

Separation of Recombinant  $\beta$ -Glucuronidase from Transgenic Tobacco by  
Aqueous Two-Phase Extraction

By

Kristin Coby Ross

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University  
in partial fulfillment of the requirements for the degree of

Master of Science  
In  
Biological Systems Engineering

Chenming Zhang  
Foster Agblevor  
Aaron Goldstein

19 May 2008  
Blacksburg, Virginia

Keywords: protein purification, transgenic tobacco, aqueous two-phase extraction,  
recombinant  $\beta$ -glucuronidase

Copyright 2008, Kristin Coby Ross

# **Separation of Recombinant $\beta$ -Glucuronidase from Transgenic Tobacco by Aqueous Two-Phase Extraction**

Kristin Coby Ross

Abstract

Biopharmaceutical manufacturing is a rigorous and expensive process. Due to the medicinal nature of the product, a high purity level is required and several expensive purification steps must be utilized. Cost-effective production and purification is essential for any biopharmaceutical product to be successful and development of the fastest, most economical, and highest-yielding purification scheme is a constant engineering challenge. Commercial-scale purification schemes currently revolve around the use of multiple chromatography steps for the purification of biopharmaceutical products. Chromatography has many shortcomings including high cost, limited throughput, and complex scale up. The goal of this research was to develop an alternative, non-chromatography purification step for the separation of an acidic model protein, recombinant  $\beta$ -glucuronidase (rGUS), from transgenic tobacco with high yield and purity.

Aqueous two-phase extraction (ATPE) is a powerful technique for separation and purification of proteins, and has the potential to replace an expensive chromatography step for the initial purification of recombinant proteins. ATPE enables high levels of target protein recovery and concentration while removing large amounts of impurities from the initial extract. Fractional factorial designs and response surface methodology were used to determine an optimized aqueous two-phase system for the purification of rGUS from transgenic tobacco. In a 13.4 % (w/w) PEG/18% (w/w) potassium phosphate system, 74% of the rGUS was recovered in the top PEG-rich phase while 90% of the native tobacco proteins were removed in the interphase and the bottom phase. A purification factor of about 20 was achieved in this process.

## ACKNOWLEDGEMENTS

I want to thank Dr. Chenming Zhang for his help and assistance throughout my graduate studies. You have given me great support throughout and I could not have accomplished all I have without your guidance. I would also like to thank Dr. Foster Agblevor and Dr. Aaron Goldstein for your willingness to serve on my committee. Your technical advice has been helpful over the past two years. I also want to thank Chris Holler, Brad Matanin, and Ray Lillie for patiently teaching me all the lab techniques I needed to complete this project.

Additionally, I want to thank Dr. Luis Fabricio-Medina Bolivar who kindly provided transgenic tobacco seeds containing the recombinant GUS gene and Dr. Carol Wilkinson for kindly providing *Nicotiana tabacum* cv. Xanthi seeds.

Finally, I want to thank my family and friends for their support and encouragement.

## CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iii
TABLE OF CONTENTS.....	iv
LIST OF FIGURES .....	vi
LIST OF TABLES.....	vii
CHAPTER 1: INTRODUCTION AND OBJECTIVES.....	1
CHAPTER 2: LITERATURE REVIEW.....	3
2.1 Production of Biopharmaceuticals.....	3
2.1.1 Biopharmaceutical Industry Overview .....	3
2.1.2 Current Problems and Limitations .....	3
2.1.2.1 Expression Systems .....	3
2.1.2.2 Purification Process .....	4
2.2 Recombinant Protein Production in Plants .....	5
2.2.1 Plant Expression Systems .....	5
2.2.2 Tobacco Expression System .....	7
2.3 Bioprocessing.....	8
2.3.1 Guidelines .....	8
2.3.2 Aqueous Two-Phase Extraction.....	10
2.3.2.1 Comparison of Aqueous Two-Phase Extraction and Chromatography .....	16
2.4 $\beta$ -Glucuronidase as a Model Protein.....	17
2.5 Summary .....	17
References.....	19
CHAPTER 3: EXPERIMENTAL.....	24
3.1 Materials .....	24
3.2 <i>Nicotiana tabacum</i> Plant Generation .....	25
3.3 Protein Extraction .....	25

3.4 Analytical Methods .....	26
3.4.1 Bio-Rad Assay .....	26
3.4.2 GUS Activity Assay.....	26
3.4.3 SDS-PAGE .....	27
3.5 Aqueous Two-Phase Extraction.....	27
3.5.1 Development of Aqueous Two-Phase System.....	27
3.5.2 Fractional Factorial Design.....	28
3.5.3 Response Surface Method Design. ....	31
References.....	34
CHAPTER 4: RESULTS AND DISCUSSION.....	36
4.1 Analysis of Screening Experiments .....	36
4.2 Response Surface Method Experiments .....	42
4.2.1 Interphase Studies .....	42
4.2.2 Response Surface Method Design Analysis .....	44
4.3. Optimum Aqueous Two-Phase System Determination .....	47
4.3.1 Confirmation of Response Surface Method Results .....	47
4.3.2 Phase Ratio Experiments .....	50
4.3.3 SDS-PAGE Results.....	53
References.....	55
CHAPTER 5: CONCLUSIONS .....	57
References.....	61
CHAPTER 6: FUTURE WORK .....	63
APPENDIX A: EXPERIMENTAL CONDITIONS.....	65
APPENDIX B: PRELIMINARY EXPERIMENTS .....	67

All images belong to the author, unless otherwise noted

## FIGURES

<b>Figure 2.1:</b> Phase diagram for polyethylene glycol-potassium phosphate system .....	12
<b>Figure 4.1:</b> Response surface plot for a PEG/potassium phosphate aqueous two-phase system. ....	47
<b>Figure 4.2:</b> Revised response surface plot for a PEG/potassium phosphate aqueous two-phase system. ....	50
<b>Figure 4.3:</b> Phase diagram for PEG/potassium phosphate at pH 8 with the published tie-lines (A, B, C, and D) and estimated tie-line (Est TL) .....	52
<b>Figure 4.4:</b> SDS-PAGE gel showing Rubisco protein removal from transgenic tobacco extract and rGUS recovery in the top phase during ATPE .....	54
<b>Figure 5.1:</b> General flow diagram for the recovery of PEG and salt from a two-stage ATPE.....	59

## TABLES

<b>Table 2.1:</b> Recombinant proteins purified using ATPE and the recovery and purification factors obtained for each.....	15
<b>Table 3.1:</b> Alias structure generated for a ½ fractional factorial design with five factors.....	30
<b>Table 3.2:</b> Factor combinations for PEG/Ammonium Sulfate, PEG/Potassium Phosphate, and PEG/Sodium Sulfate systems used to study partitioning of native tobacco proteins and GUS.....	31
<b>Table 3.3:</b> Factor combinations for PEG/Potassium Phosphate system used to study the partitioning of rGUS in transgenic tobacco extract .....	33
<b>Table 4.1:</b> Average partition coefficients (K) and standard deviations for native tobacco proteins in the PEG/ammonium sulfate system. ....	38
<b>Table 4.2:</b> Average partition coefficients (K) and standard deviations for GUS in the PEG/ammonium sulfate system. ....	38
<b>Table 4.3:</b> Average partition coefficients (K) and standard deviations for native tobacco proteins in the PEG/sodium sulfate system. ....	39
<b>Table 4.4:</b> Average partition coefficients (K) and standard deviations for GUS in the PEG/sodium sulfate system. ....	39
<b>Table 4.5:</b> Average partition coefficients (K) and standard deviations for native tobacco proteins in the PEG/potassium phosphate system. ....	40
<b>Table 4.6:</b> Average partition coefficients (K) and standard deviations for GUS in the PEG/potassium phosphate system. ....	40
<b>Table 4.7:</b> ANOVA table for two-level, ½ fractional factorial screening experiments for GUS in PEG/potassium phosphate aqueous two-phase system.....	41
<b>Table 4.8:</b> ANOVA table for two-level, ½ fractional factorial screening experiments for native tobacco proteins in PEG/potassium phosphate aqueous two-phase system.....	41
<b>Table 4.9:</b> Average selectivities ( $\alpha$ ) and standard deviations for rGUS in PEG/potassium phosphate system used in statistical analysis for RSM experiments .....	46
<b>Table 4.10:</b> ANOVA table for face-centered CCD model using PEG/potassium phosphate aqueous two-phase system.....	46

<b>Table 4.11:</b> Average selectivities ( $\alpha$ ) and standard deviations for rGUS in PEG/potassium phosphate system used in statistical analysis for RSM experiments .....	49
<b>Table 4.12:</b> ANOVA table for face-centered CCD model using PEG/potassium phosphate aqueous two-phase system.....	49
<b>Table 4.13:</b> rGUS recovery and purification factor at several different phase ratios .....	52
<b>Table A.1:</b> Conditions, mean enzymatic activity, standard deviation, and p values obtained during control experiments .....	66
<b>Table B.1:</b> Results from preliminary experiments performed to test the enzymatic activity of rGUS in an aqueous two-phase system at pH 6, pH 7, and pH 8 .....	67



# Chapter 1

## Introduction and Objectives

Biopharmaceutical manufacturing is an expensive and complicated process. Using a transgenic tobacco expression system has the potential to reduce biopharmaceutical production costs, but successful implementation depends upon the development of cost-effective purification schemes. Commercial-scale purification schemes currently use multiple chromatography steps for the purification of biopharmaceutical products. However, chromatography has many disadvantages, including high cost, limited throughput, and complex scale up. ATPE has the potential to replace chromatography for the initial purification of recombinant proteins from crude or partially purified extracts. The goal of this research was to develop an alternative, non-chromatography purification step for the separation of an acidic model protein, recombinant  $\beta$ -glucuronidase (rGUS), from transgenic tobacco with high yield and purity.

To accomplish this goal, the partitioning behavior of both GUS and native tobacco proteins will be studied using statistical analysis to determine which factors are significant in their partitioning. Once the significant factors are determined, further statistical analysis will be used to obtain an optimized ATPE conditions for partitioning of rGUS and transgenic tobacco. The phase ratio of the aqueous two-phase system will then be adjusted to determine the optimum recovery conditions for the separation of rGUS from transgenic tobacco.

This thesis contains six chapters. Chapter one contains a short introduction and the project objectives. Chapter two begins by reviewing the importance of the biopharmaceutical industry and the current problems and limitations in the production and purification of biopharmaceutical products. A comprehensive literature review follows, detailing ways to improve biopharmaceutical manufacturing through the use of plant expression systems and aqueous two-phase extraction. The advantages and disadvantages of plant expression systems for recombinant protein production are discussed, followed by general bioprocessing considerations and a comprehensive review of recombinant protein purification by aqueous two-phase extraction. Specific examples from the literature were briefly touched upon.

Chapters three and four explain the project objectives and experimental methods used during the course of the project. Chapter five focuses on the experimental work accomplished to meet the project objectives. The results obtained during development of an optimum aqueous two-phase extraction for the separation of recombinant  $\beta$ -glucuronidase from transgenic tobacco are presented. This chapter also compares the results obtained during this work to another purification procedure for purifying recombinant  $\beta$ -glucuronidase from transgenic tobacco.

Chapter six gives the conclusions for this work as well as a discussion concerning implementation of aqueous two-phase extraction into a commercial-scale biopharmaceutical purification scheme. Chapter six expands on the implementation of aqueous two-phase extraction by discussing future work to supplement the work completed during this project.

## **Chapter 2**

### **Literature Review**

#### **2.1 Production of Biopharmaceuticals**

##### *2.1.1 Biopharmaceutical Industry Overview*

Modern biotechnology is vital for the discovery and development of novel drugs to satisfy unmet medical needs. Due to rapid advances in recombinant DNA technology and genetic manipulation techniques, biopharmaceutical-based therapeutics have emerged as one of the most significant medical innovations in the past three decades. Biopharmaceuticals are complex macromolecules that cannot be directly synthesized or extracted from a biological host, and therefore must be manufactured through the use of biotechnology. Many of the biopharmaceuticals manufactured today are recombinant proteins, including monoclonal antibodies, enzymes, and hormones. Biopharmaceuticals have the potential to treat a wide range of diseases, including cancer, autoimmune diseases, and inflammatory diseases, and these treatments were previously thought to be impossible and impractical. To date, more than 150 proteins of human therapeutic value have received Food and Drug Administration (FDA) approval [1] and entered the market. At the same time, the number of biopharmaceutical companies and biological drug candidates is rapidly expanding. It is estimated that the biopharmaceutical market generates more than \$50 billion in sales annually [1] and this number will climb higher with the influx of research and development investments, acquisitions, and licensing deals. While the accomplishments of the biopharmaceutical industry are certainly impressive and inspiring, they come at an exceptionally high price. Drug discovery and development is a costly and complicated undertaking with more than 99% of the experimental compounds ultimately failing as treatment regimens [2].

##### *2.1.2 Current Problems and Limitations*

###### *2.1.2.1 Expression Systems*

The majority of the biopharmaceuticals manufactured today are produced from bacterial fermentation (*Escherichia coli*), yeast cell cultures (*Saccharomyces cerevisiae*, *Pichia pastoris*), or mammalian cell cultures. These expression systems are well-

established and well-characterized, yet still have some disadvantages inherent to each of the respective platforms. Due to the prokaryotic nature of bacteria, recombinant protein production through bacterial fermentation is limited to simple proteins where some posttranslational modifications are absent or unimportant. Biopharmaceutical products are typically complex proteins in which these modifications are essential for correct protein folding and function. Eukaryotic yeast cell cultures eliminate this particular disadvantage, but also add others. Recombinant protein production through yeast cell cultures can have a high amount of protein degradation and a low protein yield. Mammalian cell cultures are used to produce approximately 70% of manufactured biopharmaceuticals [1]. This eukaryotic expression system offers the highest yield of functional recombinant protein with correct posttranslational modifications. However, biopharmaceuticals produced through cell culture have to be produced in relatively small quantities due to the highly specialized culture conditions. Mammalian cells are susceptible to reduced productivity or death as a result of slight deviations in culture conditions [3]. Further limitations of mammalian cell culture expression systems include expensive infrastructure requirements, low scalability, and viral or oncogenic contamination issues [3]. This leads to high overall manufacturing costs, depending on the quantity of the biopharmaceutical produced.

#### *2.1.2.2 Purification Process*

Taking an FDA-approved biopharmaceutical product from the research laboratory to the patient is a long and arduous road. Due to the medicinal nature of the product, all biopharmaceuticals require a very high level of purity and manufacturing protocols are tightly regulated. These requirements make commercial-scale recombinant protein purification rigorous and expensive. Several expensive purification steps must be utilized to obtain the stringent purity requirements imposed by the FDA and the addition of multiple processing steps results in higher product losses. Scale of production can vary greatly depending on the biopharmaceutical produced and this also affects the cost of purification. For example, one dose of a therapeutic hormone is typically a few micrograms of protein, while one dose of a monoclonal antibody treatment may be a million-fold higher with doses of a gram or more being quite common [4]. Cost-effective

production and purification is essential for any biopharmaceutical product to be successful, regardless of whether or not it is an effective medicine. Cost is the major limiting factor in biopharmaceutical manufacturing and development of the fastest, most economical, and highest-yielding purification scheme is a constant engineering challenge.

A vast range of purification methods exist for the separation of recombinant proteins from the impurities inherent in their expression systems. Commercial-scale purification schemes currently revolve around the use of various chromatography steps for the purification of biopharmaceutical products. New purification technologies for large scale bioprocessing are in high demand due to the high cost and other limitations associated with the use of multiple chromatography steps. Finding alternative solutions for both biopharmaceutical production and purification schemes can make a positive impact on cost and can enhance the chance for a successful biopharmaceutical product.

## **2.2 Recombinant Protein Production in Plants**

### *2.2.1 Plant Expression Systems*

An alternative expression system that could increase efficiency and reduce costs is a transgenic plant expression system. Plant expression systems have proven to be a proficient means of producing functional therapeutic recombinant proteins and holds promise for the mass production of biopharmaceutical products. Leafy crops including lettuce, alfalfa, and tobacco, cereals including corn, rice, wheat, and barley, legumes including soybean, pea, and fruits and vegetables including potato, carrot, and tomato have all been used for biopharmaceutical production [5]. Transgenic plants have many advantages when compared with traditional bioreactors. Raw material can be produced at low cost on an agricultural scale with the possibility, in certain cases, of using edible plant material with no need for costly purification. Transgenic plant expression systems have reduced capital costs relative to fermentation methods [6] as well as rapid production scale up. Plants can also provide increased safety from contamination because plants do not serve as hosts for human pathogens and do not produce endotoxins. Transgenic plants have been shown to be capable of synthesizing a wide range of recombinant proteins including human growth hormone fusion proteins, interferons,

human serum albumin, monoclonal antibodies, antibody fragments, plasma proteins, vaccines, and blood substitutes [7].

