OPTIMIZING SILK PROTEIN PRODUCTION USING AN ENGINEERING APPROACH

A Thesis
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Kalleigh Emmalyn Meester

In Partial Fulfillment of the Requirements
for the Degree of
MASTER OF SCIENCE

Major Program:
Biomedical Engineering

April 2020

Fargo, North Dakota

North Dakota State University
Graduate School

Title

Optimizing Silk Protein Production Using an Engineering Approach

By

Kalleigh Meester

The Supervisory Committee certifies that this ***disquisition*** complies with North Dakota State University's regulations and meets the accepted standards for the degree of

MASTER OF SCIENCE

SUPERVISORY COMMITTEE:

Dr. Amanda Brooks
Chair

Dr. Diane Darland

Dr. Lokesh Narayanan

Approved:

4/23/2020
Date

Dr. Annie Tangpong
Department Chair

ABSTRACT

The acquisition of spider silk is a complex and costly process that restricts its availability. Increasing applications stemming from the biomedical and pharmaceutical sectors is driving the demand higher, necessitating the need for efficient large-scale production. This thesis investigates 1) recombinant protein expression systems, 2) major ampullate gland cell culture techniques for natural silk production, and 3) process optimization of recombinant silk protein expression. Using a process engineering analysis, the current *E.coli* system expression system was found to be a cost-effective and efficient technique for silk production. While a Box-Behnken predictive model was developed to optimize expression conditions based on small-scale *E.coli* expression data, it failed to translate to a larger-scale. Alternatively, the protein secreting cells that line the major ampullate silk gland were isolated and grown in conditions mimicking the native microenvironment, demonstrating a clear impact on growth of the cells and a potential new source of silk.

ACKNOWLEDGEMENTS

A large amount of gratitude is owed to Dr. Amanda Brooks, who presented this opportunity and continuously supported me as I explored the world of engineering and science. Always by leading through example, Dr. Brooks challenged me to approach the unknown with a positive and inquisitive attitude, which I now understand to be a true engineering mindset. To my Fargo family and my Minnesota family: thank you for your continuous advice and encouragement, I couldn't have done this without your support. Lastly, I would like to thank the willingness of my committee members, Dr. Diane Darland and Dr. Lokesh Narayanan, to create a multidisciplinary atmosphere that allowed me to explore my place in the world of research.

All figures and tables in this work are original designs and depictions by the author using the software BioRender and Microsoft PowerPoint (Version 16.35).

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANOVA	Analysis of Variance
BBD	Box-Behnken Design
CCD	Central Composite Design
cDNA	Complementary Deoxyribonucleic Acid
CHO	Chinese Hamster Ovary
CO_2	Carbon Dioxide
DMSO	Dimethyl Sulfoxide
DOE	Design of Experiment
E.coli	Escherichia coli
EDTA	Ethylenediaminetetraacetic Acid
FBS	Fetal Bovine Serum
HCl	Hydrochloric Acid
HEK	Human Embryonic Kidney
IPTG	Isopropyl β- d-1-thiogalactopyranoside
LB Broth	Luria-Bertani Miller Broth
MaSp	Major Ampullate Spidroin
N. clavipes	*Nephila clavipes*
NaOH	Sodium Hydroxide
NMR	Nuclear Magnetic Resonance
OD_{600}	Optical Density
PCR	Polymerase Chain Reaction
PTM	Post-Translational Modifications
RNA	Ribonucleic Acid
RPM	Revolutions Per Minute
RSM	Response Surface Methodology
SX2	Recombinant Silk Protein
TBST	Tris-buffered Saline & Tween20

LIST OF APPENDIX FIGURES

CHAPTER 1: INTRODUCTION

Taking a multi-variate technical approach to a biological process may help demystify the black box of optimized biological processes, such as recombinant protein expression, allowing us to harness the power of biology using engineering concepts. The production and use of recombinant proteins have become commonplace in both academic labs and commercial endeavors to study biochemical processes and produce therapeutics and other research products. The high demand has been facilitated by the availability of commercial systems for protein production and isolation. These commercial and lab specific processes have been used to produce spider silk proteins, which are desirable for their mechanical and biomedical properties. Unfortunately, success and subsequent commercial translation of spider silk proteins has been somewhat hampered by our inability to harvest the raw protein at scale. The desired protein produced in the major ampullate gland of golden orb weaving spiders (*Nephila clavipes*) can be harvested directly from the spun silk (although this form is unmodifiable and very limited), from a variety of recombinant protein production systems, or potentially, from silk gland cell culture. Each of these techniques has disadvantages and is continually being operationally revised through trial and error to produce higher protein yields with modifiable characteristics, such as the length and structure of the protein. Although the concept of revision for optimization key in biological settings, especially when time, cost, and product consistency are crucial to the success of a project, the incremental nature of the process delays the commercial translation of many processes. Applying engineering concepts to the biological process of recombinant protein expression and purification presents a better way. The **objectives** in this thesis are to (1) identify key factors necessary to yield a more optimized recombinant protein expression system, using spider silk as a model and (2) develop a feasible alternative to the production of recombinant spider silk (SX2) protein using cell culture techniques. The objectives work towards the **hypothesis** that the optimization of SX2 protein production can be more rapidly accomplished by applying process engineering, utilizing statistical methods, and analyzing biological variables.

Objective 1: Explore variables such as cost, speed, and yield that play a part in recombinant protein expression on both a small and large-scale.

- **Sub-aim 1.1:** Compile a thorough comparison of recombinant protein expression methods in relation to biological variables and process engineering.
- **Sub-aim 1.2:** Apply manufacturing process techniques to the analysis of recombinant protein expression.

Objective 2: Develop methods that mimic the native gland conditions to culture major ampullate gland cells to produce natural spider silk protein.

- **Sub-aim 2.1:** Examine the effect of continuous and changing pH on the growth of major ampullate cells.
- **Sub-aim 2.1:** Determine the differentiation state of MaSp cells after growth in cell culture.

Objective 3: Determine a set of parameters for high SX2 recombinant protein output in an *E. coli* protein expression system by testing the relationship of IPTG concentration, optical density, and incubation temperature on the protein yield.

- **Sub-aim 3.1:** Set up and conduct a Box-Behnken DOE to test the effect of the variables on each other.
- **Sub-aim 3.2:** Identify large batch and small batch considerations for optimizing variables.

CHAPTER 2: A PROCESS ENGINEERING APPROACH TO THE ANALYSIS OF RECOMBINANT PROTEIN EXPRESSION SYSTEMS

Abstract

The advent of recombinant protein expression ushered in a renaissance for biochemistry and pharmaceutics. As the demand for recombinant protein has increased, the development of versatile, easy to use, commercially available systems have facilitated their widespread application. Commonly used host organisms to express these non-native genes consist of cellular (bacterial, yeast, insect, and mammalian) systems, transgenic plants, and transgenic animals. Based on the constraints of biology, each system provides unique advantages while being plagued by equally unique disadvantages that are a combination of both the host system and inherent properties of the target protein. Regardless of the specific expression system, recombinant protein expression can be analyzed from a process-engineering viewpoint, with the entire system being seen as a manufacturing process, to optimize protein yield based on the desired production scale. Large-scale production systems are generally chosen for high efficiency, low cost, and a high-quality output. Alternatively, small-scale results are not merely results produced in less volume but instead are a fundamentally different process and should be analyzed differently due to a high variance of outputs in research. This review will describe different host systems in the context of small and large-scale production and will analyze the advantages of these production systems from an engineering perspective and discuss current breakthroughs in technology that will advance recombinant protein expression.

Introduction

There are many factors that are considered when selecting a recombinant expression system [1]–[3]. Previously, these factors were analyzed typically from a biological context and as linear advantages/disadvantages. Similarly, when examined from a strictly process engineering perspective, critical considerations are typically overlooked due to the lack of an interdisciplinary understanding. To improve recombinant protein expression, a multidisciplinary mindset is necessary to break through the historical and institutional barriers (i.e., “that’s how I was taught”, etc.) that exist, ultimately limiting progress. Process engineers offer a unique perspective that challenges these perceptions and limitations and can offer unique interdisciplinary solutions to help with streamlining. Taking a process engineering

approach to a biological system can mitigate some of the variability amongst experiments and personal while maximizing the output of a very labor-intensive process.

Manufacturing in Biology

Recombinant protein production requires various types and amounts of protein output. Bench-scale, 10mL-100mL cultures, can be classified as production that is specific to a set of conditions that produces very little output but is done to study the effects of variables or to develop the system. Academic research commonly requires recombinant protein for the study of structure and mechanism of varying proteins, therefore requiring batch-scale sizes of (1L-10L) cultures. Alternatively, scale-up production utilizes large-scale technology, such as bioreactors, to produce cultures of 10,000L or larger. Figure 2.1 compares the methods of all three size scales of production and puts into context the assumed associated costs, along with the change of the variables and the volume of output produced.

Examples of Manufacturing in Biology

Bench-Scale (Job Shop)	(10mL-100mL culture) • Development of expression system • Optimization protein expression
Batch-Scale (Batch)	(1L-10L culture) • Expressing 100mg of recombinant protein for structural study • Development of expression system
Scale-up (Flow)	(10,000L or larger culture) • Expression of insulin for pharmaceutical use

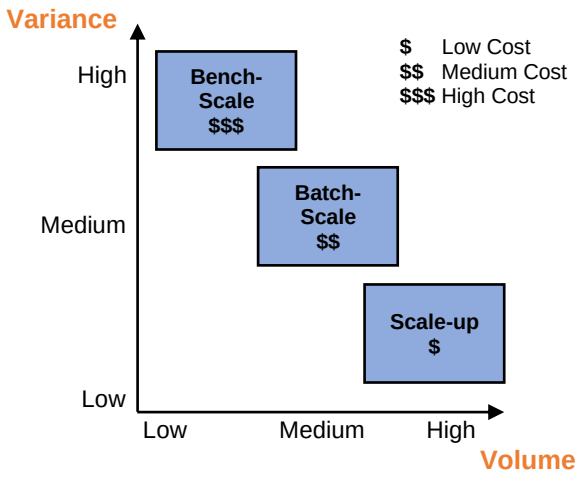

Figure 2.1. Defines bench scale, batch-scale, and scale-up processes in recombinant protein expression systems, with examples of cellular expression batch sizes, while considering the change in variance and volume with the sizes.

Bench-scale

Small-scale production of recombinant protein is crucial in the development of protein production widely used throughout biological, chemical, and biomedical sciences. Tens of thousands of proteins from Eubacteria and Archaea have been collected and purified in the past decade [4]. Conditions for these expression protocols are broad and the exact details must be tailored specifically to a single protein. There are many reasons why bench-scale expressions are crucial in the development of protein production systems for example when developing upstream processing methodology or when determining/optimizing conditions for a specific system. Manufacturing terminology defines this as “job shop” production system or a bench-scale, with a low volume of output and large array of changing variables.

When defining the overall flow of protein production, bench-scale production is an efficient and cost reducing step for the physical development of this type of plasmid containing system. The steps that occur before and during this development, as discussed in this review, are crucial to the production of protein in batch-scale or scale-up methods. The decisions that are made based on biological variables and process parameters play a critical role in the success of the future production [5], [6]. Unfortunately, the complexity of biological systems combined with a “that’s how I was taught” mindset makes it difficult to fully understand the workflow that goes into the beginning stages of protein production.

Batch-Scale

When the correct conditions and upstream processing are determined in a bench-scale setting, recombinant protein production in the lab can be increased to batch-scale, allowing the output (i.e., recombinant protein) to be produced in much larger amounts with less variability.

Although the associated costs with scale-up production are lower than batch-scale due to increased efficiency (Figure 2.1), the initial costs of the large-scale system and complexity of set-up are not feasible in traditional research lab settings. Hence, if the recombinant protein demand is not high enough, the set-up of a mature automation system equipped for large-scale production will not be as cost effective as it is in industry settings. Small-scale bioreactors have been developed to create a more cost-effective option for production at batch-scale that allows for batch size growth of the expression systems but is designed for much smaller yields than scale-up amounts. This incremental production scale

reduces the set-up cost of the system while still providing increased yield [7]. When considering the demand for protein, batch-scale is still a logical and accessible choice for an academic lab to meet the requisite amount. However, to make this method more cost-effective, approaching recombinant protein expression as a process to be optimized can significantly improve the system.

When developing methods to produce optimal amounts of protein, there are many technical approaches that have been developed to assist the creative researcher. Statistical methods offer a useful approach to biological variable optimization to eliminate bias that is unintentionally formed based on familiarity with a system [8]–[10]. These are powerful tools that are becoming more easily accessible and accurate for biological applications. Another approach to batch-scale optimization is a deeper understanding of cellular processes that have limited protein expression methods [11]–[13].

Scale-up

Recombinant protein expression systems that are scaled up are typically done so for products in the pharmaceutical industry such as insulin, growth factors, and glucagon [14]. There is minimal variance in the product as it is produced in large-scale batches, which is a major advantage for quality control output monitoring. As research in biotechnology is expanded, the demand for large-scale recombinant protein products will continue to rise.

Continuous culture methods offer a unique set of advantages when compared to traditional batch-scale methods more commonly seen in large-scale production. Besides the proven cost effectiveness of continuous production, there is potential for other benefits such as steady-state operation, reduced cycle times, and steady-state product quality that add to the allure of continuous flow manufacturing [15]. Although the benefits of continuous flow bioreactors are expanding, the technology to keep up with this development is also necessary. Some common continuous flow lab equipment such as continuously stirred tank reactors and plug flow reactors are now being introduced to handle an increase in protein concentration, productivity, and quality [16], [17].

Significant advancements have been made in industries with a high demand for the identification and production of recombinant proteins. All steps of the recombinant protein production process have benefitted from these developments. Automated methods for rapid cloning [18], multidimensional

chromatography [19], and continuous culture methods [20] are some of a few developing systems that have allowed for a more rapid turnaround time in high quality, reproducible product.

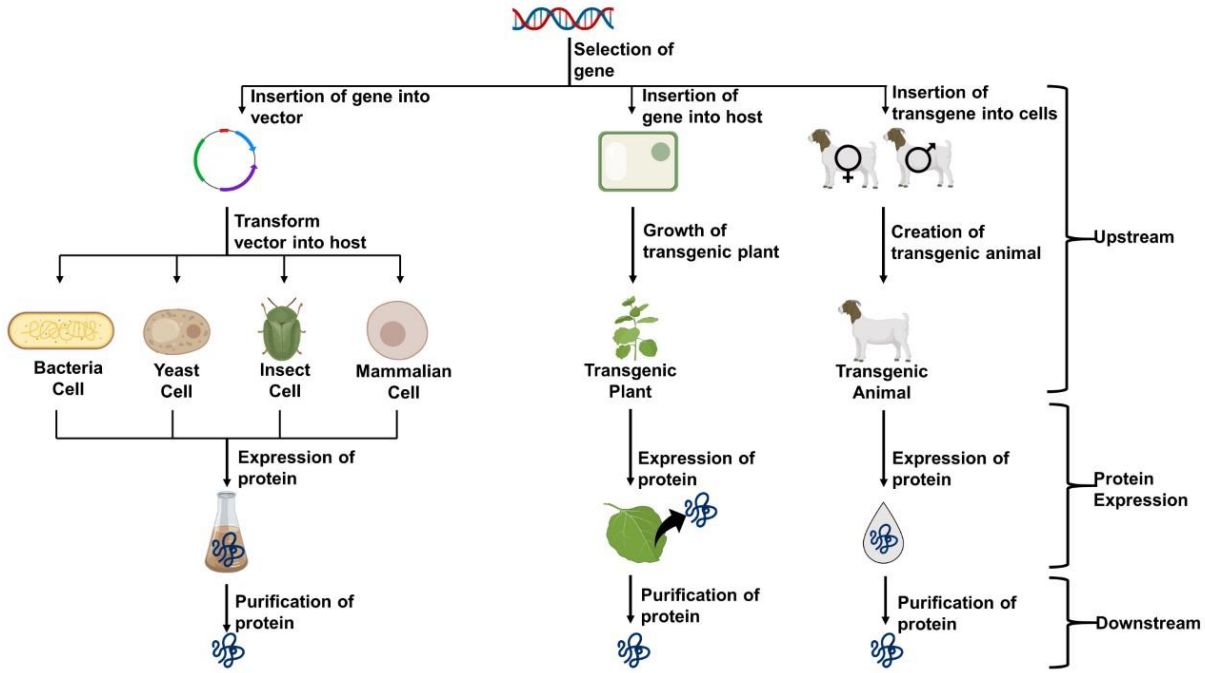

Figure 2.2. A graphical process flow of the upstream, protein expression, and downstream processes of cellular expression systems, transgenic plant expression systems, and transgenic animal expression systems.

Manufacturing Considerations and Variables

The manufacturing breakdown of the system is categorized into (1) the upstream process of selecting all the necessary genetic elements and setting up the expression system, (2) the expressionof the recombinant protein output, and (3) the downstream process of purifying the protein for the desired application. Figure 2.2 displays the flow of these processes as applied to the varying expression systems. Prior to upstream processing, the selection of the host expression system is the critical step that paves the way for the entirety of the recombinant protein production.

When selecting an expression system, there are 5 critical process-related considerations to take into account. Listed are the 5 key variables that can be used to assess and rank expression systems for optimal, application-specific, recombinant protein production:

- Upstream processing
- Cost

- Speed
- Yield
- Downstream processing

Upstream Processing

Upstream processing begins after the selection of the gene of interest and the decision of which expression system to use. Figure 2.2 displays the general differences of the systems' upstream processes. The complexity of the upstream processing varies among systems; however, some considerations that help assess the overall complexity and appropriateness of the system include how developed the gene insertion technology is, the overall accessibility of the system to research and industry, and the characterization of the genetics based on previous work done to create recombinant proteins. The assessment of these individual characteristics, along with primary literature documenting techniques and methods for the host system can provide important insight when comparing and ranking expression systems.

Cost

Initial costs are defined as the costs that are associated with the set-up of a system, and for this review, are factored into the upstream processing complexity. The more complex the upstream processes are for a system, the higher the initial costs of the system are assumed to be. Setting aside the consideration of initial costs, process costs must also be considered and included in the analysis.

Process costs can include many things in a manufacturing setting, but for simplicity they can be considered as process time, process yield, and cost of goods (i.e., materials) necessary for expression of the recombinant protein. The process time of a recombinant protein expression system is a loose consideration; as longer process times equate to a more expensive output but may also increase the yield and effect the downstream utility of the product. As the process time is also considered in the speed analysis, this factor can then be accounted for and embedded in the other considerations. Process yield, although it is intimately connected to other considerations, is a critical variable that cannot be neglected. The higher the yield, and the higher the potential for a scale-up, the more cost effective the system will be, ultimately providing a lower the cost per unit of protein output.

The last factor used for the analysis of the cost variable, is the material cost. Two considerations that can help assess the material cost of expression systems are the stability of the host system and the cost of the materials necessary to support this stability (e.g., media, food, antibiotics, etc.). Finally, a system with built in flexibility may provide more potential for cost improvements in a manufacturing setting. A system that has more flexibility for changes, such as genetic modification, can be optimized to decrease the cost.

Speed

For the purpose of this review the speed of a system is based on the amount of time it takes for recombinant protein expression to occur, after the upstream processing is completed. The amount of time between the beginning stages of growth of the system, whether it be a culture inoculation or the birth of a transgenic animal, to when that system begins producing protein that can be purified (downstream processing) is considered the process speed. With the variability of all systems, there are some generalizations that must occur for the sake of ranking the individual systems. For cellular expression systems the doubling time of the cells is a rough indicator of how long it will take to induce recombinant protein expression. Beyond single cell systems, the rate of organism proliferation and growth can be used as a rough guideline for ranking the systems according to their speed. What sets the systems apart even further is the flexibility that certain systems have to accelerate their rate of proliferation and maturation.

Yield

The yield is assessed by known yield of the host and the feasibility for scale-up. Each system varies based on biological factors that either restrict or enhance the system's ability to produce recombinant protein in scale-up size batches. Yields for whole organism systems are yields restricted simply by the amount of space that is available to grow/house the organism. Others are restricted by inherent biological processes (e.g., genetic inheritance, the protein's protease susceptibility, the native level of protein production, etc.) that make it almost impossible to grow in large batches (~10,000L for culturable systems) without the aid of genetic changes or bioreactors.

Downstream Processing

The assessment of downstream processing relates to the complexity of purifying the recombinant protein from the culture or organism. The first consideration that can help assess this is the technical and

commercial development that has already been done for the system. For example, some developed methods have altered cells to excrete the recombinant protein directly into the growth medium as opposed to inside the cell for ease in purification. This assessment works in conjunction with assessing the utility of the system for the desired application (e.g., a biological therapeutic, antibody development, etc.). Both of these factors shed light on the advancements that have been made for the systems and ease of processing. The toxicity of contaminating host proteins must also be considered as part of the downstream process, such as the use of the recombinant protein in pharmaceutical applications. The purification process accounts for 45-95% of the total cost of the manufacturing of pharmaceutical recombinant proteins [21]. It is crucial for samples to be extremely pure (95% to ~100%), and the large acceptance and use of a system for production in pharmaceuticals makes it easier to benchmark (learn techniques from others in industry), therefore improving the downstream processing.

Application of Recombinant Protein

One consideration that is only known by the selector/designer of the expression system is the desired application of the system. Some variables must be weighted more depending on the application of the recombinant protein produced, meaning that even some systems with very complex upstream processes and slow speed of protein expression are still selected for their economical scale-up potential. Likewise, these same systems will not be selected for the application in an academic lab setting due to a limited budget and small lab space.

