Optimization of Recombinant Protein Production by *Streptomyces lividans* Host

by

Keyvan Nowruzi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Chemical Engineering

Waterloo, Ontario, Canada, 2010

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Keyvan Nowruzi

Abstract

Interleukin-3 is a cytokine, which acts on many target cells within the haemopoietic system, often in synergy with the other cytokines. *Streptomyces lividans* NCIMB 11416/IL3 p002 secreting human interleukin-3 was used as the host organism in this study of improving target protein production. *Streptomyces* also produces several proteases including extracellular endoprotease that truncate the N-terminus of the recombinant protein. Federal guidelines and regulations banning animal-derived medium components necessitate the refinement or redevelopment of industrial medium formulations. The development of a defined medium without animal products is most desirable for the production of pure and safe biological products. The objective of the proposed research was the development and application of engineering methodology for the development of a defined medium and the analysis and optimization of a bacterial bioprocess for recombinant protein production. The underlying hypothesis is that a significant improvement of target protein productivity is achievable by using appropriate optimization techniques. During the first phase of this study the task was to develop a systematic procedure for the design and optimization of a chemically defined medium. The study aimed at replacing casein peptone in conventional medium for *S. lividans* with essential amino acids and determining the optimum proportion of the amino acids. To accomplish this, starvation trials with growth limiting amino acids were performed to establish the baseline for the nutritional requirement. The starvation trials revealed that essential amino acids for growth and product formation are amongst the following eight amino acids: Arg, Asn, Asp, Glu, Leu, Met, Phe, and Thr. Following these preliminary experiments, a statistically based experimental method called mixture experiments along with distance-based multivariate analysis revealed that Asp, Leu, Met, and Phe were the essential amino acids. Then, another mixture experiment design known as simplex lattice design was performed and artificial neural networks were employed to obtain the optimum proportions of the essential amino acids. The optimal medium was found to be composed of 56% Asp, 5% Met, and 39% Phe. It was found in previous studies that in complex media, several types of protease are produced during fermentation. Using the

defined medium no proteolytic activity was detected in the fermentation broth.

The second optimization method was based on metabolic flux analysis. A comprehensive metabolic network was developed for *S. lividans*. The metabolic network included carbohydrate and amino acid metabolism in both anabolic and catabolic reactions. According to the experimental results, the time course of the fermentation was divided into two phases, Phase E1 and Phase E2. In the first phase amino acids were used as a nitrogen source and in the second phase ammonia was the nitrogen source for growth and product formation. The metabolic network was used to form a set of linear algebraic equations based on the stoichiometry of the reactions by assuming pseudo-steady state for intracellular metabolites. The metabolic flux model consisted of 62 intracellular metabolites and 91 biochemical reactions. Two different objective functions were considered for optimization: maximizing the specific growth rate and minimizing the redox equivalent. A linear programming approach was used for optimizing the objective functions. The proposed model was able to predict the specific growth rate very accurately with a maximum error of 10%. The oxygen uptake rate and carbon dioxide evolution rate were evaluated with maximum error of 27% and 35%, respectively. Sensitivity analysis revealed that amino acid uptake was the growth limiting flux during the Phase E1 of the fermentation. During Phase E2 the uptake rate of ammonia had a significant effect on the specific growth rate. Sensitivity analysis of the specific growth rate and redox potential with respect to the biomass components showed that any additional supply of biomass building blocks (amino acids, nucleotides) would not significantly affect the specific growth rate and redox potential production as well as the calculated flux pattern.

Acknowledgements

I would like to express my sincere gratitude and appreciation to:

- my supervisors, Professor Ali Elkamel, Professor Jeno M. Scharer, and Professor Murray Moo-Young, not only for their encouragement and invaluable advice throughout the course of this research and thesis writing, but also in making the University of Waterloo a Place I will definitely miss
- Professor Otman Basir, Professor Raymond Legge, Professor Eric Jervis, and Professor Frank Gu for serving as my committee members and providing me with constructive comments for this thesis
- the external examiner, Professor Argyrios Margaritis for his valuable comments
- Dr. Doug Cossar from Cangene Corp., for providing starting materials including the host organism and following up with insightful discussions
- Jana Otruba and Katherine Chiang for their time and generous assistance during the experimentation phase of this study
- Ralph Dikhout for his technical assistance with the operation of the HPLC
- members of Biochemical Engineering Research Group for their critical discussions and for creating a homelike working environment
- my friends for their support and encouragement

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Recombinant protein over-production

Modern biotechnology has become a major participant in global market with an annual turnover of more than 216.3 billion dollars in 2008 and the market is forecast to have a value of 305.7 billion dollars in 2013, an increase of 41.3% since 2008 (http://www.datamonitor.com). This industry produces biopharmaceuticals (antibiotics, vaccines, and monoclonal antibodies), yeast (for beer, wine and bread production), health-care products (vitamins), nutrients (amino acids, polysaccharides) fine chemicals and bulk chemicals (alcohol), by fermentation, bioconversion, and enzymatic processes. Substrates have been transformed into a wide variety of products through the metabolism of microorganisms. Nutrients are partly used for biomass and product synthesis and energy supply, but they are also being used for the synthesis of undesirable byproducts and energy dissipation in futile cycles. In many cases, optimization can provide significant economic benefits because these processes are characterized by significant price differences between reactants and products. Important aspects in the development and optimization of these processes are the conversion yields, the product purity and the product quality. These three aspects can be optimized both at the biochemical level and at the process

operation level (Parekh S. and Strobel R. J., 2000).

One of the most important parameters which affects growth, product formation and recovery in a bacterial fermentation is the medium composition. Therefore, the design of a fermentation medium is of critical importance during the development of an industrial fermentation. The development of defined medium without animal products is most desirable for production of pure and safe biological products. Traditionally, medium design and optimization was done by varying one factor while keeping the other factors at constant level. This strategy is simple and easy to implement although it may require a relatively large number of experiments and frequently fails to achieve the optimal conditions (Kalil S.J et al., 2000).

On the other hand, higher product yields can be achieved by selection and improvement of microbial strains: for example 100-fold to 1000-fold increases has been recorded for small metabolites and antibiotics (Demain A.L., 2000; Diez B. et al., 1997). Traditionally, this task has been carried out by a series of random targeted mutagenesis in a selective environment, or by addition of external genetic material followed by selection. These methods have been very successful in the past, but required lengthly trial and error experimentation. This scenario is changing due to the evolution of an approach called metabolic engineering (Bailey, 1991). Metabolic engineering is a science that combines the benefits of molecular biology, biochemistry, genetics, chemical engineering biotechnology, and mathematical modelling (Stephanopoulos G.N. and Vallino J.J., 1991). Metabolic engineering has two major components. The first is development of strategies for better understanding the structure of metabolic systems and intracellular enzymatic reactions. The second component is to apply the results of these strategies in selected organisms (Torres N.V. and Voit E.O., 2002).

In order to optimize the capacity of microbial metabolism and increase yields, the intracellular metabolite fluxes have to be quantified. The flux distribution in metabolic pathways determines the efficiency of the production process. Metabolic flux analysis which only requires the measurement of extracellular metabolites has been proposed as a means to determine the flow through primary metabolic pathways (Vallino J.J. and Stephanopoulos G., 1990; Varma

A. and Palsson B.O., 1994; Cannizzaro C. et al., 2004). It is theoretically possible to measure and control metabolic conversions on the level of single enzymatic reactions by metabolic flux analysis methods, with the measurement of only a limited number of extracellular metabolites (van Gulik W.M. and Heijnen J.J., 1995). Other benefits of this computational method is, defined medium design, elucidation of metabolic and toxicological effects, location of metabolic control, determination of maximum theoretical yields and a quantitative prediction of biochemical phenotypes from gene data banks (Varma A. and Palsson B.O., 1993a,b,c; Pallson B. O., 1997; Heijnen J.J. et al., 2004; Ihmels J. et al., 2004).

1.2 Research objectives

The main objective of this study was to develop and apply engineering strategies for the analysis and optimization of a bioprocess based on *Streptomyces lividans* producing recombinant human interleukin-3 (rHuIL-3). Special emphasis was placed on the effect of the fermentation medium components. To accomplish this objective the following experiments were performed:

- Replace the casein peptone in the complex medium with amino acid cocktail
- Monitor the concentration profiles of the amino acids and use these data as input for statistical optimization and mathematical modeling
- Search for the most significant amino acids in the medium for recombinant protein production using mixture experiments and distance-based multivariate analysis
- Apply response surface methodology and artificial neural network to determine the optimum level of the essential amino acids for maximum production of the protein
- Construct a mathematical model to simulate the biological system
- Define and optimize plausible objective functions for achieving recombinant protein production enhancement

- Perform batch fermentation to collect experimental data for metabolic flux analysis

It is believed that this research provides a theoretical sound framework that may be used for the optimization of recombinant protein production by other bacterial hosts.

1.3 Research approach

The primary aim of this research was to develop a mathematical model for improving the production of the targeted protein. Since there was no unique mathematical model capable of capturing all features of a biochemical system, statistical methods, artificial neural networks, and metabolic flux analysis were expected as promising endeavours for achieving the objective.

1.3.1 Mixture design analysis

Mixture design was applied in a sequential manner to investigate the effect of amino acids on rHuIL-3 production in shake flask cultures. First, starvation trials were carried out to narrow down the number of probable essential amino acids. Based on the results from the starvation trials, screening mixture design was performed and the collected data were analyzed by distance-based multivariate methods to rank the essential amino acids. Then, another mixture design was used and the data used as input to response surface methodology. Finally artificial neural networks were employed to find the optimum level of the essential amino acids. Fermentation experiments were performed at the predicted optimal level of the essential amino acids and the adequacy of the model to predict the behavior of the microorganism was verified.

1.3.2 Metabolic flux analysis (MFA)

A metabolic network was proposed based on the stoichiometry of bacteria. The basic metabolic pathways, Embden Mayerhof Parnas (EMP) pathway, pentosephosphate (PP) pathway, tricarboxylic acid (TCA) cycle and anaplerotic pathways as well as biosynthetic pathways leading

to amino acid synthesis were considered. Fermentations were run using the designed defined medium in 2 and 7 liter nominal volume bioreactors. The off-line extracellular analysis of key metabolites, biomass and rHuIl-3 were carried out throughout the fermentation. The accumulated data along with the stoichiometric metabolic network were used to construct a set of linear programming constraints. Two different objective function were chosen and the metabolic flux distribution was determined using GAMS (General Algebraic Modelling System) software. The logarithmic sensitivity analysis was used to specify the nutrients which had the highest effect on the biomass production. A uniform distribution was considered for biomass components and the effect of the variations of the biomass components on biomass production was investigated.

1.4 Thesis outline

This thesis consists of 6 chapters and is organized as follows:

Chapter 1 presents introduction, research objectives and research approach.

Chapter 2 provides a comprehensive literature review including the bacterial host, cellular metabolism, mixture design, distance-based multivariate analysis, artificial neural networks and metabolic flux analysis.

Chapter 3 presents the material and methods applied in this research including microorganism, stock culture preparation, medium components and analytical methods used for the analysis of the fermentation samples.

Chapter 4 describes the application of statistical methods and artificial neural networks for maximizing the biomass and product production.

Chapter 5 presents the construction of the metabolic network and metabolic flux distribution for *S. lividans*.

Chapter 6 summarizes the significant findings of this research and provides some recommendations for future research.

Chapter 2

Literature review

Since the early 1970s advances in genetics and molecular biology have given rise to the development and progress in recombinant DNA technology. The selection of a host for production of a biological molecule has had a substantial influence on isolation and purification protocols employed along with their respective cost implications. Undoubtedly, the most expolited host to express recombinant pharmaceutical products is *Escherichia coli*. However, its known limitations such as trapped products in inclusion bodies or in the periplasmic spaces, led to the exploration of other possible hosts.

