In Silico Protein Optimization Laboratory for Bioprocess Engineers

Brian Aufderheide[1] and Jemila James Nunes[2]

[1] *Hampton University,* [2] *The University of Trinidad and Tobago*

Abstract

In the past ten years, there has been a growing movement for Chemical Engineering departments to include bioprocessing or bioengineering to their titles to cash into the hot field of biotechnology, specifically, the manufacture of heterologous proteins. However, few students will actually produce a therapeutic protein in a bioreactor or gain the tools to understand the complex interactions of not only cell growth and death but protein folding and protease activity. Authors have developed simulations in MATLAB that qualitatively match results found in the literature for different gene expression systems: yeast (*Pichia pastoris*), bacteria (*E. coli*), and mammalian (*Chinese hamster ovary*). Students are guided to change one engineering parameter at a time, then based on those results develop a heuristic optimization strategy to improve protein titers. Students have taken these as overnight labs and in some cases as in class exams. Complete labs with all Matlab programs will be provided along with some sample solutions at authors' website[1].

Keywords

Bioprocess Engineering Course, Heterologous Protein Production, Gene Expression Systems, bioreactor modeling and simulation

Introduction

The future of chemical engineering education looks bright since the inclusion of bioengineering courses in traditional chemical engineering programs is increasing. Yet, what makes the field exciting, the making of heterologous or foreign proteins in a microbe or mammalian cell that can be sold as a pharmaceutical or industrial enzyme, is limited to descriptions of the genetics and protein coding in the actual biology to standard mathematical models for bioreactor production. For most programs, the resources necessary for practical application are not available, limiting key elements necessary for deeper understanding of the field. Computer-based simulations for bioprocessing are therefore crucial for bridging the gap between bioprocess theory and its application and analysis. There are many tools available which are focused on design, process control, and web-based simulations, and modeling[2-5] but this program focuses on modeling different gene expression systems by modifying the engineering parameters substrate concentration, temperature, pH, and flow rate(s). The outputs are substrate concentration, cell concentrations (lyzed and total), protease and protein concentrations. The goal is to give students a hands-on feel for how each engineering parameter can affect the overall protein production in different ways, be it cell growth, cell death, protease activity, or protein folding. Each bioreactor run to produce proteins from bacteria, yeast or mammalian cells would require a total time of 3 to 10 days. Therefore to run various different operating conditions is very prohibitive both in time and expense.

© American Society for Engineering Education, 2019

The learning objectives are to understand how substrate concentrations, pH conditions, temperature of media, and harvest times can affect cell growth, cell death, protein folding, and protease activity that all contribute to the overall protein concentration found in the bioreactor. Data analysis tools to calculate doubling times, rate of protease activities, and yields for substrate to cell biomass or prote in to biomass are taught to the students to help discern what is actually taking place in the bioreactor. To better understand the complexity and how the models were developed, background theory on unstructured cell bioreactor modeling, Monod growth kinetics, and each design parameter is provided as well as some sample laboratory results from running the program.

Background Theory

1. Production of Heterologous Proteins

Heterologous proteins of biological interest are made by inserting a genetically modified plasmid into a host microorganism. Plasmids or cDNA are constructed by inserting the gene of interest, coding for a promotor which acts as a switch that prompts protein production, and inserting a selectable marker. The promotor is situated right before the target gene and determines when and where the gene will be expressed. Some promotors are 'always on' so that protein is made continuously while some are 'inducible', where protein is made only after an environmental cue is provided, with the latter type being preferred. In the methylotrophic yeast *Pichia pastoris*, the Alcohol Oxidase (AOX) promoter is present in two types with AOX1 being the major constituent that is induced by methanol which triggers protein production after the cells have grown on glycerol or glucose[6].

The primary sequence of the protein is transcribed into mRNA which contains codons of nucleotides that are then translated into the amino acid sequence of the protein. For active protein, the primary sequence must be folded correctly. This depends on environmental conditions to a large extent but chaperones are sometimes used to assist in achieving a stable Gibbs' free energy.