Post-Translational Modifications

The biological functions of many recombinant proteins require a system that can post-translationally modify the protein prior to purification. In some proteins, post-translational modifications (PTM) are crucial in protein folding, stability, conformation, etc. PTMs that are commonly considered in recombinant protein expression include glycosylation and phosphorylation [22]. These naturally occurring processes are only present in certain expression systems such as mammalian systems, insect systems, and transgenic animals, due to the availability of enzymes responsible for site-specific modification in eukaryotic systems. If PTM are something that a protein requires for its final function, this must be one of the first considerations when selecting the appropriate expression system. An extensive discussion of PTMs is beyond the scope of this review, but Barber et al. presents a comprehensive discussion and

elaborates on the natural phenomenon of PTMs and precise modification of proteins relative to function and applicability [22].

Expression Systems

Expression Systems

Expression system selection is a crucial decision in the process of producing a recombinant protein and can benefit from the considerations afforded by a process engineering perspective. The use of the 5 critical manufacturing variables mentioned previously along with the ranking of the systems can aid in the selection process. The systems used for production of recombinant proteins are cellular expression systems (bacterial, yeast, insect, and mammalian), transgenic plants, and transgenic animals (Fig. 2.2).

Cell Expression Systems

Bacterial Expression System (Fig 2.3)

The upstream processing of a bacterial expression system consists of the construction of a vector and insertion into the bacterial host [23]. There are many decisions that are made in this step that play a significant role in the downstream processing of the recombinant protein [24]–[26]. These decisions have been streamlined due to well-characterized genetics and molecular tools, creating a simple and easily accessible upstream process. The selection of the host strain is another decision that has been streamlined based on recommendation by the National Institutes of Health, which suggests making the *E.coli* K12 strain and its derivatives the primary system used in recombinant therapeutic production [27]. Some other hosts that are used as alternatives are *Lactoccocus* [28] and *Pseudomonas* [29].

Based on the abundance of developed tools and supplies, the low cost of bacterial expression systems makes them a great host when cost is the highest concern. As speed can also be correlated to cost, the high rate of doubling (i.e. proliferation) along with fast expression time allows processing costs to be minimized. The speed of the process falls mainly on the biological caveats that go along with the individual systems. Although changes made to the biological variables in upstream processing [30] and recombinant protein expression processing [31] can alter the expression speeds of bacterial systems; the overall speed can be correlated to the doubling times for cellular expression systems (bacterial, yeast, insect, and mammalian). *E.coli* in particular has a doubling time of ~20 minutes that creates an

environment for a quick turnaround time from inoculation to induction. Finally, the flexibility of the system due to the abundance of molecular and chemical tools for modification allows for all-around optimization of all variables, further decreasing the cost [32]. Yield is a big factor in batch-scale and scale-up methods of manufacturing and broken down logically the higher the yield the lower the cost associated with the overall process. Many studies are done to increase yield in bacterial systems, thus, decreasing the cost of the overall system [33]. The specific recombinant protein yield from bacteria depends on the genetics of the vector, the protein of interest, and the specific host strain used, but in general, the yield of protein expressed in bacterial systems is known to be high. The difficulties that come with increasing the yield in bacteria fall on insoluble and inactive proteins being co-produced for reasons such as mRNA stability and promoter strength [34]. However, the maturity of this system from the years of advancements has significantly increased the system's ability to not only be scaled-up but also to produce higher protein yields despite the biological challenges [35]–[37]. There are recent advances that challenge commonly used methods that are a great example of assessing a system based on the process design [38], [39].

Downstream processing of bacterial systems such as *E.coli* is improved by methods that have been developed throughout the years. One challenge faced in the downstream purification process is based on where the bacterial cell synthesizes, processed, and stores the expressed recombinant protein. Recombinant protein may end up in either the cytoplasm or exported into the culture medium [40]. Some argue that it is easier to purify recombinant protein that has been secreted as opposed to protein that remains in the cytoplasm; however, process improvements and new biological insights into the process that was once a burden have led to significant improvements and flexibility in terms of available techniques [41], [42]. Another concern with the downstream purification processing is application of recombinant protein to later pharmaceutical/industry uses. Bacteria excrete toxins, such as endotoxin, that require extra steps for removal in the downstream process to ensure recombinant protein of the highest quality and purity.

Lastly, there are some considerations that are not necessarily considered process improvements from a manufacturing standpoint but are significant reasons why bacterial hosts may be suboptimal. Many proteins of higher organisms are glycosylated, which is a required step for proper biological protein functioning [40]. *E.coli* have traditionally been thought to lack the mechanisms for glycosylation, but

recently have been recognized to carry out limited glycosylation on some proteins, a factor that is continuously being synthetically and genetically improved with the discovery of enzymes that perform N-linked glycosylation. The genes of glycosylation enzymes can be co-expressed in *E.coli* with the protein of interest, allowing for the glycosylation of proteins that were once unable to use *E.coli* as a host system [43], [44]. Another common problem is the production of truncated protein products due to the expression system's inability to produce full-length recombinant proteins larger than 70 kDa [45]. This limitation is also being overcome through technical advancements and genetic engineering. Despite the few downfalls of this system, advancements in microbial technology have made it possible to now produce difficult to express products that were once not possible [46], [47].

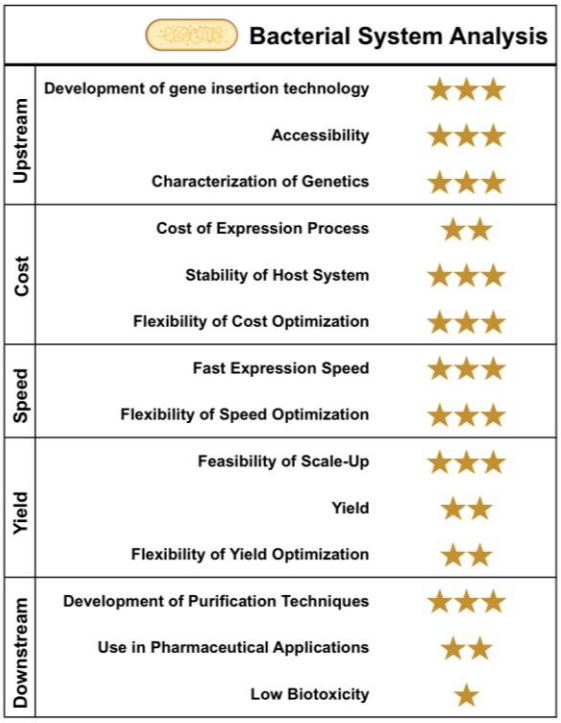

Figure 2.3. A system analysis of process-related variables for a bacterial expression system, ranking variables from one star (worst) to three stars (best).

Yeast Expression System (Fig 2.4)

The upstream processing of a yeast expression system is similar to that of a bacterial expression system. In fact, these two systems are so similar they are commonly considered as interchangeable alternatives for one another [48]. Some common systems used are: *Saccharomyces* [49], *Pichia* [50], and

Scizosacchromyces [51]. Similar to bacteria, the technology for using a yeast system is mature with a variety of vectors being developed [52]–[54]. This creates an easy upstream processing environment that is relatively accessible for all settings in need of recombinant protein.

The cost of a yeast expression system is also comparable to a bacterial expression system. The ease of yeast for batch-scale and scale-up sizes also allows for the improvement in the overall cost of the system, along with the low cost of the growth and maintenance of the system. However, with the yields of yeast being lower due to biological conditions of the system, the cost of the system is affected. Furthermore, the processing takes the same amount of time and energy with relatively lower amounts of resulting product.

The analysis of speed of a yeast system can be correlated to the doubling time of the system, or ~90 minutes, with factors such as host and media impacting the exact time. Although compared to the bacterial speeds the doubling time looks high; however, when compared to all other systems the use of yeast-based systems is still desirable in terms of speed, processing and manufacturing in scale-up batches.

Yields are lower in a yeast expression system when compared to bacteria for multiple reasons that continue to be improved: expression plasmids lost in high yields, recombinant proteins secreted in between membrane spaces instead of into culture, and excessive glycosylation of proteins [55]–[57]. For more information, the reader would be directed toward an extensive review done discussing a framework for reconstructing genome-scale metabolic models in order to increase the yield [58].

Downstream processing has been significantly improved due to engineering that controls the release of the recombinant protein into the culture medium [59]–[61]. This technique avoids the combining of toxic intracellular material with the recombinant protein during a lysing process, thereby improving the time, quality, and cost associated with the downstream purification process. Unlike bacteria, yeast is a single cell system, with a eukaryotic environment capable of post translational processing and modifications such as proteolytic processing, folding, disulfide bond formation, and glycosylation [62]. This capability plays a critical role in why yeast is a strong competitor to other higher eukaryote expression systems.

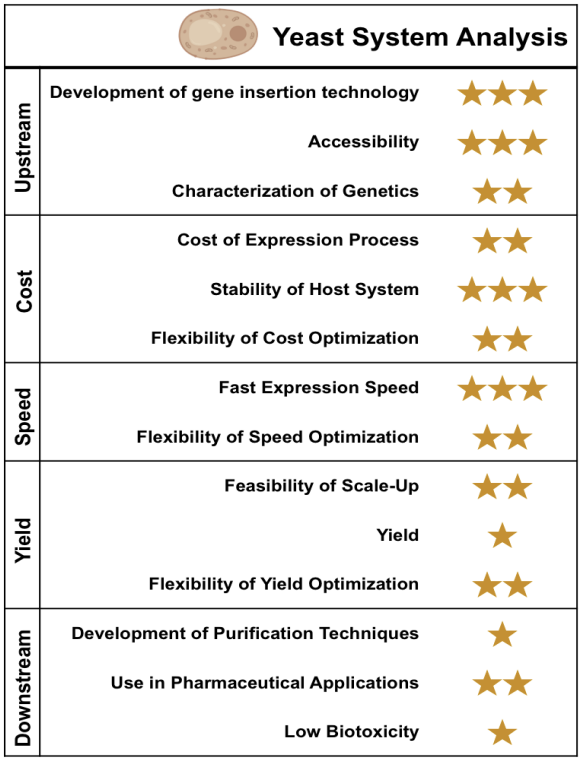

Figure 2.4. A system analysis of process-related variables for a yeast expression system, ranking variables from one star (worst) to three stars (best).

Insect Expression System (Fig 2.5)

Unlike both the bacterial and yeast expression systems, the upstream processing required for an insect host is slightly more complicated. A viral vector, construction of which requires two stages and a double-crossover event [40], is required for recombinant protein expression in an insect system. Viral construction sometimes leads to unwanted results, leading to the development of alternative insect vectors such as shuttle vectors [63]. With recent advancements in genome editing (e.g., CRISPR-Cas9, etc.), allowing for improved capabilities and utility, insect cell systems are making a comeback with their potential for use in industrial production of recombinant proteins [64], [65]. A common cell system that has been used successfully to make eukaryotic proteins is the baculovirus/insect cell system [40]. This is a mature system that has been extensively discussed in several detailed reviews [66], [67].

An insect expression system consists of cells that are far more robust than mammalian cells, and consequently can be grown in simpler and cheaper media and growth conditions. Like all other cell lines,

the process of growing and maintaining cell culture offers a unique opportunity to optimize the manufacturing process, increase the yield, reduce the use of animal's in research and decrease the overall cost of the product.

The doubling times for insect cells depend on the strain and the growth conditions but can range from 18-72 hours. This is a reasonable amount of time especially when considering other options such as transgenic animals and plants, allowing it to be ranked in the middle of the considered systems for this variable.

The yield from insect cells is typically lower than other cellular systems due to the shear sensitivity and high oxygen uptake required, making it challenging to produce in protein in large-scale batches. However, problems like this, which are not unique to insect cell systems, are more manageable with the use of genetic engineering and bioreactors [7], [68].

Downstream processing is similar to all cell expression systems, however, the development of specific techniques for insect cell recombinant proteins is lacking since it is not as widely used in industry. However, insect cells may offer some distinct advantages for protein purification based on their ability to secrete the protein of interest into the culture medium [69]–[71].

As an additional complication, government regulation is not as developed for protein produced from insect cells, so the technical development of the system is lacking when compared to other cell systems (bacterial, yeast, and mammalian). Nevertheless, this system is well suited for the eukaryotic requirements of a recombinant protein based on their ability to glycosylate as required for proper function, similar to yeast and mammalian system [72].

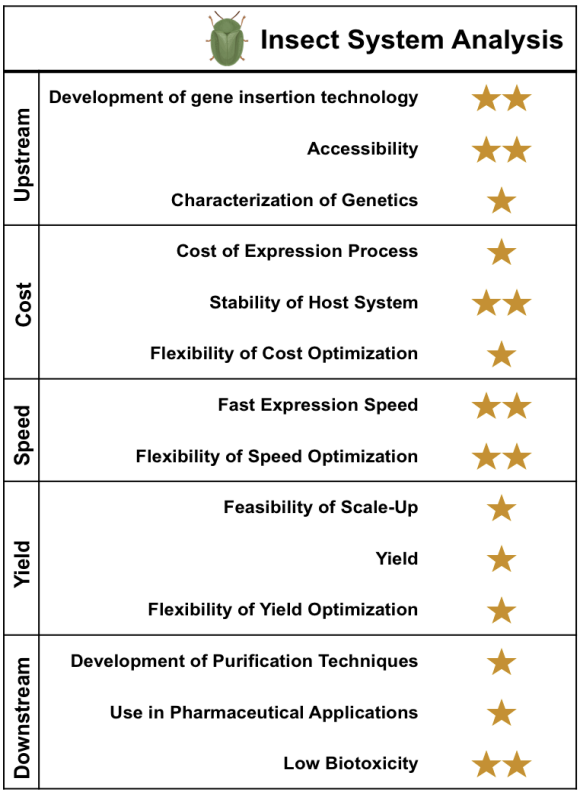

Figure 2.5. A system analysis of process-related variables for an insect expression system, ranking variables from one star (worst) to three stars (best).

Mammalian Expression System (Fig 2.6)

The most well-known mammalian cell systems are Chinese Hamster Ovary cells (CHO) [73], and Human Embryonic Kidney 293 (HEK) cells [74]; however, novel Cap-T cells that are an immortalized human amniocyte line, may be more capable of high-density growth [75]. Notably, CHO cells are the most developed and widely used due to their ease of use and long history in biopharmaceuticals [76].

Mammalian cells are by far the most sensitive of the cell expression systems in terms of stability and required nutrients [40]. The cost of the media plays a huge role in limiting the utility of mammalian cell lines for batch-scale processes. Nevertheless, producing mammalian cells capable of scale-up version to create continuous growth and protein production will lower the cost of the process [77]. Alternatively, serum-free medias have also been developed, eliminating a critical yet costly growth requirement of mammalian systems, ultimately reducing the high cost of the system [78], [79].

It is well known and documented in literature that a large drawback of mammalian expression systems is very low yields due to slow growth and capabilities of the system [80], [81]. The doubling time of mammalian cells depend on the specific host cell line chosen. As a general assessment, the commonly used CHO-K1 cells have a doubling time of ~20 hours.

The growth of mammalian cells is by far the slowest growing of all the cellular expression systems. Higher yield cultures may be established by growth in suspension culture and fed-batch cultures in bioreactors, necessary advancements to meet the high demand for vaccines and therapeutics [82], [83]. The ability to truly scale up mammalian cell culture and compete with other high yield systems is additionally challenging for reasons such as the need for a continuous CO_2 supply and expensive transfection reagents.

Although some recombinant proteins have been successfully expressed using bacteria, yeast, and insect cells; mammalian expression systems are commonly selected due to the need to produce recombinant proteins with the full functional activity due to the system's ability to introduce proper folding, post translational modifications, and product assembly. This is a biological consideration that does not relate to the process breakdown of the systems, but it is still a significant consideration when selecting an expression system. Finally, due to the heavy use of mammalian expression systems in the pharmaceutical industry, the system is considered safe with few immunogenicity concerns. Furthermore, since mammalian cell lines are commonly used there are well-developed and documented methods for more efficient purification.

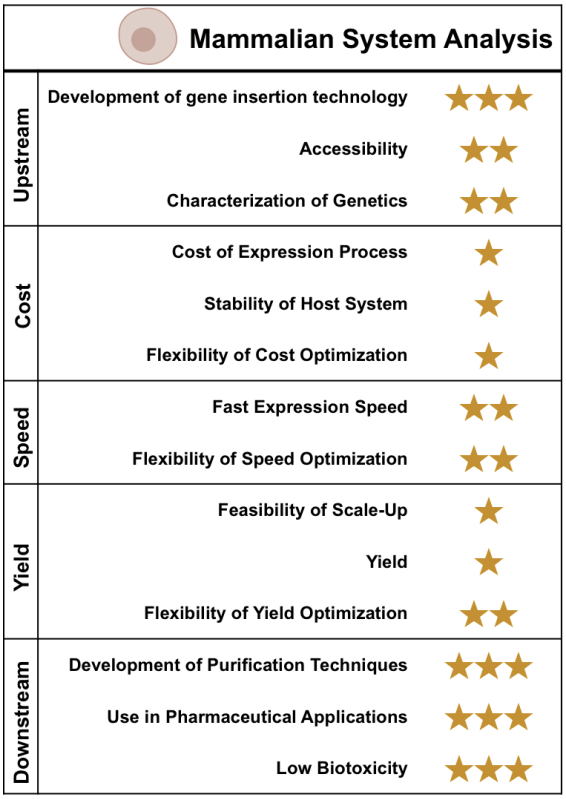

Figure 2.6. A system analysis of process-related variables for a mammalian expression system, ranking variables from one star (worst) to three stars (best).

Cell-free Expression System

Cell-free expression is a rapid and high throughput way to express the proteins using isolated components of the biological machinery of a cell without the use of a living host cell, as discussed in recent reviews of the concept [45], [84]. The growth conditions are not constrained by a cell wall or necessary homeostasis conditions required for cell culturing methods, making it an efficient way to produce cytotoxic proteins. Due to the specificity and novelty of the system, upstream complexity and cost of this system tends to be high [85]. However, beyond these initial parameters, the biological variables can be considered analogous and competitive when compared to the alternative methods. Unfortunately, one of the major pitfalls of cell free expression systems is their inability for commercial, large-scale use due to economic and technical barriers (i.e. high costs and limitations during translation). Nonetheless, cell free systems offer certain competitive advantages to traditional systems when used for bench- and batch-scale research applications. Some systems that are currently being developed include:

cell-free *E. coli* [86], cell-free wheat germ (WGE) [87], cell-free HeLa [88], and cell-free Leishmania-based (LTE) systems [85].

Whole Organism Systems

Plant Expression System (Fig 2.7)

Upstream processing for whole organism systems presents an increasing level of complexity. A significant difficulty arises because multiple interacting genes can control traits [89]. However, the construction of transgenic plant vectors and recombinant plasmids as described in the current literature [90], has led to several successful in transgenic plant lines, including tobacco [91], potato [92], and wheat [93].

Countering challenging upstream processing, the production costs of growing and maintaining transgenic plants are very low compared to the upkeep of cell lines. The concept of expressing recombinant protein in a transgenic plant system is becoming more widely accepted due to the cost of producing proteins in plants being significantly cheaper than in bioreactors [94]. Widespread development in the growth of sustainable agriculture may contribute to further optimization of the production of recombinant protein in plants [95]. While a majority of costs are incurred due to the upstream processing complexity when creating a transgenic plant line, once the transgenic plant has been cultivated the actual processing costs of the recombinant protein expression are considered as low. However, when time is factored into the cost, the speed of the plant expression system is based directly on the speed of growth of the plant, making it one of the slowest systems when compared to rapidly producing cell expression systems with fast doubling times.

Unlike cell culture expressions in which the density of the cell culture as well as the ability of the cells to process the recombinant protein limit the yield, plant expression system yields are not restricted by density and single cell processing abilities. The limit to recombinant protein production in a transgenic plant is based on simple environmental parameters, including nutrients, sunlight, and water. The unlimited potential of this variable gives transgenic plants a significant advantage over other methods of recombinant protein production. Application of these high yield possibilities is significant in the pharmaceutical world where yield is a major concern [96].

Yield without efficient downstream processing provides little advantage. The downstream processing for recombinant protein expression in a plant system ranges based on 1) the host plant system used for the gene insertion and 2) the function of the recombinant protein being expressed. Universal methods are being developed that can be applied to protein-specific systems to increase ease of the purification process [97], [98]. Regardless, when compared to other types of expression systems, expressing a protein in a transgenic plant system allows for the expression of large, toxic, and/or otherwise difficult to express proteins to be produced [99]. Reviews of the system offer unique insight into another non-manufacturing process factor, the stability of the recombinant protein for long periods of time, a significant benefit for commercial scale-up [100], [101]. Government regulation is an important common barrier to consider in whole organism systems. Proteins produced in transgenic organisms often struggle to obtain regulatory approval for their downstream use in humans. This may be a reflection of the fact these developing system have yet to break the barrier of strict quality and regulatory expectations [100]. Luckily, there are plenty of downstream applications, like animal vaccines, with less regulation that can benefit from the use of this system.