Potential disadvantages of recombinant protein production in transgenic plants are centered around environmental concerns, authenticity of plant-produced proteins, and process development. There are concerns regarding the release of genetically modified plants into the environment without crossing to wild species. However, depending upon the plant species or tissue the protein is targeted to, containment may or may not be a substantial issue. For example, in all crop species, expression of the transgene in the chloroplast will result in natural containment since functional chloroplast DNA is not transmitted through pollen [8]. On the other hand, post-translational modifications of proteins are often essential for their correct folding and function. Though plants can correctly process mammalian genes, they are unable to perform some authentic mammalian post-translational modifications, such as glycosylations. Usually this does not lead to a difference in protein activity, but could lead to problems with efficacy for a biopharmaceutical product [9]. To combat the glycosylation authenticity problems, several plant lines have been produced which have humanized plant glycosylation systems [6]. Another issue for recombinant protein production by transgenic plants is the low protein expression level. Though protein expression levels in transgenic plants are steadily increasing due to expression vector improvements, typical expression levels of 0.01% to 0.1% total soluble protein are currently reported for biopharmaceutical products [10]. This is much less than a protein produced at high level in *E. coli*, which can accumulate at the level of 20% total soluble protein [6]. Moreover, the processing of large quantities of biomass will be required to obtain necessary amounts of purified recombinant protein using a plant expression system. While this may require substantial front-end equipment, the overall production costs associated with transgenic plants are expected to be much lower when compared to the traditional expression systems [3]. In summary, plant bioreactors certainly have the potential to be more cost-effective than traditional expression systems. Despite this potential, the absence of efficient methods for recombinant protein purification has so far limited the exploitation of plant expression systems on a commercial scale.

### 2.2.2 Tobacco Expression System

Synthesis of recombinant proteins in a tobacco (*Nicotiana tabacum*) expression system is a particularly well-suited alternative when compared to current commercial-scale expression systems. The availability of robust transformation procedures and well-characterized regulatory elements for the control of transgene expression makes a tobacco expression system extremely attractive. Tobacco is neither a food nor a feed crop, so there is a reduced risk of transgenic material or recombinant proteins contaminating feed and human food crops [5]. Large amounts of biomass can be easily produced; therefore tobacco leaf tissue has been widely used as the targeted site for recombinant protein expression [9]. Tobacco has been used extensively as a model plant to investigate the feasibility of using transgenic plants as bioreactors and to demonstrate the lower overall cost attributed to biopharmaceuticals produced by plants [11]. Many recombinant proteins have been successfully produced in tobacco including  $\alpha$ -amylase, chymosin, erythropoietin, growth hormone, enterotoxin B, hepatitis B surface antigen,  $\beta$ -interferon, lysozyme, phytase, and xylanase [12].

There are some shortcomings associated with expressing recombinant proteins in transgenic tobacco concerning the recovery and purification of a target protein. Tobacco leaves contain large amounts of native phenolic compounds and toxic alkaloids that must be removed during the purification process [13, 14]. When the leaf material is processed by grinding or shearing, these compounds are released and can interfere with downstream processing through the formation of complexes with proteins in the aqueous extract [15, 16] or by fouling resin during adsorption processes such as chromatography [17]. Leaf tissue contains high levels of proteases, which can contribute to target protein instability during initial harvest and protein extraction [14]. Due to these restrictions, harvested material has a limited shelf life and must be processed immediately after harvest [5]. Tobacco extract also contains numerous other plant compounds such as carbohydrates, nucleic acids, and native proteins that must be eliminated during protein purification. Native tobacco proteins can be categorized into Fraction 1 or Fraction 2 proteins based on electrophoresis [18]. The photosynthetic chloroplast enzyme ribulose 1,5-bisphosphate carboxylase-oxygenase (Rubisco) makes up the majority of Fraction 1, and can account for up to 50% of the total soluble leaf protein [18]. Its molecular weight is approximately

560 kD and consists of eight large subunits and eight small subunits of 55 kD and 12.5 kD, respectively [19]. Rubisco is an acidic protein (isoelectric point is 6.0 for the large subunits and 5.3 for the small subunits), and creates a serious engineering challenge for acidic recombinant protein purification from tobacco.

## **2.3 Bioprocessing**

### *2.3.1 Guidelines*

For transgenic plants, the cost of downstream processing is estimated to be greater than 80% of the total cost [12]. If the cost for protein purification from a plant bioreactor is higher than the cost for purification of the same protein from mammalian cell culture, this would offset a key advantage of lower overall cost associated with the production of biopharmaceuticals in plants [11]. When designing a cost-effective purification scheme, the most important factor to consider is the ultimate goal of the purification process: the target protein at the final required purity. As mentioned previously, the requisite purity is usually set by the FDA and it is extremely important to be able to meet their standards for the biopharmaceutical product to succeed. Depending on the source of the target protein, purification strategies will vary. The expression system determines the majority of the impurities that will need to be separated from the target protein to obtain the desired final purity. Plant expression systems can introduce phenolic compounds, pigments, and mucilages into the sample [20] as well as host proteins and host DNA. Additional impurities, such as solvents and chromatography column leachables can also be introduced during any of the downstream processing steps. However, one of the largest impurities encountered through all of the purification steps is water.

All of the impurities mentioned above are process-related impurities and the majority cannot be avoided. In addition, there are also product-related impurities to overcome during purification of the target protein. Denatured forms, misfolded forms, fragments, and aggregates of the target protein occur during production and are hard to separate from the target protein due to their similar physico-chemical characteristics. Contaminants such as viruses and bacteria can also adventitiously enter the production process and must be removed. Elimination of impurities and contaminants must be very



gentle to conserve the target protein's native structure and in general, near-neutral pH values and moderate temperatures ( $\leq 30$  °C) must be maintained [21].

To remove process- and product-related impurities, the physico-chemical characteristics of the target protein including size, charge, isoelectric point (pI), hydrophobicity, and shape can be exploited. A vast range of methods exist for the isolation and purification of target proteins based on these characteristics and new techniques are continuously developed. Due to the complexity of the mixture of impurities and target proteins, several purification steps are usually needed to obtain the target protein at the required purity. Commercial-scale purification schemes revolve around the use of multiple chromatography steps. The addition of multiple processing steps results in higher product loss and increased processing time, which both subsequently increase cost.

Purification platforms are based on a three step approach: protein capture, intermediate purification, and polishing. In the capture step, the target protein is isolated and concentrated from the crude extract while also transferring the target protein into the most stabilizing environment possible. Removal of some critical impurities and/or contaminants can also be accomplished during the capture step. In the intermediate purification step, the bulk of the impurities, such as other proteins and nucleic acids are removed. During the last polishing step, trace amounts of impurities are eliminated along with substances that are closely related to the target protein (aggregates, misfolded target proteins). The objective during the polishing step is to achieve the final purity.

The key to successful and efficient protein purification is to select the most appropriate techniques, optimize their performance to suit the requirements, and combine them in a logical way to maximize yield and minimize the number of steps required [22]. Each step in a purification scheme is a compromise between three considerations: retaining biological activity of the target protein, maximizing the degree of purification at the step, and maximizing recovery of the target protein. The degree of purification and the recovery are frequently in conflict, and a compromise is typically needed. Specific activity (SA) is defined as units of target protein activity per milligram of total soluble protein:

$$\text{Specific Activity} = \frac{\text{Target Protein Activity (U)}}{\text{Total Soluble Protein (mg)}} \quad (1)$$

Specific activity should increase throughout the purification, reaching a maximum value when the protein is pure. Recovery is defined as units of target protein activity after a purification step, or for an entire purification process, divided by the units of target protein activity in the initial extract:

$$\text{Recovery (\%)} = \frac{\text{Final Target Protein Activity (U)}}{\text{Initial Target Protein Activity (U)}} \times 100 \quad (2)$$

Recovery decreases with each purification step. The main objective in protein purification is to maintain total activity units and recovery while steadily increasing specific activity. Purification factor is the final specific activity ( $SA_{\text{final}}$ ) at a given step, or for an entire purification process, divided by specific activity in the initial extract ( $SA_{\text{initial}}$ ):

$$\text{Purification Factor} = \frac{SA_{\text{final}}}{SA_{\text{initial}}} \quad (3)$$

The final purification factor may range from two-fold for a well-expressed recombinant protein to 100,000 fold for a minor protein [22].

### 2.3.2 Aqueous Two-Phase Extraction

ATPE is a powerful technique for separation and purification of proteins, as well as many other biomolecules, including nucleic acids, viruses, and cells. The characteristic feature of ATPE is the high water content, which provides a gentle, non-toxic environment for recombinant proteins. Compared to existing technology, ATPE has the capacity to process high biomass loads while circumventing many of the shortcomings present in early centrifugation and filtration steps that arise due to high viscosity and heterogeneous particle size [23]. This technique can be highly selective and, when properly optimized, offers the possibility of combining clarification, concentration, and initial purification into one purification step. Reducing the number of steps in a purification process improves the cost and target protein recovery. ATPE has the potential to become an alternative to the widely-used centrifugation, filtration, and chromatography steps for the initial purification of recombinant proteins.

ATPE utilizes immiscible mixtures of different water-soluble polymers or a single polymer and a specific salt. When both polymers or the polymer and salt are combined, two phases will be formed with the first polymer dominating one phase and the second polymer or salt dominating the other phase. ATPE involves two unit operations: equilibration and phase separation [24]. Equilibration comprises mixing of the components that constitute the aqueous two-phase system and the material to be partitioned, followed by dispersion of the phases to obtain equilibrium [24]. This is followed by phase separation of the two phases. Time for phase separation is the rate-limiting step in ATPE. Phase separation under gravity varies between a few minutes to a few hours due to the small difference in the densities and the viscosities of the two phases [25]. Centrifugation at low speed is commonly used to hasten the process [25].

The composition of an aqueous two-phase system can be represented by a phase diagram with a single binodial curve, illustrated by the polymer-salt system in Figure 1.1. Points to the right of the binodial curve represent all the mixtures of the polymer and the salt that create a two-phase system. Points T and B are called nodes, and represent the compositions of the two phases. Connecting nodes T and B is a tie-line. Any point on the tie-line, such as point M, represents the total system composition. Different points on the same tie-line result in two-phase systems with identical phase compositions but different volumes of the coexisting phases. Compositions of the coexisting phases can only be changed by the addition of another component or a change in the solvent. This would result in different nodes and a different tie-line, illustrated by the other tie-lines above the binodial curve in Figure 2.1. Point C on the binodial curve is the point where the two nodes coincide and is called the critical point. This point represents the theoretical case in which the compositions and the volumes of the coexisting phases are equal [23]. Numerous phase diagrams for both two polymer and polymer-salt systems have been reported [23, 26, 27] and are helpful when designing aqueous two-phase systems.

Systems composed of a polymer and a salt have advantages over systems composed of two polymers. Cost is reduced as phase-forming polymers are more expensive than phase-forming salts [28, 29]. Also, the phases will have a lower viscosity, which results in faster phase separation [30]. When using a polymer/salt aqueous two-phase system, polyethylene glycol (PEG) is almost exclusively used as the phase-forming

polymer [30]. The phase behavior of PEG/salt systems have been reported in detail with phase-forming salts including potassium phosphate, ammonium sulfate, magnesium sulfate, sodium sulfate, and sodium carbonate [23, 26, 27].

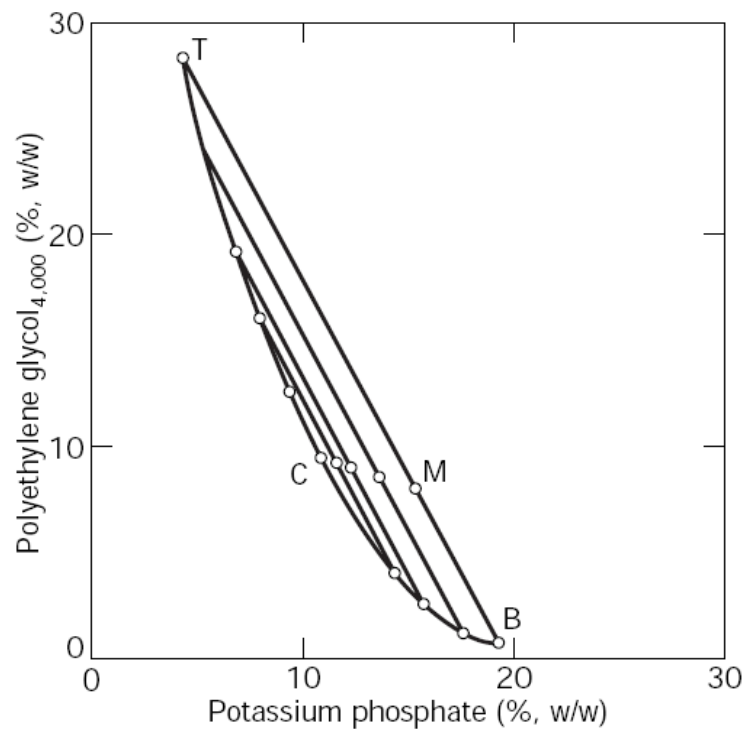


Figure 2.1 Phase diagram for polyethylene glycol-potassium phosphate system [31].

Exploitation of ATPE for protein purification uses the unique equilibrium distribution, or partitioning, of the target proteins and the impurities between the two phases. Ideally, the differences in partition behavior between the target protein and all other impurities and unwanted material should be very large. In this case, the target protein would partition to one phase and all of the impurities would partition to the other phase, resulting in excellent target protein purification. Most often this is not the case, and the difference in partition behavior is not large for some of the impurities and the target protein. Many factors affect the partition behavior of the proteins and impurities. These factors include, but are not limited to, the type, molecular weight, and concentration of phase polymers and salts, the type and concentration of ions added to the

two-phase system, temperature, and pH of the two-phase system. Manipulating these factors changes the properties of the phases and can help to increase differences in partition behavior between target proteins and impurities. When separating biomolecules using ATPE, the amounts of soluble material should not exceed 0.1 wt % of the total weight of the two-phase system used. For example, a 10 g aqueous two-phase system can accommodate up to 0.01 g total protein. It has been shown that when the amount of material being partitioned exceeds 0.1 wt % of the total system weight, the material may affect the properties of the system [23].

Several variables are used to help describe the relationships between solutes in the top and bottom phases and can be used when trying to optimize purification of a target protein. The solute partition coefficient,  $K$ , is generally defined as the ratio of the solute concentration in the top phase,  $C_T$ , to the solute concentration in the bottom phase,  $C_B$ .

$$K = \frac{C_T}{C_B} \quad (4)$$

The partition coefficient will be greater than one if the solute preferentially partitions to the top phase and less than one if the solute preferentially partitions to the bottom phase. Altering any of the aforementioned factors can significantly change the partition coefficient. The selectivity of the target protein is defined by  $\alpha$ , and it is a ratio of the partition coefficient of the target protein,  $K_P$ , and the lumped partition coefficient of all other impurities,  $K_C$ .

$$\alpha = \frac{K_P}{K_C} \quad (5)$$

Depending on which phase the target protein partitions to, the selectivity will be greater than one or less than one. The selectivity of a target protein will be constant for all two-phase system compositions on the same tie line in the phase diagram. Due to the constant selectivity, the purification factor of the recovered target protein can be improved by varying the phase ratio:

$$\text{Phase Ratio} = \frac{\text{Mass of Top Phase}}{\text{Mass of Bottom Phase}} \quad (6)$$

If the selectivity is greater than one, a smaller phase ratio will increase the purification factor of the target protein. However, the recovery of the target protein will decrease as the phase ratio decreases. An optimal phase ratio would provide a balance between the purification factor and the recovery.

Due to the advantages cited above, ATPE has been used for the purification of many recombinant proteins from bacterial and yeast expression hosts. Table 2.1 lists 14 proteins purified using ATPE, the expression host used, and the recovery and purification factors obtained. ATPE offered recovery values of 90% or greater for 11 of the 14 target proteins and the purification factor varies from 1.25 – 17 [28, 32, 33, 34, 35, 36]. ATPE has also been used to purify recombinant proteins from mammalian cell culture, although it is not as well-studied as with bacterial or yeast cell culture. Murine IgG was separated from a serum free, crude, concentrated hybridoma cell culture supernatant using a system composed of 15% PEG 1450, 14% potassium phosphate, and 12% NaCl at pH 5.5 where virtually all IgG partitioned to the top phase and the contaminants to the bottom phase [37]. Addition of 14% potassium phosphate partitioned the IgGs into the bottom salt phase and the contaminants were reduced 18 fold, resulting in 80% pure IgG and a purification factor of 5.9 [37]. Another aqueous two-phase system composed of 12% PEG, 10% phosphate, and 15% NaCl at pH 6 was used to purify human monoclonal antibodies from a Chinese hamster ovary (CHO) concentrated cell culture supernatant with an 88% recovery in the polymer-rich top phase and a purification factor of 4.3 [31]. The same aqueous two-phase system was also successfully used to purify antibodies from a hybridoma cell culture supernatant with 90% recovery and a purification factor of 4.1 [38].

Table 2.1 Recombinant proteins purified using ATPE and the recovery and purification factors obtained for each [28, 32, 33, 34, 35, 36].

