OPTIMIZATION OF BIOPROCESSES FOR MULTIPLE OBJECTIVES

LEE FOOK CHOON

NATIONAL UNIVERSITY OF SINGAPORE

2009

OPTIMIZATION OF BIOPROCESSES FOR MULTIPLE OBJECTIVES

LEE FOOK CHOON

(MSc, MBA, B.Eng.(Hons.))

A THESIS SUBMITTED

FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMICAL AND BIOMOLECULAR ENGINEERING

NATIONAL UNIVERSITY OF SINGAPORE

2009

ACKNOWLEDGEMENTS

It has been a most fruitful learning journey during the past few years of my academic research. It has been and will always be a memorable and enriching experience to stretch the envelopes into relatively unknown areas of knowledge. It is my pleasure to express my gratitude to all those who suggested diagnostics in their own ways that made the problems tractable.

My academic supervisor, Professor Gade Pandu Rangaiah has been instrumental in giving specific pointers during our numerous meetings. He has the insights and experience to use the preliminary results generated as inputs for the next step of the study. He asked probing questions which enabled me to go through the thinking process which were the harbinger of the yet to be developed answers. His timely and professional feedbacks will certainly suit a wide range of candidates, part-time or otherwise. He had connected well with candidates who were working full-time and facing the various demands of life through a well-thought balanced approach straddling sloth and torpor on one side and slavishness on the other side. This work will not materialize were Professor Ajay Kumar Ray (academic co-supervisor) not placed his confidence in my intention to pursue doctoral study. He was the initial inspiration and catalyst that convinced me to press on with this study. In particular, Professor Ray arranged for Dr. Abhijit Tarafder to set up the computer loaded with the software tools that are essential for this study.

I want to thank Dr. Lee Dong-Yup who provided the initial literature and data in Chapter 3. He gave us a glimpse into the fascinating topics being pursued in systems biotechnology. I thank Mekapati Srinivas who facilitated the logistics and Naveen Agrawal who spotted a missing program line in the early part of the study.

My parents have been very encouraging and supportive of my academic pursuit and it is a major driving force to see me through the endeavours. They understand the importance of pursuing new knowledge as the needs of employment change in Singapore where non-routine symbolic thinking and problem solving abilities are critical success factors for sustainability. My spouse has been very patient and accommodative of me spending long hours away from home. I want to thank her for her marvellous encouragement and moral support since the beginning of the study till now.

TABLE OF CONTENTS

SUMMARY

The present research focuses on the optimization of penicillin and amino acids production. Penicillin is the first microbially produced antibiotics to be discovered, and its production technology is a paradigm for the biopharmaceutical industry. It is the first and most important active pharmaceutical ingredient produced commercially by an aerobic submerged fermentation.

Amino acids such as serine and tryptophan are active pharmaceutical ingredients and nutrients for livestock. Their high commercial values are not matched by their total production rates worldwide. Engineering the enzyme kinetics of multi-product microbial cell factories such as *Escherichia coli* through gene knockout and manipulation has great potential in enhancing the biosynthesis of amino acids.

The main objective of this research is to model and optimize penicillin bioreactor train and desired biosynthesis rates in *Escherichia coli* for multiple objectives. Pareto search was successfully carried out using the non-dominated sorting genetic algorithm (NSGA-II) in conjunction with exhaustive search, interactive branch-and-bound and pattern recognition heuristics.

In the first study, modelling of the penicillin V bioreactor train was done to set the stage for optimization. One *Penicillium chrysogenum* fermentation model was carefully selected based on available industrial information and research works. The bioreactor train model was developed to allow a targeted continuous production rate where each bioreactor operates semi-continuously in a synchronized manner. There were two cases of bi-objective optimization: simultaneous maximization of yield and penicillin concentration, and simultaneous maximization of yield and minimization of batch cycle time. The tri-objective case involves simultaneous maximization of yield, profit and penicillin concentration. Pareto-optimal fronts were obtained for both bi- and tri-objective scenarios using six decision variables.

In the second study, optimization of the central carbon metabolism of *Escherichia coli* was performed for dual objectives to maximize the desired flux ratios of three enzyme kinetics − PEP carboxylase (PEPCxylase), 3-deoxy-D-arabino-heptulosonate-7-phosphate synthase (DAHPS) and serine synthesis (SerSynth). The Pareto obtained in simultaneous maximization of DAHPS and PEPCxylase fluxes, and in simultaneous maximization of DAHPS and SerSynth fluxes provided a

template for metabolic pathway recipe. The metabolic pathway recipe is a form of *a priori* knowledge for experimental research to improve the multi-product capability of microbial cell factories for conflicting objectives.

In the third study, an augmented model for optimizing serine and tryptophan flux ratios simultaneously, was developed by linking the dynamic tryptophan operon model and aromatic amino acid-tryptophan biosynthesis pathways to the central carbon metabolism. Six new kinetic parameters of the augmented model were estimated with considerations of available real data and other published works. Major differences between calculated and reference concentrations and fluxes were explained. Sensitivities and underlying competition among fluxes for carbon sources were consistent with intuitive expectations based on visual metabolic network and preceding results.

In the final study, biosynthesis rates of serine and tryptophan were simultaneously maximized using the augmented model via concurrent gene knockout and manipulation. The optimization results were obtained using NSGA-II supported by pattern recognition heuristics. Possible existence of local Paretos was discussed. One Pareto branch was obtained using NSGA-II for the wide gene multiplier range of 0-1.5. The remaining Pareto was obtained through simulations following the Pareto pattern recognition. Missing Pareto solutions have been explained wherever possible. The results obtained concur with the reported microbial cell fermentation studies and known dynamic behaviour of the tryptophan operon.

In summary, simulation and optimization of multiple bioreactors for penicillin V production for conflicting objectives provided many optimal and practicable solutions for the decision maker. Concurrent gene knockout and manipulations of *Escherichia coli* based on complex nonlinear kinetics show the feasibility of enhancing multi-product biosynthesis rates in one microorganism within certain technological and physiological limits for the first time. These findings are useful in designing new bioprocesses involving multiple products and re-configuring a complex metabolic network for valuable and novel products by probing their performance limits. The current work can be extended to four related areas in systems biotechnology of multi-product fermentation plant and microbial cell factories − modelling, optimization, Pareto ranking and decision making, and techniques to minimize numerical difficulties.

LIST OF SYMBOLS

Chapter 2

Symbols

C_L (C_L^*)	dissolved (saturated) oxygen concentration in the broth (mol/m^3)
Cost	operating cost of a bioreactor for one batch cycle ($)
$Cost_{air,batch}$	cost of sterile air in batch mode ($/h)
$Cost_{air,cont}$	cost of sterile air in continuous mode ($/h)
$Cost_{air,disch}$	cost of sterile air in discharge mode ($/h)
$Cost_{chill,batch}$	cost of chilled water in batch mode ($/h)
$Cost_{chill,cont}$	cost of chilled water in continuous mode ($/h)
$Cost_{chill,disch}$	cost of chilled water in discharge mode ($/h)
$Cost_{csl}$	cost of corn steep liquor ($/kg)
$Cost_{el,batch}$	cost of electricity in batch mode ($/h)
$Cost_{el,cont}$	cost of electricity in continuous mode ($/h)
$Cost_{el,disch}$	cost of electricity in discharge mode ($/h)
$Cost_{glu}$	cost of glucose ($/kg)
$Cost_{water}$	cost of potable water ($/1000 kg)
$C_{p,\ chilled\ water}$	specific heat capacity of chilled water (J/(kg °C))
f	overall product loss (fraction)
f_h	active fraction of the hyphal compartment
F	glucose feed volumetric flow rate (L/h)
F_{GLU}	glucose feed mass flow rate ($kg/(m^3\ h)$)
k_a, k_s, k_h	rate constant for growth reactions (h^{-1})
k_La	overall oxygen mass transfer coefficient (h^{-1})
k_{u1}	rate constant for branching reaction (h^{-1})
k_{u2}	rate constant for tip extension reaction (h^{-1})
k_{u3}	rate constant for differentiation reaction (h^{-1})
k_2	rate constant for penicillin V production reaction (h^{-1})
K_I	inhibition constant for penicillin V production reaction (g glucose/L)
K_s	saturation constant for growth reactions (g glucose/L)
K_{u3}	saturation constant for differentiation reaction (L/g glucose)
K_2	saturation constant for penicillin V production reaction (g glucose/L)

m_s	maintenance coefficient (h^{-1})
n	number of bioreactors in the train
n_p	stirring rate (revolutions per second)
P	penicillin V concentration (g/L)
P_{final}	penicillin V concentration at the end of fermentation (kg/m^3)
P_g	electric power input to the stirrer (kW)
Q_p	sterile air aeration rate (m^3 of air/m^3 of broth/minute)
Q_{vol}	volumetric flow rate of broth from a bioreactor (m^3/h)
r_{CSL}	specific uptake rate of nutrients in corn steep liquor (g/(g DW h))
r_{GLU}	specific uptake rate of glucose (g/(g DW h))
r_p	specific rate of penicillin V production (g/(g DW h))
R	targeted penicillin V production rate from the bioreactor train (kg/h)
S	substrate concentration (g/L)
S_{CSL}	concentration of nutrients in corn steep liquor (g/L)
$S_{CSL,in}$	concentration of corn steep liquor at the beginning of a batch mode (kg/m^3)
S_{GLU}	concentration of glucose (g/L)
$S_{GLU,fed}$	concentration of glucose during continuous feeding (kg/m^3)
$S_{GLU,in}$	concentration of glucose at the beginning of a batch mode (kg/m^3)
S_T	total substrate (or glucose equivalents) concentration (g/L)
$t_{batch\ cycle}$	batch cycle time (h)
$t_{continuous}$	duration for the continuous glucose feed to the bioreactor (h)
$t_{discharge}$	time needed to completely discharge the broth from the bioreactor (h)
$t_{fermentation}$	fermentation time (h)
t_{steril}	time needed to sterilise and line up a bioreactor for a new batch (h)
t_{switch}	duration for the batch mode (h)
u_1, u_2, u_3	branching, tip extension and differentiation reaction rate (h^{-1})
V, V_p	broth volume (L)
V_{final}	broth volume at the end of fermentation (m^3)
V_{in}	broth volume at the beginning of a batch mode (m^3)
X	biomass concentration (g/L)
Z_a	fraction of apical compartment in a hyphal element (g/g DW)
Z_h	fraction of hyphal compartment in a hyphal element (g/g DW)
Z_s	fraction of subapical compartment in a hyphal element (g/g DW)

Greek symbols

α	ratio of glucose feed mass flow rate (kg/(m^3/h)) to glucose concentration in the feed (kg/m^3)
α_{CSL}	conversion factor (g glucose/g corn steep liquor)
α_1	stoichiometric coefficient for glucose (biomass formation) (g glucose/g DW)
α_2	stoichiometric coefficient for glucose (penicillin formation) (g glucose/g penicillin)
μ	total specific growth rate (g/(g DW h))
μ_a, μ_a, μ_a	specific growth rate for apical, subapical, and hyphal cells, respectively (g/(g DW h))
ρ_{broth}	broth density (kg/m^3)

Chapters 3 to 5

Enzymes

ALDO	aldolase
AMT	aminotransferases
ANTAP	anthranilate phosphoribosyl transferase
ANTAS	anthranilate synthase
CHM	chorismate mutase
CHOS	chorismate synthase
ChoSynth	chorismate synthesis pooled enzyme in the common aromatic amino acids pathway
DAHPS	DAHP synthase
DHQD	3-dehydroquinate dehydratase
DHQS	3-dehydroquinate synthase
ENO	enolase
EPSS	5-enolpyruvoylshikimate 3-phosphate synthase
G1PAT	glucose-1-phosphate adenyltransferase
G3PDH	glycerol-3-phosphate dehydrogenase
G6PDH	glucose-6-phosphate dehydrogenase
GAPDH	glyceraldehyde-3-phosphate dehydrogenase
IPS	indolglycerol phosphate synthetase

MetSynth	methionine synthesis
MurSynth	mureine synthesis
PFK	phosphofructokinase
PGDH	6-phosphogluconate dehydrogenase
PGI	glucose-6-phosphate isomerase
PGK	phosphoglycerate kinase
PGluM	phosphoglycerate mutase
PDH	pyruvate dehydrogenase
PEPCxylase	PEP carboxylase
PGM	phosphoglucomutase
PPAI	phosphoribosyl anthranilate isomerase
PPDG	prephenate dehydrogenase
PPDT	prephenate dehydratase
PK	pyruvate kinase
PTS	phosphotransferase system
R5PI	ribose-phosphate isomerase
RPPK	ribose-phosphate pyrophosphokinase
Ru5P	ribulose-phosphate epimerase
SerSynth	serine synthesis
SHDG	shikimate dehydrogenase
SHK	shikimate kinase I, II
Synth1	synthesis 1
Synth2	synthesis 2
TA	transaldolase
TIS	triosephosphate isomerase
TKa	transketolase, reaction a
TKb	transketolase, reaction b
TRPS	tryptophan synthase
TrpSynth	tryptophan biosynthesis pooled enzyme in the terminal pathway

Metabolites

2pg	2-phosphoglycerate
3pg	3-phosphoglycerate
6 pg	6-phosphogluconate

3dhq	3-dehydroquinate
3dhs	3-dehydroshikimate
3php	3-phosphohydroxypyruvate
3ps	3-phosphoserine
accoa	acetyl-coenzyme A
anta	anthranilate
cdrp	1-(o-carboxyphenylamino)-1-deoxyribulose 5-phosphate
cho	chorismate
dahp	3-deoxy-D-*arabino*-heptulosonate 7-phosphate
dhap	dihydroxyacetonephosphate
dipim	diaminopimelate
e4p	erythrose 4-phosphate
epsp	5-enolpyruvoylshikimate 3-phosphate
f6p	fructose-6-phosphate
fdp	fructose-1,6-biphosphate
g1p, g6p	glucose-1-phosphate and glucose-6-phosphate
gap	glyceraldehyde 3-phosphate
glc	glucose
hpp	4-hydroxyphenylpyruvate
i3gp	indole 3-glycerolphosphate
ile	isoleucine
kival	α-ketoisovalerate
lala	L-alanine
met	methionine
mur	mureine
oaa	oxaloacetate
pep	phosphoenolpyruvate
pgp	1,3-diphosphoglycerate
phe, L-phe	phenylalanine, L-phenylalanine
ppa	prephenate
ppy	phenylpyruvate
pra	anthranilate-5-phosphoribosyl pyrophosphate
prpp	5-phosphoribosyl-α-pyrophosphate

pyr	pyruvate
rib5p	ribose-5-phosphate
ribu5p	ribulose-5-phosphate
s3p	shikimate 3-phosphate
sed7p	sedoheptulose-7-phosphate
ser	serine
shik	shikimate
trp, L-trp	tryptophan, L-tryptophan
tyr, L-tyr	tyrosine, L-tyrosine
xyl5p	xylulose-5-phosphate

Co-metabolites (unbalanced)

adp	adenosindiphosphate
amp	adenosinmonophosphate
atp	adenosintriphosphate
nad	diphosphopyridindinucleotide, oxidized
nadh	diphosphopyridindinucleotide, reduced
nadp	diphosphopyridindinucleotide-phosphate, oxidized
nadph	diphosphopyridindinucleotide-phosphate, reduced

Symbols

A	transcription attenuation function
b	asymptotic value of transcription attenuation function
c	first order time constant of transcription attenuation function
$c_{glc}^{extracellular}$	extracellular glucose concentration
c_{glc}^{feed}	glucose feed concentration
c_{g6p}	concentration of glucose-6-phosphate
c_i	i^{th}-metabolite concentration
$c_{i,ref}$	i^{th}-metabolite concentration at reference conditions
c_{nadp}	concentration of diphosphopyridindinucleotide-phosphate, oxidized
c_{nadph}	concentration of diphosphopyridindinucleotide-phosphate, reduced

c_x	biomass concentration
C_{choris}	concentration of chorismate
C_{dahp}	concentration of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate
C_{enz}	concentration of pooled enzyme in the terminal tryptophan biosythesis pathway
C_{e4p}	concentration of erythrose 4-phosphate
C_{nadph}	concentration of diphosphopyridindinucleotide-phosphate, reduced
C_{pep}	concentration of phosphoenolpyruvate
C_{prpp}	concentration of 5-phosphoribosyl-α-pyrophosphate
C_{ser}	concentration of serine
C_{trp}	concentration of tryptophan
d, e, f	constants for the tryptophan uptake
D	dilution factor or mRNA destroying enzyme concentration
E	total enzyme concentration
E_A	active enzyme concentration
EnzDegraded	degradation rate of pooled enzyme in the terminal tryptophan biosynthesis pathway
EnzSynth	synthesis rate of pooled enzyme in the terminal tryptophan biosynthesis pathway
f_{pulse}	glucose pulse
F	tryptophan uptake rate
g	maximum tryptophan consumption rate
G	tryptophan internal consumption rate
k_{choris}	rate constant of chorismate consumption
k_d	rate constant for mRNA degradation
k_P	rate constant for mRNA polymerase binding to operator site
k_ρ	rate constant for ribosome binding to a free *trpE* related site on an mRNA
k_{prpp}	rate constant of 5-phosphoribosyl-α-pyrophosphate consumption
k_{ser}	rate constant of serine consumption
k_{trp}	rate constant of tryptophan consumption
K	tryptophan biosynthesis rate constant

K_g	saturation constant for tryptophan internal consumption
K_i	enzyme inhibition equilibrium constant
K_r	dissociation constant of the holorepressor-DNA operator complex
K_t	tryptophan aporepressor activation constant at equilibrium
$K_{G6PDH,g6p}$	kinetic constant
$K_{G6PDH,nadp}$	kinetic constant
$K_{G6PDH,nadph,g6pinh}$	inhibition constant
$K_{G6PDH,nadph,nadpinh}$	inhibition constant
m	number of metabolites
M_F	free mRNA concentration
n_H	Hill coefficient
O	total operon concentration
O_F	free operon concentration
P	mRNA polymerase concentration
r_{G6PDH}^{max}	maximum reaction rate of glucose-6-phosphate dehydrogenase
r_i	i^{th}-enzymatic reaction rate
$r_{i,ref}$	i^{th}-enzymatic reaction rate at reference conditions
$r_{ChoSynth}$	chorismate synthesis flux
$r_{ChoSynth}^{max}$	maximum rate of chorismate synthesis
r_{DAHPS}	DAHP synthase flux
r_{ENO}	enolase flux
$r_{MurSynth}$	mureine synthesis flux
$r_{PEPCxylase}$	PEP carboxylase flux
r_{PK}	pyruvate kinase flux
r_{PTS}	phosphotransferase system flux
r_{RPPK}	ribose-phosphate pyrophosphokinase flux
$r_{SerSynth}$	serine synthesis flux
r_{Synth3}, ChoConsumed	chorismate consumption rate
$r_{Synth4,}$ PrppConsumed	5-phosphoribosyl-α-pyrophosphate consumption rate
$r_{Synth5,}$ SerConsumed	serine consumption rate
$r_{Synth6,}$ TrpConsumed	tryptophan consumption rate
$r_{TrpSynth}$	tryptophan synthesis flux
$r_{TrpSynth}^{max}$	maximum rate of tryptophan synthesis
R	total repressor concentration

R_A	active repressor (holorepressor) concentration
R_I	inactive repressor (aporepressor) concentration
T	tryptophan concentration
T_{ext}	extracellular tryptophan concentration in the growth medium
z	number of enzymatic sub-systems

Greek symbols

γ	enzyme degradation rate
μ	cellular specific growth rate
ρ	ribosome concentration
ρ_x	microbial cell density
τ_e	time taken by a ribosome to synthesize a *trpE* polypeptide
τ_m	time taken for an mRNA polymerase to assemble *trpE* related binding site on a ribosome
τ_P	transcription initiation periodic time
τ_ρ	translation initiation periodic time

LIST OF TABLES

LIST OF FIGURES

Chapter 1
INTRODUCTION

1.1 Multi-Objective Optimization

Multi-objective optimization (MOO) involves the search for tradeoffs (or Pareto-optimal front or equally good solutions) when there are two or more objectives. When there are conflicting objectives, it is not possible to obtain a single solution which is simultaneously optimal for all the objectives (utopia point). The concept of MOO was conceived by the economist, Pareto in 1896. A solution dominates another one if the first solution is no worse than the second solution in all objectives and it is strictly better than the second solution in at least one objective. Solutions in the non-dominated set are better than the rest of the solutions. There are tradeoffs within the non-dominated set. A compromise or tradeoff is reached when one solution cannot be made better without making another solution worse. The non-dominated set is given a special term: Pareto-optimal set or front. The approach taken here is the ideal MOO where the Pareto is obtained without assigning preferences to any of the objectives. The non-dominated set of the entire search space is the globally Pareto-optimal set. Local Pareto may exist when the objectives are non-linear and the Pareto-optimal front is discontinuous.

Available methods for MOO can be classified in several ways. One of them is based on whether many Pareto-optimal solutions are generated or not, and the role of the decision maker in solving the MOO problem. This particular classification, adopted by Miettinen (1999) and Diwekar (2003), is shown in Figure 1.1.

A classical way to make an MOO problem tractable is to combine several objective functions into a single/scalar objective function, using either arbitrary or user-preferred weight factors. Unfortunately, this "scalarization" of what is really a vector objective function suffers from several drawbacks (Bhaskar et al., 2000). Firstly, the results are sensitive to the values of the weighting factors used, which are difficult to assign on an *a priori* basis. More importantly, there is a risk of losing some optimal solutions (Chankong and Haimes, 1983; Haimes, 1977). The "scalarization" method, also known as weighted sum method, is intuitive and easy to use. For problems having a convex Pareto-optimal front, this method guarantees finding solutions on the entire Pareto-optimal set.

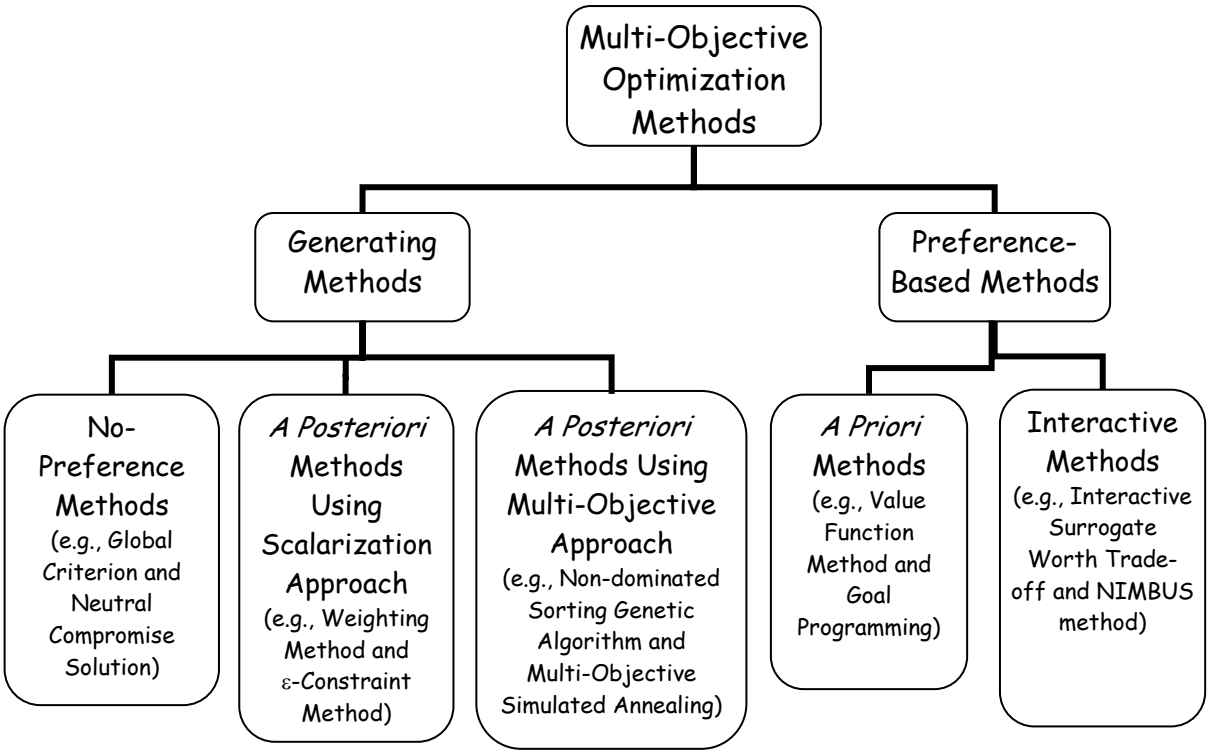


Fig. 1.1. Classification of multi-objective methods.

Non-convexity of the objective function gives rise to a duality gap. Numerically, the duality gap is the difference between the primal and dual objective values. The original mathematical problem is called the primal. Dual is another mathematical problem with the property that its objective is always a bound on the primal. This non-convexity is illustrated using Figure 1.2. The objective functions are denoted as F_1 and F_2. The task is to minimize F_1 and F_2. Multiplying F_1 and F_2 by user-specified weights, the multiple objectives are converted into a single objective function. The contour lines marked S and T represent two different weight vectors. A given weight vector results in a contour line of a particular gradient. Different weight vectors can result in different contour lines having the same gradient.

The Pareto-optimal front refers to the regions AB, BC and CD. The minimum value of the single objective function corresponds to a Pareto-optimal solution (e.g., point A for F_1 and point D for F_2 in Figure 1.2). Unfortunately, there is no contour line that will be tangent to a point in the region BC. In nonlinear MOO problems, a uniformly distributed set of weight vectors need not necessarily lead to a uniformly distributed set of Pareto-optimal solutions. The relationship between weight vectors and the distribution pattern of Pareto-optimal solutions is not usually known. Multiple

minima (or maxima) may be found for a given weight vector. Search effort can be wasted if these multiple solutions are weakly dominated to each other.

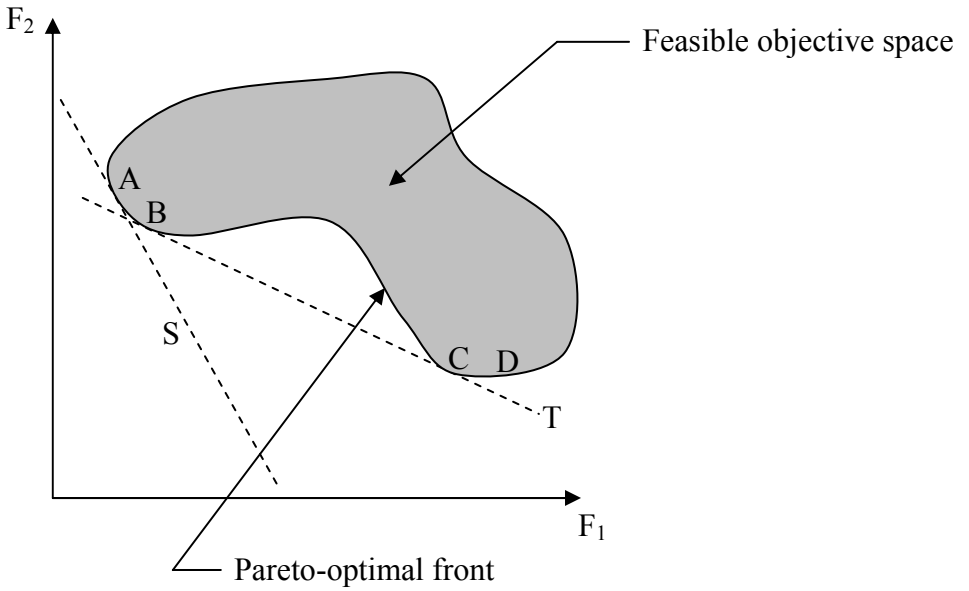


Fig. 1.2. Pareto and non-convexity in the search space.

If "scalarization" does not suffer from the risk of losing some optimal solutions, then a vast array of single objective optimization methods can be used such as direct search methods, gradient based methods and sequential quadratic programming. One common difficulty in applying the above classical methods (Deb, 2001) to MOO is the convergence to an optimal solution depends on the chosen initial guess. Most algorithms tend to get stuck to a suboptimal or local solution. Evolutionary techniques such as genetic algorithm exploit the advantages of parallel search for multiple solutions.

Non-dominated sorting genetic algorithm (NSGA) is one modified version of the simple genetic algorithm for MOO. NSGA differs from a simple genetic algorithm only in the way the selection operator works. The crossover and mutation operators remain as usual (Srinivas and Deb, 1995). NSGA uses a ranking selection method to emphasize the good chromosomes and niche method to create diversity in the population without losing a stable sub-population of good chromosomes. NSGA-II (Deb et al., 2002), an improvement of NSGA, is an elitist NSGA using an elite-preservation strategy as well as an explicit diversity-preserving mechanism.

1.2 Multi-Objective Optimization in Bioprocesses

MOO of bioprocesses particularly in the biopharmaceutical and multiproduct microbial cell fermentation industries is attracting increasing interest from researchers. A recent work in biopharmaceuticals involved the MOO of an industrial penicillin V bioreactor train for dual and triple objectives (Lee et al., 2007) focusing on fermentation using *Penicillium chrysogenum*. MOO in biopharmaceutical areas tend to be allied with batch plant design (e.g. optimizing the multiple options in equipment selection for a plant producing vitamin C by Mõsať et al., 2008) or a particular chemical unit operation (e.g. waste solvent recovery in pharmaceutical industry by Kim and Smith, 2004); there is a scarcity of MOO studying the biological reactions per se. There have been isolated studies in the area of multi-objective bioprocess synthesis (e.g. penicillin plant synthesis by Steffens et al., 1999) and multiproduct batch plant design (e.g. batch plant design for the production of insulin, vaccine, chymosin and protease subject to fuzzy demands by Dietz et al., 2008). The close association of MOO in bioprocesses with design and chemical unit operations reflects the familiarity and competencies of chemical engineers in these areas. In another area within bioprocesses, little work has been done in optimizing living micro-organism metabolic pathways for multiple objectives.

Biologists and biochemists have a solid foundation in experimental research methods of life sciences. Much of their studies rely on *a priori* knowledge, heuristics and intuition. Biochemists have compiled S-system (or synergistic system) models related to various metabolic pathways (Voit, 2000). S-system, which is similar to power laws found in generalized mass action (GMA) modelling framework, is used to represent the kinetics of various bioprocesses such as ethanol fermentation pathway in *Saccharomyces cerevisiae* (yeast) and citric acid metabolism in *Aspergillus niger* (mold). There have been a few multi-objective optimization of metabolic processes in *Saccharomyces cerevisiae* using linear programming (e.g. Link et al., 2008, Vera et al., 2003); none in *Escherichia coli* (bacteria) using highly nonlinear model. In their work (Link et al., 2008, Vera et al., 2003), the S-system representation of the original *Saccharomyces cerevisiae* kinetics was linearized. Multi-objective linear programming was used to obtain the Pareto-optimal set where ethanol production rate was maximized and various intermediate metabolite concentrations were minimized.

There is one recent study on the use of MOO in inferring biochemical

networks such as metabolic pathways modelled through the S-system (Liu and Wang, 2008). Experimentally measured data from batch fermentation of *Saccharomyces diastaticus* LORRE 316 (high-ethanol tolerance yeast) to produce ethanol were used to infer the S-system structure and its parameters by minimizing simultaneously the concentration error, slope error and interaction measure. Another recent work combined flux balance analysis and energy analysis, and applied normalized normal constraint to multiple liver-specific objectives such as ATP synthesis and urea secretion (Nagrath et al., 2007).

1.3 Motivation and Scope of Work

Continuous processes in petroleum, petrochemical and chemical manufacturing have traditionally occupied a disproportionate part of MOO studies. There have been increasing applications of process systems engineering techniques to bioprocesses. The broad objective of this study is to investigate MOO for bioprocesses and in metabolic engineering taking penicillin production and *Escherichia coli* as the respective example.

There was no attempt made to optimize the penicillin production at the fermentation stage for multiple objectives though there have been isolated studies on designing a penicillin plant conceptually using multiple economic and environmental impact criteria (e.g. Steffens et al., 1999). This provides the motivation and scope to model an existing penicillin V bioreactor train for simultaneous optimization of key performance indicators of interest to decision makers in Chapter 2.

There has been no work reported on MOO of the central carbon metabolism of *Escherichia coli* using a highly nonlinear detailed model though a few studies were carried out for the single objective cases (Schmid et al., 2004; Visser et al., 2004; Vital-Lopez et al., 2006). The complex model provides opportunities to study two types of problems separately, discrete gene knockouts and combinatorial gene manipulation, to maximize the fluxes of desired biosynthesis pathways as discussed in Chapter 3.

An augmented model was developed in Chapter 4 by integrating the aromatic amino acids biosynthesis pathway and tryptophan operon dynamics with the central carbon metabolism of *Escherichia coli*. New kinetic parameters of the aromatic biosynthesis pathway were carefully evaluated based on measured and theoretical data

and intuitive expectations of the behaviour of metabolic network in microbial cell factories. Existing tryptophan operon kinetics were adapted in order to be compatible with the specific growth rate of the central carbon metabolism model.

Serine and tryptophan synthesis rates have been optimized separately in the past (serine – Vital-Lopez et al., 2006; tryptophan – Schmid et al., 2004) but not concurrently. In Chapter 5, serine and tryptophan synthesis rates in the augmented model were maximized concurrently through simultaneous gene knockout and gene manipulation to obtain a Pareto-optimal front. This is a potentially challenging application of the augmented model which has embedded non-convexities, nonlinearities and isolated or disjointed Pareto in the entire search space.

The best objectives to consider for optimization have the most impacts on a high-level aim defined *a priori*. A high-level aim such as an environmental aspect or sustainability reflects its importance in industrial practice. The best objectives may be tacit knowledge of the decision maker. Alternatively, ranking the Pareto with respect to the high-level aim arguably identifies the best objectives to pursue. In the absence of a high-level aim as in this study, one is still able to choose several objectives that are known to be important to the decision maker.

1.4 Organization of the Thesis

Following the introductory material in this chapter, Chapter 2 critically reviews the various fermentation models before a model for penicillin V bioreactor modelling and optimization is selected, and then describes MOO of this bioreactor train for multiple objectives. Chapter 2 describes the counteractions among decision variables in generating the Pareto. MOO of various biosynthesis fluxes of the central carbon metabolism of *Escherichia coli* is discussed in Chapter 3. An iterative branch and bound technique is used as an alternative to the manual exhaustive search to generate the Pareto obtained from gene knockouts in Chapter 3. The augmented model developed in Chapter 4 provides the platform for optimization study in Chapter 5. The ability of the augmented model in Chapter 4 to channel carbon into tryptophan biosynthesis is described through a two-stage evaluation. Numerical difficulties and Pareto results consistency with reported fermentation studies are highlighted in Chapter 5. Appropriate conclusions from this research and recommendations for future study are presented in Chapter 6.

Chapter 2
OPTIMIZATION OF AN INDUSTRIAL PENICILLIN V BIOREACTOR TRAIN

2.1 INTRODUCTION

Penicillin belongs to the family of hydrophobic β-lactams. The main commercial penicillin G and penicillin V are produced by *Penicillium chrysogenum*. In 1995, the global production of penicillin G and penicillin V amounted to 24,100 and 8,100 metric tons, respectively, with an estimated value of US$ 1.06 billion (van Nistelrooij et al., 1998); and the annual global production in 2001was estimated at 65,000 metric tons (Lowe, 2001). The bulk of penicillin V is converted into 6-aminopenicillanic acid (6-APA), which is used to make amoxicillin and ampicillin. Rising demand in countries such as China and India drives the annual growth for penicillin production. Given these developments, improvement in the production of penicillin is of considerable importance to both industries and consumers. This work presents a multi-objective optimization study, carried out to find a range of better design and operating conditions for improving the performance of penicillin production units using *Penicillium chrysogenum*. This is perhaps the first study on multi-objective optimization of an industrial penicillin V bioreactor train. The rest of this section reviews the motivation and scope of this study.

Up to now, there has been little work done in multi-objective optimization of biopharmaceuticals (Chapter 1). Biochemists such as Voit (2000) has compiled a list of mathematical modelling works related to bioprocesses using S-system (or synergistic system). S-system, which is similar to power law models, is used to represent the kinetics of various bio-processes. Torres and Voit (2002) have documented the single objective optimization of citric acid production in *Aspergillus niger*, ethanol production in *Saccharomyces cerevisiae* and tryptophan production in *Escherichia coli*. Vera et al. (2003) have studied the mathematical multi-objective optimization in metabolic processes leading to the production of ethanol by *Saccharomyces cerevisiae*. In their work, multiple linear objective functions were obtained from the S-system model by applying natural logarithms to the influx and efflux terms of the equations when pseudo steady state is assumed. Ethanol production was maximized and the various intermediate metabolite concentrations

were minimized using multi-objective linear programming. Sendin et al. (2006) studied the effectiveness of using various techniques (weighted sum, goal attainment, normal boundary intersection, multi-objective indirect optimization and multi-objective evolutionary algorithm) to simultaneously maximize ethanol production and minimize five dependent metabolite concentrations for *Saccharomyces cerevisiae*. The optimal results were first evaluated with reference to total pathway enzyme concentration and biosynthetic effort efficiency for the unconstrained and constrained cases. In the latter case, homeostatic and total enzymatic flux constraints (Section 3.3) were imposed. Mandal et al. (2005) studied the bi-objective optimization of protease and catalase selectivity during *Aspergillus niger* fermentation using ε-constraint facilitated by differential evolution. Halsall-Witney and Thibault (2006) have applied evolutionary algorithms to investigate the multi-objective optimization of gluconic acid production by *Pseudomonas ovalis* in a batch stirred tank reactor. In the area of multi-protein batch plant design (Dietz et al., 2008) for producing insulin, vaccine, chymosin and protease, a fuzzy multi-objective algorithm has been applied to simultaneously optimize net present value, production delay/advance and flexibility index in terms of potential plant capacity to the actual plant capacity ratio.

Recent work by Biwer et al. (2004) dealt with the impact of uncertain model parameters on the economic and environmental performance of Penicillin V production. However, there was no attempt made to optimize the penicillin plant operation. Another recent work (Kookos, 2004) described the single-objective economic potential maximization of penicillin production in a bioreactor using simulated annealing. Araúzo-Bravo et al. (2004) investigated the use of soft sensors and an adaptive controller based on neuro fuzzy systems in a commercial penicillin production plant. Thus, there has been no study on the multi-objective optimization of an industrial penicillin plant. This motivated the thesis author to model an industrial penicillin V bioreactor train and then optimize it for multiple objectives using the non-dominated sorting genetic algorithm, which was successfully employed for many chemical engineering applications (e.g., Agrawal et al., 2006; Bhutani et al., 2006; Nandasana et al., 2003; Oh et al., 2001; Sarkar and Modak, 2005; Tarafder et al., 2005; Yee et al., 2003).

Single objective optimizations such as maximization of economic profit result in trade-offs in other aspects such as larger amount of solvent used and greater volume of biologically active wastes generated.

2.2 Process Description

Penicillin V (and G) is produced through fermentation using the mycelium known as *Penicillium chrysogenum*; a simplified process flow sheet is shown in Figure 2.1. The preferred mode for fermentation is fed-batch since it allows a far more accurate control of feed policy and operating parameters compared to simple batch fermentation (van Nistelrooij et al., 1998). Continuous culture, which in the 1950's aroused much enthusiasm among academic investigators, has failed to find its way into manufacturing due to instability of the strain (or its production potential) over longer time spans (van Nistelrooij et al., 1998). It is much more difficult to maintain a sterile fermentation environment for penicillin production using a continuous process.

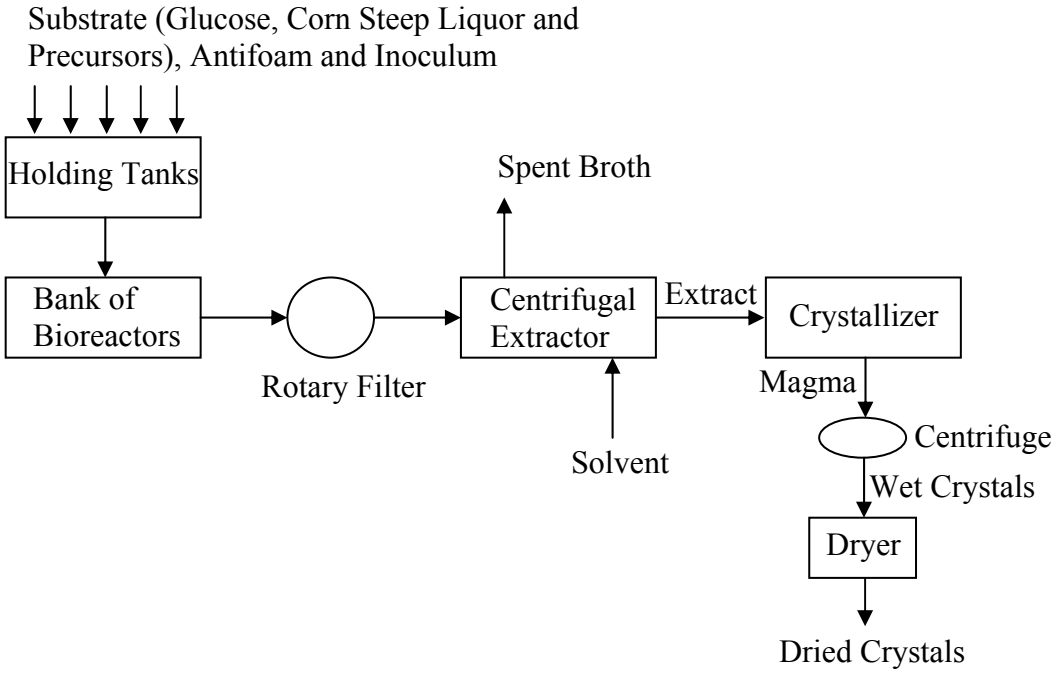


Fig. 2.1. Simplified process flowsheet of a penicillin plant.

The broth in a holding tank is sterilized by heating it with high pressure steam. Inoculum containing the initial biomass is prepared in a separate set of holding tanks. The cooled broth and inoculum are then transferred to the bioreactors. The bank of identical bioreactors is run semi-continuously in a synchronous fashion (Figure 2.2), while the downstream units are operated continuously, in order to meet the targeted production rate. There is a constant phase difference between any two bioreactors in the train to ensure a continuous stream of broth leaving the train to downstream units.

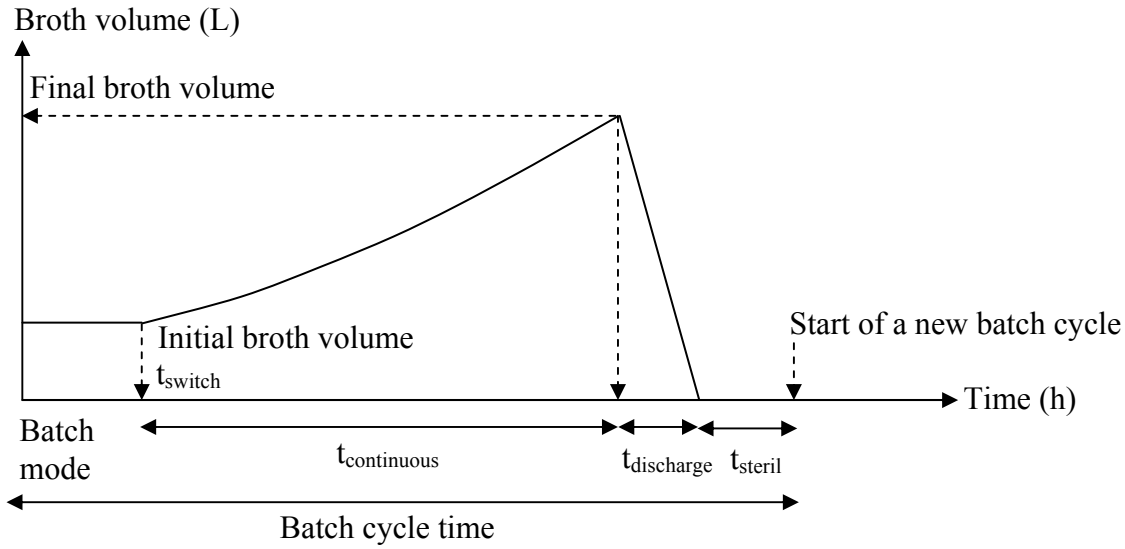


Fig. 2.2. Typical fed-batch profile.

Rotary filter removes the biomass and transfers the liquid containing penicillin V to the continuous counter-current Podbielniak centrifugal extractor. The inlet liquid stream to the extractor is mixed with sulphuric acid to obtain a pH of 2 for efficient extraction. A suitable organic solvent such as butyl acetate is used to extract the penicillin from the product stream. The penicillin-rich extract stream is sent to the carbon treatment unit (not shown in Figure 2.1) to remove pigments and other impurities. The butyl acetate in the raffinate stream is recovered via distillation. The stream from the carbon treatment unit enters the crystallizer where penicillin V sodium (or potassium) salt is formed. Wet crystals separated using the centrifuge, are then dried before packaging and storage.

2.3 Fermentation Models

Penicillin production remains a trade secret and no models are available from industrial producers. The thesis author reviewed the open literature since 1990 and identified four different models (Birol et al., 2002; Menezes et al., 1994; Paul and Thomas, 1996; Zangirolami et al., 1997) for this study. The model proposed by Paul and Thomas (1996) is a structured model comprising more than 20 parameters for hyphal differentiation and penicillin production. The thesis author was not able to validate and use the model since several parameter values were not available. The model by Birol et al. (2002) is an assembly of earlier models proposed by Bajpai and

Reuß (1980), Nielsen (1993) and Jørgensen et al. (1995). However, Birol et al. (2002) did not conduct any fermentation experiments to justify their model parameters. In particular, the parametric constants taken from the work by Bajpai and Reuß (1980) seem to be arbitrarily assigned. The thesis author's simulations of the model proposed by Birol et al. (2002) gave a maximum penicillin concentration of 0.048 g/L, which is about 500 times lower than those reported in Menezes et al. (1994) and Zangirolami et al. (1997). Such low penicillin concentrations were encountered in the past. The highest concentration reported for the year 1946 was around 220 U/ml or 0.12 g/L (Hersbach et al., 1984). Due to continual improvement in industrial strain selection, it is possible for penicillin concentration to fall within the 10 to 70 g/L range (Biwer et al., 2004; Hersbach et al., 1984).