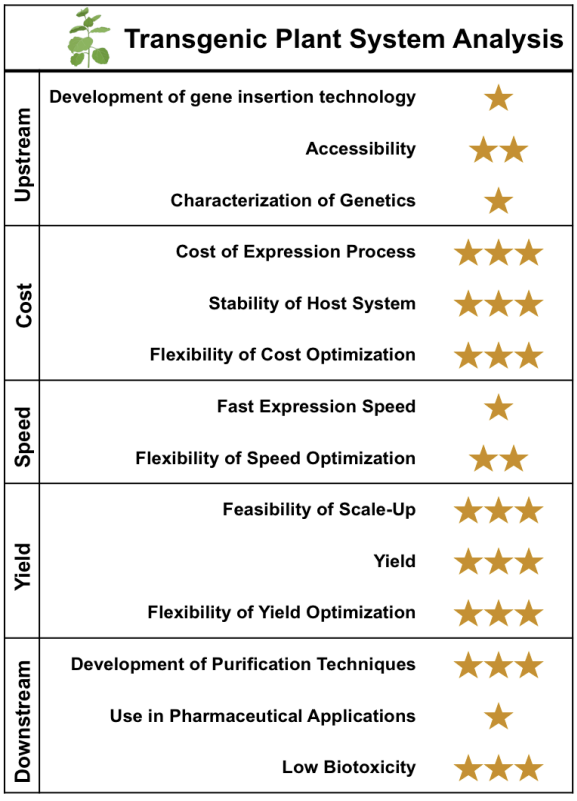

Figure 2.7. A system analysis of process-related variables for a transgenic plant expression system, ranking variables from one star (worst) to three stars (best).

Transgenic Animal Expression System (Fig 2.8)

The upstream processing for transgenic animals consists of genetic engineering where novel genetic information is introduced into the germline of the animal, allowing the gene to be perpetuated through to the progeny [102]. Unfortunately, the amount of time spent simply waiting for the gene to become stable over generations along with screening offspring for stable integration of the gene and expression of the recombinant protein creates a hefty workload for upstream processing. Since mammary glands are generally considered the organ of choice for recombinant protein expression, the discussion in this review will only focus on this particular type of system [103]. Although there are a variety of alternatives, including blood [104], eggs [105], urine [106], and silk [107].

Despite the significant cost of the upstream processing, the cost of the maintaining the transgenic animal line is low when simply considering the husbandry and collection of the recombinant protein. Hence, the cost of the system can be considered from an agriculture analysis. Fortunately, agriculture processing and cost analysis are well developed, streamlined, and understood. As discussed in other focused reviews, transgenic animals offer an economical option for pharmaceutical applications as compared to expensive large-scale bioreactors required for cellular expression systems [108], [109].

Despite a thorough characterization of agricultural transgenic animals and complex upstream processing, which decreases the overall speed of recombinant protein production, the speed of recombinant protein production is also reliant on how fast the mammary glands can ramp up recombinant protein expression. The amount of time a transgenic animal protein expression process takes relies on the animal being used to express the recombinant protein. The life cycles of the varying animals allows for the assessment of the amount of time it takes to get from a newborn transgenic animal to the point of producing the desired recombinant protein. Luckily, the low cost and ease of producing large numbers of the recombinant protein-producing animals makes large-scale a feasible option.

Similar to whole transgenic plant systems, the yield of expressed recombinant protein from a transgenic animal system is not as limited as cellular expression systems. However, yield can be further optimized through the farming process, dependent on the animal host system selected (i.e. goats, rabbits, cows, sheep and pigs). However, it is important to note the yield is not always dependent on the size of the animal, but more on the reproductive specificities such the gestational time, time until milk production,

and milk volume [110]. The details of individual expression systems and the yield of recombinant protein are discussed in more detail in the literature [103]. Ultimately, there is no doubt in the ability of the mammary glands to produce high yields.

Independent of the yield, the downstream processing of soluble recombinant protein in mammary glands is well documented in literature and not a new concept [111], [112]. The ability of upstream process decisions to improve downstream purification is proven to be an effective way to purify high quality recombinant protein [113]. Regulatory approval currently exists for some transgenic animal systems, however it is still far more complicated than cell expression systems to gain approval for use in biopharmaceuticals [109]. With this system being a whole orgamism animal system, the concept of ethical and regulatory issues is a consideration that must be considered on a case-by-case basis when selecting a system.

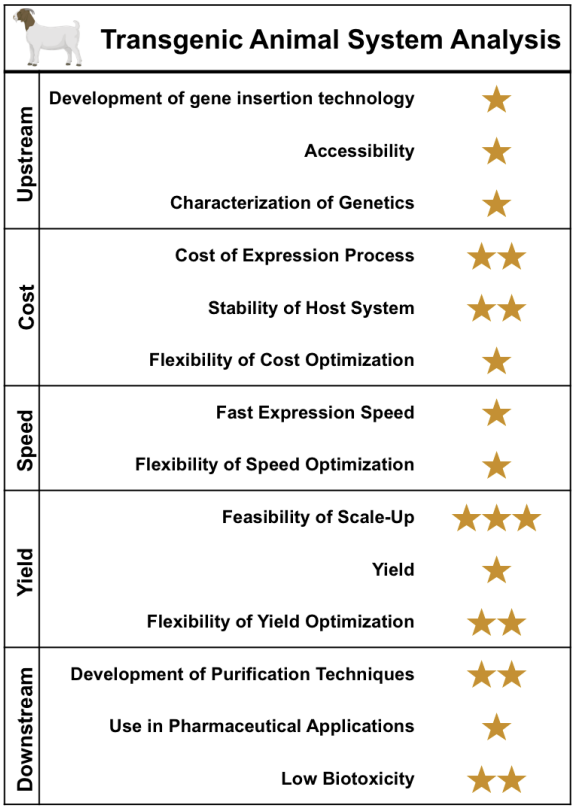

Figure 2.8. A system analysis of process-related variables for a transgenic animal expression system, ranking variables from one star (worst) to three stars (best).

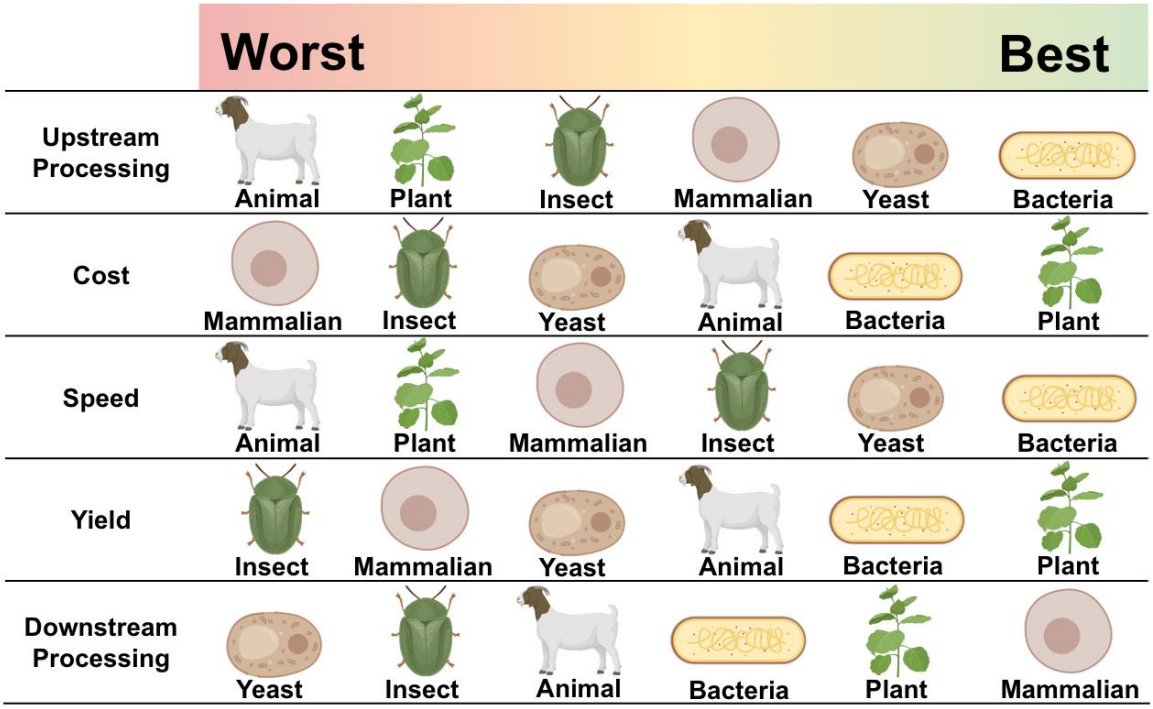

Figure 2.9. An analysis from best to worst of expression systems (cellular expressions, transgenic animals, transgenic plants) based on a manufacturing process breakdown of upstream processes, cost, speed, yield, and downstream processing.

Conclusion

A process engineering approach to comparing and contrasting recombinant protein expression systems is a new way to look at decades-old techniques (Figure 2.9). Prior to this reconstruction of variables; bacterial, mammalian, and yeast expression systems were viewed as the frontrunners for the most “desirable” systems for recombinant protein production. Setting aside the complexity and high costs of upstream processing for transgenic expression systems helps shed new light on a true process comparison between the differing systems. Active research in the world of transgenic plants and animals as well as cell-free systems for large-scale recombinant protein expression is increasingly adding convincing new evidence for the high value they bring. However, when considering the recombinant protein process in its entirety, upstream processing is perhaps the most significant factor for identifying an optimal production system. For small-scale production, cellular-based expression systems are still the frontrunners as the most logical option in terms of yield, cost, and ease of set-up.

References

[1] K. Büssow *et al.*, “Structural genomics of human proteins – target selection and generation of a public catalogue of expression clones,” *Microbial Cell Factories*, vol. 4, no. 1, p. 21, Jul. 2005, doi: 10.1186/1475-2859-4-21.

[2] L. Banci *et al.*, “First steps towards effective methods in exploiting high-throughput technologies for the determination of human protein structures of high biomedical value,” *Acta Cryst D*, vol. 62, no. 10, pp. 1208–1217, Oct. 2006, doi: 10.1107/S0907444906029350.

[3] J. Arnau, C. Lauritzen, G. E. Petersen, and J. Pedersen, “Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins,” *Protein Expression and Purification*, vol. 48, no. 1, pp. 1–13, Jul. 2006, doi: 10.1016/j.pep.2005.12.002.

[4] S. Knapp *et al.*, “Protein production and purification,” *Nature Methods; New York*, vol. 5, no. 2, pp. 135–46, Feb. 2008, doi: http://dx.doi.org.ezproxy.lib.ndsu.nodak.edu/10.1038/nmeth.f.202.

[5] P. Braun and J. LaBaer, “High throughput protein production for functional proteomics,” *Trends in Biotechnology*, vol. 21, no. 9, pp. 383–388, Sep. 2003, doi: 10.1016/S0167-7799(03)00189-6.

[6] M. Hammarström, N. Hellgren, S. van den Berg, H. Berglund, and T. Härd, “Rapid screening for improved solubility of small human proteins produced as fusion proteins in Escherichia coli,” Protein Science, vol. 11, no. 2, pp. 313–321, 2002, doi: 10.1110/ps.22102.

[7] W. Weber, E. Weber, S. Geisse, and K. Memmert, “Optimisation of protein expression and establishment of the Wave Bioreactor for Baculovirus/insect cell culture,” *Cytotechnology*, vol. 38, no. 1, pp. 77–85, Jan. 2002, doi: 10.1023/A:1021102015070.

[8] A. A. Prabhu, B. Mandal, and V. V. Dasu, “Medium optimization for high yield production of extracellular human interferon-γ from Pichia pastoris: A statistical optimization and neural network-based approach,” *Korean J. Chem. Eng.*, vol. 34, no. 4, pp. 1109–1121, Apr. 2017, doi: 10.1007/s11814-016-0358-1.

[9] S. Tarahomjoo, S. R. Tabatabaei, A. Karimi, and M. Rahbar, “Strain Selection and Statistical Optimization of Culture Conditions for 19F Polysaccharide Production from Pneumococcus,” p. 13.

[10] N. K. Lee, “Statistical Optimization of Medium and Fermentation Conditions of Recombinant Pichia pastoris for the Production of Xylanase,” *Biotechnol Bioproc E*, vol. 23, no. 1, pp. 55–63, Feb. 2018, doi: 10.1007/s12257-017-0262-5.

[11] J. Sheng, H. Flick, and X. Feng, “Systematic Optimization of Protein Secretory Pathways in Saccharomyces cerevisiae to Increase Expression of Hepatitis B Small Antigen,” *Front. Microbiol.*, vol. 8, 2017, doi: 10.3389/fmicb.2017.00875.

[12] J. Hou, K. E. J. Tyo, Z. Liu, D. Petranovic, and J. Nielsen, “Metabolic engineering of recombinant protein secretion by Saccharomyces cerevisiae,” *FEMS Yeast Res*, vol. 12, no. 5, pp. 491–510, Aug. 2012, doi: 10.1111/j.1567-1364.2012.00810.x.

[13] E.-C. Ong, P. Smidt, and J. T. McGrew, “Limiting the metabolic burden of recombinant protein expression during selection yields pools with higher expression levels,” *Biotechnology Progress*, vol. 35, no. 5, p. e2839, 2019, doi: 10.1002/btpr.2839.

[14] F. R. Schmidt, “Recombinant expression systems in the pharmaceutical industry,” *Appl Microbiol Biotechnol*, vol. 65, no. 4, pp. 363–372, Sep. 2004, doi: 10.1007/s00253-004-1656-9.

[15] J. Walther, R. Godawat, C. Hwang, Y. Abe, A. Sinclair, and K. Konstantinov, “The business impact of an integrated continuous biomanufacturing platform for recombinant protein production,” *Journal of Biotechnology*, vol. 213, pp. 3–12, Nov. 2015, doi: 10.1016/j.jbiotec.2015.05.010.

[16] A. Schmideder and D. Weuster-Botz, “High-performance recombinant protein production with Escherichia coli in continuously operated cascades of stirred-tank reactors,” *J Ind Microbiol Biotechnol*, vol. 44, no. 7, pp. 1021–1029, Jul. 2017, doi: 10.1007/s10295-017-1927-y.

[17] A. K. Sharma, H. Agarwal, M. Pathak, K. D. P. Nigam, and A. S. Rathore, “Continuous refolding of a biotech therapeutic in a novel Coiled Flow Inverter Reactor,” *Chemical Engineering Science*, vol. 140, pp. 153–160, Feb. 2016, doi: 10.1016/j.ces.2015.10.009.

[18] H. E. Klock, A. White, E. Koesema, and S. A. Lesley, “Methods and Results for Semi-automated Cloning Using Integrated Robotics,” *J Struct Funct Genomics*, vol. 6, no. 2, pp. 89–94, Sep. 2005, doi: 10.1007/s10969-005-3084-1.

[19] Y. Kim *et al.*, “Automation of protein purification for structural genomics,” *J Struct Func Genom*, vol. 5, no. 1, pp. 111–118, Mar. 2004, doi: 10.1023/B:JSFG.0000029206.07778.fc.

[20] K. Peebo and P. Neubauer, “Application of Continuous Culture Methods to Recombinant Protein Production in Microorganisms,” *Microorganisms*, vol. 6, no. 3, p. 56, Sep. 2018, doi: 10.3390/microorganisms6030056.

[21] M. Saraswat, L. Musante, A. Ravidá, B. Shortt, B. Byrne, and H. Holthofer, “Preparative Purification of Recombinant Proteins: Current Status and Future Trends,” BioMed Research International, vol. 2013, pp. 1–18, 2013, doi: 10.1155/2013/312709.

[22] K. W. Barber and J. Rinehart, “The ABCs of PTMs,” *Nat Chem Biol*, vol. 14, no. 3, pp. 188–192, Feb. 2018, doi: 10.1038/nchembio.2572.

[23] F. Baneyx, “Recombinant protein expression in Escherichia coli,” *Current Opinion in Biotechnology*, vol. 10, no. 5, pp. 411–421, Oct. 1999, doi: 10.1016/S0958-1669(99)00003-8.

[24] S. K. Gupta and P. Shukla, “Advanced technologies for improved expression of recombinant proteins in bacteria: perspectives and applications,” *Critical Reviews in Biotechnology*, vol. 36, no. 6, pp. 1089–1098, Nov. 2016, doi: 10.3109/07388551.2015.1084264.

[25] R. Chen, “Bacterial expression systems for recombinant protein production: E. coli and beyond,” *Biotechnology Advances*, vol. 30, no. 5, pp. 1102–1107, Sep. 2012, doi: 10.1016/j.biotechadv.2011.09.013.

[26] Jia Baolei and Jeon Che Ok, “High-throughput recombinant protein expression in Escherichia coli: current status and future perspectives,” *Open Biology*, vol. 6, no. 8, p. 160196, doi: 10.1098/rsob.160196.

[27] C.-J. Huang, H. Lin, and X. Yang, “Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements,” *J Ind Microbiol Biotechnol*, vol. 39, no. 3, pp. 383–399, Mar. 2012, doi: 10.1007/s10295-011-1082-9.

[28] M. S. King, C. Boes, and E. R. S. Kunji, “Chapter Four - Membrane Protein Expression in Lactococcus lactis,” in *Methods in Enzymology*, vol. 556, A. K. Shukla, Ed. Academic Press, 2015, pp. 77–97.

[29] S. Lieder, P. I. Nikel, V. de Lorenzo, and R. Takors, “Genome reduction boosts heterologous gene expression in Pseudomonas putida,” *Microbial Cell Factories*, vol. 14, no. 1, p. 23, Feb. 2015, doi: 10.1186/s12934-015-0207-7.

[30] T. M. Henkin and A. L. Sonenshein, “Mutations of the Escherichia coli lacUV5 promoter resulting in increased expression in Bacillus subtilis,” *Mol Gen Genet*, vol. 209, no. 3, pp. 467–474, Oct. 1987, doi: 10.1007/BF00331151.

[31] M. Schmidt, K. Babu, N. Khanna, S. Marten, and U. Rinas, “Temperature-induced production of recombinant human insulin in high-cell density cultures of recombinant Escherichia coli,” *Journal of Biotechnology*, vol. 68, no. 1, pp. 71–83, Feb. 1999, doi: 10.1016/S0168-1656(98)00189-8.

[32] C. Sanville Millard, L. Stols, P. Quartey, Y. Kim, I. Dementieva, and M. I. Donnelly, “A less laborious approach to the high-throughput production of recombinant proteins in Escherichia coli using 2-liter plastic bottles,” *Protein Expression and Purification*, vol. 29, no. 2, pp. 311–320, Jun. 2003, doi: 10.1016/S1046-5928(03)00063-9.

[33] V. Murray, Y. Huang, J. Chen, J. Wang, and Q. Li, “A Novel Bacterial Expression Method with Optimized Parameters for Very High Yield Production of Triple-Labeled Proteins,” *Protein NMR Techniques*, A. Shekhtman and D. S. Burz, Eds. Totowa, NJ: Humana Press, pp. 1–18, Nov. 2011, doi: 10.1007/978-1-61779-480-3_1

[34] O. Khow and S. Suntrarachun, “Strategies for production of active eukaryotic proteins in bacterial expression system,” *Asian Pac J Trop Biomed*, vol. 2, no. 2, pp. 159–162, Feb. 2012, doi: 10.1016/S2221-1691(11)60213-X.

[35] G. Hannig and S. C. Makrides, “Strategies for optimizing heterologous protein expression in Escherichia coli,” *Trends in Biotechnology*, vol. 16, no. 2, pp. 54–60, Feb. 1998, doi: 10.1016/S0167-7799(97)01155-4.

[36] N. Ferrer-Miralles, J. Domingo-Espín, J. L. Corchero, E. Vázquez, and A. Villaverde, “Microbial factories for recombinant pharmaceuticals,” *Microb Cell Fact*, vol. 8, no. 1, p. 17, Mar. 2009, doi: 10.1186/1475-2859-8-17.

[37] L. Sanchez-Garcia, L. Martín, R. Mangues, N. Ferrer-Miralles, E. Vázquez, and A. Villaverde, “Recombinant pharmaceuticals from microbial cells: a 2015 update,” *Microb Cell Fact*, vol. 15, no. 1, p. 33, Feb. 2016, doi: 10.1186/s12934-016-0437-3.

[38] K. Ukkonen, A. Neubauer, V. J. Pereira, and A. Vasala, “High Yield of Recombinant Protein in Shaken E. coli Cultures with Enzymatic Glucose Release Medium EnPresso B,” in *Heterologous*

Gene Expression in E.coli: Methods and Protocols, N. A. Burgess-Brown, Ed. New York, NY: Springer, 2017, pp. 127–137.

[39] Y. Shen, M. Morishita, and E. di Luccio, “High yield recombinant expression and purification of oncogenic NSD1, NSD2, and NSD3 with human influenza hemagglutinin tag,” *Protein Expression and Purification*, vol. 166, p. 105506, Feb. 2020, doi: 10.1016/j.pep.2019.105506.

[40] D. P. Clark and N. J. Pazdernik, “Chapter 10 - Recombinant Proteins,” in *Biotechnology (Second Edition)*, D. P. Clark and N. J. Pazdernik, Eds. Boston: Academic Cell, 2016, pp. 335–363.

[41] B. H. Johnson and M. H. Hecht, “Recombinant Proteins Can Be Isolated from E. coli Cells by Repeated Cycles of Freezing and Thawing,” *Bio/Technology*, vol. 12, no. 12, pp. 1357–1360, Dec. 1994, doi: 10.1038/nbt1294-1357.

[42] R. F. Ohana *et al.*, “HaloTag7: A genetically engineered tag that enhances bacterial expression of soluble proteins and improves protein purification,” *Protein Expression and Purification*, vol. 68, no. 1, pp. 110–120, Nov. 2009, doi: 10.1016/j.pep.2009.05.010.

