Resource Allocation and Process Improvement of Genetic Manufacturing Systems

Gregory Thomas Purdy

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

Doctor of Philosophy
in
Industrial and Systems Engineering

Jaime A. Camelio
Kimberly P. Ellis
Patrick Koelling
G. Don Taylor

October 26, 2016
Blacksburg, VA

Keywords: Discrete Event Simulation, Gene Synthesis, Genetic Manufacturing Systems, Inspection Allocation, Mathematical Modeling, Molecular Biology, Synthetic Biology

Resource Allocation and Process Improvement of Genetic Manufacturing Systems

Gregory Thomas Purdy

ABSTRACT

Breakthroughs in molecular and synthetic biology through *de novo* gene synthesis are stimulating new vaccines, pharmaceutical applications, and functionalized biomaterials, and advancing the knowledge of the function of cells. This evolution in biological processing motivates the study of a class of manufacturing systems, defined here as genetic manufacturing systems, which produce a final product with a genetic construct. Genetic manufacturing systems rely on rare molecular events for success, resulting in waste and repeated work during the deoxyribonucleic acid (DNA) fabrication process. Inspection and real time monitoring strategies are possible as mitigation tools, but it is unclear if these techniques are cost efficient and value added for the successful creation of custom genetic constructs.

This work investigates resource allocation strategies for DNA fabrication environments, with an emphasis on inspection allocation. The primary similarities and differences between traditional manufacturing systems and genetic manufacturing systems are described. A serial, multi-stage inspection allocation mathematical model is formulated for a genetic manufacturing system utilizing gene synthesis. Additionally, discrete event simulation is used to evaluate inspection strategies for a fragment synthesis process and multiple fragment assembly operation. Results from the mathematical model and discrete event simulation provide two approaches to determine the appropriate inspection strategies with respect to total cost or total flow time of the genetic manufacturing system.

Resource Allocation and Process Improvement of Genetic Manufacturing Systems

Gregory Thomas Purdy

GENERAL AUDIENCE ABSTRACT

Breakthroughs in molecular and synthetic biology through *de novo* gene synthesis are stimulating new vaccines, pharmaceutical applications, and functionalized biomaterials, and advancing the knowledge of the function of cells. This evolution in biological processing motivates the study of a class of manufacturing systems, defined here as genetic manufacturing systems, which produce a final product with a genetic construct. Genetic manufacturing systems rely on rare molecular events for success, resulting in waste and repeated work during the deoxyribonucleic acid (DNA) fabrication process. Inspection and real time monitoring strategies are possible as mitigation tools, but it is unclear if these techniques are cost efficient and value added for the successful creation of custom genetic constructs.

This work investigates resource allocation strategies for DNA fabrication environments, with an emphasis on inspection allocation. The primary similarities and differences between traditional manufacturing systems and genetic manufacturing systems are described. A serial, multi-stage inspection allocation mathematical model is formulated for a genetic manufacturing system utilizing gene synthesis. Additionally, discrete event simulation is used to evaluate inspection strategies for a fragment synthesis process and multiple fragment assembly operation. Results from the mathematical model and discrete event simulation provide two approaches to determine the appropriate inspection strategies with respect to total cost or total flow time of the genetic manufacturing system.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. Introduction

The number of companies providing commercial deoxyribonucleic acid (DNA) fabrication services, often referred to as *de novo* (from the beginning) gene synthesis, has significantly increased during the last 15 years. The scope of these services is not restricted to genes and allows the fabrication of DNA sequences from tens to thousands of base pairs. Life scientists are no longer limited to DNA extracted from biological samples but have progressively taken advantage of the services provided by the industrial DNA fabrication sector. The ability to design and fabricate custom DNA sequences that do not exist in nature has enabled new approaches to various research and development challenges in the biological sciences.

Synthetic biology, and more specifically *de novo* gene synthesis, has progressed significantly during the last two decades with advances in both the application of the science and advancements of technology that allow unprecedented breakthroughs in the creation of DNA constructs. In infectious disease research, the synthesis of viral genomes has led to new vaccine engineering strategies [1-3], vaccine production workflows [4], and the resurrection of the influenza virus strain responsible for the 1918 Spanish flu pandemic [5, 6]. Within industrial biotechnology, computational approaches can be used to design genes, which code for enzymes with new or improved catalytic activities [7, 8]. It is now common practice to redesign genes to maximize the production of the protein they code when inserted in a foreign host [7-10]. In addition, high-volume production of synthetic DNA fragments provides the building blocks to assemble synthetic versions of a bacterial genome [11] and a yeast chromosome [12].

With these recent breakthroughs, *de novo* gene synthesis is changing the face of biology. The manufacture of products created through molecular biology processing is becoming more commonplace within the industrial biotechnology sector. Molecular and synthetic biology continue to develop new tools to manipulate and control biological systems with immense precision and understanding. For example, gene synthesis techniques can create segments of DNA that do not occur naturally with precision down to the individual base pair ordering of a sequence.

1.1 Significance

Molecular and synthetic biology have spawned a new manufacturing production environment, referred to here as a genetic manufacturing system (GMS). A GMS is defined as any production or manufacturing environment in which the final product is a genetic construct. A GMS could range from a system used to create a genetic sequence of only a few base pairs or used to create an entire genome comprised of billions of base pairs. While these two product outcomes may seem drastically different, many of the process operations are foundational and employed in an iterative fashion to create the desired genetic sequence. Similar to traditional manufacturing systems, these complex process flows can be decomposed and manufacturing systems analysis tools used to better understand the dynamics of the production environment.

With the immense interest in gene synthesis research, funding and commercial opportunities in the sector have also increased. In fact the US biotechnology sector has doubled in size over a ten year period and grown into a $98.5 billion industry [13]. The amount of capital within gene synthesis research tends to increase the speed of innovation

while efficiency may not be the key concern. Thus, many of the processes and operations have not yet been optimized to minimize production costs or time.

The analysis of traditional manufacturing systems (TMS) has progressed dramatically over the past century with mathematical modeling, simulation, and data analytics providing vital contributors to the growth of this field. These techniques also provide the basis for improvements within GMS. Process flow optimization, sequencing and scheduling for genetic production facilities, inspection allocation problems, and facility layout and design are a few areas in which methods from manufacturing systems analysis can be applied to this burgeoning field. An emphasis on cost and time reductions are commonplace in traditional manufacturing settings, and a challenging question is how to translate these gains into the world of GMS?

1.2 Motivation

The motivation for this work originated from an NSF CREATIV (Award #1241328) grant, under the NSF INSPIRE program, titled *Modeling and optimization of DNA manufacturing processes* [14, 15]. Throughout this project, it quickly became apparent that the fields of synthetic biology and manufacturing systems analysis differed not only in the approach to processing, but the fundamental aspects of how a manufacturing system is perceived. These unique differences result in the need to establish a class of manufacturing systems, known as genetic manufacturing systems, to encapsulate this area of research.

To utilize manufacturing systems analysis, an understanding both the technical and process complexities of creating DNA from the base raw materials is needed. Based on this understanding, through process mapping the complex steps can be decomposed into

processes and discrete operations. This helps to not only visualize the process, but also identify areas of significant cost or time investments as well as waste. With a grasp of the basic gene synthesis processing steps and complexity of the operations, modeling and simulation opportunities became apparent specifically within the realm of inspection allocation in genetic manufacturing systems.

This work focuses on the important interfaces of molecular biology, synthetic biology, gene synthesis, and manufacturing systems. Within the realm of DNA fabrication, specifically at the academic and industrial research level, the purpose is to synthesize novel segments of DNA for a multitude of applications. Recently a significant influx of fiscal resources has emerged for these efforts. Unfortunately, this capital boon often focuses on product innovation rather than the need for process improvement and waste reduction during the process flow. Transitioning to this type of thinking when synthesizing DNA can create friction between practitioners and researchers. However, the benefits include potential reductions in cost, total process time, and waste, by reducing redundant processing. This work serves as the first known endeavor to demonstrate the gains of utilizing modeling and simulation for inspection allocation in the novel new interdisciplinary field of genetic manufacturing systems.

1.3 Research Contributions

This proposed work has four research contributions, which provide impact to both the manufacturing systems and synthetic biology communities. These contributions are:

Identification of a class of manufacturing system, referred to as genetic manufacturing systems. Through this work, the need emerged for the formalization of a class of

manufacturing systems to describe the complex processing required in the creation of custom DNA constructs. This work defines this new class of system and highlights the similarities and differences from traditional manufacturing systems. Discussion is provided on various key research questions to be analyzed utilizing techniques from manufacturing systems analysis.

Mapping of gene synthesis process flow considering both process and routing decisions. This is the first known work to develop a process flow map of the operations necessary to create a custom DNA fragment while considering the routing and logical decisions necessary to produce the desired target sequence. Utilizing the process flow map, it is possible to then model the creation of a custom DNA sequence as a manufacturing process for several potential applications.

Mathematical formulation of the inspection allocation problem for genetic manufacturing systems utilizing gene synthesis. A key contribution of this work is a mathematical description of inspection allocation decisions required to optimize the cost and time for a genetic manufacturing system utilizing gene synthesis. Through the inspection allocation formulation, gene synthesis processing has been successfully modeled as a manufacturing process.

Discrete event simulation of the inspection allocation problem for genetic manufacturing systems in academic and industrial laboratory research environments. The mathematical model provides generalizations on inspection allocation effort for a wide range of potential settings. To assist academic research gene synthesis laboratories, a discrete event simulation accounts for the additional assumptions of these settings. This simulation supports process improvement and provides practitioners with generalizations

on the appropriate inspection strategy for a common two stage process found in gene synthesis as well as a multiple fragment assembly process.

These four research contributions are described in detail throughout the remainder of this work. Together, these contributions emphasize the advantages of utilizing manufacturing systems analysis techniques in the interdisciplinary field of genetic manufacturing systems.

1.4 Dissertation Outline

This dissertation will be structured as follows. Chapter 2 provides background on different processing steps and inspection strategies in molecular biology, synthetic biology, and gene synthesis. Chapter 3 presents the key differences between traditional and genetic manufacturing systems and describes a genetic production environment using a process flow map. Chapter 4 adapts the inspection allocation mathematical model to genetic manufacturing systems. Chapter 5 introduces a discrete event simulation methodology to provide practitioners with an approach to select the best inspection allocation strategy for their specific process considerations. The resulting simulation model incorporates the complicating factors associated with these types of production environments is used to analyze a two-stage genetic manufacturing system. Chapter 6 extends the discrete event simulation methodology to a three-fragment Gibson assembly process and provides generalizations about the appropriate inspection strategies for minimizing based on cost or time for different process parameters. Chapter 7 discusses the conclusions of this dissertation as well as potential future research areas.

2. Background on Molecular Biology and Gene Synthesis

Genetic manufacturing systems encompass production environments in which the final product consists of a genetic construct. This section provides the background information necessary to analyze GMSs with specific emphasis on those which employ gene synthesis protocols.

2.1 Interactions between Molecular Biology, Synthetic Biology, and Gene Synthesis

Many of the processes in GMSs used to create genetic constructs have been employed for decades. However, new enabling technologies and continuous refinement of the protocols used to describe these process flows have been crucial in the rapid development of breakthroughs in the biological sciences. This section briefly describes how molecular biology, synthetic biology, and gene synthesis intersect.

Molecular biology is the study of various biological phenomena of the cell at the molecular scale. A specific focus of molecular biology involves the molecular interactions necessary for the replication of DNA and related processes. A number of methods were developed in molecular biology to manipulate DNA. Specific methods which are relevant for this dissertation include the polymerase chain reaction, molecular cloning, gel electrophoresis, and sequencing. More specifics on these processes are provided in Section 2.2. While these techniques were originally developed for use in molecular biology, they are still frequently used in synthetic biology and gene synthesis. Figure 1 depicts how molecular biology, synthetic biology, and gene synthesis interact and build upon the

knowledge gained from the study of processes regulating the replication of DNA in the biological sciences.

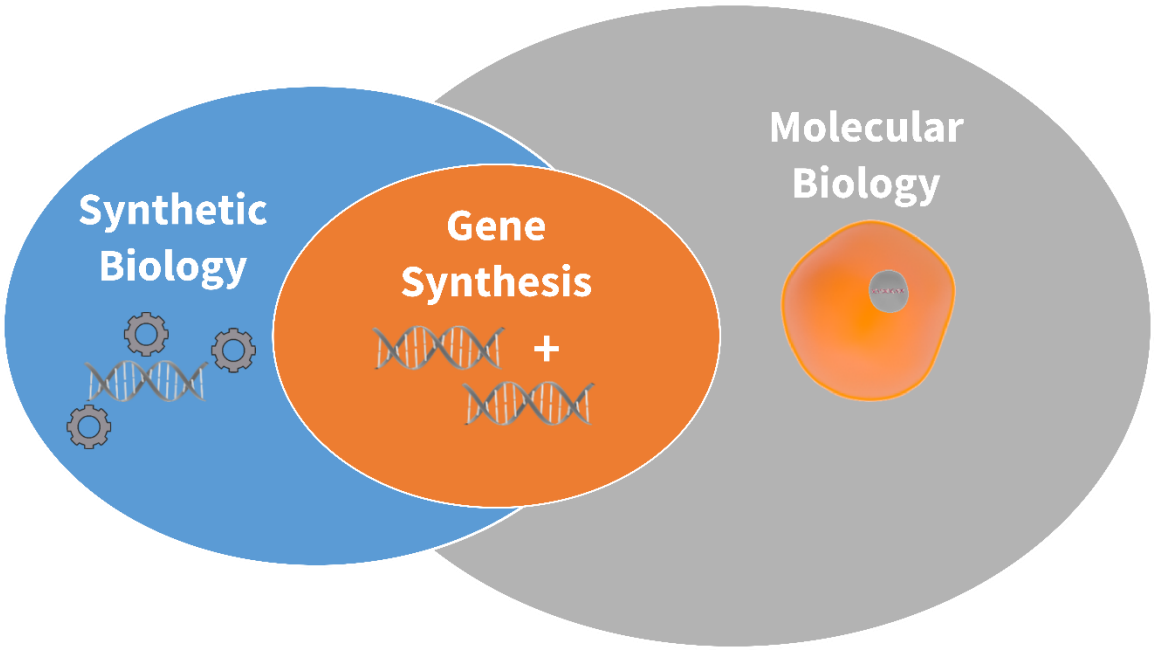

Figure 1: Intersections of Molecular and Synthetic Biology with Gene Synthesis.

Synthetic biology stemmed from molecular biology as a fusion of biology and engineering aimed at creating new biological systems or improving upon existing biological structures. A key difference between synthetic biology and molecular biology is that organisms and DNA sequences do not necessarily have to be found in the natural environment. Instead, synthetic biologists utilize the techniques of molecular biology to engineer new biological systems with natural DNA/biological components, synthetic DNA/biological components, or both. In addition to creating new systems, synthetic biologists construct and use standardized biological parts to fine tune and optimize the synthetic systems they develop.

Advances in synthetic biology have been spurred by the developments in both DNA synthesis and DNA sequencing. *De novo* gene synthesis has become a leading approach to create DNA sequences without the need for an existing DNA sequence [16]. Specific protocols have been developed in gene synthesis to allow the creation of base pair specific DNA sequences, without the need for a starting DNA template found in nature or previously synthesized through another protocol. A number of approaches can be used to create customized DNA sequences based on the desired length of the target fragment. A selection of these techniques and the basic underlying molecular biology principles used to employ these protocols are reviewed in Section 2.2.

Gene synthesis and synthetic biology have benefited from a symbiotic relationship. One of the motivations behind *de novo* gene synthesis was the need for reliable, yet highly customized, DNA sequences for use in synthetic biological systems. The techniques developed through gene synthesis allow the creation of virtually any DNA sequence without the need for any prior DNA template. Compared to other means of creating brand new sequences, the techniques of gene synthesis are relatively low in cost and significantly less time consuming. Improvements to the gene synthesis protocols have provided the opportunity for standardized DNA sequences and new biological components to be more readily available for synthetic biologists. Many commercial biotechnology companies are now offering turn-key gene synthesis services (typically under 3 kilo base (kB) pairs) which only require the customer to provide the target DNA sequence, with turnaround times of approximately two weeks. These commercial services, based on the protocols of *de novo* gene synthesis, have significantly reduced the cost and time needed to create customized DNA sequences under 5 kB.

Gene synthesis protocols for the creation of fully customized DNA sequences are a primary mechanism spurring the advancements in synthetic and molecular biology. With gene synthesis being crucial to the development of knowledge within the biological sciences, this work will focus on the processing flows described through the gene synthesis protocols. The remainder of this chapter will provide an overview of the basic processing steps utilized in many gene synthesis protocols and the inspection techniques as well as implementation strategies available in *de novo* gene synthesis.

2.2 Gene Synthesis Background

While implementations of gene synthesis processes vary across types of genetic manufacturing systems and individual organizations, the overall structure of these processes is remarkably similar. Protocols used in *de novo* gene synthesis [16-18] differ across implementations and affect the way each step is performed in a particular organization. The following subsections provide background on gene synthesis processing as well as inspection options and strategies available in these types of systems.

2.2.1 *Gene Synthesis Processing Steps*

The processing steps involved in the creation of custom DNA segments using gene synthesis are described in this sub-section. Specifically, the synthesis and assembly of oligonucleotides, molecular cloning of a DNA molecule, and DNA sequencing are covered. This section focuses on the specific types of processes used in the analysis found in this dissertation, but does not provide an exhaustive description of all gene synthesis processing operations.

2.2.1.1 Oligonucleotide Synthesis

Long double-stranded DNA molecules are assembled from pools of small single-stranded DNA molecules called oligonucleotides (oligos for short) [19, 20]. Oligos are produced in dedicated instruments using well-established chemical processes [21, 22]. Oligo synthesizers proceed through a series of cycles adding one nucleotide at a time. While the yield of each cycle is very high (typically above 99.8%), the random nature of the chemical processes results in a small error rate. These errors limit the length of DNA that can be chemically synthesized and impacts the quality of the raw materials used in gene synthesis processing. Oligos are used in a wide variety of applications outside of gene synthesis, and require expensive equipment and infrastructure rarely available in academic life science facilities. Therefore, small organizations tend to purchase oligos from external vendors while larger organizations often have an internal oligo synthesis facility. Oligonucleotide synthesis is generally considered an input used in the creation of the raw materials used in gene synthesis processing.

2.2.1.2 Gene Synthesis through Oligonucleotide Assembly

Assembling the predesigned oligos into progressively longer double-stranded DNA molecules is the core of gene synthesis processing. Several different techniques have been described through the protocols to accomplish this goal and a selection are outlined in this subsection.

2.2.1.2.1 Assembly Polymerase Chain Reaction

One of the prevalent methods for oligonucleotide assembly is polymerase chain assembly (PCA) otherwise known as an assembly polymerase chain reaction (PCR). This assembly

PCR, uses *in vitro (outside the body)* biochemical reactions to combine DNA molecules with complementary bases into longer DNA strands [16, 23-25]. These DNA strands will become progressively longer through a series of annealing and extension steps facilitated by a DNA polymerase capable of filling in gaps of complementary base pairs created by overlap sections of the initial single-stranded oligonucleotides. Many factors contribute to the complexity of the products generated by a PCR. These factors include: weak interactions between DNA molecules; the possibility of similar sequences being produced that do not match the desired product; limitations of the enzymes stitching DNA molecules together; and the overlap number between the single stranded DNA molecules. A successful PCR will combine the shorter DNA fragments in the correct order matching the reference sequence of interest. Due to the complexity and random nature of these chemical processes, it is not unusual for the desired DNA molecule to be a very small fraction of the products generated by a PCR reaction. However, the assembly PCR protocol allows the amplification of a desired target sequence with the introduction of additional primers following the initial construction phase through a subsequent PCR.

2.2.1.2.2 Ligase Chain Reaction

Other techniques have been proposed to accomplish the assembly of oligonucleotides in gene synthesis protocols including the ligase chain reaction (LCR) [16, 25, 26]. During a LCR, the DNA ligase is used to join oligonucleotides by exploiting the ability of the ligase to repair DNA sequences. The ligase is used to fuse the ends of two oligonucleotides which have a predesigned overlap with one another. LCRs also require multiple thermocycles to allow the separation of double stranded DNA (denaturation), natural pairing of complementary overlaps (annealing), and fusion of complementary segments (ligation).

Another feature of the LCR is the strict temperature requirements of the reaction. These temperature requirements limit the fusion of initial oligos containing a mutation and encourages the fusion of non-mutant oligos.

2.2.1.2.3 Thermodynamically Balanced Inside Out Synthesis

Thermodynamically balanced inside out synthesis is another novel technique used to join single-stranded oligonucleotides into longer segments of double stranded DNA [27, 28]. This method begins with an initial DNA fragment created from the TBIO designed primers. The initial inside fragment is developed into a progressively longer strand from the inside-out. The inside fragment is used as the starting template for subsequent primer sets in the TBIO reaction. This protocol differs significantly from the other procedures discussed in this subsection as it builds the fragment from the center and moves outward instead of stitching building block DNA fragments together into longer fragments.

2.2.1.3 Molecular Cloning

After an oligonucleotide synthesis process is complete, the challenge is to isolate the correct DNA sequence from the multitude of products created through the reaction. Molecular cloning can screen the different combinations of the oligonucleotides generated by the assembly reaction. An interesting property of the well-studied and understood *Escherichia coli* (E. coli) bacteria is its ability to create genetically identical clones in a short amount of time. *E. coli* bacterial cells are exposed to a solution containing the pool of DNA molecules produced by an assembly reaction. Some of the cells will absorb one individual molecule and propagate it in their offspring to form individual colonies on a Petri dish. Each colony will contain millions of identical *E. coli* cells because of the

bacteria's ability to rapidly replicate. Each of these bacterial colonies will contain one permutation of the initial oligos used in the oligo assembly reaction. Commercially available kits enable the introduction of a foreign DNA fragment to the E. coli genetic vector in a relatively simple and reliable process. More information on these types of cloning techniques in [16-18, 29, 30].

2.2.1.4 DNA Sequencing

Typically, the final step of the gene synthesis workflow is to sequence the DNA extracted from the bacterial colonies. Like oligo synthesis, DNA sequencing (reading the sequence of DNA molecules) is a technique used in a variety of applications outside of gene synthesis requiring expensive specialized equipment. Thus, DNA sequencing is generally outsourced to an external service provider or an internal core facility. There are numerous types of sequencing technologies capable of determining the DNA sequence with varying levels of precision, cost, and time. The types of technologies available for DNA sequencing will be addressed in Section 2.2.2.

Due to the wide variety of protocols available to produce different types of DNA molecules through gene synthesis, this study focuses on one of the prevalent standard processes used in synthetic biology. This study will focus on two of the core operations found in GMS utilizing *de novo* gene synthesis: the PCR reaction [16, 23, 24] followed by cloning [16-18, 29, 30]. Figure 2 shows a simplified process flow of the GMS to be considered in this work.

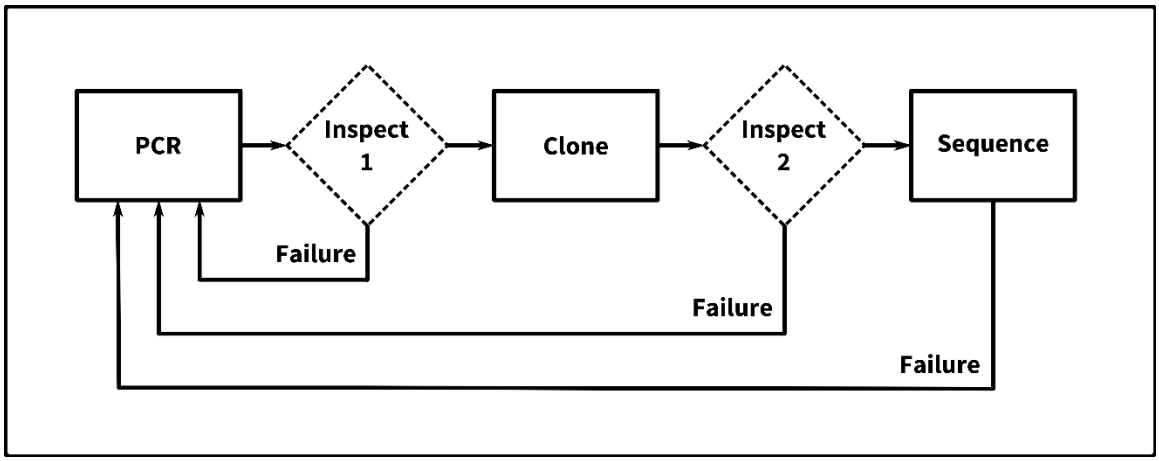

Figure 2: Simplified Gene Synthesis Process Flow with Routing Decisions.

Earlier it was noted that sequencing is typically utilized as a final verification step at the termination of processing to create the desired DNA fragment. However, this step is not universally applied especially in the academic research community. Regardless, the sequencing step has been included in the analyses of this work as sequencing after processing is the only way to confirm that the correct sequence has been produced through the process.

2.2.2 *Gene Synthesis Inspection Options*

Inspection of gene synthesis operations requires different technologies than those used in traditional manufacturing quality control. Two categories of inspection techniques are available to molecular biologists: sizing and sequencing.

2.2.2.1 DNA Sizing Inspection Techniques

Many molecular and synthetic biology laboratories utilize agarose gel electrophoresis [29-31] to determine the size of DNA fragments. This approach is relatively inexpensive, but

time-consuming and labor-intensive. After running a sample containing a mixture of DNA molecules through gel electrophoresis, the approximate size of the constituents can be determined by looking at the position of individual bands corresponding to DNA molecules of varying length. The low resolution of agarose gel electrophoresis negatively affects the precision of the data available from this type of inspection. Agarose gel electrophoresis is more suitable to approximating the contents of a DNA sample than to support a definitive quantitative analysis of its quality.