The model by Zangirolami et al. (1997) uses a penicillin V producing strain and the model by Menezes et al. (1994) uses a penicillin G producing strain. Both models were formulated based on the respective authors' experimental results. The penicillin V model (given later in this section) was formulated based on experimental work using a 41-L bioreactor with a maximum broth volume of 25 L. The penicillin G model was formulated based on a study using a 1000-L bioreactor. The penicillin V model is morphologically structured around the metamorphosis and growth reactions of the *Penicillium chrysogenum* hyphal element. It uses an inhibition constant to account for the suppression of penicillin V production at high substrate concentration. The penicillin G model is not morphologically structured and does not use an inhibition constant.

The transient profiles obtained using the models proposed by Menezes et al. (1994) and Zangirolami et al. (2002) are similar (Figure 2.3) even though the two models were formulated using different industrial *Penicillium chrysogenum* strains, mathematical structures and bioreactor sizes. Unlike the model formulated by Birol et al. (2002), the predicted penicillin concentration by both models in their original forms is comparable to what is expected of current commercial production strain. The transient profiles are similar to the typical profile of fed-batch penicillin fermentation (Hersbach et al., 1984; Lowe, 2001). Panlabs Inc., a firm which supplies its penicillin-producing clients with improved strains, has published results obtained in the late 1970's, which show penicillin concentration averaging 45,000 U/ml or 25.4 g/L (van Nistelrooij et al., 1998). Later data are not publicly available, but it seems safe to assume that a further doubling is feasible (van Nistelrooij et al., 1998). Simulations

results using models of Menezes et al. (1994) and Zangirolami et al. (2002) show penicillin concentration increasing steadily beyond the initial biomass growth phase. Continuous glucose feed sustains the metabolism needed for maintenance and penicillin formation. The model proposed by Zangirolami et al. (1997) is selected for the current work since it is formulated from experimental work using relatively recent industrial mycelium strain.

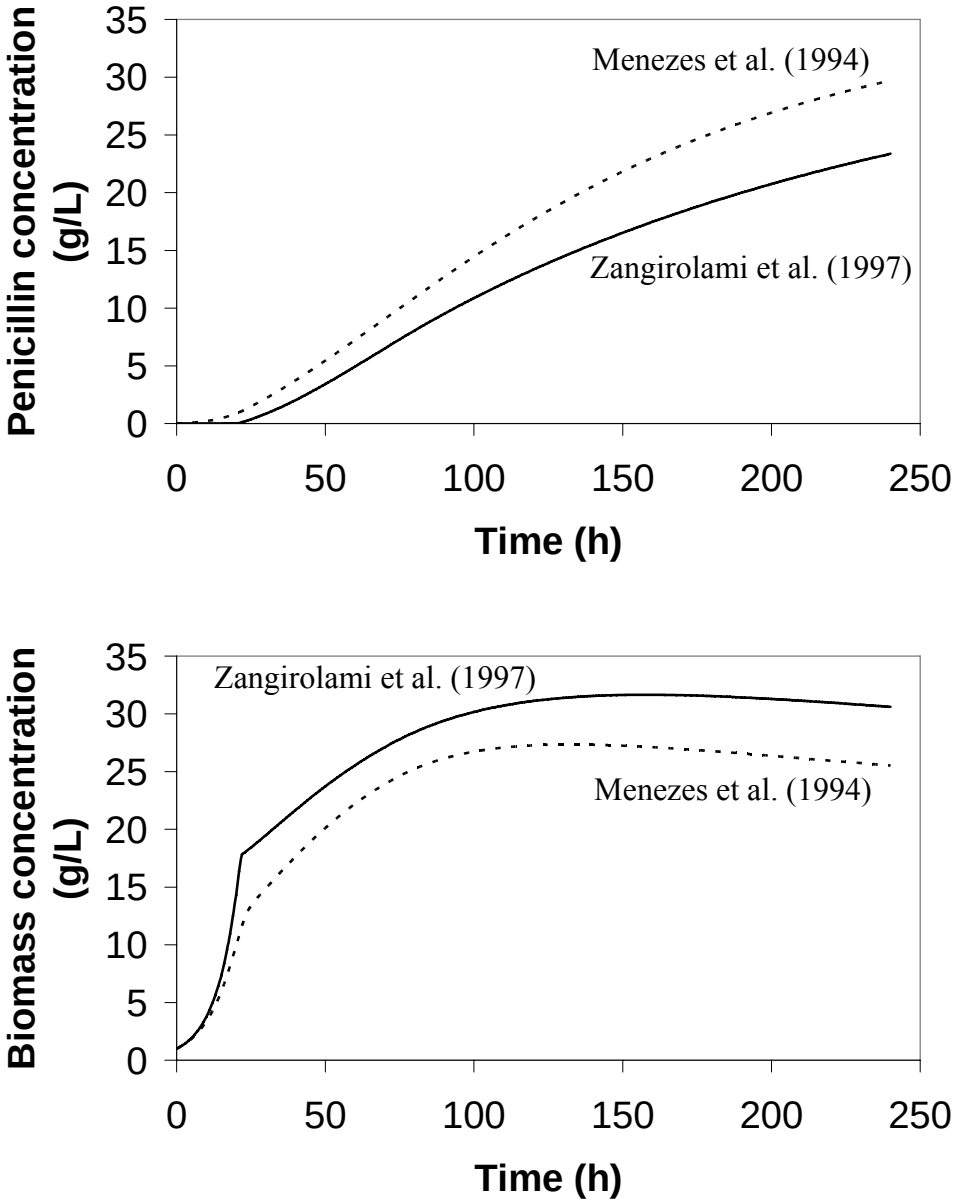


Fig. 2.3. Transient profiles obtained from simulations when the fed-batch policy and initial conditions (Menezes et al., 1994) are applied to the models of Menezes et al. (1994) and Zangirolami et al. (1997). The biomass concentration refers to live cells concentration.

Figure 2.4 shows a hyphal element of the mycelium *Penicillium chrysogenum*. In the model of Zangirolami et al. (1997), penicillin production occurs within the subapical compartment and the active part of the hyphal element.

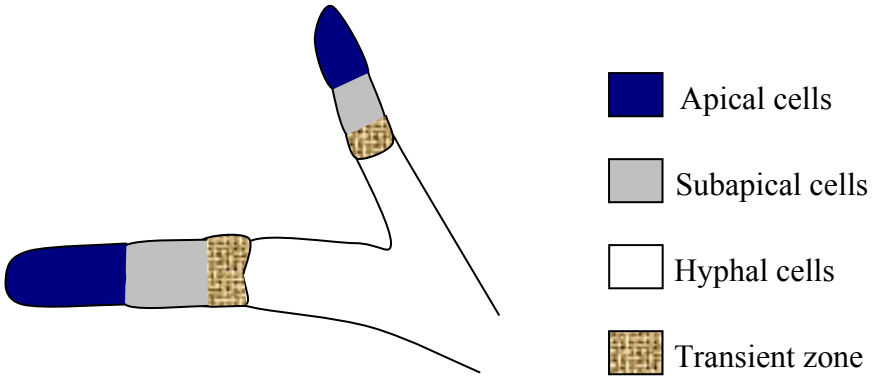


Fig. 2.4. A hyphal element of *Penicillium chrysogenum.*

The differential balance equations describing the metamorphosis and growth reactions occurring in the various morphological compartments, total biomass, penicillin V production, glucose consumption, consumption of nutrients in corn steep liquor and broth volume, are summarized below.

Morphological compartments

Apical cells: $$\frac{dZ_a}{dt} = u_1 - u_2 + Z_a\mu_a - Z_a\mu \tag{2.1}$$

Subapical cells: $$\frac{dZ_s}{dt} = u_2 - u_1 - u_3 + Z_s\mu_s - Z_s\mu \tag{2.2}$$

Hyphal cells: $$\frac{dZ_h}{dt} = u_3 + f_h Z_h\mu_h - \mu Z_h \tag{2.3}$$

Metamorphosis reactions and kinetics

Branching: $Z_s \rightarrow Z_a \quad u_1 = k_{u1}Z_s$ (2.4)

Tip extension: $Z_a \rightarrow Z_s \quad u_2 = k_{u2}Z_a$ (2.5)

Differentiation: $Z_s \rightarrow Z_h \quad u_3 = \frac{k_{u3}Z_s}{S_T K_{u3} + 1}$ (2.6)

$$S_T = S_{GLU} + \alpha_{CSL}S_{CSL} \tag{2.7}$$

Growth kinetics

The growth of apical, subapical, and the active fraction f_h of hyphal cells is described by Monod kinetics.

$$\mu_a = k_a \frac{S_T}{S_T + K_s} \tag{2.8}$$

$$\mu_s = k_s \frac{S_T}{S_T + K_s} \tag{2.9}$$

$$\mu_h = k_h \frac{S_T}{S_T + K_s} \tag{2.10}$$

$$\mu = \mu_a Z_a + \mu_s Z_s + f_h \mu_h Z_h \tag{2.11}$$

Total biomass

$$\frac{dX}{dt} = \mu X - \left(\frac{F}{V}\right) X \tag{2.12}$$

Penicillin V production

$$\frac{dP}{dt} = r_p X - \left(\frac{F}{V}\right) P \tag{2.13}$$

$$r_p = k_2 \frac{S_{GLU}}{S_{GLU} + K_2 + \frac{S_{GLU}^2}{K_I}} \left(Z_s + f_h Z_h\right) \tag{2.14}$$

Glucose consumption

$$\frac{dS_{GLU}}{dt} = S_F \left(\frac{F}{V}\right) - \left(\frac{F}{V}\right) S_{GLU} - r_{GLU} X \tag{2.15}$$

$$r_{GLU} = \alpha_1 \frac{S_{GLU}}{S_{GLU} + \alpha_{CSL} S_{CSL}} \mu + m_s + \alpha_2 r_p \tag{2.16}$$

Consumption of the nutrients in the corn steep liquor

Corn steep liquor is the major nitrogen source for the mycelium. About 42% (by weight) of the corn steep liquor consists of nutrients such as free amino acids, proteins, vitamins and lactate. The remaining 58% of the corn steep liquor cannot be metabolized during fermentation. In contrast, 100% of the glucose can be metabolized

during fermentation. Equation (2.7) represents the concentration of the total glucose equivalents. Corn steep liquor, spores (initial biomass), glucose and precursors (such as phenoxyacetic acid) are added to the bioreactor at the start of the batch mode (Figure 2.2). Glucose solution is then fed continuously to the broth to sustain the mycelium metabolism.

$$\frac{dS_{CSL}}{dt} = -\left(\frac{F}{V}\right)S_{CSL} - r_{CSL}X \tag{2.17}$$

$$r_{CSL} = \alpha_1 \frac{S_{CSL}}{S_{GLU} + \alpha_{CSL}S_{CSL}}\mu \tag{2.18}$$

Broth volume

$$\frac{dV}{dt} = \frac{F_{GLU} \cdot V}{S_{GLU,fed}} \tag{2.19}$$

Values of parameters in the above equations are taken from Zangirolami et al. (1997); they are: $k_{u1} = 2.3\ h^{-1}$; $k_{u2} = 0.7\ h^{-1}$; $k_{u3} = 0.19\ h^{-1}$; $K_{u3} = 20.0$ L/g glucose; $f_h =$ 0.13 g active Z_h/g Z_h; $\alpha_1 = 2.2$ g glucose/g dry weight; $\alpha_2 = 0.6858$ g glucose/g penicillin; $\alpha_{CSL} = 0.42$ g glucose/g corn steep liquor; $k = 0.14\ h^{-1}$; $K_s = 0.0015$ g glucose/L; $m_s = 0.0281\ h^{-1}$; $k_2 = 1.35\ h^{-1}$; $K_2 = 0.0132$ g glucose/L; and $K_I = 0.0101$g glucose/L. The ordinary differential equations in the above model equations were solved using the DIVPRK program in the IMSL software.

2.4 Formulation of the Multi-Objective Optimization Problem

2.4.1 Profit, Yield and Bioreactor Train Model

The industrial penicillin V bioreactor train comprises certain number of identical fermenters designed for a targeted penicillin production rate. The Podbielniak centrifugal extractor and crystallizer are sized to accommodate the given production rate. The key performance indicators selected in this study are batch cycle time, yield, profit and penicillin concentration. Minimizing the batch cycle time (Figure 2.2) for a bioreactor train reduces the time needed to prove the operability of the fermentation process, and also creates greater flexibility in coping with the frequent changes in process operating conditions and sequence of equipment used. Yield refers to the mass of penicillin produced per unit mass of total glucose

equivalents added. Maximizing the yield is equivalent to maximizing the mass of penicillin produced and minimizing the mass of total glucose equivalents concurrently. Maximizing the yield involves minimizing the accumulation of fermentation waste materials, which is desirable since disposal of fermentation waste material such as the fungal mycelium and the biologically active waste liquid is a real problem (Ohno et al., 2003).

The difference between revenue and cost constitutes the profit. Cost comprises the operating cost and the installed cost of the bioreactor train. Stirred tank bioreactors up to 400 m^3 in volume are used in antibiotics production (Schuler and Kargi, 2002). In this study, the volume (Hersbach et al., 1984) of each bioreactor is set at 250 m^3 of which the working volume (i.e. maximum broth volume) is 200 m^3. As shown later in section 2.6, the maximum broth volume ranges from around 170 m^3 to 200 m^3. The excess bioreactor volume is needed to accommodate rising foam during fermentation. The amortized installed cost of the bioreactor train designed to cater to a targeted penicillin production rate does not vary much, and will not be included in the profit objective function. The operating cost is for raw materials (glucose, corn steep liquor, and water) and utilities (electricity, sterile air, and chilled water). Both the penicillin price and the bioreactor train production rate are fixed in this study. Owing to the latter, the size and amortized installed cost of downstream processing units such as the Podbielniak centrifugal extractor and crystallizer do not vary much.

Maximizing penicillin concentration embeds an implicit minimization of the operating cost of downstream processing units such as the Podbielniak centrifugal extractor and crystallizer. For a given extractor size, maximizing the penicillin concentration for a targeted penicillin fraction in the penicillin-rich extract stream assists in minimizing the extractor rotational speed (lower electricity cost) and solvent consumption. Handling large amount of solvent requires one to contend with the issues of solvent recovery, handling, storage, spillage, disposal as well as impact on the environment. The installed cost of the crystallizer is dependent on its volume, which in turn depends on the volumetric flow rate of the product and the residence time. The volumetric flow rate is proportional to the targeted penicillin V production rate. The residence time is slightly dependent on the narrow temperature range of 0 to 4°C which exists within the crystallizer. In other words, the bioreactor train production rate determines the crystallizer size. From the above discussion, it then

follows that maximizing the profit is equivalent to minimizing the operating cost of bioreactors.

The price of penicillin V is \$17/kg and the targeted production rate for the train is 248 kg/h (Biwer et al., 2005). Assuming overall product loss of 15% (Lowe, 2001), the penicillin production rate of the bioreactor train is 292 kg/h. Taking into account the overall product loss, the expected revenue is \$4,216/h. The number of bioreactors in the train is 20 (Section 2.6). Profit (\$/h) can then be computed as follows:

$$\text{Profit} = 17\,\frac{n\left(P_{final}V_{final}\right)\left(1-f\right)}{t_{batch\ cycle}} - \frac{20\,\text{Cost}}{t_{batch\ cycle}} \tag{2.20}$$

The first term on the right side is the revenue, whose derivation is outlined below. The second term is the operating cost and accounts for raw materials and utilities during sterilization, batch/semi-batch operation and discharge of bioreactors. The cost components are discussed in the following sub-section. Note that amortized installed cost and operating cost of downstream units should also be subtracted from the revenue to find the actual profit. Further, prices of utilities taken from the literature (mainly, Turton et al., 2003) are not adjusted to the present time since they vary with supply and demand, geographical location, government tariffs, currency exchange rates and other factors. Overall, profit given by equation (2.20) is reasonable for use as an objective.

The yield can be computed from:

$$\text{Yield} = \frac{P_{final}V_{final}}{\left(V_{in}\right)\left(S_{GLU,in} + \alpha_{CSL}S_{CSL,in}\right) + \left(V_{final} - V_{in}\right)\left(S_{GLU,fed}\right)} \tag{2.21}$$

The broth mass and volumetric flow rate from the bioreactor train are respectively, $n(\rho_{broth}V_{final})/t_{batch\ cycle}$ and $Q_{vol} = n(V_{final})/t_{batch\ cycle}$. Assume that the batch cycle time and the discharge time are the same for all bioreactors in the train, and that the broth is transferred from the train continuously for a targeted penicillin V production rate. The discharge time is V_{final}/Q_{vol} which simplifies to $t_{batch\ cycle}/n$. The batch cycle time (Figure 2.2) is the sum of $t_{fermentation}$, $t_{discharge}$ and t_{steril}, and the fermentation time is equal to $t_{switch} + t_{continuous}$. Thus, the batch cycle time is:

$$t_{batch\ cycle} = \frac{t_{fermentation} + t_{steril}}{1 - \frac{1}{n}} \tag{2.22}$$

If n is one, the batch cycle time is infinite since a single bioreactor does not permit broth to be transferred to another unit continuously. Since both the sterilization time and number of bioreactors are constant, the batch cycle time is only dependent on the fermentation time. If R is the targeted penicillin V production rate and f is the overall product loss, then we have

$$R = \frac{n\left(P_{final} V_{final}\right)}{t_{batch\ cycle}} \cdot \left(1 - f\right).$$

Combining this with equation (2.22) gives

$$R = \frac{\left(n-1\right)\left(P_{final} V_{final}\right)\left(1-f\right)}{t_{fermentation} + t_{steril}}.$$

The model equations (2.1) to (2.19) are integrated until

$$\frac{\left(n-1\right)\left(P_{final} V_{final}\right)\left(1-f\right)}{t_{fermentation} + t_{steril}}$$

equals to or slightly exceeds R. The revenue ($/h) is simply the price of penicillin ($17/kg) times $\frac{\left(n-1\right)\left(P_{final} V_{final}\right)\left(1-f\right)}{t_{fermentation} + t_{steril}}$.

2.4.2 Cost Components

The operating cost equations are formulated based on the following premises: (1) electricity to stir the broth, sterile air to aerate the broth and chilled water to remove metabolic heat are not required during sterilization; (2) cost related to sterilizing steam and inoculum preparation is assumed to be unaffected by bioreactor train design and hence not included in the profit objective; (3) precursors are part of the corn steep liquor; (4) as the industrial glucose feed rate is expressed as kg/m^3 of broth/h, higher broth volume necessitates higher glucose volumetric flow rate, and the broth volume profile with respect to time under continuous glucose feed phase is exponential as shown in Figure 2.2; (5) no change in penicillin concentration during broth discharge at the end of fermentation.

Stirred tank bioreactors up to 400 m^3 in volume are used in antibiotics production, with electric power input up to 5 kW/m^3 (Schuler and Kargi, 2002). Generally, a bioreactor or fermenter has a height-to-diameter ratio of 2 to 3 (Schuler and Kargi, 2002). In the design of the fermenter, a ratio to 2.5 to 1 is used. The broth volume takes up 80% of the bioreactor volume (Hersbach et al., 1984). Consider a

bioreactor using flat-blade turbine under aerated non-Newtonian turbulent conditions. The electric power input (kW), P_g is derived from a correlation proposed by Wang et al. (1979) and its final form is given by: $P_g = 0.0064809 V_p^{1.698} Q_p^{-0.252} n_p^{3.15}$ where V_p is the broth volume (m^3), Q_p is the sterile air aeration rate (m^3 of air/m^3 of broth/minute) and n_p is the stirring rate (revolutions per second). Using Q_p of 1 m^3 of air/m^3 of broth/minute and a stirring rate of 2 revolutions per second, the calculated electric power input per unit volume (kW/m^3 of broth) is 2.43 for a broth volume of 100 m^3 and 3.94 for a broth volume of 200 m^3. This is comparable to the input power of 3 − 4 kW/m^3 stated by Hersbach et al. (1984). Similarly, Perry et al. (1997) states a typical value of 2.5 kW/m^3. Using the sterile air aeration rate and stirring rate stated earlier, the equation describing electric power input (kW) simplifies to: $P_g = 0.0575 V_P^{1.698}$. The price of electricity is \$0.06/kWh (Turton et al., 2003).

The broth volume at time t is given by:

$$V_p = V_{in} e^{\left(\frac{F_{GLU}}{S_{GLU,fed}}\right)t} = V_{in} e^{\alpha t}$$

where V_{in} is the initial broth volume. Here α is the ratio of glucose feed mass flow rate (F_{GLU} kg/m^3/h) to the glucose feed concentration ($S_{GLU,fed}$ kg/m^3).

The cost of electricity (\$/h) can be computed as follows:

Batch mode: $Cost_{el,batch} = (0.06)(0.0575)(t_{switch})(V_{in})^{1.698}$ (2.23)

Continuous feed mode:

$$Cost_{el,cont} = (0.06)(0.0575)\left(V_{final}^{1.698} - V_{in}^{1.698}\right)\left(\frac{1}{1.698\alpha}\right) \quad (2.24)$$

Discharge mode: $Cost_{el,disch} = (0.06)(0.0575)\left(\frac{V_{final}^{1.698}}{2.698}\right)\left(\frac{t_{batch\ cycle}}{n}\right)$ (2.25)

Note that discharge time is $t_{batch\ cycle}/n$ (Section 2.4.1).

The price of sterile air is \$5.25/1000 standard m^3, which is 50% higher than the one stated by Turton et al. (2003) for 50 psig air. The sterile air flow rate is 1 m^3/m^3 of broth/minute. The cost of sterile air (\$/h) can be computed as follows; note that aeration is needed during the discharge mode also, in order to maintain metabolism even though there is no further substrate feeding.

Batch mode: $Cost_{air,batch} = 0.00525)(60)(V_{in})(t_{switch})$ (2.26)

Continuous feed mode: $Cost_{air,cont} = \frac{(0.00525)60(V_{final} - V_{in})}{\alpha}$ (2.27)

Discharge mode: $Cost_{air,disch} = (0.00525)(30)(V_{in})\left(\frac{t_{batch\ cycle}}{n}\right)$ (2.28)

Chilled water flowing through the jacket of the bioreactor removes the metabolic heat evolved during fermentation. The oxygen uptake rate during metabolism is given by: $N_a = (C_L^* - C_L)k_La$. At 25 °C and 1 atmosphere (absolute), the saturated oxygen concentration (C_L^*) in the broth at equilibrium with the bulk gas phase is 0.263 mol/m^3. If C_L is less than 5-10% of C_L^*, irreversible damage adversely affects the productivity of the mycelium (Hersbach et al., 1984). In this study, the oxygen concentration in the broth (C_L) is taken to be 20% of C_L^*. The overall oxygen transfer coefficient (k_La) is 200 h^{-1} (Hersbach et al., 1984). Therefore, the oxygen uptake rate is 42.6 mol/(m^3.h). In aerobic fermentation, the rate of metabolic heat evolution can be correlated to the oxygen uptake rate: $q_H = 0.12N_a$ (Schuler and Kargi, 2002). The rate of metabolic heat evolution (W) can be calculated from: $Q_H = 5944V$. The chilled water is supplied at 5 °C and returned at 15 °C. The price of chilled water is $0.185/1000 kg (Turton et al., 2003). The cost of chilled water can be computed as follows:

Batch mode: $Cost_{chill,batch} = \frac{(0.000185)(5944)(V_{in})t_{switch}}{10C_{p,chilled\ water}}$ (2.29)

Continuous feed mode: $Cost_{chill,cont} = \frac{(0.000185)(5944)(V_{final} - V_{in})}{10\alpha\, C_{p,chilled\ water}}$ (2.30)

Discharge mode: $Cost_{chill,disch} = \frac{(0.000185)(0.5)(5944)(V_{final})}{10C_{p,chilled\ water}}\left(\frac{t_{batch\ cycle}}{n}\right)$ (2.31)

The price of potable water is $0.26/1000 kg (Turton et al., 2003). The cost of potable water can be computed as follows:

$Cost_{water} = 0.26\ V_{final}$ (2.32)

The price of corn steep liquor is $0.2/kg (internal estimate). The cost of corn steep liquor can be computed as follows:

$Cost_{csl} = 0.2\ V_{in}\ S_{CSL,in}$ (2.33)

The price of glucose is $0.216/kg (Biwer et al., 2005). The cost of glucose can be computed as follows:

$Cost_{glu} = 0.216\ V_{in}\ S_{GLU,in} + 0.216(V_{final}-V_{in})\ S_{GLU,fed}$ (2.34)

Summing the above equations (2.22) to (2.33) gives the operating cost of a bioreactor for one batch cycle.

$$\text{Cost} = \text{Cost}_{\text{el,batch}} + \text{Cost}_{\text{el,cont}} + \text{Cost}_{\text{el,disch}} + \text{Cost}_{\text{air,batch}} + \text{Cost}_{\text{air,cont}} + \text{Cost}_{\text{air,disch}} + \text{Cost}_{\text{chill,batch}} + \text{Cost}_{\text{chill,cont}} + \text{Cost}_{\text{chill,disch}} + \text{Cost}_{\text{water}} + \text{Cost}_{\text{csl}} + \text{Cost}_{\text{glu}} \quad (2.35)$$

The cost defined by equation (2.35) is in the second term of the right side of the profit equation (2.20).

2.4.3 Cases

Based on the above discussion on the performance indicators, the current study covers the following two cases of bi-objective optimization.

(a) Max Yield

$$= \frac{P_{\text{final}} V_{\text{final}}}{(V_{\text{in}})(S_{\text{GLU,in}} + \alpha_{\text{CSL}} S_{\text{CSL,in}}) + (V_{\text{final}} - V_{\text{in}})(S_{\text{GLU,fed}})} \quad (2.36)$$

$$\text{Max } P_{\text{final}} \quad (2.37)$$

(b) Max Yield

$$= \frac{P_{\text{final}} V_{\text{final}}}{(V_{\text{in}})(S_{\text{GLU,in}} + \alpha_{\text{CSL}} S_{\text{CSL,in}}) + (V_{\text{final}} - V_{\text{in}})(S_{\text{GLU,fed}})} \quad (2.38)$$

$$\text{Min } t_{\text{batch cycle}} = \frac{t_{\text{fermentation}} + t_{\text{steril}}}{1 - \frac{1}{n}} \quad (2.39)$$

In addition, a tri-objective optimization for maximizing yield, P_{final} and profit (equation (2.20)) simultaneously is studied. Maximizing profit for the bioreactor train may reveal possible trade-offs vis-à-vis yield and penicillin concentration.

The above objective functions are subject to bounds on decision variables given in Table 2.1 and the model equations (2.1) to (2.19). Further, broth volume is constrained to below 200 m^3 for a 250-m^3 bioreactor. When the model equations (2.1) to (2.19) are integrated, it is possible for the state variables (Z_a, Z_s, Z_h, V, X, P, S_{GLU} and S_{CSL}) to become negative or physically inadmissible. When the broth volume breaches the constraint or when a state variable is inadmissible, the objective function

value is penalized by setting it to an arbitrarily high (or low) level; under such conditions, yield, penicillin concentration and profit are set to 10^{-10}, and batch cycle time is set to 10^{10}.

Table 2.1 Decision variables and their ranges used in the multi-objective optimization of bioreactor train.

Decision variable	Range used in optimization	Basis
Switchover time	1 to 60 h	less than 100 h (Biwer et al., 2004; Hersbach et al., 1984)
Initial glucose concentration (batch mode)	10 to 600 g/L	less than 1000 g/L (Biwer et al., 2004; Hersbach et al., 1984)
Continuous glucose feed concentration	10 to 600 g/L	500 g/L (optimum) (Hersbach et al., 1984)
Initial biomass concentration (batch mode)	0.1 to 3 g/L	1 to 2 g/L (Hersbach et al., 1984)
Initial corn steep liquor concentration (batch mode)	10 to 600 g/L	less than 1000 g/L (Biwer et al., 2004; Hersbach et al., 1984)
Initial broth volume (batch mode)	5000 to 60000 L	less than 100,000 L for the industrial fermenter (Biwer et al., 2004)

2.5 Method Used in the Multi-Objective Optimization (MOO)

Evolutionary algorithms such as genetic algorithm (GA) can be applied to complex bio-processes. GAs mimic the principles of natural genetics and natural selection in solving optimization problems. These algorithms use a population of chromosomes; each chromosome is associated with a particular set of values pertaining to the decision variables and fitness (related to objective) function. The randomly-generated initial population of chromosomes undergo reproduction, crossover and mutation to create a new population, which is evaluated and tested for its fitness. One cycle of these operations and the subsequent evaluation procedure is

known as a generation. The above procedure is repeated for several generations until the termination criterion (such as maximum number of generations) is met.

Non-dominated sorting GA (NSGA) is a modified version of the simple GA, for multiple objectives. Non-domination refers to a solution being better in at least one objective than another solution. NSGA differs from the simple GA only in the way the selection operator works. The crossover and mutation operators remain as usual (Srinivas and Deb, 1995). NSGA uses a ranking selection method to emphasize the good chromosomes and niche method to create diversity in the population without losing a stable sub-population of good chromosomes. NSGA-II is a further improvement of NSGA (Deb et al., 2002); it is an elitist NSGA using an elite-preservation strategy as well as an explicit diversity-preserving mechanism.

GA operators can be implemented with binary or real coding for the decision variables in the problem. Representation of real numbers with binary coding limits the precision that can be achieved, requires alteration of many bits for slight change in decision variable value and can retard the search for the optimum. Considering these and similar to Tarafder et al. (2005) who opted for the real-coded NSGA-II for better results, this study too employed the real-coded NSGA-II for the bioreactor train optimization for multiple objectives. The parameters in this algorithm (with values used in this study in brackets) are maximum number of generations (up to 500), population size (100 chromosomes), probability of crossover (0.85), probability of mutation (0.05), distribution index for the simulated crossover operation (10), distribution index for the simulated mutation operation (20) and random seed (0.6). Except for the first and last parameter listed in the previous sentence, the rest of the NSGA-II parameter values are taken from the recent work by Tarafder et al. (2005). Values for maximum number of generations and random seed are obtained by trial and error, as discussed in next section.

In this work, each bioreactor in the train is treated as identical with broth volume constrained to below 200,000 L. The targeted production rate of the bioreactor train is 292 kg/h after taking into account an overall product loss of 15% (Lowe, 2001). The value of t_{steril} is arbitrarily set at 24h to ensure that there is no cross contamination between batches. The model equations (2.1) to (2.19) are integrated until $\frac{(n-1)(P_{final}.V_{final})(1-f)}{t_{fermentation}+t_{steril}}$

converges to or slightly exceeds 292 kg/h. The step size used in integration is 0.1 hour; the same results were obtained with a smaller step size which confirms adequacy of 0.1 hour step size without requiring large computational time. The entire study was done using Pentium (R) 4 computer with 2 GHz CPU and 1.00 GB RAM, and Windows XP Professional. Each simulation took not more than 10 seconds of CPU time; each optimization run for 500 generations required around 5 hours and 7½ hours of CPU time for the bi-objective and tri-objective scenarios, respectively.

2.6 Optimization Results and Discussion

2.6.1 Decision Variables

In the preliminary study for simultaneous maximization of yield and minimization of batch cycle time, both the number of reactors and glucose feed mass flow rate (kg of glucose/m^3 of broth/h) converge towards the respective upper bound. Also, in the study for simultaneous maximization of yield and penicillin concentration, both these variables converge towards the respective lower bound. The batch cycle becomes shorter as the number of bioreactors (see equation 2.22) and/or glucose feed mass flow rate increase; higher glucose feed mass flow rate generally shortens the fermentation time and batch cycle. Longer batch cycle time (or equivalently longer fermentation time when t_{steril} is constant) linked to a smaller number of bioreactors and/or lower glucose feed mass flow rate, results in higher penicillin concentration. Since the number of reactors is actively constrained if it were used as a decision variable, the number of reactors in the train is set at 20. Likewise, instead of postulating glucose feed rate profile with undetermined coefficients as decision variables, the glucose feed rate is set at 1.8 kg/m^3 of broth volume/hour (Hersbach, et al., 1984). With these choices, there are six decision variables: switchover time from batch mode to continuous mode, initial glucose concentration (batch mode), continuous glucose feed concentration, initial biomass concentration, initial corn steep liquor concentration and initial broth volume. Table 2.1 shows the bounds on these decision variables and the basis for them.

2.6.2 Bi-Objective Optimization

Two cases: (a) simultaneous maximization of yield and penicillin concentration and (b) simultaneous maximization of yield and minimization of batch

cycle time, are considered for bi-objective optimization. Figure 2.5 shows the results obtained for case (a) at 100, 300 and 500 generations with the random seed set at 0.6. Higher penicillin concentration driven by greater availability of glucose substrate inhibits penicillin formation as fermentation continues. Thus, the yield decreases when the selectivity with respect to penicillin formation declines at higher penicillin concentration typically associated with longer fermentation time. Results at 300 and 500 generations in Figure 2.5 indicate convergence of the Pareto within 500 generations; further, results at 400 generations (not shown) are almost identical to those of 500 generations. Another trial is made with the random seed set at 0.7, which affects the initial population of chromosomes and also subsequent generations. The solutions obtained at 100 generations with the random seed of 0.7 (not shown in Figure 2.5) are at greater distances from the Pareto-optimal front compared to the corresponding solutions obtained for a random seed of 0.6. Although random seed influences results at lower generations, the same Pareto-optimal fronts are generated for two different random seed values, as displayed in Figure 2.5.

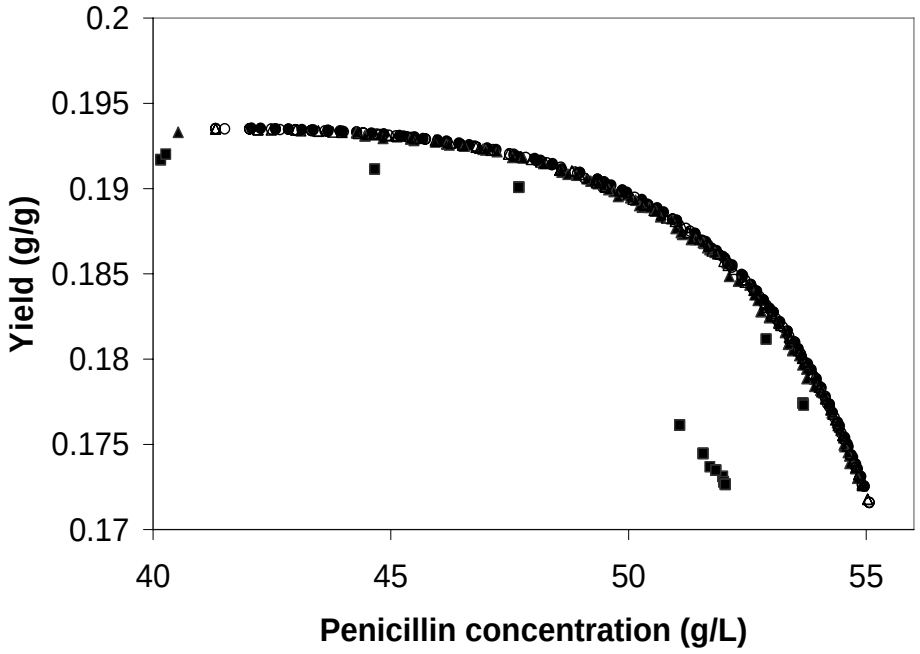


Fig. 2.5. Pareto-optimal fronts obtained at 500 generations for the simultaneous maximization of yield and penicillin concentration using random seed: 0.6 (•) and 0.7 (○). Also shown are the solutions at (a) 300 generations using random seed: 0.6 (▲) and 0.7 (Δ), and (b) 100 generations using random seed of 0.6 (■).

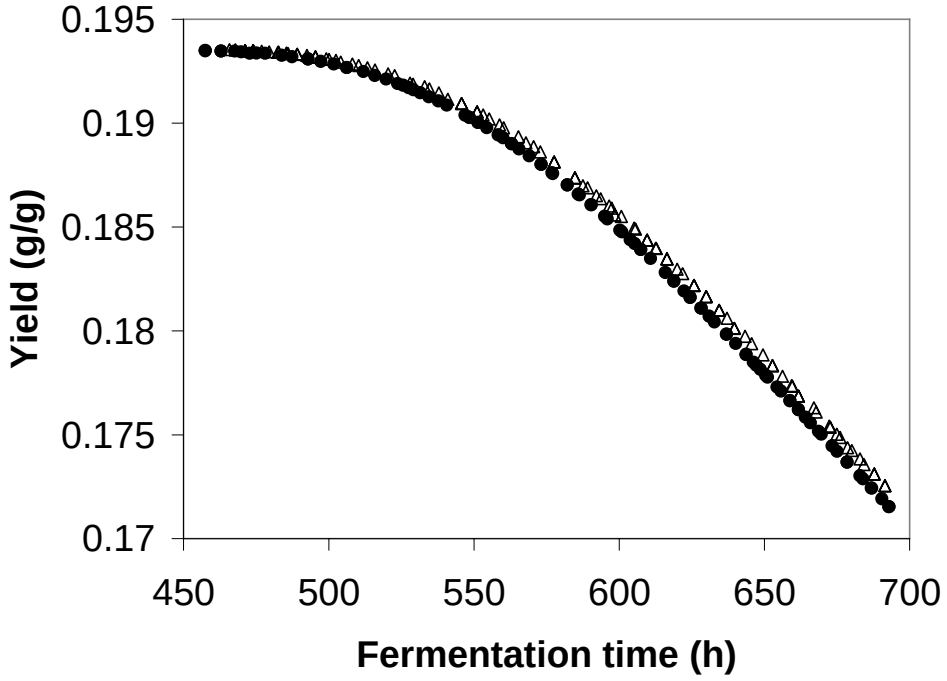


Fig 2.6. The yield versus calculated fermentation time corresponding to the Pareto-optimal fronts in Fig. 2.5 (Δ, random seed of 0.6) and Fig. 2.11 (•, random seed of 0.5).

Figure 2.6 shows the calculated fermentation time corresponding to the Pareto-optimal fronts in Figures 2.5 and 2.11. The decision variables corresponding to the Pareto-optimal fronts of Figure 2.5 are shown in Figure 2.7. Among the decision variables, glucose feed concentration has strong influence on the Pareto-optimal fronts shown in Figure 2.5. Initial broth volume is actively constrained at the lower bound of 5000 L in order to stretch the fermentation time and raise the penicillin concentration as glucose feed concentration increases from about 219 g/L to about 323 g/L. As noted earlier, the gain in penicillin concentration due to longer fermentation time is accompanied by the decline in yield (Figure 2.6). The switchover time, initial glucose concentration, initial biomass concentration and initial corn steep liquor concentration form two distinct groups for the two random seeds. Results in Figure 2.7 suggest that these variables have marginal effect on the Pareto, and that the effect of switchover time and initial glucose concentration on the yield and penicillin concentration is counter-acted by those of initial biomass and corn steep liquor concentrations. This is discussed in the next paragraph. Typically, most of the nutrients will be consumed

during the initial biomass growth phase. Beyond this phase, continuous glucose feed is required to sustain the mycelium metabolism and penicillin formation. However, excess glucose inhibits penicillin formation resulting in lower yield and declining rate of increase of penicillin concentration.

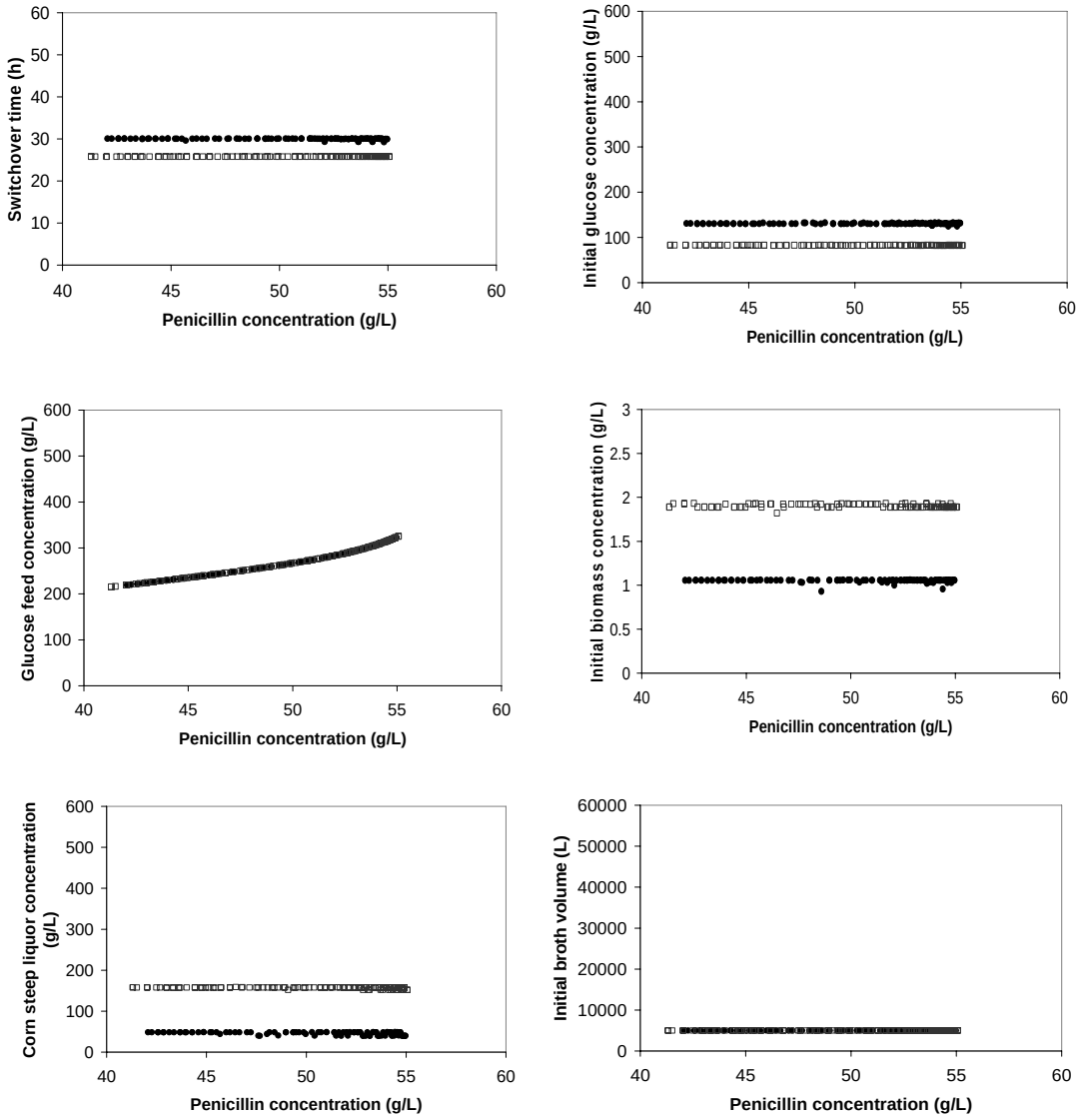


Fig. 2.7. Decision variables corresponding to the Pareto-optimal fronts in Fig. 2.5 for two random seeds: 0.6 (•) and 0.7 (□).

The four decision variables – switchover time, initial glucose concentration, initial biomass concentration and initial corn steep liquor concentration – seem to have opposing effects on the objectives resulting in identical Pareto-optimal front but with different values for them. This is confirmed as the same Pareto-optimal front was obtained when switchover time and initial glucose concentration are fixed (at 15 h and 100 g/L, respectively) in the bi-objective maximization of yield and penicillin

concentration. Also, the initial biomass concentration and initial corn steep liquor concentration are relatively constant (Figure 2.8). Similar observation applies to the total glucose equivalents (not shown) comprising glucose and 42% of corn steep liquor. A certain critical biomass built up within the batch mode is achievable through various means of manipulating the above four decision variables. Glucose feed concentration and initial broth volume have no influence within the batch mode, where no continuous feeding is involved. In contrast, they have great influence on the Pareto-optimal results due to significant product formation when glucose is fed continuously to a bioreactor. Recently, Tarafder et al. (2007) discussed multiple solution sets producing very comparable Pareto-optimal fronts in chemical engineering applications.

Results for case (b) – simultaneous maximization of yield and minimization of batch cycle time – are shown in Figure 2.9, where the yield approaches a distinct plateau as the batch cycle time rises beyond 200 hours. This is because penicillin formation is inhibited by glucose accumulation in the bioreactor; this can be seen from the yield versus calculated batch cycle time shown as crosses in Figure 2.9 corresponding to the Pareto-optimal front in Figure 2.5. Note that crosses, except the left extreme cross (yield = 0.1935, batch cycle time = 515.6 h and initial broth volume = 5000.8 L), are not Pareto-optimal solutions for maximizing yield and minimizing batch cycle time. The right extreme point on the Pareto-optimal front in the left of the plot corresponds to: yield = 0.1905, batch cycle time = 292.8 h and initial broth volume = 5002.3 L. The initial broth volume is very close to its lower bound. The gap between these two points indicates the difficulty in finding the Pareto-optimal points due to closeness to the lower bound and flat nature of the front.

Among the decision variables, initial broth volume has the greatest influence on the yield-batch cycle time Pareto (Figure 2.10). Higher (lower) initial broth volume results in shorter (longer) batch cycle time and lower (higher) yield. When the batch cycle lengthens and yield approaches the peak, the initial broth volume converges towards its lower bound. A relatively constant glucose feed concentration of 78.3 g/L drives the yield from 0.123 to 0.183 when the batch cycle time increases from 141 to 200 hours. As the batch cycle time increases from 200 to 293 hours, the glucose feed concentration rises steadily from 78.3 to 112.2 g/L. For batch cycle that lasts longer than 200 hours, the initial biomass and initial glucose concentrations generally increase to allow accumulation of sufficient biomass to a level that sustains metabolic

activities and penicillin formation in the longer run. Unlike the influence of initial broth volume and glucose feed concentration on the optimal Pareto, the impact of switchover time and initial corn steep liquor concentration is relatively short-term and confined within the initial biomass growth phase.