2. Kinetics of Growth and Unstructured Modeling

The general equations for a bioreactor that includes reactor volume, V, dry cell mass concentration, X, substrate concentration, S, and protein cell concentration, P, that includes death kinetics, k_d, are the following:

$$\frac{dV}{dt} = F_{in} - F_{out}$$

$$\frac{dX}{dt} = -\frac{F_{in}}{V}X + \mu X - k_d X$$

$$\frac{dS}{dt} = \frac{F_{in}}{V}(S_{in} - S) - Y_{S/X}\mu X$$

$$\frac{dP}{dt} = -\frac{F_{in}}{V}P + Y_{P/X}\mu X$$

© American Society for Engineering Education, 2019

where growth kinetics, μ (hr^{-1}), is typically one of the following three forms:

Monod: $\mu = \frac{\mu_{max}S}{K_s+S}$ Substrate Inhibition: $\mu = \frac{\mu_{max}S}{K_s+S+\frac{S^2}{K_i}}$ Product Inhibition: $\mu = \frac{\mu_{max}S}{K_s+S+K_iP}$

μ_{max} (hr^{-1}) is the actual maximum growth rate, and fitted parameter only for Substrate and Product Inhibition. K_s is the Monod constant (gL^{-1}). K_i is the inhibition constant (gL^{-1}). $Y_{S/X}$ and $Y_{P/X}$ are yield coefficients for substrate with respect to cell concentration and protein produced per cell concentration (gg $^{-1}$).

The difficulty with the modeling equations are that correlations of parameters affected by temperature and especially pH are not easily determined. Those parameters are from growth to death kinetics and bioprocessing yields, especially $Y_{P/X}$ which is rarely constant and is affected also by protease activity that degrades the protein. Note that the equations make it easy to go from batch to fed batch to continuous simply by putting in correct flow rates for F_{in} and F_{out}. For the authors' simulated protein production, journal articles that characterize the effects of temperature and pH are qualitatively done to mimic the published results since it is rare to find a model for their effects to simulate outright. Results shown here are for production of a protein using *Pichia pastoris* and are based on Zhang *et al.*, 2000[7], and Inan *et al.*, 1999[8].

3. Protease activity

Extracellular proteases are present in fermentations because of cell lysis. They consume protein which can be detrimental to overall productivity but they also consume other potentially inhibitory by-products of cell metabolism. The level of protease activity depends on environmental conditions such as pH and temperature. Cells have an optimum pH range for growth but so do proteases so operating at the cell optimum pH value might increase protease activity. Increasing temperature also increases the protease degradation rate. Operating in a protease free environment typically does not work because they are necessary for consuming certain toxins, which if left unchecked, can become inhibitory to cell metabolism. Use of protease inhibitors drives up downstream purification costs. Recent studies[9-11] have shown that the stability of heterologous proteins can be improved by deletion of protease genes. Ultimately, there is usually the need to strike a balance between conditions that optimize cell growth versus conditions that can minimize protease activity.

4. Effect of Environmental Parameters on Protein Production

Substrate concentration affects the number of cells produced - higher cell numbers means more protein potentially produced. An excess of substrate could result in substrate inhibition and affect the growth kinetic model while extreme substrate limitation leads to cell death and lysis, which increases proteases. The substrate, however, does not impact protein folding or the rate of gene expression.

Temperature affects cell growth since each microbe has an optimum temperature range that is preferable for optimum growth. An increased temperature increases the likelihood of protein mis-folding so lower temperatures are preferred to get active protein. Temperature increases protease activity.

© American Society for Engineering Education, 2019

pH also affects cell growth since each microbe prefers a specific range of pH values, however microbes can regulate intracellular pH, but this requires more maintenance energy. Since protein folding depends on formation of hydrogen bonds and other internal interactions, pH changes influence protein folding strongly since it can result in side group modification. Bacterial cells can withstand pH changes best, while yeast and mammalian cells are less robust under less than optimal conditions. Protease proteins can greatly be affected by the pH of the media.