[43] J. D. Valderrama-Rincon *et al.*, “An engineered eukaryotic protein glycosylation pathway in Escherichia coli,” *Nature Chemical Biology*, vol. 8, no. 5, pp. 434–436, May 2012, doi: 10.1038/nchembio.921.

[44] M. Wacker, “N-Linked Glycosylation in Campylobacter jejuni and Its Functional Transfer into E. coli,” Science, vol. 298, no. 5599, pp. 1790–1793, Nov. 2002, doi: 10.1126/science.298.5599.1790.

[45] D. Gagoski *et al.*, “Performance benchmarking of four cell-free protein expression systems,” *Biotechnology and Bioengineering*, vol. 113, no. 2, pp. 292–300, 2016, doi: 10.1002/bit.25814.

[46] L. G. Horga *et al.*, “Tuning recombinant protein expression to match secretion capacity,” *Microbial Cell Factories*, vol. 17, no. 1, p. 199, Dec. 2018, doi: 10.1186/s12934-018-1047-z.

[47] J. Tian *et al.*, “Predicting synonymous codon usage and optimizing the heterologous gene for expression in E. coli,” *Scientific Reports*, vol. 7, no. 1, pp. 1–9, Aug. 2017, doi: 10.1038/s41598-017-10546-0.

[48] P. Barbosa, G. J. Wistow, M. Cialkowski, J. Piatigorsky, and W. E. O’Brien, “Expression of duck lens delta-crystallin cDNAs in yeast and bacterial hosts. Delta 2-crystallin is an active argininosuccinate lyase.,” *J. Biol. Chem.*, vol. 266, no. 33, pp. 22319–22322, Nov. 1991.

[49] M. Geymonat, A. Spanos, and S. G. Sedgwick, “A Saccharomyces cerevisiae autoselection system for optimised recombinant protein expression,” *Gene*, vol. 399, no. 2, pp. 120–128, Sep. 2007, doi: 10.1016/j.gene.2007.05.001.

[50] S. Macauley-Patrick, M. L. Fazenda, B. McNeil, and L. M. Harvey, “Heterologous protein production using the Pichia pastoris expression system,” *Yeast*, vol. 22, no. 4, pp. 249–270, 2005, doi: 10.1002/yea.1208.

[51] S. L. Forsburg, “Comparison of Schizosaccharomyces pombe expression systems.,” *Nucleic Acids Res*, vol. 21, no. 12, pp. 2955–2956, Jun. 1993.

[52] N. Burns *et al.*, “Large-scale analysis of gene expression, protein localization, and gene disruption in Saccharomyces cerevisiae.,” *Genes Dev.*, vol. 8, no. 9, pp. 1087–1105, May 1994, doi: 10.1101/gad.8.9.1087.

[53] J. F. Senecoff, R. C. Bruckner, and M. M. Cox, “The FLP recombinase of the yeast 2-micron plasmid: characterization of its recombination site,” *PNAS*, vol. 82, no. 21, pp. 7270–7274, Nov. 1985, doi: 10.1073/pnas.82.21.7270.

[54] H. Kim, S. J. Yoo, and H. A. Kang, “Yeast synthetic biology for the production of recombinant therapeutic proteins,” *FEMS Yeast Research; Hoboken*, vol. 15, no. 1, pp. 1–16, Feb. 2015, doi: http://dx.doi.org/10.1111/1567-1364.12195.

[55] A. Idiris, H. Tohda, H. Kumagai, and K. Takegawa, “Engineering of protein secretion in yeast: strategies and impact on protein production,” Applied Microbiology and Biotechnology, vol. 86, no. 2, pp. 403–417, Mar. 2010, doi: 10.1007/s00253-010-2447-0.

[56] A. W. Murray and J. W. Szostak, “Pedigree analysis of plasmid segregation in yeast,” *Cell*, vol. 34, no. 3, pp. 961–970, Oct. 1983, doi: 10.1016/0092-8674(83)90553-6.

[57] S. R. Hamilton *et al.*, “Humanization of Yeast to Produce Complex Terminally Sialylated Glycoproteins,” *Science*, vol. 313, no. 5792, pp. 1441–1443, Sep. 2006, doi: 10.1126/science.1130256.

[58] T. Österlund, I. Nookaew, and J. Nielsen, “Fifteen years of large scale metabolic modeling of yeast: Developments and impacts,” *Biotechnology Advances*, vol. 30, no. 5, pp. 979–988, Sep. 2012, doi: 10.1016/j.biotechadv.2011.07.021.

[59] L. M. Damasceno *et al.*, “An optimized fermentation process for high-level production of a single-chain Fv antibody fragment in Pichia pastoris,” *Protein Expr. Purif.*, vol. 37, no. 1, pp. 18–26, Sep. 2004, doi: 10.1016/j.pep.2004.03.019.

[60] B. J. Hackel, D. Huang, J. C. Bubolz, X. X. Wang, and E. V. Shusta, “Production of Soluble and Active Transferrin Receptor-Targeting Single-Chain Antibody using Saccharomyces cerevisiae,” *Pharm Res*, vol. 23, no. 4, pp. 790–797, Apr. 2006, doi: 10.1007/s11095-006-9778-7.

[61] B. Poppenberger *et al.*, “Heterologous expression of Arabidopsis UDP-glucosyltransferases in Saccharomyces cerevisiae for production of zearalenone-4-O-glucoside,” *Appl. Environ. Microbiol.*, vol. 72, no. 6, pp. 4404–4410, Jun. 2006, doi: 10.1128/AEM.02544-05.

[62] J. M. Cregg, J. L. Cereghino, J. Shi, and D. R. Higgins, “Recombinant Protein Expression in Pichia pastoris,” *Molecular Biotechnology*, vol. 16, no. 1, pp. 23–52, 2000, doi: 10.1385/MB:16:1:23.

[63] D. D. Hegedus, T. A. Pfeifer, J. Hendry, D. A. Theilmann, and T. A. Grigliatti, “A series of broad host range shuttle vectors for constitutive and inducible expression of heterologous proteins in insect cell lines,” *Gene*, vol. 207, no. 2, pp. 241–249, Jan. 1998, doi: 10.1016/s0378-1119(97)00636-7.

[64] H. Mabashi-Asazuma and D. L. Jarvis, “CRISPR-Cas9 vectors for genome editing and host engineering in the baculovirus–insect cell system,” *PNAS*, vol. 114, no. 34, pp. 9068–9073, Aug. 2017, doi: 10.1073/pnas.1705836114.

[65] Z. Zhang, J. Yang, and D. Barford, “Recombinant expression and reconstitution of multiprotein complexes by the USER cloning method in the insect cell-baculovirus expression system,” *Methods*, vol. 95, pp. 13–25, Feb. 2016, doi: 10.1016/j.ymeth.2015.10.003.

[66] R. M. Patterson, J. K. Selkirk, and B. A. Merrick, “Baculovirus and Insect Cell Gene Expression: Review of Baculovirus Biotechnology,” *Environmental Health Perspectives*, vol. 103, no. 7, p. 4, 1995.

[67] M. Martínez-Solís, S. Herrero, and A. M. Targovnik, “Engineering of the baculovirus expression system for optimized protein production,” *Appl Microbiol Biotechnol*, vol. 103, no. 1, pp. 113–123, Jan. 2019, doi: 10.1007/s00253-018-9474-7.

[68] S. Wenzel, T. Imasaki, and Y. Takagi, “A practical method for efficient and optimal production of Seleno-methionine-labeled recombinant protein complexes in the insect cells,” *Protein Science*, vol. 28, no. 4, pp. 808–822, 2019, doi: 10.1002/pro.3575.

[69] C. Ladd Effio *et al.*, “Modeling and simulation of anion-exchange membrane chromatography for purification of Sf9 insect cell-derived virus-like particles,” *Journal of Chromatography A*, vol. 1429, pp. 142–154, Jan. 2016, doi: 10.1016/j.chroma.2015.12.006.

[70] M. I. P. Malgapo and M. E. Linder, “Purification of Recombinant DHHC Proteins Using an Insect Cell Expression System,” in *Protein Lipidation: Methods and Protocols*, M. E. Linder, Ed. New York, NY: Springer, 2019, pp. 179–189.

[71] L. E. Pyle, P. Barton, Y. Fujiwara, A. Mitchell, and N. Fidge, “Secretion of biologically active human proapolipoprotein A-I in a baculovirus-insect cell system: protection from degradation by protease inhibitors.,” *J. Lipid Res.*, vol. 36, no. 11, pp. 2355–2361, Nov. 1995.

[72] X. Shi and D. L. Jarvis, “Protein N-Glycosylation in the Baculovirus-Insect Cell System,” *Current Drug Targets*., vol. 8, no. 10, pp. 1116-1125, 2007, doi: 10.2174/138945007782151360

[73] G. J. Pendse, S. Karkare, and J. E. Bailey, “Effect of cloned gene dosage on cell growth and hepatitis B surface antigen synthesis and secretion in recombinant CHO cells,” *Biotechnology and Bioengineering*, vol. 40, no. 1, pp. 119–129, 1992, doi: 10.1002/bit.260400117.

[74] S. R. W. Chen, X. Li, K. Ebisawa, and L. Zhang, “Functional Characterization of the Recombinant Type 3 Ca 2+ Release Channel (Ryanodine Receptor) Expressed in HEK293 Cells,” Journal of Biological Chemistry, vol. 272, no. 39, pp. 24234–24246, Sep. 1997, doi: 10.1074/jbc.272.39.24234.

[75] S. Fisher, N. Charara, A. Gerber, J. Wolfel, G. Schiedner, B. Voedisch and S. Geisse, “Transient recombinant protein expression in a human amniocyte cell line: The CAP-T (R) cell system,” *Biotechnology and Bioengineering,* vol. 109, no. 9, pp. 2250-2261, Sep. 2012, doi: 10.1002/bit.24514

[76] J. Zhu, “Mammalian cell protein expression for biopharmaceutical production,” *Biotechnol. Adv.*, vol. 30, no. 5, pp. 1158–1170, Oct. 2012, doi: 10.1016/j.biotechadv.2011.08.022.

[77] V. V. Emmerling *et al.*, “Rational plasmid design and bioprocess optimization to enhance recombinant adeno-associated virus (AAV) productivity in mammalian cells,” *Biotechnology Journal*, vol. 11, no. 2, pp. 290–297, 2016, doi: 10.1002/biot.201500176.

[78] D. Y. Kim, J. C. Lee, H. N. Chang, and D. J. Oh, “Development of serum-free media for a recombinant CHO cell line producing recombinant antibody,” *Enzyme and Microbial Technology*, vol. 39, no. 3, pp. 426–433, Jul. 2006, doi: 10.1016/j.enzmictec.2005.11.047.

[79] V. Chromikova, M. A. Zaragoza, and F. Krammer, “Generation of a serum free CHO DG44 cell line stably producing a broadly protective anti-influenza virus monoclonal antibody,” *PLOS ONE*, vol. 12, no. 9, p. e0183315, Sep. 2017, doi: 10.1371/journal.pone.0183315.

[80] A. D. Bandaranayake and S. C. Almo, “Recent advances in mammalian protein production,” *FEBS Lett*, vol. 588, no. 2, pp. 253–260, Jan. 2014, doi: 10.1016/j.febslet.2013.11.035.

[81] V. Jadhav *et al.*, “CHO microRNA engineering is growing up: recent successes and future challenges,” *Biotechnol. Adv.*, vol. 31, no. 8, pp. 1501–1513, Dec. 2013, doi: 10.1016/j.biotechadv.2013.07.007.

[82] M. S. Sinacore *et al.*, “CHO DUKX cell lineages preadapted to growth in serum-free suspension culture enable rapid development of cell culture processes for the manufacture of recombinant proteins,” *Biotechnology and Bioengineering*, vol. 52, no. 4, pp. 518–528, 1996, doi: 10.1002/(SICI)1097-0290(19961120)52:4<518::AID-BIT7>3.0.CO;2-S.

[83] A. Amanullah *et al.*, “Novel micro-bioreactor high throughput technology for cell culture process development: Reproducibility and scalability assessment of fed-batch CHO cultures,” *Biotechnology and Bioengineering*, vol. 106, no. 1, pp. 57–67, 2010, doi: 10.1002/bit.22664.

[84] E. D. Carlson, R. Gan, C. E. Hodgman, and M. C. Jewett, “Cell-free protein synthesis: Applications come of age,” *Biotechnology Advances*, vol. 30, no. 5, pp. 1185–1194, Sep. 2012, doi: 10.1016/j.biotechadv.2011.09.016.

[85] O. Kovtun, S. Mureev, W. Jung, M. H. Kubala, W. Johnston, and K. Alexandrov, “Leishmania cell-free protein expression system,” *Methods*, vol. 55, no. 1, pp. 58–64, Sep. 2011, doi: 10.1016/j.ymeth.2011.06.006.

[86] K. M. Wilding *et al.*, “Endotoxin-Free E. coli-Based Cell-Free Protein Synthesis: Pre-Expression Endotoxin Removal Approaches for on-Demand Cancer Therapeutic Production,” *Biotechnology Journal*, vol. 14, no. 3, p. 1800271, 2019, doi: 10.1002/biot.201800271.

[87] Y. Endo and T. Sawasaki, “High-throughput, genome-scale protein production method based on the wheat germ cell-free expression system,” *J Struct Func Genom*, vol. 5, no. 1, pp. 45–57, Mar. 2004, doi: 10.1023/B:JSFG.0000029208.83739.49.

[88] K. K. Y. Ho, V. L. Murray, and A. P. Liu, “Chapter 16 - Engineering artificial cells by combining HeLa-based cell-free expression and ultrathin double emulsion template,” in *Methods in Cell Biology*, vol. 128, J. Ross and W. F. Marshall, Eds. Academic Press, 2015, pp. 303–318.

[89] D. P. Clark and N. J. Pazdernik, “Chapter 15 - Transgenic Plants and Plant Biotechnology,” in *Biotechnology (Second Edition)*, D. P. Clark and N. J. Pazdernik, Eds. Boston: Academic Cell, 2016, pp. 461–492.

[90] Z. Hui and G. Wen, “Construction and identification for the transgenic plant vector of recombinant plasmid pBI-Eg95-EgA31 of Echinococcus granulosus,” *Chinese J Zoonoses*, vol. 5, pp. 4-8, Jan. 2008.

[91] H. S. Mason, D. M. Lam, and C. J. Arntzen, “Expression of hepatitis B surface antigen in transgenic plants,” *PNAS*, vol. 89, no. 24, pp. 11745–11749, Dec. 1992, doi: 10.1073/pnas.89.24.11745.

[92] D. Chong, W. Roberts and T. Arakawa, “Expression of the human milk protein β-casein in transgenic potato plants,” *Transgenic Res,* vol. 6, pp. 289-296, Jul. 1997, doi: 10.1023/A:1018410712288

[93] M. Rubtsova, K. Kempe, A. Gils, A. Ismagul, J. Weyen, and M. Gils, “Expression of active Streptomyces phage phiC31 integrase in transgenic wheat plants,” *Plant Cell Rep*, vol. 27, no. 12, pp. 1821–1831, Dec. 2008, doi: 10.1007/s00299-008-0604-z.

[94] J. K.-C. Ma *et al.*, “Realising the value of plant molecular pharming to benefit the poor in developing countries and emerging economies,” *Plant Biotechnol. J.*, vol. 11, no. 9, pp. 1029–1033, 2013, doi: 10.1111/pbi.12127.

[95] E. L. Rech and M. R. Lopes, “Insights into Brazilian agricultural structure and sustainable intensification of food production,” *Food and Energy Security*, vol. 1, no. 2, pp. 77–80, 2012, doi: 10.1002/fes3.12.

[96] G. Giddings, G. Allison, D. Brooks, and A. Carter, “Transgenic plants as factories for biopharmaceuticals,” *Nat. Biotechnol.*, vol. 18, no. 11, pp. 1151–1155, Nov. 2000, doi: 10.1038/81132.

[97] A. J. Conley, J. J. Joensuu, A. M. Jevnikar, R. Menassa, and J. E. Brandle, “Optimization of elastin-like polypeptide fusions for expression and purification of recombinant proteins in plants,” *Biotechnology and Bioengineering*, vol. 103, no. 3, pp. 562–573, 2009, doi: 10.1002/bit.22278.

[98] J. R. Cutt, D. C. Dixon, J. P. Carr, and D. F. Klessig, “Isolation and nucleotide sequence of cDNA clones for the pathogenesis-related proteins PR1a, PR1b and PR1c of Nicotiana tabacum cv. Xanthi nc induced by TMV infection.,” *Nucleic Acids Res*, vol. 16, no. 20, p. 9861, Oct. 1988.

[99] J. Scheller, K.-H. Gührs, F. Grosse, and U. Conrad, “Production of spider silk proteins in tobacco and potato,” *Nature Biotechnology*, vol. 19, no. 6, pp. 573–577, Jun. 2001, doi: 10.1038/89335.

[100] J. K.-C. Ma, P. M. W. Drake, and P. Christou, “The production of recombinant pharmaceutical proteins in plants,” *Nature Reviews Genetics*, vol. 4, no. 10, pp. 794–805, Oct. 2003, doi: 10.1038/nrg1177.

[101] A. R. Kusnadi, Z. L. Nikolov, and J. A. Howard, “Production of recombinant proteins in transgenic plants: Practical considerations,” *Biotechnology and Bioengineering*, vol. 56, no. 5, pp. 473–484, 1997, doi: 10.1002/(SICI)1097-0290(19971205)56:5<473::AID-BIT1>3.0.CO;2-F.

[102] D. P. Clark and N. J. Pazdernik, “Chapter 16 - Transgenic Animals,” in *Biotechnology (Second Edition)*, D. P. Clark and N. J. Pazdernik, Eds. Boston: Academic Cell, 2016, pp. 493–521.

[103] Z. Bösze and L. Hiripi, “Recombinant protein expression in milk of livestock species,” *Methods Mol. Biol.*, vol. 824, pp. 629–641, 2012, doi: 10.1007/978-1-61779-433-9_34.

[104] U. H. Weidle, H. Lenz, and G. Brem, “Genes encoding a mouse monoclonal antibody are expressed in transgenic mice, rabbits and pigs,” Gene, vol. 98, no. 2, pp. 185–191, Feb. 1991, doi: 10.1016/0378-1119(91)90172-8.

[105] C. A. Penno, Y. Kawabe, A. Ito, and M. Kamihira, “Production of recombinant human erythropoietin/Fc fusion protein by genetically manipulated chickens,” Transgenic Research, vol. 19, no. 2, pp. 187–195, Apr. 2010, doi: 10.1007/s11248-009-9310-z.

[106] Z.-Y. Ryoo et al., “Expression of recombinant human granulocyte macrophage-colony stimulating factor (hGM-CSF) in mouse urine,” vol. 10, pp. 193-200, 2001, doi: 10.1023/a:1016657501149

[107] K. Tatemastu, “Utilization of Transgenic Silkworms for Recombinant Protein Production,” *Journal of Biotechnology & Biomaterials*, vol. s9, Jan. 2012, doi: 10.4172/2155-952X.S9-004.

[108] M. K. Dyck, D. Lacroix, F. Pothier, and M.-A. Sirard, “Making recombinant proteins in animals--different systems, different applications,” *Trends Biotechnol.*, vol. 21, no. 9, pp. 394–399, Sep. 2003, doi: 10.1016/S0167-7799(03)00190-2.

[109] D. A. Dunn, C. A. Pinkert, and D. L. Kooyman, “Foundation Review: Transgenic animals and their impact on the drug discovery industry,” *Drug Discov. Today*, vol. 10, no. 11, pp. 757–767, Jun. 2005, doi: 10.1016/S1359-6446(05)03452-5.

[110] Y. Wang, S. Zhao, L. Bai, J. Fan, and E. Liu, “Expression Systems and Species Used for Transgenic Animal Bioreactors,” *Biomed Res Int*, vol. 2013, 2013, doi: 10.1155/2013/580463.

[111] T. D. Wilkins and W. Velander, “Isolation of recombinant proteins from milk,” *Journal of Cellular Biochemistry*, vol. 49, no. 4, pp. 333–338, 1992, doi: 10.1002/jcb.240490403.

[112] K. E. van Cott *et al.*, “Affinity purification of biologically active and inactive forms of recombinant human protein C produced in porcine mammary gland,” *Journal of Molecular Recognition*, vol. 9, no. 5–6, pp. 407–414, 1996, doi: 10.1002/(SICI)1099-1352(199634/12)9:5/6<407::AID-JMR277>3.0.CO;2-X.

[113] M. A. Al-Ghobashy, M. A. K. Williams, B. Brophy, G. Laible, and D. R. K. Harding, “On-line casein micelle disruption for downstream purification of recombinant human myelin basic protein produced in the milk of transgenic cows,” *Journal of Chromatography B*, vol. 877, no. 16, pp. 1667–1677, Jun. 2009, doi: 10.1016/j.jchromb.2009.04.012.

