NEXT GENERATION QUANTITATIVE MEASUREMENTS TO VALIDATE A MODEL FOR SACCHAROMYCES CEREVISIAE

By

Christopher Clay Stowers

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Chemical Engineering

August, 2008

Nashville, Tennessee

Approved:

Professor Kenneth A. Debelak

Professor Erik M. Boczko

Professor Robert D. Tanner

Professor Todd D. Giorgio

Professor Rick R. Haselton

Professor Paul E. Laibinis

To Dad and Mom, Rex and Sharon Stowers

and

My Wife, Monica Stowers

ACKNOWLEDGMENTS

I am forever grateful to my mentor, Prof. Erik Boczko, for his guidance and never-ending support during by doctoral studies. It is impossible for me to ever thank him enough for the countless hours of discussion and advice he provided throughout my graduate school work. Regardless of how busy he was, he never once turned me away when I wanted to discuss something. I hope that my current and future accomplishments will bring him satisfaction.

I also want to thank my advisor, Prof. Ken Debelak, for providing me guidance and focus during my graduate studies. I am sure that it would have been much more difficult to arrive at the end of my doctoral work without his support. I must also thank the members of my committee for providing guidance throughout graduate school and ensuring that my doctoral work was the best it could be. Their thoughtful criticisms have made me a better researcher. I especially want to thank Prof. Rick Haselton for providing laboratory space for many of my experiments.

I am also very grateful to my many collaborators. Much of the work presented here would not have been possible without their contributions. Below is a list of collaborators, including the section of this work to which they contributed: Erik Boczko (II-VI), Hyunju Ban (IV-VI), Doug Hackworth (III), Tomas Gedeon (III), Konstantin Mischaikow (III), Todd Young (IV), J. Brian Robertson (IV and V), Rick Haselton (VI), Carl Johnson (IV), and Robert D. Tanner (V & C).

I also gratefully acknowledge the support of the National Science Foundation (NSF) grant # DMS 0443855.

Finally, no one has supported me more than my family. I could never express in words how grateful I am to my parents. They have believed in me and supported me in all of my endeavors. They have always been excellent role models and have never wavered in their love and support. I am also grateful for all the love and support I have received from my wife. She has been overwhelmingly supportive of my studies, despite the many late nights in the lab. She has provided me abundant support during the last few months of my doctoral work when I needed it the most.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Rationale

The study of dynamic biological systems is perhaps the most captivating of all scientific endeavors due to the observed robustness of the biological organism. The genetic networks that underpin these organisms must hold the keys to understanding their emergent behavior. However, quantitative techniques for interrogating these networks are currently in their infancy. Specifically, there exists a lack of sensitive and robust assays for measuring nucleic acids. There is also a lack of mathematical models pertaining to the population dynamics and asynchronous growth of organisms. These two complications make studying dynamic biological systems at the gene expression level impervious. Thus, the overall goal of this dissertation is to develop tools and techniques to be used to measure the quantitative behavior of biological systems at the gene expression level.

NCR Background

The specific system that provides the motivation for this work is an ostensibly simple stress response circuit in baker's yeast, *Saccharomyces cerevisiae*, that regulates the organism's genetic response to nitrogen limitation called *nitrogen catabolite repression* (NCR), see Figure 1.1. The circuitry of the network has been well studied for the last 40

years and comparatively much is known about its function; however, little is known about its dynamics [9,10]. In order to study the dynamics at the same level of sophistication at which mathematical models have been formulated, accurate quantitative biophysical and biochemical techniques are required at molecular dimensions and on physiological timescales.

It is of widespread interest to measure the genetic response of an organism to stress. The NCR circuitry of *Saccharomyces cerevisiae* provides a challenging but tractable opportunity. Minimal mathematical models of the NCR circuit predict that the mitotic cell cycle will drive the expression of the GLN3-URE2p subcircuit causing their corresponding mRNA and protein species to oscillate in phase[5,6]. The measurement of such signals is important to further our understanding of the relationship between the organization of genetic circuitry and the dynamic responses that it is capable of transducing. That is, we are trying to understand the structure theorems that underpin what adaptation has forged into robust and responsive pathways working with the limitations imposed by biophysics. Understanding such theorems and their limitations will improve the understanding of the robustness of biological systems, help us build fault tolerant computer networks, design adaptive nano devices, and design more efficient bioprocesses.

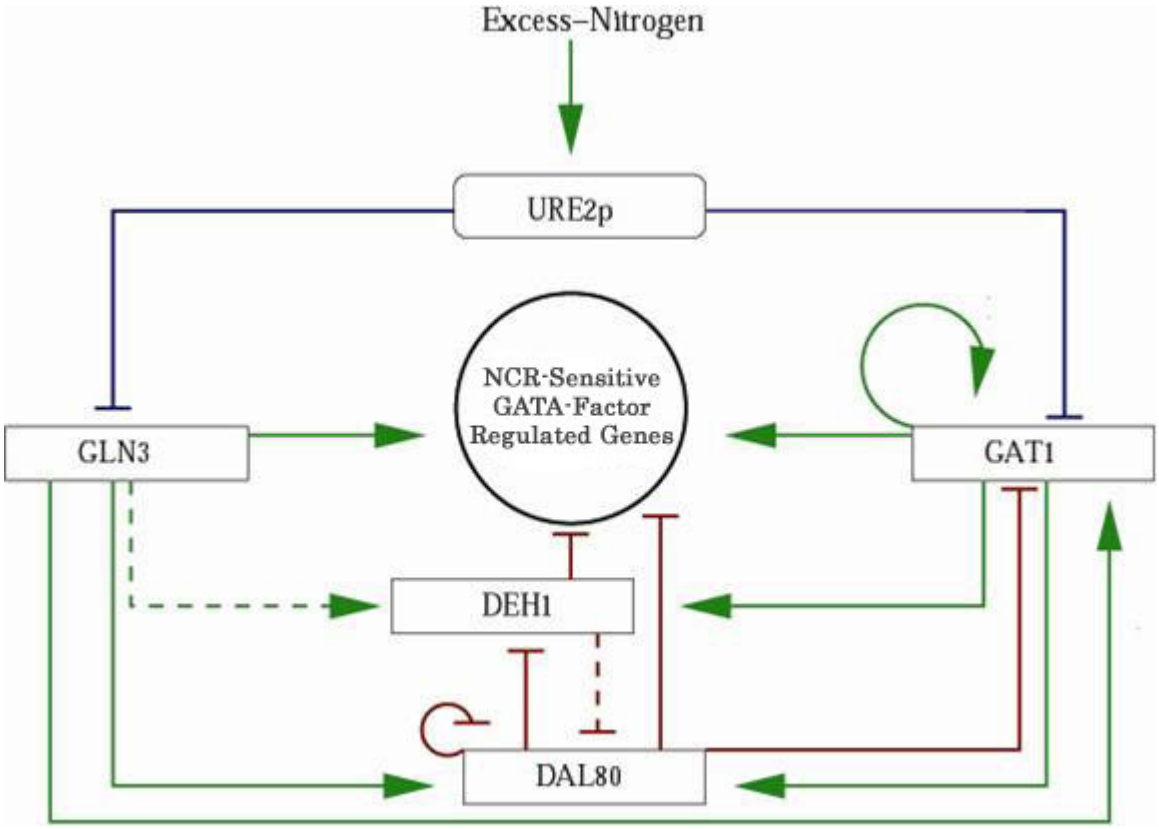


Figure 1.1: The NCR circuit. The square blocks represent mRNA of the respective NCR gene. The rounded block represents the transcribed protein of the URE2 gene. Green arrows indicate up-regulation and blunted red arrows represent down-regulation at the level of transcription. Dashed lines represent a weaker response. Blue lines represent repression that is not at the transcription level.

The existence of regulatory loops, time delays, and complex transport phenomena has brought mathematical modeling and analysis to the forefront of systems biology. To date, many signal transduction motifs have been identified in well studied regulatory networks and also by large-scale screens [10]. Carefully considered and important results about model identification for motifs and general reaction terms have appeared in the literature [4,21,22]. General results for pattern forming circuits have also been published [12]. Numerical screens have been performed suggesting that biologically adapted circuits possess exceptional robustness [24]. Rigorous analyses have been performed on these circuits to understand the mathematical origins of robustness [3]. The majority of these results come from systems that are relatively data poor. Whether or not these models and the information contained within them survive the test of time depends crucially on the ability to make careful measurements at the molecular level and on physiological timescales.

History of Biological Measurement

Unfortunately, many if not most traditional biophysical techniques were not developed with dynamics in mind and do not have the power to effectively interrogate the dynamics of genetic regulatory systems. It is important to realize that scientific progress typically hinges upon technological advancement. For example, the crowning achievement of the 1950's was the discovery of the structure of DNA that required the chemistry of crystallography, the physics of diffraction, and the mathematics of group theory to accomplish. Over the next decade, the tenets of the Central Dogma of Biology were established and gave biology its first real logical molecular underpinnings. Interestingly,

up until the early 1980's sophisticated studies of yeast physiology and biochemistry were routine [7,17,25]. However, the focus on the gene as the fundamental unit of biological explanation led to a roughly 25 year change of focus toward static gene discovery and a near abandonment of individual and population physiology. Paradoxically perhaps, the ascendancy of genomics has now reintroduced the desire to understand individual and population physiology at the molecular level. Armed with molecular information and powerful computers, there has been a steady increase in literature involving the modeling of pathways, circuits, organelles, and even whole cells [20]. In fact, it is the reductionists' dream to build a computational model of a cell that would respond to stimuli as real cells do. Systems biologists from all disciplines have fully embraced this goal. This endeavor has exposed several experimental obstructions. We and many others have been forced to focus on the problem of measurement. For example, the basic biology lab staple of Southern Blotting was developed in the 1980's and has been used countless times to provide evidence for the existence of a particular gene in a cell line; however, the technique lacks the sensitivity to detect subtle changes in gene expression. These subtle changes are often expected to be essential to the functional dynamics of gene regulatory networks[6].

Steps of Gene Expression Measurement

Faced with the need to produce quantitative measurements of the NCR circuitry dynamics, but realizing that currently available techniques fall short of the needed quantitative accuracy, we began to analyze the measurement process step by step. Quantitation of gene expression from regulatory networks can be separated into three

steps: nucleic acid harvesting through cell disruption (CHAPTER II), cell cycle synchronization/population modeling (CHAPTERS III, IV, & V), and nucleic acid quantification (CHAPTER VI). The inaccuracy of any of these processes impact the accuracy of gene expression data. In fact, depending on experimental design, other processes may be involved whose accuracy may also be essential to the overall accuracy of gene expression data. For example, an additional step such as reverse transcription is commonly used to convert harvested RNA into DNA prior to nucleic acid quantitation. However, this dissertation will focus on the three steps mentioned above.

Cell Disruption

Cell disruption is perhaps the most commonly overlooked source of quantitative error in studies where biomolecules are quantified from single cells. It fact, the measurement of biomolecules is directly proportional to the number of cells disrupted. This dependency lends the process to errors since samples are typically taken in a time-course and are disrupted independently. Variations in the percentage of disrupted cells from sample to sample produces an inherent error in the end quantitation. One could imagine that particular cells may be more likely to disrupt than others, possibly due to replicative age or cell cycle position. Therefore, gene expression dependent on age or the cell cycle may be inaccurately quantified from obtaining a biased pool of nucleic acids through typical cell disruption protocols. Many use internal controls such as actin to normalize disrupted samples, but this process could be logically circular if variations in the standard exist. CHAPTER II of this dissertation provides an in-depth analysis of cell disruption and the

process of harvesting nucleic acids from yeast. Modifications to traditional techniques are proposed in order to provide more reliable and uniform cell disruption.

Cell Cycle Synchrony

Cell cycle dependence is also a complication in quantifying gene expression values. The cell cycle poses a significant problem for the following reason. As yeast age, their volume grows. There is a volume checkpoint required of newly budded daughter cells to enter the cell cycle that immediately and completely dephases mothers and daughters making the culture notoriously asynchronous. The quantitative phase shift depends on the specific growth conditions, but in rich media it is approximately 11 minutes, or $1/12^{th}$ of the total cell cycle time. This small difference will within three cell cycles drive the entire population to stationary asynchrony that will obscure any periodic signal. In addition, since all age classes are mixed in a stationary culture, one cannot discern age differences in gene expression – an entirely open and fascinating subject. Furthermore, at stationary state, cells are expected to be uniformly distributed around the cell cycle with respect to time. Thus, if a particular gene is expressed in sinusoidal form around the cycle, but a pool of nucleic acids are harvested from a population of cells distributed uniformly around the cycle, a constant signal instead of the dynamic signal of the individual cell will be detected.

Common techniques for achieving cell cycle synchrony in yeast produce a mere 3 to 4 synchronous cycles, results from which are typically far short of validating mathematical models for gene network dynamics. Additionally, many of these techniques, such as

alpha factor arrest, starvation, and elutriation could likely produce a stress response related to the gene under study, leading to a measurement influenced by the measurement process itself. For this reason, this dissertation focuses on alternative techniques capable of producing extended cell cycle synchrony without invasive methods (CHAPTER III), possibly providing the missing link to obtaining more useful information about the expression patterns of cell cycle dependent genes. This technique utilizes continuous volume filtration on synchronous yeast populations as a mechanism for maintaining synchrony.

A further alternative for extending cell cycle synchrony is to exploit autonomous cell cycle oscillations that have been reported in the literature for over four decades [8,13,14,18,23]. Many of the details of the feedback mechanism(s) that underlie these oscillations are still poorly understood. These oscillations provide an interesting research opportunity as well as a vehicle for producing synchrony. A potential disadvantage is that these oscillations have only been achieved by carefully controlling the environment (dilution rate, pH, dissolved oxygen, nutrients, etc.) within the bioreactor. This dependency means that the environmental variables are limited in their flexibility during experiments while preserving culture synchrony. This restriction potentially limits the ability to learn about the dynamics of the NCR circuitry. CHAPTER IV of this dissertation focuses on the development of using these oscillations as a tool for measuring synchronous gene expression, particularly those genes of the NCR circuitry.

The autonomous oscillations described in Chapter IV are typically considered an obstruction within the bio-manufacturing industry [15]. In fact, just as these oscillations

can be utilized by scientists to measure synchronous gene expression, as described in Chapter IV, they could also be utilized by engineers as a mechanism of persistent synchronous production of cell cycle dependent bio-materials. Downstream bioprocessing costs constitute over 75% of the total cost of production for bio-materials[1]. These costs are generally associated with purifying a desired product away from undesired biological material, the ambient bio-complexity. By utilizing systems that permit synchronous production of bio-products, rather than the typical asynchronous production, the product concentration within the effluent of the bioreactor can be enriched relative to the biocomplexity. Chapter V of this dissertation focuses on developing these ideas.

Polymerase Chain Reaction (PCR)

Since many genes, including those of the NCR circuit, are believed to be cell cycle dependent, there is a vast research effort to detect expression levels from single cells in order to avoid population convolution of the measured signal [16]. The only readily available technique for quantifying of nucleic acids at the single cell level is the Polymerase Chain Reaction (PCR). The advantage of PCR over other existing techniques is the ability to exponentially amplify small signals into a detectable range. Unfortunately, the PCR process also amplifies any error created in sample preparation. Additionally, problems with DNA contamination become increasingly more important as samples become more dilute. Thus, the two largest problems of using quantitative PCR, exponential amplification of error and DNA contamination, are most prevalent within the realm of DNA concentrations expected within a single cell.

Although the literature has been flooded with PCR studies in virtually every realm of biology, very few studies have investigated the quantitative nature of the technique itself [2,11,19]. Over the past two decades, PCR has been used countless times to quantify nucleic acids, yet the vast majority of studies seem to avoid the topics of accuracy and reproducibility. Furthermore, no study of which we are aware has investigated the quantitative nature of PCR for the realm of dilute nucleic acid concentrations pertaining to single cell measurements. For this reason, a section of this dissertation focuses on nucleic acid quantitation using PCR, with specific emphasis on dilute nucleic acid concentrations (CHAPTER VI).

The research presented in this dissertation focuses on the development of new tools and techniques to make quantitative measurements on subtle changes in gene expression patterns. Cell disruption and the harvesting of nucleic acid molecules are explored and a quantitative model for a specific protocol is developed. Two separate strategies for combating cell cycle dependence of gene expression are explored. These two techniques have separate advantages and defects, and these are discussed later. Lastly, the direct quantitation of nucleic acids is explored through an analysis of reproducibility and accuracy of PCR.

References

1. Aldridge, S. 1996. Downstream processing needs a boost. *Genetic Engineering and Biotechnology News*. **26**:1.

2. Abramov, D., Trofimov, D. and Rebrikov, D. 2006. Accuracy of a real-time polymerase-chain-reaction assay for quantitative estimation of genetically modified food sources in food products. *Applied Biochemistry and Microbiology*. **42**:428-430.

3. Barkai, N. and Leibler, S. 1997. Robustness in simple biochemical networks. *Nature.* **387**:913-917.

4. Beck, T. and Hallman, M. 1999. The TOR signaling pathway controls nuclear localization of nutrient regulated transcription factors. *Nature.* **402**:689-692.

5. Boczko, E.M., Cooper, T.G., Gedeon, T., Mischaikow, K., Murdock, D.G., Pratap, S., and Wells, K.S. 2005. Structure theorems and the dynamics of nitrogen catabolite repression in yeast. *Proc. Natl. Acad. Sci.* **102**:5647-5652.

6. Boczko, E.M., Gedeon, T., and Mischaikow, K. 2007. Dynamics of a simple regulatory switch. *Journal of Mathematical Biology.* **55**:679-719.

7. Brewer, B.J., Chlebowicz-Sledziewska, E., and Fangman, W.L. 1984. Cell cycle phases in the unequal mother/daughter cell cycles of *Saccharomyces cerevisiae. Mol. Cell. Biol*. **4**:2529-2531.

8. Chen, Z., Odstrcil, E., Tu, B., and McKnight, S. 2007. Restriction of DNA replication to the reductive phase of the metabolic cycle protects genome integrity. *Science*. **316**:1916-1919.

9. Cooper, T. 2002. Transmitting the signal of excess nitrogen in *Saccharomyces cerevisiae* from TOR proteins to the GATA factors: connecting the dots. *FEMS Microbiology Reviews*. **26**: 223-238.

10. Cunningham,T. Rai, R., and Cooper, T. 2000. The level of DAL80 expression down-regulated GATA factor-mediated transcription in *Saccharomyces cerevisiae. Journal of Bacteriology*. **182**:6584-6591.

11. Diehl, F., Li, M., Dressman, D., He, Y., Shen, D., Szano, S., Diaz, D., Goodman, S., David, K., Juh, H., Kinzler, K., and Vogelstein, B. 2005. Detection and quantification of mutations in the plasma of patients with colorectal tumors. *Proc. Natl. Acad. Sci.* **102**:16368-16373.

12. Ermentrout, B. 1998. Neural networks as spatio-temporal pattern-forming systems. *Rep. Prog. Phys*. **61**:353-430.

13. Essajee, C. and Tanner, R. 1979. The effect of extracellular variables on the stability of continuous baker's yeast-ethanol fermentation process. *Process Biochemistry.* **25**:16-21.

14. Heinzel, E., Dunn, I., Furukawa, K., and Tanner, R. 1982. Modeling of sustained oscillations observed in continuous culture of *Saccharomyces cerevisiae*. *Proceeding of the 1st IFAC Workshop.* 57-65.

15. Jules, M., Francois, J., and Parrou, J. 2005. Autonomous oscillations in *Saccharomyces cerevisiae* during batch cultures of trehalose. *FEBS Journal*. **272**:1490-1500.

16. Longo, D. and Hasty, J. 2006. Dynamics of single cell gene expression. *Mol. Syst. Biol*. **2**:64.

17. Lord, P. and Wheals, A. 1980. Asymmetrical division of *Saccharomyces cerevisiae*. *J. of Bacteriology*. **142**:808-818.

18. Murray, D.B., Klevecz, R.R., and Lloyd, D. 2003. Generation and maintenance of synchrony in *Saccharomyces cerevisiae* continuous culture. *Exp. Cell Res*. **287**:10-15.

19. Roussel, Y., Harris, A., Lee, M., and Wilks, M. 2007. Novel methods of quantitative real-time PCR data analysis in a murine *Helibobacter pylori* vaccine model. *Vaccine*. **25**:2919-2929.

20. Sanbonmatsu, K. and Tung, C. 2006. Large-scale simulations of the ribosome: a new landmark in computational biology. *J. of Theoretical Biology and Biophysics*. **46**:334-342.

21. Santillan, M. and Mackey, M. 2004. Influence of catabolite repression and inducer exclusion on the bistable behavior of the lac operon. *Biophysical Journal*. **86**:1282-1292.

22. Swameye, I., Muller, T., Timmer, J., Sandra, O., and Klingmuller, U. 2003. Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by data-based modeling. *Proc. Natl. Acad. Sci*. **100**:1028-1033.

23. Tu, B.P., Kudlicki, A., Rowicka, M., and McKnight, S.L. 2006. Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science.* **310**:1152-1158.

24. von Dassow, G., Murno, E., and Odell, G. 2000. The segment polarity network is a robust developmental module. *Nature*. **406**:188-192.

25. Wheals, A.E. 1982. Size control models of *Saccharomyces cerevisiae* cell proliferation. *Molecular and Cellular Biology.* **2**:361-368.

CHAPTER II

RELIABLE CELL DISRUPTION IN YEAST

Introduction

Yeast is an important eukaryotic model system whose physiology, regulation, and adaptation is beginning to be understood in quantitative terms [1,2,3,4,7,17,26,28,29]. As systems biology is evolving ever more sophisticated mathematical models, the need and desire for precise quantitative measurements is increasing. It is a consequence, perhaps of some cosmic conservation that while yeast genetics are relatively straightforward, their biochemistry is complicated by their thick cell wall that can withstand extreme pressures of greater than 300 MPa [8]. This fact complicates the procedures to isolate and analyze intracellular macromolecules and metabolites.

There has been a long-standing tradition in biochemistry and molecular biology to normalize measured quantities, such as mRNA, with constitutive internal controls. Actin, for instance, is thought to be a housekeeping gene and is often used to normalize the loading and comparison of samples. A problem with this practice is that without a known gold standard this approach leads to circular reasoning. What if actin levels are instead oscillating? Then anything that appears as constant compared to actin, is not actually constant, but rather oscillating. As methods of detection increase in resolution, the problem of normalization becomes only more apparent. Our ultimate goal is to resolve a conjecture that certain yeast mRNA concnetrations are oscillating periodically with the

cell cycle as predicted by a mathematical model [2,7]. There is growing evidence for and renewed interest in periodic gene expression and its relation to the cell cycle as well as to ultradian rhythms whose causes remain uncertain [12,16,18,24]. As this interest grows, so does the need for logically sound and quantitative protocols for normalization.

In order to convincingly demonstrate the precise dynamical behavior of a given mRNA species from a culture of yeast cells, it is important to demonstrate a logically and practically sound method for normalization between sample variations. Many recent papers have discussed both the theoretical and practical issues surrounding the normalization problem [5,11,15,25]. Because the use of internal controls is inherently circular, we are drawn to the logical clarity of external controls [28]. In this practice, a known quantity of foreign RNA is introduced into the sample during extraction. The single assumption of this practice is that the introduced species is indistinguishable from the other RNA molecules from the point of view of the remainder of the extraction process. The practice depends on a single parameter: how much external control does one add to each sample?

Typically, in a time course, independent samples will contain different numbers of cells. With the use of a coulter counter, it is possible to directly measure the total volume of cells in a sample. It seems natural and reasonable that an external control proportional to the total volume of disrupted cells be used to normalize mRNA expression. Thus, a logical goal would be to spike yeast samples with a quantity of external RNA control that is proportional to the total sample volume. If, however, the volume of cells disrupted by a method for RNA extraction is not a constant and reproducible fraction of the total cell

volume, then this practice cannot work. We consulted the literature for data describing a method to quantitate cell disruption and found none. This problem is typically of the nature of laboratory folklore, and we found unreferenced statements similar to the following:

‘ Disintegration of yeast cells has been estimated by protein concentration in the supernatant’ [22].

It is important to observe that disruption for the purpose of biochemical isolation is related to cell wall integrity and not cell vitality. This work demonstrates that the dye aniline blue, that specifically binds β 1,3-Glucan, clearly distinguishes disrupted cells[13,14]. In typical yeast disruption protocols, cells suspended in a phenol chloroform mixture are disrupted by beating them against glass microbeads on a vortex mixer. The protocol suggests beating for 10 minutes. The data taken in this study demonstrate that 10 minutes is sufficient to disrupt roughly 57% +/- 13, while disruption greater than 94% +/-3 requires 30 minutes. Furthermore, agitation times short of 30 minutes are shown to be insufficient to uniformly disrupt cells across all phases of the cell cycle.

Because the percentage of cells disrupted at time *t* is a monotonically increasing function and the damage sustained by a cell over time is cumulative, it is reasonable to suspect that this process is modeled by a cumulative distribution function. The data generated in this study are used to demonstrate that the percentage of cells disrupted by the protocol after *t* minutes of agitation is well modeled by a Weibull distribution as shown below in Equation (2.1),

$$F(t) = 1 - e^{-(\frac{t}{\alpha})^{\beta}} \qquad \text{Equation (2.1)}$$

This distribution has found extensive application in engineering problems involving fatigue, fracture, and reliability modeling [19]. The shape parameter β allows one to capture processes that have time-dependent, non-constant failure rates. The data collected in this study show a decreasing failure rate that is consistent with the observation of cell cycle phase dependent disruptability.

Materials and Methods

Yeast cells of *Saccharomyces cerevisiae* strain LHY3865 (mat a- URA, LEU, bar1Δ) were grown in YNB media without ammonia or amino acids and with 100 mg/L leucine, 20 mg/L uracil, 0.2% glutamine, and 2% glucose at 30 °C. Batch cultures were grown with agitation at 225 rpm in a New Brunswick Innova 44 orbital incubator/shaker. Continuous cultures were grown in a New Brunswick BioFlow 110 3.0 L bioreactor with a dilution rate of D=0.35 hr^{-1}, the pH was maintained at 5.3 standard units by sparging CO_2, and the culture was aerated with a Rushton-type impeller run at 225 rpm. RNA quality was determined with an Agilent bioanalyzer 2100 Expert System.

Disruption Assay

Cells were harvested from batch cultures or from continuous cultures from the bioreactor at $\mathbf{OD}_{600}$=0.6 and pelleted by centrifugation at 12,000 rpm for 2 minutes. A scaled down Ambion RiboPure assay (kit #1926) was performed according to the described protocol with the following modifications: the reactions were performed in 0.2 mL PCR tubes and the reagents were scaled down 25-fold (19 μL of lysis buffer, 2 μL of 10% SDS, and 19 μL of phenol chloroform mixture (60:40) per reaction).

The lysis mixture was added to a PCR tube containing approximately 30 μL of cold zirconia beads (supplied within the Ambion kit). Each PCR tube was fit inside a 1.5 mL Epindorf tube, placed in a vortex adapter, and agitated for the specified time. The lysate was rinsed from the PCR tube with several washes of phosphate buffered saline (PBS) to ensure that all cells, particularly damaged cells, were washed free from the beads. The lysate was pelleted by centrifugation for 5 minutes at maximum speed (15,000 rpm). The RNA enriched supernatant was removed for subsequent analysis, while the remaining cell pellet was resuspended in 250 μL of GIBCO 1x PBS (#14040-133). Aniline blue Sigma (#415049) was added to the cell lysate at concentration of 1 mg/mL. The stained lysate was analyzed using a conventional microscope equipped with a DAPI filter.

Scoring Algorithm

The aniline blue dye robustly illuminates the cell wall allowing interrogation of individual cell wall integrity. Stained cells were scored as intact or disrupted by visual inspection with the following criteria: A disrupted cell has sustained sufficient damage to spill its intracellular contents to the void. The decision process is illustrated in Figures 2.1-2.3. Typically, cells judged as disrupted had collapsed intracellular compartments or holes within the cellular membrane comprising at least 20% of the surface area of the cell. Cells with minimal damage to their cellular membrane were scored as intact. Cells with substantial damage often exhibited collapsed intracellular compartments allowing them to be clearly scored as disrupted, see for instance Figure 2.1-E and Figure 2.2 (all but E) and Figure 2.3-C. Individual cells were scrutinized in several focal planes to ensure accuracy (see Figure 2.2). Cells were scored 100 at a time. Each estimate of the fraction disrupted was computed from no less than 9 independently prepared slides.

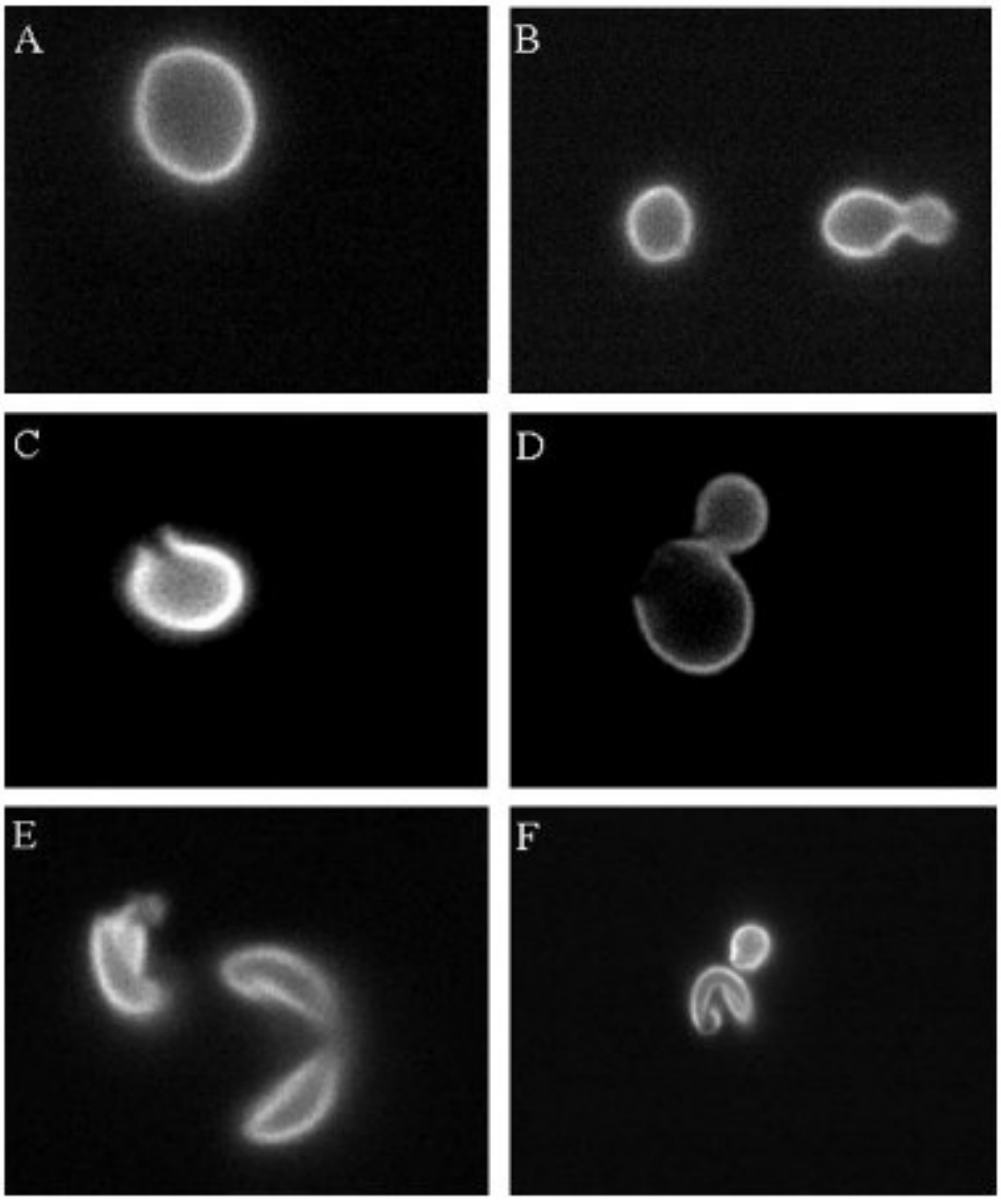


Figure 2.1: The panels illustrate the visual scoring algorithm designed to determine cell disruption. Cells were labeled with aniline blue. The cell in image (A) is completely intact. The cells in image (B) appear to have small nicks in their cell wall, but these are deemed insufficient to spill their intracellular contents. The budded cell is considered as a single cell because no septum has yet formed. The cell in image (C) has sustained more pronounced damage, but is deemed intact because its contents appear to remain intracellular. Of the two cells in image (D), the mother cell is disrupted. The daughter is considered her own cell and remains intact. The cell(s) of image (E) are considered disrupted as their cell walls are cleaved and they have released their contents to the void. In the final frame (F), one of the two cells is deemed as disrupted.

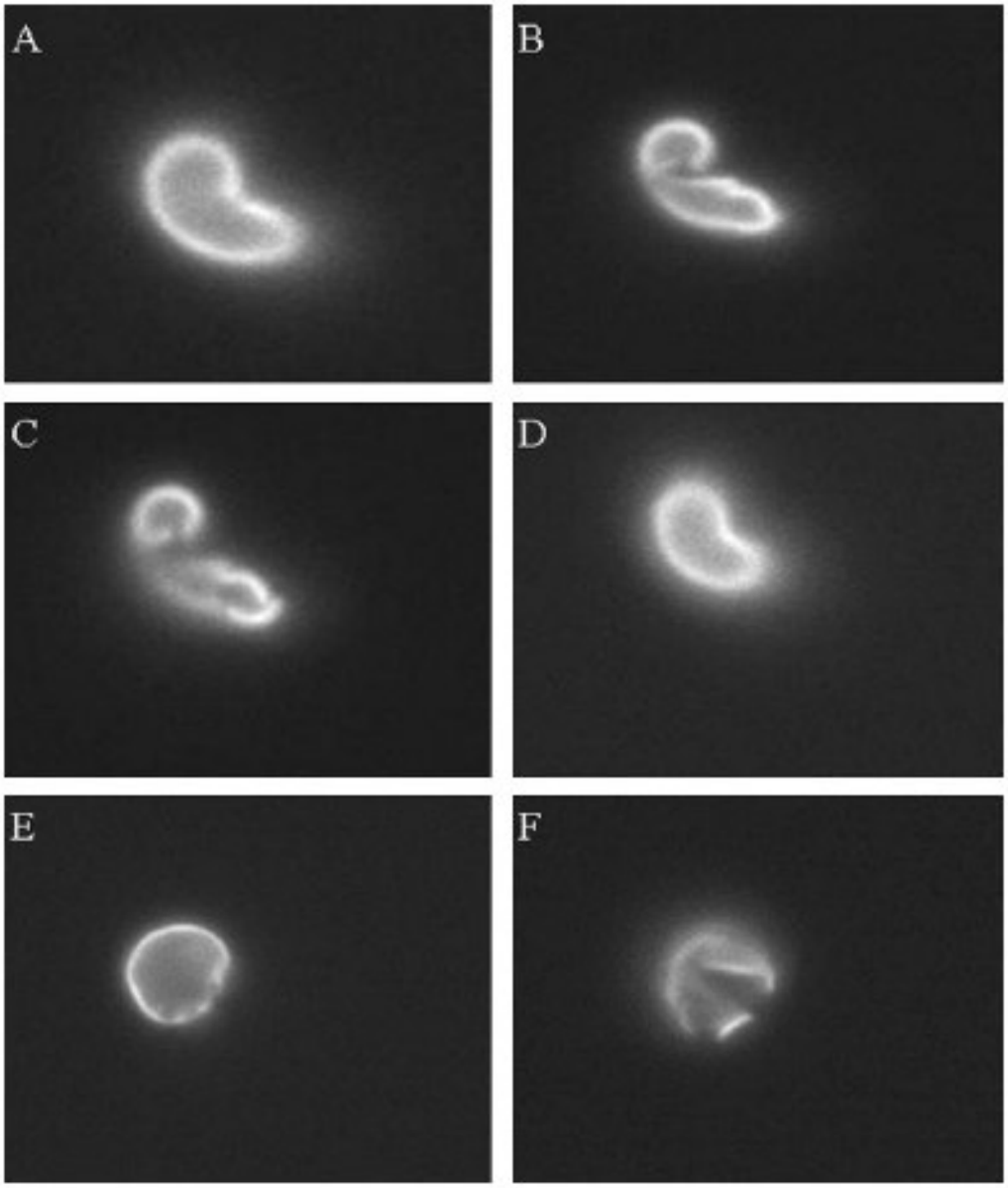


Figure 2.2: This panel of images illustrates the importance of the focal plane in determining cell disruption by the scoring algorithm. Images A-D shows four different focal planes for the same cell. Accordingly, this cell was classified as disrupted. The images in E-F support the same conclusion for a different cell. In our counting procedure, cells were scrutinized 100 at a time and each cell was individually examined for defects to ensure accuracy.

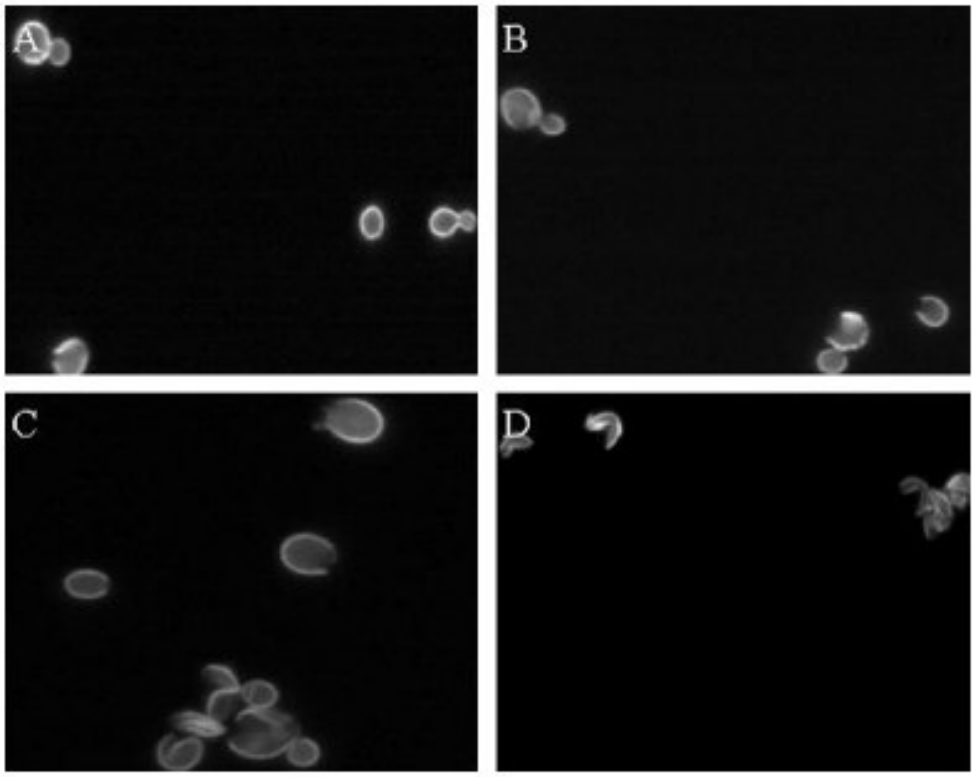


Figure 2.3: The images in this panel are taken along a time course of disruption and are representative of the process. Image (A) shows cells subjected to 2 minutes of disruption. All of the cells in this image were deemed intact. Image (B) is taken after 10 minutes of disruption, the standard time suggested by the protocol. Two of the five cells were scored as disrupted. Image (C) is taken after 20 minutes of disruption. Six of the nine cells in the field were scored as disrupted. Image (D) shows cells after 30 minutes of disruption and all of the cells in this frame were scored as disrupted.

Cell Cycle Synchronization

Yeast cells were arrested through the addition of $3x10^{-5}$ M α-factor mating pheromone (Sigma #63591) and were incubated for 3 hours. Cells were subsequently released from arrest by pelleting followed by three washes with fresh pre-conditioned media, free of α-factor, containing 0.1 mg/mL Pronase E (Sigma # P-6911). The preconditioned media was prepared by allowing LHY3865 yeast cells to grow within the media for 4 hours at $\mathbf{OD}_{600}$ = 0.4 before being removed by a 0.2 μm filter. The synchronized cells were grown in the bioreactor as described above and were harvested at six time points pre-determined to correspond to critical points within the cell cycle. The time points were determined from preliminary experiments using microscopy to establish the timing of standard cell cycle morphological landmarks characteristic of cell cycle transition [9,10]. Briefly, these landmarks are the following: G_1 is marked by the separation of a parent and daughter cell, the beginning of S and end of G_1 is marked by bud emergence, the beginning of the M phase is marked as the emergence of bud necking. It is widely accepted that the S portion of the cell cycle is 20-30 minutes for *Saccharomyces cerevisiae*. For the strain used in this study, we have determined S to be approximately 30 minutes by analyzing volume growth measurements.

Results

The disruption of the yeast cells by mechanical agitation was visualized because the dye aniline blue clearly stains the cell walls of budding yeast such that it was possible to monitor and score cell breakage as described above. It was an important consideration to find a dye that did not simply stain for cell vitality, but allowed us to assay cell wall integrity since mechanical agitation is designed to liberate the intracellular contents of cells through rupture.

Since we know of no protocol or device that is capable of directly measuring the degree to which the intracellular contents of a cell are free for extraction, an indirect measurement was utilized. The staining procedure and the scoring protocol that was employed produced consistent and reproducible results. This judgment is based on several independent observations. First and foremost, cellular disruption through mechanical agitation is a cumulative process and the fraction of disrupted cells as a function of time should be modeled by a cumulative distribution function. We found that our data are well modeled by a Weibull distribution. Second, the relative recovery of well mixed intracellular species should be proportional to the relative volume of disrupted cells. We also found this to be true.

The data in Table 2.1 summarize the results of six independent experiments. Each of the individual data points describing the degree of disruption is the result of considering over 900 independent cells. These data were used to produce the graph in Figure 2.4 that shows how the variance decreases with agitation time and how the data are well

described by a Weibull distribution with scale parameter α=10.97 min and shape β=0.77. The significance of the shape parameter is that unlike an exponential or Poisson process, the Weibull process with $\beta < 1$ implies a nonconstant disruption rate that is decreasing with time. Fewer cells are being disrupted per unit time as the agitation time increases. The data can also be accurately modeled by an exponential function. Fitting the data to a single parameter exponential function provides a similar value for the time constant, α, of 11.65 min. The cells described by the data in Table 1 were harvested at $\mathbf{OD}_{600}$=0.6 near mid-log phase. The same quantitative results were observed for cells at lower $\mathbf{OD}_{600}$=0.4 and higher $\mathbf{OD}_{600}$=1.0 densities, and also with another strain of *Saccharomyces cerevisiae* BY4743(data not shown).

Table 2.1: Data and statistics summarizing the percentage of cells disrupted from six identical and independent (iid) experiments. Each individual entry is the result of inspecting at least 900 aniline stained yeast cells. The average values and their standard deviations (SD) are plotted in Figure 2.4. Among other things, these data indicate that 30 minutes is sufficient to reproducibly disrupt greater than 90% of cells with a coefficient of variation (CV) of less than 4%.

Time (min)	**Fraction disrupted (iid) samples 1-6**						**Average**	**SD**	**CV**
2	0.230	0.375	0.346	0.374	0.215	0.060	0.267	0.124	0.463
10	0.657	0.628	0.644	0.636	0.504	0.332	0.567	0.128	0.226
20	0.853	0.838	0.863	0.832	0.720	0.594	0.783	0.106	0.136
30	0.945	0.957	0.962	0.963	0.888	0.897	0.935	0.034	0.036

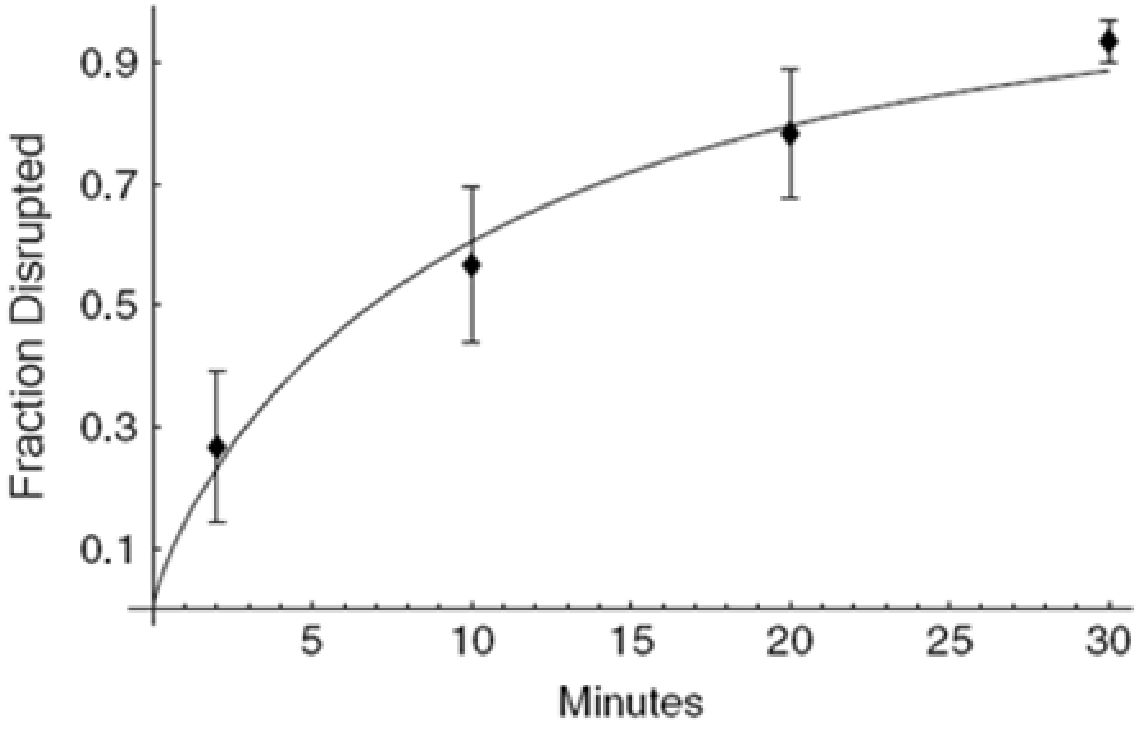


Figure 2.4: The best fit Weibull distribution to the cell disruption data tabulated in Table 2.1. The corresponding best fit parameters are α=10.97 min and β=0.77.

The data in Table 2.2 describe the quantitative yield of RNA recovered from 6 independent experiments. The high quality of the extracted RNA is inferred in part from the RIN (RNA Integrity Number) that were determined through bioanalysis[21]. All of the samples had RIN numbers between 9.8 and 10. The 260:280 ratios of all the samples lie between 2.11 and 2.29 and the 260:230 ratios lie between 2.0 and 2.47. Thus, the quality of the extracted RNA does not decrease with increasing agitation times. The data in Table 2.2 clearly indicate that the quantity of extracted RNA increases with disruption time. In order to demonstrate that the yield is proportional to fraction of cells disrupted, the ratio of the yield of RNA, Y, is plotted relative to the fraction of disrupted cells, F, as counted by the disruption assay developed in this study (see Figure 2.5). Finally, we observe that the variance in yield is markedly decreased with 30 minutes of disruption. These observations support the notion proposed above that the scoring method employed is at least consistent.

Table 2.2: RNA yield is shown as a function of disruption time. The results from six independent and identical experiments are shown.

Time(min)	**RNA yield (ng/µL)**					
2	86.2	89.5	140.8	146.3	178.6	191.1
10	202.9	228.6	258.1	283.5	290.1	305.6
20	328.8	357.6	372.4	480.2	483.2	492.7
30	543.3	557.1	558.1	564.9	566	605.7

In addition to exploring the behavior of stationary batch cultures, we explored the degree of disruption as a function of the cell cycle phase and agitation time. The cell cycle data are summarized in Table 2.3 and are plotted in Figure 2.6. The data show that agitated yeast cells are most likely to break in M-phase and that the degree of disruption approximates a saw tooth function around the cell cycle, with a linear increase in disruptability from G_1 to M and then back again. It has previously been observed that cells are heartiest in G_1[20]. The observed cell cycle dependent disruptability is consistent with the observation that FKS1 expression is cell cycle dependent and peaks in G_1[23]. FKS1 encodes a subunit of β 1,3-glucan synthase[6].

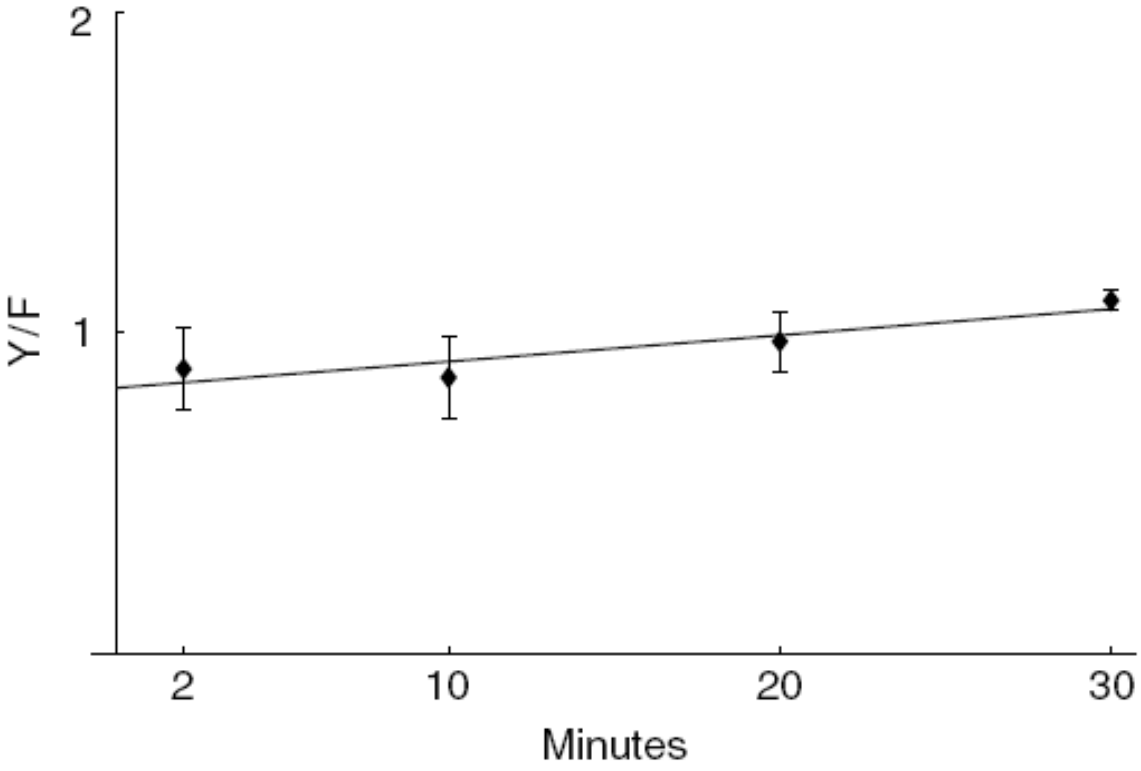


Figure 2.5: The recovered yield of RNA is proportional to the fraction of disrupted cells. The recovered yield of RNA from six independent experiments, see Table 2.2, were divided by their respective fraction disrupted and normalized by the total average RNA yield per fraction of disrupted cells. The means and standard deviations are plotted along with the best fit line. The best fit slope is 0.008 and the intercept is 0.83. These data indicate that the yield of mRNA is directly proportional to the fraction disrupted.

Table 2.3: Data summarizing the percentage of cells disrupted as a function of disruption time and as a function of the cell cycle. The data are plotted in Figure 2.6.