Example of Protein Optimization Experimental Runs

Students are given the lab assignment and all of the associated Matlab files. The main running file is the initialization file and in the top user section, students can enter the conditions to run the

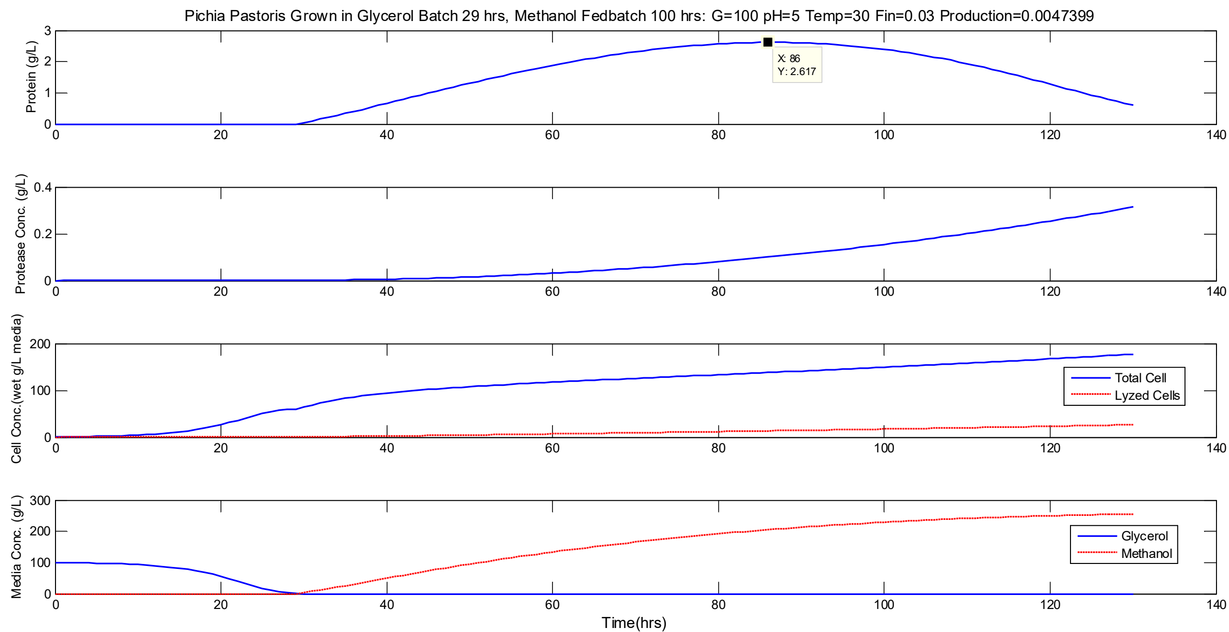

Fig 1. *Pichia pastoris* cells growing at initial conditions: G=100 g/L, pH = 5, Temp=30 ^{0}C, Fin=0.03, t_methanol=100 hours.

protein production at the user specified set of engineering parameters. In the case of *Pichia pastoris* protein production the parameters are: G, the initial glycerol concentration in seed batch phase (g/L), pH, Temp in centigrade, Fin, flow rate for methanol fed batch phase where protein is made (L/hr), and t_methanol, the harvest time (hrs). In Part I (75 points), the students typically modify one engineering parameter at a time from baseline conditions. The baseline experimental run under initial conditions is shown in Figure 1. The protein concentration shows an optimum value at roughly 86 hours. This optimum can only occur if proteases are present to degrade the protein that was secreted during the batch. The negative slope after the optimum gives an approximate idea on the amount of protease activity. The steeper the slope, the more active the proteases are and the more protein has been degraded. Students are instructed to calculate the protease degradation rate, the doubling time of the cell growth during glycerol batch phase, time taken for all glycerol to be consumed, yields of glycerol to cell concentration, final concentrations of protein, proteases, total cells, lyzed cells, substrate, and final protein production which factors in the time taken to produce the final protein concentration.

© American Society for Engineering Education, 2019

Figure 2 shows the same experimental run at initial conditions harvested at the optimal time.

