

**MATHEMATICAL MODELS BASED ON  
MOLECULAR PROCESSES OF THE CELL**

Thesis by  
Alexandros Seressiotis

In Partial Fulfillment of the Requirements  
for the Degree of  
Doctor of Philosophy

California Institute of Technology  
Pasadena, California

1988

(Submitted August 24, 1987)

© 1988

Alex Seressiotis

All Rights Reserved

## ACKNOWLEDGEMENTS

It is sobering and heartwarming to realize how many people contributed to my personal, scientific, and professional growth for the past five years. A project of such a duration and intensity clearly could not be carried out without the support and encouragement of all these people. I would like to thank them all for being my friends and colleagues, for sharing the frustration and the joy that at times accompany scientific research, and for being a part of my life.

In particular, I would like to thank Professor James E. Bailey for his continuous support and encouragement throughout my graduate studies. His knowledgeable judgment, remarks, and advice, along with the broad freedom that he entrusted on me while conducting this project are deeply appreciated.

The assistance provided by April Olson in typewriting major parts of the thesis is deeply appreciated.

My initiation and progress to the world of academic research was paved with the help and efforts of a lot of previous and current members of the group. Sun Bok Lee provided the starting point for my own research. Friedrich Srienc, Alicia Loeffler, and Karl Friehs were more than friendly faces in a "windowless" office. Always ready to get involved into conversations leading from recombinant E. coli to the meaning of life, they have added a lot of spice into the daily monotonous routine. Jin Ho Seo, Ken Reardon and Chaitan Khosla furnished constructive interactions and help with dealing with the space opposite to my office. Unfortunately, I cannot mention everybody that made the long days and especially the long nights enjoyable. However I must tell you all that your help was deeply appreciated and looking back to the past five years I can truly say that it was fun. Thank you...

These lines could never be written without the constant and dear support of two very special people. My parents in Greece provided a dream for me to build on and made every possible effort to see that I am well equipped to carry out that dream. Thanks for letting me fly...

It would be unfair not to mention the Greeks of the round (lunch) table. Unfortunately the first face that strikes my mind is that of Costa Economou. His early departure from life left a void in my life that cannot be repaired. Leonidas Paparizos was always a source of fresh ideas and a good friend from day one up to now. Panos and Evie provided many times the home away from home and provided much needed advice, help, and laughter. Kostas Tsakalis and his speculative approach on life was always eager to offer his help at the expense of at least his own convenience. Chris Pilinis cared and always provided the "other" perspective of life. Thank you all guys. I hope, some day I will be able to pay you back.

Finally, it is hard to envision this work being completed if Maria had not been a part of my life. Our whole relationship served in more than one way as a platform to inspire and drive most of the effort involved in this thesis. Words are just not enough to express the effect of our relationship not only on my work but on my whole life. I just hope that there is much more to follow...

## ABSTRACT

A necessary first step in modeling the metabolic reaction network is a systematic procedure for determining the different reaction sequences which connect two metabolites. A software system (MPS) has been designed on the principles of artificial intelligence in order to address the problem of analysis and synthesis of metabolic pathways leading from one carbon-containing metabolite to another. MPS can be used to predict on a qualitative basis the effects of adding or deleting enzymatic activities to or from the cellular environment, to extract information about metabolic regulation, and to direct experiments in metabolic engineering. The main principles that have been used for the development of MPS are described along with case studies demonstrating the capabilities and potential applications of such a software system. The examples will examine carbon catabolic pathways and amino acid biosynthetic pathways. The catabolic pathways example concentrates on the conversion of glucose 6-phosphate to pyruvate. The output from MPS, which synthesized the classical catabolic pathways along with possible variations, leads to the identification of required genotypes/sets of enzymes that convert glucose 6-phosphate to pyruvate with different ATP and NAD(P)H coupling. The amino acid examples refer to the production of L-alanine from pyruvate and the identification of alternative pathways that perform this bioconversion.

The appropriate use of controllable promoters and plasmid origins of replication provide an opportunity for identifying operating strategies that maximize productivity of unstable recombinant cultures. This is demonstrated by first developing a kinetic model for product formation in recombinant cultures that exhibit both segregational and structural instability, and subsequently by identifying operating conditions that maximize productivity with respect to the base case (uncontrollable promoter and plasmid origin of replication).

Finally, an approach is being made to examine the validity of traditional-macroscopic relationships for the description of equilibrium cellular processes. A model is described that addresses these processes from a statistical point of view and results of this model are compared with analogous results obtained from traditional methods applied to a subset of the reactions that characterize the control mechanisms of the *lac* promoter.

## TABLE OF CONTENTS

Acknowledgments . . . . .	iii
Abstract . . . . .	v
Table of Contents . . . . .	vii
List of Illustrations . . . . .	xii
List of Tables . . . . .	xvi
<b>INTRODUCTION . . . . .</b>	<b>1</b>
<b>CHAPTER 1: MPS: AN ARTIFICIALLY INTELLIGENT SOFTWARE SYSTEM FOR THE ANALYSIS AND SYNTHESIS OF METABOLIC PATHWAYS . . . . .</b>	<b>5</b>
Introduction . . . . .	6
The Design of MPS . . . . .	8
Examples of Use of MPS . . . . .	10
1 Glucose 6-Phosphate to Pyruvate . . . . .	10
1.1 The Embden-Meyerhof-Parnas Family of Pathways . . . . .	11
1.2 The Pentose Phosphate Family of Pathways . . . . .	12
1.3 The Entner-Doudoroff Family of Pathways . . . . .	13
1.4 Pathways Involving By-Products . . . . .	15

2 Pyruvate to L-Alanine . . . . .	17
Conclusion . . . . .	19
Acknowledgment . . . . .	20
References . . . . .	21
Tables . . . . .	24
Figures . . . . .	26

**CHAPTER 2: A KINETIC MODEL FOR PRODUCT FORMATION IN  
UNSTABLE RECOMBINANT POPULATIONS . . . . . 36**

Introduction . . . . .	37
Background . . . . .	38
Stability of Recombinant DNA . . . . .	38
Effect of Cloned Gene Expression on Host Cell Growth . . . . .	39
Product Formation . . . . .	42
Kinetic Model . . . . .	44
Batch Culture . . . . .	44
Continuous Culture . . . . .	47
Simulation Results . . . . .	48
Model Parameters . . . . .	48

Batch Cultures . . . . . 49

Continuous Culture . . . . . 52

Discussion . . . . . 54

Acknowledgment . . . . . 56

Nomenclature . . . . . 57

References . . . . . 59

Table . . . . . 62

Figures . . . . . 63

**CHAPTER 3: OPTIMAL GENE EXPRESSION AND AMPLIFICATION**

**STRATEGIES FOR BATCH AND CONTINUOUS CULTURES . . . . . 71**

Introduction . . . . . 72

Mathematical Model and Controls Characteristics . . . . . 73

Batch Bioreactor Simulation Results . . . . . 75

Continuous Bioreactor Simulation Results . . . . . 78

Conclusion . . . . . 80

Acknowledgment . . . . . 81

References . . . . . 82

Figures . . . . . 83

**CHAPTER 4: INTRACELLULAR EQUILIBRIUM CALCULATIONS BASED ON  
SMALL SYSTEMS THERMODYNAMICS . . . . . 91**

Introduction . . . . . 92

Thermodynamics of an Ensemble of Small Closed Systems . . . . . 93

Application to the *lac* Repressor-Operator System . . . . . 96

Acknowledgment . . . . . 98

Nomenclature . . . . . 99

References . . . . . 100

Figures . . . . . 101

**APPENDIX A: KEY FEATURES OF THE DATA BASES AND PATHWAY  
SYNTHESIS PROGRAMS OF MPS . . . . . 104**

Description of the Data Bases . . . . . 105

Enzyme Data Base . . . . . 105

Substance Data Base . . . . . 107

MPS: The Program . . . . . 107

Initialization . . . . . 107

The Search Algorithm . . . . . 108

Tables . . . . . 113

Figures . . . . . 116

**APPENDIX B: THE ENZYME DATA BASE USED TO OBTAIN THE RESULTS  
OF CHAPTER 1 . . . . . .122**

## LIST OF FIGURES

**FIGURE 1.1:** The part of the metabolic chart related to glucose 6-phosphate catabolism.

**FIGURE 1.2:** Two of the pathways produced by MPS for the conversion of glucose 6-phosphate to pyruvate. Pathway 4 has been identified as the pentose phosphate pathway and pathway 9 as the Embden-Meyerhof-Parnas pathway.

**FIGURE 1.3:** Schematic representation of the Embden-Meyerhof-Parnas pathway.

**FIGURE 1.4:** Schematic representation of the pentose phosphate pathway.

**FIGURE 1.5:** Schematic representation of a variation of the pentose phosphate pathway.

**FIGURE 1.6:** A variation of the pentose phosphate pathway that leads to the concurrent production of pyruvate and erythrose 4-phosphate.

**FIGURE 1.7:** The schematic representation of a pathway in which substances assume a pseudocatalytic role.

**FIGURE 1.8:** The schematic representation of two pathways leading from pyruvate to L-alanine.

**FIGURE 1.9:** One of the four pathways that MPS synthesized for the conversion of pyruvate to L-alanine.

**FIGURE 1.10:** The schematic representation of the pathway show in Figure 9.

**FIGURE 2.1:** Schematic diagram of basic functional elements in an expression plasmid (A) and diagram of genetic transitions in a recombinant population with structural and/or segregational instability (B).

**FIGURE 2.2:** Alternate forms of the protein synthesis rate function  $f$  [Eq. (9.1)] considered in these models.

**FIGURE 2.3:** Simulated batch bioreactor trajectories for pure segregational instability ( $\theta=0$ ;  $\phi=0.01$ ) without product inhibition ( $n=0$ ).

**FIGURE 2.4:** Calculated productive cell fraction  $\psi$  and product concentration after substrate exhaustion in a batch bioreactor as a function of cell plasmid content (pure segregational instability;  $\theta=0$ ,  $\phi=0.01$ , without product inhibition;  $n=0$ . Calculations for  $f(\mu)$  given by Eq. 2.30 ( $\bullet$ ), Eq. 2.31 ( $\diamond$ ), and Eq. 2.29 ( $\circ$ )).

**FIGURE 2.5:** Dependence of final product content and productive cell fraction after substrate exhaustion in a batch bioreactor for different instability and product inhibition models. Solid symbols correspond to inhibition based on intracellular product concentration (Model I), and open symbols correspond to Model II which employs overall broth product concentration to evaluate product inhibition. Circles are for pure structural instability ( $\theta=0.01$ ,  $\phi=0$ ), and squares were calculated for pure segregational instability ( $\theta=0$ ,  $\phi=0.01$ ).  $f(\mu)$  given by Eq. 2.29 for all cases.

**FIGURE 2.6:** Simulated final productive cell fraction and product concentration versus the gene expression parameter,  $\gamma$ , characteristic transcription and translational efficiency, for product inhibition Model I ( $\bullet$ ) and Model II ( $\circ$ ). ( $\phi=0.01$ ,  $\theta=0$ ,  $N_p=50$ )

**FIGURE 2.7:** Computed time trajectories of total cell concentration and productive cell fraction following initiation of continuous-flow bioreactor operation. Dilution rates shown:  $D=0.1, 0.2, 0.3, 0.4, 0.5, 0.6$ , and  $0.7 \text{ hr}^{-1}$ . ( $\phi=0.01$ ,  $\theta=0$ , product inhibition Model I, Eq. 2.29 for  $f(\mu)$ .)

**FIGURE 2.8:** Experimentally measured interferon activity obtained from recombinant *E.coli* fermentation as a function of inducer (nalidixic acid) concentration. Inducer concentration is correlated with cloned-gene transcriptional activity (data from Ref. 43 of Chapter 2).

**FIGURE 3.1:** Diagram of the assumed trajectories for cellular plasmid content and gene expression parameter following a switch to environmental conditions promoting amplification and gene expression at time  $t_{AMP}$  or  $t_{EXP}$ , respectively.

**FIGURE 3.2:** (A) Final cloned-gene product concentration (grams protein/liter culture) vs. the switching time (the time of plasmid amplification (o) and time of cloned-gene induction (●)). (B) Fraction of cells containing plasmids after substrate exhaustion in the batch process as a function of initiation times for amplification (o) or gene expression (●).

**FIGURE 3.3:** Illustration of product concentration at substrate exhaustion as a function of the time of amplification ( $t_{AMP}$ ) and the time of expression ( $t_{EXP}$ ) when both events can be activated independently at different points during batch operation. For the particular parameters considered here, maximum product accumulation is provided for  $t_{AMP}=10$ ,  $t_{EXP}=10$ ,  $N_{P,final}=50$ ,  $\gamma_i=1.0$ .

**FIGURE 3.4:** Final productive cell fraction vs. amplification and expression switching times for batch operation. Same parameters as Fig. 3.3.

**FIGURE 3.5:** Time trajectories (solid lines) of the productive cell fraction and cloned-gene product concentration vs. operating time in a batch reactor with  $t_{AMP}=10$ ,  $t_{EXP}=10$ . Shown for comparison (dashed lines) are trajectories for the uncontrolled case with  $N_P=50$ ,  $\gamma=1.0$ .

**FIGURE 3.6:** Final product and productive cell fraction values for different induced gene expression activity values  $\gamma$  as a function of induction time  $t_{EXP}$ ; batch reactor.

**FIGURE 3.7:** Calculated final product concentration vs. amplification initiation switching time for negligible product inhibition (o:  $N_{P,final}=50$ ; ●:  $N_{P,final}=90$ ;  $\Delta$ :  $N_{P,final}=120$ ).

**FIGURE 3.8:** Simulated effluent product concentration and productive cell fraction for three different CSTR reactor cascades. Values are shown as functions of time since startup. Conditions in tanks 1, 2, and 3 favor cell growth, cause plasmid amplification, and induce gene expression, respectively. Different symbols correspond to different  $V_1:V_2:V_3$  tank volume ratios.  $\circ=1:1:1$ ;  $\bullet=1:1.5:3$ ;  $\Delta=3:1.5:1$ .

**FIGURE 4.1:** The results of the statistical (points) and the traditional (lines) approaches for equilibrium calculations for the *lac* repressor-operator system. Here  $[O]_0=1$  molecule/cell.

**FIGURE 4.2:** Equilibrium concentrations calculated for the *lac* repressor-operator system based on the statistical (points) and the traditional (lines) approaches. In this case  $[O]_0=15$  molecules/cell.

**FIGURE 4.3:** The relative error  $[(O_{\text{trad}}-O_{\text{stat}})/O_{\text{stat}}]$  associated with the use of the traditional single-cell intracellular equilibrium relations for the calculation of the ratio  $[O]/[O]_0$ .

**FIGURE A1:** The general representation of enzymes in the data base.

**FIGURE A2:** The representation of *homoserine acyltransferase* and *phosphoglyceromutase* in the enzyme data base.

**FIGURE A3:** The representation of 3-phospho-D-glyceroyl phosphate and O-phospho-L-homoserine in the substance data base.

**FIGURE A4:** A simple search network.

**FIGURE A5:** Reaction networks incorporate multiple linear pathways in order to provide the necessary substances for reactions requiring more than one reactant.

**FIGURE A6:** Reaction networks incorporate cyclic pathways when substances that have been already consumed are regenerated.

## LIST OF TABLES

**TABLE 1.1:** The overall stoichiometry of the ten pathways converting glucose 6-phosphate to pyruvate with no carbon-carrying side products (except CO<sub>2</sub>) normalized on the basis of production of one mole of pyruvate. The stoichiometric coefficients for ADP, NAD<sup>+</sup>, NADP<sup>+</sup> and Ferricytochrome C have the same values but opposite signs as the coefficients of ATP, NADH, NADPH, and Ferrocycytochrome C, respectively.

**TABLE 1.2:** Enzyme participation in the ten pathways converting glucose 6-phosphate to pyruvate with no carbon-carrying side products (except CO<sub>2</sub>).

**TABLE 2.1:** Summary of parameter values used in model calculations.

**TABLE A1:** The fields used for enzyme descriptions.

**TABLE A2:** The fields used for substance descriptions.

**TABLE A3:** The rules involved in the network of Figure A4.

## ***INTRODUCTION***

The advent of genetic engineering technology has provided an array of techniques and methods that can be used to genetically modify microorganisms in order to biologically produce bioproducts of interest. These modifications, however, introduce an unnecessary and unnatural burden to the host cell. The host organism is not evolutionarily optimized towards the production of some bioproduct of interest to us, and having to use its enzymatic machinery to carry out the additional load of instructions (plasmid replication, foreign gene expression, etc.), its own growth and survival is inhibited. Therefore, most of the applications of recombinant DNA technology concentrate on the minimization of the interactions of the host cell with the additional artificially implanted plasmids or genes. This minimization hopefully leads to the minimization of the exhibited growth inhibition, providing a relatively healthy recombinant cell population that is capable of producing the bioproduct(s) of interest. The base level of interactions that a host organism can have with a plasmid constitutes of sharing its replication, transcription, and translation enzymatic sets to reproduce and express the plasmid and the genes on the plasmid. Therefore, an effort has been made in the past to identify host-vector systems in which the bioproduct of interest would be inactive within the host's environment.

An inactive bioproduct would clearly provide a minimum level of interactions between the host cell and the plasmid. However, this approach is not always possible or justifiable. For some bioproducts it can be extremely hard to find an organism in which these substances are biologically inactive. If, for example, one wishes to genetically modify an organism in order to overproduce an amino acid, then, one cannot find a living organism in which a given amino acid is not biologically active. On the other hand, the activity of biomolecules, and especially the activity of polypeptides, depends not only on the chemical structure but also on the conformation and tertiary structure that has been attained. The biologically active conformation is by definition achieved in an environment that the molecule is active. Therefore, it is necessary to examine the

possible interactions of a given bioproduct with the overall cellular function and especially with cellular metabolism.

The metabolic cellular network is an extensive and complex network reactions the majority of which cannot be kinetically analyzed with the current analytical techniques. A first approach to addressing this problem is to analyze the qualitative aspects of the metabolic network and examine the effects of possible addition to or deletion from the network of enzymatic activities. The motivation behind this approach is that metabolism has evolved to a network of reactions not by accident but by the necessity to offer to the living cell the adaptivity and flexibility required by life. The identification of the properties of the network, properties that can be derived by examining possible routes for the interconversion of given metabolites to target metabolites, is an important a necessary step for modeling cellular metabolism. Such a step has been taken and is described in the first chapter of the thesis. The constructed algorithm (MPS) can help the biochemical engineer to decide about the required genotypes or organisms that will best serve to the overproduction of a metabolite of interest.

A second important decision process is that of identifying the operating conditions that would lead to the optimization of the productivity of recombinant DNA cultures. Given that the existence of a foreign plasmid inhibits the growth of the host cell, then if an instability on the plasmid reproduction can lead to the appearance of two populations one carrying the plasmid and one being the wild type, the wild type cells have a growth advantage over the plasmid carrying ones. This can under certain conditions lead to the extinction of the recombinant population. The scenario becomes even more unfavorable for the plasmid carrying cells if product inhibition is also considered. Product inhibition can be present both for a biologically inactive and a biologically active product. It can be safely assumed, though, that the latter case (biologically active substance) will lead to higher growth inhibition. In order to examine these phenomena, a kinetic model has been

developed for product formation in unstable recombinant DNA populations. For the description of product formation the model is based on classical chemical engineering kinetics descriptions for the reactions characterizing gene expression. The full model and its development is described in the second chapter. In the third chapter the model is being used to derive optimal operating strategies that maximize productivity of both batch and continuous fermentors, based on the assertion that one can use controllable plasmid replicons and gene promoters.

Finally, cells are most of the time modeled as ideally stirred reactors (and this approach has been used for the development of the product formation model of chapter 2). This cannot be always assumed as being a correct approach. Key biological substances exist in the cell in very small amounts (of the order of a few molecules per cell). It is, therefore, doubtful that traditional approaches involving macroscopic descriptions of such phenomena as reactions at equilibrium are valid for modeling molecular cellular processes. A more rigid approach would consist of a statistical description of the processes. That approach has been derived in chapter 4 for the characterization of reactions at equilibrium, and has been applied to a subset of the reactions that constitute the control mechanism of the *lac* operator-promoter. The results obtained are compared with the results of the traditional approach for the same system.

## **CHAPTER 1**

***MPS: AN ARTIFICIALLY INTELLIGENT SOFTWARE SYSTEM  
FOR THE ANALYSIS AND SYNTHESIS OF METABOLIC  
PATHWAYS***

## INTRODUCTION

The objective of metabolic engineering is to redirect cellular metabolism in order to achieve higher or lower yield or rate of formation of bioproducts. The decision of which microorganism and conditions to use is the primary problem faced by the biochemical engineer. Different microorganisms incorporate different sets of enzymes for the degradation of nutrients and synthesis of bioproducts, and the same microorganism uses different subsets of its enzymes in different media. Moreover, genetic engineering may be applied for modifying the metabolic reaction network employed by a cell. The ability to specifically mutate genes that code for enzymes offers the opportunity of removing enzymatic activity from the cellular environment. On the other hand, the advent of recombinant DNA technology provides the capability of introducing through plasmids new or supplemental quantities of existing enzymes or proteins that inhibit or amplify existing enzymatic activities. A systematic tool to predict the effects of such genetic modifications on the productivity of the desired compound can be very valuable in the design of optimal bioprocesses.

Cellular metabolism has, through evolution, developed into a large and complex reaction network. Through an extensive number of intermediates and allosteric enzymatic control mechanisms, the functional adaptivity and stability required by life is maintained. The ability of a bacterial cell to survive under a variety of environmental conditions, and sometimes under major genetic modifications, is largely due to the flexibility of the metabolic reaction network to decrease or increase the rate of substrate consumption or product-intermediate formation. This freedom is expressed by choices among alternative pathways which, through different sets of reactions, utilize substrates, energy, and reducing power with different stoichiometric coupling, leading to different product, energy and reducing power yield factors.

The complexity of cellular metabolism and, mainly, the fact that the kinetic behavior of most of the enzymes is not accessible *in vivo* restrict the development of detailed models based on chemical engineering kinetics. Such models have been

developed based on a small number of key enzymatic activities and are not easily expandable.<sup>1, 2</sup>

In order to predict on a qualitative basis the effects of genetic modifications, such as deleting or adding enzymatic activities to a cell, a software system has been developed that synthesizes metabolic pathways. The system, called MPS, is based on the principles of artificial intelligence in order to analyze the metabolic network efficiently. MPS software embodies artificial intelligence concepts in the sense that a complicated search is conducted guided by rules embodied in the program to conduct the search efficiently. Solution of complex search problems is one of the major components of artificial intelligence research and application.<sup>3-17</sup>

Although MPS does include screening and pathways analysis features which aid the user to investigate and interpret information obtained from the set of pathways synthesized, it is not an "expert system" in the sense that it attempts to relieve the user of the responsibility for some knowledge and for applying at least modest levels of inductive and deductive reasoning. Thus, MPS is intended as a research and development tool to be used by those with a basic perspective on the purposes and strategies of metabolism.

MPS is designed to generate the pathways that convert a given substance to a target metabolite. It generates and reports only the genetically independent pathways, while at the same time the synthesis algorithm overcomes the cycles and other degeneracies existing in the metabolic network. Cyclic sequences of enzyme-catalyzed steps and the presence of multiple enzymatic activities that catalyze the same reaction are the major problems of analyzing the metabolic network through classical approaches, since they introduce singularities in the stoichiometric matrices used for such approaches.<sup>18, 19</sup>

Artificial intelligence and graph theory concepts have been previously applied for different purposes in chemistry,<sup>20</sup> and the potential for some applications of artificial intelligence in biotechnology has been discussed in recent articles.<sup>21, 22</sup> Research has been conducted in the past for molecular representations and enu-

meration of isomers, <sup>23</sup> for chemical documentation, <sup>24</sup> for representing chemical reactions, <sup>25, 26</sup> and for organic synthesis. <sup>27, 28, 29</sup> However, the problem of analyzing the metabolic network is fundamentally different from any problem of organic synthesis. <sup>30</sup> The heuristics that organic chemists (and, therefore, artificially intelligence programs that address the problem of organic synthesis) apply can not be used in the synthesis and analysis of metabolic pathways. The conversion of glucose to pyruvate, for example, can be performed in the laboratory environment in many different ways, but generally the actual set of reactions that has been chosen by living organisms (these will be presented in a later section of this paper) cannot be followed *in vitro*. These reactions are the only ones available within the cellular environment, and involve a set of intermediates, the importance of which is not clear when a given conversion is examined, but which is revealed when alternative pathways are also examined. Furthermore, the energetics of bioconversions are also important for metabolic pathways. For example, from a chemist's point of view, the conversion of glucose to CO<sub>2</sub> and H<sub>2</sub>O is a trivial procedure, while, from the cellular perspective, this oxidation involves both the pathways to pyruvate and the tricarboxylic acid cycle. The stoichiometry of ATP, NADH and NADPH production by a cell oxidizing glucose can vary significantly, depending on the pathways employed.

## THE DESIGN OF MPS

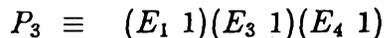
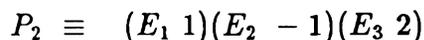
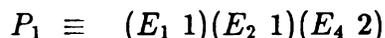
MPS consists of data bases for enzymes and substrates and a search and screening algorithm. MPS has been written in Common Lisp, <sup>31</sup> using a GCLISP LM interpreter and compiler on an IBM PC AT, equipped with 3.5Mb of CPU memory. MPS calls on two data bases, one which stores substances and the other which stores enzymes and characteristics of the corresponding enzyme-catalyzed reaction *in vivo*. Because of limited data on intracellular metabolite concentrations, Gibbs free energy changes associated with metabolic reactions *in vivo* are generally not

available. Consequently, the enzyme data base in MPS categorizes the associated reaction either as reversible or as irreversible *in vivo*. Further details on the structure of these data bases, crucial to the efficient functioning of MPS, are provided in the Appendix.

The search algorithm has been designed to generate all the independent pathways of transforming a given intermediate to a given target substance. The definition of independency of pathways employed in MPS is illustrated by the following example: Consider the reaction network set:



Given  $A$  to be the initial substance and  $F$  the target substance, how many independent ways exist for the transformation of  $A$  to  $F$ ? The description of pathways that is used by MPS consists of a series of enzyme nodes (*enzyme  $\nu$* ), where  $\nu$  is the number of times the standard reaction catalyzed by *enzyme* participates in the pathway. Then the following pathways of converting  $A$  to  $F$  exist:



The overall reaction for all three pathways is  $A \rightarrow 2F$ . Are these three pathways independent? From a traditional point of view based on linear independence of the reactions, these are not independent since:

$$P_3 = 0.5P_1 + 0.5P_2$$

However, if the question of independency is posed in the form: "Does each pathway shown define an independent genotype?", then each of these three pathways is indeed independent. This is the way that MPS defines the independence of pathways.

To reduce the computational effort and computer memory required to elaborate and store metabolic pathways, it is important to adopt efficient procedures for identifying routes to certain required enzyme co-substrates. These and other rules designed to streamline the combinatorial growth of an unwieldy set of expanding pathways are summarized and illustrated by example in the Appendix.

The search algorithm, when completed, returns the pathways performing the transformation of the initial substance to the target substance. The fact that LISP is very efficient in handling symbolic information is very helpful in order to identify subsets of the pathways that fulfill some property. In order to make the user-MPS interaction easier, a menu-driven program that performs screening of the pathways based on a limited number of criteria has also been incorporated in MPS. The criterion that the user may require for identifying a subset of pathways of interest could be any combination of requirements that specified enzymes participate or not in the pathway, or for substances to be involved in the overall reaction of the pathway as substrates, as products, or not to be involved at all. Pathways can be displayed and stored through this additional screening program in a readable manner as will be shown in the examples that follow.

## **EXAMPLES OF USE OF MPS**

In order to illustrate the performance of MPS, the conversion of glucose 6-phosphate to pyruvate and the conversion of pyruvate to alanine will be considered. The data base used was the full scale data base presently available to MPS. There are approximately 90 enzymes and 120 substances in the current enzyme and substance data bases.

### **1 GLUCOSE 6-PHOSPHATE TO PYRUVATE**

The part of the metabolic chart related to the conversion of glucose 6-phosphate to pyruvate is shown in Figure 1.1.

MPS constructed twenty independent pathways for converting glucose 6-phosphate to pyruvate. MPS synthesis of all of these took 46 seconds to be completed. Screening on the basis of no carbon-carrying by-products, except CO<sub>2</sub>, resulted in ten independent pathways. Information about these ten pathways can be found in Tables 1.1 (stoichiometry) and 1.2 (enzyme participation). The actual output from MPS for two pathways is shown in Figure 1.2. The information reported by MPS for each pathway consists of:

1. The set of enzymes that are involved in the pathway.
2. The number of times that the standard reaction for each enzyme stored in the data base is used by the pathway. In the case that a reversible reaction is used in a direction opposite direction from the one stored, a negative number is reported.
3. The overall reaction and stoichiometry of the pathway.

### 1.1 The Embden-Meyerhof-Parnas Family of Pathways

The most classical pathway for the catabolism of glucose 6-phosphate, the EMP pathway, is shown schematically in Figure 1.3. MPS reported the EMP as Pathway 9 (Tables 1.1, 1.2 and Figure 1.2).

What other variations of this pathway have been generated by MPS? Observe that in Figure 1.1, there exists a way of converting dihydroxyacetone phosphate to pyruvate that does not involve glyceraldehyde 3-phosphate. This conversion is the methylglyoxal bypass.<sup>32</sup> Therefore, when fructose 1,6-biphosphate is converted to dihydroxyacetone phosphate and glyceraldehyde 3-phosphate, one option is to convert the generated dihydroxyacetone phosphate to glyceraldehyde 3-phosphate for a yield of 2 moles of glyceraldehyde 3-phosphate per mole of glucose 6-phosphate consumed, and carry on with the conversion of glyceraldehyde 3-phosphate to pyruvate. This option generates the EMP pathway. A second option, though, is for the *triosephosphate isomerase* reaction to be followed in the opposite direction, i.e.

glyceraldehyde 3-phosphate to be converted to dihydroxyacetone phosphate, and then the methylglyoxal bypass to be followed. This variation of the EMP pathway was reported by MPS (Pathway 10).

What happens if a microorganism is not capable of performing the *triosephosphate isomerase* reaction (due to missing or inactivated enzyme)? Can pyruvate be produced via a variation of the EMP pathway and in a yield of two moles of pyruvate per mole of glucose 6-phosphate consumed? Provided that the enzymatic activities involved in the two branches originating from glyceraldehyde 3-phosphate and dihydroxyacetone phosphate, respectively, are present in the cell, then the answer to this question is yes (Pathway 2). If, however, an enzyme in one of the two branches is missing, then this conversion is not possible. In that case one mole of pyruvate will be produced from one mole of glucose 6-phosphate, and one mole of one of the two triose phosphates will be a side product in the conversion. These pathways have been also reported by MPS. The MPS search algorithms carry out the logical decisions and branching summarized in the previous discussion without user intervention in order to determine the genetically independent pathways. These comments are intended to illustrate the types of logic which have been implemented in MPS in order to synthesize these pathways and those discussed below.

## 1.2 The Pentose Phosphate Family of Pathways

Next the pentose phosphate pathway, shown as Pathway 4 in Tables 1.1 and 1.2 and Figure 1.2, and schematically in Figure 1.4, will be considered. One difference with the EMP family of pathways (2, 9 and 10) that can be immediately noticed is that NADPH, absent from the EMP overall reaction, is generated by the pentose phosphate pathway. A second evident difference is the pyruvate yield. The production of CO<sub>2</sub> that is involved in the pentose phosphate pathway leads to a lower pyruvate yield of less than two moles per mole of glucose 6-phosphate processed. Overall yields and stoichiometries are provided by MPS output; no user input or calculations are required to obtain these quantities.

Since, as shown in Figure 1.4, the same structure for the glyceraldehyde 3-phosphate conversion to pyruvate that was seen in the EMP pathway is present in the pentose phosphate pathway, MPS has reported the variations involving the methylglyoxal bypass that were mentioned earlier (Pathways 1 and 3).

By comparing the enzymatic sets of the EMP pathway and the pentose phosphate pathway (Figure 1.2), we can see that all but the first enzyme of the EMP pathway, *glucose phosphate isomerase*, are involved in the pentose phosphate pathway. What would happen if a cell did not have the ability to convert fructose 6-phosphate to fructose 1,6-biphosphate? Clearly, all the pathways of the EMP family would be eliminated. What about the pentose phosphate pathway? Is there a variation of this pathway that will make the cell capable of producing pyruvate? Observe in Figure 1.4 that, at the point of the pentose phosphate pathway where fructose 6-phosphate has been generated, glyceraldehyde 3-phosphate has been generated as well. Two molecules of fructose 6-phosphate and one molecule of glyceraldehyde 3-phosphate are produced from three molecules of glucose 6-phosphate. Then, a possible variation of this pathway is the following: two molecules of fructose 6-phosphate, instead of going through the EMP pathway, are recycled via the *glucose phosphate isomerase* to glucose 6-phosphate. MPS utilizes this logic. The overall stoichiometry incorporates the net consumption of one molecule of glucose 6-phosphate and the net production of one molecule of glyceraldehyde 3-phosphate, which can be converted to pyruvate. This variation has been reported by MPS (Pathway 5), and is schematically shown in Figure 1.5. So it is possible for the cell to convert glucose 6-phosphate to pyruvate without possessing *phosphofructokinase*, the enzyme which catalyzes the production of fructose 1,6-biphosphate. The methylglyoxal bypass variation of the pathway has been also reported (Pathway 6).

### 1.3 The Entner-Doudoroff Family of Pathways

The remaining pathways reported by MPS are the two variations of the ED pathway, one with the methylglyoxal bypass (Pathway 8) and one without the methylglyoxal

bypass (Pathway 7). Note that the ED family of pathways has a pyruvate yield of two moles per mole of glucose 6-phosphate (like the EMP family of pathways) and produces NADPH (like the pentose phosphate family of pathways).

...

Other information provided by MPS and the resulting pathways will next be examined. By inspection, it is easy to deduce what pathways are eliminated if one or more enzymes are deactivated by mutation. Consider Table 1.2, which has also been generated by MPS, and which refers to the ten pathways discussed earlier. Deletion of the enzyme *gluconolactonase*, for example, will lead to the elimination of pathways 1 and 3 through 8, and the resulting genotype will have the ability to convert glucose 6-phosphate to pyruvate and no metabolites as by-products only via pathways 2, 9 and 10. These pathways may be easily identified in user interaction with MPS by requesting all pathways which lack this enzyme and which produce no by-product metabolites. On the other hand, suppose that a cell does not have the ability to use the ED pathway, but is able to use both the EMP and the pentose phosphate pathway (wild type *E. coli* does not use the ED pathway, but uses both EMP and pentose phosphate pathways to catabolize glucose<sup>33</sup>). The enzymatic activity in the ED pathway (Pathway 7) that is not contained in either the EMP pathway (Pathway 9) or the pentose phosphate pathway (Pathway 5) consists of two enzymes: *phosphogluconate dehydratase* and *phospho-2-keto-3-deoxy-gluconate aldolase*. Therefore, in order to establish the ED pathway in *E. coli*, the genes for these two enzymes must be introduced into the organism.

