. The gradient and intercept values can be used to make a plot of diluter[10] movement vs. volume dispensed such as in Figure 7 - 4 and Figure 7 - 5. Diluter movement is found to be linear with volume dispensed.

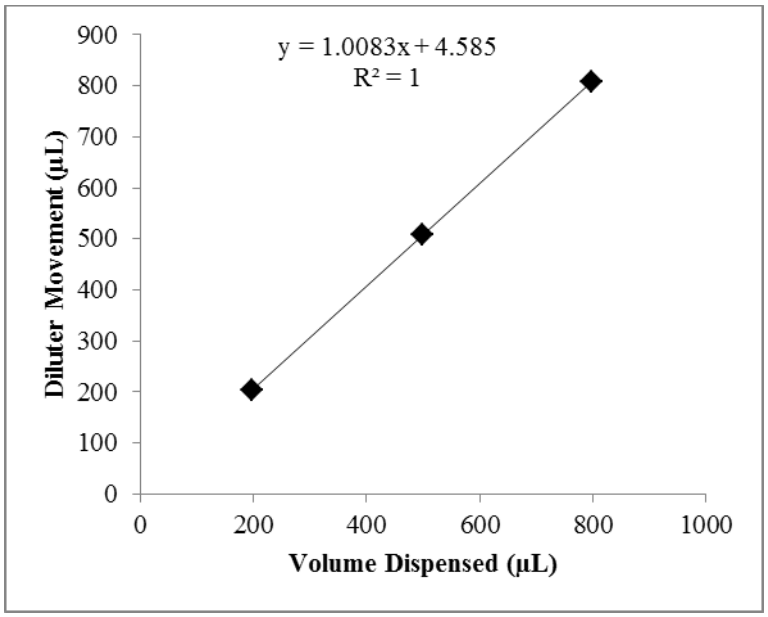


Figure 7 - 4: 50wt% PEG 2,000 calibration curve between 800µL and 200µL

[10] The diluter is the piston which causes liquid to either be dispensed or aspirated. The amount it moves is directly proportional to the amount of material pipetted.

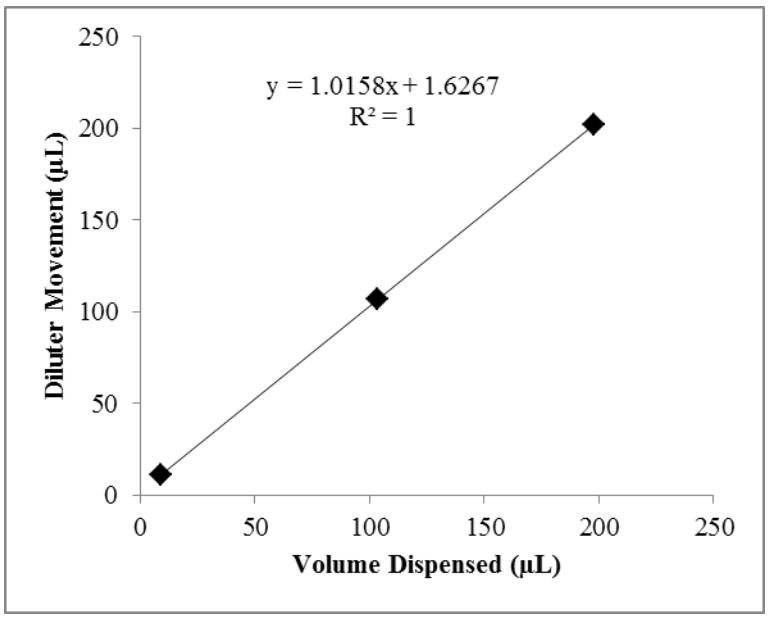


Figure 7 - 5: 50wt% PEG 2,000 calibration curve between 200µL and 10µL

Table 7 - 2: Liquid class diluter movement settings. Dilute movement = gradient x desired volume + intercept.

Liquid solution	Liquid class bracket (µL)	Density (g ml^{-1})	Gradient	Intercept	R^2
50wt% PEG 4,000	800-200.01	1.102	1.0243	0.5147	1
50wt% PEG 4,000	200.01-10	1.102	1.0334	1.9899	1
50wt% PEG 2000	800-200.01	1.082	1.0083	4.545	1
50wt% PEG 2000	200.01-10	1.082	1.0158	1.6267	1
40wt% Phosphate pH 7.00	800-200.01	1.378	1.0123	6.3947	1
40wt% Phosphate pH 7.00	200.01-10	1.378	1.0089	0.5593	1
25wt% NaCl	800-200.01	1.182	1.0094	3.9564	1
25wt% NaCl	200.01-10	1.182	1.0229	0.6026	1
Water*	1000.01-200.01	1	1.031	6.12	1
Water*	200.01-15.01	1	1.034	0.9	1
Water*	10.03-3	1	1.042	1.1	1

*** The default liquid class within the Tecan Evoware software was used for all buffer and water dispensing.**

7.4 Simulation solvers

In this work simulations were conducted to investigate dynamic aqueous two-phase extraction processes involving continuous counter-current and multi-cycle batch extraction modes of operation. Simulations in gPROMS were conducted using the DASOLV (differential algebraic solver) solver. The DASOLV solver is the default solver for solving sets of differential algebraic equations (DAEs) in gPROMS ModelBuilder 4.1.0 (Process Systems Enterprise, 2017). The solver can call upon three additional types of solvers for initialisation, linear-algebra and re-initialisation problems encountered during simulations. The DASOLV solver uses a Backward Differentiation Formulae (BDF) algorithm to solve DAEs. BDFs are useful for "stiff" differential equations and DAEs (Gear, 2007). BDF methods work by approximating the solution of the next time step by interpolating from previously calculated time steps.

7.5 Parameter estimation solvers

The parameter estimation solver used in this work from gPROMS is the nonlinear programming sequential quadratic programming (NLPSQP) solver. The systems considered are non-linear as some of the constraints and inequalities are non-linear in nature, for instance, the thermodynamic equations used to describe phase chemical potential are nonlinear (Eq.(3.21)-(3.23)). The sequential quadratic programming part corresponds to an interactive optimisation technique used to minimise the maximum likelihood objective function. Unlike simulated annealing, SQP is a deterministic gradient based algorithm which requires the objective function and corresponding constraints to be twice differentiable. A detailed explanation of the algorithm is described by Gockenbach (2003) and Philip and Elizabeth (2010).

7.6 MATLAB scripts

7.6.1 Tie-line determination

The following is an example of the code that was used to determine tie-lines based on mass ratio. The code adjusts the gradient of the tie-line such that the equality between the mass ratio and the inverse lever arm rule is satisfied. This code is used in steps 4 and 5 in the overall methodology presented in Figure 4 - 1.