CHAPTER 3: HARNESSING NATURAL SILK PRODUCTION BY OPTIMIZING SPIDER SILK GLAND CELL CULTURE

Abstract

The production of natural dragline spider silk will play a significant role in our ability to advance use of spider silk in biotechnology and pharmaceutical applications, thus realizing the long-held promise of the commercial potential of spider silk proteins. Classic, recombinant techniques are used to produce spider silk protein for research use; however, recombinant production poses issues with truncated proteins, as well as changes in the structure and behavior of the protein. Currently, there are only two recognized ways to obtain natural spider silk protein: 1) collect silk from a spun web or 2) forcibly extract silk from a captive spider, both of which offer a limited source due to the purity of the fiber, territorial temperament, and short life span of golden orb weaving spiders. An alternative approach to the collection of the natural silk fiber is to collect the protein itself through the growth of a cell line developed from the major ampullate gland that excretes the desirable spider silk protein. This study investigates the growth of these cells while exploring the impact that pH and sodium bicarbonate levels play in the cell morphology and excretion of the dragline silk protein. It is clear that mimicking the microenvironment of the major ampullate gland brings us one step closer to the growth and maintenance of a major ampullate gland cell line that will eventually lead to the large-scale production of natural spider silk protein.

Introduction

The primary structure of spider silk protein, with its structural hierarchy and repetitive structural motifs, not only endows the fiber with several desirable mechanical properties but also allows for easy genetic manipulation. The balance of tensile strength and elasticity provides extreme toughness and energy absorbance, allowing for use beyond basic research extending to commercial utility [1]. While orb-weaving spiders, particularly *Nephila clavipes*, produce a variety of solid fibers with each having mechanical properties suited to its ecological niche, dragline silk is by far the most studied for its specific mechanical properties and availability [1], [2]. Furthermore, artificially spinning native high molecular weight silk protein found in the gland has yielded promising artificial fibers both a high tenacity and a high elastic modulus rivaling the natural toughness of natural silk exemplars [3]. Unfortunately, this is not a sustainable strategy for commercial translation of spider silk protein as natural, full-length silk protein can

only be reliably obtained by either forcibly silking a spider in captivity or directly obtaining silk from the orb web. Despite the touted mastery of silk, obtaining natural protein from spiders held in captivity is not considered a feasible commercial option due to 1) their territorial behavior and rather brief life span, 2) the relatively low yield of silk, and 3) the time-consuming collection process. Alternatively, recombinant protein expression systems have been heavily exploited for the production of silk. Based on our knowledge of the genetics and structure of major ampullate spider silk proteins, a variety of recombinant host systems, ranging from goats' milk to *E.coli*, can be used for production. The challenges with protein expression vary by host system with one of the common downfalls being a significant limitation in protein size (Table 3.1).

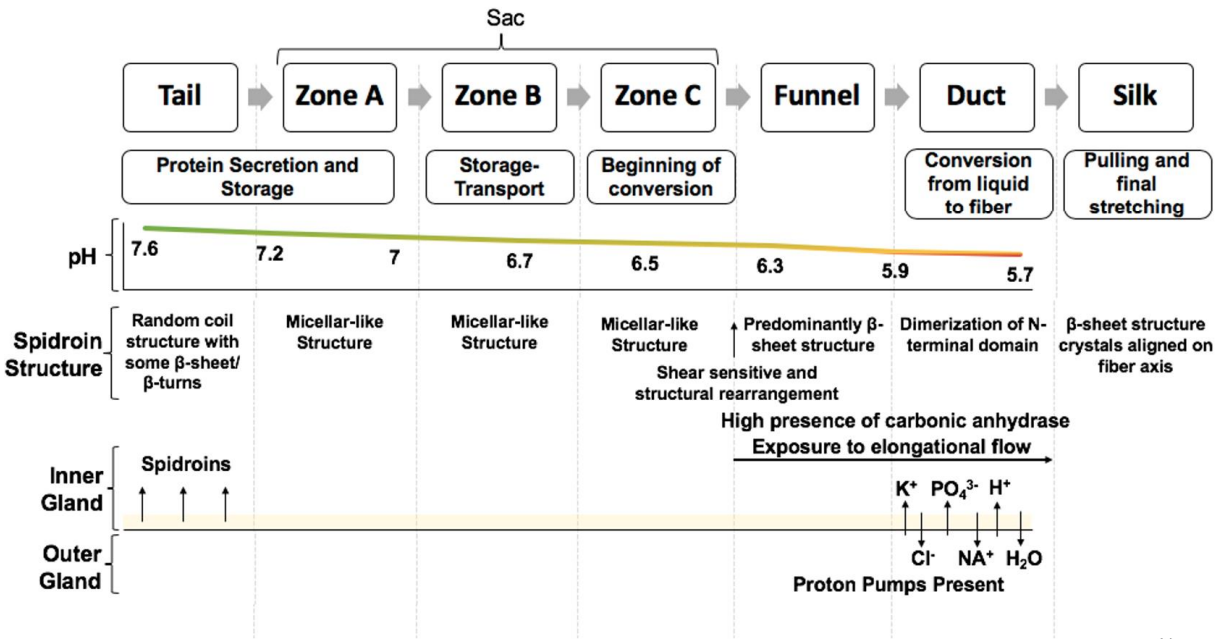

Figure 3.1. The natural spinning process of *N. clavipes* major ampullate gland divided into sections. Factors that affect the spinning process of dragline silk are mapped to the different sections.

While much is known about the sequence, structure, and function of major ampullate silk protein and natural silk spinning, integration of each element of the fiber spinning process, beginning with the natural cellular secretion of the protein, remains enigmatic. Figure 3.1 breaks down the current understanding of the natural process of major ampullate spider silk fiber formation from protein secretion and aggregation to fiber creation [4]–[8]. This existing conceptual framework involves alterations of both the pH and ion concentration as the protein traverses to the spinneret. Alterations in these key factors

may not only play a role in the protein's secretion and conformation, but also may hold the key to cell viability in culture. Exposing the cells that line the major ampullate gland to these microenvironmental conditions may prove that production of major ampullate silk protein relies on these conditions for their gene expression.

Recent insight into the silk secretion process has posed the unique option of obtaining silk protein as a secreted extracellular matrix directly from native major ampullate silk gland cell culture [9]. This novel method introduces potential for many competitive characteristics of the natural major ampullate silk produced as compared to other methods of production for recombinant silk proteins (Table 3.1). To further develop the idea of natural silk production, this study reports the methods to extract the modified columnar, epithelial cells lining the major ampullate silk gland of *Nephila clavipes,* a process previously uncharacterized. Furthermore, it explores the impact of changing the pH and sodium bicarbonate concentration present in silk gland cell culture to more closely mimic the conditions found in the native gland as the silk protein traverses the gland and is converted to a solid fiber. These methods show promising results, delivering cells that not only maintain morphological fidelity but also continue expression of major ampullate silk, potentially yielding a native quality protein.

Table 3.1. (a) Common hosts used for producing recombinant spider silk protein. (b) Methods of producing natural spider silk protein with predicted characteristics for MaSp cells.

a. Host Category	Host	Molecular Weight (kDa)	High Expression Level	Large-Scale Suitable	Ease in Use	Source
Bacteria	*E.coli*	63-71	✓	✓	✓	[10]
Yeast	*P.pastoris*	65	✓	✓	✓	[11]
Plant	Tobacco	13-100	✗	✓	✗	[12]
Plant	Potato	13-100	✗	✓	✗	[12]
Animal	Mice	31-66	✓	✓	✗	[13]
Animal	Goats	65-120	✓	✓	✗	[14]
Insect	B.mori	75-130	✗	✓	✗	[15]

b. Natural Silk Source	Molecular Weight (kDa)	High Expression Level	Large-Scale Suitable	Ease in Use	Source
N. Clavipes	300-350	✓	✗	✗	[16]
MaSp Cells	Unknown	✓	✓	✓	[9]

Materials and Methods

Spiders

12 *Nephila clavipes,* golden orb-weaving spiders, were wild caught in Florida (Tarantula Spiders, USA).

Media Preparation

Reference Appendix A for complete protocol. Complete TNM-FH media for insect cell culture was prepared using TNM-FH insect media (Fisher, USA) supplemented with gentamicin (1mg/mL, Sigma-Aldrich), amphotericin (1 mg/mL, Fisher BioReagents), tetracycline (1 mg/mL, Sigma-Aldrich), oxalacetic acid (1M, Fisher Scientific), insulin (10mg/mL, Sigma-Aldrich) and FBS (10% v/v, Fisher Scientific). Media was further supplemented with varying amounts of sodium bicarbonate (Sigma-Aldrich) (Table 3.2) and

the pH was adjusted with either 1M NaOH or 1M HCl. All media was syringe filtered (SFCA 0.22um, 30mm, CELLTREAT) for sterility directly after being prepared and stored at 2-8°C for up to 4 weeks.

Table 3.2. Sodium bicarbonate concentration and pH of insect media.

pH	Sodium Bicarbonate Concentration
5.7	0.005 M
6.2	0.011 M
7.0	0.021 M
7.6	0.021 M

Gland Extraction

Reference Appendix B for complete protocol. Major ampullate glands were extracted from *N. clavipes* after an overdose of CO_2. Briefly, the cephalothorax was removed from the abdomen with a scissors after completely immersing the spider for 3 minutes in 70% ethanol to remove any contaminants adhered to the exoskeleton. Subsequently, the cephalothorax was incised along its length and the glands were dissected and separated in a sodium citrate buffer (0.1M, pH 6.0, Thermo Fisher Scientific). The major ampullate glands (easily distinguishable from the others), shown in Figure 3.2, were separated from the other glands and placed in the media containing 10% FBS, 7.0 pH and no sodium bicarbonate, with 10% v/v DMSO and placed at -80°C.

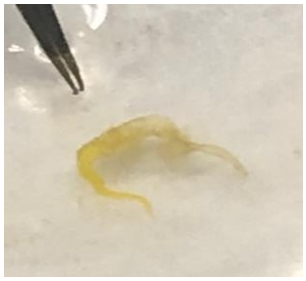

Figure 3.2. Major ampullate gland extracted from a golden orb weaver.

Cell Processing

Reference Appendix C for complete protocol. For isolation, glands were removed from the freezer and rinsed four times with sodium citrate. Subsequently, the major ampullate glands from 2 spiders were aggregated (four glands total) and were placed in fresh TNM-FH, 7.0 pH) without FBS and finely minced. A total of 16 glands were processed for the continuous media pH growth, providing four biological

replicates. A total of 8 glands were processed for the reduced media pH growth, providing 2 biological replicates. 200µL of 0.05% trypsin (Fisher Scientific) plus 0.02% EDTA (Thermo Fisher Scientific) was added to each tube containing 4 glands each and allowed to incubate in 37°C for 5 minutes. At which time, collagenase (type 4, Worthington Biochemical) was added to a final concentration of 1mg/ml and incubated at 37°C for 10 minutes. TNM-FH media, 7.0 pH, with 10% FBS (200 µL) was added to the suspension to quench the enzymatic digestion and the contents were centrifuged at 400 RPG at room temperature, for 7 minutes. After decanting the supernatant, cell pellets were gently washed by resuspending with 100µL complete TNM-FH media with FBS, centrifuged at 400 RPG at room temperature, for 7 minutes, then the supernatant was decanted.

For the continuous media pH study, the cells were resuspended in 500µL media (pH 7.0 with 10% FBS) and filtered with a 70um nylon mesh cell strainer (Fisher Scientific) into a 50 mL conical. The filtered cell-containing media was divided evenly into 8 individual wells, using two 6-well plates (TCPS, CELLTREAT). Media with 10% FBS was added of varying pH values listed in the “Media pH Day 0” column in Table 3.3.

For the decreasing media pH study, the cells were resuspended in 1000µL media (pH 7.0 with 10% FBS) and filtered with a 70um nylon mesh cell strainer (Fisher Scientific) into a 50 mL conical. 250µL of filtered media was reserved for RNA extraction, then the filtered cell-containing media was divided evenly into 4 individual wells, using two 6-well plates (TCPS, CELLTREAT). Media with 10% FBS was added of varying pH values listed in the “Media pH Day 0” column in Table 3.4.

Cell Growth

For both experiments, the cells were incubated at 27°C, no CO_2 for proliferation and growth. Each day, the pH of the culture media (5µL) in each well was assessed and replaced with 5µL fresh media of same pH as last media change, and cells were visualized at 25x under brightfield. Media was changed by removing ~1.25 mL media in each well, by placing pipette tip at a 45° angle and pulling media off of from top of the while taking care not to agitate cells on the bottom. The cells that remained adhered to well plate were fed with fresh media bringing the volume back to ~2 mL. Table 3.3 and Table 3.4 list the required media pH per experiment relative to the media change days. Sodium bicarbonate levels follow concentrations that vary by pH as indicated in Table 3.2.

Table 3.3. The media change schedule for the primary cell line with continuous pH growth conditions.

Sample	Media pH Day 0	Media pH Day 5
1	5.7	5.7
2	5.7	5.7
3	6.2	6.2
4	6.2	6.2
5	7.0	7.0
6	7.0	7.0
7	7.6	7.6
8	7.6	7.6

Table 3.4. The media change schedule for the primary cell line with continuous growth at 5.7 pH (samples 1 & 2) decreasing pH (samples 3 & 4).

Sample	Media pH Day 0	Media pH Day 2	Media pH Day 5	Media pH Day 8	Media pH Day 16
1	5.7	5.7	5.7	5.7	5.7
2	5.7	5.7	5.7	5.7	5.7
3	7.6	7.0	6.3	5.7	5.7
4	7.6	7.0	6.3	5.7	5.7

RNA Extraction

Prior to plating, total RNA was isolated from 250µL of homogenous cell suspension. Cells were spun down at 400 rpg for 7 minutes and resuspended in 1mL of TRIzol Reagent (Invitrogen) and RNA was isolated according to the manufacturer's directions. Briefly, chloroform (Sigma-Aldrich) was added to the TRIzol Reagent containing the cells to dissociate the nucleoproteins complex, then the colorless upper aqueous phase containing the RNA was transferred to a new tube. The RNA was precipitated using isopropanol (North Dakota State University, USA) forming a gel-like pellet at the bottom of the tube. The RNA was then washed using 75% ethanol (North Dakota State University, USA). Lastly, the RNA was solubilized with RNase-free water (IBI Scientific).

cDNA Processing

Isolated RNA was used to generate first-strand cDNA (SuperScript VILO cDNA Synthesis Kit, Invitrogen). The RNA concentration was first read using a spectrophotometer to ensure that the capacity of the enzyme was not exceeded. The maximum amount of RNA (14µL) was added to each reaction and

an oligo dT was used to amplify mRNA only. The remainder of the reaction was done according to the manufacturer's directions for the VILO reaction mix.

Major Ampullate Spidroin (MaSp) DNA Amplification

The production of major ampullate spidroin (MaSp) was detected using a standard Polymerase Chain Reaction (PCR). PCR was done on the first strand cDNA that represents genes being actively transcribed and processed as mRNA. Briefly, GoTaq Green Master Mix (Promega) was used to amplify MaSp2 from 3500ng of the first strand cDNA template strand. Two primer sets were used (Table 3.5) and the presence of bands was assessed using a 1% agarose gel in TBE with ethidium bromide for visualization. The gel was run at 150V for 60 minutes.

Table 3.5. PCR cycle temperatures and primers used to confirm presence of MaSp cells.

1X	95°C	2 min.
40X	95°C	30 sec.
	54°C	30 sec.
	72°C	1 min.
1X	72°C	5 min.
1X	4°C	∞
Primer Set 1 (Expected bases 717)		
Forward: 5' CCC AAT CCA TCT ACG CTT CT 3'		
Reverse: 5' CAG GTC CTG ATG GTC CAT ATA C 3'		
Primer Set 2 (Expected bases 502)		
Forward: 5' GCC CAT TTC AGT GCT TCA AC 3'		
Reverse: 5' TAG CCG AAG CTG CAC TTA TC 3'		

Results

Effect of Continuous Media pH on MaSp Cells

Figure 3.3a. graphically shows the change in pH of samples 1, 3, 5, and 7 of the continuous media pH study, note a media change returning pH values to day 0 pH values of all samples (pH values listed in Table 3.3). Sample 7, cells grown at 7.6 pH, shows a quick decrease in the pH of the media

starting on day 3, until media was replenished on day 5. Sample 5, cells grown at 7.0 pH, show consistent pH maintenance until day 6 when there is a slight decrease to ~6.5. Sample 3, cells grown at 6.2 pH, have consistent pH values with a slight increase in the pH on day 4 before the media change on day 5, then another increase in pH between day 5 and 7 reaching a peak pH of ~6.8. Sample 1, cells grown at 5.7 pH, show well-maintained pH throughout the entirety of testing. Figure 3.3b shows homogenous cells (samples 1, 3, 5, 7) redistributed into 6-well plates on day 0, compared to a change in the morphology of the cells on day 7. There are clear morphological changes in the cells in the 6.2 sample, while the 5.7 sample looks to have continuous growth looking similar to the homogenous cells plated on day 0.

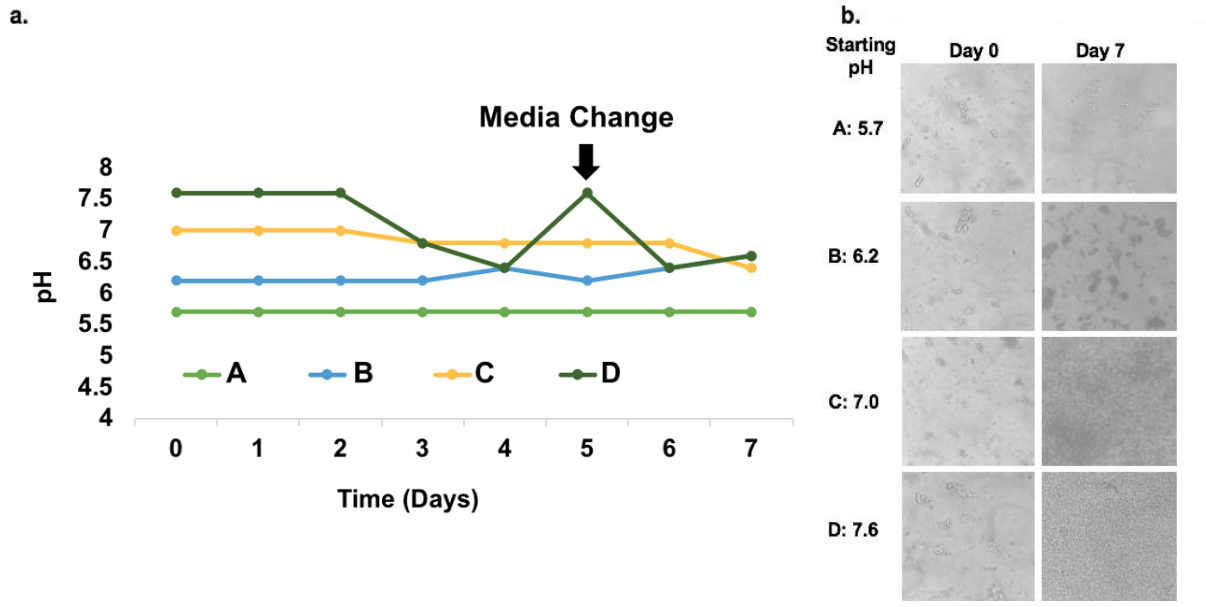

Figure 3.3. (a) The change in pH of the 4 testing groups of MaSp primary cells, testing for continuous media pH growth of cells. (b) The visual changes, using 25x brightfield, of the 4 varying media pH groups of primary cells compared at day 0 and day 7.

Effect of Decreasing Media pH on MaSp Cells

MaSp cells were processed and plated as described in methods. Cells from the homogenous mixture were reserved from processing to confirm that the correct primary cells were isolated and maintained from the tissue of major ampullate gland were being tested, discussed further in this paper. Figure 3.4 shows the visual changes of samples 1 and 3 in the cells on days 0, 2, 5, 8, and 16 (listed in Table 3.4). The row displaying a constant pH of 5.7 shows the homogenous mixture of cells growing from day 1 to day 16. The morphological characteristics of the cells appear to be unchanged. The row

displaying a decreasing pH shows a clear morphological change of the cells from day 2-5 when the pH is dropped to 7.0. By days 8 and 16 the density of the wells appears to be higher and the differences as compared to day 1 are evident.

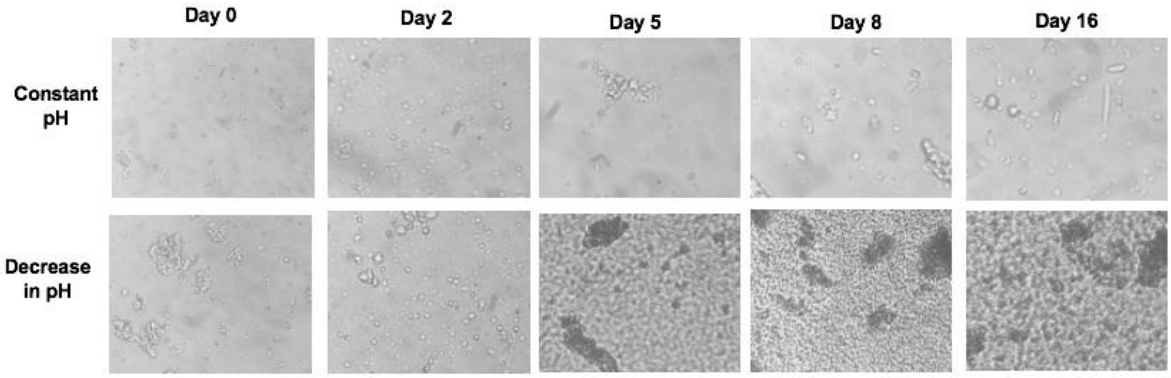

Figure 3.4. The visual changes of the MaSp cells at 25x shown on days 0, 2, 5, 8, and 16. Changes are shown by comparing continuous growth at 5.7 pH to growth of cells at decreasing pH on indicated days.