Cell Cycle Phase	Time (min)			
	2	**10**	**20**	**30**
G_1 Release	0.118	0.357	0.478	0.929
G_1	0.192	0.296	0.604	0.965
G_1-S trans	0.277	0.513	0.652	0.905
S	0.265	0.541	0.832	0.973
S-G_2 trans	0.384	0.688	0.929	0.957
M	0.462	0.708	0.947	0.945
M-G_1 trans	0.214	0.441	0.611	0.904

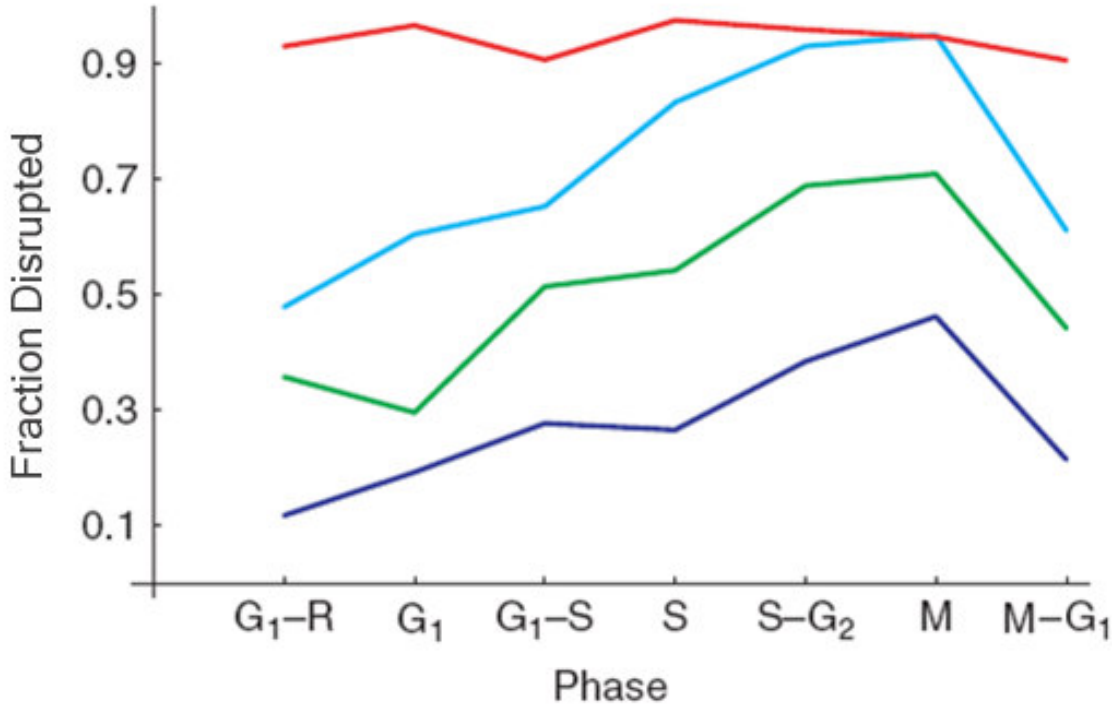


Figure 2.6: Cell disruption as a function of the cell cycle and agitation time. For a given agitation time less than 30 minutes, cell disruptability grows throughout the cell cycle and is maximized in M-phase. Thirty minutes agitation is sufficient to achieve uniform disruption. The intervals between cell cycle phases are plotted as equidistant for convenience. The blue curve shows the results for 2 minutes of agitation, the green curve is 10 minutes, the cyan curve is 20 minutes, and the red curve is 30 minutes.

The marked difference in disruptability is consistently and reproducibly seen for agitation times less than 30 minutes. At and beyond 30 minutes of agitation, cells from the different phases of the cell cycle suffer similar amounts of disruption at greater than 90%, in agreement with the batch culture data. Finally, we observe that the phase dependent disruptability is consistent with the calculated nonconstant and decreasing disruption rate implied by $\beta < 1$ as discussed previously. The phase dependent disruptability would explain why the data are better fit by the Weibull distribution with shape parameter $\beta=0.773$ than with an exponential distribution.

Conclusions

Understanding how the expression of arbitrary genes are related to and regulated by the cell cycle is a central question being pursued vigorously by systems biologists and has direct implications for diseases like cancer. The most important conclusion that can be drawn from these data is that disrupted populations, unless they are completely broken, can be biased depending on the length and degree of agitation. Supernatants from disrupted batch cultures of yeast will be biased at least with respect to the cell cycle unless the cells are agitated long enough. This bias could in fact be exploited to an advantage for the selective harvest of bio-molecules. The results of this study indicate that for shorter disruption times, cells in the M- phase of the cell cycle are most likely to be disrupted, promoting selective harvest of biomolecules from these cells.

It is not yet clear how the expression of genes are linked quantitatively to the cell cycle; however, it is clear that experimental designs will impact the results and hence the

interpretation of the data. Thus, every effort must be made to understand and properly control quantitative assays of gene expression, protein, lipid, or metabolite levels. We were not able to find data of the type reported here in the current and past literature nor were we able to find it in manufacturer's technical reports.

This study demonstrates that for a popular commercial method, 30 minutes of mechanical agitation is sufficient to disrupt cells uniformly with respect to the cell cycle and reproducibly with a coefficient of variation that is less than 4% and such that greater than 90% of the total cell volume is disrupted. Additionally, the results show that this period of time does not affect the yield nor the quality of recovered total RNA. This work also demonstrates that the Weibull cumulative distribution function provides a reasonable model to describe the process of yeast cell disruption by mechanical agitation.

References

1. Blank L., Kuepfer L., and Sauer U. 2005. Large scale analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biology* **6**:R49.

2. Boczko E., Cooper T., Gedeon T., Mischaikow K., Murdock D., Pratap S., and Wells K. 2005. Structure theorems and the dynamics of nitrogen catabolite repression in yeast. *Proc. Natl. Acad. Sci.* **102**:5647-5652.

3. Chen C., Calzone L., Ccikasz-Nagy A., Cross R., Novak B., and Tyson J. 2004. Integrative analysis of cell cycle control in budding yeast. *Mol. Biol. Cell.* **15**:3841-3862.

4. Cross F., Archambault V., Miller M., and Klovstad M. 2002. Testing a mathemtical model of the yeast cell cycle. *Mol. Biol. Cell* **13**:52-70.

5. Dheda K., Huggit F., Bustin S., Johnson M., Hook G., and Zumla A. 2004. Validation of houskeeping genes for normailzing RNA gene expression in real time PCR. *BioTechniques.* **37**:112-114.

6. Douglas C., Foor F., Marrinan J. et al. 1994. The *Saccharomyces cerevisiae* FKS1 gene encodes an integral membrane protein that is a subunit of 1,3-β-D-glucan synthase. *Proc. Natl. Acad. Sci*. **91**:12907-12911.

7. Gedeon T., Boczko E., and Mischaikow K. 2007. Dynamics of a simple regulatory switch. *Journal of Mathematical Biology*. **55**:679-719.

8. Hartman C. and Delgado A. 2004. Numerical simulation of the mechanics of a yeast cell under high hydrostatic pressure. *J. Biomechanics.* **37**:977-987.

9. Hartwell L. and Unger M. 1977. Unequal division in *Saccharomyces cerevisiae* and its implications for the control for the cell division. *J. Cell. Biol.* **75**:422-435.

10. Hartwell L. 1974. *Saccharomyces cerevisiae* cell cycle. *Bacteriol. Rev*. **38**:164-198.

11. Kanno J., Aisiki K., Igarashi K., Nakatsu N., Ono A., Kodama Y., and Nagao T. 2006. Per-cell normalization method for mRNA measurement by quantitative PCR and microarrays. *BMC Genomics.* **7**:64.

12. Keulers M., Sazuki T., Satroutdinov A., and Kuriyama H. 1996. Autonomous metabolic oscillations in continuous culture of *Saccharomyces cerevisiae* grown on ethanol. *FEMS Micro. Lett.* **142**:253-258.

13. Kipper F., and Lloyd D. 1995. The aniline blue flourochome specifically stains the septum of both live and fixed *Schizosaccheromyces pombe* cells. *FEMS Microbiol. Lett.* **132**:215-219.

14. Klis F., Mol P., Hellingwerf K., and Brul S. 2002. Dynamics of cell wall structure in *Saccharomyces cerevisiae*. *FEMS Microbiology Reviews.* **26**:239-256.

15. de Kok J., Roelofs R., Giesendorf B., Pennings J., Waas E., Feuth T., Swinkels D., and Span P. 2005. Normalization of gene expression measurements in tumor tissues:comparison of 13 endogenous control genes. *Laboratory Investigation.* **85**:154-159.

16. Lloyd, D. and Murray, D. 2005. Ultradian metronome: timekeeper for orchestration of cellular coherence. *T.I.B.S.* **30**:333-337.

17. Yeger-Lotem, E., Sattah, S., Kashtan, N., Itzkovitz, S., Milo, R, Pinter, R., Alon, U., and Margalit, H. 2004. Network motifs in integrated Cellular networks of transcription-regulated and protein-protein interaction. *Proc. Natl. Acad. Sci.* **101**:5934-5939.

18. Murray, D., Klevecz, R., and Lloyd, D. 2003. Generation and maintenance of synchrony in *Saccharomyces cerevisiae* continuous culture. *Exp. Cell Res*. **287**:10-15.

19. Murthy, D., Xie, M., and Jiang, R. 2004. Weibull Models. ISBN 0-471-36092-9, John Wiley & Sons.

20. Plesset, J., Ludwig, R., Cox, B., and McLaughlin, C. 1987. Effect of cell cycle position on thermotolerance in *Saccharomyces cerevisiae. J. Bacteriol*. **169**:779-784.

21. Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M. and Ragg, T. 2006. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Molecular Biology* **7**:3.

22. Schuster, M., Wasserbauer, E., Ortner, C., Graumann, K., Jungbauer, A., Hammerschmid, F., and Werner, G. 2000. Short cut of protein purification by integration of cell disruption and affinity extraction. *Bioseparation.* **9**:59-67.

23. Smits, G., van den Ende, H., and Klis, F. 2001. Differential regulation of cell wall biogenesis during growth and development in yeast. *Microbiology*. **147**:781-794.

24. Tu, M., Kudlicki, A., Rowicka, M., and McKnight, S. 2005. Logic of the yeast metabolic cycle: temporal compartmentalization of cellular process. *Science.* **310**:1152-1158.

25. Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. 2002.Accurate normalization of real time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Geneome Biology*. **3**:1-0034.11.

26. Walther, T., Reinsch, W., Grose, A., Ostermann, K., Deutch, A., and Bley, T. 2004. Mathematical modeling of regulatory mechanisms in yeast colony development. *J. Theor. Biol*. **229**:327-338.

27. Woldringh, C., Huls, P., and Vischer, N. 2003. Volume growth of daughter and parent cells during the cell cycle of *Saccharomyces cerevisiae* α as determined by image cytometry. *J. Bacteriology.* **175**:3174-3181.

28. Yang I. 2006. Use of external controls in microarray experiments. *Methods in Enzymology.* **411**:50-63.

29. Yi, T., Kitano, H., and Simon, M. 2003. A quantitative characterization of the yeast heterotrimeric G protein cycle. *Proc. Natl. Acad. Sci*. **100**:10764-10769.

CHAPTER III

EXTENDING CELL CYCLE SYNCHRONY THROUGH AN ANALYSIS OF VOLUME GROWTH AND VOLUME FILTRATION USING A STRUCTURED LESLIE MODEL

Introduction

Unlike the simple volume symmetric division of *E. coli*[38], an initially synchronous culture of budding yeast becomes asynchronous and stationary very rapidly. While stable, synchronous, autonomous oscillations have been observed and are of enormous interest, they do not occur generically and are far from understood[4,27,31,40]. Population synchrony is often monitored by tracking the percent of a culture that is budded as a function of time. The physiological factors influencing the rapid decay of cell cycle synchrony in budding yeast were investigated three decades ago. It was found that new daughter cells take longer to traverse the mitotic cycle than their mothers because of a volume asymmetry at division. That is, daughter cells at the time of division, are smaller in volume than their mothers. Furthermore, as mothers age, they give rise to progressively smaller daughters on average, compounding the problem[48]. Currently there is renewed interest in the physiology of replication in relation to aging and the asymmetric partitioning of biomolecules between mother and daughter cells[1,2,3].

As yeast are now routinely the subject of expression analysis, synchronous growth and division has important and largely unexplored implications for attaching meaning to commonly measured population signals[15,40]. Our interest in developing a model for the volume growth and population synchrony of budding yeast stems from our previous work

on an ostensibly simple gene regulatory circuit involved in nitrogen catabolite repression (NCR). An analysis of a minimal model of the NCR-circuit indicates that the components of the system oscillate in phase with the cell cycle[6,7]. In order to understand how a cellular oscillation is observable at the population level, and further, how one could engineer an experiment to convincingly demonstrate periodic oscillation at the cellular level from a population measurement, we undertook the development of the structured population model of yeast growth and division to be described in this chapter.

The central observations of this study can be summarized as follows. Theoretically, volume symmetric division leads to persistent synchrony. Each strain of budding yeast has a characteristic mean daughter-mother division volume asymmetry, some more and some less pronounced. Parenthetically, this asymmetry is an inversely proportional function of growth rate[30]. As the asymmetry between mother and daughter division volume increases, synchrony decays in a predictable way. For a given strain of yeast growing exponentially in a bioreactor, continuously filtering out the smallest and largest cells extends the synchrony of the apparently asymmetric system. With judicious choices of filtration cutoff volumes, synchrony can be extended by an order of magnitude. The filtration process can be conceived of as a means to restore partial symmetry. While it is true that continuous filtration will skew the population of cells under observation, it can be accomplished without inducing a generic stress response within the yeast. This trade off may be useful for certain experiments.

The cell cycle synchrony of a population of yeast, its persistence, decay, and control are essentially and ontologically dynamical systems phenomena. There is a long and fruitful

history associated with the modeling of population growth. Budding yeast and their mitotic cell cycle continue to be an interesting and important area of mathematical cell biology. We make no formal attempt to review this enormous literature but restrict our attention to those models which we are aware that have dealt with volume growth and the effects of a mixed population of cells growing with potentially different growth rates. The mixed mother-daughter model was developed based on the mathematical results of branching processes to explain the variations in the G_1 phase of the cell cycle[17,28,36]. This model was used to derive a stationary distribution of mothers and daughters as a function of the cell cycle. A model developed by Hartwell and Unger in 1977 and expanded on by Lord and Wheals in 1980, considered the properties of an asynchronous population growing exponentially[18,30]. A central result of their pioneering work was to derive a formula for the replicative age distribution at stationarity that depends on only two parameters: the culture growth rate and the parental doubling time. The formulas and analysis derived by Lord and Wheals have continued to underpin current models of cell cycle dynamics and division[11]. An admitted limitation of their work however is that it explicitly assumes that the growth rates among the age classes are the same. Their paper presented compelling evidence to support this claim. There is also a wealth of evidence to the contrary and evidence that older mothers grow larger with each division[14,48]. Age structured models that take into account this finer, but important, level of detail were proposed and utilized to analyze population signals of a critical protein in search of the still elusive link between size control and division[25,43].

Population balance models that extend that of Hartwell and Unger have been proposed to explore the links between metabolism and the cell cycle during asynchronous as well as

synchronous growth. These models were extensively reviewed by Bellgardt in 1994[4]. Recently, sophisticated population balance models have been constructed that take into account the mass changes that accompany growth and division and that can vary among distinct age classes[9]. The Leslie model presented here is a discrete version of the continuous population balance model, although the focus is explicitly on volume as opposed to mass. The obvious advantage of this class of model is that it naturally allows variations among age classes since they are explicitly represented. An important reason for utilizing and exploring a volume and age structured model is that it captures the effects that influence synchrony. Since it is a dynamical systems model, it can be used to directly examine the dynamical phenomena of synchrony and the effects of filtration as a control mechanism, which is the goal of this study.

There is a long history of elutriation as a means of preparing and examining yeast sub-populations in the biological literature[44]. There is also a long history of filtration and sedimentation as a means to separate and control the growth of micro-organisms in the chemical engineering literature[12,45,47]. These two literatures are now converging as systems biology has hit its stride and seeks to leverage every available technology to examine and understand the physiology of networks. As described in this chapter, the main result of the modeling work suggests that continuous volume filtration can maintain the synchrony of an initially synchronous population for 20 to 30 cycles: An order of magnitude improvement. This theoretical result can be put into practice utilizing current microfluidic techniques at every population scale of investigation from the nanophysiometer up to the bioreactor.

The Leslie Model

In a culture of budding yeast, the mitotic cell cycles of distinct cells need not be in phase with each other. We want to model the dynamics of the mitotic cell cycles of a population of budding yeast growing in a bioreactor. A description of the dynamics requires a model describing the rate at which individuals progress through the mitotic cell cycle. The vital rates correspond to growth, division, aging, and death. We describe the vital rates through a consideration of two variables, cell volume and replicative age, with the aid of a Leslie matrix. Leslie models are an important and well studied class of structured population models. Structured population models are commonly used to describe the life cycle of an organism or process. A comprehensive review of their mathematical properties is presented by Caswell[10]. While we wish to highlight certain aspects of the model for its utility, we in no way want to obscure or jeopardize the biological punchline: Continuous volume filtration can extend cell cycle synchrony. A heuristic understanding of our model can be obtained without recourse to equations through the process flow diagram in Figure 3.1. Figure 3.1 is analogous, but not identical, to those described by Vanoni in 1983 and Hatzis in 2006[19,43].

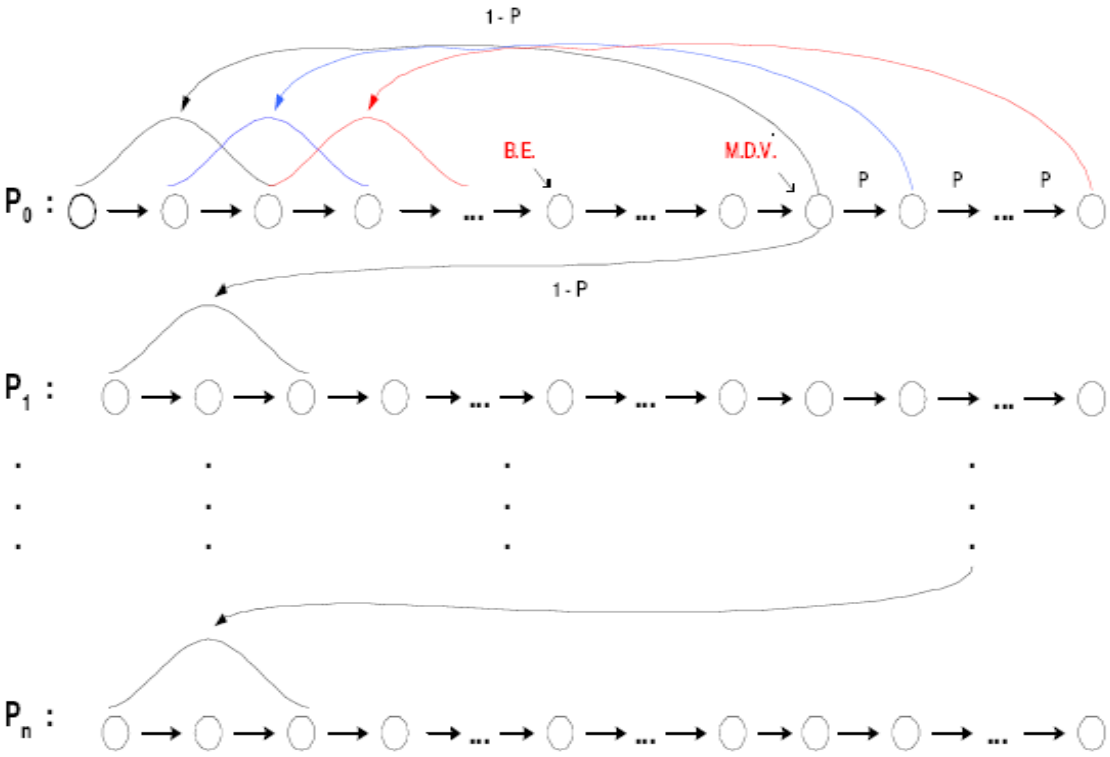


Figure 3.1: Graphical representation of the Leslie process model that involves growth within each age class, division with concomitant birth of new daughters, and replicative aging of mothers. Each age class, P_k, consists of a series of volume intervals, represented as the open circles. The volume intervals within each age class can be annotated with milestones and cell cycle phases. The milestones of bud emergence (B.E.) and mean division volume (M.D.V) are shown. The arrows between volume interval indicate the processes of growth and division. The division process can produce a single volume or a distribution of volumes as indicated by the bell shaped curves.

Variables

The model is organized around two variables, replicative cell age and the volume of an individual yeast cell. Each variable is described below.

1. *Replicative cell age*. As a yeast cell buds during the mitotic cell cycle, a chitinous bud scar is permanently formed on the mother cell. The bud scars can be visualized with calcoflour white staining, and like the rings of a tree, can be used to determine a replicative age[33]. Each generation can be quantitatively identified with the equivalence class of those yeast that carry precisely the same number of bud scars. Traditionally, generations, or bud scar equivalence classes, have been denoted by $P_0,P_1,P_2\ldots,P_k,\ldots P_n$. Replicative age has been identified as a variable that directly impacts synchrony[48]. Replicative age is a discrete variable that we will index by *k*, the number of bud scars.

2. *The volume of an individual yeast cell*. Cell volume has been observed to increase monotonically with time until division, within a given age class, and thus is often used as a proxy for progression through the mitotic cell cycle. The volume of a budded cell is taken as the total volume of both mother cell and the bud until division, at which point they become distinct. The results of this chapter confirm that volume is intimately connected with synchrony. Volume is consistently expressed in units of cubic microns throughout this chapter.

Volume and Intervals of Time

Yeast cells of a given replicative age *k*, are observed to grow in volume between well defined limits. The minimum and maximum volumes naturally delimit and define

intervals, $\boldsymbol{I}(k) := [\underline{V_k}, \overline{V_k}]$. We consider the temporal evolution of the system at a sequence of equally spaced times, $t_s := t_o + s\Delta t$. The volume intervals, $\boldsymbol{I}(k)$, are partitioned into subintervals $I(i,k) := [V(i,k), V(i+1,k)] \subset I(k)$, with $I(k) = \cup_i I(i,k)$, $i=0,1,\ldots n_k$, where $V(0,k) := \underline{V_k}$, and $V(n_k+1,k) := \overline{V_k}$ and n_k is the maximal number of age classes.

The partitions are chosen according to the growth law within each age class, such that any cell with volume in the interval $I(i,k)$ now, would have a volume in $I(i+1,k)$, precisely Δt later. The unit of time is minutes, and we have taken $\Delta t=1$ throughout this chapter. The state of the yeast population at time t_s is described by a vector,

$\rho(i,k)(t_s)$:= number of cells of generation k with volume $v \in I(i,k)$.

Each of the $\rho(i,k)(t_s)$ cells living in $I(i,k)$ at time t_s are faced with the following possibilities:

1. The cell dies
2. The volume of the cell increases
3. The cell divides

The details of each of these possibilities are described below.

Cell Death

The probability of cell death is denoted by $d_{i,k}$. Mortality curves have been measured for several strains of yeast under a variety of conditions[14,32,35]. These data can be used to determine an age class specific death rate. Some studies observe that the death rate on average amounts to 10^{-10}/*cell generation*[30].

Volume Growth

The probability of growth is denoted by $g_{i,k}$, and the fraction of cells that survive and grow is $\kappa_{i,k} := g_{i,k}(1-d_{i,k})$. Volume growth has been measured and is generally considered to increase exponentially with time, which we assume for all of the experiments and analyses in this chapter. The age class specific growth rate is denoted λ_k. The volume intervals are conveniently described by

$$I(0,k) = [\underline{V_k}, \underline{V_k}\, e^{\lambda_k \Delta t}]$$

$$I(i,k) = [V(i,k), V(i,k)\, e^{\lambda_k \Delta t}] \qquad \text{Equation (3.1)}$$

$$I(n_k,k) = [V(n_k,k), \overline{V_k}]$$

Cell Division

All cells do not divide precisely at the same volume. The probability that division occurs is denoted $c_{i,k} := 1 - \kappa_{i,k}$. The importance of including *sloppy size control* in models of growth and division has been discussed in the literature [48]. We have implemented a variety of distributions. Two of the most natural are a Poisson process[37], modeling division as time to failure, and a Brownian process using a normal distribution. As will be described in the results section, this choice makes little or no qualitative difference. The mean of $c_{i,k}$, for fixed k, is referred to as the mean division volume and denoted as k-MDV.

We assume that the division of a cell of volume v in age class P_k results in a cell of age class P_0 with volume v' and a cell of age class P_{k+1} with volume v''. Furthermore, $v = v' + v''$. We sometimes denote the division process as $P_k \rightarrow P_{k+1}$. It has been experimentally observed that after a cell has budded, the ensuing volume growth is concentrated almost

entirely in the bud[48]. This implies that there is a conditional probability distribution for v' that depends on the size and age of the mother. Let $\mu_{i,j,k}$ be the probability that after a cell division, $P_k \rightarrow P_{k+1}$, we get a cell of age class P_0 with volume in *I(i,0)* from a dividing cell in *I(j,k)*. The mean of $\mu_{i,j,k}$, for fixed k, is referred to as the mean emergent daughter volume and denoted as k-MEDV. Let, $v_{i,j,k}$, represent the probability that a parent cell of volume *I(j,k+1)* emerges from a division in *I(i,k)*. The mean emergent parent volume is denoted as k-MEPV. Generally, the distribution of division volumes has been observed to be normal [21,46].

Given these definitions, we can present the projection formula that updates the population in time.

$$\rho(l,0)(t_{s+1}) = \kappa_{l-1,0}\rho(l-1,0)(t_s) + \sum_{i,k} \mu_{l,i,k} c_{i,k} \rho(i,k)(t_s) \qquad \text{Equation (3.2)}$$

$$\rho(l,m)(t_{s+1}) = \kappa_{l-1,m}\rho(l-1,m)(t_s) + \sum_{i} v_{l,i,m} c_{i,m} \rho(i,m-1)(t_s); m > 0 \quad \text{Equation (3.3)}$$

The first summand in each equation represents the volume growth contribution while the second summation term represents the density coming from division. The term $c_{i,k}\rho(i,k)(t_s)$ represents the fraction of dividing cells in volume interval *I(i,k)* and $\mu_{l,i,k} c_{i,k}\rho(i,k)(t_s)$ is the fraction of those that end up in the volume interval *I(l,0)*. The first equation represents daughters and is distinguished because every division results in a daughter. In the higher age classes, $m > 0$, density from division arrives from only one source, namely the age class P_{m-1}.

Milestones

The parameters of the model that we have described in the previous three subsections, such as $\underline{V_k}$, λ_k, $d_{i,k}$, $g_{i,k}$, k-MDV, k-*M*EDV, and k-MEPV, are experimentally measurable quantities associated to a particular strain of yeast that often depend on growth conditions. We refer to these parameters as general volume milestones.

An experimentally important measure of cell cycle synchrony is the percent of cells in the culture that are budded, also known as the bud index. This quantity can be computed from $\rho(i,k)(t_s)$, given an age class dependent, bud emergence cumulative distribution function, $B_{i,k}$. That is, $0 \leq B_{i,k} \leq 1$, is a monotonically increasing function of i, for each k, and describes the probability that the cells in $I(i,k)$ are budded. The function is monotone because once a cell has budded, it remains that way until it divides. The mean of the bud emergence distribution for fixed k, is denoted as k—BE. The bud index at time t_s is the normalized inner product:

$$BI(t_s) = \frac{\sum_{i,k} \rho(i,k)(t_s) B_{i,k}}{\sum_{i,k} \rho(i,k)(t_s)} \qquad \text{Equation (3.4)}$$

Careful measurements of bud emergence have been made and reveal that the cumulative distribution function of buddedness relative to volume is derived from an underlying normal distribution[46].

Bud emergence is also a hallmark at the end of the G_1 phase and the beginning of the S-phase of the cell cycle. Likewise, other cell cycle phases can be demarcated within each age class. This annotation enhances the power and utility of the Leslie model. As discussed above, the general outline of the process flow in the Leslie model is similar to

that previously outlined by Hatzis and Vanoni, although there are some qualitative differences[19,43]. In their process, it is tacitly assumed that the *k*-MEDV forms a monotonically increasing series as a function of *k*. We make no such assumptions. The model can be implemented with measured or arbitrary values. In fact, the data described by Woldringh et al. indicate that *k*-MEDV forms a monotonically decreasing series as a function of age class *k*[48].

We have utilized the volume milestones of two strains in this work. To the best of our knowledge, the most comprehensive set of milestones have been measured in the diploid strain X2180. For this strain, the model was parameterized with yeast physiology data derived from experiments performed over the past four decades[3,9,18,22,23,35,36,46,48]. Among these, the data of Woldringh et al. are particularly comprehensive and well suited for our modeling[48]. The volume milestones and their description are summarized in Table 3.1.

Table 3.1: Volume milestones and growth parameters for the strain X2180.

Age(k)	$\underline{V_k}$	$\overline{V_k}$	λ_k	BE	MDV	MEDV
0	14	75	0.0062	38.5	70.7	28.5
1	40	85	0.0061	46.8	75	24.4
2	48	87	0.0044	56.1	82.4	24.2
3	56	94	0.0047	63.9	88.9	22.3
4-13	64	125	0.0047	76.3	95	22.2

Additionally, we have utilized the haploid, α-factor sensitive strain LHY3865, which is much larger than X2180, and for which we have measured many, but not all, of the volume milestones, see Table 3.2.

Table 3.2: Volume milestones and growth parameters for the strain LHY3865.

Age(k)	$\underline{V}_k$	$\overline{V_k}$	λ_k	BE	MDV	MEDV
0	30.0	105.0	0.0054	59.0	98.0	46.0
1	45.0	105.0	0.0049	69.5	97.5	43.0
2	53.0	104.0	0.0049	68.9	96.6	36.5
3	60.0	115.0	0.0049	78.8	110.4	36.5
4	73.0	140.0	0.0049	95.7	134.1	36.5
5	97.0	185.0	0.0049	155.0	179.1	36.5
6-13	129.0	190.0	0.0049	155.0	179.1	36.5

The behavior of the model can be investigated with arbitrary parameters. For instance, we were interested to examine how the mother to daughter volume asymmetry impacts synchrony, with all other factors being equal. For this part of the study, we used a data set that has no analog in nature that we are aware of, but was constructed to coincide with realistic volume values and exponential growth rates.

Table 3.3: Growth parameters for studying the impact of the daughter to mother volume asymmetry on the decay of synchrony.

Age(k)	$\underline{V}_k$	$\overline{V_k}$	λ_k	BE	MDV	MEDV
0-13	40	110	0.0047	60	100	50

Initial Conditions

In order to compare the dynamics of our model with data, we considered several natural initial conditions. For instance, most experiments that follow the bud index oscillations start from an initially synchronized population of cells. Historically, various experimental methods have been used to synchronize yeast. These include metabolic starvation, elutriation, and pheromone blocks[44]. Perhaps the most common of these is the use of mating pheromones like α-factor that arrest cells in G_1 prior to the cdc28 delimited start. Computationally, we created an initial condition to mimic this population of cells by pruning the time invariant population density of each class such that no cells exist outside of the terminal 20% of the G_1 volume intervals prior to the mean bud emergence. The pruned population density was then renormalized. We will refer to this distribution as the α-factor initial condition.

In the late 1960's, Helmstetter had the ingenious insight to create what is now referred to as the *baby machine*[20]. The concept can be made to work with virtually any dividing cells, but was conceived for yeast. Cells are adhered to a membrane and perfused with media. As the cells divide, the daughters fall into a receptacle. The collected P_0 cells can be re-adhered to a fresh membrane and the process iterated, with or without pheromones, limited only by imagination. In this way, one can experimentally create and subsequently analyze coherent populations. Other clever ways of preparing and separating cells also exist[29,34].

With the help of a *baby machine*, we collected coherent P_0 cells and analyzed these cells through a Coulter counter to measure their volume distribution. We refer to such data as a baby initial condition.

Filtration

The main objective of this study was to observe the behavior of a population of yeast undergoing continuous filtration. Here we wish to formally define what we mean by filtration. Figure 3.2 depicts how the process works.

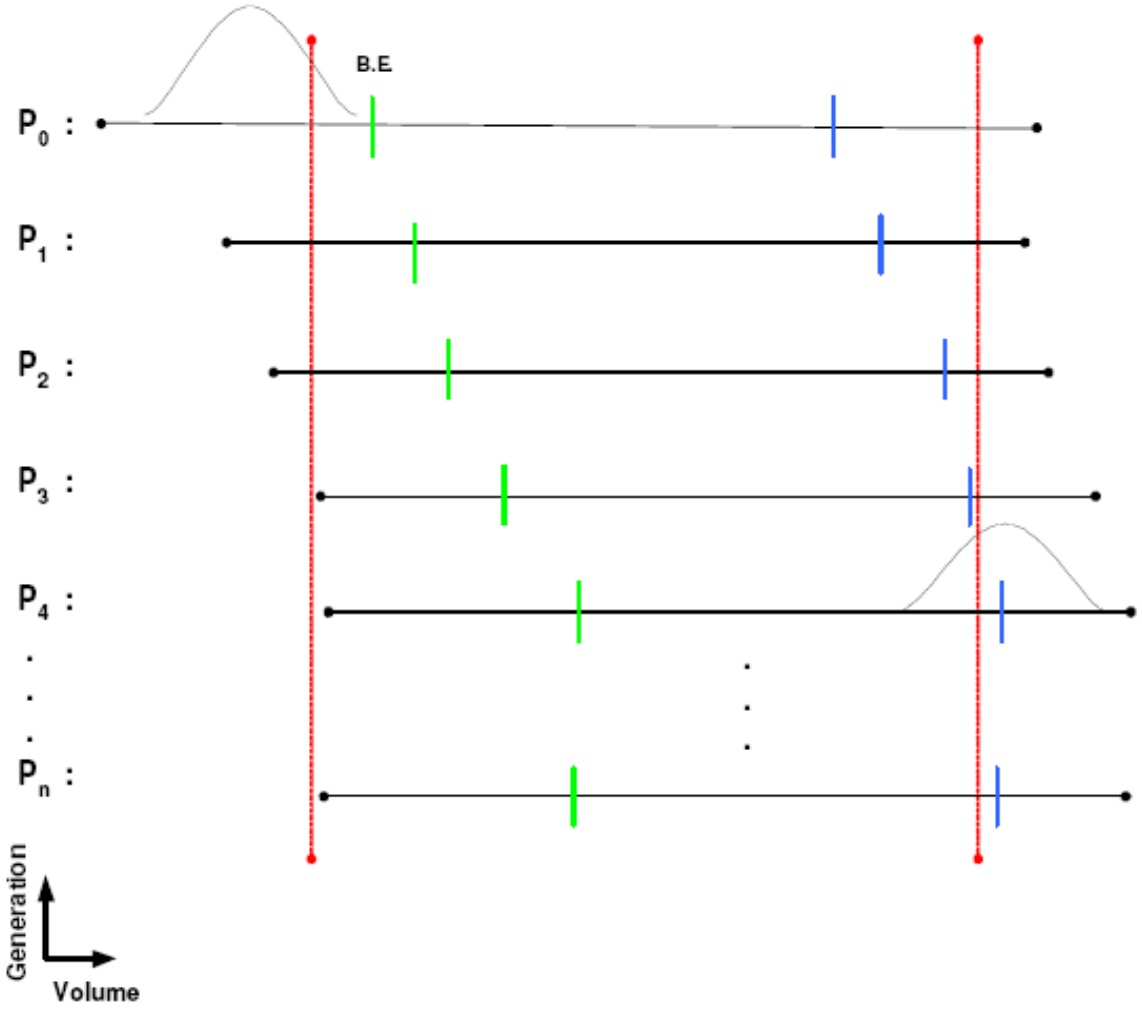


Figure 3.2: Schematic representation of a volume filtration process. The vertical red lines indicate the lower and upper volume filters. The horizontal lines represent the volume grids of each age class ascending in age from the top of the page. All cells whose volume is below the lower cutoff are removed from the system as well as those whose volume is larger than the upper cutoff. The cells with volumes in between the two red stripes remain in the system. Filtration acts to symmetrize the age classes and restore mother to daughter volume parity as best as possible. There is a parallel between the extended synchrony achieved via two stage filtration and that described in Figure 3.4.

Two volumes are specified, V_* and V^*, and together these define a volume interval, $F := (V_*, V^*) \subset U_\kappa I(k)$. In Figure 3.2, the vertical red lines indicate the volumes V_* and V^* and how they intersect the various intervals *I(k).* All cells, regardless of age, whose volume lies outside of *F* are removed from the system at every timestep.

$$V(i,k) \le V_* \quad or \quad V(i,k) \ge V^* \Rightarrow \quad \rho(i,k)(t_s) = 0 \qquad \text{Equation (3.5)}$$

This is intended to mimic what a perfect volume filter might do to a real yeast culture. In engineering practice this would be called a two stage filtration because each of the two defining inequalities would be implemented by a separate filter and the process is performed in series.

Materials and Methods

Yeast cells of *Saccharomyces cerevisiae* strain LHY3865 (mat a- URA, LEU, bar1Δ) were grown in YNB media without ammonia or amino acids and with 100 mg/L leucine, 20 mg/L uracil, 0.2% glutamine, and 2% glucose at 30° C. Batch cultures were grown with agitation at 225 rpm in a New Brunswick Innova 44 orbital incubator/shaker. Continuous cultures were grown in a New Brunswick BioFlow 110 3.0 L bioreactor with a dilution rate of D=0.35 hr^{-1}, air was sparged through the reactor at a rate of 500 mL/min, and the culture was agitated with a Rushton-type impeller run at 225 rpm.

Cell Cycle Synchronization

A 750 mL yeast culture was arrested at a cell density of $\mathbf{OD}_{600}$=0.8 through the addition of $3x10^{-5}$ M α-factor mating pheromone (Sigma # 63591) and was incubated for 3 hours. Cells were subsequently released from arrest by pelleting followed by three washes with fresh pre-conditioned media, free of α-factor, containing 0.1 mg/mL Pronase E (Sigma # P-6911). The pre-conditioned media was prepared by allowing LHY3865 yeast cells to grow within the media for 4 hours at $\mathbf{OD}_{600}$ = 0.4 before being removed by a 0.2 μm filter. The synchronized cells were then resuspended in 1.5 L of preconditioned media and grown in the bioreactor as described above. 0.5 mL samples were taken from the bioreactor at a time interval of 3 minutes and were immediately frozen in 50% glycerol by the addition to an ethanol-dry ice bath. For batch experiments, samples were taken at a time interval of 10 minutes for the first 90 minutes of the experiment and then every 20 minutes for the remainder of the experiment.

Bud Index Analysis

Samples were analyzed using a conventional microscope for bud index. Each data point consisted of more than 100 different analyzed cells. Samples were vortexed briefly and then sonicated for 1 minute prior to analysis to minimize cell clumping in order to ease analysis. 10 μL of each sample was then pipetted onto a glass slide to be analyzed with a microscope. Cells were individually interrogated using multiple focal planes and a 100X objective. Yeast cells were only considered budded if a septum did not separate the mother from the daughter cells.

Cell Volume Measurement and Baby Machine Construction

The baby machine consisted of a nitrocellulose filter supported by a stainless steel mesh sandwiched between two glass funnels approximately 3 inches in diameter[20]. The apparatus was supported by a ring stand and the inlet line was connected to a ½" ID PVC hose which is used to feed pre-conditioned medium. The outlet line was connected to a vacuum source protected by a water trap which was used to collect the harvested babies (daughter yeast cells). Prior to use, a Millepore nitrocellulose membrane was cut to fit the inside diameter of the funnel. This membrane was then coated in poly-d-lysine by filtering through 50 mL of 20 μg/mL poly-d-lysine at a rate of 1 mL/sec. The filter was then rinsed by running 100 mL of sterile DI water through the filter at a rate of 1 mL/sec. A 50 mL culture was then loaded onto the filter and profused through the filter via gravity, at a rate much less than 1 mL/sec. The filter was then inverted and preconditioned media is profused through the filter at a rate of 1mL/sec. The media profused through the filter was collected in the water trap of the vacuum source which was stored on ice so that the daughter yeast cells remain quiescent.

Daughter yeast cells were subjected to volume measurement using a Beckman Multisizer Coulter Counter. The volume distribution of older yeast generations was determined using an alternative method. This method involved time lapse microscopy. Yeast cells were loaded into MatTek dishes (#P35G0-10-C) and imaged over 8 hours (3+ cell cycles). The images were analyzed using cross-sectional area as a proxy for cell volume. The results could be used as approximate milestones for higher generations.

Results

Bud Index Dynamics

We define synchrony as the number of consecutive bud index oscillations whose amplitude is at least 60% of maximum. That is, the amplitude varies between less than 20% budded and greater than 80% budded. The Leslie model qualitatively as well as quantitatively captures the dynamics of two different yeast strains with very different volume milestones. Figure 3.3 shows good agreement between the Leslie model and the experiments described by Woldringh et al. with strain X2180[48]. We have made careful measurements of the bud index oscillations for the α-factor sensitive strain LHY3865, both in batch and in a continuous fermentation. The data shown in Figure 3.3 are typical of those described in the literature over the past 4 decades[8]. The LHY3865 cells are initially synchronized with the mating pheromone α-factor that arrests unbudded cells in G_1. The agreement of fine structural features between the experiment and simulation, such as the breadth at the top of the oscillation, indicates that the model is capturing the essential features of budding yeast volume growth and division. The model predictions tend to fit batch data (top panel of Figure 3.3) more accurately than continuous data (bottom panel of Figure 3.3). This is likely due to the fact that the volume milestones for yeast growth and division were determined from batch experiments.

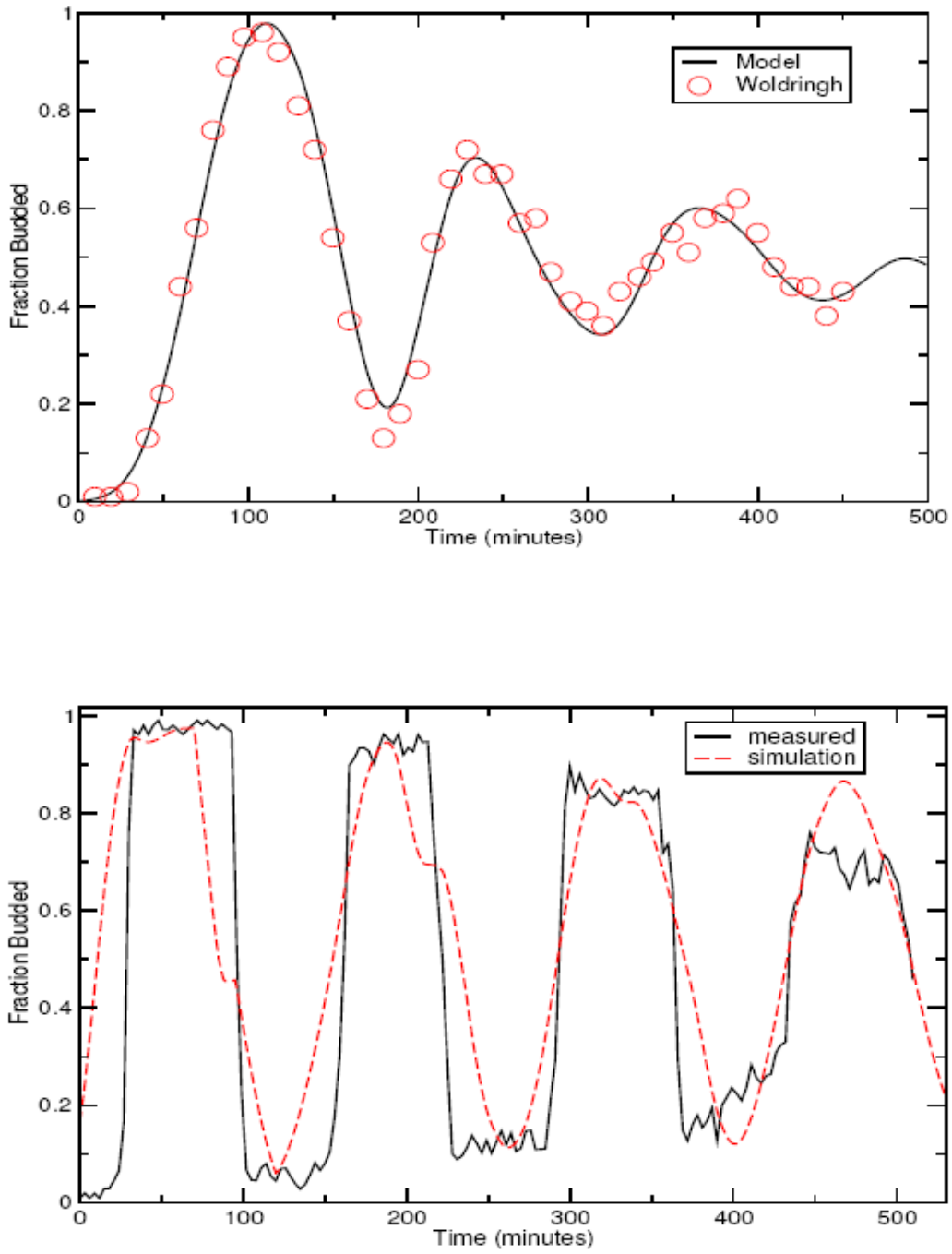


Figure 3.3: Experimental measurements of bud index synchrony and comparison with simulation. The top figure shows that the Leslie model captures the essential bud index oscillations of the X2180 strain. The experimental data are from Woldringh et al.[48]. The initial condition for the experimental data was prepared by elutriation. We were able to reconstruct the initial condition computationally from their data set[48]. The bottom figure shows the good agreement between simulation and experiment for the LHY3865 strain, grown in a continuously operated bioreactor and synchronized with α-factor.

Using the bud index experimental data, we have performed a sensitivity analysis to determine how the individual milestones affect the congruence between model and system dynamics. The results indicate that the milestones of the daughter generation are the most sensitive and the sensitivity decays monotonically with age. How well the model dynamics fit the data is most sensitive to the mean division volume of the daughter generation, followed by the mean bud emergence milestone. In general, a 10% change in the milestones produced less than a 10% change in the overall fit between model dynamics and experimental time series. This indicates that the basic processes of the model robustly capture the dynamical phenomena associated with bud index oscillations.

Stationary Properties

The model parameterized with the X2180 milestones reproduces the measured stationary values within the measured deviations where available, see Table 3.4. The measured quantities were the fractions (F) of daughters (D), parents (P), budded (B) and unbudded (U). It has been observed that a quantitative relationship exists of the form $(1-P(G_1))/\tau_D = \eta$, where $P(G_1)$ is the percentage of cells in the G_1 phase, τ_D is the observed population doubling time, and η is a constant[36]. This expression would be unremarkable save for the fact that η=1.1 hrs was observed over a wide range of growth rates, suggesting some universality. The observed population doubling time is in reality a population weighted average over all the generations and we have computed this quantity from the model using two natural ensemble averages that produce nearly the same value of η=1.2 hrs that is in close agreement with the experimental value for which no standard deviation was reported.

Table 3.4: Comparison of stationary properties of the model with experiment for the X2180 strain. The experimental data are reproduced from Table 1 of Woldringh et al. with the exception of the last two entries that are taken from Slater et al.[36,48]. F(D),F(P),F(B),F(U) are the fractions of daughters, parents, budded and unbudded respectively.

Property	Model Prediction	Experiment
F(D)	61	60.3 +/- 1.8
F(P)	39	39.7 +/- 1.7
F(B)	63	66.9 +/- 4.0
F(U)	37	33.1
F(B)<τ>	1.2	1.1
<F(B)τ>	1.2	1.1

Decay of Synchrony with Division Asymmetry

As described in the introduction, it has been well known that the volume asymmetry between mothers and daughters has a profound effect on the decay of synchrony of initially synchronized populations of budding yeast. Since budding yeast display a bewildering array of strain variation, we felt it legitimate and interesting to ask how the amplitude of the bud index oscillation decays as a function of inherent volume asymmetry between mothers and daughters at division. This volume asymmetry has a constant mean value for each strain of yeast. Essentially, this value is MEPV - MEDV, when it does not vary with *k*. From the bud index curve, we computed the envelope of the oscillation and fit the amplitude decay. As expected from the theory, the decay is exponential[10]. The initial rates of decay are described in the top panel of Figure 3.4, while the number of corresponding synchronous cycles are shown in the bottom panel. The data show that the number of cycles of synchrony declines dramatically with volume asymmetry as expected. When the daughter to mother volume ratio is 80%, the number of synchronous cycles has decayed from infinity to one for the X2180 milestones.

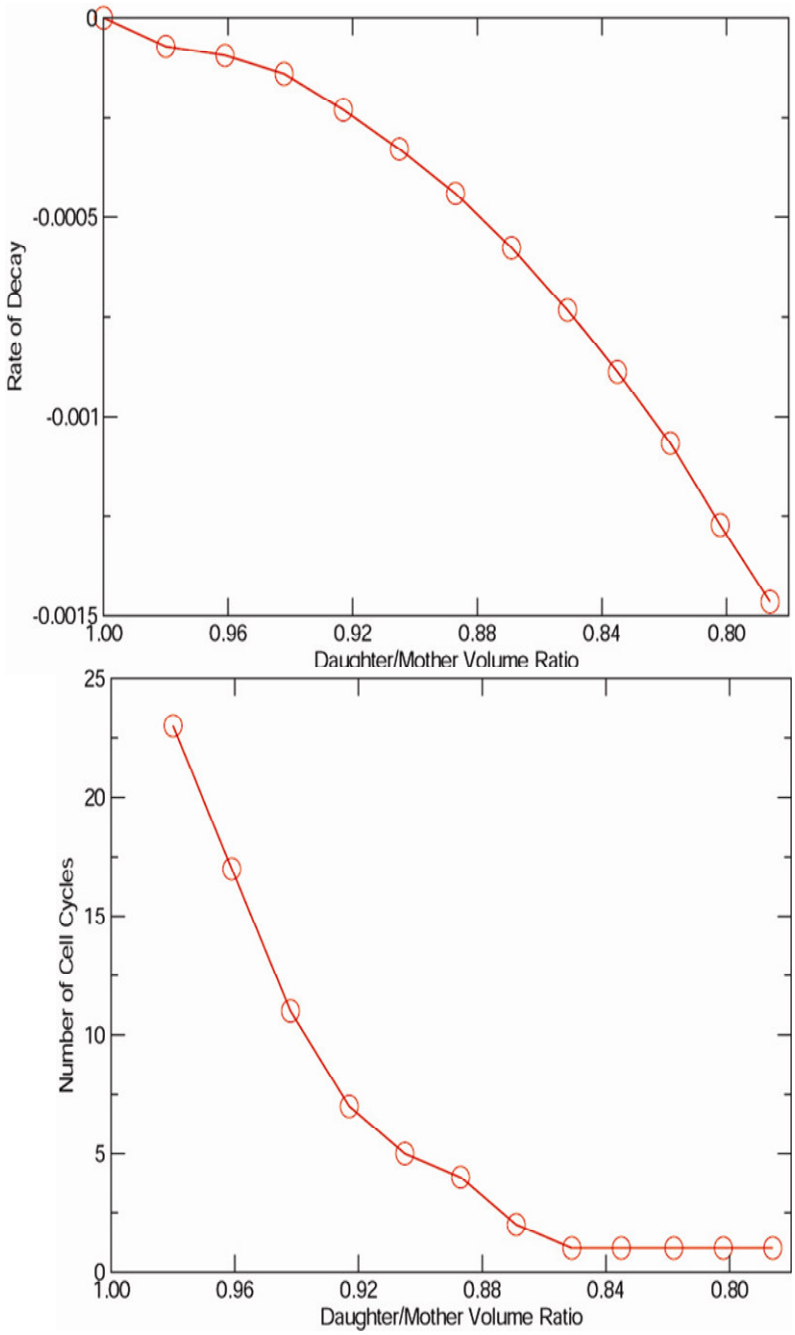


Figure 3.4: The decay of synchrony as a function of the daughter to mother volume ratio. As the asymmetry grows, the synchrony decays. Synchrony is computed from the oscillations of the bud index as a function of time, starting from a population of daughter cells narrowly distributed in G_1. In the completely symmetric case of equal volume division the system is perfectly periodic, but this is a degenerate case. In general, solutions decay exponentially to a unique equilibrium. The top panel describes the initial rate of decay of the amplitude of the bud index oscillation in arbitrary units. The bottom panel shows how the division asymmetry and the exponential decay lead to a rapid decay in the number of synchronous cell cycles that can be observed.