Protein	Recovery (%)	Purification Factor
Fumarate Hydratase ( <i>Brevibacterium ammoniagenes</i> )	83	7.5
Fumarate Hydratase ( <i>Escherichia coli</i> )	93	3.4
Aspartate-Ammonia-Lyase ( <i>Escherichia coli</i> )	96	6.6
Isoleucyl-tRNA-Synthetase ( <i>Escherichia coli</i> )	93	2.3
Penicillin Amidase ( <i>Escherichia coli</i> )	90	8.2
Human Insulin-Like Growth Factor-I and II Fusion Protein ( <i>Escherichia coli</i> )	90	2
Penicillin Acylase ( <i>Escherichia coli</i> )	85	5.7
Glucose-6-Phosphate-Dehydrogenase ( <i>Leuconostoc sp</i> )	94	1.3
L-2-Hydroxyisocaproate- Dehydrogenase ( <i>Lactobacillus casei</i> )	93	17
Leucine Dehydrogenase ( <i>Bacillus sphaericus</i> )	98	2.4
$\alpha$ -Amylase ( <i>Bacillus subtilis</i> )	80	2
Elastase ( <i>Bacillus subtilis</i> )	89.9	1.7
Endo-Polygalacturonase ( <i>Kluyveromyces marxianus</i> )	95	1.25
Alcohol Dehydrogenase ( <i>Saccharomyces cerevisiae</i> )	96	2.5

ATPE has been used to purify recombinant proteins from plant expression hosts at the laboratory scale. ATPE was successful in purifying egg white lysozyme, porcine pancreatic ribonuclease A, and equine heart cytochrome C from transgenic corn endosperm and corn germ using a PEG/sodium sulfate system with recovery of nearly 100% of the target proteins and average purification factors of 9 [39]. Spiking

experiments in which egg white lysozyme was added to tobacco extract to simulate a transgenic system have also demonstrated the effectiveness of ATPE. The egg white lysozyme was successfully separated from tobacco extract in a PEG/sodium sulfate system with a 79% recovery and purification factor of 10 [40]. ATPE was shown to be effective in the initial recovery, purification, and stabilization of Ricin B from tobacco hairy root culture [41]. Human anti-human immunodeficiency virus monoclonal antibody (MAb) 2F5 was shown to be successfully separated from transgenic tobacco extract using a PEG/potassium phosphate system with 95% recovery and 3-4-fold purification [27]. In addition, this system permitted the removal of plant-derived phenolics and toxic alkaloids. Overall, there are very few published reports describing ATPE for recombinant protein purification from plant expression systems.

It is important to note that phase separation in mixtures of two polymers or of a single polymer and a salt is not clearly understood and the mechanism controlling aqueous two-phase partitioning is largely unknown. Therefore, when designing an aqueous two-phase system, the best choice of partition conditions for the separation of target proteins remains mostly trial-and-error [23].

#### *2.3.2.1 Comparison of Aqueous Two-Phase Extraction and Chromatography*

On a large scale, separation of proteins using chromatography is not an easily applied method, despite being extremely effective and simple on a laboratory scale. Chromatographic methods like column chromatography or high-pressure liquid chromatography are highly specific, but can handle only small amounts of feed at a time. Chromatography scale up is complex, and limitations such as discontinuity in the process, slow protein diffusion, and large pressure drops in the system are seen [28]. Compared to chromatography, an aqueous two-phase system can handle larger volumes and a higher biomass load [23, 24]. If optimized correctly, ATPE can be a highly selective process [24]. Aqueous two-phase systems have no solid phase, therefore, thorough mixing of the two phases is possible and interphase mass transport is rapid. An ATPE can easily be scaled up without an appreciable change in the nature or efficiency of the process [28]. It has been demonstrated that scale-up is feasible by proportional linear increase of the amount of ingredients [42, 43, 44]. ATPE is anticipated to be scalable using PEG/salt



systems due to lower viscosity, lower costs, and fast separation time [24, 25]. Scale up of extractions in aqueous two-phase systems has been possible through the use of conventional extraction equipment used in chemical industry [24]. ATPE uses the same principles as liquid–liquid extraction, a well-known chemical process. Therefore, similar commercially-available equipment may be used for mixing and phase separation [31]. A two-stage extraction process can be used to first partition the target protein to the top polymer-rich phase and after separating the phases, a second ATPE is performed to partition the target protein to the bottom phase. The top polymer phase may be recycled in large-scale processes to lower costs, and furthermore, ATPE is suitable for continuous operation [24, 31, 45, 46]. The large-scale application of ATPS is mainly limited by the theoretical understanding of phase equilibrium and protein partitioning.

#### **2.4 $\beta$ -Glucuronidase as a Model Protein**

Beta-glucuronidase (GUS) is an enzyme that catalyzes the hydrolysis of  $\beta$ -glucuronides [47]. GUS is a homo-tetrameric enzyme and has an isoelectric point around pH 5.5 [48]. The molecular mass of each monomer is approximately 68.2 kDa [48, 49]. To examine the feasibility of using ATPE in recombinant protein purification, GUS can be used as a model protein to simulate biopharmaceutical expression, production, and purification. GUS is an ideal model acidic protein because it has well-known properties and its enzymatic activity is easily quantified using a standard assay procedure over a broad pH range, with most GUS activity occurring between pH 5.0 and pH 7.8 [47]. The enzyme is very stable when exposed to various ionic conditions and detergents and is mostly resistant to protease degradation and thermal degradation [47]. Plant extracts containing GUS may be stored at -70 °C for prolonged periods of time and at 4 °C for a few days without losing significant activity [50]. Also, GUS is not endogenous in higher plant species [47] and therefore the likelihood of background activity is very small.

#### **2.5 Summary**

Biopharmaceutical-based therapeutics have emerged as one of the most significant medical innovations in the last three decades. Treatments are now available for many conditions and diseases, such as cancer and autoimmune diseases, which

previously had very few therapeutic options. Production of an FDA-approved biopharmaceutical product is a rigorous and difficult process. Biopharmaceuticals require a very high purity and several expensive purification steps must be utilized to obtain the stringent purity requirements set by the FDA. Cost-effective recombinant protein production and purification is vital for successful biopharmaceutical manufacturing. Deviating from the traditional biopharmaceutical expression systems and purification schemes by using alternative procedures can make a positive impact on cost. An alternative expression system that could increase efficiency and reduce costs is a transgenic plant expression system. Plant expression systems have proven to be an efficient means of producing functional therapeutic proteins and hold promise for the production of biopharmaceuticals. Non-chromatography-based purification schemes can significantly reduce the cost of downstream processing. Combining a tobacco expression system and an aqueous two-phase extraction purification process has the potential to improve the cost of biopharmaceutical production by eliminating the need for an expensive mammalian cell culture expression system and initial purification using chromatography.

## References

1. Redwan, el-RM (2007) *Cumulative updating of approved biopharmaceuticals*. Hum Antibodies 16(3-4) 137-58.
2. Boehm, Robert (2007) *Bioproduction of Therapeutic Proteins in the 21st Century and the Role of Plants and Plant Cells as Production Platforms*. Ann NY Acad Sci 1102 (1), 121-134.
3. Chu, L. and D. K. Robinson (2001) *Industrial choices for protein production by large-scale cell culture*. Current Opinion in Biotechnology 12(2) 180-187.
4. Aldingtona , S. and J. Bonnerjea (2007) *Scale-up of monoclonal antibody purification processes*. J Chromatogr B 848(1), 64-78.
5. Fischer, R., E. Stoger, S. Schillberg, P. Christou, and R. M. Twyman (2004) *Plant-Based Production of Biopharmaceuticals*. Current Opinion in Plant Biology 7, 1-7.
6. Larrick, J. W., L. Yua, C. Naftzgera, S. Jaiswala, and K. Wycoffa (2001) *Production of secretory IgA antibodies in plants*. Biomolecular Engineering 18(3) 87-94.
7. Fischer, R., and N. Emans (2000) *Molecular Farming of Pharmaceutical Proteins*. Transgenic Research 9, 279-299.
8. Schillberg, S., R. M. Twyman, and R. Fischer (2005) *Opportunities for Recombinant Antigen and Antibody Expression in Transgenic Plants-Technology Assessment*. Vaccine 23, 1764-1769
9. Menkhaus, T. J., Y. Bai, C. Zhang, Z. L Nikolov, and C. E. Glatz (2004) *Considerations for the Recovery of Recombinant Proteins from Plants*. Biotechnol. Prog. 20, 1001-1014.
10. Daniell, H., S. J. Streatfield, and K. Wycoff (2001) *Medical molecular farming: production of antibodies, biopharmaceuticals and edible vaccines in plants*. Trends Plant Sci. 6(5), 219-26.
11. Twyman, R. M., E. Stoger, S. Schillberg, P. Christou, and R. Fischer (2003) *Molecular Farming in Plants: Host Systems and Expression Technology*. Trends in Biotechnology 21(12), 570-578

12. Kusnadi, A. R., Z. L. Nikolov, and J. A. Howard (1997) *Production of Recombinant Proteins in Transgenic Plants: Practical Considerations*. Biotechnology and Bioengineering 56(5), 473-484.
13. Davis, D. L. and M. T. Nielson (1999) *Tobacco: Production, Chemistry and Technology*. D.L. Davis and M.T. Nielson (Editors). Oxford: Blackwell Science, 3.
14. Moloney, M. M. (1995) *Molecular Farming in Plants - Achievements and Prospects*. Biotechnology & Biotechnological Equipment 9(1), 3-9.
15. Cheryan, M. (1980) *Phytic acid interactions in food systems*. Crit Rev Food Sci Nutr. 13(4), 297-335.
16. Jervis, L. and W. S. Pierpoint (1989) *Purification Technologies for Plant-Proteins*. J of Biotechnol 11(2-3), 161-198.
17. Witcher, D. R., E. E. Hood, D. Peterson, M. Bailey, D. Bond, A. Kusnadi, R. Evangelista, Z. Nikolov, C. Wooge, R. Mehig, W. Kappe, J. Register, and J. A. Howard (1998) *Commercial production of beta-glucuronidase (GUS): a model system for the production of proteins in plants*. Molecular Breeding 4(4), 301-312.
18. Garger, S. J., R. B. Holtz, M. J. McCulloch, and T. H. Turpen (2000) *Process for isolating and purifying viruses, soluble proteins, and peptides from plant sources*. U. S. Patent 6,033,895.
19. Kung, S. D. (1976 ) *Tobacco fraction 1 protein: a unique genetic marker*. Science. 191(4226), 429-434.
20. Platis, D. and N. E. Labrou. (2006) *Development of an Aqueous Two-Phase Partitioning System for Fractionating Therapeutic Proteins from Tobacco Extract*. J Chromatogr A 1128(1-2), 114–124.
21. Jervis, L., and W. S. Pierpoint (1989) *Purification Technologies for Plant Proteins*. J Biotech 11, 161-198.
22. Amersham Biosciences (now known as GE Healthcare) (2002) *Protein Purification Handbook*. 18-1132-29 Edition AC.
23. Zaslavsky, B. Y. (1995) *Aqueous Two-Phase Partitioning*. Marcel Dekker, Inc., New York 75–84, 221–285, 447–498, 643–645, 648–649, 660–662.
24. Hatti-Kaul, R. (2001) *Aqueous Two-Phase Systems, A General Overview*. Molecular Biotechnology 19(3), 269-277.

25. Hustedt, H., K. H. Kroner, and M.-R. Kula (1985) *Applications of Phase Partitioning in Biotechnology, in Partition in Aqueous Two-Phase Systems. Theory, Methods, Uses and Applications to Biotechnology*. H. Walter, D.E. Brooks, and D. Fisher (Editors) Academic Press, Orlando, FL 529–587.
26. Albertsson, P. A. (1986) *Partition of Cell Particles and Macromolecules, 2<sup>nd</sup> Edition*. John Wiley & Sons Inc., New York 261-287.
27. Snyder, S. M., K. D. Cole, and D. C. Sziag (1992) *Phase Compositions, Viscosities, and Densities for Aqueous Two-Phase Systems Composed of Polyethylene Glycol and Various Salts at 25 °C*. J Chem Eng Data 37, 268-274.
28. Silva, M. E. and T. T. Franco (2000) *Liquid-Liquid Extraction of Biomolecules in Downstream Processing - A Review Paper*. Braz. J. Chem. Eng. 17(1), ISSN 0104-6632.
29. Salabat, A. (2001) *The influence of Salts on the Phase Compositions in Aqueous Two-Phase Systems: Experiments and Predictions*. Fluid Phase Equilibria 187-188, 489-498.
30. van Berlo, M. K. Ch. A. M. Luyben, and L. A. M. van der Wielen (1998) *Poly(ethylene glycol)–salt aqueous two-phase systems with easily recyclable volatile salts*. J Chromatogr B 711(1-2), 61-68.
31. Flickinger, M. C. and S. W. Drew (1999) *Encyclopedia of Bioprocess Technology - Fermentation, Biocatalysis, and Bioseparation*. John Wiley & Sons 2179–2190.
32. Köhler K, B. Nilsson, and A. Veide (1992-1993) *Recovery of Extracellular Human Insulin-like Growth Factor-I and II as a Fusion Protein from Escherichia coli Culture Broth by Aqueous Two-Phase Extraction*. Bioseparation. 3(4), 241-50.
33. Zhi, W., J. Song, J. Bi, and F. Ouyang (2004) *Partial Purification of  $\alpha$ -Amylase from Culture Supernatant of Bacillus subtilis in Aqueous Two-Phase Systems*. Bioprocess Biosyst Eng. 27(1), 3-7.
34. Xu, Y., G.-Q. He, and J. –J. Li (2005) *Effective Extraction of Elastase from Bacillus sp. Fermentation Broth using Aqueous Two-Phase System*. J Zhejiang Univ Sci B. 6(11), 1087-94.

35. Wu, Y. T., M. Pereira, A. Venâncio, and J. Teixeira. (2000) *Recovery of Endo-Polygalacturonase using Polyethylene Glycol-Salt Aqueous Two-Phase Extraction with Polymer Recycling*. *Bioseparation* 9(4), 247-54.
36. Marcos, J. C., L. P. Fonseca, M. T. Ramalho, and J. M. Cabral (1999) *Partial Purification of Penicillin Acylase from Escherichia coli in Poly(ethylene Glycol)-Sodium Citrate Aqueous Two-Phase Systems*. *J Chromatogr B Biomed Sci Appl.* 734(1), 15-22.
37. Andrews, B. A., S. Nielsen, and J. A. Asenjo (1996) *Partitioning and Purification of Monoclonal Antibodies in Aqueous Two-Phase Systems*. *Bioseparation* 6(5), 303-13.
38. Azevedo, A. M., P. A. Rosa, I. F. Ferreira, and M. R. Aires-Barros (2007) *Optimization of Aqueous Two-Phase Extraction of Human Antibodies*. *J Biotechnol.* 132(2), 209-17.
39. Gu, Z., and C. E. Glatz (2007) *Aqueous Two-Phase Extraction for Protein Recovery from Corn Extracts*. *J Chromatogr B* 845(1), 38-50.
40. Balasubramaniam, D., C. Wilkinson, K. Van Cott, and C. Zhang (2003) *Tobacco Protein Separation by Aqueous Two-Phase Extraction*. *J Chromatogr A* 989(1) 119-129.
41. Zhang, C., F. Medina-Bolivarb, S. Buswell, and C. L. Cramer (2005) *Purification and Stabilization of Ricin B from Tobacco Hairy Root Culture Medium by Aqueous Two-Phase Extraction*. *J Biotechnol* 117(1), 39-48.
42. Hart, R.A., P. M. Lester, H. Riefsnyder, J. R. Ogez, and S. E. Builder (1994) *Large Scale, in situ Isolation of Periplasmic IGF-I from E. coli*. *BioTechnology* 12, 1113–1117.
43. Schutte, H., K. H. Kroner, W. Hummel, and M.-R. Kula (1983) *Recent Developments in Separation and Purification of Biomolecules*. *Ann. N.Y. Acad. Sci.* 413, 270–282.
44. Kroner, K .H., H. Schutte, W. Stach, and M.-R. Kula (1982) *Scale up of Formate Dehydrogenase by Partition*. *J. Chem. Technol. Biotechnol.* 32, 130–137.
45. Minuth, T., J. Thommes, and M.-R. Kula (1995) *Extraction of Cholesterol Oxidase from Nocardia rhodochrous using a Nonionic Surfactant-Based Aqueous Two-Phase System*. *J Biotechnol.* 38, 151-164.

46. Hustedt, H., K.-H. Kroner, U. Menge, and M.-R. Kula (1985) *Protein Recovery using Two-Phase Systems*. Trends Biotechnol. 3, 139-144.
47. Gallagher, S.R., in S.R. Gallagher (Editor) (1992) *GUS Protocols: Using the GUS Gene as a Reporter of Gene Expression*. Academic Press, San Diego, CA.
48. Jefferson, R.A., S. M. Burgess, D. Hirsh (1986)  *$\beta$ -Glucuronidase from Escherichia coli as a Gene-Fusion Marker*. Proc. Natl. Acad. Sci. U. S. A. 83, 8447.
49. Jefferson, R.A. and K.J. Wilson (1991) *The GUS gene fusion system*, in *Plant Molecular Biology Manual*. S. Gelvin, R. Schilperoort, and D.P. Verma (Editors) Kluwer Academic Publishers, the Netherlands 1-33.
50. Jefferson, R.A., M. Bevan, and T. Kavanagh (1987) *The use of the Escherichia coli beta-glucuronidase as a gene fusion marker for studies of gene expression in higher plants*. Biochem Soc Trans. 15(1), 17-8.