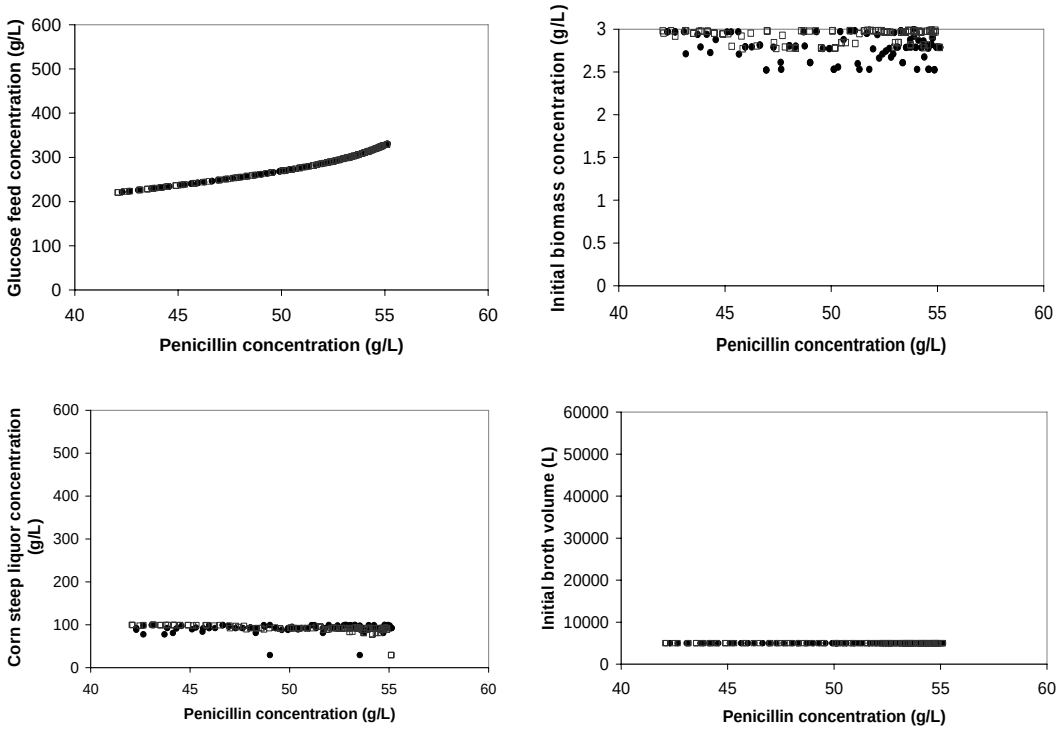


Fig. 2.8. Decision variables corresponding to the Pareto-optimal fronts in Fig. 2.5 when switchover time and initial glucose concentration are fixed at 15 h and 100 g/L, respectively, for two random seeds: 0.6 (•) and 0.7 (□).

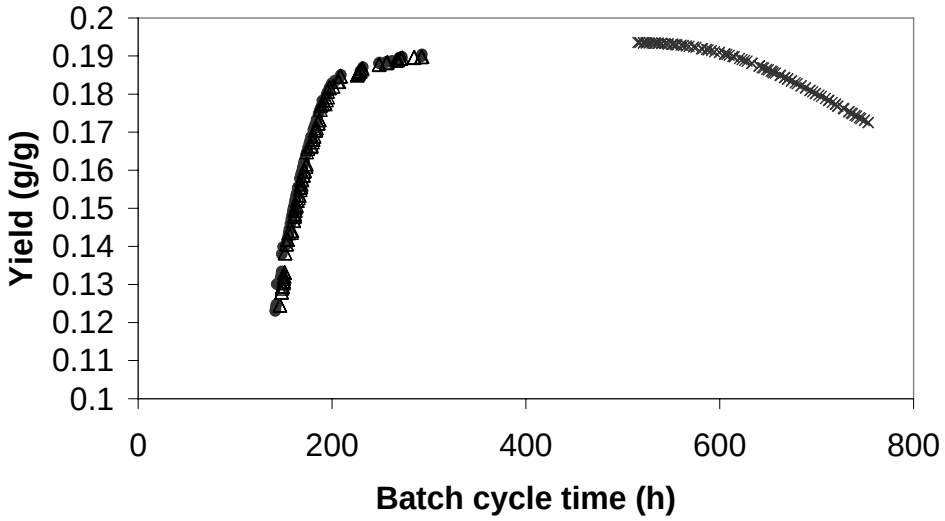


Fig. 2.9. Pareto-optimal front obtained at 500 generations (•) for the simultaneous maximization of yield and minimization of batch cycle time using random seed of 0.6. Also shown are: solutions at 400 generations (Δ), and yield and *calculated* batch cycle time corresponding to the Pareto-optimal front in Fig. 2.5 (×).

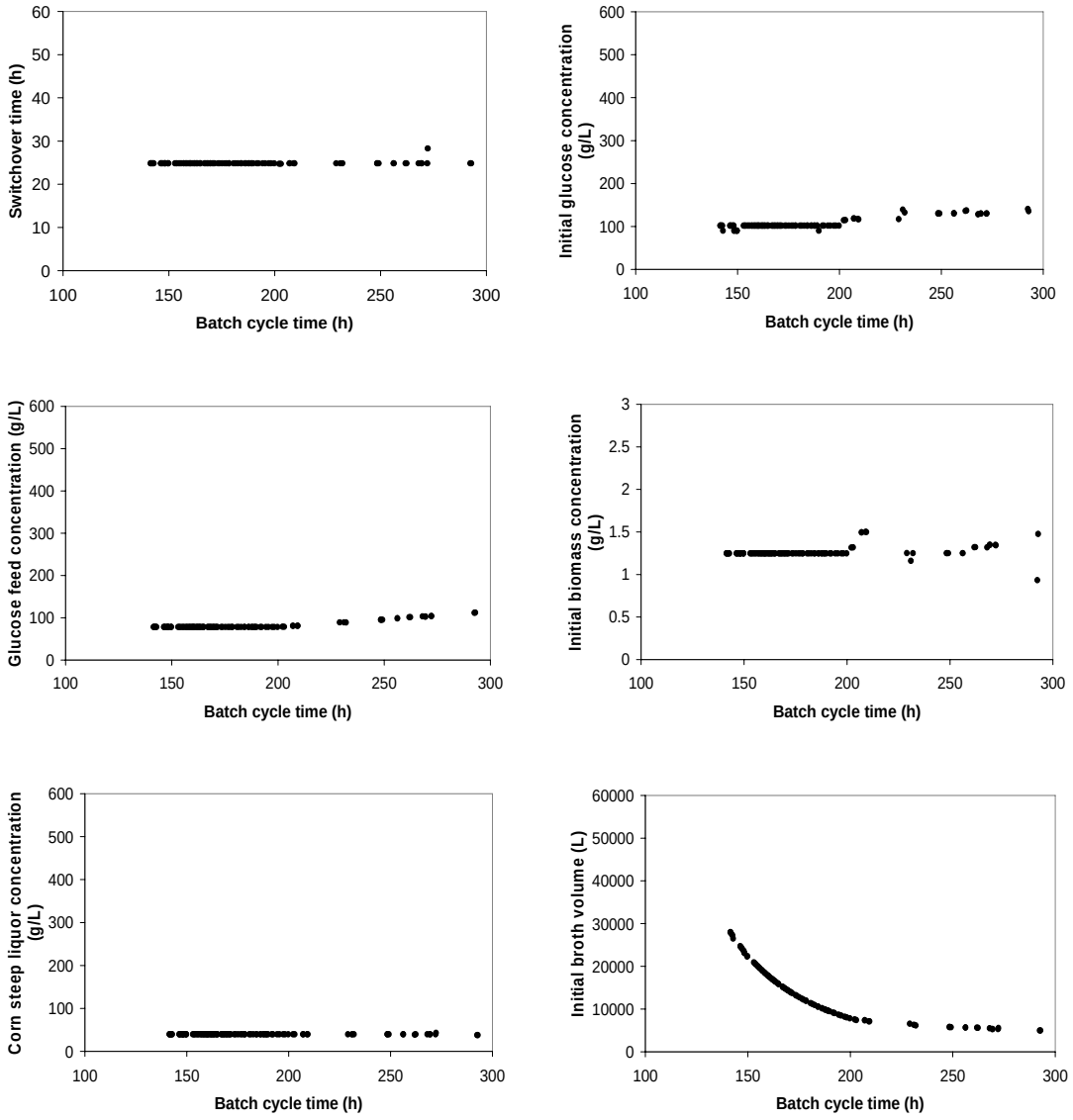


Fig. 2.10. Decision variables corresponding to the Pareto-optimal front in Fig. 2.9.

2.6.3 Tri-Objective Optimization

In the next part of the study, yield, profit and penicillin concentration are maximized simultaneously using NSGA-II. As before, almost identical Pareto-optimal fronts are obtained regardless of the random seed used. For tri-objective optimization, we decided to use 0.5 as the random seed after several preliminary runs, to balance computation time and accuracy of the solutions obtained. The results obtained at 400 generations (not shown) are very close to the Pareto-optimal front at 500 generations (Figure 2.11). Profit is understandably higher when there is a greater selectivity of

penicillin formation relative to biomass formation and maintenance, which means higher yield and lower raw material cost. Yield and profit are hence positively and strongly correlated (Figure 2.11a) for the Pareto-optimal solutions from both tri- and bi-objective optimization. As explained in the bi-objective optimization study above, higher yield is obtainable at the expense of penicillin concentration. The same pattern is observed in the Pareto-optimal front for tri-objective optimization (Figure 2.11b). Nevertheless, operating cost is driven down by shorter fermentation time (meaning lower cost) typically associated with lower penicillin concentration. Hence, lower penicillin concentration results in greater profit as shown in Figure 2.11c. On the contrary, operating cost increases and profit decreases as broth volume and penicillin concentration rise. Both yield and profit fall as penicillin concentration (Figure 2.11) and fermentation time (Figure 2.13) increases.

The Pareto-optimal front obtained in the bi-objective maximization of yield and penicillin concentration is almost identical to that obtained in the tri-objective maximization of yield, profit and penicillin concentration (Figure 2.11). For a given profit, yield of the tri-objective optimization is marginally higher than that from the bi-objective optimization (Figure 2.11a). In both cases, glucose feed concentration is the main factor contributing to the Pareto-optimal solutions (Figure 2.12), and the initial broth volume is actively constrained at the lower bound. In contrast, influence of switchover time, initial glucose concentration, initial biomass concentration and initial corn steep liquor concentration on Pareto-optimal yield, profit and penicillin concentration is relatively short term. Differences in decision variable values in Figure 2.12 from bi- and tri-objective optimization may be due to multiple solutions. The switchover time from batch to continuous mode for bi-objective optimization is longer than that for tri-objective optimization (Figure 2.12), and its marginal effect on fermentation time to achieve a particular penicillin concentration is clear from Figure 2.13. The broth volume from bi-objective optimization at a particular fermentation time is higher than that from tri-objective optimization (Figure 2.14). This could be due to non-explicit intention to maximize profit in the bi-objective optimization. Thus, for a given fermentation time in the Pareto set (Figure 2.13), tri-objective optimization requires a lower broth volume than that in bi-objective optimization.

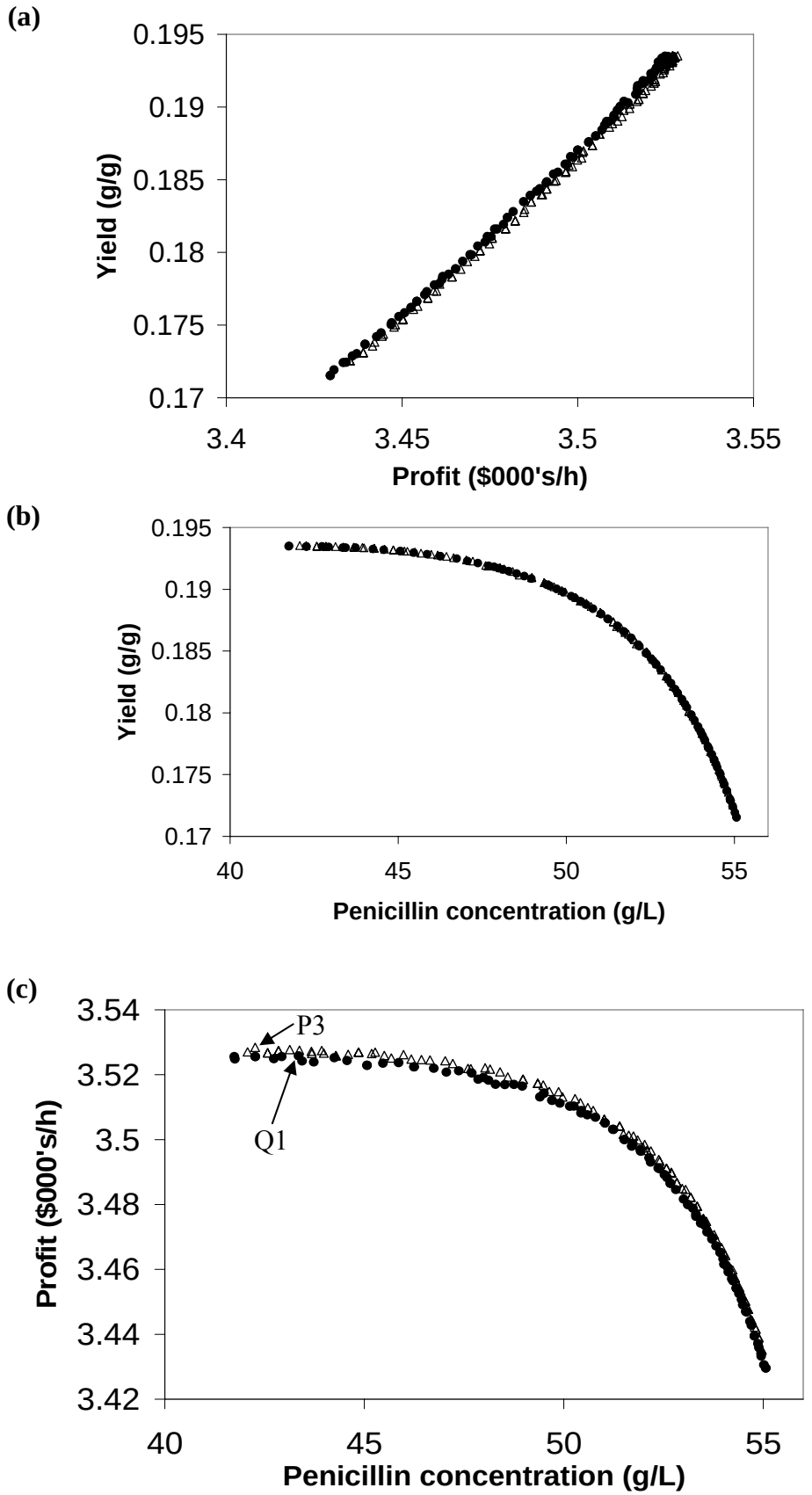


Fig. 2.11. Pareto-optimal front obtained at 500 generations for the simultaneous maximization of yield, profit and penicillin concentration (•). Also shown are Pareto-optimal front in Fig. 2.5 along with *calculated* profit for comparison (Δ).

In Figure 2.12, the initial broth volume is actively constrained at its lower bound of 5000 L. The glucose feed concentration increases steadily from 217.7 to 325.4 g/L (219.3 to 322.9 g/L for the bi-objective optimization when the random seed is 0.6). The final broth volume in a bioreactor is constrained at 200,000 L when the fermentation time is about 690 hours (Figure 2.14). The lowest broth volume occurs at 177,562 L (point M1 in Figure 2.14) when the fermentation time is 504.1 hours. However, it does not correspond to maximum profit. Although broth volume rises gradually to 178,830 L (point M2 in Figure 2.14) when the fermentation time decreases from 504.1 to 465.8 hours, the cost does not increase in line with broth volume. Detailed calculations indicate that glucose is the major cost component.

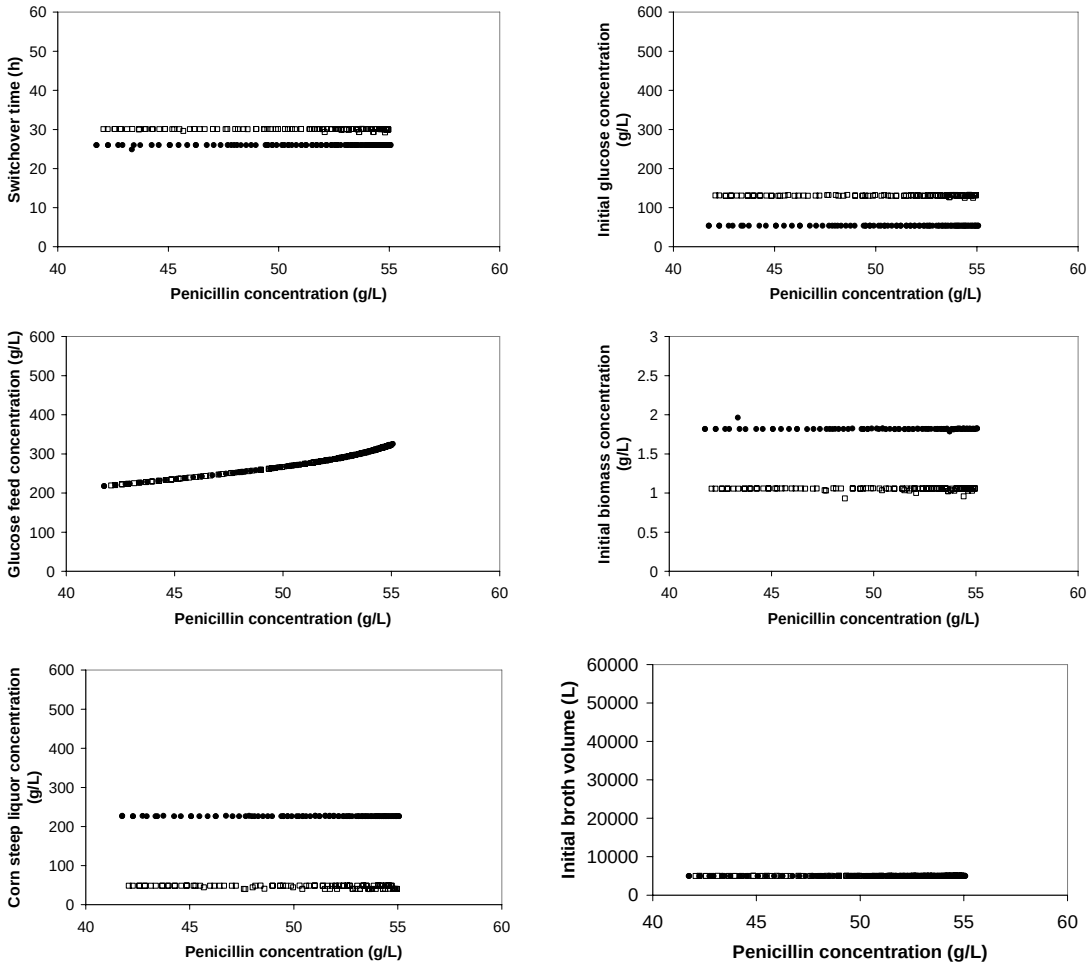


Fig. 2.12. Decision variables corresponding to the Pareto-optimal fronts in Fig. 2.11 (•) and the Pareto-optimal front in Fig. 2.5 (□).

The glucose feed concentration corresponding to M1 and M2 are 239.25 g/L (with profit of $3525.1/h) and 220.30 g/L (with profit of $3527.0/h) respectively. The

relatively lower glucose cost and to some extent shorter batch cycle time at M1 countervails the relatively higher cost encountered in batch and continuous modes. Hence, a particular broth volume (178,812 L) on the left of M1 (M3 in Figure 2.14), corresponds to a profit of $3528.4/h, as depicted by the chromosome "P3" in Figure 2.11(c). In the tri-objective optimization, the lowest broth volume (point L1 in Figure 2.14) corresponds to a profit of $3525.8/h, as depicted by the chromosome "Q1" in Figure 2.11(c). Although glucose feed concentrations pertaining to the points on the left side of L1 are lower than that of L1, the glucose cost differences are slightly overridden by other cost components linked to broth volume.

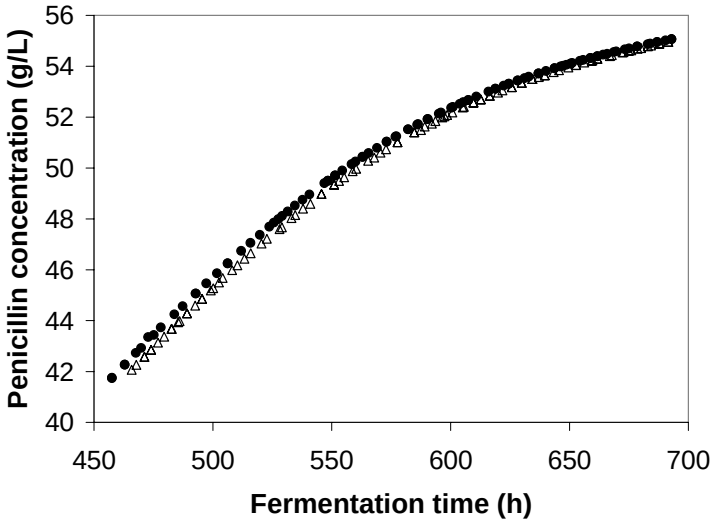


Fig. 2.13. Penicillin concentration versus the fermentation time corresponding to the Pareto-optimal fronts in Fig. 2.5 (Δ) and Fig. 2.11 (•).

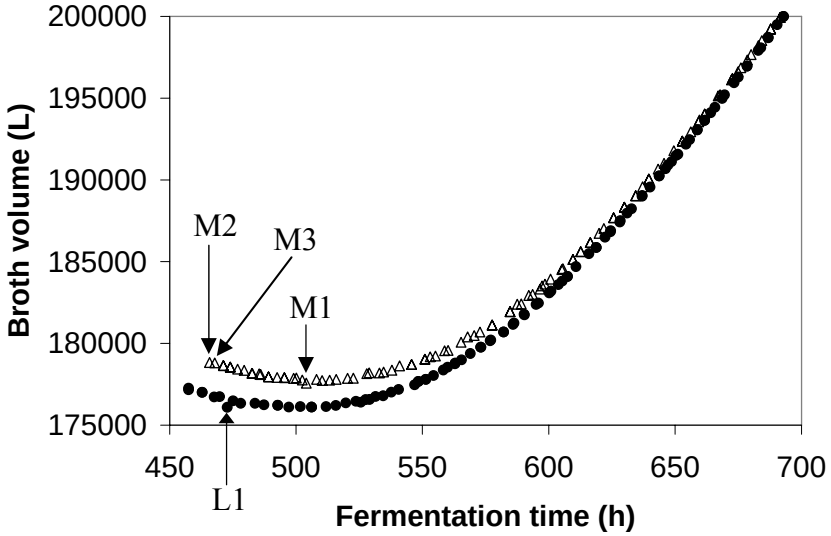


Fig. 2.14. The broth volume and fermentation time corresponding to the Pareto-optimal fronts in Fig. 2.5 (Δ) and Fig. 2.11 (•).

2.7 Conclusions

Optimization of an industrial penicillin V bioreactor train with a targeted continuous production rate was successfully performed using NSGA-II, for both bi- and tri-objective scenarios. The bi-objective maximization of yield and penicillin concentration results show that higher penicillin concentration is obtainable at the expense of lower yield by stretching the batch cycle time. In another case involving the concurrent minimization of batch cycle time and maximization of yield, higher yield is linked to longer batch cycle time. In maximizing yield, penicillin concentration and profit simultaneously, higher profit is associated with higher yield and lower penicillin concentration. Results of bi-objective maximization of yield and penicillin concentration are similar to those of tri-objective maximization of yield, profit and penicillin concentration. Glucose feed concentration has the greatest influence on the simultaneous maximization of yield and penicillin concentration and on the tri-objective optimization. In both scenarios, initial broth volume is actively constrained at its lower bound. In contrast, initial broth volume dictates the optimal Pareto of the simultaneous maximization of yield and minimization of batch cycle time. The optimization study reported here demonstrates the potential of MOO for penicillin production; but, owing to model accuracy and variation of cost data, it is necessary to fine tune model parameters and update cost data for optimizing any penicillin plant for multiple objectives.

Chapter 3
OPTIMIZATION OF A MULTI-PRODUCT MICROBIAL CELL FACTORY FOR MULTIPLE OBJECTIVES – USING CENTRAL CARBON METABOLISM MODEL

3.1 Introduction

Emerging mathematical models for multi-product microbial cell factories such as *Escherichia coli, Corynebacterium glutamicum* and *Saccharomyces cerevisiae* enable chemical engineers to extend the reach of their competencies in systems area such as optimization and process control to biotechnology and biochemical manufacturing. *Corynebacterium glutamicum* is used commercially to produce amino acids (Bongaerts et al., 2001; Eggeling et al., 2006) such as L-glutamate (leading to monosodium glutamate, MSG), L-lysine (an animal feed additive), L-phenylalanine (a nutraceutical, a flavour enhancer and an intermediate for synthesis of pharmaceuticals), tryptophan (an animal feed additive and a nutritional ingredient in milk formula for human infants) and L-aspartate (a food additive and a sweetener). Likewise, *Saccharomyces cerevisiae* (yeast) is used commercially to produce ethanol and carbon dioxide (for the baking process in the food industry) as well as bio-fuel. Depending on the bacteria strain, *E. coli* is potentially capable of producing more than twenty types of amino acids (Eggeling et al., 2006; Lee et al., 2009; Leuchtenberger et al., 2005).

Optimization of a multi-product microbial cell factory, a systems biotechnology specialty, is increasingly useful in predicting feasible outcomes that fulfill specified objectives, in tandem with rising reliability of the mathematical models for describing the microbial cell metabolic pathways (Lee et al., 2005a). This methodology complements the well-developed experimental procedures in classical strain development, genomic techniques and intra-cellular flux analysis, used in engineering a microbial cell factory targeted for industrial production. Rising demand is driving the annual growth for amino acids. Hence, improvements in the production of amino acids are of considerable importance to both industries and consumers (Scheper et al., 2003). This chapter proposes a mixed-integer multi-objective optimization (MIMOO) study to find a range of better metabolic pathway recipe for improving amino acids production using *E. coli*.

In the area of bioprocess modelling and optimization, biochemists such as Voit (2000) have compiled a list of modelling works related to bioprocesses using S-system (or synergistic system) to represent the metabolic kinetics of cell factories. The S-system models were then adopted for single objective optimization of citric acid production in *Aspergillus niger*, ethanol production in *Saccharomyces cerevisiae* and tryptophan production in *E. coli* (Torres and Voit, 2002). Conflicting objectives are commonly encountered in bioprocesses (Lee et al., 2007; Chapter 1). Up to now, there have been several works in the MOO of bioprocesses but none, to the best of the thesis author's knowledge, on the MOO of a multi-product microbial cell factory in the form of *E. coli*. See Chapter 1 for the reported applications of MOO in bioprocesses.

Almost all works on optimization of a multi-product microbial cell factory focussed on a single objective (e.g., Schmid et al., 2004; Visser et al., 2004; Vital-Lopez et al., 2006). A common feature in these works is the pseudo-stationary assumption. Enzymatic reaction kinetics in a microbial cell factory are reversible and interdependent. In reality, the fluxes due to enzymatic reactions are never stationary. Given the limitations of a model, it is necessary to assume a pseudo-stationary state where some variables fluctuate about an averaged steady state within certain bounds.

Schmid et al. (2004) used a nonlinear kinetic model to maximize tryptophan production via enzyme modulations. In their study, the results obtained from piece-wise optimization were combined selectively to form the results of the integrated optimization. Visser et al. (2004) used a lin-log kinetic model of *E. coli* to determine the optimal glycolytic (Section 3.2) enzyme modulations required to either maximize glucose uptake through phosphotransferase sub-system (PTS) or the production of serine. In this study, only ten (and eleven in the case of maximizing serine production) out of the thirty enzymatic fluxes present in the complete model were the decision variables. These two recent works of nonlinear programming (NLP) illustrate the potential applications of systems biotechnology in generating metabolic pathway recipe.

Vital-Lopez et al. (2006) used a linearized kinetic model, possibly anticipating a complex problem to be solved, as the basis for maximizing serine production.

In their study, gene overexpression/repression and knockout[1] are considered for the whole model – an example of mixed-integer nonlinear programming (MINLP). Uncertainty of their results generally increases due to approximation when the optimization search domain recedes further from the initial steady state conditions. Though they have proposed a procedure for unconstrained optimization incorporating both linearized and non-linear kinetic models, its applicability and effectiveness for multi-objective MINLP remain untested.

There has been little work on MOO of multi-product microbial cell factories. Vera et al. (2003) have studied MOO in metabolic processes leading to ethanol production by *Saccharomyces cerevisiae*. Ethanol production which is driven by the enzyme sub-system pyruvate kinase (PK) was maximized and the concentrations of various intermediate metabolites (intra-cellular glucose, g6p, fdp, pep and atp – full form of all abbreviations used is given in the List of Symbols) were minimized under pseudo-stationary conditions where the five metabolite concentrations are assumed to be time-invariant. This allows the system of six differential equations to be converted into an equivalent S-system (synergistic system) which is then linearized by applying natural logarithms to PK kinetic expression (an objective function) and the various influx and efflux terms associated with each of the five metabolites (pseudo-stationary constraints). While searching for a Pareto-optimal set, the metabolites concentrations and enzymes activities – temporally invariant under pseudo-stationary conditions – vary within their respective lower and higher bounds. Enzymes levels (equivalently genes overexpression or repression) and metabolites concentrations were manipulated in the study of Vera et al. (2003) without explicit consideration of the impact of gene knockout – a case of multi-objective linear programming (MOLP).

Lee et al. (2005b) evaluated the correlation between maximum biomass and succinic acid production for various combinatorial gene knockout strains. This sets the stage for the simultaneous maximization of biomass and succinic acid production using the ε-constraint method. In this method, an MOO problem is converted into an equivalent single objective problem by constraining all objectives except one to be within specified limits, and then the resulting single objective problems are solved using a suitable NLP or MINLP method.

[1] Gene expression is the process by which DNA sequence of a gene is converted into functional proteins. Many proteins are enzymes that catalyze biochemical reactions vital to metabolism. Gene overexpression (e.g. copying genes) creates larger quantity of an enzyme. Gene repression or knockdown reduces the quantity of an enzyme; gene knockout deletes an enzyme-producing gene.

In contrast to the widely used continuous processes to produce large quantities of a few products, the strategic importance of MOO of a microbial cell factory is elevated as the challenge to dynamically cater to various market segments and the competition increase. MOO as a first line predictor, has the potential to shorten the time to develop new commercial strains when used in conjunction with the well-established experimental procedures. Clearly, there has been very little MIMOO study of a multi-product microbial cell factory. This motivated the thesis author to optimize a multi-product microbial cell for multiple objectives using the elitist non-dominated sorting genetic algorithm (NSGA-II, Deb et al., 2002), which has been successfully employed for many chemical engineering applications.

3.2 Central Carbon Metabolism of *Escherichia coli*

Figure 3.1 shows the metabolic network of the central carbon metabolism of *Escherichia coli*. It depicts 30 enzymatic sub-systems (shown in rectangles), 18 metabolites or precursors in between the enzymatic sub-systems and 7 co-metabolites (amp, adp, atp, nadp, nadph, nad and nadh). Enzymes are shown in rectangles; precursors (balanced metabolites) are in bold between enzymes; allosteric effectors (atp, adp and fdp), activators (positive sign), inhibitors (negative sign) and regulators (without sign) are given in circles/ellipses. The glycolytic (consisting of PTS, PGI, PFK, ALDO, TIS, GAPDH, PGK, PGluMu, ENO, PK and PDH enzymatic sub-systems) and pentose-phosphate (consisting of G6PDH, PGDH, Ru5P, R5PI, TKa, TA and TKb enzymatic sub-systems) metabolic pathways are central channels of carbon fluxes. The fluxes of the enzymatic sub-systems have strong influences in the form of feedback regulation (example: changes in PEPCxylase flux affect both the serine and aromatic amino acids synthesis, which occur earlier in the pathway) and feedforward regulation (example: changes in G6PDH flux affect the aromatic amino acids synthesis, which occurs later in the pathway). Higher order effects such as cascade and combined feedback-feedforward regulations are also embedded in the model. Further, metabolites and co-metabolites regulate the enzymatic sub-systems (example: pep has negative regulatory effect on PFK flux whereas adp and amp have positive regulatory effects on PFK flux); these effects are shown in circles next to the enzymatic sub-systems.

The nonlinear dynamic model of the central carbon metabolism of *E. coli* formulated by Chassagnole et al. (2002) is selected, to study the effects of genes/enzymes knockouts, overexpression and repression on amino acids synthesis. This detailed model consists of 18 nonlinear differential equations (arising from mass balances) and 30 nonlinear rate equations for the enzymatic sub-systems, which take into account the impacts of gene expression. No differential mass balance equations are available for the 7 co-metabolites; and concentrations of these co-metabolites are assumed to be constant. The co-metabolites (also known as co-factors) contain the food and energy needed to sustain the metabolism of *E. coli*. Co-metabolites are non-protein chemical compounds that are bound to enzymes and can be considered as helpers in biochemical reactions. The microbial cells consume as well as re-generate the co-factors. Due to the biological need of the cell to maintain the concentration of the co-factors and the cyclic nature of co-factor consumption and re-generation, we assume constant concentration for co-metabolites. Amounts of co-metabolites needed are likely to change following gene manipulations, but their total concentrations remain fairly constant at given gene expression levels. Future optimization studies may account for the effects of gene manipulations on co-metabolite concentrations.

Model equations in Chassagnole et al. (2002) are not repeated for brevity. However, these equations along with values of parameters in them are available from the thesis author and the CD of the recent book (Rangaiah et al., 2009). The kinetic parameters (except for the maximum enzymatic reaction rates), experimentally measured initial steady state values of the metabolites/co-metabolites and fed-batch process parameters are available in Chassagnole et al. (2002). The maximum enzymatic reaction rates are taken from an online resource (http://jjj.biochem.sun.ac.za/database/index.html; accessed in May 2007). The thesis author solved the model equations using DIVPRK program in the IMSL software, with an integration step of 0.1 sec and a glucose pulse (height = 16 mM; width = 0.1 sec). The transient profiles of the metabolite concentrations and enzymatic reaction fluxes shown in Appendix A generally agree with those in Chassagnole et al. (2002). Initial steady state values of the 18 metabolites computed using the nonlinear equation solver, DNEQNF of the IMSL software, are shown in Table 3.1. They are close to the experimentally measured initial steady state values of Chassagnole et al. (2002). All these confirm the validity of the model equations, parameters and programs used in this study.

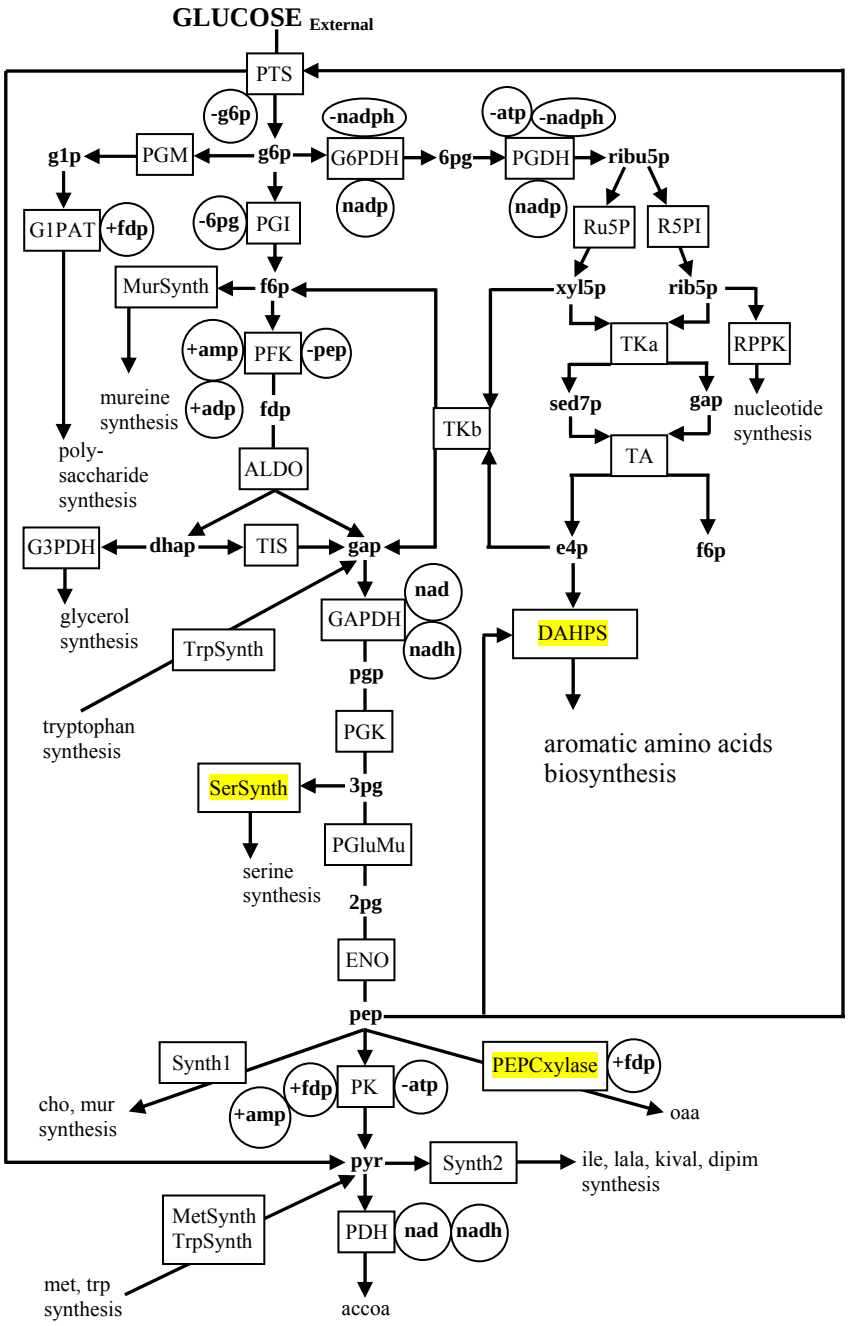


Fig. 3.1. Metabolic network of the central metabolism of *Escherichia coli*. Enzymes are shown in rectangles; precursors (balanced metabolites) are in bold between enzymes; allosteric effectors (atp, adp and fdp), activators (positive sign), inhibitors (negative sign) and regulators (without sign) are given in circles/ellipses. All abbreviations are defined in the List of Symbols.

3.3 Formulation of the MOO Problem

The DAHPS enzymatic sub-system is the first among a series of steps in the aromatic amino acids (tryptophan, phenylalanine and tyrosine) synthesis pathways. The PEPCxylase enzymatic sub-system produces the oxaloacetate (oaa precursor in Figure 3.1) to generate the aspartate precursor (not shown in Figure 3.1) needed for the lysine, methionine, threonine and isoleucine synthesis. The SerSynth enzymatic sub-system governs the steps leading to serine synthesis pathways. The complex interactions among the DAHPS, PEPCxylase and SerSynth enzymatic sub-systems exert definite and possibly conflicting influences on the synthesis of three distinct groups of amino acids (Figure 3.1).

Maximizing DAHPS, PEPCxylase and SerSynth enzymatic flux ratios are expected to enhance the desired amino acids synthesis rates. Here, enzyme flux refers to the reaction rate facilitated by that enzyme. Two bi-objective scenarios are studied in the current chapter.

Case A: Maximize DAHPS flux ratio and PEPCxylase flux ratio (3.1)

Case B: Maximize DAHPS flux ratio and SerSynth flux ratio (3.2)

The flux ratio is the ratio of a flux after genetic engineering (to knock out, overexpress and/or repress the genes regulating the enzymatic sub-systems) to that of the reference system before genetic engineering (referred to as wild strain by biotechnologists). The calculated initial metabolite/co-metabolite concentrations and steady-state fluxes (given in Table 3.1) are used as reference values in the MIMOO study, by setting the time derivatives of the metabolites to zero.

The two crucial system constraints in the optimization are homeostasis and total enzymatic flux (Schmid *et al*., 2004). First is the homeostatic constraint:

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\left|C_i - C_{i,ref}\right|}{C_{i,ref}} \leq 0.3 \qquad (3.3)$$

The summation is over all metabolites (m in number, 18 in our case). C_i is the concentration of i^{th} metabolite. $C_{i,ref}$ is the reference concentration of the i^{th} metabolite given in Table 3.1. The principle of homeostasis requires the microbial cell to maintain intra-cellular metabolite concentrations within certain bounds (± 30% in our study) - a physiological constraint so that the microbial cell does not suffer from

Table 3.1 Initial metabolite/co-metabolite concentrations and steady-state fluxes of enzymes used as reference values in the homeostasis and total enzymatic flux constraints. Experimentally measured values of Chassagnole et al. (2002) are in brackets. The co-metabolite concentrations are assumed to be constant.

Metabolite/Co-metabolite with serial number	Concentration (mM)	Enzyme with serial number	Flux (mM/s)
Metabolite		1. PTS	0.2000
1. Glucose (extracellular)	0.05549 (0.0556)	2. PGI	0.05825
2. g6p	3.4767 (3.48)	3. PFK	0.1410
3. f6p	0.5994 (0.60)	4. ALDO	0.1410
4. fdp	0.2703 (0.272)	5. TIS	0.1394
5. gap	0.2173 (0.218)	6. GAPDH	0.3199
6. dhap	0.1665 (0.167)	7. PGK	0.3199
7. pgp	0.00798 (0.008)	8. PGluMu	0.3023
8. 3pg	2.1268 (2.13)	9. ENO	0.3023
9. 2pg	0.3982 (0.399)	10. PK	0.03811
10. pep	2.6648 (2.67)	11. PDH	0.1878
11. pyr	2.6689 (2.67)	12. PEPCxylase	0.04312
12. 6pg	0.8138 (0.808)	13. PGM	0.002319
13. ribu5p	0.1108 (0.111)	14. G1PAT	0.002301
14. xyl5p	0.1378 (0.138)	15. RPPK	0.01031
15. sed7p	0.2760 (0.276)	16. G3PDH	0.001658
16. rib5p	0.3974 (0.398)	17. SerSynth	0.01749
17. e4p	0.09776 (0.098)	18. Synth1	0.01421
18. g1p	0.6520 (0.653)	19. Synth2	0.05355
		20. DAHPS	0.006836
Co-metabolite		21. G6PDH	0.1393
1. amp	0.955 (0.955)	22. PGDH	0.1393
2. adp	0.595 (0.595)	23. Ru5P	0.08370
3. atp	4.27 (4.27)	24. R5PI	0.05559
4. nadp	0.195 (0.195)	25. TKa	0.04527
5. nadph	0.062 (0.062)	26. TKb	0.03843
6. nad	1.47 (1.47)	27. TA	0.04526
7. nadh	0.1 (0.1)	28. MurSynth	0.00043711
		29. MetSynth	0.0022627
		30. TrpSynth	0.001037

toxic or inhibitory effects - to avoid impediment of cellular functions and undesirable flux diversions. Large changes in metabolite concentrations cause unforeseeable effects on gene expression (and hence kinetic rate parameters) that are not captured in the existing model.

The second is the total enzymatic flux constraint:

$$\frac{1}{z}\sum_{i=1}^{z}\frac{r_i}{r_{i,\,ref}} \le 1.0 \tag{3.4}$$

This is a technological constraint, where the summation covers all enzyme fluxes (z in number, 30 in our case). $r_{i,ref}$ is the i^{th} reference enzymatic reaction rate given in Table 3.1. Total enzymatic activity is constrained not to exceed 1.0 to avoid diffusion problem (due to increased cytoplasm viscosity), protein precipitation, secondary kinetic effects (due to steric hindrance) and excessive intracellular stress leading to unpredictable regulatory effects. When either one of these two constraints is breached, the objective function value is penalized by setting it to an arbitrarily low level; under such conditions, the DAHPS, PEPCxylase and SerSynth fluxes are set to 10^{-20}.

Metabolite concentrations and enzymatic fluxes change from one steady state to another due to gene knockouts, overexpression and/or repression. Redistribution of the fluxes presents opportunities for optimizing the metabolic pathways subject to physiological and technological constraints. Translation of gene knockouts, overexpression and repression into decision variables is described in the next section. In this chapter, gene knockouts and overexpression/repression are considered separately. Simultaneous knockouts and manipulation of genes is discussed in Chapter 5.

3.4 Techniques Used in Solving MIMOO Applications

By setting the time-derivative of each metabolite concentration to zero under pseudo-stationary assumption, the set of differential equations for mass balance equations is converted into a system of algebraic equations. Each nonlinear equation contains several rate expressions and terms. The glucose impulse term, f_{pulse}, in the mass balance equation $\left(\frac{dC_{glc}^{extracellular}}{dt} = D\left(C_{glc}^{feed} - C_{glc}^{extracellular}\right) + f_{pulse} - \frac{C_X r_{PTS}}{\rho_X}\right)$ in the work of Chassagnole et al. (2002) is used to generate transient profiles using the original system of differential equations, and hence it is not relevant to the MIMOO study.

The system of 18 algebraic equations is solved for 18 metabolite concentrations, using the DNEQNF program in the IMSL FORTRAN libraries. The entire study was done using a personal computer with 2 GHz Pentium$^{(R)}$ IV CPU, 1 GB RAM and Windows XP Professional. Each solution of the system of algebraic equations took no more than 3 seconds of CPU time on this computer; each optimization run for 500 generations (using NSGA-II as described later in this section) required less than 20 minutes of CPU time for each of the bi-objective cases.

One main difficulty encountered in the optimization of the microbial cell factory is to identify the enzymes to be knocked out. Enzymes cannot be deleted arbitrarily – certain enzymes are essential for the metabolic network integrity and stability. An attempt to delete essential enzymes results in the termination of the DNEQNF program and consequently the optimization program too. To overcome this problem, feasible sets of 1-enzyme, 2-enzyme and 3-enzyme knockouts are identified through a manual combinatorial exercise (by setting the maximum reaction rates of the selected enzymatic sub-systems to zero and solving the model equations). Although this takes considerable effort and time, identifying enzymes which can be deleted, whether singly or in groups, circumvents numerical difficulties in the MIMOO study where simultaneous gene manipulation and knockouts is applied. The number of feasible sets (with all combinations in brackets) of 1-, 2- and 3-enzyme knockouts are 15 (30), 114 (435) and 665 (4060) respectively. These sets are neither available in the literature nor known *a priori*. This manual combinatorial exercise provides the Pareto-optimal sets by enzyme knockouts as well. An alternative to the manual exhaustive search in gene knockouts optimization is the interactive branch-and-bound (Section 3.6) technique. A strategy to identify a subset of genes that could be knocked out prior to simultaneous gene manipulation and knockouts is through heuristics (Chapter 5). The subset of gene obtained by heuristics is more targeted for the simultaneous gene manipulation and knockouts optimization. Many of the genes within the larger feasible sets obtained through a manual exhaustive search breach the homeostatic and/or total enzymatic flux constraints.