Main tie-line determining script

```
%% This script determines tie-line compositions based on experimental
% experimental mass ratio. I need to supply the overall system data as
% columns of PEG, phosphate and mass ratio in that order also.

%% Loading experimental data of systems
load('massRatioData115124.mat') % The system data array needs to be
PEG, Phosphate and mass ratio

%% Which system to determine tie-line data for
noOfTielines = 6 ;

%% Assigning overall composition values
systemData = massRatioData115124 ;

% This is the overall system composition
PEGOverall = systemData( : , 1 );
PhosphateOverall = systemData( : , 2 ) ;

% This is the experiental mass ratio
ExpMassRatio = systemData( : , 3 ) ;

%% Setting up gradient and parameter data for the loops

% This is the guess for the gradient for fzero solver -- if the code
doesnt
% solve try changing the guess of the gradient.
GradientGuess = -1.6 ;

% Equilibrium composition and gradient array. First two columns are
peg and
% phosphate in top. Next two columns are peg and phosphate in bottom.
Last
% column is gradient of tieline.
GradientData = zeros( noOfTielines , 1 ) ;

% Assigning binodial curve parameters for simplicity
A = 48.5905 ;
B = -0.4068 ;
C = 0.0006 ;

%% Loop to find the gradient that corresponds to mass ratio

parfor j = 1:noOfTielines
    % Finding the gradient that matches mass ratio and assigning it to
data
    % array
    try
```

```
    GradientData( j ) = fzero( @(grad) PhaseRatioError( A , B , C ,
PEGOverall( j ) , PhosphateOverall( j ) , ExpMassRatio( j ), grad) ,
GradientGuess ) ;
    catch ME
        disp('Solver not able to find solution');
    end

end


%% Determining equilibrium compositions based on gradients previously
determined

% Equilbrium data array
EquilbriumData = zeros( noOfTielines , 4 ) ;

parfor i = 1:noOfTielines
   try
    EquilbriumData( i , : )  = EquilibriumComps( A , B, C ,
PEGOverall( i ) , PhosphateOverall( i ), GradientData( i ) ) ;
    catch ME
        disp('Solver not able to find solution');
   end
end
```

Phase ratio error function – calculates the difference between the experimental phase ratio and ratio of line OB to TO.

This function calculates the difference between the experimentally determined phase ratio and that determined using the inverse lever arm rule. The difference is minimised by a MATLAB solver (fzero) by varying the value of the gradient of the tie-line.

```
function [ PhaseError ] = PhaseRatioError( A,B,C, PEGOverall,
PhosphateOverall, ExpRatio, TielineGradient )
%PhaseRatioError outputs the error of phase ratio
%   Detailed explanation goes here

PhosphateRange=linspace(0,60,200)' ;

Intercept = PEGOverall - TielineGradient * PhosphateOverall ;

% Define equation of tie-line
Tieline = @(Phosphate) TielineGradient .* Phosphate + Intercept;

% Calculate intercepts of tie-line and binodial
[ PhosEq , PEGEq ] = intersections( PhosphateRange ,
BinodialCurve(A,B,C, PhosphateRange) , PhosphateRange , Tieline (
PhosphateRange ) );

% Calculating ratios of top to mid-point (TM) and mid-point to bottom
(MB)

TM = sqrt( (PhosEq(1) - PhosphateOverall)^2 + (PEGEq(1) -
PEGOverall)^2 ) ;
MB = sqrt( (PhosEq(2) - PhosphateOverall)^2 + (PEGEq(2) -
PEGOverall)^2 ) ;

Ratio = MB / TM ;
```

```
PhaseError = Ratio - ExpRatio ;

end
```

Equilibrium composition function – calculates equilibrium compositions based on a specified tie-line gradient and overall system composition

This function takes binodial curve parameters, overall PEG and phosphate composition and tie-line gradient to work out the equilibrium compositions at the intersection of the binodial curve and the tie-lines.

```
function [ TielineCompositions ] = EquilibriumComps( A,B,C,
PEGOverall, PhosphateOverall, TielineGradient )
%PhaseRatioError outputs the error of phase ratio
%   Detailed explanation goes here

PhosphateRange=linspace(0,60,200)' ;

Intercept = PEGOverall - TielineGradient * PhosphateOverall ;

% Define equation of tie-line
Tieline = @(Phosphate) TielineGradient .* Phosphate + Intercept;

% Calculate intercepts of tie-line and binodial
[ PhosEq , PEGEq ] = intersections( PhosphateRange ,
BinodialCurve(A,B,C, PhosphateRange) , PhosphateRange , Tieline (
PhosphateRange ) );

% Calculating ratios of top to mid-point (TM) and mid-point to bottom
(MB)

TM = sqrt( (PhosEq(1) - PhosphateOverall)^2 + (PEGEq(1) -
PEGOverall)^2 ) ;
MB = sqrt( (PhosEq(2) - PhosphateOverall)^2 + (PEGEq(2) -
PEGOverall)^2 ) ;

Ratio = MB / TM ;

TielineCompositions=zeros(1,4) ;

TielineCompositions( 1 , 1) = PEGEq(1) ;
TielineCompositions( 1 , 2) = PhosEq(1) ;
TielineCompositions( 1 , 3) = PEGEq(2) ;
TielineCompositions( 1 , 4) = PhosEq(2) ;

end
```

Binodial curve function – a function that calculates polymer wt% given values of A,B,C and salt wt%

This is a simple function that is used by many scripts to calculate the PEG composition in the binodial curve as a function of phosphate and binodial curve parameters.

```
function [ BinodialPEG ] = BinodialCurve( A , B , C ,
PhosphateBinodial )
% BinodialCurve computes the PEG composition for a given phosphate
% composition
%   Define binodial curve equation

    BinodialPEG = A .* exp( B .* PhosphateBinodial .^0.5 - C .*
PhosphateBinodial .^ 3) ;

end
```

7.7 Simulated annealing optimisation

A simulated annealing optimisation algorithm was used to generate initial guesses for interaction parameter estimates in step 7 of the overall methodology presented in Figure 4 - 1. This section explains what a simulated annealing algorithm is and how it works.

Generally speaking optimisation techniques can be categorised into deterministic or heuristic/stochastic methods. Deterministic optimisation typically uses gradients to converge to an optimal solution as well as other information specific to the related problem being solved whilst stochastic optimisation relies on the use of randomly generated variables. Typically deterministic optimisation methods are faster than stochastic methods as they converge to a solution faster, however, they can often get stuck in local optima (Cavazzuti, 2013). The use of stochastic optimisation techniques for the estimation of parameters in thermodynamic models has been reported by Ferrari et al. (2009) for liquid-liquid extraction. Stochastic optimisation techniques have been used for the optimisation of distillation (Cardoso et al., 2000), heat exchanger and water network design (Jeżowski et al., 2007) as well as many other process engineering problems.

Simulated annealing is a heuristic/probabilistic optimisation technique which is inspired by physical annealing processes used in metallurgy. Annealing is a physical process which is used to make metals easier to work with by making them softer. The metal is heated to a temperature where the metal molecules are allowed to re-crystallise which allows for any defects to be removed. After heating, the metal is cooled down very slowly to ensure a "refined microstructure". The cooled metal is softer resulting in a metal which is easier to work with (Metal Supermarkets, 2016).

To explain how a simulated annealing algorithm works, first it is useful to describe how gradient/descent algorithms work. Gradient/descent algorithms aim to reduce an objective function value by following the direction of steepest descent. When such algorithms arrive at a minimum, whether it is a local or global, they are stuck and algorithms terminate. One way to increase the likelihood of fining the global optimum using gradient based algorithms is by running the algorithms using multiple starting points. In simulated annealing, the algorithm tries to avoid becoming stuck in local optima by sometimes accepting solutions which increase the value of the objective function. The probability of accepting a solution which increases the value of the objective function is related to the annealing temperature (T) as shown in Eq.(7.1)

which is the default acceptance function used in MATLAB 2015b (MathsWorks®, 2016b):

$$Probability\ of\ acceptance = \frac{1}{1 + \exp\left(\frac{Increase\ in\ objective\ function}{T}\right)} \quad (7.1)$$

Eq.(7.1) shows that if the annealing temperature, T, is high the probability of accepting a solution which increases the objective function value increases. The probability of accepting a solution using Eq.(7.1) is between 0 and 0.5. In simulated annealing algorithms the annealing temperature is initially set high which results in many solutions being explored, however, as the algorithm proceeds the temperature is reduced slowly which reduces the likelihood of accepting solutions that increase the value of the objective function. Note that any solutions which lower the value of an objective function are always accepted. A good overview of the simulated annealing algorithm and its history is provided by Eglese (1990) as well as in the official MATLAB documentation "How Simulated Annealing Works" (MathsWorks®, 2016a). In this work, the simulated annealing algorithm within the Global Optimisation Toolbox in MATLAB 2015b was used. Table 7 - 3 shows the algorithm settings used.