## Chapter 3

### Experimental

#### 3.1 Materials

Non-transgenic *Nicotiana tabacum* cv. Xanthi seeds were kindly provided by Dr. Carol Wilkinson. Transgenic tobacco seeds containing the recombinant GUS gene regulated by Super P promoter were kindly provided by Dr. Luis Fabricio-Medina Bolivar (Arkansas State University) and were originally developed by Dr. Stanton B. Gelvin (Purdue University) [1]. Polyethylene glycol (PEG), molecular weight 3350, was purchased from Integra (Renton, WA, USA). PEG, molecular weight 8000, was purchased from Promega (Madison, WI, USA). Ammonium sulfate, potassium phosphate monobasic and dibasic, sodium phosphate monobasic and dibasic, sodium sulfate, and sodium chloride were purchased from Fisher Scientific (Pittsburgh, PA, USA). Beta-glucuronidase from *E. coli*, poly(vinylpyrrolidone) (PVPP), high molecular weight (>1,000,000), 2-mercaptoethanol, ethylenediamine-tetraacetic acid (EDTA), p-nitrophenyl  $\beta$ -D-glucuronide (PNPG), phytigel, and p-nitrophenol (PNP) were purchased from Sigma (St. Louis, MO, USA). Bio-Rad Protein Assay and Bio-Safe Coomassie stain were purchased from Bio-Rad Laboratories (Hercules, CA, USA). Bovine serum albumin was purchased from Pierce (Rockford, IL, USA). Dithiothreitol (DTT), kanamycin, and Murashige and Skoog basal salts and vitamins were obtained from Bioworld (Dublin, OH, USA). Greiner 96-well clear, flat bottom microplates were purchased from USA Scientific (Ocala, FL, USA). All sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS PAGE) products including 4-12% Bis-Tris Novex mini gels and NuPAGE LDS (lithium dodecyl sulfate) sample buffer, 3-(N-morpholino)propane sulfonic acid (MOPS) SDS running buffer (50 mM MOPS, 50 mM Tris, 0.1% SDS, 1 mM EDTA, pH 7.7), and antioxidant were purchased from Invitrogen (Carlsbad, CA, USA). Microcon centrifugal filter devices were purchased from Millipore (Bedford, MA, USA).



### **3.2 *Nicotiana tabacum* Plant Generation**

Transgenic and non-transgenic tobacco seeds were prepared according to published protocol [2]. Seeds were placed aseptically in plastic boxes (11.4 cm × 8.6 cm × 10.2 cm) containing modified Murashige and Skoog (mMS) media [3] and allowed to germinate. Media was supplemented with 200 mg/L kanamycin for transgenic plants. Plants were grown in a temperature controlled (23 °C) incubator and were used for experiments after approximately two months when the plants were large enough to fill the plastic boxes. Leaf size varied from approximately 5 cm to 10 cm. Plants were propagated after the leaves were used for experiments. After one month of growth, the propagated plants were large enough to fill the plastic boxes and were used for experiments. This propagation process was repeated periodically. After the first propagation of transgenic plants, all plants were grown on mMS media with 100 mg/L kanamycin.

### **3.3 Protein Extraction**

Fresh tobacco leaves were excised from each plant to prepare the extracts. Leaf tissue was weighed and placed in a 50 mL conical tube. Ice-cold buffer was added to each tube at a ratio of 5 mL buffer:1 g leaf tissue. The extraction buffer consisted of 50 mM sodium phosphate + 10 mM BME + 1mM EDTA at pH 7 or pH 8 depending on ATPE conditions. The samples were homogenized using a PowerGen 700 (Fisher Scientific, Pittsburgh, PA, USA) until no large particulates remained. Immediately after homogenization, 2% (w/v) pre-hydrated PVPP was added to the sample. The extract was mixed using a Genie 2 Fisher Vortex (Fisher Scientific, Pittsburgh, PA, USA) and allowed to sit at room temperature for 15 minutes. After centrifugation at 4 °C and 17,000 × g for 20 minutes, the supernatant was removed and filtered through a 0.22-µm syringe filter. The rGUS expression level in the transgenic plants ranged from 100 to 380 units of rGUS activity per milligram of total soluble protein.

### **3.4 Analytical methods**

#### *3.4.1 Bio-Rad Protein Assay*

Protein concentration was determined by the Bio-Rad protein assay with bovine serum albumin (BSA) as the standard. The Bio-Rad assay is a dye-binding assay in which a color change of the dye occurs in response to various concentrations of protein [4]. The dye binds to basic and aromatic amino acid residues [4]. All assays were carried out in 96-well clear, flat bottom Greiner microtiter plates and performed in duplicates. Ten  $\mu\text{L}$  of sample or standard was added to each empty well followed by 200  $\mu\text{L}$  of prepared Bio-Rad reagent as specified by the manufacturer for the standard assay and 160  $\mu\text{L}$  of sample or standard was added to each empty well followed by 40  $\mu\text{L}$  of prepared Bio-Rad reagent as specified by the manufacturer for the microassay. The samples were allowed to incubate at room temperature for 5 minutes. Absorbance measurements were read at 595 nm on a Bio-Tek Synergy microplate reader (Bio-Tek Instruments, Inc., Winooski, VT, USA).

#### *3.4.2 GUS Activity Assay*

The GUS activity assay used was a continuous spectrophotometric assay [5] based on the method developed by Jefferson and Wilson [6]. This assay utilizes the ability of GUS to hydrolyze p-nitrophenyl  $\beta$ -D-glucuronide (PNPG) to release chromophore p-nitrophenol (PNP). GUS activity is expressed as unit/mL [6] with one unit of GUS activity defined as the amount needed to liberate 1 nmol PNP/minute at room temperature and pH 7.0. A standard curve was developed using various concentrations of PNP solubilized in 50 mM NaPi at pH 7.0 and pH 8.0 according to published protocol [7]. All assays were carried out in 96-well microtiter plates and performed in triplicates. Four microliters of sample was added to each empty well, followed by 176  $\mu\text{L}$  of 50 mM NaPi + 10 mM BME and 20  $\mu\text{L}$  of 10 mM PNPG solubilized in 50 mM NaPi. Kinetic absorbance measurements were read at 405 nm every 50 s for a total of 8 min on a microplate reader. The results from control experiments (data and results presented in Appendix A) indicated that the GUS activity was not significantly affected ( $\pm 10\%$ ) by the varying concentrations of PEG, ammonium sulfate, potassium phosphate, or sodium sulfate encountered throughout this report.

### *3.4.3 Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis*

Samples were reduced and run on 4-12% Bis-Tris gels with MOPS as the running buffer. Thirteen microliters of sample was mixed with 5 mL LDS sample buffer and 2 mL of 500 mM DTT. The samples were vortexed and heated at 70 °C for 10 min in a water bath. The gels were run for 50-60 minutes at 200 V. After running, the gels were washed with deionized (DI) water and stained with Bio-Safe Coomassie stain or SilverXpress silver staining kit. The gels were scanned with a Bio-Rad ChemiDoc XRS imager and analyzed using Quantity One Software.

## **3.5 Aqueous Two-Phase Extraction**

### *3.5.1 Development of Aqueous Two-Phase System*

PEG/ammonium sulfate, PEG/potassium phosphate, and PEG/sodium sulfate systems were investigated to study the partitioning of native tobacco proteins, GUS, and rGUS. A stock solution of 40% (w/w) PEG in deionized water was prepared and 30% (w/w) stock solutions were also prepared for each phase-forming salt. Ammonium sulfate and sodium sulfate stock solutions were prepared in 50 mM sodium phosphate buffer and were titrated to the appropriate pH based on aqueous two-phase system conditions. Potassium phosphate stock solution was prepared with both monobasic and dibasic potassium phosphate in 50 mM sodium phosphate buffer and was titrated to the appropriate pH if necessary. Systems of 2 g (total weight) containing the required amounts of PEG, phase-forming salt (ammonium sulfate, potassium phosphate, or sodium sulfate), sodium chloride, and tobacco extract, GUS, or transgenic tobacco extract (50 µL) were prepared from the appropriate stock solutions. Sodium phosphate buffer (50 mM) was used to balance the total system weight to 2 g. The systems were thoroughly mixed by vortexing, then centrifuged at  $1150 \times g$  at room temperature for 10 minutes to accelerate the phase separation. The centrifuged systems were allowed to stand for 30 minutes at room temperature. The bottom phase was pipetted out and both phases were weighed. The density of the top and bottom phase was estimated by measuring the mass of 200 µL of each phase. When necessary, the total protein concentration in each phase was determined by the Bio-Rad assay, and the GUS or rGUS protein concentration was determined by the GUS activity assay.

### *3.5.2 Fractional Factorial Design*

Statistical design of experiments is a procedure for planning experiments that will yield valid and objective conclusions. When using ATPE, there are many factors which effect the partitioning of proteins, some of which may be critical and others which may have little or no effect. Screening experiments are an efficient way of determining the important factors in protein partitioning and is an excellent first step when the ultimate goal is to maximize protein purification using ATPE. The primary purpose of the experiment is to select, or screen out, the few important main effects from the many less important ones [8]. There are five important factors which affect the partitioning of proteins when using ATPE to separate GUS from transgenic tobacco extract: polymer type, polymer concentration, phase-forming salt concentration, pH, and the ionic strength of the aqueous two-phase system. Low and high levels of the factors were 3400 and 8000 MW for PEG molecular weight, 10 % (w/w) and 15 % (w/w) for PEG concentration, 13 % (w/w) and 18 % (w/w) for potassium phosphate concentration, 0.1 M and 1.2 M for ionic strength, and 7 and 8 for system pH. The levels for PEG concentration and phase-forming salt concentration were selected to ensure that a two-phase system would result when the polymer and phase-forming salt were mixed. This was accomplished using published phase diagrams available for each of the three systems [9, 10, 11]. Ionic strength levels were selected based on a literature review of ATPE [12, 13, 14]. The pH range was selected because rGUS extraction from transgenic tobacco is maximized at neutral to basic pH values [6, 7] and GUS enzymatic activity can be accurately assayed from pH 7 to 8 [5, 6]. A broader pH range from pH 6 to pH 8 was originally tested using transgenic tobacco plants (data and results presented in Appendix B) but the enzymatic activity of rGUS was very low and hard to quantify at pH 6 compared to pH 7 and pH 8. As a result, it was difficult to evaluate data between the different pH values. The low enzymatic activity of rGUS at pH 6 could be attributed to insufficient rGUS extraction from the tobacco plant or possible inaccuracy in the GUS enzymatic assay at pH 6.

Because there were several factors to be tested, a fractional factorial design, was used to save time and resources. A fractional factorial design uses only a portion of the factor combinations required for a complete factorial design and significant factors can be revealed when the portion of factor combinations is chosen correctly, just as with a

complete factorial design. To choose the correct factor combinations, a two-level,  $\frac{1}{2}$  fractional factorial experiment was produced using the design generator:

$$F = ABCD \quad (1)$$

A, B, C, D, and F represent the five factors used in the design: PEG molecular weight, PEG concentration, phase-forming salt concentration, ionic strength, and pH, respectively. The design generator determines the alias structure for the experiments. When the estimate for the effect of a factor includes the influence of one or more other factors, the effects are aliased. For example, if the estimate of factor A actually estimates  $A + BCDF$ , then the main effect A is aliased with the 4-way interaction BCDF [8]. A main effect is defined as the effect of a factor on the response variable which is measured without regard to other factors in the analysis. Aliasing causes no difficulty in estimating effects when the effects are aliased with higher order interactions. This is because higher order interactions are typically non-existent or insignificant [8]. Assuming higher order interactions are insignificant is what enables a fractional factorial design to be used to determine significant factors. The design determined by the generating equation for a  $\frac{1}{2}$  fractional factorial experiment with five factors has a resolution of five. Resolution five designs can estimate all main effects and two-way interaction effects. Two-way interaction effects occur when the effect of one factor is dependent on the level of one other factor [8]. The alias structure is shown in Table 3.1. Any design generator can be replaced by its negative counterpart to produce an equivalent, but different fractional factorial design [8].

Table 3.1 Alias structure generated for a  $\frac{1}{2}$  fractional factorial design with 5 factors.

Alias Structure
I + ABCDF
A + BCDF
B + ACDF
C + ABDF
D + ABCF
F + ABCD
AB + CDF
AC + BDF
AD + BCF
AF + BCD
BC + ADF
BD + ACF
BF + ACD
CD + ABF
CF + ABD
DF + ABC

Screening experiments were performed for tobacco and GUS separately in PEG/ammonium sulfate, PEG/potassium phosphate, and PEG/sodium sulfate systems. A two level,  $\frac{1}{2}$  fractional factorial design was used to study the effect of these five factors on the separation of GUS from transgenic tobacco extract. The response variable in all the screening experiments was the partition coefficient. The factor combinations were generated by JMP statistical software (SAS Institute Inc. Cary, NC, USA) for a two-level, resolution five,  $\frac{1}{2}$  fractional factorial experiment. The factor combinations are shown in Table 3.2. JMP and SAS statistical software (SAS Institute Inc. Cary, NC, USA) were used in the screening experiments for data analysis.

Table 3.2 Factor combinations for PEG/Ammonium Sulfate, PEG/Potassium Phosphate, and PEG/Sodium Sulfate systems used to study partitioning of native tobacco proteins and GUS.

System Order	PEG MW	PEG Conc. (%w/w)	Salt Conc. (%w/w)	Ionic Strength (M)	pH
1	8000	10	13	0.1	7
2	3400	15	13	0.1	7
3	3400	10	18	0.1	7
4	8000	15	18	0.1	7
5	3400	10	13	1.2	7
6	8000	15	13	1.2	7
7	8000	10	18	1.2	7
8	3400	15	18	1.2	7
9	3400	10	13	0.1	8
10	8000	15	13	0.1	8
11	8000	10	18	0.1	8
12	3400	15	18	0.1	8
13	8000	10	13	1.2	8
14	3400	15	13	1.2	8
15	3400	10	18	1.2	8
16	8000	15	18	1.2	8

### 3.5.3 Response Surface Method Design

After the significant factors were determined, those factors were used in a response surface method (RSM) design to increase rGUS recovery and purification in the PEG phase from transgenic tobacco. RSM designs can accomplish many different objectives including hitting a target, maximizing or minimizing a response, reducing variation, making a process robust, or seeking a combination of these objectives [8]. The objective for these experiments is to maximize the response. The response variable used in all RSM experiments is the selectivity of the rGUS. By maximizing the selectivity of rGUS, rGUS recovery and purification from transgenic tobacco should be increased as well. Maximizing the selectivity of rGUS can be achieved by experimenting with multiple inputs to achieve a better output [8]. A RSM experiment is able to produce a set of optimized factor levels which will generate the maximum selectivity by adjusting the

levels of the two significant factors. A central composite design (CCD) was selected as the RSM to maximize the selectivity of rGUS. A CCD contains an imbedded factorial or fractional factorial design with center points and star points that allow estimation of curvature [8]. A CCD always contains twice as many star points as there are factors in the design. If the distance from the center of the design space to a factorial point is  $\pm 1$  for each factor, the distance from the center of the design space to a star point is  $\pm\alpha$  with  $|\alpha| > 1$  [8]. Therefore, the star points represent new low and high values for each factor in the design. The value of  $\alpha$  depends on certain properties of the selected CCD and on the number of factors involved. Also, the number of center point runs required in the design depends on the type of CCD selected for the design.

A face-centered CCD was preferred as the RSM for maximizing the selectivity of rGUS. Face-centered CCDs do not require using star points outside the low and high levels (i.e.  $|\alpha| = 1$ ). Using this type of CCD is required in this instance because the factor levels for PEG concentration were selected to ensure that a two-phase system would be formed and exceeding those bounds would be detrimental. A two-level CCD with center points cannot estimate individual pure quadratic effects, but it can detect them effectively assuming the design is properly constructed. A face-centered CCD can provide reasonably high quality predictions over the entire design space but give poor precision for estimating pure quadratic coefficients due to the reduced design space used. Five center point runs are required for a face-centered CCD with two factors and are used to provide a measure of process stability and inherent variability. The number of center points required for the CCD ensures that the design has uniform precision [8].

The significant factors determined by the two-level  $\frac{1}{2}$  fractional factorial screening design for partitioning of GUS and native tobacco proteins in a PEG/potassium phosphate system were used in a RSM design to maximize the response variable. Transgenic tobacco was used in all RSM experiments. The factor combinations used in the RSM experiments were generated by JMP statistical software (SAS Institute Inc. Cary, NC, USA) for a face-centered CCD experiment. The phase conditions are shown in Table 3.3. JMP statistical software (SAS Institute Inc. Cary, NC, USA) was used in the screening experiments for data analysis.



Table 3.3 Factor combinations for PEG/Potassium Phosphate system used to study the partitioning of rGUS in transgenic tobacco extract.