Confirmation of MaSp Cells During Decreasing Media pH

To verify that major ampullate cells were isolated and were still able to express the major ampullate silk transcript over time in culture, total RNA was isolated from cells at the time of cell processing when the homogenous mixture was initially isolated as well as on days 9 and 16 of culture. Figure 3.5a shows that after first strand cDNA was synthesized, the transcript for major ampullate silk was detected with two primer sets. Both primer sets amplified the predicted transcript at the time of creating a homogenous mixture of 8 glands. Primer set 1 detected the transcript after 16 days in pH lowering culture with a noticeable lack of transcript at day 9 (Figure 3.5b) from the same culture conditions.

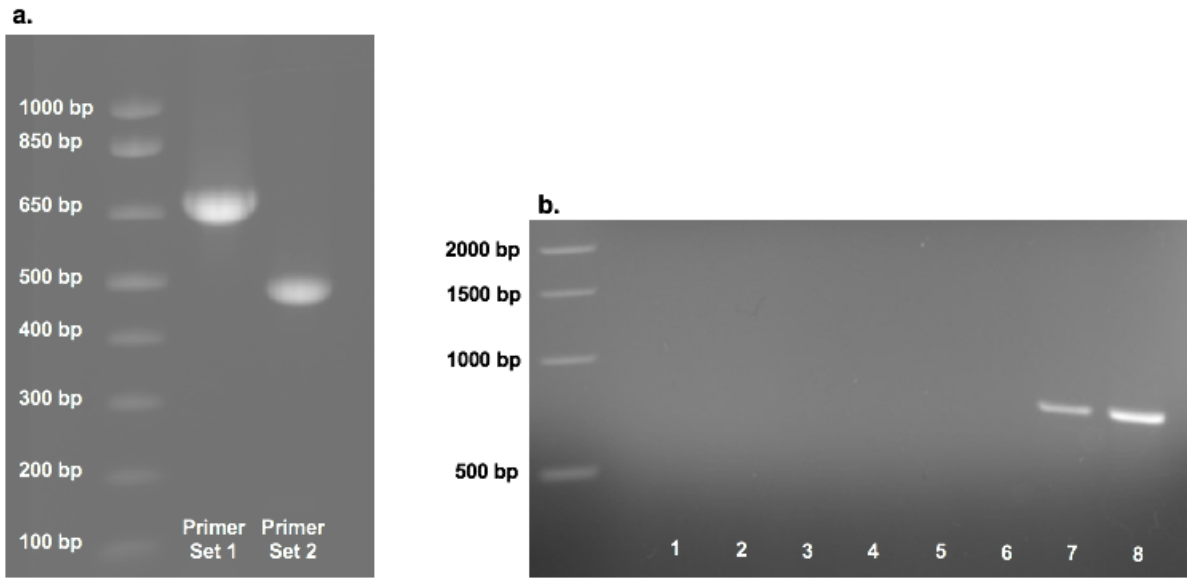

Figure 3.5.(a) Primer set 1 and primer set 2 both confirm the presence of MaSp cells in the homogenous mixture of cells at the time of seeding. (b) Column: (1) Day 9 sample 1 (2) Day 9 sample 2 (3) Day 9 sample 3 (4) Day 9 sample 4 (5) Day 16 sample 1 (6) Day 16 sample 2 (7) Day 16 sample 3 (8) Day 16 sample 4. Primer set 1 confirms the presence of silk producing MaSp cells grown at a decreasing pH after 16 days.

Discussion

To develop the growth of this unique insect primary cell system, a basic understanding of how spider silk is produced in native spider conditions is crucial, keeping in mind that a golden orb weaving spider has many glands for varying purposes: piriform, minor ampullate, major ampullate, aggregate, tubuliform, and aciniform [4]. Most studies, including the current study, focus on the major ampullate gland, which produces dragline silk not only as the structural supports of the orb web, but also as the spider's life line. Regardless of the type of gland, spider silk is a protein biomaterial and as such, it can be harnessed and manipulated. Despite many unknowns, from a simple process view, the silk protein is 1) secreted from cells that line the lumen of the gland 2) aggregated and structurally rearranged, 3) modified via phase separation through ion exchange, acidification, shear and elongational forces and 4) dehydrated as it travels through the tapered major ampullate gland to produce a fiber (Figure 3.1). Combining knowledge of the silk spinning process in golden orb weavers with the physical and biochemical properties and potential of the ampullate cells is crucial in order to advance the growth of major ampullate cells cultures. Protein aggregation and dehydration combined with the shear force

conditions and pH gradient in the major ampullate gland play a clear role in formation of dragline silk, but it is unknown the effect that they play in the silk producing cells life cycle.

One particular topic in question is how the gland controls the pH gradient throughout the spinning process, although recent evidence suggests that carbonic anhydrase may play a major role [5], [6]. Andersson et al. investigated the effects of exposing the major ampullate gland to methazolamide, a membrane-permeable carbonic anhydrase inhibitor. Exposure of the gland to the enzyme demonstrated that the collapse of the pH gradient, with the normally acidic pH stabilizing at approximately 7.0 in the duct. These results suggest that carbonic anhydrase plays a major role in the production of spider silk. One thing that can be said with certainty though is that decreasing the pH along the gland promotes a change in the conformation of the spidroins. Specifically, these conformational changes are correlated with an increase in the β-sheet structure content [17], [18]. Figure 3.1 shows a process view of all of the variables that play a role in the spinning of spider silk throughout the gland.

Sodium bicarbonate is a crucial factor in both physiological and cell culture conditions that allow the cells to buffer the environmental pH. After observing significant drops in the pH of cells cultured at a pH of 7.6 and 0.021M sodium bicarbonate (Figure 3.3a), the importance of the buffer system is clear. The high pH requires a higher concentration of sodium bicarbonate due to the growth and maintenance of cells causing a natural decrease in the pH due to the CO_2 released by the cells upon activity of carbonic anhydrase. The natural drop from pH 7.6 (Figure 3.3a) could indicate a need for an increase in the concentration of the buffer system, a change of media more often, or spiking the media with more sodium bicarbonate prior to the pH dropping.

Alternatively, growth and maintenance of cells under a pH gradient imposed and maintained by media changes has shown a significant impact on the MaSp cell growth and behavior shown in Figure 3.4. The cells showed consistent signs of growth when maintained at a consistent pH of 5.7; however, there were no signs of increased protein excretion after 7 days as indicated by a change in the viscosity of the media (data not shown). This would be a logical when considering the region of the spiders gland that is associated with the lower pH; a pH of 5.7 pH seems to be associated with the aggregation and dehydration of the silk protein as it traverses the gland toward the spinneret and fiber production [19]. This also corresponds to the region where the proteins are subjected to the biochemical gradients and

are exposed to mechanical shear elongation as shown in Figure 3.1. If the cells are grown in conditions of high acidity and low shear stress, they are not in optimal conditions to excrete the spidroins. This can be a logical explanation for the high cell viability and slow but continuous growth without signs of large amount of protein secretion.

The most significant change seen in cell morphology and growth was when the pH was maintained at 6.2. Shown in Figure 3.4, the cells begin to show the appearance of modified columnar epithelial cells similar to the appearance of characterized major ampullate cell morphology [20]. This could point to the presence of cells capable of secreting silk spidroins. At this pH in the natural spinning process, Dicko et al. showed that acidification (pH 6.35) plays a clear role in the structural transition of the spidroins to a β-rich structure. The liquid silk solution made up of spidroins at pH 6.4 has been proven to demonstrate maximum shear sensitivity [21]. Behavior of the silk protein is of high importance when dealing with the growth and maintenance of MaSp cell growth. In optimal conditions for protein excretion, the occurrence of the high volume of protein excretion is balanced with protein aggregation in the spider by passing the silk into a different zone of the gland, a process not currently possible in silk culture, potentially a significant reason the cells have traditionally been unstable in culture. It is hypothesized that significant excretion of spidroins into the media will quickly decrease cell viability due to protein aggregation. At continuous growth at pH 7.0 and pH 7.6 the cells showed consistent growth; however, a layer quickly forms on top of the cell culture media that obscures the cells, making splitting and maintenance nearly impossible and necessitating a better system to control silk protein secretion. Further adjustments to the conditions must be done to determine the optimal conditions to balance protein excretion and aggregation in vitro.

The final results of the maintenance of the MaSp cells at continuous pH levels clearly show that the pH gradient plays a crucial role in the viability and growth of the cells that line the major ampullate gland. In previous work, MaSp cells isolated as described in this study took ~6 weeks to begin excreting protein [9]. However, this past work was done with no indicated adjustment to the pH of the TNM-FH media, which is approximately pH 4.7 at time of synthesis, much lower than any condition experienced by native gland cells. The slow protein excretion in the previous study could be due to the lack of optimized conditions for the production of silk protein. It is also important to note that the protein matrix

formed will likely contain dragline silk. Further confirmation of native spider silk produced in culture should be confirmed via SDS-PAGE, amino acid analysis, and nuclear magnetic resonance (NMR). Immunohistochemistry, which is a standard protein characterization technique, is only of limited value due to the non-immunogenic nature of silk protein.

Further testing of the cell culture microenvironment on the growth of MaSp cells shows the effect of a changing pH level on the cells similar to the changing environment in the major ampullate gland. Table 3.4 shows how the morphology of MaSp cells changes as the pH is slowly decreased from pH 7.6 to pH 5.7. The growth of the cells at pH 5.7 allowed for a steady and consistent growth of the cells. The method of extracting and processing the glands produced one homogenous mixture of the cells, confirmed to contain the transcript for silk spidroin (Table 3.5, Figure 3.5a). Interestingly, the transcript was not detected after 8 days in culture but was detectable again after 16 days (Figure 3.5b), potentially indicating a lag in spidroin production as the cell culture was ramping up.

Conclusion

Future experiments can use the confirmed PCR technique to pinpoint when the cells begin to produce silk mRNA. Previous work described secreted protein as a matrix that cells attach to easily and is hard to separate, this can also be tracked in association with the presence of the silk producing mRNA to find correlations. Additionally, the presence of carbonic anhydrase is necessary in the major ampullate gland to catalyze the conversion between carbon dioxide and water and the dissociated ions. During the growth of the cells in varying conditions, the activity of carbonic anhydrase can be tested using a histochemical method to determine the varying role of the enzyme based on the conditions the cells are subjected to. Further testing can investigate the responses of these cells to other factors that play a role in the natural gland spinning process such as increasing amounts of shear stress and the inflow and outflow of ions. Finally, the development of these particular cells can be aided with the use of CRISPR-CAS9 towards immortalization and continuous production of the desirable major ampullate silk protein to produce large-scale amounts of the natural silk.

References

[1] M. Xu and R. V. Lewis, “Structure of a protein superfiber: spider dragline silk.,” *PNAS*, vol. 87, no. 18, pp. 7120–7124, Sep. 1990, doi: 10.1073/pnas.87.18.7120.

[2] M. B. Hinman and R. V. Lewis, “Isolation of a clone encoding a second dragline silk fibroin. Nephila clavipes dragline silk is a two-protein fiber.,” *J. Biol. Chem.*, vol. 267, no. 27, pp. 19320–19324, Sep. 1992.

[3] A. M. Anton, A. Heidebrecht, N. Mahmood, M. Beiner, T. Scheibel, and F. Kremer, “Foundation of the Outstanding Toughness in Biomimetic and Natural Spider Silk,” *Biomacromolecules*, vol. 18, no. 12, pp. 3954–3962, Dec. 2017, doi: 10.1021/acs.biomac.7b00990.

[4] L. Eisoldt, A. Smith, and T. Scheibel, “Decoding the secrets of spider silk,” *Materials Today*, vol. 14, no. 3, pp. 80–86, Mar. 2011, doi: 10.1016/S1369-7021(11)70057-8.

[5] M. Andersson et al., “Carbonic Anhydrase Generates CO2 and H+ That Drive Spider Silk Formation Via Opposite Effects on the Terminal Domains,” PLoS Biology, vol. 12, no. 8, p. e1001921, Aug. 2014, doi: 10.1371/journal.pbio.1001921.

[6] C. Dicko, F. Vollrath, and J. M. Kenney, “Spider Silk Protein Refolding Is Controlled by Changing pH,” *Biomacromolecules*, vol. 5, no. 3, pp. 704–710, May 2004, doi: 10.1021/bm034307c.

[7] D. Knight and F. Vollrath, “Changes in element composition along the spinning duct in a Nephila spider,” *Naturwissenschaften*, vol. 88, no. 4, pp. 179–182, Jun. 2001, doi: 10.1007/s001140100220.

[8] A. Rising, M. Widhe, J. Johansson, and M. Hedhammar, “Spider silk proteins: recent advances in recombinant production, structure–function relationships and biomedical applications,” *Cell. Mol. Life Sci.*, vol. 68, no. 2, pp. 169–184, Jan. 2011, doi: 10.1007/s00018-010-0462-z.

[9] S. K. Robinson, “Investigations of CHO 1-15 and Silk Gland Cell Line Development,” p. 133.

[10] A. E. Brooks, S. M. Stricker, S. B. Joshi, T. J. Kamerzell, C. R. Middaugh, and R. V. Lewis, “Properties of Synthetic Spider Silk Fibers Based on Argiope aurantia MaSp2,” Biomacromolecules, vol. 9, no. 6, pp. 1506–1510, Jun. 2008, doi: 10.1021/bm701124p.

[11] S. R. Fahnestock and L. A. Bedzyk, “Production of synthetic spider dragline silk protein in Pichia pastoris,” *Appl Microbiol Biotechnol*, vol. 47, no. 1, pp. 33–39, Jan. 1997, doi: 10.1007/s002530050884.

[12] J. Scheller, K.-H. Gührs, F. Grosse, and U. Conrad, “Production of spider silk proteins in tobacco and potato,” *Nature Biotechnology*, vol. 19, no. 6, pp. 573–577, Jun. 2001, doi: 10.1038/89335.

[13] H.-T. Xu et al., “Construct Synthetic Gene Encoding Artificial Spider Dragline Silk Protein and its Expression in Milk of Transgenic Mice,” Animal Biotechnology, vol. 18, no. 1, pp. 1–12, Jan. 2007, doi: 10.1080/10495390601091024.

[14] D. Williams, “Sows’ ears, silk purses and goats’ milk: new production methods and medical applications for silk.,” *Med Device Technol*, vol. 14, no. 5, pp. 9–11, Jun. 2003,

[15] F. Teule et al., “Silkworms transformed with chimeric silkworm/spider silk genes spin composite silk fibers with improved mechanical properties,” Proceedings of the National Academy of Sciences, vol. 109, no. 3, pp. 923–928, Jan. 2012, doi: 10.1073/pnas.1109420109.

[16] R. W. Work and P. D. Emerson, “An Apparatus and Technique for the Forcible Silking of Spiders,” *The Journal of Arachnology*, vol. 10, no. 1, pp. 1–10, 1982.

[17] J. M. Kenney, D. Knight, M. J. Wise, and F. Vollrath, “Amyloidogenic nature of spider silk,” European Journal of Biochemistry, vol. 269, no. 16, pp. 4159–4163, 2002, doi: 10.1046/j.1432-1033.2002.03112.x.

[18] D. P. Knight, M. M. Knight, and F. Vollrath, “Beta transition and stress-induced phase separation in the spinning of spider dragline silk,” International Journal of Biological Macromolecules, vol. 27, no. 3, pp. 205–210, Jun. 2000, doi: 10.1016/S0141-8130(00)00124-0.

[19] R. S. Wilson, “Control of Drag-Line Spinning in Certain Spiders,” *American Zoologist*, vol. 9, no. 1, pp. 103–111, 1969.

[20] M. Andersson, L. Holm, Y. Ridderstråle, J. Johansson, and A. Rising, “Morphology and Composition of the Spider Major Ampullate Gland and Dragline Silk,” *Biomacromolecules*, vol. 14, no. 8, pp. 2945–2952, Aug. 2013, doi: 10.1021/bm400898t.

[21] X. Chen, D. P. Knight, and F. Vollrath, “Rheological Characterization of *Nephila* Spidroin Solution,” *Biomacromolecules*, vol. 3, no. 4, pp. 644–648, Jul. 2002, doi: 10.1021/bm0156126.

CHAPTER 4: OPTIMIZING PROTEIN PRODUCTION USING A PROCESS ENGINEERING AND STATISTICAL APPROACH

Abstract

The production of recombinant protein has long been fundamental to the research and development of spider silk at both academic and industry scale. However, in the context of the growing variety and accessibility of commercially available recombinant expression systems, a process engineering analysis may prove beneficial to optimize production silk production and pave the way for commercial development of spider silk products. Recombinant protein production systems vary in terms of cost, complexity, and yield, with *E.coli* being the basis for one of the most established expression systems for spider silk production. *E.coli* is a desirable system due to the quick turnaround time, high yield, and relatively low cost. Although the cost of this system is low, in terms of production in an academic lab, the costs can escalate quickly, necessitating a new perspective on an old problem. Traditionally, methods of biological optimization include one variable at a time testing that works to a certain extent but tends to miss the overall process view of how the variables in the system are interacting. This study outlines the optimization of a design of experiments based on recombinant spider silk protein yield in an *E.coli* system using a Box-Behnken design, while combining elements of process engineering, statistical analyses, and biological variable breakdown. Large scale production of microbial systems introduces new variables that cause a decrease in recombinant protein yield, therefore eliminating small scale optimization as a feasible method. With the use of an Ishikawa diagram, a full process breakdown was completed to outline every variable that plays a part in the low yield of recombinant protein.

Introduction

The widespread production of recombinant protein is essential to the study of many medically important biochemical and structural processes that underlie fundamental aspects of biomedical engineering [1]. With its balance of mechanical strength and elasticity stemming from biological structural hierarchy, major ampullate silk protein is a particularly good model to explore the process of recombinant protein production in bioengineering [2]. Although native major ampullate silk can be extracted directly from the spider and despite the fact that silk is produced throughout the lifetime of the spider, the amount

of major ampullate silk produced is not sufficient to meet the commercial demand, necessitating recombinant protein expression for the high demand needed to develop biomedical engineering applications. Dissecting recombinant protein production through the lens of process engineering allows it to be divided into three techniques: genetic engineering of a plasmid, expression of the recombinant protein, and purification of the produced protein (Figure 4.1). Increasing the yield of recombinant protein production requires optimizing each of these techniques, which can be considered as sub-processes. Thus, maximizing the yield of a recombinant protein is a complex, multifactorial interaction of biological variables and processes. The specifications of the process are determined by the requirements imposed by the specific protein for production, including its size, primary sequence, and necessary post-translational modifications. Regardless, optimization of recombinant protein production is a formidable and unrelenting task.

Figure 4.1. A process flow for the production of SX2 using recombinant protein expression.

The task of protein optimization may be made more manageable when viewed as a process that can be optimized. Process engineering provides some particularly valuable tools in this regard to allow for a different method of thinking when analyzing a process or problem.

Response surface methodology (RSM) is one tool used to determine the optimum operating conditions of a system, as completed in this study. To begin, a mathematical model is selected, usually consisting of a second-order model. A few popular methods of mathematical modeling are a central composite design (CCD) and a Box-Behnken design (BBD). When comparing the two designs, one important difference is the testing of variables occurring at five different levels in the CCD. This means that testing points can fall out of range which is not feasible to test in some biological models. For

example the testing of a concentration is not possible to set at a value below zero. The five-level testing also increases the amount of test runs necessary to complete the RSM when comparing to a BBD requiring only three levels. The BBD also allows for a predetermined range to be set allowing for the control of the variables to fall in a feasible range. Once the mathematical model used in the experimental design is selected, the quality is evaluated with the application of analysis of variance (ANOVA). The optimal conditions are then fitted in a polynomial and can be visualized on a 3-D surface plot.

The advantage in a design such as RSM compared to one variable at a time testing is the ability to visualize and determine the influence of factors on one another. There are many fields of study that this can be a beneficial tool to use, and a biological setting is a prime example where time, materials, and quality is of the highest importance. The application of RSM has shown success in many studies of optimizing recombinant protein production yield [3]–[5]. Similar to these studies, the expression and purification of major ampullate spider silk was analyzed using a Box-Behnken design of experiment to visualize the effects three independent variables have on the recombinant protein yield.

Materials and Methods

Materials

LB media was obtained from IBI Scientific (USA), IPTG (100mM) was used as received from Thermo Fisher Scientific, Tween 20 was obtained from Sigma-Aldrich. All other chemicals for buffers were used as received from Thermo Fisher Scientific. Double distilled water was used throughout the experiments.

Major Ampullate Silk Protein Expression

Two repeats of the repetitive sequence of *Nephila clavipes* major ampullate silk protein (SX2) were previously cloned into the pET30a plasmid and expressed in *E.coli* BL21-DE3 with both N-terminal and C-terminal 6-histidine tags for purification. The protein identity was previously confirmed via mass spectrometry and characterized [6]. Cultures (both seeding cultures and expression cultures) were grown at 37° C in LB broth supplemented with kanamycin (50ug/µL). Expression cultures were inoculated with an overnight culture at 10% of the final expression culture volume. Initial expression conditions were selected based on procedures used to express the silk protein in *E. coli* as reported in the literature [2]. Expression of the protein was induced with 0.1mM IPTG when the OD_{600} of the expression culture read

~0.5. OD_{600} was monitored every 25 minutes during the pre-induction phase. After addition of the IPTG, the culture was incubated with shaking for 4 hours at 37° C. Subsequently, the cells were harvested by centrifugation at 3700 rpm for 12 minutes and resuspended in binding buffer (20mM Imidazole, 300mM NaCl, 50mM Na_2HPO_4, 0.05% Tween 20) before being purified and quantified.