MPS also identifies the groups of enzymes indicated in Table 1.2. Each group has the property that, when one enzyme of the group appears in one of the ten pathways, the rest of the enzymes in that group also appear. There are two enzymes, *glucosephosphate isomerase* and *triosephosphate isomerase*, defining single enzyme groups. Both of these enzymes catalyze reversible isomeration reactions and, depending on the direction that the reactions proceed, produce intermediate metabo-

lites that are consumed by other groups. There are also six multienzyme groups. The first enzyme in each group catalyzing a reaction irreversible under physiological conditions commits its substrate(s) to all of the reactions of the group. It can be hypothesized, then, that metabolic regulation has been most likely designed to affect the enzymatic activity of the first enzyme in each group catalyzing an irreversible reaction. This statement is true for amino acid biosynthetic pathways.<sup>34</sup> This heuristic approach leads to the identification of five important enzymes: *phosphogluconate dehydrogenase*, *glucose 6-phosphate dehydrogenase*, *phosphogluconate dehydratase*, *glyceraldehyde phosphate dehydrogenase*, and *6-phosphofructokinase* (according to the information in the data base, none of the enzymes in the sixth group catalyzes an irreversible reaction), that should be considered the most important of the set of twenty-two enzymes involved in the ten catabolic pathways found.

Finally, as shown in Table 1.1, the overall reaction for each of the ten pathways involves different utilization/production stoichiometry for ATP, NADH and NADPH. This information is important, and will be used in the future to further classify the pathways with respect to cellular objectives and to deduce information about the regulation of glucose catabolism.

#### 1.4 Pathways Involving By-Products

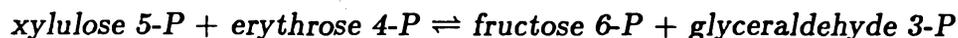
We have examined the pathways that convert glucose 6-phosphate to pyruvate and do not produce any carbon-carrying by-products. The remaining pathways generated by MPS involve the coproduction of pyruvate and some other carbon-carrying intermediate metabolite. In order to further illustrate how MPS can be used, two of these pathways will be examined next.

Assume that:

1. A given microorganism catabolizes glucose 6-phosphate via the pentose phosphate pathway, and

2. Erythrose 4-phosphate (a precursor for the biosynthesis of *L*-tyrosine) is a bio-product of interest, and we want the catabolic pathway to lead to production of erythrose 4-phosphate and pyruvate.

By screening the pathways generated to pyruvate by MPS for those that produce erythrose 4-phosphate as well, the genetic modifications required to transform the wild-type microorganism to one with the desired property are readily identified. There is one pathway, using a subset of the enzymatic activity found in the pentose phosphate pathway, leading to the concurrent production of pyruvate and erythrose 4-phosphate. The pathway is shown in Figure 1.6. Compared with the full pentose phosphate pathway (Pathway 9 in Figure 1.2), it is easily identified that the difference is deletion from the pentose phosphate pathway of the enzyme *transketolase-2*, catalyzing the reversible reaction:



Therefore, deletion of this enzyme is the genetic modification required for the concurrent production of pyruvate and erythrose 4-phosphate in this hypothetical microorganism. The user may readily draw this conclusion by comparison of the two pathways, but this result could also be determined directly by using the screening capabilities of MPS.

Consider now the case of a microorganism that catabolizes glucose 6-phosphate only by the pathway shown in Figure 1.5. The enzymatic activities involved in this pathway are presumed to be the only carbon catabolism activities present in this hypothetical microorganism. Assume further that the enzyme *glucose 6-phosphate dehydrogenase*, catalyzing the first reaction of this pathway:

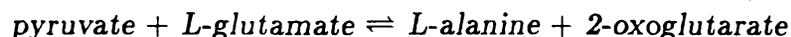


is deleted. Then the pathway is eliminated, and given the remaining enzymatic activity in the microorganism, it appears that there is no way of converting glucose 6-phosphate to pyruvate (with or without by-products), at least none that is obvious

from Figure 1.1. However, MPS has synthesized a possible pathway for this case. It produces sedoheptulose 7-phosphate in addition to pyruvate, and it is schematically shown in Figure 1.7. Note the pseudocatalytic role that glyceraldehyde 3-phosphate and erythrose 4-phosphate play in order to accommodate the production of xylulose 5-phosphate. This feature of MPS will be further examined in the following section.

## 2 PYRUVATE TO L-ALANINE

In order to further illustrate the capabilities of MPS, the conversion of pyruvate to L-alanine has also been analyzed. The classical route for the production of L-alanine from pyruvate is through a simple transamination reaction:<sup>32</sup>



Since the starting substance is pyruvate and L-glutamate is also required, in order for the above reaction, which is catalyzed by *alanine aminotransaminase*, to be carried out, L-glutamate must first be formed from pyruvate. However, the immediate precursor of L-glutamate is 2-oxoglutarate. Therefore, the pyruvate to L-glutamate pathways can be decomposed to two parts:

- A pathway converting pyruvate to 2-oxoglutarate, and
- A pathway converting 2-oxoglutarate to L-glutamate.

However, 2-oxoglutarate is produced by the alanine transamination reaction. Therefore, only the reaction(s) transforming 2-oxoglutarate to L-glutamate are required from the pyruvate to L-glutamate pathways.

MPS has been designed in such a way that an expanding pathway cannot utilize an enzyme unless all the substrates are present. The products of any given reaction can only be used for the expansion of a pathway only if the reaction itself has been already utilized by the pathway. Therefore, MPS, by design, does not “know” that a molecule of 2-oxoglutarate is produced by the transamination reaction, and, therefore, 2-oxoglutarate cannot provide the required L-glutamate. The strategy of

using 2-oxoglutarate (or any other substance in an analogous case), can be easily incorporated in the existing version of MPS, but it will greatly increase the required execution time, since for each search problem several subsearching tasks (for our example, generate L-glutamate from 2-oxoglutarate in addition to the main task of conversion of pyruvate to L-alanine). The strategy that has been actually employed by MPS was the following: it produced L-glutamate from pyruvate, and, when the transamination reaction took place, MPS observed the regeneration of 2-oxoglutarate and eliminated the pathways leading from pyruvate to 2-oxoglutarate. The drawback in such a design principle can be illustrated in the following example: Assume that a microorganism does not have the enzymatic activity to produce 2-oxoglutarate from pyruvate, but has the enzyme *L-glutamate dehydrogenase* which converts 2-oxoglutarate to L-glutamate. Then if MPS was implemented with a data base reflecting the enzymes of this hypothetical microorganism, it wouldn't be able to carry out the transamination reaction, and the program would not report that L-alanine could be produced by this reaction. This pathway, however, would be found if the enzymatic data base employed by MPS was "conservatively" built, i.e., no relevant enzymatic activities were absent from the data base even if these activities were absent from a given microorganism.

MPS synthesis for the conversion of pyruvate to L-alanine produced four pathways. The third pathway involves the set of reactions discussed in the previous paragraph. This pathway is also shown in Figure 1.8. There is, however, a second way to convert 2-oxoglutarate to L-glutamate, involving an intermediate step of glutamine production. This second way has been utilized by MPS in order to produce the fourth pathway reported which is also shown schematically in Figure 1.8, where the additional reactions are noted with dotted lines.

Two more pathways were also reported by MPS. These pathways produce hydroxypyruvate as a byproduct; the first one is shown in Figure 1.9 and schematically in Figure 1.10. The second pathway involves the alternative reaction set to convert 2-oxoglutarate to L-glutamate. These two pathways incorporate a differ-

ent transaminase, *serine pyruvate aminotransferase*, in order to produce alanine. Therefore, a cell possessing the enzymes listed in Figure 1.9 is capable of producing alanine even in the case that *alanine aminotransferase* has been inactivated.

At this point, it is also appropriate to comment on another aspect of MPS. MPS is not capable of predicting reactions which might be catalyzed by enzymes unknown to the program. The pathways synthesized by MPS depend on what enzymes and corresponding reactions are stored in the data base. The data base at this time does not have the description of the enzymes that convert hydroxypyruvate to 3-phospho-glycerate. Therefore, additional pathways that would have converted the hydroxypyruvate by-product to 3-phospho-glycerate, hence eliminating the need for the pyruvate to 3-phospho-glycerate pathway incorporated in pathways 1 and 2, have not been reported.

## CONCLUSION

MPS is an easily expandable prototype for the synthesis and analysis of metabolic pathways. The design principles used lead to the efficient use of the enzyme data base. As has been illustrated, by the given examples, MPS has the ability of handling complex reaction pathways. The first example, glucose 6-phosphate to pyruvate conversion, illustrated the ability of MPS to synthesize the catabolic pathways of glucose that have been already reported in the literature. MPS can be equally useful in exploring less examined areas of the metabolic network and in producing valuable information about the effects that addition and/or deletion of certain enzymatic activities will have on the production of metabolites and bioproducts of interest.

MPS relies totally on the information stored in the data bases. By storing more enzymatic activities, further insight into the structure and functionality of the metabolic network can be obtained by the user. The data bases can be also expanded in another way. With each enzyme can be associated a description of the

organisms in which the enzyme activity is present. That will allow the immediate identification of pathways consistent with the genotype of a single organism, and therefore we can identify organisms capable of producing a bioproduct of interest. Furthermore, for each enzyme we can also include information about the gene that codes for the enzyme. This information can include the set of organisms in which an active gene product can be expressed, enabling identification of the candidate-genes that should be introduced into a organism through a plasmid, in order to make an organism capable of using pathways that the wild-type genotype lacks.

The information stored in the substance data bases can also be expanded. By including for each substance the set of organisms that have the ability to uptake this substance from the environment or secrete this substance to the environment, then MPS could identify overall fermentation equations and appropriate substrates that should be used in order to direct metabolism to the production of a desired biochemical.

Finally, the results provided by MPS can be used to classify pathways with respect to cellular objectives and for the identification of possible control points of the metabolic network. Preliminary results in this direction have been presented recently,<sup>35</sup> and more extensive applications of MPS for this purpose will be presented in future publications.

**Acknowledgment:** This work has been supported by the Energy Conversion and Utilization Technology (ECUT) program of the U. S. Department of Energy.

## References

1. M. M. Domach, S. Leung, R. E. Cahn, G. G. Cocks, and M. L. Schuler, *Biotech. Bioeng.*, **26**, 203 (1984).
2. S. W. Perreti and J. E. Bailey, *Biotechnol. Bioeng.*, **25**, 1672 (1986).
3. A. Barr and E. A. Feigenbaum, eds., *The Handbook of Artificial Intelligence*, Vol. 1, Heuristech Press, Stanford, CA (1981).
4. C. L. Chang and J. R. Slagle, *Artificial Intelligence*, **2**, 117 (1971).
5. H. A. Simon and J. B. Kadane, *Artificial Intelligence*, **6**, 235 (1975).
6. G. Levi and F. Sirovich, *Artificial Intelligence*, **7**, 243 (1976).
7. M. M. Newborn, *Artificial Intelligence*, **8**, 137 (1977).
8. R. A. Finkel and J. R. Fishburn, *Artificial Intelligence*, **19**, 89 (1982).
9. M. P. Georgeff, *Artificial Intelligence*, **18**, 175 (1982).
10. H. A. Simon, *Artificial Intelligence*, **21**, 7 (1983).
11. M. S. Campbell and T. A. Marsland, *Artificial Intelligence*, **20**, 347 (1983).
12. R. M. Karp and J. Pearl, *Artificial Intelligence*, **21**, 99 (1983).
13. V. Kumar and L. Kanal, *Artificial Intelligence*, **21**, 179 (1983).
14. B. W. Ballard, *Artificial Intelligence*, **21**, 327 (1983).
15. R. E. Korf, *Artificial Intelligence*, **27**, 97 (1985).
16. T. Ibaraki, *Artificial Intelligence*, **29**, 73 (1986).
17. D. R. Bruce, *Artificial Intelligence*, **31**, 295 (1987).
18. J. A. Roels, *Biotech. Bioeng.*, **22**, 33 (1980).

19. R. L. Irvine and J. D. Bryers, *Compreh. Biotech.*, **41**, 757 (1985).
20. T. H. Pierce and B. A. Hohne (eds). *Artificial Intelligence Applications in Chemistry*. American Chemical Society, Washington D.C. (1986).
21. G. Stephanopoulos and D. W. Townsend, *Chem. Engng. Res. and Des.*, **64**, 160 (1986).
22. G. Stephanopoulos and G. Stephanopoulos, *Trends in Biotech.*, **4**, 241, (1986).
23. D. H. Rouvray, *Chem. Soc. Revs*, **3**, 355 (1974).
24. R. Fugmann, *Angew. Chem. Internat. Edit.*, **9** (1970).
25. M. Gordon and J. W. Kennedy, *J. Chem. Soc. Faraday Trans. II*, **69** (1973).
26. T. M. Gund, P. von R. Schleyer, P. H. Gund, and W. T. Wipke, *J. Amer. Chem. Soc.*, **97** (1975).
27. E. J. Corey and G. Petersson, *J. Amer. Chem. Soc.*, **94** (1972).
28. H. Gelrnter, N. S. Sridharan, A. J. Hart, and S. C. Yen. *Topics Curr. Chem.*, **41** (1973).
29. W. T. Wipke (editor), *Computer Representation and Manipulation of Chemical Information*, John Wiley and Sons, New York, NY (1974).
30. F. H. Westheimer, *Science*, **235**, 1173 (1987).
31. G. L. Steele Jr. *Common Lisp: The Language*, Digital Press, Billerica, MA (1984).
32. G. Gottschalk, *Bacterial Metabolism*, Springer-Verlag, New York, NY, 2nd edition (1986).
33. H. W. Doelle, *Bacterial Metabolism*, Academic Press, New York, NY, 2nd edition (1975).

34. A. L. Lehninger, *Biochemistry*, Worth Publishers, Inc., New York, NY, 2nd edition (1975).
35. A. Seressiotis and J. E. Bailey, *Biotech. Letters*, **8**, 837 (1986).
36. P. H. Winston and B. K. P. Horn, *LISP*, Addison-Wesley, Reading, MA, 2nd edition (1984).
37. L. Stryer, *Biochemistry*, W. H. Freeman and Co., San Francisco, CA, 2nd edition (1981).
38. E. Charniak and D. McDermont, *An Introduction to Artificial Intelligence*, Addison-Wesley, Reading, MA (1985).

SUBSTANCE	PATHWAY									
	1	2	3	4	5	6	7	8	9	10
GLUCOSE 6-P	-0.6	-0.5	-0.6	-0.6	-1.0	-1.0	-0.5	-0.5	-0.5	-0.5
CO <sub>2</sub>	0.6	0.0	0.6	0.6	3.0	3.0	0.0	0.0	0.0	0.0
ATP	0.8	0.5	-0.4	1.6	2.0	0.0	1.0	0.0	1.5	-0.5
NADH	0.6	0.5	0.0	1.0	1.0	0.0	0.5	0.0	1.0	0.0
NADPH	1.2	0.0	1.2	1.2	6.0	6.0	0.5	0.5	0.0	0.0
Ferrocyclochrome C	0.8	1.0	2.0	0.0	0.0	2.0	0.0	1.0	0.0	2.0

**TABLE 1.1.** The overall stoichiometry of the ten pathways converting glucose 6-phosphate to pyruvate with no carbon-carrying side products (except CO<sub>2</sub>) normalized on the basis of production of one mole of pyruvate. The stoichiometric coefficients for ADP, NAD<sup>+</sup>, NADP<sup>+</sup> and Ferricytochrome C have the same values but opposite signs as the coefficients of ATP, NADH, NADPH and Ferrocyclochrome C, respectively.

ENZYME	PATHWAY									
	1	2	3	4	5	6	7	8	9	10
P-GLUCONATE DEHYDROGENASE	✓		✓	✓	✓	✓				
RIBOSE-P ISOMERASE	✓		✓	✓	✓	✓				
RIBULOSE-P 3-EPIMERASE	✓		✓	✓	✓	✓				
TRANSKETOLASE-1	✓		✓	✓	✓	✓				
TRANSALDOLASE	✓		✓	✓	✓	✓				
TRANSKETOLASE-2	✓		✓	✓	✓	✓				
GLUCOSE-6-P DEHYDROGENASE	✓		✓	✓	✓	✓	✓	✓		
GLUCONOLACTONASE	✓		✓	✓	✓	✓	✓	✓		
PHOSPHOGLUCONATE DEHYDRATASE							✓	✓		
P-2-KETO-3-DEOXY-GLUCONATE ALDOLASE							✓	✓		
GLYCERALDEHYDE-P DEHYDROGENASE	✓	✓		✓	✓		✓		✓	
P-GLYCERATE KINASE	✓	✓		✓	✓		✓		✓	
P-GLYCEROMUTASE	✓	✓		✓	✓		✓		✓	
ENOLASE	✓	✓		✓	✓		✓		✓	
PYRUVATE KINASE	✓	✓		✓	✓		✓		✓	
GLUCOSE-P ISOMERASE		✓			✓	✓			✓	✓
6-P-FRUCTOKINASE	✓	✓	✓	✓					✓	✓
FRUCTOSE-BI-P ALDOLASE	✓	✓	✓	✓					✓	✓
TRIOSE-P ISOMERASE			✓	✓		✓		✓	✓	✓
METHYLGLYOXAL SYNTHASE	✓	✓	✓			✓		✓		✓
GLYOXYLASE I, II	✓	✓	✓			✓		✓		✓
LACTATE DEHYDROGENASE (CYTOCHROME)	✓	✓	✓			✓		✓		✓

TABLE 1.2. Enzyme participation in the ten pathways converting glucose 6-phosphate to pyruvate with no carbon-carrying side products (except CO<sub>2</sub>).

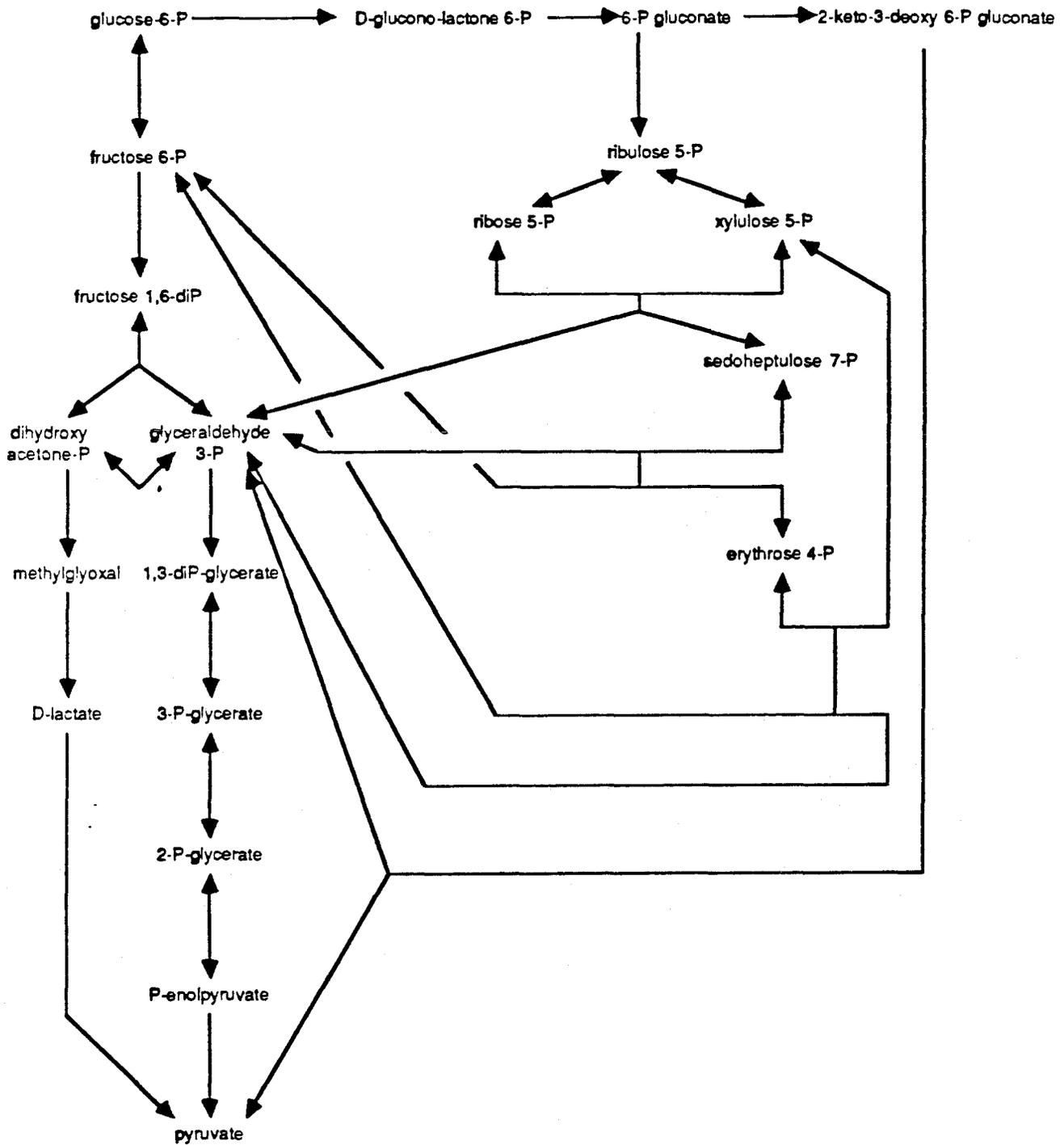


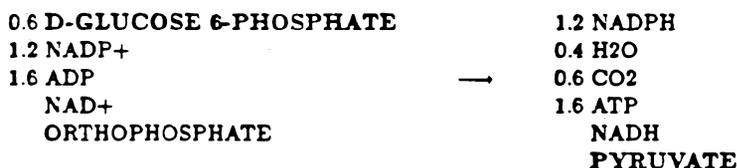
FIGURE 1.1: The part of the metabolic chart related to glucose 6-phosphate catabolism.

#### PATHWAY 4

The participating enzymes are:

1. GLUCOSE-6-PHOSPHATE DEHYDROGENASE	(v = 0.6)
2. GLUCONOLACTONASE	(v = 0.6)
3. PHOSPHOGLUCONATE DEHYDROGENASE	(v = 0.6)
4. RIBOSEPHOSPHATE ISOMERASE	(v = 0.2)
5. RIBULOSEPHOSPHATE 3-EPIMERASE	(v = 0.4)
6. TRANSKETOLASE-1	(v = -0.2)
7. TRANSALDOLASE	(v = 0.2)
8. TRANSKETOLASE-2	(v = 0.2)
9. 6-PHOSPHOFRUCTOKINASE	(v = 0.4)
10. FRUCTOSE-BIPHOSPHATE ALDOLASE	(v = 0.4)
11. TRIOSEPHOSPHATE ISOMERASE	(v = 0.4)
12. GLYCERALDEHYDE-PHOSPHATE DEHYDROGENASE	(v = 1.0)
13. PHOSPHOGLYCERATE KINASE	(v = 1.0)
14. PHOSPHOGLYCEROMUTASE	(v = 1.0)
15. ENOLASE	(v = 1.0)
16. PYRUVATE KINASE	(v = 1.0)

The overall reaction is:



#### PATHWAY 9

The participating enzymes are:

1. GLUCOSEPHOSPHATE ISOMERASE	(v = 0.5)
2. 6-PHOSPHOFRUCTOKINASE	(v = 0.5)
3. FRUCTOSE-BIPHOSPHATE ALDOLASE	(v = 0.5)
4. TRIOSEPHOSPHATE ISOMERASE	(v = 0.5)
5. GLYCERALDEHYDE-PHOSPHATE DEHYDROGENASE	(v = 1.0)
6. PHOSPHOGLYCERATE KINASE	(v = 1.0)
7. PHOSPHOGLYCEROMUTASE	(v = 1.0)
8. ENOLASE	(v = 1.0)
9. PYRUVATE KINASE	(v = 1.0)

The overall reaction is:



FIGURE 1.2: Two of the pathways produced by MPS for the conversion of glucose 6-phosphate to pyruvate. Pathway 4 has been identified as the pentose phosphate pathway and pathway 9 as the Embden-Meyerhof-Parnas pathway.

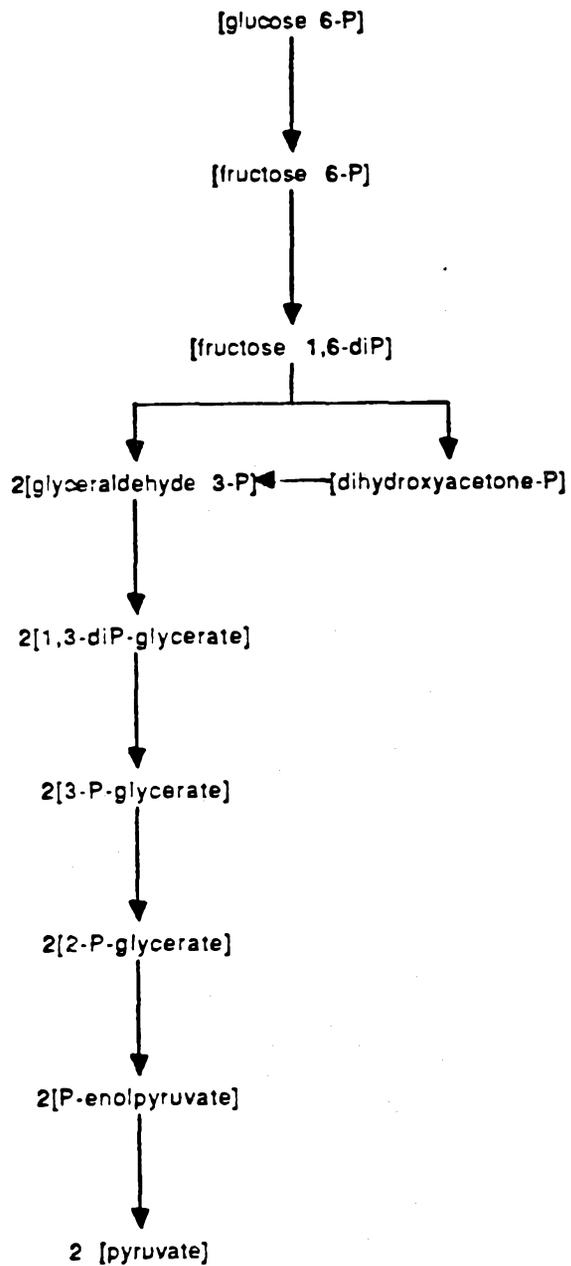


FIGURE 1.3: Schematic representation of the Embden-Meyerhof-Parnas pathway.

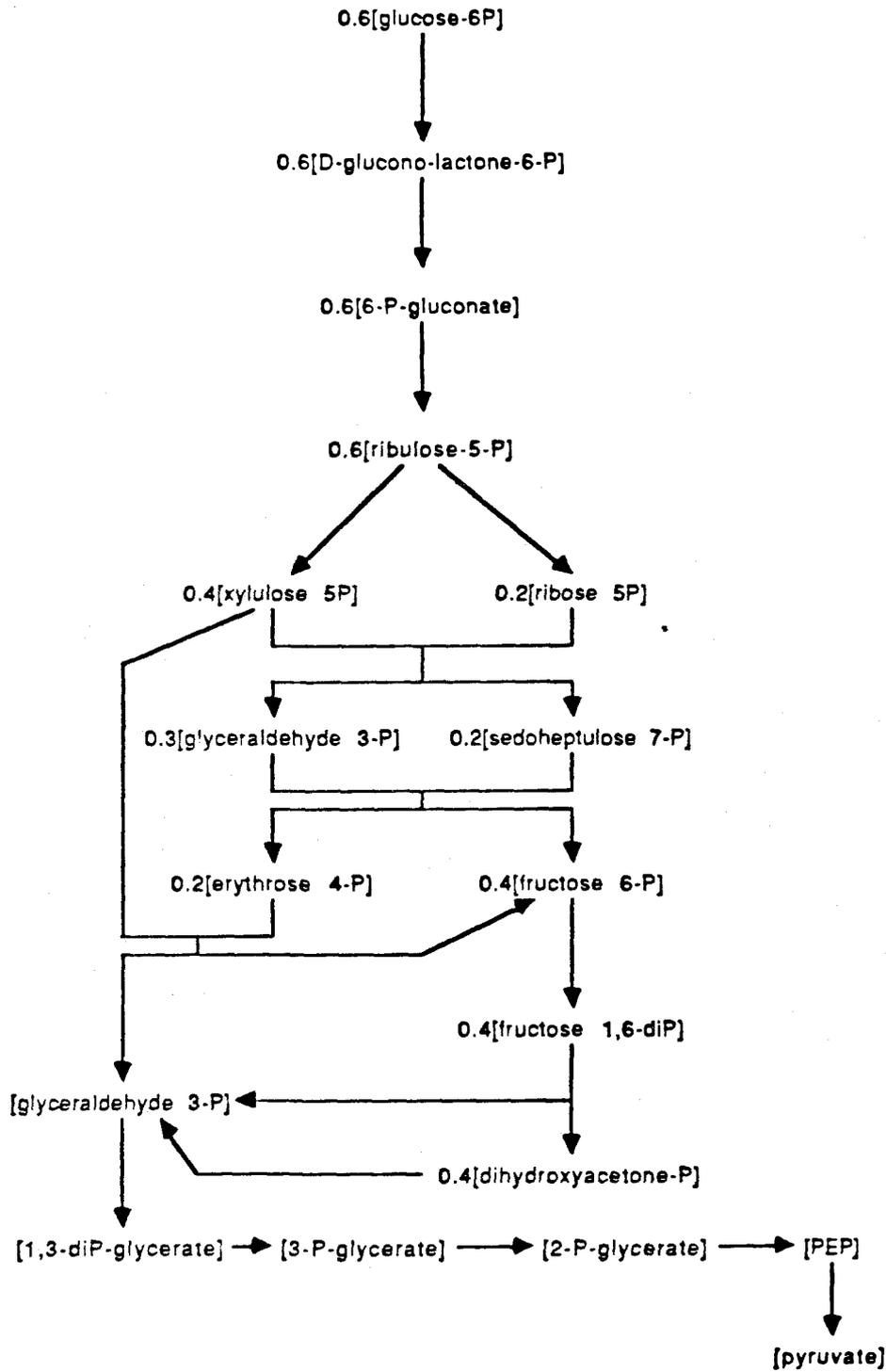


FIGURE 1.4: Schematic representation of the pentose phosphate pathway.

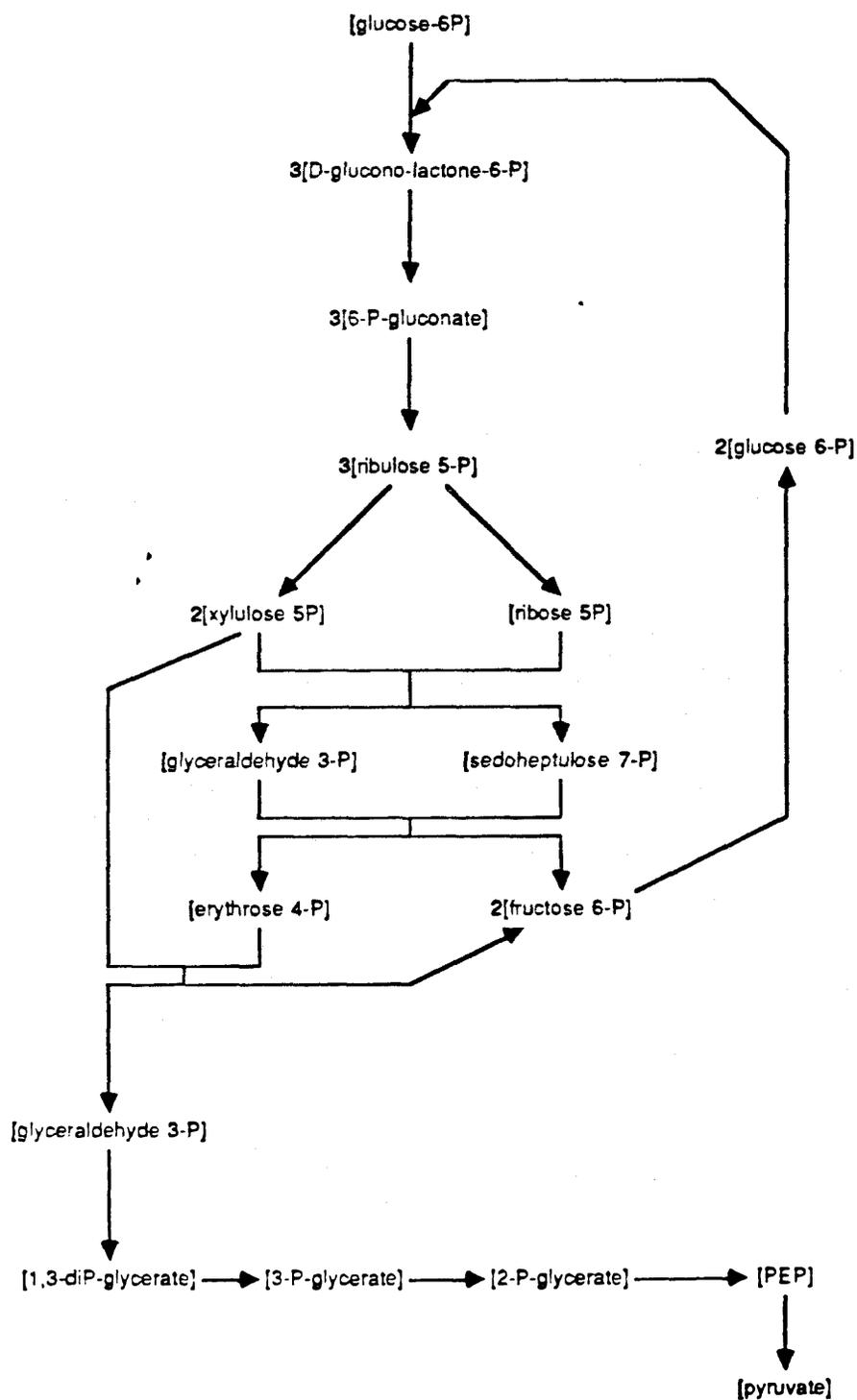


FIGURE 1.5: Schematic representation of a variation of the pentose phosphate pathway.