Table 7 - 3: Settings for simulated annealing algorithm used in MATLAB 2015b.

Simulated annealing algorithm settings in MATLAB	
Stopping Criteria	
Maximum iterations	Infinite
Maximum function	9,000
Time limit	Infinite
Function tolerance	1×10^{-6}
Objective function limit	-Infinity
Stall iterations	1,500
Annealing Parameters	
Annealing function	Fast annealing
Re-annealing interval	50
Temperature update function	Exponential temperature update
Initial temperature	100
Acceptance criteria	
Acceptance probability function	Simulated annealing acceptance

7.8 Initialisation protocols

Due to the nonlinear nature of Eq.(3.21)-(3.23) which calculated chemical potential for description of phase equilibria, a robust initialisation procedure had to be implemented in gPROMS as very good initial guesses of system compositions in the top and bottom phases were required to solve all equations simultaneously. To best describe the initialisation protocol used, consider two states A and B corresponding to two different overall compositions within the two-phase region for a given aqueous two-phase system. The equilibrium compositions for the system at state A is well known while the equilibrium composition for B is unknown. To use gPROMS to calculate the equilibrium compositions at state B required initially starting the model at state A using the gPROMS "START" command within the initialisation procedure facility. After the model initialised at state A, the "NEXT" command was used to transition the numerical solution to state B using the "MOVE_TO" command. Another option is to use the "JUMP_TO" command, however, this is less robust as it does not gradually move the system of equations and is therefore more likely to fail. Once the model reaches state B, the model is deemed to have been successfully initialised and therefore the originally defined calculations/simulations will begin. An illustration of this is shown in Figure 7 - 6.

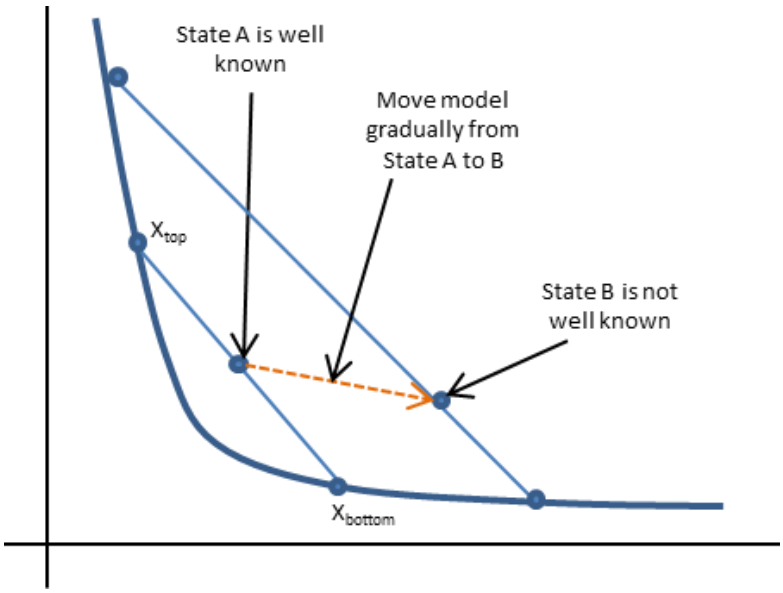


Figure 7 - 6: Illustration of how an initialisation protocol was implemented in gPROMS for simulation of aqueous two-phase extraction.

7.9 Derivation of Eq.(4.25) to quantify a combined variance by Burke (2014)

Eq.(4.25) was used to combine uncertainty in equilibrium compositions due to the binodial curve and due to density into one single representative value. A detailed description by Burke (2014) is provided below.

Combined Variance of Two Groups with Equal Numbers of Observations

Sara E. Burke

Updated January 12, 2014

Special Thanks to Nolan Esser

If we have two groups of n observations each, and we know the mean and sample variance of each group separately, how can we calculate the mean and sample variance of the combination of the two groups?

Let's refer to the groups as X and Y. The elements in the first group are $x_1 \dots x_n$ and the elements in the second group are $y_1 \dots y_n$. We are given four values:

$$\begin{aligned} \text{Mean of } X &= \bar{x} \\ \text{Variance of } X &= V_x \\ \text{Mean of } Y &= \bar{y} \\ \text{Variance of } Y &= V_y \end{aligned}$$

And we are expected to calculate two values:

$$\begin{aligned} \text{Overall Mean} &= M \\ \text{Overall Variance} &= V \end{aligned}$$

The overall mean is easy to calculate. It is the sum of all the observations divided by the total number of observations $2n$. Because each group has equally many observations, the overall mean is just the midpoint of the two means.

$$M = \frac{\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} y_i}{2n} = \frac{n\bar{x} + n\bar{y}}{2n} = \frac{\bar{x} + \bar{y}}{2}$$

The overall variance is not so simple.

The variance of X can be written as follows. As shown, from now on all sums across indices i from 1 to n will be written using only the $\sum$ symbol.

$$V_x = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1} = \frac{\sum (x_i - \bar{x})^2}{n-1}$$

The overall variance follows the same structure, but we have to sum across all of the X observations and all of the Y observations.

$$V = \frac{\sum(x_i - M)^2 + \sum(y_i - M)^2}{2n - 1}$$

Let's take the X portion of the numerator and expand it as shown below.

$$\begin{aligned}
\sum(x_i - M)^2 &= \sum(x_i - \bar{x} + \bar{x} - M)^2 \\
&= \sum((x_i - \bar{x}) + (\bar{x} - M))^2 \\
&= \sum((x_i - \bar{x})^2 + 2(x_i - \bar{x})(\bar{x} - M) + (\bar{x} - M)^2) \\
&= \sum(x_i - \bar{x})^2 + 2\sum(x_i\bar{x} - \bar{x}^2 - x_iM + \bar{x}M) + \sum(\bar{x} - M)^2 \\
&= \sum(x_i - \bar{x})^2 + 2\bar{x}\sum x_i - 2n\bar{x}^2 - 2M\sum x_i + 2n\bar{x}M + n(\bar{x} - M)^2
\end{aligned}$$

