INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

U·M·I

University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 9116445

Kinetics, modeling and optimization of recombinant yeast fermentations

Patkar, Anant Yeshawant, Ph.D.

Purdue University, 1990

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

Graduate School Form 9
(Revised 8/89)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis Acceptance

This is to certify that the thesis prepared

By Anant Yeshawant Patkar

Entitled

KINETICS, MODELING AND OPTIMIZATION OF RECOMBINANT YEAST FERMENTATIONS

Complies with University regulations and meets the standards of the Graduate School for originality and quality

For the degree of Doctor of Philosophy

Signed by the final examining committee:

__________, chair

____

____

____

Approved by: __________ Department Head __________ Date

This thesis ☐ is ☒ is not to be regarded as confidential

____
Major Professor

KINETICS, MODELING AND OPTIMIZATION OF RECOMBINANT YEAST FERMENTATIONS

A Thesis

Submitted to the Faculty

of

Purdue University

by

Anant Patkar

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 1990

To my mother and father

ACKNOWLEDGEMENTS

I would like to sincerely thank Professor J.-H. Seo for his help, guidance and encouragement during the course of my graduate studies. During the initial stages of my work Professor H. C. Lim served as a co-advisor. I appreciate his assistance and friendship during this period.

I wish to thank Professors G. T. Tsao, A. E. Emery and M. R. Okos for serving on my committee. The financial support provided by the National Science Foundation and the Department of Chemical Engineering is also appreciated.

Thanks are due to Jayakumar for helping me get acquainted with laboratory work. I also appreciate the help of David Carmichael and Stacey Clark with my frequent computer-related queries.

I am thankful to all my friends at Purdue for all the insight, friendship, and the enjoyable times we shared. Last but not the least, the constant encouragement and support by my family throughout the course of my study is sincerely appreciated.

TABLE OF CONTENTS

Page

Page

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

NOMENCLATURE

CP	Intracellular concentration of complementation product, units/g
CN	Intracellular plasmid concentration, (number of plasmids)/cell
D	Dilution rate, /hr
F	Fraction of plasmid-free cells, or feed rate L/hr.
g	Gradient of the hamiltonian with respect to the control vector.
H	Hamiltonian.
I	Inhibitor concentration, mg/L
K, k	Model parameters, or weight on the penalty function.
k_d	First order inactivation constant, /hr
N	Concentration of mRNA, (number of mRNA)/OD.L
P	Intracellular concentration of invertase, (units/OD.L)
p	Probability of forming a plasmid-free cell upon cell division.
R	Fraction of glucose that is channeled through the fermentative pathway, or metabolic flux, (g glucose)/hr.OD
r_t	Rate of transcription of SUC2 gene (number of mRNA)/hr.OD.L
S	Substrate concentration, g/L
s	Conjugate gradient direction.

t	Time, hr
u	Control variable.
V	Volume, L
w	A boundary function defined by Eq. (6.12).
X	Cell mass concentration, g/L or OD
x	State vector.
Y_{xr}	Yield of respiratory pathway, OD/(g. glucose)
Y_{xf}	Yield of fermentative pathway, OD/(g. glucose)
Y_{XS} , $Y_{X/S}$	Cell yield, (g. cell mass)/(g. substrate) or OD.L/(g. substrate)

GREEK

α	A search parameter in the conjugate gradient method.
β	A search parameter in the conjugate gradient method.
δ	A step size parameter in the conjugate gradient method.
ε	Operating cost per unit time.
θ	Probability of forming a plasmid-free cell upon cell division.
λ	Adjoint vector, or costate vector.
μ	Specific cell growth rate, /hr
Π	Performance index, or objective function.
π	Invertase formation rate, KU/hr
ϕ	Gradient of hamiltonian in singular problems.

SUBSCRIPTS AND SUPERSCRIPTS

+	Plasmid-containing cells.
–	Plasmid-free cells.
*	Optimal.
d	Decay.
F	Feed.
f	At the end of fermentation, or fermentative.
max	Maximum possible value.
min	Minimum possible value.
s	Singular.
T	Transpose.
t	Total.
r	Respiratory.

ABSTRACT

Patkar, Anant, Ph. D., Purdue University, December 1990. Kinetics, modeling and optimization of recombinant yeast fermentations. Major Professor: Jin-Ho Seo.

The objective of this research is to optimize the performance of fed-batch fermentations involving recombinant organisms. This goal is achieved by using, as a model system, the production of invertase from *Saccharomyces cerevisiae* containing the recombinant plasmid pRB58. The plasmid contains a copy of the yeast SUC2 gene which codes for invertase. The expression of the SUC2 gene is repressed at high glucose levels in the media.

Batch and fed-batch fermentations were performed with this strain in selective media. It was observed that the cell yield decreases with when glucose concentration in the medium is high. Also, the specific invertase activity is very tightly regulated by glucose levels. An unstructured mathematical model was developed to describe the experimental results. The model explains the variation in cell yield by postulating a cybernetic principle, which regulates the fluxes in the respiratory and fermentative pathway. The invertase production was modeled by a simple substrate-inhibition form.

A conjugate gradient algorithm was developed and tested for a variety of singular systems. This technique was used to determine the optimum substrate feed rate profiles in a fed-batch mode of fermentation for the model system. The optimum feed rates result in an initial cell growth phase followed by an invertase production phase.

Aside from this main research theme, two other optimization problems were solved. First, a simple model was developed to describe anaerobic fermentations involving recombinant yeast. This model was then used to determine the optimum glucose feed rate profiles in a fed-batch mode of operation. Second, a mathematical model was developed to describe a recombinant *E. coli* system. Using singular control theory the problem of determining the best feed rate in a fed-batch fermentation was reduced to a simple one-parameter optimization problem. This simpler problem was then solved using a one-dimensional search. In both these systems, optimal feed rates

resulted in an initial high cell growth phase followed by high product formation phase.

1. INTRODUCTION

1.1 Background

In the late 1970s a set of biochemical methodologies emerged which made genetic engineering possible. The discovery of highly specific enzymes for cutting and joining DNA provided the tools for making novel DNA and incorporating it into cells. For microbiologists and biochemists recombinant DNA methodology is a powerful tool for studying how genes work. Its importance for biochemical engineers, however, lies in its applicability to the production of proteins of medical and industrial importance. Developmental work has already led to the production of several biologically active human agents including insulin, interferon and growth hormone in microorganisms like *E. coli*, *B. subtilis* and *S. cerevisiae*.

The transition from gene manipulation in the laboratory to large scale commercial production of proteins is crucial. The factors affecting expression of heterologous genes in microbial hosts are complex, and maximizing expression of cloned sequences is of utmost importance. Successful scale-up of laboratory experiments to large fermentors puts demands upon the stability of the cloned DNAs and requires careful investigation of the influence of growth conditions upon gene expression. The practical problems involved include (i) the instability of foreign DNA in the host of interest, (ii) the tendency for the expression products to be intracellular, causing an increase in the production costs and a decrease in cell viability, (iii) the fact that expressed products, especially the intracellular ones, are often inactive, requiring expensive unfolding and refolding processes.

Maximizing the productivity of recombinant systems needs special attention to be given to the trade-offs involved between cloned-gene expression and the growth of the host cells. Control of cloned-gene expression levels is important, because overproduction of cloned-gene proteins is deleterious and sometimes even fatal to the host cell. The use of regulated promoters provides an efficient way of manipulating

gene expression and hence improving the product yield by uncoupling the cell growth stage and the cloned-gene expression phase during fermentations.

Earlier research involving recombinant organisms was focussed primarily on prokaryotes like *E. coli*. However, the structural organization of prokaryotes is fundamentally different from the more complex eukaryotes. A well studied eukaryote like *Saccharomyces cerevisiae* provides an ideal model system to study expression of higher eukaryotic genes, especially when post-transcriptional and post-translational modifications are essential.

1.2 Research Objectives

The objective of this research is to optimize the performance of fed-batch fermentations involving recombinant microorganisms. This goal is achieved by using, as a model system, the production of invertase from *Saccharomyces cerevisiae* (commonly known as baker's yeast) containing the recombinant plasmid pRB58. The plasmid contains a copy of the yeast SUC2 gene which codes for invertase. The expression of the SUC2 gene is repressed at high glucose concentrations in the media. This characteristic allows regulation of invertase production by manipulating glucose feed rate in a fed-batch mode of operation.

The problem of determining the best feed rate to maximize productivity falls under the realm of calculus of variations. Since the feed rate appears linearly in the mass balance equations the problem is singular in nature. Computation of optimum feed rate requires a mathematical model able to describe the cell growth and product formation. A technique for optimizing fed-batch fermentations described by four or less mass balance equations was developed by Modak *et al.*[1]. The optimization of higher dimensional systems is computationally difficult, if not impossible. This restricts the dimensionality of the model that can be chosen for optimization purposes. Structured models, which explicitly account for various cell constituents, or segregated models, which treat the cell mass as consisting of individually distinguishable cells, are not appropriate for the purpose at hand. Thus the model should be simple enough to be amenable to optimization, while detailed enough to describe the key characteristics of the process.

Considering the complexity of many biochemical processes, there is a need to develop optimization techniques applicable to higher (more than four) dimensional systems. Modak *et al.*[2] developed a nonsingular control approach, which can, in

principle, be used for this purpose. However, its applicability was not demonstrated for systems with dimensions higher than three. To overcome these shortcomings, a conjugate gradient technique for optimization of fed-batch fermentation processes is proposed. Phosphate limited growth of unstable recombinant yeast cultures was modeled by DiBiasio *et al.*[3] using five differential equations. This system provides an ideal test case for the new approach.

The research objectives can be summarized as:

1. To determine the kinetics of recombinant yeast growth and invertase production;
2. To develop a mathematical model to describe the observed experimental results;
3. To compute the best feeding policy to maximize invertase productivity;
4. To develop an optimization technique for higher dimensional fermentation processes and test it.

2. LITERATURE SURVEY

Recombinant DNA technology is a relatively new field in biochemical engineering and hence some terms, which will be used later, are explained first. Plasmid vectors need certain components in addition to the cloned gene. Sequences common to almost all plasmid vectors will be discussed. The main features of recombinant systems, as opposed to wild type systems, are highlighted next. A considerable effort has been devoted towards modeling of recombinant systems in the last several years. The next section addresses this issue. Since yeast is used as the host organism in this research, the characteristics of yeast, including its growth metabolism, important features of its gene expression and the common plasmid vectors are explained next. The aim of the research is to optimize recombinant yeast fermentations after developing a suitable mathematical model. Subsequent sections deal with previous attempts at optimizing fed-batch cultures and modeling yeast growth. The recombinant plasmid used in this study is pRB58. Its structure and the function of its important constituents are discussed in detail in the final section.

2.1 Background material and terminology[4]

2.1.1 What is genetic engineering? Before the advent of recombinant DNA methodology, several attempts were made to transform pro- and eukaryotic cells with foreign DNA. Their experiments met with little success because the exogenous DNA cannot replicate and hence can not be maintained stably in the transformed cell.

If fragments of DNA cannot replicate, the obvious solution is to attach them to a suitable replicon†. Such replicons are known as vectors. Composite molecules in

† Replicons are DNA molecules which contain an origin of replication and hence are able to replicate in the host cell.

which foreign DNA has been inserted into a vector molecule are called recombinant molecules. The construction of such composite molecules has also been termed genetic engineering or gene manipulation. The process has also been termed gene cloning because a line of genetically identical organisms, all of which contain the composite molecule, can be propagated and grown in bulk hence *amplifying* the molecule and any gene product whose synthesis it directs.

The insertion of a piece of foreign DNA into a vector requires certain techniques to be available:

1. Mechanisms for cutting and joining DNA molecules from different sources,
2. A method for monitoring the cutting and joining reactions,
3. A technique of transforming.

Currently, the operations of cutting and joining DNA molecules are done by using restriction endonucleases and DNA ligases, respectively. A standard method for monitoring the reactions is agarose gel electrophoresis. Transformation is normally achieved by permeablizing the cell walls by the use of suitable chemicals.

2.1.2 Plasmid structure. Plasmids are replicons which are stably inherited in an extra-chromosomal state. Along with bacteriophages they are the most commonly used vectors. The plasmids should contain, in addition to the genetic information for the desired protein, control sequences for efficient expression and marker genes.

A marker gene is a gene whose product gives the plasmid-containing cell a specific property distinguishing it from cells that do not contain plasmids. Typically a marker gene confers on the host cell resistance to some antibiotics or complements some auxotrophic mutations in the host cell. Selection markers can serve two important functions. First, they provide rapid positive selection procedures for identifying vector-containing colonies. Second, they allow formulation of selective media to minimize competition from any plasmid-free cells that may be born during population growth.

The term control sequences means the DNA sequences that control the expression and localization of the cloned product. Thus a typical gene consists of a promoter, an operator, a ribosome binding site, a secretion signal (if any) and a nucleotide sequence coding for the desired product.

At the beginning of transcription, RNA polymerase binds itself specifically to a starting sequence, the *promoter*[5]. It has been found that the nucleotide sequences for different promoters are not exactly the same, and hence the strength of association of the promoters with RNA polymerase varies. The stronger the promoter, the higher will be the rate of initiation of transcription.

The rate of expression of some genes is constant, independent of the environmental conditions. These genes are termed *constitutive*. On the other hand for some genes the rate of transcription may change according to the needs of the cell under different growth conditions. In many organisms, transcription is regulated by proteins that, by binding to DNA near or within the promoter, increase or decrease the rate at which RNA polymerase initiates mRNA synthesis. The regions where these regulatory proteins attach are conventionally called *operators*. The gene coding for the secreted products contains a transport signal sequence that directs the localization of products destined for secretion.

2.2 Characteristics of recombinant systems

2.2.1 Plasmid instability. Although one can, in principle, start with a culture containing only plasmid containing cells (X^+), in due course, a considerable fraction of cells in the culture would not have any plasmids (X^-). This commonly occurring problem, known as *plasmid instability*, can be due to any one or more of the following factors. First, if the replication of the plasmid is not under the control of a nuclear chromosome, it may replicate less than once per cell cycle, resulting in a decrease in the number of plasmids per cell and ultimately causing plasmid loss. Second, in some organisms asymmetric cell division leads to unequal distribution of plasmids between the newly formed cells. For example, in *Saccharomyces cerevisiae* the cell division by budding is asymmetric, leading to the formation of different sized cells, the larger one being called the mother cell and the smaller one the daughter cell[6]. It is found that plasmids belonging to a particular class segregate unequally in these cells, the mother cell getting more plasmids than the daughter cell. Third, for reasons explained later, the X^- cells generally have a higher growth rate than X^+ cells in a nonselective medium. Thus, X^- cells eventually outgrow the X^+ cells. The above reasons result in what is called *segregational instability*. The other type of instability, called *structural instability*, results in the inability of cells to synthesize an active cloned gene product. This can be caused by mutation either in the structural cloned gene or in the associated sequences which control and make possible cloned gene expression.

Several methods exist for minimizing this problem. The simplest is to grow the cells in a selective medium, to give X^+ cells a selective advantage. Even with selection pressure, segregational instability reduces the overall growth rate and productivity of the recombinant population. Also, selection pressure does not control structural instability. Another solution is the use of the fed-batch mode of operation which can be used to control the cell environment to minimize the effects of instability. The above solutions may be considered as operational solutions. A molecular biological solution is to integrate the required gene in the host chromosome. Since chromosomes are replicated once per cell cycle and are inherited equally by the daughter cells, the instability problem can be completely avoided. One problem with this approach is that it results in very few copies of the required gene per cell. At low copy numbers, the productivity can be considered to be linearly proportional to the copy number. Thus the use of chromosomal integration results in a productivity lower than that by the use of multicopy plasmids.

The importance of our understanding of plasmid stability has been recognized and emphasized by many workers in the field of genetic engineering. Imanaka and his coworkers[7,8] observed that plasmid stability can be affected by the genetic characteristics of the host cells, the copy number of the plasmid, culture conditions, and the genes contained in the plasmids. Many experimental studies of this phenomenon have been reported[9-18]. Helling, Kinney and Adams[16] pointed out unsteady and oscillatory growth patterns of *E. coli* with and without plasmid RSF2124 or its derivative when the two strains were mixed at the start of a continuous culture. Depending on the initial fraction of plasmid-containing strain of *E. coli*, the population of the plasmid-containing strain decreased conspicuously after the start of the chemostat culture, but the strain increased its population later on, reversing the population balance with respect to the plasmid-free strain. The oscillatory behavior of the plasmid-harboring cell was attributed to adaptive mutation of the plasmid. Marquet, Alouani and Brown[17] reported a continuous culture study for a double mutant of *S. cerevisiae* containing a plasmid coding for α-antitrypsin. The results indicate that the strain was maintained for 150 generations with no plasmid loss. Siegel and Ryu[19] studied plasmid instability for *E. coli* in a two-stage continuous culture system.

Aiba and Koizumi[18] examined the effect of temperature on plasmid stability for *Bacillus stearothermophilus* containing the plasmid pLP11. When the temperature was lower than 50°C, the plasmid functioned normally undergoing neither partial deletion nor total disappearance. When the temperature of cultivation exceeded 50°C, the

stability deteriorated in the late-log phase of batch culture and in continuous runs. Koizumi and coworkers[11] also reported a study of the effect of temperature and dilution rate on the copy number of plasmids in continuous culture of *Bacillus stearothermophilus*. Kim and Ryu[14] studied the instability kinetics of *trp* operon plasmid ColE1-*trp* in recombinant *E. coli*. The *trp* operon is partially derepressed by 3-β-indole acrylic acid (IAA). They found a critical IAA concentration at which the enzyme production, stability and growth rate were significantly affected.

Hopkins *et al.*[12] examined the effect of dissolved oxygen shock on the stability of recombinant *E. coli* containing plasmid pKN401. The recombinant cells were highly unstable under conditions where a dissolved oxygen shock was induced. Parker and DiBiasio[10] studied the effect of growth rate and expression level on plasmid stability in *S. cerevisiae*. As the level of plasmid expression was raised, the stability dropped markedly. It was pointed out that increased transcription might have interfered with plasmid replication, hindered segregation, or overburdened the cell's DNA repair capability. It was also observed that plasmid stability is substantially greater at higher growth rates.

2.2.2 Host-plasmid interactions. Expression of plasmid genes and binding of cell catalysts, initiation factors, and other regulatory molecules to plasmid DNA alter the normal pattern of interactions and synthetic activities in the host cell. This has stoichiometric as well as kinetic implications. In cases of highly active cloned gene expression and/or extremely large plasmid content, recombinant cells become nonviable[19,20]. The effects of the presence of plasmids and gene expression on the host are listed below.

A. Recombinant cell metabolic stoichiometry is altered by the synthesis of plasmids and their products, requiring allocation of monomer precursors and metabolic energy for plasmid-directed synthesis. This necessarily results in lower cell mass yields.

B. It is expected that the magnitude of plasmid-mediated perturbation on host-cell growth and biosynthetic activities should increase as the number of plasmids per cell increases. This was confirmed by the experimental studies of Seo and Bailey[21] using plasmids containing mutated repressors of plasmid synthesis in *E. coli*. They found that specific growth rate is a monotonic decreasing function of plasmid content.

C. The dependence of product formation rate on plasmid copy number was determined by Seo and Bailey[21] for the system described above. A 34-fold increase in copy number resulted in only a sevenfold increase in cloned-gene product activity. Although an approximate proportionality between number of genes and amount of gene product activity was found through a copy number of around 60, this proportionality subsequently broke down rapidly. It has to be noted that the plasmid content may be increased by genetic manipulation to reduce plasmid replication activity or by environmental manipulation through specific growth rate reduction. It was found that cloned-gene specific product activity increased much more rapidly for environmental change of plasmid content. This illustrates the inadequacy of plasmid content alone as a determinant of productivity.

D. The specific growth rate of recombinant *S. cerevisiae* exhibits a maximum as a function of plasmid content[15].

E. The reduction in growth rate of the host cell, because of the presence of plasmids, is more when the product gene within the plasmid is expressed.

2.3 Modeling recombinant systems

In order to interpret data, test hypotheses, and synthesize knowledge gained from experimental studies, a conceptual, quantitative framework for description of host-plasmid interactions is required. In the past few years a considerable amount of effort has been directed towards modeling of recombinant systems. Modeling and theoretical studies have addressed both stoichiometric and kinetic issues.

The models available can be categorized broadly as segregated and nonsegregated, or as structured and unstructured[22]. Segregated models recognize that the system contains many cells and hence require population balance methods for its mathematical representation. Nonsegregated models, on the other hand, are not concerned with distributions of certain properties over the entire population, and consider only the average properties of the population. Structured models assume the system to be made up of two or more parts. Unstructured models do not make such a distinction. Actual cellular systems can be considered to be structured and segregated. But structured, segregated models tend to be quite complex and do not lend themselves to easy mathematical treatment.

2.3.1 Segregated, unstructured models. Hjortso and Bailey[23,24] proposed a segregated model for recombinant *S. cerevisiae*. They considered the problem of finding plasmid content distribution in a growing population of budding yeast. A detailed cell-cycle model developed earlier[25,26] was used to find out an age distribution function. Two hypothetical models for plasmid replication regulation were explored: one assumed an increase in replication frequency with a decrease in copy number and the other assumed a constant number of replications independent of the initial number of plasmids. Analytical solutions were obtained for plasmid content distribution for steady-state growth in a selective medium and for dynamic growth following a shift to a non-selective medium. Analogous models have been proposed for control of plasmid replication in *E. coli*[14,27].

A segregated model for plasmid content and product synthesis was proposed by Seo and Bailey[27]. It was assumed that the single-cell rate of product synthesis depends only on the number of plasmids in the cell. This assumption was found necessary to avoid excessively complicated and unwieldy mathematics. Using a parabolic form for the dependence of specific product formation rate on plasmid content, they were able to show that the overall specific product synthesis rate measured is less than or equal to the single-cell rate of gene expression evaluated at the average plasmid content. Consequently, efforts to estimate the single-cell gene expression rate from population-average data will underestimate the true single-cell rate. It was also shown that the fraction of plasmid-containing cells is essentially a function of the single cell growth rate ratio α, and the probability of plasmid loss, θ. The parameter α may be controlled by adjustment of the growth environment, and θ is characterized by plasmid genetic functions.

2.3.2 Structured, nonsegregated models. These models can be classified broadly as genetically structured and genetically unstructured models. Almost all genetically structured models for recombinant systems are due to Lee and Bailey[28-31]. These models assume some mechanisms at the molecular level of DNA replication and gene expression.

Lee and Bailey[28] formulated a mathematical model describing λdv replication in a growing single cell of *E. coli*. A wealth of information is available about the regulation of replication of this plasmid. The model successfully simulates plasmid maintenance, cellular content of plasmids and important regulators of plasmid synthesis kinetics. It also describes well the influences of mutations in the replication regulation

functions. This model was extended to include expression kinetics of a cloned gene in the plasmid[29]. Detailed descriptions were employed for the plasmid regulatory functions. Empirical relationships deduced from previous experiments were employed to describe the effect of cell growth rate on host cell macromolecular synthesis activities. Analysis of simulations showed that the primary reason for higher copy number at lower growth rate is reduced synthesis of repressor protein. An important aspect of this model is that the structure of this model and the parameters involved were determined entirely from information in the molecular biology and microbiology literature. Lee and Bailey[30,31] also developed genetically structured models for *lac* promoter-operator function in the chromosome and in multicopy plasmids. The basis for this model is provided by the great amount of information available on *lac* promoter-operator system. Genetically structured models of this type should prove increasingly useful in analyzing the influence of particular genetic modifications in expression systems and ultimately in optimizing these systems.

Lee and coworkers[32] also developed models for product formation by unstable recombinant organisms in batch and in continuous flow reactors. The cell population was characterized by three different genotypes according to the absence or presence of plasmids (segregational instability) and of the active cloned gene (structural instability). Empirical growth inhibition factors due to plasmids and to product protein were assigned to the corresponding strains. Product formation kinetics was based upon a quasi-steady-state transcription-translation expression model. An approximate form of this model is identical to the empirical Leudeking-Piret expression for product formation. Simulation results showed that there exists an optimum combination of plasmid copy number and cloned-gene transcription and translation efficiencies to maximize reactor productivity. These host-vector parameters can be manipulated by both genetic engineering and by bioreactor control.

Srienc and coworkers[15] recognized that it is not the marker gene itself which endows the host cell with the selective growth type but the product of the gene, either RNA or protein, which performs the functions essential for growth in a selective medium. This fact is very important, because in unstable recombinant populations, the newly formed cells may contain the complementing product but may not have any plasmids. It has been observed experimentally that cells with plasmids typically grow more slowly than cells without plasmids, and this growth rate penalty increases as the number of plasmids per cell increases and as the level of gene expression from the plasmid increases. On the other hand, presence of complementation product increases

the growth rate of host cells relative to the cells that do not contain the complementation product. The specific growth rate of plasmid-containing cells (μ_+) was represented as the product of an activation term, which is a function of the intracellular concentration of the complementation product (CP), and an inhibition term, which depends upon the intracellular plasmid concentration (CN).

$$\mu_+ = A(CP)\, I(CN) \tag{2.1}$$

By assuming some functional forms for A and I, it was shown that the growth rate μ_+ shows a maximum with respect to the plasmid content for otherwise identical recombinant strains.

Mathematical characterization of the growth of recombinant populations in a selective medium is complicated by independent segregation of the plasmids and the complementation product. In most cases the number of plasmid molecules per cell is substantially smaller than the number of complementary product molecules per cell. Assuming that the complementing product[15] is randomly distributed over the volume of the dividing cell, when irregular plasmid partitioning occurs, such that one newborn cell receives all of the plasmids, both daughter cells will contain at birth identical concentrations of complementing product. Thus, it is to be expected that plasmid-free cells will continue to grow in a selective medium following plasmid loss. This has been observed experimentally by Murray and Szostak[33]. They observed that *S. cerevisiae* cells born without plasmids were able to grow on selective medium for five to six generations. Srienc and coworkers[15] have modeled the growth of cells in a selective medium after plasmid loss. The decrease in the complementation product concentration with time was represented by a simple mass balance with first-order inactivation.

$$\frac{d(CP)}{dt} = -k_d\, CP - \mu_- CP \tag{2.2}$$

where k_d = first order inactivation constant.

and μ_- = growth rate of plasmid-free cells.

It was concluded that greater complementing product stability and higher mother cell plasmid content enhance the growth capabilities of X^- cells.

Peretti and Bailey[34] studied the effects of copy number, promoter strength, and ribosome binding site strength on metabolic activity of host cell and plasmid gene expression. Their model details many of the reactions central to DNA replication, transcription and translation for glucose-limited balanced growth of *E. coli*. Control interactions and reaction mechanisms were formulated using as much detail as possible for the processes of transcription and translation. Relative promoter and ribosome binding strengths for the vector genes and mRNAs were used as inputs to the system. Based on these the model simulates the alterations in host cell metabolism and cloned gene product synthesis. An interesting result, unverified experimentally, is that the rate-limiting process changes as amplification increases. It appears that ribosome availability is the limiting factor at low copy numbers. However, as the copy number increases, the availability of RNA polymerase is the more significant bottleneck in product formation. The simulations also indicate that enhanced vector-specified affinity for ribosomes is a better strategy for improving productivity than enhanced promoter strength.

Plasmids affect the host cell allocation of precursors, energy reserves, catalysts, activators, and repressors. The synthesis of plasmids and their products requires that nutrients be utilized to satisfy the additional material and energy demands of the cell. The effect of this on metabolic stoichiometry can be analyzed independently of cell kinetics. Da Silva and Bailey[35] analyzed the effect of the presence of plasmids on growth yields. Theoretical growth yield factors were estimated. Their method, although developed for *E. coli*, can be extended to other organisms for which the ATP use in metabolic activities is known or can be reasonably estimated. It was found that the ATP requirement for the synthesis of the additional plasmid DNA is minimal compared to that required for the excess protein. The level of expression, however, does influence the yields on ATP.

2.3.3 Conclusions. The above brief survey of the modeling efforts for recombinant systems shows that mechanistically detailed models exist for *E. coli*, but not for *S. cerevisiae*. This is because much more is known about the metabolic and genetic aspects of *E. coli*. However, in spite of the great amount of information available most of the conclusions of the detailed models cannot be verified experimentally. Also, genetically structured models were developed for very specific systems and are difficult to modify for application to other systems. The models for yeast are more empirical in nature and assume arbitrary functional forms for growth rate and product formation.

2.4 The yeast *Saccharomyces cerevisiae*

2.4.1 Yeast as a host of choice. Earlier research in recombinant DNA technology was focussed mostly on prokaryotes like *E. coli*. However, prokaryotes possess a structural organization that is fundamentally different from the more complex eukaryotic organisms. Hence, they do not provide the best environment to study the spontaneous expression of chromosomal genes of higher eukaryotes, especially when post-transcriptional and post-translational modifications are required. In eukaryotic genes there exist certain non-coding sequences called *introns* that are copied into primary transcripts[4,5]. These introns are subsequently eliminated by a process known as splicing. Removal of introns by splicing does not take place in prokaryotes like *E. coli*.

S. cerevisiae is an attractive host for studying the expression of eukaryotic proteins. It is nonpathogenic. Since it has been a popular organism for experimental studies, there is an abundance of information available on several aspects of yeast microbiology and biochemistry. Yeast cells are organized into the same major membrane-bound compartments as all other nonphotosynthetic eukaryotic cells. In addition to the nucleus, yeast cells contain the endoplasmic reticulum, the golgi apparatus, mitochondria, the cytosol, the vacuole, etc. *S. cerevisiae* is capable of expressing directly some eukaryotic and prokaryotic genes. It performs at least some post-translational and post-transcriptional modifications characteristic of eukaryotes. The yeast secretory pathway is similar to higher eukaryotes. Also, its extensive use over the years has shown it to be a successful industrial organism. There is yet another reason for choosing yeast as a host for cloning genes from eukaryotes. Although the genetic code is redundant, the different codons for a particular amino acid are not used equally and codon preference differs from organism to organism. While the codon usage in *E. coli* is quite different from that of the animal cells, that of yeast is rather similar. Thus, because of availability of different tRNAs a protein from a cloned eukaryotic gene may be synthesized in greater amounts in yeast compared with *E. coli*.

2.4.2 Gene expression in yeast[4]. Yeast DNA, like the DNA of all eukaryotic cells, is transcribed by three different forms of RNA polymerase, each of which transcribes different sets of genes. RNA polymerase II makes all the RNA that is to serve as mRNA.

A given yeast mRNA molecule, like most other eukaryotic mRNAs, carries the genetic message for only a single polypeptide chain. Thus, regulatory molecules that coordinate the transcription of several functionally related mRNAs must act at multiple chromosomal sites. Most yeast mRNA molecules, like those of all other eukaryotic cells, contain relatively long stretches (about 200 residues) of poly A at their 3' ends. These poly A tracts are not specified by the genomic DNA, but are added after transcription to the newly made mRNA molecules released from DNA into the nucleus.

A yeast gene can be defined as a DNA segment that is transcribed into a single mRNA product together with the adjacent control elements at its two ends. These control elements are specific groups of base pairs that by binding to specific cellular proteins, direct the starting and stopping of RNA chains at correct positions along the chromosomes. Yeast promoters differ in a fundamental way from bacterial promoters through the possession of upstream activating sequences (UASs) and upstream repressing sequences (URSs), which can be several hundred base pairs 5' to the start of transcription. In addition to these upstream elements, yeast promoters have two other essential components: (1) TATAAA-like blocks (called TATA sequences) usually placed some 20 to 40 base pairs 5' to the transcription start site and (2) an initiator element encompassing the start site, which precisely determines where transcription begins. On the contrary, RNA polymerase II promoters lie immediately 5' to the sequences transcribed.

Many yeast genes are transcribed at constant rates throughout the cell cycle. The speed of such constitutive synthesis is a function of the specific base pairs at the yeast promoters. Other promoters work at variable rates that depend on the extent of their binding to specific control proteins, many of whose amounts are environmentally controlled. Many highly regulated genes that are turned on by positive regulatory proteins also maintain finite basal levels of constitutive expression that is not under environmental control. It has been found in many such cases[3] that different promoter elements are involved for constitutive and regulated expression.

In *S. cerevisiae*, highly expressed genes are almost exclusively those encoding glycolytic enzymes or enzymes otherwise involved in carbon metabolism. However, expression from promoters of genes encoding glycolytic enzymes is essentially constitutive. In contrast, the expression of genes encoding enzymes required for catabolism of fermentable carbon sources other than glucose (e.g., sucrose, galactose, and maltose) exhibits substantial alteration of expression in response to growth

conditions. Expression of genes encoding maltase (MAL1 through MAL6) and those encoding invertase (SUC1 through SUC7) show induction, by addition of maltose or sucrose to the culture medium, respectively, from an essentially undetectable basal level to a level containing 1% of the cellular RNA. Thus DNA fragments from the 5' end of regulated genes like GAL10 (coding for an enzyme required for galactose metabolism), and PHO5 (coding for one form of acid phosphatase) can be fused to other genes, resulting in similar regulated expression of those genes.

2.4.3 Plasmid vectors for yeast. The first sequences used to make a functional yeast vector were those of the 2μm plasmid, the only plasmid that has so far been found to occur naturally in yeast. The yeast 2μm plasmid is a circular double-stranded plasmid present at 60-100 copies/cell in most *S. cerevisiae* strains. When a single 2μm circle is introduced into a plasmid-free yeast, its replication, as well as that of its immediate descendants, occurs more than once per every cell cycle until the normal complement of about 60 copies per cell is reached. Plasmids carrying 2μm DNA sequences are called YEp or yeast episomal plasmids. Those which transform by integration into a yeast chromosome are called YIp or yeast integrative plasmids.

Another source of yeast sequences for making new plasmids that multiply in yeast cells are the many genomic segments at which DNA replication is ordinarily initiated along the yeast chromosome. These sequences are called autonomously replicating sequences (ARS). These yeast replicating plasmids (YRp), however, are not stably maintained in transformed cells. For unknown reasons, YRp plasmids tend to remain associated with the mother cell and are not efficiently distributed to the daughter cell.

Highly stable yeast plasmid vectors can be made by including, in addition to ARS sequences, certain yeast DNA sequences that function as centromeres. These sequences are necessary for the attachment of the chromosomes to the microtubules of the mitotic spindle. Such CEN-containing plasmids (YCp) function as true chromosomes and segregate accurately during both meiosis and mitosis.

All versatile vectors for use in yeast cells contain origins of replication active in both yeast and *E. coli* as well as one or more markers that can be used to select those *E. coli* cells into which the vectors have entered. The shuttle vectors can thus be amplified in *E. coli* before being used to transform yeast cells.

2.5 Optimization of fed-batch cultures

Once a mathematical model is available, steady-state optimization for a continuous reactor is a simple problem of maximization of a function of many variables. This type of optimization is called *static optimization* as no time trajectories are involved. Optimization of environmental conditions as a function of time in the case of a fed-batch reactor is a much tougher problem. Such optimization is called *dynamic optimization* and it involves maximization of a functional by finding the optimum function. Such problems fall under the realm of variational calculus. The solution requires the use of Pontryagin's maximum principle[36] and singular control theory[37].

Pontryagins's maximum principle does not provide solutions to the problems in which the control variable appears linearly in the differential equations for the state variables. Such problems are called singular problems and the theory which deals with them is called singular control theory. Singular control theory has found many applications in chemical engineering. Siebenthal and Aris[38] discussed the application of the maximum principle to the control of stirred tank reactors. They considered the problem of guiding the reactor to a desired steady state in a time-optimal manner. Since the control variable, the cooling rate of the reactor, appears linearly in the energy balance for the reactor, this is a singular control problem. The general policy obtained consisted of three segments: cooling reactor at maximum rate, zero cooling rate and certain intermediate cooling rates. This intermediate control is nothing but *singular control.*

Gunn[39] considered the problem of bifunctional catalyst for certain reaction networks. The optimization problem was to determine the catalyst composition along the length of the tubular reactor for maximizing the concentration of desired product.

D'Ans *et al.*[40] studied the problem of growth of bacteria in a continuous culture. The objective was to reach the desired steady state starting with an arbitrary initial state and to maximize the amount of cells produced during the transient period, using dilution rate as the control variable. They used a method based on extremization of line integral using Green's theorem. The method is applicable to problems with two dynamic state equations with an objective function expressed as a time integral.

Kelly[41] proposed a transformation technique which reduces the dimensionality of the system and converts a singular problem into a nonsingular problem, one in which the control variable appears nonlinearly. Ohno *et al.*[42] used this technique in solving

the problem of choosing the best operating pattern among batch, fed-batch and continuous fermentation for amino acid production. Constant cellular yield and negligible substrate consumption for product formation were assumed in the models. These simplifying assumptions made it possible to have an analytical solution. The applicability of the method for more complex systems is limited.

Since the linearity of the control variable (the feed rate of the substrate) is the cause of the singular nature of the problem, one approach is to use another control variable which appears nonlinearly in the kinetic model. Yamane *et al.*.[41,42] used the specific growth rate as a control variable. Guthke and Knorre[45] chose the substrate concentration as a control variable in fed-batch fermentations of antibiotics. The main reason behind these transformations is the use of well established numerical methods for solving nonsingular control problems. Once the transformed control variables are known the feed rate can be easily computed. Unfortunately, the physical constraints such as maximum and minimum feed rates and fermentor capacity are totally neglected. Also the feeding policies calculated from the transformed control variables may not be realizable physically.

Weigand *et al.*[46] solved the problem of optimizing fed-batch fermentations for cell mass growth. In the case of constant cellular yield, the singular control maintains the substrate concentration constant at a level which maximizes the specific growth rate. Comparisons were made among the performances of single cycle fed-batch, repeated fed-batch, continuous and batch modes of fermentations. It was concluded that the repeated fed-batch is superior to continuous operation in case of substrate inhibition kinetics. For a process with a constant cellular yield and the specific growth rate in Monod form, the fed-batch culture does not offer any improvement over the batch culture. San and Stephanopoulos[47] reported the effect of time delays on optimal feeding policies. It was observed that the qualitative nature of the optimal policies remains unchanged even in the presence of time delays.

Bonte[48] proposed a unidirectional search technique for solving the singular control problem. This method requires the assumption of the structure† of the control profile. The necessary conditions of optimality of singular control are used to reduce

† The sequence of maximum, minimum and singular arcs.

the search of adjoint variables at the switching point to the singular arc. The switching time from nonsingular to singular arc is an unknown to be determined iteratively to optimize the objective function. Tayeb and Lim[49] used this technique to optimize penicillin fermentations. The entire fermentation period showed two phases, a phase of maximum growth and a phase of penicillin production with very slow cell growth. It was found that the optimum feed rate which maximized the amount of penicillin was different from that which maximized the productivity.

Modak and Lim[50] solved the problem of finding out the optimum feed rate profiles in a fed-batch operation of a bioreactor. They obtained numerical solutions to many systems of interest. The method used involves guessing the general profile of the control variable using certain conjectures. Switching times between different arcs in the profile remain the only unknown parameters. These can be determined by a simple multidimensional search to optimize the objective function. However, the method requires that an explicit form for singular feed rate be available. Bell and Jacobson[51] pointed out that if the optimal control sequence and its properties are known *a priori*, then a computation method which utilizes this information is much more efficient than gradient methods. This is possible, in general, only for systems having less than five state differential equations.

Menawat *et al.*[52] used the ε-algorithm of Jacobson *et al.*[53] for optimizing various fermentations. This numerical approach converts a singular control problem to a nonsingular problem by appending a nonlinear function of the control variable to the objective function. It was found that in spite of some numerical instability, the objective function of the transformed problem compared well with that of the original problem. Stutts[54] investigated the applicability of a modified conjugate gradient algorithm for the optimization of fed-batch cultures. The optimal feed rate profiles for three dimensional problems were very close to that obtained using singular control theory. This technique, however, had convergence problems when applied to higher dimensional problems. Also different initial guesses of the feed profile resulted in different final results.

Very few experimental or theoretical studies of optimization of fed-batch operation for recombinant systems have been reported in literature. Seressiotis and Bailey[55] investigated optimal gene expression and amplification strategies for batch and continuous recombinant cultures. It was realized that the use of regulated promoters and plasmid replication controls amenable to environmental manipulation

offers the opportunity of reconciling the opposing effects of cloned-gene content and expression level on process productivity. A mathematical model developed by Lee and coworkers[32] was used for optimization. The validity of this model has not been verified experimentally as yet. However, since the qualitative features of the model are reasonable, the simulations were expected to show only qualitative trends. It was presumed that in a batch reactor operation at a particular point, say t_{AMP} or t_{EXP}, the amplification and expression processes, respectively, begin. The optimum switching times were determined. As expected, it was found that the amount of product synthesized by the recombinant culture is maximized by an operating strategy that employs low levels of gene expression during an initial growth stage of the batch culture and switches to high-product-synthesis activity at a suitable point during the batch. It was concluded that, to optimize a particular expression system, it is important to know the inhibition relationships among plasmid content, cloned-gene expression and product accumulation, and host cell growth and protein synthesis.

2.6 Mathematical modeling of yeast growth

In order to appreciate the modeling efforts it is useful to understand the metabolic steps involved in the yeast growth metabolism. This subject has been reviewed by Lievense and Lim[56]. Only the key aspects are discussed here.

2.6.1 Yeast growth metabolism.

2.6.1.1 An overview. Three major catabolic pathways are involved in the growth of yeast on glucose. *Fermentation of glucose*, which occurs when the glucose concentration is high (or when oxygen is absent), is relatively inefficient and results in an energy yield of about two ATP per mole of glucose metabolized. Ethanol and carbon dioxide are produced as by-products of this reaction.

$$C_6H_{12}O_6 \rightarrow 2C_2H_5OH + 2CO_2$$

Oxidation of glucose predominates at very low glucose concentrations (50-100 mg/L) in aerobic cultures. Glucose is completely oxidized to CO_2 and 16-18 moles of ATP are generated.

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + H_2O$$

Oxidation of ethanol occurs at very low concentrations of glucose. About 6-11 moles of ATP are formed per mole of ethanol consumed.

$$C_2H_5OH + 3O_2 \rightarrow 2CO_2 + 3H_2O$$

2.6.1.2 Major catabolic pathways. A reasonably complete description of the mechanistic basis for the metabolic behavior described above can be given. The important metabolic pathways are shown schematically in Figs. (2.1-2.3).

Oxidation of glucose. The oxidation of glucose, which is derepressed at low glucose concentrations, is characterized by a very active tricarboxylic acid (TCA) cycle and respiratory chain (55,56). The energy efficiency of glucose oxidation derives from the large number of $NADH_2^+$ produced for each mole of glucose oxidized to carbon dioxide. The reducing power of $NADH_2^+$ drives the formation of ATP in the electron transport chain in the mitochondria. The depletion of TCA cycle intermediates is compensated by the carboxylation of pyruvate which feeds oxaloacetate directly into the cycle.

Oxidation of ethanol. The oxidation of glucose similarly exhibits an active TCA cycle and respiratory chain. In the absence of glucose, polysaccharides must be synthesized from a two-carbon source. This synthesis, called *gluconeogenesis*, consists of the glyoxylate bypass and a reversal of glycolytic pathway.

Fermentation of glucose. The fermentation of glucose is induced in the presence of high glucose concentrations. The reductive pathway from pyruvate to ethanol is then important as it regenerates NAD^+, a necessary cofactor in the glycolysis. Energy is obtained only from the ATP generated during glycolytic steps.

2.6.2 Models of yeast growth. Most of the previous models of yeast cultures have concentrated on describing steady states in continuous fermentations or lag phases in batch cultures. These models, therefore, have a limited applicability in studying the fed-batch cultures.

Yoon *et al.*[57] developed an unstructured model to describe yeast growth on mixed substrates. The model is able to describe growth on glucose and accumulation of ethanol. Although, the diauxic growth is successfully simulated, the model does not take into account any of the metabolic pathways of yeast growth. For this reason the use of such a model is questionable.