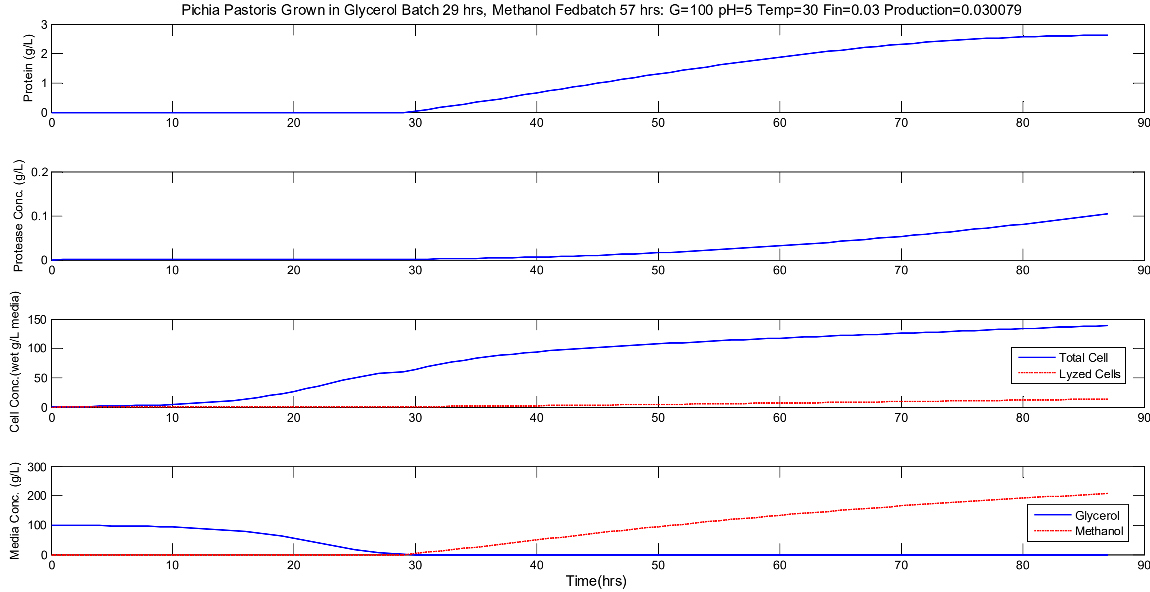

Fig 2. *Pichia pastoris* cells growing at initial conditions with reduced t_methanol: G=100 g/L, pH = 5, Temp=30 ^{0}C, Fin=0.03, t_methanol=57 hours