System Order	PEG MW	PEG Conc. (%w/w)	KPi Conc. (%w/w)	Ionic Strength (M)	pH
1	3400	10	18	0.1	8
2	3400	10	18	0.65	8
3	3400	10	18	1.2	8
4	3400	12.5	18	0.1	8
5	3400	12.5	18	0.65	8
6	3400	12.5	18	0.65	8
7	3400	12.5	18	0.65	8
8	3400	12.5	18	0.65	8
9	3400	12.5	18	0.65	8
10	3400	12.5	18	1.2	8
11	3400	15	18	0.1	8
12	3400	15	18	0.65	8
13	3400	15	18	1.2	8

## References

1. Ni, M. D. Cui, J. Einstein, S. Narasimhulu, C. E. Vergara, and S. B. Gelvin (1995) *Strength and tissue specificity of chimeric promoters derived from the octopine and mannopine synthase genes*. Plant J. 7, 661–676.
2. Neve, M.D., H.V. Houdt, H.-M. Bruyins, M.V. Montagu, A. Depicker (1998) *Recombinant proteins from plants: production and isolation of clinically useful compounds*. C. Cunningham, A.J.R. Porter (Editors) Humana Press, Totawa, NJ, pg 203.
3. Murashige, T., F. Skoog, (1962) *A revised medium for rapid growth and bioassays with tobacco tissue cultures*. Physiol Plant 15, 473-497.
4. Bio-Rad Laboratories (2008) *Bio-Rad Protein Assay* 1-24.
5. Aich S., L.T. Delbaere, R. Chen, (2001) *Continuous spectrophotometric assay for  $\beta$ -glucuronidase*. BioTechniques 30 pg 846.
6. Jefferson, R.A. and K. J. Wilson (1991)  *$\beta$ -Glucuronidase as a Gene Fusion Marker in Agricultural Molecular Biology* in *Plant Molecular Biology Manual*. S. Gelvin, R. Schilperoort, and D.P. Verma (Editors) Kluwer Academic Publishers, The Netherlands 1-33.
7. Holler, C. and C. Zhang (2008) *Purification of an acidic recombinant protein from transgenic tobacco*. Biotechnol Bioeng 99(4), 902-9.
8. NIST/SEMATECH e-Handbook of Statistical Methods (2008) <http://www.itl.nist.gov/div898/handbook/>
9. Zaslavsky, B. Y. (1995) *Aqueous Two-Phase Partitioning*. Marcel Dekker, Inc., New York 75–84, 221–285, 447–498, 643–645, 648–649, 660–662.
10. Albertsson, P. A. (1986) *Partition of Cell Particles and Macromolecules*, 2<sup>nd</sup> Edition. John Wiley & Sons Inc., New York 261-287.
11. Snyder, S. M., K. D. Cole, and D. C. Sziag (1992) *Phase Compositions, Viscosities, and Densities for Aqueous Two-Phase Systems Composed of Polyethylene Glycol and Various Salts at 25 °C*. J Chem Eng Data 37, 268-274.
12. Gu, Z., and C. E. Glatz (2007) *Aqueous Two-Phase Extraction for Protein Recovery from Corn Extracts*. J Chromatogr B 845(1), 38-50.

13. Balasubramaniam, D., C. Wilkinson, K. Van Cott, and C. Zhang (2003) *Tobacco Protein Separation by Aqueous Two-Phase Extraction*. *J Chromatogr A* 989(1) 119-129.
14. Zhang, C., F. Medina-Bolivarb, S. Buswell, and C. L. Cramer (2005) *Purification and Stabilization of Ricin B from Tobacco Hairy Root Culture Medium by Aqueous Two-Phase Extraction*. *J Biotechnol* 117(1), 39-48.

## Chapter 4

### Results and Discussion

#### 4.1 Analysis of Screening Experiments

Three systems, PEG/ammonium sulfate, PEG/potassium phosphate, and PEG/sodium sulfate, were studied to determine an appropriate aqueous two-phase system for purifying rGUS from native tobacco proteins. Preliminary experiments and a literature review demonstrated that rGUS partitions to the top, PEG-rich phase in the three PEG/salt systems. Therefore, purification was accomplished by partitioning rGUS into the top PEG phase and the tobacco proteins to the bottom salt phase. The best PEG/salt system will have the largest partition coefficient for GUS ( $> 1$ ) and the lowest partition coefficient for native tobacco proteins ( $< 1$ ). In all screening experiments, non-recombinant GUS and non-transgenic tobacco were used separately to characterize their partitioning in PEG/salt system. To meet this goal, screening experiments consisting of a two-level, resolution five,  $\frac{1}{2}$  fractional factorial design were performed to determine which salt would provide the best separation of GUS from tobacco proteins, and to help elucidate which factors were significant in GUS and native tobacco protein partitioning in ATPE. A fractional factorial design selects only a portion of the factor combinations required for a complete factorial design to save time and resources by assuming that higher order interactions are insignificant. Significant factors can be revealed when the portion of factor combinations is chosen correctly, just as with a complete factorial design. Five factors important in protein partitioning were screened, including PEG molecular weight, PEG concentration, phase-forming salt concentration, ionic strength, and system pH. Low and high levels for each of these factors were 3400 and 8000 MW, 10 % (w/w) and 15 % (w/w), 13 % (w/w) and 18 % (w/w), 0.1 M and 1.2 M, and pH 7 and pH 8, respectively. The response variable in the experiments was the partition coefficient for GUS or native tobacco proteins and the response was maximized or minimized accordingly. Three replicates of the fractional factorial screening experiments were conducted for each PEG/salt system.

It was determined that the PEG/ammonium sulfate system and PEG/sodium sulfate system did not partition native tobacco proteins to the bottom phase as efficiently

as the PEG/potassium phosphate system. As shown in Table 4.1, 4.2, 4.3, and 4.4, both systems favored partitioning of the tobacco proteins and GUS to the top phase, resulting in large partition coefficients ( $> 1$ ) for both, which is undesirable. The average  $K_{\text{Tobacco}}$  was  $6.3 \pm 1.7$  and the average  $K_{\text{GUS}}$  was  $216.4 \pm 20.0$  for the PEG/ammonium sulfate system and the average  $K_{\text{Tobacco}}$  was  $22.6 \pm 5.5$  and the average  $K_{\text{GUS}}$  was  $81.6 \pm 10.4$  for the PEG/sodium sulfate system. The partition coefficients and the standard deviations for native tobacco proteins and for GUS in the PEG/potassium phosphate system are shown in Table 4.5 and 4.6. In this system, the partition coefficients for GUS are large, average  $K_{\text{GUS}}$  was  $221.0 \pm 17.9$ , and the partition coefficients for tobacco proteins are small, average  $K_{\text{Tobacco}}$  was  $0.56 \pm 0.21$ , which is desirable. The broad range of  $K_{\text{GUS}}$  and  $K_{\text{Tobacco}}$  values in all of the aqueous two-phase systems is due to the planned variation in the factor combinations [8] and indicates that some systems are more capable of partitioning GUS to the top phase and tobacco proteins to the bottom phase. Due to the overall high  $K_{\text{GUS}}$  and low  $K_{\text{Tobacco}}$  values obtained, PEG/potassium phosphate was selected as the best aqueous two-phase system for purification of rGUS from transgenic tobacco. Full statistical analysis is reported for the PEG/potassium phosphate system. The statistical analysis consisted of standard least squares analysis by SAS (Version 9.1). In a PEG/potassium phosphate ATPE, the PEG concentration significantly affected the partitioning of GUS ( $P = 0.011$ ), and the PEG concentration and ionic strength of the system significantly affected the partitioning of native tobacco proteins ( $P = 0.006$  and  $P = 0.001$ , respectively) as shown by analysis of variance (ANOVA) at a significance level of 0.05. ANOVA is detailed in Table 4.7 for GUS and in Table 4.8 for tobacco proteins.

Table 4.1 Average partition coefficients (K) and standard deviations for native tobacco proteins in the PEG/ammonium sulfate system.

System Order	PEG MW	PEG Conc. (% w/w)	Salt Conc. (% w/w)	Ionic Strength (M)	pH	K <sub>Tobacco</sub>	Standard Deviation
1	8000	10	13	0.1	7	2.8	0.51
2	3400	15	13	0.1	7	8.8	1.02
3	3400	10	18	0.1	7	3.7	0.78
4	8000	15	18	0.1	7	4.2	1.01
5	3400	10	13	1.2	7	1.6	1.09
6	8000	15	13	1.2	7	9.9	2.31
7	8000	10	18	1.2	7	3.4	0.63
8	3400	15	18	1.2	7	16.7	4.84
9	3400	10	13	0.1	8	1.6	1.22
10	8000	15	13	0.1	8	12.9	2.67
11	8000	10	18	0.1	8	4.1	1.65
12	3400	15	18	0.1	8	10.1	3.99
13	8000	10	13	1.2	8	1.7	0.65
14	3400	15	13	1.2	8	3.9	0.81
15	3400	10	18	1.2	8	3.5	1.23
16	8000	15	18	1.2	8	11.5	2.52

Table 4.2 Average partition coefficients (K) and standard deviations for GUS in the PEG/ammonium sulfate system.

System Order	PEG MW	PEG Conc. (% w/w)	Salt Conc. (% w/w)	Ionic Strength (M)	pH	K <sub>GUS</sub>	Standard Deviation
1	8000	10	13	0.1	7	214.3	17.68
2	3400	15	13	0.1	7	164.5	15.06
3	3400	10	18	0.1	7	442.1	33.64
4	8000	15	18	0.1	7	2.5	0.72
5	3400	10	13	1.2	7	294.8	22.61
6	8000	15	13	1.2	7	140.8	26.01
7	8000	10	18	1.2	7	322.3	36.64
8	3400	15	18	1.2	7	219.2	27.01
9	3400	10	13	0.1	8	249.5	8.16
10	8000	15	13	0.1	8	176.6	13.92
11	8000	10	18	0.1	8	293.5	24.88
12	3400	15	18	0.1	8	108.8	4.97
13	8000	10	13	1.2	8	372.3	29.06
14	3400	15	13	1.2	8	85.3	12.64
15	3400	10	18	1.2	8	289.1	37.7
16	8000	15	18	1.2	8	87	9.07

Table 4.3 Average partition coefficients (K) and standard deviations for native tobacco proteins in the PEG/sodium sulfate system.

System Order	PEG MW	PEG Conc. (% w/w)	Salt Conc. (% w/w)	Ionic Strength (M)	pH	K <sub>Tobacco</sub>	Standard Deviation
1	8000	10	13	0.1	7	44.9	8.37
2	3400	15	13	0.1	7	6.7	2.39
3	3400	10	18	0.1	7	7.5	1.4
4	8000	15	18	0.1	7	12.4	3.62
5	3400	10	13	1.2	7	4.8	1.22
6	8000	15	13	1.2	7	6.6	0.98
7	8000	10	18	1.2	7	1.4	0.87
8	3400	15	18	1.2	7	17	4.04
9	3400	10	13	0.1	8	13.3	2.93
10	8000	15	13	0.1	8	19.1	6.44
11	8000	10	18	0.1	8	112.9	16.81
12	3400	15	18	0.1	8	4.6	2.84
13	8000	10	13	1.2	8	16.2	1.39
14	3400	15	13	1.2	8	5.4	3.08
15	3400	10	18	1.2	8	32.5	11.43
16	8000	15	18	1.2	8	57	19.97

Table 4.4 Average partition coefficients (K) and standard deviations for GUS in the PEG/sodium sulfate system.

System Order	PEG MW	PEG Conc. (% w/w)	Salt Conc. (% w/w)	Ionic Strength (M)	pH	K <sub>GUS</sub>	Standard Deviation
1	8000	10	13	0.1	7	41.1	8.92
2	3400	15	13	0.1	7	137.5	20.73
3	3400	10	18	0.1	7	24.8	6.99
4	8000	15	18	0.1	7	244	29.4
5	3400	10	13	1.2	7	14.6	4.51
6	8000	15	13	1.2	7	75.1	9.27
7	8000	10	18	1.2	7	10	2.21
8	3400	15	18	1.2	7	10.6	0.89
9	3400	10	13	0.1	8	226.4	22.37
10	8000	15	13	0.1	8	59.2	8.09
11	8000	10	18	0.1	8	230.2	18.76
12	3400	15	18	0.1	8	82.9	10.26
13	8000	10	13	1.2	8	38.9	5.02
14	3400	15	13	1.2	8	13.3	1.95
15	3400	10	18	1.2	8	52.1	7.39
16	8000	15	18	1.2	8	44.2	9.38

Table 4.5 Average partition coefficients (K) and standard deviations for native tobacco proteins in the PEG/potassium phosphate system.

System Order	PEG MW	PEG Conc. (% w/w)	Salt Conc. (% w/w)	Ionic Strength (M)	pH	K <sub>Tobacco</sub>	Standard Deviation
1	8000	10	13	0.1	7	0.31	0.08
2	3400	15	13	0.1	7	0.37	0.12
3	3400	10	18	0.1	7	0.81	0.34
4	8000	15	18	0.1	7	0.35	0.16
5	3400	10	13	1.2	7	0.91	0.23
6	8000	15	13	1.2	7	0.57	0.21
7	8000	10	18	1.2	7	0.78	0.36
8	3400	15	18	1.2	7	0.55	0.2
9	3400	10	13	0.1	8	0.36	0.04
10	8000	15	13	0.1	8	0.46	0.2
11	8000	10	18	0.1	8	0.41	0.18
12	3400	15	18	0.1	8	0.52	0.14
13	8000	10	13	1.2	8	0.82	0.22
14	3400	15	13	1.2	8	0.71	0.26
15	3400	10	18	1.2	8	0.74	0.15
16	8000	15	18	1.2	8	0.23	0.12

Table 4.6 Average partition coefficients (K) and standard deviations for GUS in the PEG/potassium phosphate system.

System Order	PEG MW	PEG Conc. (% w/w)	Salt Conc. (% w/w)	Ionic Strength (M)	pH	K <sub>GUS</sub>	Standard Deviation
1	8000	10	13	0.1	7	1.3	0.74
2	3400	15	13	0.1	7	358.8	19.23
3	3400	10	18	0.1	7	326.9	23.83
4	8000	15	18	0.1	7	169.5	24.61
5	3400	10	13	1.2	7	135.2	25.35
6	8000	15	13	1.2	7	62.2	5.48
7	8000	10	18	1.2	7	168.7	12.16
8	3400	15	18	1.2	7	152.6	17.55
9	3400	10	13	0.1	8	367.7	24.03
10	8000	15	13	0.1	8	226.8	26.13
11	8000	10	18	0.1	8	257.7	24.18
12	3400	15	18	0.1	8	335.5	23.4
13	8000	10	13	1.2	8	140	16.15
14	3400	15	13	1.2	8	318.1	19.77
15	3400	10	18	1.2	8	411.6	15.11
16	8000	15	18	1.2	8	103.2	8.98



Table 4.7 ANOVA table for two-level, 1/2 fractional factorial screening experiments for GUS in PEG/potassium phosphate aqueous two-phase system. A is PEG molecular weight, B is PEG concentration, C is potassium phosphate concentration, D is ionic strength, and F is pH. The factor shown in bold, PEG concentration significantly affected GUS partitioning at a significance level of 0.05.

Source	DF	Sum of Squares	F Ratio	P Value
A	1	37641.6	1.03	0.326
<b>B</b>	<b>1</b>	<b>301944.7</b>	<b>8.25</b>	<b>0.011</b>
C	1	17077.7	0.47	0.504
D	1	73176.7	2.00	0.176
F	1	154424.2	4.22	0.057
A*B	1	91959.3	2.51	0.132
A*C	1	142510.1	3.89	0.066
A*D	1	41824.0	1.14	0.301
A*F	1	11814.0	0.32	0.578
B*C	1	26814.5	0.73	0.405
B*D	1	11300.2	0.31	0.586
B*F	1	67216.4	1.84	0.194
C*D	1	27176.8	0.74	0.402
C*F	1	56042.6	1.53	0.234
D*F	1	26224.9	0.72	0.410

Table 4.8 ANOVA table for two-level, 1/2 fractional factorial screening experiments for native tobacco proteins in PEG/potassium phosphate aqueous two-phase system. A is PEG molecular weight, B is PEG concentration, C is potassium phosphate concentration, D is ionic strength, and F is pH. The factors shown in bold, PEG concentration and ionic strength, significantly affected protein partitioning at a significance level of 0.05.