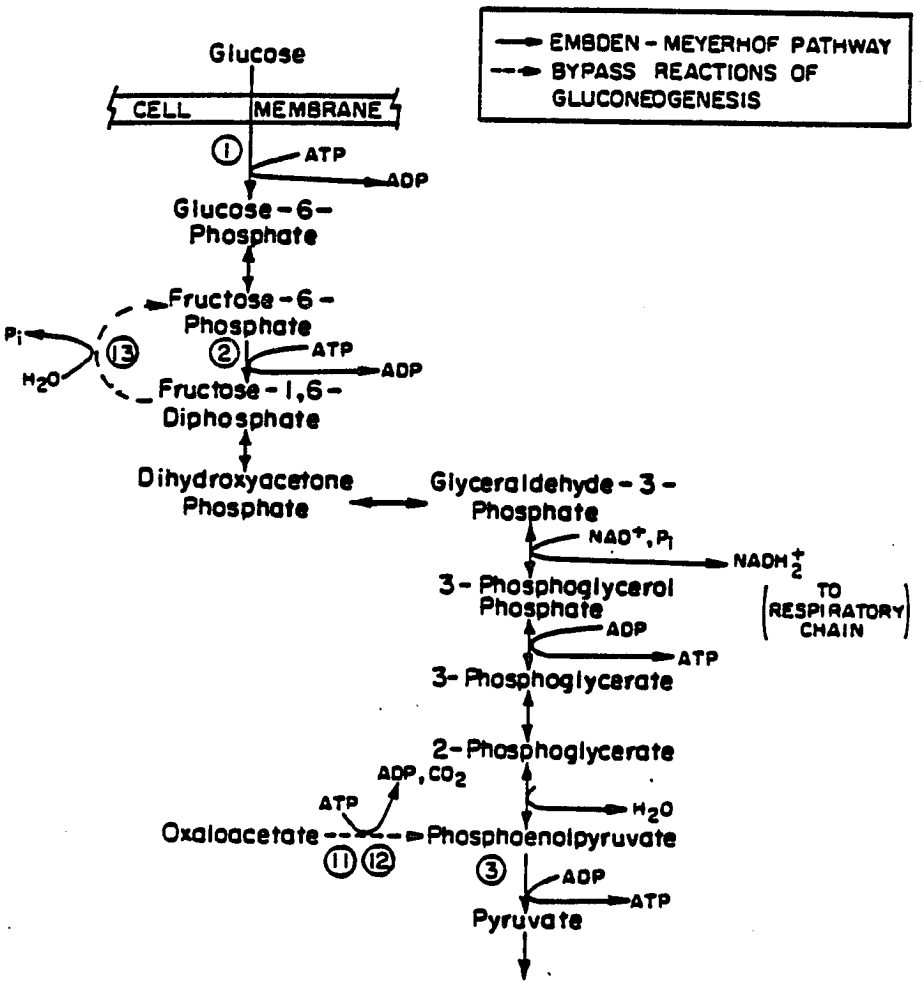

Figure 2.1 Yeast metabolic pathway - I

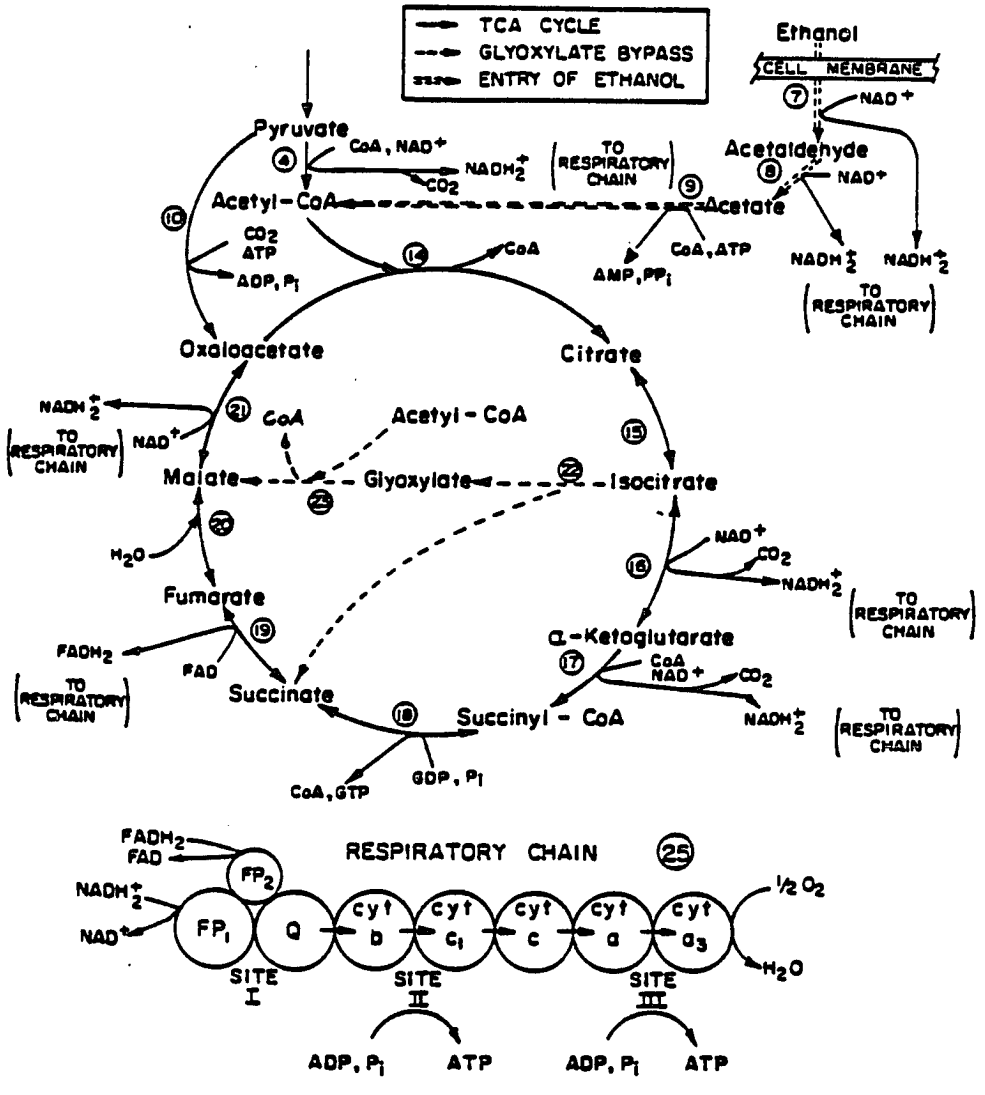

Figure 2.2 Yeast metabolic pathway - II

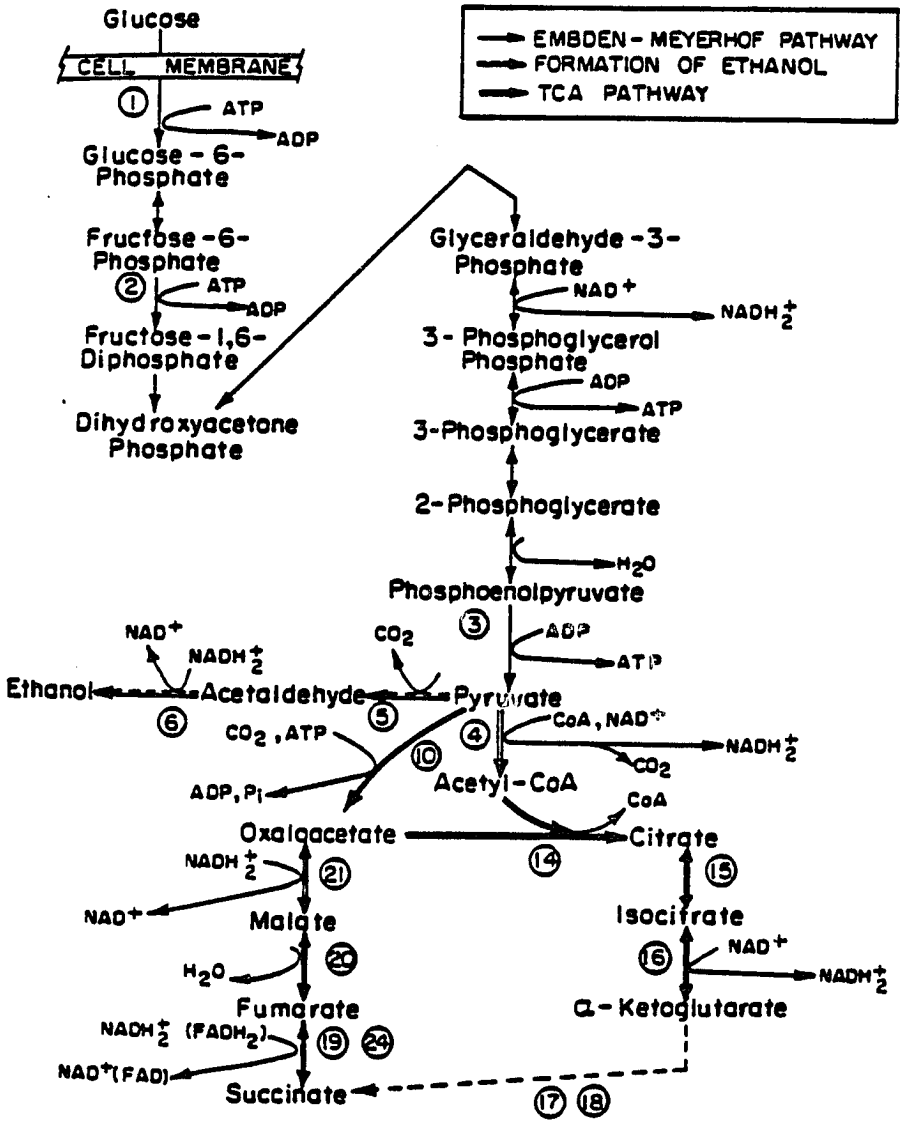

Figure 2.3 Yeast metabolic pathway - III

Bijerk and Hall[58] proposed a mechanistic model of yeast growth assuming the cell mass to consist of two fractions, A and B. 'A' mass was responsible for glucose uptake and energy production while 'B' mass carried out reproduction. Oxidative metabolism was totally neglected. This assumption is not valid in fed-batch cultures. Fukuda *et al.*[59,60] assumed an on-off type control mechanism to be responsible for glucose consumption. The fermentative pathway is turned on and the respiratory pathway shut off if glucose concentration is higher than 0.28 g/lit, and vice versa.

Hall and Barford[61,62], and Bellgardt *et al.*[63] proposed complex models to simulate internal energy metabolism and the cell cycle. The models assumed an elaborate reaction network consisting of the reactions of the TCA cycle. These models have a very large number of variables, most of which cannot be measured easily and hence have a limited value for optimization study.

Toda *et al.*[64] proposed an unstructured model for diauxic growth of the yeast *S. carlsbergensis* in a chemostat. The concentration of glucose in the fermentation medium was assumed to be responsible for controlling the respiratory and fermentative pathways. This empirical model describes the batch and continuous culture behavior quite well. Lievense[65] reported a detailed model for yeast growth, which considers the fermentative and respiratory pathways as controlled by fermentative and respiratory enzyme pools. It can predict batch and continuous behavior. The dynamic responses of a shift up or shift down in the dilution rates were also successfully described. The model, however, is too complex for optimization purposes. Modak[66] showed that under certain assumptions, Lievense's model can be simplified. The resulting model is similar to the Toda model. The new model is simpler but still able to describe fed-batch growth of yeast.

2.7 The plasmid pRB58

Recombinant yeast containing the plasmid pRB58 is used as a model system for this study. Overall structure of this plasmid is explained first, followed by detailed information about its constituents.

2.7.1 Structure of the plasmid. The recombinant plasmid pRB58 was constructed by Carlson and Botstein[67] to study the expression of SUC2 gene in yeast. The plasmid contains the entire yeast SUC2 gene including the SUC2 promoter, signal sequence and structural gene. It contains 2μm origin of replication for efficient maintenance in yeast. The SUC2 gene codes for the enzyme invertase. It also has an ampicillin marker for

selection in *E. coli*, and a URA3 marker gene which enables production of uracil within the yeast cell.

2.7.2 The yeast 2μm plasmid. This naturally occurring yeast plasmid forms the basis for many recombinant plasmids used for transforming yeast. The characteristics of 2μm plasmid have been reviewed by Futcher[68]. Only the main features will be highlighted here.

The 2μm circle is a 6 kbp double-stranded DNA plasmid found in the nucleus of nearly all strains of *S. cerevisiae* at a copy number of about 60. It is stably maintained as an extrachromosomal element through mitotic and meiotic cell divisions[69]. Plasmid replication is very similar to that of chromosomal DNA. The plasmid contains at least four genes and some *cis*-acting sites which provide the plasmid with a mitotic partition system and a copy number amplification system. The plasmid has little or no phenotype.

Its most striking features are two perfect 599 bp inverted repeats. The FLP gene on the plasmid is responsible for recombining the inverted repeats. The REP1 and REP2 genes are both required in *trans* for plasmid stability. One of the *cis*-acting sites is an autonomously replicating sequence (ARS) which is found near the boundary of one of the inverted repeats. The ARS coincides with the plasmid's origin of replication.

Replication of the 2μm plasmid is very similar to the chromosomal replication. It occurs only during the S phase of the growth cycle. Furthermore, like chromosomal DNA and unlike mitochondrial DNA or bacterial plasmids, 2μm circle molecules, there is exactly one initiation of replication per cell cycle. In spite of this the molecule can be replicated several times per cycle (probably through a double rolling circle model[70]). A copy number amplification mechanism is responsible for a tight control over the plasmid copy number.

2.7.3 Invertase expression.

2.7.3.1 Invertase enzyme. The enzyme invertase (β-D-fructofuranosidase fructohydrolase EC 3.2.1.26) catalyzes the hydrolysis of sucrose to glucose and fructose. Yeast strains containing any of the SUC1-7 genes produce this enzyme. The yeast SUC2 strains produce both an internal nonglycosylated form of invertase as well as a secreted glycosylated form of the enzyme[67].

The external, glycosylated invertase has a total weight of approximately 270,000 daltons, about half of which is carbohydrate[71]. Trimble and Maley[72] found that the external enzyme is composed of two identical 60000 dalton subunits to each of which is added an average of nine neutral oligosaccharide chains consisting of a di-N-acetylchitobiosyl core and 26 to 54 mannose residues.

Gascon *et al.*[73] determined the kinetic properties of both internal and external invertase. The K_m values for purified internal and external invertase are almost identical and were found to be 25 mM and 26 mM, respectively. The pH-activity curves of the two enzymes are almost identical and shows a maximum around pH 4-5. However, the pH-stability of the two enzymes differs considerably. The external invertase is stable at 30 °C between pH 3 and 7.5. The internal invertase, on the contrary, is stable from pH 6 to 9. The difference in stability can probably be attributed to the stabilizing effect of the carbohydrate chains present in the external form of the enzyme.

2.7.3.2 The SUC2 gene. The regulation of SUC2 gene expression in yeast is complex and only understood partially. The net effect of several regulatory genes is repression of invertase production at high glucose concentrations in the medium. SUC genes are each structural genes for both forms of invertase discussed above. The level of the secreted glycosylated form is regulated by glucose repression. The glycosylated form accounts for most of the invertase activity in derepressed cells. The intracellular invertase is made at low levels, which do not change significantly with changes in glucose concentration.

The SUC2 gene has been cloned[67] and its nucleotide sequence determined[74]. Carlson and Botstein[75] found that the mechanism of regulation of the amount of invertase mRNA is insensitive to the number of copies of the structural gene within the cell. They also proposed a model for SUC2 gene expression. According to this model, a 1.9 kb mRNA encodes the entire sequence of the precursor of secreted invertase. Translation of this mRNA begins with a methionine codon at the beginning of the secretion signal sequence. The 1.8 kb mRNA begins within the signal sequence region and lacks the start of the translation codon used for the 1.9 kb mRNA. Translation of the shorter mRNA begins at the next initiation codon. The two enzymes are found in different cell compartments simply because one has a signal sequence while the other does not. The mechanism is shown schematically in Fig. (2.4). The glucose regulation was explained by assuming that two promoter sequences are present; one sequence does

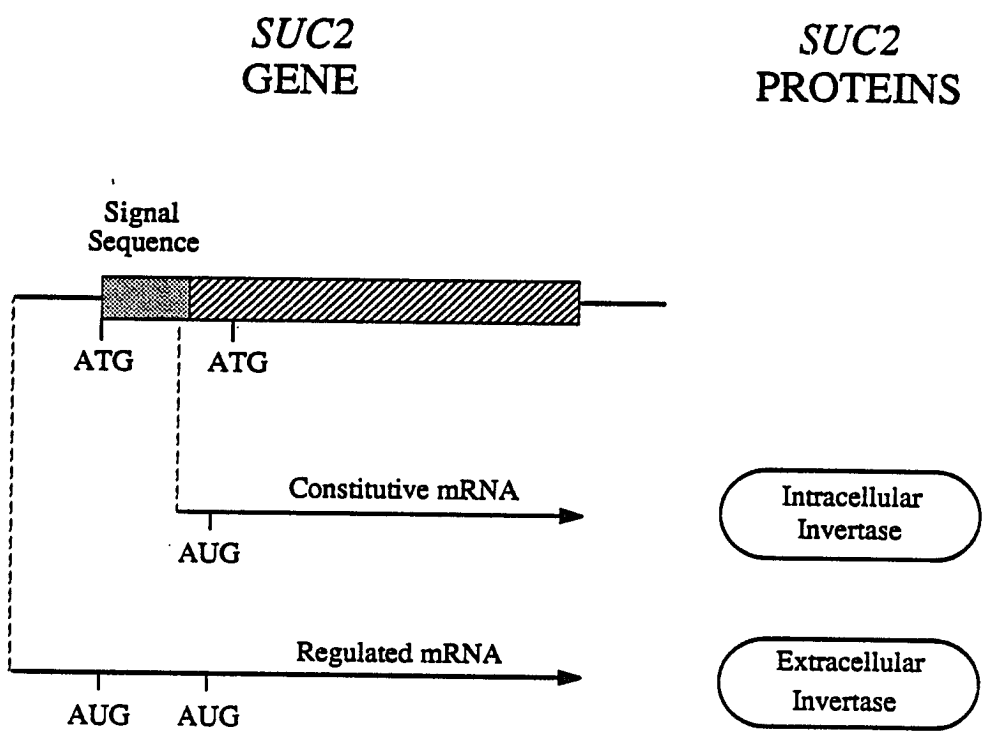

Figure 2.4 Transcription of yeast SUC2 gene.

not to respond to any signal related to glucose concentration and results in the production of the smaller mRNA, while the other promoter responds to glucose regulation and results in the synthesis of the longer mRNA.

Carlson *et al.*[75] translated the SUC2 nucleotide sequence into the corresponding amino acid sequence. There results corroborate the model of differential regulation described above. At the amino terminus end of invertase, the amino acid sequence translated from the DNA sequence includes a core of hydrophobic residues and strongly resembles prokaryotic and eukaryotic signal peptides.

Sarokin and Carlson[76] constructed a series of deletions in the 5′ noncoding region of SUC2 gene. Analysis of the effects of these deletions identified an upstream region required for derepression of secreted invertase synthesis. The 3′ boundary of this region is near -418. The 5′ boundary is about 100 base pairs upstream. No essential sequences lie between the upstream region and the TATA box at -133. The data suggest that the regulation of SUC2 gene expression operates by positive control. None of the deletion mutations caused a high-level constitutive phenotype that would be expected from the deletion of a negative regulatory site.

Sarokin and Carlson[77] demonstrated that the upstream region between -650 and -418 can confer glucose-repressible expression to a heterologous gene, a LEU2-*lacZ* gene fusion, that is not normally regulated by glucose repression. This confirms that the upstream region is responsible for regulation. Also, the regulation is at the level of transcription. They found that the SUC2 upstream region was active in the inverted orientation also. This suggests that the region resembles in some respects the enhancer elements found in higher eukaryotes.

Further investigation by Sarokin and Carlson[78] revealed that tandem copies of a 32-bp sequence from the upstream regulatory region activate expression of LEU2-*lacZ* fusion. The level of expression increased with the number of copies of this element, but was independent of their orientation. This activation was not significantly glucose repressible. The 32-bp sequence includes a 7-bp motif that is repeated at five sites within the upstream regulatory region. It was speculated that although the upstream regulatory region acts positively to derepress gene expression, the regulatory response is directly controlled by a negative *trans*-acting factor.

Hohmann and Gozalbo[79] compared the upstream sequences of the SUC, MAL and GAL genes, all of which are subject to glucose repression. Within the upstream regions of all SUC genes three regions with palindromic sequences analogous to stem and loop structures were identified. Comparable structures were also found in MAL and GAL gene upstream sequences. Palindromes are possible protein recognition sites on the DNA sequence.

Regulation of SUC2 gene expression is further complicated by the fact that it is under the control of catabolite repression. Catabolite repression, or glucose repression, is a global regulation of the genes controlling the metabolism of many carbon sources. Celenza and Carlson[80] cloned the gene SNF1 (sucrose nonfermenting). A functional SNF1 gene product is required to derepress expression of many glucose-repressible genes including SUC2. The evidence suggests that the SNF1 gene product acts as a positive activator to derepress glucose-repressible genes in response to low external glucose concentration. Investigations by Celenza and Carlson[81] revealed that the level of a 2.4-kilobase polyadenylated RNA encoded by the SNF1 gene is not regulated by glucose repression. Thus expression of the SNF1 gene is not glucose repressible. Neigeborn and Carlson[82] deduced from mutation studies that six *trans*-acting genes are required for derepression of SUC2 in response to glucose deprivation. These genes were designated SNF1 to SNF6.

Hexokinase II, in addition to its glycolytic activity, has been implicated in the control of catabolite repression. Ma and Botstein[83] constructed null mutations in both hexokinase genes, HXK1 and HXK2, and studied their effect on catabolite repression of three different genes in yeasts: SUC2, CYC1, and GAL10. Their results indicate that the null phenotype of the HXK1 gene is not much different from that of the wild type, but the null mutants of the HXK2 gene fail to show catabolite repression in all three systems. Loss of hexokinase II caused a 100-fold increase in the expression of both the SUC2 and CYC1 genes under repressive conditions. It appears that hexokinase II acts in a negative fashion in catabolite repression. Also, hexokinase II functions earlier than the SNF1 gene product in the regulatory pathway.

Neigeborn *et al.*[84] cloned SSN20 gene. They showed that mutations at the SSN20 locus partially alleviated the requirement for SUC2 upstream regulatory sequences that are normally essential for derepression of secreted invertase. The TATA box was, however, still required. It was speculated that the SSN20 gene product mediates interaction between factors associated with upstream activating sequences and

factors acting at downstream promoter sequences.

Schultz and Carlson[85] cloned and analyzed the SSN6 gene. An *ssn6* null mutation was an effective suppressor of *snf1* with respect to restoration of SUC2 expression. This also causes glucose insensitive expression of SUC2 mRNA. Several possible mechanisms were proposed to explain this behavior.

In conclusion, one can see that regulation of SUC2 gene expression is a complex process involving many genes (SNF1-6, SSN20, SSN6, HXK2). The exact nature of the interactions between these genes still remains unclear. The mechanism, however, seems to consist of a hierarchy of controls bearing similarity to the control of the GAL system and the general amino acid control in yeast.

3. MATERIALS AND METHODS

3.1 Model system

The host organism used in this study is *Saccharomyces cerevisiae* SEY2102 (MATα *ura*3-52 *leu*2-3,-112 *his*4-519). This strain is an α-haploid. It is auxotrophic for uracil, leucine and histidine, and hence will not grow in media devoid of these constituents. Furthermore, it has a deletion (*suc2*-Δ9) in the chromosomal copy of the SUC2 gene, which codes for invertase. It does not contain any other unlinked invertase structural genes (SUC1, SUC3-7). Thus the host strain cannot produce invertase.

The yeast 2μm based plasmid pRB58 was constructed by Carlson and Botstein[67], and introduced into the host SEY2102 by Emr *et al.*[86]. It contains the entire yeast SUC2 gene including the promoter, signal sequence and structural gene. The plasmid also contains the URA3 gene which serves as a marker by complementing the auxotrophic requirement for uracil. Both the host and the recombinant strains were stored in 15% glycerol at −20°C. A schematic of the plasmid pRB58 is shown in Fig. (3.1).

3.2 Fermentation medium

Most of the fermentations were performed using a synthetic minimal medium (SDc)[87] containing salts, amino acids, trace elements, vitamins, nitrogen sources, and glucose (Table 3.1). The medium was supplemented with 5 g/L casamino acids (Difco, Detroit, Michigan). Different amounts of glucose were used for different experiments. In the initial stages of the research a few preliminary experiments were done using a different growth medium (A1) (Table 3.2).

The fermentor, containing distilled water, was autoclaved at 121°C and a concentrated medium was added through a 0.2 μm filter (Gelman Sciences, Inc., Ann Arbor, Michigan) and added to the fermentor to make up the desired glucose concentration.

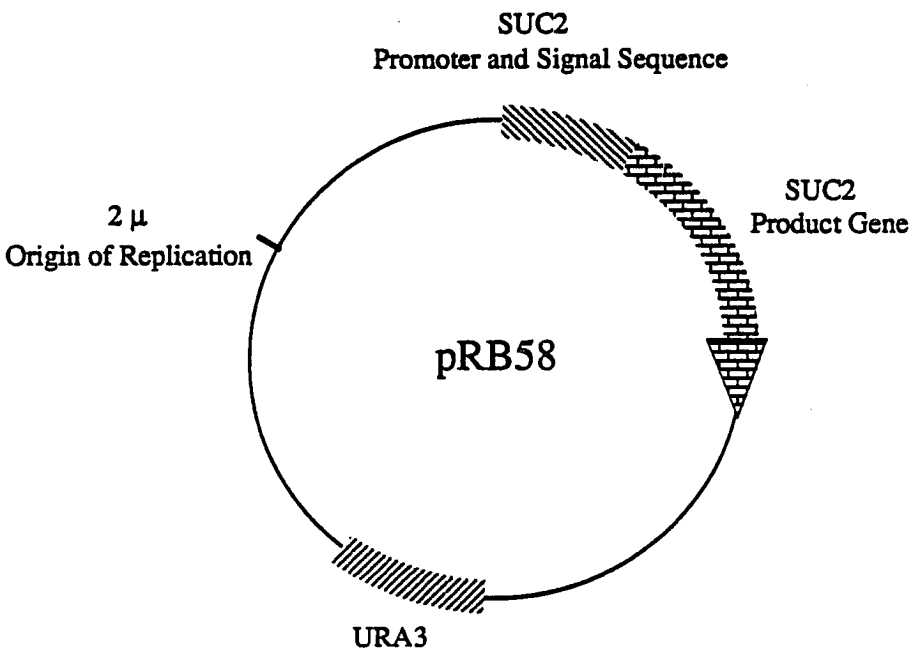

Figure 3.1 A schematic of the plasmid pRB58

Table 3.1 Composition of synthetic minimal medium (SDc)

Component	Medium Concentration (g/L)
d-Glucose	2.0-20.0
Bacto-yeast nitrogen base w/o amino acids	6.66
Casamino acids	5.0

Table 3.2 Composition of medium A1

Component	Medium Concentration (g/L)
d-Glucose	2.0-20.0
Bacto-yeast nitrogen base w/o amino acids	6.66
Leucine	3.06×10^{-2}
Histidine	2.04×10^{-2}

3.3 Analytical methods

3.3.1 Cell mass. Off-line measurements of cell mass concentration were carried out with a Baush and Lomb Spectronic 70 spectrophotometer at 600 nm wavelength. Marten[88] found that the cell mass was linearly proportional to the optical density (OD) at 600 nm below 0.5 OD. To eliminate the nonlinearity at higher OD, samples were diluted to a final OD below 0.4. Corrected OD was computed as the product of the dilution factor and the OD of the diluted sample. Cell dry weight was measured for a few samples in some of the fed-batch fermentations. During the fermentation 20-40 mL. samples were withdrawn and filtered with a preweighed 0.2 μm Gelman filter. The pellet was washed with 25 mL distilled water. The filter was dried for 24 hours in an oven, maintained at 120°C, and weighed. The conversion factor between corrected OD and dry cell weight was found to be 0.30 g/(L.OD).

3.3.2 Glucose assay. Samples were collected every 60 to 90 minutes and centrifuged in a microcentrifuge (Eppendorf centrifuge 5415 C). The cell free medium was frozen at –20°C for glucose and ethanol analysis. The glucose concentration was determined either by using a colorimetric technique based on the enzymatic reaction involving glucose oxidase and o-dianisidine as an indicator, or by using a glucose analyzer (Yellow Spring Instruments, Model 27). The enzymatic test kit is available commercially from Sigma Chemicals, St. Louis. The lowest glucose level that can be measured with sufficient accuracy with both these methods is about 10 mg/L. Appendix A describes the steps involved in these methods.

3.3.3 Ethanol determination. The ethanol concentrations were determined with a Varian 3700 gas chromatograph. The chromatograph column was a 5 ft. long, 8 inch stainless steel tubing packed with Porapak Q and conditioned at 250°C. Helium was used as a carrier gas at a flow rate of 30 mL/min. A flame ionization detector was used with hydrogen fed at 30 mL/min and air at 300 mL/min. During analysis, the injector and detector were at a temperature of 160°C, and the column was kept at 130°C. Each sample was injected twice (5 μL per injection) to get reproducible results. An ethanol standard was injected after every four or five samples. Under these conditions ethanol had a retention time of approximately 2 minutes. In some fermentations some unknown compound, with a retention time of about 7-9 minutes, was observed.

In the initial stages of the research the output of the GC was recorded on a Varian model A-25 chart recorder with an integrator. Calculating the area under the peaks using this method is quite time-consuming. As part of the research, the gas chromatograph was interfaced† with a Zenith personal computer. DAS16 (Metrabyte Corporation, Taunton, MA), a multifunction analog/digital input/output board, was used as part of the interface. It is a full length board that installs internally in an expansion slot of the computer. It uses a 12 bit successive approximation converter with a 12 microsecond conversion time. The channel input configuration is switch selectable on the board, providing a choice between 16 single ended channels or 8 differential channels. The configuration was set to 8 differential channels for the GC interface.

The A/D conversion was initiated with software commands. A simple BASIC program was written to facilitate data acquisition from the chromatograph. The board comes with some utility subroutines which greatly simplify the programming. The BASIC program is simple, user-friendly and menu-driven. It performs the following functions.

- It collects and stores chromatograph output at a frequency of once every second,
- shows the results as a output vs. time plot,
- calculates area under the ethanol peak,
- allows interactive adjustment of base line, starting and stopping points of integration, and
- monitors the output of the chromatograph without data storage (helpful during startup of the chromatograph).

A complete listing of the program is included in Appendix B.

3.3.4 Cell homogenization. Since most of the invertase resides within the periplasmic space, it is necessary to break the cell wall to perform invertase assay. This was accomplished by mechanically breaking the cells with glass beads in a cell homogenizer (B. Braun, West Germany). Liquid carbon dioxide was used as a coolant to maintain

† The help of Richard Lowe, an electronics technician in the School of Chemical Engineering, was invaluable.

the sample at 4°C. From protein assays on the homogenized samples it was determined that homogenizing the cells for 1.5 minutes was enough to break the cell wall completely. All the fermentation samples were homogenized for 2 minutes. The details of this method are listed in Appendix C.

3.3.5 Invertase assay. A modified two step assay method previously described by Goldstein and Lampen[89] was used to determine invertase activity. The assay was based on the hydrolysis of sucrose to glucose and fructose catalyzed by invertase. The enzyme was incubated with sucrose solution at 30°C and pH 4.9 for 10 minutes. The reaction was stopped by adding phosphate buffer (pH 7.0) and boiling for 3 minutes. The glucose generated was determined as described in section 3.3.2. Since sucrose hydrolyzes at a slower rate at this pH even in the absence of invertase, a blank was used to subtract this effect. Also for reproducibility sucrose solutions were prepared immediately before the assay. One unit of invertase activity was defined as the amount of enzyme which hydrolyzed sucrose to produce 1 μmole of glucose per minute at 30°C and pH 4.9. Appendix D describes the procedure in greater detail.

3.3.6 Plasmid stability. The fraction of plasmid containing cells was determined for a few batch and fed-batch fermentations. Samples from fermentation broth were diluted to about 1000 cells/mL using aseptic techniques. About 100 μL of this diluted sample was spread on a non-selective (containing uracil) agar plate. Colonies are formed after incubation at 30°C for about 24 hours. More than 100 of these colonies were transferred to a selective plate (one without uracil) by using sterile toothpicks. Only plasmid containing cells can grow on the selective plate and the required fraction can be easily calculated from the number of colonies transferred and the number of colonies grown on the selective plate.

3.4 Experimental apparatus and procedure

3.4.1 Inoculum. A vial of yeast, frozen in glycerol, was transferred to the cold room and allowed to warm up to 4°C for several hours. The vial was then transferred to room temperature until it warms up to the room temperature. About 1 mL of thawed culture was used to inoculate 10 mL of SDc media (20 g/L glucose) in a culture tube. The culture was allowed to grow in an incubator-shaker maintained at 30°C for 24 hours. About 1 mL of this overnight culture was used to inoculate a 50 mL of SDc medium (20 g/L) contained in a 250 mL Erlenmeyer flask the night before the experiment. This culture was grown to a late exponential phase and about 20-30 mL of the grown culture was used to inoculate the fermentor.

3.4.2 Fermentor setup and operation. A fully automated fermentor (Mouse, Queue Systems, Parkersburg, WVa.) was used for fermentation purposes. It allows direct and/or cascade control of many variables including pH, impeller speed, dissolved oxygen, temperature, air flow rate and substrate feed rate. Changing of the set points and control constants is done in a straightforward manner using on-screen menus.

Before the fermentation began, the pH probe (Ingold Electrodes Inc., Wilmington, Mass.) was calibrated using standard pH buffers. The fermentor was then filled with 500 mL of distilled water, assembled and autoclaved at 121°C for 20 minutes. The concentrated medium was fed in to make up the volume to the desired level. For batch experiments the starting volume was around 1.2 liters, while for fed-batch experiments it was kept at about 0.6-0.7 liters. The feed pump was also calibrated before the fermentation.

The dissolved oxygen (DO) was measured with a polarographic DO probe (Ingold Electrodes Inc., Wilmington, Mass.). This probe was calibrated when the temperature reached 30°C and the pH stabilized at 5.5. Nitrogen was sparged through the vessel with the impeller speed at around 400 rpm. When the DO probe reading was almost constant, this DO was defined as 0%. The flow of nitrogen was stopped, air flow was started at 5 L/min and the impeller speed was set to 800 rpm. After the DO reading stabilized this was defined as 100% DO.

During the fermentation the DO was maintained at 90% by adjusting airflow rate. The impeller speed was kept constant at 200 rpm until it was no longer possible to keep the DO at 90%. Then the impeller speed was gradually increased. In batch fermentations at higher cell densities it was difficult to keep the DO at this high level. Then the set point was lowered to 80%. In all the fermentations, the pH was controlled at 5.5 by adding 1.0 M acid (sulfuric or phosphoric) and 1.0 M ammonium hydroxide through a 0.2 μm sterile filter. Samples were taken through a sample tube.

4. EXPERIMENTAL RESULTS

4.1 Introduction

Experimental characterization of the model system is a necessary prerequisite for model development. Since the ultimate aim is to determine optimal glucose feed rate profiles in a fed-batch mode of operation, fed-batch experiments are essential. At the beginning, however, simple batch experiments were conducted to get some idea about the dynamics of cell growth and product formation. These were followed by fed-batch experiments using different feeding strategies.

4.2 Preliminary Results

To characterize the yeast strain, simple batch fermentations were done in shake flasks using medium A1†. The specific growth rates observed during exponential growth on glucose are listed in Table (4.1). The host strain (without the recombinant plasmid) grows about 60% faster than the recombinant strain under these conditions.

It was soon realized that the specific growth rate (0.25 hr^{-1}) on glucose for the recombinant yeast is much slower than that reported for wild type yeast strains (0.4-0.45 hr^{-1}). The growth rate on ethanol was also very small. This is probably because the amount of leucine and histidine used was not sufficient to sustain growth after the glucose phase. It was, therefore, decided to use a richer medium with 5 g/L casamino acids instead of adding leucine and histidine. Use of this medium results in a much higher growth rate on glucose and ethanol. All the experiments described in the following sections were done in an SDc medium (Table 3.1).

† The composition of this medium is listed in Table (3.2). It contains leucine and histidine, but does not contain casamino acids.

Table 4.1 Specific growth rates during preliminary experiments

μ_-	0.39 hr^{-1}
$\mu_+(1 - \theta)$	0.25 hr^{-1}

4.3 Batch fermentations

Aerobic batch fermentations were carried out for different initial glucose concentrations. The results are plotted in Figs. (4.1-4.5). Similar to wild type yeast, the recombinant yeast underwent a diauxic growth. Initially the glucose concentration was high, and after a short lag period the cell, by using glucose as the carbon source, grew rapidly in the first exponential phase. A specific growth rate of about 0.43-0.45 hr^{-1} was attained. Because the glucose concentration in the medium was high the cells utilized the fermentative pathway of growth. This resulted in production of ethanol. A higher initial glucose concentration resulted in a higher amount of ethanol. Thus during the first phase of growth glucose is consumed while ethanol is produced. When glucose is totally consumed, the cells, in the second phase of growth, utilize ethanol as the carbon source.

The invertase activities in Figs. (4.1-4.11) are normalized by a measure of the amount of cell mass (OD.mL) in the culture broth. For acceptable accuracy in the invertase assay a minimum OD of about 0.5 is required. If the cell concentration is below this level the pellet obtained after centrifuging is very small and difficult to handle. Thus for the first few samples in the fermentation invertase assay was not performed.

As noted before, the expression of the yeast SUC2 gene is repressed at high glucose concentrations. The batch fermentations show that the same is true for cloned SUC2 gene in the plasmid pRB58. At the beginning of the fermentation when the glucose concentration was kept high the specific invertase activity (SIA) was very low. Near the end of the first exponential phase, when glucose concentration dropped to a very low value, SIA increased indicating derepression of SUC2 promoter. All batch experiments show similar qualitative trends in invertase activity. However, experiments B1 and B2 show somewhat lower SIA at the end of fermentation. These were among

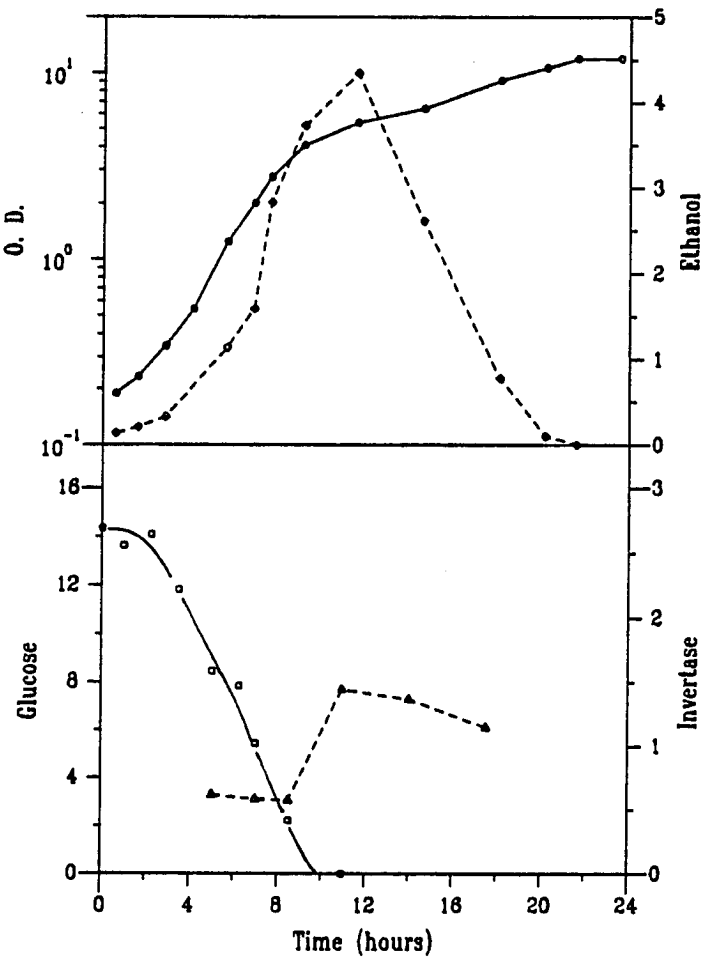

Figure 4.1 Batch fermentation with recombinant strain, B1

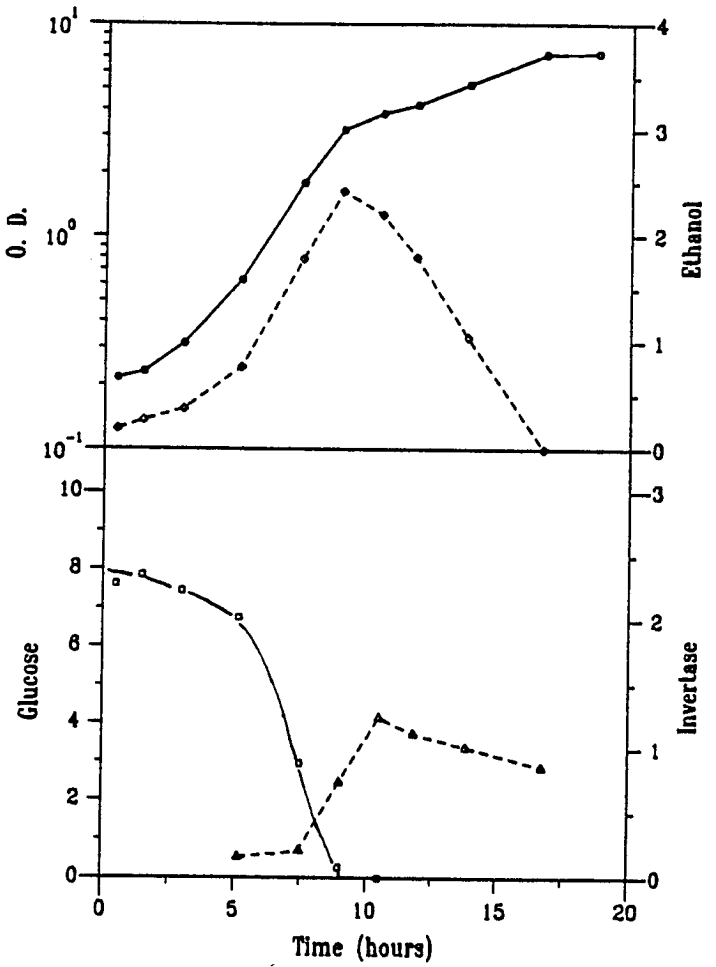

Figure 4.2 Batch fermentation with recombinant strain, B2

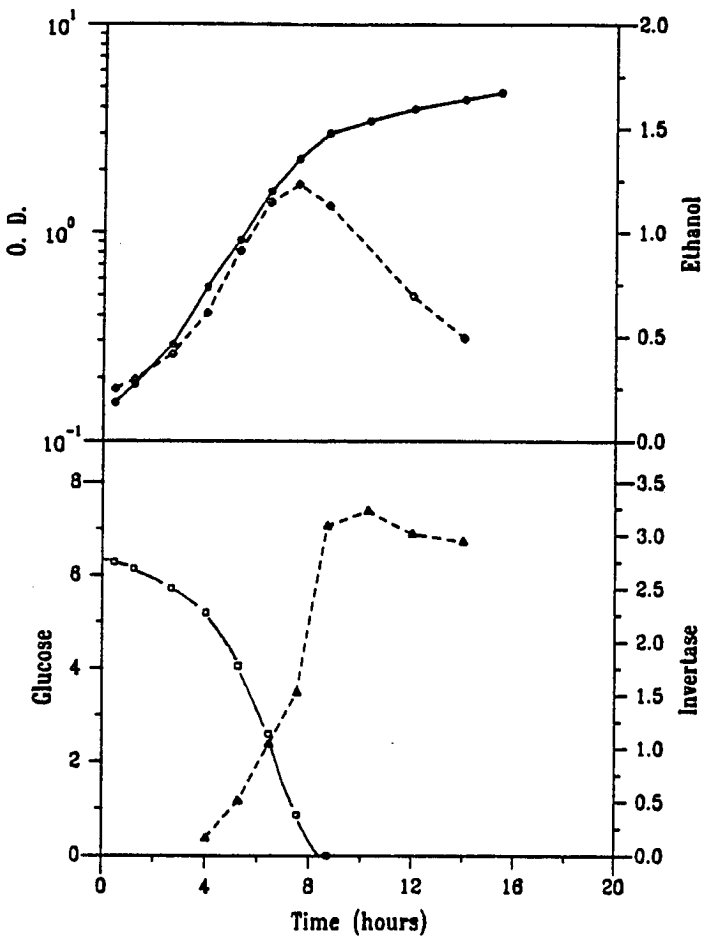

Figure 4.3 Batch fermentation with recombinant strain, B3

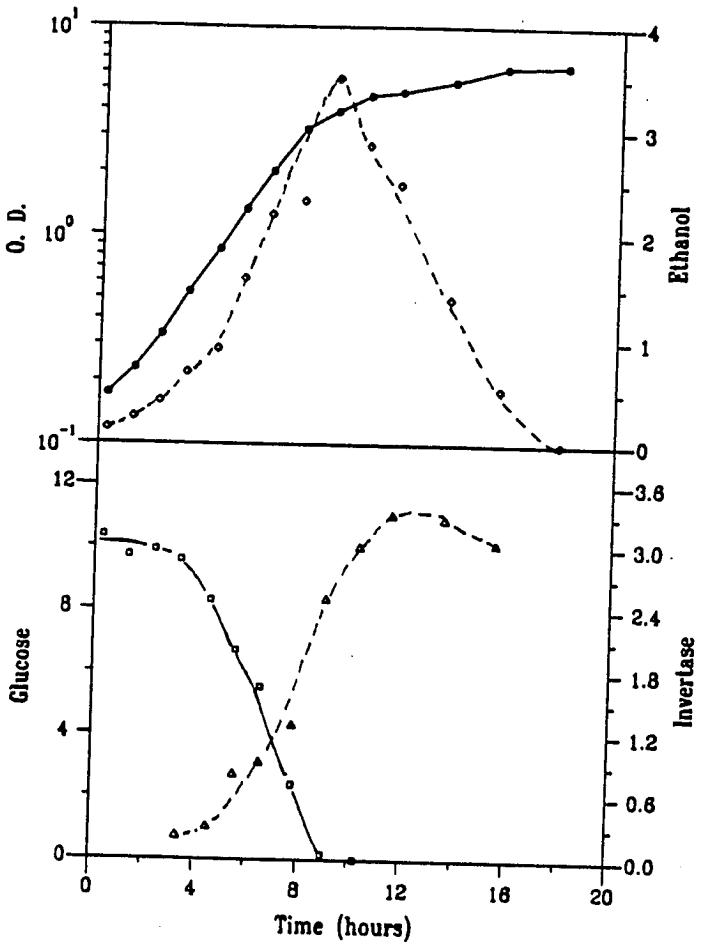

Figure 4.4 Batch fermentation with recombinant strain, B4

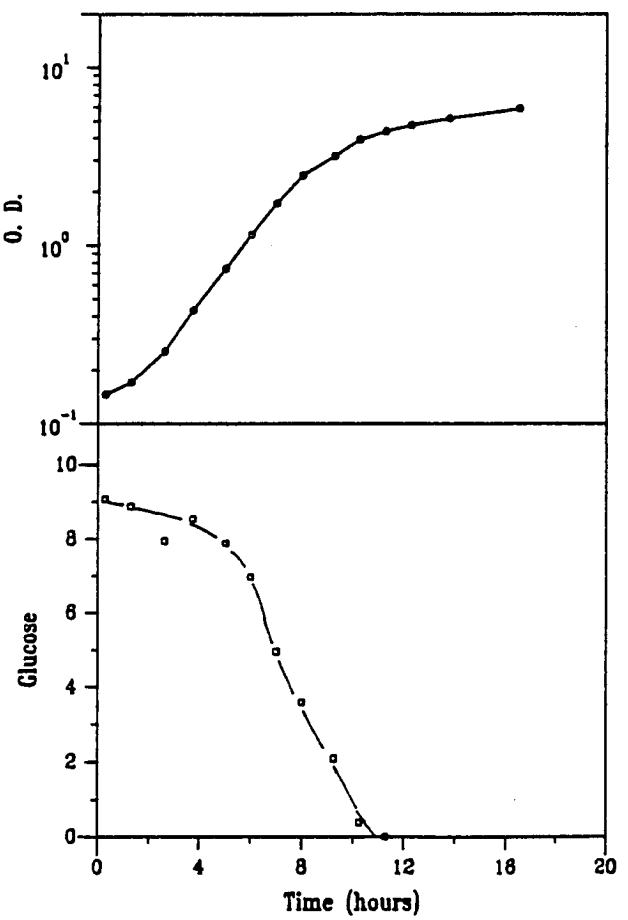

Figure 4.5 Batch fermentation with host strain, B5

the first fermentations performed with a new fermentor and the best control constants† were not determined till then. The medium pH kept changing between 5.2 and 5.8. This, apparently, did not affect cell growth and ethanol formation, but probably did affect SUC2 gene expression. Experiments B3 and B4 showed significantly higher SIA at the end of fermentation.