Gene manipulations (overexpression and/or repression) are optimized using the NSGA-II and the FORTRAN program containing the model and its solution. Decision variables can be implemented with binary or real coding in the NSGA-II program. Decision variables in the form of integers from 1 to 30 are used to denote the enzymatic sub-systems (or simply enzymes). A 5-bit binary variable which covers

integers ranging from 1 to 32 (where 31 and 32 are not used) is used as a decision variable for 1-enzyme manipulation. Two or more 5-bit variables are needed in the multi-enzyme manipulation. For gene overexpression/repression, real decision variables in the range 0.5 to 2.0 are also used to multiply the maximum enzymatic reaction rates; these bounds are selected after preliminary optimization runs with several ranges for these decision variables. In this study, the number of gene knockouts or expression for optimization is limited to a maximum of 3 due to potential difficulties in achieving more knockouts/expression experimentally. Translation of gene knockouts and expression into decision variables and their implementation in the optimization can easily be done without any concern on continuity since evolutionary algorithms such as NSGA-II are applicable to non-differentiable functions.

Using the glucose-6-phosphate dehydrogenase (G6PDH) enzymatic sub-system as an example, reaction rate of G6PDH (i.e., rate of reaction facilitated by G6PDH) is given as:

$$r_{G6PDH} = \frac{r^{max}_{G6PDH} C_{g6p} C_{nadp}}{\left(C_{g6p} + K_{G6PDH,g6p}\right)\left(1 + \frac{C_{nadph}}{K_{G6PDH,nadph,g6pinh}}\right)\left(K_{G6PDH,nadp}\left(1 + \frac{C_{nadph}}{K_{G6PDH,nadph,nadpinh}}\right) + C_{nadp}\right)} \tag{3.6}$$

where the maximum enzymatic reaction rate is r^{max}_{G6PDH}. The integer number representing an enzyme follows the sequence given in Table 3.1. Therefore, G6PDH is numbered 21. The integer number which is a discrete decision variable is selected randomly by NSGA-II. For a selected enzymatic sub-system, its maximum reaction rate is set to zero in gene knockout. In 1-enzyme manipulation study, two decision variables are involved: an integer number representing an enzymatic sub-system and a real number representing gene overexpression or repression. If G6PDH (numbered 21) were selected, its maximum reaction rate will be multiplied by a real number in the range from 0.5 to 2.0, selected by NSGA-II. Gene is overexpressed or repressed if the chosen real number is greater or less than 1.0 respectively.

NSGA-II parameters used in this study are: maximum number of generations (up to 500), population size (100 chromosomes), probability of crossover (0.85), probability of mutation (0.05), distribution index for the simulated crossover operation (10), distribution index for the simulated mutation operation (20) and random seed (0.6). Except for the first and last parameter listed here, rest of the NSGA-II parameter values are taken from Tarafder et al. (2005). Values for

maximum number of generations and random seed are obtained by trial and error. Our preliminary gene manipulation optimization runs show convergence within 500 generations for the random seed of 0.6.

3.5 Optimization of Gene Knockouts

The Pareto-optimal metabolic pathway recipe through multi-gene knockout combinations shows that the triple-gene knockout has the best non-dominated flux ratios due to greater flexibility in manipulating fluxes for the various pathways (Figure 3.2). Deleting PGM in the single-gene knockout generates the Pareto-optimal flux ratios for both the bi-objective scenarios (Figure 3.2) subject to the homeostatic and total enzyme activity constraints. No flux appears in PGM, and the fdp activation results in negligible G1PAT flux. Fluxes of the glycolytic (consisting of PTS, PGI, PFK, ALDO, TIS, GAPDH, PGK, PGluMu, ENO, PK and PDH) and pentose-phosphate pathway (starting from G6PDH) undoubtedly increase.

PGM and G6PDH are the main catalysts that channel carbon sources for the polysacharride synthesis and pentose-phosphate pathway respectively. The pentose-phosphate pathway has a higher carbon utilization rate and level than that needed in polysacharride synthesis. Under unconstrained condition, deleting G6PDH which is a major user of carbon sources, leads to a significant increase in the Pareto-optimal fluxes. However, knocking out G6PDH violates the total enzymatic flux constraint. If another gene is also knocked out, the total enzymatic flux and homeostatic constraints are breached in 54% and 77% of the cases, respectively.

Deleting PK and G1PAT (Figure 3.2 – chromosome B_1) or G6PDH and MetSynth (Figure 3.2 - chromosome B_2) in the double-gene knockout generates Pareto-optimal flux ratios for Case B of the bi-objective scenarios; deleting G6PDH and MetSynth (Figure 3.2 – chromosome A_1) generate Pareto-optimal flux ratios for Case A also. Deleting G6PDH and MetSynth increases fluxes (Table 3.2) of the glycolytic pathway. In contrast, most of reactions in the pentose-phosphate pathway were not actively utilized as exhibited by zero fluxes of G6PDH and PGDH, largely attenuated fluxes of R5PI, TKa and TA, and inverse fluxes of RU5P and TKb. The inverse (or negative) fluxes are the result of product formation rate being greater than the reactant influx rate. Inverse fluxes signify that carbon sources are being drained from the pentose-phosphate pathway by the glycolytic pathway (since the first entry

point of the pentose-phosphate pathway G6PDH is deactivated) more quickly than they are replenished via the same pathway – this is equivalent to the backflow of carbon sources. TKb is a gateway supplying carbon from the pentose-phosphate pathway to the glycolytic precursors (f6p and gap). Inverse flux of TKb allows e4p to back up in the network. Accumulations of e4p (0.1306 mM) and pep (3.4248 mM) through inverse fluxes elevate the DAHPS flux to a large extent. Constant PTS flux and to a smaller extent MetSynth deletion assist in maintaining PK, PDH and Synth2 flux levels.

The DAHPS flux ratio is the highest among the three objectives due to relatively greater increase in its precursor concentrations (28% for pep and 33% for e4p) in comparison to the enhancing effects of precursors dictating the flux of PEPCxylase (28% for pep coupled with activation via fdp) and SerSynth (28% for 3pg). Double knockout (PK and G1PAT) results in an overall increase of the concentration levels of glycolytic precursors, slightly attenuating the glycolytic fluxes. The concentration levels of 3pg and pep, precursors for the three amino acids synthesis pathways, increase by 41%; the concentration level of e4p (a precursor of DAHPS) increases by 29%. SerSynth flux ratio increases as its precursor (3pg) concentration level increases from 28% to 41%. However, both the DAHPS and PEPCxylase flux ratios decrease due to self-regulatory effects, as their precursor concentration levels increase beyond the corresponding levels obtainable from deleting G6PDH and MetSynth.

Triple knockout of PK, G1PAT and G3PDH generates Pareto-optimal flux ratios in Case B of the bi-objective scenarios while deleting G1PAT, RPPK and DAHPS generates Pareto-optimal flux ratios in Case A. Similar to that of the double-gene knockout, deleting PK, G1PAT and G3PDH results in an overall increase of the glycolytic precursors concentration levels and a slight attenuation of the glycolytic fluxes. Also, the concentration levels of the three precursors (3pg, pep and e4p) increase by almost the same percentage points as those of the double-gene knockouts. In Case A, deleting G1PAT, RPPK and DAHPS maximizes the PEPCxylase flux ratio as the DAHPS flux ratio is set to zero. The glycolytic and pentose-phosphate pathway fluxes are amplified and the carbon sources of e4p and pep are diverted from DAHPS to the glycolytic pathway and PEPCxylase, respectively. Another Pareto-optimal metabolic pathway recipe in Cases A and B is obtained by deleting G6PDH, MurSynth and TrpSynth (chromosomes adjacent to A_1 and B_2 in Figure 3.2). Similar

to that of the 2-enzyme knockout, the DAHPS flux ratio is the highest among all the three scenarios. Except for the zero concentration of 6pg, the concentration levels of the remaining metabolites are elevated. There is an overall flux increase in glycolytic pathway, polysaccharide synthesis, nucleotide synthesis and glycerol synthesis. The pentose-phosphate pathway exhibits zero fluxes (G6PDH and PGDH), largely attenuated fluxes (R5PI, TKa and TA) and inverse fluxes (RU5P and TKb).

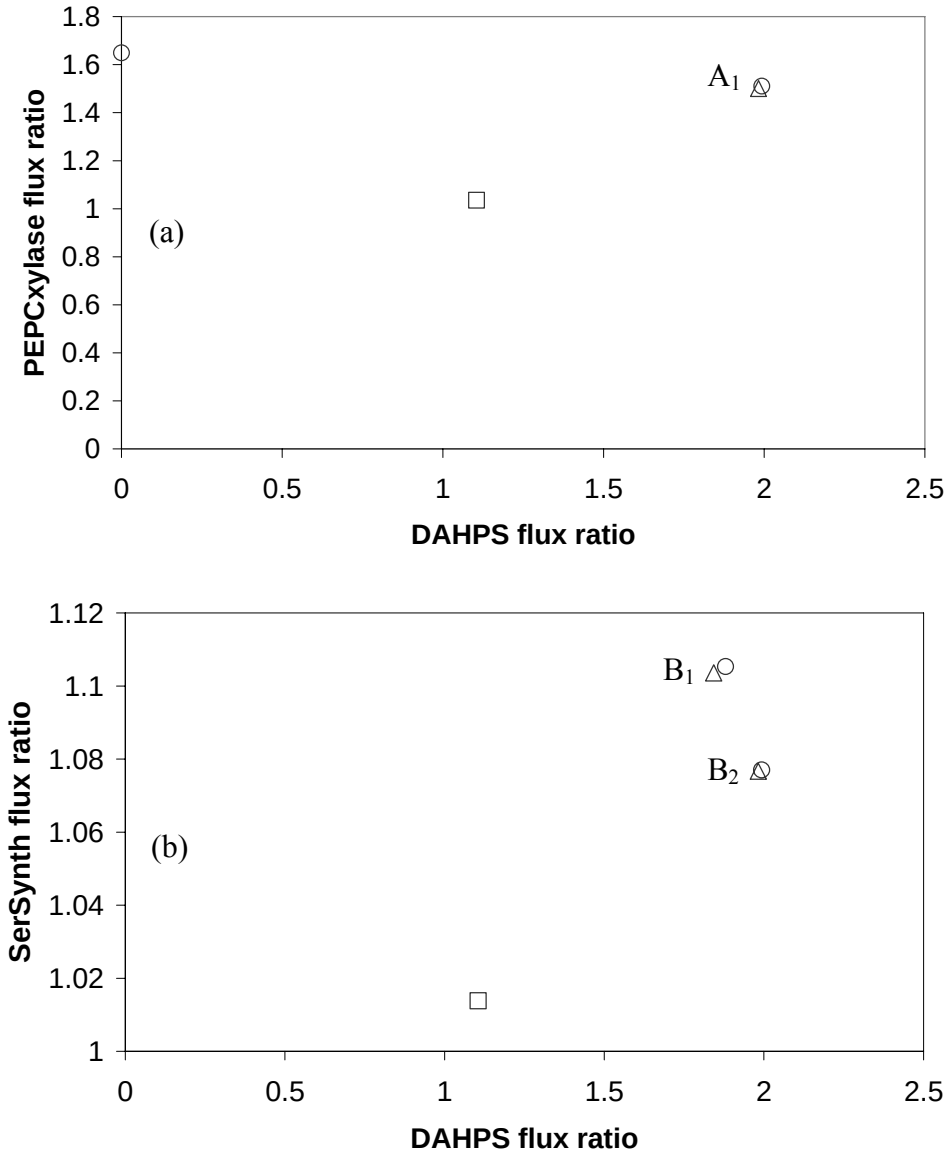


Fig. 3.2. Pareto results for gene knockouts (single-gene □; double-gene Δ; triple-gene ○) in simultaneous maximization of (a) DAHPS and PEPCxylase flux ratios, and (b) DAHPS and SerSynth flux ratios. The chromosomes in double-gene knockouts are labelled.

Table 3.2 Pareto-optimal metabolic pathway recipe for 2-enzyme knockouts represented by the three labelled chromosomes in Fig. 3.2. The flux ratios are listed in the second and third column for (chromosomes A_1, B_1 and B_2) for each enzyme. The same enzymes (G6PDH and MetSynth) are knocked out in both chromosomes A_1 and B_2.

Enzyme with serial number	Flux ratios of chromosomes A_1 and B_2	Flux ratios of chromosome B_1
1. PTS	1.000	1.000
2. PGI	3.348	0.772
3. PFK	1.294	0.964
4. ALDO	1.294	0.964
5. TIS	1.294	0.960
6. GAPDH	1.098	0.971
7. PGK	1.098	0.971
8. PGluMu	1.099	0.963
9. ENO	1.099	0.963
10. PK	1.024	0 (knocked out)
11. PDH	0.993	0.802
12. PEPCxylase	1.500	1.460
13. PGM	2.078	0.0093
14. G1PAT	2.086	0 (knocked out)
15. RPPK	1.022	1.034
16. G3PDH	1.340	1.291
17. SerSynth	1.077	1.104
18. Synth1	1.064	1.087
19. Synth2	0.999	0.983
20. DAHPS	1.982	1.843
21. G6PDH	0 (knocked out)	1.112
22. PGDH	0	1.112
23. Ru5P	− 0.138	1.098
24. R5PI	0.208	1.132
25. TKa	0.022	1.154
26. TKb	− 0.327	1.032
27. TA	0.022	1.154
28. MurSynth	1.000	1.000
29. MetSynth	0 (knocked out)	1.000
30. TrpSynth	1.000	1.000

3.6 Interactive Branch-and-Bound Facilitated By NSGA-II

By reducing the manipulation factors in steps and applying NSGA-II, it is possible to identify potential genes that can be deleted by pattern recognition. At each step of the gene manipulation, the user decides which genes are likely to be knocked out subject to the homeostatic and total enzymatic flux constraints. The user then tests this conjecture by stepping down the manipulation factors of the genes of interest. A new conjecture is formed based on the last results. The process is repeated until the genes are completely deleted. The Pareto will consist of the completely knocked out genes in the last step that fulfill the constraint requirements. Heuristics method (Chapter 5) uses similar technique to identify genes that could be deleted prior to simultaneous gene manipulation and knockouts. The triple enzyme knockout optimization for Case A is described below in this section.

Using a random seed of 0.6, the manipulation factors for all the three genes are stepped down. If the manipulation factors are decreased to below 0.32, there will be a forced termination of the IMSL program. Results suggest that G1PAT (number 14) and G6PDH (number 21) could be deleted. Similar pattern is also observed if a random seed of 0.7 were used. To test this conjecture, G1PAT (number 14) is fixed as a deleted gene by setting its manipulation factor to 0.0. The manipulation factors of the remaining two floating genes are set at 0.35 and 0.50 to identify the next gene that could be deleted. The next gene with the potential of being knocked out is DAHPS (number 20). Other intermediate results suggest that there could be two Pareto-optimal solutions. Next, fixing G1PAT (number 14) as a deleted gene and fixing DAHPS as a gene to be deleted by gradually reducing its manipulation factor will help to narrow the search domain for Pareto. One of the three gene triplets in the penultimate step is a Pareto-optimal solution. Using manual gene knockout, only the G1PAT-RPPK-DAHPS (number 14-15-20) knockout is Pareto-optimal subject to both the homeostatic and total enzymatic flux constraints. This is indeed one of the two Pareto solutions in Figure 3.2. The G1PAT-DAHPS-Ru5P knockout is infeasible while the G1PAT-SerSynth-DAHPS knockout breaches the homeostatic constraint.

It is possible to find the second Pareto using the interactive branch-and-bound technique. G6PDH can be deleted as noted above. By setting the manipulation factor of G6PDH to 0.0 and reducing the manipulation factors of the two remaining floating genes, two distinct groups of genes are obtained. By further branching of the

identified repressible genes and reducing their manipulation factors, six distinct gene triplets are obtained. Applying manual gene knockout to the six gene triplets, only the G6PDH-TrpSynth-G3PDH (number 21-30-16) knockout is feasible. The other five triplets cannot be knocked out. Knocking out G6PDH-TrpSynth-G3PDH breaches the homeostatic constraint by 0.00888. The Pareto solution (Section 3.5) G6PDH-MurSynth-TrpSynth (number 21-28-30) does not breach both homeostatic and total enzymatic flux constraints. This means that the G6PDH-TrpSynth-G3PDH (number 21-30-16) triplet is identified by NSGA-II as a non-dominated solution subject to both constraints when G6PDH and TrpSynth are deleted and the multiplier for G3PDH is set to 0.25. To obtain the second Pareto solution, the last gene MurSynth is identified manually (out of 28 genes) by fixing G6PDH-TrpSynth as deleted genes.

3.7 Optimization of Gene Manipulation

The optimization of the Pareto-optimal metabolic pathway recipe by genetic manipulations shows that the triple-enzyme manipulation has the best non-dominated flux ratios due to the flexibility in changing enzymatic reaction rates to enhance the desired flux ratios (Figures 3.3A and 3.4A). The left and right Pareto-optimal segments of single-enzyme manipulation (Figures 3.3A and 3.3B) are governed respectively by PEPCxylase and DAHPS in Case A, and SerSynth and DAHPS in Case B (Figures 3.4A and 3.4B). The single chromosome in between the two segments is the result of manipulating PFK in both Cases A and B. Fluxes of the polysaccharide and glycerol synthesis, and to a lesser extent pentose-phosphate pathway and nucleotide synthesis are continually attenuated on ascending the left Pareto-optimal segment and descending the right Pareto-optimal segment when the governing gene of each segment is being overexpressed. As a result, the fluxes are redistributed among DAHPS, PEPCxylase and SerSynth. Metabolite concentrations follow similar trend as carbon sources are diverted towards building gene molecules. The in-between chromosomes, with manipulation factors set to an upper bound of 2.0, are pivotal to the generation of the two distinct Pareto-optimal segments when a governing enzyme is being switched.

In the double-enzyme manipulation, the left and right Pareto-optimal segments (Figures 3.3A and 3.3B) in Case A are governed by PEPCxylase/G6PDH and DAHPS/G6PDH, respectively. The manipulation factor of G6PDH is actively

constrained at 0.5 (lower bound) to divert fluxes away from the pentose-phosphate pathway (the concentration of 6pg decreases by about 50%), thus resulting in higher fluxes in glycolytic pathway (PGI flux is the highest), polysaccharide synthesis, nucleotide synthesis and glycerol synthesis. The PEPCxylase flux ratio increases as PEPCxylase manipulation factor approaches 2.0 (upper bound) on ascending the left Pareto segment (Figures 3.3A and 3.3B). Similarly, the DAHPS flux ratio increases as DAHPS manipulation factor approaches the upper bound on descending the right Pareto segment. Interestingly, when the paired enzyme is switched from A_1 to A_2, fluxes from other pathways are drawn towards DAHPS instead of PEPCxylase (Table 3.3), indicating a distinct change in the metabolic pathway recipe.

The leftmost Pareto-optimal segment (Figures 3.4A and 3.4B) in Case B of the double-enzyme manipulation is governed by SerSynth and GAPDH. The manipulation factor of SerSynth hardly deviates from its upper bound of 2.0 (Figures 3.4A and 3.4B) and the manipulation factor of GAPDH changes from 2 (at chromosome B_1) to 1.27 (Figures 3.4A and 3.4B). Fluxes are drained from DAHPS, polysaccharide synthesis and glycerol synthesis to sustain high SerSynth reaction rate. The concentrations of fdp and gap, both precursors being the proximate carbon sources of 3pg (SerSynth precursor), substantially decrease by 70% and 50% (at chromosome B_1), respectively, thereby leading to the enhanced production of serine. The three leftmost chromosomes consisting of SerSynth and PFK on the same Pareto segment furthest from chromosome B_2 are pivotal to the generation of the two distinct Pareto-optimal segments containing chromosomes B_1 and B_2 which are governed by the paired enzymes SerSynth/GAPDH and DAPHS/SerSynth, respectively. The DAHPS flux ratio increases and the SerSynth flux ratio decreases (Table 3.3) as the manipulation factor of DAHPS approaches 2.0 (1.07 for SerSynth) on moving towards chromosome B_2. The concentrations of fdp and gap decrease by 23% and 13% (at chromosome B_2) respectively as serine production declines. The three chromosomes on the right side of chromosome B_2 consisting of DAHPS and G6PDH (manipulation factor of 0.5) represent high DAHPS production rate and relatively constant serine production rate. The DAHPS flux ratio increases considerably when the DAHPS manipulation factor approaches 2.0 resulting in simultaneous increase of glycolytic and decrease of pentose-phosphate fluxes.

Table 3.3 Pareto-optimal metabolic pathway recipe for 2-enzyme manipulations represented by the four labelled chromosomes in Figs. 3.3A and 3.4A. The flux ratios are listed in second, third, fourth and fifth column (chromosomes A_1, A_2, B_1 and B_2, respectively) for each enzyme.

Enzyme with serial number	Flux ratios of chromosome A_1	Flux ratios of chromosome A_2	Flux ratios of chromosome B_1	Flux ratios of chromosome B_2
1. PTS	1.000	1.000	1.000	1.000
2. PGI	2.117	2.139	1.037	1.096
3. PFK	1.143	1.137	1.024	1.008
4. ALDO	1.143	1.137	1.024	1.008
5. TIS	1.143	1.137	1.030	1.010
6. GAPDH	1.051	1.035	1.033	0.996
7. PGK	1.051	1.035	1.033	0.996
8. PGluMu	1.052	1.036	0.978	0.994
9. ENO	1.052	1.036	0.978	0.994
10. PK	1.011	1.007	0.996	0.985
11. PDH	1.002	1.001	0.999	0.997
12. PEPCxylase	1.287	1.067	0.962	0.907
13. PGM	1.383	1.242	0.455	0.782
14. G1PAT	1.385	1.244	0.450	0.781
15. RPPK	1.009	1.002	0.959	0.985
16. G3PDH	1.146	1.095	0.558	0.889
17. SerSynth	1.034	1.022	1.978	1.022
18. Synth1	1.029	1.018	0.991	0.961
19. Synth2	1.000	1.000	1.000	1.000
20. DAHPS	1.378	2.077	0.316	1.498
21. G6PDH	0.526	0.520	0.994	0.964
22. PGDH	0.526	0.520	0.994	0.964
23. Ru5P	0.463	0.437	1.015	0.947
24. R5PI	0.621	0.643	0.962	0.988
25. TKa	0.533	0.561	0.962	0.989
26. TKb	0.382	0.291	1.077	0.898
27. TA	0.532	0.561	0.962	0.989
28. MurSynth	1.000	1.000	1.000	1.000
29. MetSynth	1.000	1.000	1.000	1.000
30. TrpSynth	1.000	1.000	1.000	1.000

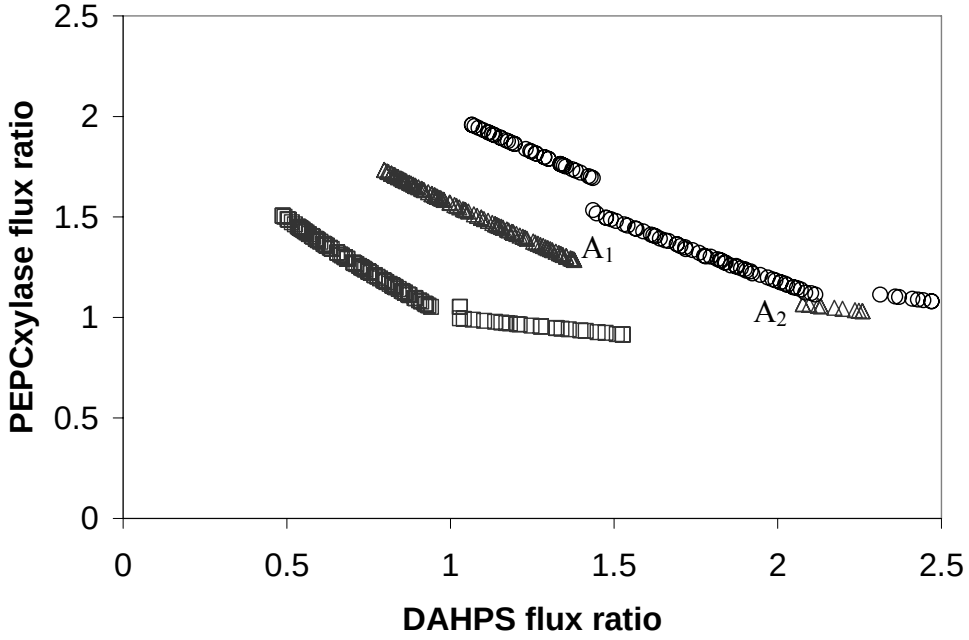


Fig. 3.3A. Pareto-optimal fronts for gene manipulations (1-enzyme □; 2-enzyme Δ; 3-enzyme ○) in simultaneous maximization of DAHPS and PEPCxylase flux ratios (Case A).

In the triple-enzyme manipulation, the left Pareto segment in Case A with comparatively high PEPCxylase flux ratio (Figures 3.3A and 3.3B) is obtained by repressing the activities of genes for PK and G6PDH and by overexpressing gene for PEPCxylase, which result in largely attenuated pentose-phosphate pathway (50% decrease in 6pg concentration) and PK fluxes. The leftmost chromosome of the central optimal Pareto segment, which is obtained by repressing genes related to GAPDH and G6PDH and overexpressing gene related to PEPCxylase, is pivotal in switching the flux control to genes related to DAHPS, PEPCxylase and G6PDH. The enzyme manipulation factors of G6PDH and DAHPS are actively constrained at 0.5 and 2.0, respectively. On descending the central segment of the Pareto-optimal front, the enzyme manipulation factor of PEPCxylase decreases from 1.79 to 1.11 (Figures 3.3A and 3.3B) to generate successively higher DAHPS flux ratio. Even higher DAHPS flux ratio is obtained by switching the flux control to genes related to SYN1, G6PDH and DAHPS. The enzyme manipulation factors of SYN1 and G6PDH are actively constrained at 0.5 to restrict competing fluxes in the chorismate and mureine synthesis, and pentose-phosphate pathways respectively. Under such conditions, the right Pareto-optimal segment is formed as the enzyme manipulation factor of DAHPS increases from 1.75 to 2.0.

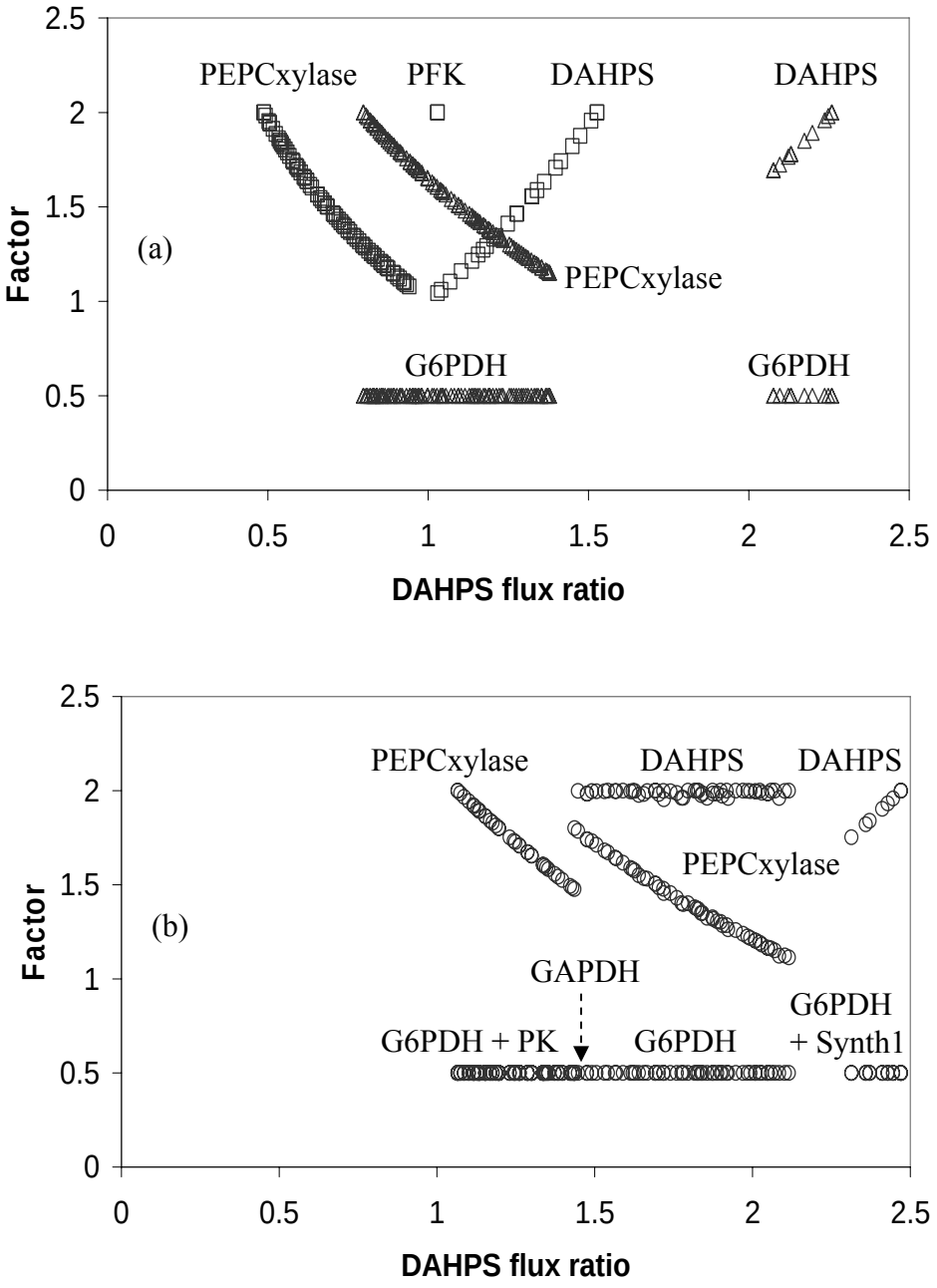


Fig. 3.3B. Pareto-optimal enzyme manipulation factors in simultaneous maximization of DAHPS and PEPCxylase flux ratios (Case A) (a) 1-enzyme (□) and 2-enzyme (Δ) manipulation factor and (b) 3-enzyme manipulation factor (○).

Except for the few leftmost chromosomes in Case B (Figures 3.4A and 3.4B) obtained by overexpressing genes related to PFK and SerSynth and repressing gene related to G6PDH, the metabolic pathway recipe on descending the Pareto-optimal segment is formed by simultaneously reducing, increasing and constraining the enzyme manipulation factors of SerSynth, DAHPS and G6PDH, respectively. Similar

to that of Case A, the relatively flat profile on the right is formed by manipulating genes related to SYN1, G6PDH and DAHPS.

In general, optimization results are as good as the model used in the study. Some uncertainty in any model and its parameters is unavoidable, particularly in case of complex biological systems such as *E. coli* and after genetic engineering of living organisms. Hence, results of optimizing multi-product microbial cell factories will have to be confirmed through experimental studies. These will be explored in future work.

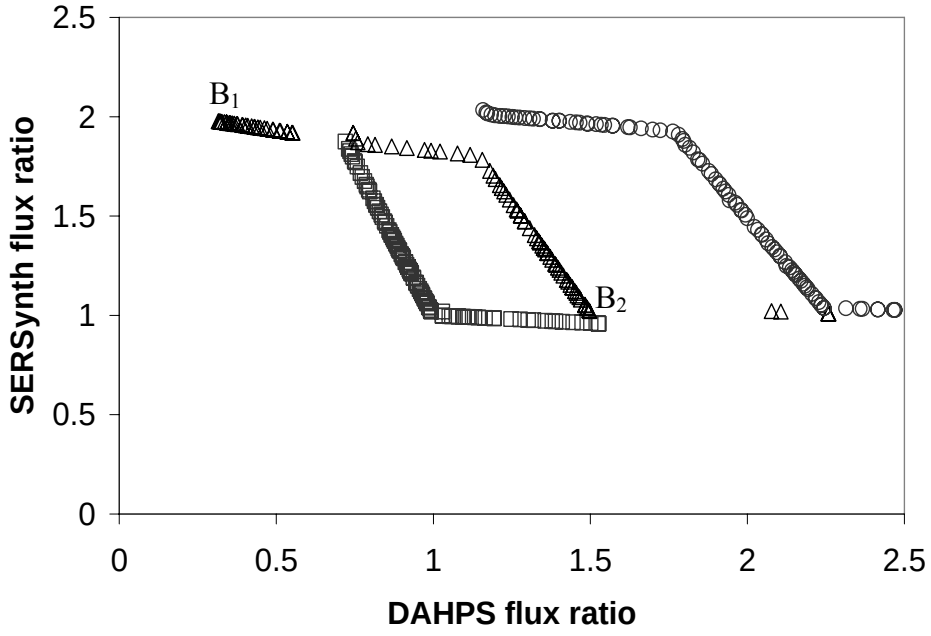


Fig. 3.4A. Pareto-optimal fronts for gene manipulations (1-enzyme □; 2-enzyme Δ; 3-enzyme ○) in simultaneous maximization of DAHPS and SerSynth flux ratios (Case B).

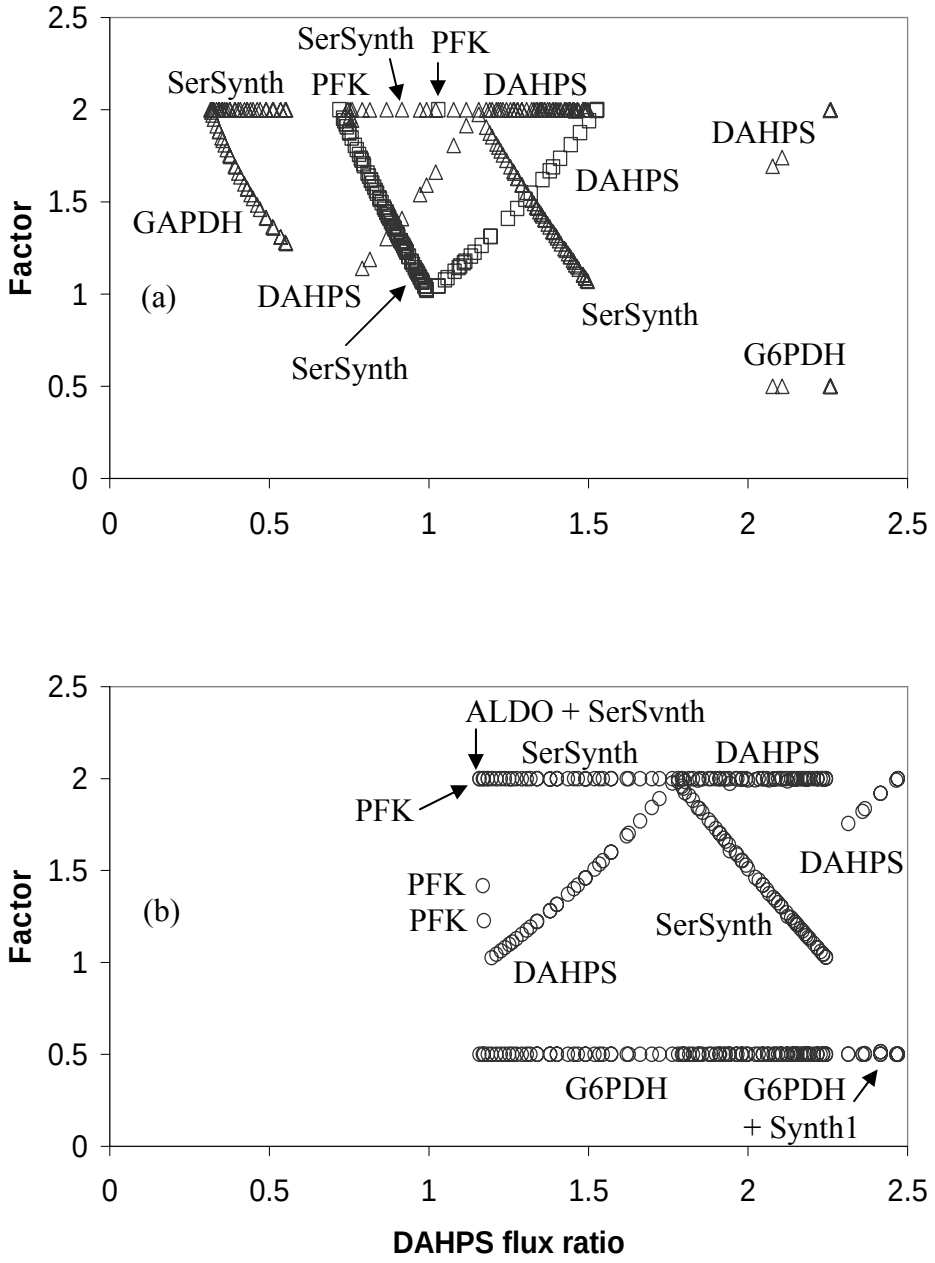


Fig. 3.4B. Pareto-optimal enzyme manipulation factors in simultaneous maximization of DAHPS and SerSynth flux ratios (case B) (a) 1-enzyme (□) and 2-enzyme (Δ) manipulation and (b) 3-enzyme manipulation (○).

3.8 Conclusions

In this chapter, MOO of fluxes of desired enzymatic sub-systems in *E. coli* was described; two cases - maximization of DAHPS and PEPCxylase fluxes, and maximization of DAHPS and SerSynth fluxes, by 1-, 2- and 3-gene knockout and manipulation[2] were considered. Optimal Pareto solutions were successfully obtained using the NSGA-II program wherein integer and/or continuous decision variables can be used; this flexibility allowed seamless use of both types of variables in the MOO of enzyme fluxes in *E. coli*. Identical Pareto sets were obtained in 1-, 2- and 3-gene knockout using interactive branch-and-bound assisted by NSGA-II and manual exhaustive search. Triple gene knockout was discussed in this study to demonstrate its efficiency in obtaining Pareto sets for gene knockout.

Knocking out PGM gives the 1-enzyme Pareto-optimal set. In 1-enzyme manipulation cases, the gene that generates the Pareto-optimal set is directly related to the desired enzymatic activity. In paired and triple enzyme knockout or manipulation, G6PDH is instrumental in diverting fluxes to the desired metabolic pathways. The triple enzyme knockout/manipulation gives the best Pareto-optimal set due to greater flexibility in redistributing fluxes to the desired pathways. In the triple enzyme knockout, DAHPS and PEPCxylase fluxes increase by 51% and 99% respectively in one case while DAPHS and SerSynth fluxes increase by around 95% and 9% respectively in another case. In the triple enzyme manipulation, the DAHPS and PEPCxylase fluxes increase up to 247% and 96% respectively in one case while DAHPS and SerSynth fluxes increase up to 247% and 203% respectively in another case.

[2] In gene knockout, a foreign sequence is inserted into an existing gene to interrupt the normal DNA sequence. The altered gene in most cases will either translate into a non-functional protein or deleted following DNA recombination. The techniques for overexpression in *E. coli* are well known. They work by either increasing the number of copies of the gene or increasing the binding strength of the promoter of an operon (Chapter 4) to assist transcription. In gene repression (also known as gene knockdown), the gene expression is reduced through either DNA modification or by treatment with a reagent such as a short DNA or RNA oligonucleotide.

Chapter 4
DEVELOPMENT OF AN AUGMENTED MODEL FOR MICROBIAL CELL FACTORY

4.1 Introduction

Increasingly sophisticated mathematical models for multi-product biosynthesis factories such as *Escherichia coli* (*E. coli*) (Chassagnole et al., 2002; Degenring et al., 2004), *Corynebacterium glutamicum* (Wendisch et al., 2006), *Saccharomyces cerevisiae* (Rizzi et al., 1997; Teusink et al., 2000) and *Penicillium chrysogenum* (Zangirolami et al., 1997) enable chemical engineers to extend the reach of their competencies in systems area such as modelling and optimization to knowledge-based biologics and fermentation industry (Lee et al., 2009). Microbial metabolomics aided by automated sub-second stopped-flow sampling technique used successfully by Chassagnole et al. (2002), advanced chromatographic techniques and mass spectrometry in stimulus response experiments using glucose pulse, have begun to unravel the intricate metabolic pathways and structures in the cell. The transient metabolite concentration data obtained through modern metabolomics enable the construction and validation of new models. Validation and optimization of the central carbon metabolism (through the kinetic parameters) obtained by combining existing models for enzymatic sub-systems of a microbial cell factory such as *E. coli* (Chassagnole et al., 2002; Degenring et al., 2004) are feasible through modern microbial metabolomics.

However, gathering transient data and modelling the aromatic amino acids (L-tryptophan, L-tyrosine and L-phenylalanine) biosynthesis pathways in *E. coli* remains a formidable challenge for modern metabolomics (Mashego et al., 2007), mainly because metabolomics research relies on the isolation of metabolites from a biological sample (i.e. in vitro analysis). Prokaryotic cells such as *E. coli* tend to leak intracellular metabolites through damaged cell membrane during the quenching process typically carried out at around −50°C. Alternatively, the simpler simultaneous quenching and direct extraction procedure creates data that are not easily decomposable into individual metabolite concentrations. Recent stimulus response experiments (Oldiges et al., 2004; Wahl et al., 2006) have generated concentration profiles of temporal metabolites from the first metabolite of the common aromatic

amino acids pathway 3-deoxy-D-*arabino*-heptulosonate 7-phosphate (dahp) to shikimate 3-phosphate (s3p). Transient metabolite concentration data further downstream of the shikimate pathway (which is also known as the common aromatic amino acids pathway) leading to the production of aromatic amino acids are not yet available (Wahl et al., 2006) due to experimental difficulties such as signal dilution and loss of intracellular metabolites (Oldiges et al., 2007). Signal dilution is a possible result of complex in vivo (i.e. intracellular) controls of enzymatic activities and gene expressions regulating aromatic amino acids biosynthesis.

Independently, data for tryptophan biosynthesis modelling in *E. coli* are mostly drawn from well-developed experimental approaches of microbiology and genomics. An example of experimental microbiology involves varying independent variables (e.g. temperature, specific cell growth rate and L-tryptophan concentration in the growth medium) to obtain a set of equilibrated data (e.g. intracellular *trp* repressor concentration) under controlled experimental conditions (Gunsalus et al., 1986). The direct measurement of the tryptophan operon transcription rate of *E. coli* using electron microscopy (Gotta et al., 1999) following addition of rifampin to log-phase cultures is an example of experimental genomics. The tryptophan biosynthesis models constructed using microbiology and genomics data focus on operon regulation and stability with hardly any consideration of the central carbon metabolism as being capable of carbon flux control through its enzymatic activities. This presents an opportunity to develop an augmented model (Section 4.2) by linking the central carbon metabolism of *E. coli* to tryptophan biosynthesis. The augmented model serves as a platform for the mixed-integer MOO study that involves concurrent gene knockouts and manipulations and a recombinant DNA technique that alters the copy number of plasmids containing operons and genes encoding aporepressors. Modern metabolomics attempt to model the entire aromatic amino acids biosynthesis semi-empirically with no regards for macromolecular synthesis mechanisms (such as transcription and translation) and gene expression that occur within the longer time horizon. Linking an operon-based model for tryptophan synthesis that accounts for the well-studied macromolecular synthesis and regulatory effects of operon repression, transcriptional attenuation and enzyme inhibition to the central carbon metabolism of *E. coli*, enables the thesis author to formulate metabolic pathway recipe for engineering a wild strain targeted for industrial production of desired amino acids via mixed-integer MOO.

4.2 Model

4.2.1 Aromatic Amino Acids Pathways

DAHPS enzymatic sub-system is the starting point of the common pathway for the aromatic amino acids biosynthesis (Figure 4.1). The condensation of phosphoenolpyruvate (pep) and erythrose 4-phosphate by DAHP synthase (DAHPS) to form dahp is tightly regulated via the transcriptional and allosteric controls exerted by the aromatic amino acid end products. DAHPS consists of three different isofunctional enzymes (isoenzymes) that catalyze the first reaction in the aromatic amino acids pathway. Each isoenzyme – encoded by the *aroF*, *aroG* and *aroH* genes – is feedback-inhibited independently by each of the three different end products consisting of L-tyrosine, L-phenylalanine and L-tryptophan, respectively. The shikimate pathway downstream of dahp (Figure 4.2) is catalyzed by various enzymes to form in sequence 3-dehydroquinate (3dhq), 3-dehydroshikimate (3dhs), shikimate (shik), shikimate 3-phosphate (s3p) and 5-enolpyruvoylshikimate 3-phosphate (epsp). Chorismate synthase which is encoded by the *aroC* gene catalyzes the conversion of epsp into chorismate (cho). This last reaction of the common pathway appears to be rate limiting. The chorismate precursor is at the crossroad of three terminal pathways leading to the biosynthesis of aromatic amino acids. Chorismate is also the substrate for the biosynthesis pathways of ubiquinone, menaquinone, folate and enterochelin (not shown in Figure 4.2).

The first reaction of the tryptophan biosynthesis terminal pathway (Figure 4.2) involves the conversion of chorismate and glutamine to anthranilate (anta), glutamate and pyruvate. The enzymatic complex that facilitates this reaction is anthranilate synthase (ANTAS). The anthranilate synthase complex is a heterotetramer composed of two molecules of each of the polypeptides encoded by the *trpE* and *trpD* genes. These are known as component I and component II, respectively. Component I contains the binding site for chorismate. In the absence of component II, component I cannot catalyze the formation of anthranilate using glutamine (produced internally) as a nitrogen source. Component II – encoded by the *trpD* gene – facilitates two activities. The first one, glutamidotransferase activity, is required to activate component I in the anthranilate synthase reaction. This activity channels the nitrogen from glutamine to the active site for anthranilate production. Only the anthranilate synthase complex exhibits this activity using chorismate as a precursor. The second

activity of Component II converts anthranilate to anthranilate-5-phosphoribosyl pyrophosphate (pra). This second catalysis is known as anthranilate phosphoribosyl transferase (ANTAP).