Volume Filtration

We examined two filtration strategies. Figure 3.2 describes the general filtering scheme. In a two stage filtration process, we impose both an upper and a lower volume limit (see Figure 3.2). All cells whose volume lies in between these limits are retained in the system, a bioreactor, and the rest are continuously removed. In the one stage filtration process, there is only a lower volume limit and all cells smaller than this are removed and those that are larger remain.

The main two stage filtration results of this study are presented in Figures 3.5 and 3.6. After inspecting the volume-time diagram constructed in Woldring et al.[48], we conjectured that it would be possible to emulate the symmetry of near equal volume division by filtering out cells that were too small or too large. We reasoned that this would have the abstract effect of making all the age class grids nearly the same. In large part, this hypothesis was conceived from the fact that the data show that a judicious choice of filtration parameters we could extend the synchrony from 1 cell cycle to close to 20 cell cycles in the X2180 strain and from 3 to 30 in the LHY3865 strain. Figure 3.7 shows the bud index profiles associated with several of the filtration parameters that describe the range from no filtering to the best that we have been able to observe at 17 cycles for the X2180 milestones.

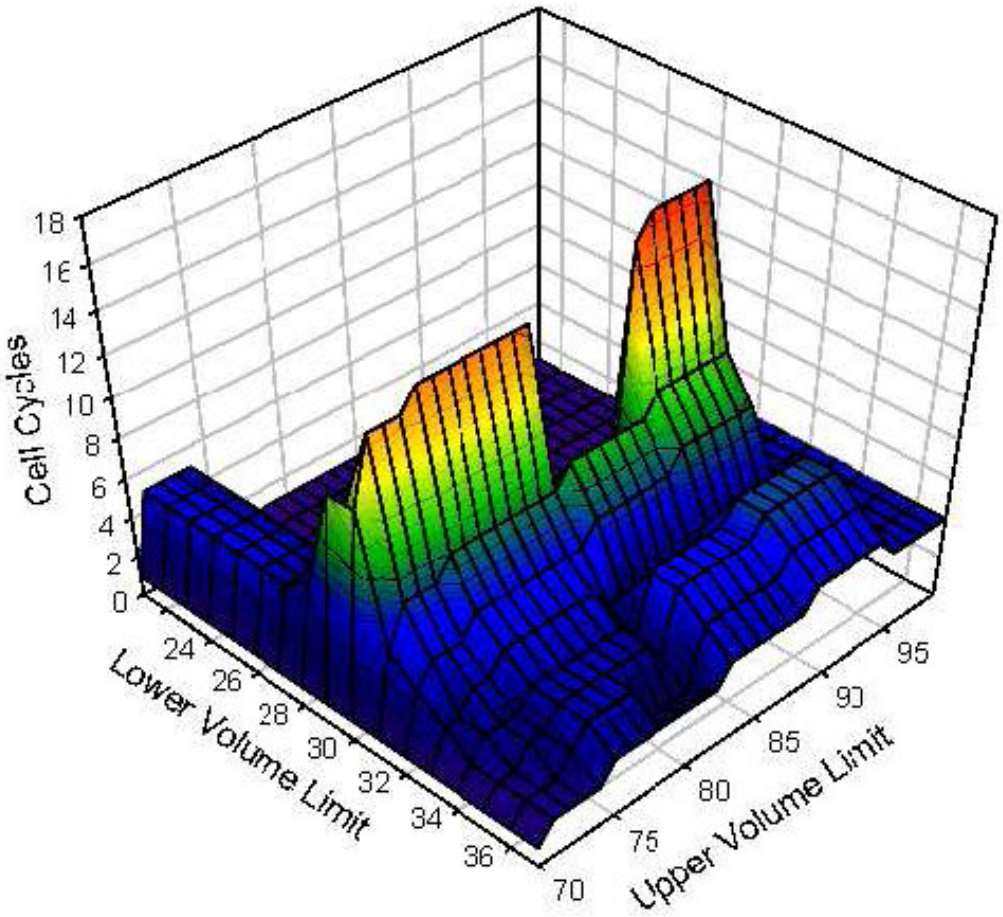


Figure 3.5: Synchrony computed as a function of two stage filtration for strain X2180 milestones. The data indicate that there is an optimal ridge of values that produce extended synchrony. As expected, the optimal values are achieved for combinations that produce age classes that are close in volume extent, as illustrated in Figure 3.2. The volume filtration was accomplished by using two volume cutoff limits. All cells below the minimum volume limit are removed from the system as are all cells that lie above the maximum volume limit. The system is populated by cells that lie in between these two limits, see Figure 3.2. The number of cycles achieved for a particular filtration was computed from bud index profiles, see Figure 3.7 for several examples.

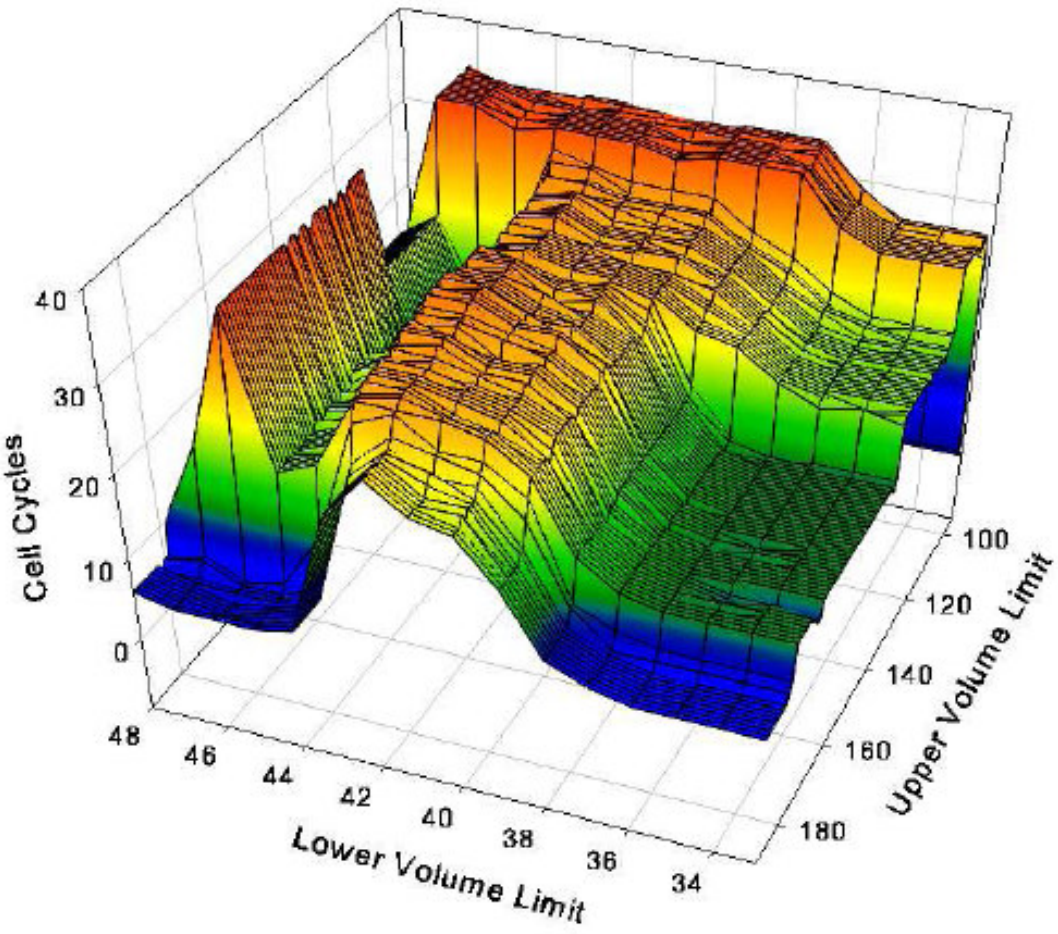


Figure 3.6: Synchrony computed as a function of two stage filtration for strain LHY3865 milestones. The data are both more complicated and show greater achievable synchrony than strain X2180. This strain shows more variation with the upper filtration limit.

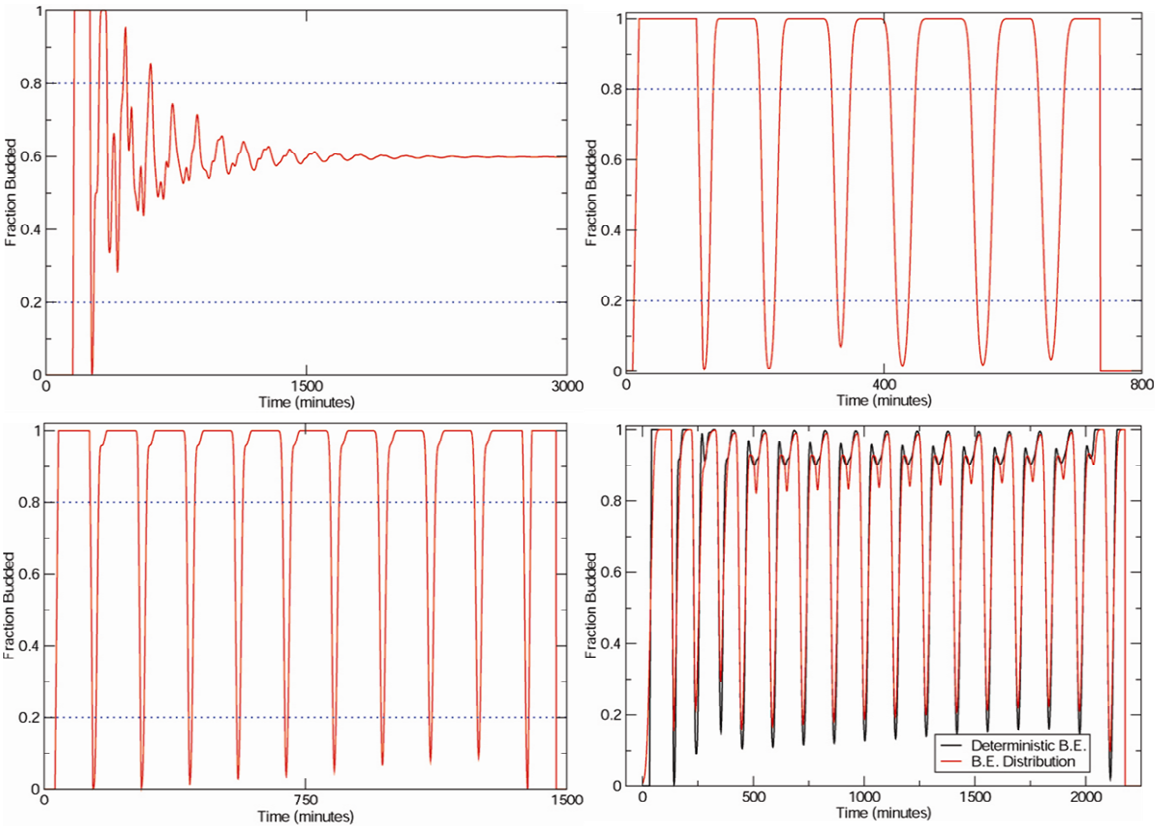


Figure 3.7: Bud index oscillations of strain X2180 milestones for a series of filtration parameters. At top left is the bud index profile for an unfiltered system. Clockwise are filtered systems that achieve 6, 16, and 10 periodic oscillations before the system decays to the trivial equilibrium. These were achieved with the following filtration limits, given in units of cubic microns: [34,91], [30,90], and [30,71] respectively. At bottom right the figure indicates that there is no qualitative difference between using a deterministic bud emergence value versus using a bud emergence distribution.

Equivalently, Figure 3.8 shows the bud index oscillations of the LHY3865 milestones subject to one and two stage filtration. The upper panels show the results of one stage filtration. Figure 3.9 codifies the behavior of the single stage filtration of the LHY3865 milestones. This figure is annotated with the k-mean daughter emergence volumes. The data show that the position of these milestones relative to the filtration volume limit determines the range of extended synchrony. The fact that there exists a broad volume range near the top of the peak ensures that the one stage filtration should in practice produce robust results.

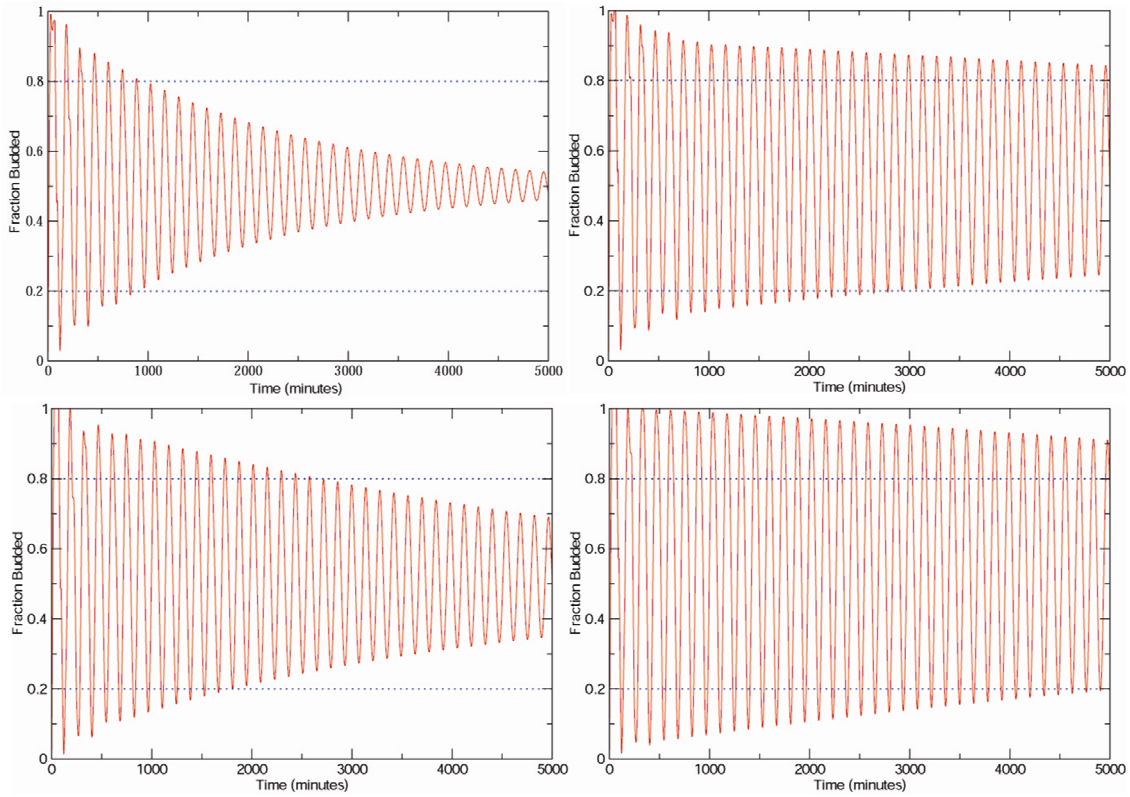


Figure 3.8: Bud index oscillations of strain LHY3865 milestones with one and two stage filtration. The top two panels correspond to single stage filtration while the bottom two panels are for two stage filtration. The two panels on the right represent the best overall synchrony achievable with the respective methods. Only consecutive oscillations that remain completely outside of the dotted lines are counted. For instance the lower left panel is considered to produce 13 consecutive cycles. Clockwise from the top left the corresponding filtration parameters, given in units of cubic microns, are: 39, 42, [37,100] and [33,178].

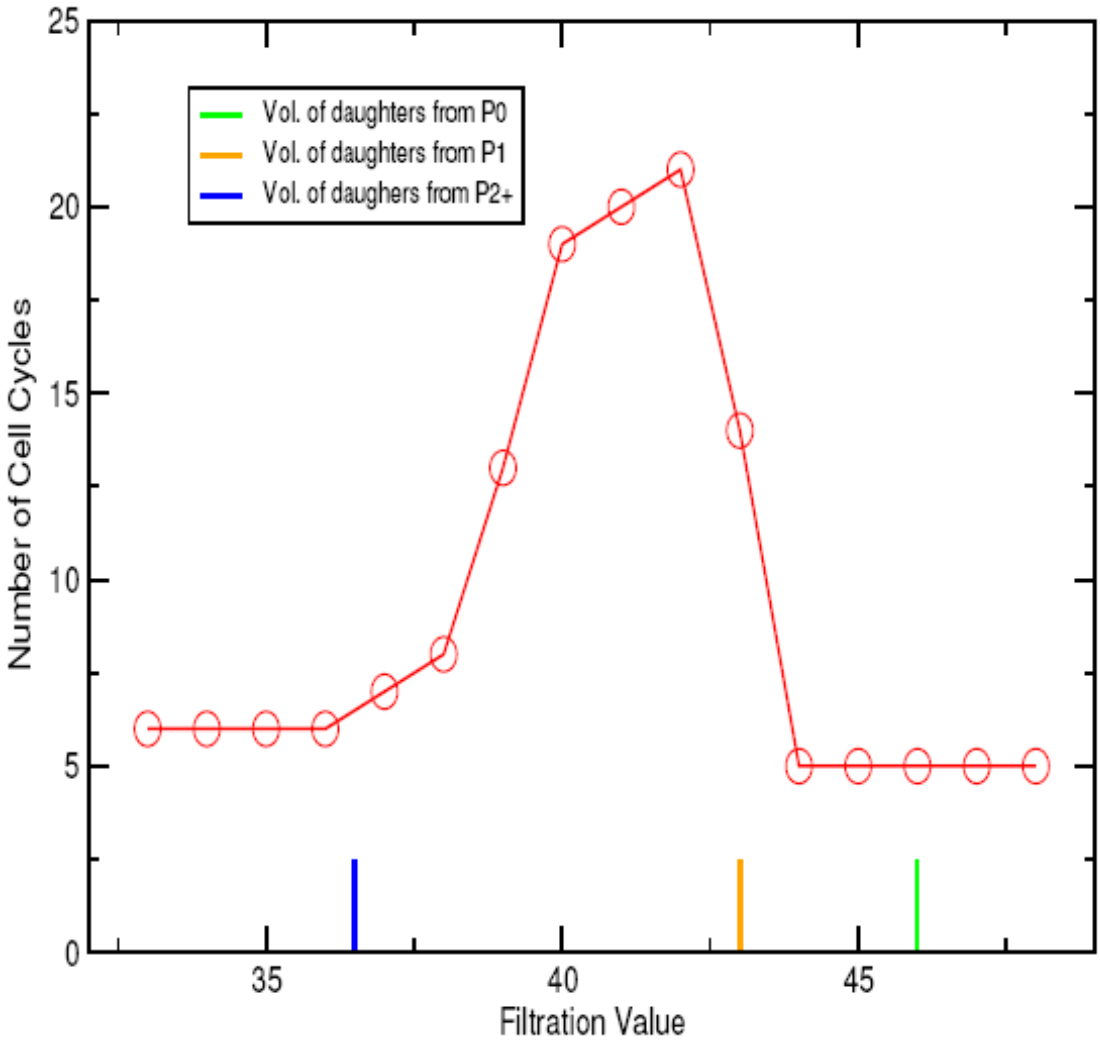


Figure 3.9: One stage filtration results for strain LHY3865 milestones. All cells below the filtration value were continuously removed from the system and synchrony is measured with respect to the number of successive cycles of the bud index oscillation that maintains at least 60% of its total amplitude. Because single one filtration is practically easier to accomplish, we investigated the optimal conditions in order to inform experimental investigation. The short vertical rays on the horizontal axis indicate the *k*-mean daughter volumes emerging from the successive generations, P_k. These are observed to be the key milestones that influence synchrony and its extension via a single stage filtration scheme.

Invariant Density

For the general volume parameters and growth kinetics of budding yeast, like those detailed in Woldringh et al. and summarized in Tables 3.1 and 3.2, the population density generically reaches a unique, non-trivial stationary state[13,48]. This behavior is observed experimentally. As a consequence of the primitivity of the Leslie Matrix and the Perron-Frobenious theorem, the invariant density can be recovered from the model as the L^1-normalized eigenvector corresponding to the unique largest eigenvalue of the matrix. The state of the system at asynchronous exponential growth is described by $\rho(t) = e^{\alpha t} X$, where α is the population growth rate, and *X* is the eigenvector that, when normalized in the L_1 norm, represents the time invariant probability density of observing a yeast cell of a given volume and age. Figure 3.10 describes the properties of the invariant density, *X*, computed for the X2180 milestones. As the figure shows, the invariant population distribution within each age class is smoothed through the use of a distribution of emergent mother and daughter volumes upon division. We examined a family of normal distributions and the qualitative features are insensitive to the specific details, such as the value of σ.

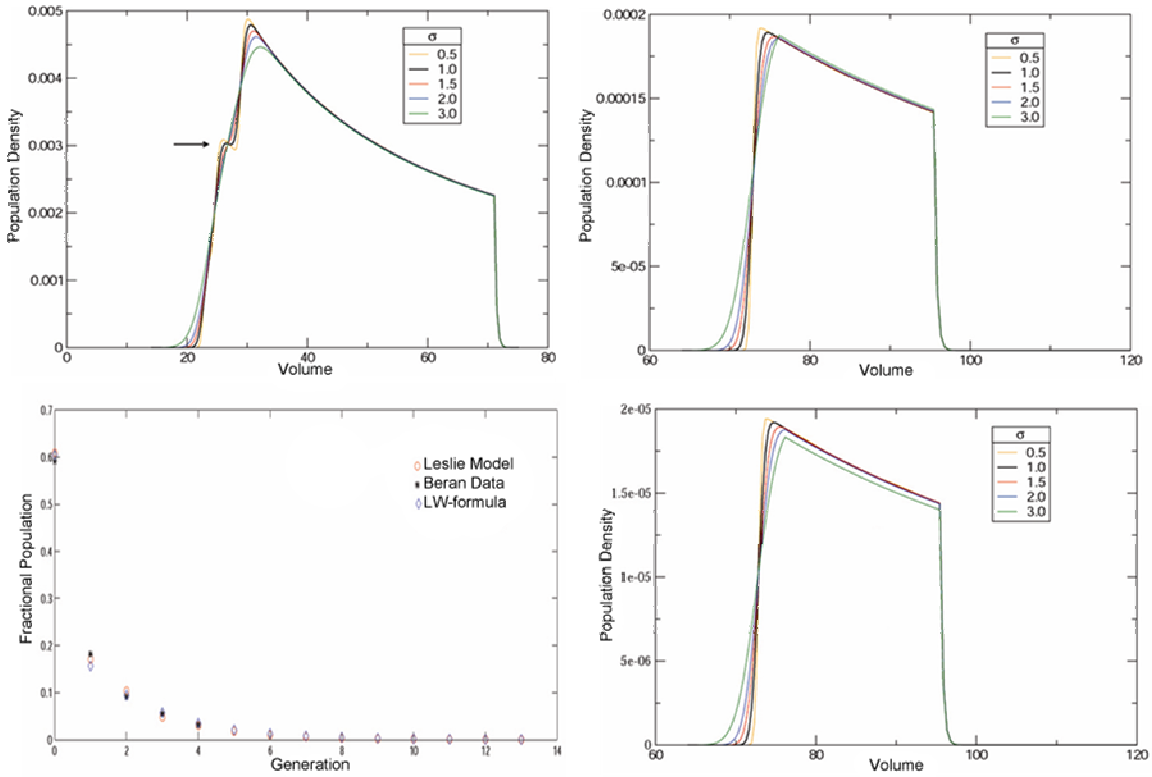


Figure 3.10: A concise description of the invariant population density distribution for the strain X2180 milestones. Volume units are cubic microns. Shown clockwise from upper left are the P_0, P_6 and P_{13} invariant density distributions. The daughter distribution shows the effect of birth and renewal as it is the only age class that exhibits bimodality. The inflection point indicated in the daughter density appears at the age distribution weighted average of the k-mean daughter birth volumes for $k > 0$. The density of successively higher generations is a decaying exponential distribution in volume. Changing the width, σ, of the division distributions, μ and ν, has only a qualitative smoothing effect on the invariant density. The lower left panel shows the age distribution of the model as compared with the experimental data of Beran et al. and with the age distribution formula derived by Lord and Wheals[5,30].

The stationary daughter distribution exhibits an inflection point at the population weighted average of the *k*-MEDV from all age classes with $k > 0$. Because of the birth of new daughters coming from all age classes, the daughter distribution is the only generation to exhibit bimodality. A local maximum appears just ahead of the 0-MEDV milestone that results from the $P_0 \rightarrow P_1$ division. The daughter density distribution decays with increasing volume after the global maximum as a linear combination of two exponentials. The structure of the invariant density is similar to those hypothesized in earlier work[3,30,36].

The invariant density within the parent age classes, P_k for $k > 0$ are similar to each other in that they achieve a global maximum that decays exponentially with increasing volume. For all age classes other than the daughter generation, the invariant density is indistinguishable from the function $A \bullet 2^{1-\theta}$, where the constant A is arbitrary and the simple linear function $\theta=(\lambda-\lambda_0)/(\lambda_1-\lambda_0)$ rescales the volume interval into the unit interval. This agrees well with the theory described previously in the literature[3,30,36,41].

Age Distribution

Since replicative age can be distinguished through bud scar analysis, it is possible to determine the age distribution of a culture of yeast. For instance, if we select a cell at random from a culture of X2180 cells during asynchronous exponential growth in a bioreactor, we will have a less than 1 in 3 chance of observing a P_1 and about a 1 in 6 chance of finding a P_2.

It is of interest to understand how each age class is weighted during oscillations as well as when the density becomes stationary[5,16,30,43]. The age distribution of a symmetrically dividing organism decays like the geometric series $\left(\frac{1}{2}\right)_{k=0}^{k+1}$. For budding yeast, the age distribution is more complicated. Lord and Wheals derived a simple formula based on the culture doubling time and the doubling time of the parents, P[30].

The age distribution computed using the X2180 milestones, shown at the lower left in Figure 3.10, shows excellent agreement with the experimental data of Beran et al. for a strain of *Saccharomyces cerevisiae* grown in a bioreactor at comparable dilution rates[5]. The formula of Lord and Wheals was fit by least squares to the Leslie model data through the variable P. The best fit value of P=88.3 minutes is however uninterpretable in relation to the X2180 parameters. For instance, the average maximum doubling time of the parent generations is calculated as 136.6 minutes, while the average minimum doubling time time is 96.8 minutes. These latter two values should realistically bookend the mean doubling times.

Based on a consideration of population flux and flux transit time, we have been able to derive a recurrence relation that explains the observed non-geometric decay of the age distribution in terms of the growth parameters that extends previous work in the literature[16,30].

Given the general exponential decay of the age distribution, we have contented ourselves to represent 14 generations computationally. Experimentally, mortality curves for

replicative age have been measured for some strains of budding yeast[14,32,34]. It has been observed that some yeast can survive upwards of 60 divisions. From the decline in the age distribution, we have observed that practically, 20-40 generations or more, need not be represented in the model to precisely capture the dynamics of the system. We know of no experimental data sets that have completely characterized more than the first 8 age classes. The precise connection between senescence and replicative aging is currently uncertain and is an interesting area of intense activity.

Conclusions

The Leslie model captures the dynamics of bud index oscillations and their decay. The measured data and the predicted bud index oscillations agree well for two different sets of strain milestones, one haploid and one diploid, of different volume extents and growth rates. The different strains of yeast display quantitatively different behavior with regard to their decay of synchrony as we have defined it. The X2180 strain exhibits 1 synchronous cycle while the LHY3865 strain displays 3. The Leslie model captures this difference, instilling confidence in the model predictions of synchrony. The strain milestones in both cases contain measurement error and are incomplete especially in generations higher than the fourth age class. Despite the measurement errors, the agreement of the model and experimental data exposes the robustness of the processes and ability of the Leslie model to capture the essentials of the asymmetric growth and division process. These claims are supported by the results of a sensitivity analysis.

It is theoretically known that volume symmetric division is a degenerate case that leads to persistent synchrony[13,42]. Several well known avenues allow the manipulation of the division volume asymmetry. Lord and Wheals observed, as have many others, that age class growth rates depend linearly on the culture doubling time and estimated that there exists a growth rate that if achievable would produce balanced and presumably synchronous growth[30]. Growth rates are most typically affected through variation of nitrogen or carbon source. Drugs such as hydroxyurea can induce nearly symmetric division[35]. It is well known that strain variations influence division volume asymmetry. We have explicitly examined the relationship between division volume asymmetry and the number of synchronous cycles of bud index oscillations. Our intentions in doing so are two-fold. First, we imagine that if a legitimate relationship exists, then it may be possible through a judicious mutation to create strains of yeast with pre-defined synchrony. Second, we see a direct relationship between the control of synchrony through continuous volume filtration and the natural synchrony that results from volume symmetric division. What this means is that a volume filter is seen in the abstract as a mechanism for restoring partial symmetry to an underlying volume asymmetric system. For instance, consider Figure 3.2. The volume grids of the different generations are not a priori commensurate; however, the volume grids that live between the filter cutoffs are more so. Those cells that are far from the symmetry conditions are removed from the system, leaving the remainder more synchronous. The intrinsic asymmetry that volume filtration cannot influence are the volume milestones such as *k*-MEDV, *k*-BE and *k*-MDV. These however, can be influenced by mutation and or nutrients. The combination of mutation, media composition, and continuous volume filtration is therefore expected to

be able to produce budding yeast that remain synchronous for long periods of time starting from a homogeneous initial condition.

We have explored both one and two stage filtration. We explored one stage filtration and because it is far simpler to implement in practice and it appears to produce results that could be observed with even a crude filtration device. The results indicate that robust windows of volume can be used to control synchrony. An example can be seen in the one stage results in Figure 3.9, there is a broad peak around 41 μm^3, approximately 4 μm^3 in width, that produces a roughly 4 fold extension in synchrony. This result, if correct, implies that even a crude filtration device should produce observable changes.

Continuous filtration is a control mechanism that will alter the population structure relative to an unfiltered population. The effected parameters are the volume distributions and the overall age distribution. The changes to the age distribution can be analyzed with the model. We believe continuous filtration can be accomplished experimentally without inducing a general stress response in the individual cells of the population.

Beyond specific application of volume filtration explained here, we observe the Leslie model can be used to explore a much broader range of questions that are of continuing interest in yeast physiology and in the larger picture of systems biology. For instance, as has been observed previously, the model can be used to investigate how signals from single cells manifest themselves at the population level[19,43]. With the addition of volume filtration, it will be possible to study cell cycle dependent protein expression more extensively.

A use that has been little explored to date is how a signal such as a gene expression profile, periodic in the cell cycle, manifests itself at the population level. When signals are routinely evaluated by grinding up large numbers of cells and pooling their mRNA, such questions seem reasonable. For example, Chapter 2 demonstrated that how one grinds up the cells has quantifiable effects that depend on the cell cycle[39].

Any extensive quantity that varies in a single cell with the cell cycle can be examined with this model. For example, oxygen consumption, glucose uptake, or mRNA production of the population can be studied given measured or putative data from single cells. Conversely, it is also possible to use the model to deconvolve population ensemble averages into individual cell signals.

Finally, a physiological component that has not been included into the current model are the putative asymmetric effects that are now emerging in the study of chronological aging and senescence[1]. It is well known that aging occurs in organisms such as *E. coli* and fission yeast that undergo morphogenically symmetric division[38]. Given the success of the Leslie model in matching the dynamics of the bud index oscillations for a few cell cycles, it is tempting to suggest that deviations between the Leslie model and the dynamics of yeast with a variety of aging phenotypes may provide new and otherwise difficult to attain insight into the rate and effects of senescence.

References

1. Ackerman, M., Chao, L., Berstrom, C., and Doebeli, M. 2007. On the evolutionary origin of aging. *Aging Cell.* **6**:235-244.

2. Agilaniu, H., Gustafsson, L., Rigoulet, M., and Nystrom, T. 2003. Asymmetric inheritance of oxidatively damaged proteins during cytokinesis. *Science.* **299**:1751-1753.

3. Barford, J.P. and Hall, R.J. 1976. Estimation of the length of cell cycle phases from asynchronous cultures of *Saccharomyces cerevisiae*. *Exp. Cell. Res*. **102**:276-284.

4. Bellgardt, K.H. 1994. Analysis of synchronous growth of bakers yeast. Part I: Development of a theoretical model for sustained oscillations. *J.of Biotechnol.* **35**:19-33.

5. Beran, K., Malek, I., Steiblova, E., and Lieblova, J. 1967. The distribution of the relative age of cells in yeast populations. In *Microbial physiology and continuous culture* : third international symposium held at the Microbiological Research Establishment (Ministry of Defence), Porton Down, Salisbury, Wiltshire.

6. Boczko, E.M., Cooper, T.G., Gedeon, T., Mischaikow, K., Murdock, D.G., Pratap, S., and Wells, K.S. 2005. Structure theorems and the dynamics of nitrogen catabolite repression in yeast. *Proc. Natl. Acad. Sci.* **102**:5647-5652.

7. Boczko, E.M., Gedeon, T., and Mischaikow, K. 2007. Dynamics of a simple regulatory switch. *Journal of Mathematical Biology.* **55**:679-719.

8.Breeden, L.L. 1997. α-Factor synchronization of budding yeast. *Methods in Enzymology.* **283**:332-342.

9. Brewer, B.J., Chlebowicz-Sledziewska, E., and Fangman, W.L. 1984. Cell cycle phases in the unequal mother/daughter cell cycles of *Saccharomyces cerevisiae. Mol. Cell. Biol*. **4**:2529-2531.

10. Caswell, H. 2001. Matrix population models, 2nd Edition. Sinauer Associates, Inc. Sunderland, MA.

11. Chen, K.C., Calzone, L., Ccikasz-Nagy, A., Cross, F.R., Novak, B., and Tyson, J.J. 2004. Integrative analysis of cell cycle control in budding yeast. *Mol. Biol. Cell.* **15**:3841-3862.

12. Davis, R.H., and Gecol, H. 1996. Classification of concentrated suspensions using inclined settlers. *International J. Multiphase Flow* **22**:563-574.

13. Diekmann, O., Heijmans, H.J.A.M., and Thieme, H.R. 1984. On the stability of the cell size distribution. *J. Math. Biol*. **19**:227-248.

14. Egilmez, N.K., and Jazwinski, S.M. 1989. Evidence for the involvement of a cytoplasmic factor in the aging of the yeast *Saccharomyces cerevisiae*. *J. Bacter.* **171**:37-42.

15. Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci*. **95**:14863–14868.

16. Gover, N.B., and Woldringh, C.L. 1995. Relationship between the fraction of cell of different genealogical ages and their cycle times in *Saccharomyces cerevisiae*: A theoretical analysis. *J. Theor. Biol*. **174**:223-225.

17. Harris, T.E. 1967. The theory of branching processes. *Die Grundlehren der Mathematischen Wissenschaften.* **119.** Springer-Verlag, Berlin; Prentice-Hall, Inc., Englewood Cliffs, N.J.

18. Hartwell, L.H., and Unger, M.W. 1977. Unequal division in *Saccharomyces cerevisiae* and its implications for the control for the cell division. *J. Cell. Biol.* **75**:422-435.

19. Hatzis, C., and Porro, D. 2006. Morphologically structured models of growing budding yeast populations. *J. Biotechnology.* **124**:420-438.

20. Helmstetter, C.A. 1991. Description of a baby machine for *Saccharomyces cerevisiae*. *The New Biologist.* **3**:1089-1096.

21. Hersh, R.T., and Kitos, P.A. 1980. Is G_1 normally distributed? *J. Theor. Biol.* **86**:117-122.

22. Johnston, G.C., Pringle, J.R., and Hartwell, L.H. 1977. Coordination of growth with cell division in the yeast *Saccharomyces cerevisiae*. *Exp. Cell. Res.* **105**:79-98.

23. Johnston, G.C., Ehrhardt, C.W., Lorincz, A., and Carter, B.L.A. 1979. Regulation of cell size in the yeast *Saccharomyces cerevisiae*. *J. Bacteriol*. **137**:1-5.

24. Johnston, G.C. and Singer R.A. 1983. Growth and the cell cycle of the yeast *Saccharomyces cerevisiae. Exp. Cell. Res*. **149**:1-13.

25. Jorgensen, P. and Tyers, M. 2004. How cells coordinate growth and division. *Curr. Biol*. **14**:R1014-R1027.

26. Kaeberlein, M. 2005. Regulation of yeast replicative life span by TOR and Sch9 in response to nutrients. *Science.* **310**:1193-1196.

27. Keulers, M., Sazuki, T., Satroutdinov, A.D., and Kuriyama, H. 1996. Autonomous metabolic oscillations in continuous culture of *Saccharomyces cerevisiae* grown on ethanol. *FEMS Micro. Lett*. **142**:253-258.

28. Kimmel, M. and Axelrod, D.E. 2002. Branching processes in biology. *Interdisciplinary Applied Mathematics.* **19**. Springer-Verlag, New York.

29. Koshwanez, J., Holl, M., Marquardt, B., Dragavon, J., Burgess, L., and Meldrum, D. 2004. Identification of budding yeast using a fiber-optic imaging bundle. *Rev. Sci. Instr.* **75**:1363-1365.

30. Lord, P.G. and Wheals, A.E. 1980. Asymmetrical division of *Saccharomyces cerevisiae. J. Bacter*. **142**:808-818.

31. Murray, D.B., Klevecz, R.R., and Lloyd, D. 2003. Generation and maintenance of synchrony in *Saccharomyces cerevisiae* continuous culture. *Exp. Cell Res*. **287**:10-15.

32. Pohley, H.J. 1987. A formal mortality analysis for populations of unicellular organisms.(*Saccharomyces cerevisiae*). *Mech. Ageing. Dev*. **38**:231-243.

33. Pringle, J.R. 1991. Staining of bud scars and other cell wall chitin with calcoflour. *Methods In Enzymology.* **194**:732-735.

34. Sinclair, D., Mills, K., and Guarente, L. 1998. Aging in *Saccharomyces cerevisiae*. *Annu. Rev. Microbiol*. **52**:533-560.

35. Singer, R.A. and Johnston, G.C. 1981. Nature of the G_1 phase of the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci*. **78**:3030-3033.

36. Slater, M.L., Sharrow, S.O., and Gart, J.J. 1977. Cell cycle of *Saccharomyces cerevisiae* in populations growing at different rates. *Proc. Natl. Acad. Sci*. **74**:3850-3854.

37. Smith, J.A. and Martin, L. 1973. Do cells cycle? *Proc. Natl. Acad. Sci*. **70**:1263-1267.

38. Stewart, E.J., Madden, R., Paul, G., and Taddei, F. 2005. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biology* **3**:295-300.

39. Stowers, C. and Boczko, E.M. 2007. Reliable Cell Disruption in Yeast. *Yeast* **24**:533-541.

40. Tu, B.P., Kudlicki, A., Rowicka, M., McKnight, S.L. 2006. Logic of the yeast metabolic cycle: Temporal compartmentalization of cellular processes. *Science.* **310**:1152-1158.

41. Tyson, C.B., Lord, P.G., and Wheals, A.E. 1979. Dependency of size in *Saccharomyces cerevisiae* cells on growth rate. *J. Bacteriology.* **138**:92-98.

42. Tyson, J.J. and Diekmann, O. 1986. Sloppy size control of the cell division cycle. *J. Theor. Biol*. **118**:405-426.

43. Vanoni, M., Vai, M., Popolo, L., and Alberghina, L. 1983. Structural heterogeneity in populations of the budding yeast *Saccharomyces cerevisiae*. *J. Bacteriology.* **156**:1282-1291.

44. Walker, G. 1999. Synchronization of yeast cell populations. *Methods in Cell Science.* **21**:87-93.

45. Walsh, T.J., and Bungay, H.R. 1979. Shallow depth sedimentation of yeast cells. *Biotechnology and Bioengineering.* **21**:1081-1084.

46. Wheals, A.E. 1982. Size control models of *Saccharomyces cerevisiae* cell proliferation. *Molecular and Cellular Biology.* **2**:361-368.

47. Wickramasinghe, S.R., Lin W.C., and Dandy D.S. 2001. Seperation of different sized particles by inertial migration. *Biotech. Lett*. **23**:1417-1422.

48. Woldringh, C.L., Huls, P.G., and Vischer, N. 2003. Volume growth of daughter and parent cells during the cell cycle of *Saccharomyces cerevisiae* a/α as determined by image cytometry. *J. Bacteriology.* **175**:3174-3181.

CHAPTER IV

COMMUNICATION, CLUSTERING, AND ENVIRONMENTAL OSCILLATION IN POPULATIONS OF BUDDING YEAST

Introduction

The periodic oscillation of physiologically relevant variables during yeast growth and division has been reported and studied for over 40 years [6,8,12.24,28,31,32,33,45,50]. The physiologically observed variables have covered the full range of biological complexity and include dissolved oxygen, pH, carbon and nitrogen sources such as glucose, ethanol and ammonia, second messengers such as cAMP, the expression level of mRNA's, the activity of metabolic enzymes, and indices of growth and division such as the percentage of cells budded and DNA content.

The observed oscillations are of broad interest in biology [24,31,33,45] and chemical engineering [12,15,18] for a number of reasons. The control of oscillation and the regulation of yeast metabolism has been an important theme in the chemical engineering literature devoted to the efficient management of bioprocesses. Oscillations are also of interest because they expose questions regarding the coordination of the cell cycle and metabolism and the possible existence of clocks and pacemakers [33]. Such oscillations and their classification have been investigated and described with regard to intercellular communication [15,30,35,38,39].

Autonomous oscillations are far from universal, however, and are only routinely observed in a few strains and within relatively narrow ranges of operating parameters. Apart from the non-universality, two central questions remain unanswered: a full understanding of the feedback mechanism that keeps the oscillations entrained and the relationship between the oscillations and cell cycle progression.

Why have questions concerning these oscillations remained challenging for so long? For one, our notions of causality are predicated on temporal precedence and these are confounded by the periodic repetition in feedback loops: for example, the chicken and the egg. A far less ethereal complication arises when population results are ascribed to individual organisms. Often the linear sequence of cell cycle phases are used as abscissa for graphs of the oscillations of environmental variables [11].

Issues are further complicated because organization can easily be obscured by projection. For instance, imagine a room filled with rows of people, three people per row, equally spaced apart. Depending on the placement of a light source, the shadow that the room full of people cast on a wall may well obscure their internal organization into rows of three, while others will reveal it.

Just as the people are organized within the hypothetical room, yeast populations in a bioreactor or in a colony display a high degree of organization. Aspects of this organization are revealed by considering the population density as a function of different variables. In a well mixed planktonic culture, it has been observed that a stratification based on cell cycle position and replicative age provides discrimination. Typically, an

extensive property such as cell volume or mass has been used as a proxy for cell cycle position [3,13,20,22,34,48].

The vast majority of operating conditions produce stationary, asynchronous, exponentially growing cultures. Under the assumption of stationary, asynchronous, exponential growth, the population structure of a budding yeast culture is described by a continuous population density that declines exponentially with cell cycle position from G_1 to M, and geometrically with age [26]. Little is known about time periodic population structures. Most experimental studies that have described population structure have been related to the effort to understand the coordination of yeast volume growth and the sensing mechanisms that gates entry to start of the cell cycle [30,49]. The experimental studies of population structure have been limited by the arduous nature of the biophysical characterizations required for which standard, high accuracy methodology does not exist. An advantage of theory and computation is that it is straightforward and it can produce population density profiles as a function of arbitrary variables [44].

This chapter is concerned with an analysis of repeatedly observed experimental results in light of new mathematical results on the nature of the population structure that develops as a result of generic models of communication. The described population structure provides a proper translation between individual behavior and population measurements. We propose that this population structure of autonomously oscillating yeast cultures can be exploited to scientific advantage. For example, the measurement of cell cycle dependent gene expression has been a persistent challenge to understanding the link between genetic networks and biological physiology. We present results regarding the

expression profiles for two genes, GLN3 and DAL80, that are essential to the function of Nitrogen Catabolite Repression (NCR) in *Saccharomyces cerevisiae.* These results provide supportive evidence to previously developed theoretical predictions [2]. We also provide experimental evidence that the population structure can be, albeit within limitation, controlled to potential advantage.

Maintenance of Synchrony

Cell cycle synchrony in budding yeast is typically measured from a time series of the percentage of the cells in the population that are budded. Buds emerge at or near the G_1 to S transition and persist until division occurs. The bud site accumulates chitin that remains as a permanent scar that can be stained to provide a quantitative experimental measure of replicative age. Bud index and age distributions are usually manually counted under a microscope and are labor intensive to obtain, although an automated fiber optic system has been described [25]. The bud index can be affected in two ways. It can rise as cells become budded and it can decline as budded cells divide and/or as newly divided cells dilute the total population. In a dense and well-mixed culture, the dilution rate should have no effect on the bud index since removal through dilution is an unbiased linear process.

Cell density, or the number of cells per unit volume of culture, can be measured using a Coulter counter. Cell density can be affected in 3 ways. First, cell density can increase by clusters of cells passing through the M-G_1 transition. Second, cells can be removed from continuous culture through dilution. Third, cells can die and disintegrate, but the rate that this occurs has been shown to be negligible [16,37,41].

Strains of budding yeast exhibit an age-dependent volume growth variation that strongly dephases cell cycle synchrony [44,49] under most nutrient conditions and at most dilution rates. Daughter cells are smaller at division than their mothers and consequently have a longer G_1 traverse. Even a small, but systematic difference in the way an age class of yeast traverses the cell cycle will eventually lead a culture to dephase. Typically no more than three consecutive decaying cycles of bud index synchrony can be achieved [4,46] in the absence of autonomous oscillation. Therefore, the maintenance of synchrony in an actively growing culture requires the existence of some mechanism to counteract the natural forces that tend to dephase cell cycle synchrony during growth and division [44]. Similar reasoning appears in slightly different language in Keulers et al.[23].

Oscillatory behavior in yeast is typically grouped into two or three classes whose designations differ depending on the author [15,38]. A description of the population structure within any of these classes remains an open and fundamental question. It is speculated that the rapid NADH oscillations produced during anaerobic glycolysis *in-vivo* could not be observed in the absence of strict population synchrony. Mixing experiments support, but do not conclusively prove this assertion [38]. It has been demonstrated that acetaldehyde acts as a synchronizing agent in this oscillation [9]. We have further investigated the effects of acetaldehyde on yeast by injecting acetaldehyde into both asynchronous and synchronous cultures and found that acetaldehyde can be used to both dephase synchronous cultures and entrain asynchronous cultures.

In two often cited papers [15,32], whose origin can be traced to a single paper [40] that cites a negative result without data, it is claimed that short period aerobic oscillations in the IFO0233 strain occur without cell cycle oscillations. Such counter intuitive claims require overwhelming evidence, and to our knowledge no data supporting these claims have been published. In contrast, we present data that indicate short period oscillations, identical to those reported, are accompanied by sizeable bud index oscillations, see Figure 4.1.

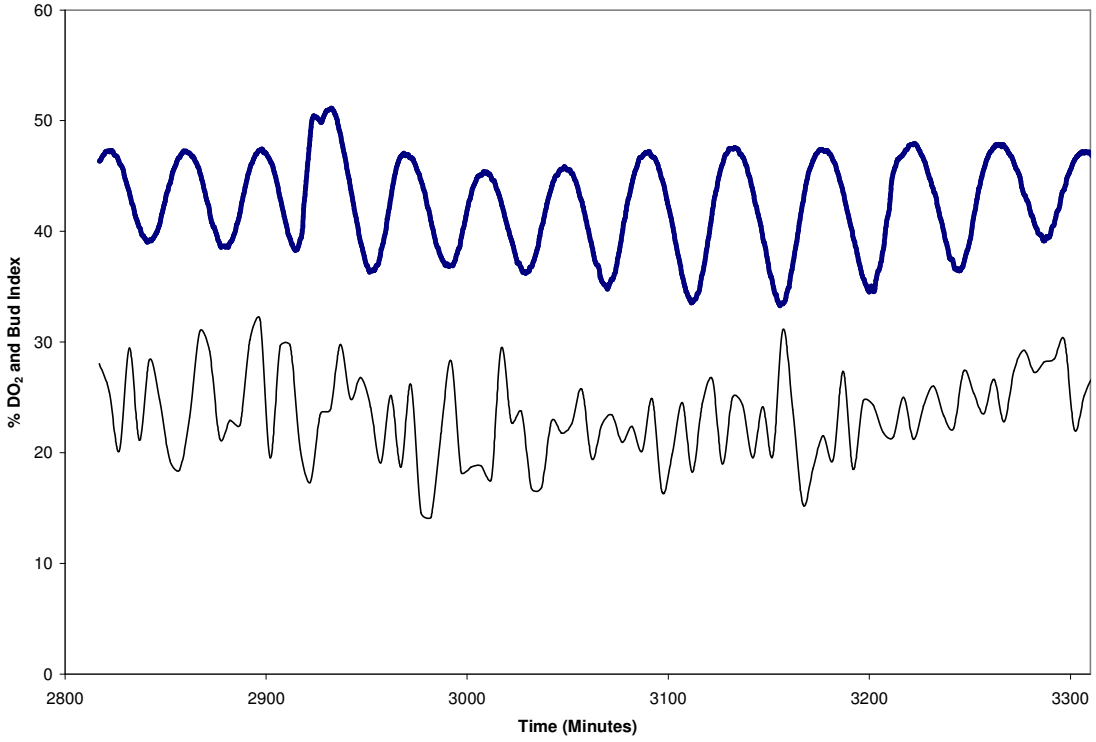


Figure 4.1: Dissolved oxygen (bold blue line) and bud index (thin black line) for yeast strain IFO0223. The presence of bud index oscillations demonstrate that the high frequency, 40 minute, oscillation of the IFO0223 strain is linked to the cell cycle, contrary to popular belief.

Several feedback mechanisms have been proposed for the maintenance of cell cycle synchrony. Mechanisms involving metabolite feedback that operate on growth rate are particularly elegant [18]. Oscillating metabolite concentrations produce oscillating growth rates and these can act to locally stabilize synchronous trajectories like a focusing billiard. Based on solid experimental evidence [29,50], it has been suggested that during long period aerobic oscillations, a burst of glycolysis follows from the liberation of glucose from glycogen and trehalose at the G_1-S boundary [11,29]. The ethanol produced by the population entering S-phase is proposed to enhance the growth rate of cells in G_1. Whether or not this proposition is true, it is an example of what we will call an *Advance* model. In this model, an agent secreted by a part of the population in one particular phase of the cell cycle advances the cell cycle progression of those cells earlier in the cycle. In the opposite case of a *Delay* mechanism, cells would slow or stop their cell cycle progression in response to a secreted factor or metabolite. In an extreme case of a delay mechanism, that might adequately be called *Blocking*, where the progression of cells would stop. Such is the response of budding yeast to pheromones such as α-factor, which stop cell cycle progression at the G_1-S boundary. We observe that advances or delays can be affected by a large number of mechanisms involving communication-the end result of which are oscillations. The existence of intercellular communication to control the metabolic state of the organism has been well documented for a wide variety of organisms and growth conditions [4,7,10,19,27].

Oscillations have been observed to exist in dilution rate windows [21]. This observation has been interpreted with respect to the interplay between the age-dependent growth rates of daughters and parents within the population [6]. Since the population density within

progressively older generations decays geometrically, typically only the first two or three generations have a measurable impact. Cultures exhibiting oscillations, while planktonic, are dense and well-mixed and phase-resetting experiments with a variety of molecules suggest that intercellular communication is at least part of the feedback mechanism that stabilizes the oscillations [31].

In this chapter, we report experimental results on oscillations in cultures of the strain IFO 0233 and Cen.PK 113. Our main contribution is to interpret these data in light of mathematical results that demonstrate that generically *Advance* or *Delay* models organize the population density into non-stationary multimodal distributions to produce temporally coherent or pseudo-synchronized groups, that we call *clusters*. The cell cycle progression of the clusters produces the observed environmental oscillations.

Clustering

We have observed that for various forms of communication among cells, which result in either advances or delays, there is robust clustering of cells within the cell cycle. Clustering is defined as the temporal pseudo-synchronization of large sub-populations within the cell cycle. In other words, a cluster is significant groups of cells passing through cell-cycle milestones, such as budding or division, at nearly the same time. A cluster is akin to a group of cars moving along a multilane highway with comparable speed.