Note that $\bar{x} = \dfrac{\sum x_i}{n}$ so $\sum x_i = n\bar{x}$. Continuing from above,

$$\begin{aligned}
\sum(x_i - M)^2 &= \sum(x_i - \bar{x})^2 + 2n\bar{x}^2 - 2n\bar{x}^2 - (\bar{x} + \bar{y})n\bar{x} + (\bar{x} + \bar{y})n\bar{x} + n(\bar{x}^2 - 2\bar{x}M + M^2) \\
&= \sum(x_i - \bar{x})^2 + n\bar{x}^2 - 2n\bar{x}M + nM^2 \\
&= \sum(x_i - \bar{x})^2 + n\bar{x}^2 - 2n\bar{x}\left(\frac{\bar{x} + \bar{y}}{2}\right) + n\left(\frac{\bar{x} + \bar{y}}{2}\right)^2 \\
&= \sum(x_i - \bar{x})^2 + n\bar{x}^2 - (n\bar{x}^2 + n\bar{x}\bar{y}) + n\left(\frac{\bar{x}^2 + 2\bar{x}\bar{y} + \bar{y}^2}{4}\right) \qquad \Big) \\
&= \sum(x_i - \bar{x})^2 + n\bar{x}^2 - n\bar{x}^2 - n\bar{x}\bar{y} + \frac{n}{4}\bar{x}^2 + \frac{n}{2}\bar{x}\bar{y} + \frac{n}{4}\bar{y}^2 \\
&= \sum(x_i - \bar{x})^2 + \frac{n}{4}\bar{x}^2 - \frac{n}{2}\bar{x}\bar{y} + \frac{n}{4}\bar{y}^2 \\
&= \sum(x_i - \bar{x})^2 + \frac{n}{4}(\bar{x}^2 - 2\bar{x}\bar{y} + \bar{y}^2) \\
&= \sum(x_i - \bar{x})^2 + \frac{n}{4}(\bar{x} - \bar{y})^2
\end{aligned}$$

The $\sum(y_i - M)^2$ portion of the numerator of V can be similarly expanded, so V can be expanded as follows.

$$\begin{aligned} V &= \frac{\sum(x_i - M)^2 + \sum(y_i - M)^2}{2n-1} \\ &= \frac{\sum(x_i - \bar{x})^2 + \frac{n}{4}(\bar{x} - \bar{y})^2 + \sum(y_i - \bar{y})^2 + \frac{n}{4}(\bar{x} - \bar{y})^2}{2n-1} \\ &= \frac{\sum(x_i - \bar{x})^2 + \sum(y_i - \bar{y})^2 + \frac{n}{2}(\bar{x} - \bar{y})^2}{2n-1} \end{aligned}$$

Note that $V_x = \frac{\sum(x_i - \bar{x})^2}{n-1}$ so $\sum(x_i - \bar{x})^2 = (n-1)V_x$. The same applies to Y.

$$V = \frac{(n-1)V_x + (n-1)V_y + \frac{n}{2}(\bar{x} - \bar{y})^2}{2n-1}$$

$$V = \frac{(n-1)(V_x + V_y) + \frac{n}{2}(\bar{x} - \bar{y})^2}{2n-1}$$

This is the simplest form.

Note that, as the sample sizes increases, the variance of the combined sample approaches the mean of the two variances plus the square of half the distance between the two means.

$$\begin{aligned} \lim_{n\to\infty}(V) &= \lim_{n\to\infty}\left(\frac{(n-1)(V_x + V_y)}{2n-1} + \frac{\frac{n}{2}(\bar{x} - \bar{y})^2}{2n-1}\right) \qquad \Big) \\ &= \frac{V_x + V_y}{2} + \frac{(\bar{x} - \bar{y})^2}{4} \\ &= \frac{V_x + V_y}{2} + \left(\frac{\bar{x} - \bar{y}}{2}\right)^2 \end{aligned}$$

7.10 Research implementation and impact

This section contains work that was conducted during the EngD during a course on bioprocess validation. This work is included to fulfil requirements for an EngD thesis.

7.10.1 Describe briefly objectives of your research

The majority of current downstream separation processes used in the manufacturing of therapeutic proteins relies heavily on chromatography. Chromatography is extremely effective; however, it is also very expensive. Downstream separations are said to account for up to 70% of total manufacturing costs (Azevedo et al., 2009). This has been exasperated by the increasing quantity of material from upstream processes that must be handled as current separation techniques must be scaled in size in order to cope.

My research focuses on using aqueous two-phase extraction as an alternative separation technology. The challenges associated with this technology include:

1. A large experimental space for aqueous two-phase systems.
2. Resource intensive process development for aqueous two-phase systems.
3. Lack of understanding in the development of aqueous two-phase systems.

These challenges can be addressed by providing a clear procedure for the development of aqueous two-phase systems. My research aims to demonstrate this using a combination of high-throughput experimentation and process modelling.

7.10.2 Describe the significance of your research; especially its (potential) contribution to the healthcare sector

The need for a clear method to develop and understand a given separation process is extremely important, especially when time to market is so important. My research aims to provide this for aqueous two-phase extraction by combining modern high-throughput experimentation with first principles mathematical models.

The experimentation results are inputs for the process models. The models are then used to quickly explore the feasibility of the separation technique. This combined approach requires fewer experiments to be conducted (than a completely experimental approach) to gain an appropriate understanding of the feasibility of a given aqueous two-phase extraction process.

The significance of this combined experimental and modelling approach to the healthcare sector is that potentially expensive chromatography separation steps could be replaced with cheaper alternatives, such as aqueous two-phase extraction, which would

result in a lower cost of biopharmaceutical production which, ideally, would be translated to patients. In addition, pharmaceutical companies would have another tool in their separations toolbox.

7.10.3 Summarise the key validation issues faced by your sponsoring company

Biomarin Pharmaceuticals faces many of the challenges encountered by the majority of companies that manufacture therapeutic proteins. Biomarin is unique in that it produces enzyme replacement therapies for orphan diseases. Orphan diseases are rare diseases which affect a small percentage of the global population. Although these proteins are different to monoclonal antibodies, similar technologies such as chromatography are still used for purification. The validation issues associated with a hypothetical chromatography purification step will be discussed where an already-in-place process must be updated due to the discontinuation of resin X. The manufacturer of resin X insists that the performance of its newer resin Y is comparable, however, this is a major validation issue that must be addressed. Biomarin would need to ensure that:

1. *The product and impurities interact the resin Y in exactly the same way as resin X*

   This can be assessed by performing scale-down studies with the new resin using high-throughput experimentation and lab-scale chromatography runs. Isotherms parameters between the two-resins can be compared. A separate issue associated with scaled down and high-throughput experimentation is that there needs to be proof that these techniques are representative of the large scale. In this situation, the data should already exist for the initial process development using resin X.

2. *The resin Y chromatography step does not influence processes downstream of it due to factors such as leachable material*

   This can be assessed by conducting a full study where the output of one unit is the input of the next. The experimental information for the setup with resin X already exists therefore only experiments with the new resin Y must be conducted. Leachable material such as polymers and ligands must be quantified.

3. *The number of reusable cycles must be re-evaluated*

   This is important because the initial performance of the resins may be the same, however, over time the performance may drift. This could be caused by different degrees of resin fouling between resin X and resin Y therefore resin degradation must be assessed by conducting appropriate resin life cycle studies.

4. *Structural integrity of the resin must be evaluated*

The resin structure could be assessed by obtaining appropriate documentation and proof from the manufacturer. If this information is not available then in-house investigations must be conducted. Key variables to investigate would be pressure-drop, resin compressibility, resin porosity, resin binding capacity, resin leaching and batch to batch variability.

5. *Resin Y does not cause modifications to the product*

Experiments need to be conducted to ensure that the same product is being produced when switching to the resin Y. The product could be modified due to the presence of different chemical groups on resin Y. Resin Y could also not remove proteases and other host cell proteins to the same degree as resin X which would degrade the quality of the product.