The fraction of plasmid free cells (F) was determined for a few samples in experiments B1 and B2. The samples were taken from different phases of cell growth - exponential phase of growth on glucose, end of glucose growth phase and exponential phase of growth on ethanol. For all the samples the fraction F was between 0.07 and 0.11. Considering the experimental error involved (estimated ±5%), the fraction F can be considered to be almost constant throughout the fermentation.

One batch fermentation was performed using the host strain (without the plasmid). The cell growth results were similar to the recombinant strain. The strain exhibited diauxic growth and the specific growth rates for both exponential phases were almost the same as those for the recombinant strain. Of course, because of the absence of SUC2 gene, it did not produce any invertase.

Although batch experiments exhibit interesting features like diauxic growth and invertase derepression, they give very little information about the behavior of the system at low glucose concentrations. At the end of first exponential phase the glucose concentration drops so fast that it is very difficult to get more than one or two samples below a concentration of, say, 2 g/L. The derepression of SUC2 promoter is known to occur at these glucose concentrations, inaccessible in the batch mode. Also, the yeast switches from fermentative to respiratory growth metabolism at low glucose concentrations. Thus, important features of cell growth and product formation remain hidden if only batch fermentations are performed.

4.4 Fed-batch experiments

Fed-batch fermentations, if used with proper feed rates, can yield long periods of relatively low glucose concentrations. Also the effect of feed switches (which is

† The proportional, derivative and integral constants of the controller can be set at the fermentor console.

characteristic of optimum feed profiles in singular systems) can be studied in a convenient way. The results of the fed-batch fermentations are shown in Figures (4.6-4.11).

In all the fed-batch experiments, the fermentation was started in a batch mode until the glucose concentration dropped to a desired value. In experiments F1, F2 and F3 the glucose concentration was monitored by analyzing the samples with a colorimetric assay described before. This procedure, however, took at least one hour and it was difficult to estimate the exact time to start the glucose feed. During experiments F4, F5 and F6 the glucose concentration was determined with a glucose analyzer which reduced the assay time to about two minutes.

In the first fed-batch experiment, the aim was to start the feed when the glucose concentration dropped below 0.5 g/L. However, because of the lag involved between sampling and the assay, the glucose concentration dropped to zero before the feed was started. This seemed to adversely affect invertase production. Even though glucose feed was started, causing an increase in the glucose concentration, the SIA remained almost constant for about four hours before it picked up. At the end of the fermentation SIA suddenly shot up. It is not clear whether this was as a result of very low glucose concentration or an experimental error.

The second fed-batch experiment showed an interesting behavior. In this fermentation, the glucose concentration was allowed to fall to about 2 g/L. The feed rate was then periodically increased so that the glucose concentration remained almost constant. During the batch growth as the glucose concentration dropped SIA started increasing slowly. During the fed-batch period, however, it remained almost constant. It appears that glucose concentration has a very tight control on invertase production. Further experiments confirmed this speculation.

In the next experiment, the feed rate was adjusted such that the glucose level initially increased and dropped towards the end of the fermentation. This resulted in expected variation in SIA. The interesting fact, however, is that the response of the invertase production machinery to changes in the glucose level seems almost immediate (at least within the limits of experimental observation). The correspondence between glucose and SIA is so good that the two profiles look almost symmetrical.

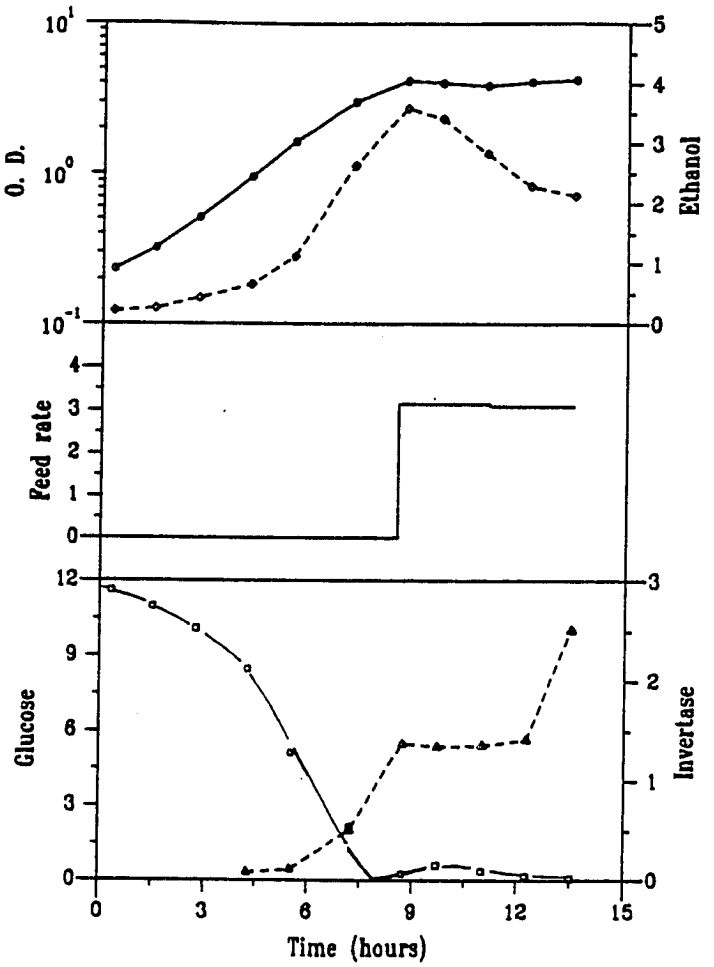

Figure 4.6 Fed-batch fermentation with recombinant strain, F1

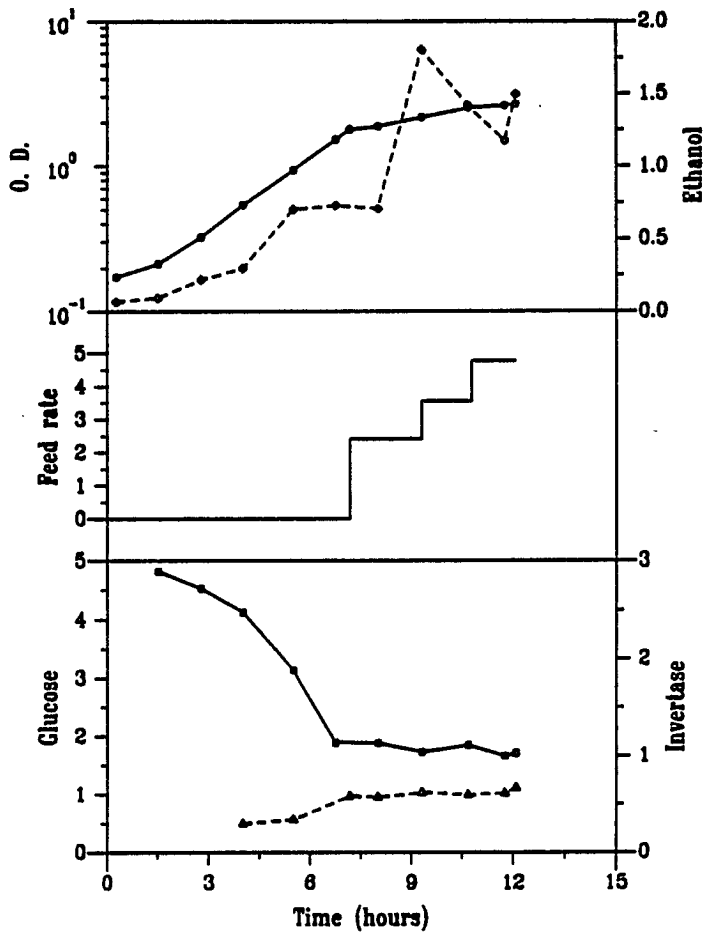

Figure 4.7 Fed-batch fermentation with recombinant strain, F2

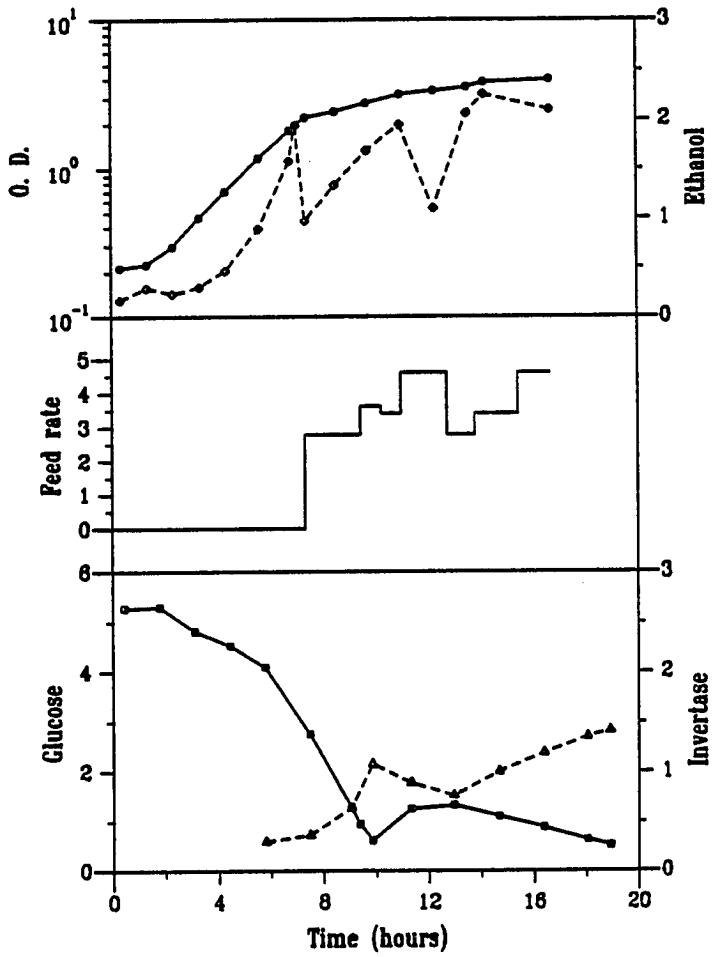

Figure 4.8 Fed-batch fermentation with recombinant strain, F3

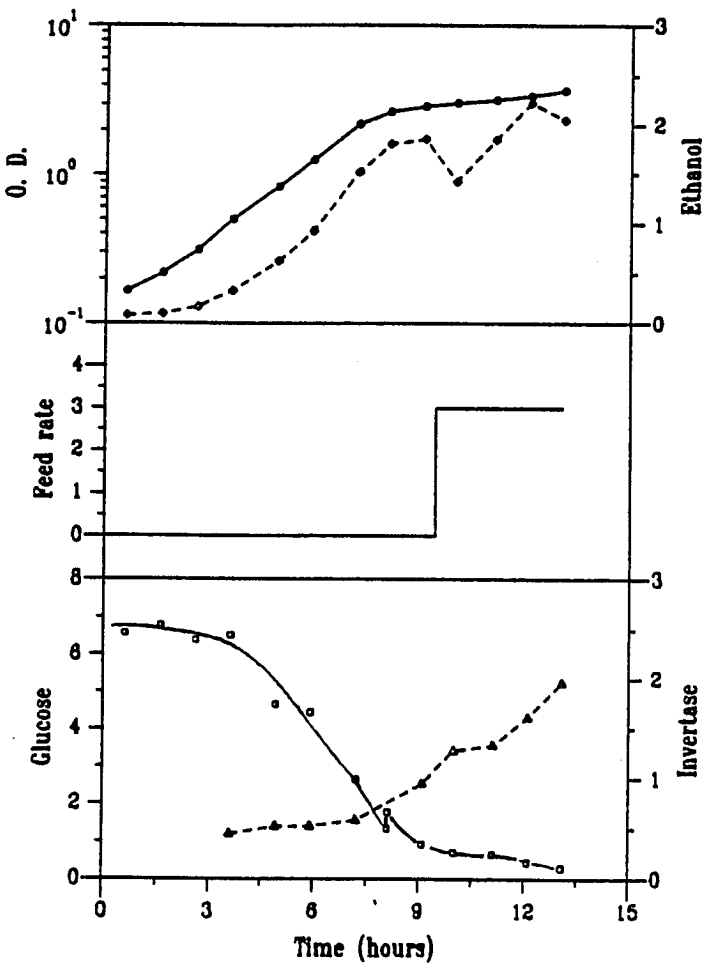

Figure 4.9 Fed-batch fermentation with recombinant strain, F4

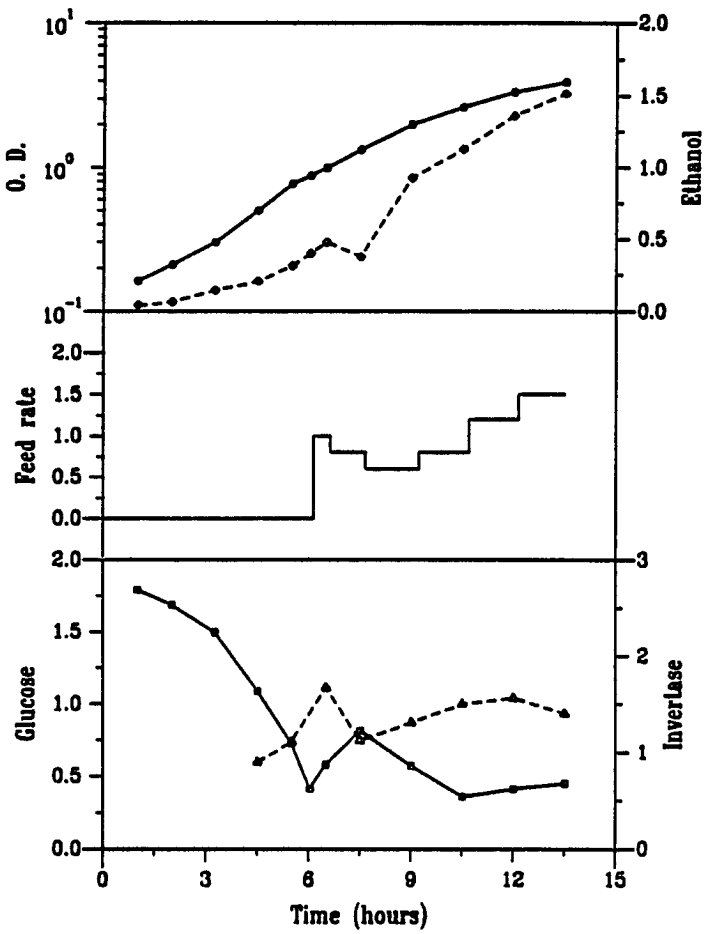

Figure 4.10 Fed-batch fermentation with recombinant strain, F5

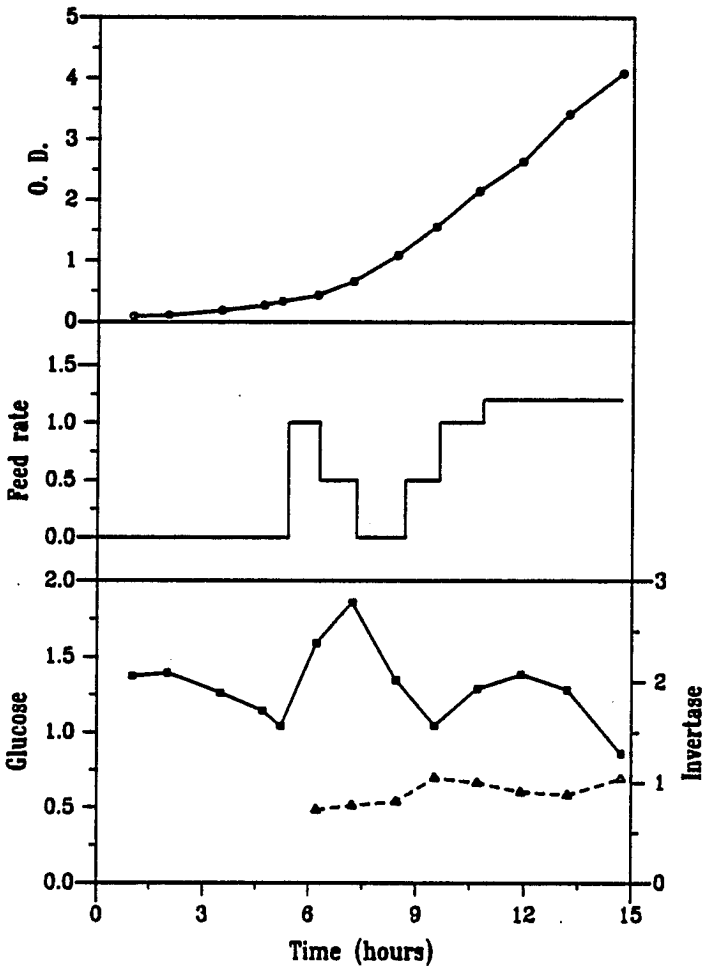

Figure 4.11 Fed-batch fermentation with recombinant strain, F6

The fermentations F3, F4 and F5 used different feeding strategies to generate different glucose profiles. In all these experiments the SIA responded very quickly to changes in glucose concentration. In experiment F6 the glucose concentration was made to oscillate. However, above a concentration of about 1.5 g/L changes in glucose levels did not affect SIA significantly.

4.5 Summary

The batch fermentations exhibited diauxic growth, first on glucose and then on ethanol. They also showed derepression of SUC2 promoter as the glucose concentration dropped. The fraction of plasmid containing cells remained almost constant during the entire fermentation. Fed-batch fermentations demonstrated the immediate response of invertase production to changes in the glucose level. The SIA appeared to be a direct one-to-one function of the glucose concentration in the medium. Also, a brief period of zero glucose concentration resulted in a constant low SIA for a long period of time, even though the glucose levels were raised subsequently.

5. MODEL DEVELOPMENT

5.1 Introduction

The fed-batch experiments described in the last chapter set the stage for development of a mathematical model to describe the system behavior. These fermentations involved shifts in glucose feed rates resulting in considerable variations in glucose concentration in the medium. Structured models have been used in past to describe such transient biological situations. However, these models, of necessity, introduce differential balance equations for intracellular constituents. This increases the dimensionality of the system as far as system optimization is concerned. As pointed out before, current numerical dynamic optimization techniques do not work well when the system dimension† is more than four (or at the most five). Thus, although the structured models may describe the system better, unstructured modeling approach was chosen in view of the final research objective. The model, however, should not be too simplistic. Yeast shows a shift from fermentative to respiratory growth metabolism at low glucose concentrations. Also, there seems to be some sort of inherent feedback control of SIA within the organism. These factors must be incorporated in the model. To reiterate, the model has to be simple enough to be optimized, but detailed enough to describe the key features of the system. Thus the first, if not the foremost, requirement on the model is:

I. The model should be described by five or less differential equations.

---

† The system dimension is the number of differential equations needed to describe the transient behavior of the system.

From the batch experiments it is clear that yeast grows at a much slower rate on ethanol than on glucose. Moreover, SIA almost stays constant in the ethanol growth phase. The fed-batch experiment F1 demonstrated that even short periods of zero glucose concentration can lead to constant SIA. Thus for maximum invertase productivity periods of zero glucose concentration are to be avoided. This slackens the requirements on the model in that the model need not describe the kinetics of cell growth and product formation during ethanol growth phase, because it will never be encountered under optimum conditions.

II. The model should correctly describe the system kinetics during growth on glucose, but need not explain kinetics during growth on ethanol.

Ethanol, at the concentrations encountered in the experiments, does not affect cell growth or invertase production rates. Thus, the optimum feed profiles will be independent of ethanol concentration in the medium.

III. Kinetics of ethanol production and consumption is irrelevant to the objective at hand.

5.2 Analysis of cell growth

To determine the functional dependence of specific growth rate and substrate uptake rate on the state variables (like glucose concentration), it is helpful to extract rate information from the observed kinetics. To do this, the cell mass and glucose vs. time data was fit with cubic splines using the IMSL routine CSAKM†. The time derivatives were then calculated using the IMSL routine CSDER. The cell yield on glucose can then be computed as,

$$Y_{xs} = -\frac{(dX/dt)}{(dS/dt)} \tag{5.1}$$

for batch fermentations, and

† The splines try to follow the data points regardless of the scatter. To avoid this, fictitious points were introduced and the goodness of the fit was judged visually.

$$Y_{xs} = -\frac{\frac{d}{dt}(XV)}{FS_F - \frac{d}{dt}(SV)} = -\frac{FX + V\frac{dX}{dt}}{F(S_F - S) - V\frac{dS}{dt}} \qquad (5.2)$$

for fed-batch fermentations. In the above equations Y_{xs} is instantaneous cell yield on glucose, X is corrected optical density, S is glucose concentration in the medium, t is time, V is culture volume, F is glucose feed rate and S_F is glucose concentration in the feed.

The results of these calculations are shown in Fig. (5.1). Lag phase points are not shown in the plot. Despite the scatter in the plot, some trends are evident. For glucose concentrations greater than about 1.5 g/L, the cell yield is relatively constant. In this concentration range the less energy efficient fermentative growth metabolism predominates and the cell yield is understandably low. A large part of the consumed glucose is directed towards ethanol production. On the contrary, at low glucose concentrations the cell yield rapidly increases. This is because under these conditions the respiratory pathway becomes dominant. This shift from fermentative to respiratory metabolism, characterized by a drastic increase in cell yield, must be incorporated as an integral part of the cell growth model. This was accomplished by using a goal oriented approach. Before describing the details of the model, though, it is interesting to examine two previous modeling efforts.

5.2.1 Lievense's model of yeast growth: Lievense[65] developed a detailed structured model (Model L) that accounts for the catabolic pathways involved in the utilization of glucose, production of ethanol and genetic level controls (induction-repression) of the synthesis of enzymes involved in these pathways.

Lievense viewed cell growth as a two step process. The first step, catabolism, converts the nutrients to the precursors and energy needed for growth. The second step, anabolism, synthesizes the biomass utilizing the precursors and the energy generates as a result of the first step. The catabolism and anabolism were assumed to be uncoupled to simulate unbalanced growth conditions. The specific growth rate and the specific enzyme synthesis rate lag behind their target values which depend on the environmental conditions during changes in dilution rate in continuous culture experiments.

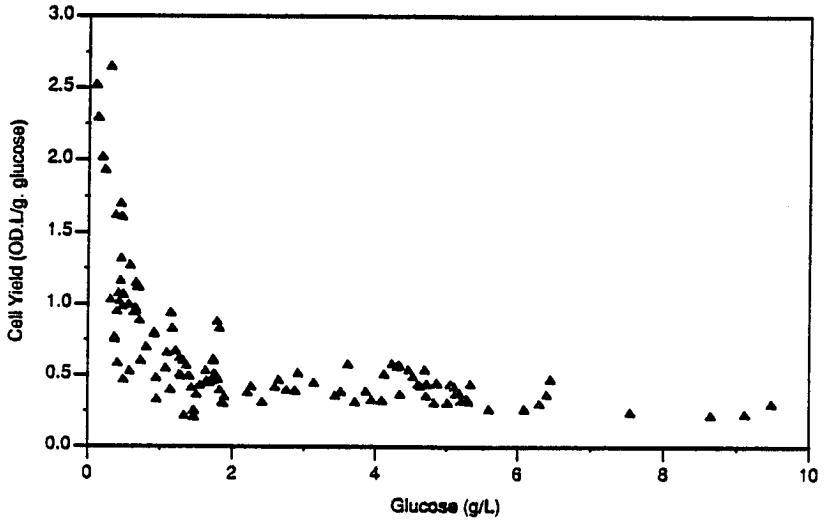

Figure 5.1 Variation of cell yield with glucose concentration.

The catabolite repression phenomenon was explained by postulating fermentative and respiratory enzyme pools which control the glucose utilization through fermentative or respiratory pathway. The synthesis of these enzymes is subject to induction and repression by glucose fluxes. The target enzyme activities were assumed to be arbitrary functions of glucose concentration. This model was quite successful in explaining a wide variety of transient situations.

5.2.2 Modak's model of yeast growth: Modak[66], by making some simplifying assumptions, converted Lievense's model to a simple unstructured model (model M). Anabolism and catabolism were treated as coupled processes, resulting in balanced growth conditions. As opposed to Lievense's approach of using regulatable enzyme pools, Modak used a direct approach. The reactions expressing the fermentative and respiratory pathways were viewed as parallel reactions in which the cell mass is competing for the same substrate, glucose.

The fraction R of glucose that is channeled through the fermentative pathway was assumed to be controlled by the glucose concentration in the medium. The functional dependence of this fraction on glucose was postulated of the following form.

$$R = \frac{1 + m_1 S^n}{m_2 + m_3 S^n} \tag{5.3}$$

Because glucose utilization in yeast occurs through two catabolic pathways branching from a common substrate, glucose, the specific growth rate μ is expressed as a sum of two Monod-type functions.

$$\mu = \frac{\mu^1_{max} S}{m_4 + S} + \frac{\mu^2_{max} S}{m_5 + S}, \tag{5.4}$$

where, m_i, n, μ^1_{max} and μ^2_{max} are model constants, μ is specific growth rate.

5.3 A goal-oriented model of yeast cell growth

5.3.1 The functional form: Although Lievense's model is quite accurate in describing many transient situations, it incorporates differential balances on enzyme pools. This leads to six differential mass balance equations and three differential equations for specific synthesis rates to describe fed-batch behavior. This model is way too complex for current numerical optimization techniques. Modak's model uses only four differential equations to describe cell growth. However, it uses an arbitrary functional form for the dependence of fraction of glucose fermented on glucose concentration. It

was felt that a simple but elegant approach can be used for modeling purposes (Model P).

One of the first experimental observations was that the fraction of plasmid-free cells does not change with time. This probably demonstrates the effectiveness of the selective medium used for the fermentations. Thus the composition of the cell population, with respect to the number of plasmid-free and plasmid-containing cells, is same throughout the fermentation. This leads to the following postulate.

P1. The cell mass can be described by a single lumped variable X identical to the corrected optical density described in Chapter 3.

Pascual *et al*[90] conducted a systematic study of the effect of ethanol on glucose transport, key glycolytic enzymes, and proton extrusion in *S. cerevisiae*. It was found that the inhibition of fermentation by ethanol is of a noncompetitive type, with a K_i value of 0.8 M (36.8 g/L). The glucose transport rate was also found to be inhibited by ethanol in a noncompetitive manner, although the K_i value was almost 10 times higher. Thus at the ethanol concentrations (maximum 3 g/L) encountered during the fermentations, described in Chapter 4, it is unlikely that ethanol will affect the glucose uptake rate (R_t) or specific growth rate (μ).

P2. Specific glucose uptake rate and specific cell growth rate are independent of ethanol concentration in the medium.

The growth characteristics of yeast are strongly affected by the regulation of the alternative pathways for energy production. The regulation of the flow in the pathways is described by Crabtree effect. Thus, the most important part of formulation of model for yeast cell growth is an adequate description of this effect. Some comments are in order here.

It was originally believed that the Crabtree effect is the result of repression of respiratory enzymes by the glucose molecule (or a catabolite derived from glucose)[109-111]. However, there is strong evidence that a high catabolic flux is the direct cause of the inhibition of respiration, with concentration of the carbon source playing a secondary role [110,112-114]. This led Lievense to propose that synthesis of respiratory and fermentative enzymes was subject to induction and repression by glucose fluxes.

In a similar vein, the approach taken here concentrates on glucose fluxes rather than glucose concentration *per se*. For modeling purpose, the complex metabolic pathways can be reduced to bare essentials considering only the respiratory (R_r) and fermentative (R_f) fluxes (Fig. 5.2). The glucose consumed by the cell first enters the glycolytic chain. There is a bifurcation point at pyruvate. Using the fermentative pathway, pyruvate can be converted to ethanol. Alternately, it can be carboxylated to oxaloacetate thus entering the tricaboxylic acid (TCA) cycle. At this bifurcation point, the cell has to make a decision about the fraction of pyruvate that should be channeled into the fermentative pathway.

It was decided to use a goal-oriented approach rather than using an empirical functional form similar to the one used by Modak. A cybernetic (goal-oriented) approach has been used successfully in the past by Dhurjati *et al* to model microbial growth[91]. The cells were viewed to be optimal control systems involved in the maximization of a performance index (e.g. cell growth rate). The key difference between our approach and Dhurjati's approach is that, he postulated certain key enzymes which the cell can control to achieve its objective, while no such enzymes are used in model P.

If maintenance requirements are neglected, all the glucose will be utilized for growth purposes. In this case, we can assume that the performance index to be maximized is the cell growth rate.

P3a. The objective of the multitude of reactions within the cell is to maximize the cell growth rate.

The energy efficiency of the two alternate metabolic pathways can be compared by comparing their theoretical ATP yield. These calculations can easily be performed by looking at individual reactions in the pathways. If all the glucose were to be diverted towards fermentation 5 ATP molecules will be generated for each molecule of glucose consumed. On the other hand, if all the glucose consumed were channeled through respiratory pathway the energy yield will be 30 ATP molecules per molecule of glucose. Since ATP is needed for most of the biosynthetic pathways, it follows that a higher ATP yield translates to a higher cell mass yield. Thus, respiratory pathway will result in higher cell mass yield. For a given substrate uptake rate this is equivalent to a higher specific growth rate. These considerations along with postulate **P3a** lead to the following result.

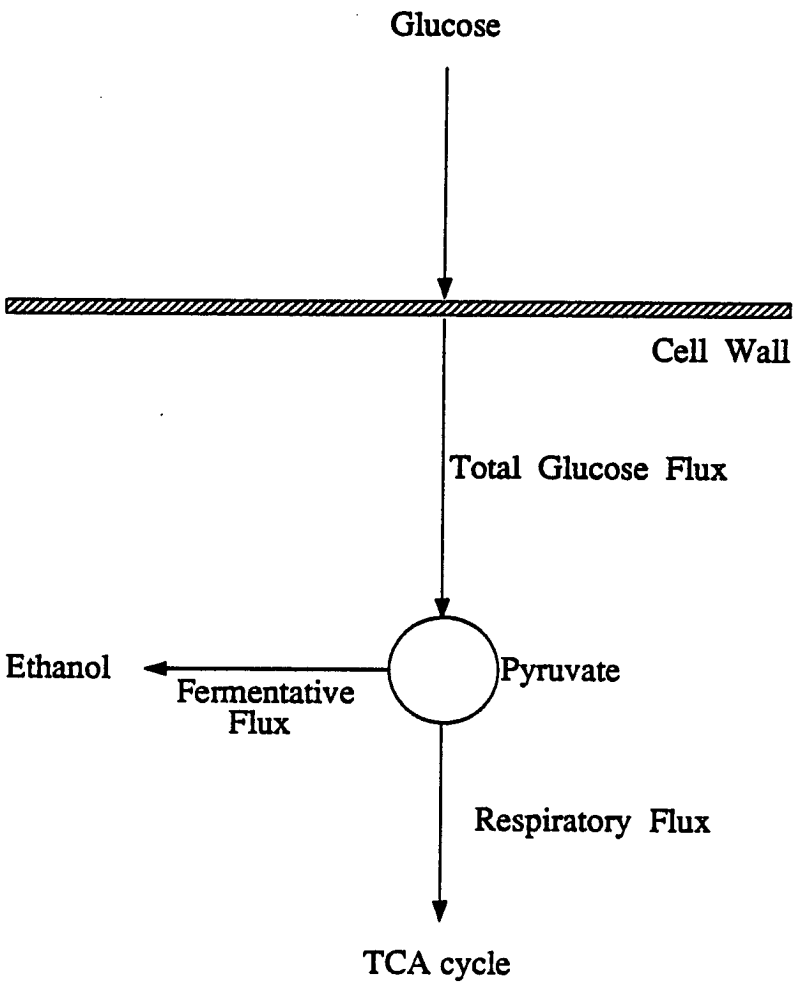

Figure 5.2 Schematic of respiratory and fermentative fluxes.

P3b. The objective of the cell is to maximize the flux through the respiratory pathway. Still some questions remain unanswered. If respiratory pathway is more efficient, why doesn't the cell, as an optimal strategist, always use respiratory pathway? At higher glucose concentrations the cell uses fermentative pathway which results in lower cell yield. To explain this apparent dilemma, one can observe that the cell needs many enzymes, cofactors and other biomolecules for sustaining the TCA cycle and electron transport chain. It is entirely possible that at high enough respiratory fluxes one or more links in the respiratory pathway becomes saturated. Similar saturation behavior is expected for glucose uptake rate. This leads us to the next postulate.

P4. The respiratory pathway gets saturated at glucose levels lower than those required for saturating glucose uptake rate. Also the saturated respiratory flux is less than the saturated glucose uptake rate.

Fig. (5.3) shows a possible functional dependence that can give rise to such behavior. Now if at high glucose fluxes the respiratory pathway is saturated, the only way the cell can satisfy postulate **P3a** is by diverting the remaining pyruvate towards ethanol production.

This provides a clear explanation of the switching between respiratory and fermentative metabolism. Thus, at low glucose concentrations the respiratory pathway is not saturated and all the pyruvate is channeled into the TCA cycle. No ethanol is formed under these conditions. At higher glucose concentrations, however, the saturation of respiratory pathway causes a part of pyruvate is converted to ethanol. As glucose concentration increases the respiratory flux remains almost constant. Thus the increased glucose uptake rate is reflected in increased fermentative flux leading to increased ethanol production rate.

Although the model framework explains the qualitative features of yeast cell growth, specific functional forms have to be used to test its quantitative capability. To begin a simple Monod-type saturation form is used for both the glucose uptake rate and the respiratory flux.

$$R_t = \frac{k_t S}{K_t + S}, \tag{5.5}$$

$$R_r = \frac{k_r S}{K_r + S}, \tag{5.6}$$

where k_t is maximum glucose uptake rate, k_r is maximum respiratory flux, K_t and K_r

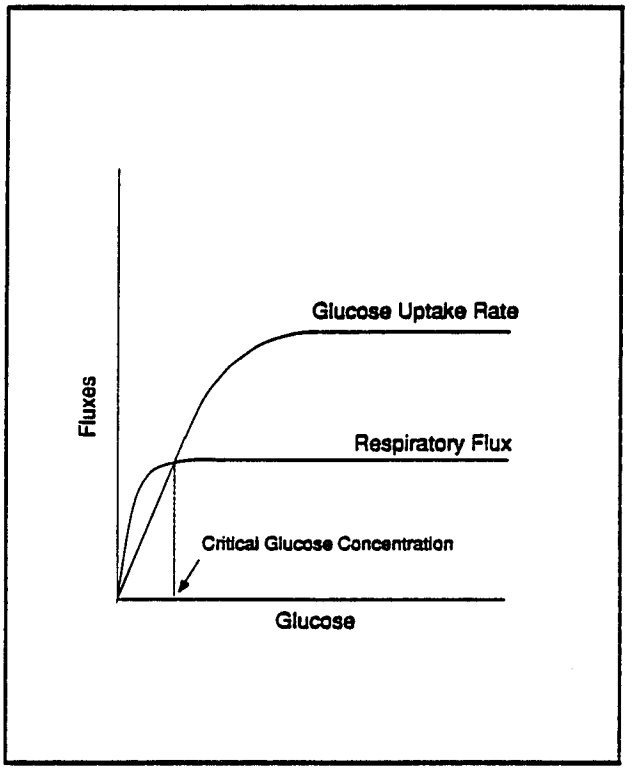

Figure 5.3 Functional dependence of respiratory and fermentative fluxes on glucose concentration.

are model parameters. In compliance with postulate **P4**, k_r has to be necessarily less than k_t. The functional form in Eq. (5.6) is valid only when the uptake rate R_t is sufficient to sustain the flux R_r. If not,

$$R_r = R_t = \frac{k_t S}{K_t + S} \tag{5.7}$$

The fermentative flux is assumed to be given by,

$$R_f = R_t - R_r \tag{5.8}$$

The logical dependence of various fluxes is shown schematically in Fig. (5.4). The specific growth rate can now be determined by assuming that the cell mass yield is constant for both the respiratory and fermentative fluxes.

$$\mu = R_r Y_{xr} + R_f Y_{xf} \tag{5.9}$$

At first glance, it might appear that this functional form (Eq. 5.9) is same as the sum of two Monod-type expressions proposed by Modak (Eq. 5.4). However, this is true only when the respiratory pathway is saturated. Below critical glucose concentration (see Fig. 5.3), however, the fermentative flux is zero and the similarity ends. The other key difference is in the causal relationship between substrate uptake and cell growth. Modak assumed a functional dependence for specific cell growth rate and calculated the substrate uptake rate based on a variable yield. The variable yield was assumed to be a function of the fraction R of glucose fermented. On the contrary, the model proposed here maintains that substrate uptake rate is the controlling factor. The distribution of substrate and the resulting variation in cell mass yield are dependent only on the substrate flux.

Based on the above analysis, the complete set of differential equations needed to describe fed-batch fermentations can be written down.

$$\frac{d}{dt}(SV) = FS_F - R_t XV \tag{5.10}$$

$$\frac{d}{dt}(XV) = \mu XV = (R_r Y_{xr} + R_f Y_{xf}) XV \tag{5.11}$$

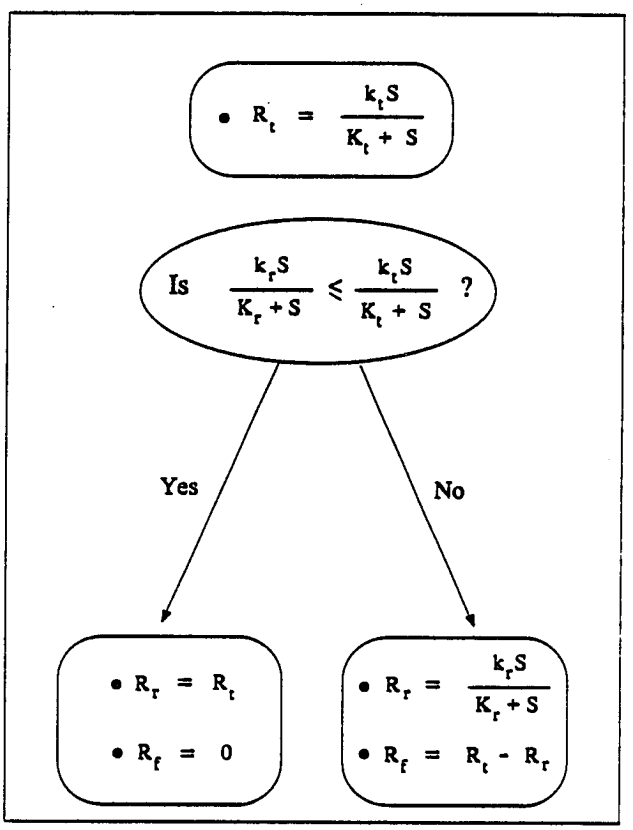

Figure 5.4 Logical dependence of metabolic fluxes.

$$\frac{dV}{dt} = F \tag{5.12}$$

5.3.2 Parameter estimation and results. In the cell growth model there are six unknown parameters and it is virtually impossible to estimate them using standard least square type methods. Thus, it is important to restrict the search set by some analysis of the experimental data.

The derivatives obtained by spline fitting (as discussed in section 5.2) were used to estimate R_t using Eq. (6.10a). There was a lot of scatter in the resulting information. However, this data served to limit the search of k_t (1.0-1.5 (g. glucose)/L) and K_t (0.8-1.2 g/L). The variation of cell yield (calculated by spline fits) with glucose (Fig. 5.1) gave some more clues. When the respiratory pathway is saturated, the overall cell yield, as predicted by the model is given by,

$$Y_{xs} = \frac{k_r (K_t + S)}{k_t (K_r + S)} (Y_{xr} - Y_{xf}) + Y_{xf}. \tag{5.13}$$

In the limit of very high glucose concentrations, when the above equation is definitely valid (within the model framework), the yield becomes independent of glucose concentration.

$$\lim_{S \to \infty} Y_{xs} = Y_\infty = \frac{k_r}{k_t} (Y_{xr} - Y_{xf}) + Y_{xf} \tag{5.14}$$

As noted before (section 5.2) such behavior was actually observed and the limiting yield Y_∞ can be assumed to be between 0.3 and 0.4 (od.lit/g). Since Eq. (6.14) is completely in terms of model parameters, this serves to limit the search set further.

Fig. (5.1) gives us some more information. One can observe that the yield starts to increase quite drastically below a glucose concentration of 0.8 g/L. This was interpreted as shift to respiratory metabolism. Thus as a good approximation we can assume that the critical glucose concentration is 0.8 g/L. Since this is the intersecting point of the curves R_t vs. S and R_r vs. S, the parameter search space is further restricted to values satisfying this condition.

