

## ABSTRACT

Stamper, Brian J. M.S. Biological Engineering, Purdue University, May, 2007. Statistically Designed Calibration For Near Infrared Spectrometric Monitoring of Mammalian Cell Culture. Major Professors: Nathan Mosier, Bernard Tao.

Near-infrared spectroscopy has been shown to effectively monitor mammalian cell cultures. However, interactions of metabolites and the effect of the background matrix have proven problematic in that they overly complicate the calibration model or lead to covariance that reduces the accuracy and precision of the model. To resolve this, a statistically designed calibration set was created that accounts for the metabolite interactions and the effect of the changing background matrix over the length of the cell culture. A three-factor response surface experimental design was executed to quantify the concentrations of glucose and glutamate while utilizing the culture age as the third factor. The resulting spectral data was pre-treated by taking the natural log of the raw data followed by mean-centering the data. This led to a model with a standard error of prediction of 2.26 mM for glucose (range tested was 30 – 65 mM) and 0.79 mM for glutamate (range tested was 6 – 13 mM). The  $R^2$  correlations of actual to predicted values for these two metabolites were 0.97 for glucose and 0.46 for glutamate. The model's prediction ability for glucose is sufficient such that NIRS may be used as the primary glucose measurement method for mammalian cell culture, but the sensitivity of the instrument limited the ability to predict glutamate such that the prediction ability is not acceptable.

## CHAPTER 1. INTRODUCTION

### 1.1. Objectives

The use of near infrared spectroscopy (NIRS) has been shown to be an effective tool for online monitoring of mammalian cell culture such that it has allowed for improved process monitoring and control. However, interactions of metabolites within the cell culture and a changing background matrix have proven problematic in the past in that they complicate the NIRS calibration models or lead to unacceptable prediction error by introducing covariance among the spectra. This researcher proposes the use of a response surface experimental design to create a Partial Least Squares (PLS) calibration set that will accurately model the concentrations of glucose and glutamate over the length of a cell culture. The experimental design will include the metabolites of interest and the effect of culture age, such that the error associated with component interactions and background matrix changes over time can be reduced.

The approach required to apply the PLS regression algorithm to NIR spectral data are shown below. The PLS algorithm consists of two sections: (1) calibration and (2) prediction. The calibration section consists of the steps listed below.

- 1. Design the calibration set of data**
- 2. Collect the spectra for each sample**
- 3. Pre-treat the spectral data to reduce bias**
- 4. Form the loading vectors and scores**
- 5. Use cross-validation to select the optimal number of loading vectors**
- 6. Determine the model coefficients**

- 7. Test the calibration model with other data**
- 8. Reiterate steps 5-7 as needed to reach optimal model**

Once the calibration is complete, the resulting model(s) can be used to predict the concentrations of the desired components based on the spectral data.

Based upon the results of this analysis, this new type of calibration model should allow for more accurate modeling over the length of a cell culture. This improved accuracy can lead to increased process understanding and more effective process control which can lead to higher productivity and improved product quality.

## 1.2. Organization

This thesis covers background information on mammalian cell culture and its monitoring methods, background information on near infrared spectroscopy, the theory defining near infrared spectroscopy, the derivation of the partial least squares method, and the detailed description of the experiments conducted. In total, the thesis has 12 chapters with 5 appendices.

## CHAPTER 2. INDUSTRY BACKGROUND AND PROJECT RELEVANCE

### 2.1. Biotechnology Industry Background

Historically, the pharmaceutical industry has been able to meet the medicinal needs of the population through the use of naturally-occurring chemical and biological compounds. Within the past few decades, however, other avenues for pharmaceutical innovation have been explored that have yielded novel targeted therapeutic molecules [1]. One such avenue was enabled by the technological advances in molecular and cellular biology that has allowed for manipulation of naturally-occurring chemicals and molecules such that novel biomolecules could be created, thus establishing modern biotechnology. These biomolecules have demonstrated lower toxicity and higher specificity as therapeutics, and have become the latest wave of medicines [2].