2.1 *Streptomyces* as a host for recombinant protein production

Streptomyces strains are non-motile, gram-positive, plasmid containing, non-pathogenic, and aerobic bacteria. They are widely distributed in nature. Their primary niche is in soil ecosystems, but they also occur in freshwater and marine environments, in salt marsh areas, in fodders, and related materials, and in the air (Ayadi D.Z. et al., 2007). Under laboratory conditions, they have the tendency to form pellets in liquid as has been observed in shake flask cultures. Most *Streptomycetes* grow optimally at mesophilic temperatures ranging 25 - 35°C . Some species, however, are known to grow at temperatures well above the mesophilic range

as high as 50 - 60°C (Demain A.L. and Solomon, 1985). *Streptomycetes* can be grouped into two general categories based on pH requirements for growth. Acidophilic streptomycetes grow in the pH range of 3.5 - 6.5 with optima near 5, while neutrophilic *Streptomycetes* grow from 5.0 - 9.0, with optima around pH 7.0. *Streptomycetes* were historically regarded as fungi, later as intermediates between fungi and bacteria, before being classified as eubacteria. They are sensitive to antibacterial antibiotics and lack a nuclear membrane, which by definition, classifies them as prokaryotes. Nutritionally, *Streptomycetes* are classified as chemoorganotrophs. In general, they do not exhibit any special nutritional requirements, most being able to grow on media consisting of inorganic salts and glucose. They produce proteolytic exoenzymes, and readily grow on complex proteins. Under certain conditions, *Streptomycetes* undergo complex morphological transformation to form spores. When encountering a suitable environment the spore germinate to produce a germ tube. The germ tube then develops into a branched mycelial network, which may penetrate and become firmly attached to a solid supporting matrix. From this mycelial network arise sporogenic hyphae, which grow up into the air and differentiate to form chains of spores (Figure 2.1). The formation of aerial hyphae often coincides with the production of antibiotics, providing protection against potential pathogens. In addition to antibiotics, *Streptomyces* are also known to secret appreciable amounts of proteinaceous enzyme inhibitors. The organism *Streptomyces lividans 66* recognizes heterologous prokaryotic promoters and is almost the sole species of choice as a host for recombinant protein production. Two additional advantages of *Streptomyces lividans*, as a host for heterologous protein production are the ability to form proper disulfide bonds, and the natural ability to synthesize a proteinase inhibitor, LEP-10 (Demain A.L. and Solomon, 1985). To date, the genes coding for polysaccharase of several *Streptomycetes* have been cloned and expressed in *Streptomyces lividans*. Successful recombinant proteins of prokaryotic origin expressed in *Streptomyces lividans* are typically produced in large quantities and fully secreted. Secretion of prokaryotic proteins can lead to levels of several grams per liter of culture broth. The yields of eukaryotic proteins are usually lower. An example is the production of T-cell receptor CD4, which has been produced in *Streptomyces lividans* fermentation at over 300 mg/l of culture fluid, which

is reportedly higher than in *Escherichia coli* or *Chinese Hamster Ovary* (CHO) cells (Demain A.L. and Solomon, 1985). Filtration or centrifugation can easily separate the target product from the biomass. In addition, the recombinant target protein is frequently the predominant protein in the culture supernatant. This further translates to significant cost saving in terms of product recovery and purification.

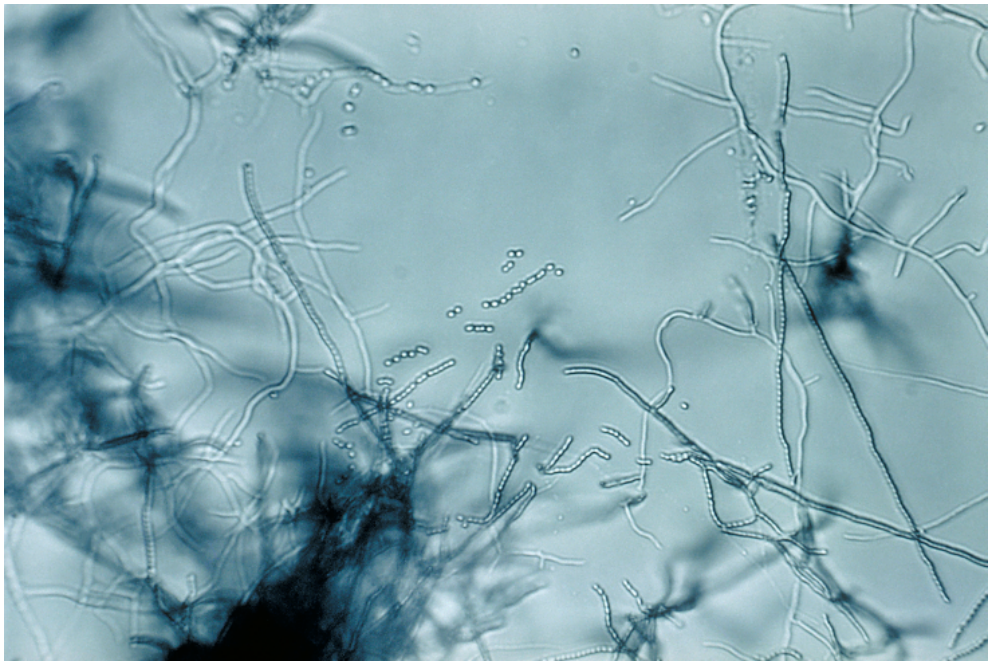

Figure 2.1: Bacterial host *Streptomyces sp.* (http://en.wikipedia.org/wiki/Streptomyces)

2.2 Interleukin-3 (IL-3)

IL-3 is a secreted glycoprotein, the peptide core comprising 140 amino acid residues in mice and 133 residues in the humans. The molecular weight of the polypeptide core is about 14,000 daltons. Like IL-2 and IL-4, IL-3 is a monomer with two intermolecular disulphide bonds. Il-3 stimulates the proliferation, differentiation, and survival of pluripotential haemopoietic cells and haemopieotic proginetor cells and their mature progeny of multiple cell lineages. Il-3

is produced and secreted primarily in response to immunological stimuli by T lymphocytes activated by specific antigens. The potential clinical utility of IL-3 lies in its ability to enhance the recovery of haemopoiesis following cytotoxic cancer therapy or bone-marrow transplantation. IL-3 also has an important role in diseases characterized by increases in mast cells and basophils, and IL-3 antagonists may provide a new approach to treatment of allergies and asthma. (Nicola N.A., 1994).

2.3 Review of cellular metabolism

Formulation of the stoichiometery of metabolic pathways is the basis for the quantitative treatment of cellular metabolism. This requires an appreciation of some basic biochemical processes along with an overview of different pathways normally present in living cells. Therefore, the basics of glycolysis, tricarboxylic acid, and anaplerotic pathways as well as amino acid biosynthesis are reviewed.

2.3.1 Glycolysis

Glycolysis is the sum total of all biochemical reactions by which glucose is converted into pyruvate. This can be accomplished by more than one pathway and the process takes place in the cytoplasm. The most frequently encountered pathways are:

1. Embden-Meyerhof-Parnas pathway (EMP).
2. Pentose phosphate pathway (PP).
3. Etner-Doudoroff pathway (ED).

The common entrance of sugars to the glycolytic pathways is through the three hexose monophosphates, glucose-1-phosphate (G1P), glucose-6-phosphate (G6P), and fructose-6-phosphate (F6P).

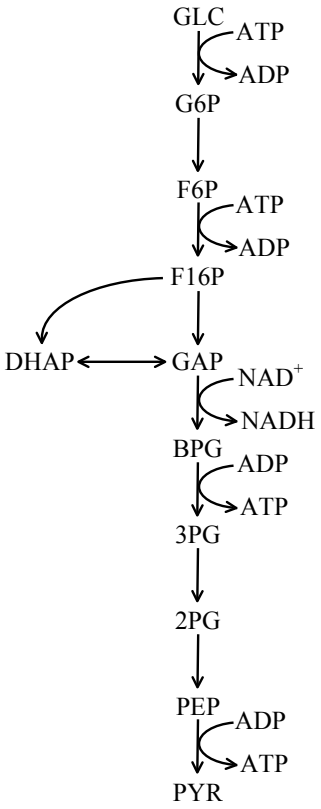

Figure 2.2: Overview of EMP pathway

In the EMP pathway, 1 mol glucose is converted into 2 mol pyruvate and it can be divided into two major parts (Figure 2.2). The first part includes six carbon components and does

not involve any oxidation-reduction reactions. The reaction sequence in this part comprises two phosphorylation reactions that lead to the production fructose-1,6-diphosphate. Then aldolase catalyzes the next cleaving reaction, and two three-carbon molecules, glyceraldehyde-3-phosphate and dihydroxy-acetone phosphate are formed from fructose-1,6-diphosphate. All components in the second part are three-carbon intermediates. The first oxidation reaction occurs in the second part of the EMP pathway where glyceraldehyde-3-phosphate is converted to 1,3-diphosphoglycerate. In this reaction, the coenzyme is reduced to NADH. The final product of this part is pyruvate which is a key intermediate in metabolism. Thus, the overall stoichiometry for the conversion of glucose to pyruvate in the EMP is:

$$GLC + 2\,ADP + 2\;P + 2\,NAD^{+} \rightarrow 2\,PYR + 2\,ATP + 2\,NADH + 2\,H_2O + 2\,H^{+} \qquad (2.1)$$

2.3.2 The pentose phosphate pathway

Many of the biosynthetic reactions require NADPH in addition to ATP. Despite their close chemical resemblance NADPH and NADH are not metabolically interchangable. Cells normally maintain their NAD^{+}/NADH ratio near 1000, which favours metabolite oxidation, while keeping their $NADP^{+}$/NADPH ratio near 0.01, which favors metabolite reduction (Voet D. and Voet J.G., 1995). NADPH is generated by the oxidation of glucose-6-phosphate via an alternative to glycoloysis, the pentose phosphate pathway (Figure 2.3). In PP pathway, glucose-6-phosphate is oxidized to 6-phosphogluconate, which is further converted to ribulose-5-phosphate and carbon dioxide. In each of these reactions, 1 mol of NADPH is formed per mole of glucose-6-phosphate entering the pathway. In the subsequent steps, ribulose-5-phosphate is converted into ribose-5-phosphate or erythrose-4-phosphate, which are both precursors for the biosynthesis of aromatic amino acids and nucleotides. In a different sequence of reactions, ribulose-5-phosphate may also be converted back to fructose-6-phosphate and glyceraldehyde-3-phosphate, thus re-entering the EMP pathway.

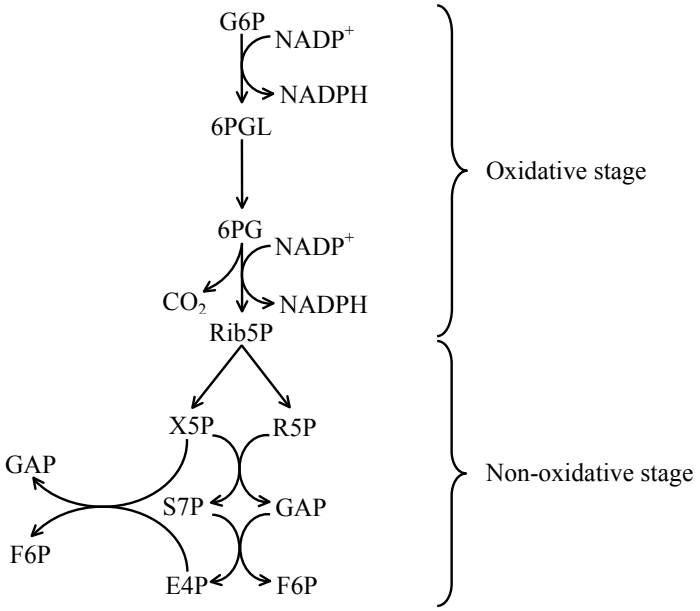

Figure 2.3: The pentose phosphate pathway

The overall stoichiometry of the PP depends on the extent to which carbon entering the PP pathway is recycled back into the EMP pathway and oxidized to form carbon dioxide

with simultaneous production of reducing power in the form of NADPH, or consumed for the formation biosynthetic precursors such as five-carbon sugars for ribonucleotide synthesis. For this reason, PP pathway has been recognized to serve an oxidative as well as an anaplerotic function, each described by the following overall stoichiometries:

Anaplerotic PP function:

$$5\,G6P + 5\,ATP \rightarrow 6\,ribose - 5 - P + 5\,ADP + 4 \sim P \tag{2.2}$$

Oxidative PP function:

$$G6P + 12\,NADP^{+} + 7\,H_2O \rightarrow 12\,NADPH + 12\,H^{+} + 6\,CO_2 + \sim P \tag{2.3}$$

The functions of PP cycle are as follows (Dagley S. and Nicholson D.E., 1970):

1. Provides reduced NADP for synthetic purposes such as the synthesis of fatty acids.
2. Provides energy. If all the reacted NADP is completely oxidized and the process is 100% efficient, 38 molecules of ATP are produced for each molecule of glucose oxidized.
3. Provides a source of pentoses for nucleotides.
4. Makes possible the interconversion of hexoses and pentoses.