Capillary electrophoresis is an inspection technique that addresses some of the limitations of agarose gel electrophoresis. Separating molecules in a microfluidic capillary is faster than conventional electrophoresis and increases the resolution. Furthermore, instruments such as the Agilent Bioanalyzer 2100 [32, 33] have built-in signal detection and processing systems that greatly improve the quality of the data they produce. Both agarose gel and capillary electrophoresis indicate the relative concentration and approximate length of the constituents within a solution, but not the exact sequence.

2.2.2.2 DNA Sequencing Inspection Techniques

Sequencing is another category of inspection in gene synthesis. Instead of merely determining the length of DNA molecules, Sanger sequencing [34-36] can resolve the exact DNA sequence of a DNA sample. Verifying the sequence of DNA molecules proceeds through two steps. First a series of reads, up to 1,000 bp in length are collected, starting from predetermined small segments of DNA (referred to as primers) regularly distributed and designed into the sequence. In a second step, the reads are assembled by searching for overlaps in the sequence resulting from the read assembly and compared to the sequence ordered by the customer [37]. This type of quality control has the highest cost and longest

total flow time of the various inspection options described. However, the results of sequencing provide the exact DNA sequence of the sample with very low error rates, both in the process itself and interpretation of the results. In gene synthesis it is considered a best practice to use DNA sequencing as the final inspection step to confirm the prior processing steps have produced the expected end product.

Next generation sequencing methods have improved the ability to sequence long sequences of DNA. High throughput sequencing technologies have directly led to the ability to sequence the entire human genome in a fraction of the time compared to the original Human Genome Project [38]. Ultimately, the goal of these technologies has been to reduce the cost and time associated with sequencing DNA both as fragments and whole genomes [39-43]. Next generation sequencing methods have been reviewed by a number of academic sources outlining the strengths and limitations of the various approaches [42-45]. One particularly popular form of next generation sequencing is referred to as massively parallel sequencing [38, 43]. This technique provides substantial reductions in cost and time by breaking the sequence into much smaller sections of DNA, conducting random reads, and utilizing bioinformatics software to reconstruct the sequence using the overlapping sections of the reads [38]. While these next generation approaches have considerably reduced the cost and time associated with DNA sequencing, for smaller sequences such as PCR products or small fragment constructs, Sanger sequencing is still the appropriate choice [46]. This work focuses on a two stage process flow (PCR and cloning), and Sanger sequencing will be used in the analysis as the final sequencing inspection option.

2.2.3 *Gene Synthesis Fragment Assembly Methods*

One of the primary strengths of gene synthesis is the ability to build progressively longer DNA sequences utilizing the various gene synthesis fragment assembly protocols. A number of techniques have been reviewed in the literature featuring different strategies to produce DNA sequences made up of constituent fragments.

2.2.3.1 Overlap Gene Synthesis Fragment Assembly Methods

Many of the methods used to combine multiple DNA fragments into longer assemblies rely on the creation of specific and discernable overlap regions in the sequence. One such approach, called overlap extension PCR, uses particular repeating C- or G-rich base pair repeats to allow the annealing of multiple fragments in a single step [47]. Overlap extension PCR can combine three fragments together simultaneously utilizing these special overlap regions made of C or G base pair combinations. Another overlap based assembly approach, sequence and ligation independent cloning (SLIC) utilizes in vitro homologous recombination [48]. Through the SLIC protocol up to five inserts can be assembled into a single vector without the need of a restriction enzyme or ligase.

Circular polymerase extension cloning (CPEC) can accomplish the combination of up to four fragments utilizing a polymerase in a single reaction. CPEC does not rely on the use of a restriction digest, ligation, or homologous recombination as the mechanism for fragment fusion [49]. One of the proprietary assembly techniques, Gateway cloning, relies on specific types of recombination attachment sites (referred to as *att*) so fragments can be combined [50]. The specific recombination sites and protocol were developed by Invitrogen and are capable of combining three fragments in one reaction. Uracil-specific excision reagent (USER) cloning is a another assembly technique which does not require

a ligation to achieve the desired result [51]. The USER technique can combine three fragments of differing sizes in a single reaction.

Recently, Gibson assembly has continued to push the abilities of synthetic biologists as a new way to combine multiple fragments in a single reaction [11, 52-54]. The Gibson assembly technique uses a cocktail of enzymes including an exonuclease, polymerase, and ligase to allow the fusion of up to six fragments with fewer steps than competing strategies.

2.2.3.2 Standardized DNA Parts Fragment Assembly Methods

Another approach to assemble multiple DNA fragments involves the use of type II restriction digest enzymes. These approaches utilize the properties of the enzyme to cleave the DNA at specific known locations. One specific approach uses standardized DNA parts called BioBricks to allow for the creation of multiple fragment assemblies [55]. These BioBricks have been created by synthetic biologists mimicking the creation of standardized components and parts common in the engineering disciplines. This allows the rapid development of new BioBricks with novel functionality as well as increasingly complex BioBrick based projects building from the registry of BioBrick parts [56, 57]. The creation of new BioBricks is possible by beginning with a brand new synthetic sequence, using natural DNA, or any combination therein utilizing established methods [58]. Additionally, multiple BioBricks can be combined to create assemblies or devices with added functionality utilizing various protocols including the In-Fusion PCR [59] and many other standard techniques [60]. An alternative called BglBricks have been proposed to overcome some of the issues associated with BioBricks [61].

2.2.3.3 Type II Restriction Enzyme Fragment Assembly Methods

Other approaches utilizing the type II restriction enzymes have also been employed to assemble longer DNA fragments. Pairwise selection assembly is capable of assembling long fragments of DNA by breaking the segment into sub-fragments which have a recyclable tag at each of the ends [62]. These tags can be excised during the assembly process resulting in an assembly of two sub-fragments which can in turn be transformed into an assembly of four sub-fragments, and the process can continue progressively. Golden Gate cloning is also capable of producing DNA assemblies made of multiple constituent parts with a high rate of correct recombination [63]. An advantage of this approach is that one tube can be used to assemble multiple fragments in a short period.

2.2.4 *Gene Synthesis Inspection Strategy*

A proper inspection strategy specifying what inspection technique to use at various stages of the gene synthesis process can positively impact the cost and time associated with the entire GMS. Referring to Figure 2, not all inspection options outlined in Section 2.2.2 are suitable for the inspection stations due to constraints of the outputs at each operation of the gene synthesis process. Certain inspection stations are not available due to the realistic limitations imposed by GMS utilizing gene synthesis processing. For example, sequencing is not available at inspection station two as it would be redundant with the final sequencing quality control step at the end of the process flow. Utilizing sequencing as a final quality check is required to ensure that the sample matches the desired sequence and is employed regardless of the other inspection strategies chosen. This procedure is consistent with best practices found in the industrial biotechnology sector and most academic research laboratories.

2.2.4.1 Number of Samples for an Inspection Station

Another aspect of the inspection strategy is the number of samples sent for inspection. Due to the diversity of products leaving the different processing steps in the gene synthesis process flow, the number of samples inspected at each station differs even if the same technique is used at inspection station one and two. The routing and decision logic presented in this section is consistent with the processing used by many gene synthesis laboratories and will be utilized throughout the remainder of this work.

2.2.4.1.1 Sample Size and Routing Decisions for Inspection Station One

Inspection station one examines a solution containing a pool of DNA molecules created during the previous PCR assembly step. A single small aliquot (portion of the sample) of the PCR products is analyzed at the first inspection station. Evaluating multiple samples would not dramatically improve the accuracy of the inspection because the measurement error of the inspection process is unlikely to affect the outcome of the pass/fail decision made at inspection station one. Rather, the main reason for failure is the random chemical reactions required to occur during the biological processes failed to succeed during the PCR process step. In the laboratory, if inspection station one yields a non-conforming sample, the PCR process is repeated. If multiple repeated PCR passes continue to yield a non-conforming sample, then the initial starting oligos may need to be adjusted for the desired reference sequence.

2.2.4.1.2 Sample Size for Inspection Station Two

The output of the cloning step is a population of individual bacterial clones. These clones may be different from each other because each can absorb a different DNA molecule from

the potential combinations found in the PCR sample. This output is significantly different than the output of the PCR which is a single solution with potentially millions of enumerative combinations of DNA sequences based on the starting oligos of the PCR.

The purpose of inspection station one is to determine if the PCR process was a success, while, the objective of the second inspection is to select a number of conforming colonies that will be sent for final inspection via sequencing. Thus, a technician chooses a number of samples from a large population and each colony in this sample is inspected individually, since each could contain the correct genetic sequence. Selecting multiple conforming colonies from the population, which has an unknown probability of non-conforming units, creates a compounded probability of successfully choosing a good clone. If all the colonies in the sample are non-conforming, a decision must be made about the entire population. Is the whole population not of the correct sequence, or did the wrong samples get selected from the population causing the appearance of a faulty process?

This problem is like acceptance sampling in traditional manufacturing, however in gene synthesis the probability of a non-conforming sample is unknown and can change over time. Additionally, since each sequence is unique there is no historical processing data to use for comparison. In addition, further preventing the direct application of acceptance sampling is the inspection options following the cloning step do not provide the precise order of the DNA sequence of the bacterial clone. Instead, these options only provide information about the relative length of the DNA sequence. This is a major difference from traditional manufacturing in which the inspection option would indicate if the product is conforming based on certain quality characteristics and prevents the direct application of acceptance sampling assumptions to GMS. A more detailed description of the variance

found in colony selection and the procedures used to simulate this nuance are outlined in Section 5.1.2.

2.2.4.1.2.1 Routing Decisions for Inspection Station Two

The unique challenges of the cloning process step directly impact the routing decisions at inspection station two. At inspection station two, a certain number of samples are inspected. The samples which appear to be conforming, based on the results of the quality check are set aside. Another set of samples are inspected until a threshold of conforming samples has been met. These samples, which appear to be conforming based on their length, are then sent to the sequencing inspection step to determine the precise order of the base pairs.

If a set of samples completely fails inspection station two because it does not include any conforming clones, then another sample of colonies produced by the cloning step will be analyzed. The new sample will have another inspection at station two in an attempt to limit the risk of sampling errors. If the second sample fails to have any conforming clones, then the probability of sampling errors is small enough that the lack of positive clones can be interpreted as a failure of the entire assembly process. The entire assembly process is repeated once to eliminate the possibility of human error. If the entire assembly process fails a second time, this points to an underlying error in the process design or raw materials used. Due to the cost and time-consuming nature of the inspection and assembly processes, failures are generally limited to two before investigating other issues that may be preventing process success.

2.2.4.1.3 Routing Decisions for Sequencing

Sequencing is an integral part of ensuring the proper DNA sequence is created through the course of a protocol. Due to the low error rate, sequencing failures have a different routing strategy. If sequencing does not show any successful clones of the promising bacterial colonies, then the sample will be discarded and the process will begin again at the PCR step. Although the size of the DNA samples may appear correct at inspection station two, sequencing can discern the true composition of the samples. For this reason, if none of the samples be the correct sequence, this suggests that the cloning was not successful in producing the desired end product. If the second attempt of the process with the existing oligos fail at the sequencing step, then this would point to a potential problem with the design of the oligos. Accordingly, it would be necessary to go back to the beginning and redesign the oligos for the desired reference sequence.

This section has provided the necessary background on the biological processes discussed in the remainder of the dissertation. The oligonucleotide assembly, fragment assembly, inspection strategies, and routing logic from this section will be used in the development of a mathematical formulation and discrete-event simulation found in later sections of the dissertation.

3. Opportunities to Apply Manufacturing Systems Analysis Techniques in Genetic Manufacturing Systems

Genetic manufacturing systems are a new class of manufacturing system which can benefit from the manufacturing systems analysis techniques which have been developed over the last few decades. GMSs are complex for a variety of reasons, including the underlying thermochemical processes required to produce custom DNA sequences which rely on the random combination of base pairs at the molecular level. Fortunately, manufacturing systems analysis techniques frequently deal with random processes and utilize modeling and simulation methods to derive process knowledge even in the most complex and challenging production environments. This section highlights the similarities and differences between a traditional manufacturing system (TMS) and its genetic counterpart.

3.1 Differences Between Traditional Manufacturing Systems and Genetic Manufacturing Systems

Traditional manufacturing analysis techniques can be applied within GMSs, however, there are key differences dictated by the characteristics of synthetic biology and more specifically, gene synthesis. Operations to create the final genetic product may be well defined, but the production process requires a rare molecular event to occur, which introduces more sources of variability into the system. This leads to highly complex and iterative process flows.

In general, protocols are used to describe many of the common processing steps in gene synthesis [16, 18]. Differences in the application of these protocols can have an impact on the success of any given processing step or the entire process flow. In addition, even if

two identical protocols are conducted in the exact same fashion, an opportunity exists for the molecular event to not occur or a random mutation to change the final product.

3.1.1 *Quality Considerations for Genetic Manufacturing Systems*

Deriving the quality of mid-stream constructs and final products is difficult, costly, and time consuming for GMS. Strategies are available for mid-stream quality assurance within biological processing, but it is unclear if these approaches reduce the cost or time of processing. Visual optical inspection of DNA is impossible due to the physical size of DNA; therefore, quality control techniques indirectly infer the quality of a sample. Both gel electrophoresis [29-31] and capillary electrophoresis [32, 33] have become common methods for deriving different types of mid-stream process data regarding a DNA sample. The interpretation of these results introduces another source of variation into the processing flow.

DNA sequencing technologies are capable of detecting the exact sequence of a construct and these technologies have considerably reduced in cost over the last ten years [39, 64]. However, sequencing is cost prohibitive to apply after every processing step and sequencing is often outsourced to a third-party vendor, increasing the lead time. Figure 3 illustrates the differences between a TMS with an inspection step and a common gene synthesis processing flow with an inspection and final sequencing quality step.

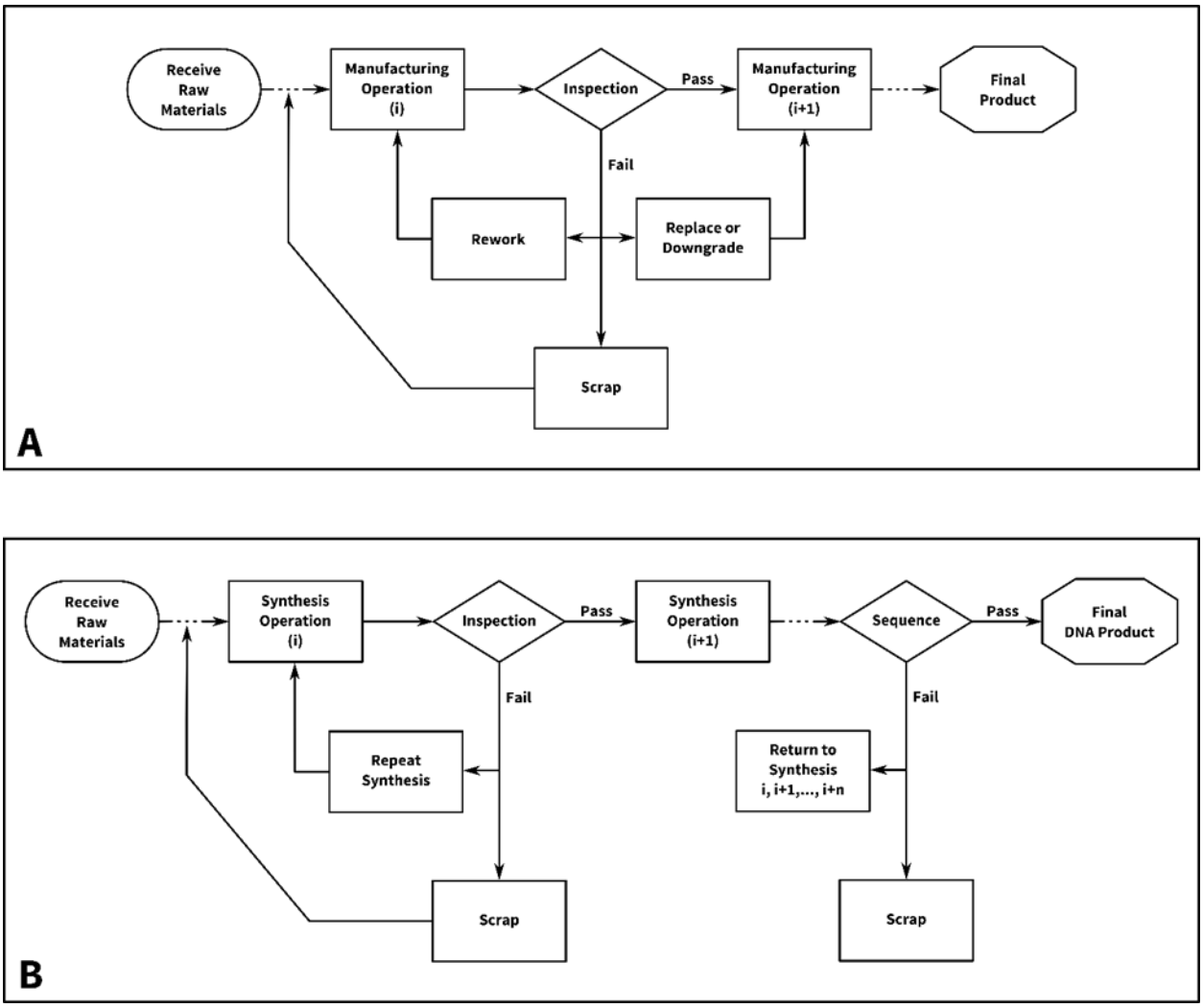

Figure 3: Process Differences Between TMS and GMS.

Panel A illustrates a traditional manufacturing process flow with an inspection step and potential options based on the result of the inspection. Panel B shows a typical gene synthesis process flow with both an inspection and sequencing step and the different options available should a failure occur at an inspection step.

3.1.2 *End Product Differences with a Genetic Manufacturing System*

The end product of GMS is a single genetic product that matches a desired reference sequence. This difference and many of the other differences discussed in this section are outlined in Figure 4.

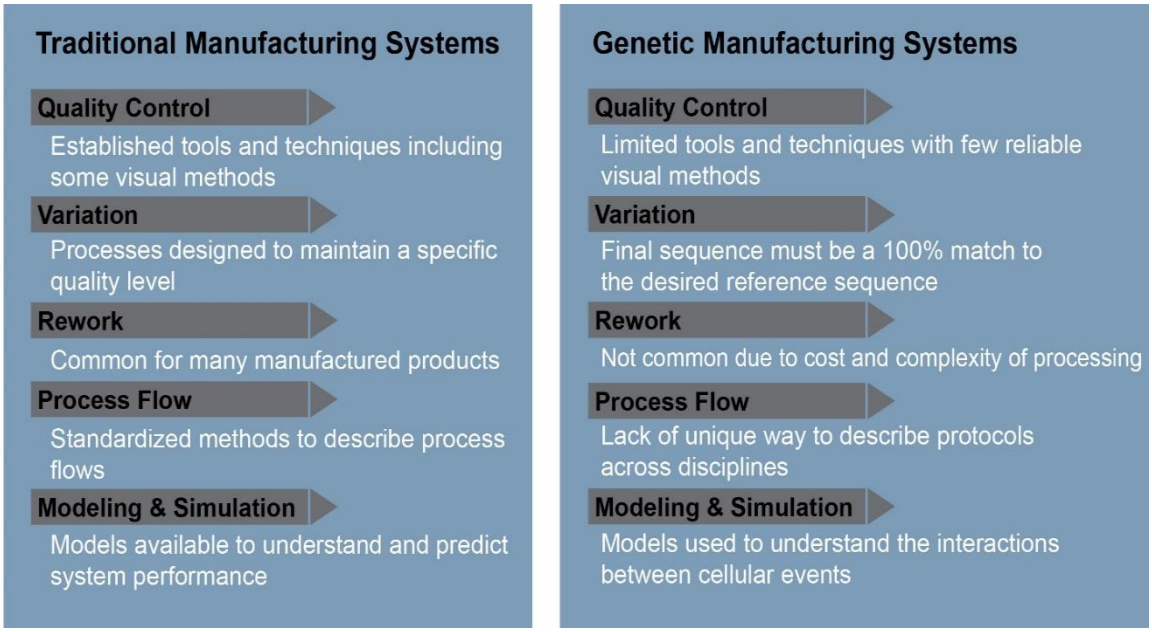

Figure 4: Major Differences Between TMS and GMS.

As shown, one of the key differences between traditional and genetic manufacturing systems is variation. There is usually no room for error or variation in GMS as even a single base pair inserted incorrectly or misplaced could drastically change the function of the sequence or final product. After a sequence has been created, perfect clones can be produced through established cloning techniques, meaning much of the time and cost of GMS is spent on creating the first perfect sequence.

3.2 Process Modeling for Genetic Manufacturing Systems

The differences between traditional and genetic production environments outlined in the previous section are important considerations when trying to model or simulate a GMS. However, before directly creating an abstraction of a GMS in the form of a model or simulation, a simplified process flow describing the steps involved in the creation of a custom genetic construct is developed in this section. Additionally, a process flow with routing logic decisions illustrates the common final process in molecular and synthetic biology which is sending out samples for Sanger sequencing. The high-level gene synthesis process flow and detailed Sanger sequencing process flow are discussed in this section.

3.2.1 *Process Flow for Genetic Manufacturing Systems Utilizing Gene Synthesis Techniques*

Techniques for GMS utilizing gene synthesis vary depending on the protocol chosen for the creation of a specific DNA fragment. Describing genetic processing as a production process is an interesting way to visually depict the process flow and routing decisions in GMS. This is especially advantageous for training purposes as there is a steep learning curve to biological processing. Skills can be taught, but representing the overall process flow visually can be helpful in workforce training for those first starting out in the field. Process mapping, including value stream mapping, can describe both the process and information flows in a production system [65]. These techniques are useful in establishing the most efficient method of conducting a protocol.

From a high level, it is possible to break down the processing of a GMS utilizing gene synthesis into two phases. As depicted in Figure 5, the synthesis phase is followed by the

assembly phase. The synthesis phase consists of four process steps. Fragment creation includes the activities necessary to produce the fragment. Typically, this would include some computational design of the desired target fragment with a breakdown of the specific oligonucleotides which should be used to create the fragment. The process also includes the ordering and receiving of the raw oligonucleotides from a third-party vendor and the assembly of the oligonucleotides utilizing one of the methods described in Section 2.2.1. Then the fragment may be inspected using the techniques reviewed in Section 2.2.2. Cloning will be used to isolate a certain number of samples to be inspected or sent out for sequencing. At this stage of processing, it is not uncommon for samples to be sent out for sequencing to verify that the fragments are correct before moving onto the assembly phase.

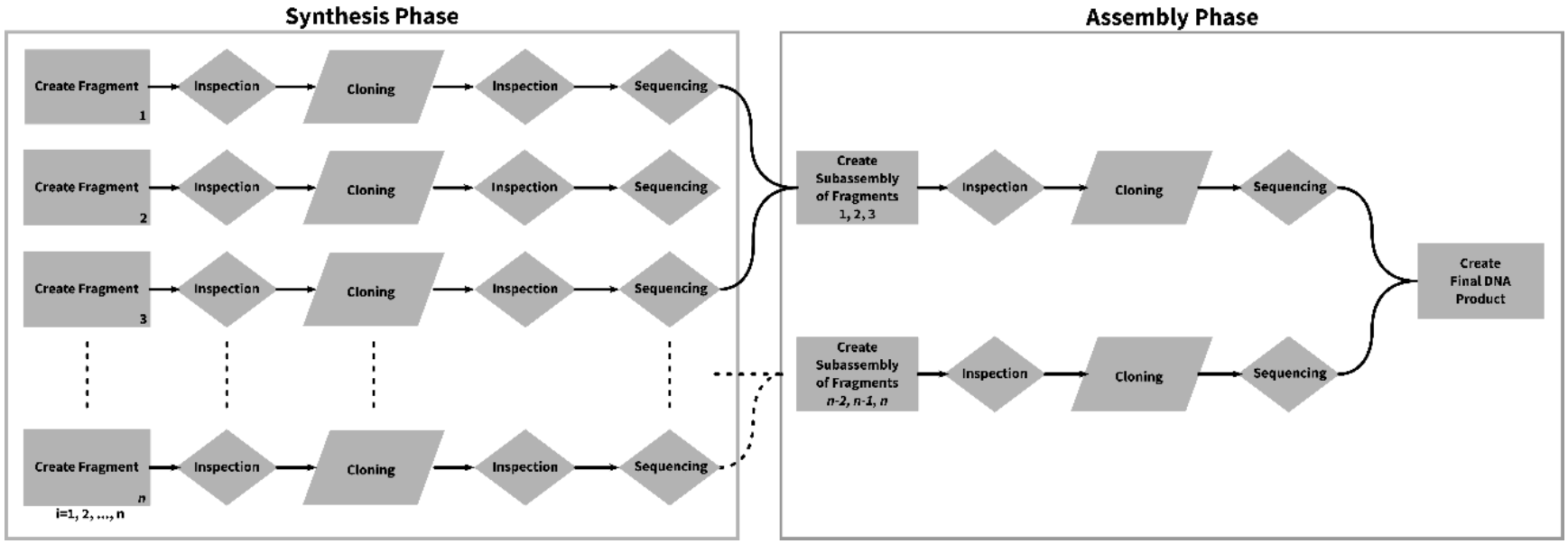

Figure 5: Synthesis and Assembly Phases in Gene Synthesis.

One key aspect of the synthesis phase is that this process can be parallelized based on the number of fragments which are necessary to produce the desired DNA molecule. Figure 6 shows a graphical view of the types of operations which occur in the synthesis phase.