The fact that advances or delays can cause spontaneous polarization in subpopulations was already appreciated by Muller et al. in their work on the dynamics of population

structure [29]. Our contribution is to demonstrate the existence of clusters through the concordance of experiment and mathematical theory. The robustness of the clustering mechanism is observed in at least the following three ways. First, clustering occurs across a wide variety of mathematical models, from extremely general to highly specific. Second, it is generated by the very large class of advance or delay feedback mechanisms and from arbitrary initial conditions. Finally, the appearance of clustering in response to these feedback mechanisms is insensitive to noise. These attributes are essential for observability.

We have considered several realizations of cell-cycle models, supposing only that the appearance of a thresholded population of cells in one cell cycle phase, S, affects the growth of cells in another, R. That is, if the integrated population density in the S phase is larger than *T*, then the cells in the R phase are either advanced or delayed depending on the operative feedback mechanism. We have labeled these two phases in analogy with the hydrogen sulfide evolution models of short term oscillatory control [33]: S for the DNA synthesis phase of the cell cycle and R for the responsive phase. The specific number of clusters that form depends on the widths of the S and R phases and on the particular feedback mechanism. Of key importance is the necessity that if clustering occurs, then there are an integer number of clusters. This observation explains the periodic-looking behaviors seen in autonomous oscillations with periods that are integer divisors of the cell-cycle length.

In the literature [45,47,50] and in the experimental data described below, the bud index oscillates with the same period as dissolved oxygen. Our data also include a

measurement of the cell density as a function of time that also oscillates with the same period. A significant contribution of our work is the interpretation of these data in terms of cell cycle clustering and mathematical theory. These data provide two points at which one can strobe the population structure of the system with respect to the cell cycle, which we believe elucidate the population structure of the system. The anti-phase relationship between the bud index and the cell density unequivocally indicate the existence of clusters in accord with the mathematical results.

Models of Growth and Division

We have analyzed the dynamics of populations of budding yeast given various generic forms of feedback between the cells. In all of these cases that we have considered, the feedback can be interpreted as resulting from communication among the cells. In fact, the structure of the feedback was chosen as a generalization of the physiologically observed mechanisms operative in real yeast cultures. In all of these cases, feedback resulted in a non-stationary, multimodal population density that contained clusters. Clustering has also been hypothesized to result from feedback of a form different than what we have considered [51]. The generality and naturality of these results provide strong evidence that *clustering exists in nature and is a robust phenomenon*.

Advance and Delay Models

The results of our analysis are summarized starting from the most general models that are devoid of all but the most basic features. The cell cycles of many strains of budding yeast are well known to differ in range and duration according to replicative age [49]. In the most basic models, we ignore these distinctions in order to simplify the analysis.

These details are reintroduced in later models. Let S and R represent portions of the cell cycle as described previously. Typically, we assume that the R phase precedes the S phase in the normal cell cycle progression. Let $\rho_S(t)$ and $\rho_R(t)$ represent the total population of cells in those respective phases of the cell cycle and τ represent a threshold of cells required for efficient communication. Since the system is dynamic, these quantities may vary with time t.

An *advance model* corresponds to the following rule: If the fraction of cells in S exceeds some threshold $\rho_S(t)>\tau$, then all cells currently in R, $\rho_R(t)$, are instantaneously promoted to the beginning of S, where they resume normal growth.

A *delay model* corresponds to the rule: If the fraction of cells in S exceeds some threshold $\rho_S(t)>\tau$, then cells at the beginning of S are blocked from entry into the S-phase. The blocked cells resume normal growth when the condition $\rho_S(t)>\tau$ becomes false at some time t.

These rules may appear too strict to be realistic. However, we note that the mating pheromone α-factor behaves as a delay model as described. Either of these rules will cause the population to become more dense in certain portions of the cell cycle and less dense in others. The introduction of this type of modal, wave-like structure within an actively growing and dividing population represents a form of synchrony. We call a group of synchronized cells a *cluster*. If the number of cells in a cluster is large enough to exceed the threshold of the model, we call it a *critical cluster*. The number and the size of the population clusters within a yeast culture will translate directly into

measurable properties. For instance, the cell density profile will increase each time a cluster of cells passes through the M-phase, and the bud index must increase each time a cluster passes into S-phase. Mathematically, we have begun to characterize the number of clusters that a particular feedback will produce. Therefore the following theorems have practical and measurable significance.

In Boczko et al. [1], the interested reader can find proofs of the following theorems that give an upper bound on the number of critical clusters that can form.

Theorem 4.1 *In an advance model, no more than* $\lfloor (|R| + |S|)^{-1} \rfloor$ *critical clusters can persist.*

Theorem 4.2 *In a delay model, no more than* $\lceil |S|^{-1} \rceil$ *critical clusters can persist.*

Here $\lfloor x \rfloor$ denotes the largest integer less than or equal to x and $\lceil x \rceil$ denotes the smallest integer larger or equal to x. Also, |S| signifies the temporal length of the S phase.

It can be shown that not only the number of clusters is bounded above, but in certain circumstances, it is also bounded below. This means that clusters **must** form.

Theorem 4.3 *In a delay model, if the cells are initially homogeneously distributed about the cell cycle and the density exceeds twice the threshold, T, then exactly* $\lceil |S|^{-1} \rceil$ *clusters will develop and persist.*

Leslie Model Simulations

At the greatest level of detail, we have considered models of budding yeast growth and division that are parameterized with experimental data and that are stratified with respect to cell cycle progression and age. The data used to parameterize the model include age class dependent growth milestones, such as the average size of a cell at bud emergence and the variance in that value. Such data are available in the literature for some strains [49], and we have made measurements for others. We have shown that these models can capture the complexity of yeast growth dynamics and account for strain variations therein[44].

In every case that we have examined, the simulation data clearly indicate that there exist open intervals of the threshold value in which synchronous cell cycle oscillations become stable in either advance or delay models. An examination of the population density shows that clustering occurs and that the cell cycle progression of the clusters with time is responsible for the oscillations.

Continuous Culture

The haploid strain Cen.PK 113-7D (mat a) of budding yeast was cultivated in a 3 L New Brunswick Scientific Bioflow 110 reactor equipped with two Rushton type impellers, operated at 550 rpm. Air was sparged at a rate of 900 mL/min. Temperature was maintained at 30 °C and pH was automatically controlled with 2 N sodium hydroxide. Overnight cultures were grown in YPD media with shaking. 20 mL were inoculated into a reactor containing 850 mL of synthetic media with the following composition: 10 gm/L anhydrous glucose (Sigma # G71528), 5 gm/L ammonium sulfate (Sigma # A2939), 0.5 gm/L magnesium sulfate heptahydrate (Sigma M2773), 1 gm/L of Yeast Extract (Becton, Dickinson and Company Cat 288620), 2 gm/L potassium phosphate (Sigma # P5379), 0.5 mL per mL of 70 % v/v sulfuric acid, 0.5 mL of antifoam A (Sigma # 10794), 0.5 mL of 250 mM calcium chloride, and 0.5 mL of mineral solution A, see Table 4.1. For nitrogen switching experiments, the appropriate nitrogen source was substituted for ammonium sulfate in the exact same amount. No other media component was altered during these experiments.

Table 4.1: Composition of the mineral solution used to prepare culture media for CEN.PK 113-7D.

Mineral Solution A	
$FeSO_4 \cdot 7\ H_2O$	40 gm/L
$ZnSO_4 \cdot 7\ H_2O$	20 gm/L
$CuSO_4 \cdot 5\ H_2O$	10 gm/L
$MnCl_2 \cdot 4\ H_2O$	2 gm/L
75% H_2SO_4	20 mL/L

The culture was subsequently grown in batch mode for 12-16 hours until it reached a cell density of approximately $5x10^9$ cells/mL. Continuous culture ensued with a dilution rate of 0.088 hr^{-1}. Samples were taken from the reactor and flash frozen using dry ice. The haploid strain IFO0223 (mat a) require slightly different conditions. Overnight cultures were grown in YPD with shaking. 20 mL were inoculated into the reactor described above, containing 900 mL of the media previously described [31] with the exception that we used 0.5 mL/L of Antifoam A. The reactor was perfused with wet air at rate of 200 mL/min, kept at pH of 3.4, and maintained an agitation speed of 750 rpm. The reactor was operated in batch mode for 16-20 hours. Operation was switched to continuous mode at a dilution rate of 0.085 hr^{-1}. Autonomous oscillations were obtained after switching to continuous mode following various hallmarks. These hallmarks include switching at the dissolved oxygen minimum, switching after the dissolved oxygen rebounded to 100% saturation, and following four hours of starvation. The cultures of IFO0223 typically achieved a maximal density of approximately $3x10^8$ cells/mL.

Measurement of Bud Index and Cell Density

Samples were collected at 15 minute intervals and immediately frozen in an ethanol-dry ice bath. Buds were counted with a 100X objective on a conventional Nikon TE-2000 microscope. A minimum of 300 cells were scored at each time point. Cells without buds were considered to be within the G_1 phase of the cell cycle.

Cell samples were thawed in a 4 ºC sonicating water bath for 5 minutes and vortexed briefly. 1 μL of sample was resuspended in 10 mL of Isoton Diluent II(Beckman Coulter # 8546719) and vortexed. The cell density of three independent 0.5 mL samples were measured using a Beckman Multisizer Coulter counter. The sizing threshold was 2-8 μm.

Gene Expression Measurements

The gene expression profiles were measured during the autonomous oscillations using luciferase reporter constructs. The cloning process including primer sequences, shuttle vectors, and cloning sites is described in Appendix B. Cells were grown in media containing 5 μM luciferin (Gold Biotech # LUCK-500-SPO) and continuously removed from the reactor using a peristaltic pump, measured using a PMT (Photomultiplier Tube), and were subsequently returned to the bioreactor. The residence time for cells outside the reactor was approximate 20 seconds. Data were recorded at 1 minute time intervals at a frequency of every other minute.

Results

Existence of Clusters

Figure 4.2 shows three successive dissolved oxygen oscillations and the periodic extension of one cycle of bud index oscillation. Two oscillations occur within a 7.8 hour period that approximately corresponds to the doubling time imposed by the dilution rate. Over the course of an oscillation, the bud index percentage falls to 4%, and in a time span of less than 1 hour rises to 50%. The sharpest part of this increase occurs within a time span of 20 minutes. This is by definition a cluster: a significant percentage of cells that arrive at a common milestone, the S phase, at approximately the same time.

The fact that the bud index reaches 50% and it occurs twice within the culture doubling time shows that half of the cells enter S-phase as a cluster, while the other half remain unbudded and reside in G_1. In the subsequent oscillation, the other half of the population density enters S-phase again as a cluster. The population density is organized into two clusters.

These conclusions are reiterated and refined by data in Figure 4.3 that describes a three and a quarter hour oscillation and a time series of the cell density. The bud index can rise only as cells pass into the S-phase, while the cell density can rise only as cells exit their cell cycle from the M-phase during division. The phase relationship of the density data to the bud index data is extremely informative. We draw the following conclusions.

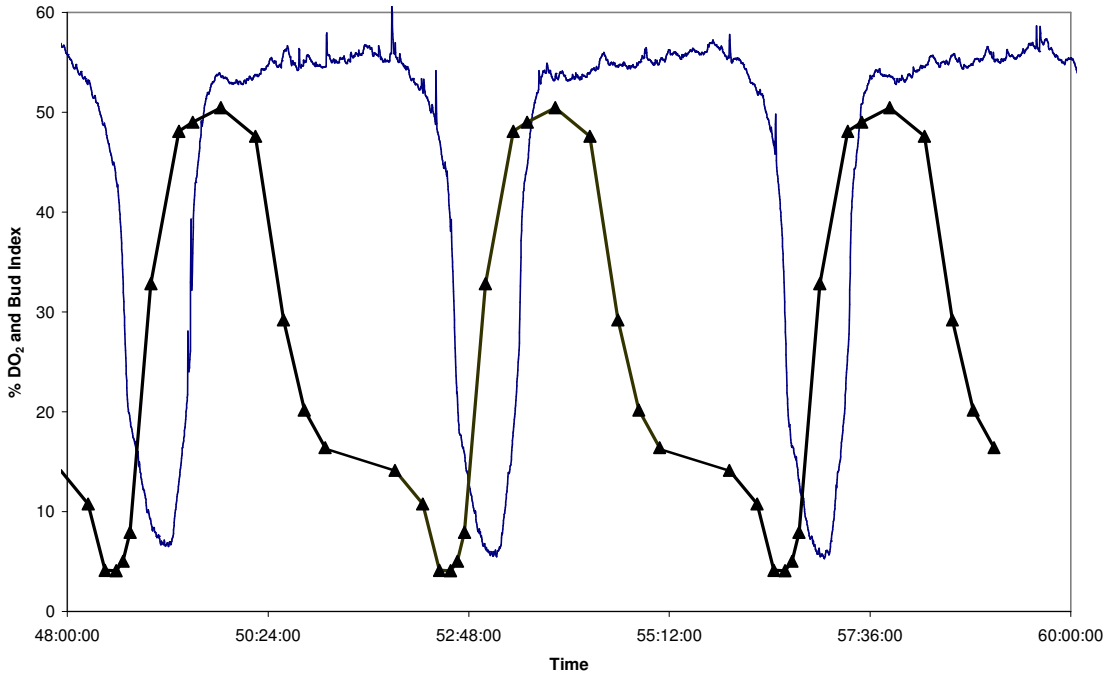


Figure 4.2: A 3 hour and 40 minute dissolved oxygen oscillation (thin blue line) and bud index measurements (thick black line) for yeast strain Cen.PK 113-7D. The dilution rate of the bioreactor was set to 0.087 hr^{-1} such that two waves of budding occur during one cell cycle time period. Time in the figure is represented as hours:minutes:seconds. The data in the figure span a time period of 12 hours.

First, as the bud index rises to its maximum, the cell density is purely in decline. From the slope of this decline we calculate a dilution rate that is within less than 8% error of the dilution rate set in the experiment. This value is well within the deviation and statistical noise of the measurements and is strong evidence that no cell division is taking place as a cluster passes into the S-phase. The subsequent drop in bud index from its maximum value corresponds to a rise in cell density, indicating that loss of buds is due to cell division of that cluster.

Second, the phase relation between the cell density maximum and the subsequent bud index minimum indicate that there exists a short period of time when both clusters coexist in the G_1 phase. Given the sampling rate, we can estimate this time to be on the order of 15 minutes. The magnitude of the bud index minimum, from 4% to 8%, is further indication of this temporal coexistance. The phase relation between these events represents the timing and spacing between the cell clusters, such that as the tail of the dividing cluster is passing through M-phase and all of its cells are returning to a G_1 phase, the second cluster's leading edge is passing into the S-phase and is begging to bud.

Third, the close temporal interplay between the bud index minimum and the cell density maximum contains more information. The cells at the cell density maximum must be in early G_1. These same cells cannot begin to bud within the next 15 minutes. However, the cells at the bud index minimum are doing just that, indicating the existence of multiple clusters of cells. The fact that two oscillations occur within one dilution rate doubling time testifies to the interleaving of two clusters.

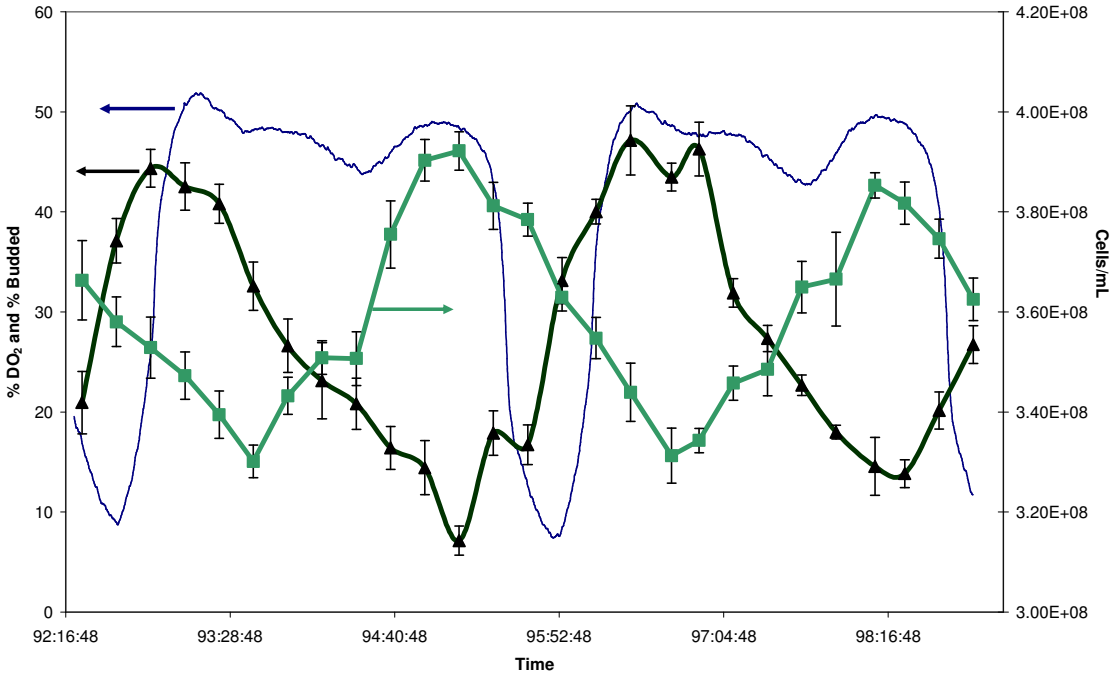


Figure 4.3: Bud index (bold black line, triangles), cell density (bold green line, squares), and dissolved oxygen (thin blue line) for a 3 hour and 15 minute oscillation. The data indicate two waves of division and budding nearly perfectly out of phase with each other per cell cycle. Error bars represent one standard deviation over three independent measurements. Time in the figure is represented as hours:minutes:seconds. The data in the figure span 390 minutes.

The temporal width of the bud index curve from its global maximum to the following global minimum indicates the width of the group of daughter cells that emerge within the daughter generation, P_0, due to the cell divisions. The ratio of that temporal width to the width from minimum to following maximum is approximately 7/4. The latter width is an upper bound on the time it takes a cluster to pass into S-phase. There are several conceivable mechanisms that could effect this change. Since all cells do not bud at precisely the same cell cycle position or volume, the simultaneity induced by the dispersion in the budding process could account for some or all of the effect. We believe that this narrowing is at least in part an indication of the feedback mechanism that produces the clustered population structure by reducing the temporal variance within the daughter population by 57% (4/7) over the course of its G_1 traverse. While dilution will progressively whittle away at a cluster of cells throughout their cell cycle traverse, it cannot theoretically alter the width of the cluster at half-height. More data are required to further sharpen these deductions.

The symmetry of the bud index and cell density curves contain meaning. The fall in the bud index is not as sharp as its rise. Simulations indicate that the trailing tail is due to the fact that the different age classes do not divide in unison. This observation is supported by the fact that the cell density appears to have three peaks, as also seen in simulations. Another observation is that the sharp rise in the bud index curve to its maximum could be indicative of a blocking mechanism near the G_1-S boundary. A blocking mechanism would cause cells to pile up along the leading edge of a cluster, while leaving a trailing tail, as observed.

From the data in Figures 4.2 and 4.3, we can construct the following description. At the low point of the bud index, one cluster of newly divided cells from the previous trailing tail are entering the early G_1 phase. Simultaneously, the second cluster is leaving G_1 and entering the S-phase in the consecutive bud index upswing. The cluster entering S-phase is highly coherent relative to the length of the cell cycle. The sharp rise in the bud index indicates that the cluster has a fairly sharp leading edge.

As the bud index declines from its maximum value, the long tail indicates a loss of coherence. This is either due to dispersion in the division times of individual cells or the staggered division of different age classes due to variations in the lengths of their cell cycles, or a combination of both. In about 1 hour, approximately 30% of the cells, or 60% of the cluster must undergo division since that is the only way to leave the budded state. The other cells in the cluster must divide in the following hour by the same reasoning. This brings us back to the state with about 4% budded in a time period ~ 3.9 hours, or half the dilution doubling time. During this time period, the other half of the cells began and ended in the unbudded state, which must be G_1, and therefore must have been in G_1 during the whole period. The next rise in bud index level can only be ascribed to these cells since the cells in the first cluster have clearly not had time to pass all the way through the G_1 phase and bud again. The second group of cells then enters S-phase as a strongly coherent cluster in the same way as the previous one. This leaves only a description of what happens to the cells during G_1. From the decline in bud index level described above, the cell cluster enters G_1 in a cluster that has a fairly sharp leading edge with 60% of its cells within 30 minutes of each other, while the remaining 40% lag

behind by as much as an hour. Then as the cells pass through G_1, coherence is apparently restored, resulting in a fairly tight cluster of cells that enter S-phase again.

Our hypothesis, consistent with the data, is that the cell cluster that is approaching the G_1-S boundary is secreting a synchronizing agent that acts on the cluster entering G_1. It is not clear if the feedback is only sensed by daughter cells or by cells in G_1 of all ages. It is possible that cohesion established among daughter cells is maintained well enough with age that the system as a whole remains cohesive, coupled with the fact that successive generations decay in population density geometrically. The fact that the population structure is robustly altered by pH in a way that is qualitatively different from dilution indicates that communication is involved. Figure 4.4 demonstrates the effect of media pH at a constant dilution rate.

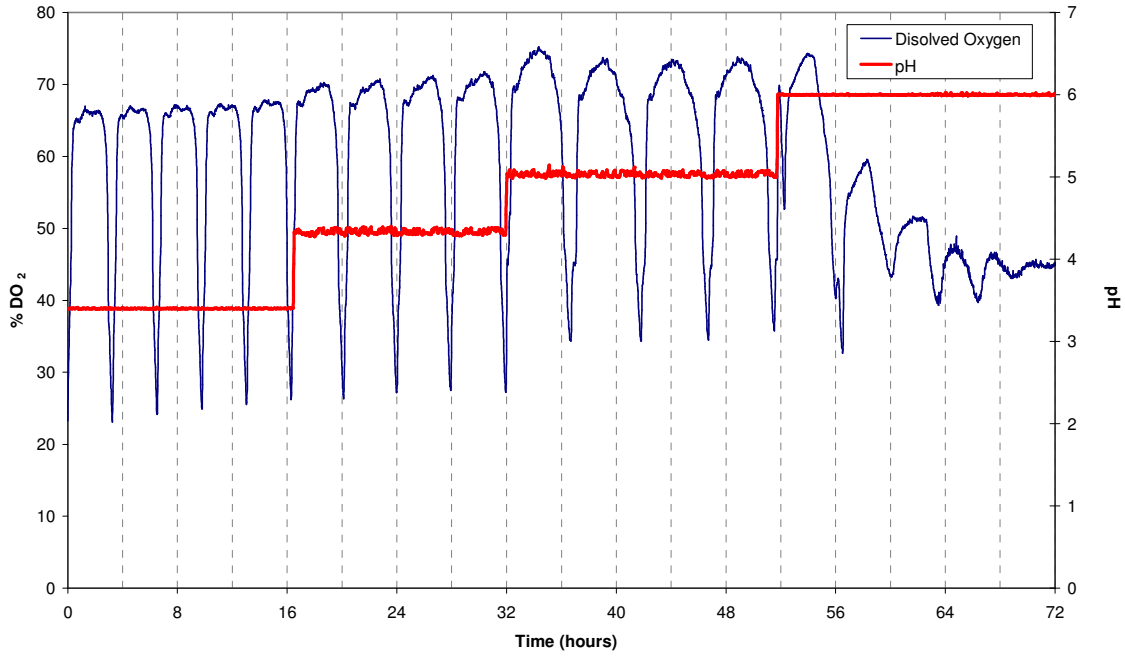


Figure 4.4: The effect of pH on the period of dissolved oxygen oscillations. The pH was increased using 2 N NaOH. The increase in pH is believed to affect inter-cellular communication and therefore alter the feedback mechanism producing the oscillations.

We have also examined the effects of acetaldehyde on an apparently asynchronous culture. Figure 4.5 depicts an initially asynchronous culture that we are compelled to believe was entrained by periodic injection of acetaldehyde. In this case, after 3 periodic injections of acetaldehyde, the system began oscillating autonomously. However, we must acknowledge the fact that these oscillations have been known to spontaneously occur, leading to the possibility that the acetaldehyde entrainment was only a coincidence. Acetaldehyde is a likely metabolic signal since it can be directly converted into ethanol and subsequently used an energy source. It has also been shown that acetaldehyde can be used to phase reset the period of the dissolved oxygen oscillations [31]. We speculate that only a fraction of cells within an asynchronous culture is responsive to pulses of acetaldehyde. For example, cells within G_1 may respond rapidly to increased nutrients, whereas cells within the S-phase, currently pre-occupied with DNA synthesis, do not respond. A response of a fraction of the cells within a culture, which causes them to accelerate through a portion of the cell cycle is an example of an Advance model and would inherently begin the formation of clusters of cells.

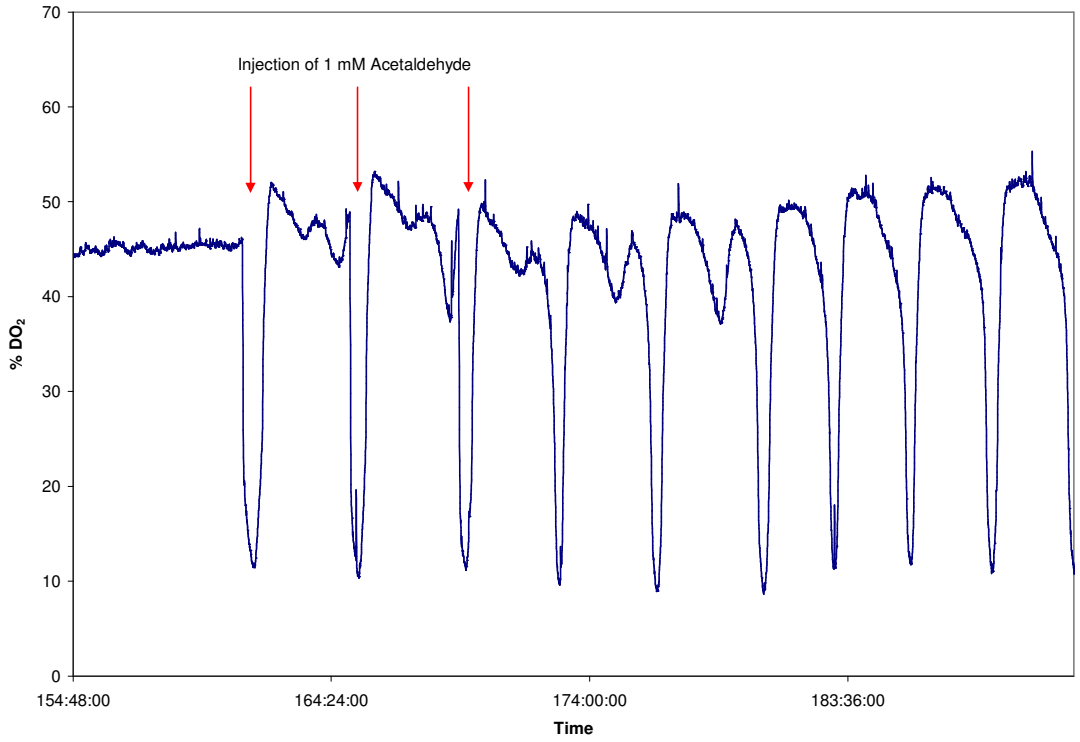


Figure 4.5: An asynchronous culture that was entrained into metabolic oscillations after the periodic addition of 1 mM acetyladehyde. Time in the figure is represented as hours:minutes:seconds. The figure spans a total time period of approximately 37 hours.

While our aim is to understand the population structure, we remark in conclusion on the connection between the population dynamics and the metabolic dynamics encoded in the dissolved oxygen oscillation. The phase of the dissolved oxygen oscillation relative to the bud index is consistent with G_1 phase oxidative metabolism and the supposed bonfire[11] of storage carbohydrates that precedes start (the beginning of the mitotic cell cycle) in slowly growing aerobic cultures [11,29,30,50]. Since G_1 is the only phase without buds, the bud index minimum and cell density maximum indicate that from 85%-95% of the cells are in G_1 prior to the large downswing in dissolved oxygen. The minimum in the bud index occurs as a cohort of newly divided cells appear as a uniform cluster in the G_1 phase. At the same time, the leading edge of the second cluster reaches bud emergence. At the dissolved oxygen minimum, slightly more than half of the cluster has budded, and the cell density profile shows only loss due to dilution, indicating that all cell division has ceased.

Factors Consistent with Clustering

The hypothesis of 2 clusters, as described above, is consistent only with cell-cycle lengths that are either nearly the same over the first few generations or if the daughter cell cycles are approximately twice as long as those of the parents. If it were otherwise, a cluster would become incoherent at cell division because generations with significantly shorter cell cycles would "get ahead" of its cluster. With more than 2 clusters, this problem could be remedied by the cells in the faster generation falling roughly into another temporally advanced cluster. For instance, if parent generations had cell cycle temporal lengths that were roughly 2/3 of the cycle length for daughter cells, then 3 clusters could be supported. Or, if the ratio was 3/4, then 4 clusters could be supported.

For the above description to be valid, the cells must spend more than half of their time in the G_1 phase. Otherwise, the two or more cluster picture, or any more dispersed profile, would not result in a bud index oscillation from nearly 0% to 50%. It has long been appreciated that metabolic alterations in cell cycle length prolong or contract G_1 [5,17,42]. This is consistent with the experiment, in which the normal cell cycle is prolonged from 2 hours to nearly 8 hours. Integration of the bud index time-series over 1 oscillation period and dividing by the period gives, averaged over time, the cells are found in the budded state 27% of the time. Taking into account the exponential decay of numbers of cells across the cell cycle, we find the average budding time to be 1/3 of the cell cycle. Precisely, we calculate the average time of bud emergence t_b from

$$\frac{\int_{t_b}^{7.8} e^{\left(-\frac{\ln(2)\,t}{7.8}\right)} dt}{\int_{0}^{7.8} e^{\left(-\frac{\ln(2)\,t}{7.8}\right)} dt} = 0.27 \qquad \text{Equation (4.1)}$$

or,

$$\frac{t_b}{7.8} = 1 - \frac{\ln(1.27)}{\ln(2)} \approx 0.66$$

Accordingly, bud emergence takes place at 5.2 hours into the cell cycle (total length of 7.8 hours). This confirms that the G_1 phase is significantly stretched.

Gene Expression Measurements

Given that we now understand that the population structure behind the metabolic oscillations impose persistent pseudo-cell cycle synchrony through clustering, these

oscillations provide a mechanism at which to make cell cycle dependent measurements. Cell cycle dependent measurements are typically confounded by population convolution introduced by asynchronous cultures. Synchronous cultures, on the other hand, provide the opportunity to detect single cell, cell cycle dependent phenomena. We first demonstrate this scenario by using a well accepted cell cycle dependent promoter called POL1. POL1 encodes for a subunit of DNA polymerase, the enzyme responsible for DNA replication during the S-phase of the cell cycle as the cell begins recruiting machinery for DNA replication. POL1 is believed to be most strongly expressed during the G_1 to S- phase transition of the cell cycle. Figure 4.6 depicts POL1 expression during a 3.75 hour oscillation. Notice that the peak of the POL1 expression corresponds to the end of the rise of the dissolved oxygen profile. It is at this point that we believe a cluster is passing from G_1 into the S phase of the cell cycle. This claim is supported by the data within Figure 4.3, which demonstrate that the bud index peaks here as well.

Having demonstrated the utility of the pseudo cell cycle synchrony imposed by metabolic oscillations, by analysis of POL1 expression, other more esoteric promoters could be investigated. We are particularly interested in the expression profiles of genes associated with NCR in yeast. Previous mathematical analysis of the connectivity of the NCR genetic circuitry produced the hypothesis that the expression of the genes GLN3 and DAL80 would oscillate in phase with the cell cycle [2]. Figure 4.7 depicts the expression profile of GLN3 measured under two separate nitrogen sources, ammonia and arginine. The first, and most trivial, observation that can be made from the GLN3 expression data is that the oscillations persist under arginine as the sole nitrogen source. Within the 40 years of literature surrounding these oscillations, we are unaware of any data

demonstrating the existence of oscillations under any nitrogen source other than ammonia. In fact, we have been able to maintain the oscillations under a variety of nitrogen sources including allantoin, glutamate, and glutamine (data not shown). The data also indicate that GLN3 is periodic in nature during both ammonia and arginine nitrogen sources. Interestingly, the expression profile under ammonia is bimodal, perhaps due to the existence of two clusters within the system, whereas under arginine feed, the expression profile appears periodic, but uni-modal. Upon nitrogen switching, the system responds quickly, indicating that nitrogen is either the limiting or the near limiting substrate of the culture. Another distinct feature in the GLN3 expression profile is the delta pulse shaped signal just after switching to arginine. This feature, although currently a mystery, may indicate key features regarding the regulation of the NCR circuitry.

Figure 4.8 depicts the expression profile for DAL80 during the autonomous dissolved oxygen oscillations. The expression profile supports the claim that DAL80 periodically oscillates in phase with the cell cycle. However, DAL80 appears to be relatively unaffected by nitrogen source. The only notable characteristic change upon switching from glutamine to arginine is the development of a tail at the end of the expression profile. This tail, indicated by the black arrow in Figure 4.9, does not exist under glutamine, a relatively good nitrogen source, but becomes increasingly more noticeable under a relatively poor nitrogen source such as arginine.

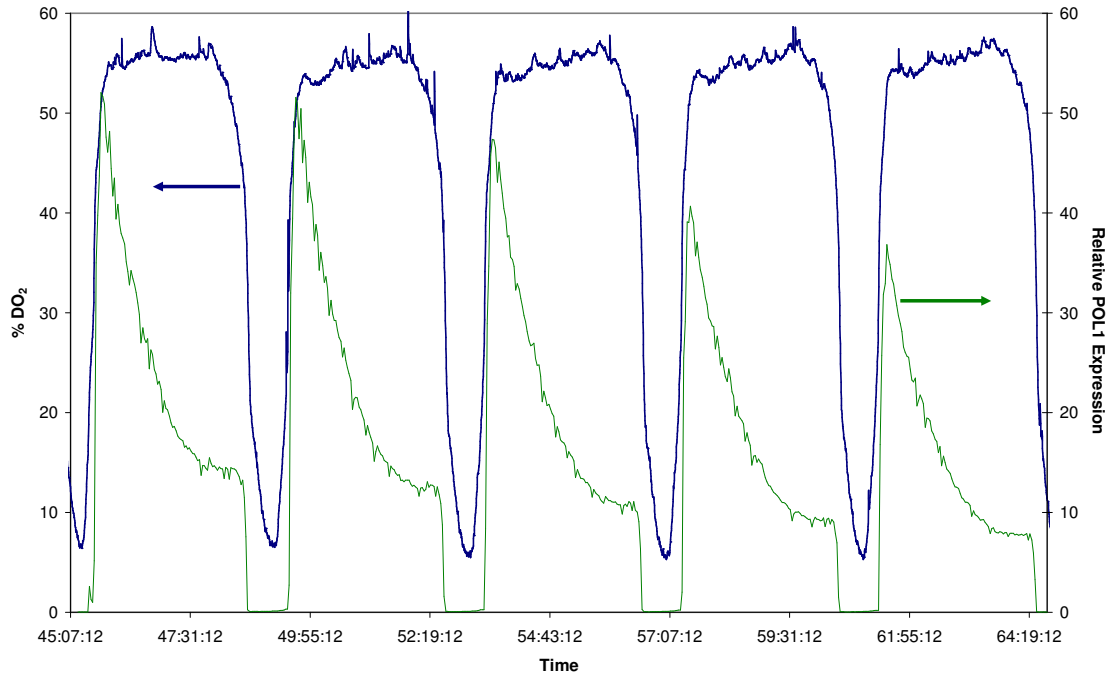


Figure 4.6: POL1 expression (green line) during autonomous dissolved oxygen oscillations (blue line). The data indicate cell cycle dependent expression, which peaks during the upswing of a dissolved oxygen oscillation. Time is shown as hours:minutes:seconds. The figure spans a time period of approximately 22.5 hours.

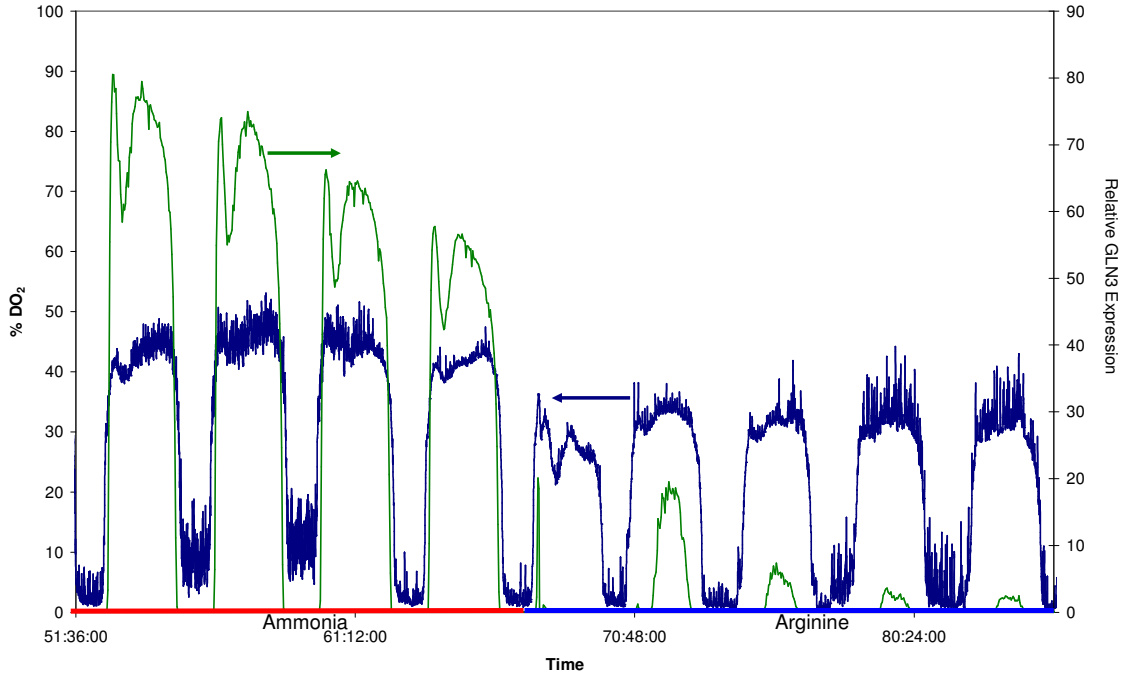


Figure 4.7: GLN3 expression (green line) during autonomous dissolved oxygen oscillations (blue line) under ammonia and arginine nitrogen sources. The nitrogen source is indicated by the color of the horizontal axis. Red represents ammonia as the nitrogen source, whereas blue represents arginine. The data support the claim that GLN3 is periodic with respect to the cell cycle. Time is shown as hours:minutes:seconds. The figure spans a time period of approximately 36.5 hours.

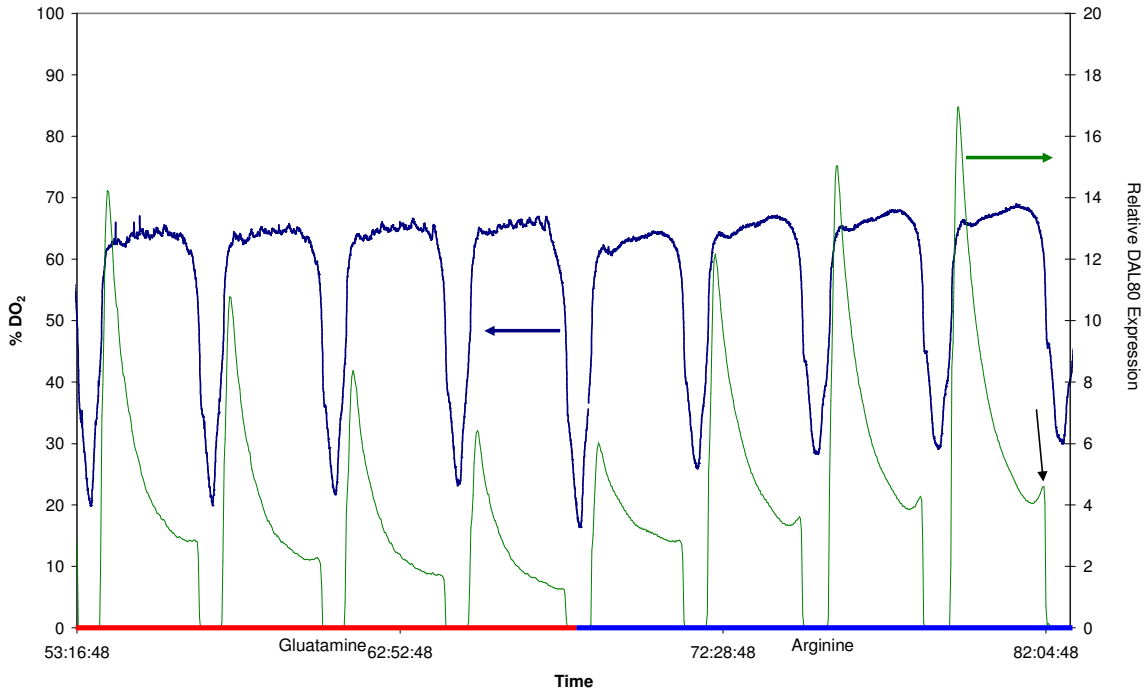


Figure 4.8: The expression profile of DAL80 (green line) during autonomous dissolved oxygen oscillations (blue line) under two different nitrogen sources, glutamine and arginine. The nitrogen source is represented by the color of the horizontal axis. Red represents glutamine and blue represents arginine. The data support the claim that DAL80 is periodic with respect to the cell cycle. Time is shown as hours:minutes:seconds. The figure spans a time period of approximately 29 hours.

Conclusions

Bud index data along with the cell density data for the Cen.PK strain conclusively show that clustering occurs. The experimental evidence supports the conclusion that 2 independent critical clusters exist with each containing roughly half the total cells. Further, the clustering of the population density and its subsequent cell cycle progression accounts for the observed dissolved oxygen oscillations. We have demonstrated that clustering, a form of cell cycle synchrony, can be exploited to obtain cell cycle dependent gene expression measurements over many cell cycles. The measurement of periodic GLN3 and DAL80 expression profiles support the predictions of mathematical theory that have otherwise remained untested. The utility of this system for cell cycle dependent measurement appears endless.

Mathematical analyses demonstrate that advance and delay feedback mechanisms can initiate clustering over a wide range of parameter values and initial conditions. The conclusion is that clustering is a robust phenomenon that can arise from intercellular communication.

What remains unresolved are the specific details of the feedback mechanism that entrains the cells. While oxygen has been shown to be a regulator of quorum sensing molecules [43], and interesting hypotheses have been developed for why it might act in that capacity in the Cen.PK oscillations [45], its involvement in the hierarchy of cause and effect has not been convincingly elucidated. We have raised the potentially contentious point that the so called short period aerobic oscillations in the IFO0223 strain that have been reported

to appear in the absence of cell cycle oscillations when the cells are grown in limiting glucose or ethanol, is unsubstantiated by the published data. We do not believe this unsubstantiated conclusion, which has gone unchallenged for too long in the literature. We have provided bud index data that clearly show that oscillations occur within the window of dilution rates for which it has been claimed otherwise without data, see Figure 4.1. It is time that this issue is resolved, lest it remain to further confound an already complex problem.

The connection between cell cycle synchrony and metabolic synchrony is a topic of great interest. Based on the evidence and our point of view, we speculate that the causal and logical connections are these. Actively dividing cells must be progressing through their cell cycle. Secreted byproducts of metabolism accumulate in the culture media and at a high enough cell density that the metabolic response of the cells to their presence induces advances and/or delays in their cell cycle progression, which depends on the cell cycle position of the receiving cell. The existence of this feedback mechanism causes a population structure to emerge that is characterized by clusters of cells that progress through the cell cycle together. If the cell cycles of the generations are comparable then an overall population synchrony can be observed as oscillations in environmental variables.

References

1. Boczko, E., Gedeon, T., Stowers, C., and Young,T. 2008. *Advance and delay models and cell cycle clustering in yeast*. Preprint

2. Boczko, E.M., Cooper, T.G., Gedeon, T., Mischaikow, K., Murdock, D.G., Pratap, S., and Wells, K.S. 2005. Structure theorems and the dynamics of nitrogen catabolite repression in yeast. *Proc. Natl. Acad. Sci.* **102**: 5647-5652.

3. Boye, E., Stokke, T., Kleckner, N., and Skarstad, K. 1996. Coordinating DNA replication initiation with cell growth: Differential roles for DnaA and SeqA proteins. *PNAS*. **93**:12206-12211.

4. Breeden, L.L. 1997. α-Factor synchronization of budding yeast. *Methods in Enzymology.* **283**: 332-342.

5. Brewer, B., Sledziewska, E., and Fangman, W. 1984. Cell cycle phases in the unequal mother/daughter cell cycles *of Saccharomyces cerevisiae*. *Mol. Cell Biol*. **4**: 2529-2531.

6. Buese,M., Kopmann, A., Diekmann, H., and Thoma, M.1998. Oxygen, pH value, and carbon source induced changes of the mode of oscillation in synchronous continuous culture of *Saccharomyces cerevisiae*. *Biotechno. Bioeng*. **63**: 410-417.

7. Chen, H., Fujita, M., Feng, Q., Clardy, J., and Fink, G. 2004. Tyrosol is a quorum sensing molecule in *Candida albicans*. *PNAS*. **101**:5048-5052.

8. Chen, Z., Odstrcil, E., Tu, B., McKnight, S. 2007. Restriction of DNA replication to the reductive phase of the metabolic cycle protects genome integrity. *Science*. **316**:1916-1919.

9. Dano, S., Sorensen, P. and Hynne, F. 1999. Sustained oscillations in living cells. *Nature*. **402**:320-322.

10. Dunny, G. Leonard, B. 1997. Cell cell communication in gram positive bacteria. *Annu. Rev. Microbiol*. **51**:527-564.

11. Futcher, B. 2006. Metabolic cycle, cell cycle and the finishing kick to start. *Genome Biology.* **7**:107-111.

12. Heinzle, E., Dunn, I., Furukawa, K., and Tanner, R. D. 1982. Modeling of sustained oscillations observed in continuous culture of *Saccharomyces cerevisiae*. *Proceedings of the 1st IFAC Workshop.* A. Halme, ed. Pergamon Press (Oxford). 57-65.

13. Helmstetter, C., Thornton, M., Romero, A., and Edward, K. 2003. Synchrony in human, mouse, and bacterial cell cultures—a comparison. *Cell Cycle*. **2**: 42-45.

14. Hense, B., Kuttler, C., Muller, J., Rothballer, M., Hartmann, A., and Kreft, J. 2007. Does efficiency sensing unify diffusion and quorum sensing? *Nat. Rev. Microbiol.* **5**:230-239.

15. Henson, M. 2005.Cell ensemble modeling of metabolic oscillations in continuous yeast cultures. *Comput. Chem. Enginer*. **29**:645-661.

16. Herker, E., Helmut, J., Lehmann, K., Maldener, C., Frohlich, K., Wissing, S., Buttner, S., Fehr, M., Sigrist, S., and Madeo, F. 2004. Chronological aging leads to apoptosis in yeast. *JCB*. **164**: 501-507.

17. Hersh, R. and Kitos, P. 1980. Is G_1 normally distributed? *J. of Theor. Biol*. **86**: 117-122.

18. Hjortso, M. and Nielsen, J. 1994. A conceptual model of autonomous oscillations in microbial cultures. *Chemical Engineering Science.* **49**:1083-1095.

19. Hornby, J., Jensen, E., Lisec, A., et al. 2001. Quorum sensing in the dimorphic fungus *Candida albacans* is mediated by farnesol. *Appl. Environ. Microbiol.* **67**:2982-2992.

20. Johnston, G.C. and Singer, R.A. 1983. Growth and the cell cycle of the yeast *Saccharomyces cerevisiae. Exp. Cell. Res*. **149**:1-13.

21. Jones, K. and Kompala, D. 1999. Cybernetic model of the growth dynamics of *Saccharomyces cervisiae* in batch and continuous culture. *J.of Biotechnol*. **71**:105-131.

22. Jorgensen, P. and Tyers, M. 2004. How cells coordinate growth and division. *Curr. Biol*. **14**:R1014-R1027.

23. Keulers, M., Suzuki, T., Satroutdinov, A., and Kuriyama, H. 1996. Autonomous metabolic oscillation in continuous culture of Saccharomyces cerevisiae grown on ethanol. *FEBS Microbiol. Lett.* **142**: 253-258.

24. Klevecz, R. and Murray, D.2001. Genome wide oscillations in expression, *Molecular Biology Reports.* **28**:73-82.

25. Koschwanez, J., Holl, M., Marquardt, B., Dragavon, J., Lloyd, B., and Melldrum, D. 2004. Identification of budding yeast using a fiber-optic imaging bundle. *Review of Scientific Instruments*. **75**: 1363-1365.

26. Lord, P.G. and Wheals, A.E. 1980. Asymmetrical division of *Saccharomyces cerevisiae. J. Bacter*. **142**:808-818.

27. Lyon, G., and Novick, R. 2004. Peptide signaling in *Staphylococcus aureus* and other gram positive bacteria. *Peptides.* **25**:1389-1403.

28. Mashego, M. Jansen, M., Vinke, J., van Gulik, W., and Heijnen, J. 2005.Changes in the metabalome of *Saccharomyces cervisiae* associated with evolution in aerobic glucose-limited chemostats. *FEMS Yeast Res*. **5**:419-430.

29. Muller, D., Exler, S., Aguilera-Vazquez, L., Guerrero-Martin, E. and Reuss, M.2003. Cyclic AMP mediates the cell cycle dynamics of energy metabolism in *Saccharomyces cervisiae*. *Yeast*. **20**:351-367.

30. Munch, T., Sonnleitner, B., and Fiechter, A. 1992. The decisive role of the *Saccharomyces cerevisiae* cell cycle behavior for dynamic growth characterization. *J. of Biotechnol*. **22**:329-352.

31. Murray, D. Klevecz, R., and Lloyd, D. 2003.Generation and maintenance of synchrony in *Saccharomyces cerevisiae* continuous culture. *Experimental Cell Research*. **287**:10-15.

32. Murray, D., Englelen, F., Lloyd, D., and Kuriyama, H. 1999. Involvement of glutathione in the regulation of respiratory oscillation during a continuous culture of *Saccharomyces cerevisiae. Microbiology.* **145**: 2739-2745.

33. Murray, D., Roller, S., Kuriyama, H., and Lloyd, D. 2001. Clock control of ultradian respiratory oscillation found during continuous yeast culture. *J. of Bacteriology*. **183**: 7253-7259.

34. Nurse, P. 1975. Genetic control of cell size at cell division in yeast. *Nature.* **256**: 547-551.

35. Palkova, Z. and Vachova, L. 2006. Life within a community: benefits to yeast long term survival. *FEMS. Microbiol. Review*. **30**:806-824.

36. Palkova, Z., and Forstova, J. 2000. Yeast colonies synchronize their growth and development. *J.of Cell Science*. **113**:1923-1928.

37. Pohley, H. 1987. A formal mortality analysis for populations of unicellular organisms (*Saccharomyces cerevisiae*). *Mech. Ageing Dev*. **38**: 231-243.

38. Richard, P., Bakker, B., Teusink, B., van Dam, K., and Westerhoff, H. 1996. Acetaldehyde mediates the synchronization of sustained gycolytic oscillations in populations of yeast cells. *Eur. J. of Biochem*. **235**: 238-241.

39. Richard, P. 2003. The rhythm of yeast. *FEMS Microbiol. Rev*. **27**:547-557.

40. Satroutdinov, A., Kuriyama, H., and Kobayashi, H. 1992. Oscillatory metabolism of *Saccharomyces cerevisiae* in continuous culture. *FEBS Microbiol. Lett.* **98**: 261-268.

41. Sinclair, D., Mills, K., Guarente, L. 1998. Ageing in *Saccharomyces cerevisiae*. *Annu. Rev. Microbiol.* **52**: 533-560.