The participating enzymes are:

1. GLUCOSE-6-PHOSPHATE DEHYDROGENASE	(v = 1.0)
2. GLUCONOLACTONASE	(v = 1.0)
3. PHOSPHOGLUCONATE DEHYDROGENASE	(v = 1.0)
4. RIBOSEPHOSPHATE ISOMERASE	(v = 0.5)
5. RIBULOSEPHOSPHATE 3-EPIMERASE	(v = 0.5)
6. TRANSKETOLASE-1	(v = -0.5)
7. TRANSALDOLASE	(v = 0.5)
8. 6-PHOSPHOFRUCTOKINASE	(v = 0.5)
9. FRUCTOSE-BIPHOSPHATE ALDOLASE	(v = 0.5)
10. TRIOSEPHOSPHATE ISOMERASE	(v = 0.5)
11. GLYCERALDEHYDE-PHOSPHATE DEHYDROGENASE	(v = 1.0)
12. PHOSPHOGLYCERATE KINASE	(v = 1.0)
13. PHOSPHOGLYCEROMUTASE	(v = 1.0)
14. ENOLASE	(v = 1.0)
15. PYRUVATE KINASE	(v = 1.0)

The overall reaction is:

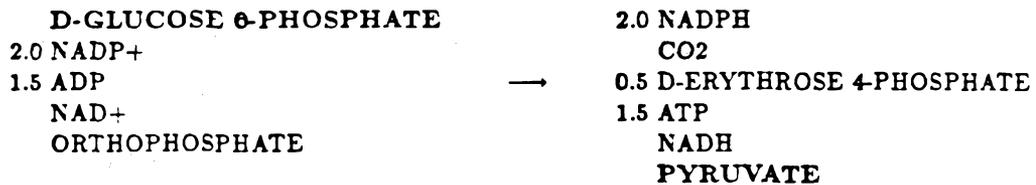
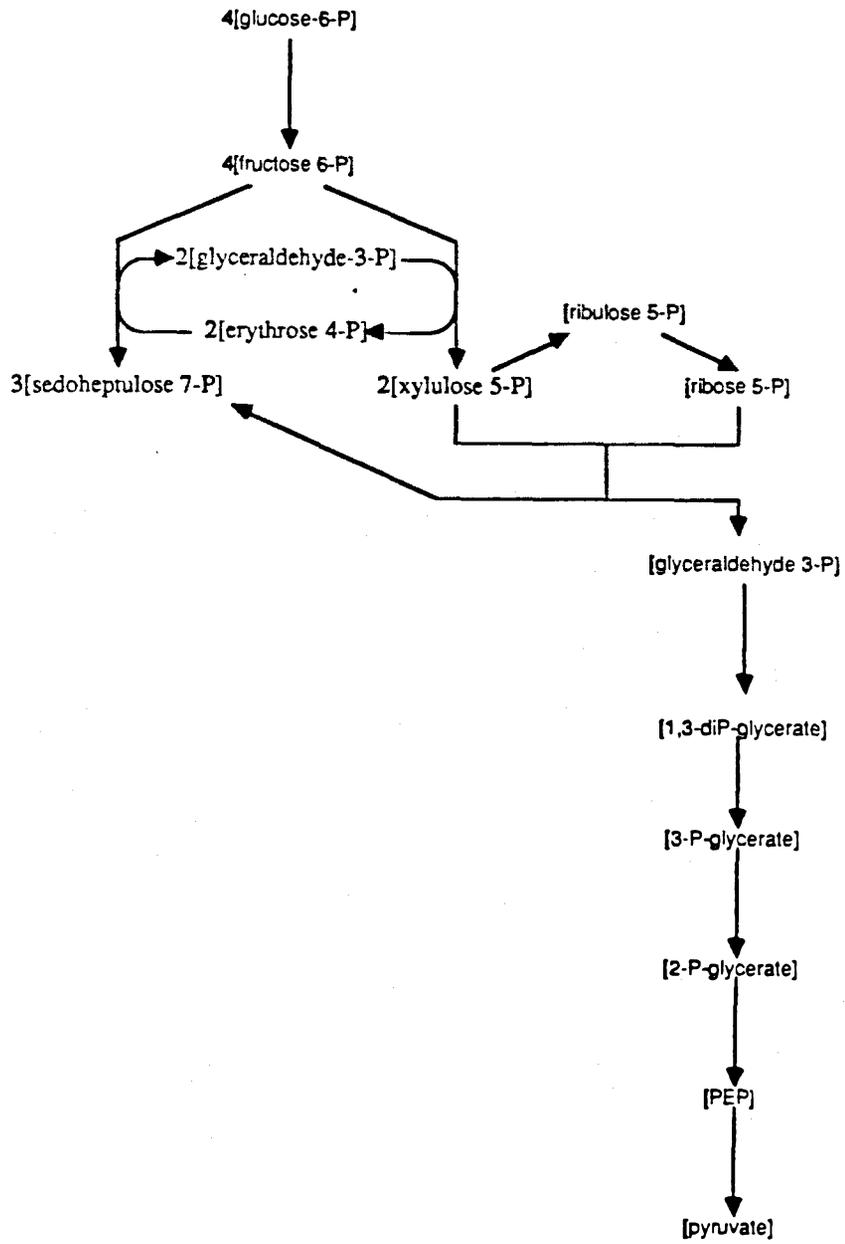
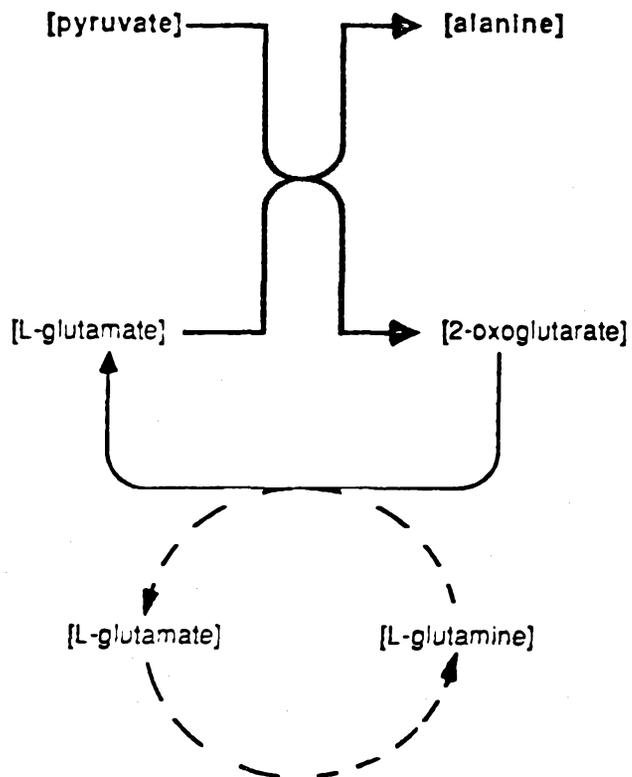


FIGURE 1.6: A variation of the pentose phosphate pathway that leads to the concurrent production of pyruvate and erythrose 4-phosphate.



**FIGURE 1.7:** The schematic representation of a pathway in which substances assume a pseudocatalytic role.



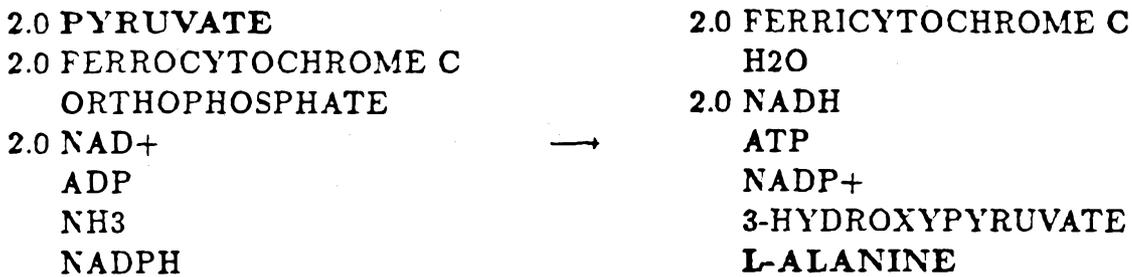
**FIGURE 1.8:** The schematic representation of two pathways leading from pyruvate to L-alanine.

### PATHWAY 2

The participating enzymes are:

- |   |            |
|---|------------|
| 1. LACTATE DEHYDROGENASE (CYTOCHROME)     | (v = -1.0) |
| 2. GLYOXYLASE I, II                       | (v = -1.0) |
| 3. METHYLGLYOXAL SYNTHASE                 | (v = -1.0) |
| 4. TRIOSEPHOSPHATE ISOMERASE              | (v = 1.0)  |
| 5. GLYCERALDEHYDE-PHOSPHATE DEHYDROGENASE | (v = 1.0)  |
| 6. PHOSPHOGLYCERATE KINASE                | (v = 1.0)  |
| 7. PHOSPHOGLYCERATE DEHYDROGENASE         | (v = 1.0)  |
| 8. L-GLUTAMATE DEHYDROGENASE (NADP+)      | (v = -1.0) |
| 9. PHOSPHOSERINE TRANSAMINASE             | (v = 1.0)  |
| 10. PHOSPHOSERINE PHOSPHATASE             | (v = 1.0)  |
| 11. SERINE-PYRUVATE AMINOTRANSFERASE      | (v = 1.0)  |

The overall reaction is:



**FIGURE 1.9:** One of the four pathways that MPS synthesized for the conversion of pyruvate to L-alanine.

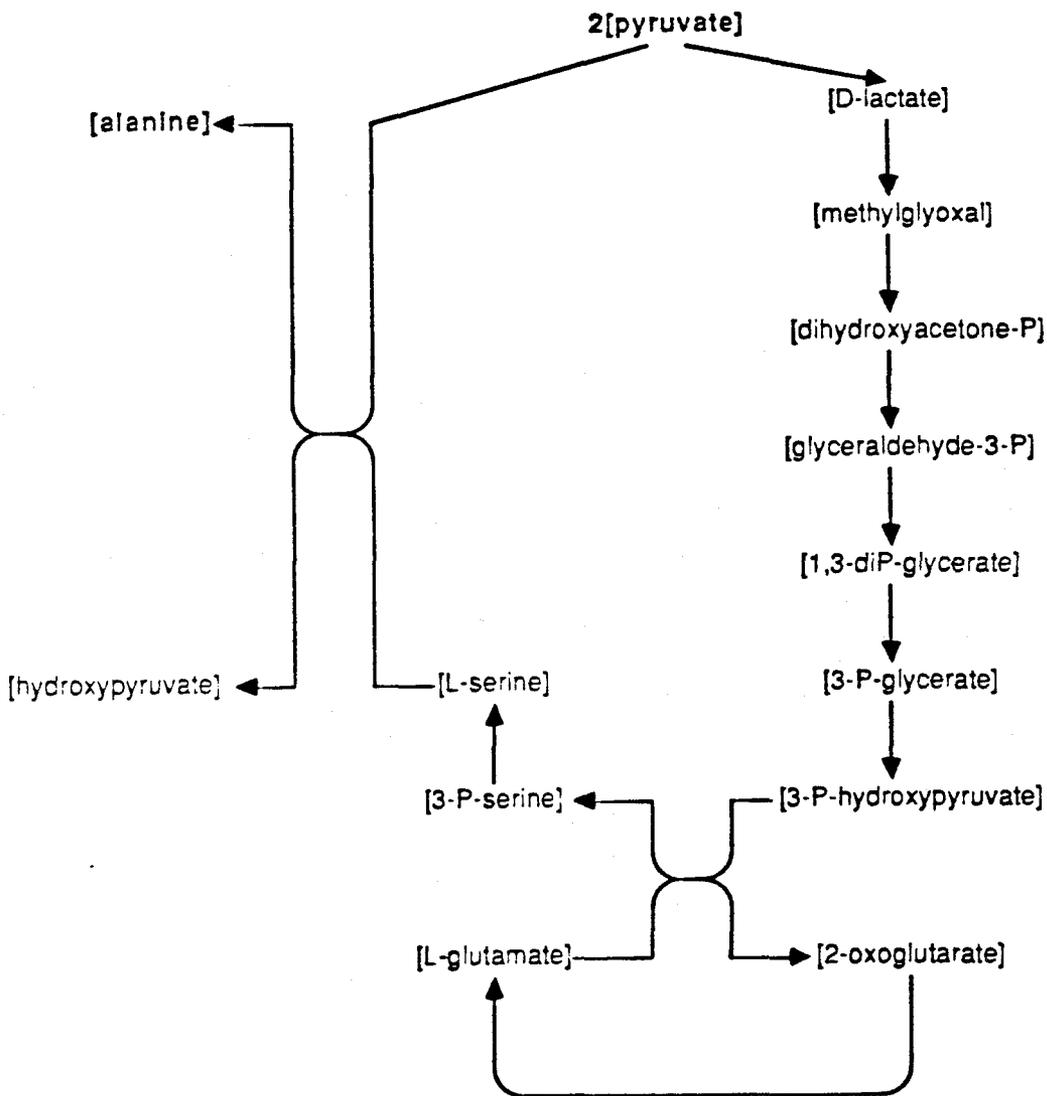


FIGURE 1.10: The schematic representation of the pathway show in Figure 9.

## **CHAPTER 2**

### **A KINETIC MODEL FOR PRODUCT FORMATION IN UNSTABLE RECOMBINANT POPULATIONS**

## INTRODUCTION

Recombinant DNA technology is now in transition from laboratory scale investigation to production scale application. Several proteins such as insulin and interferon are now being produced on significant scale.<sup>1</sup> The stability of the plasmid or other recombinant DNA vector may be one of the most important factors in large-scale production using genetically engineered cells.

A number of previous investigators have proposed and analyzed mathematical models describing genetic instability in growing cultures. General models for genetic stability and its effects during continuous cultivation of microorganisms have been developed in detail by Moser.<sup>2</sup> More recently, a number of investigators have calculated the effect on overall cell population properties of plasmid loss at cell division and of differences in growth rates of plasmid-free and plasmid-containing cells.<sup>3-8</sup>

Also important for design and analysis of reactors for recombinant organisms are descriptions of cloned-gene product formation. Previous analyses by Ollis and Chang<sup>8</sup> considered product formation in recombinant cultures using the empirical Leudeking-Piret equation for product formation kinetics. Seo and Bailey presented overall cloned-gene product formation calculations based on a segregated model and empirical single-cell product synthesis kinetics.<sup>9</sup> Recently, models for product synthesis in recombinant systems have been formulated based upon the molecular mechanisms involved.<sup>10</sup> In this work, the previously developed genetically structured model for product formation in recombinant cells is extended to describe the kinetics of cell growth and product formation in fermentation of unstable recombinant cells in batch and continuous reactors.

One purpose of this work is to develop mathematical expressions for calculating product formation which are formulated at the single-cell level and elaborated to allow calculations at the reactor level. Second, the models will be applied to determine reactor performance and to examine the sensitivity of reactor performance to certain process and

host-vector design parameters. The simulation results indicate that there may be an optimum copy number or an optimum transcription efficiency of the cloned gene for maximizing reactor productivity. Also, the apparent stability of the genetically engineered cell is determined not only by the intrinsic stability parameters associated with plasmid segregation and mutation but also by the product synthesis activity of the cell and by reactor operating conditions such as dilution rate in a continuous fermentor.

## BACKGROUND

### Stability of Recombinant DNA

Schematic representation of a recombinant DNA expression plasmid is shown in Fig. 2.1A. The essential components of recombinant DNA are a (foreign) gene fragment of interest (G in Fig. 2.1A) and a cloning vector which serves to propagate that DNA segment in the host cell. Plasmids, phages, or their derivatives are widely used for cloning vectors. For stable maintenance of recombinant DNA in proliferating host cell populations, the vector should carry a minimal replicon (R in Fig. 2.1A), that is, the genes and genetic sites required for replication and possibly also for segregation. In addition, selection markers (M in Fig. 2.1A) providing antibiotic resistance or complementation of host cell mutations are usually introduced into a cloning vector because the replicon itself often has no detectable phenotypic effect on the host cells.

During propagation of recombinant DNA in microbial cells, however, a loss of desired phenotype is often observed. This phenomenon can occur through two processes called "segregational instability" and "structural instability" (Fig. 2.1B).<sup>11</sup> Segregation instability refers to the loss of the entire recombinant DNA from the cell due to irregular segregation of the cloning vector at cell division.<sup>12,13</sup> On the other hand, a recombinant DNA is frequently subjected to various types of structural rearrangements such as deletion, insertion, and duplication, causing loss of expression of an active product of one or more genes of the recombinant DNA.<sup>14-18</sup>

Growth of a recombinant organism in selective medium aids in maintaining a population of cells containing the vector which still possesses the selection marker phenotype. Thus, application of selection pressure can be used to reduce the damage of segregational instability in a process for manufacturing a cloned-gene product. Since, however, the gene product of the selection marker is usually not the product of commercial interest and since there is typically no method for direct selection of the desired product phenotype, selection pressure does not eliminate or control structural instability. The cell population may maintain the recombinant DNA but lose the ability to express a desired gene product in active form.

### Effect of Cloned Gene Expression on Host Cell Growth

Since a recombinant DNA is composed of a cloning vector and a cloned DNA segment and since these two elements are not necessarily propagated together in the heterogeneous populations found in unstable recombinant cultures, the effect of the cloning vector and the effect of cloned gene expression on the host cell growth rate are considered here separately. In the models examined in this work, it is assumed also that functional dependence of specific growth rate on limiting nutrient concentration  $s$ , cloning vector concentration  $\hat{G}$ , and cloned gene product concentration  $\hat{p}$  is separable. Thus

$$\mu = \mu^0 f(s) g(\hat{G}) h(\hat{p}) . \quad (2.1)$$

It has been reported that the presence of certain plasmids has a deleterious effect on host cell growth.<sup>19-21</sup> Since common precursors, catalytic assemblies, and energy supplies provided by the host cell are required for biosynthetic processes of both plasmid-encoded genes and chromosomal genes, competition between the cellular activities directed by the plasmid and the bacterial chromosome may occur. Supporting this hypothesis, Nordström *et al.*<sup>22</sup> and Seo and Bailey<sup>23</sup> have reported that the growth rates of different strains containing a series of plasmid copy mutants were reduced

significantly in cases where the plasmid copy number was relatively large. Previous simulation results based on an intracellular model including competition for RNA polymerase predict that transcription efficiency of both plasmid and chromosomal genes is greatly reduced as the plasmid copy number increases.<sup>10</sup> Based on the previous model simulation and the experimental data mentioned, the following relationship between the cellular content of plasmid gene copies and host-cell growth rate is a reasonable functional form for certain recombinant systems:

$$g(\hat{G}) = \left[ 1 - \frac{\hat{G}}{\hat{G}_{\max}} \right]^m . \quad (2.2)$$

The effect of plasmid copy number on the host-cell metabolic activity depends largely on the nature of plasmid replicon, the particular genes and control loci encoded in the plasmid, and the genetic background of the host cell. As a consequence, the two parameters appearing in Eq. (2.2),  $\hat{G}_{\max}$  and  $m$ , can be affected by these genetic factors as well as by environmental conditions.

Although many genes have been successfully cloned and expressed in recombinant microorganisms, accumulated data for several of these host-vector expression systems shows that expression of the amplified, cloned gene inhibits cell growth and survival. Such inhibitory effects appear even more pronounced when the expression product of the cloned gene is a highly hydrophobic protein such as a membrane protein, a hormone, or an interferon.<sup>24-26</sup>

Practical motivations for genetic engineering include production of valuable peptides or proteins and desirable alteration or enhancement of host cell metabolic activities or functions. In both cases, it is important to consider the locus of accumulation of the product. Synthesis of cloned gene products in the most widely used host, *E. coli*, usually results in intracellular accumulation of the cloned gene product. Limited successes in secretion of cloned gene products from *E. coli* into the medium have been reported.<sup>27,28</sup>

In an ideal expression-secretion system, almost all of the cloned gene product will be found in the cell environment. Similarly, it is expected that most product metabolites will be found outside the cell in useful strains with metabolic enhancements provided by genetic engineering. However, even in this instance and for secreted proteins (and indeed for many other microbiological processes), the product concentration inside the cell will likely exceed the product concentration in the medium.

Two different models for product inhibition will be considered. In the first, the inhibitory effect is presumed to depend on the intracellular product concentration  $\hat{p}$  according to

$$h(\hat{p}) = \left[ 1 - \frac{\hat{p}}{\hat{P}_{\max}} \right]^n \quad \text{Model I} \quad (2.3)$$

The second model is based on the total broth product concentration  $p$ :

$$h(p) = \left[ 1 - \frac{p}{P_{\max}} \right]^{n'} \quad \text{Model II} \quad (2.4)$$

Model II is of the form applied previously to several fermentations,<sup>29</sup> but Model I seems intuitively more reasonable in many cases and certainly is the required form when all of the product accumulates intracellularly. One objective of this modeling study, then, is to explore the sensitivity of calculated bioreactor performance to different types of product inhibition.

Using Eqs. (2.2) and (2.3) or (2.4) to calculate the influences of plasmid and cloned-gene product content on cell growth implies (see Eq. (2.1)) that cells without plasmids and product grow more rapidly than plasmid-containing and productive cells. Thus, this model corresponds to use of non-selective medium. Extension of this overall productivity model to selective medium requires careful consideration of growth rates at small plasmid content and growth of plasmid-free cells. Experiments and model analyses of

these questions have only recently begun.<sup>30</sup>

### Product Formation

The empirical formula suggested by Leudeking and Piret<sup>31</sup> has been widely used in previous descriptions of product formation kinetics. In the case of a recombinant DNA system, however, it is now possible to describe product formation based on the corresponding molecular mechanisms. Mathematical expressions for the synthesis of messenger RNA and protein via transcription and translation processes have been presented previously (e.g., Ref. 10). Assuming that the decay of cloned gene mRNA and of the corresponding protein follow first-order kinetics, the *intracellular* concentration of cloned-gene message ( $\hat{m}$ ) and protein ( $\hat{p}$ ) can be described by the following material balances:

$$\frac{d\hat{m}}{dt} = k_p^o \eta \hat{G} - k_d \hat{m} - \mu \hat{m} \quad (2.5)$$

$$\frac{d\hat{p}}{dt} = k_q^o \xi \hat{m} - k_e \hat{p} - \mu \hat{p} \quad (2.6)$$

where  $k_p^o$ ,  $k_q^o$ ,  $k_d$ , and  $k_e$  are the overall transcription rate constant, overall translation rate constant, decay constant of mRNA, and decay constant of protein, respectively. The symbols  $\eta$ ,  $\xi$ , and  $\mu$  denote the transcription efficiency, translation efficiency, and specific growth rate, respectively. As discussed in a previous paper,  $k_p^o$  and  $k_q^o$  are functions of growth rate<sup>10</sup>;  $k_d$  and  $k_e$  are presumed relatively independent of growth rate.

Since  $(k_d + \mu) \gg (k_e + \mu)$ , the quasi steady-state approximation<sup>32</sup> can be applied to  $\hat{m}$ ; i.e.,  $\hat{m}$  may be approximated by the equation implied by setting  $d\hat{m}/dt = 0$ :

$$\hat{m} \equiv \frac{k_p^o \eta}{k_d + \mu} \hat{G} \quad (2.7)$$

Substituting Eq. (2.7) in Eq. (2.6) gives

$$\frac{d\hat{p}}{dt} = f(\mu)\gamma\hat{G} - k_e\hat{p} - \mu\hat{p} \quad , \quad (2.8)$$

where

$$f(\mu) = \frac{k_p^o k_q^o}{k_d + \mu} \quad , \quad (2.9.1)$$

and the "gene expression parameter"  $\gamma$  is defined by

$$\gamma = \eta\xi \quad . \quad (2.9.2)$$

Notice that  $\gamma$  depends on the cloned-gene promoter and ribosome binding site "strengths" as indicated in the parameters  $\eta$  and  $\xi$ , respectively. This quantity can be altered by changing the plasmid construction or, if suitable controls are incorporated in the plasmid, by changing medium composition or temperature.

Dependencies of the parameters  $k_p^o$  and  $k_q^o$  on specific growth rate for (plasmid-free) *E. coli* have been estimated previously based on literature data.<sup>10</sup> Unfortunately, little data are presently available for estimating these kinetic parameters in the presence of plasmids. Presumably both depend on plasmid content (see, for example, Ref. 33). In the absence of more information, the correlations previously described are used here for plasmid-containing cells.

In order to explore relationships between these kinetics and other overall forms used in previous studies, an approximate form of product synthesis kinetics will also be considered. In this model, the function  $f(\mu)$ , which is a complicated nonlinear function of specific growth rate  $\mu$ , will be approximated by a linear function

$$f(\mu) \equiv k(\mu + b) \quad (2.10)$$

where  $k$  and  $b$  are constants.

## KINETIC MODEL

The kinetic model developed here is based on the following assumptions:

1. The number of plasmids per  $R^+$  cell is constant during cultivation of recombinant cells (actually  $\hat{G} = f(\mu)^{10,23}$ , and an appropriate relationship can be determined from a plasmid replication model<sup>10,34</sup>).
2. The mutation frequency of the cloned gene and the frequency of plasmid loss due to irregular segregation are assumed to be constant irrespective of host cell growth rate and culture conditions.
3. The cell population is presumed to consist of three different genetic classes according to the presence or absence of plasmids and according to expression of active cloned-gene product (see Fig. 2.1B). Another possible combination [ $G^+R^-$ ] is not considered here based on the experimental observations of Imanaka *et al.*<sup>17</sup>

### Batch Culture

#### (a) Cell Growth

Let  $\theta$  and  $\phi$  denote the ratio of the mutation rate of structural genes to the host cell growth rate and the ratio of segregational plasmid loss rate to the growth rate of the host cell, respectively. Then the following equations for densities of strains with genotypes 1 through 3 ( $G^+R^+$ ,  $G^+R^-$ , and  $G^-R^-$ , respectively) can be derived from the scheme of genetic changes outlined in Fig. 2.1B:

$$\frac{dx_1}{dt} = \mu_1 x_1 (1 - \theta - \phi) \triangleq \bar{\mu}_1 x_1 \quad (2.11)$$

$$\frac{dx_2}{dt} = \mu_2 x_2 (1 - \phi) + \mu_1 x_1 \theta \quad (2.12)$$

and

$$\frac{dx_3}{dt} = \mu_3 x_3 + \mu_1 x_1 \phi + \mu_2 x_2 \phi \quad (2.13)$$

In Eq. (2.11),  $\bar{\mu}_1$  denotes the apparent (or overall) specific growth rate of the product-synthesizing subpopulation. Growth in a well-mixed, isothermal batch reactor is assumed in writing these cell balances.

The specific growth rate of each species,  $\mu_j$ , will be written:

$$\mu_1 = \mu^o \left[ 1 - \frac{\hat{G}}{\hat{G}_{\max}} \right]^m \left[ 1 - \frac{\hat{p}}{\hat{p}_{\max}} \right]^n \left[ \frac{s}{K_s + s} \right] = \mu^o f(s) g(\hat{G}) h(\hat{p}) \quad (2.14)$$

$$\mu_2 = \mu^o \left[ 1 - \frac{\hat{G}}{\hat{G}_{\max}} \right]^m \left[ \frac{s}{K_s + s} \right] = \mu^o f(s) g(\hat{G}) \quad (2.15)$$

and

$$\mu_3 = \mu^o \left[ \frac{s}{K_s + s} \right] = \mu^o f(s) \quad (2.16)$$

for Model I while  $\mu_1 = \mu^o f(s) g(\hat{G}) h(\hat{p})$  for Model II (cf. Eqs. (2.3),(2.4)). Here, for the sake of simplicity, the Monod constants,  $K_{s_j}$ , of each species ( $j = 1, 2$ , and  $3$ ) for limiting substrate,  $s$ , are assumed to be identical. The maximum specific growth rate of each species has been written to include inhibition by the vector or the cloned gene product depending on the presence of each of these factors in each of the three strains. This is a key feature of this model.

#### (b) Substrate Consumption

Presuming constant stoichiometric ratios (yield factors), the rate of substrate consumption can be obtained from a mass balance equation of the form

$$\frac{ds}{dt} = - \sum \frac{1}{Y_{sj}} \frac{dx_j}{dt} - \frac{1}{Y_p} \frac{dp}{dt} \quad (2.17)$$

For convenience, the second term is considered to be negligible in this paper. Assuming  $Y_{sj}$  is not significantly different for each cell type, Eq. (2.17) becomes

$$\frac{ds}{dt} = - \frac{1}{Y_s} \left[ \frac{dx_1}{dt} + \frac{dx_2}{dt} + \frac{dx_3}{dt} \right] \quad (2.18)$$

### (c) Product Formation Kinetics

As shown in the previous section, product formation, which occurs only in Species 1, is calculated using either Eqs. (2.8) and (2.9) or the approximate model corresponding to Eq. (2.10). In either case, the overall specific growth rate of species 1,  $\bar{\mu}_1$ , is used to calculate the growth-rate dependence of product synthesis.

Since  $\hat{p}$  is the intracellular product concentration, it is important to convert this value to product concentration in the culture volume in order to assess reactor productivity. The required relationship is

$$p = \hat{p}x_1/\rho_b \quad (2.19)$$

where  $\rho_b$  is the density of the bacterial cell (cell mass/unit volume cell). Differentiating this gives

$$\frac{dp}{dt} = \frac{1}{\rho_b} \frac{d(x_1\hat{p})}{dt} = \frac{1}{\rho_b} \left[ x_1 \frac{d\hat{p}}{dt} + \hat{p}\bar{\mu}_1x_1 \right] , \quad (2.20)$$

assuming  $\rho_b$  is independent of time. This is a reasonable assumption for plasmid-free bacteria<sup>35</sup>; little data are presently available on the density of recombinant strains and its

variability.

It is interesting to explore the connection between these product formation kinetics based on mechanism and those of traditional Leudeking-Piret form. Substituting Eqs. (2.8) and (2.10) into Eq. (2.20) gives

$$\frac{dp}{dt} = \frac{1}{\rho_b} [k\eta\xi\hat{G}(\bar{\mu}_1 + b)x_1 - k_e x_1 \hat{p}] = \frac{1}{\rho_b} \left[ k\eta\xi\hat{G} \left[ \frac{dx_1}{dt} + bx_1 \right] \right] - k_e p. \quad (2.21)$$

Eq. (2.21) is nearly identical to the empirical equation of Leudeking and Piret.<sup>31</sup> Considering only the kinetics of protein synthesis and accordingly neglecting the protein decay term, Eq. (2.21) becomes

$$\left. \frac{dp}{dt} \right|_{\text{synthesis}} = \alpha \frac{dx_1}{dt} + \beta x_1 \quad (2.22)$$

where

$$\alpha = k\eta\xi\hat{G}/\rho_b, \quad \beta = \alpha \cdot b \quad (2.23)$$

Here, the structured model provides an interpretation for the (typically empirical) Leudeking-Piret parameters.

### Continuous Culture

Material balances for cell concentration, substrate concentration, and product formation for continuous cultures with sterile feed may be developed in a manner analogous to that described above for batch cultures, with the results

$$\frac{dx_1}{dt} = -Dx_1 + \mu_1 x_1 (1 - \theta - \phi) \quad (2.24)$$

$$\frac{dx_2}{dt} = -Dx_2 + \mu_2x_2(1 - \phi) + \mu_1x_1\theta \quad (2.25)$$

$$\frac{dx_3}{dt} = -Dx_3 + \mu_3x_3 + \mu_1x_1\phi + \mu_2x_2\phi \quad (2.26)$$

$$\frac{ds}{dt} = D(s_0 - s) - \frac{1}{Y_s} (\mu_1x_1 + \mu_2x_2 + \mu_3x_3) \quad (2.27)$$

$$\frac{dp}{dt} = -Dp + \frac{f(\bar{\mu}_1)}{\rho_b} \eta \xi \hat{G}x_1 - k_e p \quad (2.28)$$

where  $D$  and  $s_0$  represent the dilution rate and inlet substrate concentration, respectively.

## SIMULATION RESULTS

### Model Parameters

The parameters used for model simulations are listed in Table 2.1. The maximum specific growth rate,  $\mu^0$ , the Monod constant,  $K_s$ , and the yield coefficient,  $Y_s$ , were taken from values for *E. coli* cells growing in glucose medium in the absence of plasmids or cloned gene expression.<sup>36</sup> Although experimental data are limited, it has been reported that the  $K_s$  and  $Y_s$  values of plasmid-containing cells are not significantly different from those of plasmid-free cells while  $\mu^0$  is significantly affected due to the presence of plasmids.<sup>20,21,23,37</sup> Accordingly, as noted earlier, it is assumed here that the  $K_s$  and  $Y_s$  values for each cell type are identical.

The exponent  $m$  was set equal to 0.5 based on the experimental data of Nordström *et al.*<sup>22</sup> In the absence of available information on the exponent for product inhibition,  $n$  was assumed to be either zero (no product inhibition) or 1. Although the values of  $\hat{G}_{max}$  and  $\hat{p}_{max}$  ( $p_{max}$ ) may be highly dependent upon the nature of the cloning vector and of the cloned gene product as well as the genetic background of the host cell, the parameters listed in Table 2.1 are reasonable based on maximal product levels reported to date and were used for these simulations. Since the purpose of these simulations is to

investigate qualitative trends influencing the productivity of recombinant microorganisms, several parameters such as  $\beta$  and  $k_e$ , which do not significantly affect the conclusions of model simulation, were assumed to be negligible.

Evaluation of  $f(\mu)$  in Eq. (2.9.1) using  $k_p^o(\mu)$  and  $k_q^o(\mu)$  from Lee and Bailey<sup>10</sup> gives

$$g(\mu) \triangleq \frac{k_p^o k_q^o}{k_d + \mu} = \frac{4.5 \times 10^{10} \mu^4}{(78\mu^2 + 233)(145\mu + 82.5)(27.6 + \mu)} \quad (\text{hr}^{-1}) \quad (2.29)$$

where  $\mu$  has units of  $\text{hr}^{-1}$ . Two different linear approximations to  $f(\mu)$ , with form given by Eq. (2.10), are considered. The first, chosen to coincide with  $g$  at  $\mu = 0$  and  $\mu = \mu^o$ , is

$$h(\mu) = 10,900\mu \quad (\text{hr}^{-1}), \quad (2.30)$$

and the second, chosen somewhat arbitrarily to represent an intermediate case better approximating growth rates for plasmid-containing cells, intersects  $g$  at a  $\mu$  value corresponding to 50 plasmids and no product, giving

$$l(\mu) = 7300\mu \quad (\text{hr}^{-1}). \quad (2.31)$$

These three different protein synthesis rate functions are plotted versus specific growth rate  $\mu$  in Fig. 2.2.

The plasmid concentration  $\hat{G}(\text{g/l})$  is calculated from the number of plasmids per cell  $N_p$  using the relationship

$$\hat{G} = 0.0001186 N_p \quad (2.32)$$

which is based upon representative plasmid size and cell volume for *E. coli*.<sup>10</sup>

## Batch Cultures

In Fig. 2.3 computer simulation results for batch cultivation of recombinant cells are shown using the model described above with  $f(\mu)$  given by the complete Eq. (2.9) and

with no product inhibition ( $n=0$ ). The changes in  $x_1$ ,  $x_3$ ,  $p$ , and  $s$  with time during fermentation were obtained from the numerical solutions of Eqs. (2.11)-(2.13), (2.18), and (2.21). The initial conditions used in these calculations are summarized below:

$$\begin{aligned} s(0) &= 30 \text{ g/l} \\ x_1(0) &= 0.001 \text{ g/l} \\ x_2(0) &= x_3(0) = p(0) = 0 \end{aligned} \quad (2.33)$$

In these calculations, the values of  $\theta$  and  $\phi$  were assumed to be 0 and 0.01, respectively, which corresponds to the case of pure segregational instability. The calculations were conducted for several different values of plasmid content  $N_p$  to show the influence of this key parameter.

These simulations show a number of interesting features. As plasmid content increases, the time required to reach stationary phase (substrate exhaustion) increases, the final concentration  $x_1$  of plasmid-containing, productive cells decreases, and there is a corresponding increase in the final concentration of plasmid-free cells ( $x_3$ ). However, batch fermentations at smaller ( $N_p = 10$ ) and larger ( $N_p = 100$ ) plasmid content give significantly less final cloned-gene product than batch cultivation of a strain with intermediate plasmid content ( $N_p = 60$ ). Thus, the calculations indicate the existence of an optimum plasmid content for maximizing process productivity.

Recently Aiba *et al.*<sup>38</sup> examined experimentally the relationship between the production of tryptophan and the the copy number of plasmids using recombinant DNAs with the *trp* operon cloned into four different plasmid vectors RP4, pSC101, RSF1010, and pBR322. They found that accumulation of tryptophan in the culture medium decreased when the copy number exceeded a certain level. According to their report,<sup>38</sup> the maximum tryptophan concentration in the culture medium was 1.7, 3.1, and 2.6 g/l for RP4,

pSC101, and RSF1010, respectively. The copy numbers corresponding to these plasmids are approximately 2, 5, and 15, respectively.<sup>38,39</sup> The time trajectories of cell growth, substrate consumption, and product formation shown in Fig. 2.3 resemble qualitatively the experimental data of Aiba *et al.* for shake flask culture of the recombinant *trp* system cloned into RP4, pSC101, and RSF1010, respectively.<sup>38</sup>

More details on the influence of plasmid content on final product concentration are shown in Fig. 2.4 which summarizes the results of many batch fermentation simulations. Circles show calculated product levels after substrate exhaustion based on the model parameters used above ( $\theta = 0.0$ ,  $\phi = 0.01$ ; pure segregational instability), and on  $f(\mu)$  given by Eq. (2.29). To evaluate the influence of the approximation introduced by using a linearized form of  $f(\mu)$  (Leudeking-Piret kinetics), batch reactor simulations were also completed for  $f(\mu)$  given by Eq. (2.30) (diamonds in Fig. 2.4) and by Eq. (2.31) (squares in Fig. 2.4). Clearly, all three functional forms for the dependence of protein synthesis rate on specific growth rate give qualitatively similar results: maximum product accumulation at intermediate plasmid content. The linearized model in Eq. (2.31) reasonably well approximates results obtained using the full model of Eq. (2.29).