The term, “biotechnology,” was created by the Hungarian engineer Karl Ereky in 1919. He defined the term as “all lines of work by which products are produced from raw materials with the aid of living organisms” [1]. However, biotechnology had been in use long before 1919. The following list describes a few biology and biotechnology milestones of importance [3]:

- 4000 BC Classical biotechnology: Dairy farming develops in the Middle East; Egyptians use yeasts to bake leavened bread and to make wine.
- 1861 French chemist Louis Pasteur develops pasteurization – preserving food by heating it to destroy harmful microbes.
- 1973 Breakthrough discovery of recombinant DNA became the platform for research in cloning, genomics, and proteomics.

- 1982 First genetically engineered product – human insulin produced by Eli Lilly and Company using *E. coli* bacteria – is approved for use by diabetics.

Biotechnology was initially applied to microbial cells such as *E. coli* to create the earliest recombinant biomolecules. The microbial cells could then synthesize the desired biomolecule despite the fact that the wild-type microbial cells did not synthesize that particular molecule. Through several years of research and experimentation, scientists and engineers determined methods to purify and formulate the synthesized biomolecules to be suitable for use as medicines. This new method to generate recombinant biomolecules in microbial cells was introduced in 1970, and the first industrial scale production of a recombinant molecule was achieved in 1979 [4].

As biotechnology developed in microbial systems, it became apparent that the types of recombinant biomolecules that could be synthesized were limited due to the potential of toxic and immunogenic effects on the patient. Microbial cells lack the ability to perform post-translational modifications of the biomolecules (i.e. glycosylation) which can affect the efficacy, immunogenicity, or half-life of the final biological drug [5],[6]. As drug discovery broadened and new types of biomolecules were considered, scientists determined that some medicines required biomolecules that were more complex than those that could be synthesized by microbial cells [7]. The use of mammalian cells was explored for their ability to post-translationally modify the biosynthetic molecule to allow for the increased efficacy and stability.

Mammalian cell culture has become widely used within the pharmaceutical and biotechnology industries for production of therapeutic proteins. As of 2005, monoclonal antibodies alone constituted > \$15 billion revenue per year, and some estimates predict that the market will increase to \$26-30 billion per year by

2010 [2],[8]. This rapidly increasing market is an indicator of the great value of mammalian cell culture to the industry, and more importantly to the industry's customers.

Due to the increased sensitivity of mammalian cells to their environment and the increased biochemical complexity compared to microbial cells, more advanced process monitoring and control strategies needed to be developed to allow for more detailed knowledge of the cellular activity.

The current state of bioprocess monitoring technology enables monitoring basic environmental conditions within the cell cultures such as pH, temperature, and dissolved oxygen, and continues to expand to include monitoring and control of other parameters. Despite these technological advances, limitations still exist for process development and manufacturing such that unacceptably large variability of the process and final product quantity and quality still occurs.

## 2.2. Project Relevance

In a highly regulated field such as pharmaceuticals, the ability to define and control the processes used to produce the drugs is essential to enable increased productivity and reduce process excursions that lead to loss of product or unacceptable product quality [9-11]. Tools must be implemented to provide this ability. As the products and processes become more complex, technologies must be investigated that can provide even greater insight into the drug production process. This insight can allow companies to increase the consistency of the product and reduce undesired patient safety incidents.

In 2004, the U.S. Food and Drug Administration defined a quality initiative titled “Quality by Design.” This initiative signifies that drug manufacturers must “design and develop manufacturing processes during the product development stage to consistently ensure a predefined quality at the end of the manufacturing process. A quality system provides a sound framework for the transfer of process knowledge from development to the commercial manufacturing processes and for post development changes and optimization” [12]. As a result, a philosophy dubbed Process Analytical Technology (PAT) was implemented to achieve Quality by Design. The PAT strategy consists of implementing tools that provide on-line, real-time data that can be used to monitor and control processes to ensure greater product consistency and therefore patient safety. These tools may also allow for reduced product and process development times, thus enabling reduced time to market and increasing availability to patients who may benefit from the drug.

One such PAT tool is near infrared spectroscopy (NIRS). NIRS is able to non-invasively monitor bioprocesses in real-time with no sample preparation. This monitoring ability could be applied to mammalian cell culture to increase the ability to develop and control processes. Due to its ability to use fiber optic cables, multiple bioreactors can be monitored simultaneously by one analyzer. The challenges associated with NIRS are the complexity of the calibration model and statistical tools required to generate the models, but these can be overcome with appropriately designed calibration models and application of statistical methods [9-11, 13-16].