6. *The design space is the same*

The design space for a chromatography step using the new resin Y must also be compared to resin X as if the design space is completely different, then the process is essentially being changed to a much greater degree than desired. More change results in more validation work that must be conducted to ensure a safe and robust product which has quality built into it.

7.10.4 Discuss the regulatory (validation) issues which may affect the take up of your research

The regulatory issues regarding this research have been broken up into two themes:

1. Validation issues associated with high-throughput experimentation
2. Validation issues associated with process modelling

High-throughput experimentation

The main issues associated with high-throughput experimentation is the scalability of data obtained using such systems. In the case of this research, 1g mass aqueous two-phase systems are formed in 2mL Eppendorf tubes. The issues regarding this scale of operation with their potential solutions are outlined below:

1. *The protein separation behaviour at the 1g is not the same as at manufacturing scale*

This can be addressed by performing scale-up studies where the concentration of products and contaminants at the different scales is compared.

2. *The equilibrium time at the 1g scale is not the same as at manufacturing scale*

This is an important issue since if equilibrium is not reached, the maximum separation capability cannot be utilised. The amount of mixing required to reach equilibrium at the 1g and manufacturing scales needs to be established.

3. *The phenomena seen at the 1g may not be the same at the manufacturing scale e.g. extent of precipitation*

Again, comparisons between the 1g and manufacturing scale must be made.

4. *How do we know the calibration of the automated liquid handling systems is correct?*

This can be addressed by having appropriate calibration protocols for the automated liquid handling which are logged and kept up to date.

5. *Material variability*

Components used in the formation of aqueous two-phase systems could potentially vary, for example, the polymer polyethylene glycol could have different polydispersity values between batches which may influence the phase characteristics. This has to be investigated to rule it out as a validation issue by performing experiments to investigate potential batch to batch variability.

Process modelling

There are many validation issues that must be address when using process models. An important question to ask is what the purpose is of the model. Will it be used to determine design space during manufacturing or a just purely a tool for process development? Depending on the use, the validation issues could be very different, for simplicity common issues are discussed below.

1. *How can you guarantee that the model is predictive?*

This can be addressed by validating the models predictive capability at experimental conditions not used to generate the model, e.g. if data set A and B from experiments A and B are used to fit the parameters of the model then they cannot be used for validation. A further experiment generating data C has to be performed for comparison.

2. *What is the required accuracy of the parameters?*

This is an important question to address because the complexity of your model will determine the number of parameters and the more parameters present, the more experimental data is required for fitting purposes which in turn reduces the confidence in those parameters. However, the question is, if you have a predictive model, do you need accurate, well defined parameters? This will depend on what the model is finally used for.

3. *How can you be sure that global optimum is found during optimisation studies?*

Firstly the model would need to be verified to ensure that it captures the complete range of experimental conditions that can feasibly be considered. The global

optimum can be verified using different optimisation techniques and starting optimisations from different starting conditions to ensure that a local optima is not found.

4. *Can validation runs be used to improve model predictions for future simulations? If so, can they be used for problem solving/process control? i.e. can the training set and validation set of experimental data be mixed and matched.*

This depends on the final intentional use of the model. It is best practice to not mix training and validation sets of experimental data. Efforts should be focused on understanding the differences between such data. If your model was used to decide operating conditions during manufacturing (i.e. determination of the operating space) then it should not be modified/improved using validation runs because changing the model would mean changing the "operating/design space".

7.11 Data from Mistry et al. (1996)

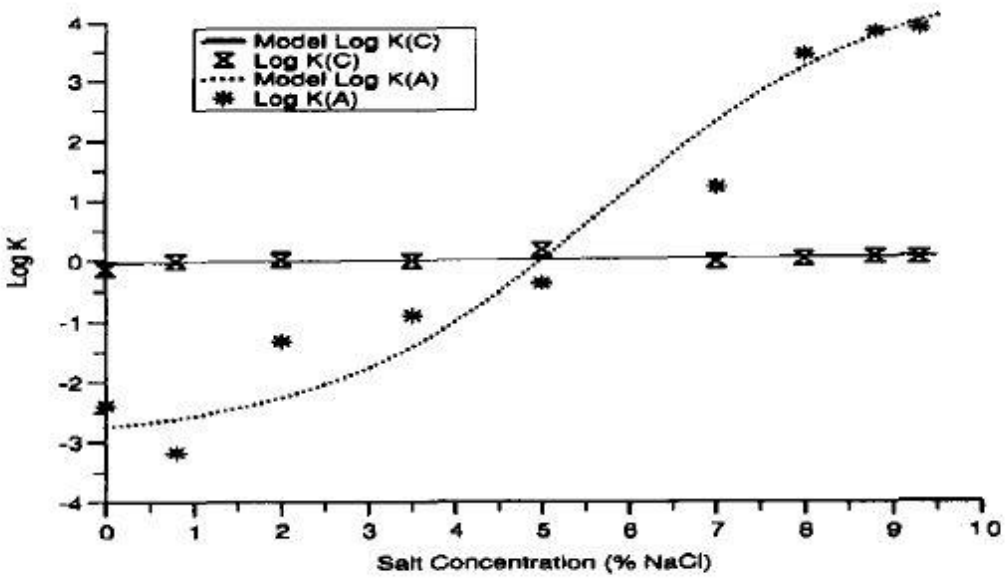


Fig. 5. Comparison of α-amylase (A) and contaminant (C) data and model fit.

Figure 7 - 7: Protein partition data from Mistry et al. (1996).

7.11.1 Partition coefficient correlation

The experimental partition coefficient data above was fitted to a linear and polynomial curve using the curve fitting tool within MATLAB.

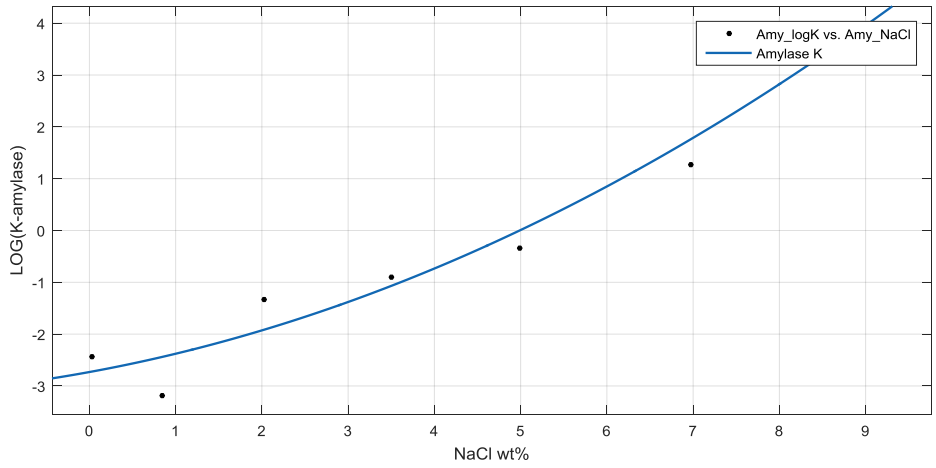


Figure 7 - 8: Experimental partition coefficient fitted to 2nd order polynomial in MATLAB. Dots represent experimental data points from Mistry et al. (1996) and blue line represent Eq.(3.16).