Also shown in Fig. 2.4 is the calculated final value of the fraction of productive cells, defined as

$$\psi(t) = \frac{x_1(t)}{x_T(t)} \quad (2.34)$$

where

$$x_T(t) = x_1(t) + x_2(t) + x_3(t) \quad (2.35)$$

versus the plasmid content in the batch fermentation. The fraction of productive cells after substrate exhaustion declines monotonically as plasmid content increases. Because no product inhibition was considered in this particular set of calculations, this relation-

ship is independent of the choice of  $f(\mu)$ .

The effects of the type of instability and of product inhibition based on intracellular concentration (Eq. 2.3) or on total broth concentration (Eq. 2.4) are summarized in the simulation results plotted in Fig. 2.5. Again, values shown correspond to the end of the batch after substrate exhaustion. All cases considered show an optimum in final product concentration with respect to plasmid content. However, the maxima are much sharper for segregational instability ( $\theta = 0$ ) than for structural instability ( $\phi = 0$ ). Final product levels are higher at every plasmid content considered for structural instability than for segregational instability independent of the mode of product inhibition.

The qualitative trends for productive cell fraction  $\psi$  vary greatly depending on the type of instability. For segregational instability,  $\psi$  approaches zero as plasmid content increases, while the productive fraction dips and then increases for larger plasmid content for pure structural instability. This is reasonable since structural genetic instability gives plasmid-containing but not productive cells which do not differ greatly in growth rate from productive cells at high plasmid content, whereas complete plasmid loss (segregational instability) gives cells with much greater specific growth rate.

Plasmid construction can influence the gene expression parameter  $\gamma$  (Eq. 2.9.2). If product inhibition is not important, it is obvious that  $\gamma$  should be made as large as possible to maximize product synthesis. However, with product inhibition, simulations with the present model indicate an optimum value of  $\gamma$  for the case of intracellular product inhibition (Fig. 2.6). Also, these calculations show that the fraction of productive cells is influenced by the cloned-gene expression efficiency as characterized here by  $\gamma$ .

## Continuous Culture

Calculations for continuous culture of a host-vector system exhibiting pure segregational instability ( $\phi = 0.01$ ,  $\theta = 0$ ) are shown in Fig. 2.7. Because of genetic instability, the behavior of the continuous flow fermentor is intrinsically unsteady. The calculation

results illustrated in Fig. 2.7 were obtained from numerical solution of the unsteady material balances for continuous culture described earlier. The simulations shown here are for product inhibition based on intracellular product concentration. Initial conditions for the transient continuous reactor, listed below, were determined by simulating batch growth from the inoculum given in Eq. 2.34 because this is the usual start-up procedure for continuous bioreactors.

$$s(0) = 29.6 \text{ g/l}$$

$$p(0) = 0.011 \text{ g/l} \quad (2.36)$$

$$x_1(0) = 0.271 \text{ g/l}$$

$$x_3(0) = 0.029 \text{ g/l} .$$

These calculations display a number of interesting features. At some dilution rates, the total cell concentration drops and then increases. These results are qualitatively consistent with the experimental observations of Aiba and Koizumi<sup>40</sup> for continuous culture of *Bacillus stearothermophilus* CU21(pLP11) at 55° C.

The model simulations indicate that the fraction of productive cells,  $\psi$ , decreases as the dilution rate decreases. Similar behavior is evident in the experimental data of Wouters *et al.*<sup>41</sup> They investigated the loss of drug resistance in *E. coli* cells harboring pBR322 in a chemostat under phosphate-limited conditions. Their experimental data show that the fraction of drug-resistant cells is lower at lower dilution rate. However, the present model does not predict the lags in responses of unstable recombinant cultures which have been observed experimentally.<sup>41,42</sup> Interestingly, the model without product inhibition gives decreasing  $\psi$  with increasing dilution rate.

## DISCUSSION

The model for product formation employed here has been derived from a structured model which provides a relationship between product formation kinetics and fundamental biological parameters such as cellular vector content, transcription efficiency, and translation efficiency. It is interesting that a linear approximation to the final expression resulting from this derivation is identical to the empirical Leudeking-Piret equation for product synthesis rate originally formulated for citric acid production. The connection may be more than coincidental. If production of a metabolite is limited by certain key enzymes and if production of a metabolite is proportional to the amount of key enzymes in the cell, the derivation given here provides a possible explanation for the Leudeking-Piret form of kinetics, even for metabolites.

The model simulation results indicate an optimum value for intracellular vector concentration  $\hat{G}$  and for the product of cloned gene transcription efficiency,  $\eta$ , and translation efficiency,  $\xi$ . These host-vector parameters can be manipulated by both genetic engineering and by bioreactor control. For example, plasmid content can be changed by modifying the replicon, and gene expression efficiency can be modified by altering the promoter, ribosome binding site, and their positions relative to the structural gene. Concentrations of replication and expression repressors or activators can be adjusted by medium composition or temperature for suitably constructed recombinant strains.

There are several experimental reports which also indicate the existence of an optimum plasmid copy number or cloned gene expression activity.<sup>38,43-45</sup> One example was mentioned earlier. Another is the study of Feinstein *et al.* who determined the  $\beta_1$ -interferon activity in a recombinant *E. coli* system.<sup>43</sup> In the expression plasmid used in their studies, the *recA* promoter-operator was employed for regulation of transcription of the interferon gene. The *recA* promoter-operator activity is induced by nalidixic acid. As shown in Fig. 2.8, it was found that an optimum inducer concentration, corresponding

to an optimum  $\eta$  value, exists. This finding is qualitatively identical to that obtained from the present model for the case of intracellular product inhibition. Comparison of the present simulation results with other reports<sup>44,45</sup> also indicate that the general trends obtained from this model showing the existence of an optimal transcription efficiency are consistent with currently available experimental information.

In contrast to several previous models concerned with genetic instability and recombinant systems, the present model indicates that the fraction of product-synthesizing cells is a function of many parameters. This fraction is influenced by genetic parameters which determine the intrinsic vector and cloned gene stabilities ( $\phi$  and  $\theta$ ), the gene expression activity  $\gamma$ , the intracellular cloned-gene product concentration  $\hat{p}$ , the nature of the cloning vector, the cloned gene and its product, and the host cell. The latter properties will influence parameters such as  $\hat{G}_{\max}$ ,  $\hat{p}_{\max}$ , and maximum specific growth rate  $\mu^0$ . Also, environmental and process operating conditions such as dilution rate and substrate concentration influence the fraction of productive cells in the culture.

Product inhibition as described in this model has two different physical interpretations. The first is actual interference of product molecules or their aggregates with reactions and morphological developments associated with cell growth. Alternately, the total amount of product accumulated may be viewed as an integral measure of the cellular biosynthetic activity which has been diverted by plasmid product gene expression away from synthesis of host cell proteins. Such diversion is expected to reduce host cell growth rate.

The success of the present model in simulating trends in fermentations of recombinant organisms is quite encouraging and suggests that this framework may prove valuable for optimizing production processes for protein manufacture using recombinant systems at both the genetic and environment/process levels. There is a clear need for greater experimental studies of the interrelationships assumed here and calculated as a result of

this model. Another contribution of this type of simulation study is identification of certain effects and trends which are sufficiently amenable to measurement and/or are of sufficient practical interest to motivate the required experiments. With refined model parameters for a particular system, detailed optimization calculations follow directly from the present modeling framework.

**Acknowledgement:** This work was supported by the Energy Conservation and Utilization (ECUT) Program of the Department of Energy and by grants from KOSEF (H-02910) and Yu-Kong Ltd. (I-10511).

### Nomenclature

D	Dilution rate
$\hat{G}$	Intracellular cloning vector concentration
$k_d$	Decay constant of mRNA
$k_e$	Decay constant of protein
$k_p^o$	Overall transcription rate constant
$k_q^o$	Overall translation rate constant
$K_{sj}$	Monod constant of j cells
m	Exponent of vector inhibition term
$\hat{m}$	Intracellular concentration of cloned-gene mRNA
n	Exponent of product inhibition term (intracellular model)
n'	Exponent of product inhibition term (extracellular model)
$N_p$	Number of plasmids per cell
p	Extracellular concentration of product
$\hat{p}$	Intracellular concentration of product
s	Concentration of limiting nutrient
$x_j$	Concentration of j cells
$Y_{sj}$	g cells/g substrate yield factor

### Greek symbols

$\alpha$	Parameter of the Leudeking-Piret product formation model
$\beta$	Parameter of the Leudeking-Piret product formation model
$\gamma$	Gene expression parameter ( $\eta\xi$ )
$\eta$	Transcription efficiency
$\theta$	Ratio of mutation rate of cloned gene to the host cell specific growth rate
$\mu_j$	Specific growth rate of j cells
$\mu^o$	Maximum specific growth rate of wild-type cells
$\bar{\mu}_1$	$\mu_1(1 - \phi - \theta)$

$\xi$	Translation efficiency
$\rho_b$	Cell density
$\phi$	Ratio of cloning vector loss rate to the specific growth rate of the host cell
$\psi$	Fraction of product producing cells

**Subscripts**

$j$	Indicates the cell genotype (1= $G^+R^+$ , 2= $G^+R^-$ , 3= $G^-R^-$ )
$T$	total concentration
$o$	feed concentration

## References

1. I. S. Johnson, *Science* **219**, 632 (1983).
2. H. Moser, *The Dynamics of Bacterial Populations Maintained in the Chemostat* (Carnegie Institution of Washington, Washington, D.C., 1958)
3. T. F. Anderson and E. Lustbader, *Proc. Natl. Acad. Sci. USA* **72**, 4085 (1975).
4. F. M. Stewart and B. R. Levin, *Genetics* **87**, 209 (1977).
5. K. Nordström, S. Molin, and H. Angaard-Hansen, *Plasmid* **4**, 215 (1980).
6. T. Imanaka and S. Aiba, *Ann. N.Y. Acad. Sci.* **369**, 1 (1981).
7. D. F. Ollis, *Phil. Trans. R. Soc. Lond.* **B297**, 617 (1982).
8. D. F. Ollis and H.-T. Chang, *Biotechnol. Bioeng.* **24**, 2583 (1982).
9. J.-H. Seo and J. E. Bailey, *Biotechnol. Bioeng.*, in press.
10. S. B. Lee and J. E. Bailey, *Biotechnol. Bioeng.* **26**, 66, 1372, 1383 (1984).
11. S. D. Ehrlich, B. Niaudet, and B. Michel, *Curr. Top. Microb. Immunol.* **96**, 19 (1982).
12. P. A. Meacock and S. N. Cohen, *Cell* **20**, 529 (1980).
13. T. Ogura and S. Hiraga, *Cell* **32**, 351 (1983).
14. J. I. Rood, M. K. Sneddon, and J. F. Morrison, *J. Bacteriol.* **144**, 552 (1980).
15. G. Grandi, M. Mottes, and V. Sgaramella, *Plasmid* **6**, 99 (1981).
16. S. H. Kim, D. D. Y. Ryu, and S. Y. Lee, *Korean Biochem. J.* **15**, 305 (1982).
17. T. Imanaka, H. Tsunekawa, and S. Aiba, *J. Gen. Microb.* **118**, 253 (1980).
18. C. P. Dwivedi, T. Imanaka, and S. Aiba, *Biotechnol. Bioeng.* **24**, 1465 (1982).
19. C. G. DiJoseph, M. E. Bayer, and A. Kaji, *J. Bacteriol.* **115**, 399 (1973).

20. D. Godwin and J. H. Slater, *J. Gen. Microb.* **111**, 201 (1979).
21. R. M. M. Klemperer, N. T. A. Ismail, and M. R. W. Brown. *J. Gen. Microb.* **115**, 325 (1979).
22. U. M. Nordström, B. Engberg, and K. Nordström, *Mol. Gen. Genet.* **135**, 185 (1974).
23. J.-H. Seo and J. E. Bailey, *Biotechnol. Bioeng.*, submitted.
24. M. Zabeau and K. K. Stanley, *EMBO J.* **1**, 1217 (1982).
25. E. Padan, T. Arbel, A. Rimon, A. B. Shira, and A. Cohen, *J. Biol. Chem.* **258**, 5666 (1983).
26. E. Remaut, P. Stanssens, and W. Fiers, *Nucleic Acids Res.* **11**, 4677 (1983).
27. T. Sako, S. Sawaki, T. Sakurai, S. Ito, Y. Yoshizawa, and I. Kondo, *Mol. Gen. Genet.* **190**, 271 (1983).
28. A. Coller and D. B. Wilson, *BIO/TECHNOLOGY* **1**, 594 (1983).
29. J. E. Bailey and D. F. Ollis, *Biochemical Engineering Fundamentals*, p.371 (McGraw-Hill, New York, 1977).
30. F. Srienc, J. L. Campbell and J. E. Bailey, *Biotechnol. Bioeng.*, submitted.
31. R. Leudeking and E. L. Piret, *J. Biochem. Microb. Technol. Eng.* **1**, 393 (1959).
32. F. G. Heineken, H. M. Tsuchiya, and R. Aris, *Math. Biosciences* **1**, 95 (1967).
33. J. E. Bailey, N. Anderson, S. Peretti, J.-H. Seo and F. Srienc, *Ann. N.Y. Acad. Sci.*, in press (1984).
34. S. B. Lee and J. E. Bailey, *Plasmid* **11**, 151, 166 (1984).
35. A.L. Koch and G. Blumberg, *Biophys. J.* **16**, 389 (1976).
36. B. Atkinson and F. Mavituna, *Biochemical Engineering and Biotechnology*

*Handbook* (Nature Press, New York, 1983).

37. J. T. M. Wouters and J. G. van Andel, *FEMS Microb. Lett.* **16**, 169 (1983).

38. S. Aiba, H. Tsunekawa, and T. Imanaka, *Appl. Environ. Microb.* **43**, 289 (1982).

39. K. Nagahari, T. Tanaka, F. Hishinuma, M. Kuroda, and K. Sakaguchi, *Gene* **1**, 141 (1977).

40. S. Aiba and J.-I. Koizumi, *Biotechnol. Bioeng.* **26**, 1026 (1984).

41. J. T. M. Wouters, F. L. Driehuis, P. J. Polaczek, M.-L. H. A. van Oppenraay, and J. G. van Andel, *Antonie van Leeuwenhoek* **46**, 353 (1980).

42. D. Noack, M. Roth, R. Geuther, G. Müller, K. Undisz, C. Hoffmeier, and S. Gaspar, *Mol. Gen. Genet.* **184**, 121 (1981).

43. S. I. Feinstein, Y. Chernajovsky, L. Chen, L. Maroteaux and Y. Mory, *Nucleic Acids Res.* **11**, 2927 (1983).

44. T. Kurokawa, M. Seno, R. Sasada, Y. Ono, H. Onda, K. Igarashi, M. Kikuchi, Y. Sugino, and T. Honjo, *Nucleic Acids Res.* **11**, 3077 (1983).

45. H. M. Shepard, E. Yelverton, and D. V. Goeddel, *DNA* **1**, 125 (1982).

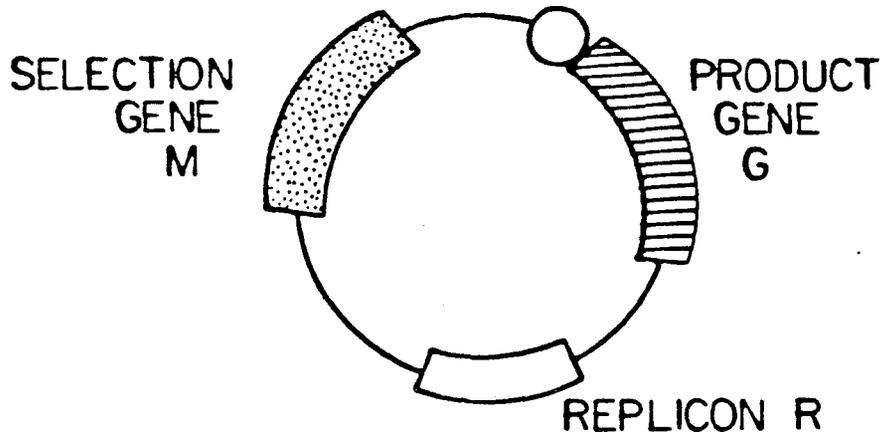
**TABLE 2.1**

Summary of Parameter Values Used in Model Calculations\*

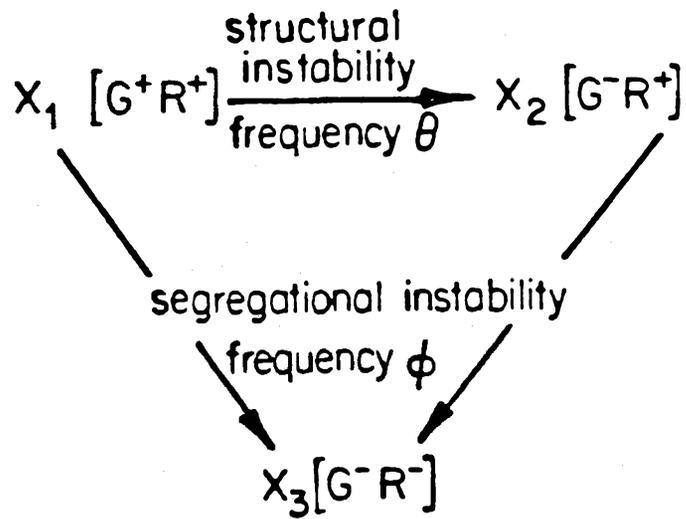
Parameter	Value
$K_s$	0.005 g/l
$m$	0.5
$n$	1.0
$n'$	1.0
$Y_s$	0.75 g cells/g substrate
$\gamma$	1.0
$\mu^o$	0.7 hr <sup>-1</sup>
$\rho_b$	1 kg/l
$\hat{G}_{max}$	0.036 g/l
$\hat{P}_{max}$	150 g/l
$P_{max}$	0.5 g/l

\* unless otherwise specified

### A. TYPICAL EXPRESSION PLASMID



### B. GENOTYPES AND THEIR TRANSITIONS



**FIGURE 2.1:** Schematic diagram of basic functional elements in an expression plasmid (A) and diagram of genetic transitions in a recombinant population with structural and/or segregational instability (B).

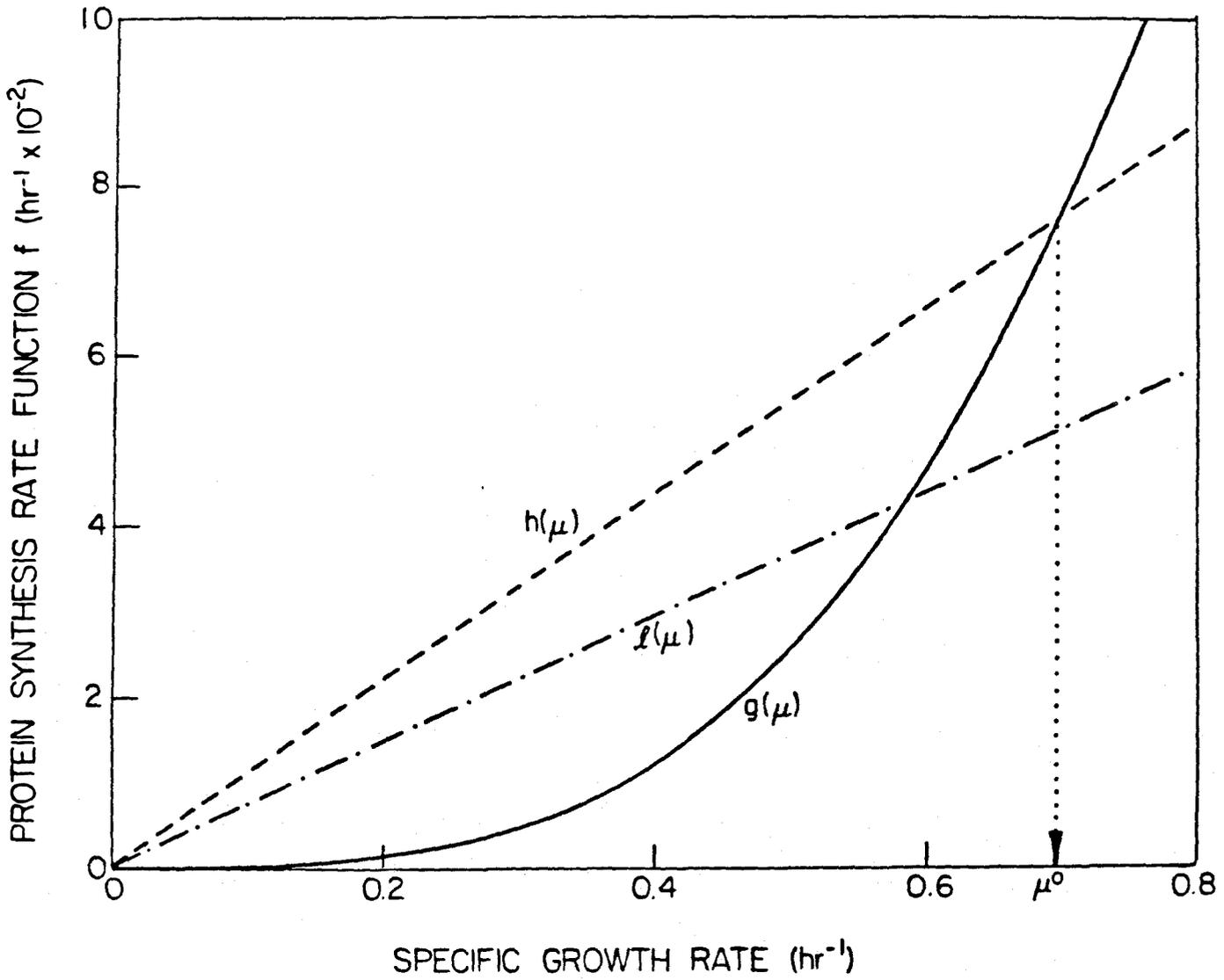
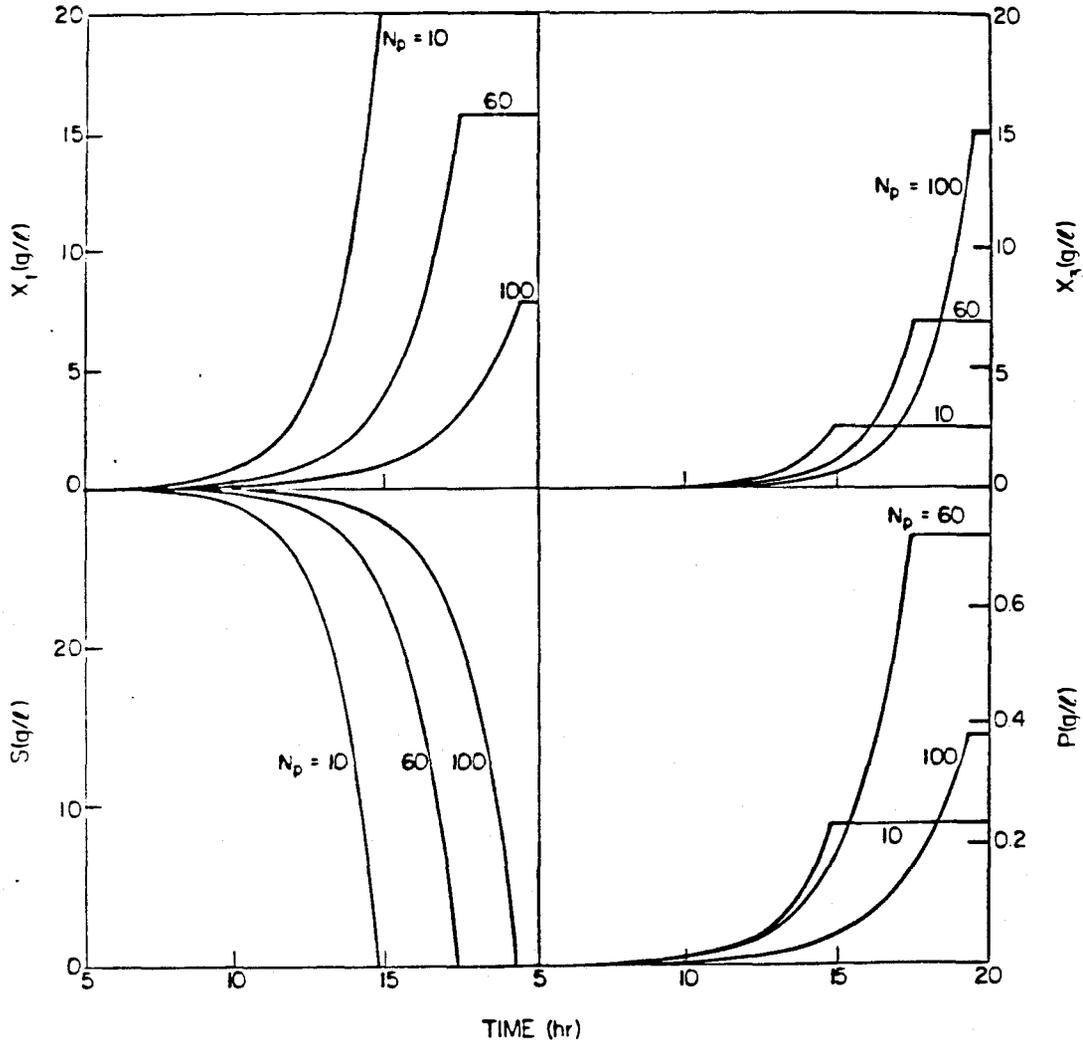
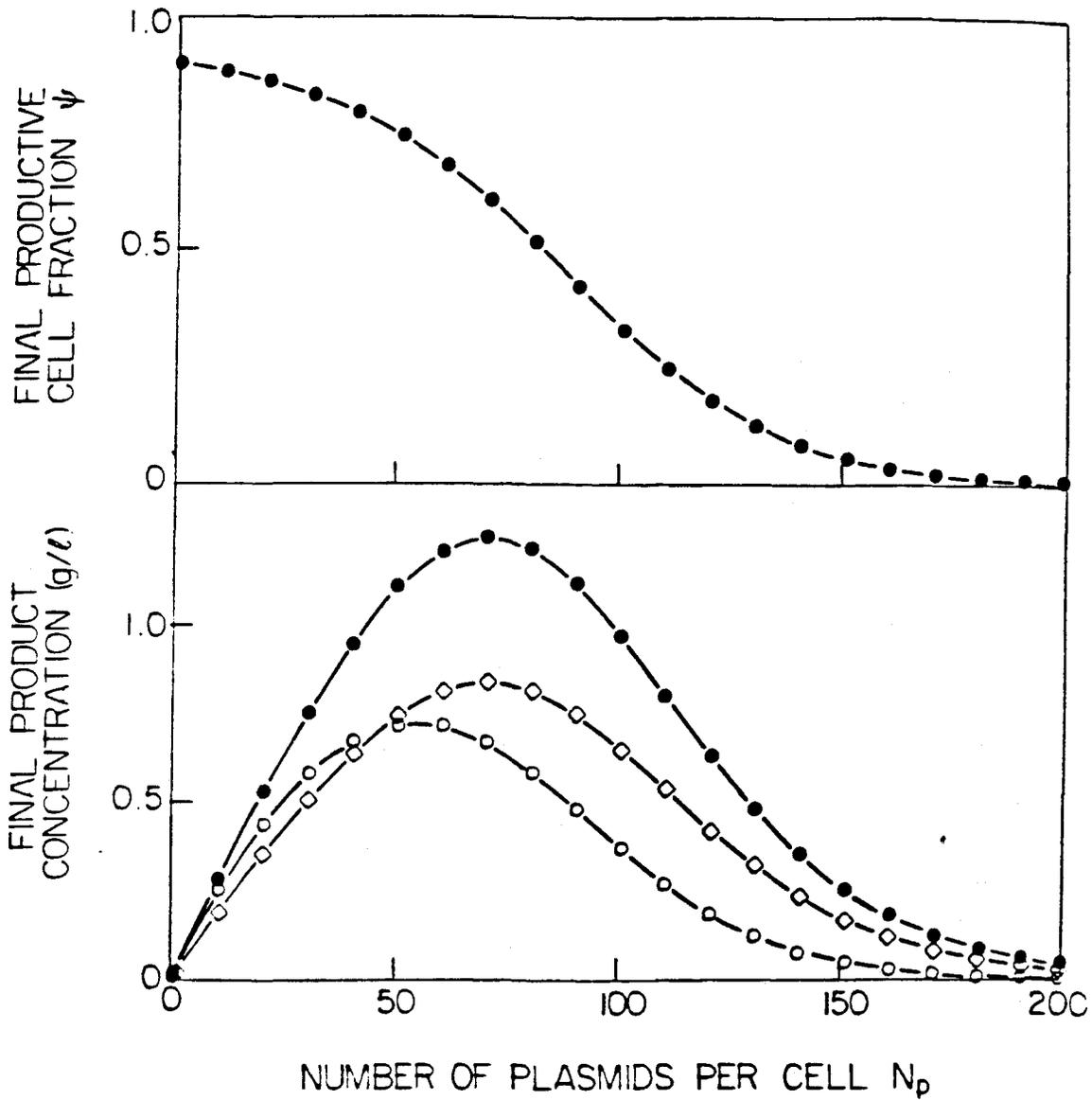


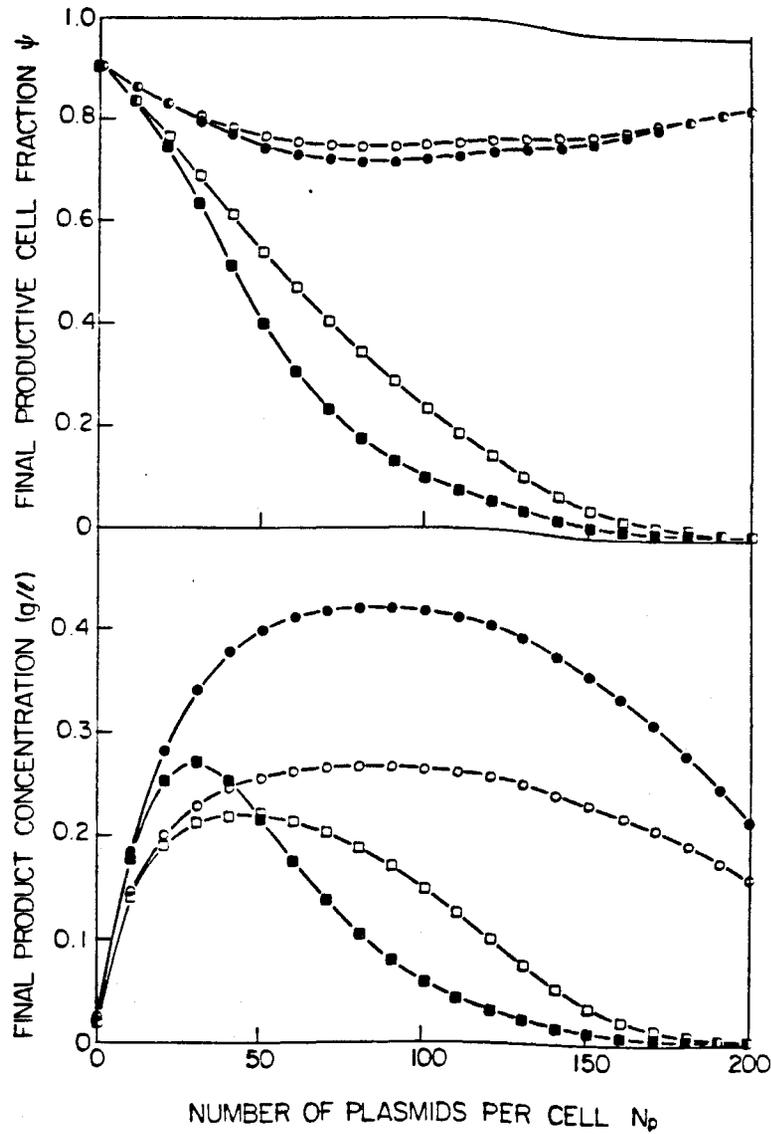
FIGURE 2.2: Alternate forms of the protein synthesis rate function  $f$  [Eq. (9.1)] considered in these models.



**FIGURE 2.3:** Simulated batch bioreactor trajectories for pure segregational instability ( $\theta=0$ ;  $\phi=0.01$ ) without product inhibition ( $n=0$ ).



**FIGURE 2.4:** Calculated productive cell fraction  $\psi$  and product concentration after substrate exhaustion in a batch bioreactor as a function of cell plasmid content (pure segregational instability;  $\theta=0$ ,  $\phi=0.01$ , without product inhibition;  $n=0$ . Calculations for  $f(\mu)$  given by Eq. 2.30 (●), Eq. 2.31 (◊), and Eq. 2.29 (○).



**FIGURE 2.5:** Dependence of final product content and productive cell fraction after substrate exhaustion in a batch bioreactor for different instability and product inhibition models. Solid symbols correspond to inhibition based on intracellular product concentration (Model I), and open symbols correspond to Model II which employs overall broth product concentration to evaluate product inhibition. Circles are for pure structural instability ( $\theta=0.01, \phi=0$ ), and squares were calculated for pure segregational instability ( $\theta=0, \phi=0.01$ ).  $f(\mu)$  given by Eq. 2.29 for all cases.