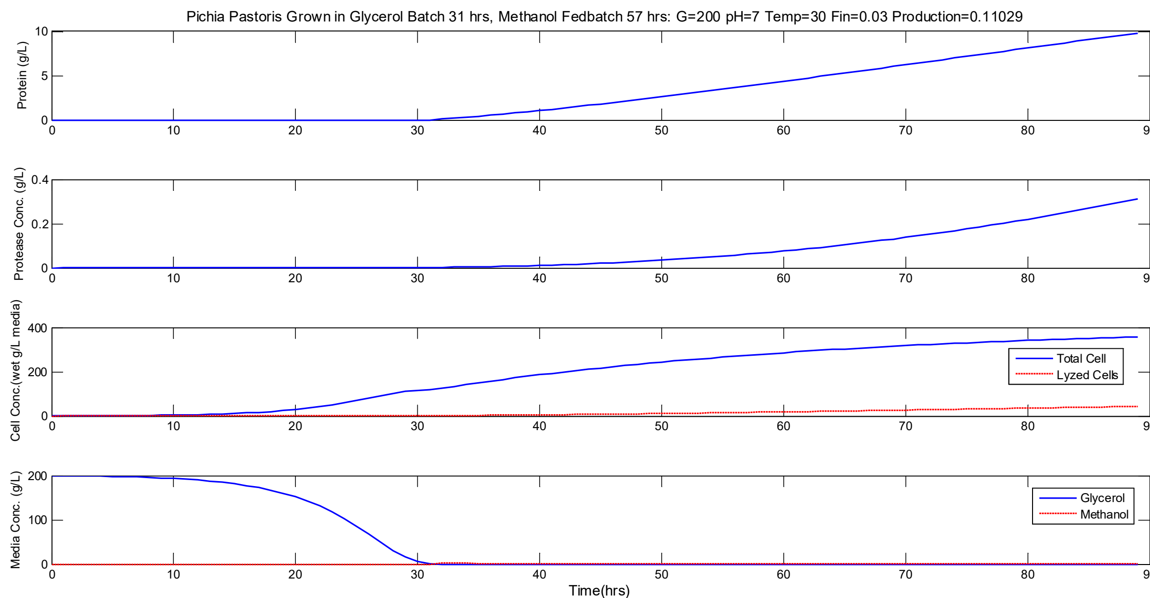

Fig 3. *Pichia pastoris* cells growing at optimized conditions: G=200 g/L, pH = 7, Temp=30 ^{0}C, Fin=0.03, t_methanol=57 hours

Students are encouraged to place all data and calculated values into a table so that comparisons between experimental runs at different engineering parameter sets can be compared. In the first part of the in silico lab, students vary one engineering parameter at a time from the baseline

© American Society for Engineering Education, 2019

conditions. They then calculate all the rates, yields, and any other relevant values and compare them to the values obtained using the baseline operating conditions. In Part 2 of the lab (25 points), the students must write a heuristic optimization strategy based on the experimental runs they already completed in Part 1. At this point, the students need to analyze the data and pinpoint which parameter(s) seem most critical and give suggestions on how they would vary these parameters to achieve the production target. They then test out their approach by running more experiments, varying one or more of the parameters as stated in their optimization plan. Note that they can change their approach based off of new data obtained. Finally an optimization goal of a certain production threshold is given that students should exceed for full credit (see Fig. 3).

Results and Discussion

Students tend to do very well when it comes to changing one design parameter at a time in the first part of the lab. Substrate concentration tends to be very well understood and student explanations for substrate behavior are the best out of all the engineering parameters. pH on the other hand tends to be the one where students struggle a bit more since it can have great effects on both cell growth and protease activity, and at times can lead to greatly enhanced protein production even though cell growth is less than baseline conditions. The write up for a heuristic optimization strategy is the weakest section of the lab for the students. Most will summarize accurately the experimental results from Part 1 but may find it a struggle to come up with an overall strategy combining all the engineering parameters to a set of optimal conditions. All students usually exceed the target optimal protein production rate – this is easily done by trial and error and hence has the lowest weighting in the lab rubric. The biggest challenge for students' success is their ability to outline a logical and reasonable strategy for optimization that is solely based on the results of their initial perturbations. Overall students in the 10 plus years of giving these labs will get an average typically of 65 – 85 with a standard deviation of 10-15 points.

Conclusions

The simulated protein optimization labs currently cover gene expression systems of *Pichia pastoris*, *E. coli*, and *Chinese Hamster Ovaries*. Each lab is very different but the underlying principles remain. This is a very novel assessment than the students have ever had before. After a lecture on heterologous protein production, a take home practice is given to the students which they complete and then discuss together as a class. There is a great deal of analysis involved, and this has to be trained to the students. The overarching goal is to provide as much experience for the students as possible to understand that in most cases, heterologous protein production is more of an art than a science. The actual lab for assessment will be protein production using a different gene expression system so students are exposed to a minimum of two different systems. Students after initially being daunted have found the exercise to be very interesting and worthwhile.

For three years, B. Aufderheide also had students complete the *in silico* protein production prior to actually making Human Growth Hormone using *Pichia pastoris* in a bioreactor lab. It was great to see students using the same data analysis tools and trying to figure out what is actually occurring in the bioreactor to affect protein production levels. B. Aufderheide also found that they helped him understand better the mechanisms of protein production and used the knowledge when doing his own contract fermentations of this nature.

© American Society for Engineering Education, 2019

References

1 Authors' web page: https://www.process2model.com

2 Van Der Schaaf H, Vermue M, Tramper J, Hartog R. "A design environment for downstream processes for Bioprocess-Engineering students". European Journal of Engineering Education, 2003 Vol 28, No 4, 507-521.