Following the synthesis phase is the assembly phase. Multiple fragments will be combined into a DNA subassembly based on the methods discussed in Section 2.2.3. Following the creation of the subassembly, there is another inspection opportunity. Cloning is used to purify the newly assembled samples in search of the desired DNA sequence. The samples will then have another opportunity for inspection or sequencing.

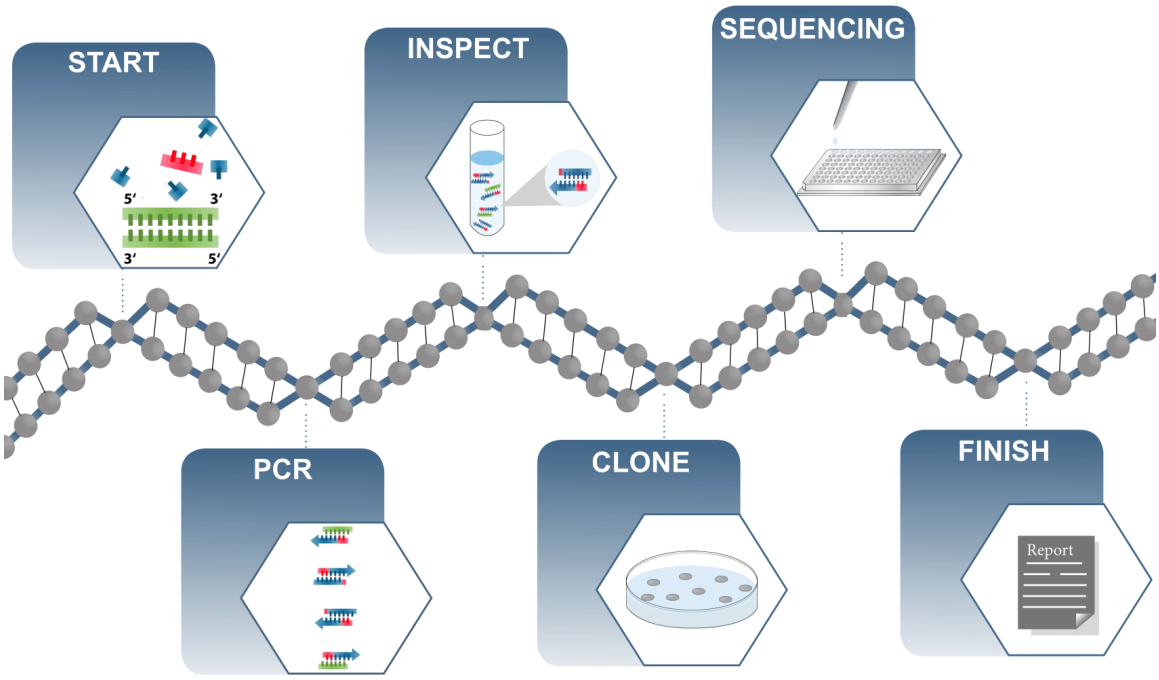

Figure 6: Process Steps in Synthesis Phase of Gene Synthesis Process Flow.

The assembly process is repeated as necessary based on the number of combinations needed before the subassemblies can be combined into the desired final DNA construct. After the assembly phase, a final sequencing operation will be conducted to confirm the DNA sequence matches the desired target sequence.

3.2.2 *Process Flow and Routing Decisions for Sequencing Processing Step*

To showcase the complexity of GMS the Sanger sequencing process has been broken down into a process flow with the routing decisions required of the technician. Figure 7 depicts the process flow and routing decisions for the sequencing process step which is commonly used in gene synthesis, synthetic biology, and the more common molecular biology.

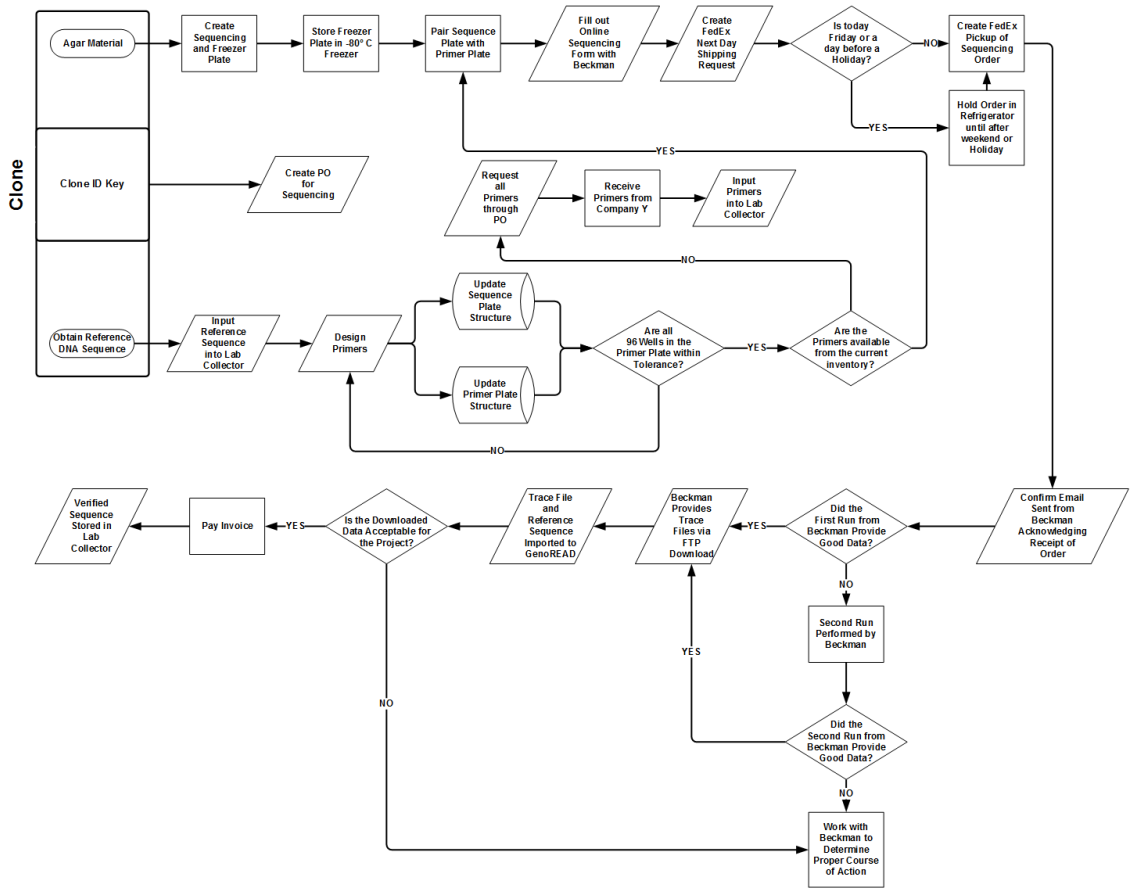

Figure 7: DNA Sequencing Process Flow and Routing Decisions.

Generating the process map and routing decisions for the sequencing process step is a compelling example for the complexity of GMS. Verifying the sequence of a series of clones by Sanger sequencing is one of the simplest and most straightforward processing steps found in a GMS utilizing gene synthesis processing. The specific aspects of the map include the preparation, shipping, and analysis of a previously synthesized sample. Despite the task's simplicity, the map contains 24 distinct process elements with six iterative decision loops. Translating a full gene synthesis protocol would require a significantly larger number of process elements. The complexity of this example illustrates the need for process understanding in gene synthesis. Grasping the level of complexity found in GMS,

the next logical step is to find ways to standardize the process operations to ideally reduce variation found in processing.

3.3 Modeling and Simulation Opportunities in Genetic Manufacturing Systems

Traditional manufacturing systems have relied on modeling and simulation for significant reductions in cost and time as well as increases in efficiency. However, given the differences between traditional and genetic manufacturing systems as outlined in Section 3.2, cases to directly apply these models are limited. Instead, there are opportunities to adapt the modeling and simulation techniques, given the constraints of genetic processing, to provide insight on how to most efficiently undertake various processing flows and operations.

One example of utilizing TMS analysis within GMS is determining the proper inspection resources to allocate for a process flow. Inspection can reduce the prevalence of errors and limit repeated work on non-conforming samples. Inspection allocation models determine if and where inspection allocation resources should be placed within a production flow. These types of models are common in TMS with different types of error being considered as well as a variety of solution techniques [66-81]. Adapting these types of models is possible for GMS by focusing on formulations geared towards the industrial biotechnology sector. Specifically, in this research, an inspection allocation model is formulated for a serial, multi-stage genetic manufacturing system.

Both mathematical modeling and discrete-event simulation serve as useful tools for modeling an academic or industrial GMS research environment. Specifically, due to the complex routing decisions and iterative nature of research projects, discrete-event

simulations provide a valuable tool for this type of analysis. The remainder of this dissertation will investigate how to mathematically model an academic and industrial research GMS as well as employ discrete-event simulation as a solution methodology to determine the appropriate inspection strategy for a practitioner.

3.4 Rationale for Utilizing Traditional Manufacturing Systems Analysis Techniques in Genetic Manufacturing Systems

The processes used in gene synthesis to generate novel genetic constructs require costly and time consuming iterative steps. Ultimately, the goal is to produce a sequence matches the desired reference sequence. After this sequence has been created, identical clones can be made in high volumes with great precision using established techniques. This major difference from traditional manufacturing processing provides a unique opportunity to investigate this new class of manufacturing system to increase their efficiency as well as reduce production cost and time.

Understanding the differences between these two types of systems will be integral to investigating new modeling and simulation opportunities in the new interdisciplinary field of genetic manufacturing systems. Combining the complexity and random nature of biological organisms with the structured examination provided by traditional manufacturing analysis yields an opportunity to provide logical processing rationale for a number of challenges facing gene synthesis and the broader fields of synthetic and molecular biology.

4. Inspection Allocation Strategies for Genetic Manufacturing Systems Utilizing Gene Synthesis

The cost of synthesizing DNA has been decreasing steadily since 1998 and today costs half as much as it did fifteen years ago [64]. However, DNA synthesis dominates the cost of producing custom fragments because DNA sequencing technologies have substantially declined in cost over the last ten years [39, 64]. As a result, there is tremendous pressure to align the costs of DNA writing technologies (DNA synthesis) with the cost of DNA reading technologies (DNA sequencing). Processes associated with DNA writing and reading technologies form the basis for a class of manufacturing production environment referred to as genetic manufacturing systems (GMS) [82]. These systems yield a final product which consists of a genetic construct. The total production cost and time is an area of improvement for genetic manufacturing systems as life scientists are demanding reduced costs and quick turnaround times while ensuring high quality levels.

4.1 Inspection Modeling Considerations in Genetic Manufacturing Systems

The overall inspection strategy for GMS affects the cost and time of processing as well as the error rate. This leads to the question of what is the ideal inspection strategy at a given stage of the process? The best inspection strategy should minimize production cost and/or time while maintaining the quality of the final product.

For this study, inspection is defined as a supplementary processing step to ensure a synthesis operation achieves a desired level of quality. Choosing the appropriate inspection strategy has been a problem tackled for traditional manufacturing systems [76].

There are also some major differences between traditional manufacturing inspection and inspection strategies for GMS that prevent the direct application of existing models. These differences are discussed in Section 3.1. For the purposes of this section, we will focus on the inspection strategies available in genetic manufacturing systems utilizing gene synthesis processing. Visual optical inspection of DNA is not realistic due to the physical size of the molecular compounds making up the DNA sequence, thus requiring indirect inference of the DNA quality through other techniques. Inspection options are expensive, time-consuming, labor-intensive, and in some cases, have high error rates. Therefore, the choice of a particular inspection strategy has a considerable effect on the performance of a gene synthesis process.

An emphasis of this dissertation is to provide justification for inspection allocation strategies in gene synthesis by viewing the problem from a manufacturing perspective. The remainder of this section presents an inspection allocation mathematical formulation for a serial, multi-stage gene synthesis system. The mathematical formulation describes the inspection allocation decisions for academic and industrial research environments. Background information on gene synthesis processing, inspection options, and inspection strategies are provided in Section 2.

4.2 Inspection Allocation Literature Review

Inspection allocation has been a common focus of improvement initiatives for traditional manufacturing settings. Many of the early works in inspection allocation formulated the problem as a mathematical model representing a serial single-line, multi-stage production process with error-free inspection and solved the problem using dynamic programming

[73-75, 79]. White [80] formulated the error-free inspection allocation model of a serial production process as a shortest path problem that was solved using a graphical network approach.

Error-free inspection is not always a realistic assumption. Multiple authors have addressed the inspection allocation problem in the presence of different errors which can occur at the inspection station. These models account for the case when a good part is classified as a defective part (Type I error) and when a defective part is identified as a good part (Type II error) with dynamic programming used as the solution methodology [66, 69]. Others have considered Type I and II errors for the inspection allocation problem using either non-linear or mixed integer programming model [67, 68, 81]. Lee and Unnikrishnan [72] apply three heuristic solution techniques to the serial multi-stage inspection allocation problem with errors and impose an additional constraint on inspection time. Most formulations assume 100% inspection of samples at any allocated inspection station. However, Heredia-Langner *et al.* [71] allow for partial lot inspection samples.

More recently, the inspection allocation literature incorporated additional complexities into mathematical models. Shiau *et al.* [77] utilize a non-linear mathematical model to simultaneously determine the processing resource (workstation) allocation considerations and inspection allocation decisions in an advanced manufacturing system with a genetic algorithm as the solution methodology. Van Volsem *et al.* [78] use both a discrete event simulation and an evolutionary optimization algorithm to not only determine if and when inspection stations should be located, but also the appropriate inspection targets at each location based on a mathematical model. Galante and Passannanti [70] created an integrated solution approach based on a mathematical model to incorporate scheduling and

inspection allocation decisions for a job-shop manufacturing environment, and solve this model with a genetic algorithm. Shetwan *et al.* [76] provide a robust review of the inspection allocation literature for serial, non-serial, and assembly manufacturing systems.

With this previous research in traditional manufacturing systems as a foundation, a mathematical formulation of the inspection allocation problem for genetic manufacturing systems is developed. The formulation considers a common gene synthesis process in a serial multi-stage configuration and the potential for both Type I and Type II errors.

4.3 Multi-stage Inspection Allocation Mathematical Model for Genetic Manufacturing System Utilizing Gene Synthesis

The inspection allocation formulation for creating custom DNA fragments in a genetic manufacturing system incorporates both Type I and Type II error while allowing for the iterative nature of genetic manufacturing systems. Specifically, the formulation is based on a serial, single-line, multistage process with the potential for inspection stations at any stage in the production process as described in Section 2.2. There are, however, major differences between GMS utilizing gene synthesis processing and more traditional manufacturing operations.

4.3.1 *Key Differences Between Traditional Manufacturing and Genetic Manufacturing Systems Related to Inspection Modeling*

Gene synthesis involves chemical and thermal processing equipment that can be used for multiple different operations. A unique physical station may not be associated with each operation in the process flow of a sequence being created. Therefore, a single station may be used for multiple types of synthesis processes. Thus, the traditional manufacturing

"stations" referenced in the inspection allocation literature will instead be referred to as "synthesis operations".

Traditional manufacturing systems typically create a variety of different types of parts or products, whereas genetic manufacturing systems utilizing gene synthesis create specific DNA sequences as the desired end product. The inspection allocation literature denotes a single entity as a part or unit. This terminology does not necessarily apply to gene synthesis, where a unit could be a chemical mixture containing millions of DNA sequences or a bacterial clone with a specific DNA sequence. Accordingly, the smallest granular unit which is considered within gene synthesis is referred to as a sample, which signifies a chemical mixture or individual bacterial clone. In either case, a sample takes on a similar meaning to a part or unit used in traditional manufacturing settings.

Gene synthesis inspection generally has fewer options when a sequence does not pass an inspection station. In traditional manufacturing, it may be possible to rework or downgrade the product to a lower quality level. When creating custom DNA fragments, however, the ultimate goal is a sample that provides a 100% positive match to the target reference sequence. Therefore, downgrading is not possible and rework is costly, unreliable, and time consuming. Because rework is generally not attempted, a non-conforming sample is scrapped and the process starts from an earlier stage.

4.3.2 *Parameter and Variable Definition*

The following section describes the parameters and decisions variable for the serial, single-line, multistage inspection allocation mathematical formulation for an academic or industrial research GMS utilizing gene synthesis.

Model Parameters

Model Indices

i Index for synthesis operation, $i = 1, 2, \ldots, I$

k Index for inspection type, $k = 1,2, \ldots, K$

Cost Parameters

rc Raw material cost at process start

sc_i Cost to synthesize samples at synthesis operation i

$ic_{i,k}$ Cost to inspect samples for inspection type k after synthesis operation i

Time Parameters

st_i Time to synthesize samples at synthesis operation i

$it_{i,k}$ Time to inspect sample for inspection station k

Error Parameters

$\rho_{i,k}$ Probability of Type I error for inspection type k after synthesis operation i, where a conforming sequence is incorrectly identified as nonconforming

$\theta_{i,k}$ Probability of Type II error for inspection type k after synthesis operation i, where a nonconforming sequence is incorrectly identified as conforming

Other Parameters

z_i Probability of a nonconforming sequence at synthesis operation i

$t_{i,k}$ $\begin{cases} 1 \text{ if inspection type } k \text{ can be used at synthesis operation } i \\ 0 \text{ otherwise} \end{cases}$

Decision Variables

$X_{i,k}$ $\begin{cases} 1 \text{ if inspection type } k \text{ is assigned to follow synthesis operation } i \\ 0 \text{ otherwise} \end{cases}$

$U_{i,k}$ Number of iterations of synthesis operation i using inspection type k to synthesize the sequence

TC Total cost of synthesizing the sequence

TFT Total flow time of synthesizing the sequence

The probability of a nonconforming sequence, z_i, is used to classify the process stability of a given synthesis operation. Values for the probability of a nonconforming sequence fall between 0 and 1. Zero represents a theoretically perfect process and one represents a process which will always fail. As values increase to one, this represents a higher likelihood that the process will be unsuccessful the first time through the system for

a given synthesis operation. Values for the probability of a Type I and II error also fall between the values of 0 and 1. Figure 8 illustrates the two-stage synthesis process considered in this section, with a final sequencing inspection to verify that the fragment matches the target sequence. The formulation developed in the remainder of this section applies to this process. However, it should be noted that additional synthesis operations could be added prior to the sequencing step and the formulation could also address these additional stages with minor alterations.

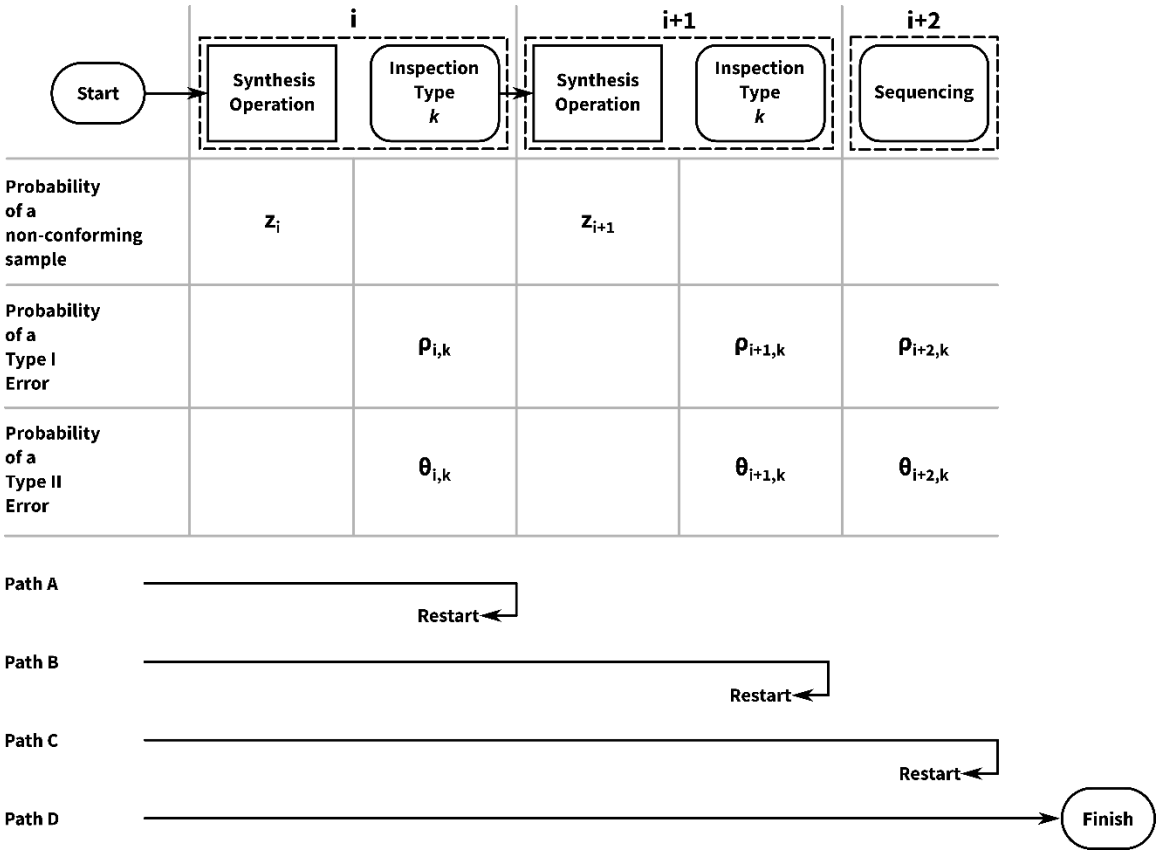

Figure 8: Serial Multi-stage Inspection Allocation Process Considered in the Mathematical Formulation.

Each synthesis operation is paired with a potential inspection opportunity. In the cases when an inspection strategy is not allocated, the inspection type will simply be no

inspection. The probability of a Type I or II error is indexed by both a synthesis operation *i* and inspection type *k* because the type of inspection may change even if it follows the same synthesis operation at different points in a process flow.

4.3.3 *Minimize Total Cost Problem Formulation*

In the mathematical formulation, the objective function minimizes the total cost to synthesize and inspect a sequence as shown in Expression (1):

$$Minimize\ TC = \sum_{i=1}^{I}\sum_{k=1}^{K}\left[sc_i + \left(ic_{i,k}\right)\cdot\left(X_{i,k}\right)\right]\cdot U_i + \sum_{k=1}^{K} rc \cdot U_1 \qquad (1)$$

The total cost is the sum of the cost to produce and inspect at a synthesis operation in addition to the required raw material costs. These costs are multiplied by the number of iterations necessary, $U_{i,k}$, for the given synthesis operation. The raw material costs, rc, for initializing the system are multiplied by the number of iterations of the first synthesis operation, $U_{1,k}$. These material costs are based upon the number of iterations needed at the first synthesis operation to eventually achieve one seemingly successful sample at the completion of the system.

An assumption of the formulation is the system will only allocate one type of inspection at each potential inspection opportunity following a synthesis operation. Additionally, only certain types of inspection strategies are available for use following a synthesis operation. An example of this would be not allowing a sequencing operation at inspection station *i+1* as it would be redundant to the final sequencing inspection shown in the two-stage process. This constraint is captured by Expression (2) as follows:

$$\sum_{k=1}^{K} t_{i,k} X_{i,k} = 1 \qquad \forall \quad i = 1, \dots, I. \qquad (2)$$

The number of iterations of a synthesis operation needs to be greater than one as each synthesis operation must be conducted at least once. However, the expected number of iterations at each inspection station is not constrained to be an integer in this formulation. Because the mathematical formulation will be used as a baseline for comparison of the discrete event simulation, not forcing an integer constraint will allow the average number of iterations reported in the simulation to be directly compared to the mathematical formulation. The constraint which ensures the number of iterations is greater than or equal to one for each synthesis operation is captured in Equation (3) as follows:

$$U_i \geq 1 \qquad \forall \quad i = 1, \dots, I. \qquad (3)$$

The decision variable related to the type of inspection strategy allocated at each synthesis operation is constrained to be a binary variable. Additionally, the total cost must be non-negative. These two constraints are represented by Equations (4) and (5), respectively, as follows:

$$X_{i,k} \in \{0,1\} \qquad \forall \quad \begin{matrix} i = 1, \dots, I \\ k = 1, \dots, K. \end{matrix} \qquad (4)$$

$$TC \geq 0 \qquad \forall \quad \begin{matrix} i = 1, \dots, I \\ k = 1, \dots, K. \end{matrix} \qquad (5)$$

To calculate the number of iterations of a synthesis operation, it is necessary to know the probability of a process following each of the paths shown in Figure 8. These paths describe the probability that a process will fail at a given synthesis operation, fail at final sequencing, or complete the process and finish with a seemingly successful sample.

Path A represents when a sample is being rejected at the first inspection station and then is restarted. This can occur through one of two different ways. First, the synthesis operation is a success, but a Type I error occurs at the inspection station. Second, the synthesis operation fails, and a Type II error does not occur. These two events are captured by Expression (6):

$$P(Path\ A) = (1 - z_i) \cdot \rho_{i,k} + z_i \cdot (1 - \theta_{i,k}). \tag{6}$$

For Path B, a sample progresses past the first synthesis operation but fails to move forward from the second inspection station. Thus, the sample is conforming after the previous station and does not incur a Type I error at the first inspection station or the sample is nonconforming after the previous station and does have a Type II error at the first inspection station, then it will erroneously continue to the second synthesis operation. This is captured by Expression (7) as follows:

$$\begin{aligned} P(Path\ B) = {} & \left[(1 - z_i) \cdot (1 - \rho_{i,k}) + z_i \cdot \theta_{i,k}\right] \\ & \cdot \left[(1 - z_{i+1}) \cdot \rho_{i+1,k} + z_{i+1} \cdot (1 - \theta_{i+1,k})\right]. \end{aligned} \tag{7}$$

The sequencing inspection verification is essentially another forced inspection step at the end of the process. To keep the model as generic as possible a “dummy” synthesis

operation is included. While synthesis operation *i+2* does not appear in Figure 8, it is reflected in the formulation through a "dummy" operation.