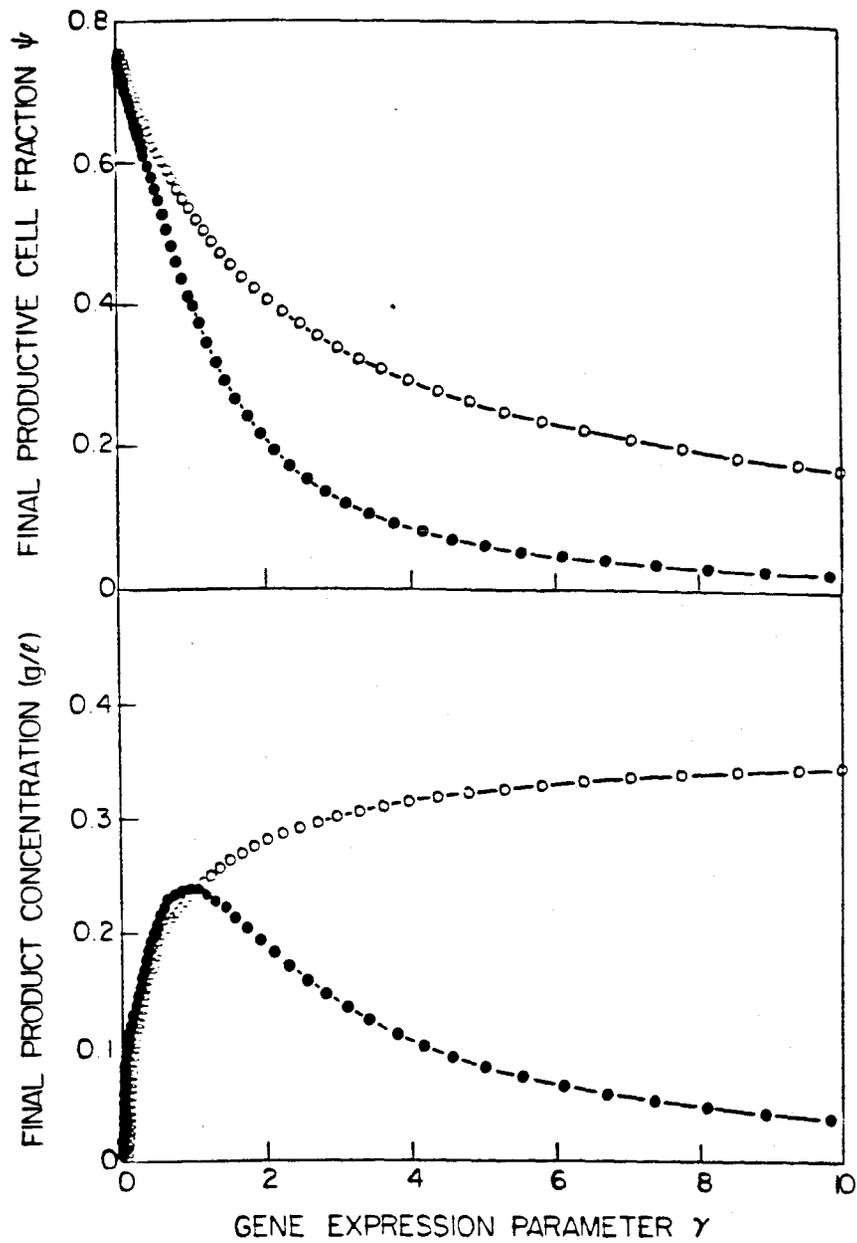
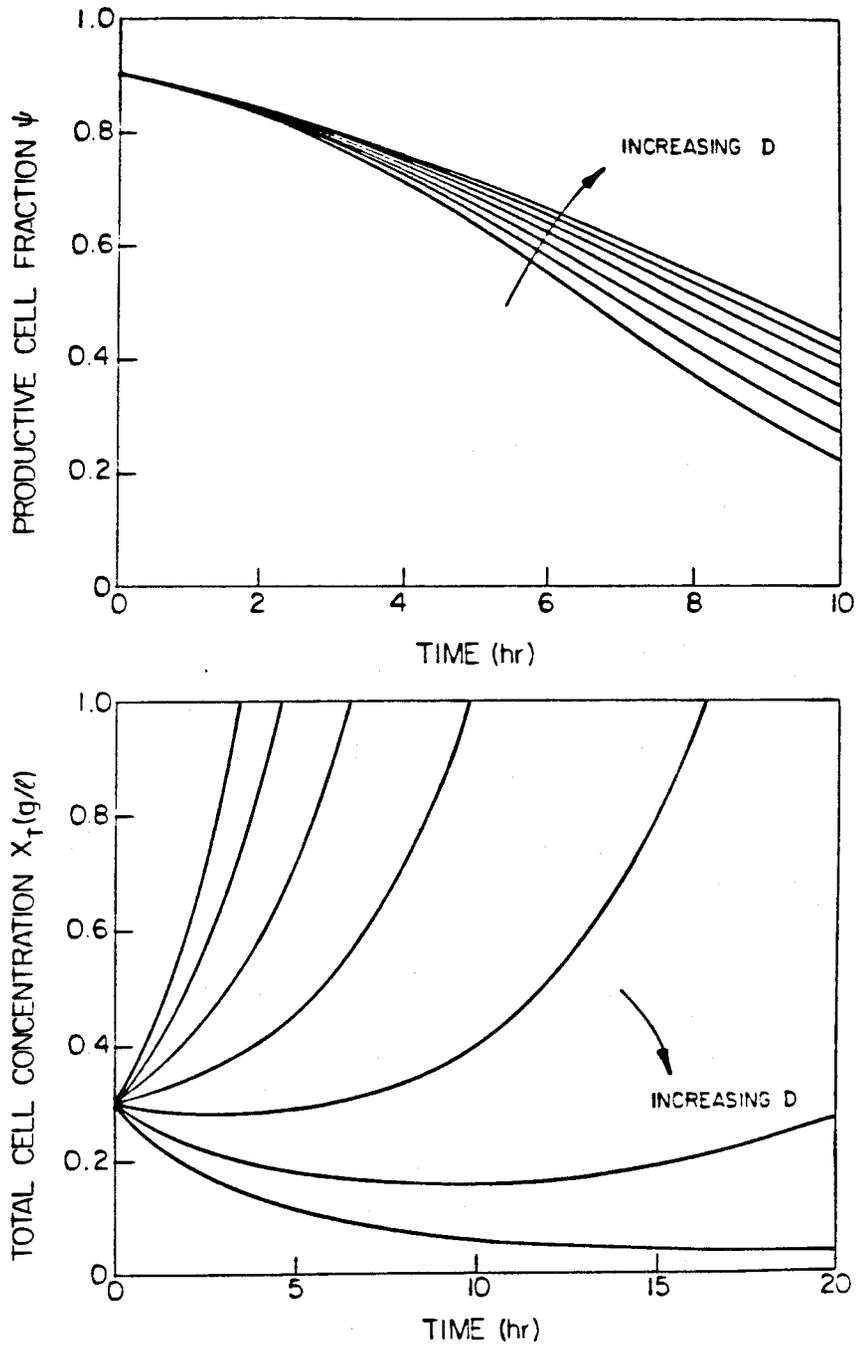
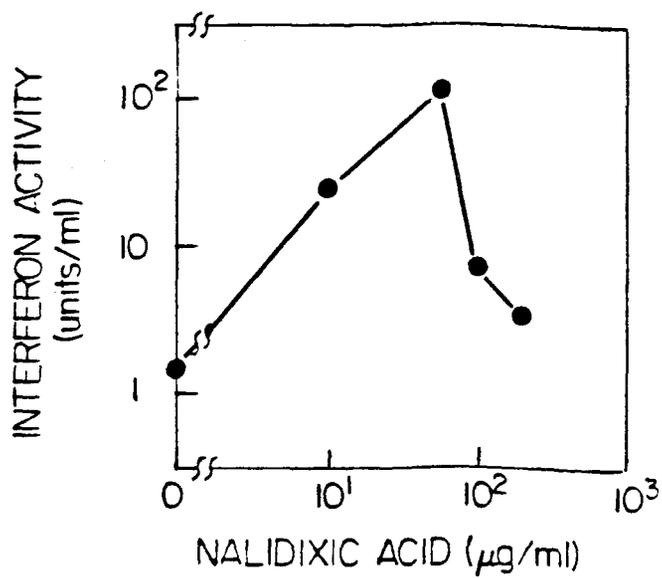


FIGURE 2.6: Simulated final productive cell fraction and product concentration versus the gene expression parameter,  $\gamma$ , characteristic transcription and translational efficiency, for product inhibition Model I ( $\bullet$ ) and Model II ( $\circ$ ). ( $\phi=0.01$ ,  $\theta=0$ ,  $N_p=50$ )



**FIGURE 2.7:** Computed time trajectories of total cell concentration and productive cell fraction following initiation of continuous-flow bioreactor operation. Dilution rates shown:  $D=0.1, 0.2, 0.3, 0.4, 0.5, 0.6,$  and  $0.7 \text{ hr}^{-1}$ . ( $\phi=0.01, \theta=0$ , product inhibition Model I, Eq. 2.29 for  $f(\mu)$ .)



**FIGURE 2.8:** Experimentally measured interferon activity obtained from recombinant *E.coli* fermentation as a function of inducer (nalidixic acid) concentration. Inducer concentration is correlated with cloned-gene transcriptional activity (data from Ref. 43 of Chapter 2).

## **CHAPTER 3**

### **OPTIMAL GENE EXPRESSION AND AMPLIFICATION STRATEGIES FOR BATCH AND CONTINUOUS RECOMBINANT CULTURES**

## INTRODUCTION

Maximizing the amount of protein produced from a cloned gene in a recombinant organism requires careful consideration of the tradeoffs involved between cloned-gene expression and host cell growth and biosynthetic activity. Cloned-gene product synthesis is favored by increasing cloned-gene content of the cell (e.g. Refs. 1-4). On the other hand, numerous experimental studies of recombinant *Escherichia coli* and *Saccharomyces cerevisiae* have shown that the presence of plasmids reduces host cell growth rate and, concomitantly, overall protein synthesis activity (e.g. Refs. 3, 4). Also, it has been observed that induction or derepression of cloned-gene expression accentuates this inhibitory effect.<sup>1</sup> Reduced host-cell growth rates and biosynthetic activity in the presence of plasmid-directed macromolecular synthesis presumably occurs because of competition between plasmid-directed and host cell-directed activity for common pools of precursors, chemical energy and electrons, activator species, repressor molecules, transport apparatus, and enzymes and other catalytic assemblies.

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. Several promoters are available for *E. coli*, *S. cerevisiae*, and other hosts which allow the expression level of the cloned gene to be switched from a relatively low to a relatively high level by a change in the organism environment.<sup>1,3</sup> Similarly, in a plasmid replicon repressed by a temperature-sensitive molecule, such as the ColE1 origin of replication for *E. coli* plasmids with a mutant RNA I gene providing temperature-sensitive replication repressor activity,<sup>5</sup> cellular plasmid content can be altered from around 25 to 700 or more copies per cell by increasing the medium temperature. Similar temperature-sensitive replicators are known for R1 plasmids.<sup>6</sup>

Such regulated replicons and promoters offer the biochemical engineer the opportunity to optimize process performance through application of optimized environmental changes which in turn generate optimal profiles of plasmid content and cloned-gene expression levels during the course of the process. In order to explore the influence of different operating strategies on recombinant culture product accumulation and also to examine the effects of different assumptions concerning host-plasmid interactions, a mathematical model previously formulated to describe product formation in unstable recombinant cultures has been used.<sup>7</sup> Responses of this model to different induction and amplification strategies have been calculated for batch and continuous bioreactors. Following presentation of these simulation results, the key factors which influence bioreactor performance and areas requiring further investigation are discussed.

## MATHEMATICAL MODEL AND CONTROLS CHARACTERISTICS

The mathematical model of unstable recombinant population growth and product synthesis employed in these calculations is presented in detail in Chapter 2. Here, only the major qualitative features of the model are summarized. Although certain details of the model are based upon data for *E. coli*, the qualitative characteristics embodied in the model are reasonable for other host organisms.

1. The specific growth rate of cells containing plasmids declines monotonically with increasing plasmid content. Cells expressing the cloned gene have decreased specific growth rates for increased accumulation of cloned-gene product.
2. The model includes both structural and segregational instability. (Only segregational instability is considered in these calculations.)

3. Cloned-gene product protein accumulation is calculated based upon a transcription-translation model which includes the growth rate dependence of transcription and translation activities.
4. Cell yields based upon substrate utilization are presumed constant and equal for all cell types.

It is presumed that the host-vector system has been constructed so that a change in culture temperature or concentration of an inducer alters cell plasmid content and cloned-gene expression activity. (To avoid cumbersome and repetitive language, gene expression regulation will be discussed henceforth only for the case of induction. Analogous considerations apply for derepression.) In the absence of detailed experimental data on the kinetics of plasmid amplification and gene expression induction, the particular ramp-plateau trajectories shown in Fig. 3.1 are assumed to result after a switch in environmental conditions causing amplification at time  $t_{AMP}$  and gene expression after time  $t_{EXP}$ . Thus, expression activity and plasmid content are not assumed to change instantaneously to new values.

Although the qualitative features to this simulation model are reasonable and reflect in some respects a substantial amount of realistic detail, certain parts of the model, including the features just discussed, are approximations based upon limited information. Thus, the model employed here is expected to be reasonable for investigating certain qualitative trends, but, in the absence of complete and more detailed experimental information for a particular host-vector expression system, this model is not expected to afford detailed and exact description of a particular production organism. Accordingly, it is appropriate at this stage to investigate a reasonable yet limited class of operating strategies. In this work, only single switch, bang-bang environmental operating strategies are considered. That is, it is presumed in batch reactor operation that, at a particular point in the batch, designated  $t_{AMP}$  or  $t_{EXP}$ , the amplification and expression processes, respectively, begin

and follow the trajectory shown in Fig. 3.1. In continuous bioreactor calculations, it is presumed that the overall reactor consists of three vessels (or relatively independent mixing domains) in series, that conditions in one domain support growth at the original, basal levels of plasmid, that cloned-gene expression are active in one of the following vessels, and that both gene amplification and expression are active on the other vessel.

### BATCH BIOREACTOR SIMULATION RESULTS

Calculations have been conducted employing the model parameters previously listed.<sup>7</sup> In these calculations, pure segregational instability was considered ( $\theta=0$ ,  $\phi=0.01$ ). Product inhibition of host cell growth was calculated based upon the intracellular product concentration (Model I, Eq. 2.3). Computational results showing the cloned-gene product concentration (grams product per liter of culture) and the fraction of productive plasmid-containing cells at the time of substrate exhaustion during the batch are plotted in Fig. 3.2. The abscissa is the time of cloned gene expression induction or plasmid amplification initiation, designated the switching time. The final maximal gene expression activity  $\gamma_f$  and number of plasmids per cell  $N_{P,final}$  are here: 1.0 and 50, respectively. In each case, it is assumed that the other parameter begins at its "on" or fully induced or amplified level. Thus, for calculations exploring the effect of amplification time, gene expression is presumed to be 50 from time zero in the case of the calculations of the effect of cloned-gene induction.

Clearly, the amount of product synthesized by the recombinant culture is maximized by an operating strategy which employs low levels of gene expression during an initial growth stage of the batch culture and which switches to high product synthesis activity at a suitable point during the batch. The fraction of productive cells in the population increases monotonically as the time of switching to high expression and amplification conditions is increased. However, this trend plateaus late in the batch. Examining product

concentration and productive fraction results together indicates that host-plasmid interaction/inhibition phenomena are entirely responsible for the sharp decline in achievable product concentration accompanying late switches to expression and amplification. Also, if the switch to product synthesis is delayed too long, a certain interval of opportunity for product accumulation has been lost and limited time for growth of the culture remains.

With the same model parameters, more extensive simulations have been undertaken considering the opportunity to accomplish both gene expression switching and plasmid amplification independently at different points during the batch. Both plasmid content and cloned-gene expression level are presumed low initially. The final concentration of product and the fraction of plasmid-containing, productive cells as a function of these two switching times are plotted in Figs. 3.3 and 3.4, respectively. These calculations show that the final product concentration is much more sensitive to the switching time choices than the final productive fraction. Although perhaps obscured by the perspective needed to show the response surface, the greatest product concentration at completion of the batch occurs when the switching times for expression and amplification are both equal to 10 hrs. Fig. 3.5 illustrates the time trajectories of product concentration and productive fraction for these optimal conditions ( $t_{AMP}=10$  hr.,  $t_{EXP}=10$  hr., solid line) relative to the uncontrolled case when plasmid content and gene expression level are  $N_p=50$  and  $\gamma=1.0$  throughout the batch. Optimal operating conditions for the particular parameters considered here provide approximately twice as much product and increase maximum productivity even greater relative to the uncontrolled case. Since the maximum product concentration is achieved sooner in the controlled case, productivity (product concentration divided by elapsed operating time) is improved even more.

The degree of product accumulation improvement afforded by optimal switching of the amplification or expression time and the absolute value of the maximum product concentration achieved is a function of the final gene expression efficiency represented in

the model by the parameter  $\gamma_f$  and the final plasmid content,  $N_{P,final}$ . By changing promoter and ribosome binding site sequences and other aspects of the expression plasmid, the activity or "strength" of fully induced gene expression can be altered,<sup>8</sup> corresponding to change in the model parameter  $\gamma_f$ . It is expected intuitively that stronger expression systems would have the direct benefit of increasing product synthesis rate but the indirect detriment of greater inhibition of normal cell metabolism and biosynthetic activity. Model simulations to explore the relationships among induced expression activity, time of expression induction, and the final product and productive fraction levels are summarized in Fig. 3.6. Interestingly, for the kinetic and yield parameters employed in these calculations, there is an optimum combination of  $\gamma_f$  and  $t_{EXP}$  which greatly increases final product concentration.

Fig. 3.7 shows calculated results for product concentration at the end of the batch (at the time of substrate exhaustion) vs. the initiation time for amplification for several different  $N_{P,final}$  values. Also, product inhibition has been assumed negligible here ( $n=0$  in Eq. 2.3). The sharpness of the maximum in final product level is greatest for the largest  $N_{P,final}$  value considered. This maximum is less steep at an intermediate value of final copy number. However, here the maximum amount of product achievable is greater in absolute terms. The benefits of using a host-vector system providing environmental control of cloned-gene content is relatively less important in the case of low  $N_{P,final}$ . These results provide interesting insights into the interrelationship of host-plasmid production strain construction, reactor operating strategies, and achievable production levels.

It is interesting to examine results of similar calculations for different situations as represented by different model parameters. First, the case of negligible inhibition of growth due to product accumulation has been considered ( $n=0$  in Eq. 2.3). Here, choosing the optimal time for initiation of amplification at approximately 8 hrs. produces a much smaller increase in product concentration at the end of the batch than does

the batch at high plasmid content characteristic of the fully amplified system ( $N_{P,final}=50$ ). If, on the other hand, the amplification involves attainment of higher plasmid content (the points for  $N_{P,final}=90$  in Fig. 3.7), the use of a switchable amplifiable system and optimal choice of the amplification time now provides substantial benefits relative to the reference case ( $N_P=90$  throughout the batch).

## CONTINUOUS BIOREACTOR SIMULATION RESULTS

Staging of continuous reactors, operating at different conditions, can be used to mimic in a continuous process the effect of changing conditions with time in a batch process. This well-known relationship in chemical process engineering has been demonstrated experimentally by Siegel and Ryu for recombinant *E. coli*.<sup>1</sup> Using a cloned promoter induced by temperature shift, they studied a two-stage continuous flow stirred tank (CSTR) system in which a lower first tank temperature favored growth while higher second tank temperature induced gene expression. Since population growth does not occur under fully induced conditions,<sup>1</sup> this two-stage arrangement, with the first tank providing an inoculant stream to the second, is essential for a successful process.

Using the intracellular simulation model previously described,<sup>7</sup> the behavior of a more elaborate, three-tank process has been examined. Here, it is assumed that different controls on cloned gene expression induction (e.g., by adding inducer) and plasmid amplification (e.g., by temperature shift) allow independent manipulation of these two processes. It is assumed that both induction and amplification are off in the first CSTR, that the second CSTR conditions cause amplification (the plasmid copy number goes from 5 copies/cell to 50 copies/cell) and that cloned gene expression is induced (the expression parameter  $\gamma$  goes from 0 to 10) in the final CSTR of the three-reactor cascade. The total volume of the three reactors is maintained constant, and different allo-

cations of this fixed capacity among growth, plasmid amplification, and gene expression are examined using model simulations.

Because of genetic instability, CSTR behavior in this case may be intrinsically dynamic; achievement of a steady state is not generally expected. Accordingly, model results are presented as time trajectories. Conditions at time zero are those at the end of a startup phase: tank 1 contains an inoculum based on previous batch growth (see Eq. 2.36), and tanks 2 and 3 contain no cells. Fig. 3.8 shows the productive cell fraction and product concentration in the effluent from the third tank.

As expected the strategy of assigning equal residence times in all three chemostats is not an optimal one. Higher product concentrations in the effluent from the third CSTR can be achieved by distributing the residence times in a decreasing order; i.e., the productive cells grow for longer time in an environment that does not favor gene amplification and product formation. The substrate concentration in the effluent of the system, in the cases shown, drops to nearly zero after the system has been operating for a while (the time of substrate exhaustion is indicated by the breaks in the productive cell fraction vs. operating time curves). This means that practically all of the substrate being fed to the system is converted to biomass, and the total concentration of cells in the effluent is constant.

The performance of the system depends on how the productive cell fraction is affected by the different distributions of residence times. Maximum growth rates of the two cell populations (productive and plasmid-free cells) are very similar when no product is present and the plasmid copy number is kept low. By having larger residence times for the first stage, growth of the productive cells is favored (given that, at time zero, the concentration of the plasmid-free cells in the first reactor is much lower than that of the productive ones). When the gene copy number is amplified, the difference on the maximum growth rates of the two populations widens, but again not as much as when product inhi-

bition is introduced by cloned-gene expression. When finally product is first formed (in the last CSTR), the fraction of productive cells is high when the residence times have been distributed in a decreasing order, and product concentration in the effluent is increased compared with the equal-residence-times case. On the other hand, if the time spent in the third reactor becomes too small, increased effluent product concentration is not obtained (in the limit that the cells spend no time in the third reactor, the product concentration in the effluent will drop to zero).

Finally, the case of an increasing distribution of residence times in the three CSTRs is also shown. Both the product concentration and productive cell fraction are lower than for the equal-residence-times case. Such a distribution favors the plasmid-free cell population, and, although the case shown does not result in total washout of the productive cells, this phenomenon, can also be observed for the different values of the dilution rates.

## CONCLUSION

The trends evident in the simulations above are extremely reasonable on intuitive grounds and probably apply to host-vector systems quite different from those for which the model was formulated and with parameter values different from those considered. The more acute the negative influence of fully implemented gene expression on cell growth, the more useful it is to employ an origin of replication and/or cloned gene promoter which permits environmental switching from low to high levels. Thus, to optimize a particular expression system with respect to these bioreactor operating variables, it is important to know the inhibition relationship among plasmid content, cloned-gene expression and product accumulation, and host-cell growth and protein synthesis.

The present model assumes reasonable trajectories for plasmid content and cloned-gene expression activity as a function of time since environmental switch. The results do

reach the final copy number was changed from 2 hours to 4 hours since environmental switch, the difference in the final product concentration was less than 10% for the batch reactor case. If, however, the transient from the value characteristic of the original, growth-optimizing conditions and the final, product formation-optimizing conditions is substantially different and occurs over a much longer time scale, significant effects on the relationship between operating strategy and total product accumulation could result.

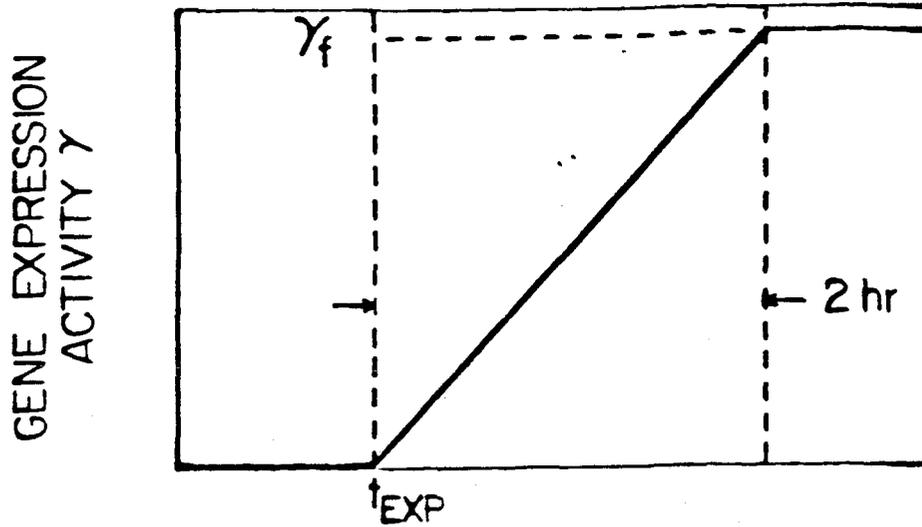
Only a limited and rather specialized class of environmental manipulations, a single switch was contemplated in this simulation study. More detailed optimization calculations allowing continuous trajectories of environmental conditions influencing gene amplification and expression levels is not deemed warranted until more detailed models are available, based upon experimental data, for cell growth inhibition and for the transients of gene expression efficiency and amplification. Given such information, it would be then interesting and worthwhile to explore the extent to which more sophisticated operating strategies and more finely tuned optimization calculations could further improve bioreactor performance.

**Acknowledgement:** This work was supported by the Energy Conversion and Utilization Technology (ECUT) program of the U.S. Department of Energy.

## References

1. R. Siegel and D. Y. Ryu, *Biotechnol. Bioeng.* **27**, 28 (1985).
2. J. Koizumi, Y. Monden, and S. Aiba, *Biotechnol. Bioeng.* **27**, 721 (1985).
3. L. Guarante, R. Ycum, and F. Gilford, *PNAS (USA)* **79**, 7410 (1982).
4. J.-H. Seo and J. E. Bailey, *Biotechnol. Bioeng.* **27**, 1668 (1985).
5. D. R. Moser, C. D. Moser, E. Sinn and J. L. Campbell, *Mol. Gen. Genet.* **192**, 95 (1983).
6. B. E. Uhlin and K. Nordström, *Mol. Gen. Genet.* **165**, 167 (1978).
7. S. B. Lee, A. Seressiotis, and J. E. Bailey, *Biotechnol. Bioeng.* **27**, 1699 (1985).
8. J. D. Windass, C. R. Newton, J. D. Maeyer-Guynard, V. E. Moore, A. F. Markham, and M. D. Edge, *Nucleic Acids Res.* **10**, 6639 (1982).

CLONED-GENE INDUCTION



PLASMID AMPLIFICATION

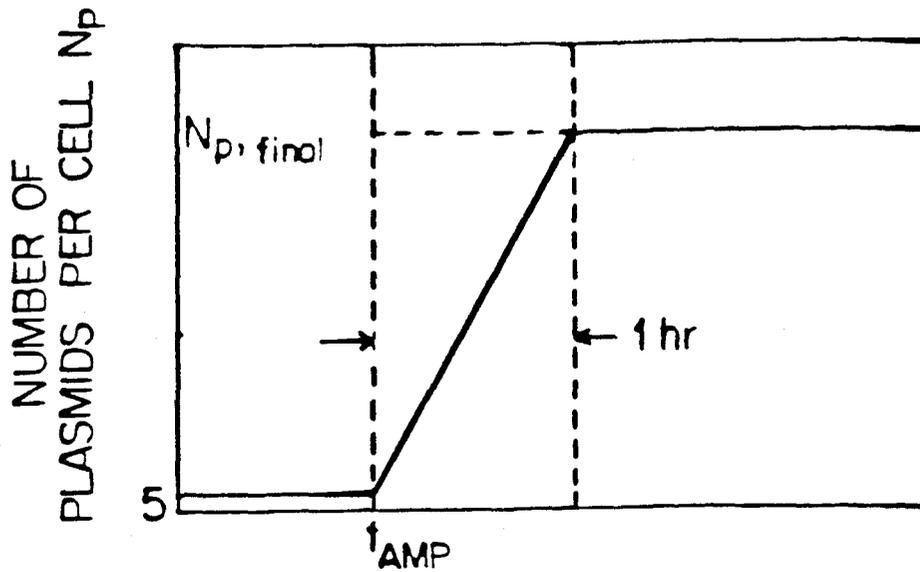


FIGURE 3.1: Diagram of the assumed trajectories for cellular plasmid content and gene expression parameter following a switch to environmental conditions promoting amplification and gene expression at time  $t_{AMP}$  or  $t_{EXP}$ , respectively.

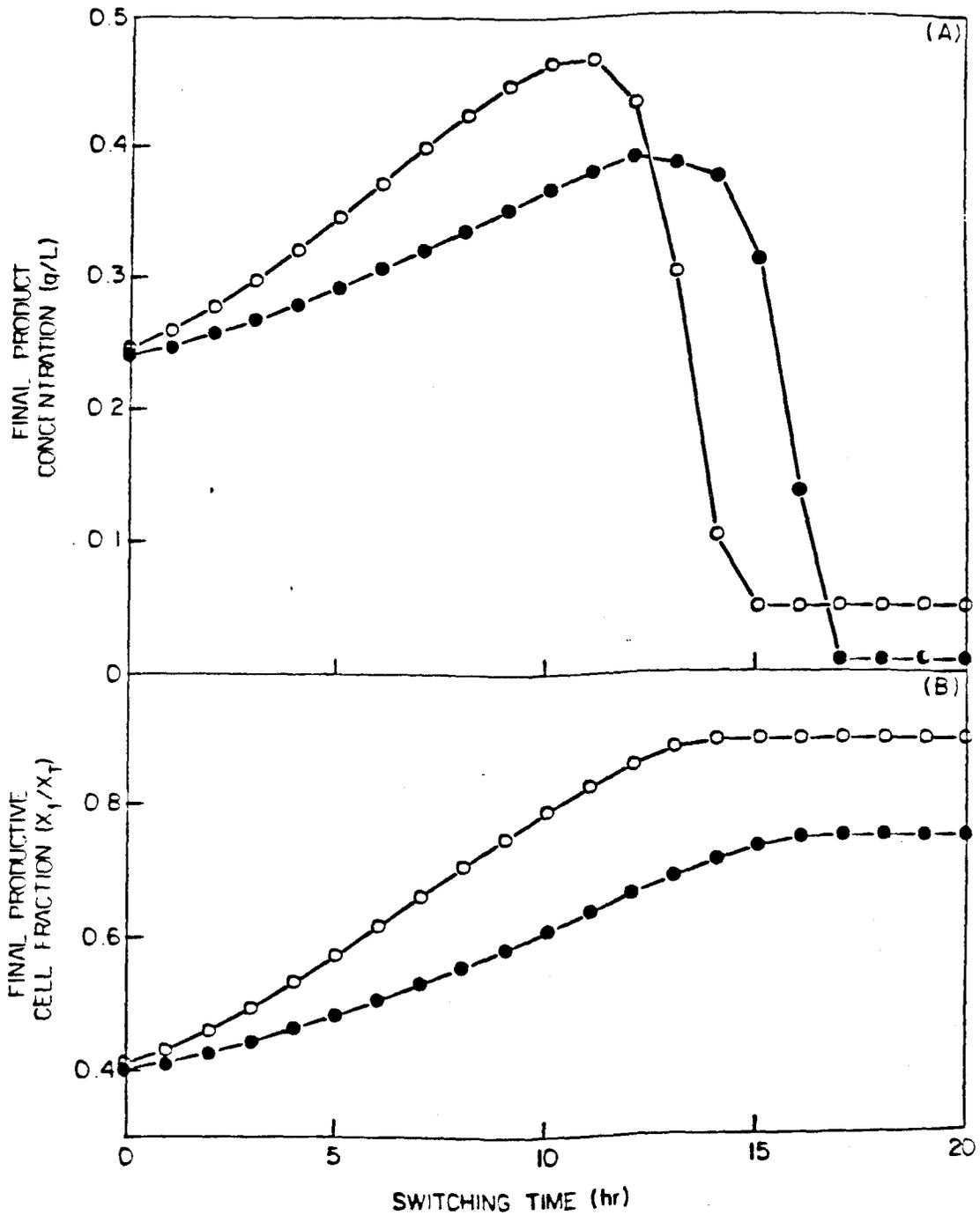
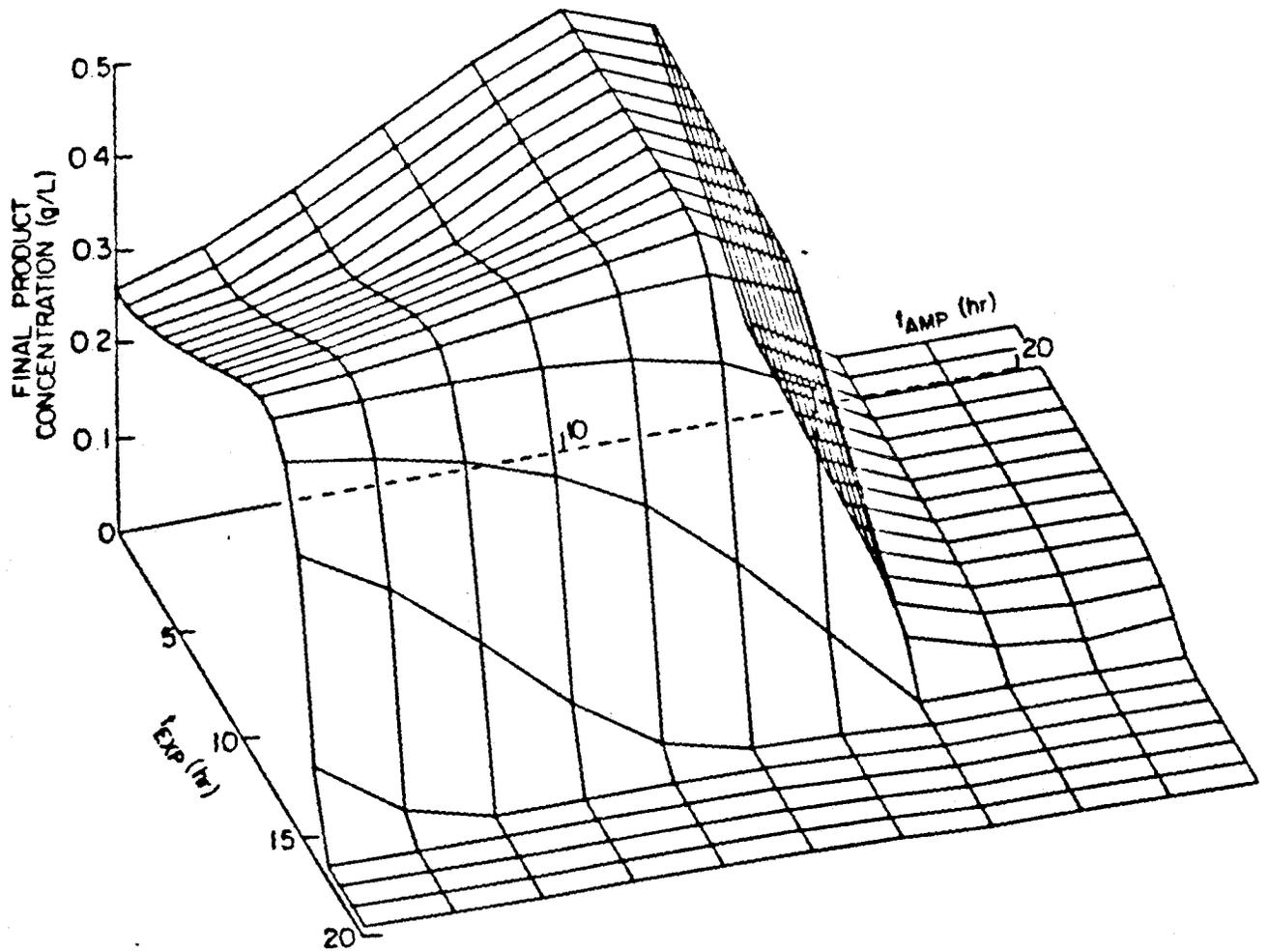


FIGURE 3.2: (A) Final cloned-gene product concentration (grams protein/liter culture) vs. the switching time (the time of plasmid amplification (o) and time of cloned-gene induction (●)). (B) Fraction of cells containing plasmids after substrate exhaustion in the batch process as a function of initiation times for amplification (o) or gene expression (●).



**FIGURE 3.3:** Illustration of product concentration at substrate exhaustion as a function of the time of amplification ( $t_{AMP}$ ) and the time of expression ( $t_{EXP}$ ) when both events can be activated independently at different points during batch operation. For the particular parameters considered here, maximum product accumulation is provided for  $t_{AMP}=10$ ,  $t_{EXP}=10$ ,  $N_{P,final}=50$ ,  $\gamma_f=1.0$ .