The Entner-Doudorof pathway is common in *Pseudomonads*, but there is no evidence of any *Streptomycete* using it (Kieser T. et al., 2000). Three intermediates of the EMP pathway, glyceraldehyde-3-phosphate, 3-phosphoglycerate, and phosphoenolpyruvate and two intermediates of the PP pathway, ribose-5-phosphate and erythrose-4-phosphate serve as precursor metabolites for biosynthesis of amino acids and nucleic acids. The relative flux through the two glycolytic pathways depends on the requirements of free energy, reducing power in the form of NADH and NADPH, and the precursor metabolites. The major control points of the EMP and PP pathways are at the entrance to pathways, i.e., at the phophofructokinase and at the glucose-6-phosphate dehydrogenase reactions (Zubay G.L., 1998). Glucose-6-phosphate dehydrogenase is regulated by the NADPH/$NADP^{+}$ ratio, whereas phosphofructokinase, which is

a complex allosteric enzyme, has several effectors. In *Saccharomces cervisiae* it is activated by AMP, ammonia, phosphate, and fructose-2,6-bisphosphate, and it is inhibited by ATP. The cofactors NADH and NADPH serve two different purposes in cellular metabolism. In aerobes, NADH is mainly involved in the generation of Gibbs free energy through the oxidative phposphorylation reaction, whereas NADPH is mainly used in the biosynthesis of building blocks. Thus, NADH serves as a substrate in the fueling reactions and NADPH serves as a substrate in the biosynthesis reactions and the ratios NAD^+/NADH and NADPH/$NADP^+$ are therefore regulated at different levels.

2.3.3 Tricarboxylic acid (TCA) cycle

TCA cycle, also called Krebs cycle or citric acid cycle, is the second stage of cellular respiration, by which living cells break down organic fuel molecules in the presence of oxygen to harvest the energy they need to grow and divide. This metabolic process occurs in most plants, animals, fungi, and many bacteria. In all eukaryotic organisms the TCA cycle is carried out in the mitochondrion. The TCA cycle plays a central role in the catabolism of organic fuel molecules i.e., glucose and some other sugars, fatty acids, and some amino acids. Before these rather large molecules can enter the TCA cycle they must be degraded into a two-carbon compound coupled to an organic carrier, called acetyl coenzyme A (acetyl CoA). Once fed into the TCA cycle, acetyl CoA is eventually converted into carbon dioxide and energy. The TCA cycle consists of eight steps catalyzed by eight different enzymes (Figure 2.4):

1. Acetyl CoA reacts with the compound oxaloacetate to form citrate and to release coenzyme A (CoA).
2. In a succession of reactions, citrate is rearranged to form isocitrate.
3. Isocitrate loses a molecule of carbon dioxide and undergoes oxidation to form alpha-ketoglutarate.
4. Alpha-ketoglutarate is decarboxylated and oxidized to form succinyl CoA.

5. Succinyl CoA is enzymatically converted to succinate.

6. Succinate is oxidized to fumarate.

7. Fumarate is hydrated to produce malate; and, to end the cycle.

8. Malate is oxidized to oxaloacetate.

Each complete turn of the cycle results in the regeneration of oxaloacetate and the formation of two molecules of carbon dioxide. Energy is produced in a number of steps in this cycle of reactions. In step 5, one molecule of ATP is produced. However, most of the energy obtained from the TCA cycle is stored in the compounds NAD^+ and FAD and converted later to ATP. Energy transfers occur through the relay of electrons from one substance to another, a process carried out through the chemical reactions known as oxidation and reduction, or redox reactions. The major regulatory sites of the TCA cycle are at the citrate synthase, isocitrate oxidoreductase (dehydrogenase), and α-ketoglutarate oxidoreductase (dehydrogenase). The activity of all three enzymes is favoured by the low level of the NAD^+/NADH ratio, while the isocitrate dehydrogenase is strongly regulated by this ratio (Stephanopoulos G.N. et al., 1998). The overall stoichiometery for the complete oxidation of pyruvate in the TCA cycle is:

$$\begin{aligned} &PYR + 3\,H_2O + GDP + 2 \sim P + 4\,NAD^+ + FAD \rightarrow \\ &3\,CO_2 + GTP + 4\ NADH + FADH_2 + 4\,H^+ \end{aligned} \tag{2.4}$$

2.3.4 Anaplerotic pathways

There is no net synthesis of α-ketoglutarate and oxaloacetate in TCA cycle and these two organic acids serve as precursor metabolites for the biosynthesis of amino acids and nucleotides. Therefore, removal of these organic acids for other cellular functions must be compensated for by other means. Reaction sequences that fulfill this role are called anaplerotic (meaning filling up in Greek) pathways. The anaplerotic pathways include the following (Stephanopoulos G.N. et al., 1998):

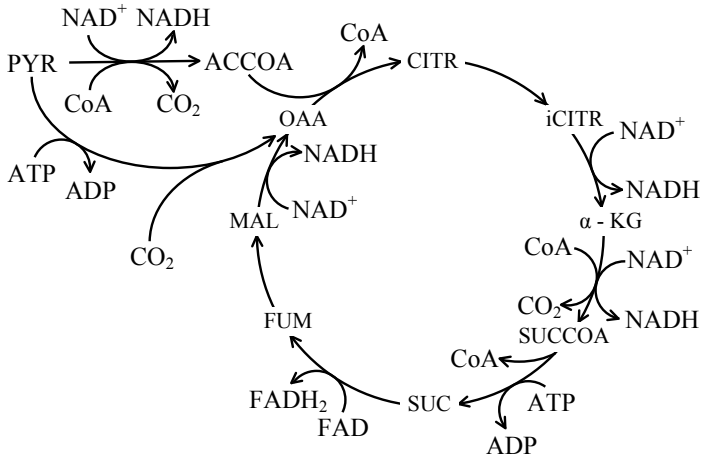

Figure 2.4: TCA cycle

1. Carboxylation of pyruvate to form oxaloacetate by pyruvate carboxylase.

2. Carboxylation of phosphoenolpyruvate by PEP carboxylase.

3. Oxidation of malate to pyruvate by the malic enzyme.

4. The glycoxylate cycle.

2.3.5 Biosynthetic reaction

The number of building blocks, coenzymes, and prosthetic groups needed for cellular synthesis is about 75 -100, and these are all synthesized from approximately 12 precursor metabolites by reactions that employ energy, reducing power, and sources of nitrogen, sulfur, and single carbon units (Neidhardt F.C. et al., 1990). Biosynthesis pathways differ markedly in complexity, some are linear, others branched or interconnected. An inspection of the metabolic routes leading to the individual amino acids reveals that common pathways are frequently employed. Figure 2.5 summarizes the family and precursors for amino acid biosynthesis. Biosynthesis of the aromatic amino acids is very complicated, and a completely independent pathway of the other amino acids is employed for histidine biosynthesis (Gottschalk G., 1979).

2.4 Mixture experiments

Statistical experimental design is a well-established concept for the planning and execution of the informative experiments (Montgomery D.C. and Runger G.C., 2006). The most common approach, factorial design, has been successfully used for the optimization of fermentation media. A two-level factorial design requires 2^n experiments if n factors have to be investigated. With $n = 20$, this would lead to 1,048,576 experiments, which is a prohibitively large number. For this reason, an alternative statistical design, known as the mixture design was employed. This experimental design application concerns the preparation and modification of mixtures, in which the response is assumed to depend only on the relative proportions of the ingredients present in the mixture and not on the amount of the mixture (Cornell J.A., 2002). In mixture

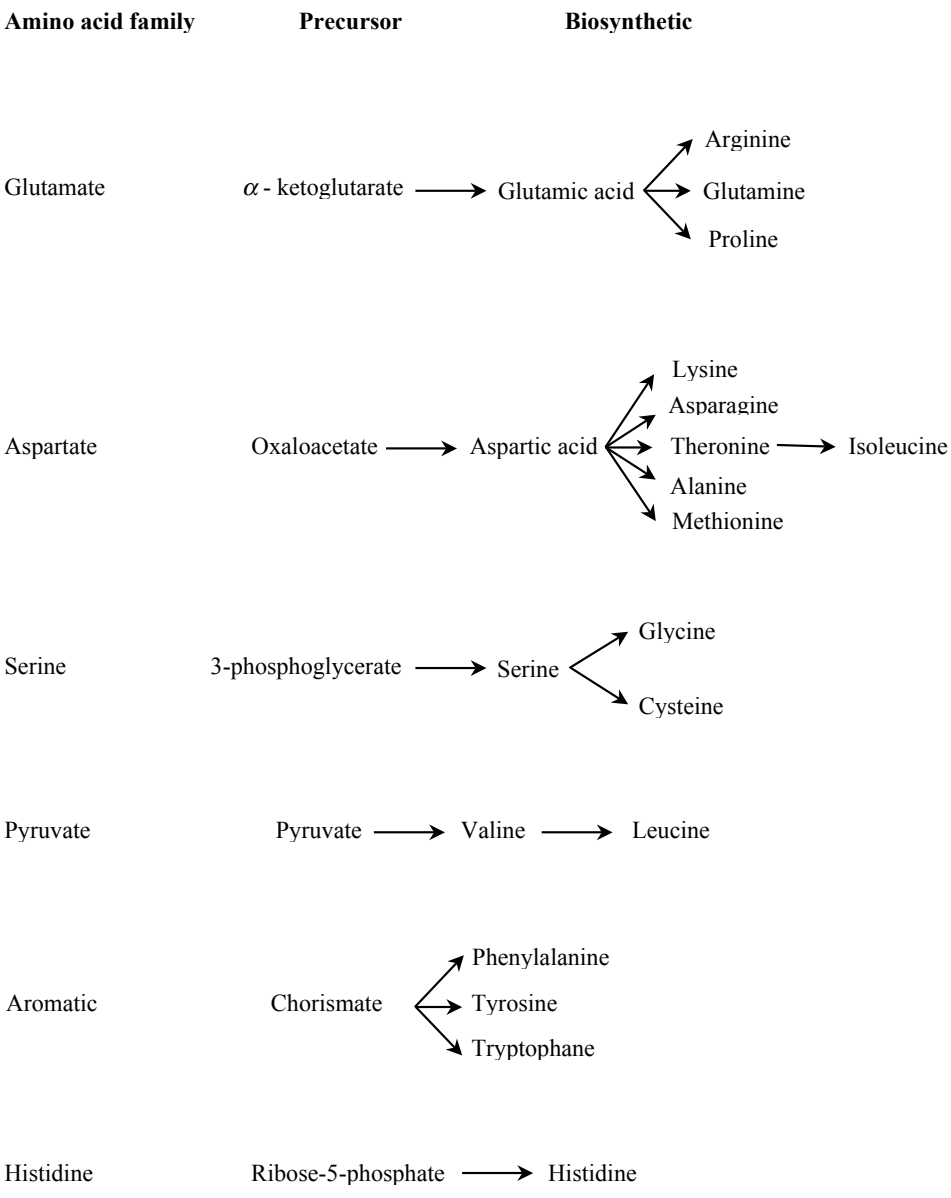

Figure 2.5: Overview of amino acid biosynthesis in bacteria

experiments, the factors are the components of a mixture, and consequently their levels are not independent. Depending on the experimental objective, the response surface over the simplex region may be mapped by different mixture designs. A {q, m} simplex lattice design for q components consists of points defined by the following coordinate settings: the proportions assumed by each component take the equally spaced values from 0 to 1,

$$x_i = 0, \frac{1}{m}, \frac{2}{m}, ..., 1; \qquad i = 1, 2, ..., q \tag{2.5}$$

and the {q, m} simplex lattice consists of all possible combinations of the components where the proportions from Eq. 2.5 are used. In general, the number of points in a {q, m} simplex lattice design is

$$N = \frac{(q+m-1)!}{m!(q-1)!} \tag{2.6}$$

An alternative to the simplex lattice design is the simplex centroid design. In a q-component simplex centroid design, there are $2^q - 1$ points, corresponding to the q permutations of (1, 0, ..., 0) , the $\begin{pmatrix} q \\ 2 \end{pmatrix}$ permutations of $(\frac{1}{2}, \frac{1}{2}, 0, ..., 0)$ the $\begin{pmatrix} q \\ 3 \end{pmatrix}$ permutations of $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, ..., 0)$ and the overall centroid $(\frac{1}{q}, \frac{1}{q}, ..., \frac{1}{q})$. Fig. 2.6 shows a simplex lattice and simplex centroid designs. For more information, the interested reader should consult Cornell (Cornell J.A., 2002) and Myers and Montgomery (Myers R.H. and Montgomery D.C., 2003).