For Path C, the sample progresses through the previous inspection stations but fails at inspection station *i+2*. After the "dummy" synthesis operation at the sequencing verification, the probability of non-conformance will always be zero meaning there will never be a failure in the "dummy" process. However, there is still the ability for an error to occur in the sequencing inspection verification. If a sample moves into the sequencing verification step, then the Type I error applied for Path C but the Type II error is not considered because the probability of non-conformance is zero. The probability of following Path C is summarized by Expression (8) as follows:

$$\begin{aligned} P(Path\ C) = &\left[(1 - z_i) \cdot \left(1 - \rho_{i,k}\right) + z_i \cdot \theta_{i,k}\right] \\ &\cdot \left[(1 - z_{i+1}) \cdot \left(1 - \rho_{i+1,k}\right) + z_{i+1} \cdot \theta_{i+1,k}\right] \\ &\cdot \left[1 \cdot \rho_{i+2,k} + 0 \cdot \left(1 - \theta_{i+2,k}\right)\right]. \end{aligned} \tag{8}$$

For a sample to complete Path D, it must move through both synthesis operations, both inspection opportunities, and then successfully through the sequencing verification. The probability that a sample follows Path D is reflected by Expression (9):

$$P(Path\ D) = \left[(1 - z_i) \cdot \left(1 - \rho_{i,k}\right) + z_i \cdot \theta_{i,k}\right] \cdot \left[(1 - z_{i+1}) \cdot \left(1 - \rho_{i+1,k}\right) + z_{i+1} \cdot \theta_{i+1,k}\right] \cdot \left[1 \cdot (1 - \rho_{i+2,k}) + 0 \cdot (1 - \theta_{i+2,k})\right]. \quad (9)$$

The decision of whether a sample is conforming following the final sequencing inspection can be represented with a Bernoulli process because the sample will either be conforming or not conforming following the conclusion of sequencing. Each of the trials of the synthesis process are assumed to be independent. With the Bernoulli assumption, the number of iterations the system needs to achieve one success of the system shown in Figure 8 is estimated by the expected value of the geometric distribution. The probability of success is a function of the probability of conformance, $1 - z_i$, and the probability of not incurring a Type I error, such that:

$$P(Success) = \left[(1 - z_i) \cdot \left(1 - \rho_{i,k}\right)\right] \cdot \left[(1 - z_{i+1}) \cdot \left(1 - \rho_{i+1,k}\right)\right] \cdot \left[(1 - \rho_{i+2,k})\right]. \quad (10)$$

The expected value of the geometric distribution is denoted by $E(Y) = \frac{1}{p}$, which is used to determine the expected number of iterations necessary for one success in the two-stage genetic manufacturing system with a sequence verification. The expected number of iterations required is captured by Expression (11):

$$U_I = \frac{1}{P(Success)} = \frac{1}{\left[(1-z_i)\cdot(1-\rho_{i,k})\right]\cdot\left[(1-z_{i+1})\cdot(1-\rho_{i+1,k})\right]\cdot\left[(1-\rho_{i+2,k})\right]}. \quad (11)$$

For the mathematical formulation, where the assignment of inspection stations is a decision variable, the expected number of iterations required is captured by Expression (12):

$$U_I = \frac{1}{P(Success)} = \frac{1}{\left[\sum_{k=1}^{K}\left((1-z_i)\cdot(1-\rho_{i,k})\cdot X_{i,k}\right)\right]\cdot\left[\sum_{k=1}^{K}\left((1-z_{i+1})\cdot(1-\rho_{i+1,k})\cdot X_{i+1,k}\right)\right]\cdot\left[\sum_{k=1}^{K}\left((1-\rho_{i+2,k})\cdot X_{i+2,k}\right)\right]}. \quad (12)$$

The number of iterations of the entire system also provides an estimate of the number of iterations necessary at the first synthesis operation which is needed to calculate the raw material cost of the system. Thus, the number of iterations necessary at the first synthesis operation is found through Equation (13):

$$U_1 = \frac{1}{P(Success)} = U_I. \quad (13)$$

Equation (13) is used to calculate the synthesis and inspection costs from the first station in the objective function.

The number of iterations of the second synthesis operation can be determined by multiplying the iteration number of the first synthesis operation by the probability that the sample will move through the first inspection station with no Type I error or a Type II error occurring. Expression (14) reflects this logic as follows:

$$U_2 = U_1 \cdot \left[(1 - z_i) \cdot \left(1 - \rho_{i,k}\right) + z_i \cdot \theta_{i,k}\right]. \quad (14)$$

For the mathematical model, where the assignment of inspection stations is a decision variable, the expected number of iterations for the second synthesis station is captured as follows:

$$U_2 = U_1 \cdot \left[\sum_{k=1}^{K}\left((1 - z_i) \cdot \left(1 - \rho_{i,k}\right) \cdot X_{i,k}\right) + \sum_{k=1}^{K}\left(z_i \cdot \theta_{i,k} \cdot X_{i,k}\right)\right]. \quad (15)$$

Finally, the number of iterations required at sequencing is the product of the iterations required for the second synthesis operation and the probability that the sample will move through the second inspection station with no Type I error or a Type II error occurring. Expression (16) provides the formulation for this value:

$$U_3 = U_2 \cdot \left[(1 - z_{i+1}) \cdot \left(1 - \rho_{i+1,k}\right) + z_{i+1} \cdot \theta_{i+1,k}\right]. \quad (16)$$

For the mathematical model, where the assignment of inspection stations is a decision variable, the expected number of iterations for the third synthesis station is captured as follows:

$$U_3 = U_2 \cdot \left[\begin{array}{c}\sum_{k=1}^{K}\left((1 - z_{i+1}) \cdot \left(1 - \rho_{i+1,k}\right) \cdot X_{i+1,k}\right) \\ +\left(\sum_{k=1}^{K} z_{i+1} \cdot \theta_{i+1,k} \cdot X_{i+1,k}\right)\end{array}\right]. \quad (17)$$

Given the estimated number of iterations for each synthesis operation, inspection station, and final sequencing verification, the total cost of the system can be calculated through Equation (1).

As captured by Expressions (1)-(17), the inspection allocation model is a non-linear, mixed-integer program, which is difficult to solve for practical sized problem instances. In addition, the model simplifies the assumptions related to the synthesis operations. Specifically, the cloning operation would typically select multiple samples for an inspection strategy. For the purposes of this model, however, the number of sample selected at cloning is limited to one. Lastly, there are certain costs not considered in the formulation which are common for a traditional manufacturing system.

Traditionally, inspection allocation problems in manufacturing include internal failure costs. These costs would be associated with reworking parts which are deemed not up to a particular quality level. Rework for gene synthesis is significantly different than the additional processing conducted in macro-scale manufacturing. Gene synthesis relies on many chemical and thermal processes to allow custom DNA sequences to be created and the process of altering the sequence, after an error has been discovered, is quite complex. While it is technically feasible to conduct rework to change a few base pairs that are incorrect, the procedures are not commonly attempted. Thus, rework has been excluded from this mathematical formulation.

The external failure costs are normally defined as the cost of a defective product being passed along to the next entity in the supply chain. External failure costs include the warranty cost (otherwise known as the replacement cost) and cost to repair an item which is defective. Repair in gene synthesis suffers from the same complexities found in reworking a defective sample and therefore repairing a sequence is also excluded from this formulation.

4.3.4 *Minimize Total Flow Time Problem Formulation*

Another important measure in GMS utilizing gene synthesis is the total flow time for a sequence to be created. This section outlines a mathematical formulation of the total flow time for a sequence to be processed using gene synthesis. The objective function for minimizing the total flow time of a genetic manufacturing system based on the process described in Figure 8 is shown in Expression (18):

$$Minimize\ TFT = \sum_{i=1}^{I}\sum_{k=1}^{K}\left[st_i + \left(it_{i,k}\right)\cdot\left(X_{i,k}\right)\right]\cdot U_i \qquad (18)$$

The objective function for the total flow time is nearly identical to the total cost formulation, except that the time to synthesize and inspect is used instead of the cost. An assumption of the total flow time formulation is raw materials are received and no time is needed to prepare or procure the raw materials. Thus, the raw material portion of the objective function has been omitted.

In addition to the objective function described in Expression (18), the model includes Equations (2)-(4) and (7)-(17) from the total cost formulation of the inspection allocation problem for a GMS. Equation (19) is also needed to fully define the total flow time formulation of the problem and insure the objective function is non-negative.

$$TFT \geq 0 \qquad \forall \begin{matrix} i = 1, \dots, I \\ k = 1, \dots, K \end{matrix} \qquad (19)$$

This chapter has defined two mathematical formulations for a serial, multi-stage inspection allocation problem for a genetic manufacturing system. Total cost and total

flow time are the objectives of interest in the formulations. To properly calculate the cost and time, the number of iterations at each synthesis operation must be known. The formulation derives the method of estimating these values assuming a Bernoulli process and using the expected value of the geometric distribution. Given the complexity and simplifying assumptions of the mathematical formulation, it is possible to motivate the use of a discrete event simulation to capture the serial, multi-stage inspection allocation problem with some additional complexities in the remainder of this dissertation.

5. Fragment Synthesis Inspection Allocation Discrete Event Simulation Methodology

In order to provide insight on the appropriate inspection strategies for practitioners of a GMS utilizing gene synthesis, a discrete event simulation is developed. The mathematical formulation presented in Section 4 describes the inspection allocation problem for GMS utilizing gene synthesis with simplifying assumptions. To properly consider the random nature of gene synthesis processing and the complex routing and cloning logic, a discrete event simulation is used to model the best inspection strategies for both cost and time objectives.

5.1 Fragment Synthesis Discrete Event Simulation Solution Approach

To support the analysis of this work and provide insights relevant for academic and industrial research practitioners, a discrete event simulation is developed for a GMS utilizing gene synthesis. Simulation can be especially useful in this application due to its flexibility and scalability. Simulation allows for highly customizable, efficient analysis and provides solutions to the question of which inspection strategies are appropriate to minimize cost or time. For this simulation study, SIMIO has been selected as the software package due to its object oriented nature and ease of handling complex processing decision and routing logic [83].

For this simulation effort, the analysis focuses on a common two-stage gene synthesis process flow including a PCR step followed by a cloning step with a final sequence verification. These techniques were introduced in Section 2.1 and are chosen due to their prevalence not only in gene synthesis, but also the wider synthetic biology community.

5.1.1 *Fragment Synthesis Process Decisions and Entity Routing for Discrete Event Simulation*

The simulation is used to investigate the effects of different inspection strategies following both the PCR and cloning steps in the process flow. Figure 9 depicts the simulation process steps and their relation to the inspection allocation decisions to be made. The inspection operations that are feasible for each inspection station are shown in the circles.

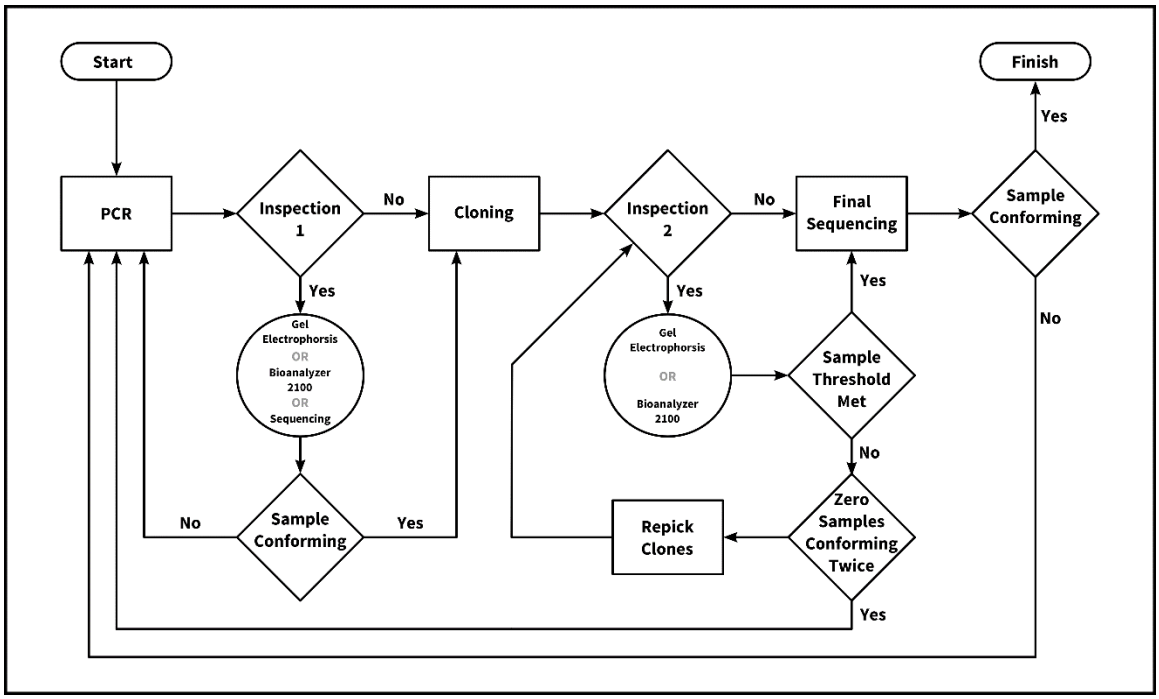

Figure 9: Fragment Synthesis Process Flow and Inspection Decisions for Discrete Event Simulation.

The simulation uses the probability of a non-conforming sample, ranging from 0 to 1, to describe the stability of a process, with zero denoting a process step that is successful and one denoting a process that is not successful. For this study, the probability of a non-conforming sample is assumed to be the same for each process in the system. Determining the true probability of a non-conforming sample for any synthesis operation has requires

further investigation within the synthetic biology community. These probabilities would likely change based on whether the sequence being created is similar in structure to other previous synthesized fragments. If previous process parameters and protocols are used as a starting point of the new sequence, then the probability of non-conformance would likely be closer to zero denoting a process more likely to succeed. However, these assertions have not been confirmed by the synthetic biology community and should be investigated in a future study. This type of information could inform future simulation models and assist practitioners in knowing the appropriate values for the probability of a non-conforming sample for their particular system. In addition to considering the true probability of a non-conforming sample for various synthesis processes, a future study could have different probabilities of a non-conforming sample for each station or time-varying probabilities.

Following the PCR or cloning step, a decision must be made about whether to assign an inspection station and which inspection strategy to employ. Figure 9 shows the available inspection options at each inspection station. For inspection station one the options are gel electrophoresis, the Bioanalyzer, or sequencing; while inspection station two has only gel electrophoresis and the Bioanalyzer as viable options. Inspection station two does not have sequencing as a viable option as it would be redundant with the final sequencing inspection which is conducted regardless of the inspection allocated at station two.

5.1.2 *Colony Picking Process Randomness*

To illustrate this point, Figure 10 shows the process flow considered in this simulation with the physical output of the PCR and cloning stages. The product of the cloning step is a

plate which should contain hundreds of bacterial colonies, with potentially different genetic sequences.

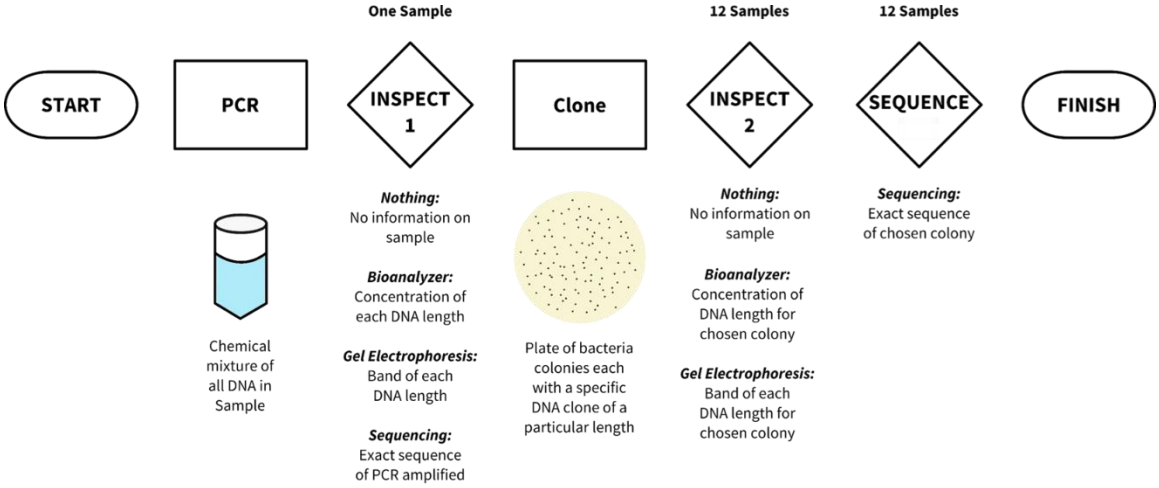

Figure 10: Simulation Process Flow with Output of Processing Steps.

Some of the colonies from the cloning step may contain the desired target sequence, but it is unlikely that all the colonies will contain the sequence of interest. Therefore, a number of colonies are picked from the plate for either inspection or final sequencing, depending on the inspection strategy.

Modeling the colony picking through the simulation is accomplished by representing the number of colonies on the cloning plate with a beta distribution *(Beta(6,6))* with a scaled lower bound of 75 and upper bound of 575. The beta distribution is used based on expert knowledge of gene synthesis practitioners, and the values are based on acceptable values commonly used in a gene synthesis laboratory. Based on their experience, the number of colonies produced during cloning approximately follows a normal distribution with an upper and lower threshold to denote when the cloning would typically not be successful. Due to these factors, the beta distribution with shape parameters (α=6, β=6)

and scaled with the appropriate factor provides similar performance to the cloning selection process. Additionally, the practitioners have found that whenever the number of colonies produced during cloning is approximately 200-300, this would suggest the cloning operation did not have a major flaw in the reaction, but does not denote if the individual samples actually contain the proper sequence of interest.

Another assumption of this simulation is the cloning step will produce some number of colonies. For the purposes of the simulation, the lower bound is chosen such that it could represent a success and the upper bound is selected to not produce too many colonies, which could denote a failure. A number of *U(0,1)* values will be created based on the value of the beta random variable.

For this simulation study, twelve samples are selected for use by the inspection stations or final sequencing to allow comparison across inspection strategies. The number of samples could take on values such as 6, 24, 48, or 96. In a future analysis, the number of samples selected during cloning could vary to determine the potential impact this decision may have on the cost and time associated with processing. To replicate this in the simulation, twelve *U(0,1)* samples are selected from the total beta random variable representing the colonies created, without replacement. Any of the twelve samples which appear to be conforming following inspection are stored and the rest discarded. Colonies are re-picked from the clone population until a threshold of twelve seemingly conforming samples can be passed along to final sequencing. Colonies are re-picked up to five times. Then the conforming samples are passed along to final sequencing even if twelve are not found. This allows for a total of six clone batches of twelve samples to be evaluated for a total of 72 colonies. If no conforming samples are identified in both the initial inspection

and first re-pick, the cloning operation is assumed to be unsuccessful and the process is restarted from the beginning.

The logic of picking colonies during the cloning step is simplified in the mathematical model. With the discrete event simulation, this process can be accounted for through dedicated logic. Replicating the colony picking randomness is important to ensure the simulation appropriately considers the additional variance found in the cloning step.

5.1.3 *Fragment Synthesis Simulation Process Parameters*

Values for the cost and time of the PCR and cloning steps are summarized in Table 1. These values are treated as deterministic and are assigned per use of an operation. Deterministic values are used because the distribution of the cost and time associated with synthetic biology processing is not established. In reality, the specific technician would affect the cost and time associated with the synthesis operation. For cost, the amount of wasted reagents or materials would increase the cost based on the skill and expertise of the technician. Likewise, the technician's knowledge of the specific protocols may offset or decrease the time associated with the various processing steps. Both synthetic biology and gene synthesis rely on expert knowledge of the technician to properly apply the protocols used as well as troubleshoot problems which occur during the DNA fabrication process. These differences would account for a portion of the process variability, but this has not been formally quantified by the community.

Table 1: Cost and Time Process Parameters for Two Stage Fragment Synthesis Discrete Event Simulation.

Process Name	Process Time (Hours)	Technician Time (Hours)	Technician Cost (USD)	Material Cost (USD)
Raw Material	0	0	$0.00	$2.00
PCR	6	1	$25.00	$2.00
Cloning	24.5	1.5	$37.50	$16.60
Repick Cloning	0.5	0.5	$12.50	$3.60
Penalty Cost	-	-	-	-

Process Name	Total Cost (USD)
Raw Material	$2.00
PCR	$27.00
Cloning	$54.10
Repick Cloning	$16.10
Penalty Cost	$50,000

Table 1 summarizes the cost and time associated with the PCR, cloning, and re-pick cloning processes based on the protocols in the gene synthesis literature. Technician time is assumed to be $25 per hour. A cost of $50,000 is included to penalize any samples which are incorrectly passed onto the customer through a Type II error at the final sequencing inspection. This value is selected to estimate the cost of losing the current and future gene synthesis business from the customer receiving an incorrect sample.

The penalty cost is only applied in the simulation when all the samples delivered to the customer are incorrectly identified as conforming, when they truly are non-conforming. If any truly successful samples be delivered to the customer, then the Type II error penalty will not be applied. This assumption and the specific way the Type II error penalty is

applied could be altered in a future analysis to determine the most detrimental scenarios for synthetic biologists.

5.1.4 *Inspection Process Parameters*

Various inspection strategies are explored through the simulation with combinations of no inspection, gel electrophoresis, Bioanalyzer, and sequencing inspection options. The cost, time, and error rates for the various inspection strategies are summarized in Table 2.

Table 2: Cost, Time, and Error Rates of the Inspection Strategies Considered in Simulation.

Inspection Type	Process Time (Hours)	Technician Time (Hours)	Type I Error Rate	Type II Error Rate
No Inspection	0	0	N/A	N/A
Gel Electrophoresis	1	0.5	0.1	0.1
Bioanalyzer	1	0.25	0.05	0.05
Sequencing	48	2	0.0001	0.0001

Inspection Type	Material Cost (USD)	Technician Cost (USD)	Total Cost (USD)
No Inspection	$0.00	$0.00	$0.00
Gel Electrophoresis	$1.50	$12.50	$14.00
Bioanalyzer	$70.00	$6.25	$76.25
Sequencing	$576.00	$50.00	$626.00

Technician time is also considered with each inspection strategy. Costs are calculated based on prices attained through academic purchases from commercial vendors. Sequencing costs are based on outsourced Sanger sequencing to a major core laboratory.

Each inspection strategy assumes 12 samples are used in the calculation. Error rates are taken from literary sources [42, 84, 85].

5.1.5 *Fragment Synthesis Simulation Termination Parameters*

With many traditional manufacturing processes, the production process needs to ensure that the end products meet a prescribed quality threshold. When making DNA in an academic or industrial research environment, the criteria of success is the accurately determining one bacterial colony corresponding to a molecular event that produced a DNA molecule perfectly matching the sequence ordered by a customer. After proper identification, the colony can be reproduced indefinitely at minimal expense with a very low error rate by growing the bacteria in a culture. Additionally, the sequence can be replicated by the industrial biotechnology sector utilizing similar cloning methods in higher volumes. These differences necessitate an alternate termination condition in the discrete event simulation to model an academic or industrial research setting. Instead of running for a prescribed period of time, the simulation terminates after a single sample has been deemed successful following the final sequencing inspection. An important clarification is that the sample is identified as successful, but may in fact be a Type II error. A penalty cost is applied at the termination of the simulation when this occurs

To accomplish this alternative simulation termination condition, the simulation has integrated process logic to continue to run until a successful sample enters the final sink module. Upon the termination condition being met, the cost across the replication, total flow time through the system, and number of PCR process iterations are recorded along with the half width of a 95% confidence interval associated with these statistics.

Twelve different inspection strategies are explored to consider each possible way of distributing the inspection resources at stations one and two. For each experiment, values for the probability of a non-conforming sample of 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, and 0.95 are investigated. As the probability of a non-conforming sample approaches values greater than 0.95, the total cost and flow time of the system increase exponentially. The greatest half-width values for cost occur in the simulation study when no inspection is allocated at either inspection station for the probability of a non-conforming sample of 0.95. Therefore, this scenario can be used to determine the appropriate replication count for the fragment synthesis discrete event simulation. The number of replications, half-width value, percent and step reduction from each replication level, and computational time for the probability of a non-conforming sample of 0.95 for the strategy when no inspection is allocated at either inspection station, is shown in Table 3.

Table 3: Lowest Probability of a Non-conforming Sample Cost Half-Width Value and Computation Time (min) for Fragment Synthesis using Inspection (0,0).

Replication #	Half-Width	% Reduction	Step Reduction	Computational Time (min)
10k	298	-	-	120
25k	188	36.91	36.91	120.00
50k	133	55.37	18.46	240.00
75k	109	63.42	8.05	360.00
100k	94	68.46	5.03	720.00

Each experiment and process stability combination is run for 75,000 replications. The replication count was chosen because an increase from 75,000 replications to 100,000 replications yields only a 5.03% reduction in the half width while doubling the simulation

run time. Due to the relatively small decrease in the worst-case scenario half width, 75,000 replications have been used for the fragment synthesis discrete-event simulation runs.