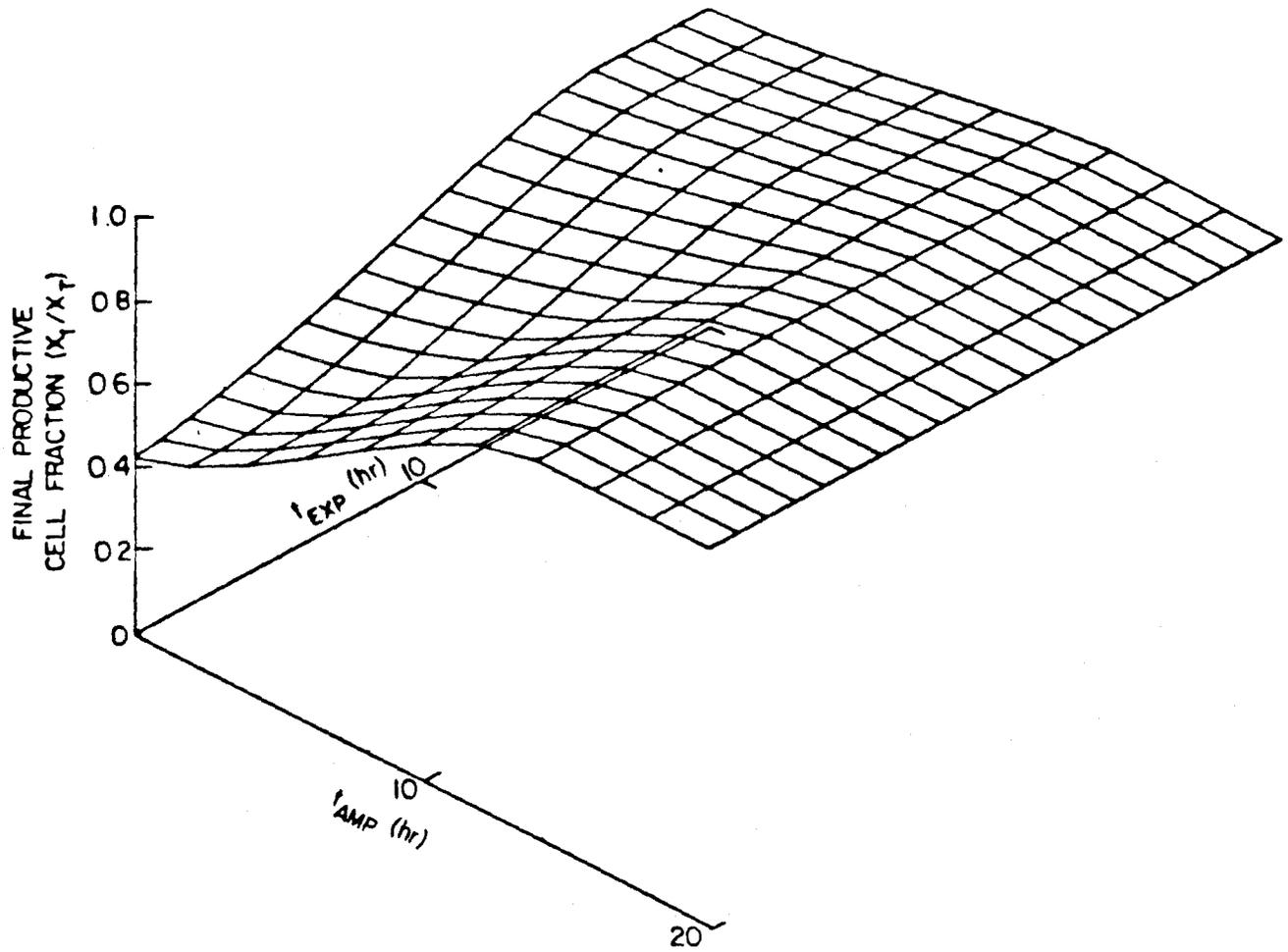
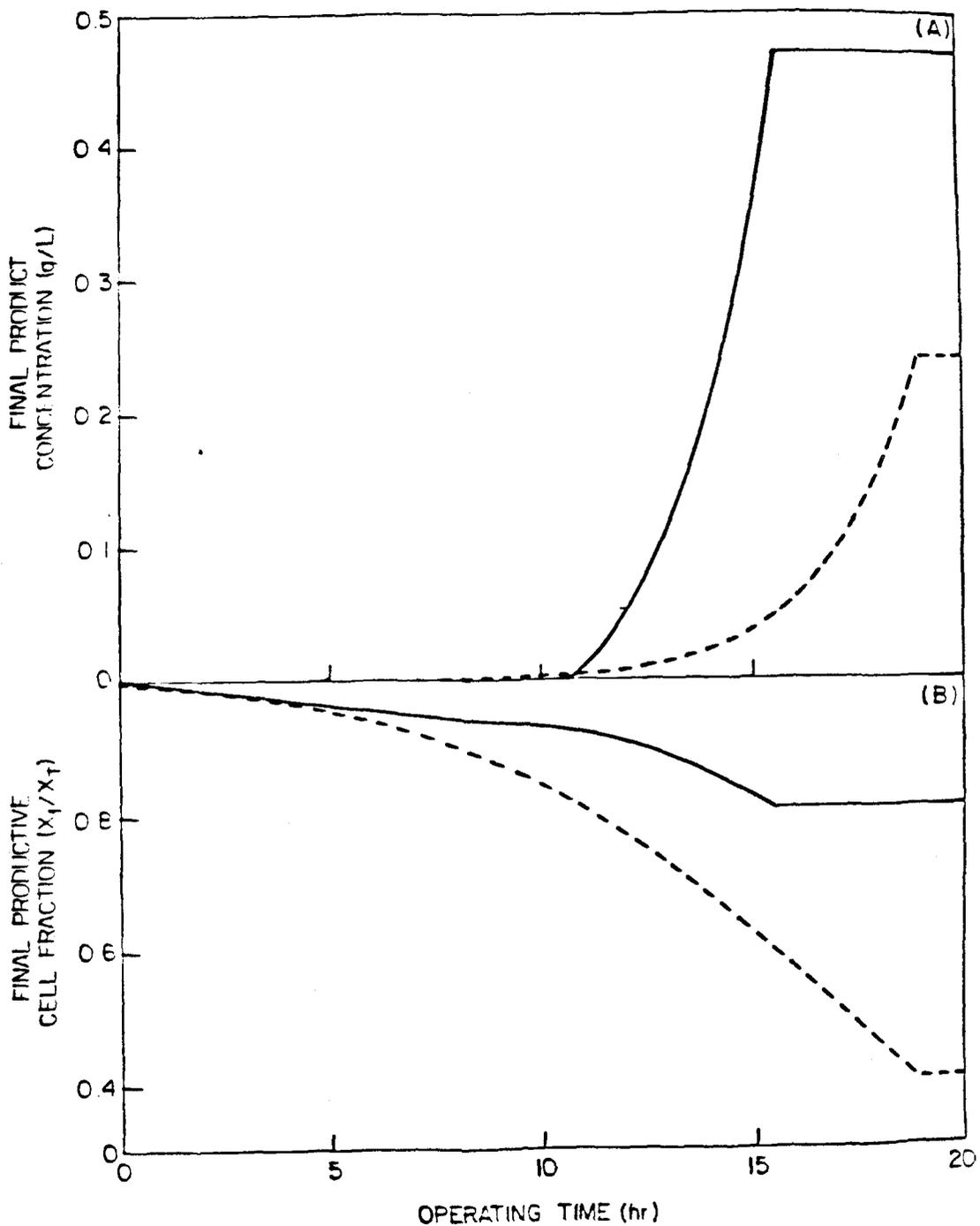


FIGURE 3.4: Final productive cell fraction vs. amplification and expression switching times for batch operation. Same parameters as Fig. 3.3.



**FIGURE 3.5:** Time trajectories (solid lines) of the productive cell fraction and cloned-gene product concentration vs. operating time in a batch reactor with  $t_{AMP}=10$ ,  $t_{EXP}=10$ . Shown for comparison (dashed lines) are trajectories for the uncontrolled case with  $N_p=50$ ,  $\gamma=1.0$ .

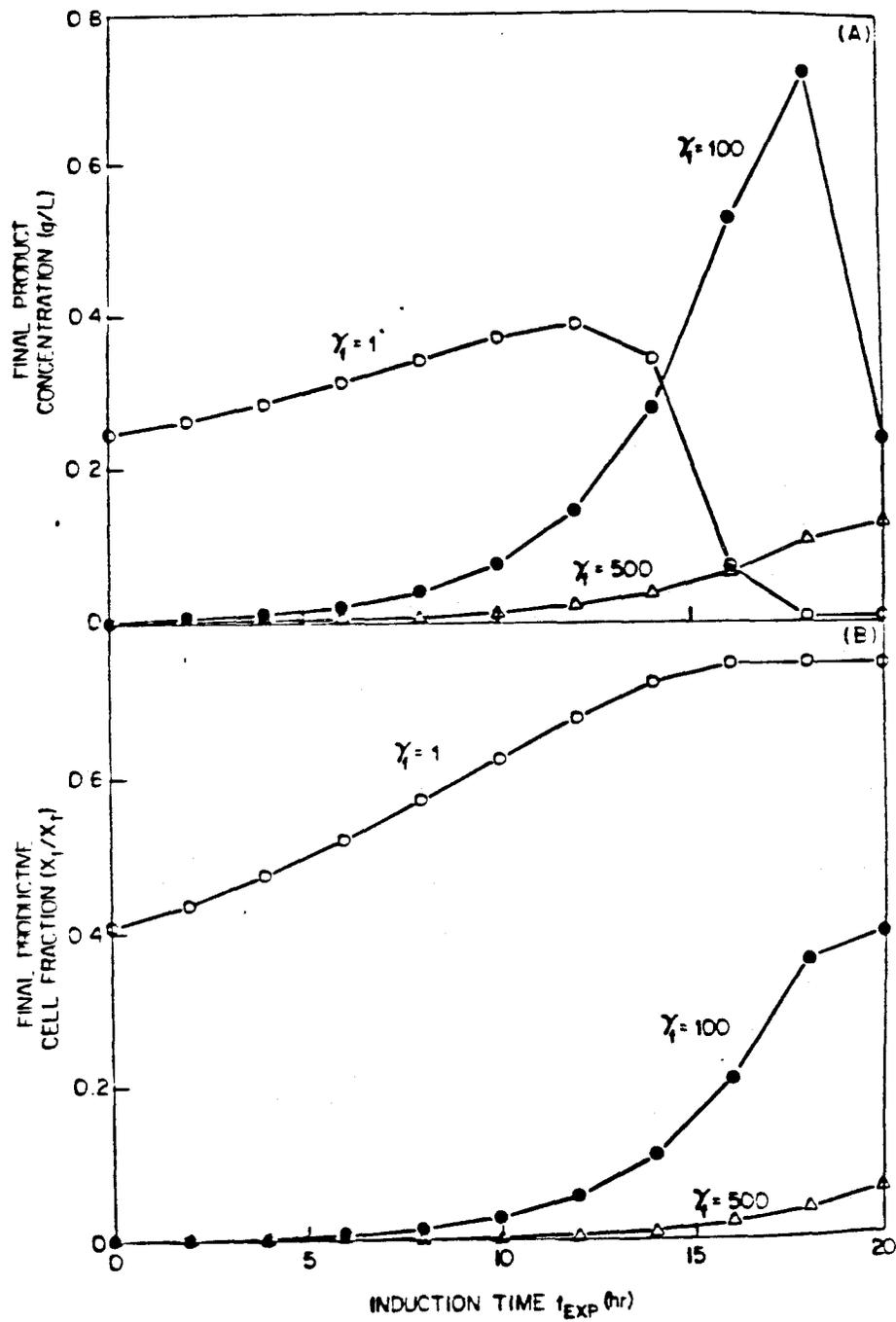


FIGURE 3.6: Final product and productive cell fraction values for different induced gene expression activity values  $\gamma$  as a function of induction time  $t_{EXP}$ ; batch reactor.

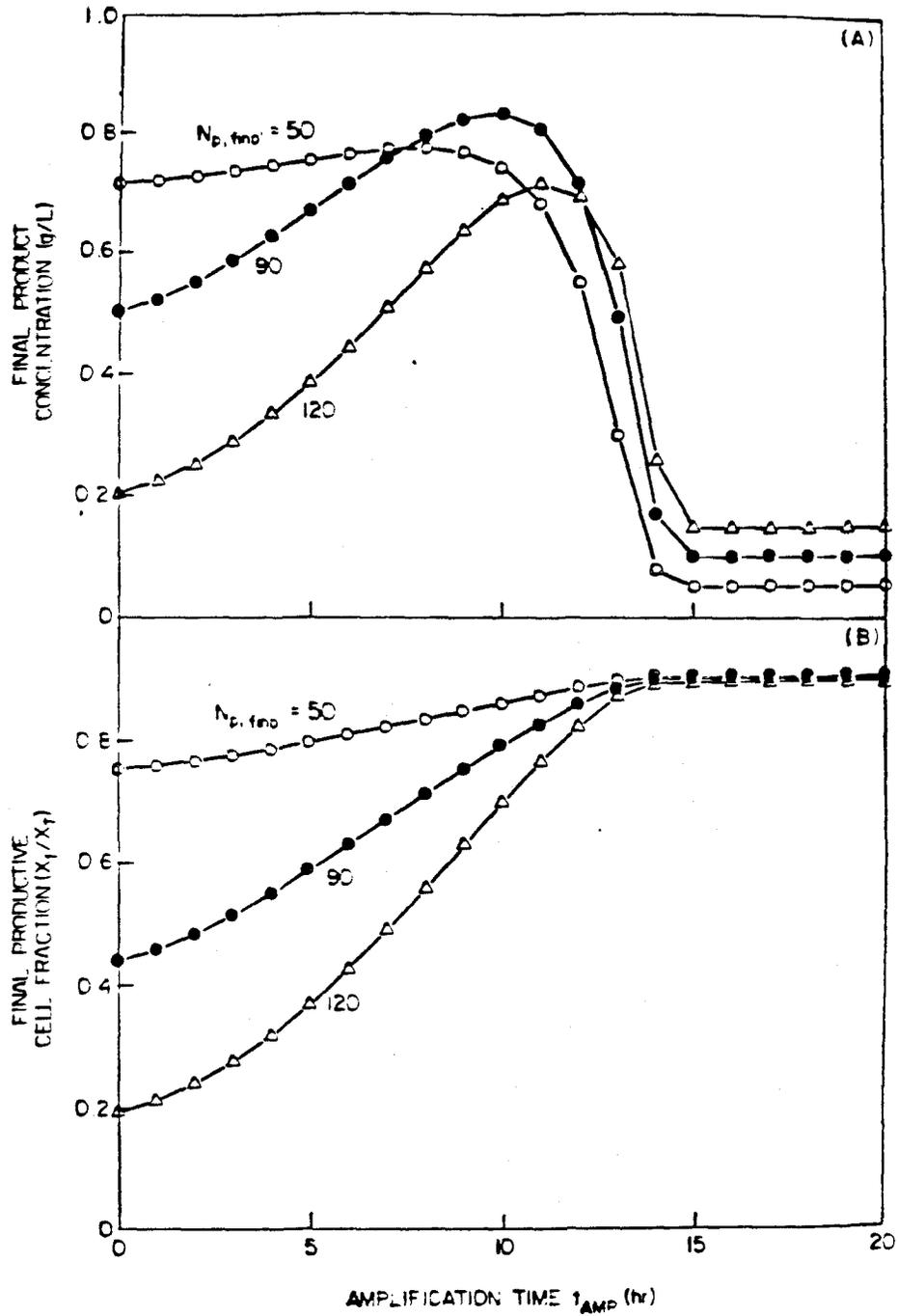
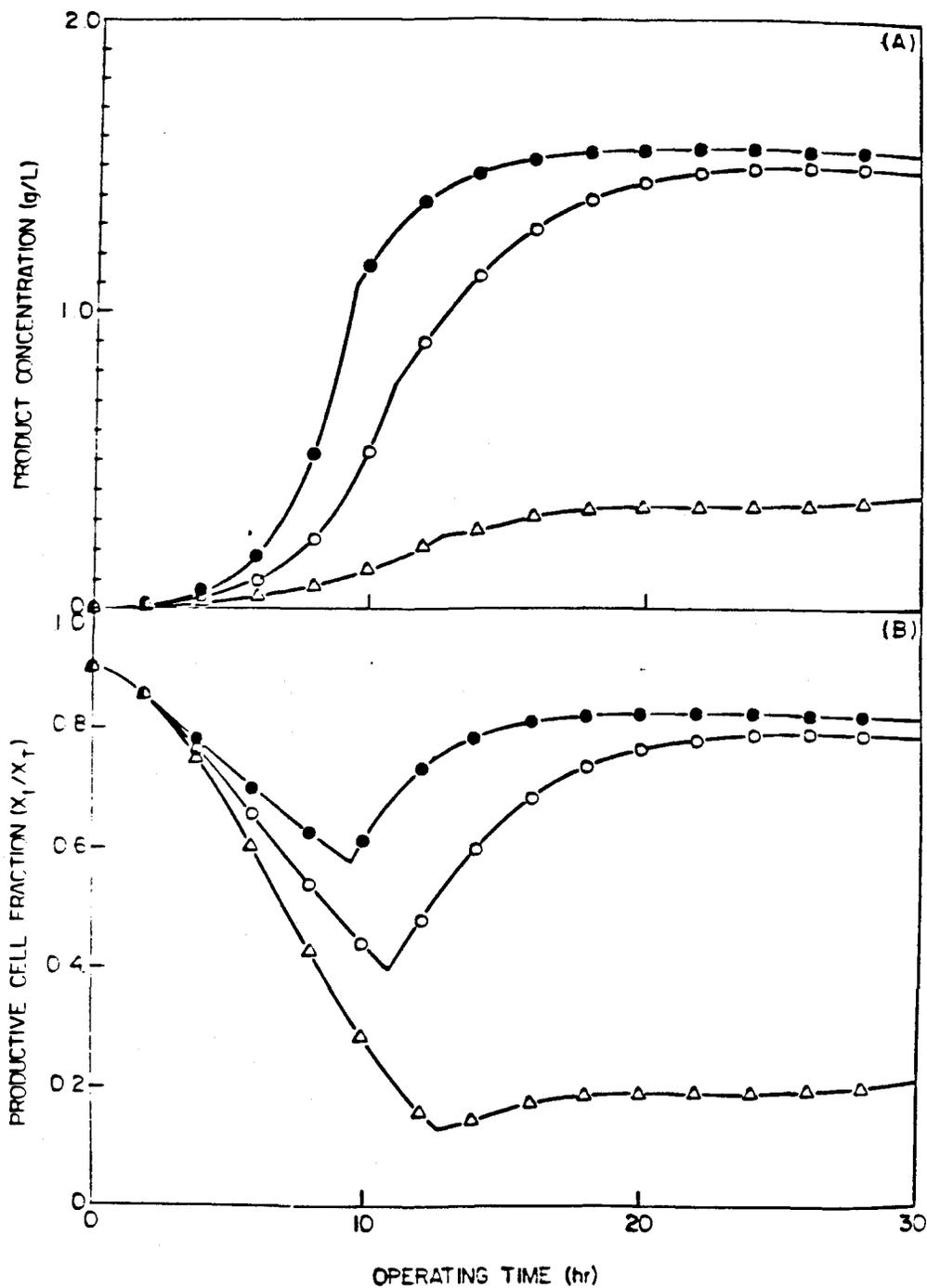


FIGURE 3.7: Calculated final product concentration vs. amplification initiation switching time for negligible product inhibition (o:  $N_{P, final} = 50$ ; ●:  $N_{P, final} = 90$ ;  $\Delta$ :  $N_{P, final} = 120$ ).



**FIGURE 3.8:** Simulated effluent product concentration and productive cell fraction for three different CSTR reactor cascades. Values are shown as functions of time since startup. Conditions in tanks 1, 2, and 3 favor cell growth, cause plasmid amplification, and induce gene expression, respectively. Different symbols correspond to different  $V_1:V_2:V_3$  tank volume ratios.  $\circ=1:1:1$ ;  $\bullet=1:1.5:3$ ;  $\Delta=3:1.5:1$ .

## **CHAPTER 4**

### **INTRACELLULAR EQUILIBRIUM CALCULATIONS BASED ON SMALL SYSTEMS THERMODYNAMICS**

## INTRODUCTION

Regulation of many processes in the living cell, such as gene expression and DNA replication, are believed to be controlled by interactions of regulatory molecules with inducers, repressors, nucleic acid sequences, protein domains, and with each other. Mathematical models of these regulatory systems have been proven quite useful for quantitative description of cell population kinetic properties (see, for example, references 1-12). In most of these model calculations, the interactions have been assumed to be in equilibrium, and equilibrium concentrations have been calculated by assuming mass-action kinetics for complex formation and dissociation. This is equivalent to writing chemical potentials for the species involved in terms of their intracellular concentrations. As noted previously by Harder and Roels<sup>7</sup> and others, the validity of this approach is questionable since the interactions considered actually occur in a collection of small closed systems, individual cells, each of which may contain a very small number of the molecules or binding sites involved in the control mechanism. On a single-cell basis, mass-action concepts which are based on properties of ensembles of large numbers of molecules are not applicable in any rigorous sense to species present in small numbers.

Binding states of such species exist only in discrete, limited numbers. For example, in a cell containing 10 molecules of species  $S_1$  which can bind to some others species, the number of bound  $S_1$  species molecules is 0, 1, 2, ..., 8, 9 or 10, and any of these binding states, and only one of them, would be observed if a single cell is examined at a particular point in time. Intracellular concentration of unbound species  $S_1$  in this case is a discrete, not a continuous variable. The "equilibrium state" of a population of such cells refers to the average of the possible single-cell states weighted by the probabilities that the molecules are found in these states. These probabilities are properly calculated according to the methods of statistical thermodynamics.

The purpose of this paper is presentation of a general algorithm for calculation of intracellular interaction equilibria in cell populations which is based on

statistical thermodynamics appropriate to the small number of molecules – intracellular situation. This algorithm is applied to a five-component model of regulation of transcription by the *lac* promoter–operator, and the results are compared to those based on the traditional single–cell equilibrium approach.

## THERMODYNAMICS OF AN ENSEMBLE OF SMALL CLOSED SYSTEMS

Consider a closed system consisting of  $n$  substances  $S_1, S_2, \dots, S_n$  that can react with each other under  $\ell$  independent reactions. Clearly  $1 \leq \ell \leq n - 1$ . Let  $\mathbf{P}_j$  denote a state of the system:

$$\mathbf{P}_j = (S_1^j, S_2^j, \dots, S_n^j) \quad (4.1)$$

where  $S_i^j$  denotes the number of molecules of substance  $S_i$  in that state. For example, states could denote unbound molecules and molecules complexed with other species. Now let  $\mathbf{R}_m$  denote one of the  $\ell$  possible reactions:

$$\mathbf{R}_m = (\nu_1^m, \nu_2^m, \dots, \nu_n^m) \quad (4.2)$$

where  $\nu_i^m$  denotes the stoichiometric coefficient of substance  $S_i$  in reaction  $\mathbf{R}_m$ .

Consider now  $\mathbf{R}_m$  to be an operator acting on  $\mathbf{P}_i$  defined as

$$\mathbf{R}_m\{\mathbf{P}_i\} = \begin{cases} (S_1^i + \nu_1^m, S_2^i + \nu_2^m, \dots, S_n^i + \nu_n^m), & \text{when } S_k^i + \nu_k^m \geq 0 \text{ for every } k, \\ \text{not defined} & \text{when there exists a value of } k \text{ such that } S_k^i + \nu_k^m < 0. \end{cases} \quad (4.3)$$

Then it is clear that the following identity holds:

$$\mathbf{R}_{m_1}\{\mathbf{R}_{m_2}\{\mathbf{P}_i\}\} = \mathbf{R}_{m_2}\{\mathbf{R}_{m_1}\{\mathbf{P}_i\}\} \quad (4.4)$$

The equilibrium of the system depends on the equilibrium of a whole set of states  $\mathbf{P}_j$  which are generated under the application of the possible reactions. Consider now two states,  $\mathbf{P}_j$  and  $\mathbf{P}_i$ , that are related to each other with a reaction  $\mathbf{R}_m$ , i.e.

$$\mathbf{P}_j = \mathbf{R}_m\{\mathbf{P}_i\} \quad (4.5)$$

or,

$$\mathbf{P}_j = (S_1^i + \nu_1^m, S_2^i + \nu_2^m, \dots, S_n^i + \nu_n^m). \quad (4.6)$$

Then the Gibbs free energy of the two states will be<sup>13</sup>:

$$G_i = \sum_{k=1}^n S_k^i \cdot \mu_k^i \quad (4.7)$$

$$G_j = \sum_{k=1}^n (S_k^i + \nu_k^m) \cdot \mu_k^j \quad (4.8)$$

where  $\mu_k^i, \mu_k^j$  are the chemical potentials (or the Gibbs free energy per molecule) of the substance  $S_k$  in the two states. These chemical potentials are given by:

$$\mu_k^i = \mu_k^0 + \frac{\hat{k}T}{S_k^i} \cdot \ln \left( \frac{(S_k^i)!}{V S_k^i} \right) \quad (4.9)$$

$$\mu_k^j = \mu_k^0 + \frac{\hat{k}T}{S_k^i + \nu_k^m} \cdot \ln \left( \frac{(S_k^i + \nu_k^m)!}{V (S_k^i + \nu_k^m)} \right) \quad (4.10)$$

where  $V$  is the volume of the system,  $\hat{k}$  is the Boltzmann's constant,  $T$  is the system's temperature, and  $\mu_k^0$  is the standard chemical potential of  $S_k$ . Then, substituting eqn. (4.9-10) into eqn. (4.7-8),

$$G_i = \sum_{k=1}^n S_k^i \cdot \mu_k^0 + \hat{k}T \cdot \ln \left( \prod_{k=1}^n (S_k^i)! \right) - \hat{k}T \cdot \left( \sum_{k=1}^n S_k^i \right) \ln V \quad (4.11)$$

$$G_j = \sum_{k=1}^n (S_k^i + \nu_k^m) \cdot \mu_k^0 + \hat{k}T \cdot \ln \left( \prod_{k=1}^n (S_k^i + \nu_k^m)! \right) - \hat{k}T \cdot \left( \sum_{k=1}^n (S_k^i + \nu_k^m) \right) \ln V \quad (4.12)$$

By subtracting now  $G_i$  from  $G_j$ ,

$$G_j - G_i = \sum_{k=1}^n \nu_k^m \mu_k^0 + \hat{k}T \cdot \ln \frac{\prod_{k=1}^n (S_k^i + \nu_k^m)!}{\prod_{k=1}^n (S_k^i)!} - \hat{k}T \cdot \left( \sum_{k=1}^n \nu_k^m \right) \ln V. \quad (4.13)$$

The first term of the right part of the above equation is easily recognized as the change in the free energy for the standard reaction  $\mathbf{R}_m$  is defined in classical thermodynamics. Accordingly,

$$\begin{aligned} \sum_{k=1}^n \nu_k^m \mu_k^0 &= \Delta G_m^0 \\ &= -\hat{k}T \cdot \ln K_m \end{aligned} \quad (4.14)$$

where  $K_m$  is the equilibrium constant of the  $\mathbf{R}_m$  reaction. If we denote now  $\hat{G}_i$  to

be  $G_i/kT$ , equation (4.13) becomes

$$\hat{G}_j = \hat{G}_i - \ln K_m + \ln \frac{\prod_{k=1}^n (S_k^i + \nu_k^m)!}{\prod_{k=1}^n (S_k^j)!} - \left( \sum_{k=1}^n \nu_k^m \right) \ln V \quad (4.15)$$

This result indicates that the Gibbs free energy of a state may be calculated knowing only the properties of its "parental" or predecessor state, and reaction data.

The overall Gibbs free energy of the closed system is:

$$\hat{G} = \sum_{i=1}^N p_i \cdot \hat{G}_i + \sum_{i=1}^N p_i \cdot \ln p_i. \quad (4.16)$$

where  $N$  is the total number of states in the system and  $p_i$  is the probability of existence of state  $\mathbf{P}_i$ . The equilibrium state of a large ensemble of such closed systems is characterized by probabilities  $p_i$  such that  $\hat{G}$  is minimized. In such a population,  $p_i$  has the alternative physical interpretation as the expected fraction of the population in state  $\mathbf{P}_i$ . Thus, the values of the probabilities  $p_i$  which minimize  $\hat{G}$  must be determined subject to the constraint

$$p_1 + p_2 + \dots + p_N = 1. \quad (4.17)$$

The solution to this problem is:

$$p_i = \frac{e^{-\Delta \hat{G}_i}}{\sum_{j=1}^N e^{-\Delta \hat{G}_j}} \quad (4.18)$$

where

$$\Delta \hat{G}_j = \hat{G}_j - \hat{G}_1 \quad (4.19)$$

The choice of the state to be named  $\mathbf{P}_1$  is arbitrary. If  $\mathbf{P}_1$  is taken to be the state that, under the action of reactions  $\mathbf{R}_m$ , does not have any "parental" states (i.e., no other states are transformed to  $\mathbf{P}_1$  under the action of any of the reactions), then  $\hat{G}_1$  can be assigned an arbitrary value. This is allowed since with  $\mathbf{P}_1$  so defined, every possible state of the system has, either explicitly (eqn. 4.3) or implicitly (eqn. 4.4),  $\mathbf{P}_1$  as its "parental" state, and so  $\hat{G}_1$  has been added to every state's free energy (eqn. 4.15).

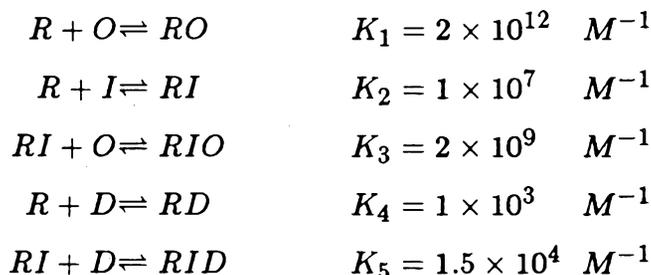
Finally, the average concentration,  $\bar{S}_k$ , of the various substances in the pop-

ulation at equilibrium can be calculated as<sup>1,14</sup>:

$$\bar{S}_k = \sum_{j=1}^N p_j \cdot S_k^j \quad (4.20)$$

## APPLICATION TO THE *lac* REPRESSOR-OPERATOR SYSTEM

The *lac* repressor-operator system can be described by the following set of reactions<sup>11,12</sup>:



where *R* denotes the *lac* repressor protein, *O* the *lac* operator, *I* the *lac* inducer (isopropyl- $\beta$ -D-thiogalactopyranoside; IPTG), *D* the nonspecific binding sites (the entire genome), *RO*, *RI*, *RIO*, *RD*, *RID* the corresponding complexes, and  $K_m$  are the corresponding association constants.

At this point, is highly convenient to eliminate the reactions involving nonspecific binding sites, because the high value of  $D_0$  (the total number of nonspecific binding sites  $\approx 10^7$  sites/cell) will otherwise lead to a numerical instability in the calculations. The elimination of the nonspecific binding of the *lac* repressor is a commonly used technique in work done on the subject<sup>15,16</sup>, but the additional complexity of the binding of *RI* to nonspecific binding sites is included in the model. The equivalent reduced system involves only the first three reactions and modified effective association constants as listed below (this reduction, although based on use of classical equilibrium relations, is justified by the high concentration of  $D_0$  which has physical meaning):

$$K'_1 = \frac{K_1}{1 + K_4 D_0} \quad (4.21)$$

$$K'_2 = \frac{K_2(1 + K_5 D_0)}{(1 + K_4 D_0)} \quad (4.22)$$

$$K_3' = \frac{K_3}{(1 + K_5 D_0)} \quad (4.23)$$

According to the previously introduced notation,  $n = 6$ ,  $\ell = 3$  for this case and

$$(S_1, S_2, S_3, S_4, S_5, S_6) \equiv (R, O, I, RO, RI, RIO)$$

$$\mathbf{R}_1 = (-1, -1, 0, 1, 0, 0)$$

$$\mathbf{R}_2 = (-1, 0, -1, 0, 1, 0)$$

$$\mathbf{R}_3 = (0, -1, 0, 0, -1, 1)$$

A suitable  $\mathbf{P}_1$  state for this system is of the form:

$$\mathbf{P}_1 = (R_0, O_0, I_0, 0, 0, 0)$$

The results of the simulation for  $\mathbf{P}_1 = (10, 1, I_0, 0, 0, 0)$  and  $\mathbf{P}_1 = (10, 15, I_0, 0, 0, 0)$ , for  $I_0$  varying from 0 molecules/cell to 10,000 molecules/cell ( $\approx 2 \times 10^{-5}$  M), are shown in Figures 4.1 and 4.2.

The *lac* repressor-operator system is mainly characterized by the ratio of free specific binding sites to total specific binding sites. Figure 4.3 shows the error in calculating this ratio by the traditional intracellular equilibrium relations versus the (total) inducer concentration. These results show errors are greatest at low inducer concentrations and approach zero as inducer concentration increases. The same trends hold with respect to effects of the number of operator sites on this error (larger errors when fewer operator sites are present). The exception to this pattern occurs when no inducer is present. Then, the error is maximized when the operator sites are equal to the *lac* repressor molecules present in the cell.

For practical situations the error introduced by application of the traditional single-cell equilibrium forms is less than 10%. This leads to the conclusion that the mass-action law is a valid approximation in cell modeling (at least in the case of the *lac* operon) since the errors involved can be considered of the same order of magnitude as the errors and uncertainties incorporated in other steps. For example, the equilibrium constants that were used in this work had been reported in the literature within an order of magnitude, which introduces an uncertainty of at least 5-10% in calculated results. Also, deviations from other assumptions in these

calculations – such as the existence of equilibrium conditions in the cell – may cause similar deviations between model and experiment.

From the practical point of view, the proposed algorithm requires much more computing time than the algorithm for solving the nonlinear algebraic system induced by the traditional single-cell equilibrium equations and mass balances (where required in this work a simple contraction method was used for the traditional approach, and quadratic convergence was always the case). Furthermore, the statistical method can result in numerical instabilities when too many states, that range from highly probable to almost improbable, are examined (a state selection is required then). It is likely that the simpler single-cell equilibrium approach is an adequate approximation in most cases and therefore preferable. When doubts arise as to the extent of the error introduced by this approximation, the algorithm presented here can be used to effect more accurate calculations for cell populations containing small numbers of interacting molecules.