3 Bojorge N, Juliao F, Valdman B. "Computer Aided Training in Bioprocess Simulation and Control". IFAC Control Education and Technology 1995

4 Sessink ODT, Van Der Schaaf H, Beeftink HH, Hartog RJM, Tramper J. "Web-based education in bioprocess engineering". TRENDS in Biotechnology Vol 25 No. 1, 16-23

5 Dolan KD, Tang YJ, Liao W. "Improvement of bioengineering courses through systems biology and bioprocess modeling". 121st ASEE Annual Conference & Exposition, 2014, Paper ID # 10178

6 Cregg JM, Madden KR, Barringer KJ, Thill GP, Stillman CA. "Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris". Mol. Cell Biol., 1989 Mar 9(3), 1316-23

7 Zhang, Wenhui, Mark A., Brevins, Bradley A., Plantz, Leonard A., Smith, and Michael M., Meagher,"Modeling *Pichia pastoris* Growth on Methanol and Optimizing the Production of a Recombinant Protein, the Heavy-Chain Fragment C of Botulinum Neurotoxin, Serotype A", Biotechnology and Bioengineering, Vol. 70 (1), 2000, 1-8.

8 Inan, Mehmet, Vijay, Chiruvolu, Kent M., Eskridge, George P., Vlasuk, Kenneth, Dickerson, Scott, Brown, and Michael M. Meagher," Optimization of temperature–glycerol–pH conditions for a fed-batch fermentation process for recombinant hookworm (*Ancylostoma caninum*) anticoagulant peptide (AcAP-5) production by *Pichia pastoris*", Enzyme and Microbial Technology,Elsevier, 24, 1999, 438-445.

9 Kamaruddin N, Storms R, Mahadi NR, Illias RMd., Abu Bakar FD, Abdul Murad AM. "Reduction of Extracellular Proteases Increased Activity and Satbility of Heterologous Protein in *Aspergillus niger"*. Arab. J. Sci. Eng, Springer, 2018 43, 3327 doi.org/10.1007/s13369-017-2914-3

10 Wang Y, Xue W, Sims AH, Zhao C, Wang A, Tang G, Qin J, Wang H. "Isolation of four pepsin-like protease genes from Aspergillus niger and analysis of the effect of disruptions on heterologous laccase expression". Fungal Genetics and Biology, 2008, 45, 17-27

11 Zhu L, Nemeto T, Yoon J, Maruyama J, Kitamoto K. "Improves heterologous protein production by tripeptidyl peptidase gene (AosedD) disruptant of the filamentous fungus *Aspergillus oryzae*". J. Gen. Appl. Microbiol., 2012, 58, 199-209.

Dr. Brian Aufderheide

Dr. Brian Aufderheide is Chairperson in Chemical Engineering at Hampton University. He completed his PhD in Chemical Engineering at Rensselaer Polytechnic Institute. His areas of expertise are in advanced control, design, and modeling of biomedical, chemical, and biological processes. He has consulted for both medical device and biotechnology companies. He was sole engineer and QC supervisor of a 40MM lb/yr custom extrusion company. He has over 15 years of experience in education developing over 25 new courses. He has supervised over 35 Industrial Design Projects. He is a returned Peace Corps Volunteer. He is dedicated in helping his students to succeed.

Jemila James Nunes holds an Hons. BSc in Chemistry from the University of Toronto and an MSc in Process & Utilities Engineering from the University of Trinidad and Tobago (UTT). As part of her PhD studies at UTT, her focus is on optimization of gaseous fermentation conditions for production of single cell protein. At UTT, she lectures in the Process Engineering Department in the areas of Bioprocess Engineering and Chemistry. She was also a Researcher for a USD $10

© American Society for Engineering Education, 2019

million dollar semi-industrial pilot project at UTT which involved microbial conversion of hydrocarbons to protein. Her research interests lie in the fields of Bioprocess Engineering and Biotechnology, specifically in the use of fermentation technology to produce high value products.

© American Society for Engineering Education, 2019