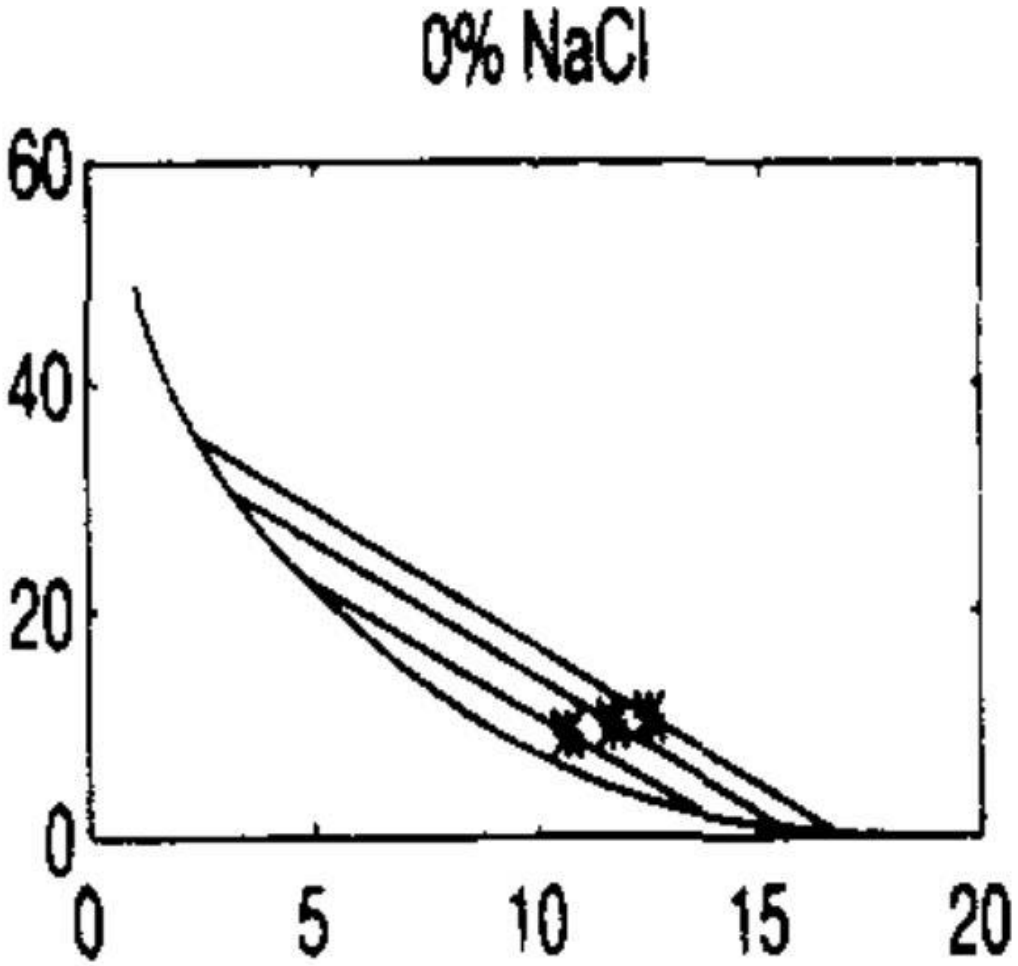


Figure 7 - 9: Experimental phase diagram from Mistry et al. (1996) for a PEG 4000-phosphate aqueous two-phase system with 0wt% NaCl.

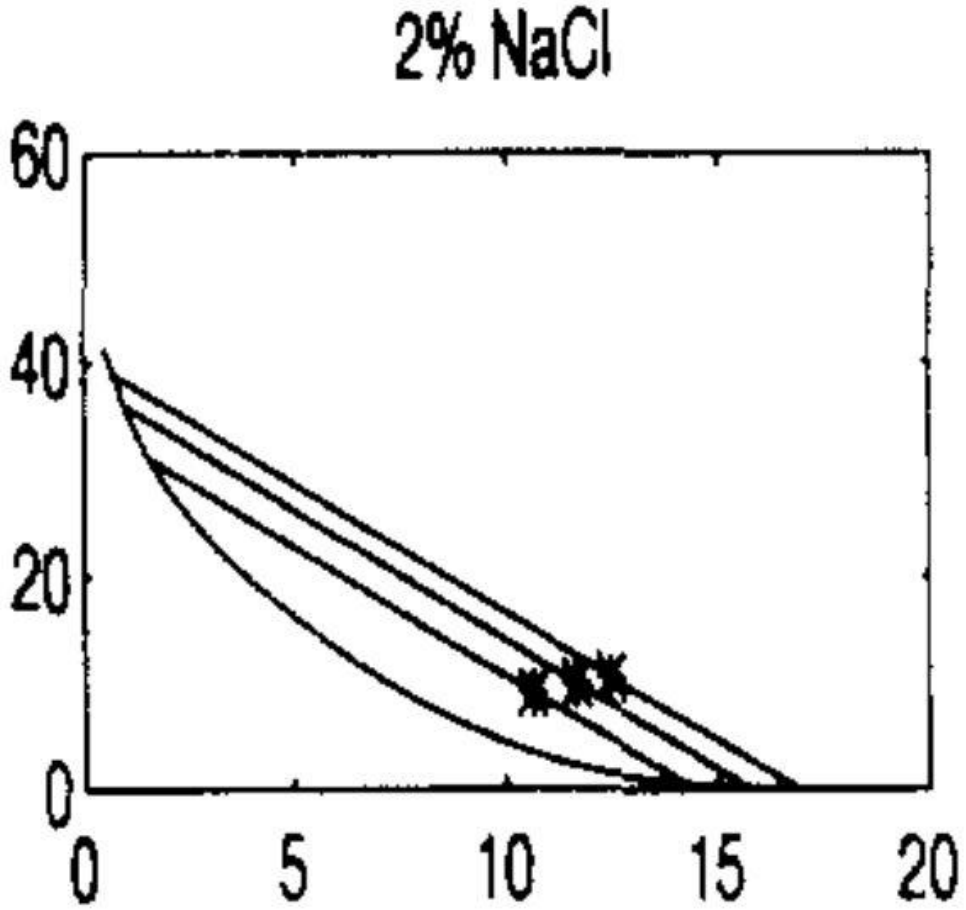


Figure 7 - 10: : Experimental phase diagram from Mistry et al. (1996) for a PEG 4000-phosphate aqueous two-phase system with 2wt% NaCl.