Recombinant Silk Protein Purification

Regardless of the scale of protein production, all protein was purified using nickel affinity chromatography. Small-scale (10mL cultures) purification was done using Promega's Maxwell 16 instrument with an in-house nickel affinity purification cartridge (Appendix D) using 30µL of HisPur Ni-NTA magnetic beads (Fisher) and the buffers shown in Table 4.1. Alternatively, large-scale (1 liter cultures) purification, which was done to confirm the optimum conditions based on the Box-Behnken analysis, used HisPur Ni-NTA resin (1mL, Fisher) per ¼ large-scale culture packed into a drip column (Bio-Rad) and the same buffers. Additionally, 24.25mL of binding buffer, 25mL wash buffer/wash, and 7.5mL elution buffer was used per ¼ large-scale culture.

Table 4.1. Buffer concentrations used in the purification process.

Purification Buffers	Imidazole Concentration	pH	Tween
Binding Buffer	20 mM	8	0.05%
Wash Buffer	100 mM	8	0.05%
Elution	500 mM	8	0 %

Protein Quantification

Purified silk protein was quantified using a Take-3 plate on an Epoch spectrophotometer (BioTek, USA). A 4-12% gradient SDS-PAGE gel (Invitrogen, USA) was run at 155V to separate the protein. Subsequently, the gel was stained with Coomassie Blue (Bulldog Bio, USA) to assess the protein's purity and size. A dot blot was completed to confirm the correct expression of the SX2 protein.

Dot Blot

Protein samples were directly spotted onto a nitrocellulose membrane (Thermo Scientific) and allowed to dry. Subsequently, the membrane was incubated in a blocking solution (30ml-TBST and 1.5g instant milk) for 1 hour. An HRP-tagged anti-6X-His Monticlonal Antibody (Invitrogen) was diluted 1:2500

in the blocking solution and the blot was allowed to incubate for 1 hour at room temperature on a rotating platform. Subsequently, three washes (25ml-TBST) were done for 15 minutes each to remove any unbound antibody before reacting with ECL Western Blotting Substrate (Promega, USA). The blot was visualized on an Aplegen Omega Lum G with a 30 second exposure time.

Experimental Design

Three independent variables with a reported impact on recombinant protein yield were identified: temperature after induction (A), the optical density at induction (OD_{600}) (B), and amount of IPTG (C) [3]–[5]. A range was determined for each variable based on either reported literature values or as a factor of the reported values. Each of these values was coded with a level (-1, 0, and 1) (Table 4.2). After identifying and coding each of these variables and their values, a matrix was designed (Figure 4.2) to test the protein yield from each combination of conditions using small-scale (10 mL) protein expression cultures (Table 4.3). The experiment number was determined according to $N = 2k(k-1) + c_p$, where k is the number of factors and c_p is the number of central points. With ($k = 3$) and ($c_p = 5$), the number of experimental designs in this study is 17 as shown in Table 4.3. The polynomial equation generated by this experimental design (using Minitab 18) is as follows:

$$Y_i = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_{11}{X_1}^2 + b_{22}{X_2}^2 + b_{33}{X_3}^2 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3$$

Where Y_i is the dependent variable (SX2 Concentration), b_0 is the intercept, b_1 to b_{33} are the regression coefficients, and X_1, X_2, and X_3 are the independent variables.

Table 4.2. Coded variables and level values used for the Box-Behnken design.

Independent Variable	Levels		
	-1	0	1
A: Temperature (°C)	32.0	34.5	37.0
B: OD_{600}	0.4	0.6	0.8
C: IPTG (mM)	0.1	1.0	1.9

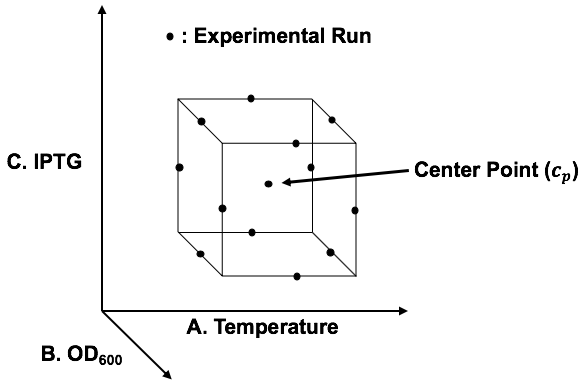

Figure 4.2. A visual representation of the experimental run testing points using a Box-Behnken design.

Table 4.3. Box-Behnken design of process variables A = Temperature, B = OD_{600} at the time of induction, C = concentration of IPTG used for induction.

Run	Coded Variables			Actual Variables		
	A	B	C	A	B	C
1	-1	-1	0	32.0	0.4	1.0
2	1	-1	0	37.0	0.4	1.0
3	-1	1	0	32.0	0.8	1.0
4	1	1	0	37.0	0.8	1.0
5	-1	0	-1	32.0	0.6	0.1
6	1	1	-1	37.0	0.8	0.1
7	-1	0	1	32.0	0.6	1.9
8	1	0	1	37.0	0.6	1.9
9	0	-1	-1	34.5	0.4	0.1
10	0	1	-1	34.5	0.8	0.1
11	0	-1	1	34.5	0.4	1.9
12	0	1	1	34.5	0.8	1.9
13	0	0	0	34.5	0.6	1.0
14	0	0	0	34.5	0.6	1.0
15	0	0	0	34.5	0.6	1.0
16	0	0	0	34.5	0.6	1.0
17	0	0	0	34.5	0.6	1.0

Results

Analysis of Variance (ANOVA)

According to the experimental design, 17 experiments were performed to optimize the recombinant protein expression method to get maximum protein yield. The obtained results (Table 4.4) were entered in Minitab 18 (Table 4.5). As show in in Table 4.5, the model F value of 18.10 implies that the model is statistically significant. According to the P value, there is a <.001% chance that the F Value occurs due to noise.

Table 4.4. Experimental runs listed with responses to change in variables.

Run	Actual Variables			Response (mg/ml)
	A	B	C	Actual
1	32.0	0.4	1.0	0.3285
2	37.0	0.4	1.0	0.2270
3	32.0	0.8	1.0	0.3385
4	37.0	0.8	1.0	0.3265
5	32.0	0.6	0.1	0.4140
6	37.0	0.8	0.1	0.2925
7	32.0	0.6	1.9	0.3265
8	37.0	0.6	1.9	0.3670
9	34.5	0.4	0.1	0.3165
10	34.5	0.8	0.1	0.3640
11	34.5	0.4	1.9	0.3015
12	34.5	0.8	1.9	0.3205
13	34.5	0.6	1.0	0.3335
14	34.5	0.6	1.0	0.3490
15	34.5	0.6	1.0	0.3345
16	34.5	0.6	1.0	0.3395
17	34.5	0.6	1.0	0.3420

Table 4.5. ANOVA Results.

Source	Sum of Squares	df	Mean Square	F Value	Prob.
Model	0.022436	9	0.002493	18.10	<0.001
A	0.005096	1	0.005096	36.99	<0.001
B	0.003685	1	0.003685	26.75	0.001
C	0.00320	1	0.003200	2.32	0.171
AB	0.001815	1	0.001815	13.17	0.008
AC	0.006656	1	0.006656	48.32	<0.001
BC	0.000123	1	0.000123	0.89	0.377
A^2	0.000209	1	0.000209	1.52	0.258
B^2	0.002877	1	0.002877	20.88	0.003
C^2	0.000687	1	0.000687	4.99	0.061
Residual	0.000157	4	0.000039		
Lack of Fit	0.000807	3	0.000269	6.84	0.047

Final Equation in Coded Factors

$$Y\ [SX2\ (mg/ml)] = 0.34041 - 0.02659 *A + 0.02112 *B - 0.00666 *C - 0.00729 *A^2 - 0.02888 *B^2 + 0.01321 *C^2 + 0.02061 *A *B + 0.04505 *A *C - 0.00536 *B *C$$

Final Equation in Actual Factors

$$SX2\ Concentration\ (Y) = 0.34041 - 0.02659 *Temperature + 0.02112 *OD_{600} - 0.00666 *IPTG\ Concentration - 0.00729 *Temperature^2 - 0.02888 *OD_{600}^2 + 0.01321 *IPTG\ Concentration^2 + 0.02061 *Temperature *OD_{600} + 0.04505 *Temperature *IPTG\ Concentration - 0.00536 *OD_{600} *IPTG\ Concentration$$

The final equation listed above is a representation of the quantitative effect the independent variables have on the response, or the values of the variables A, B, and C on the response Y.

Response Surface Methodology

Figure 4.3 shows the response surface 3-D plots and the contour maps associated with the comparison of all variables. A decrease in the IPTG along with an increase in the OD_{600} shows an increase in protein response Figure 4.3a. The decreased temperature along with an increase in the OD_{600} shows in increase in the protein response, shown in Figure 4.3b. There appears to be a large protein increase as the temperature and IPTG is decreased and OD_{600} is held constant, shown in Figure 4.3c.

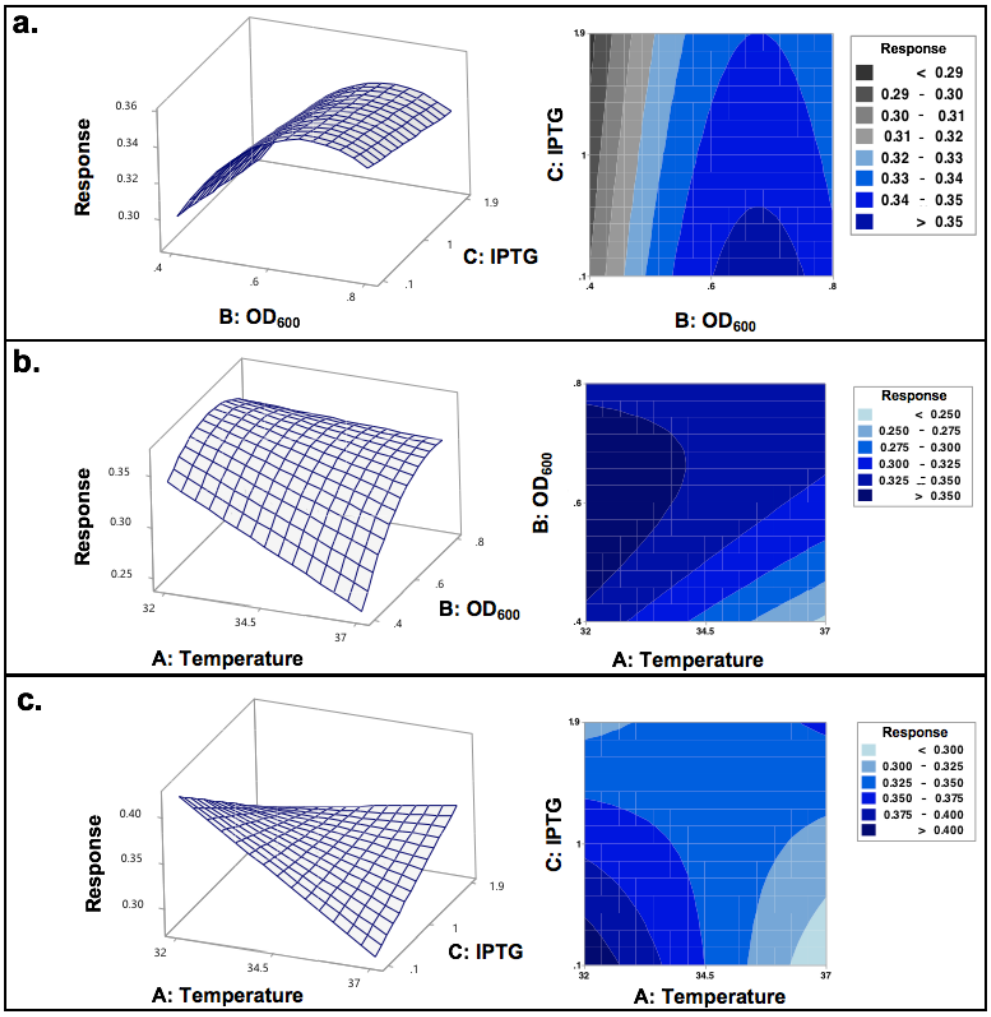

Figure 4.3. Response surface graphs and contour maps display the effects of the compared variables on the response: OD_{600} and IPTG (a), temperature and OD_{600} (b), and temperature and IPTG (c).

Large-scale Testing of Optimized Variables

The Box-Behnken results showed a clear set of optimized conditions to maximize recombinant spider silk production; however, the model was obtained using data from small-scale expression cultures, which have limited commercial utility. Hence, it was important to consider the large-scale translation of these conditions. 1L expression cultures were grown under the pre-optimization conditions and the post-optimization conditions (Table 4.6). The protein concentrations indicate no significance difference in recombinant protein yield upon scale up, as indicated by the OD_{280} as well as the dot blot (Figure 4.4b). Based on an estimate from visual results (Figure 4.4a), it is clear that the protein purified from cultures grown in the pre-optimized conditions is purer. The before optimization value shows a distinct band for the SX2 protein that is desired with minimal other proteins co-eluting. The after-optimization conditions

show a distinct band of SX2 along with multiple bands of other sizes that contribute to overall protein yield but are not the desirable protein of the expression.

Table 4.6. Variables of protein expression before and after optimization with results of large scale testing of before and after optimization variables.

Variable	After	Before
Temperature (°C)	32	37
OD600 (nm)	0.65	0.5
IPTG (mM)	0.1	0.1
SX2 Concentration (mg/mL)	0.502	0.505
Purity	85%	95%
Calculated SX2 Concentration (mg/mL)	0.427	0.480

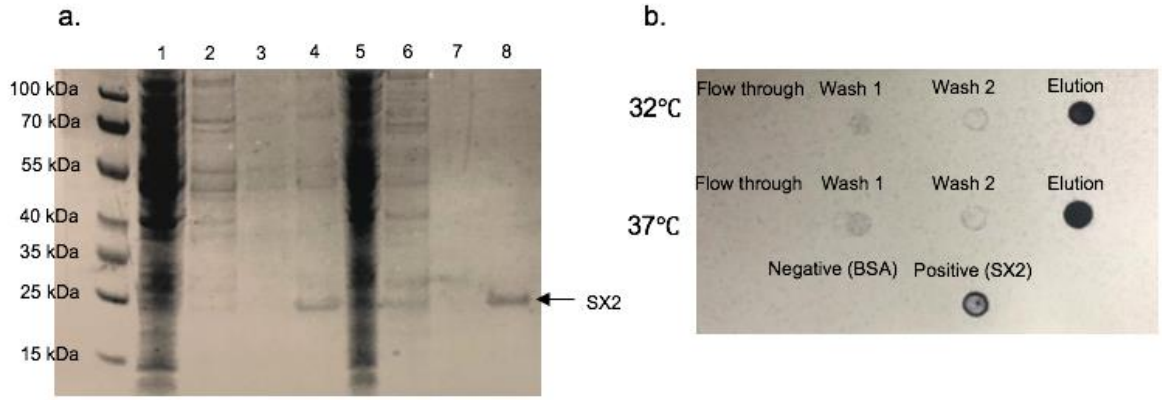

Figure 4.4. (a) SDS-PAGE corresponding to each of the conditions shown on the dot blot. “After optimization” conditions (32℃): (1) Flowthrough (2) Wash 1 (3) Wash 2 (4) Elution. “Before optimization” conditions (37℃): (5) Flowthrough (6) Wash 1 (7) Wash 2 (8) Elution. Note the apparent increase in purity of the eluted protein from 37°C. (b) Dot blot comparing large-scale expressions of the predicted optimum induction temperature (32°C) based on the small-scale results and the standard conditions reported in the literature (37°C). The flow through, washes, and elution from large-scale purification are all shown in comparison with a previously characterized SX2 protein and BSA as the negative control.

Discussion

This study aims to determine the effects that cell density (OD_{600}) prior to induction, post-induction temperature, and inducer concentration (IPTG) (Table 4.1) have on one another as part of an overall optimization process for recombinant silk protein. Although these are not the only biological variables that may have an impact, these were chosen as the easiest to isolate their effect on protein yield; hence, all other conditions of expression were kept consistent. Three levels for each variable range were identified

as low, intermediate, and high and assigned a label (Table 4.2). A matrix of small-scale expression cultures (i.e., runs) was designed and run (Table 4.4). Based on the data from these small scale-expression cultures, an ANOVA was performed to assess the quality of the mathematical model. The overall model F value was 18.10 with the P value showing there is <0.001% chance of occurrence by randomness. Based solely on P-value testing, it was concluded that variables B (OD_{600}) and C (IPTG concentration) are not significantly different. However, when visually plotting the data using a Box-Behnken analysis there is clearly a local maximum formed near the cell density value of 0.65 and IPTG concentration value of 0.1mM (Figure 4.3a). Thus, even with these results considered statistically insignificant by ANOVA, based on a deeper analysis of interacting variables, it can still be concluded that these values will lead to maximum protein production (Table 4.5). Taken together these results pointed to optimized conditions of the variables at an OD_{600} value of 0.65, an IPTG concentration of 0.1mM, and a post-induction temperature of 32°C.

Previous efforts to scale protein production have recognized that for certain proteins, particularly self-aggregating proteins like silk, a simple volumetric scaling may be inadequate and additional optimizations may be needed [7]. However, part of the power of predictive mathematical models, such as the Box-Behnken design used in this study, lies in the scalability of the model. Hence, the optimized conditions predicted by Box-Benhken based on small-scale expression cultures were applied to large-scale 1-liter cultures. Unfortunately, when comparing the protein yield from conditions commonly used before optimization with those after optimization (Table 4.6), it was clear that the overall amount of protein did not change. Interestingly, the protein purified from pre-optimization conditions appeared to be purer (Table 4.6), indicating that the amount of silk protein yielded may have in fact been less after optimized conditions were applied to larger-scale expression.

The lower yield of recombinant protein in large-scale expression cultures using the predicted optimized values does not necessarily point to a failure in the Box-Behnken analysis although it is somewhat surprising since high protein translation rates (dependent on the specific protein) have been noted at low temperatures consistent with our analysis [8], [9]. Large-scale batches introduce new considerations that affect the growth of *E.coli,* giving rise to higher cell densities that range from 10 to 100 g L^{-1} cell dry weight [10]. At these densities, the consumption of glucose and oxygen can quickly lead

to oxygen depletion leading *E.coli* to shift to anaerobic respiration or fermentative metabolism, potentially increasing the production of metabolites such as formate, acetate, lactate, ethanol, and succinate and consequently lowering the pH [11]–[14]. Future testing can consider this knowledge make changes to the upstream processing to determine if the concentration of SX2 can be improved [15], [16]. An additional potential explanation for such results may lie in the kinetics of heat diffusion and the lower induction temperature used after optimization.

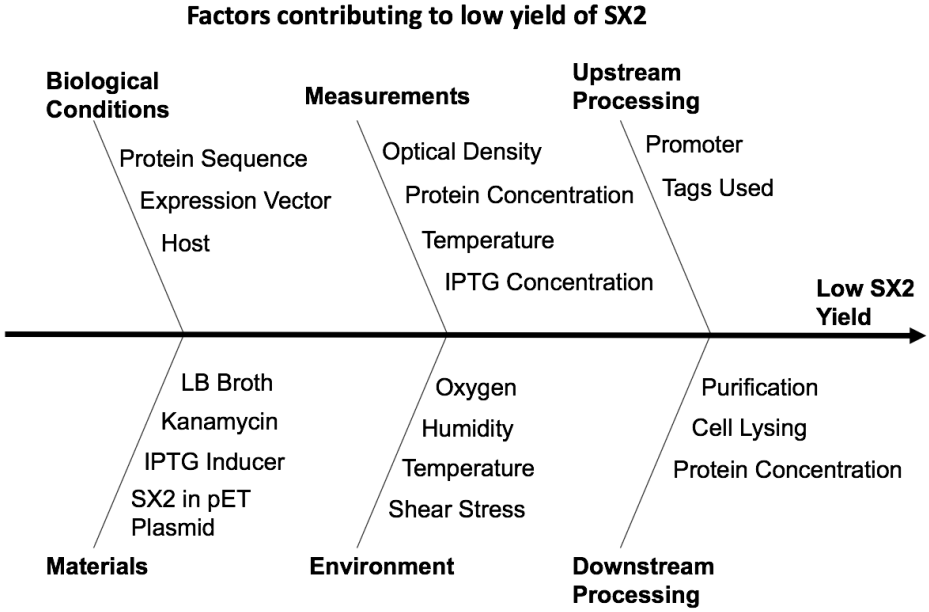

Figure 4.5. An Ishikawa diagram used to identify "root causes" in a process. The consideration of a low SX2 yield is caused by 6 main categories, and each category is further described with examples from the SX2 production process.

Although Box-Behnken methods were able to successfully analyze and provide a prediction based on the combination of biological variable contemplated in this study, the analysis is limited by the range of specific conditions considered and can point only a local maximum within the range of specific conditions considered. Further testing should be done to consider the effect of other factors that have been shown to have an influence in the overall production of the SX2 protein in large-scale batches. Factors such as the shear force experienced due to oxygenation, the pre-induction temperature, additional supplements to the media and post-induction time are further studies that can be done to determine the specific optimal values for the SX2 protein [17]–[19]. Previously, supplements, such as

lactose, were added that have been shown to change the results in large-batch testing based on dissolved oxygen (DO) concentrations and nutrient availability in high cell density conditions [20].