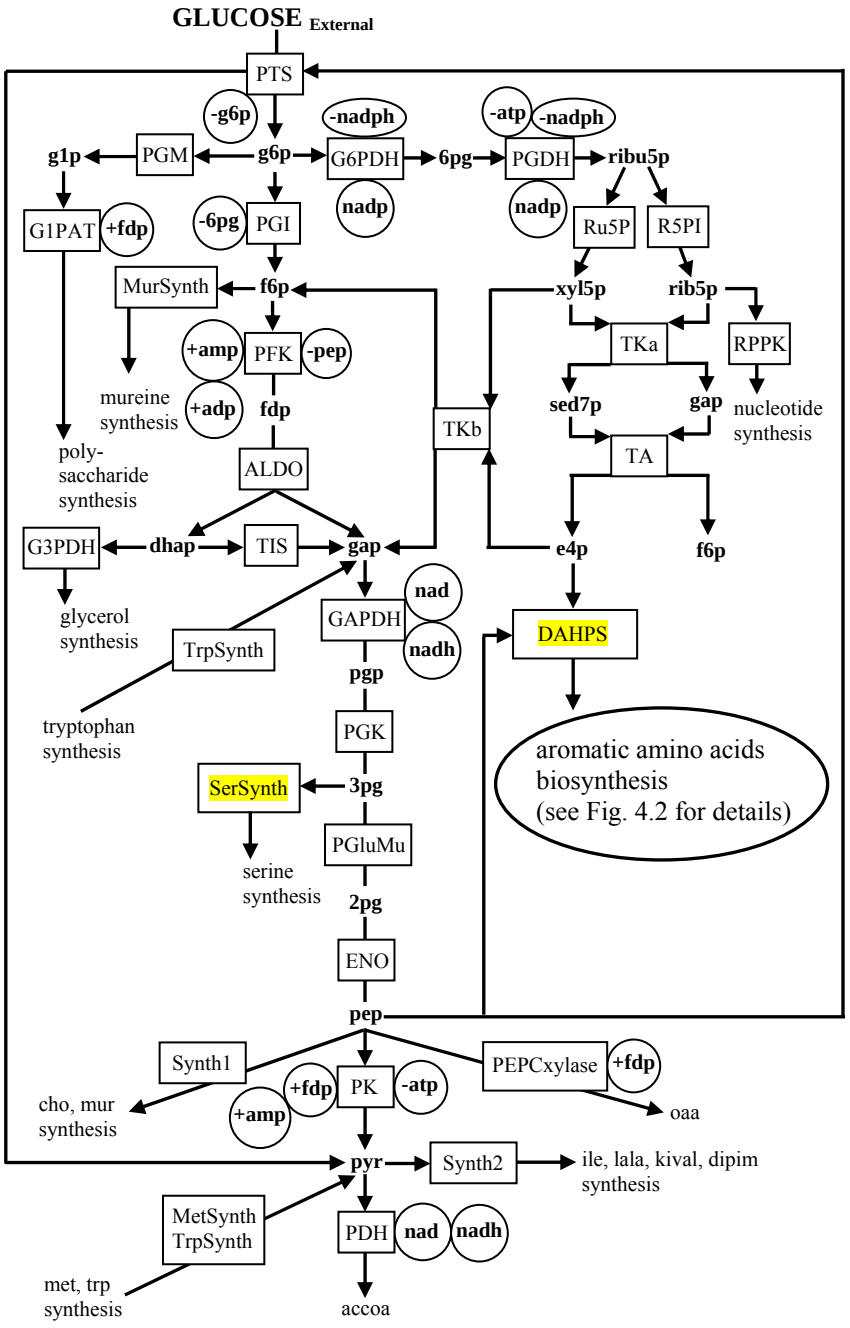


Fig. 4.1. Metabolic network of the central carbon metabolism of *Escherichia coli*. Enzymes are shown in rectangles; precursors (balanced metabolites) are in bold between enzymes; allosteric effectors (atp, adp and fdp), activators (positive sign), inhibitors (negative sign) and regulators (without sign) are given in circles/ellipses. All abbreviations are defined in the List of Symbols. Details of the aromatic amino acids pathways are given in Fig. 4.2.

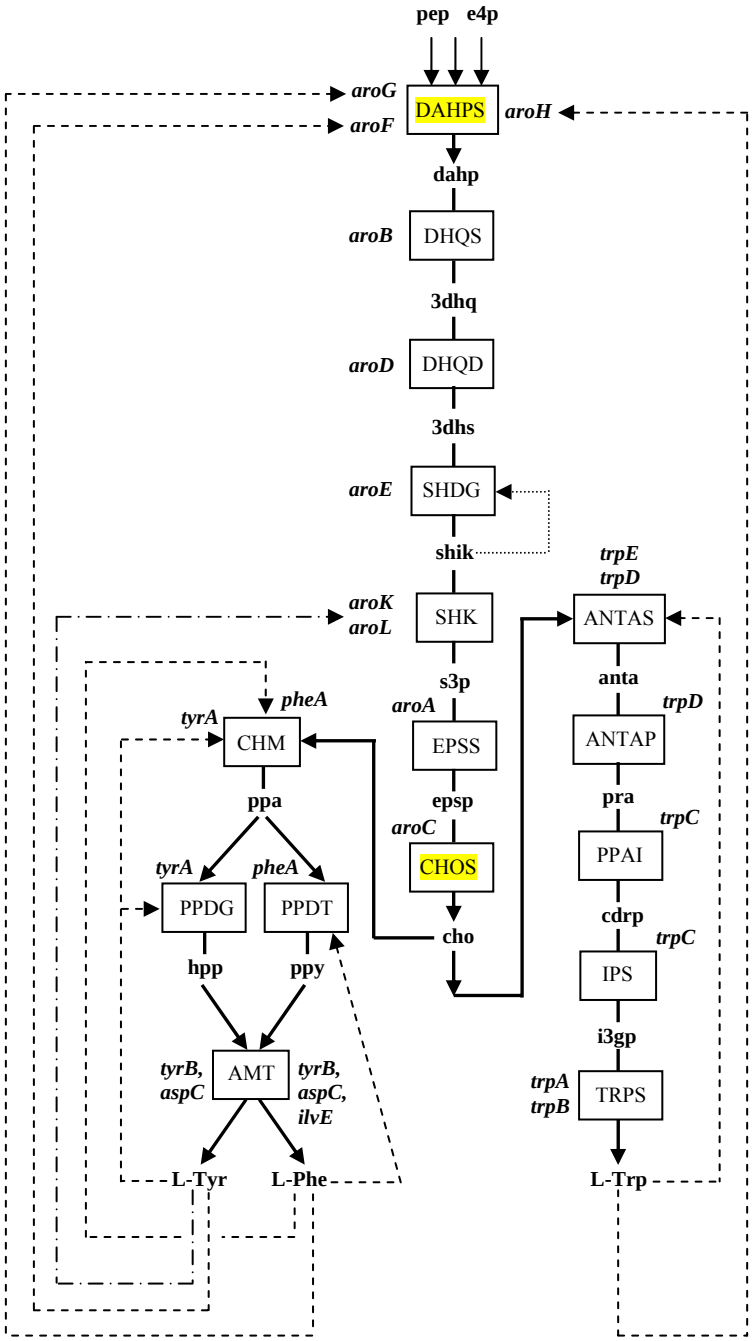


Fig. 4.2. Metabolic network of the aromatic amino acids biosynthesis of *Escherichia coli*. The shikimate pathway between dahp and chorismate is common for all the three end products. The terminal pathway for tryptophan biosynthesis starts from chorismate and ends at tryptophan. The terminal pathways of the other two end products L-tyrosine and L-phenylalanine are also shown. To indicate various types of control, different lines are used: --- transcriptional and allosteric controls exerted by the three end products; ····· allosteric controls only; – · – transcriptional controls only. The genes are in italics, enzymes are shown in rectangles and metabolites are in between the enzymes.

Tryptophan exerts negative feedback control on the activities of both anthranilate synthase (up to 100%) and anthranilate phosphoribosyl transferase (not exceeding 70%) by binding to the enzymes allosteric sites (Pittard, 1996). This action changes the geometrical conformation of the enzymes and prevents efficient binding of precursors (such as chorismate) to the active sites of the same enzymes. The binding of chorismate and tryptophan to sites present in anthranilate synthase are competitive. Feedback inhibition of anthranilate synthase is the most important control among those exerted on the terminal tryptophan pathway enzymes because it is limiting the rate of tryptophan biosynthesis. Therefore, anthranilate synthase is taken as the dominant (i.e. representative) enzyme in the tryptophan biosynthesis model.

Anthranilate-5-phosphoribosyl pyrophosphate is converted to 1-(o-carboxyphenylamino)-1-deoxyribulose 5-phosphate (cdrp) with the help of phosphoribosyl anthranilate isomerase (PPAI), and cdrp is in turn converted to indole 3-glycerolphosphate (i3gp) by a single enzyme indolglycerol phosphate synthetase (IPS). Both PPAI and IPS are encoded by the *trpC* gene. The multi-enzyme complex, tryptophan synthase (TRPS) which is encoded by the *trpA* and *trpB* genes catalyzes the conversion of i3gp and L-serine to L-tryptophan and glyceraldehyde 3-phosphate (gap). While L-serine is consumed in the final step of tryptophan biosynthesis, it is manufactured by converting the central carbon metabolite 3-phosphoglycerate (3pg) in three steps. In the first step, 3-phosphoglycerate dehydrogenase (*serA* gene product) oxidizes 3pg to 3-phosphohydroxypyruvate (3php). Following this, 3-phosphoserine aminotransferase (*serC* gene product) converts 3php to 3-phosphoserine (3ps). In the third step, 3ps is dephosphorylated to L-serine by 3-phosphoserine phosphatase (*serB* gene product). Besides being consumed during tryptophan biosynthesis, L-serine is used in the synthesis of cysteine and phospholipids.

4.2.2 Controls in Tryptophan Operon

The term “operon” first proposed in 1960 refers to a cluster of genes under the control of a single DNA operator. An operon resides in a chromosome of the wild *E. coli* strain. Multicopy plasmids containing tryptophan operons can be introduced into a cell through recombinant DNA technology. The tryptophan operon (*trp* operon) consists of promoter, DNA operator, leader sequence and structural genes (Figure

4.3a). According to the Central Dogma in molecular genetics, a gene composed of deoxyribonucleic acid (DNA) is transcribed into ribonucleic acid (RNA), which is then translated into a polypeptide that is subsequently processed to become a protein such as enzyme. The transcription of genetic information from *trp* operon to RNA is carried through the action of the enzyme mRNA (messenger RNA) polymerase. Under normal circumstances, the sigma factor of the mRNA polymerase (mRNAP) recognizes the promoter and transcription initiation site. Once the mRNAP is bound to the promoter, the process of transcription can proceed. In transcription, an mRNA molecule complementary to one of the two strands of a double-stranded DNA molecule is synthesized. Once a small portion of the mRNA has been formed, the sigma factor disengages and the rest of the mRNA chain elongation will be conducted by the core mRNA polymerase as it moves along the *trp* operon.

Feedback inhibition described in Section 4.2.1 is a form of post-translational control of pre-existing enzymes such as anthranilate synthase. Repression and transcriptional attenuation are two other types of control exerted on the *trp* operon that affect the synthesis of mRNA. The active holorepressor (Figure 4.3b) is formed when two tryptophan molecules bind independently to the two sites of an originally inactive aporepressor molecule encoded by the *trpR* gene. The *trpR* gene resides in the chromosome of the wild *E. coli* strain. Plasmids containing the *trpR* gene can be inserted into the *E. coli* cell. The inactive aporepressor protein is unable to bind to the DNA operator itself. However, once the tryptophan concentration increases, the active holorepressor molecule is formed. Transcription process is blocked at an early stage when the active holorepressor binds to the DNA operator. The tryptophan concentration resulting in repression (Koh et al., 1998) ranges from around 1 μM (practically no repression) to 100 μM (repression exceeding 90% of all the *trp* operons being considered).

In addition to the promoter and DNA operator regions, there is a sequence called the leader sequence (specified by the *trpL* gene), which codes for a polypeptide that contains tandem tryptophan codons (Figure 4.3a) near its terminus and functions as an attenuator. If the tryptophan concentration is high (greater than 5 μM), the transfer RNA will be charged (tryptophanyl-$tRNA^{Trp}$ or $tRNA^{Trp}$). The charged $tRNA^{Trp}$ is brought to the ribosome (molecular machinery for protein synthesis) by a protein factor. As the ribosome moves along the mRNA chain, the complete tryptophan-rich leader peptide is synthesized by a translation process. However, if the

tryptophan concentration is low (less than 1 μM), the tryptophan-rich leader peptide will not be synthesized (Koh et al., 1998). While transcription of the downstream DNA sequences of the *trp* operon is still proceeding, translation of sequences already transcribed has begun. As the mRNA is being released from the DNA, the ribosome binds to mRNA and translation begins. Attenuation occurs (i.e. mRNAP stops transcription) because a portion of the newly formed mRNA folds into a double-stranded loop that signals termination of mRNAP action. When tryptophan is abundant, the ribosome will translate the leader sequence until it comes to the stop codon. The remainder of the leader mRNA can then assume a terminator stem-loop, a transcription pause site, which is followed by a uracil-rich sequence that actually causes termination. In contrast, if tryptophan concentration is low, the ribosome pauses at a tryptophan codon (a sequence of three bases in mRNA encoding tryptophan) and translation to form the complete tryptophan-rich leader peptide ceases. The presence of the stalled ribosome at this position allows an alternative anti-terminator stem-loop to form. The mRNAP then moves past the non-folded termination site and begins transcription of the structural *trpEDCBA* genes.

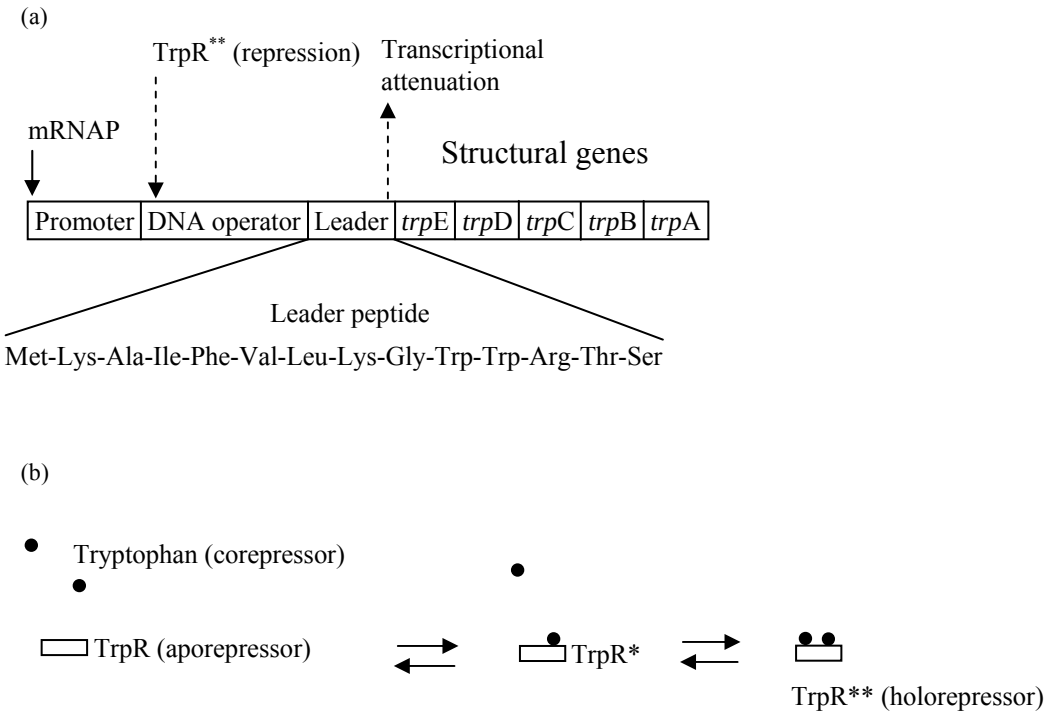


Fig. 4.3. (a) Structure of the tryptophan operon in *Escherichia coli*.
(b) Sequence showing the formation of the active holorepressor TrpR**.

4.2.3 Tryptophan Operon Model

The original central carbon metabolism model assumes tryptophan synthesis rate to be constant. In the augmented model, the tryptophan synthesis mechanism is expanded using a tryptophan operon model. Mureine and methionine synthesis rates, which are constant in the central carbon metabolism, remain unchanged in the augmented model.

Several tryptophan operon models incorporating control mechanism during the last 10 years were reviewed. Earlier models did not consider some of the control mechanisms (namely, feedback inhibition exerted on an enzyme, repression of the operator and transcriptional attenuation) described in Sections 4.2.1 and 4.2.2. A model proposed by Xiu et al. (1997) considered both feedback inhibition of the enzymes and repression of the *trp* operon by tryptophan but not transcriptional attenuation. The model takes into account the effects of growth rate and demand of tryptophan for protein synthesis. Subsequent enhancements to the model by the same researchers (Xiu et al., 2002) incorporate the switching on and off of the *trp* operon due to repression and transcriptional attenuation in response to different intracellular tryptophan concentrations.

A detailed mechanistic model proposed by Santillán and Mackey (2001a) for dynamic regulation of the *trp* operon takes into account for the first time repression, feedback enzyme inhibition and transcriptional attenuation. Bhartiya et al. (2003) used the Hill equation to model transcription and translation processes considered earlier by Santillán and Mackey (2001a). Model of Santillán and Zeron (2004) is essentially an expansion of the repression and transcriptional attenuation mechanisms incorporated earlier in the work of Santillán and Mackey (2001a). The two models give comparable fit between experimental data and simulated results following a nutritional shift (minimal + tryptophan medium to minimal medium) with wild *E. coli* strain. Santillán and Mackey (2001b) have confirmed that their models are globally stable for wild type *E. coli* as well as for three different mutant strains. The detailed model proposed by Santillán and Mackey (2001a), where special attention is given to parameter estimation based on experimental data and inherent time delays during tryptophan biosynthesis, is linked to the central carbon metabolism (Chassagnole et al., 2002) to create the augmented model.

Equations (4.1) to (4.9) describe the dynamics of a *trp* operon (Santillán and Mackey, 2001a).

$$\frac{dO_F}{dt} = \frac{K_r}{K_r + R_A(T)}\left\{\mu O - k_p P\left[O_F(t) - O_F(t-\tau_p)e^{-\mu\tau_p}\right]\right\} - \mu O_F(t) \tag{4.1}$$

$$A(T) = b\left(1 - e^{-T(t)/c}\right) \tag{4.2}$$

$$\frac{dM_F}{dt} = k_p P O_F(t-\tau_m)e^{-\mu\tau_m}\left[1 - A(T)\right] - k_\rho \rho\left[M_F(t) - M_F(t-\tau_\rho)e^{-\mu\tau_\rho}\right] - (k_d D + \mu)M_F(t) \tag{4.3}$$

$$\frac{dE}{dt} = \frac{1}{2}k_\rho \rho M_F(t-\tau_e)e^{-\mu\tau_e} - (\gamma + \mu)E(t) \tag{4.4}$$

$$E_A(E,T) = \frac{K_i^{n_H}}{K_i^{n_H} + T^{n_H}(t)}E(t) \tag{4.5}$$

$$R_A(T) = \frac{T(t)}{T(t) + K_t}R \tag{4.6}$$

$$G(T) = g\frac{T(t)}{T(t) + K_g} \tag{4.7}$$

$$F(T, T_{ext}) = d\frac{T_{ext}}{e + T_{ext}\left[1 + T(t)/f\right]} \tag{4.8}$$

$$\frac{dT}{dt} = KE_A(E,T) - G(T) + F(T, T_{ext}) - \mu T(t) \tag{4.9}$$

The parameters in equations (4.1) to (4.9) are defined in Table 4.1. Equation (4.1) represents the repression dynamics. The mRNA synthesis rate described by equation (4.3) incorporates a transcription attenuation function. Feedback inhibition is built into equation (4.4). Tryptophan biosynthesis, uptake and consumption are represented by equation (4.9). The cellular specific growth rate acts as a dilution factor in the nonlinear operon dynamics.

Table 4.1 Santillán and Mackey (2001a) model parameters.

Parameter	Description
A	Transcription attenuation function
b	Asymptotic value of transcription attenuation function
c	First order time constant of transcription attenuation function
d, e, f	Constants for the tryptophan uptake
D	mRNA destroying enzyme concentration
E	Total enzyme concentration
E_A	Active enzyme concentration
F	Tryptophan uptake rate
g	Maximum tryptophan consumption rate
G	Tryptophan internal consumption rate
k_d	Rate constant for mRNA degradation
k_P	Rate constant for mRNA polymerase binding to operator site
k_ρ	Rate constant for ribosome binding to a free *trpE* related site on an mRNA
K	Tryptophan biosynthesis rate constant
K_g	Saturation constant for tryptophan internal consumption
K_i	Enzyme inhibition equilibrium constant
K_r	Dissociation constant of the holorepressor-DNA operator complex
K_t	Tryptophan aporepressor activation constant at equilibrium
M_F	Free mRNA concentration
n_H	Hill coefficient
O	Total operon concentration
O_F	Free operon concentration
P	mRNA polymerase concentration
R_A	Active repressor (holorepressor) concentration
T	Tryptophan concentration
T_{ext}	Extracellular tryptophan concentration in the growth medium
γ	Enzyme degradation rate
μ	Cellular specific growth rate
ρ	Ribosome concentration
τ_e	Time taken by a ribosome to synthesize a *trpE* polypeptide
τ_m	Time taken for an mRNA polymerase to assemble *trpE* related binding site on a ribosome
τ_P	Transcription initiation periodic time
τ_ρ	Translation initiation periodic time

Equation (4.1) shows that the mRNAP enzyme initiates transcription at every τ_p time interval. It also indicates active repression depresses the free operon concentration. The active repressor (holorepressor) concentration modelled by equation (4.6) is dependent on the availability of both the free tryptophan and aporepressor (inactive repressor) molecules. The free mRNA synthesis rate in equation (4.3) is dependent on the mRNAP concentration, rate of mRNAP binding to free operator sites, mRNA release time interval (τ_m), free operator binding sites availability subject to transcriptional attenuation, translation initiation time interval (τ_ρ), ribosome availability to initiate translation and rate of ribosome binding to mRNA. The transcriptional attenuation is modelled as a first order system in equation (4.2). The enzyme (represented by anthranilate synthase) synthesis rate in equation (4.4) is dependent on the supply of free mRNA, translation completion time interval (τ_e), ribosome availability, rate of ribosome binding to a free *trpE* related site on an mRNA and the possibility of enzyme degradation. Tryptophan biosynthesis rate in equation (4.9) depends on the supply pool of active enzyme, internal consumption rate of tryptophan in cellular protein synthesis and availability of tryptophan in the growth medium.

Equation (4.8) is not applicable because this study uses glucose (minimal growth medium) substrate to ensure compatibility with the central carbon metabolism (Chassagnole et al., 2002) model; $F(T, T_{ext})$ is zero in the current study. Feedback inhibition of enzyme activity in equation (4.5) is modelled as a Hill equation. Binding of two tryptophan molecules to each of the subunits of anthranilate synthase inhibits the enzyme activity. A pioneer tryptophan molecule bound to an allosteric site changes the geometrical configurations of an enzyme molecule in favour of other tryptophan molecules over the chorismate molecules. Because the tryptophan binding is sequential and cooperative, the Hill coefficient (n_H) is greater than 1. The internal consumption of tryptophan is described using Michaelis-Menten model in equation (4.7). Notably, exponential biomass growth that results in protein dilution is accounted for in equations (4.1), (4.3), (4.4) and (4.9).

This study uses a specific growth rate of 0.1 doublings per h for minimal growth medium in order to be consistent with the central carbon metabolism model. The parameter estimates in Santillán and Mackey (2001a) are based on 0.6 doublings per h. Parameters dependent on the specific growth rate are affected by changes in the number of mRNA molecules in a cell, number of ribosomes in a cell, mRNA

elongation rate and dry weight of a cell. Appendix B contains the detailed method to adapt operon, aporepressor and total repressor concentrations in a cell, rate constant for mRNAP binding to a free DNA operator site, rate constant for ribosome binding to a free DNA operator site, rate constant for ribosome binding to a free *trpE* related site on an mRNA, and the time taken by a ribosome to synthesize a *trpE* polypeptide to match the specific growth rate of 0.1 doublings per h. As explained in parameter estimation (Section 4.3), equations (4.5), (4.7) and (4.9) are replaced by equations (4.20) to (4.22) (given later) because of inadequate linkage between equation (4.9) purportedly describing the terminating steps of tryptophan biosynthesis and central carbon metabolism. As pointed out in section 4.3, the terminating steps of tryptophan biosynthesis involving several other key metabolites − chorismate, 5-phosphoribosyl-α-pyrophosphate (prpp) and serine − are not reflected in equation (4.9). Equations (4.1), (4.2), (4.3), (4.4) and (4.6) are retained in the revised augmented model.

4.2.4 Augmented Model Description

Mass balance equations (4.10) to (4.14) link the tryptophan operon dynamics and metabolites in the shikimate and terminal tryptophan biosynthesis pathways (equations 4.1 to 4.9) to the nonlinear central carbon metabolism model (Chapter 3).

$$\frac{dC_{pep}}{dt} = r_{ENO} - r_{PK} - r_{PTS} - r_{PEPCxylase} - r_{DAHPS} - r_{ChoSynth} - r_{MurSynth} - \mu C_{pep} \quad (4.10)$$

$$\frac{dC_{dahp}}{dt} = r_{DAHPS} - r_{ChoSynth} - \mu C_{dahp} \quad (4.11)$$

$$\frac{dC_{choris}}{dt} = r_{ChoSynth} - r_{TrpSynth} - r_{Synth3} - \mu C_{choris} \quad (4.12)$$

$$\frac{dC_{prpp}}{dt} = r_{RPPK} - r_{Synth4} - r_{TrpSynth} - \mu C_{prpp} \quad (4.13)$$

$$\frac{dC_{ser}}{dt} = r_{SerSynth} - r_{TrpSynth} - r_{Synth5} - \mu C_{ser} \quad (4.14)$$

The central carbon metabolism model (Chassagnole et al., 2002) has 25 metabolites consisting of 18 balanced metabolites and 7 unbalanced co-metabolites. The new metabolites added in the augmented model are the enzyme for tryptophan biosynthesis, 3-deoxy-D-*arabino*-heptulosonate 7-phosphate (dahp), chorismate (cho), tryptophan (trp), 5-phosphoribosyl-α-pyrophosphate (prpp) and serine (ser).

Equation (4.10) accounts for chorismate and mureine synthesis separately, which were aggregated in the original central carbon metabolism model. Enzymatic fluxes of the shikimate pathway are condensed into equation (4.11) where dahp is the starting precursor of the common aromatic amino acid biosynthesis pathway heading towards chorismate. Equation (4.12) recognizes chorismate as the carbon flux director of the three separate terminal pathways, its pivotal role in a series of sequences leading to tryptophan biosynthesis and its internal consumption during cellular protein synthesis. The reactions involving chorismate and prpp to form precursors (leading to tryptophan production), and the internal consumption of prpp are reflected in equation (4.13). The final set of reactions between serine and indole 3-glycerolphosphate to form tryptophan and gap, and serine internal consumption are collated in equation (4.14). Serine synthesis and RPPK enzymatic sub-system to regenerate prpp are given elsewhere in Chassagnole et al. (2002).

Chorismate synthesis and the internal consumption processes are modelled as first order kinetics in equations (4.15) to (4.18).

$$r_{ChoSynth} = r^{max}_{ChoSynth} C_{dahp} C_{pep} C_{e4p} C_{nadph} \tag{4.15}$$

$$r_{Synth3} = k_{choris} C_{choris} \tag{4.16}$$

$$r_{Synth4} = k_{prpp} C_{prpp} \tag{4.17}$$

$$r_{Synth5} = k_{ser} C_{ser} \tag{4.18}$$

Liao et al. (1996) suggested that e4p is the first limiting precursor for DAHP synthase, followed by phosphoenolpyruvate (pep). Chorismate synthesis originally presented as equation (10) in Schmid et al. (2004) has been modified in this study to incorporate erythrose 4-phosphate (e4p) as shown in equation (4.15).

There are four new parameters consisting of maximum rate of chorismate synthesis $r^{max}_{ChoSynth}$, rate constant of chorismate consumption k_{choris}, rate constant of prpp consumption k_{prpp} and rate constant of serine consumption k_{ser}. These four parameters are estimated by minimizing the sum of squares of errors (equation (4.19)) where error refers to the fractional difference between reference and predicted metabolite concentrations.

$$\text{Minimize } \sum \left(\frac{C_i^{reference} - C_i^{predicted}}{C_i^{reference}} \right)^2 \tag{4.19}$$

The reference concentrations are those of the 18 metabolites taken from the central carbon metabolism model of Chassagnole et al. (2002) and serine concentration determined from Piperno and Oxender (1968). The predicted concentrations are determined by solving the steady-state equations of the augmented model comprising the entire mass balances and kinetic expressions of the central carbon metabolism and equations (4.1) to (4.18) above. The steady-state concentration of serine is estimated in Appendix C as 0.04872 mM. The parameter estimates are discussed in the following section.

4.3 Parameter Estimation

The operon model parameters given in Santillán and Mackey (2001a) are based on a specific cell growth rate of 0.6 doublings per h. To be consistent with the central carbon metabolism of Chassagnole et al. (2002), our augmented model too uses a specific cell growth rate of 0.1 doublings per h (Appendix B) for the *trp* operon kinetics. The kinetic parameters of the central carbon metabolism are unchanged since experimental metabolomics data that are essential for re-calibrating the entire set of kinetic parameters in conjunction with the four new kinetic parameters estimation are not available in the published literature.

NSGA-II is used to estimate the four new parameters (Section 4.2.4) using equation (4.19) as the objective function. No steady-state measured concentrations of 3-deoxy-D-*arabino*-heptulosonate 7-phosphate (dahp), chorismate (cho), tryptophan (trp) and 5-phosphoribosyl-α-pyrophosphate (prpp) and pooled enzyme (represented by anthranilate synthase in equation (4.4)) for the wild *E. coli* strain are available in the published literature. Preliminary estimates of the four parameters give sufficiently close values for the rate constant of serine synthesis for various random number seeds. The remaining three rate constants could not be uniquely determined. An alternative approach is to treat the chorismate and prpp internal consumption fluxes as constant; then, the two unknown parameters to be estimated are the respective chorismate and serine synthesis rate constants. While the rate constant for serine synthesis remains largely unchanged for various random number seeds, the same is not true of the rate constant for chorismate synthesis.

Preliminary simulations of the augmented model shows that DAHPS flux is invariant to changes in the gene expression levels even though some of the

manipulated genes such as G6PDH (in the pentose-phosphate pathway) and PEPCxylase (in the glycolytic pathway) are known to be capable of redirecting or reducing carbon sources (Chapter 3) to the aromatic amino acid pathways. Preliminary studies on the simultaneous maximization of tryptophan and serine flux ratios using 2-gene manipulation (multiplier range 0.5-1.5) showed that tryptophan synthesis flux ratio is almost identical to 1.0 while serine synthesis flux ratio varies between 1.0 and 2.0.

On closer examination, it can be seen that equation (4.9) incorporates the effects of tryptophan and active pooled enzyme concentrations on the tryptophan synthesis rate without regard for the other key metabolites in the terminal tryptophan biosynthesis pathway. To provide a tighter link between the central carbon metabolism and the terminal tryptophan biosynthesis, it is necessary to replace equation (4.9). The pooled enzyme of the terminal tryptophan biosynthesis pathway, chorismate, prpp and serine are important precursors for the terminating steps in tryptophan biosynthesis. This forms the basis for the terminal tryptophan biosynthesis kinetics modelled using equation (4.20).

$$r_{TrpSynth} = r_{TrpSynth}^{max} C_{enz} C_{cho} C_{prpp} C_{ser} \tag{4.20}$$

Chorismate is consumed at the beginning and serine is consumed in the final step of the terminal tryptophan biosynthesis pathway. The conflicts between tryptophan and serine synthesis provides the motivation for the subsequent application of the augmented model for MOO in Chapter 5.

The internal consumption of tryptophan for cellular metabolism is modelled using first order kinetics.

$$r_{Synth6} = k_{trp} C_{trp} \tag{4.21}$$

The mass balance for tryptophan is written as:

$$\frac{dC_{trp}}{dt} = r_{TrpSynth} - r_{Synth6} - \mu C_{trp} \tag{4.22}$$

In the revised augmented model, this mass balance for tryptophan together with kinetic rate expressions (4.20) and (4.21) replace the earlier equations (4.5), (4.7), (4.8) and (4.9).

In the revised augmented model (equations (4.1) – (4.4), (4.6), (4.10) – (4.18) and (4.20) – (4.22) and the central carbon metabolism), there are six kinetic parameters to be estimated (see Table 4.2). The two additional kinetic parameters

among these six are the maximum rate of tryptophan synthesis ($r^{max}_{TrpSynth}$) and the rate constant of tryptophan consumption (k_{trp}). Tryptophan concentration for the *E. coli* wild strain is estimated as 4.091 μM (Santillán and Mackey, 2001a). The theoretical yield of dahp without pyruvate recycling to pep is 0.43 mole per mole of glucose (Liao et al., 1996). Since the steady-state glucose concentration is 0.0556 mM (Chassagnole et al., 2001), the estimated steady-state concentration of dahp is 0.43 of 0.0556 i.e. 0.0239 mM. In estimating 6 kinetic parameters, the reference concentrations are those of the 18 metabolites in the central carbon metabolism, serine, tryptophan and dahp. The last three metabolites are part of the shikimate and terminal tryptophan biosynthesis pathways.

In estimating the 6 parameters, data on only three reference metabolite concentrations (i.e. dahp, serine and tryptophan) are available for the shikimate and terminal tryptophan biosynthesis pathways. The steady-state concentrations of chorismate and prpp are not available. The concentrations of pooled enzyme in the terminal tryptophan biosynthesis pathway, dahp, serine and tryptophan are almost constant. Though there are variations in the chorismate and prpp concentrations when different random number seeds are used, fluxes of chorismate and prpp consumptions are practically constant at 0.00749 mM/s and 0.01036 mM/s respectively. This accounts for the variations in the rate constants of chorismate and prpp consumption. Note that fluxes of chorismate and prpp consumption vary in the optimization of concurrent gene knockout and manipulation (Chapter 5) to determine Pareto-optimal front. The estimated parameter values are shown in Table 4.2. They are selected by assuming that the concentration of chorismate is around 0.1 mM (Schmid et al., 2004). The minimal objective function value is constant at 0.028 for various random number seeds.

Table 4.2 Estimated parameters of the augmented model.

Parameter	Description	Estimated Value
$r^{max}_{ChoSynth}$	Maximum rate of chorismate synthesis	17.657000 $mM^{-3}s^{-1}$
$r^{max}_{TrpSynth}$	Maximum rate of tryptophan synthesis	63.193038 s^{-1}
k_{choris}	Rate constant of chorismate consumption	0.086983 s^{-1}
k_{prpp}	Rate constant of prpp consumption	10.269372 s^{-1}
k_{ser}	Rate constant of serine consumption	0.365393 s^{-1}
k_{trp}	Rate constant of tryptophan consumption	0.000103 s^{-1}

Table 4.3 gives steady-state metabolite/co-metabolite concentrations and fluxes following the estimated parameters of Table 4.2. The calculated values are subject to the homeostasis and total enzymatic flux constraints. Experimentally measured and estimated metabolite concentrations of Chassagnole *et al*. (2002), endogeneous serine concentration measured by Piperno and Oxender (1968), tryptophan concentration estimated by Santillán and Mackey (2001a) and theoretical dahp concentration of Liao et al. (1996) are in brackets. The co-metabolite concentrations are assumed to be constant. The calculated fluxes of the original central carbon metabolism model are in brackets. The synthesis rates of mureine and methionine are assumed to be constant in both the augmented and original central carbon metabolism models. Tryptophan synthesis rate, which is assumed to be constant in the central carbon metabolism model, is calculated in the augmented model. Nevertheless, the concentrations and fluxes of the augmented model are comparable to those given by the references above except for several relatively appreciable differences (more details in the later part of this section). Synth1 flux that collectively represents chorismate and mureine synthesis is assumed to be dependent on the concentration of pep but not that of e4p in the central carbon metabolism. Synth1 flux is not pertinent in the augmented model. Chorismate synthesis flux calculated by subtracting mureine synthesis flux from Synth1, instead of using equation 4.15, causes inconsistency in the results.

The parity plot (Figure 4.4) shows good agreement between the simulated metabolite concentrations using the estimated parameters and their respective measured (Chassagnole et al., 2001; Piperno and Oxender, 1968) and source (Liao et al., 1996; Santillán and Mackey; 2001a) values. The overall goodness-of-fit measure is 0.9992. The maximum normalized deviation, which is the ratio of the absolute difference (between the calculated and reference concentrations) to the reference concentration, is 8.26% for the metabolite fructose-1,6-bisphosphate (fdp). The normalized deviations for 3-phosphoglycerate (3pg), phosphoenolpyruvate (pep), 2-phosphoglycerate (2pg) and 1,3-diphosphoglycerate (pgp) are 5.70%, 5.66%, 5.65% and 5.62% respectively. The normalized deviations for the remaining metabolite concentrations are below 5.0%. The deviations are reasonable given the parametric uncertainties and structural assumptions (such as the mass balances used in the aromatic amino acid pathways) used in developing the augmented model.

The calculated enzyme synthesis rate in Table 4.3 is of the same order of magnitude (i.e. 10^{-5} μM/s) as the one calculated using the simplified *trp* operon model in Schmid et al. (2004). The *trp* operon model in Schmid et al. (2004) ignores the mechanisms of transcription and translation, and does not incorporate feedback inhibition of enzyme in the terminal tryptophan biosynthesis pathway. The augmented model developed here provides a platform for wider and deeper study (Chapters 5 and 6) of flux engineering interventions. Tryptophan synthesis flux shows the greatest normalized deviation (i.e. ratio of the absolute difference between the calculated and reference fluxes to the reference flux). The DAHPS flux deviates by 9.64% since dahp is the starting precursor of the aromatic amino acid biosynthesis pathways instead of being accumulated as were the case in the central carbon metabolism model. G1PAT flux deviates by 9.30%; varying DAHPS flux when tryptophan biosynthesis (TrpSynth) is considered in the augmented model has rather strong impact on G1PAT flux. PEPCxylase flux deviates by 4.27%. SerSynth flux deviates by 1.77% and its significance is not immediately obvious until the next chapter. The relatively appreciable differences in the fluxes obtained from the augmented model and central carbon metabolism model reflect the sensitivities of various pathways with respect to tryptophan biosynthesis. Such sensitivities point to underlying competition for carbon sources and provide opportunities for optimization study involving multiple objectives in the next chapter. The synthesis and consumption rates for chorismate, prpp, serine and tryptophan are relatively balanced as this is critical for the cellular metabolism.

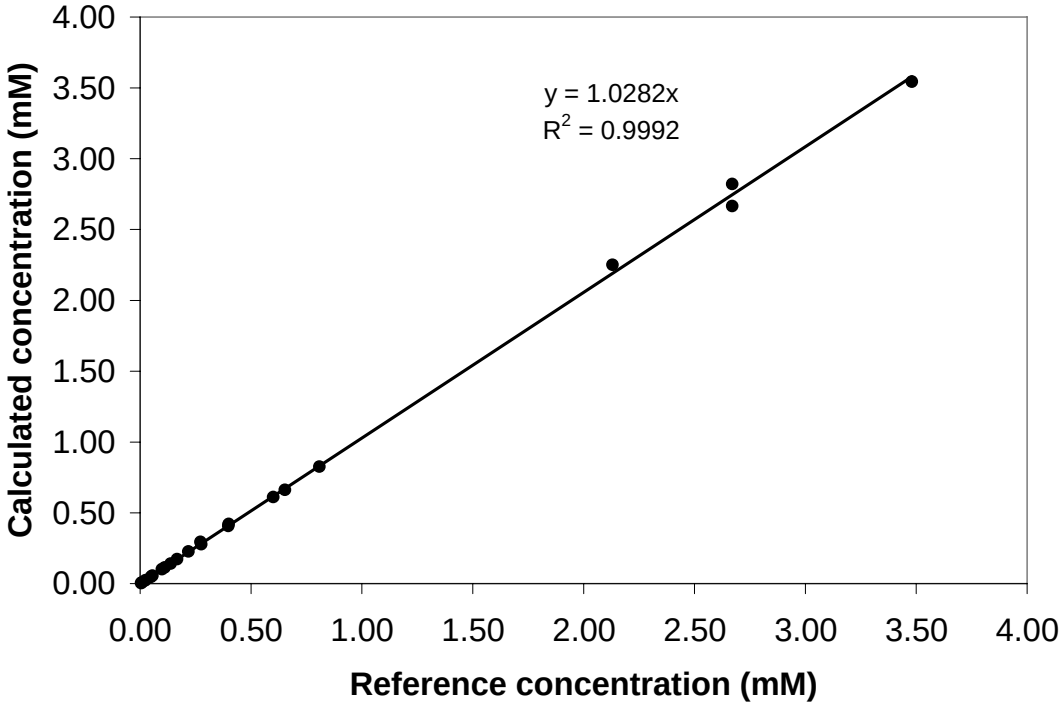


Fig. 4.4. Parity plot of the metabolite concentrations.

Table 4.3 Steady-state metabolite/co-metabolite concentrations and fluxes of the augmented metabolic network formed by integrating Figs. 4.1 and 4.2. Reference concentrations are in brackets: measured (meas), estimated (est) and theoretical (theo) are indicated. Calculated fluxes (except for MurSynth, MetSynth and TrpSynth which are constant) of the original central carbon metabolism are in brackets. See Section 4.3 for more details.

Metabolite/ Co-metabolite with serial number	Concentration (mM)	Enzyme with serial number	Flux (mM/s)
Metabolite		1. PTS	0.2000 (0.2000)
1. Glucose (extracellular)	0.05612 (0.0556, meas)	2. PGI	0.05584 (0.05825)
2. g6p	3.5447 (3.48, meas)	3. PFK	0.1398 (0.1410)
3. f6p	0.6112 (0.60, meas)	4. ALDO	0.1398 (0.1410)
4. fdp	0.2945 (0.272, meas)	5. TIS	0.1381 (0.1394)
5. gap	0.2276 (0.218, meas)	6. GAPDH	0.3166 (0.3199)
6. dhap	0.1740 (0.167,est)	7. PGK	0.3166 (0.3199)
7. pgp	0.00845 (0.008, est)	8. PGluMu	0.2988 (0.3023)
8. 3pg	2.2513 (2.13, est)	9. ENO	0.2988 (0.3023)
9. 2pg	0.4216 (0.399, est)	10. PK	0.03833 (0.03811)
10. pep	2.8212 (2.67, meas)	11. PDH	0.1870 (0.1878)
11. pyr	2.6657 (2.67, meas)	12. PEPCxylase	0.04496 (0.04312)
12. 6pg	0.8268 (0.808, meas)	13. PGM	0.002533 (0.002319)
13. ribu5p	0.1136 (0.111, est)	14. G1PAT	0.002515 (0.002301)
14. xyl5p	0.1414 (0.138, est)	15. RPPK	0.01036 (0.01031)
15. sed7p	0.2783 (0.276, est)	16. G3PDH	0.001722 (0.001658)
16. rib5p	0.4077 (0.398, est)	17. SerSynth	0.01780 (0.01749)
17. e4p	0.1015 (0.098, est)	18. Synth1	0.01443 (0.01421)
18. g1p	0.6622 (0.653, meas)	19. Synth2	0.05354 (0.05355)
19. pooled enzyme	0.00200	20. DAHPS	0.007495 (0.006836)
20. dahp	0.0239 (0.0239, theo)	21. G6PDH	0.1415 (0.1393)
21. cho	0.08612	22. PGDH	0.1415 (0.1393)
22. trp	0.004091 (0.004091, est)	23. Ru5P	0.08490 (0.08370)
23. ser	0.04872 (0.04872, meas)	24. R5PI	0.05657 (0.05559)
24. prpp	0.0010087	25. TKa	0.04620 (0.04527)
		26. TKb	0.03870 (0.03843)
Co-metabolite		27. TA	0.04620 (0.04526)
1. amp	0.955 (0.955, meas)	28. MurSynth	0.00043711 (0.00043711)
2. adp	0.595 (0.595, meas)	29. MetSynth	0.0022627 (0.0022627)
3. atp	4.27 (4.27, meas)	30. TrpSynth	0.000000535 (0.001037)
4. nadp	0.195 (0.195, meas)	31. TrpConsumed	0.000000421364
5. nadph	0.062 (0.062, meas)	32. EnzSynth	0.000000055562
6. nad	1.47 (1.47, meas)	33. EnzDegraded	0.0
7. nadh	0.1 (0.1, meas)	34. ChoSynth	0.007494
		35. ChoConsumed	0.007491
		36. PrppConsumed	0.01036
		37. SerConsumed	0.01780

4.4 Conclusions

Kinetic parameters of enzymatic sub-systems of the central carbon metabolism obtained through experiments and supported by theories in microbiology and biochemistry have been calibrated using stimulus response technique where a glucose pulse is injected into a culture of *E. coli*. The central carbon metabolism model has been used successfully in MOO in an earlier study (Chapter 3). What is lacking is the integration of the central carbon metabolism and the dynamic *trp* operon model. This provides the prime motivation for developing the augmented model presented in this chapter. The ability of the augmented model to channel carbon fluxes through the common aromatic amino acid and terminal tryptophan biosynthesis pathways is achieved by substituting the tryptophan biosynthesis kinetics of the dynamic *trp* operon model with one which reflects the key metabolites and representative enzyme of the terminal tryptophan biosynthesis pathway. Since it is not possible to uniquely estimate the six new kinetic parameters of the augmented model due to certain unavailable reference concentrations in the published literature, estimates are selected from the various parameter estimation runs to minimize the sum of squares of fractional errors. Since the existing *trp* operon model is independent of the central carbon metabolism, there is a need to carefully evaluate the augmented model in terms of reproducibility of metabolite concentrations and enzymatic fluxes. Steady-state concentrations and fluxes were calculated using the newly developed augmented model. Calculated and reference metabolite concentrations were compared via parity plot. Larger deviations between calculated and reference concentrations and fluxes are explained wherever possible. It is possible to discern underlying competition among DAHPS, TrpSynth, G1PAT and PEPCxylase fluxes for carbon sources. Concurrent gene knockout and manipulation for multiple objectives using the augmented model are studied in the next chapter.

Chapter 5
OPTIMIZATION OF A MICROBIAL CELL FACTORY FOR MULTIPLE OBJECTIVES – USING AUGMENTED MODEL

5.1 Optimization Studies of Microbial Cell Factories

Strain improvement traditionally relies on well-developed iterative classical mutagenesis and screening procedures. While these efforts have resulted in the creation of potent producer strains, the mutagenic approach has limited usefulness since desirable gene knockouts and manipulations are difficult to achieve and undesirable mutations are unavoidable. The availability of recombinant DNA techniques allows genetic interventions that will further improve existing producer strains. Optimization of a multi-product microbial biosynthesis factory such as *E. coli* via a suitable mathematical model, a systems biotechnology specialty, has the potential to complement the classical procedures used in improving producer strains. The metabolic pathway recipe formulated via mixed-integer multi-objective optimization can suggest desirable gene knockouts and manipulations and optimal mix of plasmid and aporepressor copy number for maximizing targeted amino acids production rates. Although amino acids are now among the classical products in biotechnology, rising demand is driving their annual growth enormously. Improvements in the production of amino acids, which are widely used in food, pharmaceutical and livestock sectors, are of significant interest to industries and consumers (Scheper et al., 2003).