Mixture models differ from usual polynomials employed in response surface work because of the constraint $\sum_{i=0}^{q} x_i = 1$. The standard forms of mixture models that are in widespread use are:

Linear:

$$E(y) = \sum_{i=1}^{q} \beta_i x_i \tag{2.7}$$

Quadratic:

$$E(y) = \sum_{i=1}^{q} \beta_i x_i + \sum \sum_{i<j}^{q} \beta_{ij} x_i x_j \tag{2.8}$$

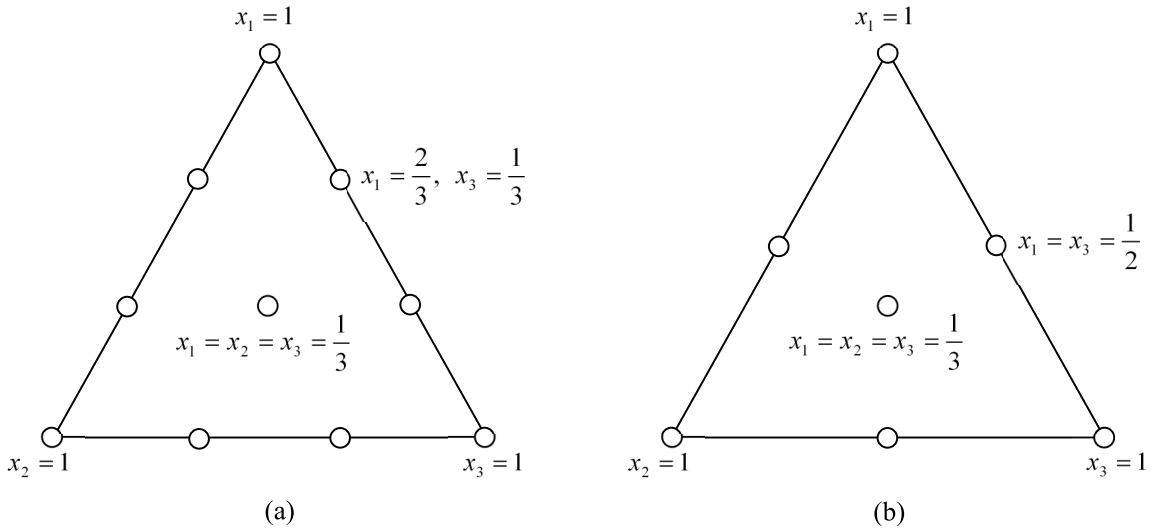

Figure 2.6: (a) {3, 3} simplex lattice design, (b) simplex centroid design

Full Cubic:

$$E(y) = \sum_{i=1}^{q} \beta_i x_i + \sum \sum_{i<j}^{q} \beta_{ij} x_i x_j + \sum \sum_{i<j}^{q} \delta_{ij} x_i x_j (x_i - x_j) + \sum \sum \sum_{i,j,k}^{q} \beta_{ijk} x_i x_j x_k \tag{2.9}$$

Special Cubic:

$$E(y) = \sum_{i=1}^{q} \beta_i x_i + \sum \sum_{i<j}^{q} \beta_{ij} x_i x_j + \sum \sum \sum_{i<j<k}^{q} \beta_{ijk} x_i x_j x_k \tag{2.10}$$

The terms in these models have relatively simple interpretations. In Eqs. 2.7 through 2.10, the parameter β_i represents the expected response to the pure blend $x_i = 1$ and $x_j = 0$ when $j \neq i$. The portion $\sum_{i=1}^{q} \beta_i x_i$ is called the linear blending portion. When there is a curvature arising from nonlinear blending between component pairs, the parameter β_{ij} represents either synergistic or antagonistic blending. Higher-order terms are frequently necessary in mixture models because of two reasons. First, the phenomena studied may be complex and secondly, the experimental region is frequently the entire operability region and is therefore large, requiring an elaborate model. When the experimental objective is screening, the design supporting a linear model is useful, but for optimization designs supporting quadratic or special cubic

models are relevant (Eriksson L. et al., 1998). Screening experiments are used in the beginning of the experimental work for investigating a large number of components to screen out the unimportant components or single out the important ones. To do so, it is necessary to know how to measure the effect of individual components. According to Cornell (Cornell J.A., 2002) the effect of component i on the response is the change in the value of the response resulting from a change in the proportion of component i while holding the relative proportions of the other components constant. The largest change that can be made in the proportion of x_i is one unit. If the change in x_i is made along the x_i-axis, which is the imaginary line extending from the base point x_i=0, $x_j=\frac{1}{q-1}$ for all j≠i, to the vertex where $x_i = 1$, $x_j = 0$ for all $j \neq i$, the proportions x_j, $j \neq i$ of the other components will remain constant. In Fig. 2.7 the three x_i-axis inside the three-component triangle are shown. The values of the proportions of each of the other $q-1$ components along the x_i-axis are $x_j = \frac{(1-x_i)}{(q-1)}$, for $j \neq i$. To measure the effect of all q components simultaneously, using Cox's model (Cornell J.A., 2002), the data should be collected at points located on all the axes to permit the estimation of all of the q parameters in the first-degree model.

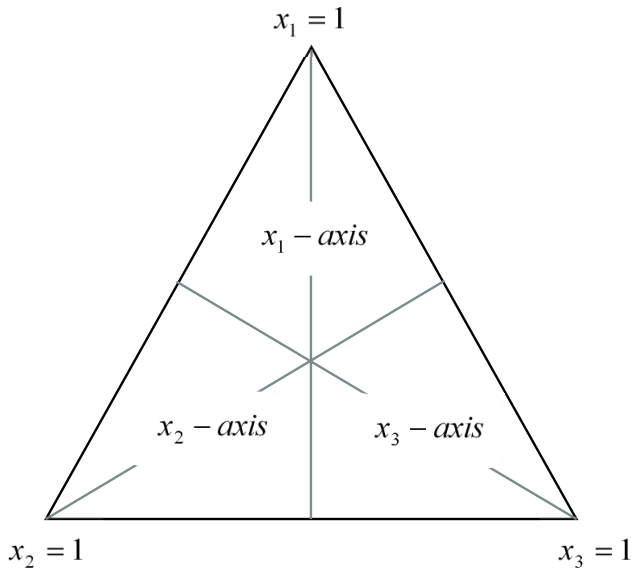

Figure 2.7: The x_i-axes, i = 1, 2, and 3

2.5 Distance-based multivariate analysis

Tests for interactions among factors form a very important component of multifactorial experiments. A significant interaction between two factors indicates that the effects of the one of the factors are not consistent across the levels of the other factor. The individual importance of single factors may be difficult or even impossible to isolate or interpret in a system where a number of interactions are present (Hilborn R. and Stearns S.C., 1982). Several nonparametric multivariate statistical methods are available (Clarke K.R., 1993; Mantel N. and Valand R.S., 1970) which can be used in an analysis-of-variance approach, but these methods have the drawback of not generally allowing tests of multivariate interactions between factors in an ANOVA design. In particular, these methods are not capable of differentiating components of multivariate variation in structured models. The reason for this is twofold (Legendre P. and Anderson M.J., 1999). First, the distance used may not be metric, so a linear model cannot be applied directly. And second, the above methods rely on permutations based on multivariate normality. Distance-based redundancy analysis (db-RDA) which is described by Legendre and Anderson (Legendre P. and Anderson M.J., 1999) is a comparatively new approach which has been developed for testing multispecies responses in multifactorial ecological experiments. This method is applicable to situations where the test is based on a distance measure of the researcher's choice. The special advantages of db-RDA are:

1. it can be used with distance measures that are non-Euclidean,
2. it can be used to test interaction terms, or any other term in a structured ANOVA model,
3. it uses nonparametric permutation methods which do not rely on assumptions of multivariate normality.

Steps in the procedure include:

1. calculating a matrix of distances among replicates using a distance measure of choice,

2. determining the principal coordinates, which preserve the distances,

3. creating a matrix of dummy variables corresponding to the design of the experiment

4. analyzing the relationship between the principal coordinates and dummy variables using RDA,

5. implementing a test by permutation for particular statistics corresponding to the particular terms in the model.

For a more detailed discussion one may consult Legendre and Anderson (Legendre P. and Anderson M.J., 1999) and McArdle and Anderson (McArdle B.H. and Anderson M.J., 2001). A computer program called DISTLM *forward* (Anderson M.J., http://www.stat.auckland.ac.nz/∼mja/Programs.htm (2003; McArdle B.H. and Anderson M.J., 2001) that does a multivariate multiple regression on the basis of any distance measure and does a forward selection of the predictor variables with the tests by permutations was used for the analysis of our experimental data. The program uses the proportion of the total sum of squares that is explained by the individual variable as the criterion for the forward selection. Output from the program includes the results of the marginal tests individually, ignoring other variables followed by the results of forward selection procedure with the conditional tests. Also information on the correlation among all pairs of explanatory variables, which provides a further check on issues of multi-co-linearity is included in the output file.

2.6 Artificial neural networks

Artificial neural networks are so named because they can be used to simulate at least partially the behavior of brain and biological neurons and learn by trial and error. The goal of a neural network is to map a set of input patterns onto a corresponding set of output patterns. Networks are first subjected to sets of input and output correspondences for a given system. After a sufficient number of training iterations, the network learns the pattern in the data fed

to it. The neural network then applies what it has learned to a new input pattern to predict the appropriate output. Neural networks are extensively interconnected parallel structures containing simple processing elements known as neurons or nodes shown in Fig. 2.8. Neurons are arranged in parallel layers, with each neuron forming a weighted connection to all layers. There is an input layer where data are represented to the neural network, and an output layer that holds the response of the network to the input. In between lies at least one layer known as a hidden layer, that enables neural networks to compute complicated associations between input and output patterns. According to universal approximation theory, a network with a single hidden layer with a sufficiently large number of neurons can map any input to any output, with an arbitrary degree of accuracy. Fig. 2.9 shows a typical three-layered neural network with N input neurons, H hidden neurons and M output neurons. The number of input and output neurons is determined by the nature of the problem. The number of neurons in the hidden layer is often determined by the required accuracy, and is therefore a parameter in formulating a neural network model. Each neuron receives an input vector, $\mathbf{X}$,

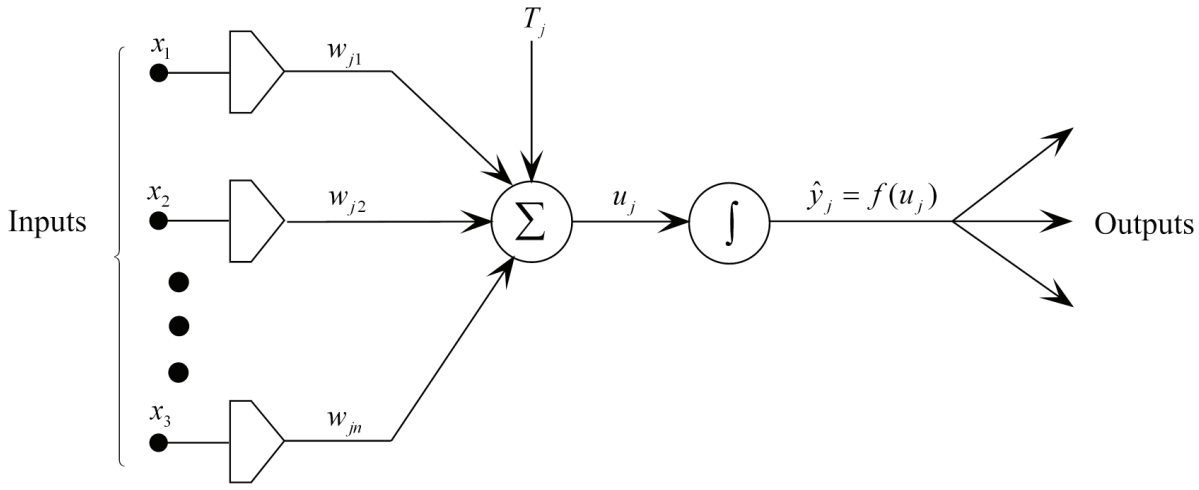

Figure 2.8: A typical neuron j in a layer

with components X_i (i = 1 to N). The output of each neuron is being determined by several factors other than the inputs. One of those factors is the weight factor, x_{ij} for the i^{th} input, X_i, corresponding to the j^{th} node. The other important factor determining the output of a