42. Singer, R. and Johnston, G. 1981. Nature of the G_1 phase of the yeast *Saccharomyces cerevisiae*. *PNAS*. **78**: 5030-5033.

43. Smith, R. and Iglewski, B. 2003. *Psuedomona aeruginosa* quorum sensing as a potential antimicrobial target. *J. Clin. Invest.* **112**: 1460-1465.

44. Stowers, C., Hackworth, D., .Gedeon, T., Mischaikow, K., and Boczko, E. 2008. Extending cell cycle synchrony and deconvolving population effects in budding yeast through an analysis of volume growth with a structured Leslie model. *Theoretical Population Biology*. Submitted.

45. Tu, B., Kudlicki, A., Rowicka, M., and McKnight, S. 2005. Logic of the yeast metabolic cycle: temporal compartmentation of cellular processes. *Science*. **310**:1152-1158.

46. Walker, G. 1999. Synchronization of yeast cell populations. *Methods in Cell Science.* **21**:87-93.

47. Wang, J., Liu, W., Uno, T., Tonozuka, H., Mitsui, K., and Tsurugi, K. 2000. Cellular stress response oscillate in synchronization with the ultridian oscillation of energy metabolism in the yeast *Saccharomyces cerevisiae*. *FEMS Microbiol. Lett*. **189**:9-13.

48. Wheals, A.E. 1982. Size control models of *Saccharomyces cerevisiae* cell proliferation. *Molecular and Cellular Biology.* **2**:361-368.

49. Woldringh, C.L., Huls, P.G., and Vischer, N. 2003. Volume growth of daughter and parent cells during the cell cycle of *Saccharomyces cerevisiae* a/α as determined by image cytometry. *J. Bacteriology.* **175**:3174-3181.

50. Xu, Z. and Tsurugi, K. 2006. A potential mechanism of energy metabolism oscillation in an aerobic chemostat culture of the yeast *Saccharomyces cerevisiae*. *FEBS Journal*. **273**:1696-1709.

51. Zhu, G., Zamamiri, A., Henson, M., and Hjortso, M. 2000. Model predictive control of continuous yeast bioreactors using cell population balance models. *Chemical Engineering Science.* **55**:6155-6167.

CHAPTER V

PERIODIC FERMENTOR YIELD AND ENHANCED PRODUCT ENRICHMENT FROM AUTONOMOUS OSCILLATIONS

Introduction

Because yeast and other fungi are genetically tractable, double rapidly, and have the ability to post-translationally modify proteins as required for their efficacy as human drugs, they have become an increasingly attractive platform for bioprocessing. Yeast are being used to manufacture insulin and hydrocortisone as pharmaceuticals[10,18,27,29], and to produce invertase and amylase for the biofuels industry[7,31]. Despite the evolved efficiency and robustness of cellular processes, most of the biotechnological processes that utilize living cells operate at modest efficiencies [3, 23]. In fact, single steps of the bio-production process have been shown to be less than 50% efficient [18], whereas the highly developed commodity chemical industry routinely operates at efficiencies greater than 90%.

Inefficiencies arise in protein purification and isolation[24]. For example, products of interest may be produced, but not secreted from the cell. Harvesting cellular proteins is challenging. While the cell cycle itself is relatively well understood in yeast, only a limited understanding of its quantitative impact on bioprocess-related parameters exists. For instance, yeast cell disruption has only recently been shown to have strong cell cycle dependence[25]. During and following cell disruption, protein products may be degraded due to the liberation of normally sequestered proteases. Protease knockout strains have

been extensively utilized for this reason[12]. Moreover, many proteins share primary through quaternary structural similarities that can confer upon them similar physical-chemical properties that confound their separation and efficient isolation. Most proteins perform their biological functions in complexes; and accordingly, have adapted surface interaction domains. These adaptive features work against efficient isolation.

For the purposes of this chapter, the protein background that a protein product of interest is expressed among is designated, its *ambient biocomplexity*. The larger and more diverse this background, the higher is its' ambient biocomplexity. The confounding effects of proteins that are physically and chemically similar to the desired product motivate this definition. The more proteins in an admixture that are physically and chemically similar to the desired product, the more steps will be required to isolate the product away from its mimics or contaminating biomolecules. It stands as self-evident that increased ambient biocomplexity can only reduce bioprocess efficiency. This chapter is concerned with exploring optimal strategies to reduce a desired products ambient biocomplexity and consequently increasing its efficient harvest by maximizing its enrichment ratio relative to impurities.

Many proteins can be and are produced in a cell cycle dependent manner. The cyclin genes that regulate cell cycle progression are an example. Currently, a detailed proteomic catalog of the type and magnitude of expressed yeast proteins as a function of the cell cycle is not available. It has been shown that total protein production increases throughout the cell cycle [1,32]. As a consequence, the lowest biocomplexity might occur near the beginning of the G_1 phase of the cell cycle. Such data suggest that

biocomplexity might fluctuate as a function of the cell cycle or could be manipulated to do so.

In an asynchronously doubling culture, cell cycle dependent products are produced asynchronously, although continuously. It has been observed that periodically harvesting a periodic product can improve yield relative to continuously harvesting from an asynchronous culture[16]. Conventional methods to produce cell cycle synchrony include metabolic blocks involving starvation, mating pheromones such as alpha-factor, chemical treatment such as nocodozole, and the preparation of homogeneous initial populations by physical means such as elutriation [33]. Typically only 3-4 decaying cycles of synchrony can be achieved with these methods. In Chapter 3, we developed theoretical continuous filtration methods that guarantee the maintenance of synchrony over 30 consecutive cell cycles, but these are not yet generally implemented[26]. None of these methods are cost-effective however, and likely the reason why synchronous yeast cultures are rarely used in industry. However, there is another option.

A few strains of yeast exhibit sustained oscillations under a limited variety of conditions[22]. Autonomously oscillating systems are attractive for at least three reasons. First, the oscillations reflect a form of cell cycle synchrony whose population structure we believe we understand. This level of understanding allows for the opportunity to interleave expression out of phase with ambient biocomplexity. The population synchrony induced via autonomous oscillations ensures that the synchronous fabrication of products occurs with no additional cost. The value of synchronous production of products has been demonstrated with *Bacillus subtilis*[9]. Finally, since autonomous

oscillations are stable, often for months, costs associated with quality controlling lots can be reduced, as well as costs associated with production interruptions and restarts incurred during batch processing.

Microbial oscillations have generally been regarded with trepidation due to their perceived instability [17]. Forty years of data and experience show that in fact autonomous oscillations are highly controllable [4,5,6,8,11,15,19,30,34]. In fact, we show that they can be used to great advantage. We show that regardless of a product's ambient biocomplexity there exists an expression profile that produces an accumulation of the product that dramatically improves product yield and its enrichment ratio. Moreover, we show how to utilize parallel, phased, periodic sampling to create continuous enrichment in a scheme we have dubbed "CLOCKS". We believe this point of view is novel. These developments are important since down-stream processing costs account for over 75% of cost of manufacturing a drug that in turn translates into 1525% of the retail cost of a bio-product[2]. Enrichment of bio-products within the effluent of fermenters directly increases their value exponentially, see Appendix A. Appendix A of this dissertation demonstrates that nominal increases in the concentration of a bio-product can lead to dramatic increases in the monetary value of a product. Non-growth associated products, such as the production of alpha amylase from *Bacillus licheniformis* [21], are not subject to the analysis that we present. Non-growth assosciated products are produced during the stationary phase of the growth curve during which cells become arrested in the cell cycle and are therefore no longer subject to cell cycle dependent production of bioproducts.

Materials and Methods

Autonomous Oscillations

The haploid (mat a) strain CEN.PK 113-7D of *Saccharomyces cerevisiae* were cultivated in a 3L New Brunswick Scientific Bioflow 110 reactor equipped with two Rushton type impellers, operated at 550 rpm. Air was sparged in at a rate of 900 mL/min. Temperature was maintained at 30 °C and pH was automatically controlled with 2 N sodium hydroxide.

Overnight cultures were grown in YPD media with shaking. 20 mL were innocculated into a reactor containing 850 mL of synthetic media with the following composition: 10 gm/L anhydrous glucose (Sigma #G71528), 5 gm/L ammonium sulfate (Sigma # A2939), 0.5 gm/L magnesium sulfate heptahydrate (Sigma # M2773), 1 gm/L of Yeast Extract (Becton, Dickinson and Company Cat # 288620), 2 gm/L potassium phosphate (Sigma # P5379), 0.5 mL per L of 70% v/v sulfuric acid, 0.5 mL of antifoam A (Sigma # 10794), 0.5 mL of 250 mM calcium chloride, and 0.5 mL per L of mineral solution A, see Table 4.1.

The culture was grown in batch for 12-16 hours to a cell density of approximately $5x10^9$ cells/mL. Continuous culture ensued with a dilution rate of 0.095 hr^{-1}. Samples were taken from the reactor and flash frozen using dry ice.

Measurement of Cell Density

Cell samples were thawed in a chilled (4 °C) sonicating water bath for 5 minutes and vortexed briefly. A 1 µL sample was resuspended in 10 mL of Isoton Diluent (Beckman Coulter # 8546719) and vortexed. The cell density of three independent 0.5 mL samples was measured using a Beckman Multisizer Coulter counter. The sizing threshold was 2-8µm.

Measurement of Bud Index

Samples were collected at 15 minute intervals and immediately frozen in an ethanol-dry ice bath. Buds were counted with a 100X objective on a conventional Nikon TE-2000 microscope. A minimum of 300 cells were scored at each time point. Cells without buds were considered to be within the G_1 phase of the cell cycle.

Calculation of Yield

Yield was computed by solving the mass balance for substrate:

$$Y(X,S) = \frac{\mu X}{D[S_F - S] - \frac{dS}{dt}} \qquad \text{Equation (5.1)}$$

Where *Y* represents the yield of cells per unit of substrate, µ is the population growth rate, *X* is the cell density, S is the substrate concentration, S_F is the concentration of the substrate in the feed, and *D* is the dilution rate. A similar equation can be derived to represent the yield of cells in terms of the dissolved oxygen content:

$$Y(X, O_2) = \frac{\mu X}{D[O_{2F} - O_2] - \frac{dO_2}{dt}}$$ Equation (5.2)

The population growth rate, μ, was determined from cell biomass and cell number data assuming the model.

$$\mu = \frac{1}{X}\frac{dX}{dt} + D$$ Equation (5.3)

Table 5.1: Parameter values used to evaluate Equations (5.1), (5.2), and (5.3).

Parameter Values		
Parameter	Value	Units
$S_{F,Glucose}$[15]	20	Gm/L
$S_{F,Glucose}$[current exp.]	10	Gm/L
$S_{F,Oxygen}$	7.6	Mg/L

The data of Heinzle et al.[15] were linearly interpolated to account for differences in sampling times and frequency. All numerical manipulations were performed using Mathematica version 5.2 from WolframResearch. The dissolved oxygen trace was recorded every 12 seconds. The dissolved oxygen measurements displayed both high frequency measurement noise and rapid oscillation. Five measurements per minute were averaged to produce a data set of 391 measurements covering one cell cycle. These data points were fit with a cubic spline using the Interpolation function. The analytical derivative of the interpolation was contaminated with high frequency noise. Two options were considered: Further smooth the raw oxygen data or smooth the derivative. Since

the derivative correctly captured the global features of the oxygen oscillations, and further smoothing of the raw data could lead to small temporal shifts of relevant features, we present the latter solution. Comparison of the methods showed no qualitative difference. The dissolved oxygen derivative data were smoothed using a Gaussian kernel in the 'ListConvolve' function of Mathematica, with a standard deviation corresponding to a 2.5 minute time window and no overhangs.

The cell density is a slowly varying function of the cell cycle. The density data were collected every 15 minutes and were not contaminated with high frequency noise. The data were interpolated with a cubic spline. The growth rate was calculated according to Equation 5.3, using the analytical derivative of the interpolating function. A dilution rate of 0.095 hr^{-1}, and an oxygen feed of 100 percent were assumed using the yield formula in Equation (5.2)

Results

Analysis of Fermentor Yield

Yield describes the gain in a product per unit input. Yield coefficients have been an important benchmark of fermentor efficiency for over 30 years [8]. The yield of cell mass in terms of glucose substrate, $Y_{X/S}$, and dissolved oxygen, $Y_{X/O2}$, were computed from the data of Heinzel et al. [15] for the yeast strain LBGH-1022, see Figures 5.1 and 5.2. The yield, $Y_{X/O2}$, was calculated independently for the strain Cen.PK from data collected in our laboratory, see Figure 5.3 . The combined results indicate that the yield oscillates with a period equal to that of the dissolved oxygen. We are unaware of any published

results of this nature. Since substrate concentration is linked to growth rate, large substrate oscillations are likely to entrain cell cycle synchrony [16] and oscillating yield. The magnitude of the yield is influenced by the phase of the substrate oscillation relative to that of product/biomass, with the largest amplitude occurring when the signals are perfectly out of phase. In the case of the Heinzle et al. data[15], the glucose profile is nearly in phase with the biomass, whereas the oxygen profile is far out of phase, as shown in Figure 5.1. As a result, $Y_{X/O2}$ is larger in amplitude than $Y_{X/S}$. Many compounds, such as ethanol, glycogen, NADPH, and acetaldehyde, have been shown to oscillate during long period aerobic oscillations in several strains [15, 34]. Therefore, it is expected that the yield with respect to these compounds oscillates periodically as well.

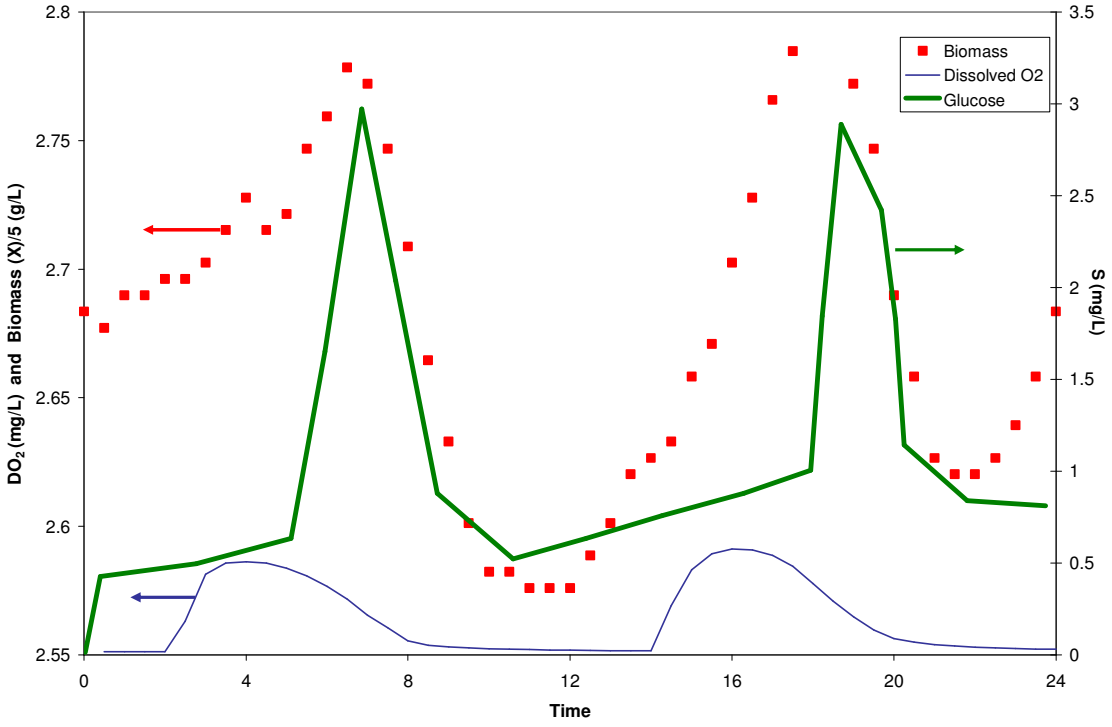


Figure 5.1: Bioprocess data from Heinzle et al [15]. The dissolved oxygen concentration (thin blue line) peaks before the glucose concentration (bold green line). The biomass (red squares) peaks roughly in phase with the substrate glucose. The biomass data are scaled to accommodate the data within the same figure.

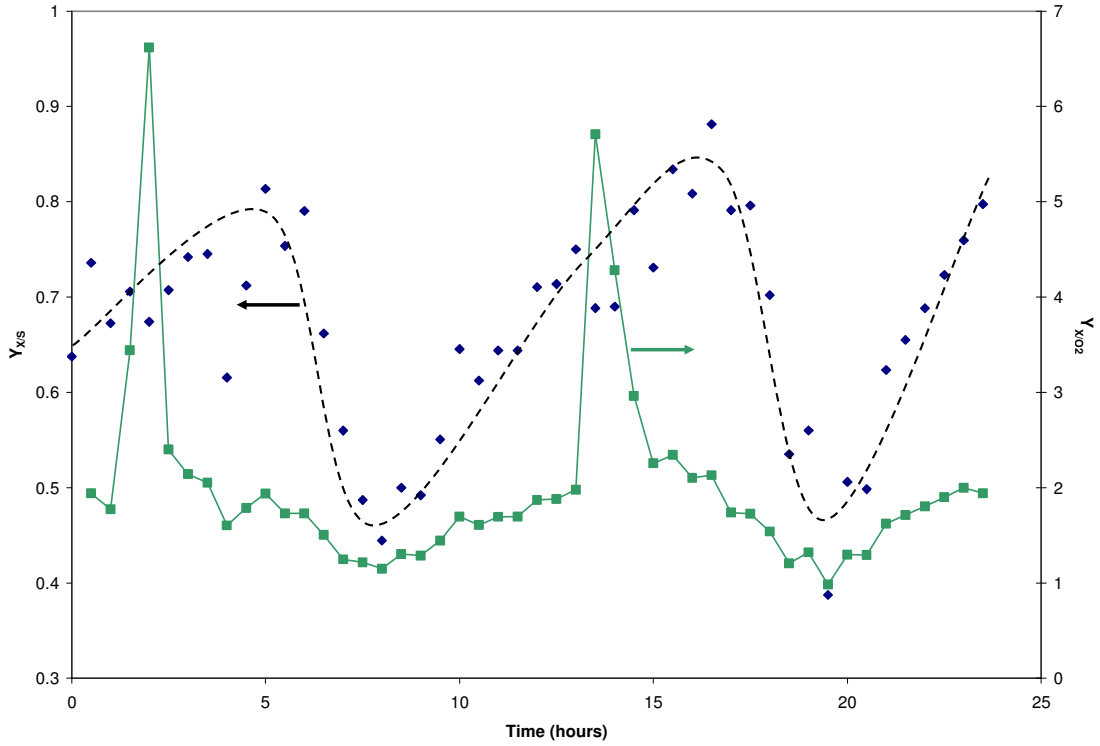


Figure 5.2: The time-dependent biomass yield relative to glucose consumption computed from the data of Heinzle et al[15], shown in Figure 5.1, using equations (5.1) and (5.3). The blue data points and dashed black line represent the yield of biomass relative to the glucose concentration. The green data points and trace represent the biomass yield with respect to oxygen concentration. These data imply product yield oscillation.

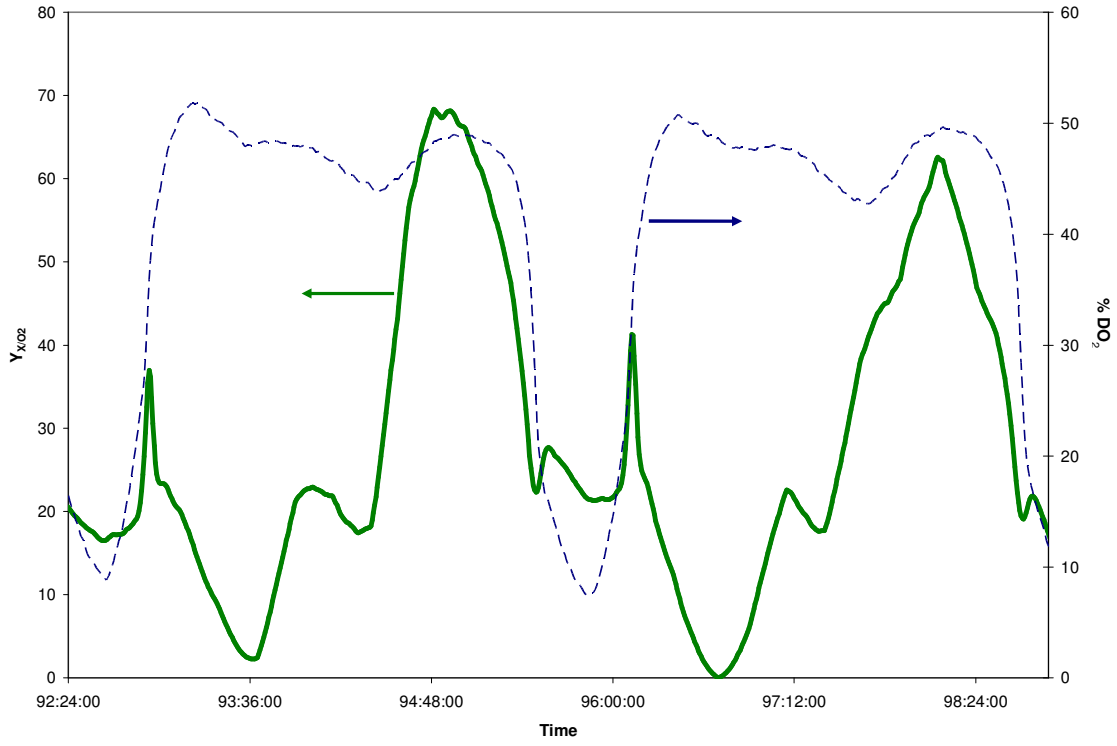


Figure 5.3: Complex, periodic, biomass yield oscillation as a function of dissolved oxygen consumption. $Y_{X/O2}$ (green) computed using data shown in Figure 5.5 and Equation (5.3) as described in the text. In this case, $Y_{X/O2}$ is not dimensionless and has the units of % DO_2 per cell. Dissolved oxygen is shown in dashed blue. Time is shown as hours:minutes:seconds elapsed since culture inception. The figure spans a total of 390 minutes.

Population Structure, Biocomplexity and the Product Enrichment Ratio

An analysis of the bud index and cell number time series demonstrates that autonomous oscillations in strain Cen.PK 113-7D are integrally linked to the cell cycle, see Figures 5.4 and 5.5. The link being that the cell population is divided into temporally coherent clusters of cells that travel around the cell cycle almost exactly out of phase with regard to bud emergence and division, see Figure 5.6. Figure 5.4 demonstrates that during the low portion of a dissolved oxygen oscillation nearly all of the cells exist within the G_1 portion of the cell cycle. This follows a maximum in the cell density, indicating that a wave of cell division has just ended, see Figure 5.5. With no new increase in cell number, the percent of cells within G_1 subsequently falls to 50%, indicating that one of the two clusters has passed into the S-phase of the cell cycle. The steepness of the transition indicates the temporal cohesion of the cell population that comprises half the total number of cells. Since each culture doubling time is marked by exactly two oscillations in the bud index and the cell density, it is concluded that the population structure under autonomous oscillations is as depicted in Figure 5.6.

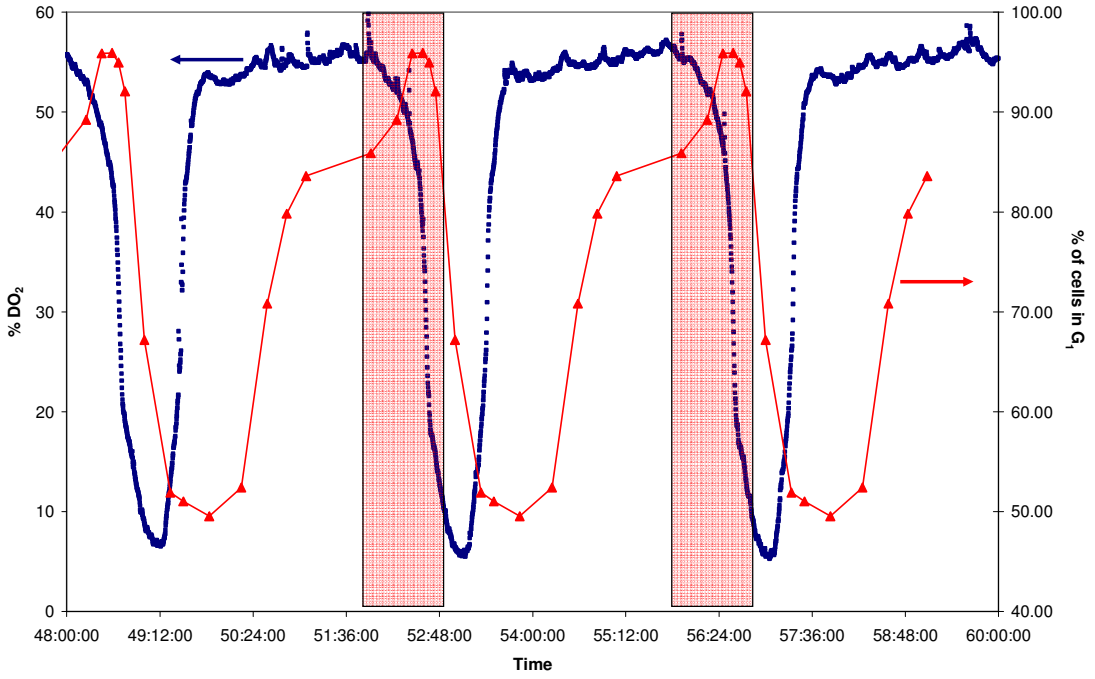


Figure 5.4: Autonomous oscillations of Cen.PK 113 in a 3L bioreactor operated at D=0.088 hr^{-1}. Dissolved oxygen, as the percentage of saturation, is shown in blue. The fraction of cells in the G_1 phase of the cell cycle, are plotted as red triangles. The red shaded regions represent a maximal G_1 harvesting window within the oscillation. As described in the text, this harvesting window may represent a minimum in total biocomplexity. The complement of the bud index data represents the percentage of cells in G_1 [13,14]. The bud index data were recorded over the central dissolved oxygen oscillation and its periodic extension is shown.

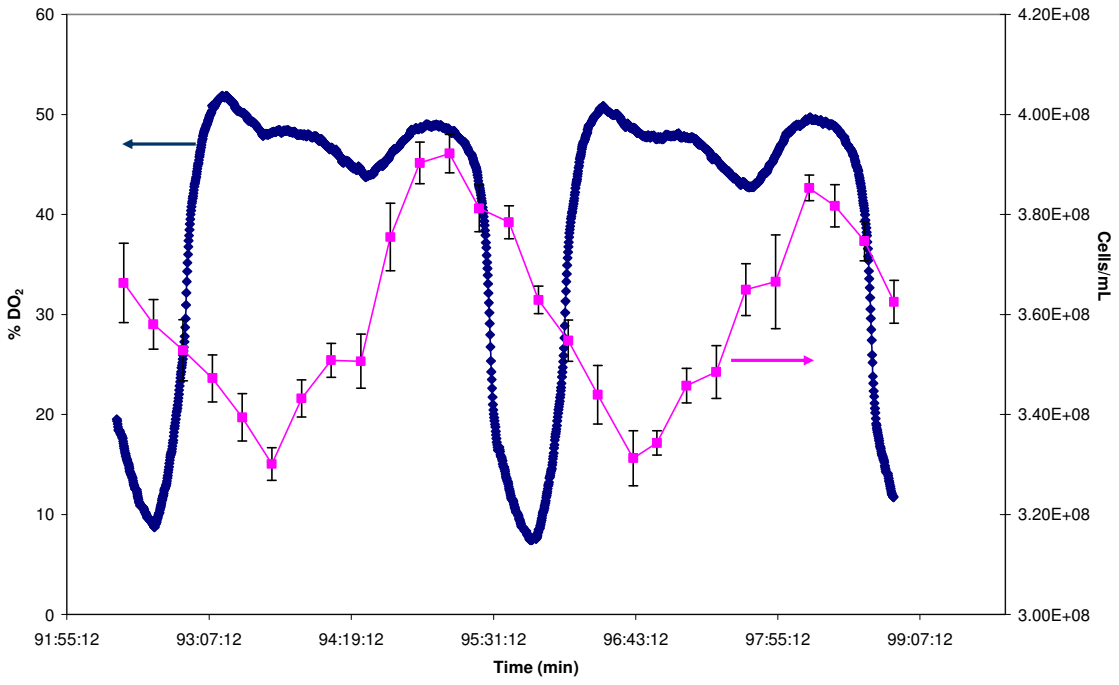


Figure 5.5: Cell density time series. Cell density (pink squares) was measured at regular 15 minute intervals across two 3.5 hour dissolved oxygen oscillations that together constitute a doubling time under an imposed dilution rate of D=0.095 hr^{-1}. Cell density as measured with a Coulter counter changes as a function of division and dilution. The data indicate a concerted division event precedes each wave of respiration. Error bars represent the standard deviation of the measurement. Time is shown as hours:minutes:seconds elapsed since culture inception. The figure spans a total of 390 minutes.

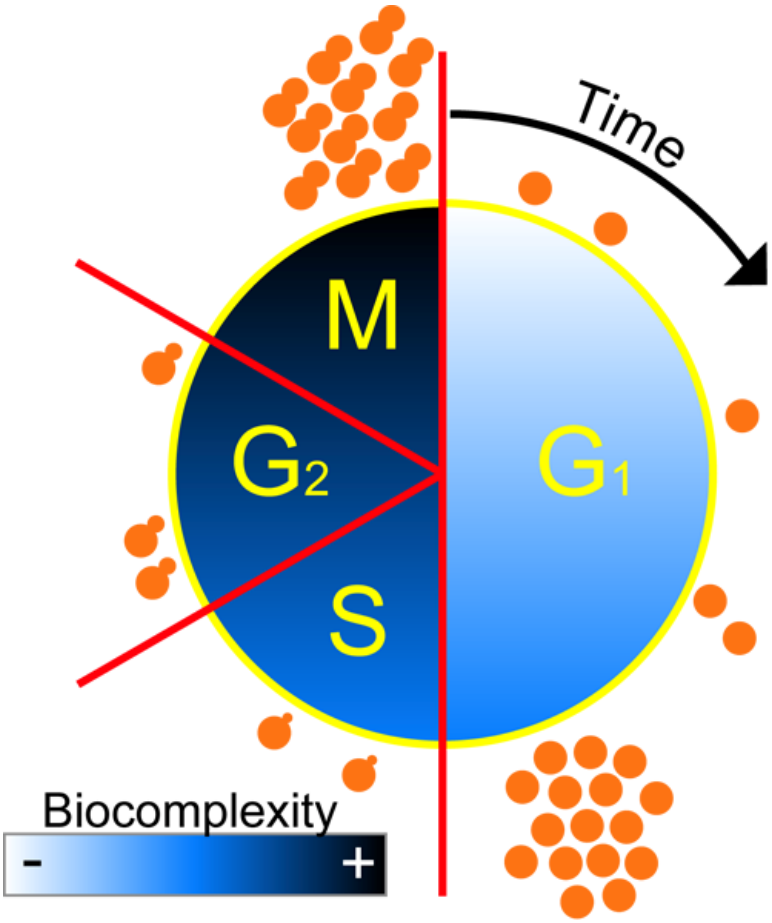


Figure 5.6: Population structure and biocomplexity of autonomous oscillations. Two clustered cell populations are arranged around a schematic of the budding yeast cell cycle in accordance with the data from Figures 5.4 and 5.5. Cell cycle progression is clockwise, with cell division punctuating the M to G_1 transition. A model of biocomplexity is indicated through shading according to the data of Alberghina et al. [1,32].

Figure 5.7 depicts a phase portrait of biomass versus glucose (substrate) concentration. The oscillatory nature of the system produces a counter clockwise orbit in the phase plane space, that has been annotated with respect to suspected total bio-complexity. The annotation is based on the observation that biomass increases exponentially around the cell cycle. The G_1 portion of the cell cycle may be a time of low biocomplexity.

The term ambient biocomplexity of a product is meant to highlight the problem that product separation costs dominate bioproduction. In any scheme to isolate a product from an admixture, each successive step is designed to enrich the product relative to background. In the case of a protein product, it is reasonable to equate the ambient biocomplexity with the total protein content of the cell. Some work indicates that the internal protein complexity of a cell increases with cell cycle position [1,32]. Isolation of a protein product from a cell sample at a time of its lowest biocomplexity is by definition guaranteed to maximize its enrichment ratio and minimize costs. We present a theoretical analysis that illustrates that periodic systems can be managed to produce a product of interest out of phase with its biocomplexity. This analysis is based on our experimental understanding of the population structure that becomes entrained during autonomous oscillations.

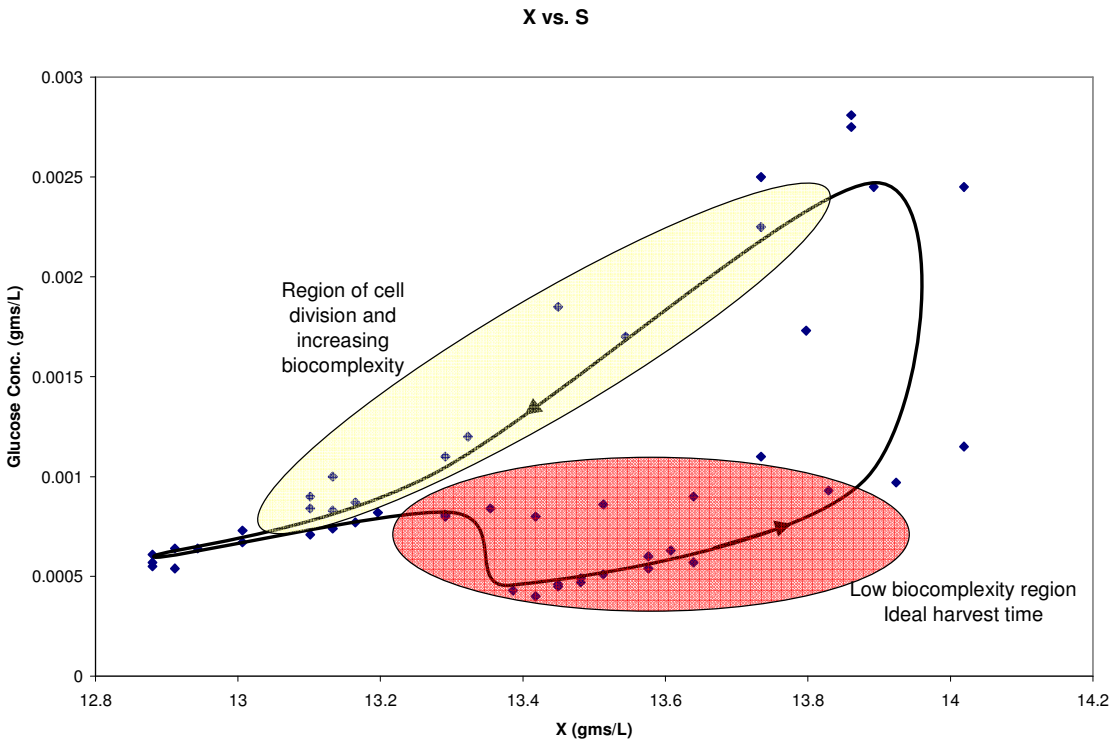


Figure 5.7: Phase portrait of a periodic orbit. Biomass (X) versus glucose concentration (S) corresponding to the Heinzle et al. data of Figure 4. The red lower shaded region encompasses the G_1 period that is characterized by increasing, but low total biocomplexity, see also Figure 6. The yellow upper shaded region encompasses the highest biocomplexity. The arrows indicate increasing time.

Accumulation of Biocomplexity

Let *p(t,s)* represent a population density stratified according to time t and cell cycle position *s*. Suppose that the cells of the population are producing a protein or an admixture of proteins in a cell cycle dependent manner. Let the function *f(s)* specify a single cell production profile that depends only on the cell cycle position, *s*. See for instance the sketch in the top panel of Figure 5.8. It is well known, and has been repeatedly shown, that the vast majority of transcripts are cell cycle dependent [30]. Given the single cell production *f(s)*, and the population density, *p(t,s)*, the instantaneous production is given by their inner product:

$$I(t) = \int p(t,s),\ f(s)ds \qquad \text{Equation (5.4)}$$

Furthermore, the products, *F*, will accumulate within a reactor with dilution rate *D*, according to a mass balance equation.

$$\frac{dF}{dt} = I(t) - D\ F(t) \qquad \text{Equation (5.5)}$$

Given the assumption that the population density is a periodic function of time, see Figures 5.4, 5.5, and 5.6, we show that the function *F* is asymptotically periodic.

Demonstration of Asymptotic Periodicity

Let the population density be periodic with period *T*. This means that $p(t+T,s)=p(t)$ for all *s*. From this it follows that *I(t)* is periodic with period *T*:

$I(t+T) = \langle p(t+T,s), f(s) \rangle = \langle p(t,s), f(s) \rangle = I(t).$

where we have used the inner product notation to represent the integrals from Equation (5.4). Now consider the function *F(t)*. Using the well known variation of parameters solution of equation (5.5) we have that

$$F(t) = e^{-Dt}\left[\int_0^t e^{Ds} I(s)ds + C\right]$$ where C is an integration constant.

Using the periodicity of I we have:

$$F(t+T) = e^{-Dt}\left[\int_0^{t+T} e^{D(s-T)} I(s)ds + C\right]$$

$$= e^{-Dt}\left[\int_0^{t+T} e^{D(s-T)} I(s-T)ds + C\right]$$

$$= e^{-Dt}\left[\int_{-T}^{t} e^{Ds} I(s)ds + C\right]$$

$$= e^{-Dt}\left[\int_0^{t} e^{Ds} I(s)ds + \int_{-T}^{0} e^{Ds} I(s)ds + C\right]$$

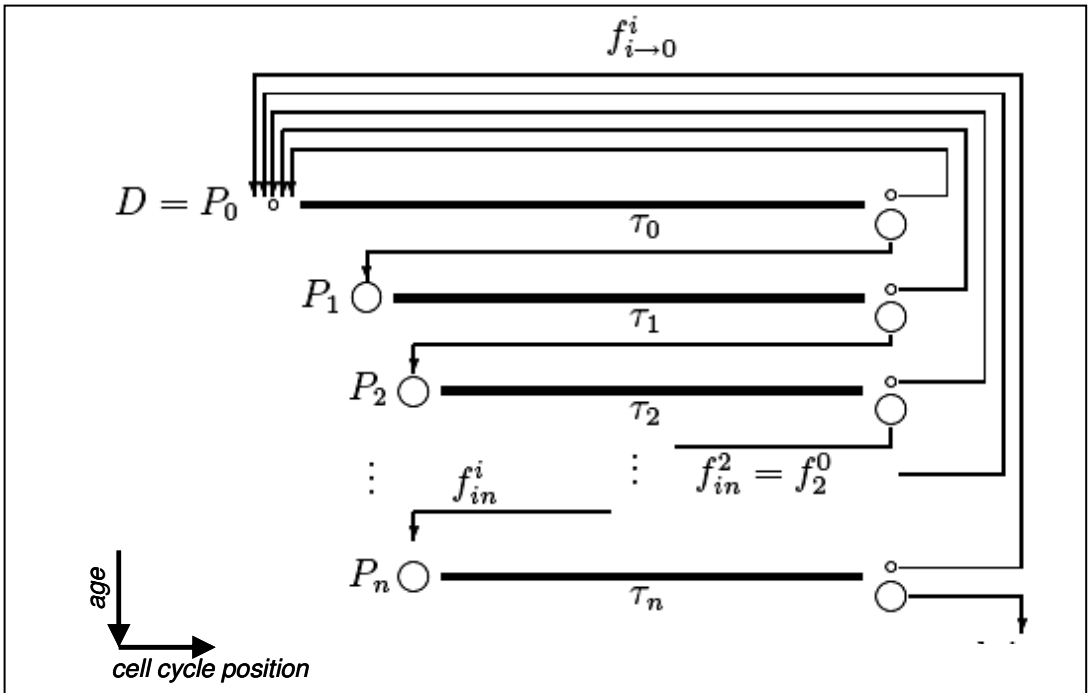
$$= F(t) + e^{-Dt}\int_{-T}^{0} e^{Ds} I(s)ds$$

Since the last integrand is finite and e^{-Dt} limits to 0 exponentially as t goes to ∞, *F(t)* is an asymptotically periodic with period *T*.

Furthermore, we demonstrate that regardless of the phase profile of the accumulation function *F*, we can design a product expression function *g(s)*, to create a product, *G*, that is maximally out of phase with the complexity produced by *f*. The conclusion is two-fold. First, ambient biocomplexity *F* will oscillate; second, that regardless of the ambient biocomplexity we can design a production scheme that is maximally out of phase with it. This implies that we can maximally enhance the enrichment ratio of *G* relative to *F*.

It has been suggested in the literature that the total protein complexity of the cell increases exponentially with the cell cycle [1,32]. Therefore, as a demonstration of enrichment ratio enhancement we consider the case where ambient biocomplexity is represented as an increasing exponential function over the cell cycle, see Figure 5.8. We compute the evolution of a population of yeast cells with a Leslie type model that is the discrete analog of the popular population balance equations. The Leslie model is extremely flexible in that it allows us to represent a finite but arbitrary number of age classes, is easily parameterized with experimental data of yeast growth relevant parameters, and is computationally tractable. We have previously shown that the Leslie models that we have employed reproduce the dynamics observed in experimental time series [26]. A brief description of the model and model parameters is provided below.

Leslie model of Yeast Growth and Division



$$\begin{bmatrix} f_0^0(t+\tau_0) \\ f_1^0(t+\tau_1) \\ f_2^0(t+\tau_2) \\ \vdots \\ f_n^0(t+\tau_n) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} f_0^0(t) \\ f_1^0(t) \\ f_2^0(t) \\ \vdots \\ f_n^0(t) \end{bmatrix}$$

where:

$$\left\{ \begin{array}{c} f_{in}^i(t) = Total\ \ flux\ entering\ the\ i-th\ generation\ at\ time\ t \\ f_i^0(t) = Daughter\ cells\ \ produced\ \ from\ a\ division\ in\ age\ class\ i\ at\ time\ t \\ \tau_i = Flux\ residence\ time, i.e.\ cell\ \ cycle\ length \end{array} \right\}$$

Figure 5.8: The top schematic represents the process of yeast growth and division. Each age class is represented by a bold horizontal line and corresponding residence time, τ. The age classes are organized by replicative age, with the daughter cells, denoted P_0, as the first grid. Higher generation parent cells are listed below. Each cell division can be characterized by producing a daughter cell back at the beginning of the P_0 grid and a parent of the next highest age class. This process can be represented by a Leslie matrix as shown by the bottom panel.

The Leslie model[26] used to calculate yeast growth and division encodes as a mathematical object, the process flow depicted at the top of Figure 5.8. The process flow has age descending from daughter cells with no scars, P_0, at the top to cells with arbitrarily many scars, P_n, at the bottom. Cells belonging to each of these age classes traverse their separate cell cycles, shown as the horizontal thick black lines. The time it takes a cell of age, k, to traverse their cell cycle is τ_k. The values of τ used are shown in Table 5.3, along with the temporal position of bud emergence. The thin black lines with arrows indicate the directions of possible flux due to cell divisions. Each arrow corresponds to a 1 in the matrix representation of the process at the bottom of Figure 5.8.

Given an initial population distribution, the matrix model is iterated to produce the dynamics. Typically, such a system would produce asynchronous exponential growth. Autonomous oscillations result from an additional feedback. We have introduced a *delay model* to produce population oscillation. In the delay model, a threshold, T_R, is introduced. Once the density of cells in the S-phase of the cell cycle reaches T_R or above, any cells in a 10% strip proximal to S become delayed in their cell cycle progression. In the simulations described in Figure 5.10, T_R was 20 % of the total culture density. The S phase is delimited on the left by bud emergence (BE) as described in Table 5.2.

Table 5.2: Parameters used in the Leslie model simulations. The parameters are listed in order of replicative age. Cell cycle duration and progression can be described in terms of time. Parameters are described in units of minutes.

Leslie Model Parameters				
Age Class	τ	BE	S-Phase End	G_2-Phase End
P_0	522	261	362	457
P_1	483	241	302	423
P_2	437	219	273	383
P_3	398	199	249	349
P_4	347	174	217	304
$P_{5\text{-}9}$	277	138	173	242

Simulations generated from the Leslie model are shown in the bottom panel of Figure 5.9, corresponding to the single cell expression profiles shown in the upper panel. The ambient biocomplexity is depicted in the upper panel of Figure 5.9, by the red cell cycle expression profile and in the bottom panel by the red accumulation signal. All of the cells in the population are assumed to be expressing the single cell expression profiles in a cell cycle dependent manner as outlined by the equations above and in the Appendices to produce the corresponding accumulation profiles. The green and blue single cell expression profiles in the upper panel and their matching accumulation signals in the bottom panel, represent different choices for the desired product. As the simple square well expression profile is shifted along the cell cycle from green to blue, a maximum enrichment ratio occurs in the product relative to the ambient biocomplexity. This occurs when the accumulation profile of the ambient biocomplexity (red) is maximally out of phase with the accumulation profile of the product (blue to green).

The series of simulations shown in Figure 5.9 amount to the equivalent of design equations. The theory shows that punctate expression profiles, such as the green and blue in the top panel of Figure 5.9, tile the spectrum of possible phases. Therefore, if the ambient biocomplexity can be measured or accurately modeled then a promoter design equation can be produced to circumvent it. The product or the product producing cells would then be harvested during appropriately chosen time windows within this periodic oscillation when the enrichment ratio is largest. Time windows corresponding to low complexity are shown in the bottom panel. Periodic harvesting is only the beginning of a far more comprehensive scheme.

Periodic harvesting can also be managed to produce a continuous and enriched product concentration from multiple oscillating fermentors. For example, let T be the length of an oscillation. Assume that the harvesting windows have duration γ, that is an integral divisor of the cell cycle period, $n\gamma = T$. Consider an entire period of the cell cycle tiled with n, equal time windows of the duration of γ. This concept is illustrated in Figure 5.10. Given n-reactors, each of which is sequentially out of phase by an amount gamma, the product accumulation can occur with maximal enrichment continuously. Imagine the autonomous, periodic oscillation, as a clock and the reactors as the numbers on a clock face. When the harvesting portion of the trajectory enters a given reactor its effluent is collected into an ancillary reactor or collection vessel. As the harvesting trajectory leaves a reactor it enters the next reactor in the sequence by design. Much has been made of the relationship of the autonomous oscillations to ultradian rhythms and clocks[19]. This continuous harvesting scheme brings that analogy to life as an engineering tool. We might consider the initialism: Continuous Long-term Oscillating Collection Scheme: CLOCKS.

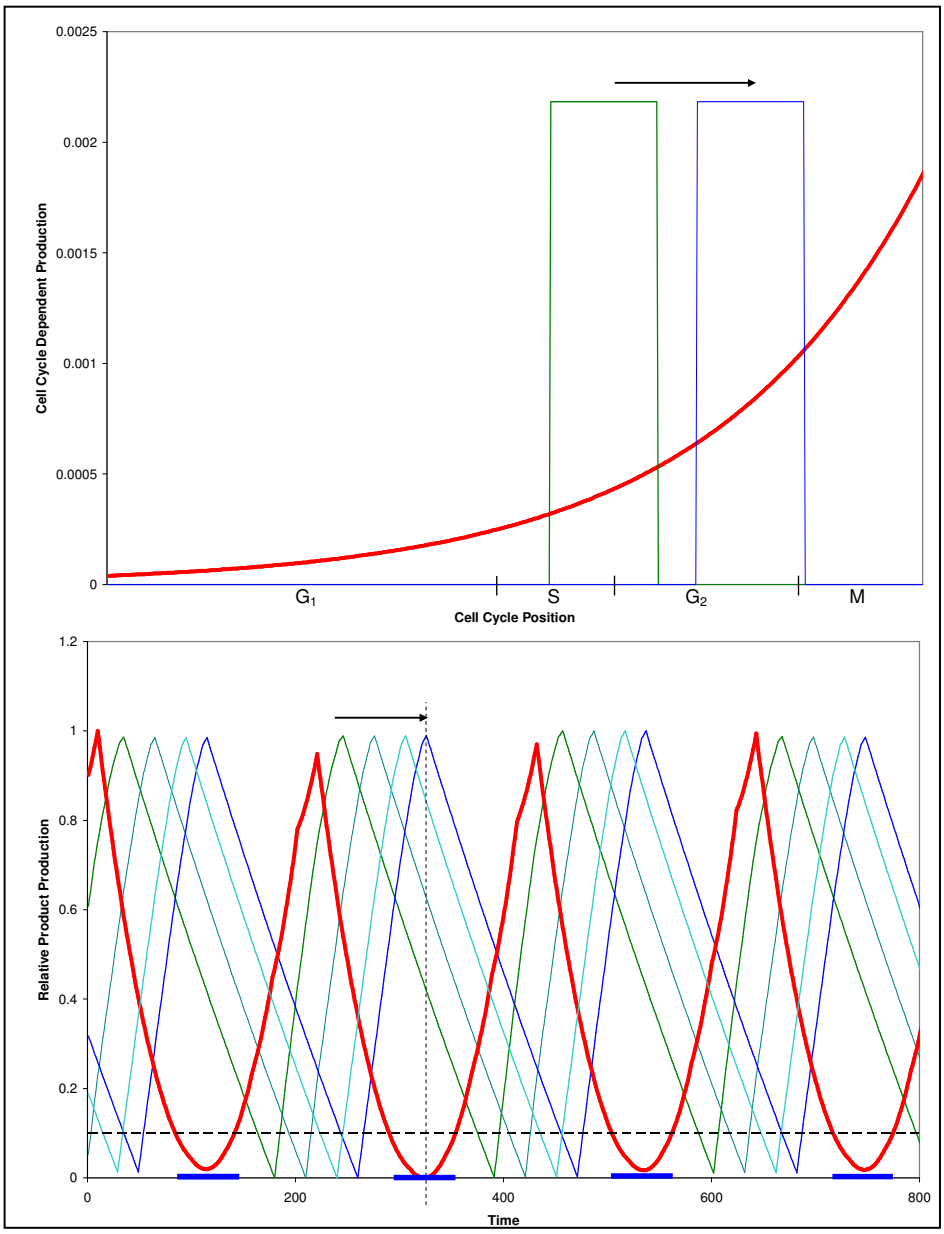


Figure 5.9: Maximization of the enrichment ratio. The upper panel shows single cell expression profiles as a function of a single cell cycle. The red curve represents a model of ambient biocomplexity that increases exponentially along the cell cycle. The green and blue punctate expression patterns represent different choices for product expression. We call these promoter models. The bottom panel shows the normalized accumulation of these signals in a population. The population dynamics have been simulated by the Leslie model as previously described. The bottom panel shows that as the punctate windows of the promoter models are shifted, they tile the possible phases of the oscillation, see arrows in the top and bottom panels. Of the promoter models, there exists one that is maximally out of phase with the biocomplexity, highlighted by the dashed vertical black line. The bold blue intervals along the abscissa correspond to a choice of harvesting windows, induced by imposing a complexity threshold indicated by the dashed horizontal line.