5.2 Results for Fragment Synthesis Discrete Event Simulation

Based on the simulation model described in Section 5.1, the results of the simulation analysis are presented. Specifically, the following sections present the best inspection strategies to use when cost, time, or the PCR iteration count are the objective of the fragment synthesis simulation.

5.2.1 *Total Cost of Inspection Strategies for Fragment Synthesis*

The discrete event simulation model is utilized to investigate the best inspection strategies to employ for minimizing cost or time. Table 4 summarizes the total cost for an inspection strategy for varying levels of the probability of a non-conforming sample. The inspection options have been abbreviated as no inspection (No), gel electrophoresis (Gel), Bioanalyzer (Bio), and sequencing (Seq); such that (Gel,No) indicates a gel electrophoresis used at inspection station one and inspection station two uses no inspection. The best solution for each value of the probability of a non-conforming sample is bold and italicized.

Table 4: Fragment Synthesis Total Cost (USD) for an Inspection Strategy for Probabilities of a Non-conforming Sample.

Inspection Strategy	Probability of a non-conforming unit							
	0	0.1	0.2	0.3	0.4	0.5	0.75	0.95
(No,No)	***709.1***	796.0	899.9	1,034.6	1,220.0	1,476.3	3,117.4	31,617.7
(No,Gel)	730.6	828.9	946.0	1,096.1	1,299.9	1,584.7	3,266.8	17,617.6
(No,Bio)	751.4	884.7	1,021.3	1,184.3	1,408.3	1,732.5	3,657.3	20,207.4
(Gel,No)	727.8	***743.9***	***762.1***	***784.2***	***815.7***	***857.5***	***1,159.2***	6,985.6
(Bio,No)	790.7	807.5	828.3	854.3	891.3	940.2	1,283.1	7,990.6
(Seq,No)	1,335.1	1,407.6	1,498.8	1,615.1	1,774.6	1,990.8	3,398.3	30,075.5
(Gel,Gel)	749.4	769.8	790.6	816.0	852.8	904.5	1,218.5	***3,694.2***
(Gel,Bio)	770.2	820.6	851.6	881.7	928.3	1,003.2	1,440.8	4,293.6
(Bio,Gel)	812.4	834.7	858.0	886.2	927.1	985.6	1,329.5	4,082.6
(Bio,Bio)	833.1	884.9	917.9	950.2	999.5	1,078.5	1,539.6	4,673.5
(Seq,Gel)	1,356.8	1,434.9	1,528.1	1,646.0	1,809.4	2,033.3	3,373.0	14,828.0
(Seq,Bio)	1,377.6	1,484.7	1,587.1	1,708.4	1,879.4	2,122.5	3,574.5	15,995.0

For this set of parameters, the results suggest no inspection at station one and two when the process is guaranteed to succeed. This result is expected because the system never rejects any samples, unless there is an error in the inspection, and thus reaches the theoretical minimum for the cost. However, as the probability of a non-conforming sample increases the best inspection strategy shifts towards those options involving gel electrophoresis. The error rate of running a gel is relatively low while the cost is almost five times less than the Bioanalyzer. The strategy of running a gel electrophoresis at inspection station one and no inspection at station two is the lowest cost for probability of a non-conforming unit of 0.1 to 0.75. As the probability of a non-conforming unit begins to approach one, the lowest inspection strategy shifts to utilizing a gel electrophoresis at

inspection station one and inspection station two. This result is also expected as the additional cost of running the gel electrophoresis at inspection station two is offset by reducing the amount of repeated work since there are so many failed samples at higher probabilities of a non-conforming unit. Using a gel at inspection station one and inspection station two has performance close to the minimum values as the probability of a non-conforming unit increases between 0.1 and 0.75.

Utilizing no inspection for either inspection station one, inspection station two, or both inspections stations one and two is only close to the lowest total cost for scenarios in which the probability of a non-conforming sample is 0.1. As the probability of a non-conforming sample increases, any strategy which has no inspection at inspection station one begins to be vastly inferior in terms of cost. This is especially apparent at a value of 0.95 for the probability of non-conformance where the total cost is between 2 and 10 times larger for the strategies which do not use inspection at either inspection stations one or two.

Using sequencing at inspection station one, as a preliminary quality control strategy, has consistently higher cost across all values of the probability of a non-conforming sample. This reinforces the need for affordable inspection strategies for mid-stream performance monitoring.

Figure 11 illustrates the total cost of an inspection strategy as a function of the probability of a non-conforming unit for the six strategies with competitive cost values. A 95% confidence interval has been included at the data point for each of the inspection strategies show in Figure 11.

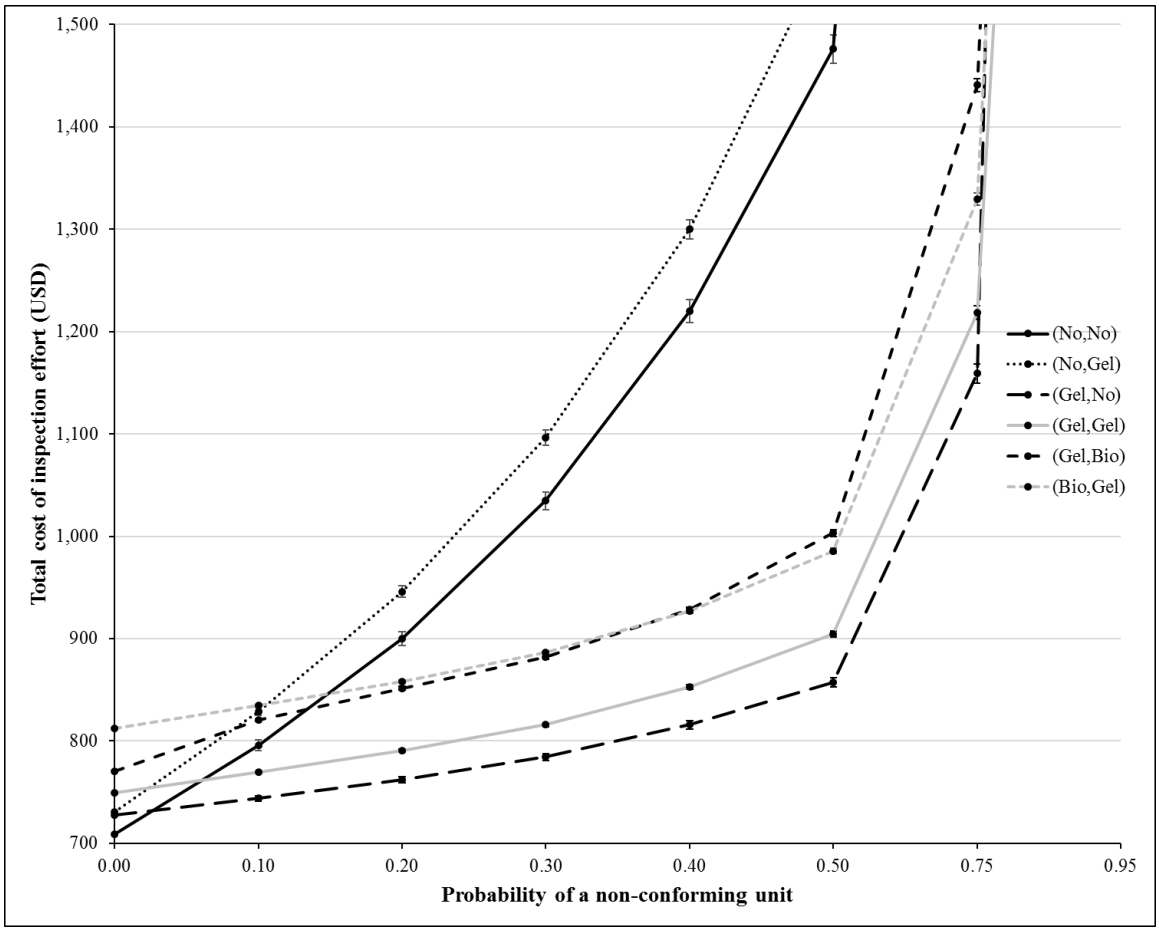

Figure 11: Fragment Synthesis Total Cost of Inspection Effort as a Function of the Probability of a Non-conforming Sample.

The inflection point when the (No,No) strategy switches to the (Gel,No) strategy can be clearly seen. Additionally, the inspection strategies which do not inspect at station one begin to dramatically increase as the probability of a non-conforming samples begins to approach one. Based on Figure 11, it is apparent that some of the inspection strategies still perform relatively well and could be considered if the best strategy is not available for an existing system.

Table 5 summarizes the percentage increase for total cost of each inspection strategy over the best inspection strategy identified through the fragment synthesis discrete event simulation.

Table 5: Fragment Synthesis Percentage Increase of Total Cost (USD) of an Inspection Strategy for Probabilities of a Non-conforming Sample.

		Probability of a non-conforming unit							
		0	0.1	0.2	0.3	0.4	0.5	0.75	0.95
Inspection Strategy	**(No,No)**	-	7.0	18.1	31.9	49.6	72.2	168.9	755.9
	(No,Gel)	3.0	11.4	24.1	39.8	59.4	84.8	181.8	376.9
	(No,Bio)	6.0	18.9	34.0	51.0	72.6	102.0	215.5	447.0
	(Gel,No)	2.6	-	-	-	-	-	-	89.1
	(Bio,No)	11.5	8.6	8.7	8.9	9.3	9.6	10.7	116.3
	(Seq,No)	88.3	89.2	96.7	105.9	117.6	132.2	193.2	714.1
	(Gel,Gel)	5.7	3.5	3.7	4.0	4.5	5.5	5.1	-
	(Gel,Bio)	8.6	10.3	11.7	12.4	13.8	17.0	24.3	16.2
	(Bio,Gel)	14.6	12.2	12.6	13.0	13.7	14.9	14.7	10.5
	(Bio,Bio)	17.5	19.0	20.4	21.2	22.5	25.8	32.8	26.5
	(Seq,Gel)	91.3	92.9	100.5	109.9	121.8	137.1	191.0	301.4
	(Seq,Bio)	94.3	99.6	108.2	117.8	130.4	147.5	208.4	333.0

Table 5 provides some interesting context about the types of inspection strategies which are still competitive when trying to minimize the total cost of inspection effort for the fragment synthesis process. For perfect processes, both the (No,Gel) and (Gel,No) inspection strategies have less than a 3% increase in cost over the option of utilizing no inspection at either station. While the options utilizing sequencing at the first inspection station have at least an 88% increase over the best inspection strategy.

As the probability of a non-conforming sample increases to 0.75, the (Gel,Gel) strategy has consistently strong performance and improves over the (Gel,No) strategy as the best inspection for probability of a non-conforming sample of 0.95. The (Gel,Gel) strategy is only a 5.1% increase over the (Gel,No) strategy even at the probability of non-conformance of 0.75. Additionally, the (Bio,No) strategy is less than a 10.7% increase over the (Gel,No) strategy up to the probability of a non-conforming sample of 0.75. At the probability of a non-conforming sample of 0.95, the (Bio,Gel) and (Gel,Bio) strategies are an increase 10.5% and 16.2% over the (Gel,Gel) strategy, respectively.

5.2.2 *Total Flow Time of Inspection Strategies for Fragment Synthesis*

Table 6 shows the total flow time in hours for various inspection strategies at specific values for the probability of a non-conforming unit. The best solution for each value of the probability of a non-conforming sample is italicized.

As expected, using no inspection at stations one and two is best for a perfect process. However, as the probability of a non-conforming unit increases, the best strategies incorporate the Bioanalyzer. Although the Bioanalyzer takes the same amount of time as running a gel, the error rate is lower than gel electrophoresis. This helps to explain why the strategies with a Bioanalyzer tend to be best while running a gel remains relatively close. For values of the probability of a non-conforming sample between 0.1 and 0.75, utilizing the Bioanalyzer at inspection station one and no inspection at inspection station two has the least total flow time. As the probability of a non-conforming sample approaches 0.95, the best strategy switches to using the Bioanalyzer at inspection station one and a gel electrophoresis at inspection station two.

Table 6: Fragment Synthesis Total Flow Time (hours) of an Inspection Strategy for Probabilities of a Non-conforming Sample.

Inspection Strategy	Probability of a non-conforming unit							
	0	0.1	0.2	0.3	0.4	0.5	0.75	0.95
(No,No)	***78.5***	87.1	97.9	112.0	131.1	157.3	324.6	3,235.6
(No,Gel)	80.6	90.0	101.5	116.5	136.6	164.5	330.6	1,752.3
(No,Bio)	80.2	88.7	98.7	111.6	128.8	152.9	297.3	1,606.9
(Gel,No)	80.3	82.0	84.2	87.0	90.9	96.1	131.6	821.9
(Bio,No)	79.9	***81.1***	***82.6***	***84.6***	***87.3***	***91.0***	***116.9***	634.9
(Seq,No)	126.5	132.5	140.0	149.6	162.7	180.6	297.0	2,509.0
(Gel,Gel)	82.3	84.4	86.7	89.7	93.9	99.6	134.0	423.9
(Gel,Bio)	82.0	84.2	86.4	89.1	93.0	98.4	130.7	413.9
(Bio,Gel)	81.9	83.5	85.1	87.2	90.2	94.3	118.9	***323.1***
(Bio,Bio)	81.5	83.3	84.9	86.9	89.7	93.7	117.6	326.4
(Seq,Gel)	128.6	134.8	142.5	152.1	165.5	183.7	292.8	1,236.1
(Seq,Bio)	128.2	134.7	142.4	152.1	165.4	183.7	293.3	1,306.8

Strategies which include using any combination of the gel electrophoresis or the Bioanalyzer at either inspection station one and two have consistently strong performance close to the lowest value. It should be noted that any strategy using sequencing increases the flow time of the system compared to other strategies.

Figure 12 shows the total flow time of an inspection effort as a function of the probability of a non-conforming sample for the six strategies with competitive total flow time values. A 95% confidence interval has been included for each data point. For values of 0.1 and 0.2 for the probability of a non-conforming sample, the strategies of using the Bioanalyzer and a gel or using the Bioanalyzer for both inspections, are almost identical.

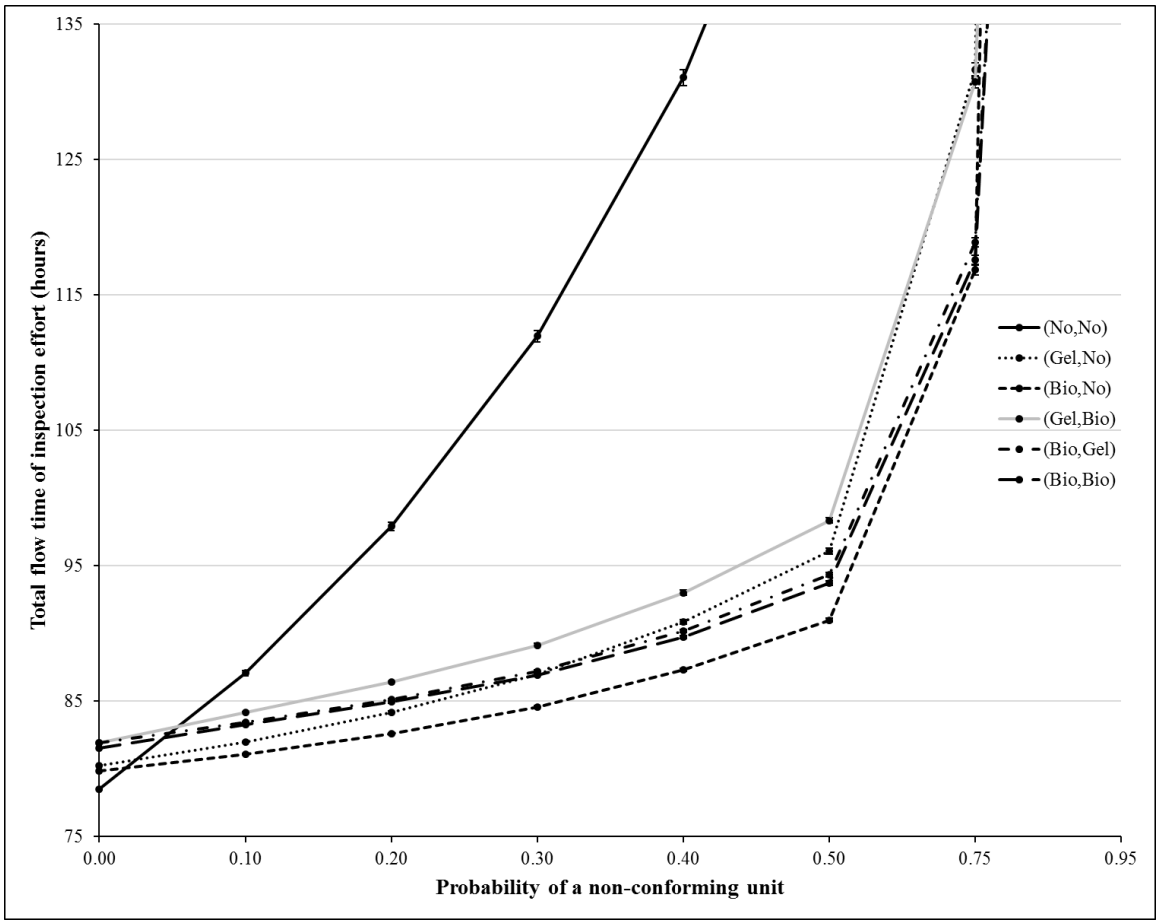

Figure 12: Fragment Synthesis Total Flow Time of Inspection Effort as a Function of the Probability of a Non-conforming Sample.

Utilizing the Bioanalyzer at either inspection station one, inspection station two, or both has strong performance for values of the probability of a non-conforming sample below 0.75. The exception to this result is the case when no inspection is used at inspection station one. It is clear that utilizing some inspection at station one can reduce the time of the fragment synthesis considerably.

Another important finding consistent with both the total cost and flow time is the detrimental impact sequencing has when conducted at station one. Utilizing sequencing causes the cost or time to nearly double over the best solution, regardless of the probability

of a non-conforming unit. Sequencing is crucial as a final check to ensure a sample matches the reference sequence, but is not beneficial at other inspection stations.

Table 7 shows the percentage increase of the time in system of each inspection strategy over the best inspection strategy identified through the fragment synthesis discrete event simulation. For the probability of non-conformance of zero, the strategies involving using gel electrophoresis or the Bioanalyzer have similar performance. Only the strategies involving sequencing have a serious detrimental affect over the best performing strategy. However, for value of the probability of a non-conforming sample between 0.1 and 0.3, the (Gel,No) option has particular strong performance followed closely by the (Bio,Gel) and (Bio,Bio) strategies. This result again supports the strategy of using a gel electrophoresis at inspection station one and utilizing the Bioanalyzer at inspection station one or two can perform well even if it is not the best strategy.

Table 7: Fragment Synthesis Percentage Increase of Time in System (hours) of an Inspection Strategy for Probabilities of a Non-conforming Sample.

Inspection Strategy	Probability of a non-conforming unit							
	0	0.1	0.2	0.3	0.4	0.5	0.75	0.95
(No,No)	-	7.4	18.5	32.3	50.1	72.9	177.8	901.3
(No,Gel)	2.6	11.0	22.9	37.7	56.5	80.7	182.9	442.3
(No,Bio)	2.1	9.4	19.5	32.0	47.5	68.1	154.4	397.3
(Gel,No)	2.2	1.1	1.9	2.9	4.1	5.6	12.6	154.3
(Bio,No)	1.7	-	-	-	-	-	-	96.5
(Seq,No)	61.1	63.4	69.5	76.8	86.4	98.4	154.2	676.4
(Gel,Gel)	4.9	4.1	5.0	6.1	7.6	9.4	14.6	31.2
(Gel,Bio)	4.4	3.8	4.6	5.4	6.5	8.1	11.9	28.1
(Bio,Gel)	4.4	2.9	3.0	3.1	3.3	3.7	1.7	-
(Bio,Bio)	3.9	2.7	2.8	2.8	2.8	3.0	0.6	1.0
(Seq,Gel)	63.8	66.3	72.5	79.8	89.5	101.8	150.6	282.5
(Seq,Bio)	63.3	66.2	72.4	79.8	89.5	101.8	151.0	304.4

For values of the probability of a non-conforming sample of 0.5 and 0.75, utilizing the (Bio,Bio) and (Bio,Gel) strategy show especially strong performance, respectively. As the probability of a non-conforming sample increases to 0.95, the (Bio,Gel) and (Bio,Bio) strategies have similar flow times. It would be more important to consider the availability of either strategy for a given system when deciding between the two competing alternatives. Additionally, the percentage increase continues to show the degradation of utilizing the (No,No) strategy specifically at high values of the probability of a non-conforming sample. At high values, the percentage increase falls between 177% and 901%, making this strategy severely detrimental over any of the viable alternatives.

5.2.3 *PCR Iteration Count for Inspection Strategies for Fragment Synthesis*

Another interesting statistic tracked through the discrete event simulation is the number of times the process is restarted and repeated. This statistic is captured in the simulation by tracking the number of times the PCR process step is conducted. This type of statistic is useful if the costs or times associated with certain synthesis operations were significantly different and repeated work became prohibitive from a cost or time perspective. Table 8 shows the PCR iteration count for various inspection strategies at specific values for the probability of a non-conforming unit. The best solution for each value of the probability of a non-conforming sample is italicized.

Table 8: Fragment Synthesis PCR Iteration Count of an Inspection Strategy for Probabilities of a Non-conforming Sample.

		Probability of a non-conforming unit							
		0	0.1	0.2	0.3	0.4	0.5	0.75	0.95
Inspection Strategy	**(No,No)**	***1.000***	***1.109***	***1.247***	***1.426***	1.670	2.004	4.135	41.218
	(No,Gel)	***1.000***	1.111	1.249	1.430	1.671	2.003	3.989	***21.168***
	(No,Bio)	***1.000***	1.111	1.249	1.430	***1.669***	2.001	3.996	22.382
	(Gel,No)	1.109	1.233	1.387	1.587	1.855	2.224	4.572	48.003
	(Bio,No)	1.051	1.168	1.315	1.502	1.756	2.105	4.335	45.710
	(Seq,No)	1.000	1.111	1.250	1.427	1.671	2.001	4.112	43.541
	(Gel,Gel)	1.109	1.233	1.387	1.585	1.851	2.217	4.425	23.764
	(Gel,Bio)	1.109	1.233	1.387	1.584	1.853	2.221	4.436	24.972
	(Bio,Gel)	1.051	1.168	1.315	1.501	1.756	2.103	4.187	22.440
	(Bio,Bio)	1.051	1.168	1.315	1.501	1.756	2.104	4.194	23.839
	(Seq,Gel)	1.000	1.111	1.250	1.427	1.671	***2.000***	***3.983***	21.323
	(Seq,Bio)	1.000	1.111	1.250	1.427	1.671	***2.000***	3.984	22.610

For probability of a non-conforming sample of zero, the inspection strategies which yield the lowest PCR iteration count are those which do not use an inspection strategy at the first inspection station. This is due to a perfect process which limits the repetition of the PCR processing step. As the probability of a non-conforming sample increases to between 0.1 and 0.3, the best inspection strategy for limiting the PCR iteration count is no inspection at both stations one and two. At a probability of non-conformance of 0.4 the best strategy changes to the (No,Bio) option. As the value of the probability of a non-conforming sample increases to between 0.5 and 0.75, the best inspection strategy becomes the (Seq,Gel) or (Seq,Bio) option. At the highest value of the probability of a non-conforming sample of 0.95 the best strategy switches to the (No,Gel) option. One important observation from the PCR iteration count statistic is that many of the other strategies also have strong performance in relation to the best inspection strategy. To highlight this, Table 9 shows the percentage increase of PCR iteration count for an inspection strategy based on the various probabilities of a non-conforming sample. Based on Table 9, many competing inspection strategies are viable for reducing the PCR iteration count.

Table 9: Fragment Synthesis Percentage Increase of PCR Iteration Count of an Inspection Strategy for Probabilities of a Non-conforming Sample.

Inspection Strategy	Probability of a non-conforming unit							
	0	0.1	0.2	0.3	0.4	0.5	0.75	0.95
(No,No)	-	-	-	-	0.048	0.187	3.814	94.720
(No,Gel)	-	0.108	0.119	0.272	0.118	0.141	0.148	-
(No,Bio)	-	0.126	0.136	0.240	-	0.069	0.328	5.738
(Gel,No)	10.920	11.136	11.222	11.299	11.131	11.209	14.789	126.770
(Bio,No)	5.101	5.286	5.451	5.311	5.215	5.227	8.826	115.940
(Seq,No)	0.003	0.106	0.197	0.090	0.127	0.028	3.247	105.691
(Gel,Gel)	10.920	11.125	11.201	11.114	10.910	10.861	11.108	12.265
(Gel,Bio)	10.920	11.107	11.178	11.089	11.017	11.022	11.361	17.973
(Bio,Gel)	5.101	5.282	5.397	5.251	5.215	5.167	5.128	6.008
(Bio,Bio)	5.101	5.263	5.379	5.243	5.204	5.200	5.285	12.618
(Seq,Gel)	0.003	0.106	0.197	0.090	0.123	-	-	0.733
(Seq,Bio)	0.003	0.106	0.197	0.090	0.124	0.003	0.018	6.812

Instead of looking at the average number of the PCR iteration count, Table 10 shows the maximum value of the PCR iteration count for the inspection strategies at the varying values for the probability of a non-conforming sample. Based on the maximum value of the PCR iteration count, the (No,Gel) inspection strategy performs well for all values of the probability of a non-conforming sample. Many other inspection strategies match the lowest value for the maximum PCR iteration count until the probability of a non-conforming sample increases to the higher values. At these higher values, the spread of the maximum value of the PCR iteration count begins to increase.