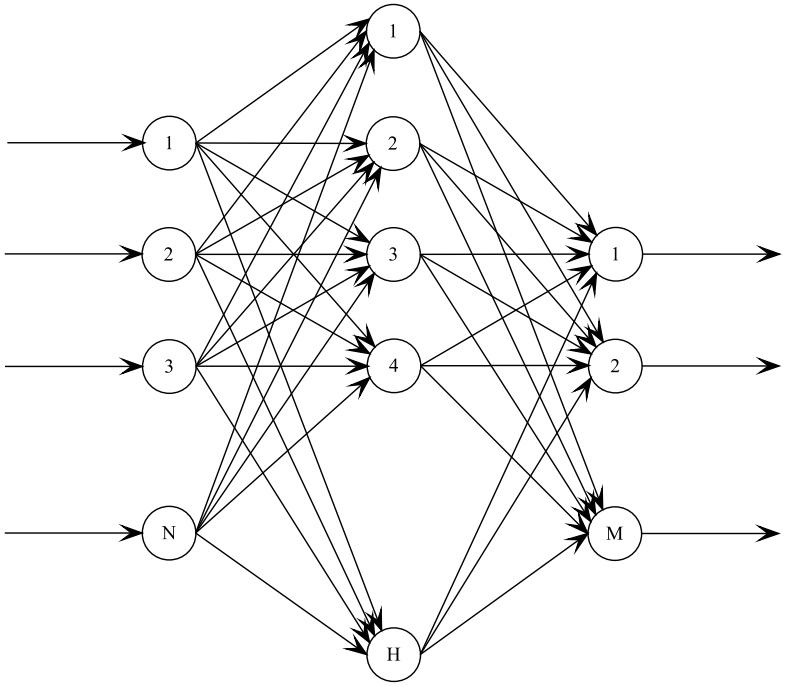

Figure 2.9: A one hidden layer feed forward neural network

neuron is the bias denoted by T_j for the j^{th} neuron. The final factor governing the output of a neuron is the transfer function. The most commonly used transfer function is the S-shaped logistic sigmoid function, which is continuously differentiable, monotonic, symmetric and bounded between 0 and 1, i.e.,

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2.11}$$

Another useful transfer function is the hyperbolic tangent function, which is differentiable, continuous and bounded between -1 and 1:

$$f(x) = \tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \tag{2.12}$$

The hyperbolic tangent transfer function outperforms sigmoid transfer function in prediction networks, and this is because of its greater slope compared to sigmoid transfer function (Baughman D.R. and Liu Y.A., 1995). The input-output data should be normalized so that they are in the same range of the transfer function used. The normalization of the inputs also

avoids overflows due to very large or very small weight factors. Data are normalized using:

$$\widehat{X} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{2.13}$$

where $\widehat{X}$ is the normalized value and X_{min} and X_{max} are the minimum and maximum values of $\mathbf{X}$, respectively. The output from neuron j, in the hidden and the output layers is given by Eq. 2.14.

$$\widehat{y}_j = f(u_j) = f(\sum_{i=1}^{n}(w_{ji}\widehat{X}_i) + T_j) \tag{2.14}$$

Neural network training is an optimization process, in which an error function is minimized by adjusting the weight factors and the bias. The error function to be minimized is the individual sum-squared error between the actual output values and the predicted values, which is given by (Elkamel A. et al., 2001):

$$E_i = \sum_{i=1}^{M}(\widehat{y}_{ji} - y_{ji})^2 \tag{2.15}$$

where $\widehat{y}_{ji}$ and y_{ji} are the i^{th} measured and predicted network outputs that correspond to the i^{th} input, respectively. Therefore, the training process requires a forward pass to calculate an output and a backward pass to update the weight factors and the bias. One pass through the set of training patterns along with updating the weight factors and the bias is called a cycle or epoch (Goh A.T.C., 1995). The details of the algorithm for adjusting the weight factors and bias to minimize the sum-squared error are described in Baughman and Liu (Baughman D.R. and Liu Y.A., 1995). The development of a neural network model consists of two steps. The first step is a training stage, where the network is subjected to a training set of input-output patterns. The second step is a testing stage, where the performance of the network is tested on patterns that have not been seen by the network during the training stage. The contribution of each predictor variable on the response could be determined by variable importance measure, VI, according to the procedure developed by Garson (Garson G. D., 1991). The procedure essentially involves partitioning of the weight factors of each hidden neuron into components associated with each input neuron. The equation proposed by Garson (Garson G. D., 1991)

for determining the variable importance measure of an input is given by:

$$VI_{Xp} = \frac{\sum_{j=1}^{n_H}(\frac{|I|_{P_j}}{\sum_{k=1}^{n_P}|I|_{P_{j,k}}}|O|_j)}{\sum_{i=1}^{n_P}(\sum_{i=1}^{n_P}\frac{|I|_{P_{i,j}}}{\sum_{k=1}^{n_P}|I|_{P_{i,j,k}}}|O|_j)} \tag{2.16}$$

where n_P is the number of input variables, n_H the number of neurons in hidden layer,$|I|_{P_j}$ is the absolute value of the hidden layer weight factor corresponding to the P^{th} input variable in j^{th} neuron in hidden layer, and $|O|_j$ is the absolute value of the output layer weight factor corresponding to the j^{th} neuron in hidden layer. For more details on the Garson's methodology consult Glorfeld (Glorfeld L.W., 1996). Neural networks have been used in section 4.5.2 as an alternative to a regression model for optimizing the defined medium.

2.7 Metabolic network modelling

In 1991 Bailey discussed the emergence of a new science called metabolic engineering, which he defined as the improvement of cellular activities by manipulations of enzymatic, transport, and regulatory functions of the cell with the use of recombinant DNA technology (Nielsen J., 2001). Metabolic flux analysis, which is based on intracellular metabolite balancing, is a powerful methodology for the determination of metabolic fluxes along the pathways (Schuster S.and Klamt S. et al., 2002; Stephanopoulos G.N. et al., 1998). In this approach, the intracellular fluxes are calculated by using a stoichiometric model (metabolic pathway map) for the major intracellular reactions and applying mass balances around intracellular metabolites. No kinetic or regulatory information is necessary to build a metabolic network. A set of measured extracellular fluxes, typically uptake rates of substrates and secretion rates of metabolites, is used as input to the calculations. The outcome of flux calculation is a metabolic flux map showing a diagram of the biochemical reactions included in the calculations along with an estimate of the steady state rate, the flux, at which each reaction in the diagram occurs.

2.7.1 Metabolic flux analysis

Metabolic flux analysis is a powerful methodology for the determination of the metabolic pathway fluxes. Intracellular fluxes are calculated by using a stoichiometric model for the major intracellular reactions and applying a mass balance around intracellular metabolites. No kinetic or regulatory information is necessary to build the metabolic network. The pathway fluxes are defined as the rate at which input metabolites are processed to form output metabolites.

For a metabolic network that contains n metabolites and m metabolic fluxes, assuming pseudo-steady state condition and negligible dilution effects for growth, the material balance is given as follows:

$$\mathbf{S} \cdot \boldsymbol{\nu} = \mathbf{R} \tag{2.17}$$

where $\mathbf{S}$ is an $m \times n$ stoichiometric matrix $\boldsymbol{\nu}$ is the vector of m metabolites and $\mathbf{R}$ is the vector of the uptake and production rates of all the intracellular and extracellular metabolites. This equation simply states that over long time periods, the formation fluxes of a metabolite must be balanced by the degradation fluxes and is the basis for determining the unknown internal fluxes $\boldsymbol{\nu}$ (Lee S.U. and Papoutsakis E.T., 1999).

To determine the metabolic capabilities of a defined metabolic network one needs to solve Eq. 2.17. Three fundamentally different situations are may be encountered. Mathematically, these situations are as follows:

1. Over-determined system
2. Determined system
3. Under-determined system

Over-determined system

In over determined systems the number of measurements available is more than the number of equations. The redundancy of the measurements can be used to calculate the rates of non-measured metabolites, increase the accuracy of the available measurements through the application of a least square calculation and identify the main sources of measurement errors (Stephanopoulos G.N. et al., 1998; Wittmann C. et al., 2004).

Determined systems

If the number of fluxes is equal to the number of metabolites or material balances (i.e. equations) the system is a called determined system and the solution to Eq. 2.17 is unique and is easily obtained by standard methods for solving linear equations.

Under-determined system

If the number of metabolic fluxes is greater than the number of mass balances the system is called under-determined and the number of feasible solutions to Eq. 2.17 is infinite but all the solutions lie in a restricted region defined as the metabolic genotype of a given organism (Varma A. and Palsson B.O., 1994). This is the most commonly encountered situation. In this case linear programming could be used to determine the intracellular flux distributions, provided that a suitable objective function can be specified. With this approach it is possible to obtain a unique solution for the intracellular fluxes by optimizing the objective function subject to the constraints of the metabolic balances (Klamt S. et al., 2002). Because all variables have to be positive in a linear programming formalism, one may need to rewrite the model such that the stoichiometry includes forward and reverse reactions. Mathematically, this optimization can be stated as:

$$\begin{aligned} Max/Min \quad & \mathbf{Z} = \mathbf{C} \cdot \boldsymbol{\nu} \\ s.t. \quad & \mathbf{S} \cdot \boldsymbol{\nu} = \mathbf{R} \end{aligned} \tag{2.18}$$

where **C** is a row vector, which specifies the influence of each individual fluxes on the objective function (Stephanopoulos G.N. et al., 1998).

A number of different objective functions have been used for metabolic analysis. These include the following:

1. *Maximize metabolite production.* This objective function is used for robustness analysis of the *Escherichia coli* metabolic network (Edwards J.S. and Palsson B.O, 2000).

2. *Maximize biomass and metabolite production.* By weighing these two conflicting objective functions appropriately, one can explore the trade-off between cell growth and forced metabolite production in a producing strain (Hochreiter S. and Wagner R., 2007).

3. *Minimize nutrient uptake.* This objective function is used to determine the conditions under which the cell will perform its metabolic functions while consuming the minimum amount of available nutrients (Savinell J.M and Palsson B.O, 1992).

4. *Minimizing ATP production or ATP balance error.* This objective function is stated to determine conditions of optimal metabolic energy efficiency (Savinell J.M and Palsson B.O, 1992).

5. *Minimizing net redox production.* This objective function finds conditions where the cells operate to generate the minimum amount or redox potential (Savinell J.M and Palsson B.O, 1992).

6. *Minimize the Euclidean norm.* This objective has been applied to satisfy the strategy of a cell to minimize the sum of flux values, or to channel the metabolites as efficiently as possible through the metabolic pathways (Bonarius B.P.J. et al., 1996).

2.7.2 Sensitivity analysis

After obtaining the unique solutions to the mass balance equations of over-determined, determined, and under-determined systems it is important to measure the sensitivity of the solution

with respect to small perturbations in the measurements. The so-called shadow prices can be addressed from the dual problem of the linear programming:

$$\pi_i = \frac{\partial Z}{\partial b_i}|_{boundary} \tag{2.19}$$

These are the derivatives of the objective function at the boundary. The shadow price can be used to determine whether the cell is limited by a particular constraint (Lee S.U. and Papoutsakis E.T., 1999).

Chapter 3

Materials and methods

3.1 Microorganism and medium

Recombinant *Streptomyces lividans* NCIMB 11416/IL3 p002, was kindly provided by Cangene Corporation, Mississuaga, Ontario, Canada. This strain was derived from *Streptomyces lividans 66* as designated by the John Innes Institute and contains multiple copies of plasmid encoding for the rHuIL-3 gene with a thiostrepton-resistance marker. The seed culture was prepared in thiostrepton containing complex medium as described previously (Yun S. et al., 2001). The basal medium used for growth studies had the following composition (per liter of de-ionized water): 10.6 g K_2HPO_4, 5.3 g KH_2PO_4, 2.5 g $(NH_4)_2SO_4$, 226 μl polypropylene glycol (PPG) 2025, 100 ml of 16% (w/v) glucose, 5 ml of 20% (w/v)$MgSO_4$, and 1 ml trace element solution. The pH was adjusted to 6.7 before autoclaving for 20 min at 121°C. Glucose and magnesium sulfate solutions were autoclaved separately. Trace element solution consisted of (per liter of de-ionized water): 40 g $Na_2EDTA.2H_2O$, 11 g $CaCl_2.2H_2O$, 7 g $FeSO_4.7H_2O$, 2 g $MnCl_2.4H_2O$, 2 g $ZnSO_4.7H_2O$, 0.4 g $CuSO_4.5H_2O$, 0.4 g $CoCl_2.6H_2O$. The amino acid mixtures were dissolved in deionized water and the resulting solutions were sterilized by filtration through a 0.2 μm filter (Millipore) at room temperature.