From these observations, a method of choosing the parameters was adopted. To start with, values of k_t and K_t were assumed within accepted ranges. Then k_r and K_r were chosen so that the resultant critical glucose concentration is close to 0.8 g/L. The yields, which satisfied the limiting condition at high glucose concentration, were then

assumed. These parameters were then tested by running simulations and comparing the results with experimental data.

To simplify the model testing, a simple simulation program was written using the IMSL routine IVPAG (a differential equation solver) and the CRC graphics package on the Engineering Computer Network. It reads the parameter guesses, integrates the differential equations and displays the plots of resulting state variable profiles along with observed experimental data. A listing of the simulation program is included in Appendix E. The final parameter values are listed in Table (5.1). Figs. (5.5-5.15) compare the results of model simulation with experimental data. Figs. (5.5-5.9) show the comparison for batch fermentations while Figs. (5.10-5.15) illustrate the results for fed-batch fermentations.

Eqs. (5.10-5.12) were used to simulate cell growth on glucose. An extended model (described in Appendix F) was used to simulate the ethanol phase results shown in Figs. (5.5-5.9). The model as developed in the section (5.3.1) can not describe cell growth using ethanol as substrate. However, as explained before, modeling of formation and consumption of ethanol is not necessary for optimization purposes. No special attempt was made to find the best parameters to simulate the growth phase on ethanol.

Results for experiments B9-B12 show that the model simulations agree very well with the experimental results. Because the model is not structured, it is unable to predict the lag period before starting to grow on glucose. The glucose data of B7 had some scatter and the model does not perform that well. Model predictions for glucose consumption rate are a little lower and cell growth rate a little higher than the experimentally observed values. The agreement is still quite satisfactory.

The real strength of the model, however, lies in the description of glucose and cell mass profiles for a variety of glucose feeding strategies in fed-batch experiments. In the first fed-batch experiment (F1), the glucose concentration dropped to zero before the feed was started. Thus for a short while the yeast cell was synthesizing enzymes needed for growth on ethanol. This causes structural changes within the cell which can be fully accounted for only in a structured model. Thus the proposed unstructured model, in trying to fit the whole glucose profile, does not do a good job of fitting the initial stages of glucose consumption. The aim in experiment F2 was to keep the glucose level constant during the fed-batch period by increasing the glucose feed rate at discrete intervals of time. The model does very good job of simulating this behavior.

Table 5.1 Model parameters

k_t	1.25	L/(hr.OD)
K_t	0.95	(g glucose)/L
k_r	0.55	L/(hr.OD)
K_r	0.05	(g glucose)/L
Y_{rx}	0.60	OD/(g glucose)
Y_{fx}	0.15	OD/(g glucose)
k_p	6.2	units/(g glucose.OD.hr)
k_p	0.1	(g glucose)/L
K_i	2.0	L/(g glucose)
k_d	1.85	/hr

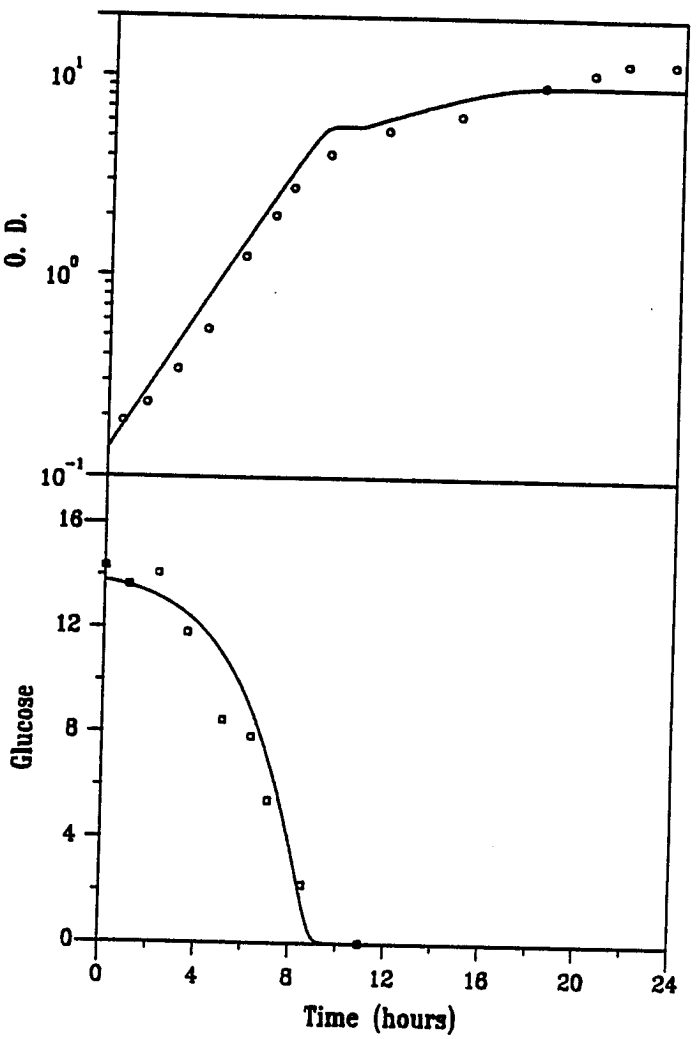

Figure 5.5 Model simulations for batch experiment B1.

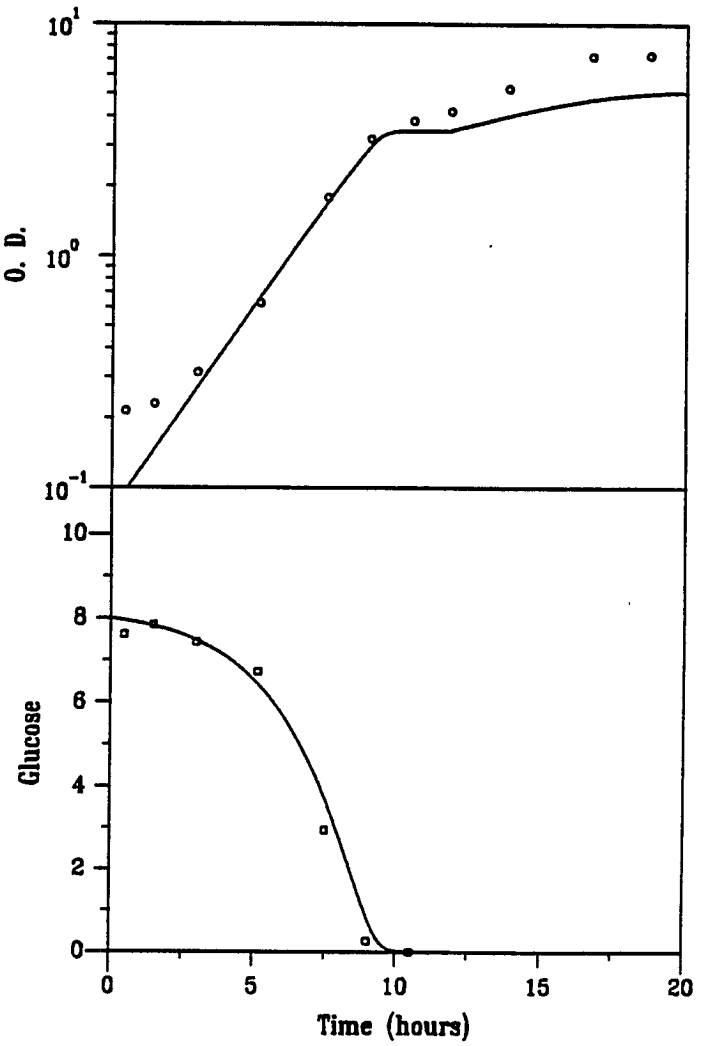

Figure 5.6 Model simulations for batch experiment B2.

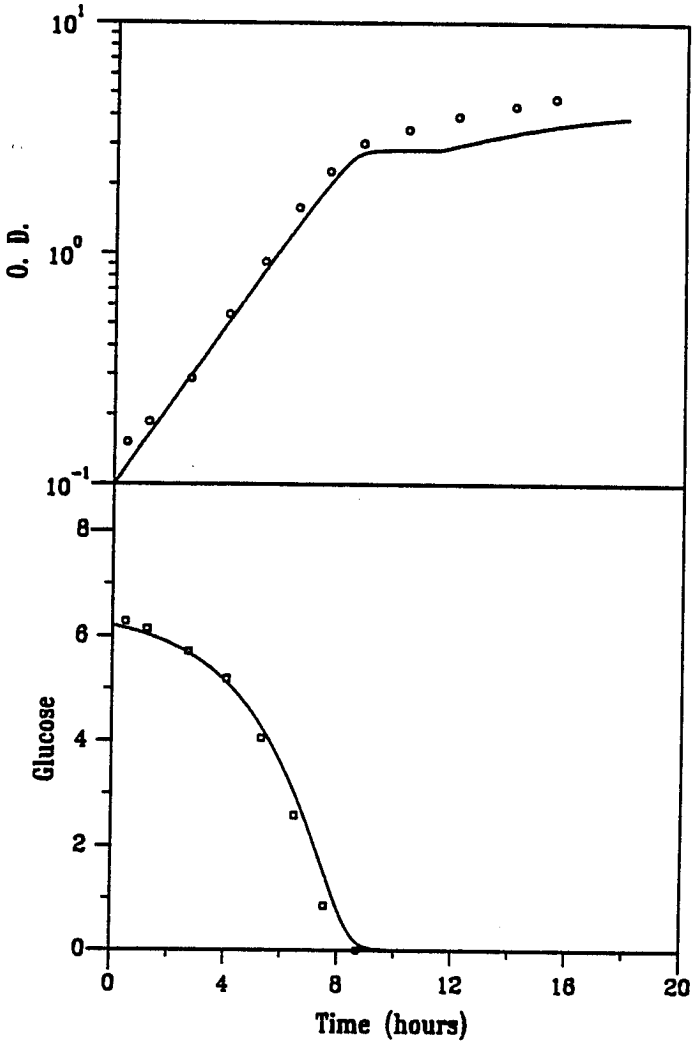

Figure 5.7 Model simulations for batch experiment B3.

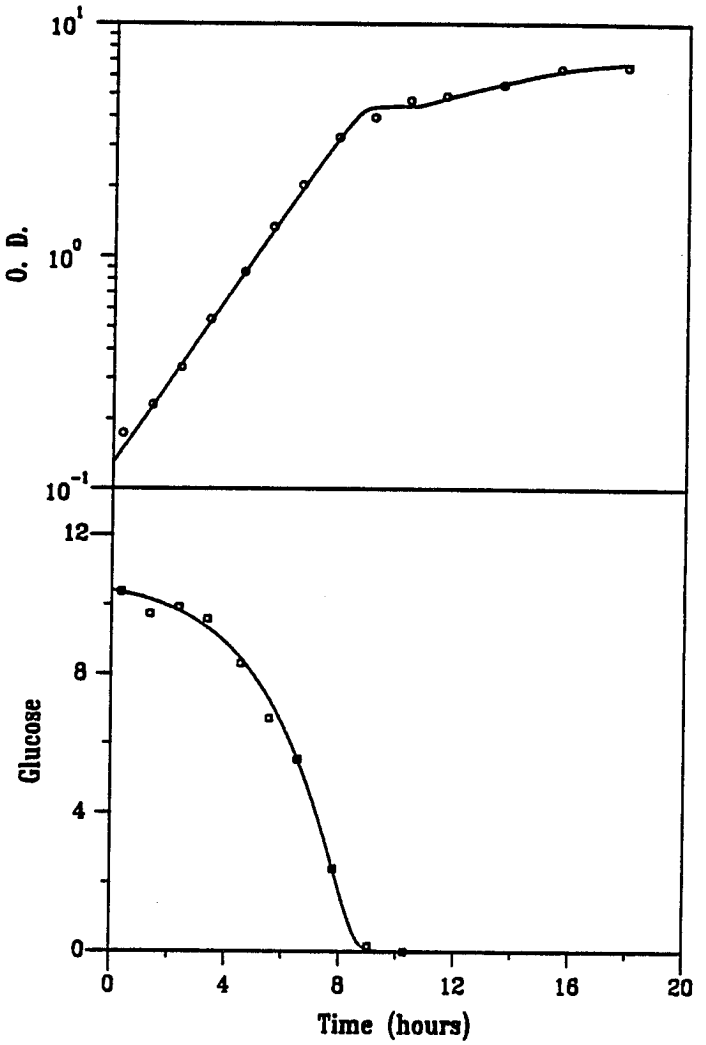

Figure 5.8 Model simulations for batch experiment B4.

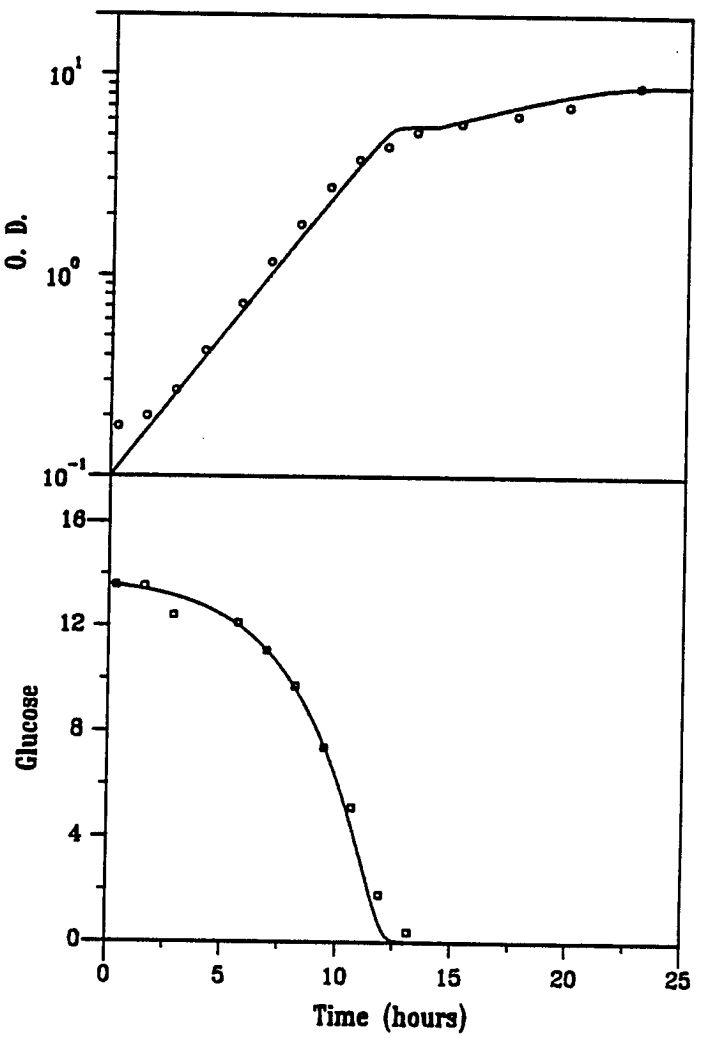

Figure 5.9 Model simulations for batch experiment B5.

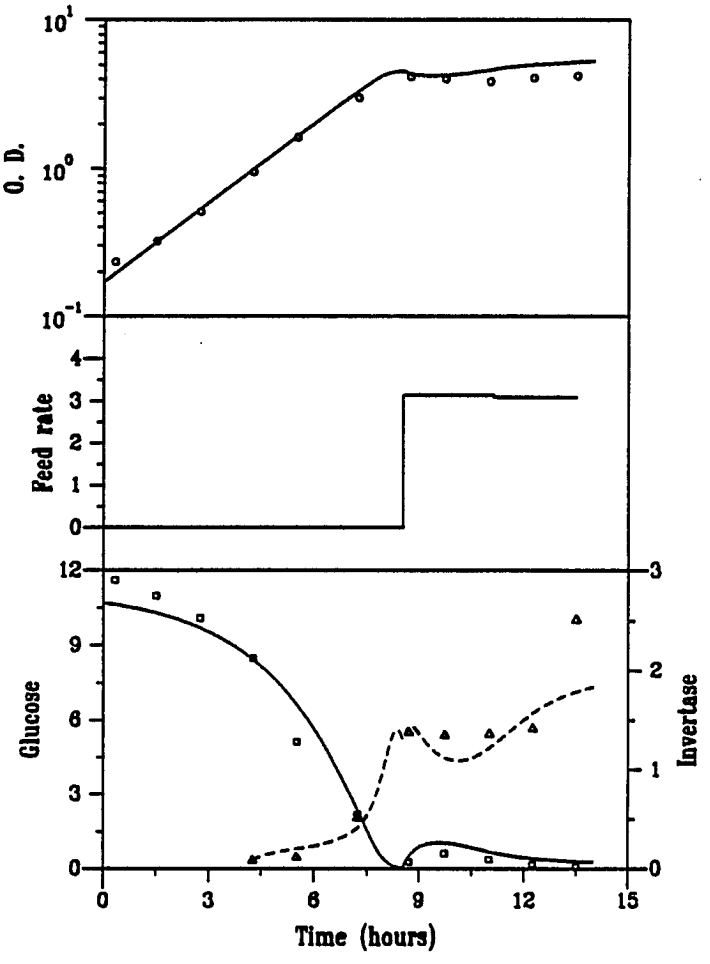

Figure 5.10 Model simulations for fed-batch experiment F1.

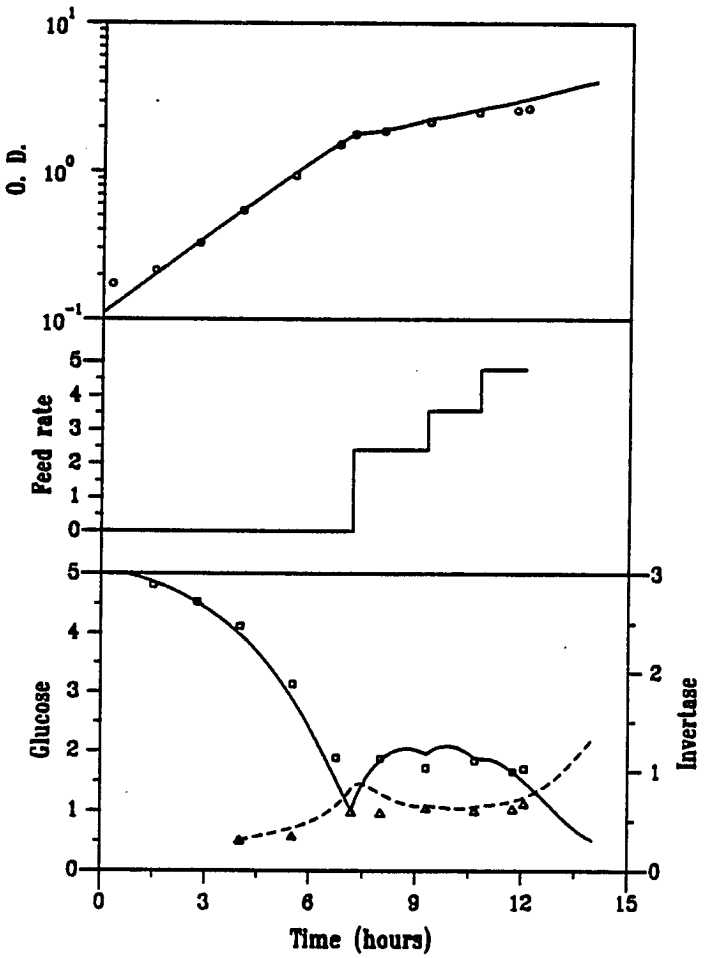

Figure 5.11 Model simulations for fed-batch experiment F2.

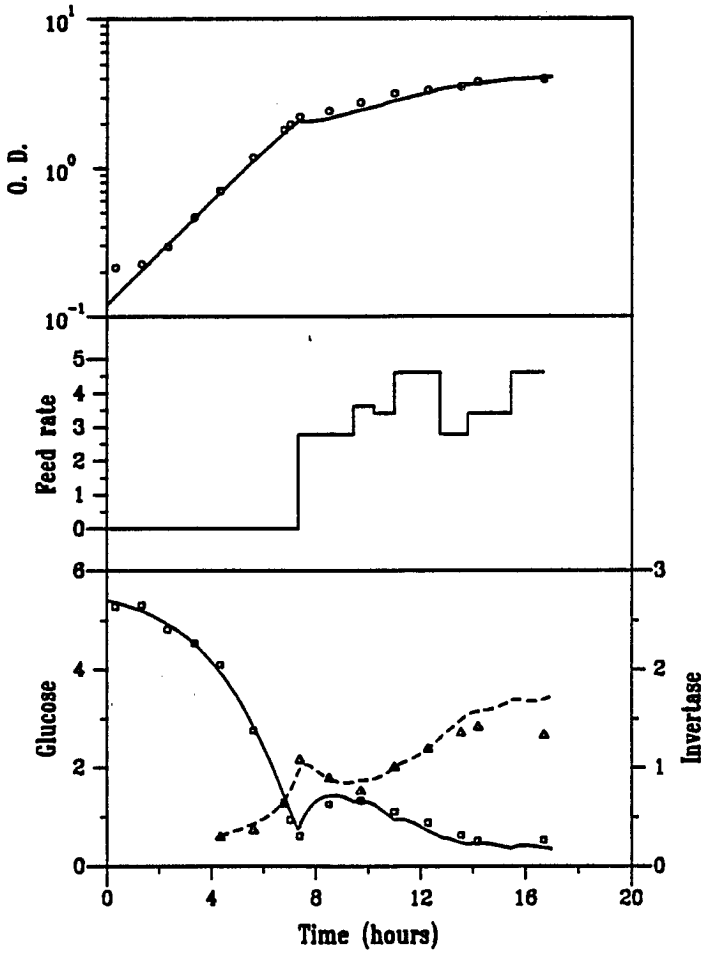

Figure 5.12 Model simulations for fed-batch experiment F3.

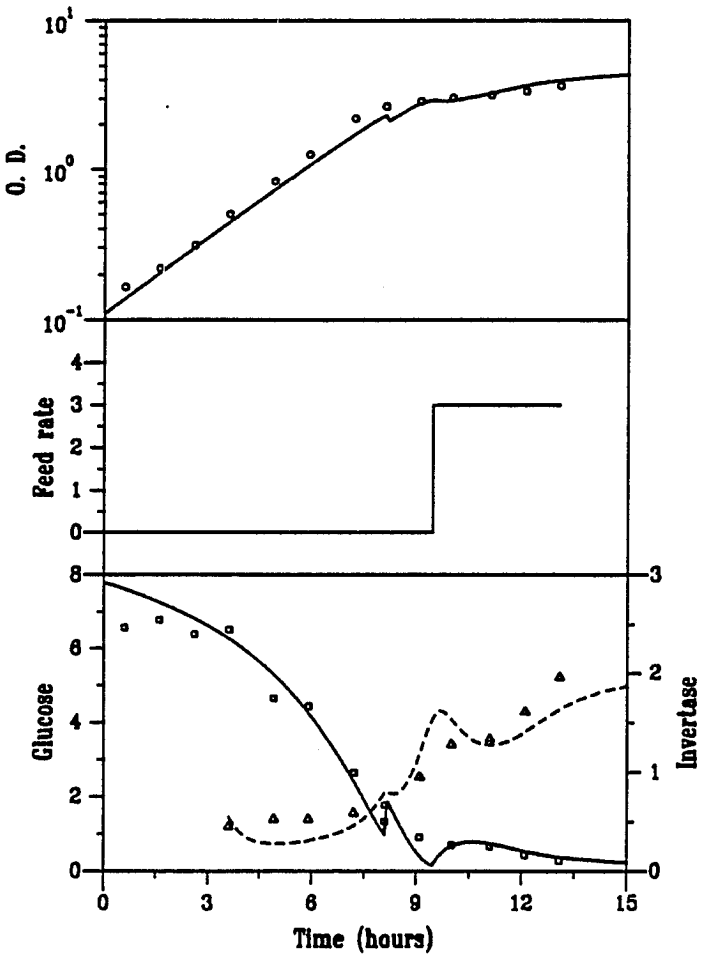

Figure 5.13 Model simulations for fed-batch experiment F4.

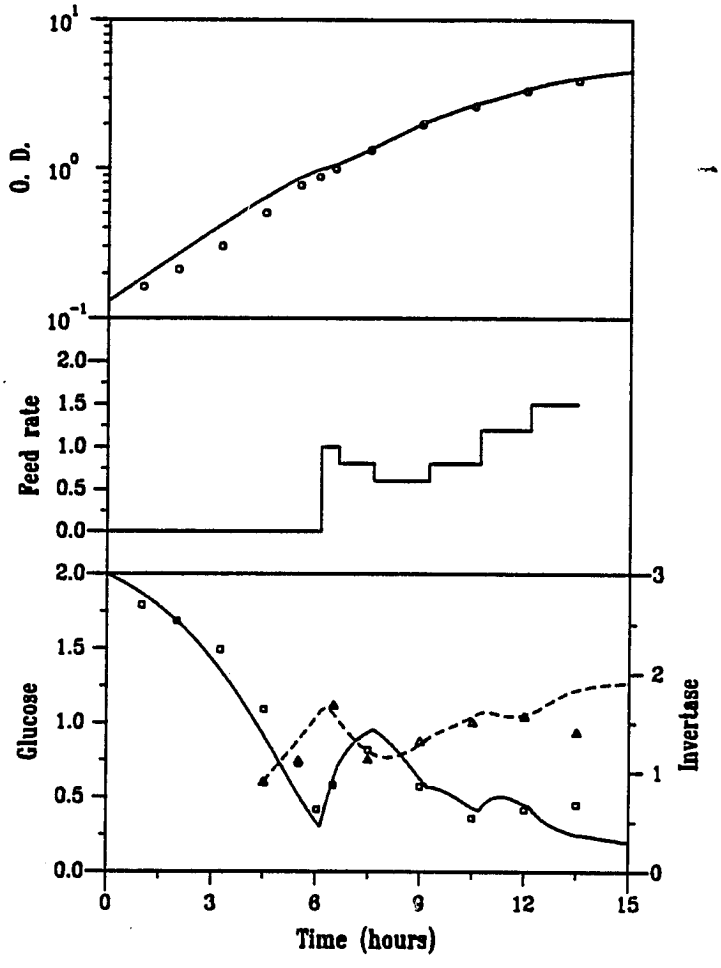

Figure 5.14 Model simulations for fed-batch experiment F5.

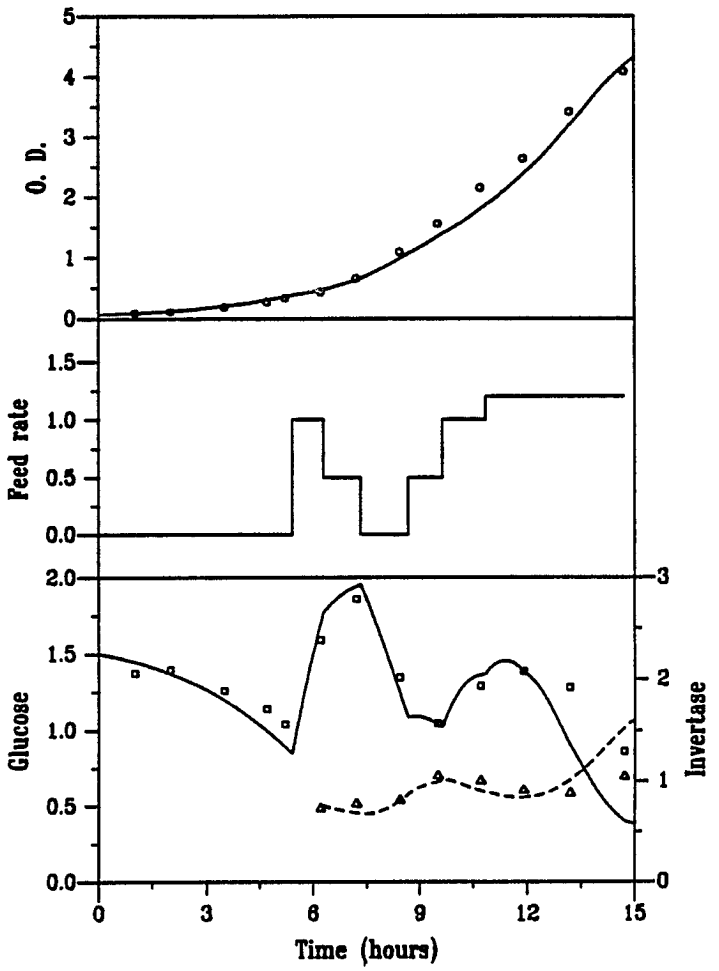

Figure 5.15 Model simulations for fed-batch experiment F6.

A more complicated feeding sequence is used in experiment F3 to induce a hump in the glucose profile. Model simulations follow this trend in glucose really well. Near the end of the fermentation, the model shows slightly lower glucose concentration values compared to experimental results. During experiment F4, about two hours before fed-batch was started, an oversight resulted in a sudden pulse of glucose feed entering the fermentor. This reflects in the glitch in the glucose profile at around 8 hours. In spite of this sudden pulse, the model is able to simulate the data quite well.

In experiments F5 and F6, the glucose feed was varied to cause two humps in the glucose profile. The model still works really well, except near the very end of the fermentation. A possible explanation for this can be given as follows. To explain dynamic situations with drastic changes in the medium requires the use of structured models. This is because continuously varying medium composition results in considerable change in the cell composition. It is quite surprising that despite these limitations, the proposed unstructured model performs as well as it does. The model simulations follow the oscillations in glucose really well.

5.4 A model for invertase production

5.4.1 Functional form In the literature survey (Chapter 2) it was pointed out that the process of regulation of SUC2 gene expression is very complex. At least ten genes are involved and the nature of the interactions between these genes is not clear. Because the exact mechanism of regulation is unknown, it is impossible to formulate a detailed mathematical model which takes into account the complete interplay among the genes. Even if the mechanism were known, the number of parameters in such a detailed structured model will be quite large. Also such a model would involve so many differential equations that it would not be amenable to optimization. These considerations suggest that an empirical approach is most suitable for the purpose at hand.

The first step in modeling invertase production is to set up a mass balance for invertase mRNA. The mRNA is generated by transcription of SUC2 gene on the plasmid pRB58. It is reasonable to assume that this mRNA is subject to decay by RNAases in the cell. Thus the rate of change of specific mRNA concentration inside the cell is governed by the following differential equation.

$$\frac{dN}{dt} = r_t - \mu N - r_d \qquad (5.15)$$

where N is the concentration of SUC2 mRNA inside the cell, r_t is the rate of transcription of SUC2 gene, r_d is the rate of decay of SUC2 mRNA.

The term μN arises because of the dilution effect caused by cell growth. Elorza*et al.*[92] studied the kinetics of the decay of invertase mRNA in a *S. cerevisiae* strain. It was concluded that the decay kinetics is of first-order in mRNA concentration. The half-life of the invertase mRNA at 37°C was estimated to be 30-35 minutes. Although the data by Elorza *et al.* was obtained at a slightly higher temperature, for lack of better assumption, we assume that at 30°C the decay follows first order kinetics.

$$r_d = -k_d N \tag{5.16}$$

Park *et al.*[93] modeled the expression of SUC2-s2 from yeast chromosome at 27°C as,

$$r_t = \frac{S\, e^{-5.0S}}{0.1 + S} \tag{5.17}$$

Experimental evidence in support of this formulation, however, was not shown. Moreover, the functional form contains somewhat peculiar exponential dependence on the glucose concentration. A different formulation of invertase production is hence desired.

The copy number of the recombinant plasmid pRB58 may be different under different growth conditions. Since the transcription rate depends on the number of gene copies inside the cell, this introduces an additional complication in modeling invertase production. A simplifying assumption is made to circumvent this problem.

P5. The plasmid copy number stays constant throughout the course of the fermentation.

The plasmid pRB58 is a 2μm based plasmid. The 2μm plasmid is known to possess a tight copy number controlling machinery[70]. In view of this fact, postulate **P5** appears reasonable.

The overall effect of the complex regulating machinery is that invertase production is derepressed at low glucose concentrations. Another important observation can be made from the fed-batch fermentation results. The specific invertase activity responds almost instantaneously to the changes in the extracellular glucose concentration. From modeling viewpoint, this observation eliminates the need for postulating, say, some repressor molecule(s) to explain the dynamics of regulation of invertase production.

P6. The rate of transcription of SUC2 gene is an explicit function of the extracellular glucose concentration.

Fig. (5.16) shows the experimentally observed dependence of SIA on glucose levels. The data points represent the results of all the fed-batch fermentations. Despite the scatter, it supports the theory that invertase levels are tightly controlled by glucose concentration. The functional form chosen to express this dependence is of simple substrate-inhibition form.

$$r_t = \frac{k_t S}{K_p + S + K_i S^2} \tag{5.18}$$

form is To determine the rate of invertase production, it is postulated that the rate of transcription is the rate controlling step. The translation step is assumed to be at a pseudosteady state. If the rate of translation is considered to be directly proportional to the mRNA concentration, we can write

$$\frac{dP}{dt} = k_{trans} N - k_{pd} P \approx 0, \tag{5.19}$$

where P is specific invertase activity. The enzyme is assumed to decay in a first-order manner. Combining Eq. (5.15, 5.16, 5.18, 5.19) results in a simple expression for the variation of total invertase activity with respect to time.

$$\frac{dPXV}{dt} = \left[\frac{k_p S}{K_p + S + K_i S^2} - k_d P\right] XV \tag{5.20}$$

5.4.2 Parameter estimation and results: In Eq. (5.20) there are four unknown parameters. The data of Elorza *et al* was obtained using a different medium and at a different temperature. Thus the value of the mRNA decay constant they reported (about 1.3-1.4 hr^{-1}) can be used only as an order of magnitude estimate. The parameters were determined by a trial and error procedure to provide the best fit to the experimental results. Invertase assay cannot give accurate values for optical densities below 0.4. Thus the experimental data does not give any idea about the specific invertase activity during the first few hours of fermentation. For simulation purposes, the initial time was chosen as the time for the first invertase data point. The initial value for integration of Eq. (5.20) is the SIA for this data point. The values of model parameters used for simulation are listed in Table (5.1). Figs. (5.10-5.15) compare the results of model simulation with experimental data.

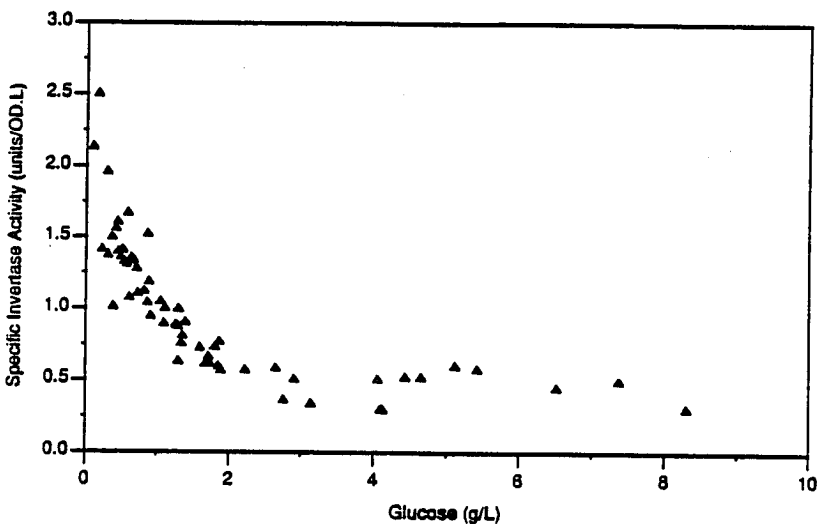

Figure 5.16 Dependence of specific invertase activity on glucose levels.

In the first fed-batch experiment (F1), the specific invertase activity stayed almost constant after the cells were exposed to a brief period of zero glucose concentration. To accurately predict the effects of this shock requires the use of structured models. However, model **P** still does a satisfactory job of duplicating the experimental behavior. The invertase activity does not stay constant as observed experimentally. Also at the end of the fermentation model simulation shows much lower SIA than that observed experimentally. However, the abrupt increase in SIA indicates that the discrepancy is more likely due to an experimental error than because of model inadequacy. The second fed-batch experiment (F2) is relatively simple to model because the glucose level during the fed-batch period of the fermentation stays relatively constant. The simulation shows a corresponding constant SIA in agreement with the experiment.

In experiment F3, the model shows a slightly lower glucose levels compared to experimental results for the last two data points. At such low glucose concentrations, SIA is very sensitive to glucose levels. Thus a slight discrepancy in the glucose level translates into a relatively higher discrepancy in the SIA. The model predicts a higher SIA at the end of the fermentation. The agreement is really good for the previous data points though. Similar considerations apply to experiment F4, in which the model shows slightly higher glucose concentrations near the end of the fermentation. This causes a corresponding decrease in the SIA values. The agreement between the model and the experiments is still satisfactory.

The glucose profiles in experiments F5 and F6 contain two humps. This results in two depressions in the SIA profiles. The model does a good job of predicting these variations in SIA. In these experiments also, the success in simulating the SIA profiles is closely related to the goodness of the glucose profile fit.

5.5 Summary

The final objective of the research (optimization) severely restricts the dimensionality of the mathematical model to be used for describing experimental data. On the other hand certain key features of the system should not be ignored. A simple model consisting of four differential equations is developed to describe fed-batch fermentations. The cell mass yield was found to decrease with an increase in the glucose concentration in the medium. To explain this phenomenon a goal-oriented approach is used. It is assumed that the objective of the multitude of reactions within the cell is to maximize the cell growth rate. This objective is equivalent to maximizing the flux through the respiratory pathway, it being the most energy efficient. Based on

this assumption a simple expression for cell yield was derived. The model so derived agrees with the experimental data quite well. To describe invertase production kinetics, a substrate-inhibition form was used. Invertase mRNA was assumed to decay in a linear first-order manner. This functional form does a good job of simulating the experimentally observed behavior.

At low glucose concentrations the invertase activity is very sensitive to the glucose levels. Thus accurately predicting the glucose consumption rate is the key to successfully model invertase production. It is possible that a detailed model involving some internal components of the cell would do a better job at describing the observed glucose profiles. However, for the purpose at hand the model proposed here is more than satisfactory.

6. OPTIMIZATION: A CONJUGATE GRADIENT APPROACH

6.1 Introduction

In the previous chapter a mathematical model was developed to explain the kinetics of cell growth and product formation. This model forms the basis for determining the optimal glucose feed rate profile to minimize some performance index (as yet unspecified). Before proceeding with optimization of this system though, we will digress to develop and validate a numerical algorithm for optimization of singular systems.

The problem of determining the best feed rate in a typical fed-batch fermentation can be stated as a problem in the calculus of variations.

$$\min_{F(t)} \Pi = \pi(\mathbf{x}(t_f)) \tag{6.1}$$

In the above equation Π represents a suitably chosen performance index (sometimes referred as the objective function), $\mathbf{x}(t_f)$ is the value of the state vector $\mathbf{x}$ at final time t_f (when fermentation is over). The performance index is to be optimized by choosing different feeding strategies F(t). Typically, in fed-batch optimization problems, the substrate feed rate is constrained as,

$$0 = F_{min} \leq F \leq F_{max} \tag{6.2}$$

where F_{max} and F_{min} represent the maximum and minimum allowed feed rates, respectively. The state variables satisfy the following differential equations.

$$\dot{\mathbf{x}} = \mathbf{a}(\mathbf{x}) + \mathbf{b}(\mathbf{x})F, \quad \mathbf{x}(0) = \mathbf{x}_0, \tag{6.3}$$

where $\mathbf{a}$ and $\mathbf{b}$ are vector functions of the state vector $\mathbf{x}$, $\mathbf{x}_0$ represents the vector of specified initial conditions.

Pontryagin's minimum principle[36] states that the above minimization problem is equivalent to minimization of the Hamiltonian defined as,

$$H = \lambda^T [\mathbf{a}(\mathbf{x}) + \mathbf{b}(\mathbf{x})] \tag{6.4}$$

where the *adjoint* (or *costate*) vector, $\lambda(t)$, is defined by

$$\dot{\lambda} = -\frac{\partial H}{\partial \mathbf{x}} \tag{6.5}$$

with the specified terminal conditions given by

$$\lambda_i(t_f) = \frac{\partial \Pi}{\partial x_i(t_f)}, \quad i = 1, 2, \cdots, n \tag{6.6}$$

Here n represents the dimension of the system of differential Eqs. (6.3). Eqs. (6.6) are valid only if the final values of all the state variables are unconstrained. If the final value of a state variable is fixed, corresponding costate variable becomes unconstrained at the final time.

This is a singular control problem, so called because the control variable (substrate feed rate) appears linearly in the Hamiltonian so that the minimum principle of Pontryagin[36] does not provide a solution for the problem. The necessary conditions for optimality of singular arcs have been studied by Jacobson *et al.*[94-97] and discussed by Bryson and Ho[37].

The linear dependence of the Hamiltonian on feed rate enables one to determine the optimal feed rate by examining the coefficient of F, $\lambda^T \mathbf{b} = \phi$. If ϕ is identically zero over a finite time interval, the minimum principle fails to give F(t) during this interval. Such intervals are called singular intervals and the corresponding feed rates, singular feed rates, F_s. Thus,

$$F^*(t) = \begin{cases} F_{max} & \phi < 0 \\ F_s(t) & \phi = 0 \\ F_{min} = 0 & \phi > 0 \end{cases} \tag{6.7}$$

The singular feed rate can be obtained by observing that during the singular interval ϕ is identically zero. This means that all its time derivatives must vanish.

$$\frac{d^j \phi}{dt^j} = 0, \quad j = 1, 2, \cdots \tag{6.8}$$

From these equations, the form of singular feed rate can be deduced [50].

$$F_s = F_s(x, \lambda) \tag{6.9}$$

The singular control theory just tells us that the optimal control arcs can be either on the boundary or singular. It does not give the sequence and exact duration of these arcs. Numerical methods are hence necessary to further elucidate the form of the optimal control.

6.2 Previous optimization attempts

Modak *et al.*[1] developed a computational algorithm for solving optimal control problems of dimension less than five. They argued that in some cases a physical insight into the problem can reveal the sequence of maximum, minimum and singular arcs. For fed-batch fermentations, they proposed the following conjecture: the feed flow rate must be manipulated to first grow the cells optimally and then to force the cells to produce the desired product optimally. For all the cases studied, this conjecture leads to one of the two possible optimal sequences: maximum-zero-singular-zero or zero-maximum-singular-zero. Of course, degenerate forms of the above general sequences are also likely†. Even though the optimal sequence can be deduced, the problem of determining the duration of each of the arcs is a difficult boundary value problem. This is because determination of optimal control necessitates the integration of state and adjoint differential equations. This requires guessing either the initial conditions on the adjoint variables, or the final conditions on the state variables. Modak *et al.* modified this two-point boundary value problem to make it easier to solve numerically. Their approach requires the use of only two adjoint variables. Also these variables need to be integrated from the start of the singular period rather than the initial time. They proposed an algorithm which essentially iterates over four variables, viz. the durations of the first two control arcs (maximum-minimum or minimum-maximum) and the values of two adjoint variables at the start of the singular period. The duration of the singular period can be determined by invoking the final condition

† A degenerate form is a sequence in which one or more of the parts of the general sequence are missing, e.g. maximum-singular-zero is a possible degenerate form.

$$V(t_f) = V_f \tag{6.10}$$

Thus the singular period is over when the reactor is full. The final time, if free, can be determined by using an appropriate final condition[50].

Thus the complex boundary value problem is reduced to a four-dimensional iteration problem. Modak *et al.* applied this technique to a variety of systems[1,98]. The major limitation of this technique, however, is that it is applicable only for systems which can be described by less than five differential equations. For higher dimensional systems two difficulties arise. First, the number of possible switches in the control profile increases†. This increases the possible permutations of maximum, minimum and singular arcs. In fact more than one singular arcs are possible. This makes it very difficult, if not impossible, to guess the sequence of the control arcs *a priori*. Second, it increases the number of variables to be guessed.

Modak and Lim[2] realized these limitations and proposed a transformation approach. The original singular problem was converted to a nonsingular problem by proper choice of a new set of state variables and the use of culture volume as the control variable instead of the feed rate. This new problem can now be solved by gradient techniques commonly used for nonsingular problems. In principle, this approach is very general and can be used to optimize systems of any dimension. However, this technique suffers from numerical difficulties for high dimensional systems. In fact, for systems with dimension greater than three, no results have been reported using this method. Another difficulty arises at the junction of two control arcs. Modak's results for three dimensional systems indicate a deviation of the final converged profile from the optimal profile, calculated by singular control theory, near the switching time between singular and batch period. Also the approach shows convergence problems at the beginning and end of singular arcs, during which the constraints on the control variable and its time derivative are active. There can be two possible reasons for these difficulties. An inequality constraint on the control variable can results in a discontinuity in the adjoint variables[37]. Also there is no satisfactory way, in the optimal control theory, to incorporate constraints on the time derivative of the control variable. The gradient

† For linear systems, with coefficients **a** and **b** independent of the state variables, the number of switches equals $(n-1)$. In general, no such formula exists for nonlinear systems.

method used by Modak ignores both these factors and thus forces the control trajectory to be suboptimal.