Conflicting objectives are commonly encountered in bioprocesses (Halsall-Whitney et al., 2006; Lee et al., 2007; Sendin et al., 2006). Most of the optimization of a multi-product microbial cell factory focussed on a single objective (e.g., Schmid et al., 2004; Visser et al., 2004; Vital-Lopez et al., 2006). Numerical difficulties associated with the highly nonlinear central carbon metabolism model is circumvented through piecewise optimization (Schmid et al., 2004), linlog approximation (Visser et al., 2004) and Lagrange linearization (Vital-Lopez et al., 2006). By setting the time derivatives of the metabolite concentration to zero in a system of differential mass balances, stationary fluxes and metabolite concentrations can be calculated when engineering interventions in the form of gene knockouts

and/or manipulations are introduced. In all cases, desirable interventions divert carbon sources to maximize the production of a targeted end-product.

Schmid et al. (2004) used a nonlinear central carbon metabolism model for *E. coli* linked to several first order kinetic expressions for tryptophan biosynthesis pathway and an equation for enzyme synthesis to maximize tryptophan production via enzyme modulations. They also studied the flux distributions and flux control coefficients. Metabolic control analysis (MCA) can suggest enzymes to be modulated to drive the metabolic pathway recipe towards a given objective function. As noted by Schmid et al. (2004), there are trends in their calculated flux control coefficients that do not correlate with the optimal enzyme activity changes and thus MCA can never replace optimization of a detailed model. The number of flux and concentration control coefficients accessible through MCA in a complex nonlinear central carbon metabolism model is so large that the challenge of identifying the Pareto-optimal result through MCA is yet to be surmounted.

Visser et al. (2004) used a linlog approximation of the central carbon metabolism model of *E. coli* to determine the optimal glycolytic enzyme modulations required to maximize glucose uptake through phosphotransferase sub-system (PTS) in one case and the production of serine in another case. Tryptophan production rate is assumed to be constant. Serine is a precursor for tryptophan biosynthesis, and the linkage between tryptophan and serine production rates is subsumed under the central carbon metabolism. Only ten (and eleven in the case of maximizing serine production) out of the thirty enzymatic fluxes present in the central carbon metabolism model were used as decision variables. It is not known whether the single objective optima obtained by Schmid et al. (2004) and Visser et al. (2004) are global or local when piecewise optimization results were combined.

Multi-objective optimization of multi-product microbial cell factories is relatively new in chemical engineering. Vera et al. (2003) have studied multi-objective optimization in metabolic processes leading to ethanol production by *Saccharomyces cerevisiae*; this was most recently reiterated by Link et al. (2008). Sendin et al. (2006) have compared the multi-objective optimization of ethanol production in *Saccharomyces cerevisiae* applied separately using weighted sum method, goal attainment method, normal boundary intersection (NBI), multi-objective indirect optimization method (MIOM) and multi-objective evolutionary algorithm (MOEA). The MIOM approach is workable when a nonlinear model is converted into

an equivalent S-system (synergistic systems) by aggregating all reactions contributing to generation or consumption of a given metabolite as a product of power functions. The power functions of the generation (influx term) or consumption (efflux term) of a given metabolite are formed by multiplying the rate constant and each contributing metabolite concentration raised to its own real exponent. The power-law approximation in the form of S-system and generalized mass action (GMA) provide the modelling framework under biochemical systems theory (BST). The S-system is converted into a system of linear equations at steady-state by logarithmic transformation of the influxes and effluxes. Applying multi-objective linear programming (MOLP) to S-system is numerically less daunting than applying other multi-objective optimization methods to the original highly nonlinear model. BST model is not available for the central carbon metabolism of *E. coli*.

Producing tryptophan and serine in one organism rather than two different organisms is a case for flexible multi-product biosynthesis factory. The flexibility is scalable through engineering interventions of the metabolic pathways to respond to changes in types of products and their quantities demanded in a highly competitive market. On a broader perspective, manufacturers will be able to shorten the research and development time using a multi-product biosynthesis factory serving as a re-configurable biocatalytic template to produce novel and useful products of the future. Organisms such as *Escherichia coli* and *Corynebacterium glutamicum* have similar central carbon metabolism and branched pathways such as those leading to aromatic amino acids and serine biosynthesis. Therefore single-product technologies accumulated through the studies of an individual organism can be channeled towards the singular purpose of studying a multi-product biosynthesis factory. Our current research is a step in exploring the challenges of designing a multi-product factory via *Escherichia coli* from metabolic engineering viewpoint. The thesis author selects *Escherichia coli* motivated primarily by the availability of detailed models and extensive literature. Advances in post-fermentation separations will be less of an issue when costs and substrate availability for fermentation are relatively more important.

L-serine and L-tryptophan have a symbiotic relationship in *Escherichia coli*. In the final step of the terminal tryptophan biosynthesis pathway, the conversion of L-serine and indole into L-tryptophan and water is catalyzed by the β_2 subunit of tryptophan synthase. The manufacturer determines the relative amounts of serine and tryptophan to be produced using a single fermenter based on a high-level aim such as

economic demand. Serine is soluble in the aqueous fermentation broth. Tryptophan is sparingly soluble in the aqueous solution and appears as solid crystals towards the end of fermentation. Post-fermentation steps for tryptophan include filtration and crystallization. Ion-exclusion chromatography can be used to extract serine from the same post-fermentation solution. Using two fermenters rather than one fermenter potentially increases the post-fermentation separation steps needed to achieve products of desired quality.

Similar to the earlier study (Chapter 3) on the multi-objective optimization of the central carbon metabolism of *E. coli* (Figure 3.1), the present study uses the highly nonlinear augmented model developed in Chapter 4. The next section describes the optimization problem and solution strategy.

5.2 Optimization Problem and Solution

5.2.1 Problem Formulation

Gene repression and overexpression (i.e. gene manipulation) as well as gene knockouts help to redistribute the various metabolic fluxes in the central carbon metabolism (Chapter 3). The challenge lies in identifying the genes to be manipulated or knocked out so as to simultaneously optimize the desired fluxes leading to the production of useful amino acids as end-products. Concurrent gene knockout and manipulation will naturally amplify the targeted fluxes to a greater extent than using either one technique due to the increased system-wide flexibility in debottlenecking and enhancing fluxes of various enzyme-catalyzed reactions. In the targeted tryptophan and serine bi-objective optimization through concurrent gene knockout and manipulation using the augmented model, maximizing tryptophan flux ratio replaces that of the DAHPS flux ratio in the earlier study (Chapter 3). Maximizing serine flux ratio remains as the other objective.

Besides concurrent gene knockout and manipulation, inserting multicopy plasmids carrying tryptophan operon into a wild *E. coli* host cell, which is a recombinant DNA technology, has the potential of increasing tryptophan biosynthesis rate. Assuming one DNA operator site per tryptophan operon in an independent plasmid, the number of DNA operator sites can be multiplied by inserting multicopy plasmids into the wild *E. coli* strain. Preliminary bi-objective optimization trials show that multiplying the plasmids carrying tryptophan operon in the augmented model

leads to severe numerical difficulties. This suggests that the stability of the microbial cell system cannot be maintained or regulated in the absence of repressors due to difficulties in controlling the dynamic expression of genes. This is slightly alleviated by inserting multiple copies of plasmids carrying *trpR* gene encoding the aporepressors into the wild *E. coli* host cell. Difficulties in controlling the microbial cell system when plasmids were inserted into the wild *E. coli* host cell have been reported in various experiments (Azuma et al., 1993; Ikeda, 2006).

The numerical difficulties associated with multicopy plasmids in the augmented model are attributed to complex cellular regulatory actions in the presence of inserted plasmids carrying tryptophan operon and *trpR* genes. Carefully constructed multicopy plasmids carrying tryptophan operon with other modifications (such as deregulated *trpE* and *trpD* genes) have been inserted into wild *E. coli* and *Corynebacterium glutamicum* (*C. glutamicum*) strain (Ikeda, 2006) to produce higher tryptophan titre at the end of fermentation. In the pursuit of ever higher tryptophan titre, sustainability of the inserted plasmids in the *E. coli* host cell throughout the entire course of fermentation to produce tryptophan remains a practical challenge till now (Ikeda, 2006). Thus parallels between experimental and computational work are manifested in the need for a microbial cell to maintain metabolic network integrity and stability.

Tryptophan operon and *trpR* plasmids multiplications are not decision variables in the current optimization study though preliminary simulations (data not shown) show that multicopy plasmids are capable of enhancing tryptophan biosynthesis rates several times; this is supported by fermentation experiments (Azuma et al., 1993; Chan et al., 1993) using various growth media. The two decision variables are not included in the current study as they merit special metabolic network integrity and stability considerations that form part of future study. The augmented model does not incorporate detailed mathematical mechanisms of the various gene deregulations by the aromatic amino acids due to lack of relevant published kinetics. Another potential decision variable is the cellular specific growth rate. Tryptophan flux ratio increases as specific growth rate increases since more tryptophan is needed for internal cellular consumption at higher biomass growth rate. On the other hand, tryptophan concentration decreases as specific growth rate increases due to the dilution effect of higher biomass growth rate. The specific growth rate of the augmented model is fixed at 0.1 h^{-1} in order to maintain kinetic parameters

consistency throughout. Tryptophan synthesis and consumption rates remain coupled although the specific growth rate is constant.

A summary of the objectives, constraints and decision variables are given below.

Objectives

$$\text{Max } r_{TrpSynth} \tag{5.1}$$

$$\text{Max } r_{SerSynth} \tag{5.2}$$

This is equivalent to maximizing their respective flux ratios (Chapter 3).

Steady-state constraints

They are obtained by setting the time derivatives of the mass balances of the augmented model (equations 4.1, 4.3, 4.4, 4.10 to 4.14, 4.22 and equations found in Table 1 (Chassagnole et al., 2001)) to zero.

Homeostatic constraint

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\left|C_i - C_{i,ref}\right|}{C_{i,ref}} \leq 0.3 \tag{5.3}$$

This is a physiological constraint. The summation is over all metabolites (m in number) and C_i is the concentration of i^{th} metabolite. The value of m is 23, consisting of 18 central carbon metabolites (Table 4.3) and 5 metabolites (3-deoxy-D-*arabino*-heptulosonate 7-phosphate (dahp), chorismate (cho), tryptophan (trp), 5-phosphoribosyl-α-pyrophosphate (prpp) and serine (ser)) of the common aromatic amino acids and terminal tryptophan biosynthesis pathways (Figures 4.1 and 4.2). The constraining value of 0.3 is the same as the one used in the earlier optimization of the central carbon metabolism.

Total enzymatic flux constraint

$$\frac{1}{z}\sum_{i=1}^{z}\frac{r_i}{r_{i,ref}} \leq 1.0 \tag{5.4}$$

This is a technological constraint, where the summation covers all enzyme fluxes (z in number)**.** The value of z is 36 consisting of fluxes in the central carbon metabolism (except Synth1) and common aromatic amino acid and terminal tryptophan biosynthesis pathways. The constraining value of 0.1 is the same as the one used in

the earlier optimization of the central carbon metabolism. The initial steady-state (reference) values for the concentrations and fluxes are given in Table 4.3.

Decision variables

There are 30 genes (PTS, PGI, PFK, ALDO, TIS, GAPDH, PGK, PGluMu, ENO, PK, PDH, PEPCxylase, PGM, G1PAT, RPPK, G3PDH, SerSynth, Synth2, DAHPS, G6PDH, PGDH, Ru5P, R5PI, TKa, TKb, TA, MurSynth, MetSynth, ChoSynth and TrpSynth – refer to the List of Symbols for the definition of these abbreviations and Table 4.3) subjected to manipulation using a multiplier in the range of 0.8-1.25 during the gene identification process and a multiplier in the range of 0-1.5 during concurrent knockout and manipulation.

5.2.2 Solution Strategy

It is not practical to use a manual exhaustive search method for concurrent gene knockout and manipulation due to the large number of possibilities. A branch - and-bound method (Section 3.6) works relatively well for up to three gene knockouts where only integer decision variables are involved. It is not suitable where real decision variables in the form of gene multiplier, whether continuous or discontinuous, are involved. Identification of potential genes for knockout and manipulation solely by intuition risks missing parts of the search domain with high nonlinearity.

A heuristics method is applied to the current optimization study by drawing on the strengths of the branch-and-bound (pattern recognition) and manual exhaustive search (checking of metabolic network integrity and exclusion of genes that breach constraints) techniques discussed in Chapter 3. The strategy is to first use NSGA-II to identify potential genes for concurrent knockout and manipulation. In the first stage of gene manipulation, two (or three) genes are randomly selected from the universal set of 30 genes, via a relatively narrow multiplier range of 0.8-1.25 to identify a subset that is responsible for the Pareto. The finally selected subset of genes from the first stage is then considered with multiplier in the range between 0 and 1.5 to generate Pareto-optimal front via concurrent knockout and manipulation. Single objective optimization is used to check the upper bounds of the Pareto-optimal front. By recognizing the Pareto patterns and genes obtained via the narrow multiplier range of

0.8-1.25, simulations are used to obtain part of the Pareto which cannot be generated using NSGA-II. The main focus is on two and three gene manipulations and knockouts. The computer used and the NSGA-II parameters (except for random seeds) in this chapter are the same as those in Chapters 2 and 3.

5.3 Gene Identification

5.3.1 Two-gene Identification Using Multiplier of 0.8 to 1.25

Pareto-optimal results obtained using NSGA-II (Figure 5.1a) reveal the existence of two separate segments and two isolated chromosomes C_1 and C_2. Discontinuity between the two separate Pareto segments indicates an on-off switching mechanism with forward and backward paths favouring tryptophan and serine biosynthesis respectively. The sharp on-off pattern is a result of switching between two sets of genes (Figure 5.1b) responsible for enhancing tryptophan and serine biosynthesis. Transcription termination by attenuation is relaxed to allow higher tryptophan synthesis rate when tryptophan concentration decreases to a sufficiently low value. Attenuation is more severe at higher tryptophan concentration in order to retard tryptophan synthesis rate. One reported tryptophan concentration range for effective attenuation is 1 to 5 μM (Koh et al., 1998). It is interesting to note that the initial steady-state concentration of tryptophan is 4.091 μM (Table 4.3). Regulation by repression at the DNA operator site of the tryptophan operon is effective over a wider tryptophan concentration range of 1 to 100 μM (Koh et al., 1998).

Unlike the sharp on-off switch function of attenuation, repression uses a gentler variable gain method (relatively more effective for tryptophan concentration less than 10 μM) to control tryptophan synthesis rate. In the concurrent gene knockout and manipulation (Section 5.3.2), tryptophan synthesis rate and concentration vary in the same direction. While the regulatory actions of attenuation, repression and feedback inhibition of enzyme maintain tryptophan synthesis rate and concentration within precise bands for given gene expression levels, it is the gene manipulation and/or knockouts that largely determine the tryptophan flux ratio and concentration at various gene expression levels. As a whole, the augmented model is able to demonstrate the balancing of contrasting needs of synthesis and internal consumption at various gene expression levels to sustain a given specific growth rate. The small difference (less than 1%) between synthesis and internal consumption rates accounts

for the dilution effect of biomass growth. Mass transport of tryptophan and serine across the cell membrane is not part of this study.

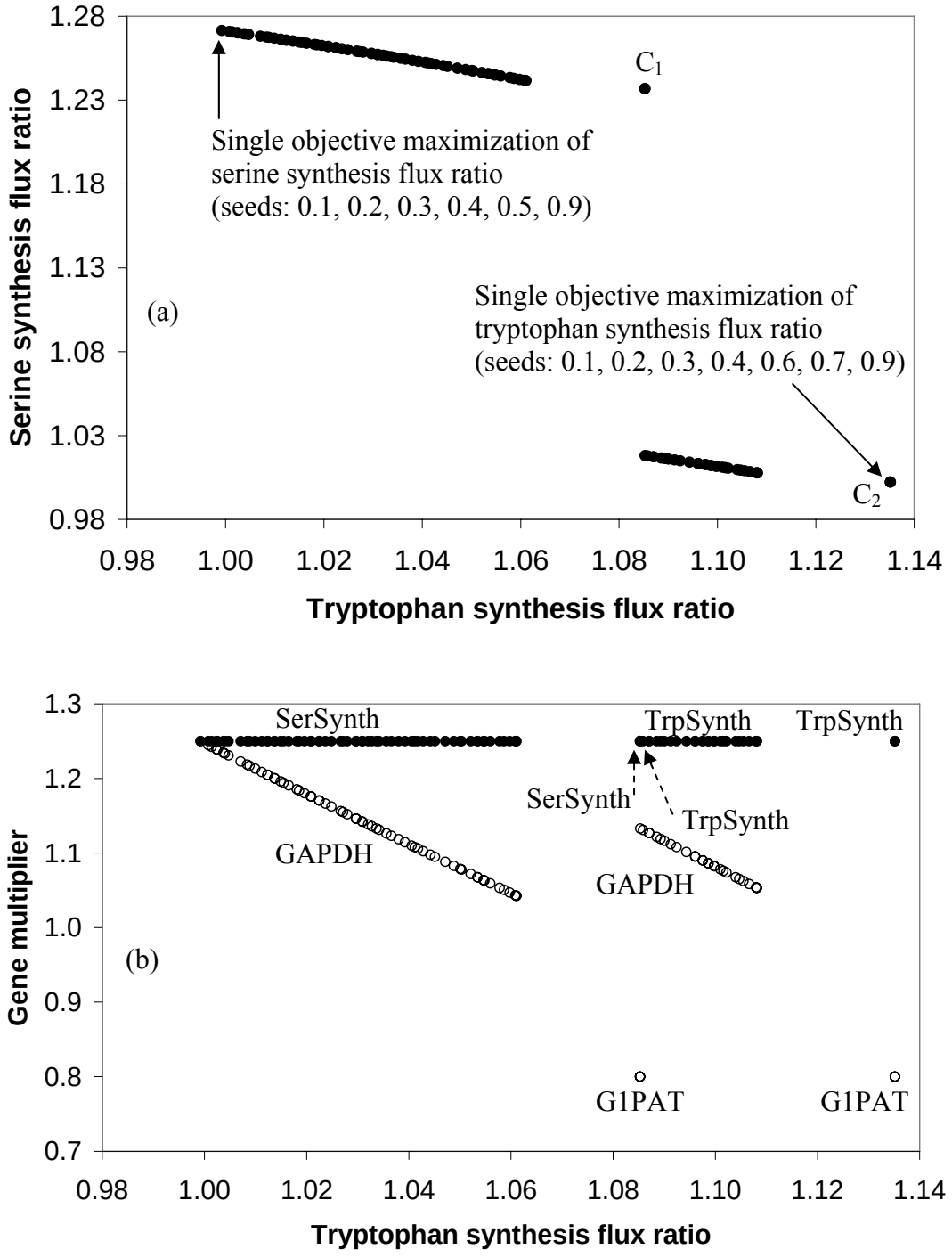


Fig. 5.1. (a) Pareto-optimal front obtained by two-gene manipulation with seed 0.6 and multiplier range of 0.8 to1.25.
(b) Optimal gene multipliers for the Pareto-optimal front in (a).

Maximum serine synthesis rate (serine flux ratio = 1.271541; tryptophan flux ratio = 0.999243) is obtained when the multipliers of SerSynth and GAPDH are

actively constrained at 1.25 (Figure 5.1b). Maximum tryptophan synthesis rate (tryptophan flux ratio = 1.135159; serine flux ratio = 1.002157) is obtained when the multipliers of TrpSynth and G1PAT are actively constrained at 1.25 and 0.8 respectively. Not all of the random number seeds used with NSGA-II are capable of finding the respective global upper bounds of serine and tryptophan synthesis rates (Figure 5.1(a)). One such case (seed = 0.5) is when both RPPK and R5PI are multiplied 1.25 times (not shown in Figure 5.1), and one would expect that greater availability of carbon sources in the pentose-phosphate pathway results in higher tryptophan synthesis. The resulting tryptophan flux ratio is 1.076906. It is less than that obtained through manipulation of genes encoding TrpSynth and G1PAT. Repressing gene encoding G1PAT by 0.8 times effectively draws carbon sources from polysaccharides synthesis and directs them to pentose-phosphate and glycolytic pathways (Figure 4.1). Overexpressing gene encoding TrpSynth by 1.25 times increases carbon fluxes in the terminal tryptophan biosynthesis pathway. In the earlier study (Chapter 3), repressing gene encoding G6PDH (multiplier = 0.5) and overexpressing gene encoding DAHPS (multiplier = 2.0) maximizes DAHPS flux ratio at 1.498. G6PDH flux ratio decreases slightly to 0.964. In the augmented model, overexpressing gene encoding DAHPS which is at the start of the common aromatic amino acid pathway does not increase tryptophan flux ratio to the same extent as the one obtained by overexpressing gene encoding TrpSynth. Note that the gene encoding TrpSynth collectively represents all the genes (after ChoSynth) in the terminal tryptophan biosynthesis pathway.

In another case (seed = 0.8; not shown in figure), repressing gene encoding DAHPS by 0.8 times and overexpressing gene encoding SerSynth by 1.25 times gives a serine flux ratio of 1.250442. While repressing gene encoding DAHPS is intuitive, higher flux ratio is possible via overexpressing gene encoding GAPDH. In the augmented model, repressing gene encoding DAHPS reduces demand for serine as an immediate precursor in the final step of tryptophan biosynthesis of the longer branched aromatic amino acid pathway. This suggests that GAPDH is more effective in increasing carbon sources to the short branched serine synthesis pathway due to its greater proximity. This is congruent with the earlier results (Chapter 3) where maximum serine flux ratio is obtained by overexpressing genes encoding SerSynth and GAPDH at the multiplier upper bound.

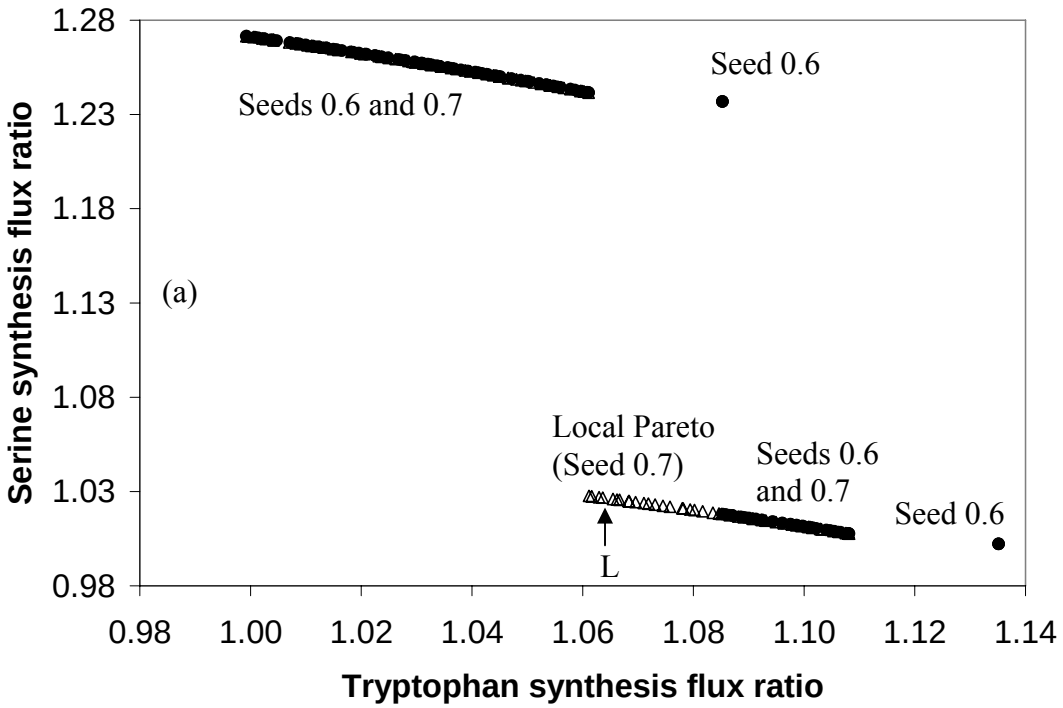
GAPDH (Figures 4.1 and 5.1(b)) is central in channelling carbon sources to serine synthesis. Repressing gene encoding G1PAT certainly increases the carbon availability for serine and tryptophan synthesis through the respective glycolytic and aromatic amino acid pathways, as indicated by chromosomes C_1 and C_2. Knocking out gene encoding G1PAT is a possibility in the concurrent gene knockout and manipulation (Section 5.4). While it appears intuitive to manipulate genes encoding DAHPS and ChoSynth so as to enhance tryptophan flux ratio, they do not contribute to the Pareto (Figure 5.1(b)). This suggests that the limiting precursors reside in the central carbon metabolism rather than the branched aromatic amino acid pathway.

It is likely that a local Pareto (Figure 5.2(a)) is obtained for seed 0.7 with the two isolated Pareto chromosomes C_1 and C_2 absent. This means, when a chromosome such as L (Figure 5.2a) in the local Pareto set is perturbed in the decision variable space, no dominating solution in the neighbourhood can be found. The local Pareto is perpetuated by extending the multiplier ranges of TrpSynth and GAPDH (Figure 5.2(b)) beyond those of the Pareto-optimal front. Similar local Pareto is also obtained for seed 0.8 (results not shown for brevity). An isolated local Pareto chromosome on the left of the local Pareto segment is obtained by manipulating PK (multiplied by 0.8 times) and GAPDH (multiplied 1.25 times). Other seeds used in NSGA-II result in either incomplete parts of the Pareto-optimal front or a local Pareto.

Genes that are potentially capable of generating the Pareto-optimal front in the concurrent knockout and manipulation are SerSynth, TrpSynth, GAPDH, G1PAT and PK. The subset of genes consisting of SerSynth, TrpSynth, GAPDH and G1PAT are being manipulated with a multiplier range of 0.8 to 1.25. Using this subset of genes only, NSGA-II (with seeds 0.6, 0.7 and 0.8) reproduces the same Pareto-optimal front as that obtained using the universal set of 30 genes. Further optimization runs were made for another subset of genes consisting of PK, SerSynth, TrpSynth, GAPDH and G1PAT. Pareto-optimal front is obtained for seeds 0.6 and 0.8. Local Pareto identical to the one shown in Figure 5.2(a) is obtained for seed 0.7.

The four genes consisting of SerSynth, TrpSynth, GAPDH and G1PAT will be used in the concurrent gene knockout and manipulation with a wider multiplier range of 0 to 1.5. While it is intuitive to manipulate SerSynth and TrpSynth to increase serine and tryptophan synthesis rates, manipulating GAPDH and G1PAT are not obvious. This is due to the complex dynamic interactions of the flux channels existing in the central and branched pathways. The heuristics method successfully identifies

the genes and their respective multipliers contributing to global and local Pareto. The reduced set of genes and Pareto patterns provide more specific guidance to the search for converged Pareto-optimal solution in concurrent gene knockout and manipulation (Section 5.4). Inability to find the global Pareto could be due to various reasons such as early convergence that reduces the probability of finding non-dominated solution(s) in other parts of the feasible search space resulting in bypass of isolated solution or disjointed Pareto segment, difficulties in accessing certain parts of the feasible search space, and non-convexities in the objective and decision variable spaces. Therefore, there is a need to try a number of values for the random number seed in the optimization runs to obtain the global Pareto-optimal front. In all the optimization runs, the Pareto-optimal solution converges within 500 generations. Heuristics technique alleviates (though not completely) the daunting numerical difficulties in the single step search for the Pareto when concurrent knockout and manipulation is applied to 30 genes.



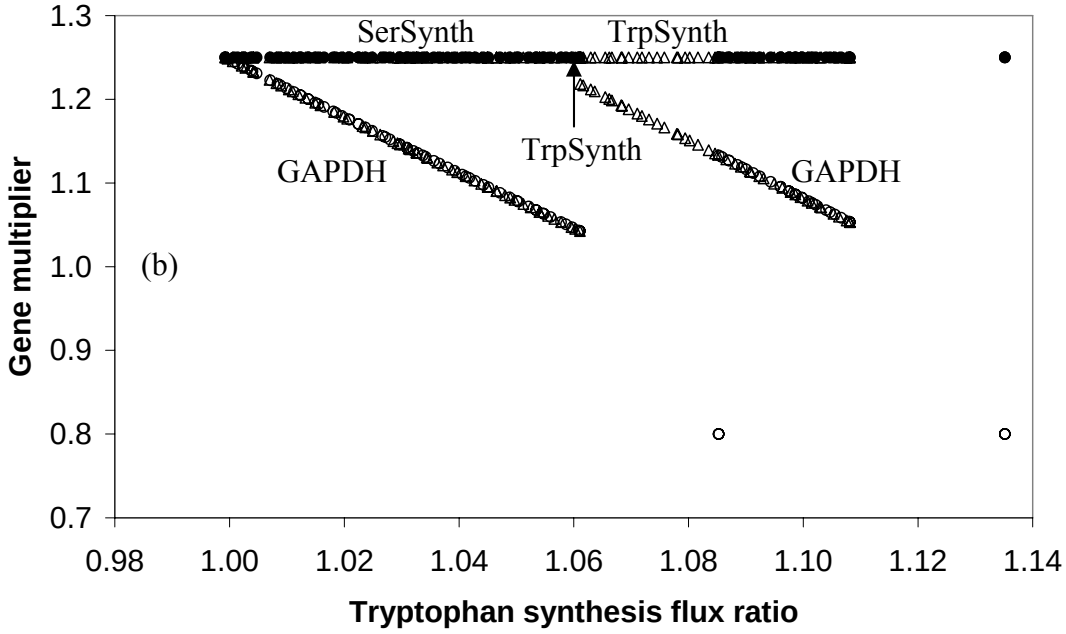


Fig. 5.2. (a) Pareto front obtained by two-gene manipulation with multiplier in the range 0.8 to1.25 and seeds: 0.6 (●) and 0.7 (Δ).
(b) Optimal gene multipliers for the Pareto-optimal front in (a). Gene labelling is for seed 0.7.

5.3.2 Three-gene Identification Using Multiplier of 0.8 to 1.25

The Pareto-optimal front for three-gene manipulation (Figure 5.3a) is generated by genes encoding SerSynth, TrpSynth, GAPDH, RPPK, G1PAT and PK (Figure 5.3b). The maximum serine synthesis rate (serine flux ratio = 1.300026; tryptophan flux ratio = 1.088622) indicated by chromosome C_3 is obtained by overexpressing genes encoding GAPDH and SerSynth (multiplier 1.25) and repressing gene encoding PK (multiplier 0.835276). Repressing another gene besides the one encoding PK either breaches the total enzymatic flux constraint (e.g. multiplying gene encoding PEPCxylase by 0.8 times) or does not result in the highest serine flux ratio. The maximum tryptophan synthesis rate (serine flux ratio = 0.992493; tryptophan flux ratio = 1.214460) indicated by chromosome C_4 is obtained by overexpressing genes encoding RPPK and TrpSynth (multiplier 1.25) and repressing gene encoding G1PAT (multiplier 0.8). Although higher tryptophan flux ratios (e.g. repressing gene encoding PEPCxylase by 0.8 times and overexpressing genes encoding RPPK and TrpSynth by 1.25 times) than those of chromosome C_4 are obtainable, they violate the total enzymatic flux constraint.

Overexpressing gene encoding TrpSynth creates high demand for serine as a precursor in the final step of tryptophan biosynthesis. This is achieved by overexpressing genes encoding SerSynth and GAPDH to increase carbon availability in the glycolytic pathway (Segment 1 in Figure 5.3a). An analogous collaborative demand and supply of precursor occurs when experimental amplification of gene encoding SerSynth increases tryptophan production in *Corynebacterium glutamicum* (Ikeda, 2006) where serine is a limiting precursor. As overexpression of gene encoding GAPDH slides, greater carbon flux available for TrpSynth helps to increase tryptophan flux ratio. Continual decrease in gene expression encoding SerSynth combined with continual increase in gene expression encoding TrpSynth and repression of gene encoding G1PAT enhance the tryptophan flux ratio. The competition between TrpSynth and SerSynth supported by diversion of carbon fluxes away from polysaccharide synthesis generate segment 2 in Figure 5.3a.

Further increase in tryptophan flux ratio, indicated by segment 3, is possible through repressing gene encoding G1PAT (to divert carbon fluxes to glycolytic and pentose-phosphate pathways), maximizing gene expression encoding TrpSynth and continually increasing gene expression encoding RPPK. This is different from the objective of increasing DAHPS flux ratio (Chapter 3) where repressing gene encoding G6PDH, decreasing gene expression encoding SerSynth and maximizing gene expression encoding DAHPS is more effective. In the augmented model, increasing RPPK flux helps to increase the production rate of prpp which is a precursor of ANTAP enzymatic sub-system (Figure 4.2) in the terminal tryptophan biosynthesis pathway. Repressed gene encoding G6PDH (to increase DAHPS flux in Chapter 3) does not appear in the Pareto-optimal front for seed 0.6 in Figure 5.3a. Repressing gene encoding G6PDH (multiplier 0.845845) and overexpressing genes encoding GAPDH (multiplier 1.25) and SerSynth (multiplier 1.214800) are responsible for an isolated Pareto chromosome between chromosome C_3 and segment 1 (not shown in Figure 5.3a) favouring serine production (serine flux ratio = 1.288708; tryptophan flux ratio = 1.101746) if random number seed 0.3 were used.

Besides gene encoding G6PDH, other genes pertaining to the pentose-phosphate pathway do not contribute to the Pareto-optimal fronts for dual and triple gene manipulation. Several independent studies (Ikeda and Katsumata, 1999; Kim et al., 2000) using tryptophan producing *C. glutamicum* or *E. coli*, have shown that overexpressing genes encoding transketolase (TK_a and TK_b) using high copy number

plasmids imposes a metabolic burden leading to retarded growth and segregation of plasmids. When glycolysis is blocked by deleting gene encoding PGI in the glycolytic pathway and carbon flux is diverted into the pentose-phosphate pathway (Mascarenhas et al., 1991), tryptophan production rate doubled but the cellular growth rate is reduced. Since the specific growth rate is constant in the augmented model, the ability to manipulate genes in the pentose-phosphate pathway is limited to moderate gene overexpression. This suggests that repressing or even knocking out gene encoding G6PDH (Chapter 3) which is at the start of the pentose-phosphate pathway, does not limit the availability of erythrose-4-phosphate (e4p) as a precursor for DAHPS to synthesize 3-deoxy-D-*arabino*-heptulosonate 7-phosphate (dahp). This is possible since glyceraldehyde-3-phosphate (gap), a precursor for synthesizing e4p, is produced in glycolytic pathway and the final step of tryptophan biosynthesis.

Local Pareto (not shown in Figure 5.3) is observed for various seeds (i.e. 0.1, 0.2, 0.4, 0.5, 0.7, 0.8 and 0.9) or variations of the multiplier range such as 1.0 to 1.25. The subset of genes that are potentially capable of generating the Pareto-optimal front in the concurrent three-gene manipulation and knockouts are SerSynth, TrpSynth, GAPDH, RPPK, G1PAT, PK and G6PDH. Numerical difficulties are significantly greater in the case of concurrent three-gene manipulation and knockouts than those of the two-gene case due to larger number of metabolic pathways subjected to engineering interventions that result in infeasible solutions. This is also true in practice (personal comments from experts e.g. Christian Wandrey; Ikeda and Katsumata, 1999) where maintaining metabolic control and stability in *E. coli* subjected to gene manipulation and/or deletion is a challenge, particularly during fermentation. Concurrent two-gene manipulation and knockout is discussed in the next section.

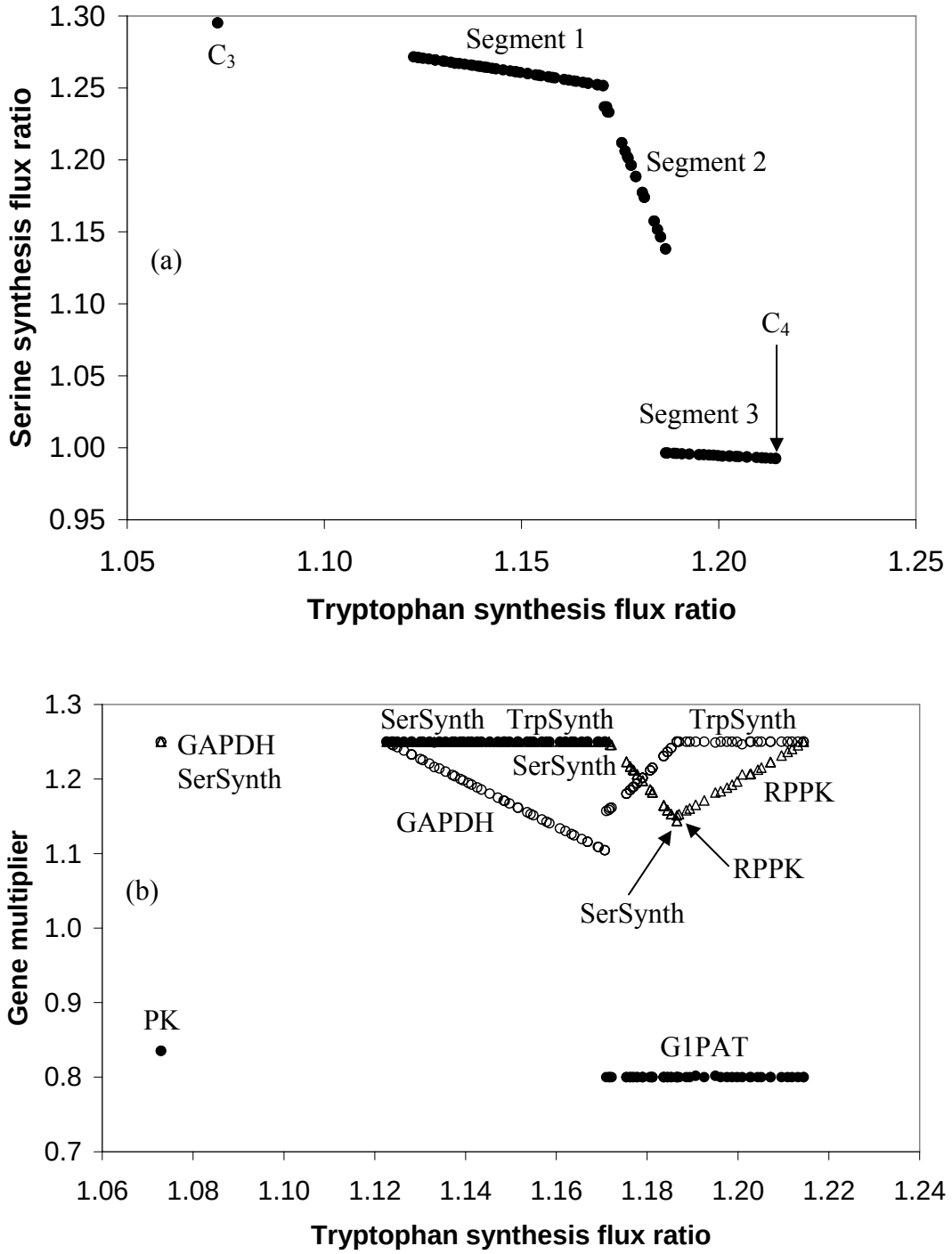


Fig. 5.3. (a) Pareto-optimal front obtained by triple gene manipulation with seed 0.6 and multiplier range of 0.8 to1.25.
(b) Optimal gene multipliers for the Pareto-optimal front in (a).

5.4 Concurrent Two-gene Knockouts and Manipulations Using Multiplier 0 to 1.5

5.4.1 Pareto and Practical Issues

In this case, two randomly selected genes from the subset consisting of GAPDH, G1PAT, SerSynth and TrpSynth, are simultaneously deleted and manipulated for optimizing the objectives. The optimizer, NSGA-II generates the leftmost Pareto segment bounded by chromosomes A1 and A2 (Figure 5.4a) for random number seeds 0.1, 0.6 and 0.7; local Pareto is absent in all the three optimization runs. The rest of the Pareto results (A3, B1-B2 and B3) are obtained through selective simulation built upon the Pareto patterns and decision variables behaviour observed during the gene identification step. The homeostatic and total enzymatic flux constraints are satisfied in all the Pareto results in Figure 5.4. Similar to the Pareto trends in the earlier study (Chapter 3), there is an overall increase in flux ratios due to greater gene amplification, where serine and tryptophan flux ratios increase up to 53.5% and 30.5% respectively. An on-off switching mechanism allows switchover to TrpSynth to increase tryptophan synthesis rate in the forward path. The backward path allows switchover to SerSynth if greater serine synthesis rate is desired. The gene encoding GAPDH regulates the amount of carbon sources through the glycolytic pathway.

Neither genes in the pentose-phosphate pathway nor gene encoding DAHPS are responsible for the Pareto (Figure 5.4b). Though overexpressing genes encoding transketolase (TKa and TKb) in the pentose-phosphate pathway can increase the production of e4p in practice (Bongaerts, 2001), there is a conjecture supported by experimental fermentation data (Ikeda and Katsumata, 1999) that accumulation of small amount of pentose phosphates within a microbial cell inhibits cell growth. Experimentally, it was demonstrated (Patnaik and Liao, 1994) that overexpressing genes encoding DAHPS, transketolase and PEP synthase (to convert pyruvate (pyr) back to phosphoenolpyruvate (pep)) to increase the supply of pep and e4p lead to production of dahp close to the theoretical yield (obtained from metabolic flux analysis by the same authors) of 0.86 mole per mole of glucose. While this strategy works well for high-density cell resuspension cultures under non-growth conditions in laboratories (Ikeda and Katsumata, 1999), such genetic modifications leading to growth impairment during fermentation in industrial bioreactors are practically

undesirable. The augmented model which is based on a constant specific growth rate of 0.1 doublings per hour has demonstrated its ability to balance synthesis and consumption to give Pareto results that generally concur with the work described above and the earlier study using the central carbon metabolism model (Chapter 3).

Chromosome A1 marks the upper bound of serine flux ratio and it is also the optimal point for the single objective maximization of serine synthesis rate. Chromosome A2 is actively constrained by the total enzymatic flux requirement. The Pareto segment will extend uninterrupted to the right of chromosome A2 if not for the total enzymatic flux constraint. Chromosome A3 is obtained from simulating simultaneous manipulation of SerSynth (by 1.5 times) and knockout of G1PAT. Chromosome A2 and its several adjacent chromosomes are dominated by chromosome A3. The number of dominated chromosomes depends on the seeds – four chromosomes for seed 0.1, two chromosomes for seed 0.6 and five chromosomes for seed 0.7. Chromosome A3 will be dominated due to simultaneous decline in serine and tryptophan flux ratios if a non-zero multiplier is applied to gene encoding G1PAT while maintaining the multiplier for SerSynth at 1.5. The domination of A2 and its adjacent chromosomes is caused by the inability of NSGA-II to generate chromosome A3 and the proximity of the numerical values. It is likely that other chromosomes are rejected once sufficient number of lower (better) rank chromosomes are obtained through non-dominated sorting leaving out the chromosomes in the less crowded region. This happens when the emerging region which eventually becomes the left most Pareto segment is so densely populated with lower (better) rank chromosomes that no other solution can be accommodated within a fixed population size. Effectively, NSGA-II simply rejects the remainder of the solutions including the short Pareto segment bounded by chromosomes B1 and B2.

In the optimization runs using three seeds (0.1, 0.6 and 0.7), the skeletal Pareto segment bounded by A1 and A2 was constructed within the first 10 generations. The sets of genes encoding GAPDH/TrpSynth, G1PAT/SerSynth and G1PAT/TrpSynth easily breach the total enzymatic flux constraint during the first three generations due to various reasons such as repression of either one of the genes encoding GAPDH and TrpSynth, and simultaneous repression of genes encoding G1PAT and TrpSynth. The available slots are quickly filled through non-dominated sorting by genes encoding GAPDH and SerSynth after the third generation. The elitist preservation mechanism, which precedes the crowding distance sorting, rapidly increases the density of better

rank chromosomes that fulfilled the homeostatic and total enzymatic flux constraints to form the offspring population. In NSGA-II procedure, each parent and offspring population is of size N, for instance. The combined parent and offspring population size is 2N. Chromosomes of non-dominated rank 1 (best) fill the first slot. Chromosomes of non-dominated rank 2 (next to best) fill the second slot. This continues until the last slot results in a population size greater than or equal to N. No other chromosomes can be accommodated beyond the last slot. Following this, the crowding distance sorting is applied to the chromosomes which belong to the last slot where a solution that is located in a lesser crowded region is preferred. This creates a new population of size N. The new population (which is now the parent) is used for selection, crossover, and mutation to create an offspring of size N. After the third generation, the combined parent and offspring population consists entirely of local neighbouring chromosomes and the crowding distance sorting mechanism to increase diversity of the search is rendered ineffective. For the narrow multiplier range of 0.8 to 1.25, the probability of breaching the total enzymatic flux constraint is lower during the first few generations. It is likely that the density of the local Pareto increases rapidly during the first few generations when manipulating genes encoding the yet to be developed isolated chromosome G (Figure 5.2a) results in breaching of the total enzymatic flux constraint.

Extension of the Pareto segment bounded by chromosomes B1 and B2 towards the left results in domination. The same Pareto segment will extend uninterrupted after chromosome B2 if not for the total enzymatic flux constraint. Chromosome B3 is obtained by simulating simultaneous manipulation of TrpSynth (multiplied 1.5 times) and knockout of G1PAT. This Pareto point also maximizes the TrpSynth flux ratio for the single objective. Chromosome B3 will be dominated due to simultaneous decline in TrpSynth and SerSynth flux ratios if the G1PAT multiplier is non-zero while maintaining the TrpSynth multiplier at 1.5.

5.4.2 Flux Distribution and Tryptophan Operon Control

The key flux ratios and concentrations (extracted from Tables 5.1 to 5.6) show several (Figures 5.5 and 5.6) interesting patterns. In generating the leftmost Pareto branch (Figure 5.4a), greater serine flux ratio is obtained by reducing the flux ratios of PGM and G1PAT (Figure 5.5) to 0.6566-0.8282 and 0.6538-0.8270 respectively.