Table 10: Fragment Synthesis Maximum Value of the PCR Iteration Count of an Inspection Strategy for Probabilities of a Non-conforming Sample.

Inspection Strategy	Probability of a non-conforming unit							
	0	0.1	0.2	0.3	0.4	0.5	0.75	0.95
(No,No)	***2***	6	9	11	14	17	***40***	408
(No,Gel)	***2***	***5***	***7***	12	***13***	***15***	43	***222***
(No,Bio)	***2***	***5***	***7***	11	***13***	20	47	233
(Gel,No)	5	7	9	12	14	20	50	597
(Bio,No)	4	7	9	10	***13***	19	43	543
(Seq,No)	***2***	5	8	***9***	***13***	20	46	556
(Gel,Gel)	5	7	9	12	16	24	47	273
(Gel,Bio)	5	7	8	12	17	27	43	281
(Bio,Gel)	4	7	9	10	17	24	47	304
(Bio,Bio)	4	7	9	10	***13***	27	43	277
(Seq,Gel)	***2***	***5***	8	***9***	***13***	20	45	283
(Seq,Bio)	***2***	***5***	8	***9***	***13***	20	45	283

This chapter has investigated the use of a discrete event simulation to model the synthesis of a DNA fragment in an academic or industrial research environment. The various assumptions, parameters, and results are reported. Three different measures are investigated through the simulation including the total cost, total flow time, and PCR iteration count of the fragment synthesis process given two different inspection opportunities with various inspection types available at these stations. Additionally, the simulation has quantified the process stability of the system through the probability of a non-conforming sample and investigated the effect of these parameters on the three metrics of interest.

6. Fragment Assembly Inspection Allocation Discrete Event Simulation Methodology

Chapter 5 described a simulation model to describe the fragment synthesis process of a genetic manufacturing system utilizing gene synthesis processing. Building on this foundation, a discrete event simulation is developed to analyze the fragment synthesis and multiple fragment assembly protocols used to create larger DNA sequences of interest. The discrete-event simulation presented in this section incorporates three fragment synthesis processes conducted in parallel and combines the fragments into a longer fragment assembly based on the Gibson assembly protocol introduced in Section 2.2.3. The complex routing and cloning logic for each of the fragment synthesis operations is identical to the logic described in Chapter 5. Following the Gibson assembly protocol, a final sequencing check is used to ensure assembly quality. The discrete-event simulation allows for Type I and II inspection errors and incorporates a penalty cost for passing only non-conforming samples to the customer. In addition to describing the assumptions and parameters utilized in the fragment assembly simulation, this section also provides results to assist academic and industry research practitioners in determining the best inspection strategy when the goal is to minimize cost or time.

6.1 Fragment Assembly Discrete Event Simulation Solution Approach

A discrete event simulation approach is used to handle the complex routing and cloning logic found in fragment assembly operations for genetic manufacturing systems utilizing gene synthesis. Based on the success of this technique for the fragment synthesis process, a similar approach will be used to model the fragment assembly processes for a genetic

manufacturing systems utilizing gene synthesis techniques. Simio is selected as the software package employed for the solution procedure for the fragment assembly process due to the object-oriented nature of the software [83]. The object-oriented structure allows parent objects to be employed in the new model reducing the need for manually duplicating new instances of an object. The following section presents the simulation assumptions and results for the discrete event simulation for a fragment assembly process in an academic or industrial research environment.

For the fragment assembly discrete event simulation, three equal length DNA fragments are combined using the Gibson protocol introduced in Section 2.2.3.1. The three DNA fragments will mimic the logic for the fragment synthesis process found in Chapter 5, which includes a PCR operation followed by a cloning operation with a final sequencing to confirm each fragment matches the desired target sequence. Following the PCR operation and cloning operation, an inspection opportunity is considered. After the three fragments have been created through fragment synthesis, they are combined into a longer DNA sequence using the Gibson assembly protocol. The Gibson assembly protocol was chosen as the fragment assembly technique due to its prevalence in the synthetic biology and gene synthesis communities. Specific routing strategies and the underlying logic for the fragment assembly discrete event simulation are discussed in the following section.

6.1.1 *Fragment Assembly Process Decisions and Entity Routing for Discrete Event Simulation*

The purpose of the fragment assembly simulation is to investigate the effects of different inspection strategies used in the fragment synthesis process on the total cost and time of the system. Because the three fragment synthesis operations are conducted in parallel, the

fragment assembly occurs after three seemingly conforming fragments have been created. Figure 13 depicts the process steps found in the fragment assembly simulation including the different inspection options available during the fragment synthesis processes. The synthesis fragment blocks reference the synthesis fragment process logic, routing decisions, and inspection options from Figure 9. The fragment assembly simulation models the behavior of a system combining three fragments of similar size utilizing the Gibson assembly protocol [11, 52-54]. Each fragment is created in parallel, but the Gibson assembly cannot occur until the three fragments synthesis process steps have been completed.

The simulation does allow for a non-conforming fragment to be passed into the Gibson assembly. This can happen through two specific scenarios. First, if no inspection is allocated at inspection station two in a fragment synthesis block, and a Type II error occurs for all successful clones at the final sequencing inspection for the synthesis operation, then a non-conforming sample will be passed to the Gibson assembly process. Second, if there is an inspection strategy chosen at station two, then all the samples passed through inspection and final sequencing must have a Type II error occur for a non-conforming sample to be passed to the Gibson assembly process. These scenarios model the final sequence or fragment confirmation process step.

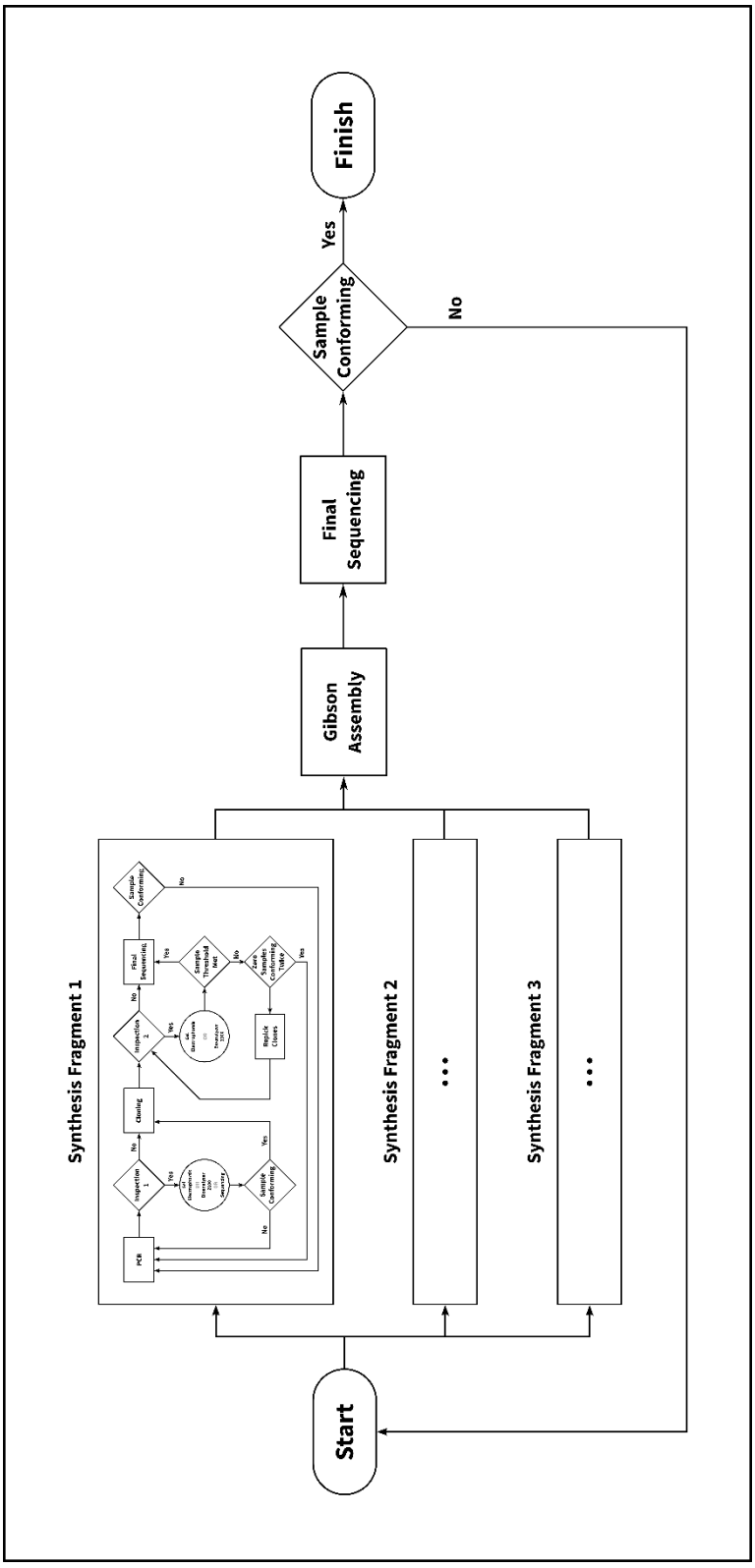

Figure 13: Fragment Assembly Process Flow and Inspection Decisions for Discrete Event Simulation.

Following the fragment synthesis processes, the Gibson assembly protocol combines the fragments. A final sequencing verification confirms the assembled fragments match the target reference sequence. If the sequence verification is successful, the fragment assembly finishes and the total cost and total flow time for the replication is recorded. Additionally, the PCR iteration count for each of the three fragment synthesis operations is recorded.

The probability of a non-conforming sample is used to describe the process stability of each of the synthesis operations found in the fragment assembly simulation. As with the fragment synthesis simulation, the value of the probability of a non-conforming sample varies between 0 and 1 using values of 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, and 0.95. This value is applied across the entire system for a given replication in a similar fashion as the fragment synthesis simulation. A future study could relax this assumption such that the probability of a non-conforming sample could vary by synthesis operation or change as a function of time.

Each of the fragment synthesis processes found in the fragment assembly simulation follow the cloning logic described in Section 5.1.1. For the fragment assembly simulation, twelve samples are selected for use in an inspection strategy or the sequencing operation. As was mentioned in Section 5.1.2, utilizing twelve samples is an assumption of the simulation which could be relaxed in a future analysis.

Each fragment has a sequencing final inspection prior to entering the Gibson assembly process. This is considered a best practice to evaluate the final sequence of a fragment to be included in the fragment assembly process, although it may not be consistently employed by all academic or industrial research laboratories. A future study could

investigate the impact of omitting this sequencing inspection prior to a fragment assembly operation.

This subsection has described the routing and decision logic for the three-fragment assembly simulation. The remaining subsections describe the different process parameters associated with the synthesis operations and inspection strategies considered in the three-fragment Gibson assembly process.

6.1.2 *Fragment Assembly Simulation Process Parameters*

Process parameter values for the fragment assembly discrete event simulation are similar to those used in the fragment synthesis simulation with two additions. Table 11 shows the various cost and time parameters for the synthesis operations in the fragment assembly discrete event simulation.

Gibson assembly is used as the fragment assembly process and is based on the New England BioLabs Inc. protocol [86]. Part of the process time and cost associated with the Gibson assembly process is a cloning transformation of the Gibson product. This is consistent with the specific protocol used for the process time and cost values. The Gibson assembly material cost is the single reaction cost based on the prices provided by New England BioLabs Inc. [87] for a ten reaction Gibson assembly cloning kit. Following the Gibson fragment assembly process, a final sequencing operation is performed to determine if the assembly produced the desired target sequence.

Table 11: Cost and Time Process Parameters for Three Fragment Assembly Discrete Event Simulation.

Process Name	Process Time (Hours)	Technician Time (Hours)	Technician Cost (USD)	Material Cost (USD)
Raw Material	0	0	$0.00	$2.00
PCR	6	1	$25.00	$2.00
Cloning	24.5	1.5	$37.50	$16.60
Repick Cloning	0.5	0.5	$12.50	$3.60
Gibson Assembly	25.5	2.5	$62.50	$18.50
Penalty Cost	-	-	-	-

Process Name	Total Cost (USD)
Raw Material	$2.00
PCR	$27.00
Cloning	$54.10
Repick Cloning	$16.10
Gibson Assembly	$81.00
Penalty Cost	$50,000

Fragment synthesis sequencing confirmations utilize twelve samples, while Gibson fragment assembly final sequencing has a different confirmation threshold. Based on the best practice of the Gibson assembly technique, two times the number of fragments used in the assembly are sequenced. Because three fragments are assembled through the Gibson protocol in this simulation, the sequencing threshold is set to six samples.

6.1.3 *Fragment Assembly Inspection Process Parameters*

Each fragment synthesis process allows inspection opportunities following both the PCR and cloning synthesis operations. The types of inspection strategies available following the PCR and cloning synthesis steps are consistent with the strategy allocations found in

Section 5.1.1.1. The specific values for the cost, time, and error rates for each of the potential inspection opportunities are consistent with those used in the fragment synthesis simulation found in Table 2.

A final sequencing inspection is used following the Gibson assembly protocol. Thus, an inspection option is not allowed following Gibson assembly. Instead Due to the high combination efficiency of the Gibson assembly procedure and the low error rate of sequencing, if the sequencing inspection yields no seemingly conforming samples, then the entire fragment synthesis and assembly process is restarted from the beginning. This means all three fragments are resynthesized as well.

One key assumption of the fragment assembly simulation is that the same inspection strategy will be used for each of the fragment synthesis processes for each replication run. This is a logical assumption to make since each of the fragment synthesis blocks are creating similar sized fragments in parallel. If the fragments were producing drastically different sized fragments, then different inspection strategies may be used for each block. Investigating the best inspection strategies to use for each fragment of a differing length in a fragment assembly process would be an interesting system to investigate in a future analysis.

6.1.4 *Fragment Assembly Simulation Termination Parameters*

As with the fragment synthesis simulation, termination occurs when a seemingly conforming fragment assembly is created and successfully sequenced. The simulation does allow for a Type II error to occur and imposes a penalty cost when this occurs. The simulation continues to run until a successful sample enters the final sink module after the

sequencing inspection operation following the Gibson assembly. Upon the termination condition being met, the cost across the replication, total flow time through the system, and the number of PCR process iterations for each fragment synthesis are recorded along with the half width of a 95% confidence interval associated with these statistics.

Twelve different inspection strategies are explored which is each possible way of distributing the inspection resources at stations one and two. For each experiment, values for the probability of a non-conforming sample of 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.75, and 0.95 are investigated. As the probability of a non-conforming sample approaches values greater than 0.95, the total cost and flow time of the system increase exponentially. Each experiment and process stability combination is run for 100,000 replications. This replication number was selected as an increase to 125,000 replications only decreases the worst case half-width value by 10.4% but doubles the computational time requirement.

6.2 Results for Fragment Assembly Discrete Event Simulation

The following section presents the results from the fragment assembly discrete event simulation model presented in Section 6.1. Specifically, the following sections present the best inspection strategies to use when cost, time, or the PCR iteration count are the objective of the fragment assembly simulation.

6.2.1 *Total Cost of Inspection Strategies for Fragment Assembly*

To understand the best inspection strategies for the upstream fragment synthesis processes in a three-fragment assembly, the discrete event simulation minimizes cost or time. Table 12 summarizes the total cost of an inspection strategy for varying levels of the probability of a non-conforming sample for the assembly of three fragments using the Gibson assembly protocol. The inspection options have been abbreviated as no inspection (No), gel electrophoresis (Gel), Bioanalyzer (Bio), and sequencing (Seq). The best solution for each value of the probability of a non-conforming sample is bold and italicized.

Table 12: Fragment Assembly Total Cost (USD) of an Inspection Strategy for Probabilities of a Non-conforming Sample.

		Probability of a non-conforming unit							
		0	0.1	0.2	0.3	0.4	0.5	0.75	0.95
Inspection Strategy	**(No,No)**	***2,834.3***	3,071.6	3,368.6	3,747.6	4,274.2	5,052.5	11,638.5	386,202.9
	(No,Gel)	2,941.2	3,229.6	3,573.7	4,019.4	4,626.3	5,528.3	12,775.5	204,183.5
	(No,Bio)	3,190.2	3,612.2	4,042.6	4,566.2	5,306.1	6,424.1	15,263.9	251,870.2
	(Gel,No)	2,890.6	***2,932.0***	***2,983.5***	***3,051.4***	***3,150.4***	***3,312.7***	***5,024.8***	79,539.8
	(Bio,No)	3,079.6	3,128.7	3,190.0	3,270.6	3,387.5	3,576.5	5,514.9	91,184.7
	(Seq,No)	4,712.5	4,932.0	5,207.0	5,559.6	6,048.7	6,777.5	13,299.6	343,125.1
	(Gel,Gel)	2,997.3	3,060.2	3,122.5	3,201.5	3,320.8	3,524.5	5,352.0	***44,112.1***
	(Gel,Bio)	3,246.2	3,402.2	3,498.1	3,595.1	3,752.3	4,034.6	6,493.4	54,143.0
	(Bio,Gel)	3,186.4	3,254.7	3,324.3	3,412.4	3,544.7	3,767.3	5,769.9	48,758.6
	(Bio,Bio)	3,435.4	3,594.0	3,694.0	3,795.6	3,957.6	4,251.9	6,825.1	57,198.1
	(Seq,Gel)	4,819.3	5,055.9	5,338.2	5,693.3	6,194.3	6,958.0	13,311.8	169,878.1
	(Seq,Bio)	5,068.3	5,392.5	5,701.9	6,068.3	6,595.3	7,423.6	14,244.4	184,596.1

For the total cost of a three fragment assembly process, the results presented in Table 12 are consistent with the fragment synthesis case for a near perfect process. When the probability of a non-conforming sample is zero, the best inspection strategy for the fragment synthesis is no inspection at either station one and two. This result is reasonable because the system will only ever reject a sample due to an error in one of the inspection opportunities or final sequencing, and thus reaches the theoretical minimum for the cost of the system.

However, as the probability of a non-conforming sample increases the best inspection strategy shifts towards those options involving gel electrophoresis. The strategy of running a gel electrophoresis at inspection station one and no inspection at station two is the lowest cost for probability of a non-conforming unit of 0.1 to 0.75. This result is consistent with the single fragment synthesis simulation. Since the system now considers three fragment synthesis operations in parallel, similar results are achieved, but the magnitude of the cost is amplified by the complexity of the system.

As the probability of a non-conforming sample increases, the lowest inspection strategy shifts to utilizing a gel electrophoresis at inspection station one and two. This result is consistent with the fragment synthesis simulation as the additional cost of running the gel electrophoresis at inspection station two is offset by reducing the amount of repeated work. The importance of utilizing inspection at both stations is emphasized in the fragment assembly case due to the significant costs associated with longer fragment creation. This is likely due to the added complexity of creating multiple fragments coupled with the detrimental effect if a non-conforming fragment reaches the Gibson assembly stage. Using

a gel electrophoresis at inspection station one and two has similar performance close the best strategy as the probability of a non-conforming sample increases between 0.1 and 0.75.

Utilizing no inspection at inspection station one and two is best for the probability of nonconformance of zero. This strategy is only close to the lowest total cost scenario when the probability of a non-conforming sample is 0.1. As the probability of a non-conforming sample increases, any strategy which has no inspection at inspection station one becomes inferior in terms of cost. When the probability of a non-conforming sample is 0.95 the total cost is between 3 and 7 times larger for the strategies not using inspection at station one. Using sequencing at inspection station one, as a preliminary quality control strategy, has consistently higher cost across all values of the probability of a non-conforming sample.

Figure 14 illustrates the total cost of an inspection strategy for a three fragment assembly as a function of the probability of a non-conforming unit for the six strategies with competitive cost values. A 95% confidence interval has been included at the data point for each of the inspection strategies shown in Figure 14.

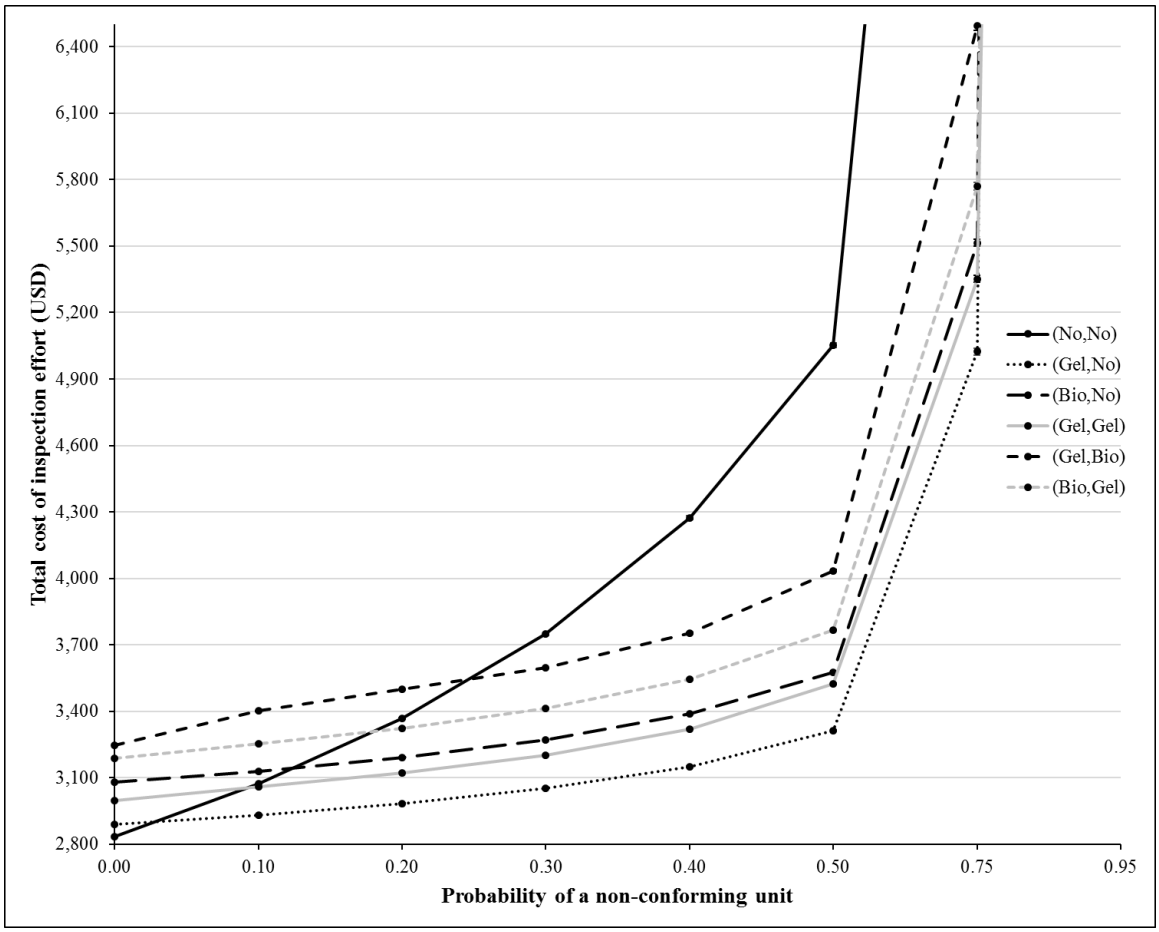

Figure 14: Fragment Assembly Total Cost of Inspection Effort as a Function of the Probability of a Non-conforming Sample.

Figure 14 is truncated at the probability of a non-conforming sample of 0.75 to increase readability. The best strategy switches from the (No,No) option for a perfect process to the (Gel,No) strategy as the probability of non-conformance values increase. Inspection strategies involving a gel electrophoresis at inspection station one also have strong performance. Table 13 summarizes the percentage increase for total cost of each inspection strategy over the best inspection strategy for a three-fragment assembly discrete event simulation.

Table 13: Percentage Increase of Total Cost (USD) for an Inspection Strategy with Probabilities of a Non-conforming Sample for Fragment Assembly.

Inspection Strategy	Probability of a non-conforming unit							
	0	0.1	0.2	0.3	0.4	0.5	0.75	0.95
(No,No)	-	4.8	12.9	22.8	35.7	52.5	131.6	775.5
(No,Gel)	3.8	10.2	19.8	31.7	46.8	66.9	154.2	362.9
(No,Bio)	12.6	23.2	35.5	49.6	68.4	93.9	203.8	471.0
(Gel,No)	2.0	-	-	-	-	-	-	80.3
(Bio,No)	8.7	6.7	6.9	7.2	7.5	8.0	9.8	106.7
(Seq,No)	66.3	68.2	74.5	82.2	92.0	104.6	164.7	677.8
(Gel,Gel)	5.8	4.4	4.7	4.9	5.4	6.4	6.5	-
(Gel,Bio)	14.5	16.0	17.2	17.8	19.1	21.8	29.2	22.7
(Bio,Gel)	12.4	11.0	11.4	11.8	12.5	13.7	14.8	10.5
(Bio,Bio)	21.2	22.6	23.8	24.4	25.6	28.4	35.8	29.7
(Seq,Gel)	70.0	72.4	78.9	86.6	96.6	110.0	164.9	285.1
(Seq,Bio)	78.8	83.9	91.1	98.9	109.3	124.1	183.5	318.5

For a perfect process, consistent with the fragment synthesis results, both the (No,Gel) and (Gel, No) strategies have less than a 3.8% increase over the (No,No) strategy. The inspection strategies not using inspection at the first station have poor performance compared to any other inspection option for perfect processes. As the probability of a non-conforming sample increases to 0.75, the (Gel, Gel) and (Bio,No) strategies are comparable to the best strategy, (Gel, No). For the probability of a non-conforming sample of 0.75, any option not using inspection at station one or a strategy employing sequencing at station one are over 100% greater than the best strategy.