6.3 A conjugate gradient approach

A conjugate gradient method based on the Fletcher-Reeves method for function extremization was proposed by Lasdon *et al.*[99]. The method was able to handle problems with unconstrained problems. Penalty functions were added to the performance index when terminal state variable constraints were present. It was proved that the directions in function space generate by the conjugate gradient method are such that the objective function is decreased at each step. Numerically, the conjugate gradient technique was found to converge much better, with little additional computation per iteration, than steepest descent technique for linear-quadratic problems†. However, the usefulness of the method was demonstrated only for nonsingular systems.

Pagurek and Woodside[100] extended this technique to handle directly saturation constraints on the control variables. Terminal constraints were still handled by means of a quadratic penalty function. A second algorithm, which uses second order terms, but requires more computation, was also proposed. This algorithm showed substantially improved convergence in some examples. Using a simple singular system as an example, it was shown that the conjugate gradient algorithm can work for singular systems also. The applicability of this technique to higher dimensional singular systems was not tested.

Stutts[54] studied the applicability of conjugate gradient method to solve singular fed-batch optimization problems. The first problem he solved involved the determination of the optimal feed rate for maximum productivity of a methylotroph to be used as a source of a single cell protein (SCP). This is a three dimensional singular problem and the method performed really well. The switching times between different control arcs were not clearly defined as a result of the incapability of the conjugate gradient algorithm of putting corners into the solution for F. A more complex (four-

† The term refers to systems governed by linear state equations and the performance index contains a quadratic term in the control variable.

dimensional singular) problem—maximizing penicillin productivity using Bajpai and Reuβ's model[101]—was then attempted. It is difficult to interpret the optimal feed rate profile he obtained. There are no regions of maximum or minimum flow rates, but there are a few regions of constant intermediate flow rate. Also, the method was very sensitive to the initial guesses of the feed rate. The third problem attempted by Stutts was optimization of a six-dimensional model for penicillin fermentation. Again the solution did not indicate any periods of maximum or minimum flow rate. Stutts performed his calculations on a VAX 11/780 which is considerably slower than the modern day computers. For example, for the four-dimensional penicillin model optimization it took about about 5 minutes of CPU time for each iteration. A simple test on an Ardent Titan-P3 machine indicated that the CPU time required would be about 2-3 seconds/iteration. Thus, there is an incentive to carry on the computations for larger number of iterations to test the validity of this algorithm.

6.4 Algorithm

Although Pagurek and Woodside obtained better results using a second order conjugate gradient method for some systems, it was felt that this method required determination of a lot of complex derivatives. Also the memory requirements are quite high compared to the simple first order conjugate gradient method. The later method is used exclusively in this numerical study.

In conjugate gradient method, it is not easy to deal with equality constraints on the final values of state variables in a direct manner. The hard constraint on final volume, normally imposed when solving a singular control problem using the method of Modak *et al*, is converted to a soft constraint by modifying the objective function by adding a penalty function.

$$\min_{F(t)} \Pi = \pi(x(t_f)) + K\,(V(t_f) - V_f)^2 \qquad (6.11)$$

Here $V(t_f)$ is the final volume obtained by simulation, while V_f is the desired final volume. This method does not guarantee that the optimal profile would result in the desired value for the final volume. However, by choosing a proper value for the constant K, the difference $\left| V(t_f) - V_f \right|$ can be made arbitrarily small. One need not add any penalty functions for control variable constraints of the type given by Eq. (6.2). The technique used in this study can take care of these constraints in a simple and direct manner. The algorithm consists of the following steps.

1. The procedure begins with an initial guess for the feed rate profile $F_0(t)$. This profile may contain boundary arcs. At the same time a function w is initialized,

$$w_0(t) = \begin{cases} 0, & \text{for } t \in W \\ 1, & \text{elsewhere} \end{cases} \tag{6.12}$$

Here W represents the control boundary region.

2. The state differential Eqs. (6.3) are integrated using the guessed feed rate profile.

3. The adjoint Eqs. (6.5) are integrated using Eqs. (6.6) as final conditions.

4. Initial gradient direction is calculated using the following equation.

$$s_0 = g_0 = \frac{\partial H}{\partial F} \tag{6.13}$$

5. The control profile for next iteration is computed as

$$F_1 = F_0 - \alpha_0 s_0 w_0 \tag{6.14}$$

where α_0 is chosen using a one dimensional search to minimize Π. However, before Π is computed in each trial of the α-search, F_1 is truncated at the upper and lower bounds for the feed rate (6.2). When the best α_0 is determined, the function $w_0(t)$ is updated to $w_1(t)$ using the procedure shown in Eq. (6.12).

6. The state and adjoint equations are integrated again as before using the improved feed rate profile.

7. The following quantities are evaluated.

$$I_i = \int_0^{t_f} w_i g_i^2(t)\, dt \tag{6.15}$$

$$\beta_i = \frac{I_i}{I_{i-1}} \tag{6.16}$$

Here i refers to the iteration number. If β turns out to be negative, a steepest descent step was taken by setting it to zero.

8. The conjugate gradient direction is determined using

$$s_i = g_i + \beta_i s_{i-1} \tag{6.17}$$

9. The control profile is modified by an α-search as before.

$$u_{i+1} = u_i - \alpha_i s_i w_i \tag{6.18}$$

10. Steps 6-9 are repeated till the algorithm converges.

The procedure as outlined above is quite straightforward. The direction g_i is the one of steepest descent, while s_i represents a direction conjugate to the direction of steepest descent. Thus, in essence, the α-search determines the best linear combination of these two conjugate directions. If β_i is set to zero, one recovers the steepest descent method.

Some comments are in order here. The algorithm is somewhat sensitive to the initial control profile chosen. Some initial guesses can cause the procedure to diverge, i.e. the performance index keeps increasing with each iteration. For the three-dimensional system studied, the technique converged for all the chosen initial guesses. For more complex systems one has to more careful though. An important consideration is that the guessed feed rate profile should satisfy the final condition on volume. This assures that the penalty function term in Eq. 6.11 is zero at the beginning. If this term is very large compared to $\pi(x(t_f))$, the algorithm concentrates more on decreasing the penalty function, without much consideration towards increasing π. This causes the control profile to go farther away from the optimum.

It was found that after every few iterations, the conjugate gradient technique tends to stagnate. Under these circumstances the same control profile gets repeated after two or three iterations. After some trial and error, it was realized that, if the method is reinitialized by setting β_i to zero for one iteration, it starts working again. As pointed out before, this changes the method to a steepest descent method for one iteration. This reinitialization was, therefore, incorporated into the general algorithm. The optimal number of iterations, before reinitialization, were found to be between 15 and 30 depending on the system.

To implement the α-search, the maximum of the absolute value of the gradient s(t) for non-boundary control arcs was determined after each iteration. Based on this maximum, $s_{max}(t)$, a parameter δ was calculated.

$$\delta = \frac{\Delta F}{s_{max}(t)} \tag{6.19}$$

Here ΔF represents the maximum allowed change in the feed rate profile after each iteration. The performance index, Π, is then determined for α values equal to $\delta/2$ and δ.

The best α is then determined by quadratic interpolation. The parameter ΔF needs to be varied as the number of iterations increases. At the beginning ΔF is kept high to make the feed profile move towards the optimum very fast. As the optimum approaches, a high ΔF value causes the algorithm to overshoot the optimum. In this numerical study, ΔF was progressively decreased as optimum was approached.

An important feature of this method is that it does not need any variable transformation to convert the singular problem into a nonsingular one. Thus the difficulties associated with differential constraints on control variables, present in the transformation approach, are completely avoided. The convergence of this method near the optimum is somewhat slow, but the only measure needed, to ensure decrease in performance index at each iteration, is a smaller step size ΔF.

Stutts[54] continuously changed the value of the parameter K, which determines the degree of softness, so to speak, of the volume constraint, with increase in the number of iterations. In this work, however, K was kept constant throughout the optimization. It was felt that it is not very important to keep the final volume exactly at the required value. In all the calculations, the difference between the calculated final volume and the desired final volume was not more than 2%.

6.5 Optimization of a simple 3-dimensional model

6.5.1 Formulation To test the effectiveness of the conjugate gradient method, an optimization of a simple 3-dimensional model was attempted. This system is one of the few solved by Modak and Lim[2] by a variable transformation approach. The governing equations for this model can be written as,

$$\frac{d}{dt}(XV) = \mu XV \tag{6.20}$$

$$\frac{d}{dt}(SV) = FS_F - \frac{\mu XV}{Y} \tag{6.21}$$

$$\frac{d}{dt}(V) = F \tag{6.22}$$

The objective is to maximize the amount of cell mass after a fixed final time t_f.

$$\min_{F(t)} \Pi = -(XV)_{t_f} + K\,(V(t_f) - V_f)^2 \tag{6.23}$$

The specific growth rate is substrate inhibited, the maximum occurring at S = 0.24 g/L and the cell yield is assumed to be constant.

$$\mu(S) = \frac{S}{0.03 + S + 0.5S^2}, \qquad Y = 0.5 \tag{6.24}$$

6.5.2 Results The problem has been solved by singular control theory[46] and it is known that the optimal singular feed rate maintains the substrate concentration constant at 0.24 g/L, thus maintaining the specific growth rate at its maximum value. The initial conditions and other parameters used for optimization are listed in Table (6.1).

Three different cases were studied. In Case A, the initial substrate concentration was set to zero. According to singular control theory, the optimum feed rate profile should consist of an initial period of maximum feed rate followed by singular and batch periods. This operating policy brings the substrate concentration to 0.24 g/L and maintains it there till the end of the fermentation. When the fermentor is full, there is a very short batch period. Results of using conjugate gradient technique are shown in Figs. (6.1-6.4). Figure (6.1) shows the improvement in the guessed feed rate profiles as a function of the number of iterations. The initial guess is a constant feed rate. After 83 iterations the shape of the feed rate profile is almost right, although there are many bumps. It took considerable time to get a smooth profile. The variation of performance index with iterations is depicted in Fig. (6.2). After about 200 iterations the increase in the performance index is quite small. However, the feed rate profile does change significantly. Thus the final profile is sharper, at the transition from maximum to singular, and from singular to batch periods, than the profile after 534 iterations.

The final control and state variable profiles obtained after convergence are shown in Fig. (6.3). As expected the substrate concentration is kept constant at 0.24 g/L. As predicted by the singular control theory, there is a short initial period of maximum feed rate. The optimality of the converged feed profile can be established by looking at the gradient of the Hamiltonian ($\partial H/\partial F$). The gradient is zero over the singular region, positive during the batch period, and negative during the initial period of maximum feed rate. This proves that the feed rate profile is optimal.

If a higher value of K is used, it is expected that the final volume would be closer to the desired value. This was tested using K = 1000 (Case B). The final converged feed rate profile in Case A was chosen as the initial guess for Case B. The results are shown in Fig. (6.5-6.6). The converged feed rate profiles in these two cases are not

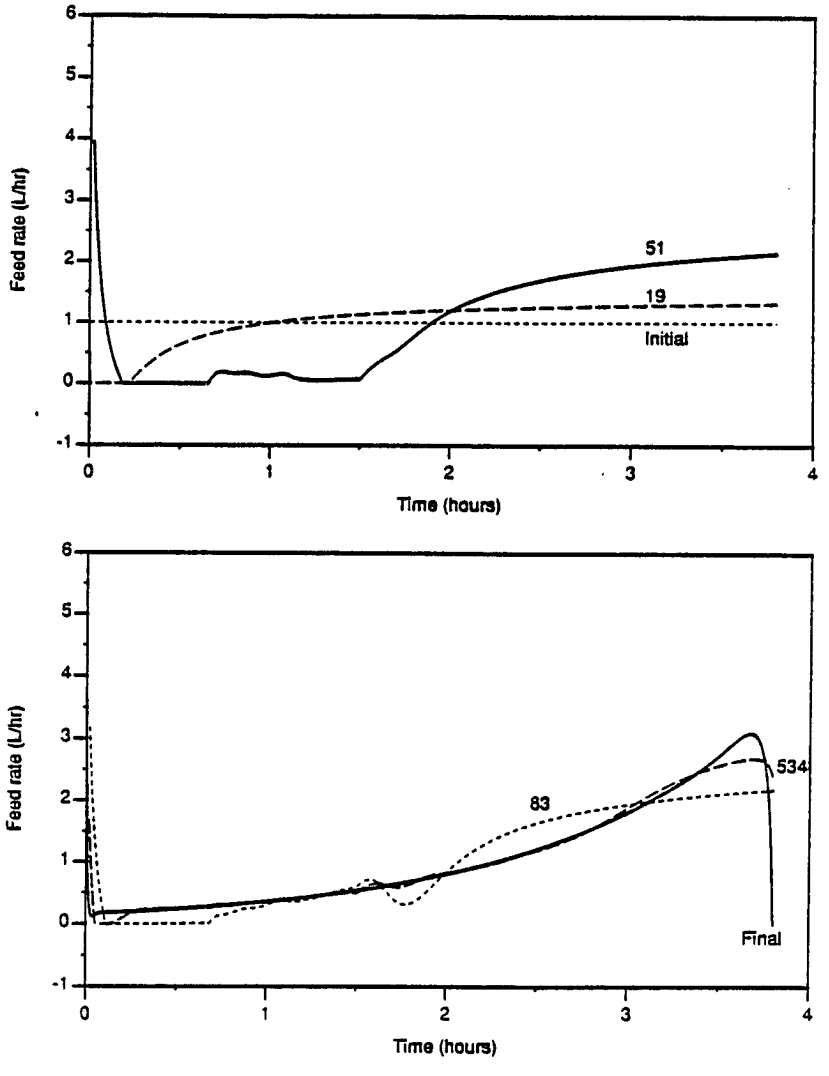

Figure 6.1 Three-dimensional model (Case A): Variation of feed rate profiles with number of iterations.

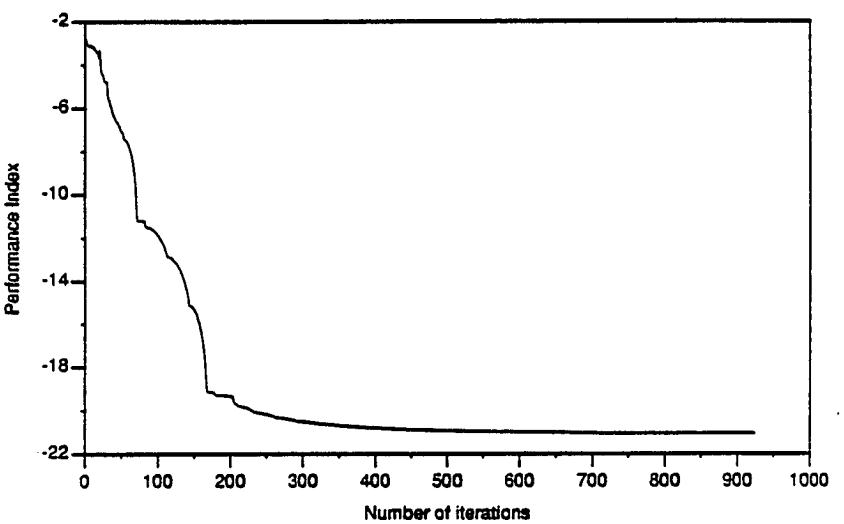

Figure 6.2 Three-dimensional model (Case A): Variation of performance index with number of iterations.

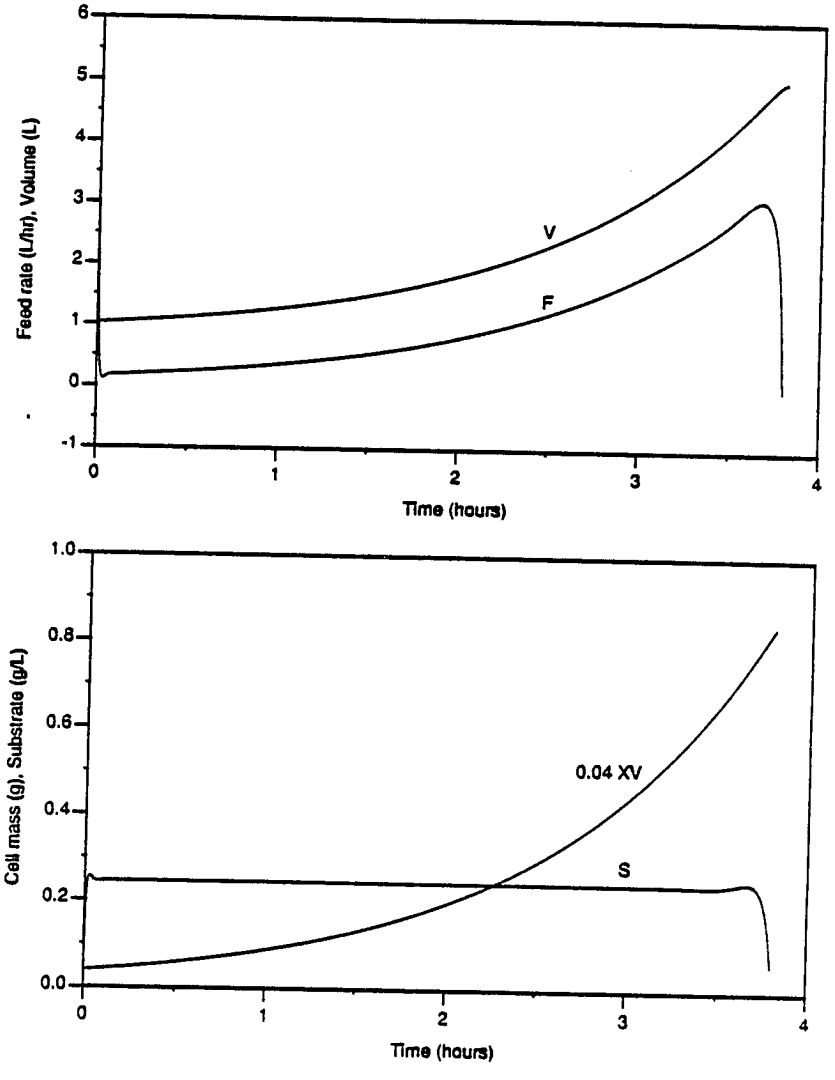

Figure 6.3 Three-dimensional model (Case A): Optimal feed rate and state variable trajectories.

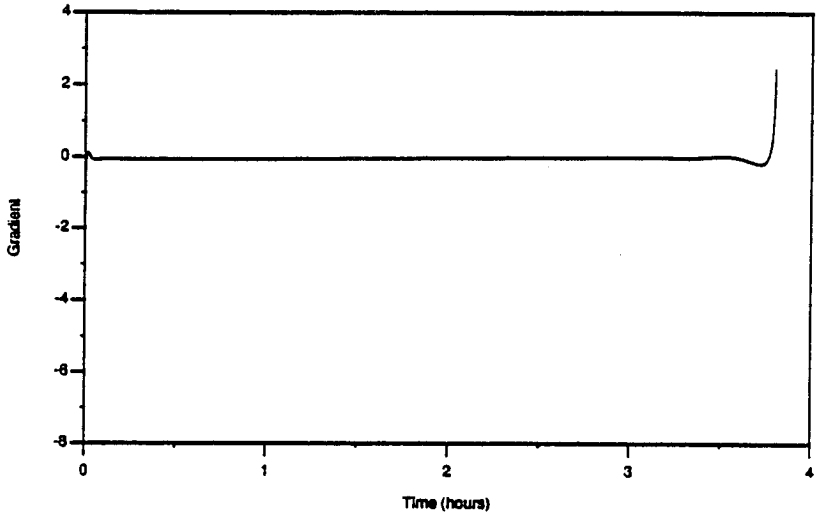

Figure 6.4 Three-dimensional model (Case A): Gradient of the Hamiltonian ($\partial H/\partial F$).

Table 6.1 Parameters for optimization of Modak's three-dimensional model

Case	A	B	C
$(XV)_0$, g	1.0	1.0	1.0
$(SV)_0$, g	0.0	0.0	1.0
V_0, L	1.0	1.0	1.0
V_f, L	5.0	5.0	5.0
S_F, g/L	10.0	10.0	10.0
F_{max}, L/hr	4.0	4.0	4.0
t_f, hr	3.8	3.8	3.8
K	10	1000	10

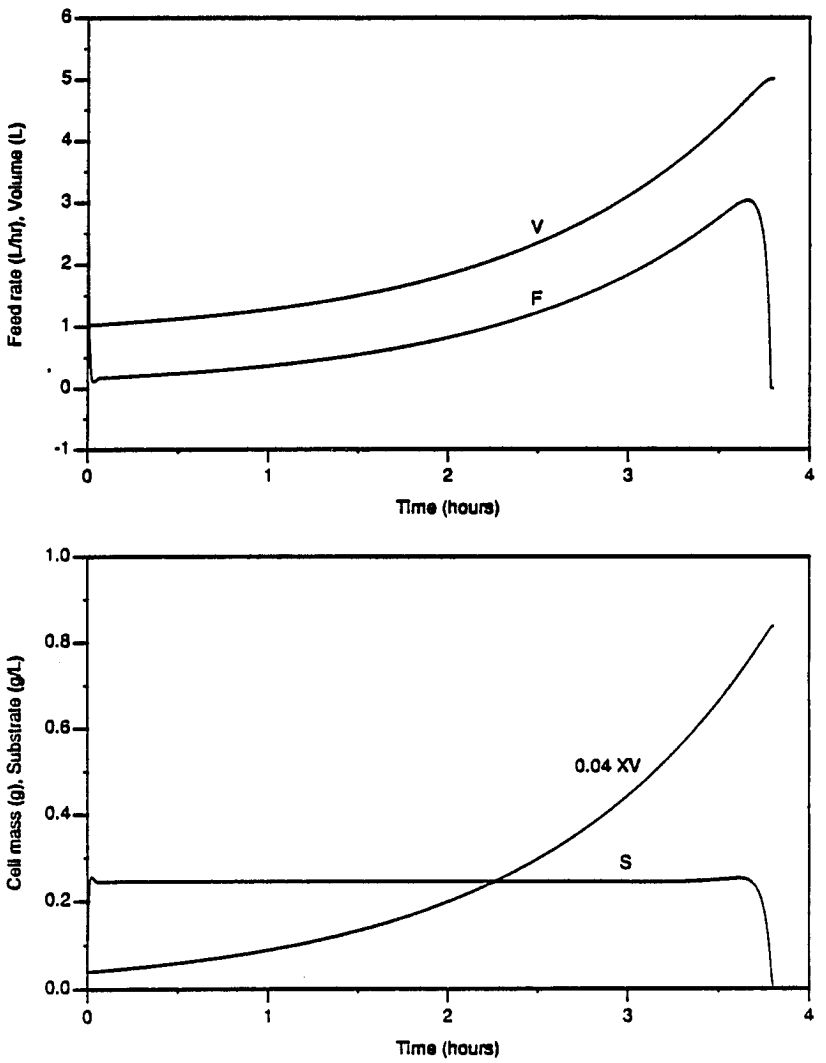

Figure 6.5 Three-dimensional model (Case B): Optimal feed rate and state variable trajectories.

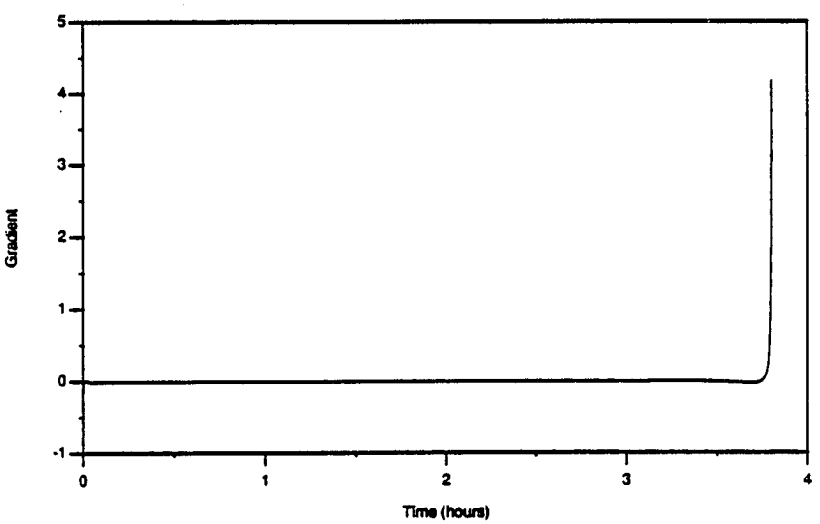

Figure 6.6 Three-dimensional model (Case B): Gradient of the Hamiltonian (∂H/∂F).

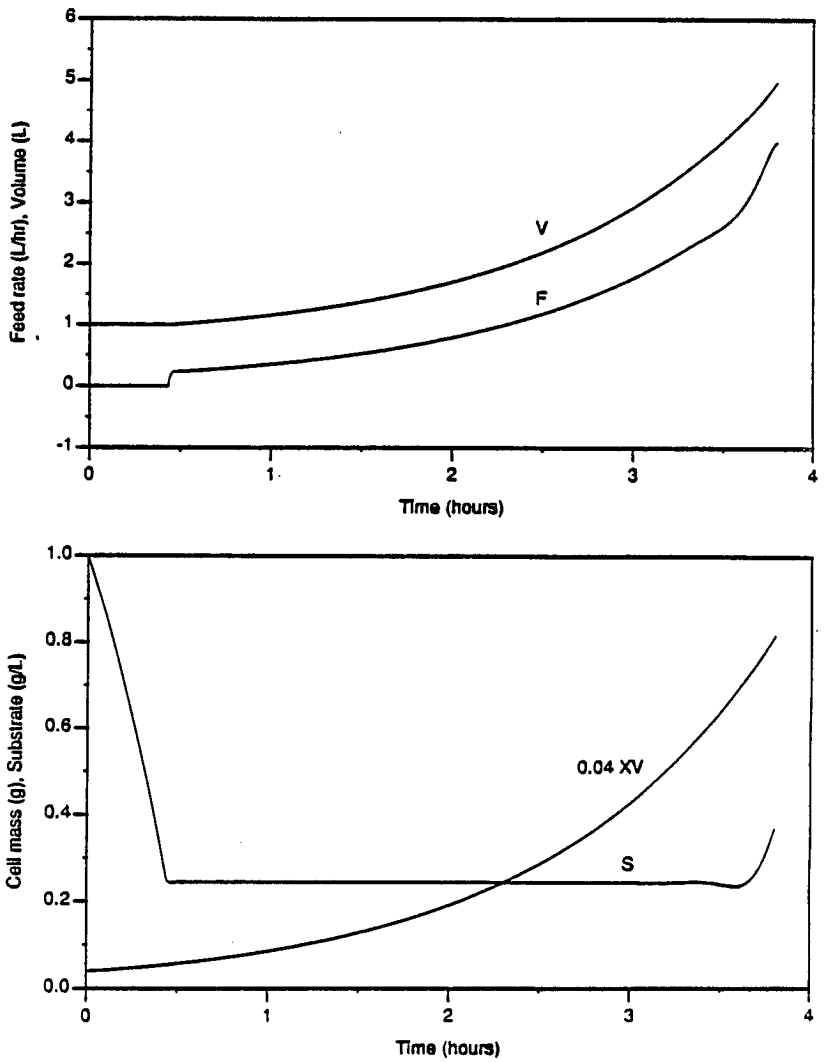

Figure 6.7 Three-dimensional model (Case C): Optimal feed rate and state variable trajectories.

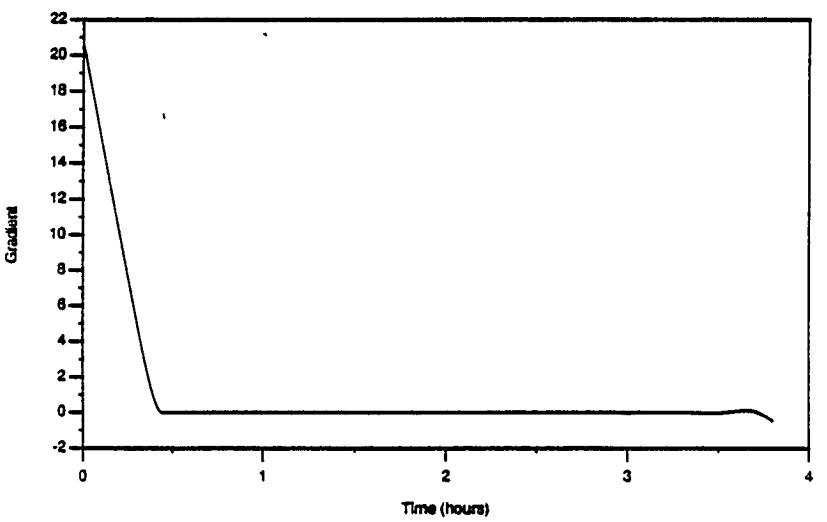

Figure 6.8 Three-dimensional model (Case C): Gradient of the Hamiltonian ($\partial H/\partial F$).

Table 6.2 Parameters used for optimization of DiBiasio's model.

Model parameters	μ_{max}, /hr		0.43
	K_S, g/L		1.1
	K_M, g/L		0.21
	p		0.14
	Y_S, (g. cell)/(mg. phosphate)		0.13
	Y_M, (g. cell)/(g.metabolite)		0.03
	k (g. metabolite)/(g. cell)		13.0
Optimization parameters	Case	A	B
	$(X^+V)_0$, g	0.1	0.1
	$(X^-V)_0$, g	0.01	0.01
	$(SV)_0$, g	0.1	0.1
	$(MV)_0$, g	0.05	0.5
	V_0, L	1.0	1.0
	V_f, L	5.0	5.0
	S_F, mg/L	6.8	6.8
	F_{max}, L/hr	0.5	0.5
	t_f, hr	20.0	20.0

much different. The gradient is a little smoother in Case B. The final volume was indeed closer to the desired value (5.002 in Case B compared to 5.056 L in Case A).

The only difference between Cases A and C is, that in Case C, the initial substrate concentration was chosen to be 1.0 g/L. Fig. (6.7-6.8) show the final converged profiles. As expected, there is an initial batch period which brings the substrate concentration down to 0.24 g/L. It is followed by a singular period, which maintains the glucose concentration constant. Near the end of the fermentation the feed profile deviates a little from the optimum. The substrate concentration shows a corresponding change from the optimal value. The gradient of the Hamiltonian shows expected trends (Fig. 6.8).

6.6 Optimization of DiBiasio's model

6.6.1 Formulation Encouraged by the success of the optimization technique with a simple three-dimensional system, the technique was tried on two complex five-dimensional systems. First test system was a model developed by Sardonini and DiBiasio[3] to explain the growth kinetics of a plasmid-carrying strain of *S. cerevisiae*. It was noted that under phosphate-limited growth conditions in a selective medium, the fraction of plasmid-free cell population was much larger than expected. This phenomenon was explained by assuming that the plasmid-free cells could grow in the selective medium by using a metabolite, that is excreted into the medium by the plasmid-carrying strain, for growth. The model equations describing fed-batch fermentations are given below.

$$\frac{d}{dt}(X^+V) = (1 - p)\mu_+ X^+V \tag{6.25}$$

$$\frac{d}{dt}(X^-V) = p\mu^+ X^+V + \mu_- X^-V \tag{6.26}$$

$$\frac{d}{dt}(SV) = -\frac{\mu_+ X^+V}{Y_S} - \frac{\mu_- X^-V}{Y_S} + FS_F \tag{6.27}$$

$$\frac{d}{dt}(MV) = k\mu_+ X^+V - \frac{\mu_- X^-V}{Y_M} \tag{6.28}$$

$$\frac{dV}{dt} = F \tag{6.29}$$

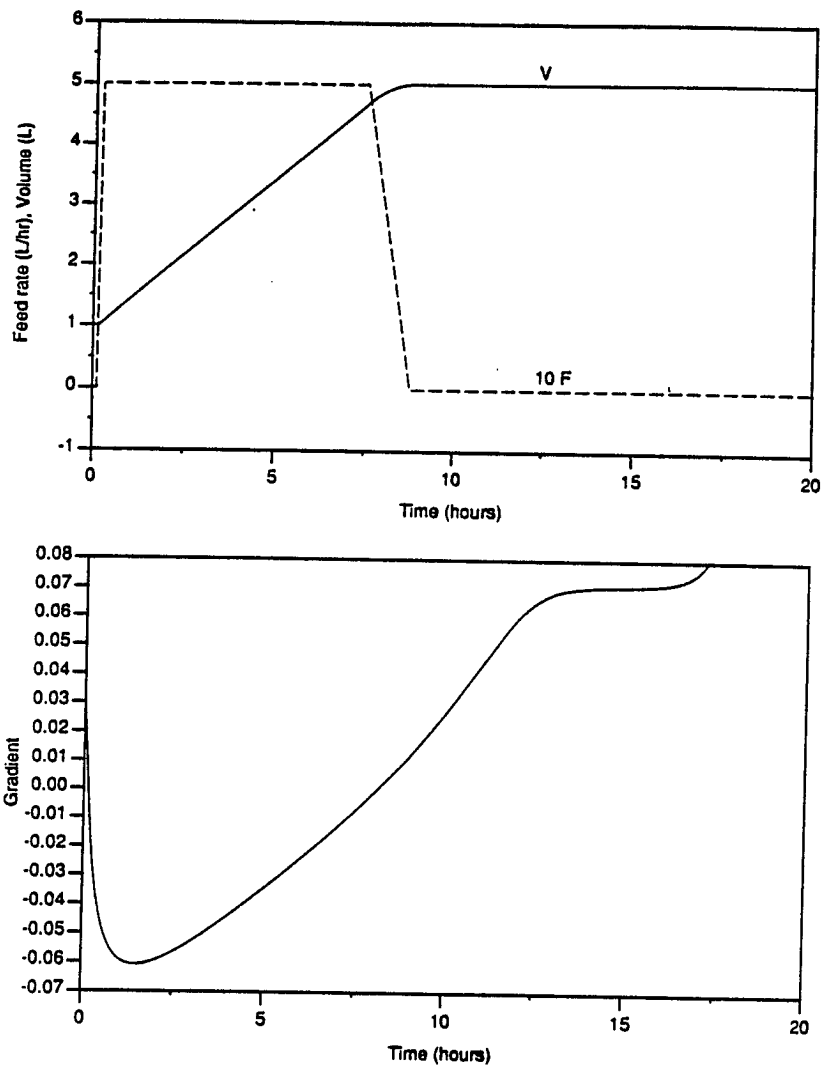

Figure 6.9 Optimization of DiBiasio's model (Case A): Feed rate, volume and gradient of the Hamiltonian.

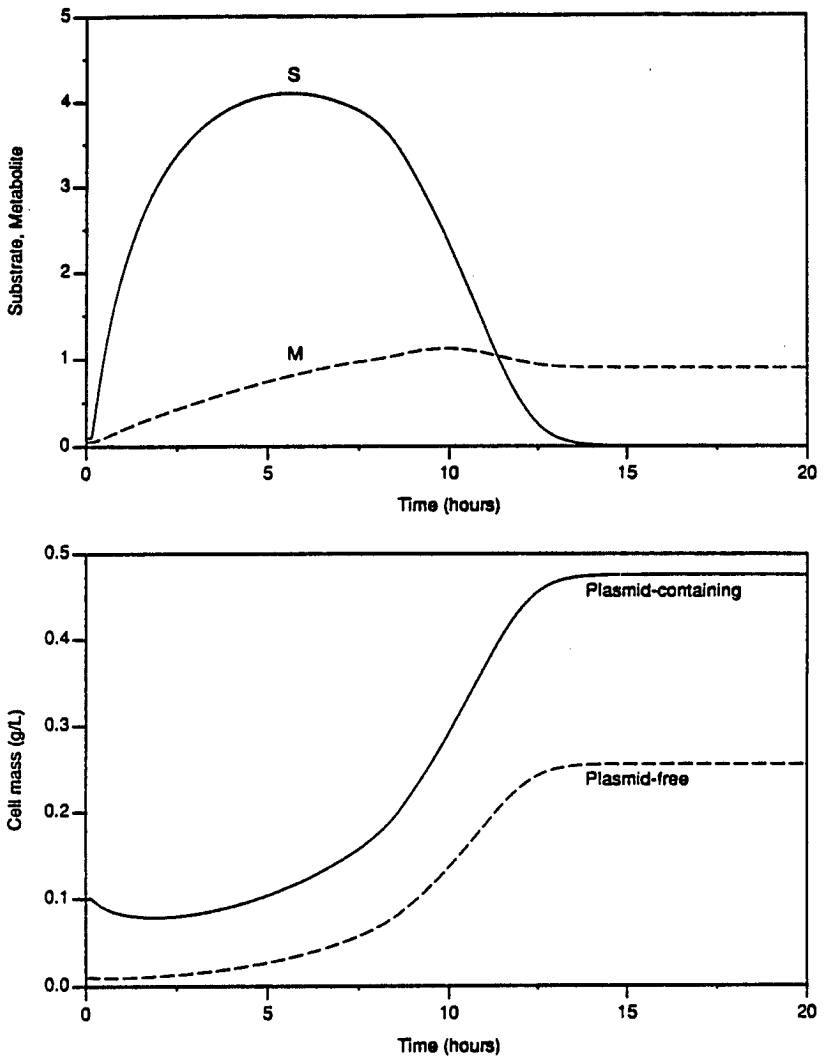

Figure 6.10 Optimization of DiBiasio's model (Case A): Cell mass, substrate and metabolite.

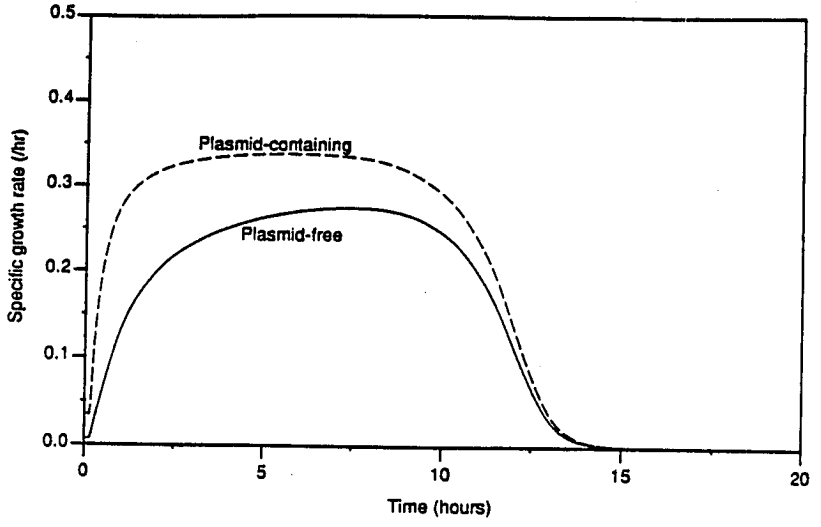

Figure 6.11 Optimization of DiBiasio's model (Case A): Specific growth rates of plasmid-containing and plasmid-free cells.

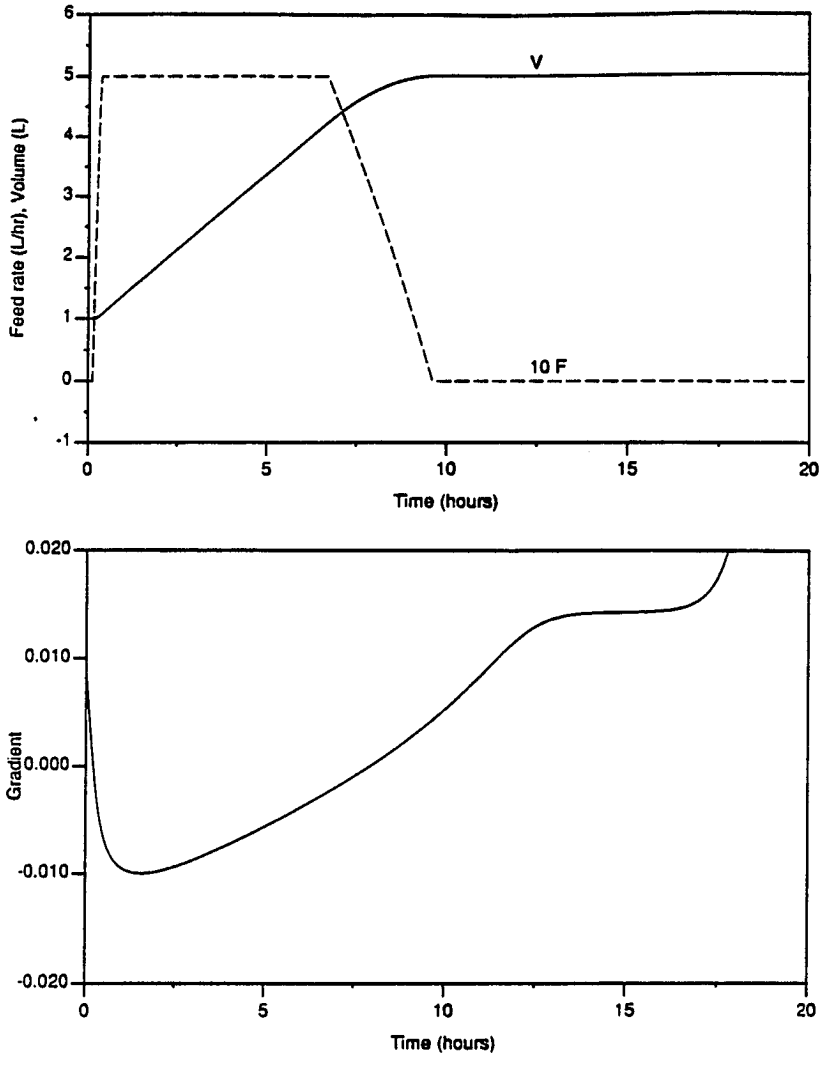

Figure 6.12 Optimization of DiBiasio's model (Case B): Feed rate, volume and gradient of the Hamiltonian.

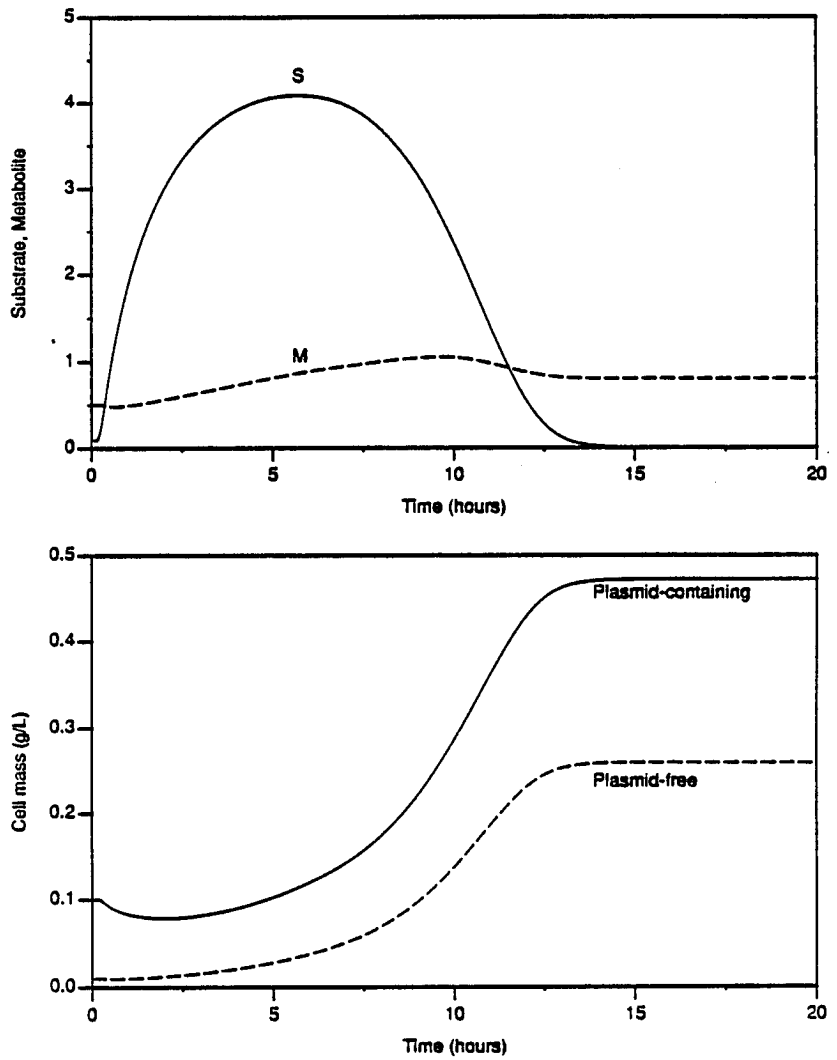

Figure 6.13 Optimization of DiBiasio's model (Case B): Cell mass, substrate and metabolite.