3.2 Analytical methods

Samples were analyzed for concentration of cells, glucose, ammonia, amino acids, organic acids, and protease activity. Each analysis was carried out in triplicate and the average was reported and used in this study.

3.2.1 Gravimetry

For the estimation of biomass, 10 ml culture samples were centrifuged (5,000 g for 20 min) in centrifuge tube. The pellet was washed twice with distilled water and vacuum filtered through a pre-dried and pre-weighed 0.45 μm filter (Millipore). The biomass cake was dried to a constant weight in a convection oven at 85°C. The first supernatant was filtered through 0.2 μm filters (Millipore) and stored at -35°C for further analysis of extracellular medium.

3.2.2 Optical density

For optical density (OD) measurements, approximately 5 ml samples were homogenized in a 5 ml cell grinder (Safe-Grind, Wheaton Industries Inc., Millville, NJ) by moving the pestle 20 times. The OD of homogenized mixture was measured at 600 nm in a DU 520 UV/Vis spectrophotometer (Beckman Coulter Inc., Fullerton, CA). The cell grinder was rinsed with known volume of dilution water to minimize the cell mass loss in the grinder. Since the OD of the homogenized mixture must be less than 0.3 absorbance unit to retain in linear range of the calibration curve, the ground samples were diluted by de-ionized water, accordingly. The supernatant fermentation broth served as blank. The OD was linearly correlated with the gravimetric biomass measurements.

3.2.3 Glucose assay

The glucose content of the supernatant was analyzed enzaymatically by a glucose kit following the procedure of the manufacturer (Megazyme Glucose Test Kit). The method is based on

oxidase/peroxidase reaction.

3.2.4 Ammonia Assay

Ammonia concentrations were measured by a pH/ISE meter model 710 A equipped with ammonia gas-sensing ion-selecting electrode (Beckman). The pH of all the standard solutions and samples was adjusted to be higher than 11 by adding a 10 M sodium hydroxide solution. This is necessary to convert the ammonium ion to ammonia. The hydrophobic membrane of the electrode allows the ammonia pass through the inner chamber of the electrode where it is converted back to ammonium ion that is detectable as pH changes in the internal filling solutions. Different but known concentrations of ammonium chloride solution were used as the standard solution to calibrate the meter.

3.2.5 Interleukin-3 assay

Total interleukin-3 concentration in fermentation broth samples were estimated using the Direct Enzyme Linked Immunosorbent Assay (Direct ELISA) and Sodium Dodecyl Sulfate PolyAcrylamide Gel Electrophoresis (SDS-PAGE).

Direct Enzyme Linked Immunosorbent Assay (Direct ELISA)

The entire assay was performed in a 96-well, high binding polystyrene microtiter plate (Immolun 2 HB Flat Bottom Microtiter Plates). The primary and secondary antibodies for this assay were rabbit polyclonal Ab specific to IL-3 (Abcam, Inc., Cambridge, MA), and anti-rabbit IgG alkaline phosphate conjugate (Sigma-Aldrich, Canada Ltd., Oakville, ON), respectively. *p*-Nitrophenyl phosphate (Sigma-Aldrich, Canada Ltd., Oakville, ON) was used as the substrate for the secondary antibody. The chemical composition of the solutions is listed in Table 3.1. The procedure used for the test is as follows:

1. Aliquot 100 μl of coating buffer into each well of a microtiter except wells in the first column.

2. Aliquot 150 μl of standard into wells A1 and B1, dilution buffer into wells C1 and D1, test sample one into wells E1 and F1, and test sample two into wells G1 and H1.

3. Aliquot 50 μl from each well in first column into the corresponding well in second column and repeat for the rest of the columns. Remove 50 μl of the contents of the last column (12^{th} column) and discard it. This will end up with eleven 3-fold dilutions.

4. Cover the plate with an adhesive film and incubate for one hour at 37°C.

5. Remove the coating solution and wash the plate four times by filling the wells with 250 μl of washing buffer. The solutions or washes are removed by flicking the plate over a sink. The remaining drops are removed by patting the plate on a paper towel.

6. Dissolve one vial of rabbit polyclonal Ab specific to IL-3 (Abcam, Inc., Cambridge, MA) in 1 ml of dilution buffer. Dilute 100 μl of the resulting solution in 11 ml of dilution buffer. Add 100 μl of the primary antibody solution to each well.

7. Cover the plate with an adhesive film and incubate for one hour at 37°C.

8. Wash the plate four times with the washing buffer.

9. Dilute 10 μl of the anti-rabbit IgG alkaline phosphate conjugate (Sigma-Aldrich, Canada Ltd., Oakville, ON) in 11 ml of dilution buffer and add 100 μl of the secondary antibody solution to each well.

10. Cover the plate with an adhesive film and incubate for half an hour at 37°C.

11. Wash the plate four times with the washing buffer.

12. Dispense 100 μl of alkaline phosphate substrate solution per well.

13. After sufficient colour development (if necessary add 100 μl of stop solution (1M EDTA solution) to the wells.

14. Read the absorbance of each well at 405 nm in a DU 520UV/Vis spectrophotometer (Beckman Coulter)

Table 3.1: ELISA formulations

10 X PBS (Posphate Buffered Saline pH 7.4)	
NaCl	80.0 g
KCl	2.0 g
Na_2HPO_4	14.4 g
KH_2PO_4	2.4 g
Deionized water (final volume)	1000 ml
pH adjusted to 7.4 $\pm$ 0.1 with HCl	
Sterilized by filtration and stored at 4°C	
20% (v/v) Tween 20	
Tween 20	20 ml
Deionized water (final volume)	100 ml
Sterilized by filtration and stored at 4°C	
Coating Buffer	
10 X PBS	100 ml
Deionized water (final volume)	1000 ml
Sterilized by filtration and stored at 4°C	
Dilution Buffer	
10 X PBS	10 ml
Bovine Serum Albumin (BSA)	1.0 g
Deionized water (final volume)	100 ml

Sterilized by filtration and stored at 4°C	
Washing Buffer	
10 X PBS	100 ml
20% Tween 20	5 ml
Deionized water (final volume)	1000 ml
Sterilized by filtration and stored at 4°C	
Alkaline Phosphate Buffer	
Diethanolamine	97 ml
NaN_3	0.2 g
$MgCl_2.6\,H_2O$	0.1 g
Deionized water (final volume)	1000 ml
pH adjusted to 9.8 ± 0.1 with HCl	
Sterilized by filtration and stored at 4°C	
Alkaline Phosphate Substrate	
p-Nitrophenyl phosphate	1 tablet
Alkaline Phosphate Buffer	10 ml

SDS-PAGE method

The assay was carried out in a Bio-Rad Mini-Protean 3 (Bio-Rad Laboratories, Inc.). Proteins in supernatant samples were separated by denaturing discontinuous SDS-PAGE, under reducing conditions, according to the method of Laemmli (Laemmli U.K., 1970). All the reagents were purchased from Bio-Rad laboratories Inc. (Mississauga) unless stated to the contrary.

Resolving gel Each resolving (separating) gel was $80 \times 60 \times 1.5$ mm. A mixture of 7.5 ml resolving gel stock solution and 7.5 ml 30% Acrylamide/Bis solution, 29:1 (3.3% C) and 100 μl ammonium sulphate (APS) were mixed prior to adding 20 μl N, N, N', N'- tertramethylethylenediamine (TEMED) to initiate the polymerization. Two gels were simultaneously

prepared in Bio-Rad Mini-Protean 3 multicasting chamber. The solution was poured into the multicasting chamber and ethanol overlay was applied. Polymerization was typically complete in 20 min. The ethanol overlay was discarded and the top of the resolving gel was rinsed with de-ionised water.

Stacking gel 3.2 ml stacking gel solution and 0.8 ml 30% Acrylamide/Bis solution 29:1 (3.3% C) and 33.3 μl APS was mixed. 7 μl TEMED was then added to initiate the polymerization. The stacking gel solution was poured on top of the polymerized resolving gels and the combs were immediately applied.

Sample preparation 233 μl of 50% trichloroacetic acid (TCA) was added to 1.0 ml of supernatant sample in an Eppendorf vial, then vortex mixed and chilled for 10 min prior to centrifugation for 20 min at 5000 g and 4°C. The liquid solution was discarded to waste collection to obtain a protein pallet. The pallet was dried. For 10 X concentration of samples 100 μl of 1 X Laemmeli sample buffer with β-mercaptoethanol (BME) and for 5 X concentration 200 μl of BME was add to each vial. Any remaining TCA residue can turn the sample buffer to yellow. To neutralize TCA and recover the colour of sample buffer known volume of 1 M sodium hydroxide solution was added to each sample. The samples were heated for 10 min at boiling water and then loaded to the gels.

Running conditions Both anode and cathode chambers were filled with 1 X running buffer. Each well on the gel was loaded with 15 μl sample mixture. Protein electrophoresis was performed in Bio-Rad Mini-Protean 3 at constant voltage of 100 V for 150 min (PowerPac 1000 power supply, Bio-Rad). After electrophoresis, the gels were carefully removed from the glass plates and left gently shaken in SDS-PAGE fixing and staining solution for about half an hour. Destaining was achieved by multiple washes with SDS-PAGE de-staining solution.

Estimation of rHuIL-3 concentration Gels were placed between clear cellophane supports (Bio-Rad) and dried. Dried gels were then scanned with a flatbed colour scanner at 800 dpi. Purified rHuIL-3 provided by Cangene Corp. (Mississauga, Canada) was used as standard. The calibration curves were constructed by plotting the intensity standard solutions versus the concentration of the standard solutions. The amount of rHuIL-3 in each sample was estimated by comparing the intensity of its band with the calibration curve. A Bio-Rad Quantity One software package was used to measure the trace intensities of the bands on the gels. To ensure the accuracy, each gel was loaded with at least four lanes of rHuIL-3 standards. The 15 kDa bands on these gels have been confirmed to be rHuIL-3 by running Western Blots specific to rHuIL-3 (Yahya, 2003).

3.2.6 Protease assay

Protease activity was measured using the method described by Sarath et al.(Beynon R.J. and Bond J.S., 2000). This method provides an estimate of general proteolytic activity of the sample. The assay is based on the absorbance measurement of the low molecular weight coloured peptides, which are the product of the proteolytic action on the protein substrate. All the samples and reagents were incubated at 32°C prior to the assay. Azocasein substrate 2% w/v was prepared by dissolving azocazein in 100 mM phosphate buffer of pH of 6.4 and clarified by centrifugation at 10000 g for 10 min. Assay was initiated by pipetting 80 μl of azocasein into a 1.5 ml eppendorf vial containing 48 μl of the cell free supernatant. The reaction mixtures were vortex mixed and incubated at 32°C for 24 hours in the presence of 5 mM sodium azide solution to prevent microbial growth. Following incubation 384 μl of 10% TCA added to each sample and mixed thoroughly to terminate the assay. The contents of the vial were allowed standing for 15 min to ensure complete precipitation of the remaining azocasein and azocasein fragments. The sample blanks were prepared by pippetting 48 μl cell free supernatant to a vial containing 384 μl of 10% TCA. The reaction mixture was vortex mixed and incubated with samples. After incubation for 24 hours, 80 μl of azocasein was added to the sample blank. The

contents of the vial were allowed standing for 15 min. All samples and sample blanks were centrifuged at 8000 g for 5 min and 384 μl of the supernatant fluid was transferred to another vial containing 448 μl of sodium hydroxide. The absorbance of these solutions was determined at 440 nm using a Multiskan Ascent (Labsystems, Finland). One unit protease activity was defined to be the amount of enzyme required to produce an absorbance change of 1.0 in a 1 cm cuvette, under the conditions of the assay. The enzyme activity was calculated using the following equation:

$$Activity = \frac{\alpha_{Sample} - \alpha_{Blank}}{0.3} \times \frac{832}{348} \times \frac{512}{48} \tag{3.1}$$

where, α_{Sample} and α_{Blank} are the absorbance of the sample and the blank, respectively. Specific activity of the protease was defined as units per millilitre per milligram of cell dry weight.