High serine concentration (Figure 5.6) of chromosome A1 at 0.07477 mM (SerSynth flux ratio = 1.5346) is concomitant with slightly lowered tryptophan concentration at 0.003950 mM (TrpSynth flux ratio = 0.9657; DAHPS flux ratio = 0.6178; ChoSynth flux ratio = 0.6178). This results in slight relaxation of tryptophan operon repression with a small increase in the free DNA operator sites and free mRNA concentration. As serine concentration of chromosome A2 (Figure 5.6) slides to 0.07195 mM (SerSynth flux ratio = 1.4767) and tryptophan concentration rises to 0.004498 mM (TrpSynth flux ratio = 1.0996; DAHPS flux ratio =0.8192; ChoSynth flux ratio = 0.8192), the applied repression of tryptophan operon results in a small decrease in the free DNA operator sites and free mRNA concentration. The pooled enzyme concentration decreases due to feedback inhibition as tryptophan concentration increases. Since e4p concentration is lower than its initial steady-state value and the opposite is true of pep concentration, it implies that e4p (a central carbon metabolite, refer to section 5.3.1) is the first limiting precursor for DAHPS enzymatic sub-system (Liao et al., 1996) followed by pep.

High serine concentration (Figure 5.6) of chromosome A3 at 0.07204 mM (SerSynth flux ratio = 1.4786) is needed to meet and sustain the demand (Section 5.3.2) from tryptophan synthesis (TrpSynth flux ratio = 1.1901; DAHPS flux ratio = 0.9474; ChoSynth flux ratio = 0.9474). Otherwise, overflow of indole (a last precursor in the final step of the terminal tryptophan biosynthesis) due to severe shortage of serine results in decline of glucose consumption, loss of plasmids and microbial cell death (Ikeda et al., 1994) in the last stage of fermentation using *C. glutamicum* where the common aromatic amino acid and terminal tryptophan biosynthesis pathways are very similar to those in *E. coli*. The flux ratios of DAHPS and ChoSynth being less than 1.0 for chromosome A3 suggest that the precursors dahp and chorismate (cho) in the common aromatic amino acid pathway are not retarding the production of tryptophan. While GAPDH flux ratios are largely resilient to its gene amplification, GAPDH acts as a significant conduit primarily for the glycolytic pathway with influential cascading effect on the fluxes of the pentose-phosphate and aromatic amino acid pathways.

The serine concentration for the Pareto segment bounded by chromosomes B1 and B2 is maintained above initial steady state value when it is supported by serine synthesis fluxes which are between 1.5% and 2% higher than the initial steady state flux. This ensures a steady availability of serine as a precursor together with indole for

the final conversion to tryptophan. The flux ratios of PGI are relatively well regulated at between 3.5% (Table 5.5) and 4.5% (Table 5.4) below the initial steady state flux. Cellular growth inhibition discussed earlier (Sections 5.3.2 and 5.4.1) is avoided. The slight increases in the fluxes of the pentose-phosphate pathway (Tables 5.4 and 5.5) are desirable for enhancing tryptophan production (Ikeda and Katsumata, 1999) without affecting the stability of entire metabolic network. The concentration of e4p is relatively constant at 0.0978 mM which is slightly below that of the initial steady state. The flux ratios of DAHPS and ChoSynth increase gradually from 0.9142 to 0.9342 as tryptophan biosynthesis rate rises in moving from chromosome B1 to chromosome B2.

The maximum tryptophan flux ratio of 1.3052 exhibited by chromosome B3 (Figure 5.5 and Table 5.6) is obtained through very moderate amplification of the fluxes in the pentose-phosphate pathway of less than 2% (e4p and pep concentrations are moderately above those of initial steady state), maintaining SerSynth flux ratio slightly above 1.0 and increased DAHPS and ChoSynth flux ratios at 1.0946. The dahp, cho and serine concentrations are higher than the ones at initial steady state. The net result is a continual non-rate limiting supply of carbon to the pathways leading to tryptophan biosynthesis. Since the tryptophan concentration is highest at 0.005339 mM, repression of the tryptophan operon and feedback inhibition (Table 5.6) of the pooled enzyme are greatest.

5.5 Conclusions

The heuristics method uses pattern recognition and multiple random number seeds to identify genes for the concurrent manipulation and knockout, and alleviate the daunting numerical difficulties encountered in the optimization study using the complex augmented model. NSGA-II alone may not always generate the global Pareto or the complete Pareto-optimal front. There are opportunities to combine other means with NSGA-II or modify NSGA-II to improve its efficacy. In the concurrent two-gene knockout and manipulation, genes encoding SerSynth, TrpSynth, GAPDH and G1PAT have been identified using the narrow multiplier range of 0.8-1.25. Existence of local Pareto was noted. In the next stage, the wide multiplier range of 0-1.5 creates one part of the Pareto-optimal front when genes encoding GAPDH and SerSynth were manipulated. Genes that cannot be manipulated desirably using a

narrow multiplier range are not likely candidates for concurrent gene knockouts and manipulations. The probability for sub-optimality in the detailed optimization using a wider multiplier range is remote after initial genes screening using a narrow multiplier range. The maximum SerSynth flux ratio of 1.5346 is obtained when genes encoding GAPDH and SerSynth are amplified 1.5 times. The remainder of the Pareto consisting of two isolated chromosomes due to one gene knockout and another gene manipulation, and a Pareto branch for high tryptophan biosynthesis rate range could not be found using NSGA-II. It is obtained through simulations following the Pareto patterns recognized during the gene identification stage. Missing Pareto solutions are possibly due to the sorting procedure and elitist preservation mechanism of NSGA-II, and the probability of breaching homeostatic and total enzymatic flux constraints during the first few generations. The maximum TrpSynth flux ratio of 1.3052 is obtained when genes encoding GAPDH and TrpSynth are amplified 1.5 times. The results obtained by optimization are consistent with the fermentation studies in the literature and with the dynamic behaviour of the tryptophan operon. In the concurrent three-gene knockout and manipulation, genes encoding SerSynth, TrpSynth, GAPDH, RPPK, G1PAT, PK and G6PDH have been identified using a narrow multiplier range of 0.8-1.25. The maximum serine and tryptophan flux ratios are 1.3000 and 1.2145 respectively. The various parts of Pareto solutions have been explained in terms of collaborative and competitive supply and demand among the contributing fluxes. The triple-gene manipulation could not be extended to the wide multiplier range of 0-1.5 due to the existence of large number of infeasible solutions leading to severe numerical computing difficulties.

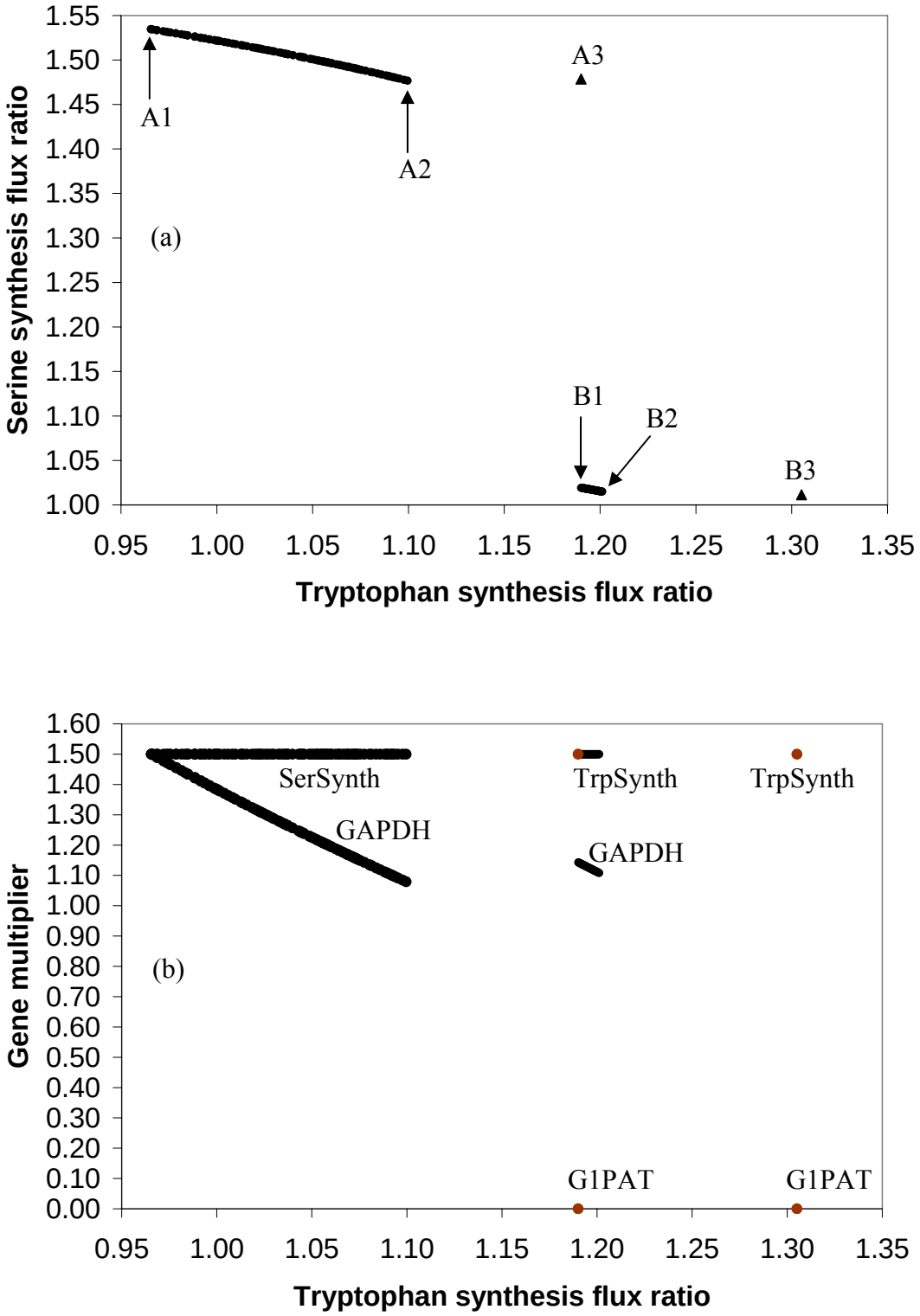


Fig. 5.4. (a) Pareto-optimal front obtained by concurrent two-gene manipulation and knockouts with multiplier range of 0.0 to1.25.
(b) Optimal gene multipliers for the Pareto front in (a).

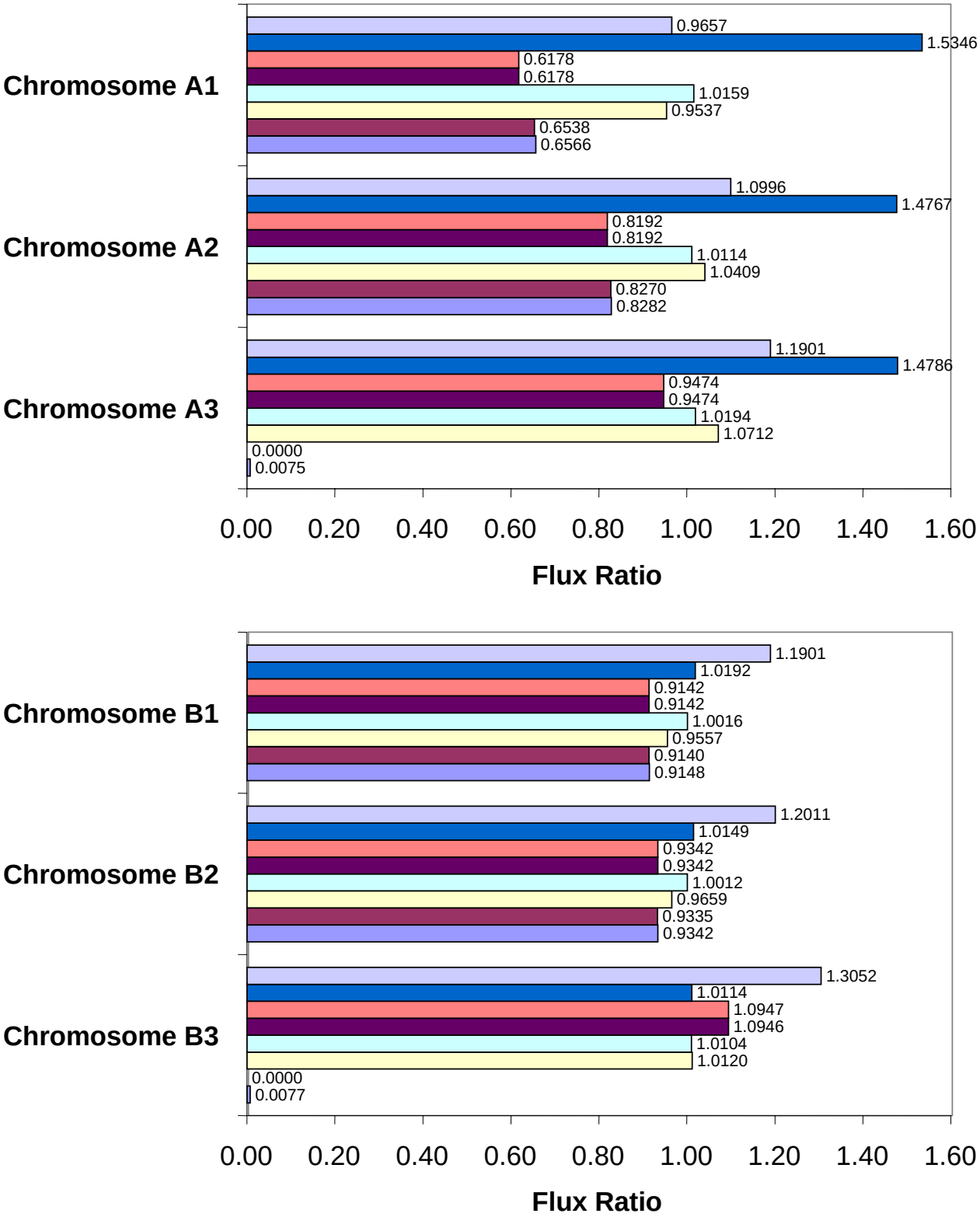
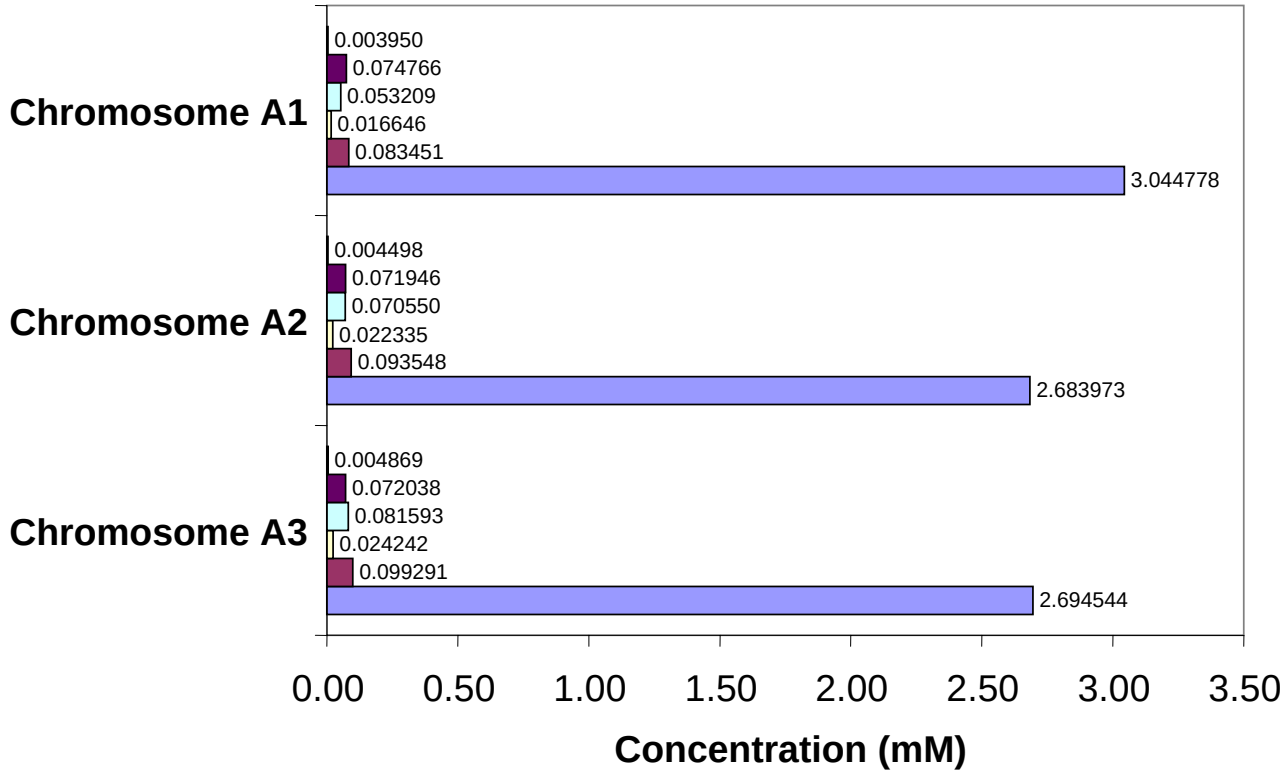


Fig. 5.5. Flux ratios of the chromosomes depicted in Figure 5.4. The flux ratios indicated for each chromosome, in descending order of bar position; correspond to TrpSynth, SerSynth, ChoSynth, DAHPS, GAPDH, PGI, G1PAT and PGM. Refer to List of Symbols for the definitions of the abbreviations. Refer to Tables 5.1 to 5.6 for complete flux data.



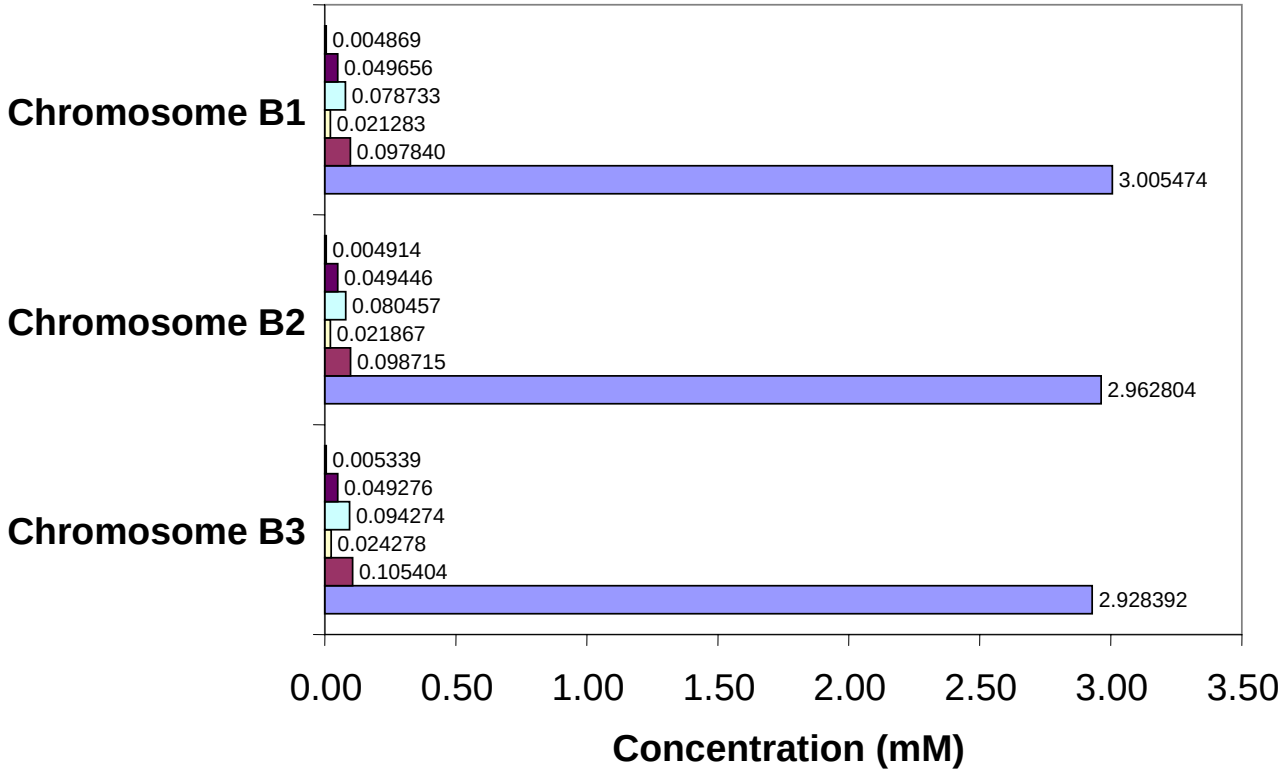


Fig. 5.6. Concentrations of the chromosomes depicted in Figure 5.4. The concentrations indicated for each chromosome, in descending order of bar position; correspond to trp, ser, cho, dahp, e4p and pep. Refer to List of Symbols for the definitions of the abbreviations. Refer to Tables 5.1 to 5.6 for complete concentration data.

Table 5.1 Concentrations and fluxes of chromosome A1 in Fig. 5.4. The initial steady-state values of the augmented model in Table 4.3 are in brackets.

Metabolite/Co-metabolite with serial number	Concentration (mM)	Enzyme with serial number	Flux ratio (flux, mM/s)
Metabolite		1. PTS	1.0000 (0.2000)
1. Glucose (extracellular)	0.05796 (0.05612)	2. PGI	0.9537 (0.05584)
2. g6p	3.6531 (3.5447)	3. PFK	1.0054 (0.1398)
3. f6p	0.6299 (0.6112)	4. ALDO	1.0055 (0.1398)
4. fdp	0.1664 (0.2945)	5. TIS	1.0084 (0.1381)
5. gap	0.1666 (0.2276)	6. GAPDH	1.0159 (0.3166)
6. dhap	0.1290 (0.1740)	7. PGK	1.0159 (0.3166)
7. pgp	0.0091(0.00845)	8. PGluMu	0.9850 (0.2988)
8. 3pg	2.4293 (2.2513)	9. ENO	0.9850 (0.2988)
9. 2pg	0.4549 (0.4216)	10. PK	1.0073 (0.03833)
10. pep	3.0448 (2.8212)	11. PDH	1.0015 (0.1870)
11. pyr	2.6668 (2.6657)	12. PEPCxylase	1.0211 (0.04496)
12. 6pg	0.8474 (0.8268)	13. PGM	0.6566 (0.002533)
13. ribu5p	0.1071 (0.1136)	14. G1PAT	0.6538 (0.002515)
14. xyl5p	0.1316 (0.1414)	15. RPPK	0.9866 (0.01036)
15. sed7p	0.3260 (0.2783)	16. G3PDH	0.7708 (0.001722)
16. rib5p	0.3814 (0.4077)	17. SerSynth	1.5346^{+} (0.01780)
17. e4p	0.08345 (0.1015)	18. Synth1	Not applicable (0.01443)
18. g1p	0.6936 (0.6622)	19. Synth2	1.0001 (0.05354)
19. pooled enzyme	0.00206 (0.00200)	20. DAHPS	0.6178 (0.007495)
20. dahp	0.01665 (0.0239)	21. G6PDH	1.0244 (0.1415)
21. cho	0.0532 (0.08612)	22. PGDH	1.0244 (0.1415)
22. trp	0.003950 (0.004091)	23. Ru5P	1.0394 (0.08490)
23. ser	0.07477 (0.04872)	24. R5PI	1.0018 (0.05657)
24. prpp	0.0009952 (0.0010087)	25. TKa	1.0052 (0.04620)
		26. TKb	1.0802 (0.03870)
Co-metabolite		27. TA	1.0052 (0.04620)
1. amp	0.955 (0.955)	28. MurSynth	1.0000 (0.00043711)
2. adp	0.595 (0.595)	29. MetSynth	1.0000 (0.0022627)
3. atp	4.27 (4.27)	30. TrpSynth	0.9657^{+} (0.000000535)
4. nadp	0.195 (0.195)	31. TrpConsumed	0.9657 (0.000000421364)
5. nadph	0.062 (0.062)	32. EnzSynth	1.0323 (0.000000055562)
6. nad	1.47 (1.47)	33. EnzDegraded	Not applicable (0.0)
7. nadh	0.1 (0.1)	34. ChoSynth	0.6178^{+} (0.007494)
		35. ChoConsumed	0.6177 (0.007491)
Tryptophan operon		36. PrppConsumed	0.9866 (0.01036)
Molecule or binding site with serial number	Concentration (μM)	37. SerConsumed	1.5346 (0.01780)
1. Total repressor	0.8509 (0.8527)		+ The chorismate, serine and tryptophan synthesis rates are greater than their respective internal consumption rates.
2. Active repressor	0.05228 (0.05414)		
3. Free DNA operator site	0.0001547 (0.0001499)		
4. Free mRNA	0.0003446 (0.0003338)		

Table 5.2 Concentrations and fluxes of chromosome A2 depicted in Fig. 5.4. The initial steady-state values of the augmented model in Table 4.3 are in brackets.

Metabolite/Co-metabolite with serial number	Concentration (mM)	Enzyme with serial number	Flux ratio (flux, mM/s)
Metabolite		1. PTS	1.0000 (0.2000)
1. Glucose (extracellular)	0.05561 (0.05612)	2. PGI	1.0409 (0.05584)
2. g6p	3.4873 (3.5447)	3. PFK	1.0112 (0.1398)
3. f6p	0.6013 (0.6112)	4. ALDO	1.0112 (0.1398)
4. fdp	0.2385 (0.2945)	5. TIS	1.0125 (0.1381)
5. gap	0.2030 (0.2276)	6. GAPDH	1.0114 (0.3166)
6. dhap	0.1559 (0.1740)	7. PGK	1.0114 (0.3166)
7. pgp	0.00804 (0.00845)	8. PGluMu	0.9837 (0.2988)
8. 3pg	2.1415 (2.2513)	9. ENO	0.9837 (0.2988)
9. 2pg	0.4010 (0.4216)	10. PK	0.9949 (0.03833)
10. pep	3.6840 (2.8212)	11. PDH	0.9990 (0.1870)
11. pyr	2.6650 (2.6657)	12. PEPCxylase	0.9562 (0.04496)
12. 6pg	0.8158 (0.8268)	13. PGM	0.8282 (0.002533)
13. ribu5p	0.1091 (0.1136)	14. G1PAT	0.8270 (0.002515)
14. xyl5p	0.1353 (0.1414)	15. RPPK	0.9915 (0.01036)
15. sed7p	0.2843 (0.2783)	16. G3PDH	0.9102 (0.001722)
16. rib5p	0.3907 (0.4077)	17. SerSynth	1.4767^{+} (0.01780)
17. e4p	0.0935 (0.1015)	18. Synth1	Not applicable (0.01443)
18. g1p	0.6567 (0.6622)	19. Synth2	0.9999 (0.05354)
19. pooled enzyme	0.00183 (0.00200)	20. DAHPS	0.8192 (0.007495)
20. dahp	0.0223 (0.0239)	21. G6PDH	0.9869 (0.1415)
21. cho	0.07055 (0.08612)	22. PGDH	0.9869 (0.1415)
22. trp	0.004498 (0.004091)	23. Ru5P	0.9915 (0.08490)
23. ser	0.07195 (0.04872)	24. R5PI	0.9801 (0.05657)
24. prpp	0.0010002 (0.0010087)	25. TKa	0.9775 (0.04620)
		26. TKb	1.0082 (0.03870)
Co-metabolite		27. TA	0.9775 (0.04620)
1. amp	0.955 (0.955)	28. MurSynth	1.0000 (0.00043711)
2. adp	0.595 (0.595)	29. MetSynth	1.0000 (0.0022627)
3. atp	4.27 (4.27)	30. TrpSynth	1.0996^{+} (0.000000535)
4. nadp	0.195 (0.195)	31. TrpConsumed	1.0996 (0.000000421364)
5. nadph	0.062 (0.062)	32. EnzSynth	0.9168 (0.000000055562)
6. nad	1.47 (1.47)	33. EnzDegraded	Not applicable (0.0)
7. nadh	0.1 (0.1)	34. ChoSynth	0.8192 (0.007494)
		35. ChoConsumed	0.8192 (0.007491)
Tryptophan operon		36. PrppConsumed	0.9915 (0.01036)
Molecule or binding site with serial number	Concentration (μM)	37. SerConsumed	1.4766 (0.01780)
1. Total repressor	0.8581 (0.8527)		+ The chorismate, serine and tryptophan synthesis rates are greater than their respective internal consumption rates.
2. Active repressor	0.05953 (0.05414)		
3. Free DNA operator site	0.0001374 (0.0001499)		
4. Free mRNA	0.0003060 (0.0003338)		

Table 5.3 Concentrations and fluxes of chromosome A3 depicted in Fig. 5.4. The initial steady-state values of the augmented model in Table 4.3 are in brackets.

Metabolite/Co-metabolite with serial number	Concentration (mM)	Enzyme with serial number	Flux ratio (flux, mM/s)
Metabolite		1. PTS	1.0000 (0.2000)
1. Glucose (extracellular)	0.05608 (0.05612)	2. PGI	1.0712 (0.05584)
2. g6p	3.4992 (3.5447)	3. PFK	1.0226 (0.1398)
3. f6p	0.6033 (0.6112)	4. ALDO	1.0226 (0.1398)
4. fdp	0.2811 (0.2945)	5. TIS	1.0231 (0.1381)
5. gap	0.2217 (0.2276)	6. GAPDH	1.0194 (0.3166)
6. dhap	0.1698 (0.1740)	7. PGK	1.0194 (0.3166)
7. pgp	0.00807 (0.00845)	8. PGluMu	0.9920 (0.2988)
8. 3pg	2.1501 (2.2513)	9. ENO	0.9920 (0.2988)
9. 2pg	0.4026 (0.4216)	10. PK	0.9954 (0.03833)
10. pep	2.6945 (2.8212)	11. PDH	0.9991 (0.1870)
11. pyr	2.6650 (2.6657)	12. PEPCxylase	0.9686 (0.04496)
12. 6pg	0.8181 (0.8268)	13. PGM	0.0075 (0.002533)
13. ribu5p	0.1119 (0.1136)	14. G1PAT	0.0 (0.002515)
14. xyl5p	0.1392 (0.1414)	15. RPPK	0.9969 (0.01036)
15. sed7p	0.2763 (0.2783)	16. G3PDH	0.9796 (0.001722)
16. rib5p	0.4013 (0.4077)	17. SerSynth	1.4786^{+} (0.01780)
17. e4p	0.0993 (0.1015)	18. Synth1	Not applicable (0.01443)
18. g1p	0.6856 (0.6622)	19. Synth2	0.9999 (0.05354)
19. pooled enzyme	0.00170 (0.00200)	20. DAHPS	0.9474 (0.007495)
20. dahp	0.0242 (0.0239)	21. G6PDH	0.9897 (0.1415)
21. cho	0.08159 (0.08612)	22. PGDH	0.9897 (0.1415)
22. trp	0.004869 (0.004091)	23. Ru5P	0.9903 (0.08490)
23. ser	0.07204 (0.04872)	24. R5PI	0.9887 (0.05657)
24. prpp	0.0010056 (0.0010087)	25. TKa	0.9868 (0.04620)
		26. TKb	0.9945 (0.03870)
Co-metabolite		27. TA	0.9868 (0.04620)
1. amp	0.955 (0.955)	28. MurSynth	1.0000 (0.00043711)
2. adp	0.595 (0.595)	29. MetSynth	1.0000 (0.0022627)
3. atp	4.27 (4.27)	30. TrpSynth	1.1901^{+} (0.000000535)
4. nadp	0.195 (0.195)	31. TrpConsumed	1.1901 (0.000000421364)
5. nadph	0.062 (0.062)	32. EnzSynth	0.8522 (0.000000055562)
6. nad	1.47 (1.47)	33. EnzDegraded	Not applicable (0.0)
7. nadh	0.1 (0.1)	34. ChoSynth	0.9474 (0.007494)
		35. ChoConsumed	0.9474 (0.007491)
Tryptophan operon		36. PrppConsumed	0.9968 (0.01036)
Molecule or binding site with serial number	Concentration (μM)	37. SerConsumed	1.4786 (0.01780)
			+ The chorismate, serine and tryptophan synthesis rates are greater than their respective internal consumption rates.
1. Total repressor	0.8630 (0.8527)		
2. Active repressor	0.06443 (0.05414)		
3. Free DNA operator site	0.0001277 (0.0001499)		
4. Free mRNA	0.0002845 (0.0003338)		

Table 5.4 Concentrations and fluxes of chromosome B1 depicted in Fig. 5.4. The initial steady-state values of the augmented model in Table 4.3 are in brackets.

Metabolite/Co-metabolite with serial number	Concentration (mM)	Enzyme with serial number	Flux ratio (flux, mM/s)
Metabolite		1. PTS	1.0000 (0.2000)
1. Glucose (extracellular)	0.05734 (0.05612)	2. PGI	0.9557 (0.05584)
2. g6p	3.6290 (3.5447)	3. PFK	0.9967 (0.1398)
3. f6p	0.6257 (0.6112)	4. ALDO	0.9967 (0.1398)
4. fdp	0.2591 (0.2945)	5. TIS	0.9973 (0.1381)
5. gap	0.2126 (0.2276)	6. GAPDH	1.0016 (0.3166)
6. dhap	0.1628 (0.1740)	7. PGK	1.0016 (0.3166)
7. pgp	0.00900 (0.00845)	8. PGluMu	1.0005 (0.2988)
8. 3pg	2.3983 (2.2513)	9. ENO	1.0005 (0.2988)
9. 2pg	0.4491 (0.4216)	10. PK	1.0061 (0.03833)
10. pep	3.0055 (2.8212)	11. PDH	1.0012 (0.1870)
11. pyr	2.6666 (2.6657)	12. PEPCxylase	1.0266 (0.04496)
12. 6pg	0.8429 (0.8268)	13. PGM	0.9148 (0.002533)
13. ribu5p	0.1134 (0.1136)	14. G1PAT	0.9140 (0.002515)
14. xyl5p	0.1407 (0.1414)	15. RPPK	0.9992 (0.01036)
15. sed7p	0.2946 (0.2783)	16. G3PDH	0.9448 (0.001722)
16. rib5p	0.4062 (0.4077)	17. SerSynth	1.0192^{+} (0.01780)
17. e4p	0.0978 (0.1015)	18. Synth1	Not applicable (0.01443)
18. g1p	0.6807 (0.6622)	19. Synth2	1.0000 (0.05354)
19. pooled enzyme	0.00170 (0.00200)	20. DAHPS	0.9142 (0.007495)
20. dahp	0.0213 (0.0239)	21. G6PDH	1.0190 (0.1415)
21. cho	0.07873 (0.08612)	22. PGDH	1.0190 (0.1415)
22. trp	0.004869 (0.004091)	23. Ru5P	1.0237 (0.08490)
23. ser	0.04966 (0.04872)	24. R5PI	1.0119 (0.05657)
24. prpp	0.0010080 (0.0010087)	25. TKa	1.0148 (0.04620)
		26. TKb	1.0343 (0.03870)
Co-metabolite		27. TA	1.0148 (0.04620)
1. amp	0.955 (0.955)	28. MurSynth	1.0000 (0.00043711)
2. adp	0.595 (0.595)	29. MetSynth	1.0000 (0.0022627)
3. atp	4.27 (4.27)	30. TrpSynth	1.1901^{+} (0.000000535)
4. nadp	0.195 (0.195)	31. TrpConsumed	1.1901 (0.000000421364)
5. nadph	0.062 (0.062)	32. EnzSynth	0.8522 (0.000000055562)
6. nad	1.47 (1.47)	33. EnzDegraded	Not applicable (0.0)
7. nadh	0.1 (0.1)	34. ChoSynth	0.9142 (0.007494)
		35. ChoConsumed	0.9142 (0.007491)
Tryptophan operon		36. PrppConsumed	0.9992 (0.01036)
Molecule or binding site with serial number	Concentration (μM)	37. SerConsumed	1.0192 (0.01780)
			+ The chorismate, serine and tryptophan synthesis rates are greater than their respective internal consumption rates.
1. Total repressor	0.8630 (0.8527)		
2. Active repressor	0.06443 (0.05414)		
3. Free DNA operator site	0.0001277 (0.0001499)		
4. Free mRNA	0.0002845 (0.0003338)		

Table 5.5 Concentrations and fluxes of chromosome B2 depicted in Fig. 5.4. The initial steady-state values of the augmented model in Table 4.3 are in brackets.

Metabolite/Co-metabolite with serial number	Concentration (mM)	Enzyme with serial number	Flux ratio (flux, mM/s)
Metabolite		1. PTS	1.0000 (0.2000)
1. Glucose (extracellular)	0.05704 (0.05612)	2. PGI	0.9659 (0.05584)
2. g6p	3.6095 (3.5447)	3. PFK	0.9974 (0.1398)
3. f6p	0.6224 (0.6112)	4. ALDO	0.9974 (0.1398)
4. fdp	0.2671 (0.2945)	5. TIS	0.9979 (0.1381)
5. gap	0.2161 (0.2276)	6. GAPDH	1.0012 (0.3166)
6. dhap	0.1654 (0.1740)	7. PGK	1.0012 (0.3166)
7. pgp	0.00887 (0.00845)	8. PGluMu	1.0004 (0.2988)
8. 3pg	2.3643 (2.2513)	9. ENO	1.0004 (0.2988)
9. 2pg	0.4427 (0.4216)	10. PK	1.0048 (0.03833)
10. pep	2.9628 (2.8212)	11. PDH	1.0009 (0.1870)
11. pyr	2.6664 (2.6657)	12. PEPCxylase	1.0202 (0.04496)
12. 6pg	0.8392 (0.8268)	13. PGM	0.9342 (0.002533)
13. ribu5p	0.1134 (0.1136)	14. G1PAT	0.9335 (0.002515)
14. xyl5p	0.1409 (0.1414)	15. RPPK	0.9994 (0.01036)
15. sed7p	0.2907 (0.2783)	16. G3PDH	0.9577 (0.001722)
16. rib5p	0.4066 (0.4077)	17. SerSynth	1.0149^{+} (0.01780)
17. e4p	0.0987 (0.1015)	18. Synth1	Not applicable (0.01443)
18. g1p	0.6764 (0.6622)	19. Synth2	1.0001 (0.05354)
19. pooled enzyme	0.00169 (0.00200)	20. DAHPS	0.9342 (0.007495)
20. dahp	0.0219 (0.0239)	21. G6PDH	1.0146 (0.1415)
21. cho	0.08046 (0.08612)	22. PGDH	1.0146 (0.1415)
22. trp	0.004914 (0.004091)	23. Ru5P	1.0182 (0.08490)
23. ser	0.04945 (0.04872)	24. R5PI	1.0092 (0.05657)
24. prpp	0.0010082 (0.0010087)	25. TKa	1.0114 (0.04620)
		26. TKb	1.0263 (0.03870)
Co-metabolite		27. TA	1.0114 (0.04620)
1. amp	0.955 (0.955)	28. MurSynth	1.0000 (0.00043711)
2. adp	0.595 (0.595)	29. MetSynth	1.0000 (0.0022627)
3. atp	4.27 (4.27)	30. TrpSynth	1.2011^{+} (0.000000535)
4. nadp	0.195 (0.195)	31. TrpConsumed	1.2011 (0.000000421364)
5. nadph	0.062 (0.062)	32. EnzSynth	0.8450 (0.000000055562)
6. nad	1.47 (1.47)	33. EnzDegraded	Not applicable (0.0)
7. nadh	0.1 (0.1)	34. ChoSynth	0.9342 (0.007494)
		35. ChoConsumed	0.9342 (0.007491)
Tryptophan operon		36. PrppConsumed	0.9994 (0.01036)
Molecule or binding site with serial number	Concentration (μM)	37. SerConsumed	1.0149 (0.01780)
1. Total repressor	0.8636 (0.8527)		+ The chorismate, serine and tryptophan synthesis rates are greater than their respective internal consumption rates.
2. Active repressor	0.06503 (0.05414)		
3. Free DNA operator site	0.0001266 (0.0001499)		
4. Free mRNA	0.0002821 (0.0003338)		

Table 5.6 Concentrations and fluxes of chromosome B3 depicted in Fig. 5.4. The initial steady-state values of the augmented model in Table 4.3 are in brackets.

Metabolite/Co-metabolite with serial number	Concentration (mM)	Enzyme with serial number	Flux ratio (flux, mM/s)
Metabolite		1. PTS	1.0000 (0.2000)
1. Glucose (extracellular)	0.05731 (0.05612)	2. PGI	1.0120 (0.05584)
2. g6p	3.6024 (3.5447)	3. PFK	1.0116 (0.1398)
3. f6p	0.6211 (0.6112)	4. ALDO	1.0116 (0.1398)
4. fdp	0.3225 (0.2945)	5. TIS	1.0113 (0.1381)
5. gap	0.2388 (0.2276)	6. GAPDH	1.0104 (0.3166)
6. dhap	0.1824 (0.1740)	7. PGK	1.0104 (0.3166)
7. pgp	0.00877 (0.00845)	8. PGluMu	1.0104 (0.2988)
8. 3pg	2.3371 (2.2513)	9. ENO	1.0104 (0.2988)
9. 2pg	0.4376 (0.4216)	10. PK	1.0037 (0.03833)
10. pep	2.9284 (2.8212)	11. PDH	1.0007 (0.1870)
11. pyr	2.6663 (2.6657)	12. PEPCxylase	1.0342 (0.04496)
12. 6pg	0.8378 (0.8268)	13. PGM	0.0077 (0.002533)
13. ribu5p	0.1163 (0.1136)	14. G1PAT	0.0 (0.002515)
14. xyl5p	0.1449 (0.1414)	15. RPPK	1.0047 (0.01036)
15. sed7p	0.2793 (0.2783)	16. G3PDH	1.0406 (0.001722)
16. rib5p	0.4176 (0.4077)	17. SerSynth	1.0114^{+} (0.01780)
17. e4p	0.1054 (0.1015)	18. Synth1	Not applicable (0.01443)
18. g1p	0.7058 (0.6622)	19. Synth2	1.0001 (0.05354)
19. pooled enzyme	0.00156 (0.00200)	20. DAHPS	1.0946 (0.007495)
20. dahp	0.0243 (0.0239)	21. G6PDH	1.0130 (0.1415)
21. cho	0.09427 (0.08612)	22. PGDH	1.0130 (0.1415)
22. trp	0.005339 (0.004091)	23. Ru5P	1.0113 (0.08490)
23. ser	0.04928 (0.04872)	24. R5PI	1.0156 (0.05657)
24. prpp	0.0010134 (0.0010087)	25. TKa	1.0181 (0.04620)
		26. TKb	1.0032 (0.03870)
Co-metabolite		27. TA	1.0181 (0.04620)
1. amp	0.955 (0.955)	28. MurSynth	1.0000 (0.00043711)
2. adp	0.595 (0.595)	29. MetSynth	1.0000 (0.0022627)
3. atp	4.27 (4.27)	30. TrpSynth	1.3052^{+} (0.000000535)
4. nadp	0.195 (0.195)	31. TrpConsumed	1.3052 (0.000000421364)
5. nadph	0.062 (0.062)	32. EnzSynth	0.7823 (0.000000055562)
6. nad	1.47 (1.47)	33. EnzDegraded	Not applicable (0.0)
7. nadh	0.1 (0.1)	34. ChoSynth	1.0947 (0.007494)
		35. ChoConsumed	1.0947 (0.007491)
Tryptophan operon		36. PrppConsumed	1.0047 (0.01036)
Molecule or binding site with serial number	Concentration (μM)	37. SerConsumed	1.0114 (0.01780)
1. Total repressor	0.8693 (0.8527)		+ The chorismate, serine and tryptophan synthesis rates are greater than their respective internal consumption rates.
2. Active repressor	0.07066 (0.05414)		
3. Free DNA operator site	0.0001172 (0.0001499)		
4. Free mRNA	0.0002611 (0.0003338)		

Chapter 6
CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

The present research has been driven by the scarcity of published works on the MOO of bioprocesses particularly in biopharmaceuticals and microbial cell factories, with the intention of developing multiple optimal solutions for penicillin V bioreactor train and metabolic pathway recipe in microbial cell factories. Of critical importance to the present research is the availability of models validated through experiments, and published or industrial fermentation results to check the Pareto-optimal results obtained in the current research.

Four *Penicillium chrysogenum* fermentation models since 1994 were considered and one from Zangirolami et al. (1997) was finally selected based on published industrial information and academic work. Optimization of an industrial penicillin V bioreactor train with a targeted continuous production rate of 248 kg/h (Biwer et al., 2004) was successfully obtained using NSGA-II, for both bi- and tri-objective scenarios using six decision variables. In the bi-objective maximization of yield and penicillin concentration, higher penicillin concentration was obtainable by stretching the batch cycle time at the expense of lower yield due to higher glucose and nutrients consumption. In the concurrent minimization of batch cycle time and maximization of yield, batch cycle time and yield vary in the same direction before biomass degradation sets in. In maximizing yield, penicillin concentration and profit simultaneously, higher profit was obtained from higher yield and lower penicillin concentration.

Bi-objective and tri-objective optimization results were very similar for penicillin V bioreactor; hence, bi-objective optimization may be adequate for this application. Glucose feed concentration has the most effect on the simultaneous maximization of yield and penicillin concentration, and also in the tri-objective case. Multiple optimal values of the decision variables were obtained for the same Pareto, which was also seen in chemical engineering applications (Tarafder et al., 2007). When two of the six decision variables – switchover time from batch mode to continuous glucose feeding and initial glucose concentration − were fixed, only one

set of optimal decision variables was obtained. This showed the interdependence of certain decision variables in contributing to the Pareto.

Optimization of fluxes of desired enzymatic sub-systems in the central carbon metabolism of *E. coli* (Chassagnole et al., 2002), a microbial cell factory was performed for two cases − maximization of DAHPS and PEPCxylase fluxes, and maximization of DAHPS and SerSynth fluxes, by 1-, 2- and 3-gene knockout or manipulation. Deleting PGM gene gives the 1-enzyme Pareto-optimal set for both bi-objective cases. In paired and triple enzyme knockout or manipulation, G6PDH diverts fluxes to the desired metabolic pathways. Pareto sets obtained in 1-, 2- and 3-gene knockout using a manual exhaustive search were identical to those obtained via an interactive branch-and-bound technique facilitated by NSGA-II. The triple enzyme knockout/manipulation gives the best Pareto-optimal set. Triple enzyme manipulation enhances the DAHPS/PEPCxylase and DAHPS/SerSynth fluxes more than the triple gene knockout. Except for SerSynth flux which has a more immediate impact on serine biosynthesis rate, the effects of DAHPS and PEPCxylase fluxes on the respective biosynthesis rates of aromatic amino acids (tryptophan, phenylalanine and tyrosine) and aspartate-threonine pathway products (lysine, methionine, threonine and isoleucine) were less discernible. Partitioning of carbon flux into diverging pathways such as the aromatic amino acids branch is assumed to be constant. Expanding the aromatic amino acid pathways and linking it to the central carbon metabolism permitted an in-depth study of conflicting objectives in serine and tryptophan biosynthesis.