Treating recombinant protein production as a manufacturing process allows us to apply “root cause” analysis tools not traditionally used for biological techniques, which leads to not only a unique and enlightening perspective on the problem but also produces an Ishikawa diagram (Figure 4.5). With the main concern being a low SX2 yield, the factors that can cause low SX2 yield are listed in the main branches of the figure; biological conditions, materials, measurements, environment, upstream processing, and downstream processing. These main branches are then expanded on with specific examples that are applicable to the specific SX2 production processes. This tool is commonly used to identify other variables, or “root causes”, to consider when solving for the problem of low SX2 yield. The variables tested using statistical methods in this paper are seen as sub-categories in the diagram. However, as is clear from the Ishikawa diagram, only a small subset of impacting variables were considered in this study. Ultimately, the optimization of a biological process requires consideration of both biological and manufacturing constraints. It is also important to note that in the assessment of protein yield both expression and purification were considered as a single process; however, all of the variables considered in this study lie in the process of expression. Thus, the measurement of protein yield after purification is an indirect measure of the efficiency of the process.

Conclusion

In the world of manufacturing, it is common to approach the production of an output as a linear process from beginning to end. These processes can first be broken down at a high-level view, then can further be broken down into sub-processes. When problem solving in a process, considerations such as the environment, materials, and people involved in the process are a few of the factors that are considered. Biological processes can use this same logic to consider every variable and process step that can have an effect on the process/output. A combination of statistical methods, process breakdowns, and knowledge of biological techniques can be used to make improvements to the overall system. This paper focuses on a statistical analysis, Response Surface Methodology using a Box-Behnken design, for the optimization of protein yield, but it is clear that a full breakdown and understanding of the processes will benefit from using many different tools to fully optimize the production of silk protein.

References

[1] T. Lefèvre, S. Boudreault, C. Cloutier, and M. Pézolet, “Conformational and Orientational Transformation of Silk Proteins in the Major Ampullate Gland of Nephila clavipes Spiders,” *Biomacromolecules*, vol. 9, no. 9, pp. 2399–2407, Sep. 2008, doi: 10.1021/bm800390j.

[2] F. Teulé *et al.*, “A protocol for the production of recombinant spider silk-like proteins for artificial fiber spinning,” *Nat Protoc*, vol. 4, no. 3, pp. 341–355, 2009, doi: 10.1038/nprot.2008.250.

[3] C. P. Papaneophytou, V. Rinotas, E. Douni, and G. Kontopidis, “A statistical approach for optimization of RANKL overexpression in Escherichia coli: Purification and characterization of the protein,” *Protein Expression and Purification*, vol. 90, no. 1, pp. 9–19, Jul. 2013, doi: 10.1016/j.pep.2013.04.005.

[4] A. L. Larentis, A. P. C. Argondizzo, G. dos S. Esteves, E. Jessouron, R. Galler, and M. A. Medeiros, “Cloning and optimization of induction conditions for mature PsaA (pneumococcal surface adhesin A) expression in Escherichia coli and recombinant protein stability during long-term storage,” *Protein Expression and Purification*, vol. 78, no. 1, pp. 38–47, Jul. 2011, doi: 10.1016/j.pep.2011.02.013.

[5] Y. Wang, C. Jing, B. Yang, G. Mainda, M. Dong, and A. Xu, “Production of a new sea anemone neurotoxin by recombinant Escherichia coli: Optimization of culture conditions using response surface methodology,” *Process Biochemistry*, vol. 40, no. 8, pp. 2721–2728, Jul. 2005, doi: 10.1016/j.procbio.2004.12.024.

[6] J. T. Prince, K. P. McGrath, C. M. DiGirolamo, and D. L. Kaplan, “Construction, Cloning, and Expression of Synthetic Genes Encoding Spider Dragline Silk,” *Biochemistry*, vol. 34, no. 34, pp. 10879–10885, Aug. 1995, doi: 10.1021/bi00034a022.

[7] L. Baldi, D. L. Hacker, M. Adam, and F. M. Wurm, “Recombinant protein production by large-scale transient gene expression in mammalian cells: state of the art and future perspectives,” *Biotechnology Letters*, vol. 29, no. 5, pp. 677–684, Mar. 2007, doi: 10.1007/s10529-006-9297-y.

[8] J. A. Vasina and F. Baneyx, “Expression of Aggregation-Prone Recombinant Proteins at Low Temperatures: A Comparative Study of theEscherichia coli cspAandtacPromoter Systems,”

Protein Expression and Purification, vol. 9, no. 2, pp. 211–218, Mar. 1997, doi: 10.1006/prep.1996.0678.

[9] S. Phadtare and K. Severinov, “Extended -10 motif is critical for activity of the cspA promoter but does not contribute to low-temperature transcription,” *J. Bacteriol.*, vol. 187, no. 18, pp. 6584–6589, Sep. 2005, doi: 10.1128/JB.187.18.6584-6589.2005.

[10] K. Hellmuth, D. J. Korz, E. A. Sanders, and W.-D. Deckwer, “Effect of growth rate on stability and gene expression of recombinant plasmids during continuous and high cell density cultivation of Escherichia coli TG1,” Journal of Biotechnology, vol. 32, no. 3, pp. 289–298, Feb. 1994, doi: 10.1016/0168-1656(94)90215-1.

[11] J. Soini, K. Ukkonen, and P. Neubauer, “High cell density media for Escherichia coli are generally designed for aerobic cultivations – consequences for large-scale bioprocesses and shake flask cultures,” *Microb Cell Fact*, vol. 7, p. 26, Aug. 2008, doi: 10.1186/1475-2859-7-26.

[12] A. Castan and S.-O. Enfors, “Formate accumulation due to DNA release in aerobic cultivations of Escherichia coli,” *Biotechnology and Bioengineering*, vol. 77, no. 3, pp. 324–328, 2002, doi: 10.1002/bit.1198.

[13] R. G. Sawers, “Formate and its role in hydrogen production in Escherichia coli,” *Biochem Soc Trans*, vol. 33, no. 1, pp. 42–46, Feb. 2005, doi: 10.1042/BST0330042.

[14] M. De Mey, S. De Maeseneire, W. Soetaert, and E. Vandamme, “Minimizing acetate formation in E. coli fermentations,” *J Ind Microbiol Biotechnol*, vol. 34, no. 11, pp. 689–700, Nov. 2007, doi: 10.1007/s10295-007-0244-2.

[15] M. Kim, C. Elvin, A. Brownlee, and R. Lyons, “High yield expression of recombinant pro-resilin: Lactose-induced fermentation in E. coli and facile purification,” *Protein Expression and Purification*, vol. 52, no. 1, pp. 230–236, Mar. 2007, doi: 10.1016/j.pep.2006.11.003.

[16] Z. Fang, C. Sha, Z. Peng, J. Zhang, and G. Du, “Protein engineering to enhance keratinolytic protease activity and excretion in Escherichia coli and its scale-up fermentation for high extracellular yield,” *Enzyme and Microbial Technology*, vol. 121, pp. 37–44, Feb. 2019, doi: 10.1016/j.enzmictec.2018.11.003.

[17] A. L. Larentis *et al.*, “Evaluation of pre-induction temperature, cell growth at induction and IPTG concentration on the expression of a leptospiral protein in E. coli using shaking flasks and microbioreactor,” *BMC Res Notes*, vol. 7, no. 1, p. 671, Sep. 2014, doi: 10.1186/1756-0500-7-671.

[18] F. R. Schmidt, “Recombinant expression systems in the pharmaceutical industry,” *Appl Microbiol Biotechnol*, vol. 65, no. 4, pp. 363–372, Sep. 2004, doi: 10.1007/s00253-004-1656-9.

[19] D. C. Andersen and L. Krummen, “Recombinant protein expression for therapeutic applications,” *Current Opinion in Biotechnology*, vol. 13, no. 2, pp. 117–123, Apr. 2002, doi: 10.1016/S0958-1669(02)00300-2.

[20] C. Khosla and J. E. Bailey, “Heterologous expression of a bacterial haemoglobin improves the growth properties of recombinant Escherichia coli,” *Nature*, vol. 331, no. 6157, pp. 633–635, Feb. 1988, doi: 10.1038/331633a0.

CHAPTER 5: CONCLUDING REMARKS

Future Work

Recombinant Silk Protein Production

After a thorough analysis of the methods to feasibly obtain spider silk, there are two paths that one could take to further develop the systems. On one hand, there is a developed system in place that can currently produce recombinant spider silk (~25 kDa) for low demand techniques such as drug delivery and blood clot reducing technology. The amounts produced in the current system are enough for small-scale research, but would benefit greatly with large-scale optimization. This master's thesis explored the possibility of optimizing the current system using small-scale batches to reduce processing costs and create better control of the testing replicates. It was determined that the calculated optimal values do not translate over to large-scale production, leading to future work in translation of optimization techniques with large-scale replicates. However, the caveat to this is the problems with large-scale production will still be significant in the results. Biological optimization using biological variables (IPTG concentration, temperature after induction, OD_{600} etc.) should occur after tackling the problem of biological constraints imposed by large batch growth (i.e. formation of inclusion bodies, O_2 deprivation, etc.). Furthermore, all variables assessed using the Box-Behnken DOE can be re-assessed by pushing the "limits" when assigning level values of the variables to be tested. While the specific values used in this work were based on cited literature and common practice in the lab, it can be noted that the testing levels are meant to depict the wide ranges of resulting outputs and the range can be expanded in future work to better visualize the limitations of the variables and the impact they have on the output. To consider all other variables that cause low yield of the recombinant silk protein, the use of the Ishikawa diagram is beneficial to include all factors of the process that may be contributing to the low yield. The separation of three individual processes that contribute to the production of recombinant protein sheds light on the fact that there are three different outputs that must be used to assess every individual process. Since the first process, development and insertion of the plasmid into the host, was already completed, the last two processes should be optimized using the techniques described in this thesis in which the output from the protein purification step was used to optimize the recombinant protein expression process. The use of the output from one process for the analysis of another can cause discrepancies in the analysis, due to the

number of variables that can be contributing to the process yield. One final process that occurs after protein purification is the assessment of the final quality of the recombinant protein. This process is a heavy consideration in both biological and engineering concepts; however, changing biological variables may unintentionally impact this. The quality of the final product is defined by the application of the final recombinant protein, and can be assessed and optimized with assays specific to the application (i.e. purity, biological activity, endotoxin levels, and protein sequence accuracy). In order to properly optimize the recombinant protein expression system in future work, the following techniques can be used: the use of large-scale batches after accounting for loss of protein due to growth at large densities, quality control of the final protein output, and analysis of the efficacy of expression prior to protein purification (i.e. densitometry of dot blot from flow through) to analyze the efficacy of expression prior to protein purification.

Natural Silk Protein Production

The ability to create a system to produce natural spider silk entails its own advantages, primarily the natural length and structure of the protein. Additionally, the major ampullate silk gland is known to produce and hold highly concentrated silk solutions, a good omen for high yield cultures. Hence, the development of a cell line that excretes the desirable dragline silk would be a huge step forward for future pharmaceutical and industrial applications of spider silk. This thesis determined that replicating the pH and sodium bicarbonate conditions of the major ampullate gland in *N.clavipes* significantly affects the growth of the MaSp cells. The biological conditions that are known to cause biochemical and mechanical changes of spider silk proteins (i.e. pH, ionic and fluid flow gradients) can be further tested with cell culture growth of primary cells extracted from the tissue to determine if they are variables that cause MaSp cells to secrete the spider silk protein. Furthermore, such insight into the biology of silk secretion may allow the future immortalization of the cells and the development of new culturing techniques. Under conditions where the cells secreted the presumptive spider silk protein matrix, cell culture became particularly difficult (e.g., inability to split, count, and remove the protein from the wells) using traditional cell culture techniques. The presence of carbonic anhydrase in the growth of the MaSp cells is necessary to catalyze the conversion between carbon dioxide, water, and dissociated ions. This activity can be tested using immunohistochemistry to ensure that culture conditions that affect the varying role the

enzyme plays in the growth of the cells are appropriate to produce silk. As with most primary cell culture, the long-term growth of these cells presents the challenge of potential de-differentiated; however, this thesis confirmed that after 16 days of growth isolated MaSp cells were still capable of producing the transcript for silk protein. The production of silk mRNA can be explored in future work by extracting cells on a daily basis to better understand when the mRNA is formed, under the varying biological growth conditions. This understanding of the mRNA production along with the presence of the silk protein matrix in the media will lead to additional biologic insight into spiders and their silk.

APPENDIX A: TNM-FH MEDIA PREPARATION PROTOCOL

Purpose: To prepare 100mL sterile media used for major ampullate cell extraction and growth.

Materials:

Gentamicin (10 ug/mL)

Amphotericin (0.25 ug/mL)

Tetracycline (10 ug/mL)

Oxalacetic Acid (9.5 mM)

Insulin (8.3 ug/mL)

Fetal Bovine Serum (FBS)

TNM-FH Insect Media

Sodium Bicarbonate (1M stock in solution)

NaOH (1M)

HCL (5M)

SFCA .22um 30mm diameter syringe filter non-pyrogenic

70µm nylon mesh cell strainer

Methods:

1. Measure out 80mL of distilled H_2O into a beaker, put a small stir bar in and set on a stir plate.
2. Thaw all the frozen aliquoted stocks from the -20^{o}C freezer in warm water (Gentamicin, Amphotericin, Tetracycline, Oxalacetic Acid, and 10% FBS).
3. Add 1000µL of Gentamicin (1mg/mL) to the beaker.
4. Add 25µL of Amphotericin (1mg/mL) to the beaker.
5. Add 1000µL of Tetracycline (1mg/mL) to the beaker.
6. Add 950µL of Oxalacetic (1 M) Acid to the beaker.
7. Add 83µL of Insulin (10mg/mL) to the beaker.
8. Add 5.06g of TNM-FH Insect Media to the beaker.
9. If adding FBS, add 10 mL of 10% FBS to the beaker.

10. Then pour the media into a 100mL graduated cylinder and add distilled water until the media is at the 100mL mark on the graduated cylinder. Pour the media back into the beaker and set back onto the stir plate.
11. Aliquot media to desired amount of 50mL tubes.
12. Add sodium bicarbonate (1M solution) to reach desired concentration.
13. Use 1M NaOH or 1M HCL to titrate the media to the designated pH while the media is mixing.
14. Use a syringe and syringe filter (0.22µm SFCA) to filter the media into 50mL conicals. Then label the conicals with: TNM-FH with or without media depending on what type was made, the pH level of the media, Sterile, Date, and Initials.
15. Store media in the fridge.

APPENDIX B: GLAND EXTRACTION PROTOCOL

Purpose: To extract major ampullate glands from golden orb weaving spiders

Materials:

Sterile equipment (tweezer, fine point tweezer, wide tweezer, and scissors)

Petri Dish

Cryovials

70% Ethanol

Sterile Sodium Citrate

TNM-FH Media with FBS

DMSO

Methods:

1. After knocking out the spider with CO_2, disinfect the spider by submerging in a beaker of 70% Ethanol for 3-5 mins.
2. Remove the cephalothorax from abdomen of the spider and place the abdomen in the dry petri plate.
3. Cut down the center of the abdomen.
4. Remove the contents of the abdomen onto petri dish.
5. Use the tweezer and wide tweezer to look for the glands and pull out the glands from the abdomen and place in sodium citrate on petri dish.
6. Locate the major ampullate glands in the sodium citrate and place both of them in one cryovial.
7. If freezing glands for later use: add 900µLof TNM-FH media with FBS and add 100µL of DMSO, store glands at -80^{o}C.
8. If immediately processing glands: refer to gland processing protocol.

APPENDIX C: GLAND PROCESSING PROTOCOL

Purpose: To process major ampullate cells for growth in plates

Materials:

Sterile scissors

70 µm Nylon Mesh Cell Strainer

50mL conical

6 well plate

Microcentrifuge tube

TNM-FH media with FBS

TNM-FH media without FBS

0.05% Trypsin with 0.02% EDTA

Collagenase

Sodium Citrate

Methods:

1. Obtain glands

If frozen:

a. Thaw the frozen glands quickly.

b. Pull the glands out with a tweezer and set them into a different microcentrifuge tube and add 500µL of sodium citrate to the tube.

c. Pull off the 500µL of sodium citrate and add another 500µL of sodium citrate to rinse a second time to take off any residual DMSO.

If fresh from a spider:

a. Follow the Extraction of Major Ampullate Glands in Appendix B.

2. With a sterile tweezers take the glands out of the sodium citrate and set them into a clean microcentrifuge tube and add 400µL of TNM-FH media without FBS.

3. Use the sterile scissors to mince the glandular tissue.

4. After the tissue is minced add 200µL of 0.05% trypsin with 0.02% EDTA and incubate at 37^{o}C for 5 minutes.

5. Add 400µL of collagenase to the microcentrifuge tube so the final concentration is 1 mg/mL and incubate at 37^{o}C for 10 minutes.

6. Add 200µL of TNM-FH media with FBS and then centrifuge the cells at 400 RPG for 7 mins.

7. Discard the supernatant and the gently wash the cell pellet with 100µL of TNM-FH media with FBS.

8. Then centrifuge the cells again at 400 rpg for 7 mins.

9. Discard the supernatant and then add 500µL of TNM-FH media with FBS to resuspend the cell pellet.

10. Put the cell strainer on the 50mL conical to strain the cells into the conical and make them sterile.

11. Move the cells from the 50mL conical to a 6 well plate.

12. Add 1.5mL of TNM-FH media with FBS so the final volume of medium is 2mL.

13. Store the cells at 27^{o}C with no CO_2

APPENDIX D: IN-HOUSE MAXWELL PURIFICATION

Materials

Empty Maxwell cartridge (Promega)

Elution tube (Promega)

Plunger (Promega)

Binding buffer (20mM Imidazole, .05% tween, 8.0 pH)

Wash buffer (100mM Imidazole, .05% tween, 8.0 pH)

Elution buffer (500mM imidazole, 8.0 pH)

HisPur NI-NTA magnetic beads (Thermo Fisher Scientific)

DNase

1X Fastbreak cell lysis reagent (Promega)

Methods

Buffer concentrations were determined prior to purification protocol. Binding buffer imidazole concentrations were tested at 5mM, 10 mM, and 20mM, with highest protein yield being observed with 20mM imidazole concentrations. Wash buffer imidazole concentrations were tested at 60mM and 100mM, with 100mM showing highest protein yield. Elution buffer imidazole concentrations were tested at 250mM and 500mM, with 500mM being selected to ensure no protein remained binded to the Ni-NTA beads. Binding and wash buffers were used with 0.05% Tween20 and without Tween20. Protein purification ran with buffers containing Tween20 showed a protein yield in wash 2 (Figure D1) as opposed to no protein being eluted in wash/elution buffers with no Tween20. All buffers (binding, wash, and elution) were adjusted to a pH of 8.0 with HCL or NaOH. Volume of Ni-NTA beads were tested at 100µL, 40µL, and 30µL, where 30µL was determined to be the optimal amount of beads used per in-house cartridge when considering binding capacity and high cost of the beads. Processing of bacterial pellets from a 10mL culture SX2 protein expression was determined to require 100µL 1X Fastbreak cell lysis reagent after comparing to protein yield using Lysozyme (from chicken egg white) to lyse bacteria. To determine location of SX2 protein after protein purification, the flow through, washes, elution, and stripped beads (250µL 0.5M EDTA) were ran on a 4-12% gradient SDS-PAGE gel at 155V to separate the protein then stained with Coomassie Blue.

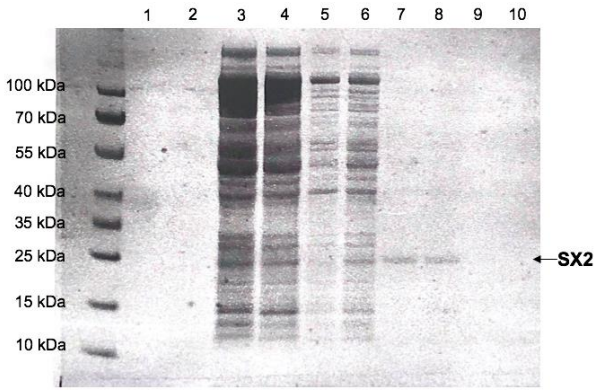

Figure D1. SDS-PAGE from in-house protein purification comparing 1X Fastbreak Cell Lysis Reagent to Lysozyme as cell lysing methods. (1) Elution with Fastbreak (2) Elution with Lysozyme (3) Flow through with Fastbreak (4) Flow through with Lysozyme (5) Stripped beads from Fastbreak (6) Stripped beads from Lysozyme (7) Wash 1 Fastbreak (8) Wash 1 Lysozyme (9) Wash 2 Fastbreak (10) Wash 2 Lysozyme

In-House Maxwell Purification

1. After 10mL culture SX2 protein expression in *E.coli*, spin down the bacterial pellet at 3700rpm for 12 minutes at room temperature, remove supernatant and resuspend in 100µL cell lysis reagent, 900µL binding buffer, 1µL protease inhibitor cocktail, and 5ul DNase
2. Load cartridge according to Figure D2
3. Place cartridge, elution tube, and plunger into the Maxwell
4. Run Maxwell on protein setting
5. Collect samples from all wells and freeze to store

	1mL bacteria
	470µLbinding, 30µL Ni-NTA beads
	1mL wash
	1mL wash
	1mL wash
	1mL wash
	Plunger

Figure D2. A diagram of how to load the in-house Maxwell cartridge.