Source	DF	Sum of Squares	F Ratio	P Value
A	1	0.09768	3.77	0.070
<b>B</b>	<b>1</b>	<b>0.37101</b>	<b>8.75</b>	<b>0.006</b>
C	1	0.00317	0.07	0.786
<b>D</b>	<b>1</b>	<b>0.54145</b>	<b>12.77</b>	<b>0.001</b>
F	1	0.02679	0.63	0.433
A*B	1	0.00031	0.01	0.932
A*C	1	0.08201	1.93	0.174
A*D	1	0.00003	0.00	0.980
A*F	1	0.00864	0.20	0.655
B*C	1	0.12322	2.91	0.098
B*D	1	0.12350	4.12	0.053
B*F	1	0.05824	1.37	0.250
C*D	1	0.08315	4.07	0.068
C*F	1	0.11741	2.77	0.106
D*F	1	0.00946	0.22	0.640

## 4.2 Response Surface Method Experiments

### 4.2.1 Interphase Studies

When first performing the initial response surface method (RSM) experiments with transgenic tobacco, there was an unexpected outcome that was not detected from the screening experiments using GUS and the non-transgenic tobacco separately. After adding the transgenic tobacco extract to the aqueous two-phase system and letting the system equilibrate and the phases separate, the interphase between the top PEG phase and the bottom potassium phosphate phase appeared slightly cloudy and green. Also, all the native tobacco protein and rGUS could not be recovered in the top and bottom phases unless the interphase was mixed with the top phase or bottom phase. Both of these events lead to the conclusion that the green plant pigments (chlorophyll), and more importantly, some of the rGUS and tobacco proteins were partitioning to the interphase. This phenomenon may be accredited to the salting-out effects, which cause proteins to partition to the interphase. For PEG/salt systems, salting-out effects occur with increasing tie-line length, shifting proteins from the salt phase into the PEG-rich phase [8]. Then, if protein solubility in the PEG phase is low, the proteins tend to partition to the interphase [8]. (However, solubility and salting-out limits are dependent on the properties of individual proteins and a differential response is expected when a combination of proteins is used [9, 10]. For example, partition curves were studied for  $\beta$ -galactosidase and four other samples of  $\beta$ -galactosidase fused to small proteins in PEG/potassium phosphate systems, and it was determined that at longer tie-line lengths, the proteins were less soluble in the phosphate phase and were adsorbed at the interphase [11].) For all the conditions studied in the RSM design, the average recovery for rGUS activity in the top phase, interphase, and bottom phase was 70%  $\pm$ 9%, 9%  $\pm$ 2%, and 14%  $\pm$ 4% respectively. The average recovery of native tobacco proteins in the top phase, interphase, and bottom phase was 8%  $\pm$ 3%, 75%  $\pm$ 11%, and 16%  $\pm$ 5% respectively. Protein denaturation at the interphase was most likely not occurring as an increase in rGUS activity could be measured once the interphase was mixed with either the top or bottom phase. Because an average of only 9% of the rGUS was lost at the interphase, the removal of 75% of the native tobacco proteins at the interphase was considered desirable and no further research was completed at shorter tie-lines to eliminate the partitioning of the proteins to the

interphase. Protein partitioning to the interphase suggests that instead of an aqueous two-phase system, an aqueous three-phase system is being used. While this is not incorrect, the convention in the published literature states that a system with only two phase-forming components is an aqueous two-phase system. This convention will be used for the remainder of this research.

For all RSM experiments, the interphase was not mixed with either the top phase or the bottom phase when determining partition coefficients and selectivity values. Transgenic tobacco protein and rGUS partitioning in the RSM experiments differed from the non-transgenic tobacco protein and non-recombinant GUS partitioning in the screening experiments. In the screening experiments, there was no GUS partitioning to the interphase and hardly any partitioning to the bottom phase, as indicated by the high  $K_{\text{GUS}}$  obtained (data shown in Table 4.3). In the RSM experiments, rGUS partitioning to the interphase occurred as well as partitioning to the bottom phase. This caused  $K_{\text{rGUS}}$  values to be lower than the  $K_{\text{GUS}}$  values in the screening experiments. Also, the majority of the native tobacco proteins partitioned to the bottom phase in the screening experiments, shown by the low  $K_{\text{Tobacco}}$  values obtained (data shown in Table 4.3). In the RSM experiments, the majority of the transgenic tobacco proteins partitioned to the interphase and was consequently excluded in the partition coefficient calculations, resulting in  $K_{\text{Tobacco}}$  values that were higher than in the screening experiments. It is important to point out that even though the majority of rGUS was partitioned to the top phase just as the screening experiments predicted with non-recombinant GUS, the major differences in partitioning of the tobacco proteins suggest that studying GUS and tobacco proteins separately in the PEG/potassium phosphate aqueous two-phase system may not be the most accurate way to determine their partitioning behavior when the ultimate goal is to use transgenic tobacco. The differences in the partitioning behavior of rGUS and transgenic tobacco proteins could be due to slight differences in the physico-chemical characteristics between transgenic tobacco proteins and non-transgenic tobacco proteins and rGUS and non-recombinant GUS. Further experimentation would be needed to determine the exact reason the proteins partitioned so differently. Nevertheless, in order to use ATPE for recombinant protein separation from tobacco, transgenic tobacco, if possible, should be used in the process development stage.

#### *4.2.2 Response Surface Method Analysis*

Once the significant factors that affect the response of interest are known, a number of additional objectives can be accomplished by using an RSM design. These objectives include hitting a target, maximizing or minimizing a response, reducing variation, making a process robust, or seeking a combination of these objectives [8]. The significant factors determined by the two-level  $\frac{1}{2}$  fractional factorial screening design for partitioning of GUS and native tobacco proteins in a PEG/potassium phosphate system were used in a RSM design to maximize the response variable. The response variable used in all RSM experiments is the selectivity of the rGUS and transgenic tobacco was used in all RSM experiments. Increasing rGUS recovery and purification from transgenic tobacco by maximizing the selectivity of rGUS is the overall desired outcome. Maximizing the selectivity of rGUS can be achieved by experimenting with multiple inputs to achieve a better output [8]. By adjusting the levels of the two significant factors, PEG concentration and ionic strength, a RSM experiment is able to produce a set of optimized factor levels which will generate the maximum selectivity. A face-centered CCD was selected as the RSM to maximize the selectivity of rGUS to ensure that the high and low factor levels were not exceeded. Five center point runs are required for this type of design to guarantee uniform precision.

The PEG MW, potassium phosphate concentrations, and pH were determined to be insignificant ( $P > 0.05$ ) in the partitioning of GUS in the PEG/potassium phosphate during the screening experiments. Therefore, the levels of these factors were selected based on which level, high or low, consistently resulted in the highest partition coefficients for GUS during the screening experiments. The levels selected were 3400 for PEG molecular weight, 18% (w/w) for potassium phosphate concentration, and pH 8. Six replicates of the entire RSM experiment were conducted. The average selectivity values and standard deviations obtained during RSM experiments are shown in Table 4.9. The selectivities were all above 1, which is desirable for rGUS partitioning. However, there is a large variation for some of the factor combinations, most notably in the selectivity values for a few of the center points. The variation could be due to unanticipated errors introduced during experimentation. The total system weight was only 2 mL and small

errors in measuring the correct amount of PEG, potassium phosphate, and NaCl could have occurred and negatively affected the reproducibility of rGUS selectivity. Also, for each RSM experiment, a different transgenic plant was used. Each plant had a different level of rGUS expression, varying from approximately 100 to 380 units of rGUS activity per milligram of total soluble protein. The differences in the expression levels may have caused the variation in rGUS selectivity. In addition, there could have been inaccuracies associated with the rGUS activity assay or the Bio-Rad assay when protein activity and concentration were low in either of the phases. Statistical analysis was performed by JMP (Version 7) using the selectivity values in Table 4.9. JMP offers standard response surface methodology that helps find the optimal response within the specified ranges of the factors [12]. A response surface design is capable of fitting a second order prediction equation for the response. The mathematical model equation used is as follows

$$Y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2 + b_{11}x^2 + b_{22}x^2 \quad (1)$$

Standard least squares regression analysis was conducted to fit the response function with the experimental data. The value of  $R^2$  for the model ( $R^2=0.42$ ) showed that 42% of the variability in the response could be explained by the model, as follows:

$$Y = 78.3 + 19.7x_1 + 3.38x_2 + 2.48x_1x_2 - 27.8x^2 - 8.51x^2 \quad (2)$$

The large percentage of variability not explained by the model is not unexpected due to the large variation in selectivity for the center points and some of the other factor combinations. In order to check the statistical significance of the second-order model equation, ANOVA was performed and the data is shown in Table 4.10. The P-value obtained ( $P = 0.481$ ) at a significance level of 0.05 indicates that the response surface model is not better than the sample mean (mean selectivity = 61.6) when predicting selectivity of rGUS. Again, this result is expected due to the  $R^2$  value of 0.418. Nevertheless, the signs of  $b_{11}$  and  $b_{22}$  were negative as shown in Equation (2); therefore the parabola will open downward and have a maximum point. The response surface plot generated by JMP is shown in Figure 4.1. The maximum selectivity was obtained at a PEG concentration of 13.4 % (w/w) and an ionic strength of 0.789 M. The predicted selectivity at these factor levels was 82.4.

Table 4.9 Average selectivities ( $\alpha$ ) and standard deviations for rGUS in PEG/potassium phosphate system used in statistical analysis for RSM experiments.

System Order	PEG MW	PEG Conc. (% w/w)	KPi Conc. (% w/w)	Ionic Strength (M)	pH	$\alpha$	Standard Deviation
1	3400	10	18	0.1	8	41.5	4.43
2	3400	10	18	0.65	8	20.8	3.78
3	3400	10	18	1.2	8	13.1	3.53
4	3400	12.5	18	0.1	8	50.3	4.42
5	3400	12.5	18	0.65	8	63.4	5.97
6	3400	12.5	18	0.65	8	48.7	11.9
7	3400	12.5	18	0.65	8	79.4	27.6
8	3400	12.5	18	0.65	8	112.3	73.1
9	3400	12.5	18	0.65	8	59.9	8.96
10	3400	12.5	18	1.2	8	117.3	79.8
11	3400	15	18	0.1	8	51.9	5.97
12	3400	15	18	0.65	8	108.2	48.9
13	3400	15	18	1.2	8	33.5	7.13

Table 4.10 ANOVA table for face-centered CCD model using PEG/potassium phosphate aqueous two-phase system.

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Model	5	5738.783	1147.76	1.0006	0.4810
Error	7	8029.169	1147.02		
C. Total	12	13767.952			

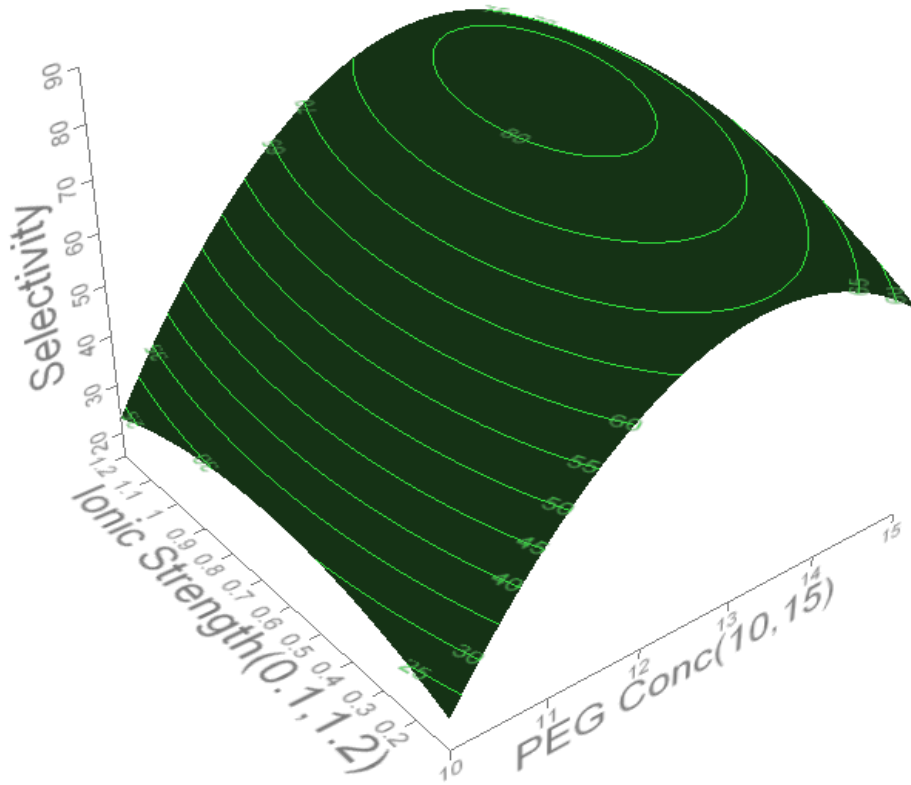


Figure 4.1 Response surface plot for a PEG/potassium phosphate aqueous two-phase system.

### 4.3 Optimum Aqueous Two-Phase System Confirmation

#### 4.3.1 Confirmation of Response Surface Method Results

Six experiments were carried out at the factor levels determined by the RSM experiments to confirm the predicted model and selectivity. The overall system weight was kept at 2 g and the amount of transgenic tobacco extract added remained at 50  $\mu$ L, just as in the screening experiments and the RSM experiments. The average selectivity obtained was  $70.8 \pm 6$ . This is lower than the predicted selectivity of 82.4 and is most likely a result of model inaccuracy. As stated previously, model inaccuracy could be due to unanticipated errors introduced during experimentation. The average recovery for rGUS activity obtained for the top phase, interphase, and bottom phase was  $74\% \pm 8\%$ ,  $10\% \pm 2\%$ , and  $12\% \pm 3\%$  respectively. The average tobacco protein recovery for the top

phase, interphase, and bottom phase was 3%  $\pm$ 1%, 69%  $\pm$ 10%, and 25%  $\pm$ 6% respectively. The average purification factor was 20  $\pm$ 4.

Due to the minimal optimization performed, the maximum selectivity value may not be the absolute maximum [8]. Instead, the maximum selectivity value most likely lies within certain boundaries of the design space around the predicted maximum selectivity [8]. To improve the model retrospectively, further statistical analysis was performed using selectivity values in which the outliers had been removed. Selectivity values that were outside the range of  $\pm$ 2 standard deviations around the mean were eliminated. The average selectivity values and standard deviations with the outliers removed are shown in Table 4.11. Standard response surface methodology was performed by JMP to fit the second order prediction equation for the response. The value of  $R^2$  for the model ( $R^2=0.71$ ) illustrates that 71% of the variability in the response could be explained by the model, as follows:

$$Y = 63.7 + 14.6x_1 - 5.15x_2 + 2.5x_1x_2 - 18.0x_1^2 - 9.01x_2^2 \quad (3)$$

This model explains more of the variability in selectivity values than the model developed previously. To check the statistical significance of the second-order model equation, ANOVA was performed and the data is shown in Table 4.12. The P-value obtained ( $P = 0.0731$ ) at a significance level of 0.05 indicates that the sample mean is only marginally better than the response surface model (mean selectivity = 51.2) when predicting selectivity of rGUS. The signs of  $b_{11}$  and  $b_{22}$  were negative as shown in Equation (3) meaning the parabola will open downward and have a maximum point. The predicted maximum selectivity was 67.2 and can be obtained at a PEG concentration of 13.5 % (w/w) and an ionic strength of 0.52 M. The maximum selectivity predicted using this response surface model is very close to the selectivity obtained during the first RSM analysis.



Table 4.11 Average selectivities ( $\alpha$ ) and standard deviations for rGUS in PEG/potassium phosphate system used in revised statistical analysis for RSM experiments.

System Order	PEG MW	PEG Conc. (% w/w)	KPi Conc. (% w/w)	Ionic Strength (M)	pH	$\alpha$	Standard Deviation
1	3400	10	18	0.1	8	41.5	4.43
2	3400	10	18	0.65	8	20.8	3.78
3	3400	10	18	1.2	8	13.1	3.53
4	3400	12.5	18	0.1	8	50.3	4.42
5	3400	12.5	18	0.65	8	63.4	5.97
6	3400	12.5	18	0.65	8	52.9	5.64
7	3400	12.5	18	0.65	8	68.3	6.02
8	3400	12.5	18	0.65	8	67.5	8.40
9	3400	12.5	18	0.65	8	59.9	8.96
10	3400	12.5	18	1.2	8	66.4	9.46
11	3400	15	18	0.1	8	51.9	5.97
12	3400	15	18	0.65	8	77.9	7.53
13	3400	15	18	1.2	8	33.5	7.13

Table 4.12 ANOVA table for face-centered CCD experiments in PEG/potassium phosphate aqueous two-phase system.

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Model	5	3171.3312	634.266	3.3514	0.0731
Error	7	1324.7857	189.255		
C. Total	12	4496.1169			

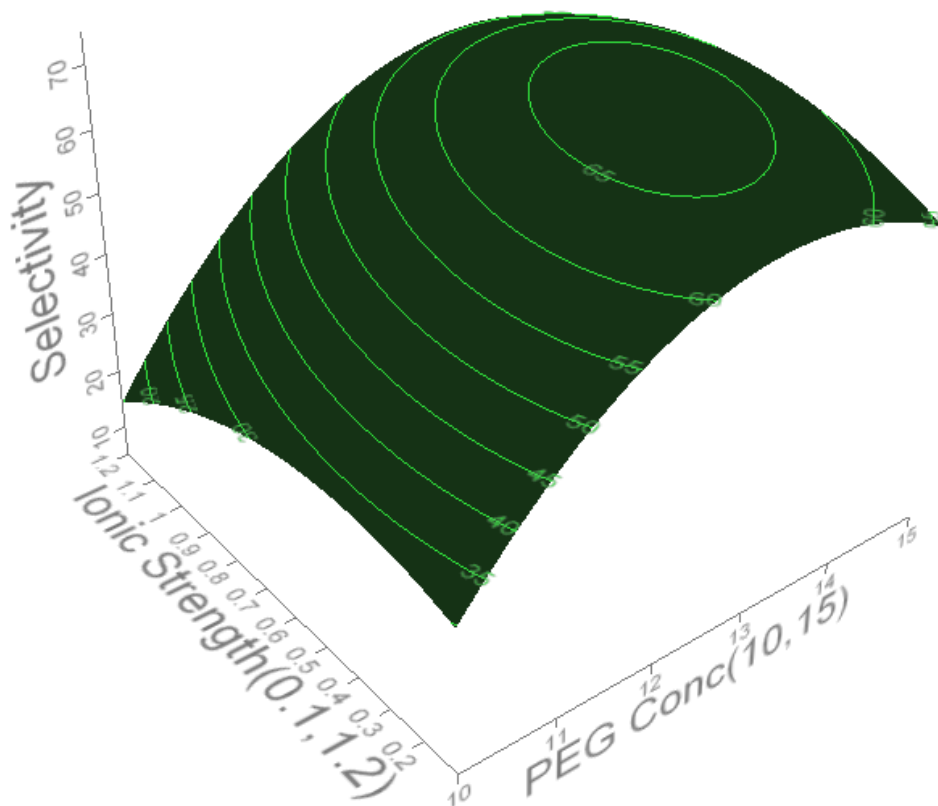


Figure 4.2 Revised response surface plot for a PEG/potassium phosphate aqueous two-phase system.