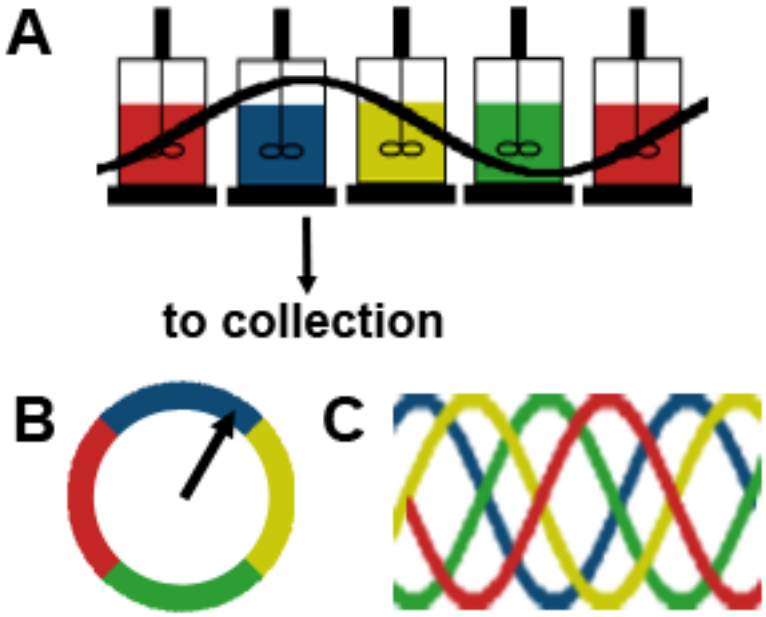


Figure 5.10: The CLOCKS harvesting scheme. Suppose that the duration of an optimum harvesting window, γ, is one quarter of an oscillation. Then four systems prepared out of phase tile a period. This is shown in (A). Since a periodic oscillation can be thought of as a clock face, the sequence of color-coded reactors correspond to the color-coded clock face in (B). The maximum enrichment ratio is depicted to occur at the peak of the oscillation shown to run through the system of reactors as a black curve. As the signal passes through the reactors, currently the blue reactor, the systems are bled in sequence as shown on the clock face. The sequence of windows, blue, yellow, green, red, blue...,
is reiterated in (C) to indicate that in this scheme, product is continuously being harvested with a maximum enrichment ratio from the reactors in sequence.

Conclusions

Downstream processing costs constitute the majority of the cost in product formation and accounts for over a 10-fold increase in the price of biomanufactured goods[2]. Therefore, methods and methodologies of product enrichment are directly relevant to bioprocessing cost. We have argued that the ambient biocomplexity of a product within an admixture is the dominant contribution to processing cost. Our data and ideas imply that product enrichment can be achieved by co-opting the properties of autonomously oscillating yeast systems to minimize biocomplexity.

We have demonstrated that autonomously oscillating systems exhibit periodic oscillation in yield. We have shown that through an understanding of the population structure of periodically oscillating systems we can increase product enrichment by interleaving product accumulation with the accumulation of its biocomplexity. Furthermore, we have described a novel phased, parallel, periodic harvesting method that is the direct engineering analogy of ultradian oscillations themselves: A continuous clock.

The ability to interleave product expression with that of its ambient biocomplexity rests on two pillars. The first is that the population structure of autonomous oscillations is marked by two observable and temporally coherent cell populations that traverse the cell cycle periodically. The second is the general flexibility afforded by the ability to genetically engineer cell cycle specific promoter expression. In mathematical abstraction, we have provided the proof of principle. What remains is a practical demonstration, towards which the following issues exist. Promoters are not all of equal

strength, and while phase is the dominant issue, the production of product above a necessary threshold may be challenging to engineer as a function of the cell cycle. If the cell is producing biomolecular mimics of the desired product at a rate several orders of magnitude larger at the maximum phase difference then, for that product, the interleaving strategy will have limited effect. The cell cycle and its implications for bioprocess control has already been recognized. Yeast have been engineered to produce amylase during the M phase of the cell cycle [31]. It has been shown that heat shock arrest of the cells in M-phase, coupled with M-phase expression increased the yield of product. It remains to be seen if autonomously oscillating populations can outperform those in heat shock.

Product enrichment from autonomous oscillations could also be used as a mechanism for simultaneous production of multiple products. If two or more products were produced in different phases of the cell cycle, each product could be harvested during separate time windows and subjected to potentially different downstream processing. This idea could lead to savings on production equipment by eliminating the need for multiple fermentation setups.

Periodic sampling from autonomously oscillating systems was previously proposed as a means to increase yield above that provided by an equilibrium [16]. It could be argued that the value of such periodic sampling for product may be offset by the idle time of down-stream processing equipment as opposed to continuous harvesting of a slightly smaller amount. This intermitant feed of crude products to downstream processing equipment could be problematic for a continuous process that is not amenable to continual shut-

down and start-up of equipment. However, the CLOCKS strategy that consists of a simple parallel array of oscillating systems whose phases are prepared to tile a period, as described in Figure 5.9, completely and elegantly overcomes these obstacles.

The physical chemistry of proteins is relatively well understood, and mature algorithms for their analysis exist along with large databases of known structures[28]. It is currently an active research interest in both basic science and in industry to develop theoretical tools to predict protein-protein interactions from an admixture[20]. With a solid understanding of cell cycle related protein expression it should be possible over the course of the next decade to predict and manipulate the ambient biocomplexity of yeast.

References

1. Alberghina, L., Mariani, L., Martegani, E., and Vanoni, M. 1983. Analysis of protein distribution in budding yeast. *Biotechnology and Bioengineering*. **15**:1295-1310.

2. Aldridge, S. 1996. Downstream processing needs a boost. Genetic Engineering and Biotechnology News. **26**:1.

3. Bansal, V., Roychoudhurry, P., Mattiasson, B., and Kumar, A. 2006. Recovery of urokinase from integrated mammalian cell culture cryogel bioreactor and purification of the enzyme using p-aminobenzamidine affinity chromatography. *J. of Molecular Recognition*. **19**:332-339.

4. Bellgardt, K.H. 1994. Analysis of synchronous growth of bakers yeast. Part I: Development of a theoretical model for sustained oscillations. *Journal of Biotechnology* **35**:19-33.

5. Bellgardt, K.H. 1994. Analysis of synchronous growth of bakers yeast. Part II: Development of a theoretical model for sustained oscillations. *Journal* of *Biotechnology* **35**:35-49.

6. Beuse, M., Bartling, R., Kopmann, A., Deikmann, H. and Thoma, M. 1998.Effect of the dilution rate on the mode of oscillation in continuous cultures of *Saccharomyces cerevisiae*. *J. of Biotechnology*. **61**:15-31.

7.Chan, E., Chen, C., and Chen, L. 1992. Recovery of invertase from ethanol fermentation broth. *Biotechnology Letters*. **14**:573-576.

8. Crooke, P., Tanner, R., and Park, D. 1986. Time dependent differential yield as a scale-up parameter in enzyme and fermentation reactors. *Biotechnology Progress*. **2**:40-47.

9. Dawson, P. 1972. Continuous synchronous culture-the production of extracellular enzymes by continuous phased culture of Bacillus subtilis. *Proc. IV IFS:Ferment. Technol. Today.* 121-128.

10. Dumas, B., Masson, C., Lebrun, K., and Achstetter, T. 2006. Hydrocortisone made in yeast: metabolic engineering turns a unicellular microorganism into a drug-synthesis factory. *Biotechnology Journal.* **1**:299-307.

11. Essajee, C. and Tanner, R. D. 1979. The effect of extracellular variables on the stability of the continuous baker's yeast-ethanol fermentation process. *Process Biochemistry.* **25**:16-21.

12. Gleeson, M., White, C., Meininger, D., and Komives, A. 1998. Generation of protease-deficient strains and their use in heterologous protein expression. *Method in Mol. Bio*. **103**:81-94.

13. Hartwell L. 1974. *Saccharomyces cerevisiae* cell cycle. *Bacteriol. Rev*. **38**:164-198

14. Hartwell L., and Unger M. 1977. Unequal division in *Saccharomyces cerevisiae* and its implications for the control of the cell division. *J. Cell. Biol* **75**:422-435.

15. Heinzle, E., Dunn, I., Furukawa, K., and Tanner, R. D. 1982. Modeling of sustained oscillations observed in continuous culture of *Saccharomyces cerevisiae*. *Proceedings of the 1st IFAC Workshop.* A. Halme, ed. Pergamm Press (Oxford). 57-65.

16. Hortso, M. 1996. Population balance models of autonomous periodic dynamics in microbial cultures: their use in process optimization. *Canadian J. of Chem. Eng*. **74**: 612-620.

17. Jules, M., Francois, J., and Parrou, J. 2005. Autonomous oscillations in *Saccharomyces cerevisiae* during batch cultures of trehalose. FEBS Journal. **272**:1490-1500

18. Kjeldsen, T., Ludvigsen, S., Diers, I., Balshmidt, P., Sorensen, A., and Kaarshold, N. 2002. Engineering-enhanced protein secretory expression in yeast with applications to insulin. *Journal of Biological Chemistry*. **277**:18245-18248.

19. Murray, D.B., Klevecz, R., and Lloyd, D. 2003. Generation and maintenance of synchrony in *Saccharomyces cerevisiae* continuous culture. *Exp. Cell Res*. **287**:10-15.

20. Qiu, J. and Noble, W. 2008. Predicting co-complexed protein pairs from heterogeneous data. *PLOS Computation Biology*. **4**.

21. Ramesh, M., Charyulu, N., Chand, N., and Lonsane, B. 2004. Non-growth assosciated production of enzymes in solid state fermentation systems: Its mathematical description for two enzymes produced by *Bacillus licheniformis. Bioprocess and Biosystems Engineering*. **15**:289-294.

22. Richard, P. 2003. The rhythm of yeast. *FEMS Microbiology Reviews*. **27**:547-557.
Sajc, L., Grubisic, D., Novakovic, G. 1999. Bioreactors for plant engineering: an outlook for further research. *Biochemical Engineering Journal*. **4**:89-99.

23. Sajc, L., Grubisic, D., and Novakovic, G.2000. Bioreactors for plant engineering: an outlook for further research. *Biochemical Engineering Journal*. **4**:89-99.

24. Smith, C. 2005. Striving for purity: advances in protein purification. *Nature.***2**: 71-77.

25. Stowers, C., and Boczko, E.M. 2007. Reliable Cell Disruption in Yeast. *Yeast* **24**:533-541.

26. Stowers, C., Hackworth, D., Mischaikow, K., Gedeon, T., and Boczko, E. 2008. Extending synchrony and deconvolving population effects in budding yeast through an analysis of volume growth with a structured Leslie model. *Theoretical Population Biology.* Submitted.

27. Szczebara, F., Changelier, C., Villeret, C., Masurel, C, Masurel, A, Bourot, S., Duport, C. et al. 2003. Total biosynthesis of hydrocortisone from a simple carbon source in yeast. *Nature*. **21**:143-149.

28. Taufer, M., An, C., Kerstens, A., and Brooks, C. 2006. Predictor @ home: A "protein structure prediction supercomputer" based on global computing. *IEEE Transactions on Parallel and Distributed Systems*. **17**: 786-796.

29. Thim, L., Hansen, M., Norris, K., Holegh, I., Boel, E., Forstrom, J., Ammerer, G., and Fiil, N.1986. Secretion and processing of insulin precursors in yeast. *PNAS*. **83**:6766-6770.

30. Tu, B.P., Kudlicki, A., Rowicka, M., McKnight, S.L. 2006. Logic of the yeast metabolic cycle: Temporal compartmentalization of cellular processes. *Science.* **310**:1152-1158.

31. Uchiyama, K., Morimoto, M., Yokoyama, Y., and Shioya, S.1996. Cell cycle dependency of rice α-amylase production in a recombinant yeast. *Biotechnology and Bioengineering*. **54**:262-271.

32. Vanoni, M. Vai, M., Popolo, L., and Alberghina, L. 1983. Structural heterogeneity in populations of budding yeast *Saccharomyces cerevisiae*. *J. of Bacteriology*. **156**: 1282-1291.

33. Walker, G. 1999. Synchronization of yeast cell populations. *Methods in Cell Science.***21**:87-93.

34. Xu, Z. and Tsurugi, K. 2006. A potential mechanism of energy-metabolism oscillation in an aerobic chemostat culture of the yeast *Saccharomyces cerevisiae*. FEBS Journal. **273**: 1696-1709.

CHAPTER VI

QUANTITATIVE PCR FOR DILUTE SOLUTIONS OF DNA

Introduction

Real-time PCR (rt-PCR) is the most widely accepted tool for quantitative gene expression analysis [6,13,26]. Rt-PCR is commonly used as a quantitative assay in a variety of clinical and research areas including the study of genetically modified foods, vaccine efficacy, and systems biology [1,2,5,20]. However, few studies have emerged regarding the robustness and reproducibility of using real-time PCR as a quantitative assay. The notable advantage of rt-PCR over more traditional techniques such as riboprotection assays is the ability to amplify extremely small, typically undetectable, signals to a measurable range. This promotes the ability for one to quantitatively measure gene expression signals from small populations, or even single cells. Coincidentally, this range of low DNA concentration is also where rt-PCR is most quantitatively unreliable since minute errors in sample preparation are exponentially amplified within the process. This error amplification process can be demonstrated with the following equation:

$$\xi = \varepsilon\, 2^{n} \qquad \text{Equation (6.1)}$$

where ξ is the total amount of amplified error after n amplification cycles and ε is the initial amount of error.

Arguably, the most common method of PCR data analysis is based on some form of determining the Ct or Cp (cycle threshold or crossing point) value from reaction amplification curves. This process typically involves using a pre-determined horizontal fluorescence threshold value of which the crossing point (Cp) of the reaction amplification curve is computed [19]. For absolute quantitation, a standard curve of Ct values is generated from known concentrations of DNA [2,18]. This curve, typically in the form of a linear regression, can be used to determine DNA concentration of unknown samples. For relative quantification, such as in the ΔΔCt Method, the relative change in Cp values for the target genes are compared to that of a reference or housekeeping gene [14]. This analysis assumes that the housekeeping gene is constant throughout the analysis and that the efficiencies of the two separate reactions are nearly identical. Despite the differences between absolute and relative quantitation methodologies, the important point to note is that both assays rely heavily on the robustness and reliability of the Ct values. More sophisticated methods for relative quantitation have emerged that utilize reference to multiple controls or housekeeping genes, providing additional statistical infrastructure to minimize the variances within Ct values [17]. These methods may in fact be more trouble than their worth since many housekeeping genes have shown to be non-constant [8]. Others have developed statistical algorithms or mathematical models to minimize variances associated with Ct values [4,9,12,13,14,18,20,22,26]. However, the root of quantitative error in all of these practices is the error introduced through variations in the Ct values, that may be minimized but not eliminated through statistics. For this reason, we base our analysis of rt-PCR variations on Ct values, with the intentions of developing an understanding of the robustness of the Ct values.

Recently, reports of statistical analysis of Ct values have appeared; however, these analysis are confined to only a few orders of magnitude in sample concentration and in only one study are more than a few replicates analyzed [13, 20, 26]. Furthermore, these studies are generally performed at DNA concentrations within the "sweet" spot of PCR, Ct values ranging between 16 and 24, naturally avoiding possible conflict with assay reproducibility. Due to the lack of extensive experimental data, these reports provide little, if any, analysis regarding the overall reproducibility and robustness of the PCR process. Typically, these reports focus on developing statistical methods for achieving more accurate results using only a few replicates over a narrow DNA concentration range. In this study, 184 replicates were performed for DNA concentrations ranging over nine orders of magnitude with the intentions of generating an extensive data set that could be used to evaluate the robustness and reproducibility of rt-PCR. The large data set provides opportunity for statistical analysis that can be used to determine the value of performing sample replicates and the results of the analysis can serve as a benchmark for evaluating the comparative accuracy of novel PCR strategies.

Several novel and promising methods for performing quantitative PCR have appeared in the literature, of which we have attempted many [3,15,16,21]. We are particularly interested in these methods because of the potential ability to accurately quantify gene expression values from single cells. We are unaware of any conventional method of PCR capable of accurately quantifying dilute solutions of DNA, as would be obtained from single cells. Novel methods for performing PCR may have advantages compared to conventional PCR. Perhaps the most interesting of these techniques involves the amplification of single molecules of DNA immobilized within a gel [16]. This strategy is conceptually

similar to the historic process of determining the microbial cell density in a solution by plating the microbes on an agar gel and subsequently counting of the number of grown colonies after some time period. This type of assay is an endpoint analysis which avoids dependence on the kinetics of the process itself. We have noticed that the kinetics of rt-PCR is far less reproducible for dilute samples of DNA (see Figure 6.1) making an endpoint analysis a highly attractive PCR platform. However, we have been unable to achieve reproducible results using the gel immobilization technology, despite using several different supermixes and gel types such as agarose, acrylamide, and gelatin.

Another promising technology for performing PCR is coined *emulsion PCR* [24]. This technology exploits the compartmentalization of micelles in an emulsion, permitting thousands of independent reactions to occur simultaneously within a single tube, with each reaction confined to its own micelle. The advantage of this method is the ability to run nano-liter sized reactions individually encapsulated within hydrophobic oil such that even a single piece of DNA within a micelle produces a relatively high DNA concentration, in theory minimizing mass transfer limitations. This method has shown promise for amplifying complex genetic libraries [24]. We attempted to perform quantitative endpoint PCR using reversed micelles; however, problems abound, especially regarding the detection of amplified micelles using spectroscopy.

There exists one method that appears to lack the technological difficulty associated with running PCR reactions within gels or emulsions. In this method, reactions are performed within the aqueous phase, similar to conventional methods, but are scaled down in size such that each reaction only receives one piece of DNA. Of course this method relies

heavily on the ability to successfully amplify DNA solutions containing only one or a few copies of DNA. In fact, it has been demonstrated that PCR can be used to reliably amplify single copies of DNA [25]. Since the reaction matrix is identical to conventional PCR methods, many of the problems related to mass transfer are avoided. Originally, this method was published using a microchamber array, where each reaction received only nanoliters of supermix with only a alight probability of containing a piece of DNA [15]. We have attempted this method in our laboratory using PDMS (polydimethylsiloxane) and glass chips, but have been unable to generate reliable data for two reasons. The first problem stems from the effects of hydrophobicity at the nano scale. Despite our attempts at using several different materials for chip fabrication, the majority of micro-wells could never be reproducibly filled. Secondly, regardless of the adhesive used to seal the chip, evaporation during thermocycling resulted in nearly half of the filled wells completely evaporating during the process.

Although technological hurdles remain that have prevented us from pursuing PCR reactions on the nano scale, the concept of developing an endpoint PCR assay to quantify gene expression is still within reach. To avoid many of these obstructions, we implemented this technique on a larger scale using a 384 well plate format. Plate assays are broadly valuable to scientists for use in high-throughput assays [10,27] as biological equipment such as spectrophotometers and even PCR machines have been developed to utilize this format. More importantly, assays on the plate format scale avoid technological hurdles associated with micro-fluidics and nano-scale measurement. The disadvantage of the plate format, compared to nanoscale formats, is the lack of resolution. Nano-scale formats can provide higher assay resolution since thousands of reactions can

be perfomed within a fraction of the space required for a plate. This higher density permits quantitation of more concentrated samples, since the resolution of the endpoint analysis is directly proportional to the number of reactions performed in a single assay. The 384 well plate format significantly limits sample concentration range; however, the format can still be used as a platform for evaluating the usefulness of the endpoint strategy compared to traditional rt-PCR.

The first part of this study involves a statistical analysis of an extensive rt-PCR data set extending over nine orders of magnitude in sample concentration. The analysis identifies both strengths and weaknesses of traditional methods and outlines the effect of running replicates of identical samples. This analysis can also serve as a benchmark for comparison to the new technology of using PCR as an endpoint analysis. The relative errors associated with various quantiles of the distributions of Ct data are computed by converting variances in Ct values into DNA concentrations using a Ct Method standard curve, see Figure 6.3. A statistic, that we call the misclassification probability, is developed and used to evaluate the quantitative ability of PCR as a function of sample concentration. The analysis of misclassification answers the following question: What is the probability of obtaining a particular Ct value that was then classified to an incorrect DNA concentration? The misclassification probability can be computed directly from Ct data and used as a statistic for evaluating the robustness of the Ct Method. This probability can be simply represented as the probability of two separate events: 1. choosing a particular Ct value and 2. misclassifying that particular Ct value to a class of Ct values with the incorrect DNA concentration. This second probability is conditional

since some Ct values may be unique to a particular DNA concentration and would therefore have no probability of misclassification.

The misclassification frequency was computed from the following formula:

$$P^{misclass}(I) = \sum_{i,j}^{T,n} P_{i,j}^{choose} \cdot P_{i,j}^{misclass} \qquad \text{Equation (6.2)}$$

where $P^{misclass}(I)$ is the probability of misclassifying a particular Ct value. $P_{i,j}^{choose}$ is the probability of obtaining a given Ct value from a sample (class), while $P_{i,j}^{misclass}$ is the conditional probability of misclassifying that particular Ct value. T is the entire spectrum of Ct values over all samples, n represents the number of Ct values in each age class, i represents the selected Ct value, i, within class I, and j represents all other Ct values within class I. We intend for the total misclassification probability of a given DNA concentration to represent a proxy for the robustness of the Ct method, where lower misclassification probabilities represent higher robustness.

Materials and Methods

Rt-PCR results were generated using linearized double stranded EC3 plasmid DNA containing the ybdO gene. The plasmid was linearized by digestion with the restriction enzyme BamH1 prior to PCR. The primer sequences used were the following: 5'-AAT TAT TCT AAA ACC AGC GTG TC-3' (F) and 5'-TTT GGG ATT GAA TCA CTG TTT C-3'. The PCR supermix was prepared as described in Karsai et al.[11], with the exception that we used Qiagen HotStarTaq Cat # 203203, Roche dNTP's Cat# 13583000, DMSO

Sigma # D8418 at 2%, and Sybr Green (Sigma # 86205) at 5-times the recommended concentration. Primers were used at a concentration of 1 μM. All samples were run on the 384 well plate platform using an Applied Biosystems 7900HT thermocycler and the SDS 2.3 software. The Ct value threshold was set at 5.0 RFU (Relative Fluorescence Units) for all samples. The DNA concentrations of concentrated stocks were measured using a Nano-Drop 100 spectrophotometer prior to use. Subsequent dilutions were performed using sterile, nuclease-free water from Ambion # AM9937.

The thermo-cycling program used was as follows:

1. 2' at 50ºC Initial Warmup Phase
2. 15' at 95ºC Initial TAQ Activation Step
3. 1' at 95ºC DNA Denaturation
4. 1' at 50ºC Primer Annealing
5. 1' at 72ºC DNA Extension
6. 0.25' at 80ºC Fluorescence Measurement

Steps 3-6 were repeated forty times.

Preparation of Identical Replicates

Careful consideration was taken so that each sample loaded onto the plate would be statistically independent, but identical in composition. To ensure this condition, PCR supermix was prepared in large well-mixed batches within a 14-mL conical tube. This approach ensures that all samples within a batch contain similar concentrations of reagents, despite pippetting errors. Samples were prepared and run 4 at a time. Each sample consisted of 184 replicates and 8 negative controls, requiring exactly half of a 384

well plate. All of the components except for DNA were loaded into a 14 mL conical tube in the following order: 800 μL PCR buffer, 5.6 mL of nuclease free water, 160 μL DMSO, 320 μL $MgCl_2$ (Qiagen Cat # 124113012), 160 μL of a primer mix (a 1:1 mix of the forward and reverse primer stored at a concentration of 50 μM each), 160 μL Sybr Green (100X stored in DMSO), 160 μL of dNTPs, and lastly 80 μL of Taq polymerase. We noticed that the order that these are added affects the reproducibility of the assay. This mixture was vortexed at high speed for 1 minute. 335 μL of supermix was then removed to be used as a negative control and placed into a 1 mL epindorf tube and 25 μL of water was added. This mixture was then briefly vortexed to ensure well mixing. The remaining 7.105 mL of supermix was then split equally four ways into 2 mL cryostat tubes, and 134 μL of water plus the amount of desired DNA was added to each cryostat tube. Each tube was then briefly vortexed. For each reaction contained within a single well of the plate, 10 μL of the respective reaction mix was loaded into a well of the 384 well plate.

TAQ Polymerase Pre-Wear

The PCR supermix was prepared as described above but without any template DNA. Samples were pre-worn by thermocycling the supermix as described in steps 3 through 6. Samples were pre-worn at intervals of 5 cycles from 5 to 40 cycles. After the pre-wear phase, 10^8 copies of DNA were added and the samples were run as described above. The relative efficiency was computed by the average derivative over the amplification curve.

Bead Plate Filling Experiments

Experiments were performed using 20 μm latex beads from Beckman Coulter (#PN6602798) using flat bottom 96 well plates from Becton Dickinson Labware. 96 well plates were used in place of 384 well plates for ease of analysis. Various dilutions of beads were prepared using a Beckman Multisizer Coulter Counter 2. 25 μL of each dilution was loaded into each well of the 96 well plate. Wells of the plate were then analyzed using conventional microscopy and a Nikon TE-2000 microscope and the number of beads in each well was recorded.

Misclassification Analysis

The probability of misclassification was computed using the Mathematica 6.0 script provided in Appendix B. The misclassification data computed from Equation 6.2 were fit to the following three parameter Hill's function model.

$$P(I) = \frac{a\, b^k}{b^k + D^k} \qquad \text{Equation (6.3)}$$

where D is 1+log(# DNA copies per sample), a represents the saturation value of misclassification probability which occurs as the value of D approaches 1, b represents the value of D corresponding to half the saturation probability, and k controls the steepness of the probability curve. The larger the value of k, the steeper the probability curve.

Results

Ct values generated from nearly 2,400 real time PCR reactions are shown in Figure 6.1. This figure demonstrates that the reliability of the Ct method as a proxy for sample concentration diminishes as sample concentration decreases. Distinct step functions of Ct values occur for sample concentrations greater than 12,000 copies of DNA per reaction, whereas for sample concentration below 12,000 copies, the mean of the Ct values for each sample concentration becomes less distinct. Although the tight bands of data for higher DNA concentrations rapidly diffuse into smears of data as the DNA concentration approaches zero, a simple t-test (data not shown) shows that the sample mean for each of the replicates of each class is statistically different as long as the entire cohort of data for each DNA concentration is used. In every sample, no Ct values were recorded for the respective 8 negative controls, indicating neglible contamination levels for each sample.

The same observation can be made by analyzing the raw data distributions (binned at 0.1 Ct) as shown in Figure 6.2. The Ct value distributions of higher DNA concentrations are punctate and distinct, whereas the low concentration distributions are disperse and overlapping. The transition from punctate Ct values to disperse values occurs over the 100-10,000 copy number range. However, the distributions do not show a completely monotone behavior as shown in the 10^6 copy distribution. This could be due to experimental error or non-monotone behavior within the PCR process.

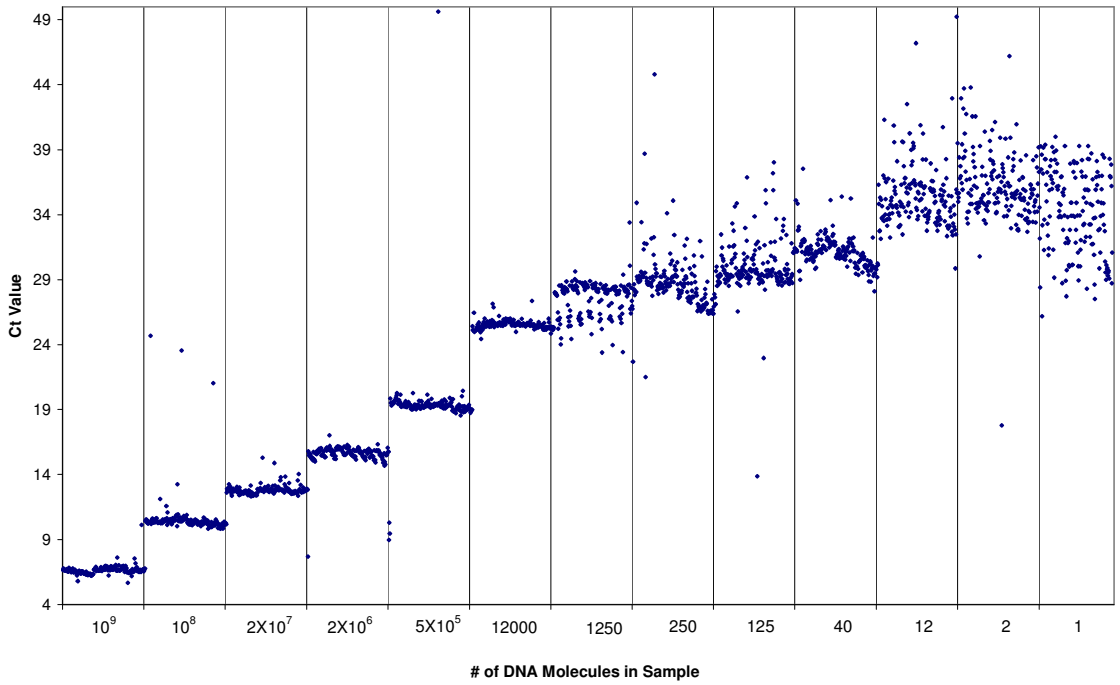


Figure 6.1: Ct values binned in order of decreasing DNA concentration. Each DNA concentration obtains Ct values from 184 identical replicates. DNA concentrations range over nine orders of magnitude.

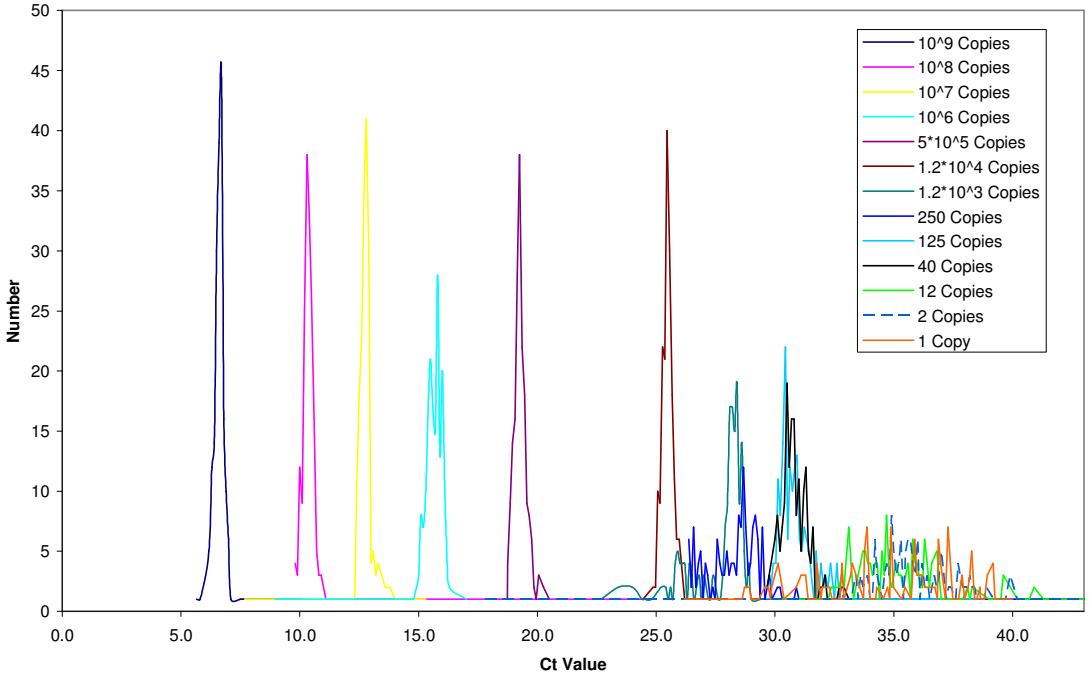


Figure 6.2: The data distribution functions for each class of Ct values associated with a particular DNA concentration. DNA concentrations are represented by color.

The entire set of Ct values for all DNA concentrations can be used to generate a standard curve as shown in Figure 6.3 as would be typically done using only a few replicates [1,2,20]. This standard curve was subsequently used to convert variances in Ct values into variances of DNA concentrations. This permits the analysis of the variance associated with using quantiles of Ct value data as depicted in Figure 6.4. In this case, the 80%, 60%, and 20% inner quantiles of the entire class of Ct data were determined and the standard curve was used to convert the variance of Ct values into the variance of DNA. The variance of predicted DNA concentrations could then be used to compute the relative error associated with each sample. Interestingly, Figure 6.4 shows that the expected relative errors for samples greater than 10^6 copies per reaction in concentration is less than 1%, demonstrating the enormous power of running many replicates of PCR samples. On the other hand, samples containing less than 10^5 copies of DNA per reaction have relative errors greater than 100%, making reliable quantitation from single samples virtually impossible. This analysis explains why others have found quantitation of dilute solutions of DNA difficult [7,23]. One should notice that the linear trend of Figure 6.4 is likely imposed by the normalization process, which involves dividing the actual errors by the perceived DNA concentration of each sample. For this reason, the actual error associated with each sample was computed for the 60% quantile and plotted in Figure 6.5. This figure demonstrates that the actual error follows a nearly linear trend as well. Higher DNA concentrations, 10^9 copies, display a variance in DNA concentration of about 100 molecules (10^{-7} %) and can be quantified with miraculous precision by utilizing the large number of replicates used in this study. However, more dilute samples, less than 10^5 copies, display enormous error making accurate quantitation of sample concentration far less reliable.

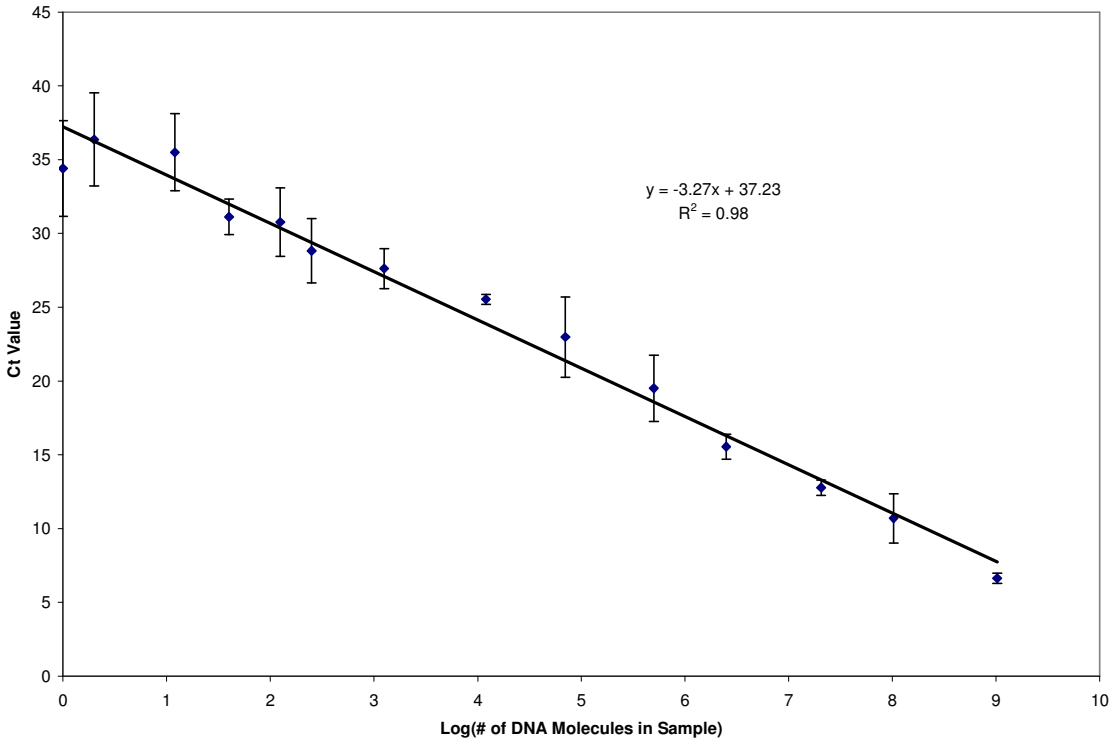


Figure 6.3: Average Ct values and ± 1 standard deviation error computed from the data of Figure 6.1. The black line is a least-squares linear regression of the data similar to those typically used in absolute gene expression quantitation.

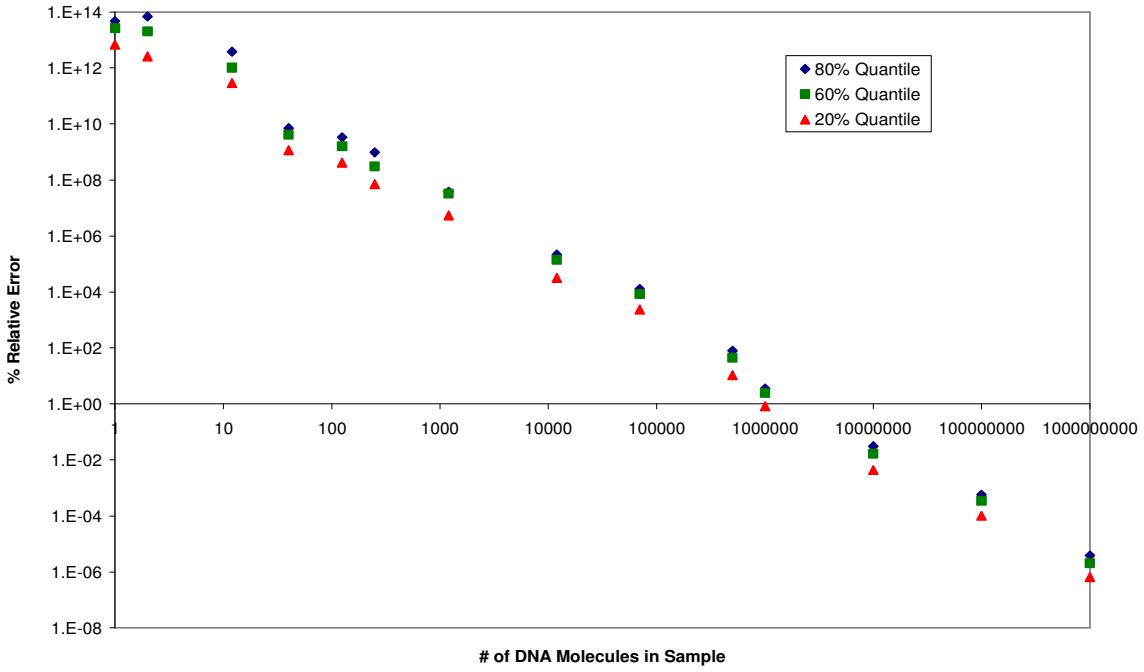


Figure 6.4: The relative error associated with each DNA concentration. Errors were computed using the range of Ct values associated with each quantile and the linear regression formula of Figure 6.3. The relative error of each DNA concentration was then computed by dividing the error by the perceived DNA concentration.

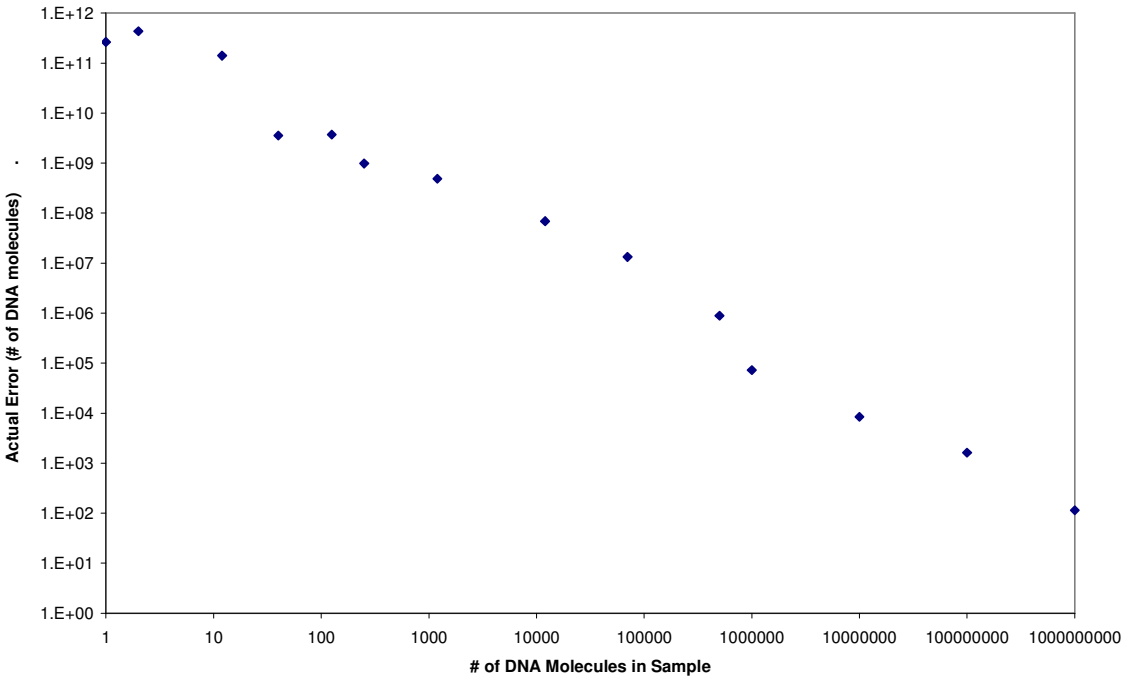


Figure 6.5: The actual error associated with each sample concentration using the 60% quantile of the data depicted in Figure 6.1. Variances in Ct values were converted to variances in DNA concentration using the linear regression in Figure 6.3.

Ct data can also be used to compute misclassification probability. We have developed a Mathematica script, see Appendix B, that computes the misclassification frequencies associated with each DNA concentration. This algorithm can also be used to mimic the process of running multiple independent sample replicates. In this process, multiple Ct values are chosen from a class of Ct data, their average is computed, and lastly the average Ct value is used for misclassification analysis. The total probability of misclassification for each class of Ct values is represented as the sum of the misclassification probability for each Ct value with that particular class. After computing the misclassification probabilities, the results were subsequently fit to a three parameter Hill's function. This allows an enormous amount of data to be represented in a simple tabular form as shown in Table 6.1. Figure 6.6 represents the misclassification probability for the data of Figure 6.1.

Table 6.1: Parameter values obtained from the least-squares fit to the Hill's function model, Equation (6.3), to the probability of misclassification data computed from Equation (6.2). These values were obtained from the 'FindFit' algorithm of Mathematica 6.0.

Hill's Function Parameter Values			
# of averaged replicates	*a*	*b*	*k*
1	0.52	4.46	9.58
3	0.59	4.08	6.80
6	0.58	3.92	6.91
8	0.54	3.89	7.71
12	0.45	3.60	9.40

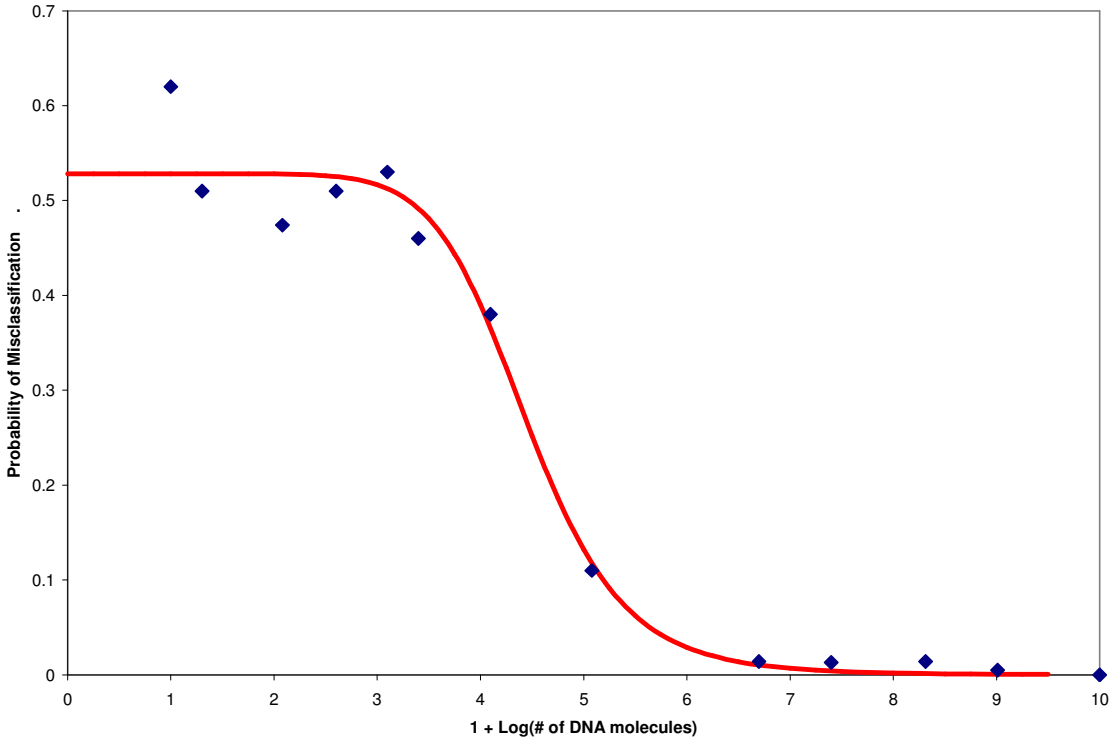


Figure 6.6: The misclassification frequency for each DNA concentration computed from Equation (6.2). These data can be modeled by a three parameter Hill's function, see Equation (6.3) shown in red. 0 represents no probability of misclassification and 1 represents that misclassification is certain.

The experimental data and Hill's function model in Figure 6.6 show that for concentrations of DNA greater than 10^6 copies, the chance of misclassification is virtually zero. However, for concentrations less than 1,000 copies of DNA, the chance of misclassification is roughly 50%, making the process ironically similar to coin flipping. The transition from low to high misclassification probabilities occurs rapidly over 2 to 3 orders of magnitude in sample concentration. Ironically, the upper portion of this transition corresponds to a Ct value of 26 to 28 cycles, which is the upper limit of the vast majority of reported Ct values within scientific literature. There is a long standing tradition of running replicates of rt-PCR samples to minimize errors. Typically, reactions are performed in triplicate and the end result is considered to be the average Ct value of the independent samples. The misclassification probability can also be used to analyze the benefit of running sample replicates. This analysis was implemented by randomly choosing multiple Ct values and averaging them prior to misclassification analysis. The misclassification probabilities were then fit to Equation (6.2) as was done in the case of running a single sample. Analysis of these functions demonstrates an advantage to running independent replicates of identical samples, as shown in Figure 6.7. Surprisingly, the effects of running replicates are not dramatic until at least twelve replicates are used. The decrease in misclassification probability for running 3, 6, 8 and 12 replicates is marginal, see Figures 6.7 and 6.8.

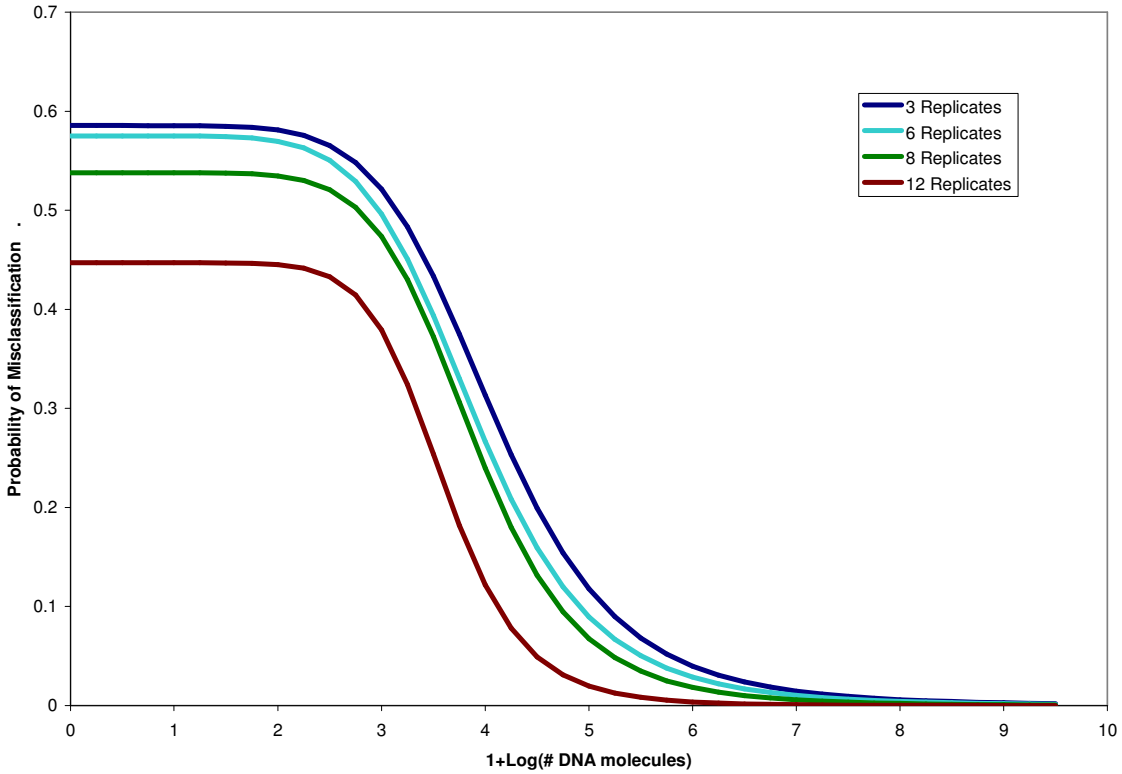


Figure 6.7: The probability of misclassification of Ct values was computed from Equation (6.2) and modeled using the three parameter Hill's function of Equation (6.3). In each case, simulations were performed by randomly choosing *n*, the number of identical replicates, Ct values from a given class. These values were then averaged to obtain the Ct value used for determining misclassification frequency.

Even running 12 independent replicates produces a misclassification frequency greater then 0.3 for samples more dilute than 1,000 copies per sample. In some cases, sample size limitations may make running many replicates impossible. This situation would be typically encountered when analyzing the gene expression from a single cell. For example, say 100 RNA molecules are harvested from a single cell for gene expression analysis and are subsequently transformed into 100 DNA molecules via reverse transcription. If these molecules were split into 12 replicates, each sample would contain approximately 8 molecules of DNA, making each sample subject to enormous error (see Figure 6.4). Since we and others have interests in the idea of measuring gene expression from single cells, we have implemented strategies utilizing alternative formats for performing PCR on dilute samples with the intentions of developing an accurate and robust assay.

Samples run under the alternative format were prepared identically to those prepared for traditional rt-PCR. However, in this case, the DNA concentration of the supermix was such that each sample only had a probability of receiving a piece of DNA. Of course, there also exists the probability that a particular reaction obtains multiple pieces of DNA. This issue will be discussed later in this chapter. In the alternative format, samples are assayed by whether or not a particular reaction crosses a minimum fluorescence threshold before 40 cycles. By measuring TAQ polymerase efficiency, we have obtained data indicating that reactions containing DNA will amplify before 40 cycles if they are going amplify at all, see Figure 6.8. Figure 6.8 shows results from 16 samples identical in composition, but pre-worn through thermal cycling for various time periods. The data indicate that after 40 cycles of thermal cycling, the vast majority of amplification ability

of TAQ polymerase is lost. Thus, we set the cycle threshold at 40 cycles for the endpoint PCR assay.

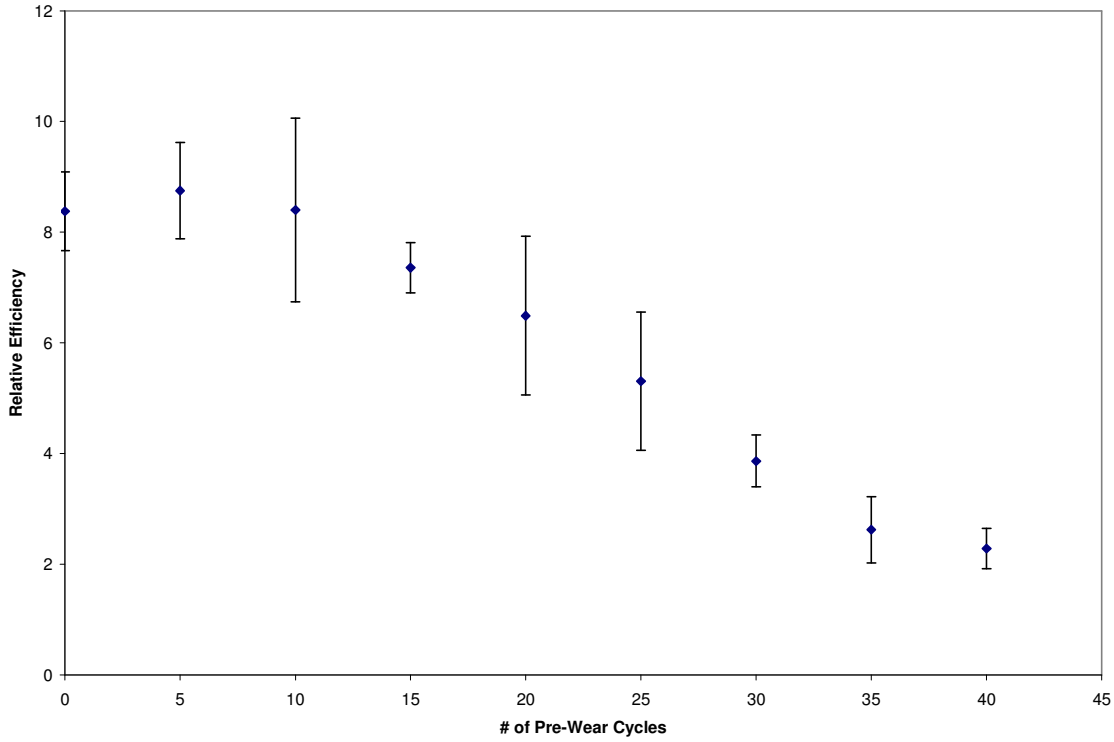


Figure 6.8: Relative TAQ polymerase efficiency as a function of thermo-cycling pre-wear. The horizontal axis represents the number of cycles each sample was thermo-cycled prior to the addition of sample DNA. 0 represents the case where fresh TAQ polymerase was used without pre-wear. The relative efficiency is represented as the average derivative of relative fluorescence with respect to change in cycle number over the amplification curves. Error bars represent ± 1 standard deviation over three independent experiments.