This thesis focuses on the application of near-infrared spectroscopy (NIRS) to improve the monitoring and control capabilities of mammalian cell culture processes that synthesize therapeutic proteins (including monoclonal antibodies) over those methods that have been used previously. Specifically, a new method to create the statistical model needed for NIRS implementation will be examined.

The statistical model is essential for NIRS use because it enables the user to predict concentrations of analytes based on spectral results. To date, use of NIRS in mammalian cell cultures has been limited due to the large number of analytes present in the cell culture medium that lead to increased calibration complexity and difficulty. NIRS may allow for more robust processes and for shorter (and cheaper) process development timelines.

## CHAPTER 3. MAMMALIAN CELL CULTURE OVERVIEW

### 3.1. Types of Mammalian Cells used in Industry

Several types of mammalian cell lines are in use today, both in industry and academia. These cell lines are stored in cell banks from which investigators can purchase the various cell lines. In the United States, the standard cell bank used by industry is the American Type Culture Collection (ATCC) located in Rockville, Maryland. ATCC's Cell Biology Collection consists of more than 3,400 cell lines from over 80 species. This includes over 950 cancer cell lines and 1000 hybridoma cell lines used for production of monoclonal antibodies [17].

Table 3-1 lists some of the most common cell types used in industry for production of therapeutic biomolecules. Each cell line has its benefits and disadvantages, such that the cell line must be carefully chosen to satisfy the demands of each project. One of the most common mammalian cell lines in use today for production of pharmaceuticals is the Chinese Hamster Ovary (CHO) cell line [18].

Table 3-1. Examples of continuous cell lines in regular use [7]

Name	Morphology	Origin	Characteristic
A9	Fibroblast	Mouse subcutaneous	HGPRT-; derivative of L929 1-negative
BHK21 C13	Fibroblast	Syrian hamster kidney	Transformable by polyoma virus
BRL3A	Epithelial	Rat liver	Produce multiplication stimulating activity
CHO-K1	Fibroblast	Chinese hamster ovary	Simple karyotype
EB-3	Lymphocytic	Human	EB virus positive
GH1, GH3	Epithelial	Rat	Produce growth hormone
HeLa	Epithelial	Human	Glucose-6-phosphate dehydrogenase Type A
L1210	Lymphocytic	Mouse	Rapidly growing suspension
L5178Y	Lymphocytic	Mouse	Rapidly growing suspension
L929	Fibroblast	Mouse	Clone of L-cell
LS	Fibroblast	Mouse	Grow in suspension; derivative of L929
MCF-7	Epithelial	Human breast pleural effusion	Oestrogen receptor positive
NRK49F	Fibroblast	Rat kidney	Induction of suspension growth by transforming growth factors
P388D1	Lymphocytic	Mouse	Grow in suspension
S180	Fibroblast	Mouse	Cancer chemotherapy screening
Vero	Fibroblast	Monkey kidney	Viral substrate and assay
3T3-L1	Fibroblast	Mouse (Swiss)	Adipose differentiation

3T3-A31	Fibroblast	Mouse (BALB/c)	Contact inhibited; readily transformed
---------	------------	----------------	--

The cell line used for the experimentation for this thesis is a variant of the CHO-K1 line called GS-CHO. This cell line was chosen due to its widespread use in industry. Its use is widespread due to its ability to consistently produce therapeutic proteins at high concentrations with the correct post-translational modifications and its increased resistance to shear compared to other mammalian lines.

### 3.2. Basic Mammalian Cell Culture Metabolism

The biochemistry and metabolism of all mammalian cell lines are essentially the same. Each cell line has its own specific characteristics, but they all follow the same basic biochemical pathways. The differences in cell line behavior arise from the presence or absence of catalytic enzymes and the resulting biochemical regulatory activities that are associated with those enzymes.

Two basic biochemical pathways form the core of mammalian cell culture metabolism: glycolysis and the tricarboxylic acid (TCA) cycle. Glycolysis converts glucose to pyruvate and a small amount of energy. From this cycle, pyruvate is converted to Acetyl-CoA and is then routed through the tricarboxylic acid cycle, which produces a large amount of energy Figure 3.1 [19].