**Acknowledgement:** This work was supported by the National Science Foundation.

## Nomenclature

$D$	Non specific binding sites for the <i>lac</i> repressor
$G$	Gibbs free energy
$\hat{G}$	$G/\hat{k}T$
$I$	<i>lac</i> inducer
$\hat{k}$	Boltzmann's constant
$K_i$	Equilibrium constant for reaction $i$
$\ell$	Total number of reactions
$n$	Total number of substances
$O$	<i>lac</i> operator
$p_i$	Probability of existence of a state $i$
$P_i$	Representation "vector" of state $i$
$R$	<i>lac</i> repressor protein
$R_i$	Representation "vector" of reaction $i$
$S_i$	Substance $i$
$S_i^j$	Number of molecules of substance $i$ in state $j$
$\bar{S}_i$	Average number of molecules of substance $i$ in the system
$T$	Temperature of the system
$V$	Volume of the system
	Greek Letters.
$\mu_i^j$	Chemical potential of substance $i$ in state $j$
$\mu_i^0$	Standard chemical potential of substance $i$
$\nu_i^j$	Stoichiometric coefficient of substance $i$ in reaction $j$

## References

1. G. Yagil and E. Yagil, *Biophys. J.*, **11**, 11 (1971)
2. G. van Dedem and M. Moo-Young, *Biotechnol. Bioeng.*, **17**, 1301 (1975)
3. K. Toda, *Biotechnol. Bioeng.*, **18**, 1117 (1976)
4. I. Imanaka and S. Aiba, *Biotechnol. Bioeng.*, **19**, 757 (1977)
5. S. Condo, K. Venkatasubramanian, W. R. Vieth, and A. Constantinides, *Biotechnol. Bioeng.*, **20**, 1797 (1978)
6. W. R. Vieth, K. Kaushik, and K. Venkatasubramanian, *Biotechnol. Bioeng.*, **24**, 1455 (1982)
7. A. Harder and J. A. Roels, *Adv. Bioch. Eng.*, **21**, 55 (1982).
8. S. B. Lee and J. E. Bailey, *Biotechnol. Bioeng.*, **26**, 66 (1984).
9. S. B. Lee and J. E. Bailey, *Plasmid*, **11**, 151 (1984).
10. S. B. Lee and J. E. Bailey, *Plasmid*, **11**, 166 (1984).
11. S. B. Lee and J. E. Bailey, *Biotechnol. Bioeng.*, **26**, 1372 (1984).
12. S. B. Lee and J. E. Bailey, *Biotechnol. Bioeng.*, **26**, 1383 (1984).
13. O. G. Berg and C. Blomberg, *J. Theor. Biol.*, **67**, 523 (1977).
14. G. Yagil, *Current Topics In Cellular Regulation* (Edited by B. L. Horecker, E. R. Stadtman), **9**, 220 (1975)
15. S. Lin and A. D. Riggs, *J. Theor. Biol.*, **16**, 166 (1972)
16. S. Lin and A. D. Riggs, *Cell*, **4**, 107 (1975)

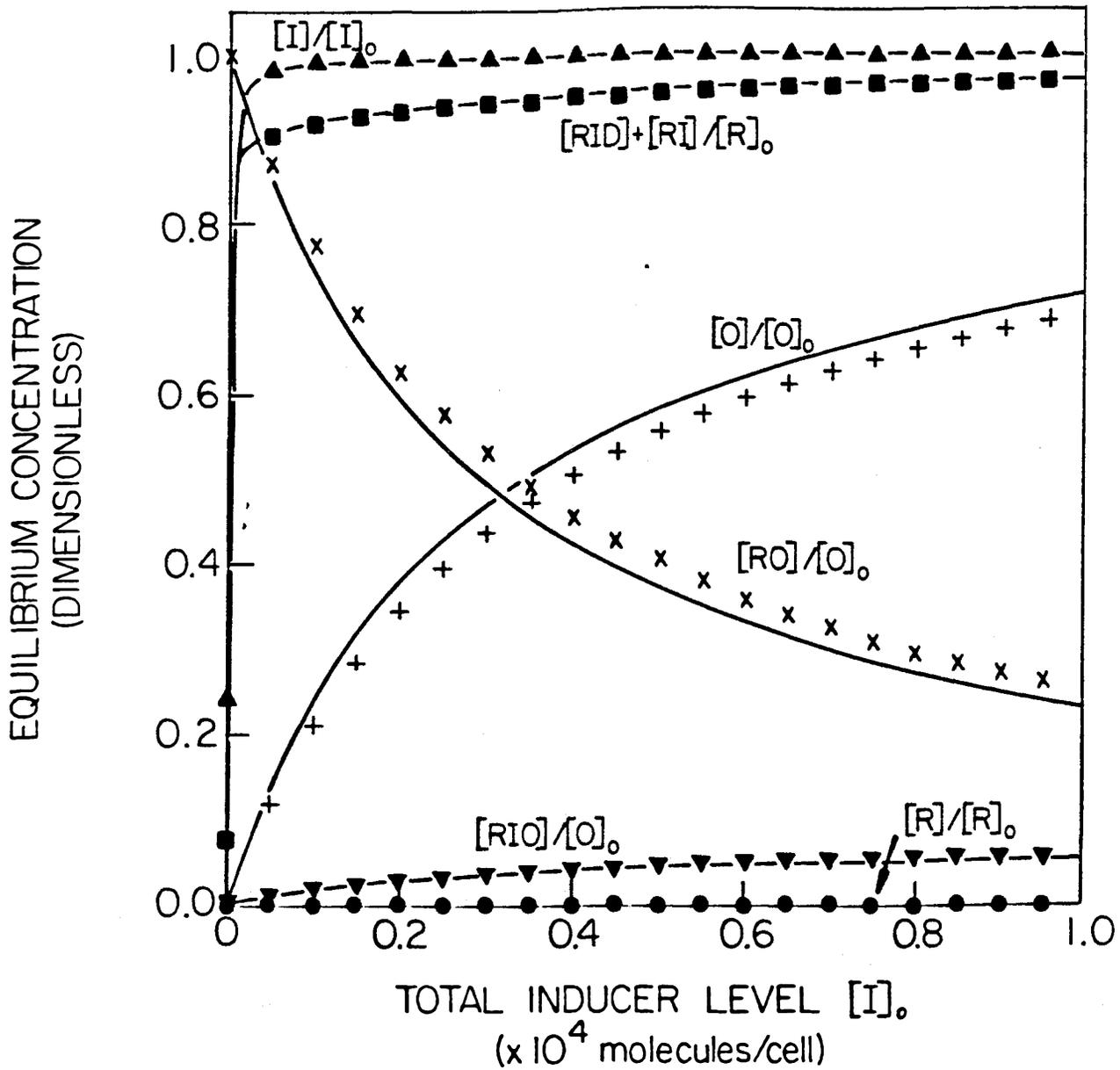


FIGURE 4.1: The results of the statistical (points) and the traditional (lines) approaches for equilibrium calculations for the *lac* repressor-operator system. Here  $[O]_0=1$  molecule/cell.

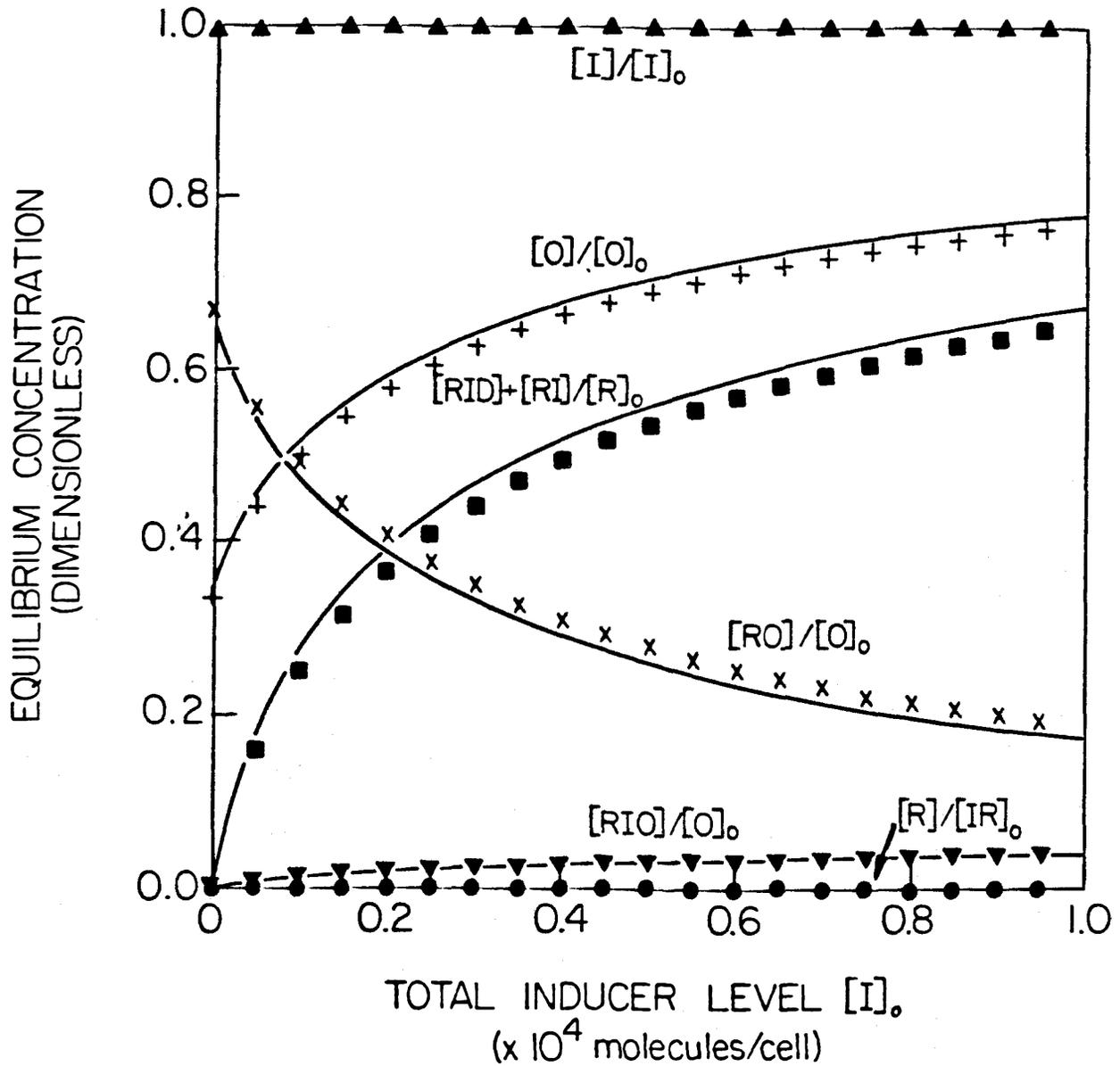


FIGURE 4.2: Equilibrium concentrations calculated for the *lac* repressor-operator system based on the statistical (points) and the traditional (lines) approaches. In this case  $[O]_0=15$  molecules/cell.

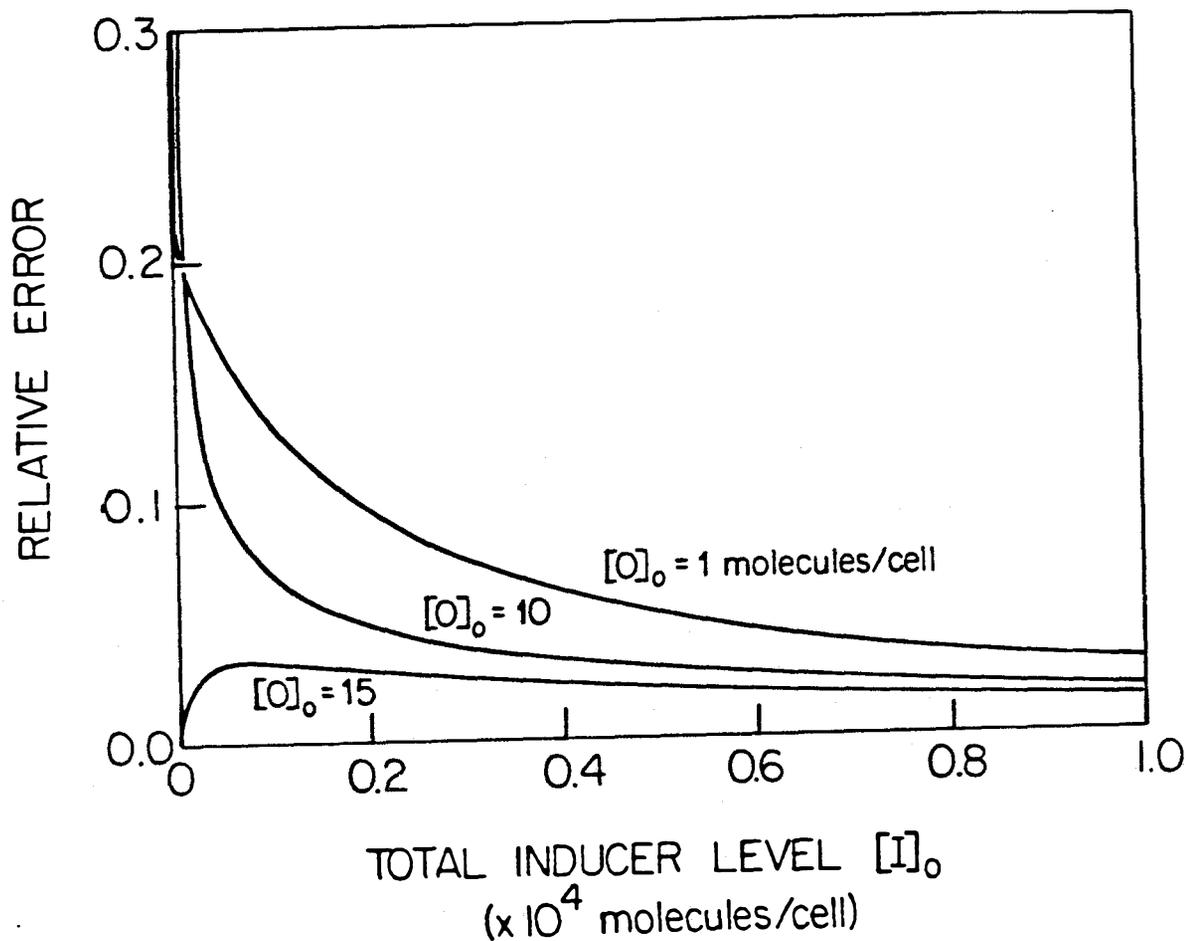


FIGURE 4.3: The relative error  $[(O_{\text{trad}} - O_{\text{stat}})/O_{\text{stat}}]$  associated with the use of the traditional single-cell intracellular equilibrium relations for the calculation of the ratio  $[O]/[O]_0$ .

## **APPENDIX A**

### **KEY FEATURES OF THE DATA BASES AND PATHWAY SYNTHESIS PROGRAMS OF MPS**

## 1 DESCRIPTION OF THE DATA BASES

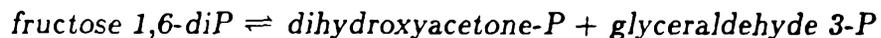
There are two data bases used by the program, one for enzymes and one for substances. Both data bases implement *structures* for storing information.

### 1.1 The Enzyme Data Base

The general representation for an enzyme in the data base is given in Figure A1. The description of the various fields is given in Table A1.

The fields that are actually used by the search algorithm are the "reaction" and "reversible" fields. The rest of the fields do not affect the progress of the search algorithm but merely provide either additional information about the enzyme (such as the "coupling reaction" field) or ways of calling up sets of pathways that fulfill some desired property. For example, the user may want to identify the subset of the synthesized pathways that contain at least one enzyme inhibited by ATP.

The characterization of reactions as reversible or irreversible under physiological conditions is based on information reported in standard biochemistry books. An alternative examined was to introduce in that field the standard Gibbs free energy change,  $\Delta G^{\circ}$ , of the reaction. Lack of data on the intracellular concentrations of the majority of intermediates is then a major problem. In order for a reaction to be characterized as reversible or not, the value of the Gibbs free energy change,  $\Delta G$ , is required.  $\Delta G$  can be calculated, given  $\Delta G^{\circ}$ , only when the concentrations of both reactants and products are known. For example consider the reaction:

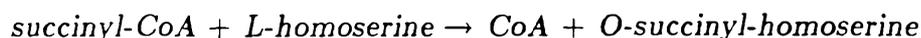


for which  $\Delta G^{\circ} = +5.7$  kcal/mol, a value high enough to assume that this reaction occurs only in the reverse direction under standard conditions ( $\Delta G = \Delta G^{\circ}$ ). Under physiological conditions, however,  $\Delta G = -0.3$  kcal/mol, so this reaction is

reversible in the cellular environment. Since the intracellular concentrations cannot be calculated or measured for the majority of metabolic intermediates, the value of  $\Delta G^{\circ}$  provides less information than the *a priori* characterization of a reaction as reversible or irreversible. When in doubt, a reaction can be characterized as reversible (at the expense of increased execution time), and possible scenarios of the form: "What if this reaction proceeds in only one direction?" can be examined after the search has been completed.

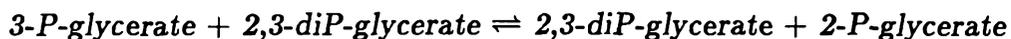
The "coupled-reaction" field serves the role of distinguishing between "important" and "non-important" substances in the reaction, from the point of view of the search algorithm. Consider for example glycolysis. Cofactors, such as ATP and NAD(P)H, are used or consumed by the glycolytic pathways. In order to maximize the efficiency of MPS for biotransformations, the search algorithm should not be concerned with the problem of how these cofactors are synthesized and how the cofactors can be converted to the target products. Instead, the algorithm should be focused on the carbon-carrying intermediates (except CO<sub>2</sub>). Thus, ATP, NAD(P)H, etc., are "non-important" reactants or products of the glycolytic pathways. The same does not hold when other areas of the metabolic network are considered. For an enzyme that belongs to the ATP biosynthetic paths, ATP is considered an "important" substance. This distinction between "important" and "non-important" substances is not always clear or possible, as, for example, in the case of a transamination reaction. Then, all the substances participating in the reaction are considered "important" (at the expense of higher execution time).

The representations of two enzymes stored in the data base are shown in Figure A2. The enzyme *homoserine acyltransferase*, (EC2.3.1.31), which is traditionally classified in the aminoacid biosynthetic pathways, catalyzes the irreversible reaction:



and is inhibited by methionine. The enzyme *phosphoglyceromutase*, (EC2.7.5.3),

which is traditionally classified in the EMP pathway, catalyzes the reversible reaction:



and does not have any inhibitors or activators.

## 1.2 The Substance Data Base

The substance data base is supportive for the enzyme data base and the search algorithm. For the representation of substances analogous data structures as the ones used for enzyme descriptions were used. The fields defined for substance descriptions are identified in Table A2.

The formal name of the substance is used by the program to refer to this substance during output. The user may refer to a substance by its formal name, or by any of its synonyms or pseudonyms. For example, although the program will always refer to *glyceraldehyde 3-phosphate* by using the formal name, the user may refer to it as "GA3P" when running the program or building an enzyme data-base. The "carrier" field contains the chemical elements that this substance carries through the metabolic network. Finally, the two last fields, which are characterized as internal in Table A2, are calculated by MPS. In Figure A3, the representations of 3-phospho-D-glyceroyl phosphate and O-phospho-L-homoserine are shown.

## 2 MPS: THE PROGRAM

### 2.1 Initialization

The first thing that MPS does is read the substance and enzyme data base. The user is asked for the name of these files, so enzymatic activities and substance descriptions can be organized in multiple data bases for particular metabolic processes or particular organisms, for example. However, since the program checks for unidentified substances, i.e. substances that are referenced in the enzyme data base but are not described in the substance data base, it is important that the enzyme

and substance data bases are compatible. Alternatively, all substance descriptions may be contained in a common data base used by every enzyme data base. The drawback of such an approach is that if, for example, a small enzyme data base is used for studying just the pathways for ethanol production in a microorganism, the full set of substances, whether relevant to the enzymes loaded or not (such as the intermediates for amino acid biosynthesis, for example), will be loaded and occupy memory without being needed.

## 2.2 The Search Algorithm

A search problem is described by an initial state, a goal defining set of criteria that the target state should satisfy, and a set of rules or operators that define the transformation of a given state to another. Consider, for example, the network shown in Figure A4. Assume that  $S$  is the initial state of the search problem. The operators are the lines connecting two nodes (e.g.,  $L$  and  $M$ ) of the network. Then a possible scenario for the search problem can be "What are the final states which can be produced starting from  $S$ ?". A final state is defined by the following property: any further application of the operators on the final state produces successor states which were already considered as parental states. In this version of the search problem, the "products" of a state are accorded primary emphasis. Most of the time, however, the produced states are equally important as the ways (paths) that were used to produce them. Consider, for example, a more restrictive goal-defining criterion: "Generate  $F$ ". One possible path solving this problem is traced in bold lines in Figure A4. It is described as a set of intermediate states connecting  $S$  to  $F$ . It is obvious that the answer ( $SOPF$ ) is not the only one that can be derived from this network. Other paths also exist. If each rule of transformation was assigned a "cost" then the goal criterion could have been formulated as "Generate the path to  $F$  that is optimal; i.e., minimizes the overall cost of the transformation."

So far the description of a search network (or search space) was based on a description of states. The underlying information in Figure A4 is the description of

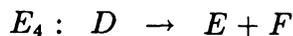
the transformation rules. The corresponding operators are listed in Table A3. The search problem in this case is defined as "Starting from a rule that transforms  $S$ , (either  $E_1$  or  $E_2$ ), end up with a rule that produces  $F$  (one of  $E_7$ ,  $E_8$ ,  $E_9$ )." The solution, ( $E_2$ ,  $E_5$ ,  $E_7$ ), is traced in bold lines in Figure A4.

The metabolic network can be viewed as a collection of operators, i.e., enzymes. Each enzyme catalyzes a reaction that can be viewed as a transformation of the substrate state to the product state. The metabolic network, however, is more complex than the picture already given. After starting from a substrate, say  $A$ , and extending a pathway, eventually an enzyme will be considered that requires multiple substrates. Consider the case in Figure A5. In order for the enzyme  $E_3$  to be incorporated in the path ( $E_1$   $E_2$ ), not only  $C$  must be produced, but  $Y$  as well. Moreover the stoichiometric ratio of  $C$  to  $Y$  is fixed by the reaction description of  $E_3$ . If  $Y$  can be produced from  $A$ , say by a single reaction catalyzed by  $E_4$  in Figure A5, then  $E_3$  can be used, and the path extension will continue. Otherwise, the path will be abandoned.

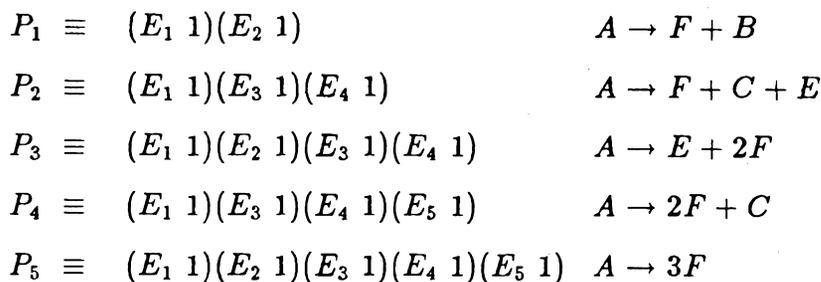
Another "peculiarity" of metabolic reaction network synthesis is introduced when substances that have been consumed at some part of the pathway are regenerated later. Consider the case shown in Figure A6. In order for the enzyme  $E_5$  to be incorporated in the extending pathway, a pathway from  $A$  to  $E$  has been added and this pathway consumed  $D$ . After the reaction catalyzed by  $E_6$  was added to the pathway, however,  $D$  has been synthesized. That means that it was not necessary to produce  $D$  directly from  $A$  in the first case. This has to be considered by the search algorithm. Before further elongation of the pathway takes place, the transformation shown in Figure A6 must take place. This transformation is not always possible considering the stoichiometric requirements of the participating enzymes, in which case it is not carried out by MPS.

The design of MPS's search algorithm is such that if the target product has been reached and carbon-containing side products have also been formed, then the search continues in order to convert the side products to the target substance. This is not

always possible. In any case, the first time that the target product is encountered in a pathway is always reported, while the subsequent steps are recorded only in the case that there is successful conversion of the side products to the end product. For example consider the following set of reactions:



Then the pathways reported by MPS will be:



Note that the pathway  $P \equiv (E_1 \ 1)(E_2 \ 1)(E_3 \ 1)$  with overall reaction  $A \rightarrow D + F$  is not reported, since it includes pathway  $P_1$  and requires additional enzymes without eliminating by-product formation.

A major restriction must be imposed in order to obtain independent pathways from MPS's search algorithm: if an enzyme requires a substrate already used by the pathway to that enzyme, then this substance must be produced by the same set of enzymes used in the pathway. To illustrate this restriction, consider the following network:



How many independent ways exist for the conversion of  $A$  to  $F$ ? There are two ways of producing  $C$ :

$$P_1 \equiv (E_1 1)(E_3 1) \quad A \rightarrow C$$

$$P_2 \equiv (E_2 1) \quad A \rightarrow C$$

and so, there are two ways of producing  $D$ :

$$P'_1 \equiv (E_1 1)(E_3 1)(E_4 1) \quad A \rightarrow D$$

$$P'_2 \equiv (E_2 1)(E_4 1) \quad A \rightarrow D$$

$E_5$  requires both  $C$  and  $D$  as substrates. So in order  $E_5$  to be incorporated in a pathway, pathways producing both  $C$  and  $D$  have to be combined. The possible combinations are  $(P_1, P'_1)$ ,  $(P_1, P'_2)$ ,  $(P_2, P'_1)$ ,  $(P_2, P'_2)$ , and the resulting pathways are:

$$P_a \equiv (E_1 2)(E_3 2)(E_4 1)(E_5 1) \quad A \rightarrow F$$

$$P_b \equiv (E_1 1)(E_2 1)(E_3 1)(E_4 1)(E_5 1) \quad A \rightarrow F$$

$$P_c \equiv (E_1 1)(E_2 1)(E_3 1)(E_4 1)(E_5 1) \quad A \rightarrow F$$

$$P_d \equiv (E_2 2)(E_4 1)(E_5 1) \quad A \rightarrow F$$

These pathways are not linearly independent, since  $P_b = P_c$ , and  $P_b = 0.5P_a + 0.5P_d$ . How about genetic independence, i.e., do  $P_a$ ,  $P_b$ ,  $P_c$ ,  $P_d$  define independent sets of enzymes? Clearly  $P_b$  and  $P_c$  do not since they are identical. Moreover, an organism that uses the pathway  $P_b$  also has the necessary enzymatic activity to carry out both  $P_a$  and  $P_d$ . That means that there are just two genetically independent pathways for transforming  $A$  to  $F$ , and MPS, by applying the restriction that substrates have to be generated in the same way throughout a pathway, will report just  $P_a$  and  $P_d$ .

The rules mentioned are part of the design of MPS's search algorithm. They have been gradually introduced in the program in order to achieve efficient management and processing handling of a large data base. Their purpose is to eliminate in early stages the combinatorial effect of having pathways constructed by just putting enzymes together. By searching for the independent pathways, instead for searching for all pathways and then extracting the minimum set of pathways that characterize the conversion, both memory and execution time requirements are drastically

reduced. If in the last example there were  $N$  pathways transforming  $A$  to  $C$  and one pathway transforming  $C$  to  $D$ , then, by not incorporating the rule that is used by MPS for this case, enough memory would need to be allocated to allow the algorithm to handle  $N \times N$  pathways. By adopting the rule, on the other hand,  $N$  pathways are tracked from the beginning of the search.

<b>FIELD</b>	<b>DESCRIPTION</b>
<b>Code</b>	The EC number of the enzyme as it appears in Enzyme Nomenclature.
<b>Name</b>	The name of the enzyme. It can either be the one recommended by the Nomenclature Committee of the International Union of Biochemistry, or the systematic name based on the reaction catalyzed.
<b>Reaction</b>	The reaction catalyzed by the enzyme. In this field only the substances associated with the search algorithm appear.
<b>Reversible</b>	If the reaction catalyzed is reversible under physiological conditions, then <b>T</b> appears in this field; otherwise, <b>NIL</b> appears.
<b>Coupling-Reaction</b>	The reaction coupled to the main reaction, involving substances that should not be involved in the search.
<b>Inhibited-by</b>	Reported inhibitors of the enzyme.
<b>Activated-by</b>	Reported activators of the enzyme.
<b>Pathways</b>	The pathways in which the enzyme is traditionally classified.

**TABLE A1:** The fields used for enzyme descriptions.

<b>FIELD</b>	<b>DESCRIPTION</b>
<b>Name</b>	The formal name of the substance.
<b>AKA</b>	Other names or pseudonyms of the substance.
<b>Carrier</b>	The elements carried by this substance through the metabolic network.
<b>Consumed-by (internal)</b>	Enzyme nodes that use this substance as a substrate.
<b>Produced-by (internal)</b>	Enzymes nodes that produce this substance.

**TABLE A2:** The fields used for substance descriptions.

Name	Rule
$E_1$	$S \rightleftharpoons L$
$E_2$	$S \rightleftharpoons O$
$E_3$	$L \rightleftharpoons M$
$E_4$	$L \rightleftharpoons F$
$E_5$	$O \rightleftharpoons P$
$E_6$	$O \rightleftharpoons Q$
$E_7$	$P \rightleftharpoons F$
$E_8$	$Q \rightleftharpoons F$
$E_9$	$M \rightleftharpoons N$

**TABLE A3:** The rules involved in the network of Figure A4.

```
(ENZYME :CODE enzyme-code
        :NAME enzyme-name
        :REACTION ((substance coefficient)
                  (substance coefficient)
                  ...
                  )
        :REVERSIBLE T or NIL
        :COUPLING-REACTION ((substance coefficient)
                             (substance coefficient)
                             ...
                             )
        :INHIBITED-BY (substance ... )
        :ACTIVATED-BY (substance ... )
        :PATHWAYS (pathway ... ))
```

**FIGURE A1:** The general representation of enzymes in the data base.

(ENZYME :CODE (2 3 1 31)  
:NAME "HOMOSERINE ACYL TRANSFERASE"  
:REACTION (("SUCCINYL CoA" -1)  
("L-HOMOSERINE" -1)  
("O-SUCCINYL-HOMOSERINE" 1))  
:REVERSIBLE NIL  
:COUPLING-REACTION (("CoA" 1))  
:INHIBITED-BY ("METHIONINE")  
:PATHWAYS ("Amino Acid Biosynthesis"))

(ENZYME :CODE (2 7 5 3)  
:NAME "PHOSPHOGLYCEROMUTASE"  
:REACTION (("3-PHOSPHO-D-GLYCERATE" -1)  
("2-PHOSPHO-D-GLYCERATE" 1))  
:REVERSIBLE T  
:COUPLING-REACTION (("2,3 bi-P-D-GLYCERATE" -1)  
("2,3 bi-P-D-GLYCERATE" 1))  
:PATHWAYS ("EMP"))

FIGURE A2: The representation of *homoserine acyltransferase* and *phosphoglyceromutase* in the enzyme data base.

```
(SUBSTANCE :NAME "3-PHOSPHO-D-GLYCEROYL PHOSPHATE"  
:AKA ("1,3 bi-P-GLYCERATE"  
"1,3 bi-P-D-GLYCERATE"  
"1,3 DPG"  
"D-3-P-GLYCEROYL-P")  
:CARRIER (C P))
```

```
(SUBSTANCE :NAME "O-PHOSPHO-L-HOMOSERINE"  
:AKA ("HOMOSERINE PHOSPHATE"  
"P-HOMOSERINE")  
:CARRIER (C P N))
```

**FIGURE A3:** The representation of 3-phospho-D-glyceroyl phosphate and O-phospho-L-homoserine in the substance data base.

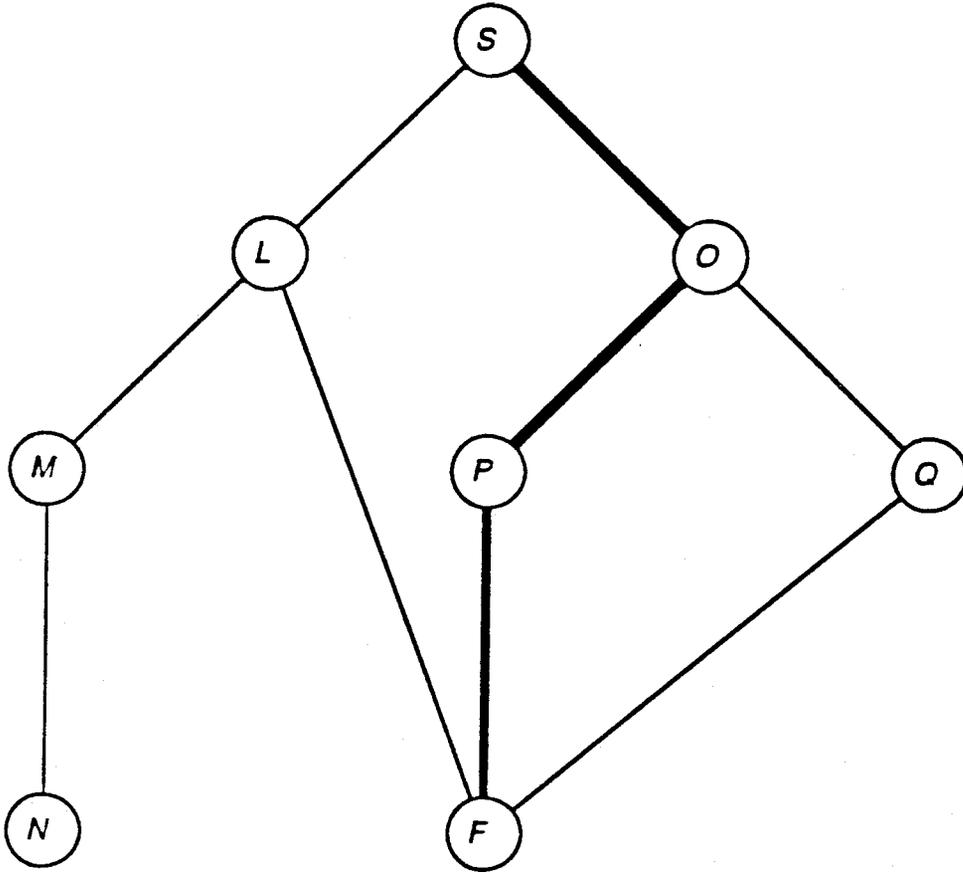
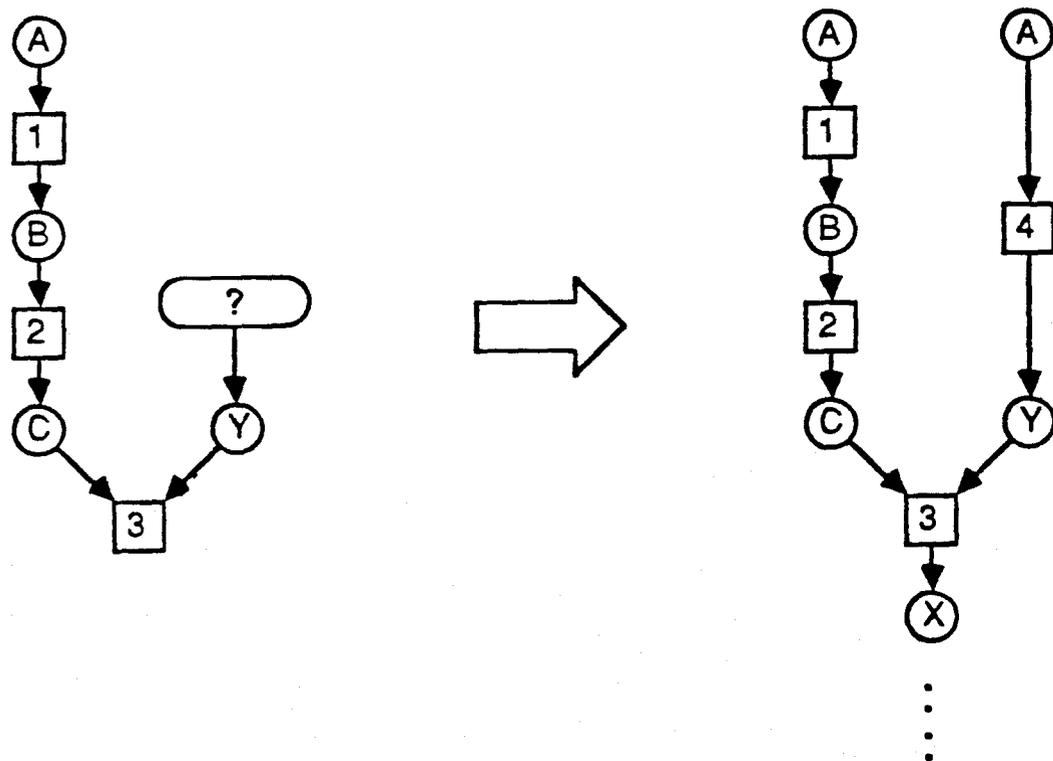
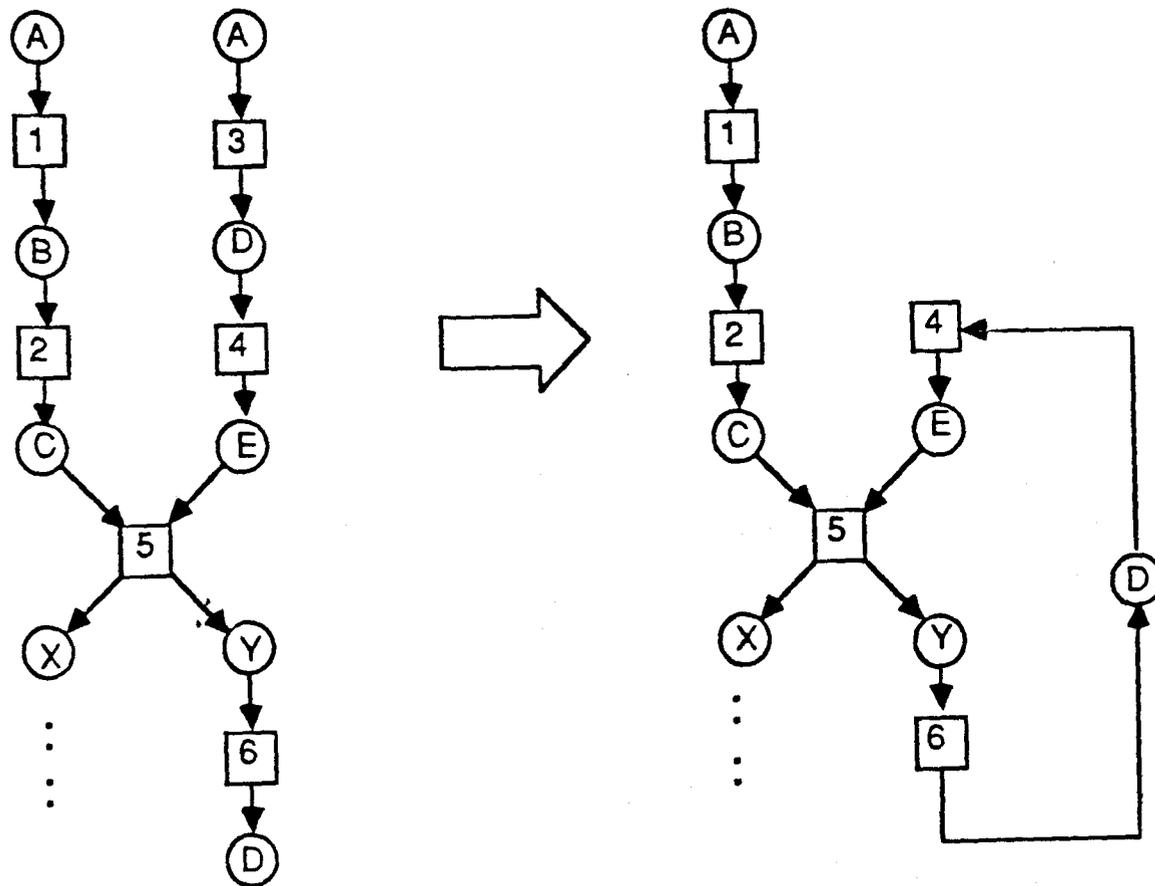


FIGURE A4: A simple search network.