Results from 16 independently prepared supermixes containing as few as 5 molecules of DNA and as many as 184 pieces of DNA are shown by the green triangles in Figure 6.9. Each supermix was spread evenly over 184 wells of a 384 well plate so that each well had only a probability of obtaining DNA. The *perceived* DNA concentration is plotted on the horizontal axis and the number of un-amplified wells is plotted on the vertical axis. Interestingly, the data are monotone and saturate at a DNA concentration of approximately 200 copies. Since the steepness of the slope of the experimental data is in essence the sensitivity of the assay, DNA concentrations below 50 copies of DNA are subject to little resolving power. However, DNA concentrations between 50 and 200 copies appear to occur within a region of reasonable sensitivity, see the green triangles in Figure 6.9.

The process of plate filling can be modeled as a random process. We chose to do this using a Mathematica 6.0 script using the 'Random Integer' function, see Appendix D. The result from this simulation is shown in of Figure 6.9. It is fairly obvious that the raw data (green triangles) do not fit the theoretical model. However, if the *perceived* DNA concentrations are skewed by a multiplicative factor of 2.4, the data fit the theory reasonably well. The normalized data are represented by the red dots in Figure 6.9. It is conceivable that such an error may be generated through serial dilutions, see Table 6.2. This table demonstrates the errors encountered when performing 1:100 serial dilutions with a 200 nL carry-over. In our experiments, some of the serial dilutions were prepared by placing 1 μL of a concentrated DNA stock into 99 μL of nuclease free water. It is feasible that an additional 200 nL is carried over through material wetting of the pipette tip from the concentrated DNA stock into the new dilution. In the test case shown in

Table 6.2, an error factor of 2.2 is obtained as the end result. However, this error does not account for pippetter error (2-5%), human error, measurement error of the spectrophotometer (2.5%), or DNA adsorption onto the surface of the pipette tip. All of these errors could increase the overall error factor.

The inherent error of the plate filling process can also be investigated via simulation. Figure 6.10 represents the standard deviation in terms of the number of DNA molecules obtained by repetitive simulation of the plate filling process. These simulations indicate that variations in the plate filling process are minimal for low DNA concentrations and maximize near the point where the number of DNA molecules on the plate equal the number of wells on the plate. In fact, variations in the plate filling process decrease as DNA concentration decreases for samples containing fewer molecules of DNA than there are wells on the plate. It is notable that the plate filling process is extremely robust since the maximal error of the plate filling process is approximately 4 molecules of DNA. The sensitivity of the assay can also be described by the slope of the plate filling curve in Figure 6.9. This slope is maximized for a single molecule and decays as the number of molecules on the plate increases. Since the assay sensitivity is maximized for lower DNA concentrations, say 100 molecules or less, the robustness and accuracy of the endpoint PCR assay should be maximized for case where the number of molecules in the DNA solution is equivalent to half the number of wells on the plate or less. The assay is most sensitive in this region because the probability of a well containing multiple copies of DNA is low. As the number of DNA molecules in the sample increases, the probability of a well containing multiple copies of DNA also increases. Since the assay

cannot distinguish a well that contains a single copy of DNA from one that contains multiple copies, wells containing multiple copies introduce variability into the assay.

Table 6.2: A test case demonstrating the compiled error through serial dilutions. Four serial dilutions are shown. In each case, 1 μL of a concentrated stock with 200 nL of incidental carry-over is placed into 99 μL of water. The end result is an error fold of 2.2. This analysis does not include systematic pippetting error, human error, or errors in measurement of the initial stock concentration.

Perceived Conc.	Actual Conc.	Error Fold
1000000000	1000000000	1.0
10000000	12000000	1.2
100000	144000	1.4
1000	1728	1.7
10	22	2.2

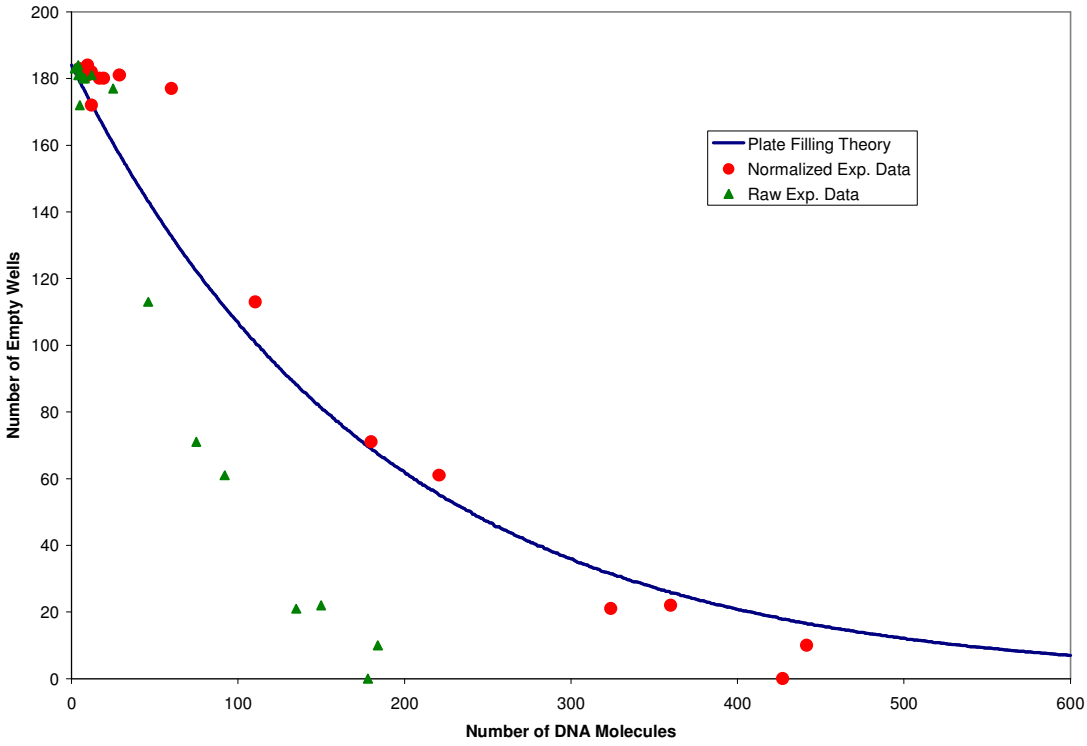


Figure 6.9: The number of amplified wells as a function of the number of DNA molecules spread over 184 wells. The green triangles represent experimental data. The solid blue line is the theoretical prediction for modeling plate filling as a random process. The red dots represent the data normalized by a multiplicative factor of 2.4 to make the data more accurately fit theoretical predictions.

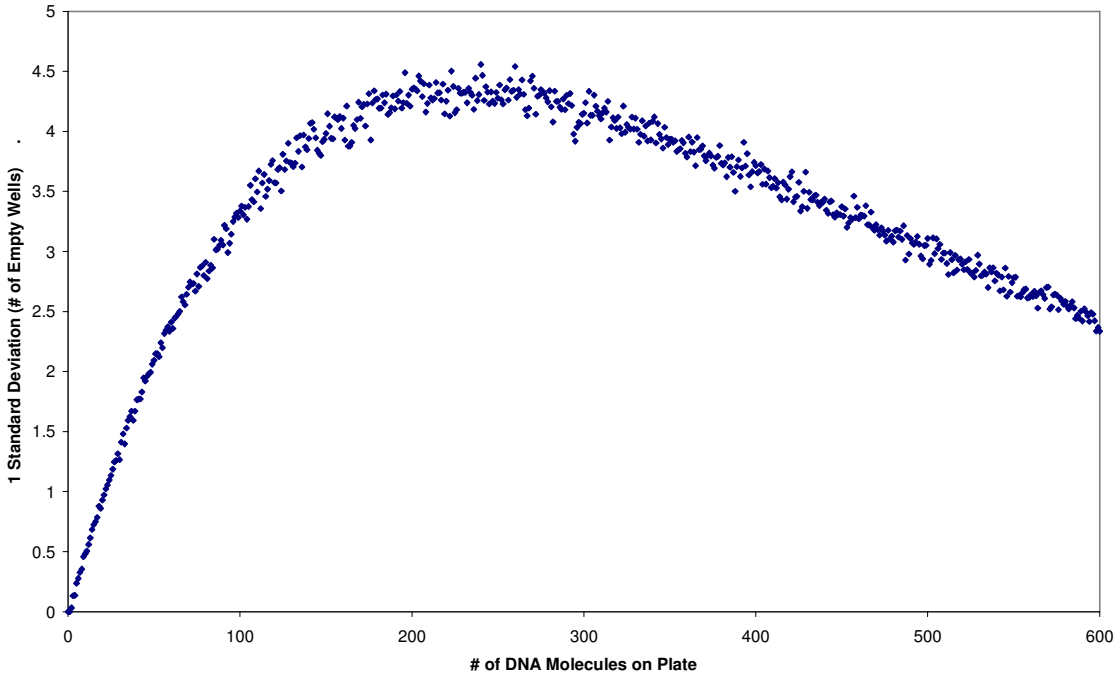


Figure 6.10: The standard deviation of the plate filling process in terms of the number of empty wells. These results were obtained by performing 1,000 simulations of the plate filling process using the algorithm in Appendix D.

Although the theoretical model for plate filling could not be directly verified for DNA, the model could be experimentally tested using larger, detectable objects such as silicon beads. Dilutions of 20 μm beads were prepared in the same manner as for DNA solutions. The beads were then loaded into 96 well, flat bottom plates and the plates were analyzed using conventional microscopy to determine the number of empty wells (wells that did not contain any beads). The data and random plate filling model is shown in Figure 6.11. In this case, the theoretical model accurately captures the dynamics of the plate filling process. However, in nearly every case, the experimental data fall below the theoretical predictions. This is likely due to human error. It is unlikely that a non-existent bead could be counted while visually surveying each well. However, it is likely that some fraction of beads is overlooked during the analysis, since the field of view of a microscope objective is much smaller than the total area of a well. This requires the total area of a well to be systematically panned so that the entire surface area of a well can be analyzed during viewing, leading to the possibility that some areas may be overlooked during analysis. Overlooking a bead could have two separate affects on the data. If the overlooked bead is within a well with beads which were noticed, the overlooked bead would simply shift that particular data point to the left, moving it further away from the theory, see Figure 6.11. However, if the overlooked bead was within a well that is scored as empty, the data point would be shifted upward, moving it closer to the theoretical curve. Nevertheless, there appaears to be a systematic error within the process since nearly all data points lie below the theoretical curve and seem to have the same shape as the theoretical predictions.

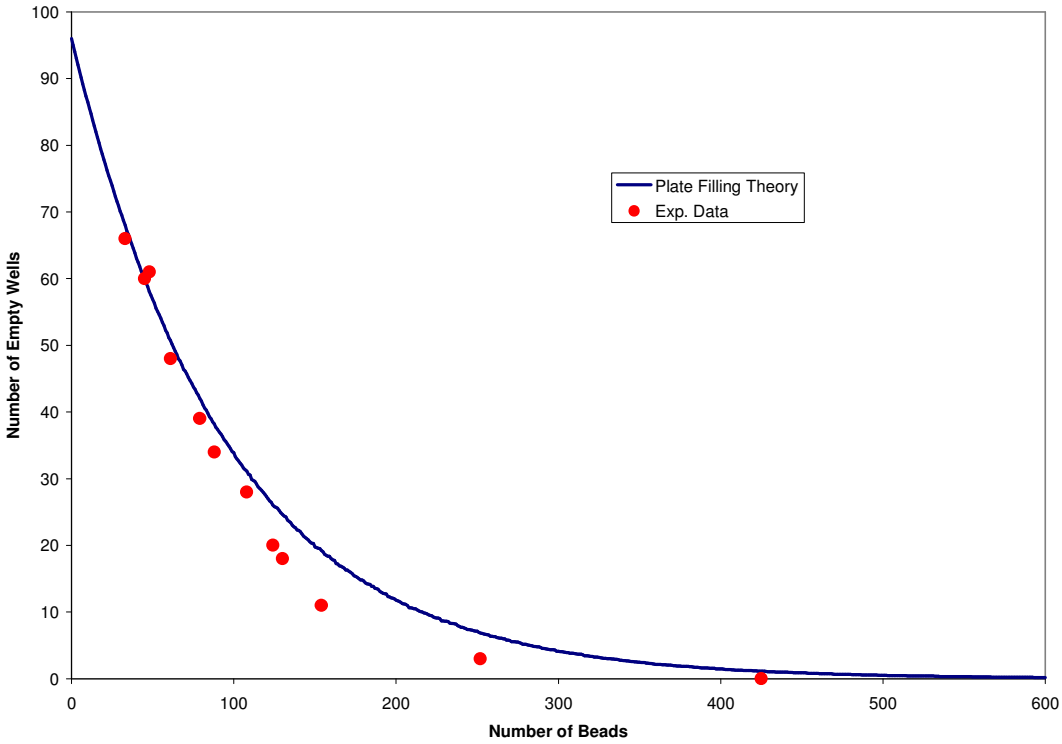


Figure 6.11: The plate filling process for a 96 well plate. Experimental data, shown by the red dots, were obtained using 20 μm beads and the theoretical prediction, solid blue line, was obtained using the algorithm in Appendix D.

Conclusions

Our data indicate that the robustness of the Ct Method for analyzing rt-PCR data is strongly dependent on the sample concentration. An analysis of Ct value variance demonstrates that the variation of Ct values is larger than the sample DNA concentration for sample concentrations less than 10^5 copies per reaction. On the other hand, samples containing more than 10^7 copies per reaction can be quantified with extremely high accuracy, see Figure 6.4. The data within this study indicate that the misclassification probability is nearly 0 for sample concentrations greater than 10^7 copies per reaction, but is larger than 0.3 for samples less than 10^3 copies per reaction. The number of replicates performed can decrease the chance of misclassification, but only marginally for samples containing less than 10^3 molecules of DNA. In general, it appears that the number of replicates should be increased for more dilute samples of DNA. However, if the original DNA sample, obtained from a cell lysate, contains only 1,000 DNA molecules or less, splitting the sample to perform replicates will likely only decrease the accuracy of the Ct Method and increase the probability of misclassification. The data of this study indicate that for DNA samples containing less than 1,000 molecules of DNA, the Ct method will likely not provide accurate results, see Figure 6.4, 6.5, and 6.6.

The process of using an endpoint PCR analysis shows significant promise, especially for dilute samples of DNA. Statistical simulations show the process of plate filling can be performed such that the number of empty wells can provide a robust statistic for endpoint PCR quantitation. Experimentally, this process was demonstrated using silicon beads. However, experimental data for plate filling with DNA and monitoring by PCR

amplification did not fit theoretical predictions. We believe this failure stems from dilution errors and the inability to accurately quantify concentrated stock solutions of DNA. In most cases, the number of amplified wells was over-predicted by theoretical models. It appears that the perceived DNA concentration of serial diluted stocks were about 2.4 –fold off. Since, a minimum of 4 pippetting steps were required in the creation of these stocks, such an error is conceivable. Typical pippetting errors range from 2 to 5%, but this does not account for any additional material carried over on the outside of the pipette tip due to material wetting. For example, if 200 nL of sample volume is carried over on the outside of the tip during a 1 μL pippetting step, a 20% error is introduced at every step. Another concern is DNA adsorption onto the surface of pipette tips. This would promote the transfer of additional DNA molecules beyond the perceived amount calculated assuming well-mixing.

This study has demonstrated that the Ct Method is extremely powerful tool for quantifying concentrated samples of DNA, but will likely provide inaccurate results for samples containing less than 1,000 copies of DNA per reaction. The original intent of this study was to evaluate the usefulness of an alternative method, which we call endpoint PCR analysis, compared to the Ct Method. However, the two processes are completely different in nature, making it difficult to derive a fair statistic for comparing the two processes. Based on the experimental data of this study, samples containing less than 1,000 copies of DNA would be more accurately quantified using endpoint analysis based on the fact that our data show the Ct Method produces errors greater than the sample concentration in this realm.

References

1. Abramov, D., Trofimov, D., and Rebrikov, D. 2006. Accuracy of a real-time polymerase-chain-reaction assay for quantitative estimation of genetically modified food sources in food products. *Applied Biochemistry and Microbiology*. **42**: 485-488.

2. Booth, C., Griffith, E., Brady, G., and Lydall, D. 2001. Quantitative amplification of single-stranded DNA (QAOS) demonstrates that cdc13-1 mutants generate ssDNA in a telomere to centromere direction. *Nucleic Acids Research*. **29**: 4414-4422.

3. Cady, N., Stelick, S., Kunnavakkam, M., Lui, Y., and Batt, C. 2004. A microchip-based DNA purification and real-time PCR biosensor for bacterial detection. 2004. *Proceedings of IEEE Sensors*. 1191-1194.

4. Cook, P., Fu, C., Hickey, M., Han, E., and Miller, K. 2004. SAS programs for real-time RT-PCR having multiple independent samples. *Biotechniques.* **37**: 990-995.

5. Diehl, F., Li, M., Dressman, D., Yiping, H., Shen, D., Szabo, S., Diaz, L., Goodman, S., David, K., Juhl, H., Kinzler, K., and Vogelstein, B. 2005. Detection and quantification of mutations in the plasma of patients with colorectal tumors. *PNAS*. **102**: 16368-16373.

6. Freeman, W., Walker, S., and Vrana, K. 1999. Quantitative rt-PCR: pitfalls and potential. *Biotechniques*. **26**: 112-125.

7. Gilliland, G., Perrin, S., Blanchard, K., and Bunn, F. 1990. Analysis of cytokine mRNA and DNA: detection and quantitation by competitive polymerase chain reaction. *PNAS*. **87**: 2725-2729.

8. Ishii, T., Wallace, A., Zhang, X., Gosselink, J., Abboud, R., English, J., Pare, P., and Sanford, A. 2006. Stability of housekeeping genes inalveolar macrophages from COPD patients. *European Respiratory Journal*. **27**: 300-306.

9. Jagers, P. and Klebaner, K. 2003. Random variation and concentration effects in PCR. *Journal of Theoretical Biology*. **224**: 299-304.

10. Kariv, I., Fereshteh, M., and Oldenburg, K. 2001. Development of a miniaturized 384-well high throughput screen for the detection of substrates of cytochrome P450 2D6 and 3A4 metabolism. *J. of Biomolecular Screening*. **6**:91-99.

11. Karsai, A., Muller, S. Platz, S., and Hauser, T. 2002. Evaluation of a home-made SYBR Green 1 reaction mixture for real-time PCR quantification of gene expression. *Biotechniques-Short Technical Reports*. **32**: 790-796.

12. Lalam, N., Jacob, C., and Jagers, P. 2004. Modeling the PCR amplification process by a size-dependent branching process and estimation of the efficiency. *Advanced Applied Probability*. **36**: 602-615.

13. Larionov, A., Krause, A., and Miller, W. 2005. A standard curve based method for relative real time PCR data processing. *BMC Bioinformatics*. **6**: 62.

14. Liu, W. and Saint, D. 2002. A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. *Analytical Biochemistry*. **302**: 52-59.

15. Matsubara, Y., Kerman, K., Kobayashi, M., Yamamura, S., Morita, Y., and Tamiya, E. 2004. Microchamber assay based DNA quantification and specific sequence detection from a single copy via PCR in nanoliter volumes. *Journal of Biosensors and Electronics*. **20**: 1482-1490.

16. Mitra, R. and Church, G. 1999. In situ localized amplification and contact replication of many individual DNA molecules. *Nucleic Acids Research*. **27**: e34.

17. Nedelman, J., Haegerty, P., and Lawrence, C. 1992. Quantitative PCR: procedures and precision. *Bulletin of Mathematical Biology*. **54**: 477-502.

18. Peirson, S., Butler, J., and Foster, R. 2003. Experimental validation and conventional approaches to quantitative real-time PCR data analysis. *Nucleic Acids Research*. **31**: e73.

19. Pfaffl, M. 2001. A new mathematical model for relative quantification in real-time PCR. *Nucleic Acid Research*. **29**: e00.

20. Roussel, Y., Harris, A., Lee, M., and Wilks, M. 2007. Novel methods of quantitative real-time PCR data analysis in murine *Helicobacter pylori* vaccine model. *Vaccine*. **25**: 2919-2929.

21. Samatov, T., Chetverina, H., and Chetverin, A. 2006. Real-time monitoring of DNA colonies growing in a polyacrylamide gel. *Analytical Biochemistry*. **356**: 300-302.

22. Schnell, S. and Mendoza, C. 1997. Closed form solutions for time-dependent enzyme kinetics. *Journal of Theoretical Biology.* **187**: 207-212.

23. Vaerman, J., Saussoy, P., and Inargiola, I. 2004. Evaluation of real-time PCR data. *Journal of Biological Regulators and Homeostatic Agents*. **18**: 212-214.

24. Williams, R., Peisajovish, S., Miller, O., Magdassi, S., Tawfik, D., and Griffiths, A. 2006. Amplification of complex gene libraries by emulsion PCR. *Nature Methods*. **3**: 545-550.

25. Xu, X. and Ikai, A. 1998. Retrieval and amplification of a single-copy genomic DNA from a nanometer region of chromosomes: a new and potential application of atomic force microscopy in genomic research. *Biochemical and Biophysical Research Communications*. **248**: 744-748.

26. Yuan, J., Reed, A., Chen, F., and Stewart, C. 2006. Statistical analysis of real-time PCR data. *BMC Bioinformatics*. **7**: 85.

27. Zhang, H., Cheng, X., Richter, M., and Greene, M. 2006. A sensitive and high-throughput assay to detect low-abundance proteins in serum. *Nature Medicine*. **12**: 473-477.

CHAPTER VII

CONCLUSIONS

This dissertation work focuses on developing quantitative techniques to improve the quality of biological measurement. Currently, there is a vast interest in interrogating the genetic circuitry of living organisms. However, most techniques of biological measurement were not developed with the concept of accurate quantitation in mind. Thus, there is a demand for novel quantitative measurement strategies and comprehensive studies that evaluate the accuracy and reproducibility of existing techniques.

The work presented here focuses on developing quantitative techniques to robustly harvest and measure biological molecules that drive genetic circuits. By applying engineering intuition empowered by mathematical and biological concepts, advances were made on many of the problems that plague the accuracy of traditional biological measurements. In fact, it was only by interdisciplinary approaches that such advances were made. Notable advances were made in a variety of areas including: reliable cell disruption, cell cycle synchrony modeling, and reliable gene expression quantitation via PCR. These improvements have vast utility that span many areas of study such as gene expression analysis, yeast physiology, and industrial bioprocesses. Chapters IV and V of this dissertation demonstrate the widespread value of this work by describing how a 40-year old biological phenomena could be exploited to make accurate cell cycle dependent gene expression measurements and increase the efficiency of industrial bioprocesses.

The principal findings of this research are:

- A traditional protocol for cell disruption produces a mere 57% disruption, while an extension to the protocol can produce over 90% disruption.
- Cell disruption via bead beating is a cell cycle dependent process. Ignoring this cell cycle dependence will impact the accuracy of cell cycle dependent biological measurement. Recognizing the cell cycle dependence of the process permits its efficient use for the selective harvest of cell cycle dependent biological molecules.
- The dynamics cell cycle synchrony of *Saccharomyces cerevisiae* can be accurately captured using an age and volume structured Leslie model.
- Continuous volume filtration can extend cell cycle synchrony through judicious selection of filtration parameters.
- Autonomous oscillations during continuous yeast cultures are integrally linked to the population structure of the culture and are therefore a mechanism of persistent cell cycle synchrony.
- Bud index oscillations and cell density measurements can be used to unveil the population structure during autonomous oscillations. The population structure may elucidate the mechanisms driving autonomous oscillations ultimately leading to advancements in understanding intercellular communication and cell cycle control.
- Autonomous oscillations can be exploited to measure cell cycle dependent gene expression. These oscillations also have enormous potential within industry for the synchronous production of biological molecules.
- The reliability of the Ct Method for the quantitation of gene expression depends crucially on the concentration of DNA within the sample.

• The misclassification probability is a useful statistic for evaluating the reliability of Ct data.

• Endpoint PCR has the potential for quantifying subtle gene expression profiles.

• The process of plate filling single molecules is robust and suitable for use in endpoint assays used for quantitation.

Future Work

Much of this work has yet to come to complete fruition. For example, the utility of the endpoint PCR assay needs further investigation. This dissertation demonstrates that this assay may be empowered to more accurately quantifying subtle gene expression profiles. However, the assay needs a more rigorous evaluation that relies heavily on the ability to create accurate dilute DNA standards, which is currently technologically difficult. Perhaps future technological breakthroughs in this area will permit the full demonstration of the advantage of endpoint PCR quantitation.

We have investigated the idea of using single cells as carriers of DNA in order to alleviate the difficulties in accurately preparing dilute suspensions of DNA. With the use of a Coulter counter, accurate stocks of dilute suspensions of cells can be prepared. If these cells are haploid yeast, then each cell will contain exactly one copy of a particular gene per cell. Individual yeast cells can also be easily visualized once loaded onto a plate, unlike single DNA molecules. However, the use of cells as carriers of DNA introduces a new problem. The TAQ polymerase and primers need acess to the DNA inside the cells in order to efficiently amplify the DNA. Thus, the cells must be

efficiently and robustly disrupted. We have performed numerous experiments of this nature, and verified that yeast cells prepared as spheroplasts, cells lacking cell walls, can be efficiently disrupted inside the plate. However, PCR reactions still tend to fail unless several cells are confined to a single well of the plate. This may be due to inefficient disruption of the nuclear membrane or supercoiling of the DNA, both of which would inhibit TAQ polymerase from gaining access to the DNA. Another potential problem may be deoxyribonucleases (DNAses) or proteases released during the disruption which may damage the DNA or the TAQ polymerase enzyme. This problem needs to be further investigated with the use of different cell types, protease and DNAse inhibitors, and different primer sets.

The complete value of the Leslie model has not been fully demonstrated. This model could be used to deconvolve population signals, potentially answering countless questions regarding cell cycle and age dependent gene expression profiles. Understanding the behavior of these expression profiles in concordance with physiology will encourage understanding the adaptive control and robustness of the living organism. We have theoretically demonstrated with the Leslie model that autonomous oscillations could be used to increase the efficiency of cell cycle dependent bioprocesses. The luciferase gene expression data for POL1, DAL80, and GLN3 also indicate that periodic and cell cycle dependent production of proteins during autonomous oscillations is reality. However, the true value of this system for bioprocessing needs to be investigated more thoroughly through experimentation. The degradation rate of the specific proteins products both inside the cell and in the media needs to be determined. If the protein product will simply accumulate inside the cell with minimal degradation, then it is likely

that the oscillations in product concentration, induced by the population structure of the autonomous oscillations, will be negligible.

APPENDIX A

LESLIE MODEL MATLAB CODE

```
Params = ReadParams('Input.txt');
GrowthRates = dlmread(Params.GrowthRatesFile);
VolumeIntervals = dlmread(Params.VolumeIntervalsFile);
VolumeMilestones = dlmread(Params.VolumeMilestonesFile);
SizeGrowthRates = size(GrowthRates);
SizeVolumeIntervals = size(VolumeIntervals);
SizeVolumeMilestones = size(VolumeMilestones);
if (Params.NumberOfGenerations ~= SizeGrowthRates)
   error('number of generations and number of growth rates do not match')
end
if (Params.NumberOfGenerations ~= SizeVolumeIntervals(1))
   error('number of generations and number of volume intervals do not match')
end
if (Params.NumberOfGenerations ~= SizeVolumeMilestones(1))
   error('number of generations and number of volume milestone sets do not match')
end
for i=1:Params.NumberOfGenerations
NumberOfTimeIntervals(i,1)=[ceil((1/GrowthRates(i,1))*log(VolumeIntervals(i,2)/VolumeIntervals(i,1)))+
1];
end
for i=1:Params.NumberOfGenerations
   for j=1:NumberOfTimeIntervals(i,1)
      VolumeValues(i,j)=VolumeIntervals(i,1)*exp((j-1)*GrowthRates(i,1));
      SineValues(i,j) = sin((2*pi/(NumberOfTimeIntervals(i,1) - 1)) * (j - 1));
   end
end
NewParentVolume(1) = VolumeMilestones(1, 3);
Period(1) = log(VolumeMilestones(1, 2)/NewParentVolume(1)) / GrowthRates(1,1);
for i=2:Params.NumberOfGenerations
   NewParentVolume(i) = VolumeMilestones(i-1, 2) - VolumeMilestones(i-1, 3);
   Period(i) = log(VolumeMilestones(i, 2)/NewParentVolume(i)) / GrowthRates(i,1);
end
N = ceil(Period);
for i=1:Params.NumberOfGenerations
   for j=1:N(i)
      SineValues2(i,j) = sin((2*pi/(Period(i) - 1)) * (j - 1));
   end
end

if (Params.CalcGeneExpression == 1)
   for i=1:Params.NumberOfGenerations
      MeanBirthIndex = ClosestVolumeIndex(i, NewParentVolume(i), GrowthRates, VolumeIntervals);
      P = NumberOfTimeIntervals(i,1) - MeanBirthIndex;
      for j=1:MeanBirthIndex
         SineValues_GrowthInterval(i,j) = 0;
      end
      for j=(MeanBirthIndex+1):NumberOfTimeIntervals(i,1)
         SineValues_GrowthInterval(i,j) = sin(((2*pi/(P)) * (j-1)) - ((2*pi/P)*MeanBirthIndex) );
      end
```

```
    end
end
for i=1:Params.NumberOfGenerations
DivisionVolumeIndex(i,1)=ClosestVolumeIndex(i,VolumeMilestones(i,2),GrowthRates,VolumeIntervals);
MeanBudPointIndex(i,1)=ClosestVolumeIndex(i,VolumeMilestones(i,1),GrowthRates,VolumeIntervals);
end
if (Params.ParentSpread == 1)
    for i=1:(Params.NumberOfGenerations - 1)
        MeanMotherVolume = VolumeMilestones(i,2)-VolumeMilestones(i,3);
        MeanParentIndex(i+1) =
ClosestVolumeIndex(i+1,MeanMotherVolume,GrowthRates,VolumeIntervals);
        SumPPWeight = 0;
        for j=1:(MeanBudPointIndex(i+1,1) - 1)
            PPWeight(i+1,j) = exp(-(VolumeValues(i+1,j) -
VolumeValues(i+1,MeanParentIndex(i+1)))^2/(Params.ParentSpreadSigma)^2);
            SumPPWeight = SumPPWeight + PPWeight(i+1,j);
        end
        for j=1:(MeanBudPointIndex(i+1,1) - 1)
            PPWeight(i+1,j) = PPWeight(i+1,j)/SumPPWeight;
             if (PPWeight(i+1,j) < eps)
               PPWeight(i+1,j) = 0.0;
            end
        end
    end
end
TenPercentWidth = ones(Params.NumberOfGenerations,1);
for j=1:Params.NumberOfGenerations
    TenPercentWidth(j,1) = (VolumeIntervals(j,2) - VolumeIntervals(j,1))/10;
end
TenPctStart = ones(Params.NumberOfGenerations,1);
for j=1:Params.NumberOfGenerations
    TenPctStart(j,1) = ClosestVolumeIndex(j, (VolumeValues(j,MeanBudPointIndex(j,1)) -
TenPercentWidth(j,1)), GrowthRates, VolumeIntervals);
end
StaticData.GrowthRates = GrowthRates;
StaticData.VolumeIntervals = VolumeIntervals;
StaticData.NumberOfTimeIntervals = NumberOfTimeIntervals;
StaticData.VolumeValues = VolumeValues;
StaticData.VolumeMilestones = VolumeMilestones;
StaticData.DivisionVolumeIndex = DivisionVolumeIndex;
StaticData.MeanBudPointIndex = MeanBudPointIndex;
if (Params.ParentSpread == 1)
    StaticData.PPWeight = PPWeight;
end
StaticData.SineValues = SineValues;
StaticData.TenPercentWidth = TenPercentWidth;
StaticData.TenPctStart = TenPctStart;
StaticData.ParentEmergentVolume = NewParentVolume';
for i=1:Params.NumberOfGenerations
    Generations(i) = struct('State', zeros(StaticData.NumberOfTimeIntervals(i,1),1), 'Update',
zeros(StaticData.NumberOfTimeIntervals(i,1),1), 'Promote',
zeros(StaticData.NumberOfTimeIntervals(i,1),1), 'NewDaughters',
zeros(StaticData.NumberOfTimeIntervals(1,1),1)   );
end
if (Params.SaveData == 1)
    SaveParams(Params);
    SaveStaticData(StaticData, Params);
end
```

```
    for j=1:Params.NumberOfGenerations
        XXa=1:1:StaticData.NumberOfTimeIntervals(j);
        pdfNormaldist=round(normpdf(XXa,StaticData.MeanBudPointIndex(j),100)*10000/(2^j));
        for i=1:StaticData.NumberOfTimeIntervals(j)
        Generations(j).State(i,1)=pdfNormaldist(i);
        end
    end
    SumALL=0;
    for j=1:Params.NumberOfGenerations
      SumALL = SumALL+sum(Generations(j).State);
    end
    for j=1:Params.NumberOfGenerations
      Generations(j).State = Generations(j).State/SumALL;
    end
clear GrowthRates VolumeIntervals NumberOfTimeIntervals VolumeValues VolumeMilestones
DivisionVolumeIndex MeanBudPointIndex PPWeight SizeGrowthRates SizeVolumeIntervals
SizeVolumeMilestones SineValues SineValues_GrowthInterval SineValues2 MidPoint SumDist
InitialSigma NewParentVolume i;
disp(strcat('Job Name -- ', blanks(4), Params.JobName))
tic;
Die=zeros(1,1);
GraveYard = zeros(StaticData.NumberOfTimeIntervals(Params.NumberOfGenerations,1),1);
PBNJ=zeros(1,Params.NumberOfTimeSteps);
PBNJ_cdf=zeros(1,Params.NumberOfTimeSteps);
if (Params.CalcGeneExpression == 1)
    GeneExpression = zeros(Params.NumberOfTimeSteps, 1);
end
M = 1;
N = 1;
if (Params.ShowPopulationPlots == 1)
    hFigGens = figure('Name', 'Generations Evolving');
    if (Params.NumberOfGenerations <= 6)
        M = 2;
        N = 3;
    else
        Factors = factor(Params.NumberOfGenerations);
        HowManyFactors = size(Factors,2);
        switch HowManyFactors
            case 2
                M = Factors(1); N = Factors(2);
            case 3
                M = Factors(1)*Factors(2); N = Factors(3);
            case 4
                M = Factors(1)*Factors(2); N = Factors(3)*Factors(4);
        end
    end
    clear Factors HowManyFactors
end
TotalPopulationTimeSeries = zeros(Params.NumberOfTimeSteps, 2);
O2Consumption = zeros(Params.NumberOfTimeSteps,1);
O2Concentration = zeros(Params.NumberOfTimeSteps,1);
O2Concentration(1)=8;
DaughterPopulationTimeSeries = zeros(Params.NumberOfTimeSteps, 2);
ParentPopulationTimeSeries = zeros(Params.NumberOfTimeSteps, 2);
SingleElementTimeSeries = zeros(Params.NumberOfTimeSteps, 2);
SingleElementOrigValue = Generations(1).State(100,1);
P0_in_timeseries = zeros(Params.NumberOfTimeSteps, 2);
P0_out_timeseries = zeros(Params.NumberOfTimeSteps, 2);
```

```
PK_in_timeseries = zeros(Params.NumberOfTimeSteps, Params.NumberOfGenerations);
PK_out_timeseries = zeros(Params.NumberOfTimeSteps, Params.NumberOfGenerations);
GridPopulation_timeseries = zeros(Params.NumberOfTimeSteps, Params.NumberOfGenerations);
PopGrowthTimeSeries=zeros(Params.NumberOfTimeSteps,Params.NumberOfGenerations+1);
recordstart=0;
DensityalongVolume_timeseries=zeros(max(StaticData.NumberOfTimeIntervals),Params.NumberOfGener
ations,Params.NumberOfTimeSteps);
PreTotal=ComputeTotalPopulation(Generations, Params);
CycleCount=0;
resetcount=0;
for i=1:Params.NumberOfTimeSteps
   if (mod(i,50)==0)
     fprintf('%d -th timesteps\n', i)
   end
   TotalPopulationTimeSeries(i,1) = i;
   TotalPopulationTimeSeries(i,2) = ComputeTotalPopulation(Generations, Params);
   DaughterPopulationTimeSeries(i,1) = i;
   DaughterPopulationTimeSeries(i,2) = sum(Generations(1).State);
   ParentPopulationTimeSeries(i,1) = i;
   ParentPopulationTimeSeries(i,2) = ComputeParentPopulation(Generations, Params);
   if(Params.Normalize==0)
     POPTimeSeries(i,1)= TotalPopulationTimeSeries(i,2);
     for j=2:Params.NumberOfGenerations+1
       POPTimeSeries(i,j)=sum(Generations(j-1).State);
     end
   end
  for j=1:Params.NumberOfGenerations
      GridPopulation_timeseries(i,j) = sum(Generations(j).State);
   end
   PBNJ(:,i)=FractionBudded(Generations,StaticData,Params);
   PBNJ_cdf(:,i)=FractionBuddedWithCDF(Generations,StaticData,Params);
     H2SAmount = 0.0;
   for j=1:Params.NumberOfGenerations
     SPhaseStart = StaticData.MeanBudPointIndex(j,1);
     SPhaseEnd = StaticData.MeanBudPointIndex(j,1) + Params.CellSPhaseWidth;
     SPhaseEnd = round((1/2 + 15/31)*StaticData.NumberOfTimeIntervals(j,1));
     H2SAmount = H2SAmount + sum(Generations(j).State(SPhaseStart:SPhaseEnd));
   end
   SFAmount = 0.0;
   for j=1:Params.NumberOfGenerations
     SFAmount = SFAmount + sum(Generations(j).State(1:StaticData.MeanBudPointIndex(j,1)));
   end
   spankingORblocking=0;
   if ((Params.H2SEntrainment == 1)&&(spankingORblocking==1)&&(H2SAmount >=
Params.H2SThreshold))
      [Generations] = H2SCellPhaseEntrainment(Generations, StaticData, Params);
   end
   if ((Params.H2SEntrainment == 1)&&(spankingORblocking==0)&&(H2SAmount >=
Params.H2SThreshold)&&(i>7000))
     for j=1:(Params.NumberOfGenerations - 1)
       if (Params.DivideWithCDF == 1)
          [Generations(j).Update, Generations(j).NewDaughters, Generations(j+1).Promote] =
Growth_With_CDF_blocking(j, Generations(j).State, Params, StaticData);
       else
          [Generations(j).Update, Generations(j).NewDaughters, Generations(j+1).Promote] =
growth_blocking(j, Generations(j).State, Params, StaticData);
       end
     end
```

```
    else
      for j=1:(Params.NumberOfGenerations - 1)
        if (Params.DivideWithCDF == 1)
          [Generations(j).Update, Generations(j).NewDaughters, Generations(j+1).Promote] =
Growth_With_CDF(j, Generations(j).State, Params, StaticData)
else
          [Generations(j).Update, Generations(j).NewDaughters, Generations(j+1).Promote] = growth(j,
Generations(j).State, Params, StaticData);          %%%%%%%% 'growth.m' %%%%%%%%
        end
      end
    end
    if (Params.DivideWithCDF == 1)
       [Generations(Params.NumberOfGenerations).Update,
Generations(Params.NumberOfGenerations).NewDaughters, GraveYard] =
Growth_With_CDF(Params.NumberOfGenerations, Generations(Params.NumberOfGenerations).State,
Params, StaticData);
    else
       [Generations(Params.NumberOfGenerations).Update,
Generations(Params.NumberOfGenerations).NewDaughters, GraveYard] =
growth(Params.NumberOfGenerations, Generations(Params.NumberOfGenerations).State, Params,
StaticData);
    end
       Generations(1).State = Generations(1).Update;
    for j=1:Params.NumberOfGenerations
       Generations(1).State = Generations(1).State + Generations(j).NewDaughters;
    end
    for j=2:Params.NumberOfGenerations
       Generations(j).State = Generations(j).Update + Generations(j).Promote;
    end
    for j=1:Params.NumberOfGenerations
      for k=1:StaticData.NumberOfTimeIntervals(j,1)
       DensityalongVolume_timeseries(k,j,i)=Generations(j).State(k);
      end
    end
       PreTotal=TotalPopulationTimeSeries(i,2);
       if (Params.FilterBySize == 1)
          if (Params.DelayFilter == 1)
             if (i > Params.FilterDelayTime)
                for j=1:Params.NumberOfGenerations
                  [Generations(j).State]= Filter(j, Generations(j).State, Params, StaticData);
                end
             end
          else
             for j=1:Params.NumberOfGenerations
                Generations(j).State = Filter(j, Generations(j).State, Params, StaticData);
             end
          end
       end
    for j=1:Params.NumberOfGenerations
        for k=1:StaticData.NumberOfTimeIntervals(j)
          if (Generations(j).State(k) < eps)
             Generations(j).State(k) = 0.0;
          end
       end
    end
    TotalPopulation = ComputeTotalPopulation(Generations, Params);
    [Ti,ADi] = IntegrateDensities(Generations, Params);
    if (i>7000)
```

```
      for j=1:Params.NumberOfGenerations
     GridFraction=zeros(1,StaticData.NumberOfTimeIntervals(j));
        if (sum(Generations(j).State)~=0)
       for k=1:StaticData.NumberOfTimeIntervals(j)
            GridFraction(k)=(Generations(j).State(k))/sum(Generations(j).State);
            Generations(j).State(k)=Generations(j).State(k)-
TotalPopulation*0.00163*ADi(j)*GridFraction(k);
          end
        end
      end
    end
    if (TotalPopulation < eps)
     fprintf('!!! *** All cells are dead at %d-th timesteps *** !!! \n', i)
      break
    end
    if (Params.Normalize == 1)
            if (TotalPopulation > eps)
          for j=1:Params.NumberOfGenerations
            Generations(j).State = Generations(j).State / TotalPopulation;
          end
       else
          break
       end
    end
    P0_in = 0;
    for rrr=1:Params.NumberOfGenerations
       P0_in = P0_in + sum(Generations(rrr).NewDaughters);
    end
    P0_in_timeseries(i,1) = i;
    P0_in_timeseries(i,2) = P0_in;
    P0_out_timeseries(i,1) = i;
    P0_out_timeseries(i,2) = sum(Generations(2).Promote);
    for rrr=1:(Params.NumberOfGenerations - 1)
       PK_out_timeseries(i,rrr) = sum(Generations(rrr+1).Promote);
    end
    PK_out_timeseries(i,Params.NumberOfGenerations) = sum(GraveYard);
    PK_in_timeseries(i,1) = P0_in;
    for rrr=2:Params.NumberOfGenerations
       PK_in_timeseries(i,rrr) = sum(Generations(rrr).Promote);
    end
    if (Params.CalcGeneExpression == 1)
       G1 = 0.0;
       G2 = 0.0;
       for j=1:Params.NumberOfGenerations
          for k=1:StaticData.NumberOfTimeIntervals(j)
            G1 = G1 + (Generations(j).State(k) * StaticData.VolumeValues(j,k) *
StaticData.SineValues(j,k));
            G2 = G2 + (Generations(j).State(k) * StaticData.VolumeValues(j,k));
          end
       end
       if (TotalPopulation > eps)
          GeneExpression(i,1) = G1/G2;
       else
          GeneExpression(i,1) = 0.0;
       end
    end
    for j=1:Params.NumberOfGenerations
       for k=1:round(StaticData.MeanBudPointIndex(j)/2)
```

```
      O2Consumption(i) = O2Consumption(i)+(1/40)*Generations(j).State(k);
       end
   for k=round(StaticData.MeanBudPointIndex(j)/2)+1:StaticData.MeanBudPointIndex(j)
      O2Consumption(i) = O2Consumption(i)+Generations(j).State(k);
       end
       for k=StaticData.MeanBudPointIndex(j)+round((1/4)*StaticData.NumberOfTimeIntervals(j)-
StaticData.MeanBudPointIndex(j)):StaticData.MeanBudPointIndex(j)+round((1/2)*StaticData.NumberOfT
imeIntervals(j)-StaticData.MeanBudPointIndex(j))
      O2Consumption(i) = O2Consumption(i)+(1/40)*Generations(j).State(k);
       end
   end
   if (i>1)
      O2Concentration(i)=O2Concentration(i-1)+1.0015*(8-O2Concentration(i-1)) - O2Consumption(i-
1)/150 - 0.0015*(7-O2Concentration(i-1));
   end
   if (mod(i,Params.PlotFrequency)==0)
      if (Params.ShowPopulationPlots == 1)
        for j=1:Params.NumberOfGenerations
            subplot(M,N,j);
            kk = j;
plot(StaticData.VolumeValues(kk,1:StaticData.NumberOfTimeIntervals(kk,1)),Generations(kk).State(:,1));
            title(strcat('P', int2str(j)));
            xlabel('Vol');
            yy = [0 max(Generations(kk).State)];
            hLine = line([StaticData.VolumeMilestones(kk,1) StaticData.VolumeMilestones(kk,1)], yy,
'LineWidth',1,'Color',[1 0 0 ]);
            hLine = line([StaticData.VolumeMilestones(kk,2) StaticData.VolumeMilestones(kk,2)], yy,
'LineWidth',1,'Color',[0 0.75 0 ]);
            if (Params.FilterBySize == 1)
               hLine = line([Params.FilterMinSize Params.FilterMinSize], yy, 'LineWidth',1,'Color','black');
               hLine = line([Params.FilterMaxSize Params.FilterMaxSize], yy, 'LineWidth',1,'Color','black');
            end
        end
         if (Params.SaveData == 1)
            SaveFigureByIndex(hFigGens, i, Params);
         end
      end
   end
    if ((PBNJ(i)>=0.8)&(resetcount==0))
      CycleCount=CycleCount+1;
      resetcount=1;
   end
   if ((resetcount==1)&(PBNJ(i)<0.8))
      resetcount=0;
   end
end
fid=fopen('NumberOfCycles_alongFilterWvolumes.dat','a+');
fprintf(fid,'%d %d %d \n',Params.FilterMinSize, Params.FilterMaxSize, CycleCount);
fclose(fid);
fidc=fopen('NumberOfCycles_alongFilterWOvolumes.dat','a+');
fprintf(fidc,'%d \n',CycleCount);
fclose(fidc);
figure
plot(1:Params.NumberOfTimeSteps,PBNJ);
hold on;
line([1 Params.NumberOfTimeSteps],[0.2 0.2],'Color','r');
hold on;
line([1 Params.NumberOfTimeSteps],[0.8 0.8],'Color','r');
```

```
if (Params.CalcGeneExpression == 1)
    hFigGene = figure('Name', strcat('Gene Expression Plot -- ', Params.JobName));
    plot(GeneExpression);
    title(strcat('Gene Expression -- ', Params.JobName));
    xlabel('Time (minutes)');
    if (Params.SaveData == 1)
        SaveFigureByName(hFigGene, 'GeneExpression', Params)
    end
end
if (Params.SaveData == 1)
    SaveDynamicData(Generations, PBNJ, PBNJ_cdf, Params)
    SaveStateSnapshotWithVolume(Generations, i, StaticData, Params)
    if (Params.CalcGeneExpression == 1)
      SaveGeneExpression(GeneExpression, Params);
    end
end
if (Params.ShowPopulationPlots == 1)
    close(hFigGens)
end
if (Params.CalcGeneExpression == 1)
    close(hFigGene)
end
dlmwrite(strcat(Params.JobName,'/ts_total_population.dat'), TotalPopulationTimeSeries, 'delimiter', '\t')
dlmwrite(strcat(Params.JobName,'/O2consumption.dat'), O2Consumption, 'delimiter', '\t')
dlmwrite(strcat(Params.JobName,'/O2concentration.dat'), O2Concentration, 'delimiter', '\t')
dlmwrite(strcat(Params.JobName,'/ts_daughter_population.dat'), DaughterPopulationTimeSeries,
'delimiter', '\t')
dlmwrite(strcat(Params.JobName,'/ts_single_element.dat'), SingleElementTimeSeries, 'delimiter', '\t')
dlmwrite(strcat(Params.JobName,'/ts_P0_rate_in.dat'), P0_in_timeseries, 'delimiter', '\t')
dlmwrite(strcat(Params.JobName,'/ts_P0_rate_out.dat'), P0_out_timeseries, 'delimiter', '\t')
dlmwrite(strcat(Params.JobName,'/ts_PK_rate_out.dat'), PK_out_timeseries, 'delimiter', '\t')
dlmwrite(strcat(Params.JobName,'/ts_PK_rate_in.dat'), PK_in_timeseries, 'delimiter', '\t')
dlmwrite(strcat(Params.JobName,'/ts_grids_population.dat'), GridPopulation_timeseries, 'delimiter', '\t')
dlmwrite(strcat(Params.JobName,'/ts_grids_DensityalongVolume.dat'), DensityalongVolume_timeseries,
'delimiter', '\t')
if (Params.Normalize==0)
  dlmwrite(strcat(Params.JobName,'/ts_grids_population_nonNOR.dat'), POPTimeSeries, 'delimiter', '\t')
end
DeltaP = PK_in_timeseries - PK_out_timeseries;
dlmwrite(strcat(Params.JobName,'/ts_deltaPK_in-out.dat'), DeltaP, 'delimiter', '\t')
[T,P] = IntegrateDensities(Generations, Params);
dlmwrite(strcat(Params.JobName,'/Age_distribution.dat'), P)
fid=fopen('AD_alongTauD_H2Son.dat','a+');
fprintf(fid,'Daughter BudEmergencyVolume %f\n', StaticData.VolumeMilestones(1,1));
for i=1:Params.NumberOfGenerations
  fprintf(fid,' %f\n', P(i,1));
end
fclose(fid);
if ((Params.Normalize==0)&(Params.NumberOfTimeSteps>4000))
  DoublingTime=zeros(1,Params.NumberOfGenerations+1);
  alphaP=zeros(1,Params.NumberOfGenerations+1);
  for i=1:Params.NumberOfGenerations+1
     LinearFitP=polyfit(4000:size(POPTimeSeries,1),
log(POPTimeSeries(4000:size(POPTimeSeries,1),i)'),1); %'
     alphaP(i)=LinearFitP(1,1);
     DoublingTime(i)=log(2)/alphaP(i);
  end
  dlmwrite(strcat(Params.JobName,'/POPgrowthrates.dat'), alphaP, 'delimiter', '\t')
```

```
  dlmwrite(strcat(Params.JobName,'/POPdoublingtimes.dat'), DoublingTime, 'delimiter', '\t')
end
if ((Params.H2SEntrainment == 1)&&(spankingORblocking==1))
   amplitude=max(PBNJ(Params.NumberOfTimeSteps-500:Params.NumberOfTimeSteps))-
min(PBNJ(Params.NumberOfTimeSteps-500:Params.NumberOfTimeSteps));
   fid=fopen('amplitudes_spank_10per_S_equilini.dat','a+');
   fprintf(fid,'%f %f\n', Params.H2SThreshold, amplitude);
   fclose(fid);
end
if ((Params.H2SEntrainment == 1)&&(spankingORblocking==0))
   amplitude=max(PBNJ(Params.NumberOfTimeSteps-500:Params.NumberOfTimeSteps))-
min(PBNJ(Params.NumberOfTimeSteps-500:Params.NumberOfTimeSteps));
   fid=fopen('amplitudes_block_10per_S_equilini.dat','a+');
   fprintf(fid,'%f %f\n', Params.H2SThreshold, amplitude);
   fclose(fid);
end
recordcount=zeros(Params.NumberOfGenerations,1);
for j=1:Params.NumberOfGenerations
   for i=2400:Params.NumberOfTimeSteps
      if ((DensityalongVolume_timeseries(1,j,i)==0))
         pre=0;
      for k=1:StaticData.NumberOfTimeIntervals(j,1)
        if ((pre==0)&&(DensityalongVolume_timeseries(k,j,i)~=0))
           recordcount(j,1)=recordcount(j,1)+1;
              pre=DensityalongVolume_timeseries(k,j,i);
        elseif ((pre~=0)&&(DensityalongVolume_timeseries(k,j,i)==0))
           pre=0;
           end
         end
       break;
       end
   end
end
if ((Params.H2SEntrainment == 1)&&(spankingORblocking==0))
   fidc=fopen('NumberOfClumps_block_10per_0p3th_S_equilini.dat','a+');
   fprintf(fidc,'R = 10percent  and  S-phase width = %d\n', Params.CellSPhaseWidth);
   for j=1:Params.NumberOfGenerations
      fprintf(fidc,' %d\n', recordcount(j,1));
   end
   fclose(fidc);
end
if ((Params.H2SEntrainment == 1)&&(spankingORblocking==1))
   fidc=fopen('NumberOfClumps_spank_10per_0p3th_S_equilini.dat','a+');
   fprintf(fidc,'R = 10percent  and  S-phase width = %d\n', Params.CellSPhaseWidth);
   for j=1:Params.NumberOfGenerations
      fprintf(fidc,' %d\n', recordcount(j,1));
   end
   fclose(fidc);
end
for i=1:Params.NumberOfTimeSteps
  inTOT(i)=sum(PK_in_timeseries(i,:));
  outTOT(i)=sum(PK_out_timeseries(i,:));
end
tau = zeros(Params.NumberOfGenerations+1, 1);
if (Params.NumberOfTimeSteps>3000)
    FITin=polyfit(Params.NumberOfTimeSteps-
3000:Params.NumberOfTimeSteps,log(inTOT(Params.NumberOfTimeSteps-
3000:Params.NumberOfTimeSteps)),1);
```

```
   FITout=polyfit(Params.NumberOfTimeSteps-
3000:Params.NumberOfTimeSteps,log(outTOT(Params.NumberOfTimeSteps-
3000:Params.NumberOfTimeSteps)),1);
   tau(Params.NumberOfGenerations+1)=(FITin(2)-FITout(2))/FITout(1);
   for i=1:Params.NumberOfGenerations-1
      FITin=polyfit(Params.NumberOfTimeSteps-
2000:Params.NumberOfTimeSteps,log(PK_in_timeseries(Params.NumberOfTimeSteps-
2000:Params.NumberOfTimeSteps,i)'),1);
      FITout=polyfit(Params.NumberOfTimeSteps-
2000:Params.NumberOfTimeSteps,log(PK_out_timeseries(Params.NumberOfTimeSteps-
2000:Params.NumberOfTimeSteps,i)'),1);
      tau(i)=(FITin(2)-FITout(2))/FITout(1);
   end
   tau(Params.NumberOfGenerations)=tau(Params.NumberOfGenerations-1);
   dlmwrite(strcat(Params.JobName,'/taus.dat'), tau, 'delimiter', '\t')
end
toc;
```

**NOTE: Sub-routine code not included.*

APPENDIX B

LUMINESCENCE REPORTER MATERIALS AND METHODS

The plasmid pGL3-*ADH1*term was generated by replacing the SV40 late terminator of pGL3-Basic (Promega) with the *S. cerevisiae ADH1* terminator from pFA6a-GFP(S65T)-HIS3MX [2]. The *ADH1* terminator was PCR amplified from pFA6a-GFP(S65T)-HIS3MX using the forward primer TTCATCTCTAGAGGCGCGCCACTTCTAAAT that contained an *Xba*I site overhang and the reverse primer GGACGAGGCAAGCTAAAC that annealed shortly downstream from the terminator and a *Bgl*II site. The PCR product was digested with *Xba*I and *Bgl*II and ligated into pGL3-Basic which had been digested with *Xba*I and *BamH*I to remove its SV40 late terminator.