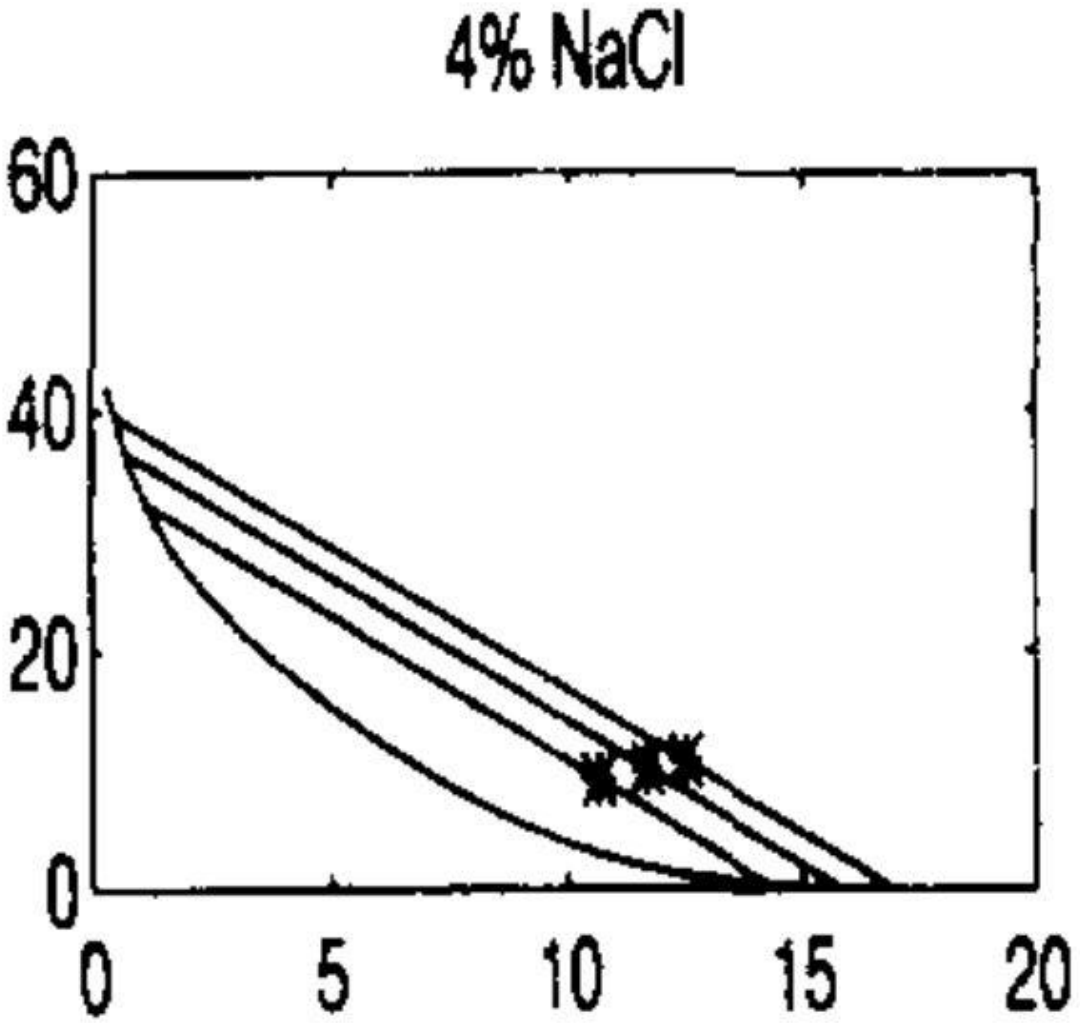


Figure 7 - 11: : Experimental phase diagram from Mistry et al. (1996) for a PEG 4000-phosphate aqueous two-phase system with 4wt% NaCl.

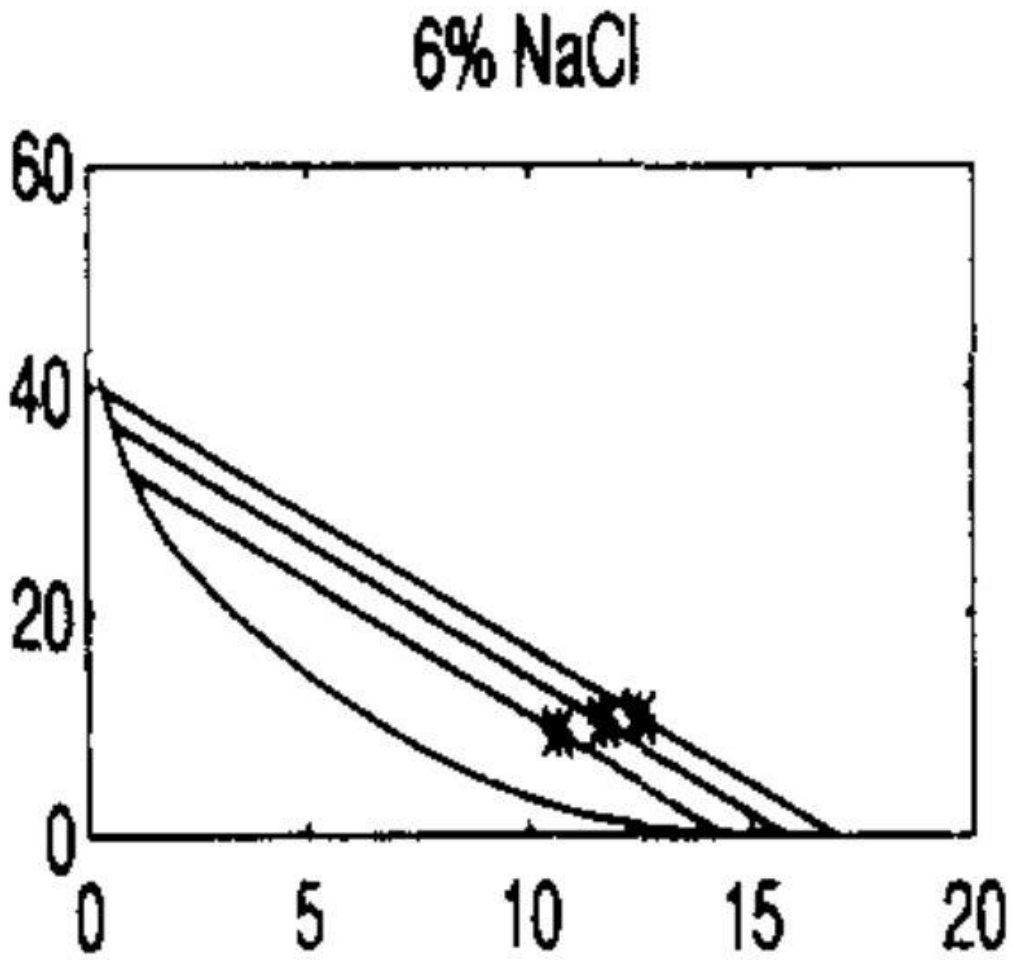


Figure 7 - 12: : Experimental phase diagram from Mistry et al. (1996) for a PEG 4000-phosphate aqueous two-phase system with 6wt% NaCl.