**FIGURE A5:** Reaction networks incorporate multiple linear pathways in order to provide the necessary substances for reactions requiring more than one reactant.



**FIGURE A6:** Reaction networks incorporate cyclic pathways when substances that have been already consumed are regenerated.

## **APPENDIX B**

**THE ENZYME DATA BASE USED TO OBTAIN  
THE RESULTS OF CHAPTER 1**

```
#s(ENZYME :CODE      (5 3 1 9)
:NAME        "GLUCOSEPHOSPHATE ISOMERASE"
:REACTION    (("D-GLUCOSE 6-PHOSPHATE" -1)
              ("D-FRUCTOSE 6-PHOSPHATE" 1))
:REVERSIBLE  T
:COUPLING-REACTION  ()
:INHIBITED-BY  ()
:ACTIVATED-BY  ()
:PATHWAYS    ("EMP"))
#s(ENZYME :CODE      (2 7 1 11)
:NAME        "6-PHOSPHOFRUCTOKINASE"
:REACTION    (("D-FRUCTOSE 6-PHOSPHATE" -1)
              ("D-FRUCTOSE 1,6-BIPHOSPHATE" 1))
:REVERSIBLE  NIL
:COUPLING-REACTION  (("ATP" -1)
                    ("ADP" 1))
:INHIBITED-BY  ("ATP"
                "CITRATE"
                "Mg2+")
:ACTIVATED-BY  ("NH3"
                "K+"
                "ORTHOPHOSPHATE"
                "ADP"
                "3'-AMP"
                "5'-AMP"
                "D-FRUCTOSE 6-PHOSPHATE")
:PATHWAYS    ("EMP"))
#s(ENZYME :CODE      (4 1 2 13)
:NAME        "FRUCTOSE-BIPHOSPHATE ALDOLASE"
:REACTION    (("D-FRUCTOSE 1,6-BIPHOSPHATE" -1)
              ("DIHYDROXYACETONE PHOSPHATE" 1)
              ("D-GLYCERALDEHYDE 3-PHOSPHATE" 1))
:REVERSIBLE  T
:COUPLING-REACTION  ()
:INHIBITED-BY  ()
:ACTIVATED-BY  ()
:PATHWAYS    ("EMP"))
#s(ENZYME :CODE      (5 3 1 1)
:NAME        "TRIOSEPHOSPHATE ISOMERASE"
:REACTION    (("DIHYDROXYACETONE PHOSPHATE" -1)
              ("D-GLYCERALDEHYDE 3-PHOSPHATE" 1))
:REVERSIBLE  T
:COUPLING-REACTION  ()
:INHIBITED-BY  ()
:ACTIVATED-BY  ()
:PATHWAYS    ("EMP"))
#s(ENZYME :CODE      (1 2 1 12)
:NAME        "GLYCERALDEHYDE-PHOSPHATE DEHYDROGENASE"
:REACTION    (("D-GLYCERALDEHYDE 3-PHOSPHATE" -1)
              ("3-PHOSPHO-D-GLYCEROYL PHOSPHATE" 1))
:REVERSIBLE  NIL
:COUPLING-REACTION  (("NAD+" -1)
                    ("ORTHOPHOSPHATE" -1)
                    ("NADH" 1))
```

```
: INHIBITED-BY      ("ATP"  
                    "L-SERINE")  
: ACTIVATED-BY     (  
: PATHWAYS         ("EMP"))  
#s( ENZYME : CODE   (2 7 2 3)  
: NAME            "PHOSPHOGLYCERATE KINASE"  
: REACTION        (( "3-PHOSPHO-D-GLYCEROYL PHOSPHATE" -1)  
                  ("3-PHOSPHO-D-GLYCERATE" 1))  
: REVERSIBLE      T  
: COUPLING-REACTION (( "ADP" -1)  
                   ("ATP" 1))  
: INHIBITED-BY     (  
: ACTIVATED-BY     (  
: PATHWAYS         ("EMP"))  
#s( ENZYME : CODE   (2 7 5 3)  
: NAME            "PHOSPHOGLYCEROMUTASE"  
: REACTION        (( "3-PHOSPHO-D-GLYCERATE" -1)  
                  ("2-PHOSPHO-D-GLYCERATE" 1))  
: REVERSIBLE      T  
: COUPLING-REACTION (( "2,3-BIPHOSPHO-D-GLYCERATE" -1)  
                   ("2,3-BIPHOSPHO-D-GLYCERATE" 1))  
: INHIBITED-BY     (  
: ACTIVATED-BY     (  
: PATHWAYS         ("EMP"))  
#s( ENZYME : CODE   (4 2 1 11)  
: NAME            "ENOLASE"  
: REACTION        (( "2-PHOSPHO-D-GLYCERATE" -1)  
                  ("PHOSPHOENOLPYRUVATE" 1))  
: REVERSIBLE      T  
: COUPLING-REACTION (( "H2O" 1))  
: INHIBITED-BY     (  
: ACTIVATED-BY     (  
: PATHWAYS         ("EMP"))  
#s( ENZYME : CODE   (2 7 1 40)  
: NAME            "PYRUVATE KINASE"  
: REACTION        (( "PHOSPHOENOLPYRUVATE" -1)  
                  ("PYRUVATE" 1))  
: REVERSIBLE      NIL  
: COUPLING-REACTION (( "ADP" -1)  
                   ("ATP" 1))  
: INHIBITED-BY     (  
: ACTIVATED-BY     (  
: PATHWAYS         ("EMP"))  
#s( ENZYME : CODE   (9 9 9 1)  
: NAME            "PYRUVATE DEHYDROGENASE COMPLEX"  
: REACTION        (( "PYRUVATE" -1)  
                  ("ACETYL-CoA" 1))  
: REVERSIBLE      NIL  
: COUPLING-REACTION (( "NAD+" -1)  
                   ("CoA" -1)  
                   ("CO2" 1)  
                   ("NADH" 1))  
: INHIBITED-BY     ("ATP"  
                    "ACETYL-CoA"
```

```

                                "NADH" )
:ACTIVATED-BY                  ( "ADP"
                                "PYRUVATE"
                                "Ca2+" )
:s(ENZYME :PATHWAYS            ( "EMP" ))
:CODE                          ( 4 1 3 7 )
:NAME                          "CITRATE SYNTHASE"
:REACTION                      ( ( "ACETYL-CoA" -1)
                                ( "OXALOACETATE" -1)
                                ( "CITRATE" 1) )
:REVERSIBLE                    T
:COUPLING-REACTION            ( ( "H2O" -1)
                                ( "CoA" 1) )
:INHIBITED-BY                 ( "ATP" )
:ACTIVATED-BY                 ( )
:s(ENZYME :PATHWAYS            ( "TCA" ))
:CODE                          ( 4 2 1 3 )
:NAME                          "ACONITATE HYDRATASE"
:REACTION                      ( ( "CITRATE" -1)
                                ( "threo-Ds-ISOCITRATE" 1) )
:REVERSIBLE                    T
:COUPLING-REACTION            ( ( "H2O" 1) )
:INHIBITED-BY                 ( )
:ACTIVATED-BY                 ( )
:s(ENZYME :PATHWAYS            ( "TCA" ))
:CODE                          ( 1 1 1 42 )
:NAME                          "ISOCITRATE DEHYDROGENASE"
:REACTION                      ( ( "threo-Ds-ISOCITRATE" -1)
                                ( "2-OXOGLUTARATE" 1) )
:REVERSIBLE                    T
:COUPLING-REACTION            ( ( "NADP+" -1)
                                ( "CO2" 1)
                                ( "NADPH" 1) )
:INHIBITED-BY                 ( )
:ACTIVATED-BY                 ( "ADP" )
:s(ENZYME :PATHWAYS            ( "TCA" ))
:CODE                          ( 9 9 9 2 )
:NAME                          "a-KETOGLUTARATE DEHYDROGENASE COMPLEX"
:REACTION                      ( ( "2-OXOGLUTARATE" -1)
                                ( "SUCCINYL-CoA" 1) )
:REVERSIBLE                    T
:COUPLING-REACTION            ( ( "NAD+" -1)
                                ( "CO2" 1)
                                ( "CoA" -1)
                                ( "NADH" 1) )
:INHIBITED-BY                 ( "SUCCINYL-CoA"
                                "NADH" )
:ACTIVATED-BY                 ( )
:s(ENZYME :PATHWAYS            ( "TCA" ))
:CODE                          ( 6 2 1 4 )
:NAME                          "SUCCINYL-CoA SYNTHETASE"
:REACTION                      ( ( "SUCCINATE" -1)
                                ( "SUCCINYL-CoA" 1) )
:REVERSIBLE                    T
```

```
: COUPLING-REACTION (( "GTP" -1)
                    ("CoA" -1)
                    ("GDP" 1)
                    ("ORTHOPHOSPHATE" 1))
: INHIBITED-BY      ()
: ACTIVATED-BY      ()
: PATHWAYS          ("TCA"))
#s(ENZYME : CODE      (1 3 99 1)
: NAME             "SUCCINATE DEHYDROGENASE"
: REACTION         (( "SUCCINATE" -1)
                    ("FUMARATE" 1))

: REVERSIBLE       T
: COUPLING-REACTION (( "FAD" -1)
                    ("FADH2" 1))

: INHIBITED-BY      ()
: ACTIVATED-BY      ()
: PATHWAYS          ("TCA"))
#s(ENZYME : CODE      (4 2 1 2)
: NAME             "FUMARATE HYDRATASE"
: REACTION         (( "L-MALATE" -1)
                    ("FUMARATE" 1))

: REVERSIBLE       T
: COUPLING-REACTION (( "H2O" 1))
: INHIBITED-BY      ()
: ACTIVATED-BY      ()
: PATHWAYS          ("TCA"))
#s(ENZYME : CODE      (1 1 1 37)
: NAME             "MALATE DEHYDROGENASE (1 1 1 37)"
: REACTION         (( "L-MALATE" -1)
                    ("OXALOACETATE" 1))

: REVERSIBLE       T
: COUPLING-REACTION (( "NAD+" -1)
                    ("NADH" 1))

: INHIBITED-BY      ()
: ACTIVATED-BY      ()
: PATHWAYS          ("TCA"))
#s(ENZYME : CODE      (4 1 3 8)
: NAME             "ATP-CITRATE LYASE"
: REACTION         (( "CITRATE" -1)
                    ("ACETYL-CoA" 1)
                    ("OXALOACETATE" 1))

: REVERSIBLE       NIL
: COUPLING-REACTION (( "ATP" -1)
                    ("CoA" -1)
                    ("ADP" 1)
                    ("ORTHOPHOSPHATE" 1))

: INHIBITED-BY      ()
: ACTIVATED-BY      ()
: PATHWAYS          ("TCA"))
#s(ENZYME : CODE      (6 4 1 1)
: NAME             "PYRUVATE CARBOXYLASE"
: REACTION         (( "PYRUVATE" -1)
                    ("OXALOACETATE" 1))

: REVERSIBLE       NIL
```

```
: COUPLING-REACTION (( "ATP" -1)
                    ("CO2" -1)
                    ("H2O" -1)
                    ("ADP" 1)
                    ("ORTHOPHOSPHATE" 1))
: INHIBITED-BY      ()
: ACTIVATED-BY     ()
: PATHWAYS         ("GOC"))
#s(ENZYME : CODE    (1 1 1 40)
: NAME           "MALATE DEHYDROGENASE (1 1 1 40)"
: REACTION       (( "L-MALATE" -1)
                  ("PYRUVATE" 1))
: REVERSIBLE     T
: COUPLING-REACTION (( "NADP+" -1)
                    ("CO2" 1)
                    ("NADPH" 1))
: INHIBITED-BY   ()
: ACTIVATED-BY   ()
: PATHWAYS       ("GOC"))
#s(ENZYME : CODE    (4 1 3 1)
: NAME           "ISOCITRATE LYASE"
: REACTION       (( "threo-Ds-ISOCITRATE" -1)
                  ("SUCCINATE" 1)
                  ("GLYOXYLATE" 1))
: REVERSIBLE     T
: COUPLING-REACTION ()
: INHIBITED-BY   ()
: ACTIVATED-BY   ()
: PATHWAYS       ("GOC"))
#s(ENZYME : CODE    (4 1 3 2)
: NAME           "MALATE SYNTHASE"
: REACTION       (( "L-MALATE" -1)
                  ("ACETYL CoA" 1)
                  ("GLYOXYLATE" 1))
: REVERSIBLE     T
: COUPLING-REACTION (( "CoA" -1)
                    ("H2O" 1))
: INHIBITED-BY   ()
: ACTIVATED-BY   ()
: PATHWAYS       ("GOC"))
#s(ENZYME : CODE    (1 1 1 49)
: NAME           "GLUCOSE-6-PHOSPHATE DEHYDROGENASE"
: REACTION       (( "D-GLUCOSE 6-PHOSPHATE" -1)
                  ("D-GLUCONO-d-LACTONE 6-PHOSPHATE" 1))
: REVERSIBLE     NIL
: COUPLING-REACTION (( "NADP+" -1)
                    ("NADPH" 1))
: INHIBITED-BY   ()
: ACTIVATED-BY   ()
: PATHWAYS       ("PHP"))
#s(ENZYME : CODE    (3 1 1 17)
: NAME           "GLUCONOLACTONASE"
: REACTION       (( "D-GLUCONO-d-LACTONE 6-PHOSPHATE" -1)
                  ("6-PHOSPHO-D-GLUCONATE" 1))
```

```
:REVERSIBLE          NIL
:COUPLING-REACTION  (( "H2O" -1))
:INHIBITED-BY       ( )
:ACTIVATED-BY       ( )
:PATHWAYS           ( "PHP" ) )
#s(ENZYME :CODE      (4 2 1 12)
:NAME              "PHOSPHOGLUCONATE DEHYDRATASE"
:REACTION          ( ( "6-PHOSPHO-D-GLUCONATE" -1)
                    ( "6-PHOSPHO-2-KETO-3-DEOXY-D-GLUCONATE"

:REVERSIBLE          NIL
:COUPLING-REACTION  (( "H2O" 1))
:INHIBITED-BY       ( )
:ACTIVATED-BY       ( )
:PATHWAYS           ( "PHP" ) )
#s(ENZYME :CODE      (1 1 1 44)
:NAME              "PHOSPHOGLUCONATE DEHYDROGENASE"
:REACTION          ( ( "6-PHOSPHO-D-GLUCONATE" -1)
                    ( "D-RIBULOSE 5-PHOSPHATE" 1))

:REVERSIBLE          NIL
:COUPLING-REACTION  (( "NADP+" -1)
                    ( "CO2" 1)
                    ( "NADPH" 1))
:INHIBITED-BY       ( )
:ACTIVATED-BY       ( )
:PATHWAYS           ( "PHP" ) )
#s(ENZYME :CODE      (5 1 3 1)
:NAME              "RIBULOSEPHOSPHATE 3-EPIMERASE"
:REACTION          ( ( "D-RIBULOSE 5-PHOSPHATE" -1)
                    ( "D-XYLULOSE 5-PHOSPHATE" 1))

:REVERSIBLE          T
:COUPLING-REACTION  ( )
:INHIBITED-BY       ( )
:ACTIVATED-BY       ( )
:PATHWAYS           ( "PHP" ) )
#s(ENZYME :CODE      (5 3 1 6)
:NAME              "RIBOSEPHOSPHATE ISOMERASE"
:REACTION          ( ( "D-RIBULOSE 5-PHOSPHATE" -1)
                    ( "D-RIBOSE 5-PHOSPHATE" 1))

:REVERSIBLE          T
:COUPLING-REACTION  ( )
:INHIBITED-BY       ( )
:ACTIVATED-BY       ( )
:PATHWAYS           ( "PHP" ) )
#s(ENZYME :CODE      (2 2 1 1)
:NAME              "TRANSKETOLASE-1"
:REACTION          ( ( "SEDOHEPTULOSE 7-PHOSPHATE" -1)
                    ( "D-GLYCERALDEHYDE 3-PHOSPHATE" -1)
                    ( "D-RIBOSE 5-PHOSPHATE" 1)
                    ( "D-XYLULOSE 5-PHOSPHATE" 1))

:REVERSIBLE          T
:COUPLING-REACTION  ( )
:INHIBITED-BY       ( )
:ACTIVATED-BY       ( )
:PATHWAYS           ( "PHP" ) )
```

```
#s(ENZYME :CODE      (2 2 1 2)
          :NAME      "TRANSALDOLASE"
          :REACTION  (( "SEDOHEPTULOSE 7-PHOSPHATE" -1)
                    ( "D-GLYCERALDEHYDE 3-PHOSPHATE" -1)
                    ( "D-ERYTHROSE 4-PHOSPHATE" 1)
                    ( "D-FRUCTOSE 6-PHOSPHATE" 1))
          :REVERSIBLE T
          :COUPLING-REACTION ()
          :INHIBITED-BY  ()
          :ACTIVATED-BY  ()
          :PATHWAYS     ("PHP"))
#s(ENZYME :CODE      (2 2 1 1)
          :NAME      "TRANSKETOLASE-2"
          :REACTION  (( "D-XYLULOSE 5-PHOSPHATE" -1)
                    ( "D-ERYTHROSE 4-PHOSPHATE" -1)
                    ( "D-GLYCERALDEHYDE 3-PHOSPHATE" 1)
                    ( "D-FRUCTOSE 6-PHOSPHATE" 1))
          :REVERSIBLE T
          :COUPLING-REACTION ()
          :INHIBITED-BY  ()
          :ACTIVATED-BY  ()
          :PATHWAYS     ("PHP"))
#s(ENZYME :CODE      (4 1 2 14)
          :NAME      "PHOSPHO-2-KETO-3-DEOXY-GLUCONATE ALDOLAS
          :REACTION  (( "6-PHOSPHO-2-KETO-3-DEOXY-D-GLUCONATE"
                    ( "PYRUVATE" 1)
                    ( "D-GLYCERALDEHYDE 3-PHOSPHATE" 1))
          :REVERSIBLE NIL
          :COUPLING-REACTION ()
          :INHIBITED-BY  ()
          :ACTIVATED-BY  ()
          :PATHWAYS     ("PHP"))
#s(ENZYME :CODE      (1 4 1 4)
          :NAME      "L-GLUTAMATE DEHYDROGENASE (NADP+)"
          :REACTION  (( "L-GLUTAMATE" -1)
                    ( "2-OXOGLUTARATE" 1))
          :REVERSIBLE T
          :COUPLING-REACTION (( "NADP+" -1)
                             ( "H2O" -1)
                             ( "NH3" 1)
                             ( "NADPH" 1))
          :INHIBITED-BY  ("GTP")
          :ACTIVATED-BY  ()
          :PATHWAYS     ("AAB"))
#s(ENZYME :CODE      (6 3 1 2)
          :NAME      "GLUTAMINE SYNTHETASE"
          :REACTION  (( "L-GLUTAMATE" -1)
                    ( "L-GLUTAMINE" 1))
          :REVERSIBLE NIL
          :COUPLING-REACTION (( "ATP" -1)
                             ( "NH3" -1)
                             ( "ADP" 1)
                             ( "ORTHOPHOSPHATE" 1))
          :INHIBITED-BY  ()
```

```
:ACTIVATED-BY      ( )
:PATHWAYS          ("AAB"))
#s(ENZYME :CODE    (1 4 1 13)
:NAME             "GLUTAMATE SYNTHASE (NADPH)"
:REACTION        (( "L-GLUTAMATE" -2)
                  ( "L-GLUTAMINE" 1)
                  ( "2-OXOGLUTARATE" 1))

:REVERSIBLE       T
:COUPLING-REACTION (( "NADP+" -1)
                   ( "NADPH" 1))

:INHIBITED-BY    ( )
:ACTIVATED-BY    ( )
:PATHWAYS        ("AAB"))
#s(ENZYME :CODE    (2 6 1 2)
:NAME             "ALANINE AMINOTRANSFERASE"
:REACTION        (( "L-ALANINE" -1)
                  ( "2-OXOGLUTARATE" -1)
                  ( "PYRUVATE" 1)
                  ( "L-GLUTAMATE" 1))

:REVERSIBLE       T
:COUPLING-REACTION ( )
:INHIBITED-BY    ( )
:ACTIVATED-BY    ( )
:PATHWAYS        ("AAB"))
#s(ENZYME :CODE    (2 6 1 1)
:NAME             "ASPARTATE AMINOTRANSFERASE"
:REACTION        (( "L-ASPARTATE" -1)
                  ( "2-OXOGLUTARATE" -1)
                  ( "OXALOACETATE" 1)
                  ( "L-GLUTAMATE" 1))

:REVERSIBLE       T
:COUPLING-REACTION ( )
:INHIBITED-BY    ( )
:ACTIVATED-BY    ( )
:PATHWAYS        ("AAB"))
#s(ENZYME :CODE    (6 3 1 4)
:NAME             "ASPARAGINE SYNTHETASE"
:REACTION        (( "L-ASPARTATE" -1)
                  ( "L-ASPARAGINE" 1))

:REVERSIBLE       NIL
:COUPLING-REACTION (( "ATP" -1)
                   ( "NH3" -1)
                   ( "ADP" 1)
                   ( "ORTHOPHOSPHATE" 1))

:INHIBITED-BY    ( )
:ACTIVATED-BY    ( )
:PATHWAYS        ("AAB"))
#s(ENZYME :CODE    (2 7 2 4)
:NAME             "ASPARTATE KINASE"
:REACTION        (( "L-ASPARTATE" -1)
                  ( "4-PHOSPHO-L-ASPARTATE" 1))

:REVERSIBLE       NIL
:COUPLING-REACTION (( "ATP" -1)
                   ( "ADP" 1))
```

```

: INHIBITED-BY      ("THREONINE" )
: ACTIVATED-BY      ( )
: PATHWAYS          ("AAB" )
#s(ENZYME : CODE      (1 2 1 11)
: NAME              "ASPARTATE-SEMIALDEHYDE DEHYDROGENASE"
: REACTION          (( "L-ASPARTATE b-SEMIALDEHYDE" -1)
                    ("4-PHOSPHO-L-ASPARTATE" 1))

: REVERSIBLE        T
: COUPLING-REACTION (( "ORTHOPHOSPHATE" -1)
                    ("NADP+" -1)
                    ("NADPH" 1))

: INHIBITED-BY      ( )
: ACTIVATED-BY      ( )
: PATHWAYS          ("AAB" )
#s(ENZYME : CODE      (1 1 1 3)
: NAME              "HOMOSERINE DEHYDROGENASE"
: REACTION          (( "L-HOMOSERINE" -1)
                    ("L-ASPARTATE b-SEMIALDEHYDE" 1))

: REVERSIBLE        T
: COUPLING-REACTION (( "NADP+" -1)
                    ("NADPH" 1))

: INHIBITED-BY      ( )
: ACTIVATED-BY      ( )
: PATHWAYS          ("AAB" )
#s(ENZYME : CODE      (2 7 1 39)
: NAME              "HOMOSERINE KINASE"
: REACTION          (( "L-HOMOSERINE" -1)
                    ("O-PHOSPHO-L-HOMOSERINE" 1))

: REVERSIBLE        NIL
: COUPLING-REACTION (( "ATP" -1)
                    ("ADP" 1))

: INHIBITED-BY      ( )
: ACTIVATED-BY      ( )
: PATHWAYS          ("AAB" )
#s(ENZYME : CODE      (4 2 99 2)
: NAME              "THREONINE SYNTHASE"
: REACTION          (( "O-PHOSPHO-L-HOMOSERINE" -1)
                    ("THREONINE" 1))

: REVERSIBLE        NIL
: COUPLING-REACTION (( "H2O" -1)
                    ("ORTHOPHOSPHATE" 1))

: INHIBITED-BY      ( )
: ACTIVATED-BY      ( )
: PATHWAYS          ("AAB" )
#s(ENZYME : CODE      ( )
: NAME              "GLUTAMATE KINASE"
: REACTION          (( "L-GLUTAMATE" -1)
                    ("2-PYRROLIDONE 5-CARBOXYLATE" 1))

: REVERSIBLE        NIL
: COUPLING-REACTION (( "ATP" -1)
                    ("NADH" -1)
                    ("ORTHOPHOSPHATE" 1)
                    ("ADP" 1)
                    ("H2O" 1)

```

```
      ("NAD+" 1))
#s(ENZYME : INHIBITED-BY ("PROLINE")
: ACTIVATED-BY ()
: PATHWAYS ("AAB"))
: CODE (1 5 1 2)
: NAME "PYRROLINE-5-CARBOXYLATE REDUCTASE"
: REACTION (("1-PYRROLINE 5-CARBOXYLATE" -1)
("L-PROLINE" 1))
: REVERSIBLE NIL
: COUPLING-REACTION (("NADPH" -1)
("NADP+" 1))
: INHIBITED-BY ()
: ACTIVATED-BY ()
: PATHWAYS ("AAB"))
#s(ENZYME : CODE (1 14 16 1)
: NAME "PHENYLALANINE 4-MONOOXYGENASE"
: REACTION (("L-PHENYLALANINE" -1)
("L-TYROSINE" 1))
: REVERSIBLE NIL
: COUPLING-REACTION (("NADPH" -1)
("O2" -1)
("NADP+" 1)
("H2O" 1))
: INHIBITED-BY ()
: ACTIVATED-BY ()
: PATHWAYS ("AAB"))
#s(ENZYME : CODE (1 1 1 95)
: NAME "PHOSPHOGLYCERATE DEHYDROGENASE"
: REACTION (("3-PHOSPHO-D-GLYCERATE" -1)
("3-PHOSPHOHYDROXYPYRUVATE" 1))
: REVERSIBLE NIL
: COUPLING-REACTION (("NAD+" -1)
("NADH" 1))
: INHIBITED-BY ()
: ACTIVATED-BY ()
: PATHWAYS ("AAB"))
#s(ENZYME : CODE (2 6 1 52)
: NAME "PHOSPHOSERINE TRANSAMINASE"
: REACTION (("3-PHOSPHOHYDROXYPYRUVATE" -1)
("L-GLUTAMATE" -1)
("2-OXOGLUTARATE" 1)
("3-PHOSPHOSERINE" 1))
: REVERSIBLE T
: COUPLING-REACTION ()
: INHIBITED-BY ()
: ACTIVATED-BY ()
: PATHWAYS ("AAB"))
#s(ENZYME : CODE (3 1 3 3)
: NAME "PHOSPHOSERINE PHOSPHATASE"
: REACTION (("3-PHOSPHOSERINE" -1)
("L-SERINE" 1))
: REVERSIBLE NIL
: COUPLING-REACTION (("H2O" -1)
("ORTHOPHOSPHATE" 1))
```

```

:INHIBITED-BY      ( )
:ACTIVATED-BY     ( )
:PATHWAYS         ("AAB"))
#s(ENZYME :CODE    (1 1 1 81)
:NAME            "HYDROXYPYRUVATE REDUCTASE"
:REACTION       (( "D-GLYCERATE" -1)
                 ("HYDROXYPYRUVATE" 1))
:REVERSIBLE      T
:COUPLING-REACTION (( "NAD+" -1)
                   ("NADH" 1))
:INHIBITED-BY   ( )
:ACTIVATED-BY   ( )
:PATHWAYS       ("AAB"))
#s(ENZYME :CODE    (2 6 1 45)
:NAME        "SERINE-GLYOXYLATE AMINOTRANSFERASE"
:REACTION   (( "L-SERINE" -1)
             ("GLYOXYLATE" -1)
             ("3-HYDROXYPYRUVATE" 1)
             ("GLYCINE" 1))
:REVERSIBLE  T
:COUPLING-REACTION ( )
:INHIBITED-BY ( )
:ACTIVATED-BY ( )
:PATHWAYS     ("AAB"))
#s(ENZYME :CODE    (2 6 1 51)
:NAME        "SERINE-PYRUVATE AMINOTRANSFERASE"
:REACTION   (( "L-SERINE" -1)
             ("PYRUVATE" -1)
             ("3-HYDROXYPYRUVATE" 1)
             ("L-ALANINE" 1))
:REVERSIBLE  T
:COUPLING-REACTION ( )
:INHIBITED-BY ( )
:ACTIVATED-BY ( )
:PATHWAYS     ("AAB"))
#s(ENZYME :CODE    (2 1 2 1)
:NAME        "SERINE HYDROXYMETHYLTRANSFERASE"
:REACTION   (( "GLYCINE" -1)
             ("L-SERINE" 1))
:REVERSIBLE  T
:COUPLING-REACTION (( "5,10 METHYLENETETRAHYDROFOLATE" -1)
                   ("H2O" -1)
                   ("TETRAHYDROFOLATE" 1))
:INHIBITED-BY ( )
:ACTIVATED-BY ( )
:PATHWAYS     ("AAB"))
#s(ENZYME :CODE    (2 3 1 30)
:NAME        "SERINE ACETYLTRANSFERASE"
:REACTION   (( "ACETYL-CoA" -1)
             ("L-SERINE" -1)
             ("O-ACETYL-L-SERINE" 1))
:REVERSIBLE  T
:COUPLING-REACTION (( "CoA" 1))
:INHIBITED-BY ( )

```

```
:ACTIVATED-BY      ( )
:PATHWAYS          ("AAB"))
#s(ENZYME :CODE      (9 9 9 3)
:NAME              "CYSTEINE SYNTHASE"
:REACTION          (( "O-ACETYL-L-SERINE" -1)
                   ("L-CYSTEINE" 1)
                   ("ACETATE" 1))

:REVERSIBLE        T
:COUPLING-REACTION (( "H2S" -1))
:INHIBITED-BY      ( )
:ACTIVATED-BY      ( )
:PATHWAYS          ("AAB"))
#s(ENZYME :CODE      (2 3 1 31)
:NAME              "HOMOSERINE ACYLTRANSFERASE"
:REACTION          (( "SUCCINYL-CoA" -1)
                   ("L-HOMOSERINE" -1)
                   ("O-SUCCINYL-HOMOSERINE" 1))

:REVERSIBLE        NIL
:COUPLING-REACTION (( "CoA" 1))
:INHIBITED-BY      ("METHIONINE")
:ACTIVATED-BY      ( )
:PATHWAYS          ("AAB"))
#s(ENZYME :CODE      (4 2 99 9)
:NAME              "O-SUCCINYLMHOMOSERINE(THIOL)-LYASE"
:REACTION          (( "O-SUCCINYL-L-HOMOSERINE" -1)
                   ("L-CYSTEINE" -1)
                   ("CYSTATHIONINE" 1)
                   ("SUCCINATE" 1))

:REVERSIBLE        T
:COUPLING-REACTION ( )
:INHIBITED-BY      ( )
:ACTIVATED-BY      ( )
:PATHWAYS          ("AAB"))
#s(ENZYME :CODE      (4 4 1 8)
:NAME              "CYSTATHIONINE b-LYASE"
:REACTION          (( "CYSTATHIONINE" -1)
                   ("PYRUVATE" 1)
                   ("L-HOMOCYSTEINE" 1))

:REVERSIBLE        T
:COUPLING-REACTION (( "H2O" -1)
                   ("NH3" 1))
:INHIBITED-BY      ( )
:ACTIVATED-BY      ( )
:PATHWAYS          ("AAB"))
#s(ENZYME :CODE      (2 1 1 10)
:NAME              "HOMOCYSTEINE:N5-METHYLTETRAHYDROFOLATE M
:REACTION          (( "HOMOCYSTEINE" -1)
                   ("METHIONINE" 1))

:REVERSIBLE        T
:COUPLING-REACTION (( "METHYLTETRAHYDROFOLATE" -1)
                   ("TETRAHYDROFOLATE" 1))
:INHIBITED-BY      ( )
:ACTIVATED-BY      ( )
:PATHWAYS          ("AAB"))
```

```
#s(ENZYME : CODE      (4 2 99 11)
: NAME      "METHYLGLYOXAL SYNTHASE"
: REACTION  (( "DIHYDROXYACETONE PHOSPHATE" -1)
             ( "METHYLGLYOXAL" 1))
: REVERSIBLE T
: COUPLING-REACTION (( "PHOSPHATE" 1))
: INHIBITED-BY  ( )
: ACTIVATED-BY  ( )
: PATHWAYS     ( "METHYLGLYOXAL BYPASS" ))
#s(ENZYME : CODE      ((4 4 1 5) (3 1 2 6))
: NAME      "GLYOXYLASE I, II"
: REACTION  (( "METHYLGLYOXAL" -1)
             ( "D-LACTATE" 1))
: REVERSIBLE T
: COUPLING-REACTION (( "H2O" -1))
: INHIBITED-BY  ( )
: ACTIVATED-BY  ( )
: PATHWAYS     ( "METHYLGLYOXAL BYPASS" ))
#s(ENZYME : CODE      (1 1 2 3)
: NAME      "LACTATE DEHYDROGENASE (CYTOCHROME)"
: REACTION  (( "D-LACTATE" -1)
             ( "PYRUVATE" 1))
: REVERSIBLE T
: COUPLING-REACTION (( "FERRICYTOCHROME C" -2)
                    ( "FERROCYTOCHROME C" 2))
: INHIBITED-BY  ( )
: ACTIVATED-BY  ( )
: PATHWAYS     ( "METHYLGLYOXAL BYPASS" ))
```