The PEST sequence from *CLN2* of *S. cerevisiae* was amplified by PCR from genomic DNA of strain SEY6210 [3] using the forward primer GAATAAGCTTGCATCCAACTTGAACATT that contained a *Hind*III site overhang and the reverse primer GAAGTTCTAGACTATATTACTTGGGTATTGCC that contained an *Xba*I site overhang. These sites were used to clone the PEST PCR product into pFA6a-GFP(S65T)-HIS3MX for maintenance (designated pFA6a-*CLN2*PEST).

Gene SOEing [1,5] was used to fuse the PEST sequence from pFA6a-CLN2PEST to the 3' end of the luciferase ORF of pGL3-*ADH1*term just prior to the stop codon. Gene SOEing was performed as described by Wersh et al [5]. Specifically, the luficerase component of

the fusion was generated from pGL3-*ADH1*term using the forward primer GCTAGCCCGGGCTCGAGATC and the reverse primer AATGTTCAAGTTGGATGCCACGGCGATCTTTCC. The forward primer annealed upstream from the luciferase start codon and included an *Xho*I site. The reverse primer included an overhang with homology to the 5' end of the PEST sequence. The PEST portion of the fusion was generated from pFA6a-*CLN2*PEST using the forward primer GGAAAGATCGCCGTGGCATCCAACTTGAACATTTCG and the same reverse primer that was used to amplify the PEST sequence from the genomic source mentioned above. The forward primer contained an overhang with homology to the 3' end of luciferase, and the reverse primer contained an *Xba*I site overhang. The two PCR products were gel-purified (Qiagen), mixed in equimolar (10 nM) concentrations, and were fused using 10 rounds of PCR without primers. The Luc-PEST fusion was amplified by a final PCR reaction involving flanking primers (i.e. the *Xho*I-containing forward primer of the previous luciferase-generating PCR and the *Xba*I-containing reverse primer of the previous PEST-generating PCR) and cloned into pGL3-*ADH1*term using *Xho*I and *Xba*I, replacing the original luciferase ORF to create the plasmid pGL3-PEST-*ADH*1term.

The region of DNA containing the luciferase-PEST fusion and the *ADH1* terminator was amplified by PCR from pGL3-PEST-*ADH*1term using the forward primer AAGTAACTGCAGATGGAAGACGTCAAAAACATAAAGAAAGGCCCG and the reverse primer GACGATAGTCATGCCCGGG. The forward primer annealed at the beginning of the start codon of luciferase and included a mismatch that mutated the fourth amino acid from adenine to valine introducing an *Aat*II site there. The forward

primer also included a *Pst*I overhang. The reverse primer annealed downstream of *ADH1*term and a *Sal*I site. This PCR product was cloned into the yeast shuttle vector pRS315[4] using *Pst*I and *Sal*I to create the versatile destabilized-luciferase expression vector pRS315-Luc(A4V)PEST for *S. cerevisiae*. A similar luciferase expression vector (pRS315-Luc(A4V)) without the PEST sequence was constructed in the same way by using pGL3-*ADH1*term as the PCR template instead of pGL3-PEST-*ADH*1term.

Promoters of interest (*POL1*, *GLN3*, and DAL80) were amplified by PCR using yeast genomic template from strain SEY 6210 and primer pairs that targeted nearly all of the intergenic region upstream of the gene of interest. The forward primers contained an overhang that included either the *Pst*I or *Xma*I restriction site. The reverse primers contained an overhang that included an *Aat*II site and the first 4 amino acids of luciferase. Specific primers used to amplify the promoters are listed in Table B.1. The resulting PCR products (promoters) were individually cloned into pRS315-Luc(A4V)PEST and/or pRS315-Luc(A4V) to make complete luciferase reporter constructs that could be maintained in yeast. The "promoter-Luc(PEST)-terminator" portion of the constructs were also moved to other pRS expression vectors (e.g. pRS314, pRS303, pRS306, pRS424) when different selection markers were needed or when the luficerase reporter needed to be stably integrated into the host's genome or over expressed on 2-micron plasmids [4]. These rearrangements were made with *Pst*I or *Xma*I and *Sal*I.

The yeast integration vectors pRS303 and pRS306 were modified for selection on G418 antibiotic by introducing the kanamycin resistance gene from pFA6a-KanMX6 [2]. This was done in a three step process. First, the Kan resistance gene was cloned from pFA6a-

KanMX6 into pRS315-P_{GPH1}-Luc(A4V)PEST using *Not*I and *BamH*I, creating pRS315-P_{GPH1}-Luc(A4V)PEST-Kan. This plasmid served as a backbone for the introduction of several promoters of interest by swapping out the *GPH1* promoter with the one of interest using *Xma*I and *Aat*II. Then the entire region containing the Kan resistance gene, promoter of interest, destabilized luciferase, and the *ADH1* terminator was cloned into either pRS303 or pRS306 using *Not*I and *Sal*I or *EcoR*I and *Sal*I. This pRS315-Kan intermediate was necessary at first due to an inconvenient *Aat*II site in pRS303 and pRS306 that prevented the direct introduction of promoters of interest into them. This was later overcome by removing the native *Aat*II site from pRS303 and pRS306 by digesting the plasmids with AatII, blunting the digestion with Klenow fragment, and ligating the blunt ends back together.

Table B.1: A list of primer sequences used for the development of luminescent constructs. Each primer contains appropriate cloning sites.

Promoter		**Primer Sequence (5' to 3')**
POL1	Forward Primer	AAGTAACTGCAGTGCATTTTTCTTAAAGAAATATAAC
	Reverse Primer	ATGATTGACGTCTTCCATTTTCCACTGTTTATTATATGCCT
GLN3	Forward Primer	AAGTAACCCGGGCAATACGAGCAGCAAAGAAATTG
	Reverse Primer	ATGATTGACGTCTTCCATTTGTTTGTTGGTGGGGGAAAAG
DAL80	Forward Primer	AAGTAACCCGGGCACCCTTGTTTATCTATCCTAC
	Reverse Primer	ATGATTGACGTCTTCCATTCTCTTATATATAATATGATATAATATAATG

References

1. Horton R., Hunt H., Ho S., Pullen J., Pease L. 1989. Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension. *Gene.* **77**: 61-68.

2. Longtine M., McKenzie A., Demarini D., Shah N., Wach A., Brachat A., Philippsen P., and Pringle J. 1998. Additional modules for versatile and economic PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**: 953-961.

3. Robinson J., Klionsky D., Banta L., Emr S. 1988. Protein sorting in *Saccharomyces cerevisiae*: Isolation of mutants defective in the delivery and processing of multiple vacuolar hydrolases. *Molecular and Cellular Biology*. **8**: 4936-4948.

4. Sikorski R. and Hieter P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics*. **122**: 19-27.

5. Wurch T., Lestienne F., and Pauwels P. 1998. A modified overlap extension PCR method to create chimeric genes in the absence of restriction enzymes. *Biotechnology Techniques.* **12**: 653–657.

APPENDIX C

DEVELOPMENT OF ACTIVITY BASED COST FUNCTIONS FOR CELLULASE, INVERTASE, AND OTHER ENZYMES

Abstract

As enzyme chemistry plays an increasingly important role in the chemical industry, cost analysis of these enzymes becomes a necessity. In this paper, we examine aspects that affect the cost of enzymes based upon enzyme activity. The basis for this study stems from a previously developed objective function that quantifies the tradeoffs in enzyme purification via the foam fractionation process [1]. A generalized cost function is developed from our results that could be used to aid in both industrial and lab scale chemical processing. The generalized cost function shows several non-obvious results that could lead to significant savings. Additionally, the parameters involved in the operation and scaling up of enzyme processing could be optimized to minimize costs [1, 2, 3]. We show that there are typically three regimes in the enzyme cost analysis function: the low cost pre-linear region, the moderate cost linear region, and high cost power-law region. The overall form of the cost analysis function appears to robustly fit the power law form.

Introduction

Since the rapid development of the chemical industry in the 1940's, the chemical industry has morphed from a commodity chemical market into a specialty chemical market where bio-related products and processes have become increasingly important. Enzymes are now commonly used as catalysts to produce products such as proteins, sugars, and lipids as well as processing tools to enhance more traditional products, such as paper pulp. In fact, enzyme industry sales are expected to increase to 2.2 billion dollars by 2010 [4]. It has been predicted that the kinetics of cost reduction for industrial enzymes will control the extent of how the enzyme market will grow [5]. Thus, cost analysis of enzymes is becoming an important factor for engineers and scientists dealing with mainstream bio-products.

The price of enzymes is often controlled by the method of production and more importantly the purification method [6]. Many enzymes can be isolated from common plants and microorganisms while others must be harvested from animals, particularly mammals such as pigs, rabbits, and even humans. Generally, enzymes isolated from common plants are less costly than those isolated from animals. Likewise, enzymes that require little purification are less costly. Some enzymes require additional processing beyond traditional standard processing which includes foam fractionation, salting out, and liquid-liquid extraction [7]. These additional processing steps can become extremely costly and can lead to increased market cost of the enzyme. These processing steps often include standard gel chromatography and/or HPLC (High Performance/Pressure Liquid Chromatography).

In this study, we will explore whether an objective function, relating enzyme activity to cost, can be developed to establish the cost function of cellulase and other enzymes. We will determine the cost increase of crude cellulase (in terms of activity per mass) following processing used to enhance the purity and concentration of this protein. We shall start with a more generalized objective function comprised of measurable purification process responses to create our cost model and reduce that model to the specific cost function for this study. A previously developed objective function quantifies the tradeoff between maximizing the enzyme concentration in a separation process such as a foam fractionation process and minimizing the loss of enzyme mass and enzyme activity in that process [1] is shown below.

$$\Phi = (AR)^a (MR)^b (ER)^c \qquad \text{Equation (C.1)}$$

where AR = activity recovery = Afoam/Ai

MR = mass recovery = Mfoam/Mi

ER = enrichment recovery = Cfoam/Ci

Typically, a is positive, b is negative, and c is positive. b and c will be both positive in the case where Φ is a generalized, desired performance factor. On the other hand, as we shall see in the results section, where MR is replaced by purchase mass and Φ is replaced by \$/mg, b becomes negative. Here we explore whether a generalized relationship between cost, activity, and recovered mass alone can characterize a cost function, and if so, determine the appropriate coefficients.

This approach to the development of cost-based processing can result in the lowering of product cost at each processing step by selecting the control variables which can maximize the respective Φ or minimize the cost based functions at each step. In a foam-

fractionation process, in particular, these control parameters are typically the pH and the foaming-gas superficial velocity [1]. In this approach, it is convenient to define Φ as a generalized value (or price) function. In this study, we define price as the price published in leading biochemical catalogs used in this study. Additional processing to generate higher purity enzyme will result in incremental increases in the value of Φ which can be represented as the first derivative of the objective function with respect to activity. In particular, in this study, we shall compare the catalog values to a parameter fitted model for industrial enzymes to determine whether there are general trends and quantitative similarities between classes of enzymes. Our goal is to develop an objective function based on market values, making the model most useful from the purchasing (consumer) point of view. However, the model could also be used from a manufacturing point of view to determine whether the production of certain products would be profitable by determining the economic value of a potentially new product prior to investment in production. In the initial part of this study, we will assume the second and third terms of the objective function model (Equation (C.1)), Φ, are constant and can be lumped into a new parameter γ. This permits the model to be simplified to the following equation, where the activity term is defined as activity per unit mass, as is often expressed in enzyme sales catalogs.

$$\Phi = \gamma(A)^a \qquad \text{Equation (C.2)}$$

With this framework, our underlying assumption for the remainder of the work is that for a given enzyme, the processing cost is captured by the power law framework and, by difference production (fermentation) costs are included within the constant γ. This is a good assumption since the majority of industrial enzymes are produced in a microbial environment or directly harvested from plants. These enzymes are often processed

directly for activity, which requires increasingly larger costs in order to reach higher activity. Enzymes generally used for medical applications, which are produced in a mammalian cell line or from animals, may tend to deviate from this assumption. Additionally, the cost specialty enzymes which are produces in only small quantities for R&D purposes may be dictated by investment cost. We shall focus on industrial enzymes here.

The cost of storage for enzymes can also influence the cost of these chemicals at the customer level. It can be readily established that significant savings (greater than 25%) can be achieved by simply buying in bulk from an enzyme supplier. However, bulk storage costs of these enzymes are not negligible considering that most enzymes must be kept at -20°C. The additional cost of storage resulting from bulk purchases needs to be considered by engineers and scientists in industry. Thus, we present a slightly modified cost function, Φ, that accounts for the cost effects of buying in bulk quantity and the resulting enzyme storage.

$$\Phi = \gamma (A)^a \delta (1/M)^b (1+\alpha * t) \qquad \text{Equation (C.3)}$$

where Φ= enzyme cost in \$/mg

M=mass purchase amount

t=storage time in weeks

a and b are dimensionless constants

γ, δ, and α have appropriate dimensions

Two key observations can be made regarding this improved cost function. First, the cost per unit mass is directly proportional to activity and inversely proportional to the amount of mass purchased. This observation is fairly trivial knowing that if a customer wants a

higher activity per unit mass product, it will cost more and the more purchased will warrant a decrease in price per unit. Secondly, the parameter b must be negative and range from [-1,0] knowing that the price per unit should decrease as more product is purchased. The larger the absolute value of b, the more discount a consumer will receive for buying in bulk. Thus, the parameter b and the storage cost function (third term) must be balanced to attain the most cost effective purchase amount for a given purchase. α has the units of 1/weeks (inverse time) and the parameters γ and δ have units dependent on the values of the coefficients a and b and the units activity and mass is expressed, respectively. Of course, activity and the mass purchase amounts could be expressed in dimensionless form to generalize Equation (C.3). It is also important to notice that the parameters γ and δ can be lumped into one single parameter. However, for our purposes it is easier to now look at the two parameters separately so that we can investigate the proposed model terms individually. The first term can be investigated as an independent activity dependent cost function, while the second term can be investigated to determine the effects of mass purchase amount on the cost function. This framework will allow us to empirically determine whether the simple models shown in Equation (C.2) and Equation (C.3) can accurately relate enzyme activity and purchase amount to cost for industrial enzymes.

Materials and Methods

The enzyme market was investigated for the following enzymes: cellulase, invertase, collagenase, papain, alpha-amylase, and elastase. The market was surveyed by determining the cost per mg of enzyme from Fisher Scientific, Sigma-Aldrich, Carolina Biochemical, and Worthington Biochemical, and Elastin Products Company [8, 9, 10, 11, 12]. If these suppliers sold a particular enzyme in different allotment sizes, the cost per mg was averaged over all allotments to determine an average price per mg from that particular supplier. This averaging was introduced here to reduce the market scattering of the data in order to enable us to more clearly observe the trend of enzyme activity on cost. Data was also collected for invertase from an available industrial market report [13]. Only enzymes that were sold in allotments in terms of mass were considered (those sold in terms of enzyme units were ignored). The activity per mg of each product was recorded along with the respective cost per mg. When the supplier gave an activity range, the lower limit of the range was selected to be the finite activity used in this study. Only enzymes that had reported activities with similar units were compared. A complete list of the data collected is shown in Table C.1. The activity units between different enzymes varied because the enzyme activity was generally expressed as the amount of substrate utilized per unit time, however, the amount of substrate and time often varied between enzymes. A complete list of activities for enzymes used in this study is also shown in Table C.1.

The data for price per mg for each enzyme was then plotted versus the corresponding activity per mg, and subsequent least-squares regression analysis was then performed.

When more data were available for the lower activity purity range, the high purity data was weighted such that the number of high purity and low purity data points had equivalent power in the regressions, meaning each data set consisted of the same number of data points in the upper and lower purity range. If this normalization is not performed, the low purity data points will not allow the regression function to capture the dynamics associated with the high purity enzyme. Since data cannot be homogenously sampled over the entire activity domain, some form of normalization is necessary in order to fit the data over the entire activity domain. Performing a regression on the raw data, simply provides an excellent fit to the more numerous low range activity data points, but completely misses the scarce high activity range data. For the alpha amylase data, the regression technique was modified to compensate for an excessive number of data points available in the low purity regime. These data points were so low in the activity domain that they were not allowing the regression function to capture the dynamics associated with the high purity enzyme. In other words, the numerous data points for amylase in the low activity range were controlling the dynamics of the regression in the high activity range such that the regression did not fit the moderate to high activity data points. Thus, the three lowest data points were not included in the regression to provide the best qualitative fit to the data over the entire activity domain. All regression analysis was performed with Microsoft Excel Version XP using the built in least squares algorithm. Several regressions were attempted including power law, exponential, polynomial, and linear fits. Power law regressions seemed to fit the best across all enzymes. This regression expression also passes through point (0,0).A similar algorithm was used to determine the enzyme cost as a function of purchase amount. The data for pricing and

mass were normalized to cost per mg and inverse mass in order to suit the regression type limitations of Microsoft Excel.

Table C.1: A list of raw data collected from available market reports, vendors, and industrial sources.

Enzyme	Activity (U/mg)	Price ($/mg)	Supplier	Activity Unit Definition
Invertase	20	0.00005	13	1 Unit hydrolyzes 1 μmole of saccharose per min at pH=4.65 and T= 25°C
	100	0.01018	10	
	200	0.05700	10	
	300	0.50600	10	
Papain	1.5	0.00072	10	1 Unit hydrolyzes 1 μmole of N-benzoyl-L-arganine-ethyl ester per min at pH=6.2 and T= 25°C
	3	0.00097	10	
	12	0.64983	10	
Elastase	3	1.04333	11	1 Unit cleaves 1 μmole of N-succinly-L-alanyl-L-alanyl-L-analine-p-nitanilide per min at pH=8.0 and T= 25°C
	8	4.00000	11	
	18	1910.0	12	
	18	1928.7	9	
Alpha Amylase	1.5	0.00023	10	1 Unit liberates 1 mg of maltose from starch in 3 min at pH=6.9 and T= 20°C
	20	0.09240	10	
	30	0.00250	10	
	150	0.41700	10	
	380	0.16800	10	
	500	8.28000	10	
	1000	97.200	10	
Cellulase	1	0.00223	10	1 Unit releases .01 mg of glucose per hour from microcrystalline cellulose at pH=5.0 and T= 37°C
	6	0.00500	9	
	25	0.02650	11	
	45	0.10800	11	
	50	0.91000	8	

Results

We found that there appears to be a generalized model for enzyme cost per mg in terms of enzyme activity. From the analyzed data, this simple two parameter model follows a power law trend as depicted in the previously described Equation (C.2). Here it is found that a is approximately 3.7 +/- 0.6 and tends to vary only slightly from enzyme to enzyme. The parameter a describes the separation (purification) cost of these enzymes, since an increase in activity results in a power law growth in the cost function (the rate at which is controlled by the value of the parameter a). γ is dependent on the method of enzyme production and can vary greatly between enzymes, as shown in Table C.2. Indeed, higher production cost, such as mammalian cell culture, will result in a larger value for γ, whereas lower production cost, such as observed in enzymes isolated from plants, generally results in a lower value for γ. This can be seen in Table C.2 below, which shows elastase, which was isolated from a pig pancreas has a γ value of .005, whereas cellulase, isolated from a fungus has a γ value of 3.00E-08. The generalized behavior of the enzymes quantified by γ and a is depicted in Figures C.1-C.5.

Table C.2: A list of parameter values and regression R^2 values for the simplified two parameter price per mass versus activity per mg objective function model given in Equation (C.2). The data used to compute these parameters and generate the subsequent plots were collected from references 6-10.

Parameter Values			
	A	**γ**	**R^2**
Papain	3.48	8.00E-05	0.92
Elastase	4.32	0.005	0.95
Invertase	3.29	3.00E-09	0.99
Cellulase	4.18	3.00E-08	0.77
Alpha Amylase	3.45	3.00E-09	0.81

The equation relating the bulk price per mg is:

$$\Phi = \delta (M)^b \qquad \text{Equation (C.4)}$$

where b=-0.2 and δ= 2.049 per $mg^{1.2}$ for papain (Figure not shown).

The negative parameter value for b depicts the savings achieved by purchasing in bulk. Significant savings can be achieved when buying bulk enzymes; but this gain may be offset by storage costs. We have not quantified the tradeoffs associated with buying bulk and storage requirements , in Equation (C.3). We expect that the parameters for capital costs of storage will vary greatly from facility to facility.

Figure C.6 further elaborates on the enzyme cost relationship by segregating the enzyme activity domain into three regimes characterized by the method of separation.

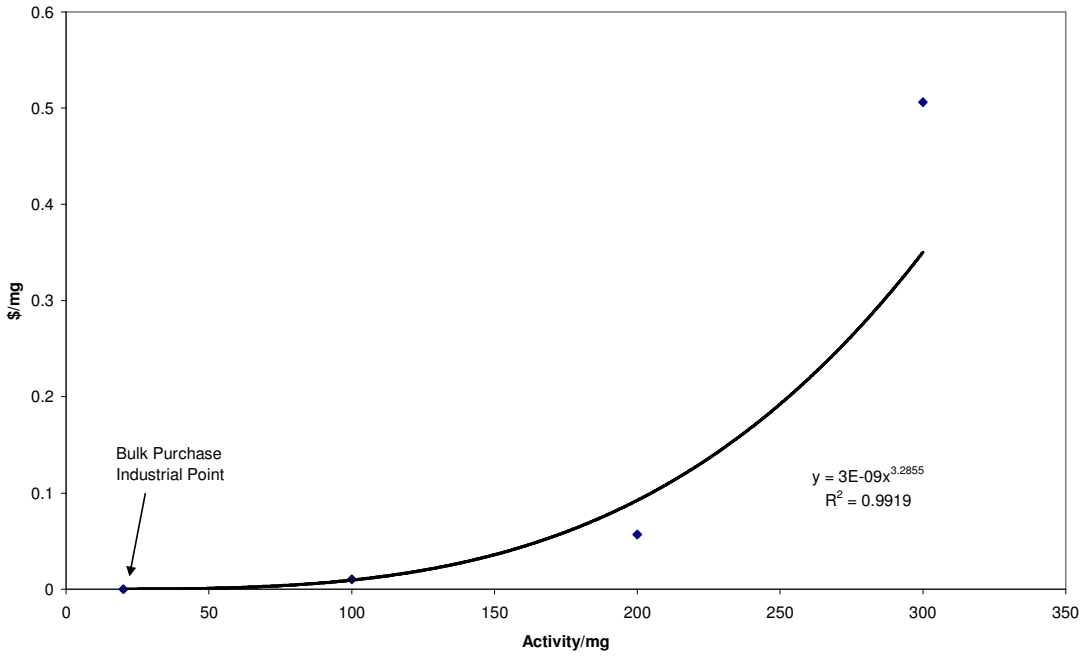


Figure C.1: Least squares regression fit of invertase data provided in Table C.1 to Equation (C.2). The bulk industrial point denoted on the plot was obtained from reference 13.

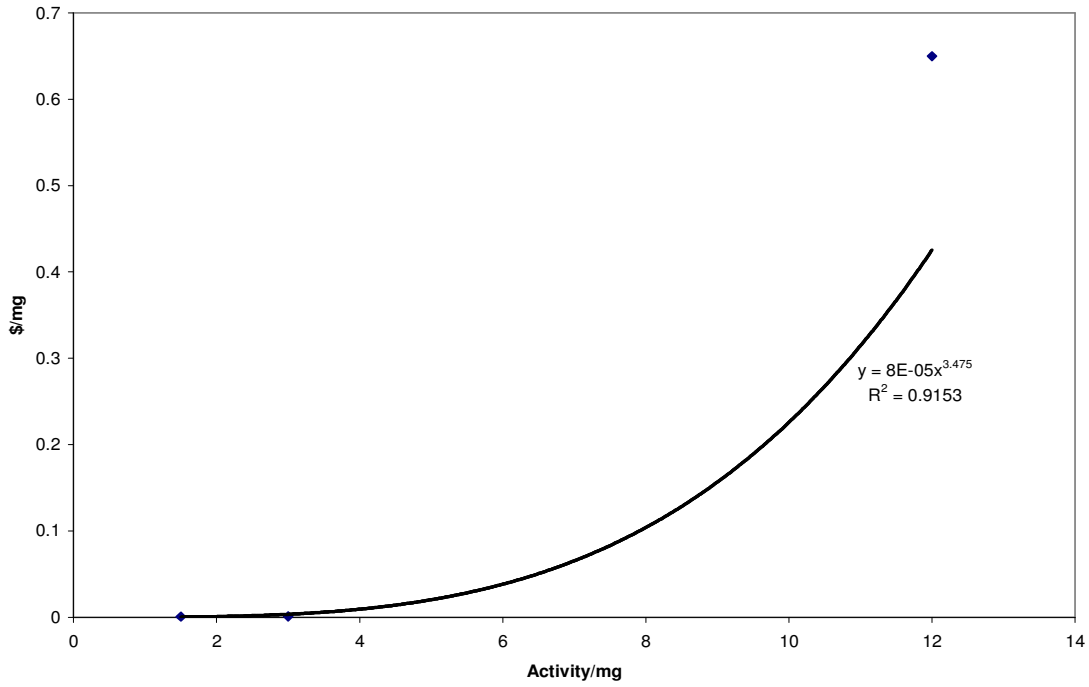


Figure C.2: Least squares regression fit of papain data provided in Table C.1 to Equation (C.2).

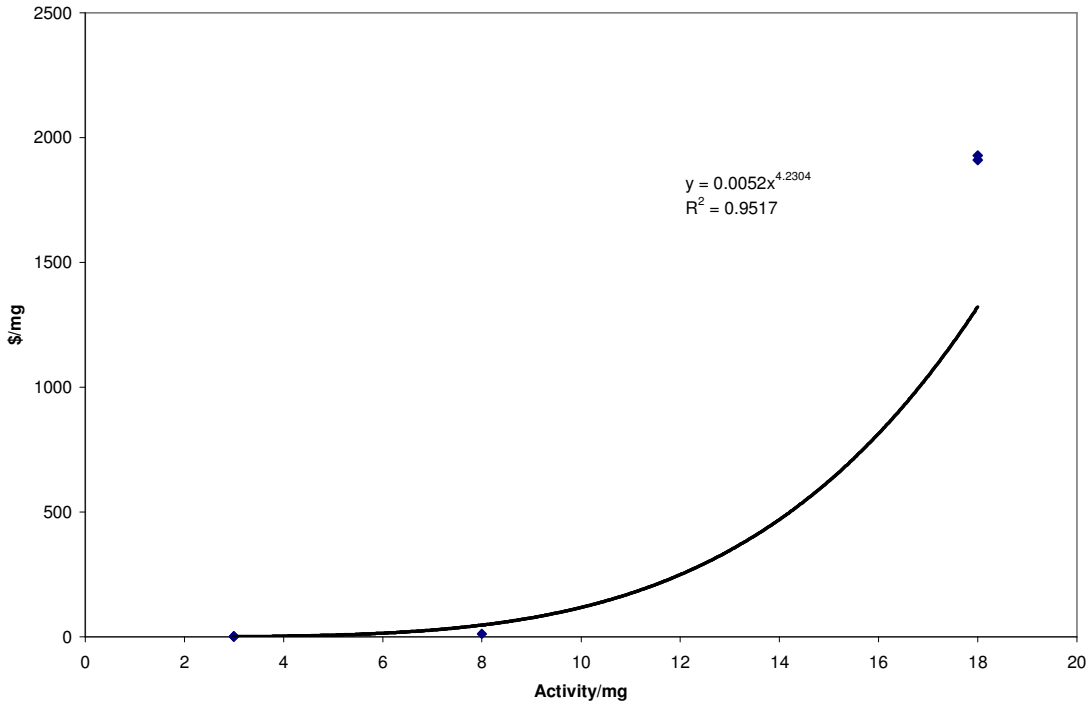


Figure C.3: Least squares regression fit of elastase data provided in Table C.1 to Equation (C.2). The bulk industrial point denoted on the plot was obtained from reference 13.

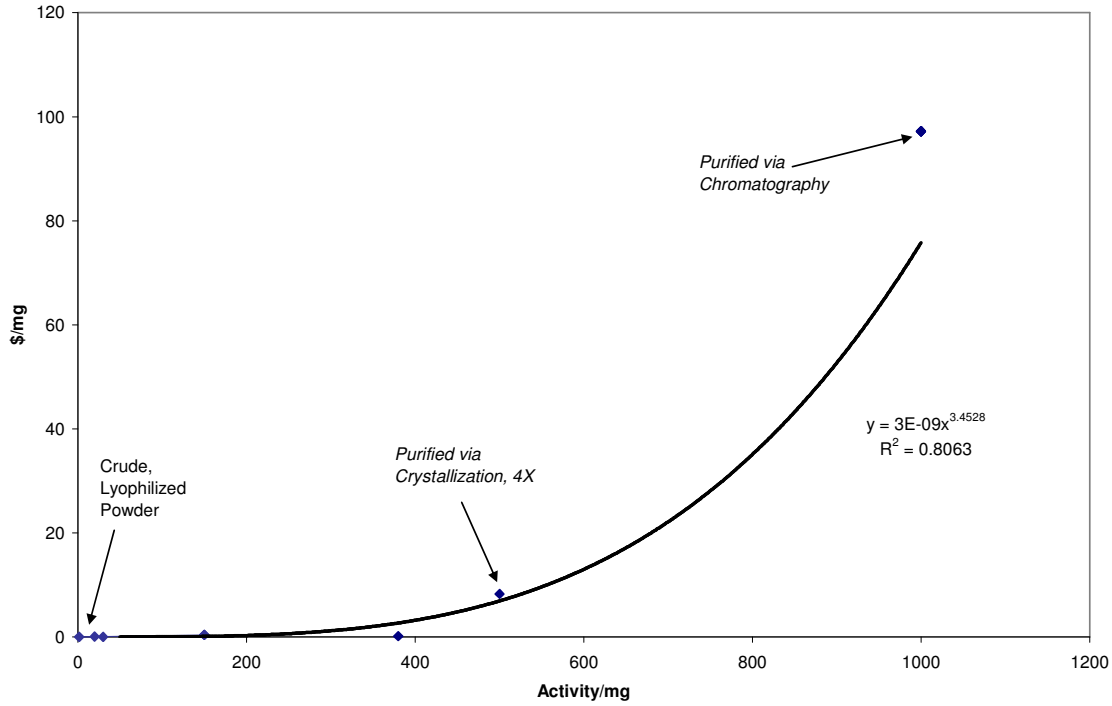


Figure C.4: Least squares regression fit of alpha amylase data provided in Table C.1 to Equation (C.2). The most costly purification method utilized in the processing of each data point is indicated where available.

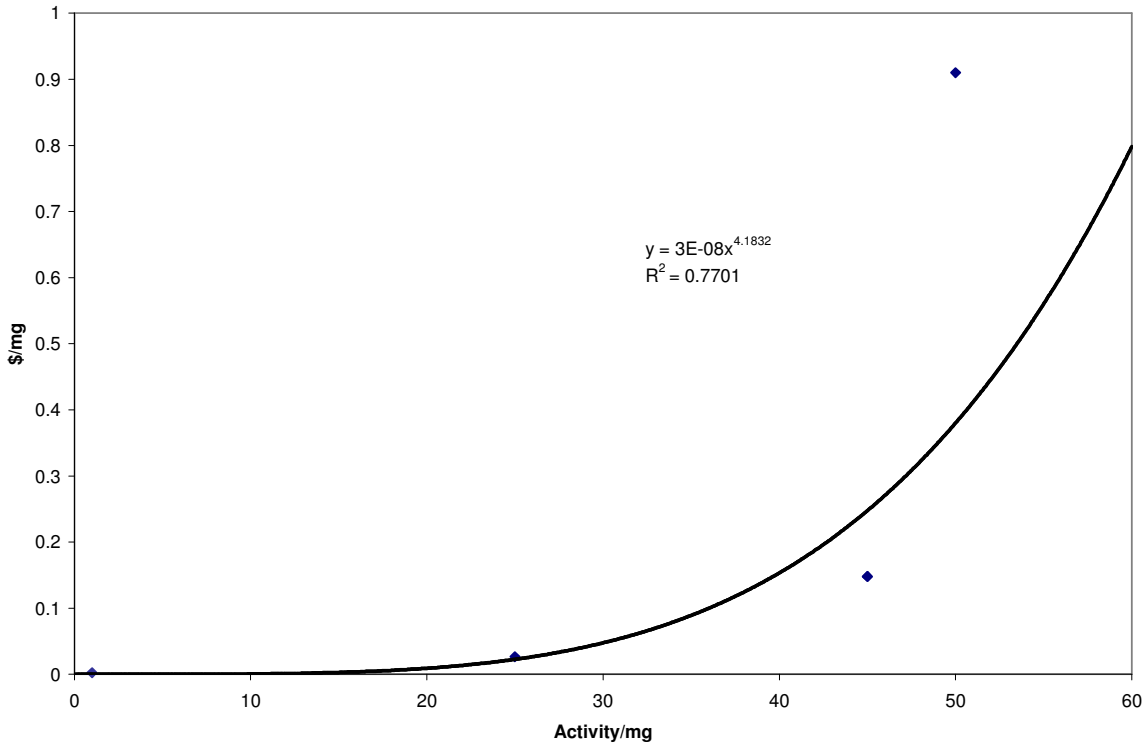


Figure C.5: Least squares regression fit of cellulase data provided in Table C.1 to Equation (C.2). The bulk industrial point denoted on the plot was obtained from reference 13.

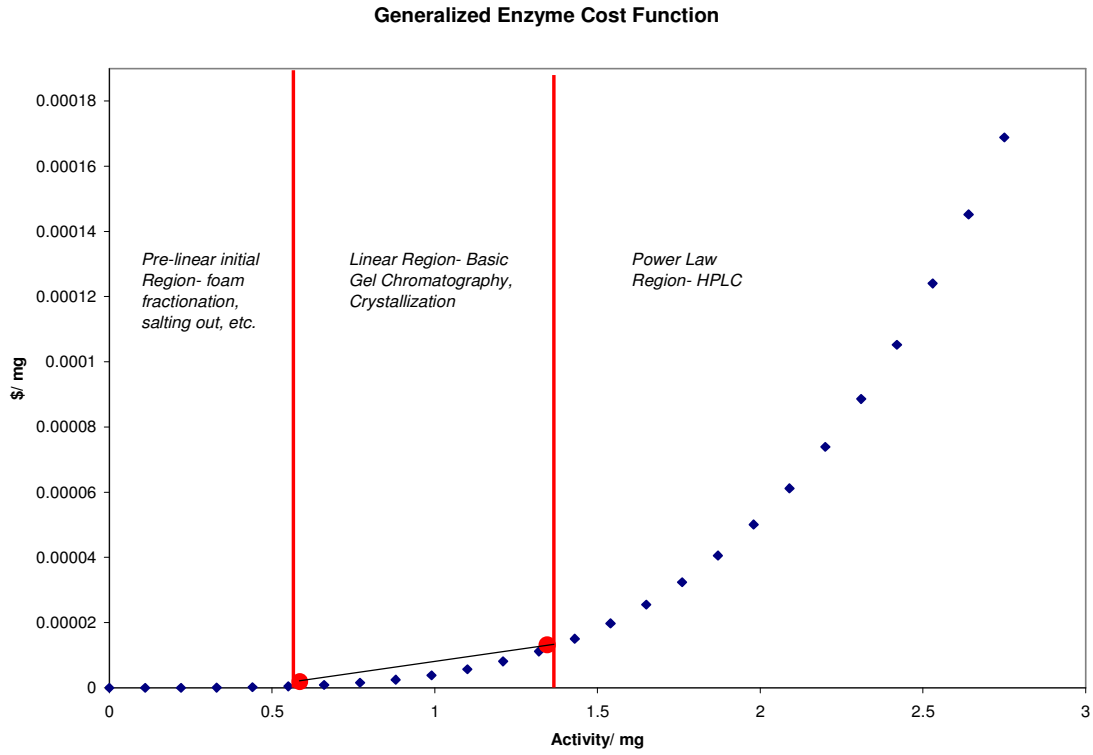


Figure C.6: The generalized function used was $y=4*10^{-6}*x^{3.7}$, where x is the activity per mg and y is the cost per mg of the generalized enzyme. The parameter values used, $\gamma=4*10^{-4}$ and a=3.7 were chosen arbitrarily, but lie within the realm of the parameters fitted to the collected data, see Table C.2.

We now further interpret the possible meaning, in terms of separation cost, of Equation (C.3). To do this analysis we break up the enzyme activity domain into three parts: the low activity-prelinear region, the moderate activity linear region, and the high activity exponential region. We do this based on a general consensus seen within all the collected data that the activity (and price) of each enzyme correlates with the separation techniques used. This correlation can be most easily seen for actual amylase data in Figure C.4.

The initial low activity pre-linear region, is characterized by little change in cost per mg as activity per mg is increased. This region appears to be characterized by low purity enzymes harvested from plants (or other readily available sources such as fermentation broths) and purified via traditional low cost processes such as foam fractionation or salting out. These purification processes can be extended at little cost to generate increased enzyme activity. Since the additional processing within this region requires little additional capital cost and only modest operating costs, the change in enzyme cost per mg may be marginal.

The second region, the moderate activity region (linear), is characterized by nearly a linear change, over the narrow band of the $4*10^{-6}*x^{3.7}$ function, in cost per mg as activity is increased. This region is generally characterized by moderately pure enzymes that have been purified by more costly processing than the pre-linear region. It appears that the majority of enzymes considered in this domain are purified via gel chromatography. The capital and operating costs of this processing are typically higher than the costs associated with foam fractionation and salting out processes[7]. Additional processing, for additional enzyme activity in this middle region where there is moderate activity, requires

larger chromatography columns or larger affinity gels whose costs tend to increase linearly in price.

The third domain is characterized by the observable (empirical) power law growth in enzyme cost per mg with respect to activity. This domain extends to the highest purity of the desired enzyme and is generally comprised of enzymes that require expensive processing. The overall processing is thus generally a combination of salting out, foam fractionation, and multi-stage gel chromatography, but often includes a final purification step, which is generally HPLC. The capital and operating costs for HPLC are significantly higher than that of the other mentioned purification processes. Additional purification through HPLC generally requires running a sample multiple times on additional columns. During each step of HPLC, a large fraction of the enzyme is lost due to the inefficiencies of the process, making the increased cost per mg behave like a power law function when several steps are used in series.

The derivative of the generalized cost function shows the incremental change in cost as activity is increased. This derivative function, represented as $y=1.5*10^{-5}*x^{2.7}$ leads to some key observations. For example, the power law exponent is still larger than one. This observation holds for every enzyme analyzed in this study, which increases our confidence in the analysis these enzymes. Even with a power greater than one, the cost will only change minimally with activities smaller than unity. Above unity, there will be a small region where price will increase only marginally as a function of activity. However, regardless of the value of the pre-exponential parameter, at some threshold

where the activity per mass is greater than unity, the enzyme cost will begin increasing significantly and exponentially as a function of enzyme activity.

Conclusions

The significant conclusion and most interesting observation of this study is that there appears to be one simple function that estimates the cost of an enzyme in $ per mg as a function of activity, $\Phi = \gamma (A)^a$, where a appears to be a generalizable exponent of the order 3.7. The generalizable function appears to fit best over the low to moderate activity range, but often under-predicts enzyme cost in the high activity range. This implies that there may not be a simple two parameter model that accurately captures the dynamics of enzyme activity related to cost over the entire activity domain. Additional data in the moderate to high activity range is needed to further characterize the limitations of Equations (C.2). However, the function developed in this analysis is still useful as an estimation tool where little or no real data may be known.

The generalizable function can furthermore be analyzed in terms of three characteristic domains describing the enzyme cost with respect to activity. These domains can be useful when purchasing enzymes for industrial or laboratory purposes. For example, if a low purity enzyme is being purchased, it is logical to purchase the enzyme at that purity equivalent to the end of the pre-linear phase of the generalized enzyme curve, denoted by the critical point located near an activity per mass ratio of 0.5 on the generalized enzyme cost function plot (Figure C.6), giving the buyer the most amount of active enzyme at a very low cost. A complimentary strategy for high purity enzymes would be to choose the

activity of the upper end of the moderate activity linear range (the lower end of the high activity exponential domain). If processing costs is dependent on enzyme impurities, the cost of processing can be compared to the cost of increasing enzyme activity, by using Equation (C.2) in a manner to minimize overall costs. Since the parameter a generally ranges between 3-4 between enzymes, a two-fold increase in activity is generally associated with an eight-fold increase in cost.

Since the parameters determined within this study were empirically computed from current market prices, these parameters are likely to change over long time-scales given the dynamic nature of any economic market. The enzymes used for this analysis are generally widely used proteins that, therefore, have fairly stable, well-developed markets. Enzymes used in small quantities for less industrial purposes may not follow similar trends since market fluctuations can be significant. This allows the industrial enzyme enzymes to be best suited for this type of study since their market will be less dynamic than other enzymes that are still within market development. However, economies of scale will still play a factor, especially in enzymes that are expected to see a wide increase in use and production scale-up such as cellulase. These changes will likely result in a scaling factor that will linearly decrease the value of γ. This observation is drawn from analyzing scale-up cost for typical bioprocessing operations involved in enzyme purification which show linear cost reductions as scale increases [14]. Similarly, this observation can be made by looking at typical equipment scale-up cost. One of the most common protein purification methods is liquid chromatography [15]. Chromatography columns tend to scale up linearly since the capital cost for the column is minimal compared to the high resin cost, which is required in directly proportional

amounts to column throughput. Scale-up of enzyme production is generally sublinear, but the data presented in this analysis demonstrates that purification costs tend to dominate overall product cost[15]. Regardless of changes in scaling due to market growth and scale-up, the functional forms determined for the relationship of enzyme activity and purchase amount to price should hold. The independent variables within Equation (C.2) were shown to be directly related to the process variables for enzyme production. Although the model under-predicts enzyme cost for the high activity range, Equation (C.2) still provides a useful tool for engineers to estimate enzyme cost based on activity.

References

1. Cherry, J. Ko, S., Grainger, R., Prokop, A., and Tanner, R. 2000. Developing An Objective Function to Characterize the Tradeoffs in Salting Out and The Foam and Droplet Fractionation Processes. *Brazilian Journal of Chemical Engineering*. **17**: 233-238.

2 Zhang, Q.; Lo, C-M.; Ju., L-K. 2007. Factors Affecting Foaming Behavior in Cellulase Fermentation by *Trichoderma reesei* Rut C-30. *Bioresource Technology.* **98**: 753-760.

3. Lo, C-M.; Zhang, Q.; Lee, P.; Ju, L-K. 2005.Cellulase Production by *Trichoderma reesei* using Sawdust Hydrolysate. *Applied Biochemistry and Biotechnology* **121**: 561-573.

4. Graff, G. 2007. *Pharma Market Spur Increased Enzyme Demand* http://www.purchasing.com/article/CA6419082.html

5. Poulsen, P. 1987. Trends in Industrial Applications of Enzymes. *Annals of the New York Academy of Sciences, Enzyme Engineering.* **501**: 413-419.

6. Nilesh; A., Kamat, M.; and Arvind, L. 2004. Expanded Bed Affinity Purification of Bacterial α-amylase and Cellulase on Composite Substrate Analogue- Cellulose Matrices. *Process Biochemistry*. **39**: 565-570.

7. Zhang, Q., Lo, C-M., and Ju, L-K. 2006. Affinity Foam Fractionation of Trichoerma Cellulase. *Applied Biochemistry and Biotechnology.* **129**: 1051-1065.

8. Carolina Biochemical Corporation. http://www.carolina.com. (catalog information on enzymes). 4/1/07

9. Fisher Scientific. http://new.fishersci.com. (catalog information on enzymes). 4/1/07

10. Sigma-Aldrich Corporation. http://wwwsigmaaldrich.com. (catalog information on enzymes). 4/1/07

11. Worthington Biochemical Corporation. http://www.worthington-biochem.com. (catalog information on enzymes). 4/1/07

12. Elastin Products Company, Inc. http://www.elastin.com (catalogue information on enzymes) 4/1/07.

13. "Enzymatic Production of Invert Sugar", Ensymm Consulting for Biotechnology (2007). http://www.ensymm.com/pdf/ensymmProjectstudyreportInversugarproduction.pdf

14. Cacciuttolo, M., and Arunakumari, A. , " Scale-Up Considerations for Biotechnology- Derived Products" http://biomedical.rutgers.edu/doc/Scale%20Up%20of%20Biotechnology%20Products.pdf

15. Blanch, H. and Clark, D. (1997). *Biochemical Engineering*. New York, New York. Marcel Dekker, Inc. pp. 678-682.

APPENDIX D

MATHEMATICA SCRIPT FOR COMPUTING THE PROBABILITY OF MISCLASSIFYING CT VALUES

```
Functions
<<Statistics`DiscreteDistributions`
<<Graphics`Graphics`
<<Graphics`MultipleListPlot`
<<Statistics`DataManipulation`

{\pard{{\f1 }}}
These functions take a List of data ( L_ ) and an interval size (q_) and make a histogram
H
f[L_,z_,q_]:=Count[Map[Ceiling[(#-Min[L])/q]&,L],x_?(#<=z&)] (* Set L , cut into q
size intervals *)
Bc[L_,z_,q_]:=f[L,z,q]-f[L,(z-1),q]/;z>1  (* 0 and 1 are an exceptional case *)
Bc[L_,z_,q_]:=f[L,z,q]/;z==1   (* bin counts *)
H[L_,q_]:=Interpolation[Join[{{Min[L]-q,0},{Min[L],0}},Table[{Min[L]+(q/2) + j
q,Bc[L,j,q]/Length[L]},{j,0,Ceiling[10 (Max[L]-Min[L])]-
1}],{{Max[L],0},{Max[L]+q,0}}],InterpolationOrder→1]
Read Data
SetDirectory["/data/PCR"];
Fn=RotateLeft[FileNames["sample*.txt"],4];
ReSample[k_,t_]:=Module[{l=Length[L],i,j,S,q,m,n,Data,R={},M,in, Pmiss={}},

 (* k = number of PCR runs taken together
 L = original ct value data set
 t = number of iterations of resampling
 S = set of resampled mean values
 *)

 For[j=1,j≤Length[Fn],
  {
   Data=Import[Fn[[j]],"List"];
   Clear[U];
   U=DiscreteUniformDistribution[Length[Data]];
   S={};
   For[i=1,i≤t,
    {
     RA=RandomArray[U,k];
     AppendTo[S,Mean[Map[Data[[#]]&,RA]]];
     };
    i++
    ];
```

```
 m=(Max[S]-Min[S]);
 q=m/Ceiling[10 m];
 AppendTo[R,{H[S,q],{Min[S],Max[S]}}]
 };
j++
];

For[j=1,j≤Length[Fn],
 {
 M={};
 msc=0;
 For[i=R[[j]][[2,1]],i≤R[[j]][[2,2]],

  {
  (* Compute Misclassification *)

  in=Drop[Range[Length[Fn]],{j}];
  While[Length[in]>0,
   {
    n=in[[1]];
    in=Drop[in,1];
    msc+=R[[n]][[1]][i];
    }
   ];

  If[(R[[j]][[1]][i]+msc)>0,AppendTo[M, R[[j]][[1]][i] msc/(R[[j]][[1]][i]+msc)]];

  msc=0;

  };
 i+=0.1
 ];

 AppendTo[Pmiss,Plus@@M];

 M={};

 };
j++
];
Pmiss
]
```

APPENDIX E

PLATE FILLING MODEL

```
Map[Mean[Table[Length[Complement[Range[1,96],RandomInteger[{1,96},#]]],{1000}]
]&,Range[0,600]]//N

Map[StandardDeviation[Table[Length[Complement[Range[1,184],RandomInteger[{1,18
4},#]]],{1000}]]&,Range[0,600]]//N
```