The dynamic tryptophan operon model (Santillán and Mackey, 2001a) and the common aromatic amino acid-terminal tryptophan biosynthesis pathways were linked to the central carbon metabolism (Chassagnole et al., 2002) to form the augmented model. The ability of the central carbon metabolism to channel carbon fluxes through the common aromatic amino acid and terminal tryptophan biosynthesis pathways was achieved by substituting the tryptophan biosynthesis kinetics of the dynamic *trp* operon model with certain key metabolites (3-deoxy-D-*arabino*-heptulosonate 7-phosphate (dahp), chorismate (cho), tryptophan (trp), 5-phosphoribosyl-α-pyrophosphate (prpp) and serine (ser)) and representative enzyme (i.e. anthranilate synthase) of the terminal tryptophan biosynthesis pathway.

Since it was not possible to obtain unique estimates of the six new kinetic parameters of the augmented model due to limited published data, estimates were

selected with justification from a number of parameter estimation runs to minimize the sum of squares of fractional errors. The augmented model was carefully evaluated in terms of goodness-of-fit of calculated metabolite concentrations with the reference data. Differences between the calculated enzymatic fluxes of the augmented model and those of the central carbon metabolism are more appreciable for DAHPS (9.64%), G1PAT (9.30%) and PEPCxylase (4.27%) due to diversion of carbon fluxes into aromatic amino acid pathways. Steady-state concentrations and fluxes were calculated using the newly developed augmented model to form the reference values in the optimization study. Larger deviations between calculated and reference concentrations and fluxes were due to kinetic parameter uncertainties and structural assumptions (such as mass balances in the aromatic amino acid pathways) used in developing the augmented model. However, the augmented model captures the sensitivities of fluxes to tryptophan biosynthesis and underlying competition among DAHPS, TrpSynth, G1PAT and PEPCxylase fluxes for carbon sources.

Pattern recognition and multiple values for the random number seed were used in the heuristics approach to identify genes for the concurrent knockout and manipulation, for optimizing serine and tryptophan flux ratios. In the two-gene manipulation, genes encoding SerSynth, TrpSynth, GAPDH and G1PAT were identified using a narrow multiplier range of 0.8-1.25. Not all random seeds were capable of generating the global Pareto, and existence of the local Pareto was attributed to domination by an isolated global Pareto solution. One Pareto branch was obtained using NSGA-II for a wider multiplier range of 0-1.5 when genes encoding GAPDH and SerSynth were manipulated. The remaining Pareto was obtained through simulations following the analysis of Pareto patterns in both objective and decision variable space. Missing Pareto solutions were attributed to the rapid accumulation of certain non-dominated solutions at the expense of others. The results obtained were in line with the reported fermentation studies (Ikeda and Katsumata, 1999; Ikeda et al., 1994) and consistent with the known dynamic behaviour of the tryptophan operon.

In the concurrent three-gene knockout and manipulation, genes encoding SerSynth, TrpSynth, GAPDH, RPPK, G1PAT, PK and G6PDH were identified and the various Pareto-optimal solutions were explained in terms of collaborative and competitive supply and demand among the contributing fluxes. DAHPS and ChoSynth fluxes were not part of the Pareto for both the dual and triple gene cases thus suggesting the efficient transport of carbon sources to the terminal steps of

tryptophan biosynthesis. Central carbon precursors such as erythrose 4-phosphate (e4p) and phosphoenolpyruvate (pep) rather than DAHPS or ChoSynth, are limiting the partitioning of carbon flow into the aromatics pathways; this is supported by fermentation experiments (Berry, 1996; Ikeda and Katsumata, 1999) where recombinant DNA technology was applied to the microbial cell factory. The triple gene application could not be extended to the wider multiplier range of 0-1.5 due to the existence of large number of infeasible solutions.

The present study brings out the many options available for optimizing bioprocesses, to the decision maker to make the best judgements subject to additional information (such as economic conditions and environmental impact) which are impossible to quantify or could not be quantified accurately. It also highlights challenges in modelling and optimizing complex bioprocesses. In general, core competencies in chemical engineering can be extended to systems biotechnology to complement experimental work in bioprocesses.

6.2 Recommendations for Future Study

The current work can be extended to four related areas in systems biotechnology − modelling, optimization and stability considerations, Pareto ranking and decision making, and techniques for minimizing numerical difficulties. Experimental validation of the augmented model and optimization results obtained would provide insightful feedback to future systems biotechnology study. In the following, recommendations for future studies on these are presented outlining the motivation, scope and challenges based on the experience gained from the doctoral research and prior knowledge.

Besides tryptophan biosynthesis, threonine biosynthesis can be added to the augmented model. The aspartate-threonine pathway diverging from PEPCxylase leading to threonine biosynthesis can be developed taking into account the key metabolites such as oxaloacetate, aspartate, aspartyl phosphate, aspartate semialdehyde, homoserine and homoserine phosphate and their respective enzyme kinetics. New mass balances are required for the aspartate-threonine pathway. The existing mass balance for phosphoenolpyruvate (pep) can be modified to cater for the conversion of pep into oxaloacetate. The kinetic parameters can be estimated by

concurrent minimization of the sum of squares of the fractional errors of concentrations and fluxes.

Stability of the microbial cell that involves multicopy plasmids carrying *trp* operon and *trpR* genes inserted through recombinant DNA technology, can be studied starting with copy number between 3 and 10 (Ikeda, 2006). The plasmid copy number is an integer decision variable. Stability of the model should be evaluated through the characteristic polynomials using suitable stability tests (Chang et al., 2005). By drawing on the insights gained from metabolic network stability analysis and the present research, optimization for two objectives (serine and tryptophan biosynthesis rates) and three objectives (serine, tryptophan and threonine biosynthesis rates) can be performed where multicopy plasmids are introduced with an additional constraint to enforce stability within a prespecified neighbourhood of a given solution. Optimization is likely to be more challenging when the range of the plasmid copy number is widened. Optimization for dual and triple objectives can also be performed without considering stability in the neighbourhood of a solution. The two sets of results, with and without stability constraints, should be analyzed in terms of metabolic network integrity.

Pareto ranking can be performed using the net flow and rough set methods (Thibault, 2008). A comparative study of net flow and rough set methods could lead to interesting insights of the Pareto obtained through optimization with and without stability considerations, similar to chemical process applications in Renaud et al. (2007). In another study, the four parameters of the net flow method – weighting of each objective, indifference threshold, preference threshold and veto threshold – could be stochastic instead of fixed values selected by the decision maker. The entire Pareto set can be transferred to Microsoft Excel and analyzed via Monte Carlo simulation using Crystal Ball software (originally supplied by Decisioneering – http://www.oracle.com/crystalball/index.html; accessed in March 2009). Crystal Ball combined with Excel allows definition of the probability functions of the four stochastic variables, generates random numbers based on these probability distributions, and stores the results for each trial. The probability distribution need not be Gaussian; other distributions such as Poisson could be used to reflect the stochastic process of decision making in which events occur continuously and independently of each other. An example of the Poisson process is the arrival of "customers" in a queue. With automated Pareto ranking, the progression of Pareto ranking can be

observed and studied. Interesting but currently unknown insights into decision making process would be the prime motivation for this study. The nature of decision making need not necessarily follow Gaussian distribution since selection of the best Pareto chromosome is affected by many factors such as economic demand and supply for the multiple end-products.

The existing NSGA-II algorithm gave disjointed Pareto-optimal front in the present study using the augmented model subject to homeostatic and total enzymatic flux constraints. This provides the motivation to modify the NSGA-II algorithm. One possibility is to accept dominated chromosomes as part of the new parent population provided that the dominated chromosomes create diversity. Dominated chromosomes which remain so after certain number of generations can be deleted from the parent population and substituted with another set of dominated chromosomes. Elitist preservation mechanism retains good chromosomes which are not dominated by others. This continues until a given criterion such as the final generation number is met. The optimizer will report whether the final solutions are completely non-dominated or not. The modified sorting procedure in NSGA-II can be tested using augmented model for bi-objective maximization of serine and tryptophan flux ratios when concurrent 2- and 3-gene knockout and manipulation is applied. There is a risk that the suggested strategy may not work as well as expected. It is not known whether other algorithms are able to give whole Pareto-optimal front in one optimization run, which is a great advantage and prime motivation for future work, without resorting to other means as were the case for heuristics technique.

Experimental validation of the augmented model and/or optimization results provides real data that could be used to develop new kinetic rate equations and improve the accuracy of the kinetic parameters of the augmented model. Though there are limitations to “black box” approach in stimulus response experiments, they complement the experimental insights of microbiology and genetics. Future applications should be targeted at novel and valuable products in the biopharmaceutical industry.

REFERENCES

Agrawal, N., Rangaiah, G. P., Ray, A. K., and Gupta, S. K. (2006). Multi-objective optimization of the operation of an industrial low-density polyethylene tubular reactor using genetic algorithm and its jumping gene adaptations. *Industrial and Engineering Chemistry Research, 45*(9), 3182-3199.

Araúzo-Bravo, M. J., Cano-Izquierdo, J. M., Gómez-Sánchez, E., López-Nieto, M. J., Dimitriadis, Y. A., and López-Coronado, J. (2004). Automatization of a penicillin production process with soft sensors and an adaptive controller based on neuro fuzzy systems. *Control Engineering Practice*, *12*(9), 1073-1090.

Azuma, S., Tsunekawa, H., Okabe, M., Okamoto, R., and Aiba, S. (1993). Hyper-production of L-tryptophan via fermentation with crystallization. *Applied Microbiology and Biotechnology, 39*(4-5), 471-476.

Bajpai, R. K., and Reuß, M. (1980). A mechanistic model for penicillin production. *Journal of Chemical Technology and Biotechnology, 30*(6), 332-344.

Berry, A. (1996). Improving production of aromatic compounds in *Escherichia coli* by metabolic engineering. *Trends in Biotechnology, 14*(7), 250-256.

Bhartiya, S., Rawool, S., and Venkatesh, K. V. (2003). Dynamic model of *Escherichia coli* tryptophan operon shows an optimal structural design. *European Journal of Biochemistry, 270*(12), 2644-2651.

Bhaskar, V., Gupta, S. K., and Ray, A. K. (2000). Applications of multiobjective optimization in chemical engineering. *Reviews in Chemical Engineering, 16*(1), 1-54.

Bhutani, N., Ray, A. K., and Rangaiah, G. P. (2006). Modeling, simulation, and multi-objective optimization of an industrial hydrocracking unit. *Industrial and Engineering Chemistry Research, 45*(4), 1354-1372.

Birol, G., Űndey, C., Parulekar, S. J., and Çinar, A. (2002). A morphologically structured model for penicillin production. *Biotechnology and Bioengineering, 77*(5), 538-552.

Biwer, A., Griffith, S., and Cooney, C. (2004). Uncertainty analysis of penicillin V production using Monte Carlo simulation. *Biotechnology and Bioengineering, 90*(2), 167-179.

Bongaerts, J., Krämer, M., Müller, U., Raeven, L., and Wubbolts, M. (2001). Metabolic engineering for microbial production of aromatic amino acids and derived compounds. *Metabolic Engineering, 3*(4), 289-300.

Bremer, H., and Dennis, P. P. (1996). Modulation of chemical composition and other parameters of the cell by growth rate. In: Niedhardt, F. C., Curtiss, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., editors. *Escherichia coli and Salmonella: Cellular and Molecular Biology*, vol. 2, 1553-1569. Washington, D. C.: ASM Press.

Chan, E.-C., Tsai, H.-L., Chen, S.-L., and Mou, D.-G. (1993). Amplification of the tryptophan operon gene in *Escherichia coli* chromosome to increase L-tryptophan biosynthesis. *Applied Microbiology and Biotechnology, 40*(2-3), 301-305.

Chang, Y., and Sahinidis, N. V. (2005). Optimization of metabolic pathways under stability considerations. *Computers and Chemical Engineering, 29*(3), 467-479.

Chankong, V., and Haimes, Y. Y. (1983). *Multiobjective decision making: theory and methodology*. New York: North Holland.

Chassagnole, C., Noisommit-Rizzi, N., Schmid, J. W., Mauch, K., and Reuss, M. (2002). Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnology and Bioengineering, 79*(1), 53-73.

Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms*. Chichester, UK: Wiley.

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation, 6*(2), 182-197.

Dengering, D., Froemel, C., Dikta, G., and Takors, R. (2004). Sensitivity analysis for the reduction of complex metabolism models. *Journal of Process Control, 14*(7), 729-745.

Dietz, A., Aguilar-Lasserre, A., Azzaro-Pantel, C., Pibouleau, L., and Domenech, S. (2008). A fuzzy multiobjective algorithm for multiproduct batch plant: application to protein production. *Computers and Chemical Engineering, 32*(1-2), 292-306.

Diwekar, U. M. (2003). *Introduction to applied optimization.* Norwell, Mass: Kluwer Academic Publishers.

Eggeling, L., Pfefferle, W., and Sahm, H. (2006). Amino acids. In: Ratledge, C., and Kristiansen, B., editors. *Basic biotechnology*. Cambridge, UK: Cambridge University Press.

Gotta, S. L., Miller, O. L., and French, S. L. (1991). rRNA transcription rate in *Escherichia coli*. *Journal of Bacteriology, 173*(20), 6647-6649.

Gunsalus, R. P, Miguel, G., and Gunsalus, G. L. (1986). Intracellular trp repressor levels in *Escherichia coli*. *Journal of Bacteriology, 167*(1), 272-278.

Haimes, Y. Y. (1977). *Hierarchical analyses of water resources systems: modeling and optimization of large-scale systems.* New York: McGraw Hill.

Halsall-Whitney, H., and Thibault, J. (2006). Multi-objective optimization for chemical processes and controller design: approximating and classifying the Pareto domain. *Computers and Chemical Engineering, 30*(6-7), 1155-1168.

Hersbach, G. J. M., C. P. van der Beek, and P. W. M. van Dijck. (1984). The penicillins: properties, biosynthesis, and fermentation. In: Vandamme, E. J., Editor. *Biotechnology of industrial antibiotics*. New York: M. Dekker.

Ikeda, M. (2006). Towards bacterial strains overproducing L-tryptophan and other aromatics by metabolic engineering. *Applied Microbiology and Biotechnology, 69*(6), 615-626.

Ikeda, M., and Katsumata, R. (1999). Hyperproduction of tryptophan by *Corynebacterium glutamicum* with the modified pentose phosphate pathway. *Applied and Environmental Microbiology, 65*(6), 2497-2502.

Ikeda, M., Nakanishi, K., Kino, K., and Katsumata, R. (1994). Fermentative production of tryptophan by a stable recombinant strain of *Corynebacterium glutamicum* with a modified serine-biosynthetic pathway. *Bioscience, Biotechnology and Biochemistry, 58*(4), 674-678.

Jørgensen, H., Nielsen, J., Villadsen, J., and Møllgaard, H. (1995). Metabolic flux distributions in *Penicillium chrysogenum* during fed-batch cultivations. *Biotechnology and Bioengineering, 46*(2), 117-131.

Kim, K.-J., and Smith, R. L. (2004). Parallel multiobjective evolutionary algorithms for waste solvent recycling. *Industrial and Engineering Chemistry Research, 43*(11), 2669-2679.

Kim, T.-H., Namgoong, S., Kwak, J. H., Lee, S.-Y., and Lee, H.-S. (2000). Effects of *tkt*A, *aro*F^{FBR}, and *aro*L expression in the tryptophan-producing *Escherichia coli*. *Journal of Microbiology and Biotechnology, 10*(6), 789-796.

Koh, B. T., and Yap, M. G. S. (1993). A simple genetically structured model of *trp* repressor- operator interactions. *Biotechnology and Bioengineering, 41*(7), 707-714.

Koh, B. T., Tan, R. B. H., and Yap, M. G. S. (1998). Genetically structured mathematical modeling of *trp* attenuator mechanism. *Biotechnology and Bioengineering, 58*(5), 502-509.

Kookos, I. K. (2004). Optimization of batch and fed-batch bioreactors using simulated annealing. *Biotechnology Progress, 20*(4), 1285-1288.

Kusher, S. R. (1996). mRNA delay. In: Niedhardt, F. C., Curtiss, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., editors. *Escherichia coli and Salmonella: Cellular and Molecular Biology*, vol. 1, 849-860. Washington, D. C.: ASM Press.

Landick, R., Turnbough, C. L., and Yanofsky, C. (1996). Transcription attenuation. In: Niedhardt, F. C., Curtiss, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., editors. *Escherichia coli and Salmonella: Cellular and Molecular Biology*, vol. 1, 1263-1286. Washington, D. C.: ASM Press.

Lee, F. C., Rangaiah, G. P., and Lee, D.-Y. (2009). Optimization of a multi-product microbial cell factory for multiple objectives – a paradigm for metabolic pathway recipe. In: Rangaiah, G. P., Editor. *Multi-objective optimization: techniques and applications in chemical engineering.* Singapore: World Scientific.

Lee, F. C., Rangaiah, G. P., and Ray, A. K. (2007). Multi-objective optimization of an industrial penicillin V bioreactor train using non-dominated sorting genetic algorithm. *Biotechnology and Bioengineering, 98*(3), 586-598.

Lee, S. Y., Lee, D.-Y., and Kim, T. Y. (2005a). Systems biotechnology for strain improvement. *Trends in Biotechnology, 23*(7), 349-358.

Lee, S. J., Lee, D-Y., Kim, T. Y., Kim, B. H., Lee, J., and Lee, S. Y. (2005b). Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and in silico gene knockout simulation. *Applied and Environmental Microbiology, 71*(12), 7880-7887.

Leuchtenberger, W., Huthmacher, K., and Drauz, K. (2005). Biotechnological production of amino acids and derivatives: current status and prospects. *Applied Microbiology and Biotechnology, 69*(1), 1-8.

Liao, J. C., Hou, S.-Y., and Chao, Y. P. (1996). Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnology and Bioengineering, 52*(1), 129-140.

Link, H., Vera, J., Weuster-Botz, D., Darias, N. T., and Franco-Lara, E. (2008). Multi-objective steady state optimization of biochemical reaction networks using a constrained genetic algorithm. *Computers and Chemical Engineering, 32*(8), 1707-1713.

Liu, P.-K., and Wang, F.-S. (2008). Inference of biochemical network models in S-system using multiobjective optimization approach. *Bioinformatics, 24*(8), 1085-1092.

Lowe, D. A. (2001). Antibiotics. In: Ratledge, C., and Kristiansen, B., editors. *Basic biotechnology*. Cambridge, UK: Cambridge University Press.

Mandal, C., Gudi, R. D., and Suraishkumar. G. K. (2005). Multi-objective optimization in *Aspergillus niger* fermentation for selective product enhancement. *Bioprocess and Biosystems Engineering, 28*(3), 149-164.

Mascarenhas, D., Ashworth, D. J., and Chen, C. S. (1991). Deletion of *pgi* alters tryptophan biosynthesis in a genetically engineered strain of *Escherichia coli*. *Applied and Environmental Microbiology, 57*(10), 2995-2999.

Mashego, M. R., Rumbold, K., De Mey, M., Vandamme, E., Soetaert, W., and Heijnen, J. J. (2007). Microbial metabolomics: past, present and future methodologies. *Biotechnology Letters, 29*(1), 1-16.

Menezes, J. C., Alves, S. S., Lemos, J. M., and Feyo de Azevedo, S. (1994). Mathematical modelling of industrial pilot-plant penicillin-G fed-batch fermentations. *Journal of Chemical Technology and Biotechnology*, *61*(2), 123-138.

Miettinen, K. (1999). *Nonlinear multiobjective optimization.* Boston: Kluwer Academic Publishers.

Mŏsaťˇ, A., Cavin, L., Fischer, U., and Hungerbühler, K. (2008). Multiobjective optimization of multipurpose batch plants using superequipment class concept. *Computers and Chemical Engineering, 32*(3), 512-529.

Nagrath, D., Avila-Elchiver, M., Berthiaume, F., Tilles, A. W., Messac, A., and Yarmush, M. L. (2007). Integrated energy and flux balance based multiobjective framework for large-scale metabolic networks. *Annals of Biomedical Engineering, 35*(6), 863-885.

Nandasana, A. D., Ray, A. K., and Gupta, S. K. (2003). Applications of the non-dominated sorting genetic algorithm (NSGA) in chemical reaction engineering. *International Journal of Chemical Reactor Engineering , 1*(R2).

Neidhardt, F. C., and Umbarger, H. E. (1996). Chemical composition of *Escherichia coli*. In: Niedhardt, F. C., Curtiss, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., editors. *Escherichia coli and Salmonella: Cellular and Molecular Biology*, vol. 1, 13-16. Washington, D. C.: ASM Press.

Nielsen, J. (1993). A simple morphologically structured model describing the growth of filamentous microorganisms. *Biotechnology and Bioengineering, 41*(7), 715-727.

Oh, P. P., Ray, A. K., and Rangaiah, G. P. (2001). Triple-objective optimization of an industrial hydrogen plant. *Journal of Chemical Engineering of Japan, 34*(11), 1341-1355.

Ohno, M., et al. (2003). Antibiotics. In: *Ullmann's Encyclopedia of Industrial Chemistry*. Weinheim: Wiley-VCH.

Oldiges, M., Lütz, S., Pflug, S., Schroer, K., Stein, N., and Wiendahl, C. (2007). Metabolomics: current state and evolving methodologies and tools. *Applied Microbiology and Biotechnology, 76*(3), 495-511.

Pareto, V. (1896). *Cours d'economie Politique.* F. Rouge, Lausanne, Switzerland.

Patnaik, R., and Liao, J. C. (1994). Engineering of *Escherichia coli* central metabolism for aromatic metabolite production with near theoretical yield. *Applied and Environmental Microbiology, 60*(11), 3903-3908.

Paul, G. C., and Thomas, C. R. (1996). A structured model for hyphal differentiation and penicillin production using Penicillium chrysogenum. *Biotechnology and Bioengineering, 51*(5), 558-572.

Perry, R. H., Green, D. W., and Maloney, J. O. (1997). *Perry's chemical engineers' handbook.* New York: McGraw-Hill.

Piperno, J. R., and Oxender, D. L. (1968). Amino acid transport systems in *Escherichia coli* K12. *The Journal of Biological Chemistry, 243*(22), 5914-5920.

Pittard, A. J. (1996). Biosynthesis of the aromatic amino acids. In: Niedhardt, F. C., Curtiss, R., Ingraham, J. L., Lin, E. C. C., Low, K. B., Magasanik, B., Reznikoff, W. S., Riley, M., Schaechter, M., and Umbarger, H. E., editors. *Escherichia coli and Salmonella: Cellular and Molecular Biology*, vol. 1, 458-484. Washington, D. C.: ASM Press.

Rangaiah, G. P. (2009). *Advances in process systems engineering: multi-objective optimization - techniques and applications in chemical engineering, 1*. Singapore: World Scientific.

Renaud, J., Thibault, J., Lanouette, R., Kiss, L. N., Zaras, K., and Fonteix, C. (2007). Comparison of two multicriteria decision aid methods: net flow and rough set methods in a high yield pulping process. *European Journal of Operational Research, 177*(3), 1418-1432.

Rizzi, M., Baltes, M., Theobald, U., and Reuss, M. (1997). In vivo analysis of metabolic dynamics in *Saccharomyces cerevisiae:* II. mathematical model. *Biotechnology and Bioengineering, 55*(4), 592-608.

Santillán, M., and Mackey, M. C. (2001a). Dynamic regulation of the tryptophan operon: a modeling study and comparison with experimental data. *Proceedings of the National Academy of Sciences of the United States of America, 98*(4), 1364-1369.

Santillán, M., and Mackey, M. C. (2001b). Dynamic behaviour in mathematical models of the tryptophan operon. *Chaos, 11*(1), 261-268.

Santillán, M., and Zeron, E. S. (2004). Dynamic influence of feedback enzyme inhibition and transcription attenuation on tryptophan operon response to nutritional shifts. *Journal of Theoretical Biology, 231*(2), 287-298.

Sarkar, D., and Modak, J. M. (2005). Pareto-optimal solutions for multi-objective optimization of fed-batch bioreactors using nondominated sorting genetic algorithm. *Chemical Engineering Science, 60*(2), 481-492.

Scheper, T., Faurie, R., and Thommel, J. (2003). *Advances in biochemical engineering/biotechnology: microbial production of L-amino-acids, 79*. Berlin, Germany: Springer-Verlag.

Schmid, J. W., Mauch, K., Reuss, M., Gilles, E. D., and Kremling, A. (2004). Metabolic design based on a coupled gene expression – metabolic network model of tryptophan production in *Escherichia coli*. *Metabolic Engineering, 6*(4), 364-377.

Schmitt, T. H., Zeng, Z., and Jardetzky, O. (1995). Dynamics of tryptophan binding to *Escherichia coli* trp repressor wild type and AV77 mutant: an NMR study. *Biochemistry, 34*(40), 13183-13189.

Schuler, M. L., and Kargi, F. (2002). *Bioprocess engineering: basic concepts.* New Jersey, USA: Prentice Hall.

Sendin, O. H., Vera, J., Torres, N. V., and Banga, J. R. (2006). Model based optimization of biochemical systems using multiple objectives: a comparison of several solution strategies. *Mathematical and Computer Modelling of Dynamical Systems, 12*(5), 469-487.

Srinivas, N., and Deb, K. (1995). Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation, 2*(3), 221-248.

Steffens, M. A., Fraga, E. S., and Bogle, I. D. L. (1999). Multicriteria process synthesis for generating sustainable and economic bioprocesses. *Computers and Chemical Engineering, 23*(10), 1455-1467.

Tarafder, A., Lee, B. C. S., Ray, A. K., and Rangaiah, G. P. (2005). Multiobjective optimization of an industrial ethylene reactor using a nondominated sorting genetic algorithm. *Industrial and Engineering Chemistry Research, 44*(1), 124-141.

Tarafder, A., Rangaiah, G. P., and Ray, A. K. (2007). A study of finding many desirable solutions in multiobjective optimization of chemical processes. *Computers and Chemical Engineering, 31*(10), 1257-1271.

Teusink, B., Passarge, J., Reijenga, C. A., Esgalhado, E., van der Weijden, C. C., Schepper, M., Walsh, M. C., Bakker, B. M., van Dam, K., Westerhoff, H. V., and Snoep, J. L. (2000). Can yeast glycolysis be understood in terms of *in vitro* kinetics of the constituent enzymes? Testing biochemistry. *European Journal of Biochemistry, 267*(17), 5313-5329.

Thibault, J. (2008). Net flow and rough sets: two methods for ranking the Pareto domain. In: Rangaiah, G. P., Editor. *Multi-objective optimization: techniques and applications in chemical engineering.* Singapore: World Scientific.

Torres, N. V., and Voit, E. O. (2002). *Pathway analysis and optimization in metabolic engineering.* New York: Cambridge University Press.

Turton, R., Bailie, R. C., Whiting, W. B., and Shaeiwitz, J. A. (2003). *Analysis, synthesis, and design of chemical processes*. New Jersey, USA: Prentice Hall.

van Nistelrooij, H. J. M., Krijgsman, J., E. de Vroom, and Oldenhof, C. (1998). Penicillin update: industrial. In: Mateles, R. I., editor. *Penicillin: a paradigm for biotechnology*. Chicago, Illinois: Candida Corporation.

Vera, J., de Atauri, P., Cascante, M., and Torres, N. V. (2003). Multicriteria optimization of biochemical systems by linear programming: application to production of ethanol by *Saccharomyces cerevisiae. Biotechnology and Bioengineering, 83*(3), 335-343.

Visser, D., Schmid, J. W., Mauch, K., Reuss, M., and Heijnen, J. J. (2004). Optimal re-design of primary metabolism in *Escherichia coli* using linlog kinetics. *Metabolic Engineering, 6*(4), 378-390.

Vital-Lopez, F. G., Armaou, A., Nikolaev, E. V., and Maranas, C. D. (2004). A computational procedure for optimal engineering interventions using kinetic models of metabolism. *Biotechnology Progress, 22*(6), 1507-1517.

Voit, E. O. (2000). *Computational analysis of biochemical systems: a practical guide for biochemists and molecular biologists.* Cambridge: Cambridge University Press.

Wahl, S. A., Haunschild, M. D., Oldiges, M., and Wiechert, W. (2006). Unravelling the regulatory structure of biochemical networks using stimulus response experiments and large-scale model selection. *IEE Systems Biology, 153*(4), 275-285.

Wang, D. I. C., Cooney, C. L., Demain, A. L., Dunnill, P., Humphrey, A. E., and Lilly, M. D. (1979). *Fermentation and enzyme technology.* New York: John Wiley.

Wendisch, V. F., Bott, M., Kalinowski, J., Oldiges, M., and Wiechert, W. (2006). Emerging *Corynebacterium glutamicum* systems biology. *Journal of* Biotechnology, *124*(1), 74-92.

Xiu, Z.-L., Chang, Z.-Y., and Zeng, A.-P. (2002). Nonlinear dynamics of regulation of bacterial *trp* operon: model analysis of integrated effects of repression, feedback inhibition, and attenuation. *Biotechnology Progress, 18*(4), 686-693.

Xiu, Z.-L., Zeng, A.-P., and Deckwer, W.-D. (1997). Model analysis concerning the effects of growth rate and intracellular tryptophan level on the stability and dynamics of tryptophan biosynthesis in bacteria. *Journal of Biotechnology*, *58*(2), 125-140.

Yee, A. K. Y., Ray, A. K., and Rangaiah, G. P. (2003). Multiobjective optimization of an industrial styrene reactor. *Computers and Chemical Engineering*, *27*(1), 111-130.

Zangirolami, T. C., Johansen, C. L., Nielsen, J., and Jørgensen, S. B. (1997). Simulation of penicillin production in fed-batch cultivations using a morphologically structured model. *Biotechnology and Bioengineering*, 56(6), 593-604.

Appendix A

Transient Enzymatic Reaction Fluxes and Metabolite Concentrations Profiles

The transient profiles shown here are calculated following an injection of glucose pulse (height = 16 mM; width 0.1 sec) into steady-state wild strain *E. coli* culture. The "f_{pulse}" term in $\left(\frac{dC_{glc}^{extracellular}}{dt} = D\left(C_{glc}^{feed} - C_{glc}^{extracellular}\right) + f_{pulse} - \frac{C_x r_{PTS}}{\rho_x} \right)$ has a height of 16 mM. Integration step of 0.1 sec is used throughout the simulations to match the sampling period of 0.2 second (Chassagnole et al., 2002) in the automated stopped-flow techniques for measuring metabolite concentrations at fixed points in time. No measurements of the fluxes were done. Constant co-metabolite concentrations are assumed in all the simulations shown below.

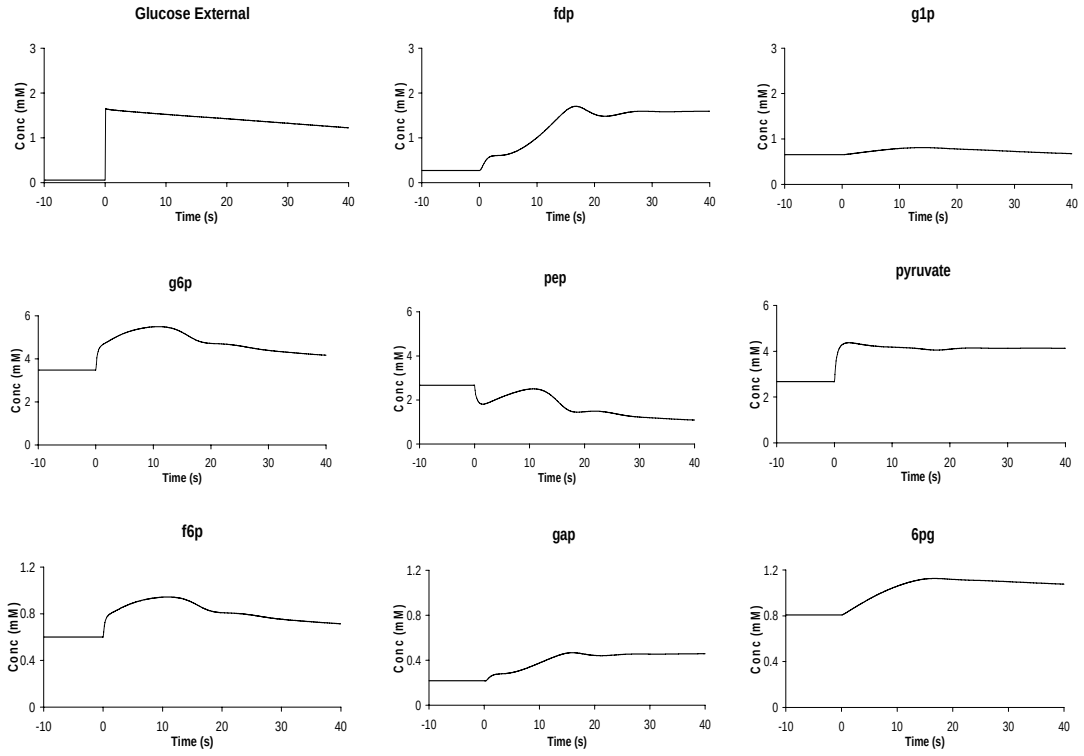


Fig. A.1. Simulated metabolite concentrations in a steady-state *E. coli* culture after a glucose pulse. They are comparable to the measured concentrations (Chassagnole et al., 2002). Refer to List of Symbols for the definitions of the abbreviations.

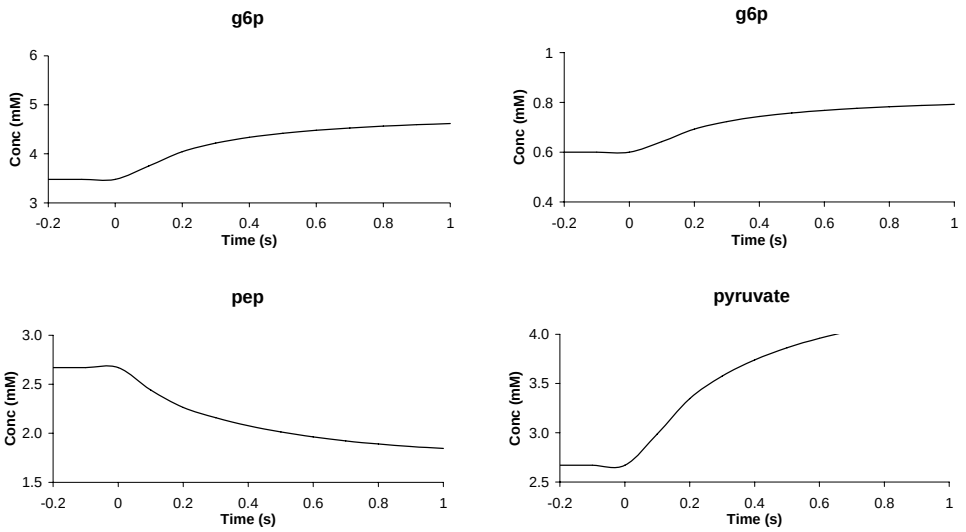


Fig. A.2. Simulated sub-second metabolite concentrations in a steady-state *E. coli* culture after a glucose pulse. They are comparable to the measured concentrations (Chassagnole et al., 2002) using stopped-flow techniques. Refer to List of Symbols for the definitions of the abbreviations.

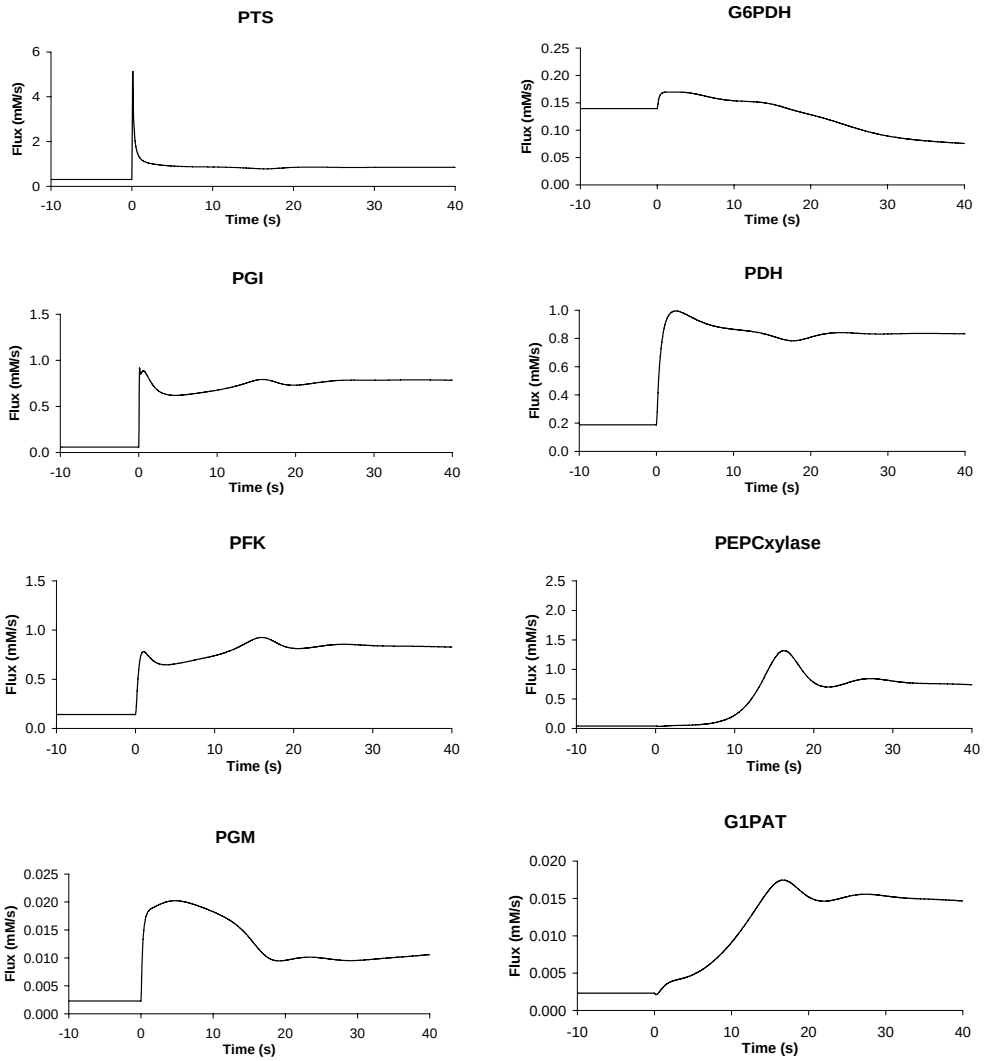


Fig. A.3. Simulated fluxes in a steady-state *E. coli* culture after a glucose pulse. Refer to List of Symbols for the definitions of the abbreviations.

Appendix B

Tryptophan Operon Model Parameters Adaptation

Specific Growth Rate

The specific growth rate of the central carbon metabolism (Chassagnole et al., 2002) is adopted in the augmented model. This appendix shows the adjustments needed for the parameters in equations (4.1)-(4.4) and (4.6) when a specific cell growth rate of 0.1 doublings per h is used instead of the original 0.6 doublings per h used by Santillán and Mackey (2001a).

Cell Volume

E. coli are rod-like bacteria 3-5 μm long and 0.5 μm in diameter (Santillán and Mackey, 2001a). They have a volume in the range from 5.9×10^{-16} litres to 9.8×10^{-16} litres. Mean volume is 7.8×10^{-16} litres (compared to 8.0×10^{-16} litres in Santillán and Mackey, 2001a).

Specific cell growth rate = μ = 0.1 doublings per h = 0.1/60 doublings per min (compared to 0.6/60 doublings per min in Santillán and Mackey, 2001a).

Rate Constant for mRNA Polymerase (mRNAP) Binding to a Free Operon DNA Operator Site

Polynomial curve fitting is applied to the data in Table B.1 (Bremer and Dennis, 1996).

Table B.1 Number of mRNA polymerase molecules in a cell and specific growth ate.

Specific growth rate (doublings per h)	0.6	1.0	1.5	2.0	2.5
Number of mRNA molecules (10^3)	1.5	2.8	5.0	8.0	11.4

The number of mRNA polymerase molecules in a cell is given by:

$-191.1(0.1)^3 + 2161.5(0.1)^2 + 57.804(0.1) + 738.77 = 765.97$

mRNA polymerase concentration in a cell = $\frac{765.97}{\frac{6.02x10^{23}}{7.8x10^{-16}}}$ = 1.63126 μM. (compared to 2.6 μM in Santillán and Mackey, 2001a).

Tryptophan operon allows transcription initiation every 0.1 minute (Landick et al., 1996).

The rate constant $k_p = \frac{1}{\tau_p P} = \frac{1}{(0.1)(1.63126)}$ = 6.13023 per μM per min (compared to 3.9 per μM per min in Santillán and Mackey, 2001a).

Rate Constant for Ribosome Binding to a Free *trpE*-related Site on an mRNA

Polynomial curve fitting is applied to the data in Table B.2 (Bremer and Dennis, 1996).

Table B.2 Number of ribosomes in a cell and specific growth rate.

Specific growth rate (doublings per h)	0.6	1.0	1.5	2.0	2.5
Number of ribosomes (10^3)	6.8	13.5	26.3	45.1	72.0

The number of ribosomes in a cell is given by:

$2238.6(0.1)^3 + 2415.7(0.1)^2 + 8674.8(0.1) + 224.96 = 1118.8356$

Ribosome concentration in a cell = $\frac{\frac{1118.8356}{6.02x10^{23}}}{7.8x10^{-16}}$ = 2.3827 μM (compared to 2.9 μM in Santillán and Mackey, 2001a).

Efficient mRNAs have been observed to initiate translation every 0.05 minute (Kusher 1996).

The rate constant $k_\rho = \frac{1}{\tau_\rho \rho} = \frac{1}{(0.05)(2.3827)}$ = 8.39384 per μM per min (compared to 6.9 per μM per min in Santillán and Mackey, 2001a).

Time Taken by a Ribosome to Synthesize a *trpE* Polypeptide

Polynomial curve fitting is applied to the data in Table B.3 (Bremer and Dennis, 1996).

Table B.3 mRNA elongation rate and specific growth rate.

Specific growth rate (doublings per h)	0.6	1.0	1.5	2.0	2.5
mRNA chain elongation rate (Nucl./s)	39	45	50	52	55

The mRNA chain elongation rate is given by:
$2.9741(0.1)^4 - 15.485(0.1)^3 + 22.79(0.1)^2 + 2.4152(0.1) + 32.306 = 32.76$ nucleotides per s (compared to 39 nucleotides per s in Santillán and Mackey, 2001a).

The *trpE* polypeptide (a subunit of the enzyme anthranilate synthase) is 520 amino acids long. This means that the length of the *trpE* gene is 1560 nucleotides long. τ_e is the time it takes for a ribosome to synthesize a *TrpE* polypeptide. Therefore, $\tau_e = 1560/32.76 = 47.6$ s = 0.79 min (compared to 0.66 min in Santillán and Mackey, 2001a).

Operon, Aporepressor and Total Repressor Concentrations in a Cell

Let R, R_I and T be the total repressor, aporepressor (inactive repressor) and tryptophan concentrations, respectively. R_I reacts with T to form the active repressor. The total repressor concentration is the sum of the aporepressor and active repressor (holorepressor) concentrations. The two binding sites on an aporepressor molecule are independent and identical. Binding of two tryptophan molecules to an aporepressor molecule is modelled as Michaelis-Menten type where the Hill coefficient is 1.0.

At equilibrium, $R = \left(\frac{K_t + T}{K_t}\right) R_I$ where K_t is the ratio of the backward rate constant to the forward rate constant. The tryptophan concentration (T) in the wild type *E. coli* is taken to be 4.1 μM (Santillán and Mackey, 2001a). The equilibrium constant K_t is obtained experimentally by Schmitt et al. (1995).

From Gunsalus et al. (1986), the normal concentration of aporepressor in a tryptophan free culture medium is 375 molecules per cell, which is equivalent to $\frac{375}{\frac{6.02 \times 10^{23}}{7.8 \times 10^{-16}}}$ or 0.7986 μM (compared to 0.75 μM in Santillán and Mackey, 2001a).

$R = \left(\frac{60.34 + 4.1}{60.34}\right)(0.7986) = 0.85$ μM (compared to 0.8 μM in Santillán and Mackey, 2001a).

In normal *E. coli*, there is only one tryptophan DNA operator site per genome. The average number of genome equivalents per cell, according to Bremer and Dennis

(1996), is around 1.6. Therefore, normal operon concentration is $\frac{1.6}{\frac{6.02 \times 10^{23}}{7.8 \times 10^{-16}}}$ i.e. 3.41 x 10^{-3} μM (compared to 3.32 x 10^{-3} μM in Santillán and Mackey, 2001a).

Rate Constants of Tryptophan Production and Internal Consumption

The operon model parameters in equations (4.5), (4.7), (4.8) and (4.9) are not required since these equations are replaced by equations (4.20), (4.21) and (4.22). Reasons for this are given in Section 4.3.

Appendix C

Estimating Steady-state Concentration of Serine

The total dry weight of a cell is 2.8 x 10^{-13} g based on an average of measurements (Neidhart and Umbarger, 1996). Assuming that 70% of the cell is water, the water content is 6.7 x 10^{-13} g. Total wet weight of one cell is 9.5 x 10^{-13} g. Density of a cell = 9.5 x 10^{-13} g/7.8 x 10^{-16} L = 1.218 kg wet weight per litre of cell volume. It is unlikely that the assumed 2.2 kg wet weight per litre of cell volume in Schmid et al. (2004) is applicable in practice. The intracellular serine concentration, which is among the 14 amino acids (and tryptophan is not one of them), measured by Piperno and Oxender (1968), is 0.04 mmoles per kg wet weight. Serine concentration in wild *E. coli.* K12 strain is calculated as 0.04 mmoles/kg wet weight = 0.04 x 1.218 mmoles/L cell volume = 0.04872 mM.