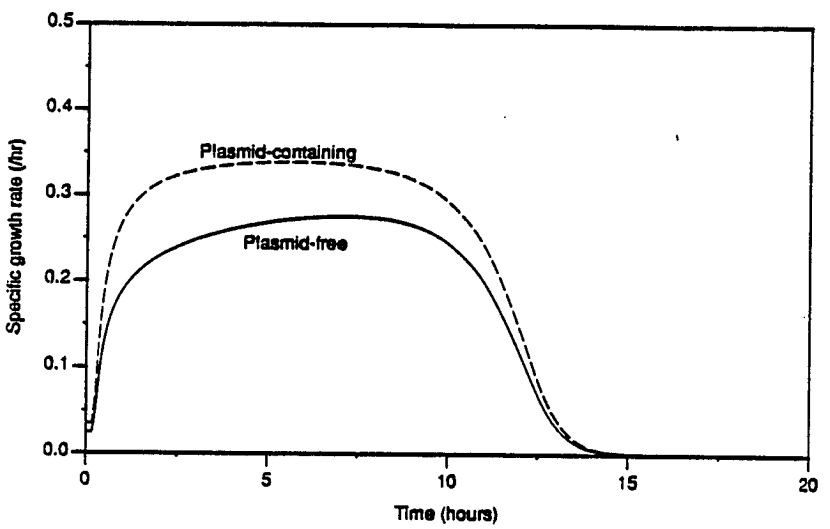

Figure 6.14 Optimization of DiBiasio's model (Case B): Specific growth rates of plasmid-containing and plasmid-free cells.

where

$$\mu_+ = \frac{\mu_{max} S}{K_S + S}, \quad \mu_- = \frac{\mu_+ M}{K_M + M} \tag{6.30}$$

In the above equations p represents the probability of plasmid loss, μ_i represent the specific growth rates, X^+ and X^- represent the cell mass concentration for the two types of cells, and Y_S and Y_M are yield coefficients. The growth of plasmid-carrying cells is limited by a single limiting substrate (S), and was described using a Monod form. Growth of plasmid-free cells is limited by this substrate as well as an additional metabolite (M). The rate of metabolite generation by plasmid-carrying cells was described by assuming it to be associated with cell growth.

For this system a suitable performance index can be to maximize the amount of plasmid-containing cells at the end of fermentation. This might be the case, for example, if plasmid-containing cells constitutively produce an intracellular product.

$$\min_{F(t)} \Pi = -(X^+V)_{t_f} + K(V(t_f) - V_f)^2 \tag{6.31}$$

6.6.2 Results This optimization problem was solved using the conjugate gradient technique outlined before. The parameters used for optimization are listed in Table (6.2). Two different cases were studied. In Case B, the concentration of the metabolite M is assumed to be ten times higher than that in Case A. Other than that, the cases are identical. Results of optimization are shown in Figs. (6.9-6.14).

In both cases, the optimal feed rate profile consists of an initial batch period, followed by a period of maximum feed rate and a batch period at the end. No singular period was observed. Although the final time was chosen to be 20 hours, the cells stop growing at around 15 hours. Thus, for any final time between 15 and 20 hours, the optimal feed rate profiles should be identical. The gradient of the Hamiltonian shows expected trends, which confirm the optimality of the converged feed rate profile.

As the substrate concentration increases, the specific growth rate of plasmid-containing cells increases. However, this results in an increase in the rate of production of the metabolite M, which increases the specific growth rate of plasmid-free cells. Figs. (6.11) and (6.14) illustrate this phenomenon. An increase in the initial concentration of the metabolite (Case B) did not have much effect on the feed rate profiles.

6.7 Optimization of invertase production

After successful application of the conjugate gradient method to two different fermentation models of differing complexity, we return to the main objective of this research: optimization of invertase production from recombinant yeast fermentations.

6.7.1 Formulation The model equations, developed in Chapter 4, can be rewritten here for ready reference.

$$\frac{d}{dt}(XV) = \mu XV = (R_r Y_{xr} + R_f Y_{xf}) XV \tag{6.32}$$

$$\frac{d}{dt}(SV) = FS_F - R_t XV \tag{6.33}$$

$$\frac{d}{dt}(PXV) = \left[\frac{k_p S}{K_p + S + K_i S^2} - k_d P\right] XV \tag{6.34}$$

$$\frac{dV}{dt} = F \tag{6.35}$$

The aim of optimization if to maximize the total invertase activity at the end of the fermentation. The final time t_f is assumed to be fixed.

$$\min_{F(t)} \Pi = -(PXV)_{t_f} + K(V(t_f) - V_f)^2 \tag{6.36}$$

This is a four-dimensional singular control problem, and was solved using the conjugate gradient technique, described in detail in the preceding sections.

6.7.2 Results Three distinct cases were considered to study the effect of changes in the initial glucose concentration, allowed maximum feed rate, and the specified final time. The parameters used for optimization are listed in Table (6.3). The optimal state and control profiles are shown in Figs. (6.15-26).

In Case A, the initial glucose concentration was chosen to be 2 g/L. The optimal feed rate contains an initial maximum feed rate, which increases the glucose concentration to more than 5 g/L, thus increasing the specific cell growth rate. At these glucose levels, invertase production is repressed and the specific invertase activity is almost constant. This period is followed by a batch period, in which the glucose

Table 6.3 Parameters used for optimization of invertase production

Case	A	B	C
X(0), OD	0.15	0.15	0.15
S(0), g/L	2.0	2.0	5.0
P(0), (units/OD.L)	0.1	0.1	0.1
V(0), L	0.6	0.6	0.6
V_f, L	1.2	1.2	1.2
S_F, g/L	10.0	10.0	10.0
F_{max}, L/hr	0.6	0.2	0.6
t_f, hr	10.0	10.0	9,10,12,14

Table 6.4 Productivity vs. final time (Case C)

Final time, hr	Invertase productivity, units/hr
9.0	0.3560
10.0	0.6157
12.0	0.7032
14.0	0.6057

concentration slowly drops, resulting in a gradual increase in the specific invertase activity. During the singular period, the glucose concentration stays almost constant around 0.225 g/L. Invertase production rate is very high during this singular phase. A small batch period follows, when the fermentor is full and the singular feed rate can no longer be implemented. Thus the optimal feed rate profile clearly results in an initial high cell growth rate phase followed by a high invertase production rate phase.

The effect of change in the maximum allowed feed rate was studied by decreasing its value to 0.2 L/hr (Case B). The only effect was a change in the duration of the maximum, batch and singular periods. The initial maximum feed rate period was longer to allow high enough glucose concentration to maximize cell growth at the beginning of the fermentation. Again the singular period maintains the glucose concentration at around 0.225 g/L.

In all the cases discussed in this chapter, the final time was always kept fixed. It is not easy to incorporate variable final time in the conjugate gradient approach in a simple and direct manner. Thus to determine the optimum final time for maximum invertase productivity, it is necessary to carry out the optimization at different fixed final times (Case C). The initial glucose concentration was chosen to be 5 g/L. At very short final times (9 hr.), no singular feed rate period was observed. A possible explanation for this can be given. Because of the final volume constraint, the fermentor has to be completely filled during a shorter time. Thus allowing a singular feed rate will result in a severe penalty in the performance index, because the actual final volume will be much lower than the desired value. Because of the absence of singular period, the invertase productivity is very low.

At a higher value of final time (10 hr.), the feed rate profile is very similar to those in Cases A and B. Presence of singular period results in an increased invertase productivity. A still higher invertase productivity is obtained at a final time of 12 hr. A distinctive feature of the singular period here is that the glucose concentration is initially constant at a little high value. After about three hours of singular feed rate, the glucose value drops to 0.225 g/L and stays constant there.

As the final time is increased further (14 hr), the invertase productivity drops because of another phenomenon. In this case, the final time is so long that the singular period cannot maintain the glucose concentration at its optimum value of about 0.225 g/L during the initial stages. The optimal policy results in a period in which the glucose concentration is very low (about 0.02 g/L). This results in a sharp drop in

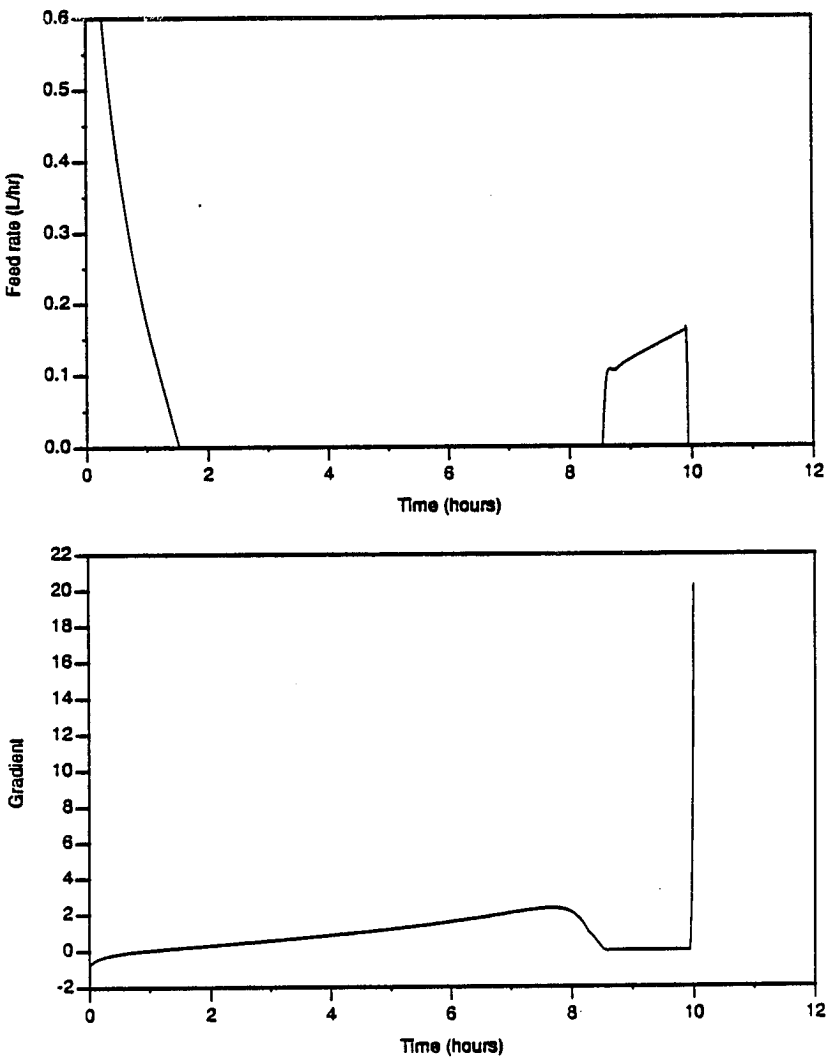

Figure 6.15 Optimization of invertase production (Case A): Optimal flow rate and gradient of the Hamiltonian.

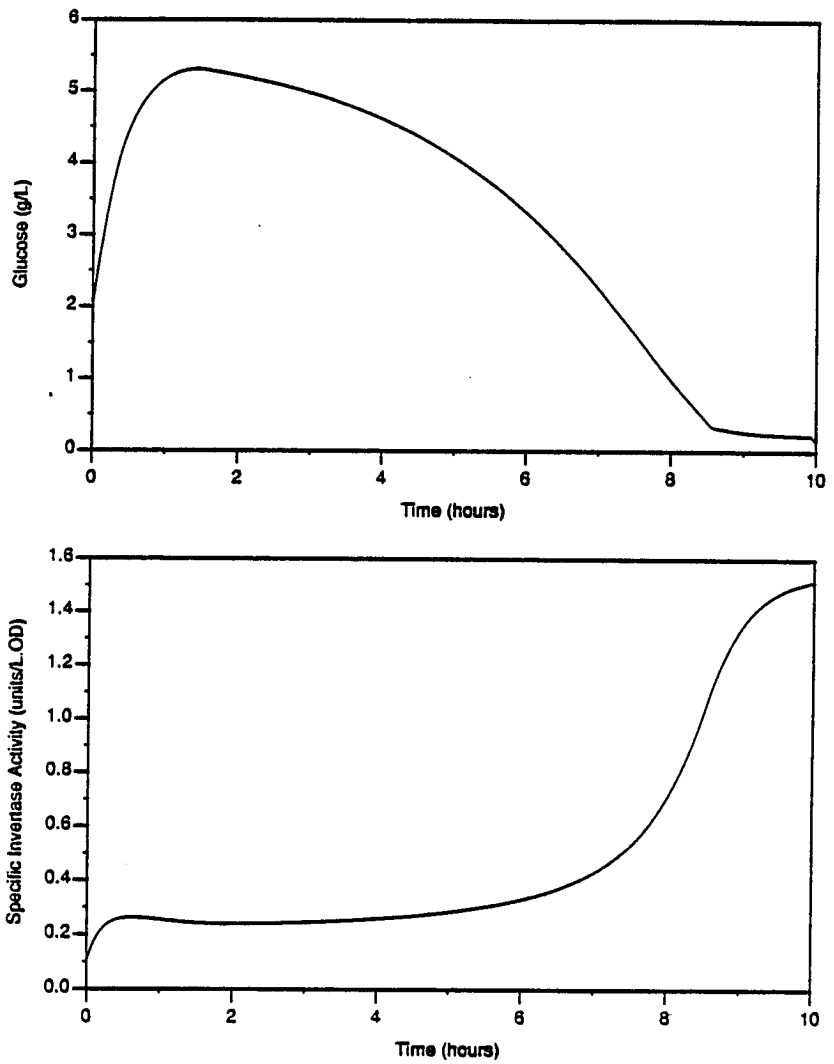

Figure 6.16 Optimization of invertase production (Case A): Optimal substrate and invertase concentration.

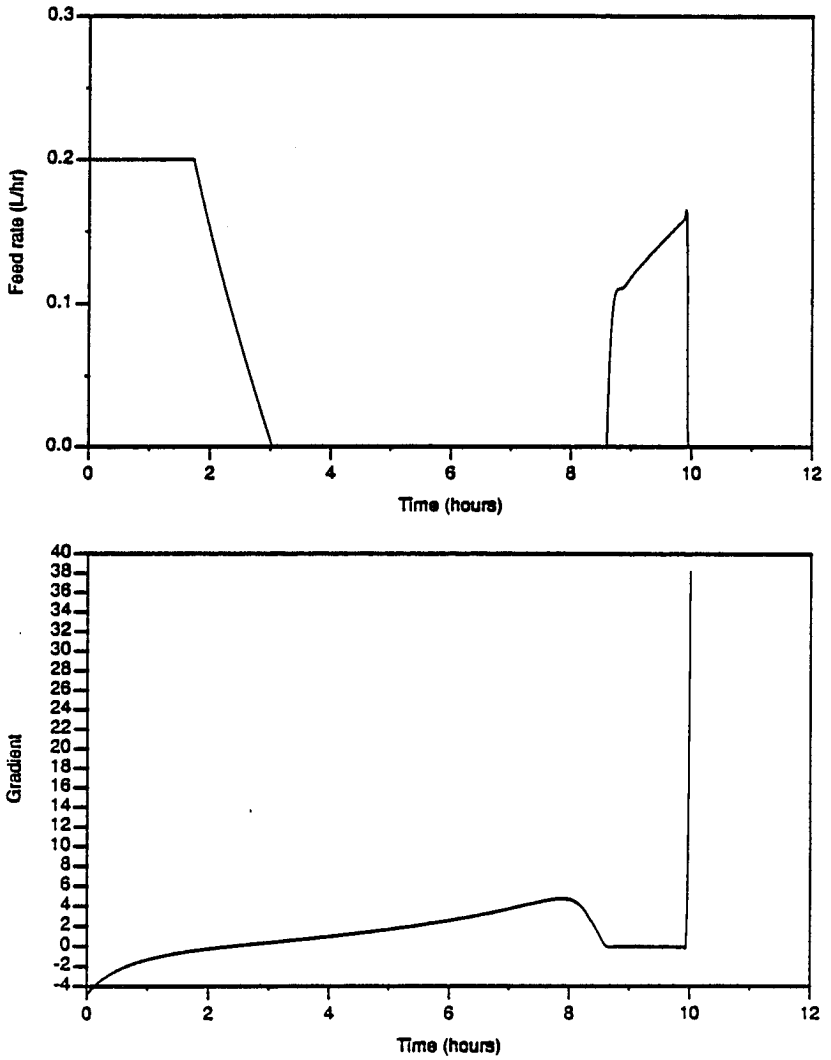

Figure 6.17 Optimization of invertase production (Case B): Optimal flow rate and gradient of the Hamiltonian.

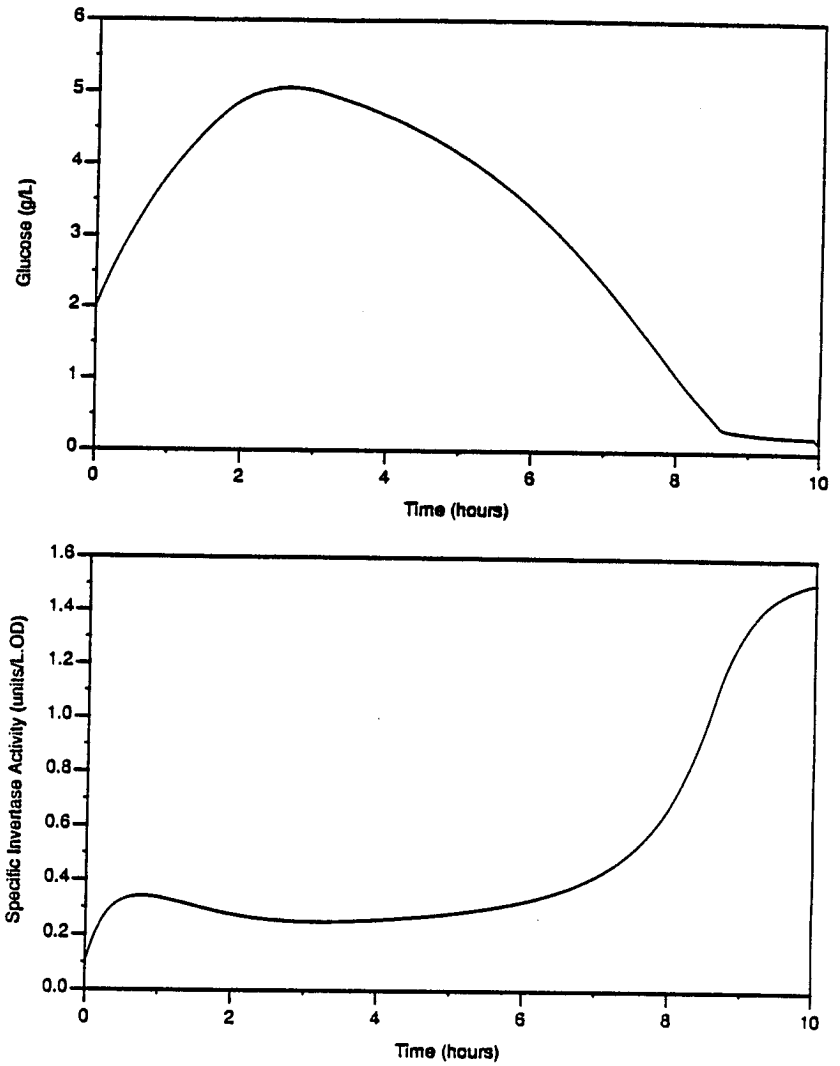

Figure 6.18 Optimization of invertase production (Case B): Optimal substrate and invertase concentration.

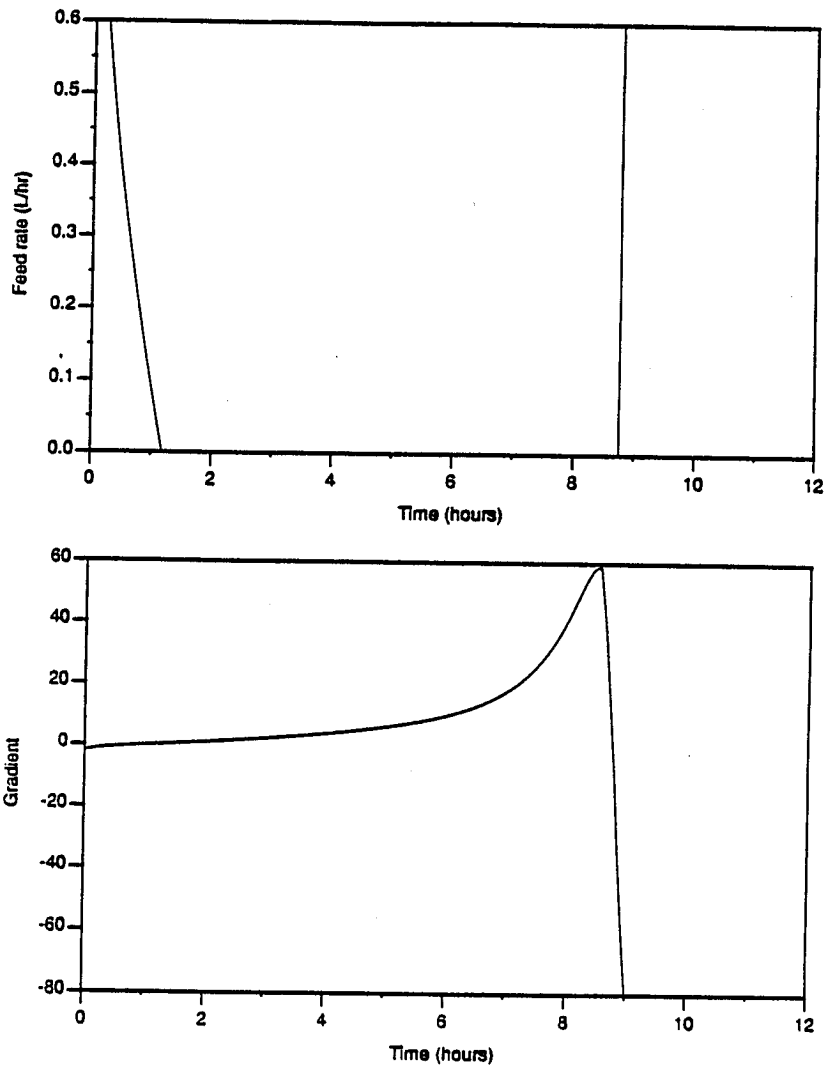

Figure 6.19 Optimization of invertase production (Case C): Optimal flow rate and gradient of the Hamiltonian (Final time = 9 hr.)

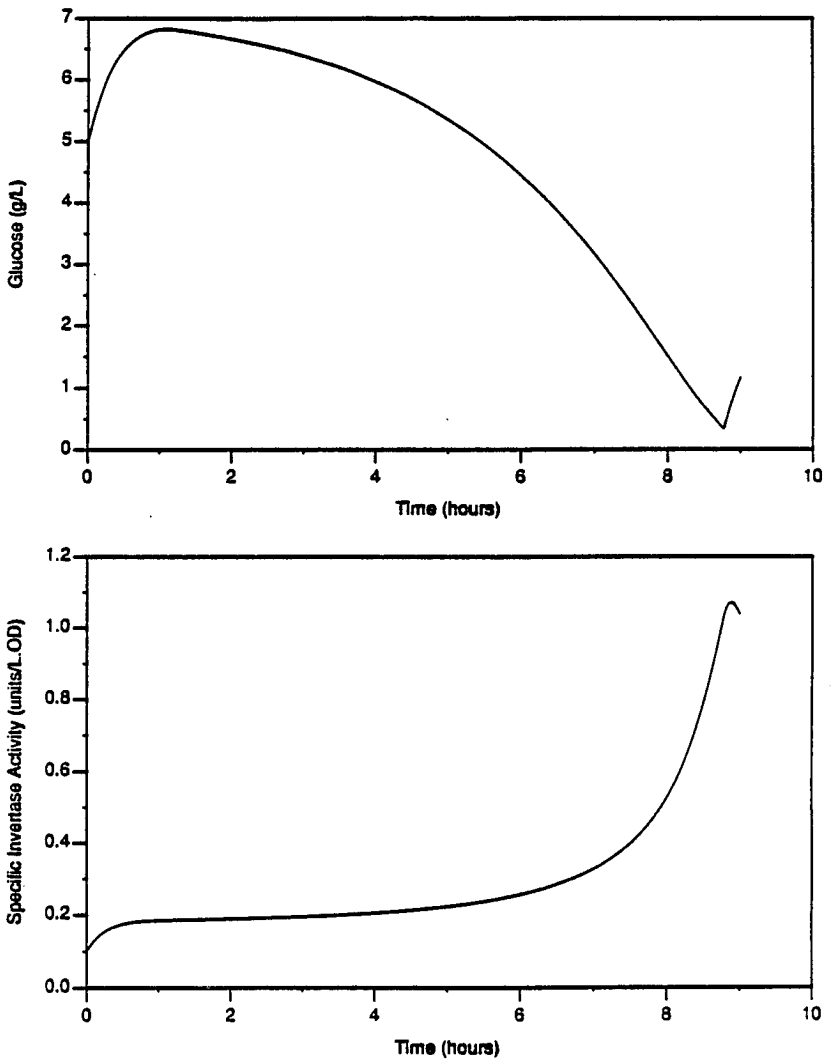

Figure 6.20 Optimization of invertase production (Case C): Optimal substrate and invertase concentration (Final time = 9 hr.)

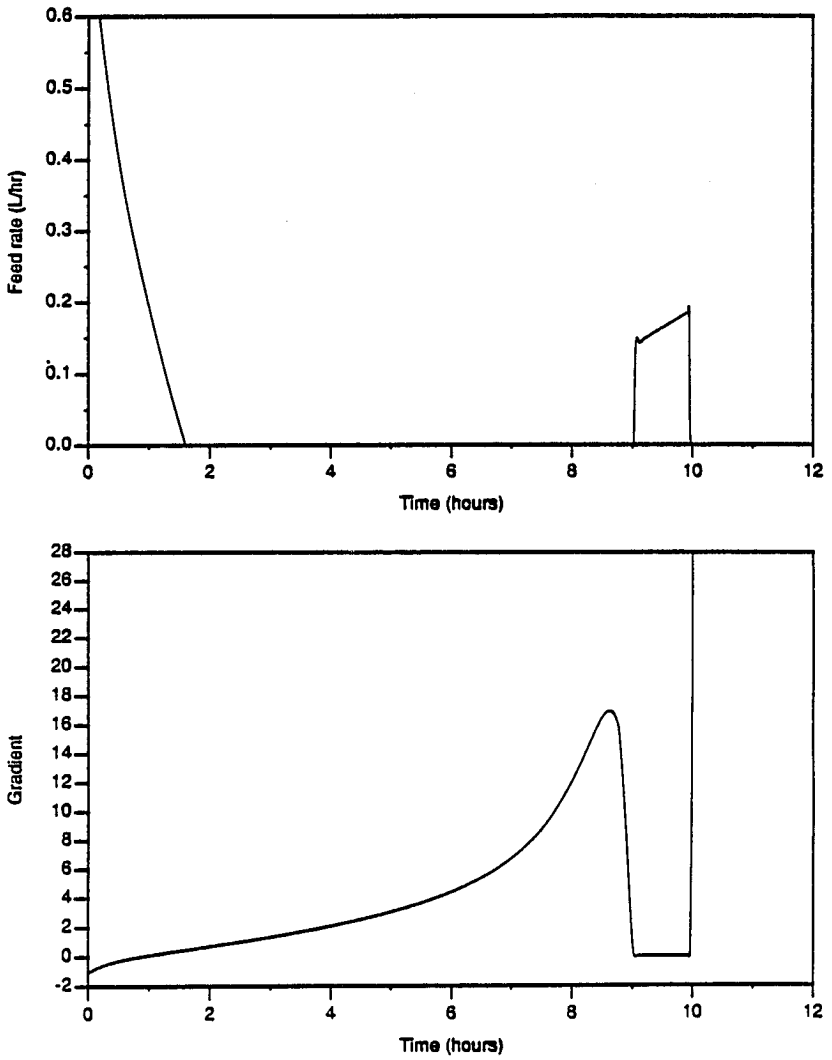

Figure 6.21 Optimization of invertase production (Case C): Optimal flow rate and gradient of the Hamiltonian (Final time = 10 hr.)

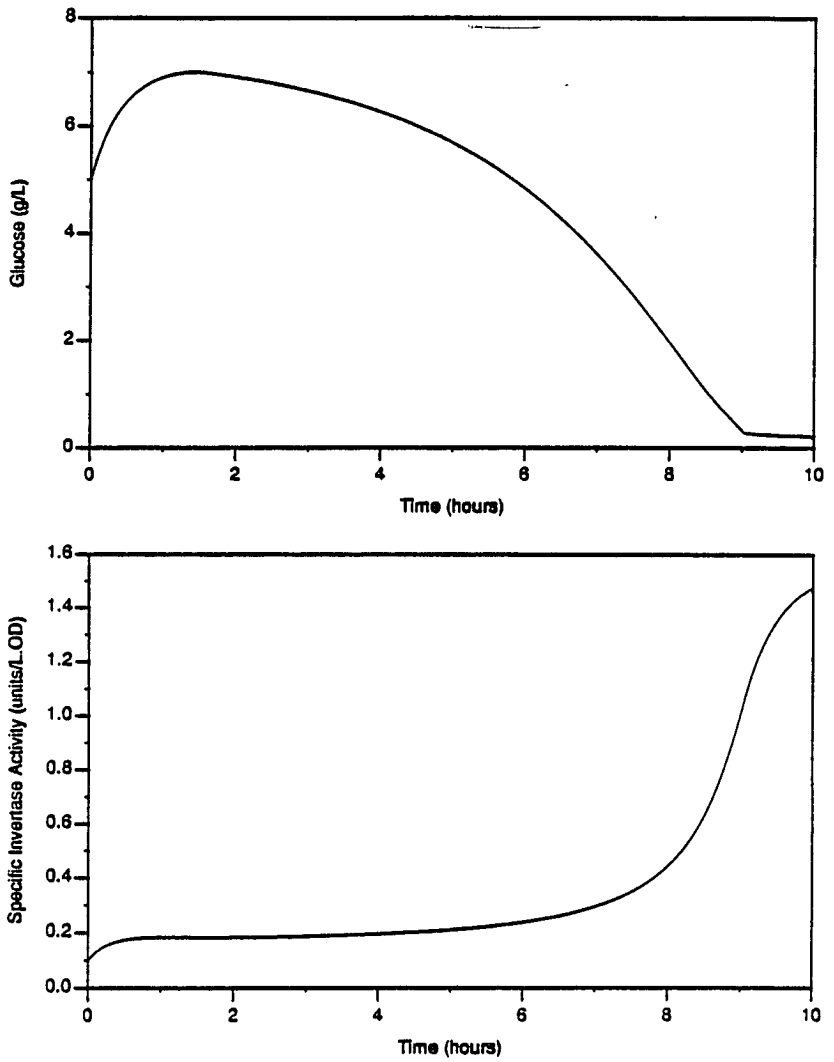

Figure 6.22 Optimization of invertase production (Case C): Optimal substrate and invertase concentration (Final time = 10 hr.)

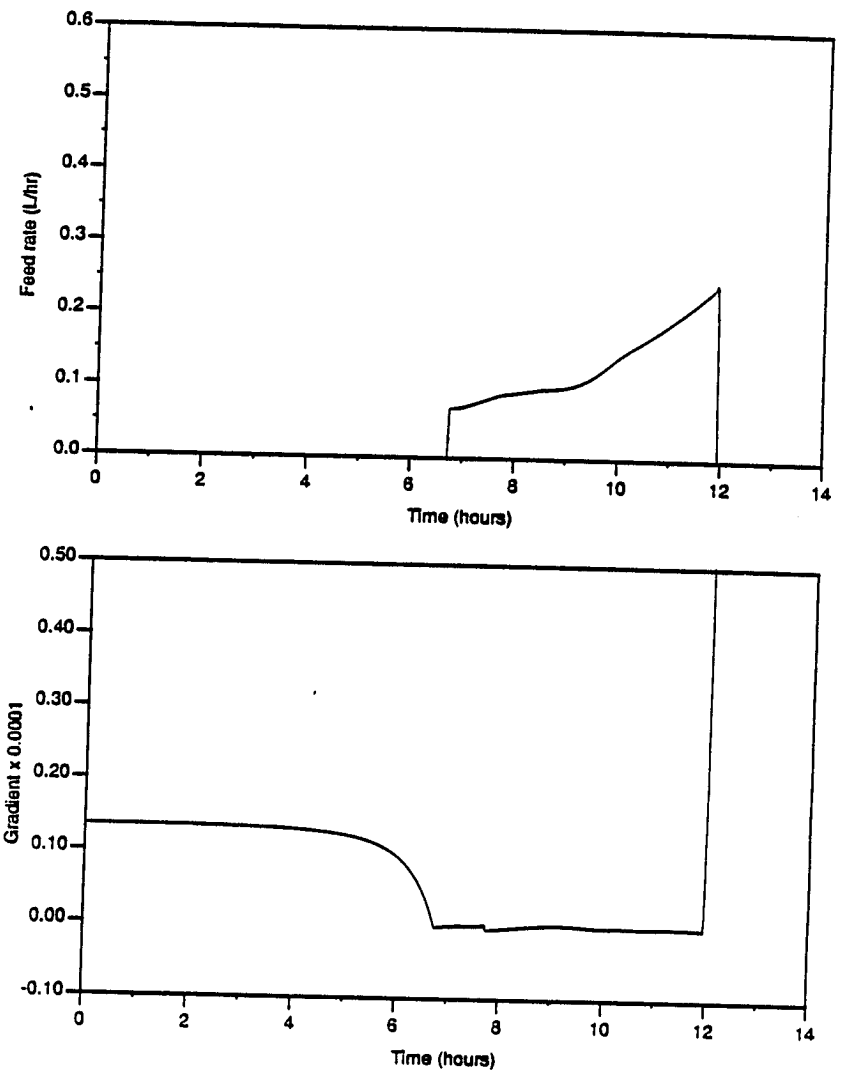

Figure 6.23 Optimization of invertase production (Case C): Optimal flow rate and gradient of the Hamiltonian (Final time = 12 hr.)

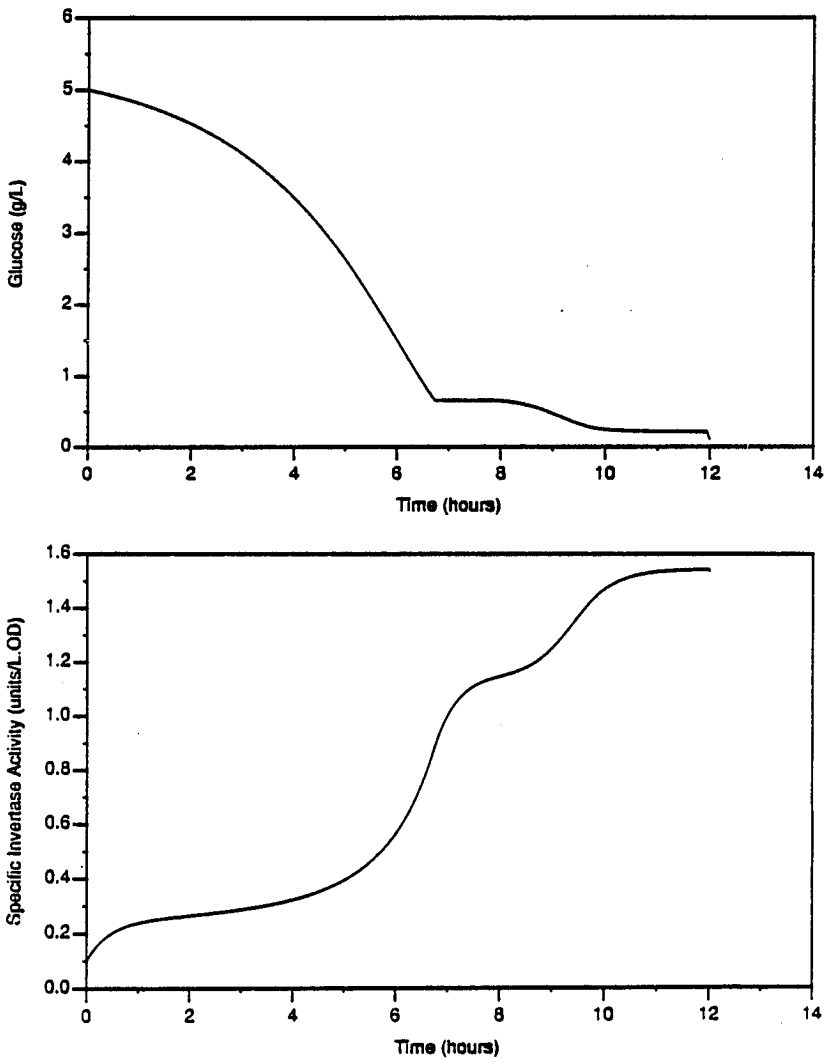

Figure 6.24 Optimization of invertase production (Case C): Optimal substrate and invertase concentration (Final time = 12 hr.)

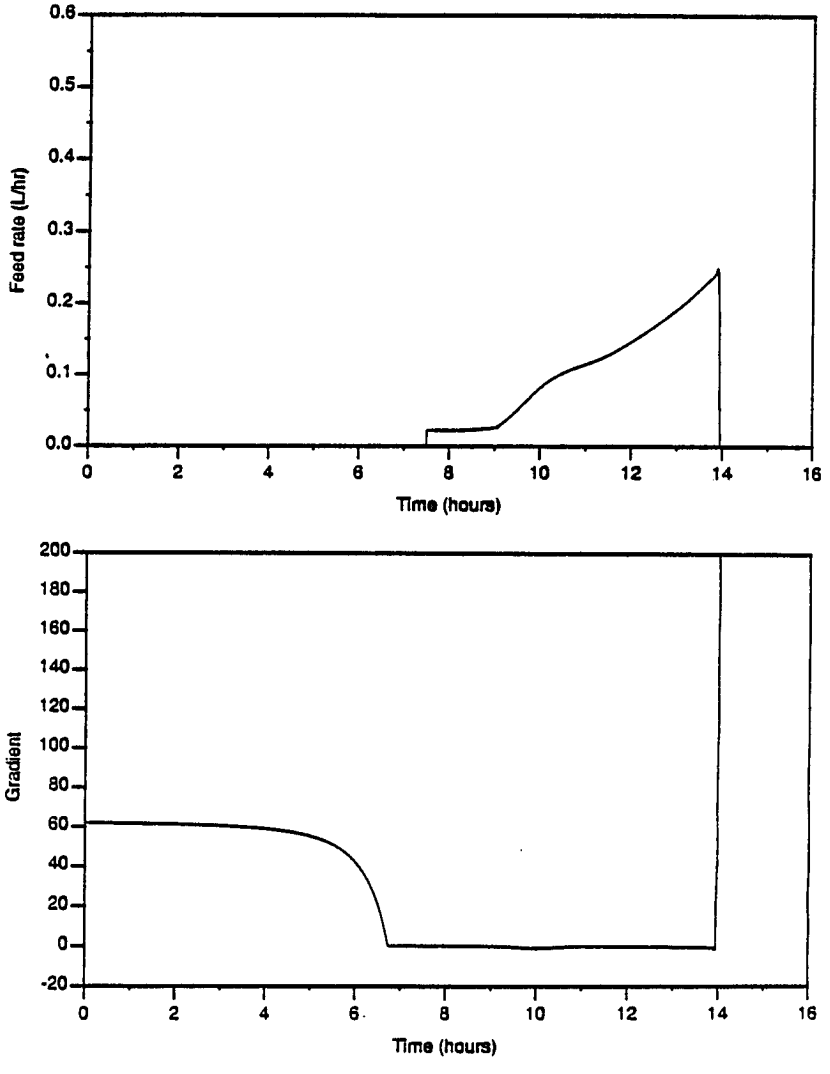

Figure 6.25 Optimization of invertase production (Case C): Optimal flow rate and gradient of the Hamiltonian (Final time = 14 hr.)

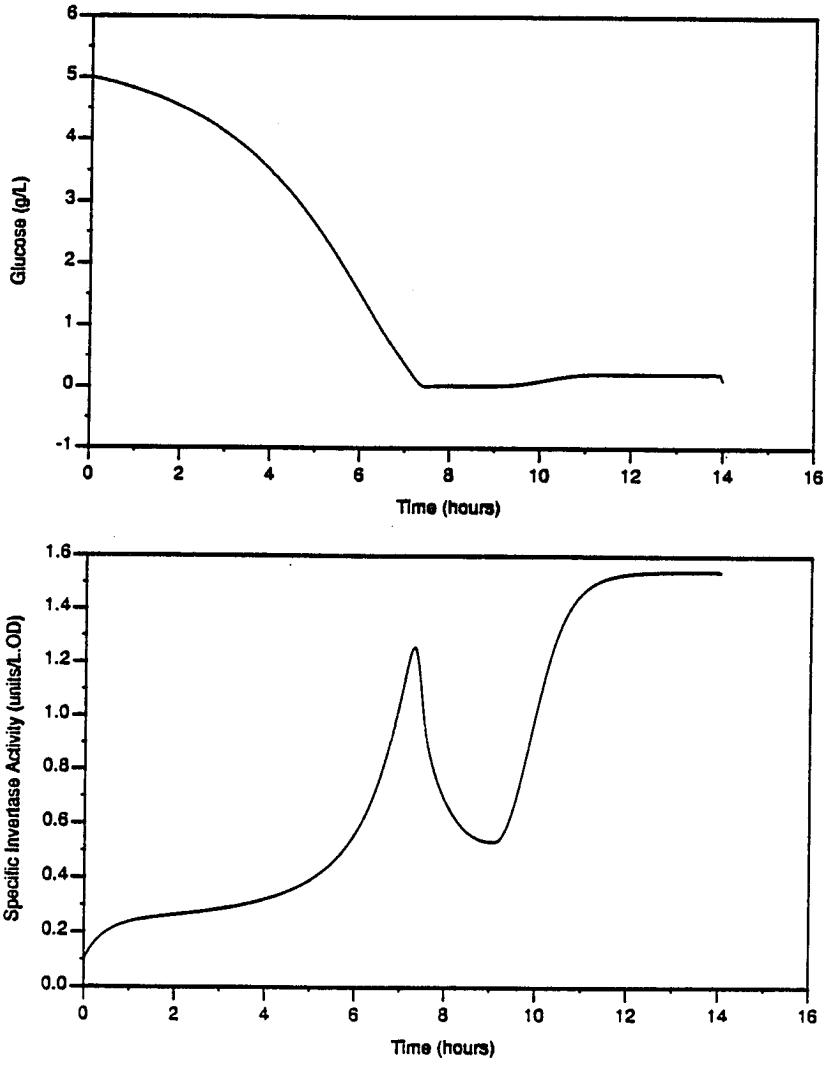

Figure 6.26 Optimization of invertase production (Case C): Optimal substrate and invertase concentration (Final time = 14 hr.)

invertase production rate. The specific invertase activity starts decreasing for a short while. This results in a low invertase productivity.

6.8 Conclusions

A first-order conjugate gradient technique was demonstrated to be effective in solving a variety of optimization problems. Although the convergence of this technique is somewhat slow when the optimum is approached, the method converged to the correct optimal profile. The correctness of the optimal profile can be judged by the variation in the gradient of the Hamiltonian. The gradient was found to be zero during singular periods, and had appropriate sign on boundary control arcs.

A boundary condition iteration method, previously developed by Modak, fails for high dimensional systems, because of two reasons. First, it becomes more and more difficult to guess the control arc sequence as the system dimension increases. Second, more adjoint variables need to be guessed at the junction points. The conjugate gradient, on the other hand, does not need *a priori* guesses of control arc sequences. After a simple initial guess of the control variable profile, the method proceeds in a smooth manner.

One of the attractive features of this method is its simplicity—the method requires only a little more computation than the steepest descent approach. The presence of boundary arcs in the control profile actually results in a slightly reduced computation time per iteration.

The method proposed here does not need any variable transformation to convert the singular problem to a nonsingular one. This is an advantage, because the variable transformations result in constraints on the rate of change of control variable. Appropriate theory to deal with such constraints is not currently available.

7. OPTIMIZATION OF ANAEROBIC YEAST FERMENTATIONS

7.1 Introduction

Pyun *et al.*[102] performed continuous culture experiments using *Saccharomyces cerevisiae* SEY2102 containing the plasmid pRB58. They used selective medium to restrict the growth of plasmid-free cells. It was found that under anaerobic conditions specific invertase activity obtained was 18-times that under aerobic conditions. Also the specific invertase activity in the fermentor indicated a maximum as a function of the dilution rate. These results suggest that an optimum operating policy for maximizing invertase productivity can be determined. The aim of this study is to determine the optimum glucose feed rate to maximize invertase productivity in a fed-batch mode of operation. To achieve this objective a mathematical model was developed to describe anaerobic fermentation kinetics.