Once the probability of non-conformance reaches 0.95, the (No,No) strategy is 775% worse than the (Gel,Gel) strategy. This continues to show the detrimental nature of using

no inspection at stations one and two for processes with low probability of a non-conforming sample values. For the probability of non-conformance of 0.95, the (Bio, Gel) option also has comparable performance.

6.2.2 *Total Flow Time of Inspection Strategies for Fragment Assembly*

Table 14 shows the total flow time in hours for various inspection strategies at specific values for the probability of a non-conforming unit. The best solution for each value of the probability of a non-conforming sample is italicized.

Table 14: Fragment Assembly Total Flow Time (hours) of an Inspection Strategy for Probabilities of a Non-conforming Sample.

		Probability of a non-conforming unit							
		0	0.1	0.2	0.3	0.4	0.5	0.75	0.95
Inspection Strategy	**(No,No)**	***152.0***	175.9	201.8	231.8	270.9	326.0	773.9	26,091.1
	(No,Gel)	154.5	179.7	206.9	239.5	280.5	338.2	787.3	12,632.0
	(No,Bio)	154.3	176.3	200.1	228.0	264.5	315.4	705.5	11,443.0
	(Gel,No)	155.1	159.8	165.4	172.2	181.4	195.4	322.9	5,931.7
	(Bio,No)	154.0	***157.4***	***161.2***	***166.0***	***172.6***	***182.8***	283.7	4,506.6
	(Seq,No)	200.0	216.5	234.4	255.2	282.0	320.3	650.4	17,360.3
	(Gel,Gel)	157.5	162.5	168.2	175.4	185.2	200.2	321.7	3,003.4
	(Gel,Bio)	157.2	162.0	167.3	173.8	183.0	196.7	312.1	2,942.5
	(Bio,Gel)	156.5	159.9	163.9	168.9	175.9	186.8	281.9	***2,289.9***
	(Bio,Bio)	156.2	159.7	163.5	168.1	174.6	185.0	***277.6***	2,301.3
	(Seq,Gel)	202.5	219.0	236.9	257.7	284.9	324.1	637.4	8,443.3
	(Seq,Bio)	202.3	218.9	236.9	257.7	284.9	324.4	636.5	8,968.7

For a perfect process, the best inspection strategy is using no inspection at either station one and two. Thus, the theoretical minimum flow time for the fragment assembly simulation, while still allowing for inspection errors, would be 152 hours. As the probability of a non-conforming sample increases, the best inspection strategies are those using the Bioanalyzer. The slightly lower error rates of the Bioanalyzer offset the additional cost when time is the primary objective of the system.

As the probability of a non-conforming sample increases to between 0.1 and 0.5, the (Bio,No) strategy becomes the best inspection option. This is consistent with the fragment synthesis simulation results. However, when the probability of non-conformance is 0.75, there is a divergence from the fragment synthesis results. For this value, the inspection strategy with the Bioanalyzer at both inspection station one and two requires the least total flow time. Due to the complexity of the three-fragment assembly process, the additional cost of the secondary inspection effort is offset by the reduction in time created by a redundant final sequencing on non-conforming clones. Once the probability of a non-conforming sample increases to 0.95, the best strategy switches to using a Bioanalyzer at inspection station one and a gel electrophoresis at inspection station two. However, using the Bioanalyzer at both inspection stations one and two has near identical performance. The fragment assembly half-width values for the total flow time of an inspection strategy for probabilities of a non-conforming sample are shown in Table 15. The best inspection strategies identified from the total flow time are italicized to allow comparison of the half-width values based on a 95% confidence interval.

Table 15: Fragment Assembly Total Flow Time (hours) Half-width Values of an Inspection Strategy for Probabilities of a Non-conforming Sample.

Inspection Strategy	Probability of a non-conforming unit							
	0	0.1	0.2	0.3	0.4	0.5	0.75	0.95
(No,No)	***0.001***	0.262	0.379	0.494	0.650	0.883	3.075	150.308
(No,Gel)	0.002	0.279	0.402	0.532	0.691	0.929	3.100	71.836
(No,Bio)	0.003	0.251	0.364	0.472	0.617	0.828	2.729	64.723
(Gel,No)	0.023	0.098	0.143	0.186	0.239	0.319	1.057	33.501
(Bio,No)	0.016	***0.070***	***0.102***	***0.135***	***0.179***	***0.249***	0.874	25.128
(Seq,No)	0.007	0.179	0.259	0.342	0.453	0.621	2.320	98.393
(Gel,Gel)	0.022	0.105	0.153	0.199	0.254	0.338	1.031	16.709
(Gel,Bio)	0.022	0.095	0.138	0.180	0.234	0.313	0.987	16.377
(Bio,Gel)	0.016	0.074	0.108	0.143	0.189	0.262	0.853	***12.737***
(Bio,Bio)	0.016	0.068	0.099	0.130	0.173	0.244	***0.830***	12.755
(Seq,Gel)	0.007	0.179	0.260	0.341	0.451	0.626	2.258	47.504
(Seq,Bio)	0.007	0.179	0.260	0.342	0.451	0.631	2.224	50.453

At the probability of a non-conforming sample of 0.95, the half-width of the measurements overlaps and it is not clear which strategy is best. Even though 100k replications are used for the three-fragment assembly process, there is still a significant amount of variation for high values of non-conformity. If the true value of non-conformance in the system is 0.95, both the (Bio,Gel) and (Bio,Bio) inspection strategies have strong performance if time is the primary objective.

Figure 15 shows the total flow time of an inspection effort as a function of the probability of a non-conforming sample for the six strategies with competitive values for

the fragment assembly process. The values for both the (Bio,Bio) and (Bio,Gel) strategies are almost identical for the majority of values of a non-conforming sample.

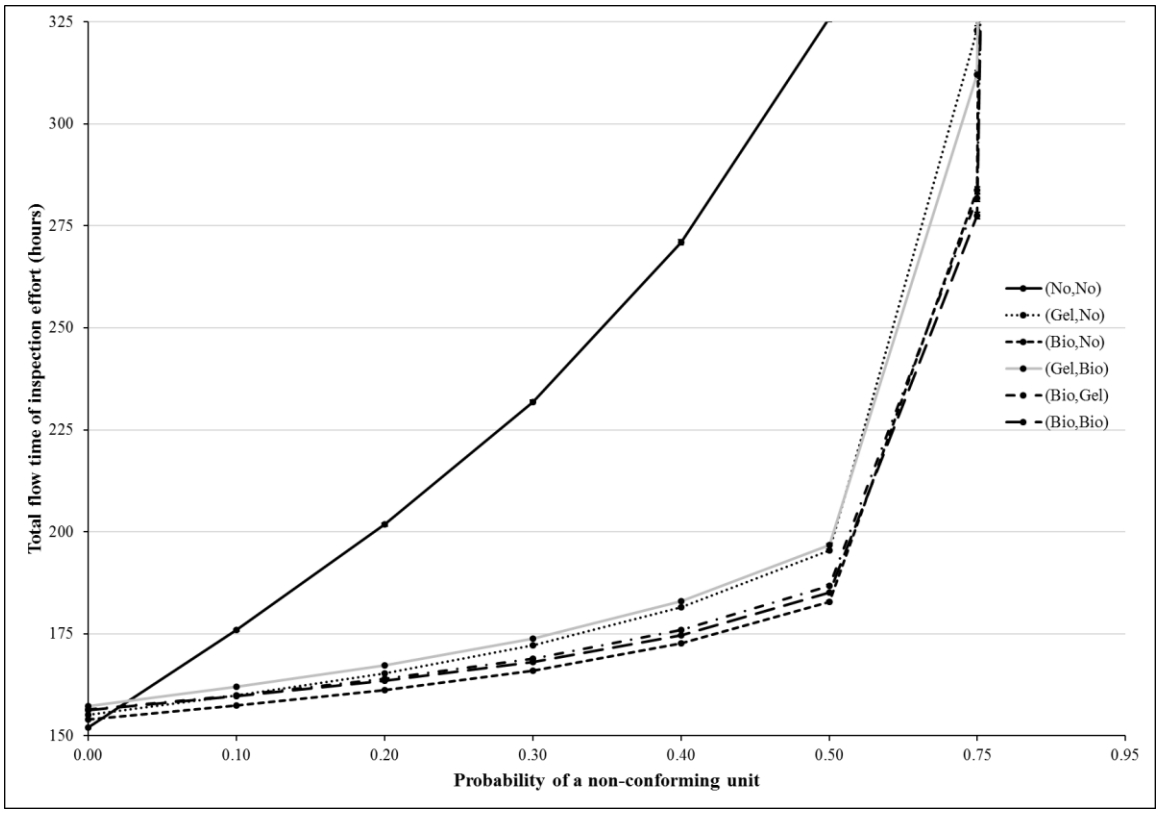

Figure 15: Fragment Assembly Total Flow Time of Inspection Effort as a Function of the Probability of a Non-conforming Sample.

The best inspection strategies when time is the primary objective use the Bioanalyzer at the first inspection station. The gap between the (Bio,No), (Bio,Bio), and (Bio,Gel) inspection strategies is small for most values of the probability of a non-conforming sample. Utilizing the Bioanalyzer at inspection station one has strong performance regardless of the strategy chosen at inspection station two.

Consistent with the fragment synthesis results, sequencing has a detrimental impact when used as a final verification technique. None of the six competitive strategies use

sequencing at inspection station one. The high time requirement of sequencing cannot be offset by the decreased error rate found through this type of inspection process.

Table 16 shows the percentage increase of the time in system of each inspection strategy over the best inspection strategy identified through the fragment assembly discrete event simulation.

Table 16: Percentage Increase of Time in System (hours) for an Inspection Strategy with Probabilities of a Non-conforming Sample for Fragment Assembly.

		Probability of a non-conforming unit							
		0	0.1	0.2	0.3	0.4	0.5	0.75	0.95
Inspection Strategy	**(No,No)**	-	11.8	25.1	39.6	57.0	78.3	178.8	1,039.4
	(No,Gel)	1.6	14.2	28.3	44.3	62.5	85.0	183.6	451.6
	(No,Bio)	1.5	12.0	24.1	37.4	53.3	72.5	154.2	399.7
	(Gel,No)	2.0	1.6	2.6	3.7	5.1	6.9	16.3	159.0
	(Bio,No)	1.3	-	-	-	-	-	2.2	96.8
	(Seq,No)	31.6	37.5	45.4	53.7	63.4	75.2	134.3	658.1
	(Gel,Gel)	3.6	3.2	4.3	5.7	7.3	9.5	15.9	31.2
	(Gel,Bio)	3.4	2.9	3.7	4.7	6.0	7.6	12.4	28.5
	(Bio,Gel)	2.9	1.6	1.6	1.7	1.9	2.2	1.5	-
	(Bio,Bio)	2.8	1.5	1.4	1.2	1.2	1.2	-	0.5
	(Seq,Gel)	33.2	39.1	46.9	55.2	65.0	77.3	129.6	268.7
	(Seq,Bio)	33.1	39.1	46.9	55.2	65.1	77.4	129.3	291.7

Less than a 2.2% difference exists between the (Bio,No), (Bio,Gel), and (Bio,Bio) strategies for values of the probability of a non-conforming sample of 0.1 and 0.75. Additionally, the sequencing strategies are a minimum of 31% greater than the lowest time in system inspection strategy and are worse at higher values of a non-conforming sample.

Any strategy involving the use of the Bioanalyzer at the first inspection station will be a sufficient choice for a three fragment Gibson assembly process. If the probability of a non-conforming sample is closer to 0.95, the Bioanalyzer at the first inspection station and either gel electrophoresis or the Bioanalyzer is best at the second inspection station. When the probability of a non-conforming sample is 0.95, the percent increase over the best strategy when no inspection is performed at either station is over 1000%. This further reinforces the need for some sort of inspection strategy with high values of non-conformance in a fragment assembly system.

6.2.3 *PCR Iteration Count for Inspection Strategies for Fragment Assembly*

Through the fragment assembly discrete event simulation, the PCR iteration count is tracked for each of the fragment synthesis processes. Table 17 shows the average PCR iteration count for various inspection strategies at specific values for the probability of a non-conforming sample. The values for each of the three-fragment synthesis PCR iteration counts are averaged and reported as one statistic. The best solution for each value of the probability of a non-conforming sample is italicized. It should be noted that some of the values appear to be equivalent. However, the italicized values are smaller due to the number of decimal places being rounded to increase the readability of the table.

Table 17: Fragment Assembly Average PCR Iteration Count of an Inspection Strategy for Probabilities of a Non-conforming Sample.

Inspection Strategy	Probability of a non-conforming unit							
	0	0.1	0.2	0.3	0.4	0.5	0.75	0.95
(No,No)	1.000	1.111	1.251	1.428	1.675	2.036	5.060	180.066
(No,Gel)	***1.000***	1.111	***1.250***	1.430	1.673	2.029	4.888	82.317
(No,Bio)	***1.000***	***1.111***	1.250	***1.427***	***1.672***	2.030	4.874	85.921
(Gel,No)	1.110	1.234	1.388	1.588	1.859	2.260	5.611	185.096
(Bio,No)	1.052	1.170	1.317	1.505	1.762	2.141	5.300	173.534
(Seq,No)	1.000	1.112	1.252	1.430	1.674	***2.029***	5.024	164.653
(Gel,Gel)	1.110	1.234	1.389	1.588	1.859	2.259	5.407	89.309
(Gel,Bio)	1.110	1.234	1.389	1.588	1.860	2.258	5.417	94.262
(Bio,Gel)	1.052	1.170	1.317	1.505	1.761	2.140	5.127	84.528
(Bio,Bio)	1.052	1.170	1.317	1.505	1.761	2.140	5.131	89.507
(Seq,Gel)	1.000	1.112	1.252	1.430	1.674	2.032	4.884	***80.042***
(Seq,Bio)	1.000	1.112	1.252	1.430	1.674	2.033	***4.872***	85.217

For values of the probability of a non-conforming sample between zero and 0.4, the best strategies for the PCR iteration count are either the (No,Gel) or (No,Bio) strategies. When the probability of a non-conforming sample begins to increase to larger values, the ideal strategies shift to those utilizing sequencing at inspection station one. This is a significant divergence from minimizing for cost or time. When trying to limit the number of recursive processes at an upstream operation, sequencing at the first inspection opportunity is a good choice especially for high values of the probability of a non-conforming sample.

To illustrate the similarity between different inspection strategies for the probability of a non-conforming sample, the percent difference is shown in Table 18.

Table 18: Percentage Increase of PCR Iteration Count for an Inspection Strategy with Probabilities of a Non-conforming Sample for Fragment Assembly.

Inspection Strategy	Probability of a non-conforming unit							
	0	0.1	0.2	0.3	0.4	0.5	0.75	0.95
(No,No)	0.0003	0.0414	0.0915	0.123	0.123	0.355	3.867	124.965
(No,Gel)	-	0.0162	-	0.211	0.211	0.014	0.337	2.843
(No,Bio)	-	-	0.0016	-	-	0.054	0.042	7.345
(Gel,No)	11.0476	11.0626	11.0947	11.277	11.277	11.413	15.165	131.250
(Bio,No)	5.2413	5.3160	5.3804	5.501	5.501	5.556	8.798	116.805
(Seq,No)	0.0117	0.0729	0.1560	0.242	0.242	-	3.133	105.709
(Gel,Gel)	11.0469	11.0731	11.1118	11.275	11.275	11.354	10.995	11.579
(Gel,Bio)	11.0469	11.0806	11.1577	11.320	11.320	11.320	11.199	17.767
(Bio,Gel)	5.2406	5.3172	5.3839	5.495	5.495	5.477	5.246	5.605
(Bio,Bio)	5.2406	5.3196	5.4105	5.491	5.491	5.512	5.313	11.826
(Seq,Gel)	0.0113	0.0687	0.1912	0.212	0.212	0.149	0.259	-
(Seq,Bio)	0.0113	0.0726	0.1896	0.229	0.229	0.233	-	6.467

Inspection strategies utilizing no inspection at station one followed by either a gel electrophoresis or the Bioanalyzer at inspection station two, perform well. Additionally, the strategies utilizing sequencing at inspection station one also perform well. The low error rate of sequencing causes the system to lower the average number of PCR iteration by reducing redundant processing at downstream operations.

This chapter investigates the best inspection strategies in a three fragment Gibson assembly process. The logic, process parameters, assumptions of the discrete event simulation, and results are reported. Staying consistent with the fragment synthesis simulation, three primary objectives are investigated. These objectives include minimizing the total cost, total flow time, and PCR iteration count for the system. To understand the

stability of the process, the probability of a non-conforming sample is considered to quantify the relative maturity of the system. Using probability of a non-conforming sample coupled with one of the system objectives, a practitioner in an academic or industrial research environment can select an appropriate inspection strategy for their process.

7. Contributions and Future Work

This dissertation describes a new application of inspection allocation and discrete event simulation within genetic manufacturing systems utilizing gene synthesis. A mathematical formulation has been adapted from traditional manufacturing approaches and applied to the inspection allocation problem in genetic manufacturing systems. A discrete event simulation has been developed to investigate the preferred inspection strategy for a two-stage gene synthesis process with multiple inspection options and a three-fragment assembly process. These simulations account for many of the unique considerations found in genetic manufacturing systems utilizing gene synthesis and provide insight on the best strategies for minimizing cost or time.

7.1 Conclusions

Different conclusions are discovered using a mathematical formulation and two discrete event simulations to understand the inspection allocation decisions found in genetic manufacturing systems utilizing gene synthesis. The conclusions from these different aspects of the work are described in this section.

Through this dissertation, a mathematical formulation has been developed for the serial, multi-stage inspection allocation problem of a genetic manufacturing system with a final sequencing inspection check. Both cost and time have been used as the objectives for the mathematical formulation. The number of iterations at each of the synthesis operations is derived to calculate the total cost or time of the system. This is accomplished by treating the production of a conforming sample in a genetic manufacturing system as a Bernoulli

process and applying the expected value of the geometric distribution to estimate the number of trials of the system to achieve a success.

The mathematical formulation can capture many of the aspects of the inspection allocation problem for a genetic manufacturing system, but discrete event simulation is capable of modeling some of the more complicated logical pieces of genetic production. Discrete event simulations to understand the inspection allocation decisions for both fragment synthesis and assembly is employed in this dissertation. The conclusions for each of these simulations are presented in the following subsections.

7.1.1 *Conclusions of the Fragment Synthesis Discrete Event Simulation*

For the parameters considered in the discrete event simulation for inspection allocation in a GMS utilizing gene synthesis it is possible to generalize about the best strategies for different objectives. If the goal is to minimize total cost in the fragment synthesis two stage process, then using a gel electrophoresis at inspection station one and no inspection at station two is the best approach for a probability of nonconformance between 0.1-0.75. As the probability of a non-conforming sample increases beyond these values, the best approach becomes a gel electrophoresis at both inspection stations one and two.

When minimizing total flow time, the best strategy is the Bioanalyzer at inspection station one and no inspection at station two for values between 0.1 and 0.75 for the probability of a non-conforming sample. When the probability of a non-conforming sample increases to 0.95, the best strategy is using the Bioanalyzer followed by a gel electrophoresis. For both total cost and flow time, when the probability of a non-conforming unit is zero, the best strategy is using no inspection at both station one and two.

Conducting no inspection at both stations causes the cost and time of the system to rapidly grow as the probability of a non-conforming sample increases. Even if the best strategy is not employed, virtually any inspection strategy involving the use of gel electrophoresis or the Bioanalyzer at inspection station one is better than no inspection, except as the probability of a non-conforming sample approaches zero. Another global observation is using sequencing as an inspection strategy is not effective from a cost or time perspective. Sequencing is important as a final inspection to confirm the sequence is an identical match to the reference sequence. However, it should not be utilized as a mid-stream inspection technique for the PCR and cloning two stage process.

One important conclusion based on Table 5 relates to the inspection strategy used when trying to minimize cost when the probability of a non-conforming sample is between 0.1 and 0.75. Even though the (Gel,Gel) strategy has comparable performance, it realistically shouldn't be used. Since the (Gel,No) option is superior and requires one less inspection to be used, logically this is the strategy which should be chosen. This same principle applies in Table 7 when looking at the percent increase in the total flow time for different strategies. While the (Bio,Gel) and (Bio,Bio) strategies have similar performance to the best inspection strategy, they realistically shouldn't be used because they require an extra processing step. However, the (Gel,No) inspection strategy could be considered since it is only a small percent increase over the best strategy and does not require an extra processing step. Due to limitations of the simulation if additional sources of variation were included within the model, it would likely increase the performance gap between the best option and other competing options requiring a subsequent inspection process.

While the specific results of the simulation apply to the case of inspection stations after a PCR and cloning step, the approach can be more widely applicable. The simulation methodology can be expanded to more complex problems involving additional processes or inspection opportunities by adding serial processes to the model or multiple parallel processes in the case of genetic assembly operations. Given the flexibility of simulating the system, an academic or industrial practitioner could estimate their costs or flow times based on their specific processing parameters and routing logic. To illustrate the ease of incorporating additional complexities using discrete event simulation for GMS, the fragment assembly process is modeled and conclusions are presented in the next subsection.

7.1.2 *Conclusions of the Fragment Assembly Discrete Event Simulation*

Analysis of the three fragment Gibson assembly process through discrete-event simulation yields similar types of generalities for both the cost and time objectives. When the objective is to minimize cost, using a gel electrophoresis at inspection station one and no inspection at station two is the best strategy for the probability of a non-conforming sample between 0.1 and 0.75. The best strategy when the probability of non-conformance increases to 0.95 becomes using a gel electrophoresis at inspection stations one and two.

When time is the key objective, the best strategy is the Bioanalyzer at inspection station one followed by no inspection for probabilities of a non-conforming sample between 0.1 and 0.5. The Bioanalyzer is the best inspection strategy at both stations one and two when the probability of non-conformance is 0.75. Once the probability of non-conformance increases to 0.95, the (Bio,Gel) and (Bio,Bio) strategies have competing performance due

to the half-width of the strategies from the simulation. The best strategy cannot be discerned due to the size of the half width. For both cost and time, the best strategy for probability of non-conformance of zero is using no inspection at either station one or two.

The detrimental impact of sequencing when used as anything other than a final sequence verification step is also confirmed in the fragment assembly simulations. The performance for both cost and time is consistently inferior to most other inspection options available. Additionally, utilizing no inspection at station one also have significantly poor cost or time performance especially for high values for the probability of a non-conforming sample. Lastly, using no inspection at stations one and two is a severe detriment to the system except under near perfect processing conditions.

Many of the conclusions from the fragment synthesis and fragment assembly simulations are similar in nature. Even for a more complex system combining multiple fragments, the best strategies are still those which provide ideal performance for the fragment synthesis sub process. Due to the assumptions of the fragment assembly simulation, the fragment synthesis processes follow the same inspection strategy. This is consistent with how the laboratory would operate as there would be no need for differing inspection strategies for similar sized fragments. However, it would be interesting to investigate the effect of multiple different sized fragments being combined through a fragment assembly process. Would there be differences across the fragments if the protocols for separate fragment synthesis processes were drastically different? Additionally, how would the upstream inspection allocation decisions change if multiple subassemblies were required to create a much larger fragment final assembly? These types

of questions would likely have a significant impact on the inspection strategies chosen at the fragment synthesis phase of the process.

7.2 Future Work

Genetic manufacturing systems offer a rich are for future research across a variety of issues. The following sections highlight four potential areas for future work.

7.2.1 *Resource Allocation in Genetic Manufacturing Systems*

Process routing and resource considerations are complex issues for GMS. Many research laboratories have one technician conducting all the work for a given project, completing all the processing operations and steps to achieve a desired final sequence over the course of days, weeks, or even months. Typically, a technician may also be working on multiple projects, leading to the question of how resources and equipment should be allocated in the laboratory.

Should a single worker be assigned to a project or should that worker oversee specific process operations and steps? Is it possible to increase equipment utilization by combining batches from two different technicians? In addition, how do these considerations change when looking at more automated production processes within the industrial biotechnology sector? How is process routing and resource allocation affected by ramping up the volume of production in the industrial biotechnology sector? These are all questions that have been addressed for many traditional manufacturing systems, but have yet to be investigated for the creation of genetic products. Therefore, a future extension would be capturing

additional resources within the mathematical formulation and simulation to investigate their impact on both the cost and time related to the creation of a custom DNA sequence.

7.2.2 *Rework in Genetic Manufacturing Systems*

Traditional manufacturing systems rely on downgrading and rework to retain value for products that do not meet a certain quality threshold. Rework is technically possible for gene synthesis as individual base pairs can be changed following the results of certain inspection steps or when dealing with imperfect initial raw materials [88, 89]. Investigating the optimal rework threshold for DNA fragments is a question which has not been answered in the literature. How many erroneous base pairs are too much to warrant the cost of conducting the rework operation? Additionally, an evaluation of the benefits of conducting rework on a fragment versus starting at the beginning of the processing sequence is another potential research direction that could be analyzed through mathematical modelling and simulation. Determining the answers to these types of questions would provide practitioners with a guideline of when to conduct rework based on the cost or time considerations of their processing requirements.

7.2.3 *Standardization of Resources in Genetic Manufacturing Systems*

One method to mitigate the complexity of intricate and iterative processing in GMSs, is through standardization. For GMS, there are opportunities for standardization in the way protocols are described and how processes are depicted for training purposes.