#### 4.3.2 Phase Ratio Experiments

In theory, selectivity relative to individual proteins should not change along a particular tie-line. Therefore, target protein recovery and the purification factor can be manipulated by adjusting the phase ratio. The recovery of the target protein will decrease as the phase ratio decreases while the purification factor will increase with a decrease in phase ratio. A published tie-line was not available for the PEG and potassium phosphate compositions of 13.4% (w/w) PEG and 18% (w/w) potassium phosphate. Consequently, the tie-line was estimated using a published phase diagram under the assumption that all the tie-lines are parallel [1]. Figure 4.3 shows the published tie lines [1] available for a PEG/potassium phosphate aqueous two-phase system and the estimated tie-line generated

from the optimized PEG and potassium phosphate compositions. An experiment was performed to observe the effect of manipulating the phase ratio on rGUS recovery and purification factor. The conditions determined by the RSM statistical analysis gave a phase ratio of 0.379. Two experiments were conducted at phase ratios above a phase ratio of 0.379 and three experiments were conducted below a phase ratio of 0.379. Table 4.13 shows the phase ratios used, the concentrations of PEG and potassium phosphate, the recoveries for rGUS in the top phase, and the purification factors obtained in the experiment. The recovery of rGUS improved as the phase ratio increased and the purification factor increased as the phase ratio decreased, as expected. However, the increase in both recoveries was only a few percent while the purification factor decreased to from 20 to 8. The purification factor increased from 20 to 29 when the phase ratio decreased to 0.064, but the recovery suffered and decreased from 74% to 43%. The increase in recovery was not dramatic enough to warrant reducing the degree of purification. Also, the increase in purification factor was not enough to warrant the 31% loss of rGUS recovery. Consequently, the aqueous two-phase system developed through the screening and RSM experiments, which had a phase ratio of 0.379, rGUS recovery of 74%, and purification factor of 20 was maintained as the optimum system for rGUS purification from transgenic tobacco.

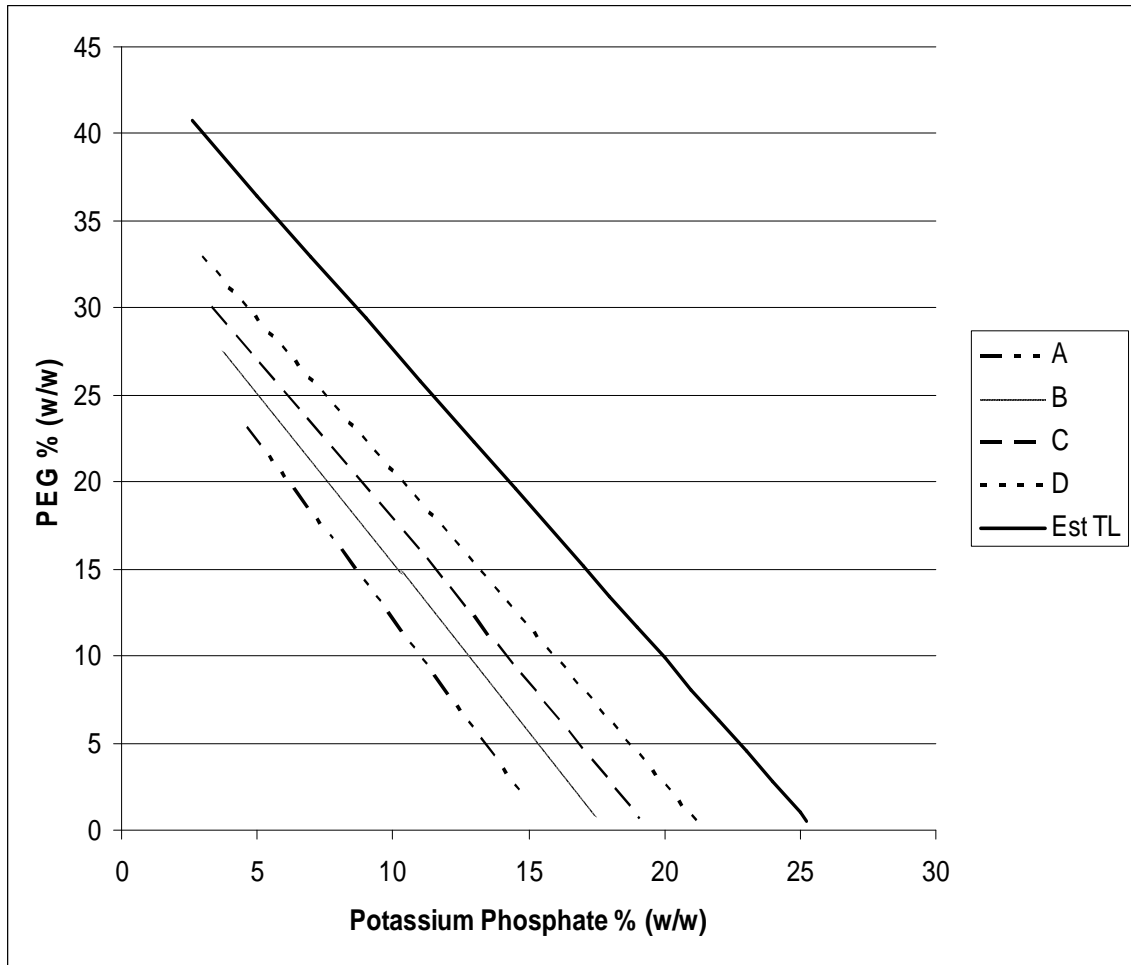


Figure 4.3 Phase diagram for PEG/potassium phosphate at pH 8 with the published tie-lines (A, B, C, and D) and estimated tie-line (Est TL) [1].

Table 4.13 rGUS recovery and purification factor at several different phase ratios. The system generated by the screening design and RSM design is shown in bold.

Phase Ratio	PEG Conc. (% w/w)	Salt Conc. (% w/w)	rGUS Recovery (top phase only)	Purification Factor
0.649	25.8	11	77%	8
0.514	19.6	14.5	76%	11
<b>0.379</b>	<b>13.4</b>	<b>18</b>	<b>74%</b>	<b>20</b>
0.256	9.87	20	65%	22
0.153	6.32	22	51%	27
0.064	2.77	24	43%	29

#### 4.3.3 SDS-PAGE Results

SDS-PAGE was also used to confirm the results of the rGUS purification using a 13.4% (w/w) PEG/18% (w/w) potassium phosphate system. The aqueous two-phase system used was the same as in the RSM confirmation studies. The purification of rGUS from transgenic tobacco extract by ATPE can be seen using SDS-PAGE in Figure 4.4. Lane 2 shows the initial transgenic tobacco extract, while lane 3 shows the top phase and lane 4 shows the top phase with the interphase mixed in. The bottom phase is not shown due to difficulties experienced during SDS-PAGE, which can most likely be attributed to the high potassium phosphate concentration present in the bottom phase. One of the strongest bands shown in the top phase in lane 3, although very faint, is the rGUS band, located at approximately 68 kDa and marked by the arrow in Figure 4.4. There are also faint bands corresponding to where the large and small Rubisco subunits are located at approximately 55 and 15 kDa, and faint bands for other native tobacco proteins. As shown in lane 4, which is the top phase mixed with the interphase, the majority of the interphase contains Rubisco protein, with notable bands occurring at approximately 55 and 15 kDa, which make up a large percentage of native proteins seen in the initial rGUS extract. The interphase also contained some rGUS, which is shown by the increased band intensity, and other native tobacco proteins.

Purification of rGUS from transgenic tobacco has been previously studied using a three-step procedure of polyelectrolyte precipitation using polyethyleneimine, hydrophobic interaction chromatography, and hydroxyapatite chromatography [6]. The precipitation step resulted in recovering more than 90% of the initial rGUS activity, but failed to eliminate Rubisco [6]. Total rGUS recovery from the optimized hydrophobic interaction chromatography step was nearly 80% of the initial rGUS activity but, once again, this step did not provide good separation of rGUS from Rubisco [6]. The third and final step chosen in the purification scheme was hydroxyapatite chromatography. Hydroxyapatite chromatography provided good separation of rGUS and Rubisco, but the final recovery for rGUS was only 40% of the initial rGUS activity [6]. This purification scheme depicts the serious engineering challenge imposed by removing Rubisco during purification of an acidic target protein. While Rubisco can be removed using multiple

processing steps, it is at the expense of rGUS recovery. Using ATPE, it is possible to remove a large percentage of Rubisco in just one processing step while retaining and more than of 70% of initial rGUS activity, as shown by SDS-PAGE. These results indicate that ATPE has the potential to be an excellent initial purification step for separation of acidic recombinant proteins from transgenic tobacco.

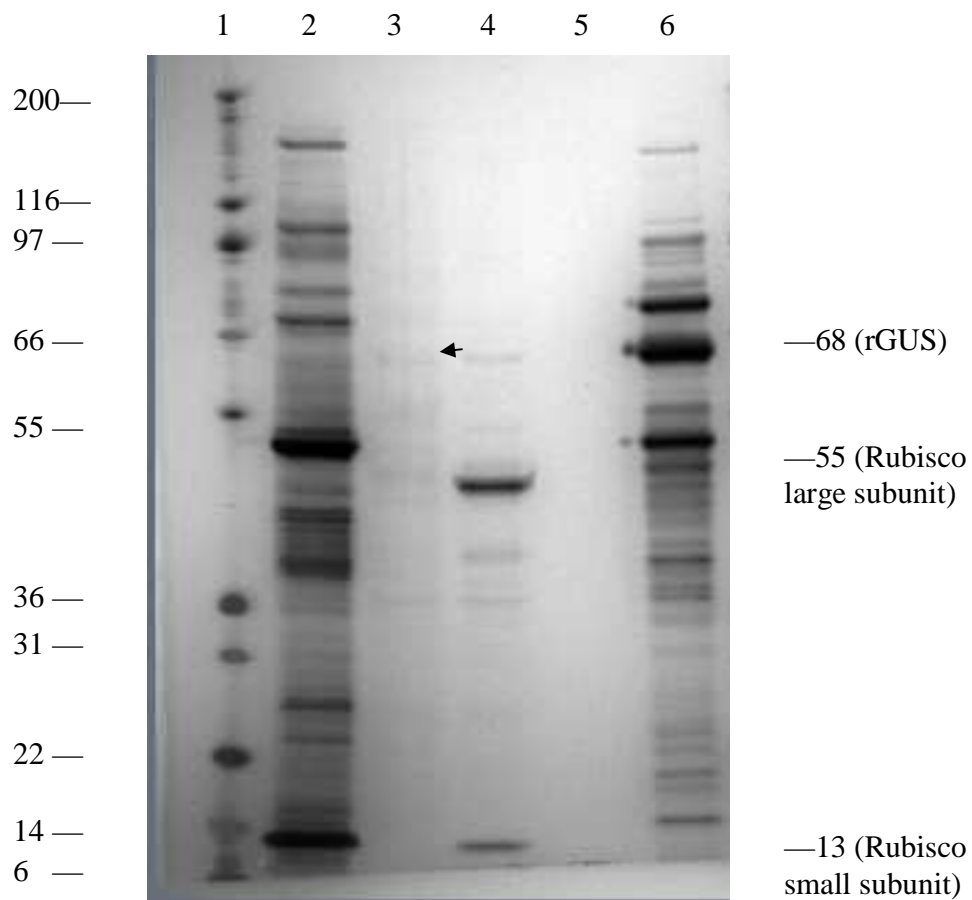


Figure 4.4 SDS-PAGE showing native tobacco proteins and Rubisco protein removal from transgenic tobacco extract and rGUS recovery in the top phase during ATPE. Gel was stained with SilverXpress silver stain. Lane 1: broad-range molecular marker; Lane 2: transgenic tobacco extract, diluted 10X; Lane 3: top phase, diluted 10X, rGUS band marked with arrow; Lane 4: top phase mixed with interphase, diluted 10X; Lane 5: blank; Lane 6: Sigma GUS from *E. coli*. The relative sizes of the molecular marker, rGUS, and the large and small subunits of Rubisco are marked.

## References

1. Zaslavsky, B. Y. (1995) *Aqueous Two-Phase Partitioning*. Marcel Dekker, Inc., New York 75–84, 221–285, 447–498, 643–645, 648–649, 660–662.
2. Snyder, S. M., K. D. Cole, and D. C. Szlag (1992) *Phase Compositions, Viscosities, and Densities for Aqueous Two-Phase Systems Composed of Polyethylene Glycol and Various Salts at 25°C*. J. Chem. Eng. Data 37(2), 268-274.
3. Balasubramaniam, D. C. Wilkinson, K. Van Cott, and C. Zhang (2003) *Tobacco Protein Separation by Aqueous Two-Phase Extraction*. J Chromatogr A 989(1) 119-129.
4. Zhang, C., F. Medina-Bolivar, S. Buswell, and C. L. Cramer (2005) *Purification and Stabilization of Ricin B from Tobacco Hairy Root Culture Medium by Aqueous Two-Phase Extraction*. J Biotech 117(1) 39-48.
5. Gu, Z. C. E. Glatz (2006) *Aqueous Two-Phase Extraction for Protein Recovery from Corn Extracts*. J Chromatogr B 845(11) 38-50.
6. Holler, C. and C. Zhang (2008) *Purification of an acidic recombinant protein from transgenic tobacco*. Biotechnol Bioeng 99(4), 902-9.
7. Gallagher, S.R., in S.R. Gallagher (Editor) (1992) *GUS Protocols: Using the GUS Gene as a Reporter of Gene Expression*. Academic Press, San Diego, CA.
8. NIST/SEMATECH e-Handbook of Statistical Methods (2008)  
<http://www.itl.nist.gov/div898/handbook/>
9. Silva, M. E. and T. T. Franco (2000) *Liquid-Liquid Extraction of Biomolecules in Downstream Processing - A Review Paper*. Braz. J. Chem. Eng. 17(1), ISSN 0104-6632.
10. Kula, M.-R., K. L. Kroner, and H. Hustedt (1982) *Purification of Enzymes by Liquid-Liquid Extraction*. Advances in Biochemical Engineering 24, 73-118.
11. Andrews, B. and J. A. Asenjo (1989) *Protein Purification Methods: A Practical Approach*. IRL Press, Oxford.

12. Köhler, K., A. Veide, and S.-O. Enfors (1991) *Partitioning of  $\beta$ -galactosidase Fusion Proteins in PEG-Phosphate Aqueous Two-Phase Systems*. *Enzyme and Microbial Technology* 13, 204-209.
13. SAS Institute Inc. (2008) *JMP Design of Experiments (DOE)*  
[http://statview.com/software/pdf/103044\\_doe.pdf](http://statview.com/software/pdf/103044_doe.pdf)



## Chapter 5

### Conclusions

Cost-effective recombinant protein production and purification is vital for successful biopharmaceutical manufacturing. Using a plant expression system could reduce biopharmaceutical production costs, but successful implementation hinges on efficient and cost-effective recombinant protein purification. Commercial-scale purification schemes currently revolve around the use of multiple chromatography steps for the purification of biopharmaceutical products. In a biopharmaceutical purification strategy, an initial purification step is used to isolate, concentrate, and stabilize the target product from crude or partially purified extract [1]. Initial purification is often achieved using ion exchange or affinity chromatography. Ion exchange chromatography is less expensive than affinity chromatography, but lacks the high selectivity that affinity chromatography offers. For example, Protein A affinity chromatography is a commonly used initial purification step in monoclonal antibody manufacturing because it can selectively and efficiently bind antibodies in complex solutions and removes >99.5% of impurities in a single step with a high yield of 97% [2]. The largest pitfall with this technique is the high cost of the Protein A resin, which can be up to 10 times as expensive as conventional chromatographic supports [3]. Ion exchange chromatography is a cheaper alternative used for the capture of monoclonal antibodies but is not nearly as effective, removing only 79% of impurities in a single step with a yield of 96% [2]. In addition to high cost, chromatography has other drawbacks including limited throughput, and a complex scale up process.