7.2 Model development

Even though selective medium was used, continuous fermentations show that a significant (10-17%) fraction of cells had lost the recombinant plasmid. However, considering the experimental errors involved in determining fraction of plasmid-containing cells, the composition of the cell population can be considered to be fairly constant over the dilution rates used. For modeling purposes, the cell mass concentration was described by a single variable X (g. cells/lit). Thus cell mass balance can be written as

$$\frac{dX}{dt} = (\mu - D)X \tag{7.1}$$

Specific growth rate μ is assumed to follow the Monod form.

$$\mu = \frac{\mu_{max} S}{K_S + S} \tag{7.2}$$

Since at steady state the specific growth rate equals the dilution rate, a plot of dilution rate vs. glucose concentration was used to estimate the growth parameters (Fig. 7.1). The parameters are listed in Table (7.1). Since under aerobic conditions growth is completely fermentative it is reasonable to assume constant cell yield.

$$\frac{dS}{dt} = -\frac{\mu X}{Y} + D(S_F - S) \tag{7.3}$$

A balance on the amount of invertase in the reactor results in

$$\frac{d}{dt}(PX) = \pi X - DPX \tag{7.4}$$

where P is specific invertase activity (KU/g.cell) and π is specific invertase formation rate. From the steady state results π can be determined as a function of glucose concentration in the fermentor. The results are plotted in Fig. (7.2). At high glucose concentrations invertase production is repressed. The rate of invertase production also goes down as the glucose concentration drops below a critical value. Any model for this system must be able to describe this bell-shaped dependence.

Sentandreu *et al.* have studied kinetics of invertase production in *S. cerevisiae*[103]. Their data indicate that rate of translation of invertase mRNA can be described by saturation kinetics. If P is the specific concentration of invertase, a balance on invertase for a batch operation gives

$$\frac{dP}{dt} = \frac{k_{transl}N}{K_{transl} + N} - (\mu + D)P, \tag{7.5}$$

where N is intracellular invertase mRNA concentration. Degradation of invertase is assumed to be negligible. Elorza *et al.*[92] have investigated the stability of invertase mRNA. They found that the decay rate can be described by a first-order kinetics. Thus a balance on mRNA can be written as

$$\frac{dN}{dt} = r_n - (\mu + D + k_d)N \tag{7.6}$$

The rate of transcription of SUC2 gene, r_N, decreases with an increase in the glucose concentration in the media. A simple way to describe this will be,

$$r_n = \frac{k_{transc}}{K_{transc} + S} \tag{7.7}$$

In the above analysis we have assumed that both k_{transl} and k_{transc} are independent of the

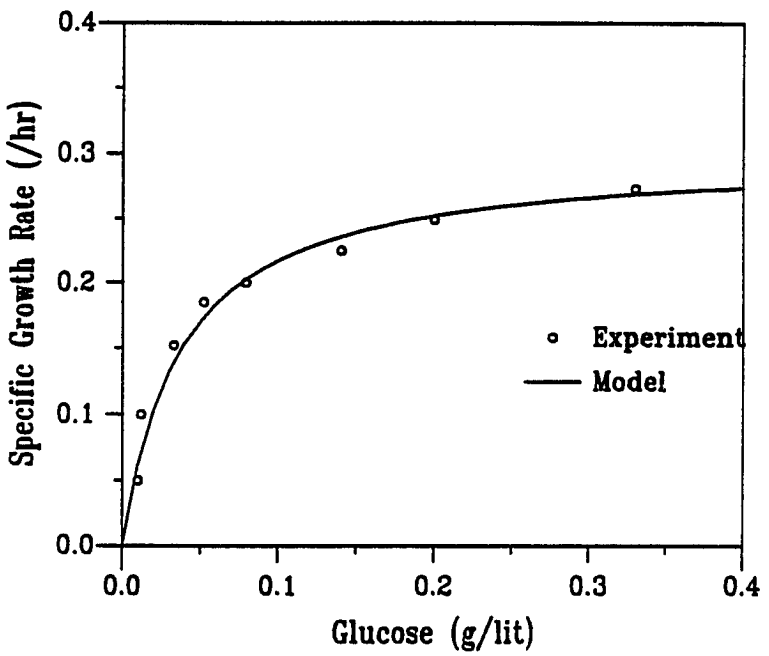

Figure 7.1 Determination of μ_{max} and K_S by fitting the experimental dependence of specific growth rate on glucose concentration.

Table 7.1 Model parameters

μ_{max}	0.3 /hr
K_S	0.038 g/L
Y	0.16 (g. cell)/(g. glucose)
k_p	190 KU.(g. glucose)/(g. cell)
K^p	0.01 (g. glucose)/hr.L
k_d	1.3 /hr
K_q	0.013 g/L

cell growth rate. However, a cell needs resources for translation and transcription. These resources, like the concentration of ribosomes inside the cell, are known to decrease at low growth rates. Thus it appears reasonable to assume that both k_{transc} and k_{transl} are proportional to the specific growth rate μ. If we assume pseudo-steady state for mRNA, equations (5-7) can be combined to give,

$$\pi = \frac{k_p \mu^2}{K_p + (k_d + \mu)(K_q + S)} \tag{7.8}$$

The decay constant k_d was assumed to be 1.3 hr^{-1} as determined by Elorza *et al.*[92]. The other two constants in the above expression were determined by fitting the model to the experimental results. The model is able to describe the experimental data (Fig. 7.3).

7.3 Differences between models for aerobic and anaerobic growth

The model described in the previous section differs in some important respects from the model developed in Chapter 5. The reason for this is primarily that the model described here was developed before the aerobic growth model. The functional forms used in anaerobic growth model were found to be inadequate to describe aerobic fermentations.

Under aerobic conditions, yeast can use both respiratory and fermentative metabolism for growth. In fact, the regulation of respiratory and fermentative fluxes, resulting in variable cell mass yield, forms the backbone of the aerobic growth model. On the other hand, under anaerobic conditions, yeast, for lack of oxygen, cannot use respiratory pathway. Since only one growth metabolism is available, it is reasonable to assume constant cell yield.

Under anaerobic conditions, it was found necessary to assume that the rates of transcription and translation were dependent on the cell growth rate. No such assumption was involved in explaining invertase production under aerobic conditions. A possible reason for this is that the amount of invertase produced under anaerobic conditions is significantly (18 times) higher than that under aerobic conditions. Thus the translation and transcription machinery of the yeast cell is under more strain to supply the necessary ingredients like ribosomes and amino acids. This effect becomes more prominent because invertase is derepressed at low glucose concentrations, when the specific growth rate is very low. Thus it is reasonable to assume that at very low glucose levels the rate of supply of components of translation and transcription machinery becomes the rate limiting step.

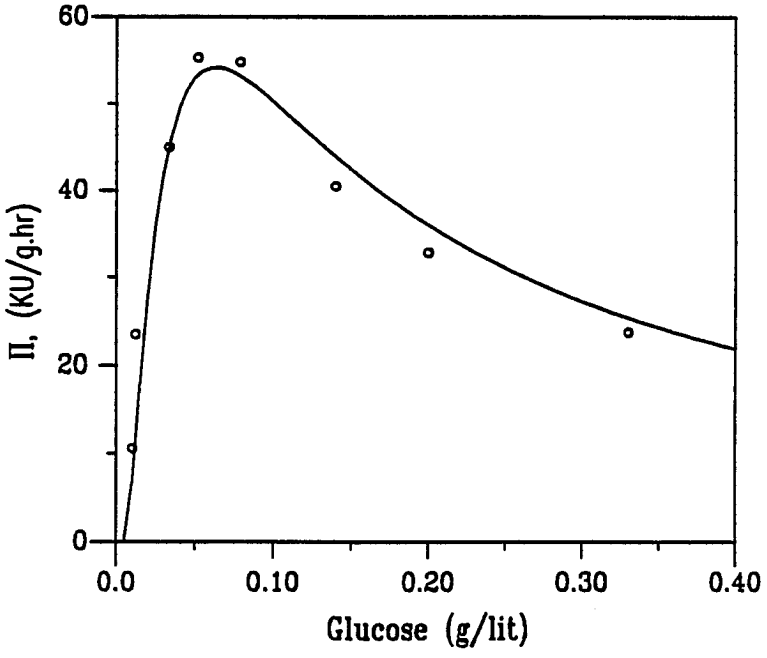

Figure 7.2 Rate of invertase production: Comparison between model and experimental results.

The experiments done by Pyun *et al.* involved much lower glucose concentrations compared to the aerobic experiments performed during this research. Thus the glucose repression of invertase production observed at high glucose levels was not that prominent under the conditions of anaerobic experiments. It was, therefore, not necessary to incorporate the S^2 factor in the denominator of Eq. (7.8).

7.4 Optimization

7.4.1 Problem definition Modak *et al.* have discussed the general theory of optimal control as applied to fermentation processes[50], as well as a computational algorithm to calculate optimal feed rates for fed-batch fermentations[1]. These studies form the basis for the numerical results presented in this chapter.

Here we solve the problem of determining the optimal feed rate profile which will minimize a given performance function in fed-batch mode of operation. First, state variables are defined as follows.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} XV \\ SV \\ PXV \\ V \end{bmatrix} \tag{7.9}$$

The differential mass balance equations are

$$\dot{\mathbf{x}} = \mathbf{f} = \begin{bmatrix} \mu x_1 \\ -\dfrac{\mu x_1}{Y} + FS_F \\ \pi x_1 \\ F \end{bmatrix}, \quad \mathbf{x}(0) = \mathbf{x}_0 \tag{7.10}$$

The objective is to minimize the performance function Π by varying substrate feed rate,

$$\min_{F(t)} \Pi = x_3(t_f) - \varepsilon t_f \tag{7.11}$$

The final time t_f is assumed to be free. The final reactor volume and substrate feed rate are constrained.

$$V(t_f) = V_f, \qquad 0 = F_{min} \leq F(t) \leq F_{max} \tag{7.12}$$

This optimization problem falls under the realm of the calculus of variations. Pontryagin's minimum principle[36] asserts that the Hamiltonian defined below must be minimized.

$$H(x, \lambda, F) = \lambda^T f = \psi + \phi F, \tag{7.13}$$

where

$$\psi = (\lambda_1 \mu - \frac{\lambda_2 \mu}{Y} + \lambda_3 \pi) x_1 - \varepsilon, \qquad \phi = \lambda_2 S_F + \lambda_4$$

The adjoint vector λ is defined by

$$\dot{\lambda} = -\frac{\partial H}{\partial x}, \qquad \lambda(t_f) = \frac{\partial \Pi}{\partial x}(t_f) \tag{7.14}$$

As shown by Modak *et al*[50], for minimizing the Hamiltonian H, the substrate feed rate F must be given by

$$F(t) = \begin{cases} F_{max} & \text{if } \phi < 0 \\ F_s & \text{if } \phi = 0; \text{ t member } [t_1, t_2] \\ 0 & \text{if } \phi > 0 \end{cases} \tag{7.15}$$

where $[t_1, t_2]$ is a finite closed interval of time. For the system discussed here, the singular control F_s can be obtained by[50]

$$F_s = \frac{1}{S_F - x_2/x_4} \left[\frac{\mu x_1}{Y} - \frac{x_4 \mu'(\pi'\mu - \pi\mu')}{(\pi''\mu' - \pi'\mu'')} \right], \tag{7.16}$$

where the primes denote differentiation with respect to S.

7.4.2 Computational Procedure Computation of optimum feed rate requires the solution of the two-point boundary value problem defined by Eqs. (7.10) and (7.14) with the optimal feed rate given by Eqs. (7.15) and (7.16). A modification of the procedure developed by Lim *et al.*[1] is used here.

As outlined in the previous section, the optimal feed rate can contain only the maximum, minimum and singular arcs. However, the current theory can not *a priori* determine the exact sequence of these arcs. However, an educated guess can be made based on a conjecture proposed by Modak *et al*[50]. Using this conjecture the optimal sequence can be either maximum-minimum-singular-minimum (Type A) or minimum-maximum-singular-minimum (Type B) depending upon the specified initial conditions.

Thus the problem now reduces to the determination of optimal switching times between the different arcs. From Eq. (7.16), it is easy to see that the singular feed rate is independent of the adjoint variables. This observation allows us to completely dispense with the solution of adjoint equations, thus reducing the computation time considerably. The numerical procedure is outlined below.

1. Assume a sequence of minimum, maximum and singular arcs.
2. Guess the time of switching t_1 between minimum and maximum feed rates.
3. Guess the switching time t_2 between the minimum/maximum and singular feed rate.
4. Integrate the state differential Eqs. (7.10) till the fermentor is full ($V = V_f$).
5. Integrate the state differential Eqs. (7.10) with zero feed rate till the stopping condition (derived below) is satisfied within a desired tolerance.
6. Store the value of the performance function and return to step (2).
7. The switching times which result in maximum performance function are the required optimal switching times.

Since the final time t_f is assumed to be free, the Hamiltonian H must be zero for t member [0, t_f]. In particular,

$$H(t_f) = [-\pi x_1 + \varepsilon]_{t_f} = 0 \tag{7.17}$$

Thus in step (5) we use the stopping condition

$$(\pi x_1)_{t_f} = \varepsilon \tag{7.18}$$

This condition stops the fermentor operation when the increase in the performance index because of product formation balances the decrease in the same because of the operating cost.

7.4.3 Numerical Results Using the numerical procedure outlined in the previous section, optimum substrate feed rates were determined for different initial cell mass and substrate concentrations. The results are summarized in Table (7.2)

Intuitively, the optimum feed rate should result in an initial fast growth of cells at high glucose concentration followed by increased rate of invertase production at low glucose concentrations. In case of low initial cell mass and glucose concentrations, the optimum profile was found to consist of an initial high feed rate to increase the glucose concentration followed by a batch period. The singular feed rate brings the glucose concentration to a level which maximizes the ratio (π/μ). Similarly for high initial glucose concentration the optimum profile consists of an initial batch period to lower the glucose concentration followed by a very short maximum feed rate period. Fig. (7.3) shows the optimal feed rate profile for a case of high initial glucose concentration. The optimal cell mass and glucose concentration profiles are shown in Fig (7.4). During the initial batch period the glucose concentration falls to a low value. The singular feed rate keeps the glucose concentration almost constant to keep the π/μ ration near its maximum.

Varying the maximum allowed feed rate does not affect the switching times and the performance function significantly. However, if the operating cost is increased by increasing ε the switching times change drastically. At higher values of ε the initial period of maximum feed rate is prolonged considerably so that the time taken for the fermentation cycle is reduced.

The performance of optimized fed-batch operation is compared with a batch operation using the same total amount of glucose (Table 7.2). For low operating costs (ε = 0.1), about 60% improvement is achieved. However, as the operating cost increases (ε = 1.0) the improvement is reduced to 35%. This is reasonable because the batch operation takes less time than the optimized fed-batch operation.

Table 7.2: Results of optimization.

A. Parameters common to all sets

P(0)	0.0	KU / g. cell
S_F	0.3	g / lit
V(0)	1	lit
V_f	5	lit

B. Switching times and performance functions for different sets of parameters

Conditions				Results						
X(0)	S(0)	F_{max}	ε	Type*	t_1	t_2	t_3	t_f	P_f	P_b
0.02	0.02	1.0	0.1	A	0.043	0.049	18.58	20.42	65.35	41.12
0.02	0.02	0.8	0.1	A	0.055	0.065	18.58	20.42	65.35	41.12
0.02	0.02	1.0	1.0	A	0.34	0.76	16.01	17.21	48.61	36.09
0.02	0.2	1.0	0.1	B	4.06	4.07	16.73	18.38	72.20	44.36
0.2	0.2	1.0	0.1	B	0.63	0.64	5.39	6.37	73.41	44.86

* Type A refers to the sequence maximum-minimum-singular-batch, while type B refers to the sequence minimum-maximum-singular-batch

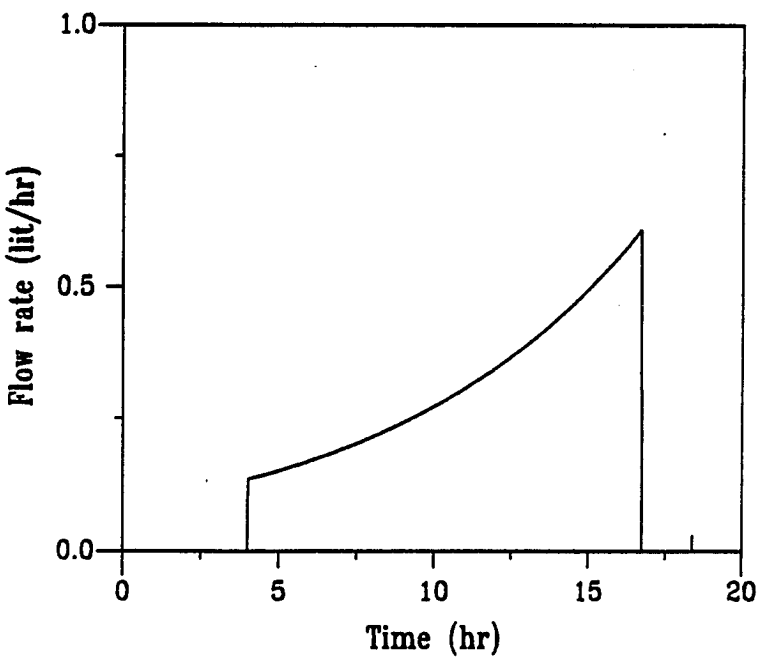

Figure 7.3 Optimal feed rate profile for high initial glucose concentration.

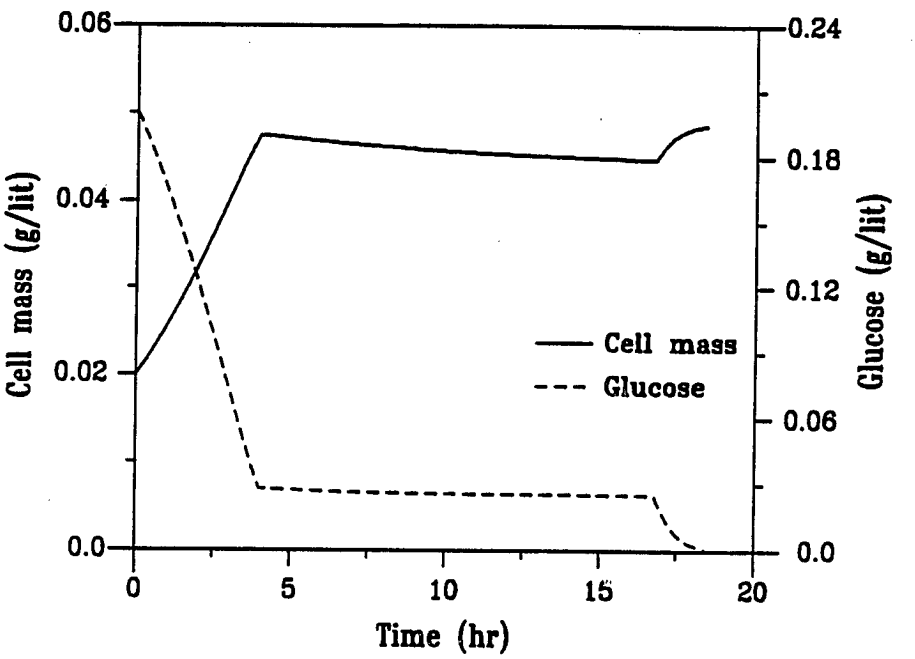

Figure 7.4 Optimal cell mass and glucose concentration profiles for high initial glucose concentration.

8. OPTIMIZATION OF A RECOMBINANT *E. COLI* FERMENTATION

8.1 Introduction

In this work, the problem of maximizing a weighted performance measure was solved for recombinant *Escherichia coli* strains by using the concentration of a growth inhibitor as the control variable. A simple unstructured mathematical model was used to explain the key features of this system. Using optimal control theory for this model, the maximization problem is reduced to numerical determination of optimal switching time between two control arcs.

8.2 Mathematical model

It is necessary, in general, to take into account the effects of plasmid instability when modeling recombinant cell growth. However, in some recombinant systems plasmids are stably maintained even in non-selective media[21]. The growth of stable recombinant cells in the presence of a growth inhibitor can be modeled by a single differential equation,

$$\frac{dX}{dt} = \mu(S,I)\, X \tag{8.1}$$

The yield of cell mass $Y_{X/S}$ is assumed to be constant. Hence,

$$\frac{dS}{dt} = -\frac{\mu(S,I)\, X}{Y_{X/S}} \tag{8.2}$$

The growth rate is assumed to follow the following functionality.

$$\mu(S,I) = \frac{\mu_{max}\, S}{K_s + S + I/K_I}, \tag{8.3}$$

where I is the concentration of the inhibitor. Such functional dependence has previously been proposed by Seo and Bailey[21] for recombinant *E. coli* strains containing closely

related copy number mutant plasmids. They used α-methylglucoside, a competitive inhibitor of glucose transport, to change the specific growth rate of the strains so that the effect of specific growth rate on plasmid copy number and cloned gene expression could be studied.

For describing the product formation kinetics an empirical form suggested by Leudeking and Piret[104] has been quite popular. In this form, the specific product formation rate is considered as the sum of growth-associated and growth-independent terms. A variation of this has been used here to account for the inverse relationship between specific growth rate and product formation rate.

$$\frac{dP}{dt} = (-\alpha\mu + \beta)X, \tag{8.4}$$

where α and β are positive constants. The negative coefficient of μ in the above expression assures this inverse dependence. For the recombinant *E. coli* strains described above, Seo and Bailey[21] observed that with an increase in the specific growth rate, (1) the plasmid content decreases almost linearly, and (2) the specific activity of the cloned-gene product (β-lactamase) decreases. The product formation data obtained by them can be fit quite well using this dependence.

At this point, it is important to realize some limitations of this model. First, the product formation rate does not go to zero, when the substrate concentration is zero as one would expect intuitively. But, because of unavailability of data at low growth rates it is difficult to guess the functional dependence at low substrate concentrations. The calculations described later were done within the limits imposed by the model. Second, the substrate consumption rate is assumed to be independent of the product formation rate.

8.3 Optimization problem

For optimization purposes, the problem can be restated in state formulation as

$$\dot{\mathbf{x}} = \mathbf{f}, \tag{8.5}$$

where

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}. \tag{8.6}$$

and

$$\mathbf{f} = \begin{bmatrix} \mu x_1 \\ -\dfrac{\mu x_1}{Y_{X/S}} \\ (-\alpha\mu + \beta)\, x_1 \end{bmatrix} \tag{8.7}$$

The initial conditions for state variables are specified.

$$x(0) = x_0 \tag{8.8}$$

The aim is to maximize the performance index Π defined by

$$\max_{0 \le I(t) \le I_{max}} \Pi = x_{3f} - \frac{k_1 x_{1f}}{x_{3f}} \tag{8.9}$$

The first and the second terms in Eq. (8.9) signify the value of the product and separation costs respectively. The separation cost is assumed to be inversely proportional to the intracellular product concentration (x_{3f}/x_{1f}). The concentration of the growth inhibitor is used as the control variable to manipulate specific growth rate. It is also assumed that the inhibitor is not metabolized[105] and that the addition of inhibitor does not change the reactor volume significantly. Assuming the final time t_f to be fixed, the Hamiltonian can be defined as

$$H(x, \lambda, I) = \lambda^T \mathbf{f} = (z\mu + \lambda_3\beta)\, x_1, \tag{8.10}$$

where λ^T given in Eq. (8.11)

$$\lambda^T = \begin{bmatrix} \lambda_1, & \lambda_2, & \lambda_3 \end{bmatrix} \tag{8.11}$$

represents the adjoint variables, and

$$z = \lambda_1 - \frac{\lambda_2}{Y_{X/S}} - \lambda_3\alpha \tag{8.12}$$

The adjoint differential equations can now be written.

$$\dot{\lambda} = -\frac{\partial H}{\partial x} = \begin{bmatrix} -z\mu - \lambda_3\beta \\ -z\mu_2 x_1 \\ 0 \end{bmatrix} \tag{8.13}$$

where

$$\mu_2 = \frac{\partial \mu}{\partial x_2} \tag{8.14}$$

The transversality conditions are given by the following set of final conditions on the adjoint variables.

$$\lambda(t_f) = \frac{\partial \Pi}{\partial x(t_f)} = \begin{bmatrix} -\frac{k_1}{x_{3f}} \\ 0 \\ 1 + \frac{k_1 x_{1f}}{x_{3f}^2} \end{bmatrix} \tag{8.15}$$

To maximize the performance index Pontryagin's maximum principle can now be used. Thus, for given x and λ values at an instant of time, the control I should be picked so as to maximize the Hamiltonian within the constraints imposed on the control variable. The gradient of the Hamiltonian is,

$$\frac{\partial H}{\partial I} = \mu_I z x_1, \tag{8.16}$$

where

$$\mu_I = \frac{\partial \mu}{\partial I} = -\frac{\mu_{max} S K_I}{(K_s K_I + S K_I + I)^2} < 0 \tag{8.17}$$

Clearly, the gradient is non-zero for non-zero substrate concentrations. Hence, the optimal control profile must lie on the boundary of the allowed region for the control variable, unless z becomes zero for a finite interval of time. The optimal control sequence I^* is determined by the sign of the gradient of the Hamiltonian.

$$\text{If } z > 0, \text{ then } \frac{\partial H}{\partial I} < 0, \text{ and } I^* = 0$$

$$\text{If } z < 0, \text{ then } \frac{\partial H}{\partial I} > 0, \text{ and } I^* = I_{max}$$

To get more information about the control profile, the time derivative $\dot{z}$ is considered.

$$\dot{z} = \frac{dz}{dt} = (-\mu + \frac{\mu_2 x_1}{Y_{X/S}})\, z - \lambda_3 \beta \tag{8.18}$$

This is a linear differential equation in z, with final conditions given by,

$$z(t_f) = -\frac{k_1}{x_{3f}} - \alpha - \frac{\alpha k_1 x_{1f}}{x_{3f}^2} < 0 \qquad (8.19)$$

To check for the possibility of sign change of z, setting z to zero in Eq. (8.18) we get

$$\dot{z} = -\lambda_3 \beta \qquad (8.20)$$

Since λ_3 is a positive constant (by Eqs. (8.13) and (8.15)), $\dot{z}$ is negative at the point of switching. It follows that z cannot be zero over a finite interval of time. Also, there can be at the most one switch in the sign of z. Thus there are only two possibilities for the optimal inhibitor concentration profile: (1) a switch from initial zero value to a constant maximum value, and (2) a constant maximum value (Fig. 8.1). It is difficult to get an explicit analytical expression for the optimum switching time, which needs to be determined numerically.

8.4 Numerical results

It was shown from theoretical arguments that the only control variable profiles that are candidates for being optimum are those which have a single switch from zero to the maximum allowed value. Calculations were performed, using the IMSL integration routine IVPAG, to determine the optimal switching time. Two cases corresponding to different α and β values were considered. The cases thus differ in the sensitivity of specific product formation rate to specific growth rate. The parameters used for simulations are summarized in Table (8.1).

8.4.1 Case (A) $\alpha = 460$, $\beta = 250$ Since the value of β (representing the growth-independent part of specific product formation rate) is smaller than that of α, product formation rate is somewhat sensitive to changes in specific growth rate. Hence, the optimal operating strategy should result in a significant improvement over a simple batch fermentation.

In Fig. (8.2), the profiles of cell mass and product for the operating strategy which gives maximum final product are compared with those for a simple batch fermentation without the addition of the growth inhibitor. The final product obtained using the optimal inhibitor feeding policy is about 60% greater. Moreover, the amount of cell mass at the end of fermentation is 49% less. The effect of changing the maximum inhibitor concentration was also investigated. A change in the maximum inhibitor concentration from 1 g/L to 12 g/L results in less than 1% change in the optimum switching time. The reason for this insensitivity is that even at a concentration of 1 g/L

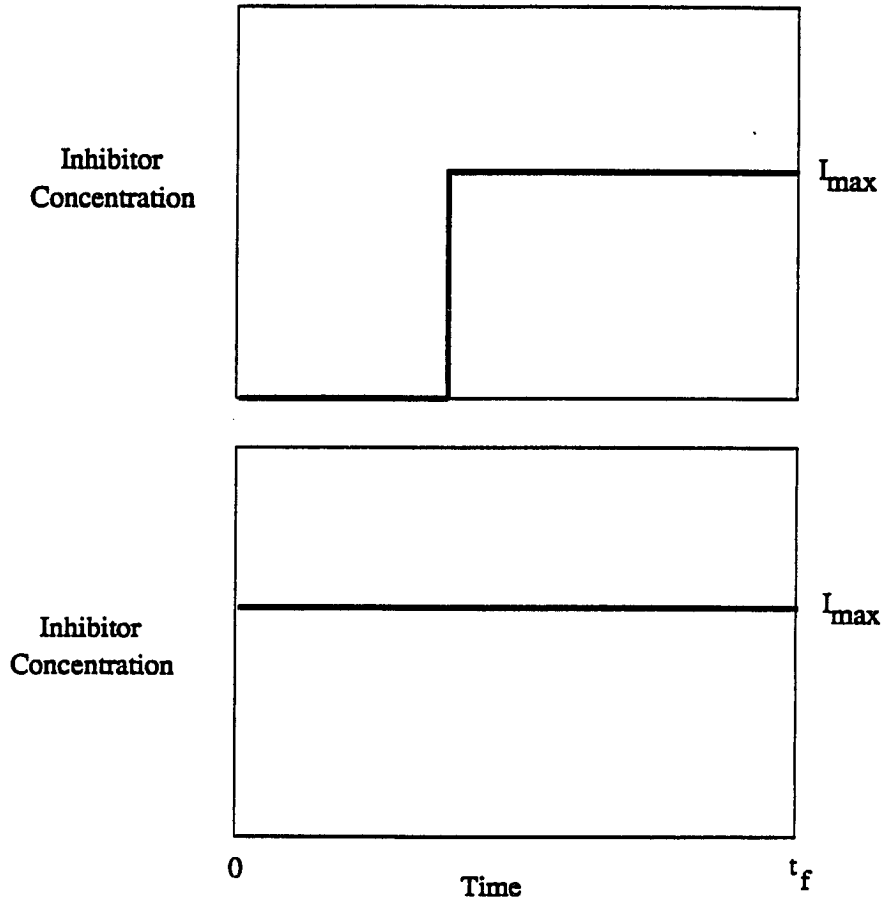

Figure 8.1 Two possible optimal inhibitor concentration profiles.

Table 8.1 Parameter values and constants used for computer simulation

Kinetic Parameter		Simulation Parameters	
μ_{max}, /hr	0.38	x_0, g/L	0.01
K_s, g/L	0.0119	s_0, g/L	2.0
K_I, (g. substrate)/(mg. inhibitor)	0.00911	p_0, units/L	0.0
$Y_{X/S}$, (g. cell)/(g. substrate)	0.32	I_{max}, mg/L	1.0-12.0
		t_f, hr	11.0

Table 8.2 Effect of operating cost and maximum inhibitor concentration on optimal switching time for growth-sensitive strain

k_1	Optimal switching time (hr) I_{max} = 1g/L	Optimal switching time (hr) I_{max} = 12g/L
0	9.16	9.16
1000	9.14	9.14
5000	9.07	9.07

the inhibitor decreases the specific growth rate quite drastically, thus increasing the specific product formation rate. Further increase in the inhibitor concentration does not cause a significant increase in the product formation rate.

The concentrations of cell mass and product at the final time as a function of the switching time are shown in Fig. (8.3). The product concentration profile shows a maximum at a switching time of 9.16 hours. The cell mass concentration, however, increases monotonically with an increase in switching time. If the aim is to maximize the cell mass, it is intuitively evident, and can be proved theoretically using arguments similar to those described in the previous section, that the best performance is achieved when the inhibitor is not added at all. If the separation cost is higher, it is expected that the optimal switching time should be lower yielding a shorter cell growth phase. Calculations were performed for different k_1 values, which reflect differing importance of separation cost. The optimum switching time does not change significantly for small k_1 values, but starts decreasing as k_1 is increased further (Table 8.2).

8.4.2 Case (B) $\alpha = 460$, $\beta = 1500$ This corresponds to a relatively low sensitivity of product formation rate to specific growth rate. In fact, this is the case for recombinant *E. coli* strain with the plasmid pDM247. Calculations show that the optimal operating strategy results in very little improvement compared to a simple batch fermentation, as expected. However, the cell concentrations do change significantly (Fig. 8.4). When separation costs are sufficiently higher compared to the product value, use of the optimal control profile is still necessary.

8.5 Discussion

Pontryagin's maximum principle was used to determine optimal profiles of growth conditions for batch cultures of recombinant *E. coli*. Theoretical analysis indicates that the optimal operating strategy consists of an initial high growth rate stage followed by a low growth but high product formation rate stage. Simple numerical calculations were performed to determine the exact optimal switching time and to confirm the theoretical argument. When product formation rate is sensitive to specific growth rate, the optimal operation policy yields more than five-fold increase in the final product concentration compared with a simple batch fermentation. For the case of a relatively low sensitivity of product formation rate to specific growth rate, the optimal strategy there is little improvement in the process performance.

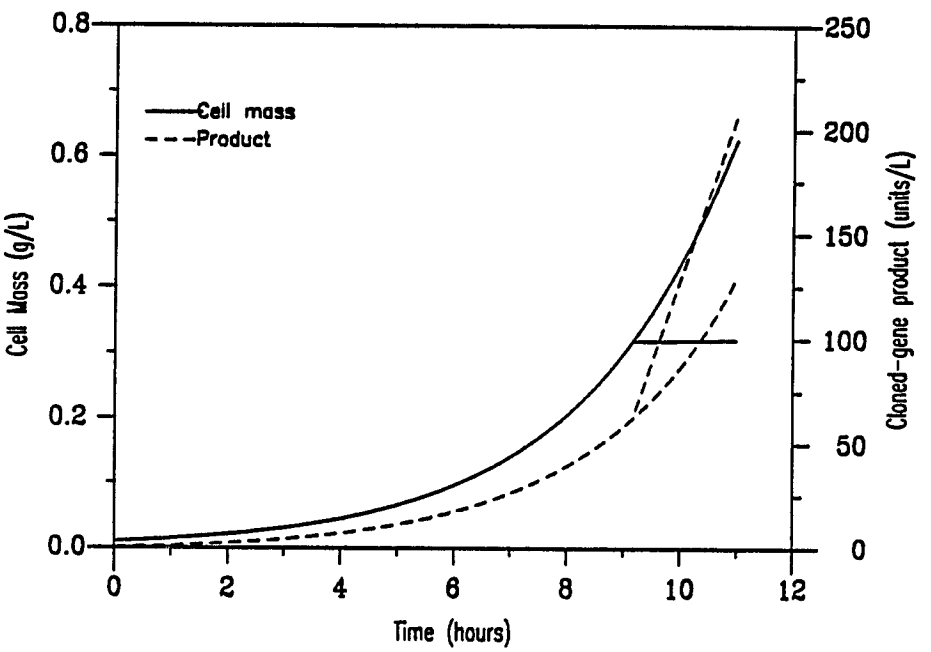

Figure 8.2 Comparison of optimal and batch state profiles (Case A)

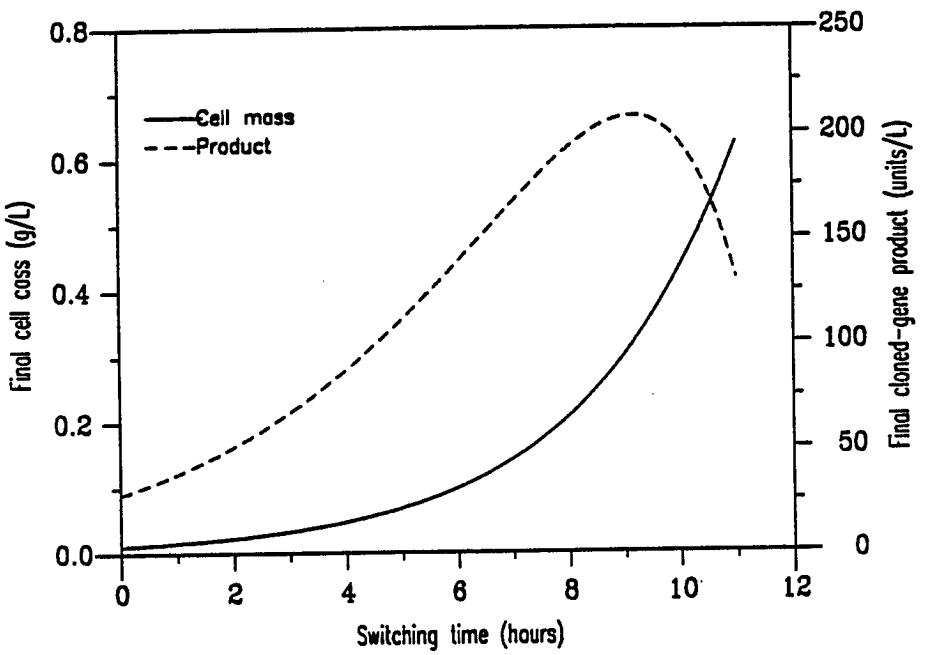

Figure 8.3 Concentrations of cell mass and product at the final time as a function of the switching time (Case A).

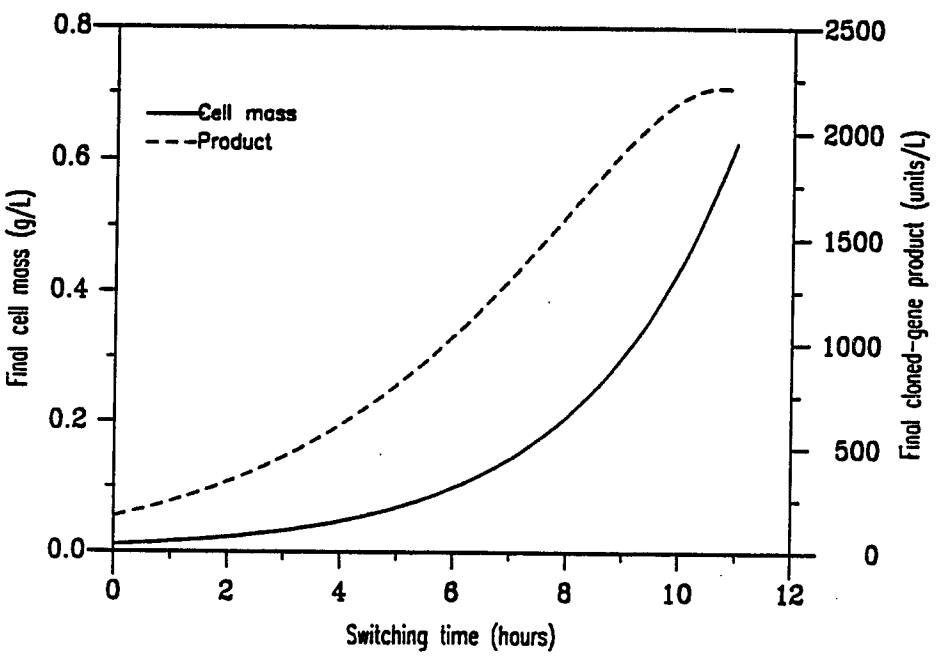

Figure 8.4 Comparison of optimal and batch state profiles (Case B)

This work employed a growth inhibitor to separate the cell growth phase from the cloned gene expression phase. Uncoupling of these two phases can also be achieved by using regulated promoters, which allow control of product levels by manipulation of environmental parameters such as medium composition or temperature. Previous studies[106,107] indicate that even in these systems an initial growth stage followed by product formation stage should be optimal. The optimization method developed in this work would be applied to recombinant systems containing a cloned regulated promoter if model equations describing cell growth and product formation in terms of a control parameter are available.

For the system described herein, the specific product formation rate is a linear decreasing function of specific growth rate, resulting in a bang-bang type of control. For some recombinant systems, however, the specific product formation rate shows a maximum at a certain specific growth rate[108]. For these systems intermediate (off-boundary) control arcs will, in general, be part of the complete optimal control trajectory.

9. CONCLUSIONS AND FUTURE WORK

9.1 Research summary

The aim of this research was to optimize the performance of fed-batch fermentations involving recombinant organisms. To achieve this goal, a recombinant *S. cerevisiae* strain was used as a model system. This yeast strain contains a recombinant plasmid, pRB58, which contains the yeast SUC2 gene. The SUC2 gene codes for the enzyme invertase, and is regulated by glucose levels in the fermentation medium. The research was carried out in three distinct, although not necessarily sequential, steps: kinetics, modeling and optimization.

Batch and fed-batch fermentations were performed under selective growth conditions. The fraction of plasmid-free cells was found to be almost constant (around 0.09) throughout the entire fermentation. During the batch experiments, yeast exhibited diauxic growth, first on glucose and then on ethanol. The results also showed derepression of SUC2 promoter as the glucose concentration dropped. In the fed-batch fermentations, different feeding strategies were used to generate different glucose profiles. These experiments demonstrated the immediate response of invertase production to changes in the glucose levels in the medium. The specific invertase activity was found to be a direct one-to-one function of the glucose concentration. Also, a brief period of zero glucose concentration resulted in a constant low specific invertase activity, even though the glucose levels were raised subsequently. The host strain specific growth rate was found to be almost identical to that of the recombinant strain.

The dimensionality of the mathematical model, to be used for describing experimental data, is restricted because of the limitations of the numerical methods used for optimization. A simple model consisting of four differential equations was developed to describe the experimental results. Analysis of the experimental data revealed that the cell mass yield decreases with an increase in the glucose concentration in the medium. To explain this variable yield, a goal-oriented approach was used. The

key assumption is that objective of the multitude of reactions within the cell is to maximize the cell growth rate. It was realized that this objective is equivalent to maximizing the flux through the respiratory pathway, it being the most energy efficient. The model is able to describe the experimental data quite well. A substrate-inhibition form was used to to describe invertase production kinetics. The invertase mRNA was assumed to decay in a first-order manner. The resulting model does a good job of simulating the experimentally observed behavior.

The objective of optimization is to maximize the total amount of invertase at the end of a fed-batch fermentation by choosing an appropriate feed rate profile. This is a singular control problem because the feed rate appears linearly in the state differential equations. Modak *et al* [1] developed a computational algorithm for solving optimal control problems. However, this method can be successfully applied to systems which can be described by less than five differential equations. Many alternatives, which in principle can be applied to higher dimensional systems, have been suggested in literature. In this research, the applicability of a conjugate gradient technique towards solving this problem was tested. The method was successfully applied for optimization of five-dimensional systems. Although the convergence of the method near the optimum is somewhat slow, it does provide a solution to all the problems tested. The methods based on boundary condition iteration, e.g. Modak's method, need to guess the sequence of control arcs *a priori*. This can be difficult, especially for high-dimensional systems. The conjugate gradient technique does not need any such guess. It was found that the method converged, regardless of the initial guess of the feed rate profile, to the same optimal trajectory. The only restriction on the initial guess was, that it should integrate to the correct final volume.

Using the conjugate gradient technique, the problem of maximizing invertase productivity from recombinant yeast was solved. In all the cases studied, the optimal feed rate profile resulted in an initial high glucose concentration, which favors high cell growth rate, followed by a low glucose concentration, which results in high invertase production rate. The final time was kept fixed for all the optimization runs. It was found that the dependence of productivity of invertase per unit time on final time shows a maximum. At very low values of final time, no singular period was observed in the optimal feed rate profile.

Aside from this main research theme, some other optimization problems were solved. First, a simple model was developed to describe the anaerobic fermentations of recombinant yeast containing the plasmid pRB58. Using this model, optimal feed rate profiles were determined to maximize invertase productivity. In this problem, an analytical expression for optimal singular feed rate can be derived in terms of the state variables. Thus the problem reduces to a simple two-dimensional parameter search problem, the two parameters being switching times between control arcs. The optimal feed rate results in an initial cell growth phase followed by a product formation phase. Second, a model was postulated to describe the inverse dependence of product formation rate on specific growth rate for some recombinant *E. coli* strains. The growth rate can be changed by addition of a growth inhibitor. By using singular control theory, it was shown that the optimal control profile cannot contain a singular arc. The problem was reduced to the numerical determination of optimal time of addition of growth inhibitor.