3.2.7 Amino acid assay

Amino acid analysis of the supernatant samples was performed by High Performance Liquid Chromatography (HPLC) following pre-column derivatization with phenylisothiocyanate (PITC) as described by Bidlingmeyer et al. (Bidlingmeyer B.A. et al., 1984). The samples containing up to 50 nmol of each amino acid were dried under high vacuum. Then, 20 μl of ethanol:water:triethylamine (TEA) (2:2:1) was added to the samples and dried again under vacuum. The residual amino acids were derivitized by adding 20 μl of derivitization reagent. The derivitization reagent was made fresh consisting of ethanol:water:TEA:PITC (7:1:1:1). PITC was added to this solution under nitrogen atmosphere. The derivatized form of the amino acids, were formed by adding 20 μl of the derivatization reagent to the dried samples under nitrogen environment and sealing them at room temperature for 20 min. The samples were dried under high vacuum, once more. Reverse phase HPLC was conducted at 55°C with a Waters 2690 separation module. The UV detector was controlled at 250 nm. Separation was performed on an Intersil C8-3 column (150 mm $\times$ 4.6 mm, I.D.) with the particle size of 5 μm. The solvent consisted of two streams: solvents A and B at the total flow rate of 1.0 ml/min.

Solvent A was aqueous buffer with 0.14 M sodium acetate and 0.5 ml/l of TEA. The pH of the solvent A was adjusted to 6.35 by adding glacial acetic acid. Solvent B was a 60% solution of acetonitrile in water. The separation gradient started with 10% solvent B increasing to 53% solvent B in 19 min using the convex curve number 5. Solvent B at the end of each injection was used to wash out any residual sample components. The derivetized samples were dissolved in 1.0 ml of solvent A and 15 μl of the resulting solutions were injected into the column. The concentration of each amino acid in the samples was determined by comparing the area under the curve with the calibration curves.

3.2.8 Assay of organic acids

Low molecular weight organic acids were determined with a Waters 2690 separation module. Separation was performed on an Intersil C8-3 column (150 mm $\times$ 4.6 mm, I.D.) with the particle size of 5 μm at 27°C using the method described by Cawthray (Cawthray G.R, 2003). All the samples and standards were acidified to pH 2.7 with *ortho*-phosphoric acid. The mobile phase composed of 93% 25 mM KH_2PO_4 and 7% methanol at a flow rate of 1 ml/min. The UV detector was controlled at 210 nm.

3.2.9 Off-gas analysis

Carbon dioxide and oxygen concentrations were monitored on line at the inlet and outlet of the fermentor with a solid-state infrared CO_2 sensor and an electrochemical oxygen sensor, respectively (Model 902, Quantek Instruments, USA).

Chapter 4

Medium design and optimization

4.1 Introduction

Federal guidelines and regulations banning animal-derived medium components necessitate the refinement or re-development of industrial medium formulations (Zhang J. and Greasham R., 1999). Medium composition has been known to have a significant effect on product concentration, yield, and volumetric productivity. Therefore, the design of a fermentation medium is of critical importance during the development of an industrial fermentation. The development of defined medium without animal products is most desirable for the production of pure and safe biological products. The employment of a defined medium often leads to better control strategies, higher growth rates, improved growth yields, and higher cell densities (Yee L. and Blanch H.W., 1992). Several medium design strategies have been proposed (Kennedy M. and Krouse D., 1999). Traditionally, medium design and optimization was done by varying one factor while keeping the other factors at a constant level. This one-factor-at-a-time technique (Chary C.V.K. et al., 1989; Monot F. et al., 1982) is simple, but interactions between components are ignored. In addition, it involves a relatively large number of experiments and the optimal medium composition can be missed completely. Alternatively, statistically designed experiments allow the evaluation of more than one factor at a time and the interactions be-

tween the factors (Gheshlaghi R. et al., 2005; Gouveia E.R. et al., 2001; Ooijkaas L.P. et al., 1999; Techapun C. et al., 2002), but they require a large number of experiments and are not suited for minimal medium development. There have been also some attempts on the use of artificial neural networks and genetic algorithms have been used to model and optimize fermentation media (Baishan F. et al., 2003; Nagata Y. and Chu K.H., 2003).

A unique combination of neural networks with statistical techniques is presented as a novel approach to identify the essential amino acids and determine their optimum concentrations for the growth of *Streptomyces lividans* and production of Recombinant Human Interleukin-3 (rHuIL-3).

A schematic outline of the technique for design and optimization of a medium is given in Fig.4.1. Initially, starvation trials with growth limiting amino acid levels were performed to establish the baseline for nutritional requirements. Following these preliminary experiments, a screening mixture experiment was designed and distance based multivariate analysis was performed to establish the rank of the amino acids. A mixture design known as a simplex lattice design (Cornell J.A., 2002; Myers R.H. and Montgomery D.C., 2003) was performed to obtain the optimum proportions of the essential amino acids. Finally, a neural network model was employed as an alternative to the statistical model.

The task at hand was to develop a systematic procedure for the design and optimization of a chemically defined medium. The study aimed at replacing casein peptone in conventional medium for *Streptomyces lividans* (strain 1326) with essential amino acids and determining the optimum proportion of the amino acids. The organism used secretes fully bioactive recombinant human interleukin-3 (rHuIL-3). The proposed methodology is believed to be robust and can be employed for the optimal medium design for any recombinant platform.

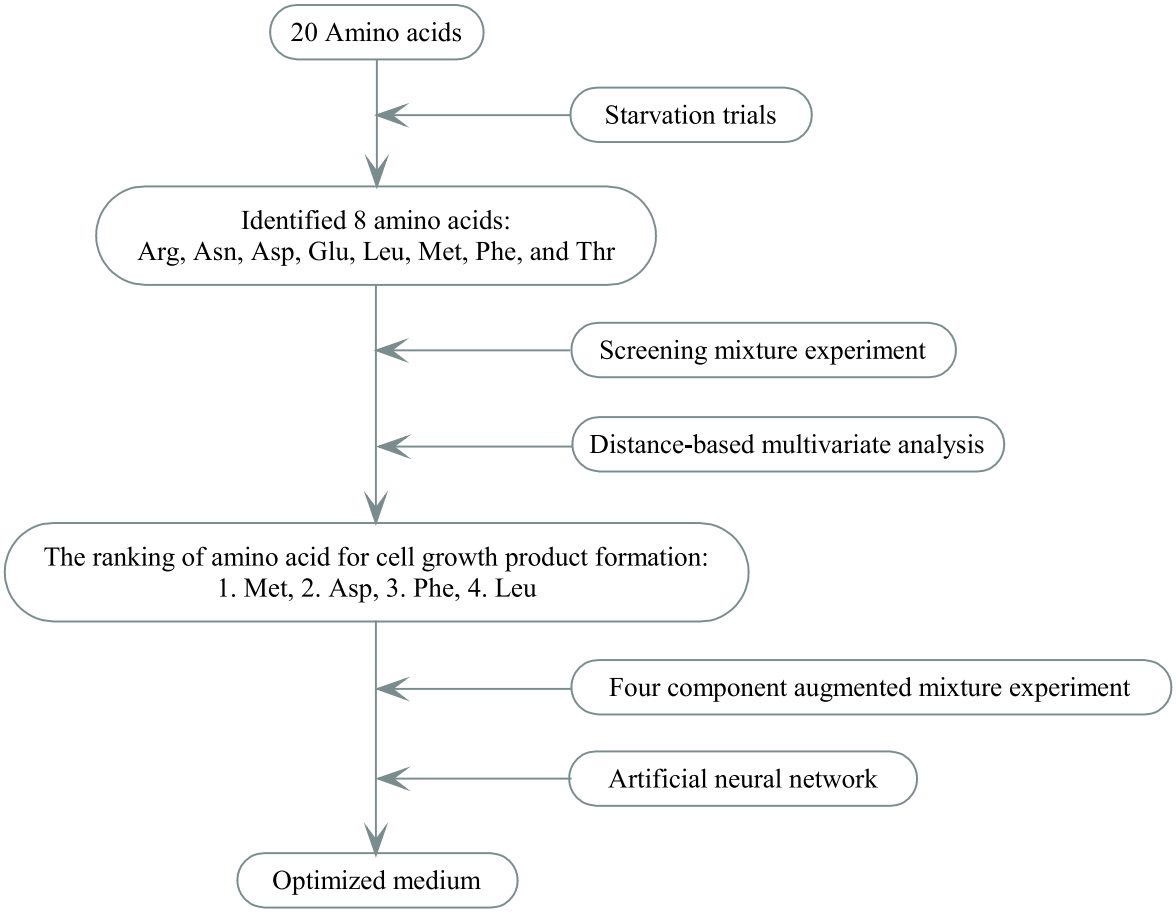

Figure 4.1: Outline of the technique used for design and optimization of medium

4.2 Materials and Methods

Medium composition used is given in section 3.1 except for the amino acid cocktail. Amino acid concentrations used were different for different experiments and are described in sections 4.3, 4.4, and 4.5.

4.2.1 Inoculation and Incubation

Streptomyces lividans cultures were grown in 500 ml baffled flasks at a volume ratio of 5:1 in medium containing 10 μg/l thiostrepton. The concentration of inoculum in each flask was

100 μl. The temperature was maintained at 32°C and agitation at 240 rpm on an Innova 4300 incubation shaker (New Brunswick Scientific Co., USA).

4.2.2 Sampling

Samples of approximately 10 ml were taken aseptically after 24 h of fermentation. The samples were centrifuged (5000 g for 20 min) in centrifuge tubes. The supernatant were stored at -35°C for later analysis.

4.3 Starvation Trial

To establish growth limiting conditions, growth studies in equimolar amino acid solutions at suboptimal concentrations were carried out. For brevity, these tests are referred to as starvation trials. The results are shown in Fig.4.2. Growth was measured both in terms of optical density and biomass. The results in Fig.4.2 indicate that reducing the total concentration of amino acids below 0.25 g/l limits growth noticeably. Thus, an initial amino acid concentration of 0.25 g/l was employed as a minimum nutrient requirement for *Streptomyces lividans*. The supernatant samples corresponding to the stoppage point of growth were analyzed for amino acids. The results were overlaid on the initial amino acid profile to show the changes of amino acid concentration during the fermentation. A typical diagram of amino acid concentrations in fresh medium and the supernatant is shown in Fig.4.3. This figure corresponds to an initial concentration of 0.25 g/l of amino acids in the medium. Amino acid analysis of the fresh medium and the supernatant were carried out in triplicates and the reported data are the average values. The results of amino acid analysis for the rest of the fermentation with different amounts of initial amino acid were the same as in Fig.4.3. It is evident that eight of the twenty amino acids were completely or almost completely consumed when growth ceased. These amino acids included: Arginine, Asparagine, Aspartic acid, Glutamic acid, Leucine, Methionine, Phenylalanine, and Threonine. However this is not a definitive indication of ab-

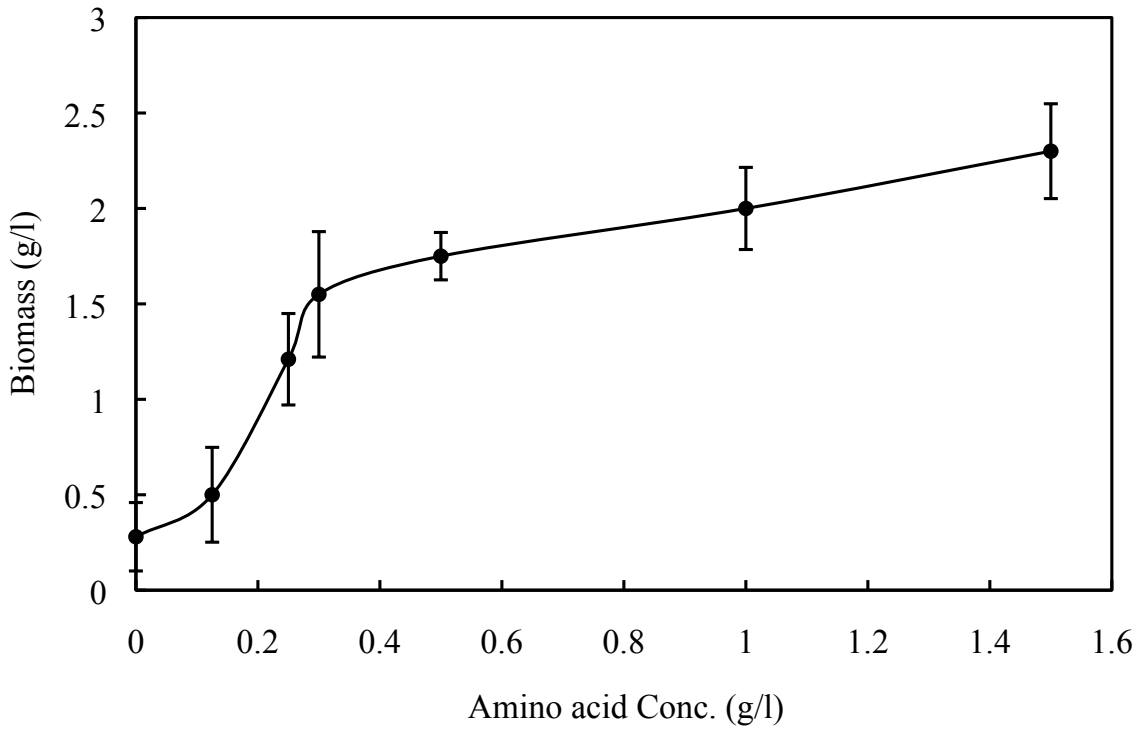

Figure 4.2: The effect of total amino acid concentration on biomass production

solute requirement of these amino acids by the bacterial culture. To clarify this ambiguity, fermentations were run with these identified eight amino acids only, the complementary twelve that were not exhausted, and all twenty amino acids in the medium. The results of these fermentations are given in Table 4.1. These results indicate that deleting the complementary twelve amino acids has no significant effect on growth. Accordingly, the identified eight amino acids were chosen for further study as potential essential amino acids and a screening mixture experiment was designed to rank the importance of the amino acids.