ATPE has the potential to combine the advantages of both affinity and ion exchange chromatography and become a highly selective, less expensive option for initial purification in large-scale biopharmaceutical applications. In comparison to chromatography purification techniques, ATPE can handle large volumes and a high biomass load [4, 5, 6]. ATPE is anticipated to be easily scalable using PEG/salts systems due to lower viscosity, lower costs, and fast separation time [7, 8, 9]. Scale up of extractions in aqueous two-phase systems has been possible through the use of conventional extraction equipment used in chemical industry [4] and by a proportional

linear increase of the amount of phase-forming components [10, 11, 12]. In practical implementation of large-scale ATPE, the crude sample and the phase-forming components are added together, the pH is adjusted, and mixing occurs to reach equilibrium [5]. Typically, mixing is accomplished using a stirred vessel with baffles or using static mixers in large-scale processes [13]. Due to the low interfacial tension of the system, an extremely low energy input is required [13]. The approach to equilibrium is fast ( $\leq 30$  seconds) when turbulent mixing is provided [9]. After equilibration, the phases separate into a top and bottom phase. Phase separation can be done under unit gravity or by centrifugation to speed up the process. Centrifugation is recommended if the viscosity is high or the density difference between the phases is low [8]. Optimal feed rates up to 3 L/min with residence times of about 16 seconds have been reported using small disk stack separators [8].

Two extraction steps are commonly used to purify proteins from a crude sample when using ATPE. In the first extraction, large particulates and the bulk of nucleic acids, polysaccharides, and expression host proteins are collected in the lower phase while the target protein is partitioned to the top phase. Both phases are separated and the addition of a fresh bottom phase creates a second ATPE in which the target protein is partitioned to the new bottom phase. Ultrafiltration or diafiltration is used in large-scale processing to remove the target protein from the secondary bottom phase, yielding a target protein concentrate and a salt solution as waste or recyclable permeate [5, 6, 8]. In some instances, the secondary bottom phase containing the target protein may be applied directly to a chromatography support [5]. Alternatively, the target proteins may be purified using a single ATPE. The target proteins can be partitioned to the bottom phase in the first extraction step when a partially purified sample with no particulates is used. Target proteins can also be recovered directly from the top polymer-rich phase by ultrafiltration with no need for a second extraction.

ATPE is typically performed in batch mode, but often it may be more cost-effective to use a continuous process [5]. Several publications have shown that ATPE is suitable for continuous operation [14, 15, 16]. Co-current, cross-current, and countercurrent flow are three possibilities for continuous processing [5]. The countercurrent method has been investigated experimentally in pilot scale [15, 16], and is

expected to be the most efficient because of a higher yield and lower consumption of chemicals [14]. Except for the highest value products, implementation of ATPE will require recycling of phase components to lower chemical and waste disposal costs. Recycling of potassium phosphate by crystallization at 6 °C has been described from solid-free process streams [8] and the successful recycling of PEG has been demonstrated [17]. Figure 5.1 shows a general flow diagram for the recovery of PEG and salt from a two-stage ATPE. The salt is recovered from the first and second ATPE while PEG is recovered from the second ATPE [18]. By recycling the phase components, the material balance of the phase-forming components introduced into the process stream can be closed by more than 90% [5], yielding a cost-effective and environmentally acceptable process.

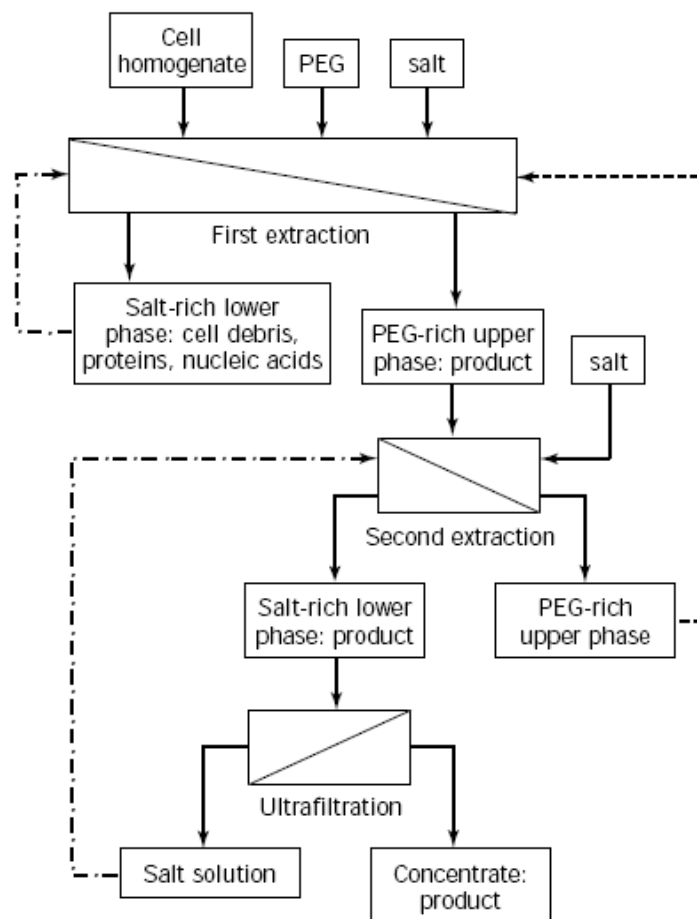


Figure 5.1 General flow diagram for the recovery of PEG and salt from a two-stage ATPE [18].

Despite its advantages over standard chromatography techniques, aqueous two-phase extraction has not yet been applied to large-scale manufacturing of biopharmaceuticals. Due to the complexity of predicting the partitioning of each protein in an aqueous two-phase system, experimentation will be required to select the correct aqueous two-phase system and phase conditions. Process development can often be the rate-limiting step when introducing biological drug candidates into clinical trials [19]. Developing an ATPE method is wholly empirical, and this might increase the time and resources needed for process development.

In this research, an ATPE was developed at a laboratory scale to separate an acidic model protein, rGUS, from transgenic tobacco. The majority of native tobacco proteins are acidic and purification of acidic recombinant proteins expressed in tobacco has been shown to be challenging [20]. Using a fractional factorial design followed by a RSM design, a minimally optimized aqueous two-phase system was determined to consist of PEG with a molecular weight of 3400 and a concentration of 13.4 % (w/w), a potassium phosphate concentration of 18 % (w/w), an ionic strength of 0.789 M, and a system pH of 8. This aqueous two-phase system provided a recovery of 74% of rGUS in the top phase and a removal of 94% of native tobacco proteins in the interphase and the bottom phase. ATPE enabled high levels of recovery and concentration of rGUS while removing large amounts of native tobacco proteins from the initial extract, most notably Rubisco. Further statistical optimization concluded that a system of PEG with a molecular weight of 3400 and a concentration of 13.5% (w/w), a potassium phosphate concentration of 18% (w/w), an ionic strength of 0.52 M, and a system pH of 8 could be more accurate in determining the maximum rGUS selectivity predicted by the statistical model. This aqueous two-phase system has the potential to be applied to other acidic recombinant proteins produced by transgenic tobacco and to reduce time and resources needed during process development.

## References

1. Amersham Biosciences (now known as GE Healthcare) (2002) *Protein Purification Handbook*. 18-1132-29 Edition AC.
2. Follman, D. K. and R. L. Fahrner (2004) *Factorial Screening of Antibody Purification Processes Using Three Chromatography Steps Without Protein A*. *J Chromatogr A* 1024(1-2) 79-85.
3. Shukla, A. A., B. Hubbard, T. Tressel, S. Guhan, and D. Low (2007) *Downstream processing of monoclonal antibodies—Application of platform approaches*. *J Chromatogr B* 848(1) 28-39.
4. Hatti-Kaul, R. (2001) *Aqueous Two-Phase Systems, A General Overview*. *Molecular Biotechnology* 19(3), 269-277.
5. Flickinger, M. C. and S. W. Drew (1999) *Encyclopedia of Bioprocess Technology - Fermentation, Biocatalysis, and Bioseparation*. John Wiley & Sons 2179–2190.
6. Papamichael, N. and H. Hustedt (1994) *Enzyme Recovery by Continuous Crosscurrent Extraction*. *Methods Enzymol.* 228, 573-584.
7. Fauqueux, P.-F., H. Hustedt, and M.-R. Kula (1985) *Phase Equilibration in Agitated Vessels during Extractive Enzyme Recovery*. *J. Chem. Technol. Biotechnol.* 35B, 51-59.
8. Hustedt, H., K. H. Kroner, and M.-R. Kula (1985) *Applications of Phase Partitioning in Biotechnology, in Partition in Aqueous Two-Phase Systems. Theory, Methods, Uses and Applications to Biotechnology*. H. Walter, D.E. Brooks, and D. Fisher (Editors) Academic Press, Orlando, FL 529–587.
9. Hustedt, H., K.-H. Kroner, U. Menge, and M.-R. Kula (1985) *Protein Recovery using Two-Phase Systems*. *Trends Biotechnol.* 3, 139-144.
10. Hart, R.A., P. M. Lester, H. Riefsnyder, J. R. Ogez, and S. E. Builder (1994) *Large Scale, in situ Isolation of Periplasmic IGF-I from E. coli*. *BioTechnology* 12, 1113–1117.
11. Schutte, H., K. H. Kroner, W. Hummel, and M.-R. Kula (1983) *Recent Developments in Separation and Purification of Biomolecules*. *Ann. N.Y. Acad. Sci.* 413, 270–282.

12. Kroner, K .H., H. Schutte, W. Stach, and M.-R. Kula (1982) *Scale up of Formate Dehydrogenase by Partition*. J. Chem. Technol. Biotechnol. 32, 130–137.
13. Minuth, T., J. Thommes, and M.-R. Kula (1995) *Extraction of Cholesterol Oxidase from Nocardia rhodochrous using a Nonionic Surfactant-Based Aqueous Two-Phase System*. J Biotechnol. 38, 151-164.
14. Papamichael, N. and H. Hustedt (1994) *Enzyme Recovery by Continuous Crosscurrent Extraction*. Methods Enzymol. 228, 573-584.
15. Hustedt, H., B. Borner, K. H. Kroner, and N. Papamichael (1987) *Fully Automated Continuous Crosscurrent Extraction of Enzymes in a Two-Stage Plant*. Biotechnol. Tech. 1, 49-54.
16. Hustedt, H. and N. Papamichael (1988) *Automated continuous crosscurrent extraction of proteins*. Enzyme Eng. 9, 135-139.
17. Papamichael, N., B. Borner, and H. Hustedt (1992) *Continuous Aqueous Phase Extraction of Proteins: Automated Processing and Recycling of Process Chemicals*. J. Chem. Technol. Biotechnol. 54, 47-55.
18. Greve, A. and M.-R. Kula (1991) *Recycling of Salts in Partition Protein Extraction Processes*. J. Chem. Technol. Biotechnol. 50, 27-42.
19. Shukla, A. A., B. Hubbard, T. Tressel, S. Guhan, and D. Low (2007) *Downstream processing of monoclonal antibodies—Application of platform approaches*. J Chromatogr B 848(1) 28-39.
20. Holler, C. and C. Zhang (2008) *Purification of an acidic recombinant protein from transgenic tobacco*. Biotechnol Bioeng 99(4), 902-9.

## Chapter 6

### Future Work

Large-scale application of ATPE has been hindered due to the complexity of predicting the protein partitioning in an aqueous two-phase system. Developing an ATPE method is wholly empirical, and an increased amount of time and resources could be needed for process development compared to well-studied chromatography techniques. An attempt to lessen the extent of experimentation needed during process development was made by studying ATPE for purification of an acidic model protein from transgenic tobacco. Study of ATPE was aided by fractional factorial designs and response surface methodology. An ATPE purification step could be potentially applicable for the purification of other acidic recombinant proteins expressed in tobacco. However, the process should be carried out with one or more actual therapeutic acidic recombinant proteins expressed in transgenic tobacco to verify the process. Additional optimization would most likely be needed depending on the protein purification scheme and overall purity required.

Increasing the number of recombinant proteins studied in ATPE will help to reduce the amount of trial and error that is necessary when designing an aqueous two-phase system. Further experimentation with rGUS in aqueous two-phase extractions could reveal the exact physico-chemical properties which define its partitioning behavior. If those properties are known, one might be able to predict the partitioning of other acidic recombinant proteins with similar physico-chemical characteristics. The ability to predict protein partitioning in aqueous two-phase systems would greatly increase the chances for ATPE to be integrated into a biopharmaceutical purification strategy.

As described previously, it is anticipated that both transgenic tobacco production and ATPE are scalable for commercial biopharmaceutical manufacturing. However, additional experiments would need to be performed on a pilot scale to determine the effectiveness of this particular application. While eliminating approximately 70% of the native tobacco proteins at the interphase is convenient at a laboratory scale, it could potentially hinder the process on a large scale. The interfacial tension in aqueous two-phase systems is very low, which results in fast phase separation with a low energy input.

If volume of the aqueous two-phase system and biomass added to the aqueous two-phase system are both increased, the majority of the tobacco proteins will most likely partition to the interphase just as in the laboratory-scale experiments. This might increase the interfacial tension and make interphase transport the rate-limiting step in ATPE. A change in the selectivity of rGUS could occur due to the presence of the interphase. Pilot scale experiments could determine whether or not protein partitioning to the interphase will limit the ATPE on a larger scale. Further experimentation and optimization using the aqueous two-phase system could identify phase conditions which partition the native tobacco proteins to the bottom phase, instead of to the interphase.

Commercial-scale economic evaluation of the overall process would be beneficial in determining the estimated production costs associated with using ATPE for the purification of acidic recombinant proteins from transgenic tobacco. In addition, methods must be developed to process the transgenic waste material collected from each step in the purification process to eliminate contamination issues. With further experimentation, process development, and process optimization, ATPE could prove to be valuable addition to a biopharmaceutical manufacturing scheme.



## **Appendix A**

### **Experimental Conditions**

Appendix A presents the results of control experiments used to study the enzymatic activity of rGUS under the different experimental conditions used throughout this project. The data from three experiments using 13% (w/w) ammonium sulfate, potassium phosphate, and sodium sulfate, 18% (w/w) ammonium sulfate, potassium phosphate, and sodium sulfate, and 10% (w/w) PEG at molecular weights of 3400 and 8000, and 15 % (w/w) PEG at molecular weights of 3400 and 8000 is presented in Table A.1. The weight percentages for these experiments were selected to correspond with the low and high levels used during all other experiments performed throughout this project. A standard t-test was used to determine that all the means from each experimental condition were not statistically significantly different from the control mean at a significance level of 0.05. The means, standard deviation, and p values are listed in Table A.1.

Table A.1 Conditions, mean enzymatic activity, standard deviation, and p values obtained during control experiments.

Condition	Replicate	Activity (U)	Mean Activity (U/mL)	Standard Deviation (U/mL)	P Value
Ammonium Sulfate 13% (w/w)	1	385.88	380.50	5.78	0.433
	2	374.39			
	3	381.22			
Potassium Phosphate 13% (w/w)	1	392.63	382.70	8.93	0.631
	2	375.34			
	3	380.12			
Sodium Sulfate 13% (w/w)	1	392.04	384.22	6.95	0.738
	2	378.74			
	3	381.87			
Ammonium Sulfate 18% (w/w)	1	386.23	380.61	6.84	0.453
	2	372.99			
	3	382.62			
Potassium Phosphate 18% (w/w)	1	382.74	381.54	1.05	0.712
	2	381.08			
	3	380.8			
Sodium Sulfate 18% (w/w)	1	390.67	381.45	8.31	0.532
	2	379.14			
	3	374.55			
PEG 3400 10% (w/w)	1	383.94	382.92	8.97	0.618
	2	373.49			
	3	391.34			
PEG 8000 10% (w/w)	1	386.63	380.78	5.91	0.463
	2	374.81			
	3	380.9			
PEG 3400 15% (w/w)	1	387.32	381.15	6.29	0.501
	2	374.75			
	3	381.39			
PEG 8000 15% (w/w)	1	389.73	382.02	7.66	0.57
	2	374.41			
	3	381.92			
Control	1	382.54	381.14	3.30	N/A
	2	377.38			
	3	383.51			

## Appendix B

### Preliminary Experiments

Appendix B presents the results of initial aqueous two-phase extractions performed using to test the enzymatic activity of rGUS in the range of pH 6 to pH 8. Four identical experiments were performed with a 10% (w/w) PEG/13% (w/w) potassium phosphate aqueous two-phase system at pH 6, pH 7, and pH 8. As shown in Table B.1, the enzymatic activity of rGUS was very hard to quantify at pH 6 compared to pH 7 and pH 8. The enzymatic activity should have been comparable due to the three systems being identical except for the pH of the system. The decision was made after the results of these experiments were analyzed that the pH range used during all aqueous two-phase extractions would be pH 7 to pH 8 where the enzymatic activity of rGUS could be easily quantified and compared.

Table B.1 Results from preliminary experiments performed to test the enzymatic activity of rGUS in an aqueous two-phase system at pH 6, pH 7, and pH 8.

	pH 6 Activity (U/mL)	pH 7 Activity (U/mL)	pH 8 Activity (U/mL)
Top Phase	2.43	27.79	27.32
Bottom Phase	1.19	1.48	2.03