9.2 Recommendations

In this research no attention was paid towards study of the kinetics of secretion of invertase. About 90% of the yeast invertase is secreted to the periplasmic space. The remaining invertase is found in the intracellular and extracellular regions. Experimental techniques for determination of invertase activity in these compartments are known[88]. It should be quite interesting to study the kinetics of transport of invertase between the cell compartments. It is possible that at high product formation rates, the secretion steps would be rate-determining steps. Thus a detailed study of secretion kinetics may result in a better model of invertase production. If such detailed model is available, it would be very helpful to compare the effectiveness of different secretion signal sequences.

The unstructured model developed in this study satisfactorily described the fed-batch experimental results. However, in case of experiments with highly oscillatory glucose profiles, the model predictions were a little off towards the end of the fermentation. This is probably because the model is unstructured and hence not able to handle rapidly changing conditions. This discrepancy in the glucose predictions translates to a discrepancy in the specific invertase activity predictions near the end of the fermentations. In fact, it was found that if an accurate model for glucose is used[†],

the invertase activity profiles can be described quite accurately. Hence it appears that the functional form for invertase production is quite accurate. Thus, if optimization is of secondary importance and a very accurate model is desired, it is better to concentrate on structured modeling of growth on glucose.

The conjugate gradient approach studied in this research was quite successful in optimization of a variety of biological systems. Although, in principle, this technique can be applied for high-dimensional systems, the rate of convergence decreases when system dimension becomes high. Tests on more complex systems —of dimension higher than five—are necessary to check the generality of this approach. Only the first-order conjugate gradient technique was tested in this numerical study. However, for more complex systems, use of the secon-order method may give better results.

Experimental results showed that short periods of zero glucose levels in the medium resulted in a constant specific invertase activity. Thus it was argued in Chapter 5 that periods of zero glucose concentration cannot be part of an optimal glucose profile. During this numerical study, the feed rate profiles never resulted in such glucose profiles. However, it is probably a good idea to incorporate such a state constraint in the optimization technique. The method described in Chapter 6 cannot be easily adapted to state variable inequality constraints.

The conjugate gradient technique can be used in another interesting manner to compute the optimal feed rates. Because this method does not use an explicit functional form for the singular feed rate, the converged feed rate profiles are not smooth, in general. But in most cases, starting from an arbitrary initial guess of the feed rate, the method is able to get quite close to the optimal feed rate quite fast. Thus within a couple of hundred iterations, the general shape—sequence of maximum, minimum and singular control arcs— of the optimal profile can be deduced without almost no *a priori* information. This sequence can then be used to calculate the optimal profile by Modak's boundary condition iteration method[1]. The conjugate gradient method can also provide good initial guesses for the switching times and adjoint variables to the boundary condition iteration method. Thus these two powerful optimization methods

† This idea was tested by fitting splines through glucose data and using that as a model for glucose.

can be combined to minimize the time to compute the optimal control profile.

LIST OF REFERENCES

LIST OF REFERENCES

1. H. C. Lim, Y. J. Tayeb, J. M. Modak and P. Bonte, *Biotech. Bioeng.*, **28**, 1408-1420 (1986).

2. J. M. Modak and H. C. Lim, *Biotech. Bioeng.*, **33**, 11-15 (1989).

3. C. A. Sardonini and D. DiBiasio, *Biotech. Bioeng.*, **29**, 469-475 (1987).

4. R. W. Old and S. B. Primrose, "**Principles of Gene Manipulation: An Introduction to Genetic Engineering,**" University of California Press, Second Edition (1981).

5. J. D. Watson, N. H. Hopkins, J. W. Roberts, J. A. Steetz, A. M. Weiner, "**Molecular Biology of the Gene,**" The Benjamin/Cummings Publishing Company Inc., Fourth Edition (1987).

6. S. Bartnicki-Garcia and I. McMurrough, in "**The Yeasts (vol. 2): The Physiology and Biochemistry of Yeasts**", (A. H. Rose and J. S. Harrison, eds.) 441, Academic Press, New York (1971).

7. T. Imanaka, H. Tsunekawa and S. Aiba, *J. General Microbiol.*, **118**, 253 (1980).

8. T. Imanaka and S. Aiba, *Ann. NY Acad. Sci.*, **369** (1981).

9. C. D. Dwivede, T. Imanaka and S. Aiba, *Biotech. Bioeng.*, **24**, 1465 (1982).

10. C. Parker and D. DiBasio, *Biotech. Bioeng.*, **26**, 66 (1984).

11. J. Koizumi, Y. Monden and S. Aiba, *Biotech. Bioeng.*, **27**, 721 (1985).

12. D. J. Hopkins, M. J. Betenbaugh and P. Dhurjati, *Biotech. Bioeng.*, **31**, 85 (1987).

13. J. O. Vehmaanpera and M. P. Korhola, *Appl. Microbiol. Biotechnol.*, **23**, 456 (1986).

14. S. H. Kim and D. D. Y. Ryu, *Biotech. Bioeng.*, **26**, 497 (1984).

15. F. Srienc, J. L. Campbell and J. E. Bailey, *Biotech. Bioeng.*, **28**, 996 (1986).

16. R. B. Helling, T. Kinney and J. Adams, *J. Gen. Microbiol.*, **123**, 129 (1981).

17. M. Marquet, S. Alouani and S. W. Brown, *Biotechnology Letters*, **8**, 535 (1986).

18. S. Aiba, J. Koizumi, *Biotech. Bioeng.*, **26**, 1026 (1984)

19. R. Siegel and D. Y. Ryu, *Biotech. Bioeng.*, **27**, 28 (1985).

20. B. E. Uhlin, S. Molin, P. Gustafsson and K. Nordstrom, *Gene*, **6**, 91 (1979).

21. J.-H. Seo and J. E. Bailey, *Biotech. Bioeng.*, **27**, 1668 (1985).

22. D. Ramkrishna, *Adv. Biochem. Eng.*, **11**, 1 (1979).

23. M. A. Hjortso and J. E. Bailey, *Biotech. Bioeng.*, **26**, 528 (1984).

24. M. A. Hjortso and J. E. Bailey, *Biotech. Bioeng.*, **26**, 814 (1984).

25. M. A Hjortso and J. E. Bailey, *Math. Biosci.*, **60**, 235 (1982).

26. M. A. Hjortso and J. E. Bailey, *Math. Biosci.*, **63**, 121 (1983).

27. J.-H. Seo and J. E. Bailey, *Biotech. Bioeng.*, **27**, 156 (1985).

28. S. B. Lee and J. E. Bailey, *Plasmid*, **11**, 166 (1984).

29. S. B. Lee and J. E. Bailey, *Biotech. Bioeng.*, **26**, 66 (1984).

30. S. B. Lee and J. E. Bailey, *Biotech. Bioeng.*, **26**, 1372 (1984).

31. S. B. Lee and J. E. Bailey, *Biotech. Bioeng.*, **26**, 1383 (1984).

32. S. B. Lee, A. Seressiotis and J. E. Bailey, *Biotech. Bioeng.*, **27**, 1699 (1985).

33. A. Murray and J. Szostak, *Cell*, **34**, 961 (1983).

34. S. W. Peretti and J. E. Bailey, *Biotech. Bioeng.*, **31**, 316 (1987).

35. N. A. Da Silva and J. E. Bailey, *Biotech. Bioeng.*, **28**, 741 (1986).

36. L. S. Pontryagin, V. G. Boltiyanskii, R. V. Gamkrelidze and E. F. Mishchenko, **"The Mathematical Theory of Optimal Processes"**, Interscience Publishers (1962).

37. A. E. Bryson and Y. C. Ho, **"Applied Optimal Control"**, Hemisphere, Washington, D.C. (1975).

38. C. D. Siebenthal and R. Aris, *Chem. Eng. Sci.*, **19**, 729 (1964).

39. D. J. Gunn, *Chem. Eng. Sci.*, **22**, 963 (1967).

40. G. D'Ans, P. Kokotovic and D. Gottlieb, *IEEE Trans. Aut. Control*, **AC-16**, 341 (1971).

41. H.J. Kelly, R. E. Kopp and A. G. Moyer in **"Topics in Optimization"** (edited by G. Leitmann), Chap. 3, Academic Press, New York (1966).

42. H. Ohno, E. Nakanishi and T. Takamatsu, *Biotech. Bioeng.*, **20**, 625 (1978).

43. T. Yamane, T. Kume, E. Sada, and T. Takamatsu, *J. Ferment. Tech.*, **55**, 587 (1977).

44. T. Yamane, E. Sada, and T. Takamatsu, *Biotech. Bioeng.*, **21**, 111 (1979).

45. R. Guthke and W. A. Knorre, *Biotechnol. Bioeng.*, **23**, 2771 (1981).

46. W. A. Weigand, H. C. Lim, C. C. Cregan and R. D. Mohler, *Biotech. Bioeng. Symp.*, **9**, 335 (1979).

47. K. Y. San and G. Stephanopoulos, *Biotech. Bioeng.*, **28**, 356 (1986).

48. P. Bonte, M. S. Thesis, Purdue University (1983).

49. Y. J. Tayeb and H. C. Lim, *Ann. N. Y. Acad. Sci.*, **469**, 382 (1986).

50. J. M. Modak, H. C. Lim and Y. J. Tayeb, *Biotech. Bioeng.*, **28**, 1396 (1986).

51. D. J. Bell and D. H. Jacobson, Academic Press, New York (1975).

52. A. Menawat, R. Muthurasan and D. R. Coughanowr, *AIChE J.*, **33**, 776 (1987).

53. D. H. Jacobson, S. B. Gershwin and M. M. Lele, *IEEE Trans. Autom. Control*, **67** (1970).

54. B. Stutts, Ph. D. Thesis, Purdue University (1983).

55. A. Seressiotis and J. E. Bailey, *Biotech. Bioeng.*, **31**, 392 (1987).

56. J. C. Lievense and H. C. Lim, in *Annual Reports on Fermentation Processes* vol. 5, ed. G. T. Tsao, 211, Academic Press, New York (1982).

57. H. Yoon, G. Klinzing, and H. W. Blanch, *Biotech. Bioeng.*, **19**, 1193 (1977).

58. A. H. E. Bijerk and R. J. Hall, *Biotech. Bioeng.*, **19**, 267 (1977).

59. H. Fukuda, T. Shiotani, W. Okada and H. Morikawa, *J. Ferment. Tech.*, **56**, 361 (1978).

60. W. Okada, H. Fukuda and H. Morikawa, *J. Ferment. Tech.*, **59**, 103 (1981).

61. R. J. Hall and J. P. Barford, *Biotech. Bioeng.*, **23**, 1763 (1981).

62. J. P. Barford and R. J. Hall, *Biotech. Bioeng.* **23**, 1735 (1981).

63. K. H. Bellgardt, W. Kuhlamann, and H. D. Meyer, in *First IFAC Conf. on Modeling and Control of Biotechnical Processes*, ed. A. Halme, 67, Helsinki, Finland (1982).

64. K. Toda, I. Yabe, and T. Yamagata, *Biotech. Bioeng.*, **22**, 1805 (1980).

65. J. Lievense, Ph. D. Thesis, Purdue University (1985).

66. J. M. Modak, Ph. D. Thesis, Purdue University (1988).

67. M. Carlson and D. Botstein, *Cell*, **28**, 145 (1982).

68. A. B. Futcher, *Yeast*, **4**, 27 (1988).

69. M. Guerineau, "Virus and plasmids in fungi," P. Lemke Ed., Marcel Dekker, New York (1979).

70. A. B. Futcher, *J. Bacteriol.*, **154**, 612 (1983).

71. N. P. Neumann and J. O. Lampen, *Biochemistry*, **6**, 468 (1967).

72. R. B. Trimble and F. Maley, **252**, 4409 (1977).

73. S. Gascon, N. P. Neumann and J. O. Lampen, *J. Biol. Chem.*, **243**, 1573 (1968).

74. R. Taussig and M. Carlson, *Nucleic Acids Res.*, **11**, 1943 (1983).

75. M. Carlson, R. Taussig, S. Kustu and D. Botstein, *Mol. Cell. Biol.*, **3**, 439 (1983).

76. L. Sarokin and M. Carlson, *Mol. Cell. Boil.*, **4**, 2750 (1984).

77. L. Sarokin and M. Carlson, *Mol. Cell. Biol.*, **5**, 2521 (1985).

78. L. Sarokin and M. Carlson, *Mol. Cell. Biol.*, **6**, 2324 (1986).

79. S. Hohmann and D. Gozalbo, *Mol. Gen. Genet.*, **211**, 446 (1988).

80. J. L. Celenza and M. Carlson, *Mol. Cell. Biol.*, **4**, 49 (1984).

81. J. L. Celenza and M. Carlson, *Mol. Cell. Biol.*, **4**, 54 (1984).

82. L. Neigeborn and M. Carlson, *Genetics*, **112**, 845 (1984).

83. H. Ma and D. Botstein, *Mol. Cell. Biol.*, **6**, 4046 (1986).

84. L. Neigeborn, J. L. Celenza and M. Carlson, *Mol. Cell. Biol.*, **7**, 672 (1987).

85. J. Schultz and M. Carlson, *Mol. Cell. Biol.*, **7**, 3637 (1987).

86. S. Emr, R. Schekman, M. Flessel and J. Thorner, *Proc. Natl. Acad. Sci.*, **80**, 7080 (1983).

87. F. Sherman, G. R. Fink and J. B. Hicks, "Methods in yeast genetics", p. 164, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (1986).

88. M. Marten, M. S. Thesis, Purdue University (1988).

89. A. Goldstein and J. O. Lampen, *Methods in Enzymology*, **42**, 504 (1975).

90. C. Pascual, A. Alonso, I. Garcia and C. Romay, *Biotech. Bioeng.*, **32**, 374 (1988).

91. P. Dhurjati, D. Ramkrishna, M. C. Flickinger and G. T. Tsao, *Biotech. Bioeng.*, **27**, 1 (1985).

92. M. V. Elorza, C. M. Lostau, J. R. Villanueva and R. Sentandreu, *Biochim. Biophys. Acta*, **475**, 638 (1977).

93. S. Park and W. F. Ramirez, *AIChE J.*, **34**, 1550 (1988).

94. D. H. Jacobson, *SIAM J. Control*, **7**, 578 (1969).

95. D. H. Jacobson, *SIAM J. Control*, **8**, 403 (1970).

96. J. L. Speyer and D. H. Jacobson, *J. Math. Anal. Applic.*, **33**, 163 (1971).

97. D. H. Jacobson and J. L. Speyer, *J. Math. Anal. Applic.*, **34**, 239 (1971).

98. Y. R. Pyun, J. M. Modak, Y. K. Chang and H. C. Lim, *Biotech. Bioeng.*, **33**, 1 (1989).

99. L. S. Lasdon, S. K. Mitter and A. D. Waren, *IEEE Trans. Automatic Control*, **AC-12**, 132 (1967).

100. B. Pagurek and C. M. Woodside, *Automatica*, **4**, 337 (1968).

101. R. K. Bajpai and M. Reuβ, *J. Chem. Tech. Biotechnol.*, **30**, 332 (1980).

102. Y. R. Pyun, *Personal communications*

103. S. Mormeneo and R. Sentandreu, *Antonie van Leeuwenhoek*, **52**, 15 (1986).

104. R. Leudeking and E. L. Piret, *J. Biochem. Microbiol. Technol. Eng.*, **1**, 393 (1959).

105. M. T. Hansen, M. L. Pato, S. Molin, N. P. Fill, and K. von Meyenburg, *J. Bacteriol.*, **122**, 585 (1975).

106. R. Siegel and D. Y. Ryu, *Annals N. Y. Acad. Sci.*, **469**, 73 (1986).

107. T. H. Park, J.-H. Seo and H. C. Lim, *Biotech. Bioeng.*, in press (1989).

108. J.-H. Seo and J. E. Bailey, *Biotech. Bioeng.*, **28**, 1590 (1986).

109. A. Sols, C. Gancedo and G. Dela Fuente, in **The Yeasts (Vol. 2): The Physiology and Biochemistry of Yeasts**, (A. H. Rose and J. S. Harrison, eds.) 271, Academic Press, London (1971).

110. C. Beck and H. K. von Meyenburg, *J. Bacteriol.*, **96**, 479 (1968).

111. C. P. M. Gorts, *Antonie van Leeuwenhoek*, **37**, 161 (1971).

112. R. H. DeDekken, *J. Gen. Microbiol.*, **44**, 149 (1966).

113. E. S. Polakis and W. Bartley, *Biochem. J.*, **97**, 284 (1965).

114. E. S. Polakis, W. Bartley and G. A. Meek, *Biochem. J.*, **97**, 298 (1965).

APPENDICES

Appendix A - Glucose assay

Reagents

1. Reagent A: one capsule of PGO enzymes added to 100 mL of distilled water
2. Reagent B: 50 mg o-dianisidine dihydrochloride added to 20 mL of distilled water
3. Reagent C: combine 100 mL reagent A with 1.6 mL reagent B.
4. Glucose standard solution: 100 mg/dL β-glucose in benzoic acid solution.

Colorimetric method

1. Label three or more 13x100 mm test tubes as follows: Blank, standard, test1, test2, etc.
2. To blank, add 0.5 mL water.
3. To standard add 0.5 mL water and 25 μL of glucose standard solution.
4. To each test, add 0.5 mL water and 25 μL of sample.
5. To each tube, add 5.0 L reagent C and mix thoroughly.
6. Incubate all tubes at 37 °C for 30 minutes.
7. At the end of 30 minutes remove all tubes from bath and read absorbance at 450 nm using blank as a reference.

Calculate glucose concentration as

$$\text{Glucose(mg/dL)} = \frac{A_{test}}{A_{standard}} \times 100$$

Appendix B - Program for data acquisition from GC

```
100 '*****************************************************************
110 '*
120 '*      INTEG.BAS - To acquire data from GC
130 '*      written by Anant Patkar (08/08/89)
135 '*
140 '*****************************************************************
150 '
155 screen 0,0,0 : cls : key off : width 80
160 '
170 'First load DAS16.BIN routine by contracting BASIC to 48K workspace
172 '
175 locate 13, 30 : PRINT "Wait! Initializing ..."
180 clear, 49152!                    'reduce workspace to 48K
190 def seg = 0                      'find BASIC's segment
200 sg = 256 * peek(&H511) + peek(&H510)
210 sg = sg + 49152!/16
220 def seg = sg                     'SG = load location
230 bload "das16.bin", 0               'for DAS16.BIN
240 '
250 'Initialize using mode 0
260 '
270 dim dio%(4), xind!(5), fx!(5)          'declare data array
273 open "das16.adr" for input as #1       'reading board address
275 input #1, b% : close #1
280 dio%(0) = b%                       'base I/O address
290 dio%(1) = 2                        'interrupt level
300 dio%(2) = 1                        'D.M.A. level
330 md% = 0                            'initialize mode
340 flag% = 0                          'declare error variable
350 das16 = 0                         'CALL offset = 0
360 call das16 (md%, dio%(0),flag%)        'initialize
370 if flag%<>0 then print"Initializing error, # " ;flag% : stop
380 '
410 locate 13, 30 : print space$(22)
450 locate 15, 5  : input "Enter your initials : "; init$
500 'Find out what you want for scan limits:-
510 '
540 dio%(0) = 0              'lower limit
550 dio%(1) = 0              'upper limit
560 md% = 1
570 call das16 (md%, dio%(0),flag%)
580 if flag%<>0 then print"Error in setting scan limits, # ";flag% : stop
590 '
```

```
600 open "color.dat" for input as #1
610 input #1, c1%, c2%, c3%, c4%, c5%, c6%, c7%
620 arxl% = 60 : arxr% = 150 : ary% = 148
640 close #1 : gosub 2000              'calls main-menu subroutine
645 opt$ = "" : while opt$ = "" : opt$ = inkey$ : wend : opti% = asc(opt$) - 48
650 on opti% goto 670, 800, 9000, 900
670 open "test.out" for output as #1 : gosub 2070
680 while inkey$ <> "b" : wend
710 nconv% = 0 : gosub 3230
720 close #1 : goto 640
800 '
810 cls : input "Enter file suffix : "; suf$
820 nam$ = init$ + suf$ + ".prn"
830 open nam$ for output as #1 : gosub 2070
840 while inkey$ <> "b" : wend
850 nconv% = 0 : gosub 3230
860 close #1 : locate 24, 10 : print "Area? (y/n) :"; : cyn$ = ""
865 while cyn$ = "" : cyn$ = inkey$ : wend
870 if cyn$ = "n" then goto 640 else gosub 3510
880 goto 640
900 cls : input "Enter suffix : "; suf$
905 nam$ = init$ + suf$ + ".prn"
910 open nam$ for input as #1
920 gosub 2070
930 xbeg! = 44.0 : ybeg! = 148.0
940 for ij% = 1 to 360
950    input #1, t!, ic%
960    if ic% = -1 goto 1010
970    xplot% = int (xbeg! + 0.4 * t!)
980     yplot% = int (ybeg! - 0.029304 * ic%)
990    line - (xplot%,yplot%), c7%
1000 next ij%
1010 close #1 : gosub 3510
1020 goto 640
2000 screen 0,1,0 : width 80 : cls : color c1%, c2%
2010 locate 5,36 : print "MAIN MENU"
2020 locate 11,25 : print "1. Checking without injection"
2030 locate 14,25 : print "2. Sample injection"
2040 locate 17,25 : print "3. Quit"
2042 locate 20,25 : print "4. Calculation of area"
2045 locate 2,2 : print "Enter option : ";
2060 return
2070 screen 1,0 : cls : key off : color c3%,c4%
2080 locate 2, 15 : print "GC Recorder"
3010 locate  4,4 : print "10" : locate  7,5 : print "8"
3020 locate 10,5 : print "6" : locate 13,5 : print "4"
```

```
3030 locate 16,5 : print "2" : locate 19,5 : print "0"
3040 xorg% = 44  : xtic% = xorg% :yorg% = 148 :ytic% = yorg% : tic% = 3
3050 xlef% = xorg% - tic%  : ybot% = yorg% + tic%
3060 line (44,28) - (284,148) ,c5% , bf
3065 line (44,28) - (284,148) ,c6% , b
3070 for i% = 1 to 11
3080    line (xtic%, ytic%) - (xlef%, ytic%), c6% : ytic% = ytic% - 12
3090 next i%
3100 ytic% = yorg%
3110 for i% = 1 to 11
3120   line (xtic%, ytic%) - (xtic%, ybot%), c6% : xtic% = xtic% + 24
3130 next i%
3140 locate 23,16 : print "Time, min" : locate 21, 6 : print "0"
3150 locate 21, 12 : print "2" : locate 21, 18 : print "4"
3160 locate 21, 24 : print "6" : locate 21, 30 : print "8"
3170 locate 21, 36 : print "10" : locate  7,  2 : print "G"
3180 locate  8,  2 : print "C" : locate 10,  2 : print "O"
3190 locate 11,  2 : print "u" : locate 12,  2 : print "t"
3200 locate 13,  2 : print "p" : locate 14,  2 : print "u"
3210 locate 15,  2 : print "t"
3220 locate 24, 10 : print "b - begin"; : pset (44,148) , c5% : return
3230 '
3240 'start data collection using mode 3
3250 '
3260 locate 24, 10 : print "s - stop ";
3270 check! = timer
3280 xbeg! = 44.0 : ybeg! = 148.0
3300 md% = 3
3310 call das16 (md%, dio%(0), flag%)
3320 count! = timer
3330 t! = count! - check!
3420 xplot% = int (xbeg! + 0.4 * t!)
3430 yplot% = int (ybeg! - 0.029304 * dio%(0))
3440 line - (xplot%, yplot%) , c7%
3450 if t! <= 360.0 then print #1, t!, dio%(0)
3460 if t! > 600 then xbeg! = -196.0 : ybeg! = 124.0
3470 if inkey$ = "s" or t! >= 1200.0 then goto 3500
3480 while timer < count! + 0.5 : wend
3490 goto 3300
3500 print #1, "-1.0 -1" : print #1, date$, time$ : return
3510 '
3520 ' To calculate the area under the curve
3540 locate 24, 2 : print "Use <hHjklL> to move. d - done";
3550 cmove$ = ""
3560 line (arxl%,28) - (arxl%, 148), c7%
3570 line (arxr%,28) - (arxr%, 148), c7%
```

```
3580 line (44, ary%) - (284, ary%), c7%
3590 while cmove$ = "" : cmove$ = inkey$ : wend
3600 if cmove$ = "h" then line (arxl%,28) - (arxl%,148), c5% : arxl% = arxl% - 1 : goto 3550
3610 if cmove$ = "H" then line (arxr%,28) - (arxr%,148), c5% : arxr% = arxr% - 1 : goto 3550
3620 if cmove$ = "l" then line (arxl%,28) - (arxl%,148), c5% : arxl% = arxl% + 1 : goto 3550
3630 if cmove$ = "L" then line (arxr%,28) - (arxr%,148), c5% : arxr% = arxr% + 1 : goto 3550
3640 if cmove$ = "j" then line (44,ary%) - (284,ary%), c5% : ary% = ary% + 1 : goto 3550
3650 if cmove$ = "k" then line (44,ary%) - (284,ary%), c5% : ary% = ary% - 1 : goto 3550
3660 if cmove$ = "d" then goto 3670
3670 tc1! = 2.5 * (arxl% - 44) : tc2! = 2.5 * (arxr% - 44)
3680 yc! = 34.125 * (148 - ary%)
3690 open nam$ for input as #1
3700 i% = 0 : icount% = 0
3710 input #1, xind!(1), fx!(1)
3720 if xind!(1) < tc1! then icount% = icount% + 1 : goto 3710 else i% = i% + 1
3725 xsavl! = xind!(1) : ysavl! = fx!(1)
3730 input #1, xind!(1), fx!(1) : xsavr! = xind!(1) : ysavr! = fx!(1)
3740 if xind!(1) > tc2! then goto 3750 else i% = i% + 1 : goto 3730
3750 close #1 : open nam$ for input as #1
3760 inum% = i%  5 : irem% = i% mod 5 : area! = 0.0
3770 for j% = 1 to icount%
3780    input #1, xind!(1), fx!(1)
3790 next j%
3800 if inum% > 0 then goto 3810 else goto 3880
3810 for j% = 1 to inum%
3820    for k% = 1 to 5
3830       input #1, xind!(k%), fx!(k%)
3840    next k%
3850    dum! = fx!(1)+4.0*fx!(2)+2.0*fx!(3)+4.0*fx!(4)+fx!(5)
3860    area! = area! + (xind!(2)-xind!(1))*dum!/3.0
3870 next j%
3880 if irem% > 0 then goto 3890 else goto 3930
3890 for j% = 1 to irem%
3900    input #1, xind!(1), fx!(1)
3910    area! = area! + 0.5 * fx!(1)
3920 next j%
3930 area! = area! - 0.5 * (yc! + ysavr!) * (xsavr! - xsavl!)
3940 locate 24, 2 : print "Area = :", area!, space$(10);: ckey$ = ""
3950 while ckey$ <> "s": ckey$ = inkey$ : wend
4000 return
9000 cls : end
```

Appendix C - Yeast cell lysis using glass beads

1. Centrifuge 20 mL of culture (in the log-phase or stationary phase) at 12,000 rpm for 20 minutes.
2. Remove the supernatant carefully.
3. Add 20 mL of distilled water to the centrifuge tube containing the pellet and vortex.
4. Centrifuge the tube again at 12,000 rpm for 20 min.
5. Remove the supernatant, resuspend in 4 mL potassium phosphate buffer (pH 7.0) and freeze at –20°C.
6. At the beginning of the assay, take the tube out of the freezer and allow it to thaw. At the same time, start carbon-dioxide flow through the cell homogenizer.
7. Transfer the sample tube contents to homogenizer tubes. Add 6 grams of acid-cleaned, pre-cooled glass beads (0.45 mm diameter) to each tube.
8. Wait till ice forms on the carbondioxide line. This indicates that the homogenizer assembly is cold enough.
9. Put the homogenizer tube inside the cell homogenizer, clamp the system properly and start homogenization. The cell homogenization is done in cycles of 30 seconds alternating with 30 seconds of cooling. Actual homogenization time is 2 minutes.
10. Take out the homogenizer tube, vortex the contents and remove the homogenized sample using Pasteur pipettes.

Appendix D - Invertase assay

Reagents

1. Sodium acetate buffer, 0.2M, pH 4.9
2. Sucrose, 0.5M
3. Potassium phosphate buffer, 0.5M, pH 7.0

Procedure

1. For each enzyme sample assayed, combine 100 µL of the sample with 300 µL of acetate buffer in a 13 x 100 mm test tube.
2. Add 400 µL of acetate buffer to a 13 x 100 mm test tube, as a control.
3. Prewarm all tubes in a 30°C bath.
4. At time = 0 minutes, add 300 µL sucrose solution to control tube, vortex and replace in bath.
5. At time = 1 minute, add 300 µL sucrose solution to sample tube number 1, vortex and replace in batch. Continue this procedure in one minute intervals for all sample tubes.
6. Ten minutes after sucrose is added to each tube, remove the tube, add 2.0 mL of phosphate buffer and vortex.
7. Immediately after adding phosphate buffer, place tube in boiling water for 3 minutes.
8. Remove tube and allow to cool to room temperature.
9. Determine the concentration of glucose in the sample as described in Appendix A. Some dilutions may be necessary depending on the concentration of invertase in the sample.

Appendix E - Program for simulation and parameter estimation

```
C-----------------------------------------------------------------
C
C  Program to test different models for yeast growth
C  Written by Anant Patkar (03/21/89)
C
C-----------------------------------------------------------------

      parameter(nvar = 8, maxd = 1001)
      implicit double precision (a-h, o-z)
      integer ndata(nvar), yes(nvar)
      dimension xdata(nvar,maxd), ydata(nvar,maxd), ymin(nvar),
     !    ymax(nvar), y(5), param(50), ysave(5), ytemp(5),
     !    tempy(maxd)
      double precision k1, k2, k3, k4, k5, k6
      character*1 test
      external fcn, fcnj
      character*10 xlabel, ylabel(nvar)
      common /par/ k1, k2, k3, ck1, ck2, ck3, yrx, yfx, yex, yxe, sf
      common /par2/ k4, ck4, k5, k6, ck6, rmu, ichf

      open (1, file = 'labels')
      rewind (1)
      open (2, file = 'modelparam')
      rewind (2)
      open (3, file = 'temp.out')
      rewind (3)
      open (4, file = 'temp1.out')
      rewind (4)

      read (1,11) xlabel
      naxis = nvar / 2
      do 10 i = 1, naxis
        read (1,11) ylabel(i)
10    continue
11    format(a10)
      close (1)

      read (5,*)
      read (5,*) xmin, xmax
      read (5,*)

      do 20 i = 1, nvar
```

```
      read (5,*) ymin(i), ymax(i), ndata(i)
20    continue
      read (5,*)

      do 70 i = 1, nvar
        yes(i) = 0
70    continue
      narg = iargc()
      ivolume = 0
      do 80 i = 1, narg
        call getarg (i,test)
        if (test.eq.'x') then
          yes(1) = 1
          yes(2) = 1
        else if (test.eq.'g') then
          yes(3) = 1
          yes(4) = 1
        else if (test.eq.'e') then
          yes(5) = 1
          yes(6) = 1
        else if (test.eq.'i') then
          yes(7) = 1
          yes(8) = 1
        else if (test.eq.'v') then
          ivolume = 1
        endif
80    continue
      read (5,*) tol, tf
      read (5,*) x0, g0, e0, ri0, v0, rm0
      read (2,*) k1, ck1, k2, ck2, k3, ck3, yrx, yfx, yex, yxe
      read (2,*)
      read (2,*) k4, ck4, k5, k6, ck6
C
C  Reading  in experimental data.
C
      read (5,*)
      read (5,*) sf, istart, ichf
      do 40 i = 1, nvar, 2
        read (5,*)
        do 50 j = 1, ndata(i)
          read (5,*) xdata(i,j), ydata(i,j)
          if (i.eq.1) ydata(i,j) = dlog10(ydata(i,j))
50      continue
40    continue
C
C. Generation of simulation data
```

```
C
      step = tf / dfloat(maxd-1)
      param(1) = 1.d-4 * step
C      param(2) = 1.d-10
      param(3) = step / 2.d0
      param(4) = 10000
      param(12) = 2
      param(13) = 0
      param(19) = 0
      time = 0.d0
      y(1) = x0 * v0
      y(2) = g0 * v0
      y(3) = e0 * v0
      y(4) = 0.d0
      y(5) = v0
      xdata(2,1) = time
      xdata(4,1) = time
      xdata(6,1) = time
      ydata(2,1) = dlog10(x0)
      tempy(1)   = x0
      ydata(4,1) = g0
      ydata(6,1) = e0
      ido = 1
      neq = 5
      nstep = maxd - 1

      do 60 i = 1, istart
        time = step * dfloat(i-1)
        xend = step * dfloat(i)
        xdata(2,i+1) = xend
        xdata(4,i+1) = xend
        xdata(6,i+1) = xend
        call divpag (ido, neq, fcn, fcnj, a, time, xend, tol,
     !              param, y)
        ydata(2,i+1) = dlog10 (y(1) / y(5))
        tempy(i+1) = (y(1) / y(5))
        ydata(4,i+1) = y(2) / y(5)
        ydata(6,i+1) = y(3) / y(5)
C        if (ivolume.eq.1) write (3,*) xend, y(5)
60    continue

      do 90 is = 1, 5
        ysave(is) = y(is)
90    continue

      ido = 3
```

```
      call divpag (ido, neq, fcn, fcnj, a, time, xend, tol,
     !            param, y)

C
C     This small construct is just to get mu values for
C     initial value of mRNA.
C
      rmu = 0.d0
      call fcn(neq, time, ysave, ytemp)
C
      param(1) = 1.d-4 * step
C      param(2) = 1.d-10
      param(3) = step / 2.d0
      param(4) = 10000
      param(12) = 2
      param(13) = 0
      param(19) = 0
      time = step * dfloat(istart)
      y(1) = ysave(1)
      y(2) = ysave(2)
      y(3) = ysave(3)
      y(4) = ri0 / k6
      y(5) = ysave(5)
      xdata(8,1) = time
      ydata(8,1) = ri0
      ido = 1
      neq = 6

      do 100 i = istart+1, nstep
        time = step * dfloat(i-1)
        xend = step * dfloat(i)
        xdata(2,i+1) = xend
        xdata(4,i+1) = xend
        xdata(6,i+1) = xend
        ir0 = i - istart + 1
        xdata(8,ir0) = xend
        call divpag (ido, neq, fcn, fcnj, a, time, xend, tol,
     !               param, y)
        ydata(2,i+1) = dlog10 (y(1) / y(5))
        tempy(i+1) = (y(1) / y(5))
        ydata(4,i+1) = y(2) / y(5)
        ydata(6,i+1) = y(3)/y(5)
C        call fcn(neq, time, y, ytemp)
        ydata(8,ir0) = k6 * y(4)
C        if (ivolume.eq.1) write (3,*) xend, y(5)
 100    continue
```

```
C
      rewind (3)
      do 110 ij = 1, 1001
        write (3,21) xdata(2,ij),ydata(2,ij), ydata(4,ij),
     !             xdata(6,ij)
  110 continue
   21 format (4(2x, f12.6))
      call flush(3)
C
      rewind (4)
      inum = 1001 - istart
      do 120 ij = istart, nstep
        ir0 = ij - istart + 1
        write (4,31) xdata(8,ir0), ydata(8,ir0)
  120 continue
   31 format (2(2x, f12.6))
      call flush(4)

C
C Call to the plotting routine.
C
      call myplot(nvar,ndata,maxd,xdata,ydata,xlabel,ylabel,
     !           naxis, xmin, xmax, ymin, ymax, yes)

      stop
      end

C-------------------------------------------------------
C
C  Subroutine required by the IMSL integration routine DIVPAG
C  to determine the differential rates.
C
      subroutine fcn (neq, time, y, yprime)
      implicit double precision (a-h, o-z)
      dimension y(neq), yprime(neq)
      double precision  k1, k2, k3, k4, k5, k6
      common /par/ k1, k2, k3, ck1, ck2, ck3, yrx, yfx, yex, yxe, sf
      common /par2/ k4, ck4, k5, k6, ck6, rmu, ichf

      x = y(1)/y(5)
      s = y(2)/y(5)
      e = y(3)/y(5)
      if (s.le.1.d-5) s = 0.d0
      rt = k1 * s / (ck1 + s)
      rr = k2 * s / (ck2 + s)
      if (rt.gt.rr) then
```

```
      rf = rt - rr
      unsat = 0.d0
    else
      rr = rt
      rf = 0.d0
      unsat = 1.d0 - s / (ck2 + s)
    endif
    if (s.le.1.d-4) unsat = 1.d0
    re = k3 * unsat * e / (ck3 + e)
    fbase = 0.d0
    if (ichf.eq.0) then
      call flow0 (time, fin, fout)
    else if (ichf.eq.1) then
      call flow1 (time, fin, fout)
    else if (ichf.eq.2) then
      call flow2 (time, fin, fout)
    else if (ichf.eq.3) then
      call flow3 (time, fin, fout)
    else if (ichf.eq.4) then
      call flow4 (time, fin, fout)
      fbase = 17.d-3
    else if (ichf.eq.5) then
      call flow5 (time, fin, fout)
    else if (ichf.eq.6) then
      call flow6 (time, fin, fout)
    endif
    if (s.le.0.d0) then
      rmu = re * yxe
      yprime(1) = y(1) * rmu - fout * x
      yprime(2) = fin * sf
    else
C       rmu = rr * yrx + rf * yfx + re * yxe
      rmu = rr * yrx + rf * yfx
      yprime(1) = y(1) * rmu - fout * x
      yprime(2) = - y(1) * rt + fin * sf - fout * s
    endif
    yprime(3) = y(1) * (yex * rf - re) - fout * e
    yprime(5) = fin - fout + fbase
C
C   Invertase production model
C
    yprime(4) = k4 * s /(ck4 + s + ck6*s**2) - (k5 + rmu)*y(4)
    check4 = dabs(yprime(4))
    if (check4.lt.1.d-5) yprime(4) = 0.d0

    return
```

```
      end

C-----------------------------------------------
C
C  Dummy subroutine required by DIVPAG
C
      subroutine fcnj(neq, x, y, dypdy)
      implicit double precision (a-h, o-z)
      dimension y(neq), dypdy(neq,neq)
      return
      end

C-----------------------------------------------------------------
C  Subroutine to plot simulation results.
C  Written by Anant Patkar (01/29/89)
C
C  Variables:
C  ---------
C  nvar = number of independent variables * 2
C       Variables with odd indices are assumed to be experimental
C       data, and ones with even indices are simulation results.
C  ndata = number of data points for each independent variable.
C  maxd  = maximum number of data points in all data sets
C  xdata = data for independent variables (all sets)
C  ydata = data for dependent variables (all sets)
C  xlabel, ylabel  = axis labels
C  naxis = number of axis = nvar / 2
C  xmin  = value corresponding to left end point of abscissa
C  xmax  = "" right ""
C  ymax,  ymin = similar
C  yes   = Array nvar x 1, the value 1 corresponds to plot
C        0 means do not plot
C
      subroutine myplot(nvar, ndata, maxd, xdata, ydata, xlabel,
     !        ylabel, naxis, xmin, xmax, ymin, ymax,
     !        yes)
      implicit double precision (a-h,o-z)
C
C ndata is a vector of number of data points in each dataset.
C maxd = max(ndata(i)), i = 1, ..., nvar.
C
      integer ndata(nvar), digits, where, flag, odd, yes(nvar)
      dimension xdata(nvar,maxd), ydata(nvar,maxd), ymin(nvar),
     !     ymax(nvar)
      character*10 xlabel, ylabel(nvar), y1, y2, y3, y4, y5, sym*1
```

```
      if (naxis.gt.5) then
         write (6,11)
 11       format(2x, 'Plotting program is limited to a maximum of ',
     !            'five dependent variables.')
         stop
      endif
C
C  Setting some parameters.
C
      xorigin = 3.0
      yorigin = 3.0
      xsize = 7.0
      ysize = 6.0
      ticdis = 1.0
      digits = 4
      xdis = 1.6
      ht = 0.12

C      call site("-kp")
      call plots(6,0)
C      call plots(5,0, "plot.out")
      call factor (1.0)
      call plot(xorigin, yorigin, -3)
C
C Change parameters of axis
C
      call axisv(ticdis, digits)
C
C Draw main dataspace window.
C
      flag = 0
      where = 0
      call axis(0.,0., xlabel, where, xsize, xmin, xmax, flag)
      where = 1
      y1 = ylabel(1)(1:10)
      logmin = ymin(1)
      logmax = ymax(1)
      call laxis(0.,0., y1, where, ysize, logmin, logmax, flag)
      call plot(0.0, ysize, 3)
      call plot(xsize, ysize, 2)
      call plot(xsize, 0.0, 2)
      call plot(0.0, 0.0, 3)
C
C Draw other axis (if required).
C
      if (naxis.gt.1) then
```

```
          where = 3
          flag = 0
          y2 = ylabel(2)(1:10)
          call axis(xsize, 0., y2, where, ysize, ymin(3),
     !          ymax(3), flag)
        endif
        if (naxis.gt.2) then
          where = 1
          flag = 0
          posit = - xdis
          y3 = ylabel(3)(1:10)
          call axis(posit, 0., y3, where, ysize, ymin(5),
     !          ymax(5), flag)
        endif
        if (naxis.gt.3) then
          where = 3
          flag = 0
          posit = xsize + xdis
          y4 = ylabel(4)(1:10)
          call axis(posit, 0., y4, where, ysize, ymin(7),
     !          ymax(7), flag)
        endif
        if (naxis.gt.4) then
          where = 1
          flag = 0
          posit = -2.d0 * xdis
          y5 = ylabel(5)(1:10)
          call axis(posit, 0., y5, where, ysize, ymin(9),
     !          ymax(9), flag)
        endif
C
C Data scaling and plotting.
C
        xfact = xsize / (xmax - xmin)
        do 10 i = 1, nvar
          if (yes(i).eq.1) then
            yfact = ysize / (ymax(i) - ymin(i))
            xsca = (xdata(i,1) - xmin) * xfact
            ysca = (ydata(i,1) - ymin(i)) * yfact
            odd = mod(i,2)
            if (odd.eq.1) then
              idummy = (i - 1)/2 + 1 + ichar('0')
              sym = char(idummy)
              call plot (xsca, ysca, 3)
              call symbol (xsca, ysca, ht, sym, 0.0)
              do 20 j = 2, ndata(i)
```

```
            xsca = (xdata(i,j) - xmin) * xfact
            ysca = (ydata(i,j) - ymin(i)) * yfact
            xsca = xsca - 2.0 * ht/7.0
            ysca = ysca - ht / 2.0
            call symbol (xsca, ysca, ht, sym, 0.0)
20          continue
          else
            call plot(xsca, ysca, 3)
            do 30 j = 2, ndata(i)
              xsca = (xdata(i,j) - xmin) * xfact
              ysca = (ydata(i,j) - ymin(i)) * yfact
              call plot(xsca, ysca, 2)
30          continue
          endif
        endif
10    continue
C
C  Quitting the plotting routine and returning the terminal
C  to alpha-numeric mode.
C
      call plot (-3.0, -2.0, -3)
      call plot(0.0, 0.0, 999)
      call alpha()
      return
      end
```

Appendix F - A modification to the cell growth model

One of the major assumptions of the cell growth model is that the cell tries to maximize flux through the respiratory pathway. However, at low glucose concentrations, the respiratory pathway is unsaturated, i.e. it is not using its full capacity. Under such unsaturated conditions, if the medium contains ethanol, it is likely that the cell will consider uptake of ethanol as a plausible solution to maximization of flux through the respiratory pathway. Under aerobic conditions, ethanol can be utilized through respiratory metabolism. Using this reason, it was postulated that the rate of uptake of ethanol is proportional to the degree of unsaturation of the respiratory pathway. Rate of production of ethanol can be assumed to be proportional to the flux through the fermentative pathway.

$$\frac{d}{dt}(EV) = R_f Y_{ef} - \frac{k_E}{K_E + E}\left[1 - \frac{S}{K_r + S}\right]$$

The yield of cell mass based on ethanol uptake was assumed to be constant.

VITA

VITA

Anant Yeshawant Patkar was born on August 14, 1964 in Khanavli, India to Keshav and Sunita Patkar. He completed his primary and secondary education in Bombay, India and graduated from Parle College in June 1981. From August 1981 to May 1985 he attended the Department of Chemical Technology of Bombay University, Bombay, India. He received his Bachelor's degree in Chemical Engineering in May 1985. In August 1985, he joined the School of Chemical Engineering at Purdue University as a graduate student. He should receive his Ph. D. in December 1990. He will work as a postdoctoral research assistant with Dr. James Piret at the University of British Columbia, Vancouver, Canada. He is interested in pursuing an academic career in India.