Protocols are used to describe the sequence of processing steps required to achieve a desired chemical reaction for a certain type of DNA sequence. Many of the protocols use

similar equipment or techniques which are consistent across molecular and synthetic biology. Unfortunately, the protocols are typically described in numbered lists or long text descriptions for the reagents, preparation, and methods of achieving the desired results. This format leads to many opportunities in which a key step or piece of information could be missed and the entire process failing. Therefore, a standard presentation of protocols would allow other laboratories and research groups to replicate results. One potential way to standardize protocols is to create a common language and formatting like computer coding. This approach could benefit gene synthesis due to its iterative nature and many of the steps being common across protocols from multiple disciplines. For biological protocols in general, Biocoder is one example language to express protocol steps in a standardized fashion [90]. This is a great first step and can be a framework for translating many of the gene synthesis protocols into a clear, concise, and sharable format.

Another interesting opportunity is describing genetic processing as a production process. This is advantageous for training purposes and may reduce the learning curve for biological processing. Skills can be taught, but depicting the overall process flow visually can be helpful to those first starting out in the field. Process mapping, including value stream mapping, can describe both the process and information flows in a production system [65, 91]. These techniques allow for the decisions which must be made during the process flow to be shown and could be useful in defining the most efficient method of conducting a protocol.

With the speed at which gene synthesis research is progressing, communication between groups is pivotal to the continued progress in the field. Both standardization of

protocol formatting and depicting gene synthesis processing as a production system are ways to improve clarity and communication across the field.

7.2.4 *Cyber-physical Security in Genetic Manufacturing Systems*

In 2014, the manufacturing sector was the most targeted industry for spear-phishing based cyber-attacks which is a notable increase from previous years [92, 93]. Attacks on the physical infrastructure of a system, such as the Stuxnet virus in 2010 [94], present a whole new range of problems for manufacturing systems [95-97]. The issue of securing DNA synthesis processes has already been raised within the molecular biology community [98, 99]. Given the differences of GMS, investigations are important to determining how the attack vectors differ in these processes from traditional manufacturing. Appropriate mechanisms to secure these attack vectors may differ from those used in critical infrastructure security due to the complex processing and random nature of DNA fabrication.

7.3 Contributions and Final Thoughts

Four specific research contributions have emerged from this dissertation research. First, a class of manufacturing production environments, genetic manufacturing systems, has been identified. These types of systems are defined and the key differences between traditional and genetic manufacturing systems are provided. Second, this is the first known effort to map a gene synthesis process flow considering both process and routing decisions. Specifically, the high-level processes have been mapped into the fragment synthesis and assembly phases. Third, the inspection allocation problem for genetic manufacturing

systems utilizing gene synthesis has been formalized mathematically. This problem is modeled to minimize the total cost or total flow time of a serial, multi-stage genetic manufacturing system. Finally, two discrete event simulation models are developed to investigate inspection allocation decisions in an academic or industrial research environments. These simulation models are based upon a serial, two-stage fragment synthesis process and a three-fragment assembly process.

Genetic manufacturing systems aim at improving and understanding processes within gene synthesis as well as in synthetic and molecular biology with techniques found in traditional manufacturing systems. With the immense effort dedicated to research on synthetic genetic material, opportunities reducing waste, cost, and time have emerged. The emphasis on process improvement for synthetic and molecular biology is in its infancy. Within the academic and industrial research environments, research funding is often available for conducting novel research where the emphasis is more on new products and less on the improvements to bolster efficiency or performance. A shift in focus towards process improvement has recently been highlighted in the pharmaceutical research domain. In a publication about the inverse of Moore's law being observed in pharmaceutical research, the authors highlight the need for more statistical design and control of experiments to encourage reproducible results [100].

Moving forward, the ideas presented in this dissertation motivate the importance of modeling and analyzing complex and random genetic manufacturing systems to improve performance. Inspection allocation and more broadly resource allocation are just one area of this interdisciplinary field. The opportunities are rich if the manufacturing and biologic communities can collaborate to develop new models and understanding of how the

manufacture of genetic components can be improved through the systematic analysis of genetic manufacturing systems.

ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation (Grant Number 1241328).

The author would like to thank the members of his committee for their time and assistance with this project and the work presented in this dissertation. Their guidance has been pivotal to the success of this project and helping to identify an interesting interdisciplinary field within the manufacturing domain.

The members of the Virginia Tech Center for Innovation based Manufacturing (CIbM) laboratory have provided a multitude of assistance and feedback on the papers generated through this work as well as the dissertation itself. The author is grateful for their help throughout this process.

Dr. Lee Wells at Western Michigan University has provided valuable guidance and feedback for the author throughout the course of this project.

The author would like to thank Samantha Roscher for her assistance with several of the figures and tables which appear in this work.

The author would also like to thank his family for their support during this process. Specifically, the author would like to thank James and Linda Purdy, for their encouragement, support, and tireless love through this process. The author would like to thank Anjelica Smith for her love, guidance, and support through this process. Finally, the author would like to thank Scott Rieders for his friendship, support, and love through this process.

REFERENCES

[1] J. R. Coleman, D. Papamichail, S. Skiena, B. Futcher, E. Wimmer, and S. Mueller, "Virus attenuation by genome-scale changes in codon pair bias," *Science,* vol. 320, pp. 1784-1787, 2008.

[2] S. Mueller, J. R. Coleman, D. Papamichail, C. B. Ward, A. Nimnual, B. Futcher, *et al.*, "Live attenuated influenza virus vaccines by computer-aided rational design," *Nature Biotechnology,* vol. 28, pp. 1723-1729, 2010.

[3] S. Mueller, J. R. Coleman, and E. Wimmer, "Putting synthesis into biology: a viral view of genetic engineering through de novo gene and genome synthesis," *Chemistry & Biology,* vol. 16, pp. 337-347, 2009.

[4] P. R. Dormitzer, P. Suphaphiphat, D. G. Gibson, D. E. Wentworth, T. B. Stockwell, M. A. Algire, *et al.*, "Synthetic generation of influenza vaccine viruses for rapid response to pandemics," *Science Translational Medicine,* vol. 5, 2013.

[5] T. M. Tumpey, C. F. Basler, P. V. Aguilar, H. Zeng, A. Solorzano, D. E. Swayne, *et al.*, "Characterization of the reconstructed 1918 Spanish influenza pandemic virus," *Science,* vol. 310, pp. 77-80, 2005.

[6] J. K. Taubenberger, A. H. Reid, R. M. Lourens, R. Wang, G. Jin, and T. G. Fanning, "Characterization of the 1918 influenza virus polymerase genes," *Nature,* vol. 437, pp. 889-93, 2005.

[7] L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Rothlisberger, A. Zanghellini, *et al.*, "De novo computational design of retro-aldol enzymes," *Science,* vol. 319, pp. 1387-91, 2008.

[8] J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St Clair, *et al.*, "Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction," *Science,* vol. 329, pp. 309-13, 2010.

[9] H. M. Salis, E. A. Mirsky, and C. A. Voigt, "Automated design of synthetic ribosome binding sites to control protein expression," *Nature Biotechnology*, vol. 27, pp. 946-50, 2009.

[10] J. B. Plotkin and G. Kudla, "Synonymous but not the same: the causes and consequences of codon bias," *Nature Reviews Genetics,* vol. 12, pp. 32-42, 2011.

[11] D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R. Y. Chuang, M. A. Algire, *et al.*, "Creation of a bacterial cell controlled by a chemically synthesized genome," *Science,* vol. 329, pp. 52-6, 2010.

[12] N. Annaluru, H. Muller, L. A. Mitchell, S. Ramalingam, G. Stracquadanio, S. M. Richardson, *et al.*, "Total synthesis of a functional designer eukaryotic chromosome," *Science,* vol. 344, pp. 55-8, 2014.

[13] J. Phillips, "Blooming efficiency: Healthcare reform and agricultural demand will drive growth," IBISWorld November 2014.

[14] A. Blatecky, M. Dilworth, K. Erb, J. Ferrini-Mundy, W. E. Ward, M. Gutmann, *et al.*, "Dear Colleague Letter - CREATIV: Creative Research Awards for Transformative Interdisciplinary Ventures," INSPIRE, Ed., ed: National Science Foundation, 2011.

[15] J. Peccoud, K. Ellis, and J. Camelio, "INSPIRE: Modeling and optimization of DNA manufacturing processes," ed. Virginia Polytechnic Institute and State University: National Science Foundation, 2012.

[16] M. J. Czar, J. C. Anderson, J. S. Bader, and J. Peccoud, "Gene synthesis demystified," *Trends in Biotechnology,* vol. 27, pp. 63-72, 2009.

[17] R. A. Hughes, A. E. Miklos, and A. D. Ellington, "Chapter twelve - gene synthesis: methods and applications," in *Methods in Enzymology*. vol. 498, V. Christopher, Ed., ed: Academic Press, 2011, pp. 277-309.

[18] J. Peccoud, *Gene synthesis: Methods and Protocols*. New York: Humana Press: Springer, 2012.

[19] H. G. Khorana, "Nucleic acid synthesis," *Pure and Applied Chemistry,* vol. 17, pp. 349-382, 1968.

[20] G. Scheuerbrandt, A. Duffield, and A. Nussbaum, "Stepwise synthesis of deoxyribo-oligonucleotides," *Biochemical and Biophysical Research Communications,* vol. 11, pp. 152-155, 1963.

[21] L. McBride and M. Caruthers, "An investigation of several deoxynucleoside phosphoramidites useful for synthesizing deoxyoligonucleotides," *Tetrahedron Letters,* vol. 24, pp. 245-248, 1983.

[22] R. B. Merrifield, "Solid phase peptide synthesis. I. The synthesis of a tetrapeptide," *Journal of the American Chemical Society*, vol. 85, pp. 2149-2154, 1963.

[23] J. S. Dymond, L. Z. Scheifele, S. Richardson, P. Lee, S. Chandrasegaran, J. S. Bader, *et al.*, "Teaching synthetic biology, bioinformatics and engineering to undergraduates: the interdisciplinary Build-a-Genome course," *Genetics,* vol. 181, pp. 13-21, 2009.

[24] W. P. Stemmer, A. Crameri, K. D. Ha, T. M. Brennan, and H. L. Heyneker, "Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides," *Gene,* vol. 164, pp. 49-53, 1995.

[25] H. O. Smith, C. A. Hutchison, C. Pfannkoch, and J. C. Venter, "Generating a synthetic genome by whole genome assembly: φX174 bacteriophage from synthetic oligonucleotides," *Proceedings of the National Academy of Sciences,* vol. 100, pp. 15440-15445, 2003.

[26] J. Cello, A. V. Paul, and E. Wimmer, "Chemical synthesis of poliovirus cDNA: generation of infectious virus in the absence of natural template," *Science,* vol. 297, pp. 1016-1018, 2002.

[27] X. Gao, P. Yo, A. Keith, T. J. Ragan, and T. K. Harris, "Thermodynamically balanced inside-out (TBIO) PCR-based gene synthesis: a novel method of primer design for high-fidelity assembly of longer gene sequences," *Nucleic Acids Research,* vol. 31, 2003.

[28] A.-S. Xiong, R.-H. Peng, J. Zhuang, F. Gao, Y. Li, Z.-M. Cheng, *et al.*, "Chemical gene synthesis: strategies, softwares, error corrections, and applications," *FEMS Microbiology Reviews,* vol. 32, pp. 522-540, 2008.

[29] B. Alberts, *Molecular Biology of the Cell: Reference Edition*, 5 ed.: Garland Science, 2008.

[30] S. Metzenberg, *Working with DNA*: Taylor & Francis Group, 2007.

[31] D. S. T. Nicholl, *An Introduction to Genetic Engineering*: Cambridge University Press, 2008.

[32] A. Masotti and T. Preckel, "Analysis of small RNAs with the Agilent 2100 Bioanalyzer," *Nature Methods,* 2006.

[33] N. J. Panaro, P. K. Yuen, T. Sakazume, P. Fortina, L. J. Kricka, and P. Wilding, "Evaluation of DNA fragment sizing and quantification by the Agilent 2100 Bioanalyzer," *Clinical Chemistry,* vol. 46, pp. 1851-1853, 2000.

[34] F. Sanger and A. R. Coulson, "A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase," *Journal of Molecular Biology,* vol. 94, pp. 441-448, 1975.

[35] F. Sanger, S. Nicklen, and A. R. Coulson, "DNA sequencing with chain-terminating inhibitors," *Proceedings of the National Academy of Sciences,* vol. 74, pp. 5463-5467, 1977.

[36] L. M. Smith, J. Z. Sanders, R. J. Kaiser, P. Hughes, C. Dodd, C. R. Connell, *et al.*, "Fluorescence detection in automated DNA sequence analysis," *Nature,* vol. 321, pp. 674-679, 1986.

[37] M. L. Wilson, Y. Cai, R. Hanlon, S. Taylor, B. Chevreux, J. C. Setubal, *et al.*, "Sequence verification of synthetic DNA by assembly of sequencing reads," *Nucleic Acids Research,* vol. 41, 2013.

[38] D. A. Wheeler, M. Srinivasan, M. Egholm, Y. Shen, L. Chen, A. McGuire, *et al.*, "The complete genome of an individual by massively parallel DNA sequencing," *Nature,* vol. 452, pp. 872-876, 2008.

[39] R. Carlson, "Time for New DNA Synthesis and Sequencing Cost Curves," in *Synthesis*, ed, 2014.

[40] E. R. Mardis, "Anticipating the $1,000 genome," *Genome Biology,* vol. 7, p. 112, 2006.

[41] S. C. Schuster, "Next-generation sequencing transforms today's biology," *Nature Methods,* vol. 5, pp. 16-18, 2008.

[42] J. Shendure and H. Ji, "Next-generation DNA sequencing," *Nature Biotechnology,* vol. 26, pp. 1135-1145, 2008.

[43] T. Tucker, M. Marra, and J. M. Friedman, "Massively Parallel Sequencing: The Next Big Thing in Genetic Medicine," *American Journal of Human Genetics,* vol. 85, pp. 142-154, 2009.

[44] L. Liu, Y. Li, S. Li, N. Hu, Y. He, R. Pong, *et al.*, "Comparison of next-generation sequencing systems," *Journal of Biomedicine and Biotechnology,* vol. 12, 2012.

[45] M. L. Metzker, "Sequencing technologies - the next generation," *Nature Review Genetics,* vol. 11, pp. 31-46, 2010.

[46] R. L. Dorit, O. Ohara, C. B. C. Hwang, J. B. Kim, and S. Blackshaw, "Direct DNA Sequencing of PCR Products," in *Current Protocols in Molecular Biology*, ed: John Wiley & Sons, Inc., 2001.

[47] K. Cha-aim, T. Fukunaga, H. Hoshida, and R. Akada, "Reliable fusion PCR mediated by GC-rich overlap sequences," *Gene,* vol. 434, pp. 43-49, 2009.

[48] M. Z. Li and S. J. Elledge, "Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC," *Nature methods,* vol. 4, pp. 251-256, 2007.

[49] J. Quan and J. Tian, "Circular Polymerase Extension Cloning of Complex Gene Libraries and Pathways," *PLoS ONE,* vol. 4, 2009.

[50] Y. Sasaki, T. Sone, S. Yoshida, K. Yahata, J. Hotta, J. D. Chesnut, *et al.*, "Evidence for high specificity and efficiency of multiple recombination signals in mixed DNA cloning by the Multisite Gateway system," *Journal of Biotechnology*, vol. 107, pp. 233-243, 2004.

[51] H. H. Nour-Eldin, F. Geu-Flores, and B. A. Halkier, "USER cloning and USER fusion: the ideal cloning techniques for small and big laboratories," *Plant Secondary Metabolism Engineering: Methods and Applications*, pp. 185-200, 2010.

[52] D. G. Gibson, G. A. Benders, C. Andrews-Pfannkoch, E. A. Denisova, H. Baden-Tillson, J. Zaveri, *et al.*, "Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome," *Science*, vol. 319, pp. 1215-20, 2008.

[53] D. G. Gibson, G. A. Benders, K. C. Axelrod, J. Zaveri, M. A. Algire, M. Moodie, *et al.*, "One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome," *Proceedings of the National Academy of Sciences*, vol. 105, pp. 20404-20409, 2008.

[54] D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison, and H. O. Smith, "Enzymatic assembly of DNA molecules up to several hundred kilobases," *Nature Methods*, vol. 6, pp. 343-345, 2009.

[55] T. Knight, "Idempotent vector design for standard assembly of biobricks," Massachusetts Institute of Technology 2003.

[56] Y. Cai, B. Hartnett, C. Gustafsson, and J. Peccoud, "A syntactic model to design and verify synthetic genetic constructs derived from standard biological parts," *Bioinformatics*, vol. 23, pp. 2760-2767, 2007.

[57] J. Peccoud, M. F. Blauvelt, Y. Cai, K. L. Cooper, O. Crasta, E. C. DeLalla, *et al.*, "Targeted Development of Registries of Biological Parts," *PLoS ONE,* vol. 3, 2008.

[58] O. Ho-Shing, K. H. Lau, W. Vernon, T. T. Eckdahl, and A. M. Campbell, "Assembly of standardized DNA parts using BioBrick ends in E. coli," *Gene Synthesis: Methods and Protocols,* pp. 61-76, 2012.

[59] S. C. Sleight, B. A. Bartley, J. A. Lieviant, and H. M. Sauro, "In-Fusion BioBrick assembly and re-engineering," *Nucleic Acids Research,* 2010.

[60] G. Røkke, E. Korvald, J. Pahr, O. Øyås, and R. Lale, "BioBrick Assembly Standards and Techniques and Associated Software Tools," in *DNA Cloning and Assembly Methods,* S. Valla and R. Lale, Eds., ed Totowa, NJ: Humana Press, 2014, pp. 1-24.

[61] J. Anderson, J. E. Dueber, M. Leguia, G. C. Wu, J. A. Goler, A. P. Arkin, *et al.*, "BglBricks: A flexible standard for biological part assembly," *Journal of Biological Engineering,* vol. 4, p. 1, 2010.

[62] W. J. Blake, B. A. Chapman, A. Zindal, M. E. Lee, S. M. Lippow, and B. M. Baynes, "Pairwise selection assembly for sequence-independent construction of long-length DNA," *Nucleic Acids Research,* vol. 38, pp. 2594-2602, 2010.

[63] C. Engler, R. Kandzia, and S. Marillonnet, "A One Pot, One Step, Precision Cloning Method with High Throughput Capability," *PLoS ONE,* vol. 3, 2008.

[64] R. Carlson, "The changing economics of DNA synthesis," *Nature Biotechnology,* vol. 27, pp. 1091-1094, 2009.

[65] M. Braglia, G. Carmignani, and F. Zammori, "A new value stream mapping approach for complex production systems," *International Journal of Production Research,* vol. 44, pp. 3929-3952, 2006.

[66] D. S. Bai and H. J. Yun, "Optimal allocation of inspection effort in a serial multi-stage production system," *Computers & Industrial Engineering,* vol. 30, pp. 387-396, 1996.

[67] D. P. Ballou and H. L. Pazer, "Process improvement versus enhanced inspection in optimized systems," *International Journal of Production Research,* vol. 23, p. 1233, 1985.

[68] D. P. Ballou and H. L. Pazer, "The impact of inspector fallibility on the inspection policy in serial production systems," *Management Science,* vol. 28, pp. 387-399, 1982.

[69] G. D. Eppen and E. G. Hurst, "Optimal location of inspection stations in a multistage production process," *Management Science,* vol. 20, pp. 1194-1200, 1974.

[70] G. Galante and G. Passannanti, "Integrated approach to part scheduling and inspection policies for a job shop manufacturing system," *International Journal of Production Research,* vol. 45, pp. 5177-5198, 2007.

[71] A. Heredia-Langner, D. C. Montgomery, and W. M. Carlyle, "Solving a multistage partial inspection problem using genetic algorithms," *International Journal of Production Research,* vol. 40, pp. 1923-1940, 2002.

[72] J. Lee and S. Unnikrishnan, "Planning quality inspection operations in multistage manufacturing systems with inspection errors," *International Journal of Production Research,* vol. 36, pp. 141-156, 1998.

[73] G. F. Lindsay and A. B. Bishop, "Allocation of screening inspection effort: a dynamic-programming approach," *Management Science,* vol. 10, pp. 342-352, 1964.

[74] P. M. Pruzan and J. R. Jackson, "A dynamic programming application in production line inspection," *Technometrics,* vol. 9, pp. 73-81, 1967.

[75] M. Raghavachari and G. K. Tayi, "Inspection configuration and reprocessing decisions in serial production systems," *International Journal of Production Research,* vol. 29, pp. 897-911, 1991.

[76] A. G. Shetwan, V. I. Vitanov, and B. Tjahjono, "Allocation of quality control stations in multistage manufacturing systems," *Computers & Industrial Engineering,* vol. 60, pp. 473-484, 2011.

[77] Y.-R. Shiau, M.-H. Lin, and W.-C. Chuang, "Concurrent process/inspection planning for a customized manufacturing system based on genetic algorithm," *The International Journal of Advanced Manufacturing Technology,* vol. 33, pp. 746-755, 2007.

[78] S. Van Volsem, W. Dullaert, and H. Van Landeghem, "An evolutionary algorithm and discrete event simulation for optimizing inspection strategies for multi-stage processes," *European Journal of Operational Research,* vol. 179, pp. 621-633, 2007.

[79] L. S. White, "The analysis of a simple class of multistage inspection plans," *Management Science,* vol. 12, pp. 685-693, 1966.

[80] L. S. White, "Shortest route models for the allocation of inspection effort on a production line," *Management Science,* vol. 15, pp. 249-259, 1969.

[81] B. J. Yum and E. D. McDowellj, "Optimal inspection policies in a serial production system including scrap rework and repair: an MILP approach," *International Journal of Production Research,* vol. 25, pp. 1451-1464, 1987.

[82] G. T. Purdy, J. A. Camelio, K. P. Ellis, N. R. Adames, and J. Peccoud, "Opportunities to apply manufacturing systems analysis techniques in genetic manufacturing systems," *Manufacturing Letters,* Submitted and under review.

[83] C. D. Pegden, "SIMIO: a new simulation system based on intelligent objects," in *Proceedings of the 39th Conference on Winter Simulation*, Washington D.C., 2007, pp. 2293-2300.

[84] A. M. Bland, M. G. Janech, J. S. Almeida, and J. M. Arthur, "Sources of variability among replicate samples separated by two-dimensional gel electrophoresis," *Journal of Biomolecular Techniques,* vol. 21, pp. 3-8, 2010.

[85] F. Jian, B. Yan-hong, W. Yun-long, and J. Jian-zhou, "Detection and identification of enterohemorrhagic Escherichia coli O157: H7 using Agilent 2100 Bioanalyzer," in *4th International Conference on Bioinformatics and Biomedical Engineering (iCBBE)*, 2010, pp. 1-3.

[86] D. Gibson. (2016). *Gibson Assembly Protocol (E5510)*. Available: https://www.neb.com/protocols/2012/12/11/gibson-assembly-protocol-e5510

[87] D. Gibson. (2016). *Gibson Assembly® Cloning Kit*. Available: https://www.neb.com/products/e5510-gibson-assembly-cloning-kit

[88] G. Linshiz, T. B. Yehezkel, S. Kaplan, I. Gronau, S. Ravid, R. Adar, *et al.*, "Recursive construction of perfect DNA molecules from imperfect oligonucleotides," *Molecular Systems Biology,* vol. 4, pp. 191-191, 2008.

[89] S. Ma, I. Saaem, and J. Tian, "Error Correction in Gene Synthesis Technology," *Trends in Biotechnology,* vol. 30, pp. 147-154, 2012.

[90] V. Ananthanarayanan and W. Thies, "Biocoder: A programming language for standardizing and automating biology protocols," *Journal of Biological Engineering,* vol. 4, pp. 1-13, 2010.

[91] M. Rother and J. Shook, *Learning to See: Value Stream Mapping to Add Value and Eliminate Muda*: Lean Enterprise Institute, 2003.

[92] Symantec. (2014). *Internet Security Threat Report 2014, Volume 19*. Available: www.symantec.com/content/en/us/enterprise/other_resources/b-istr_main_report_v19_21291018.en-us.pdf

[93] Symantec. (2015). *Internet Security Threat Report 2015, Volume 20*. Available: https://www4.symantec.com/mktginfo/whitepaper/ISTR/21347932_GA-internet-security-threat-report-volume-20-2015-social_v2.pdf

[94] D. Albright, P. Brannan, and C. Walrond, "Did Stuxnet Take Out 1,000 Centrifuges at the Natanz Enrichment Plant?," ed: Institute for Science and International Security, 2010.

[95] H. Turner, J. White, J. A. Camelio, C. Williams, B. Amos, and R. Parker, "Bad Parts: Are Our Manufacturing Systems at Risk of Silent Cyberattacks?," *Security & Privacy, IEEE,* vol. 13, pp. 40-47, 2015.

[96] H. Vincent, L. Wells, P. Tarazaga, and J. Camelio, "Trojan Detection and Side-channel Analyses for Cyber-security in Cyber-physical Manufacturing Systems," *Procedia Manufacturing,* vol. 1, pp. 77-85, 2015.

[97] L. J. Wells, J. A. Camelio, C. B. Williams, and J. White, "Cyber-physical security challenges in manufacturing systems," *Manufacturing Letters,* vol. 2, pp. 74-77, 2014.

[98] H. Bugl, J. P. Danner, R. J. Molinari, J. T. Mulligan, H.-O. Park, B. Reichert, *et al.*, "DNA synthesis and biological security," *Nature Biotechnology,* vol. 25, pp. 627-629, 2007.

[99] A. Nouri and C. Chyba, "DNA Synthesis Security," in *Gene Synthesis*. vol. 852, J. Peccoud, Ed., ed: Humana Press, 2012, pp. 285-296.

[100] M. I. Sadowski, C. Grant, and T. S. Fell, "Harnessing QbD, Programming Languages, and Automation for Reproducible Biology," *Trends in Biotechnology*, vol. 34, pp. 214-227, 2016.