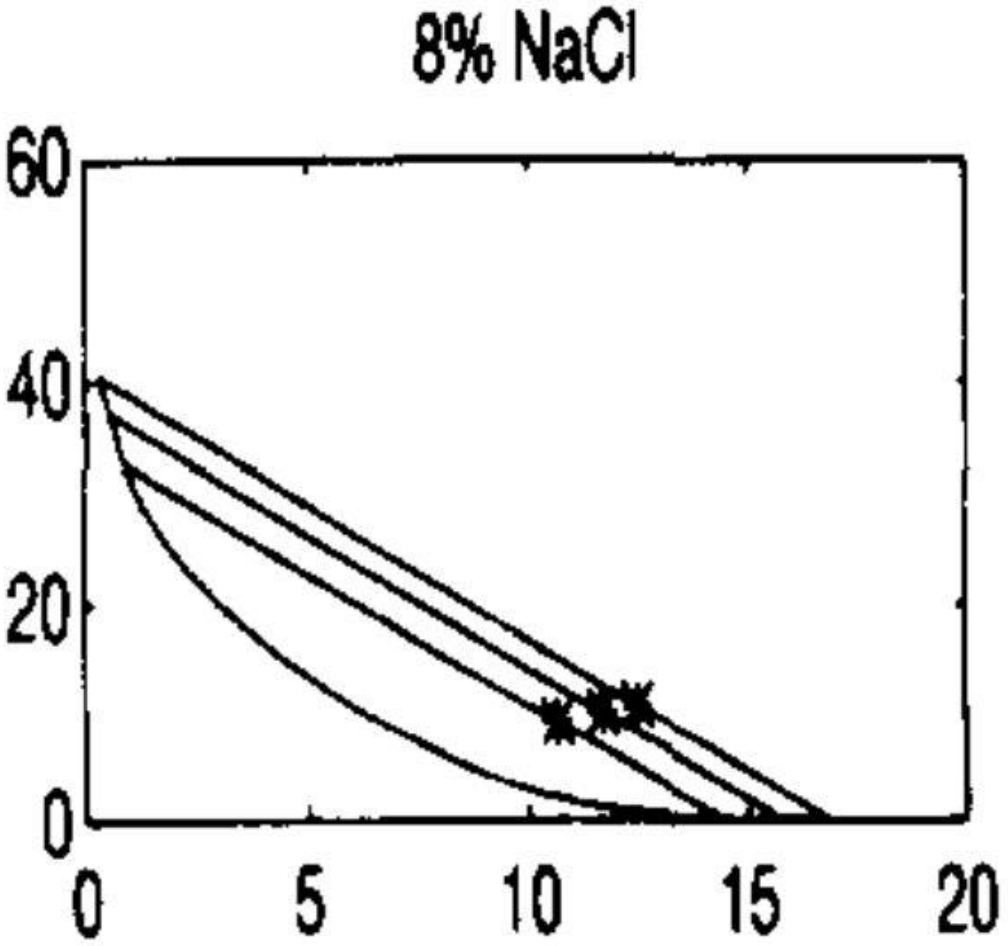


Figure 7 - 13: : Experimental phase diagram from Mistry et al. (1996) for a PEG 4000-phosphate aqueous two-phase system with 8wt% NaCl.

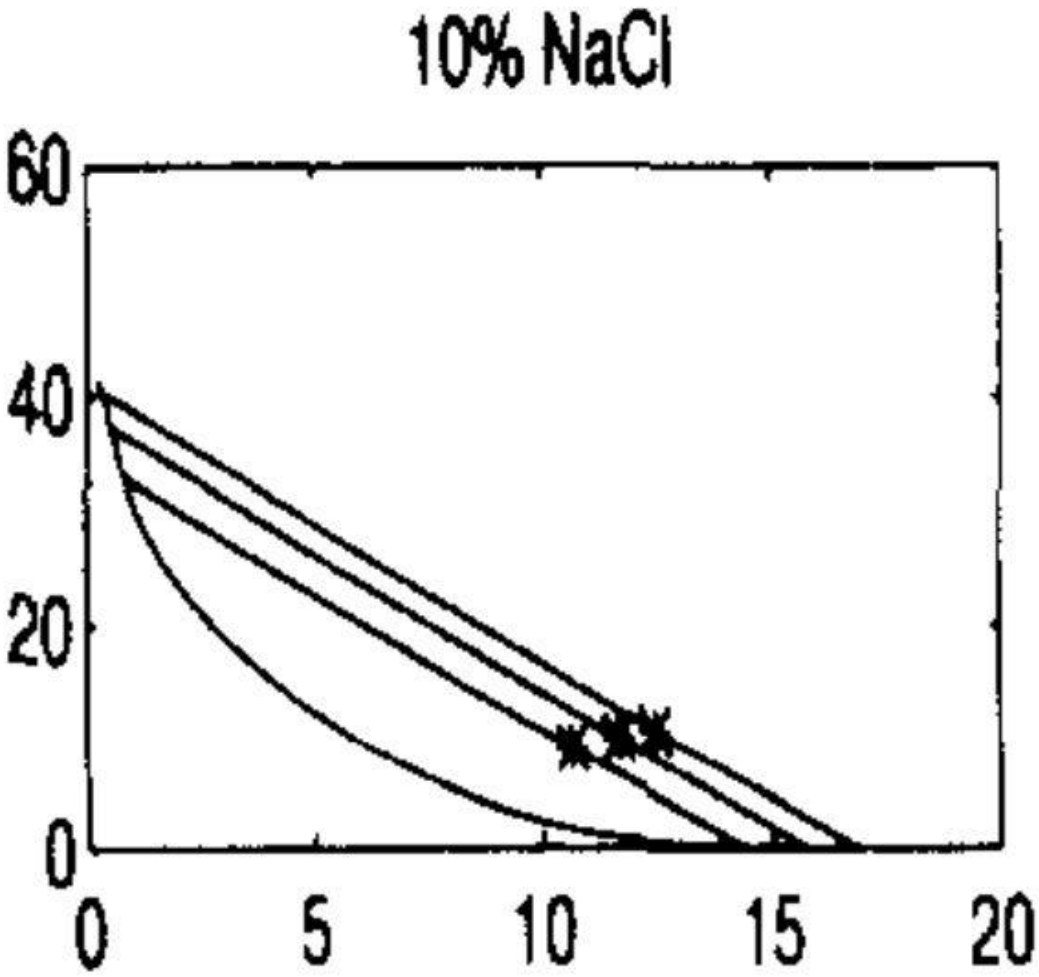


Figure 7 - 14: : Experimental phase diagram from Mistry et al. (1996) for a PEG 4000-phosphate aqueous two-phase system with 10wt% NaCl.

7.12 Parity plot from Samatou (2012) generated for empirical relationship relating mass fraction of salts and polymers to density

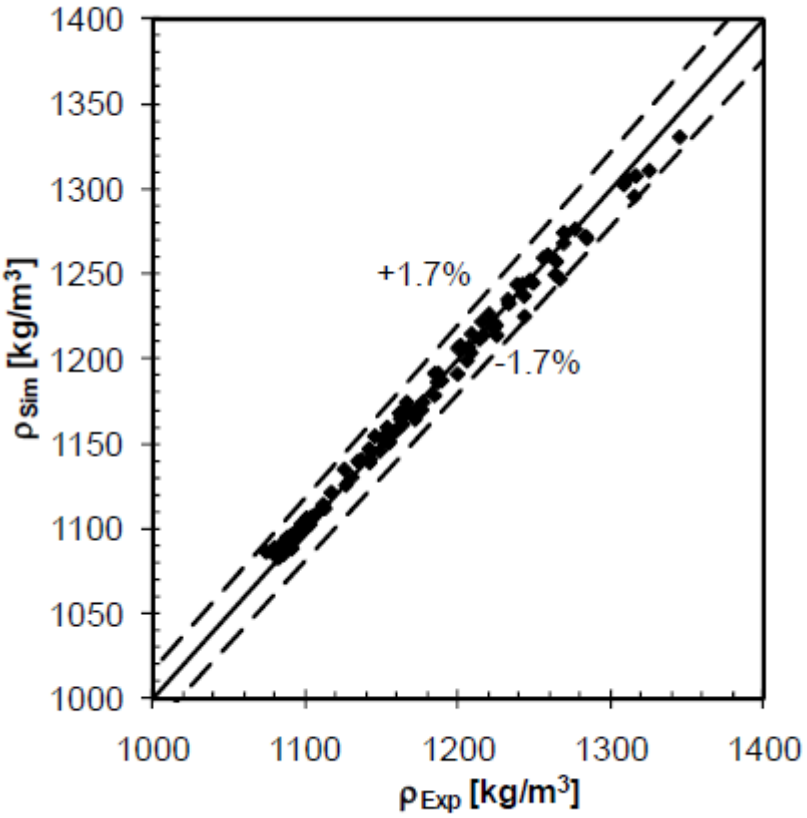


Figure 7 - 15: Parity plot obtained fromSamatou (2012) for density correlation (Eq.(3.20)).