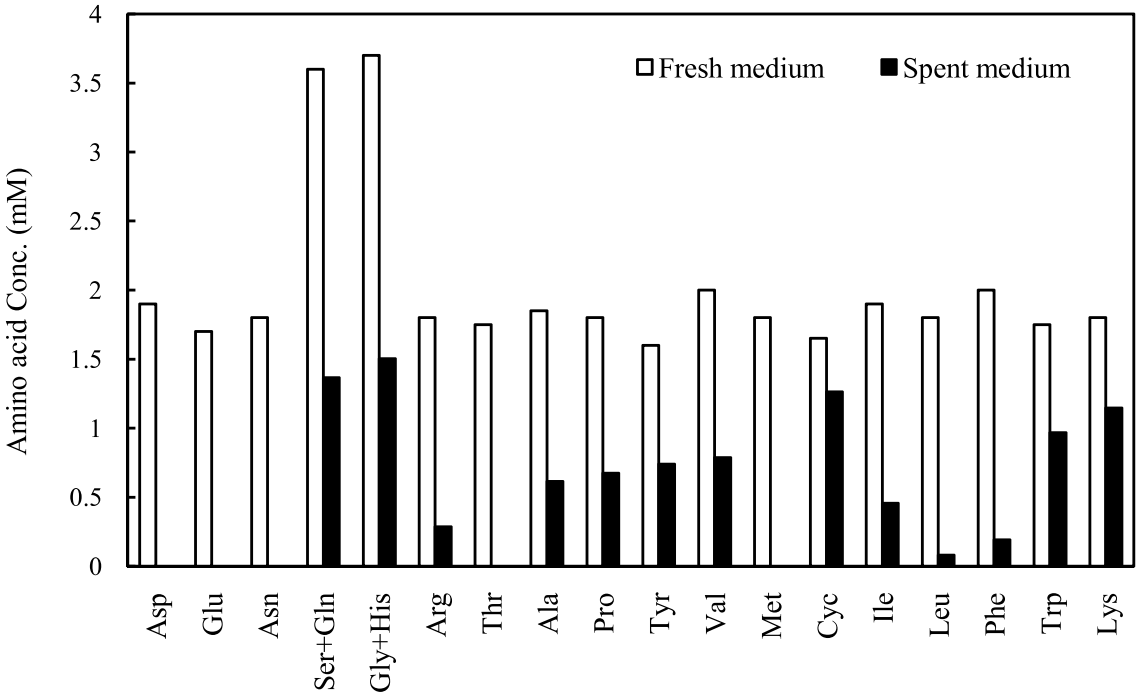

Figure 4.3: Amino acid concentration in fresh medium and first supernatant

Table 4.1: Test of growth dependence on amino acid combinations

Number of Amino Acids	Total weight of Amino Acids (g/l)	Biomass (g/l)
Identified 8	0.630	1.85
Complementary 12	0.870	0.64
All 20	1.500	1.81
None	0.000	0.28

4.4 Screening mixture experiment

A screening mixture experiment was designed to rank the importance of the identified eight amino acids. The design was performed with the aid of Design Expert version 6.0.4 statistical software (Stat-Ease Inc., Minneapolis, MN). The corresponding design and results are shown in Table 4.2. A total amino acid concentration of 0.25 g/l was used in the experimental design. Then the DISTLM *forward* program (Anderson M.J., http://www.stat.auckland.ac.nz/∼mja/Programs.htm (2003) was employed for distance-based multivariate analysis using forward selection (McArdle B.H. and Anderson M.J., 2001) to rank the identified amino acids. This program needs two input files in ASCII*.txt format. The first input file contains the data matrix of the response variables **Y** (rHuIL-3 concentration and biomass) normalized by dividing each column by the greatest value in the corresponding column, and the other input file is an X matrix containing 36 predictor variables. The predictor variables are 8 amino acid mole fractions in the experimental design (primary predictor variables) and the remaining 28 predictor variables are the binary interaction of those 8 primary predictor variables. Euclidean distance was the choice of the distance measure. After running the program 18 out of 36 variables explained one hundred percent of the variability in the data which meant that the degrees of freedom was eighteen.

Table 4.2: Screening mixture experiment design and results

	Mole fraction								Biomass	rHuIL-3
Run	Arg	Asp	Asn	Glu	Leu	Met	Phe	Thr	(g/l)	(mg/l)
1	0.169	0.128	0.129	0.142	0.127	0.145	0.160	0.000	0.953	47.7
2	0.173	0.132	0.133	0.000	0.131	0.149	0.165	0.119	1.224	52.6
3	0.070	0.053	0.054	0.059	0.053	0.060	0.601	0.048	0.839	37.9
4	0.079	0.539	0.060	0.066	0.059	0.068	0.075	0.054	0.848	25.2
5	0.171	0.130	0.131	0.143	0.000	0.146	0.162	0.117	1.247	36.8

Continued on next page

Table 4.2 – continued from previous page

	Mole fraction								Biomass	rHuIL-3
Run	Arg	Asp	Asn	Glu	Leu	Met	Phe	Thr	(g/l)	(mg/l)
6	0.616	0.052	0.052	0.057	0.052	0.059	0.065	0.047	0.727	35.1
7	0.151	0.115	0.116	0.127	0.114	0.130	0.144	0.104	1.178	41.2
8	0.171	0.000	0.131	0.144	0.129	0.147	0.162	0.117	1.103	45.5
9	0.079	0.060	0.541	0.066	0.059	0.067	0.075	0.054	1.332	42.4
10	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.423	0.0
11	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.396	0.0
12	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.112	0.0
13	0.174	0.132	0.133	0.146	0.131	0.000	0.165	0.119	0.142	0.0
14	0.000	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.206	0.0
15	0.083	0.063	0.063	0.069	0.062	0.071	0.079	0.510	0.813	34.5
16	0.074	0.056	0.057	0.062	0.056	0.573	0.070	0.051	0.805	38.2
17	0.151	0.115	0.116	0.127	0.114	0.130	0.144	0.104	0.963	42.1
18	1.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.053	0.0
19	0.000	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.116	0.0
20	0.177	0.134	0.135	0.148	0.133	0.151	0.000	0.121	0.852	37.1
21	0.075	0.057	0.057	0.567	0.057	0.064	0.071	0.051	1.011	39.8
22	0.171	0.130	0.000	0.144	0.129	0.147	0.162	0.117	0.772	24.6
23	0.151	0.115	0.116	0.127	0.114	0.130	0.144	0.104	0.866	49.0
24	0.000	0.135	0.136	0.150	0.134	0.153	0.169	0.122	0.915	32.7
25	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	0.246	4.9
26	0.000	0.000	0.000	0.000	0.000	1.000	0.000	0.000	0.495	25.4
27	0.079	0.060	0.061	0.066	0.537	0.068	0.075	0.054	1.050	31.0
28	0.151	0.115	0.116	0.127	0.114	0.130	0.144	0.104	1.005	40.0
29	0.151	0.115	0.116	0.127	0.114	0.130	0.144	0.104	0.985	40.7

A summary of the computer output is shown in Fig. 4.4. This figure presents the percent variability of the response matrix (normalized rHuIL-3 and biomass concentration) as explained by the individual amino acids and their binary combinations. Using this information, the probable amino acid requirement (starting with the highest ranked single amino acid, methionine) was calculated by summing the corresponding probability values.

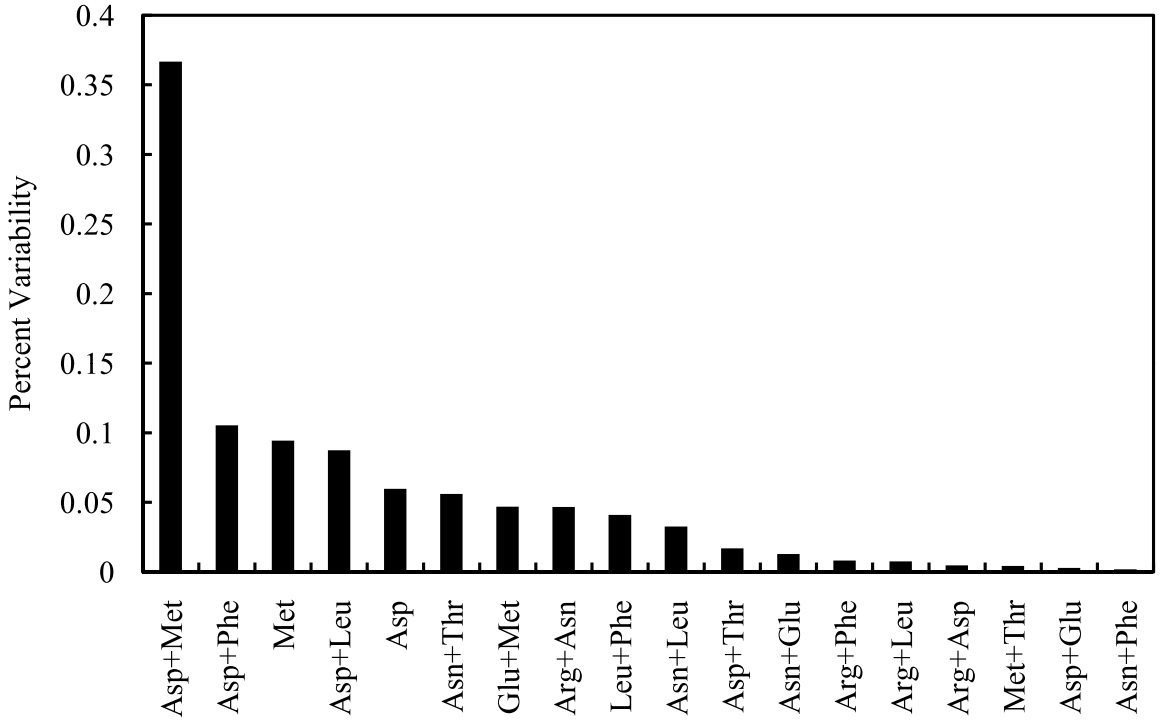

Figure 4.4: Percent variability

The predicted rank is shown in Table 4.3. Verification of the rank predicted by the computer program was done by another series of fermentations, in which the total amino acid concentration was 0.25 g/l. The first run contained only methionine (Met), the second Met and aspartic acid (Asp) with equimolar concentration, the third one Met, Asp, and phenylalanine (Phe) with equimolar concentration, and so on. The results of these fermentation runs are

shown in Table 4.3. It is obvious that after leucine (Leu), addition of the rest of the identified amino acids decreases the growth. This is due to dilution of the essential amino acids. As mentioned previously, the total amount of the amino acids concentration was fixed, therefore, by adding more amino acids to the solution the essential amino acids the concentrations decreased. Based on these findings, Met, Asp, Phe and Leu were singled out as the essential amino acid candidates.

Table 4.3: Predicted rank of amino acid combinations and corresponding biomass concentration (ranking obtained by DISTLM *forward* computer program)

Variable	Probability	Biomass (g/l)
Met	0.09	0.52
Met+Asp	0.47	1.23
Met+Asp+Phe	0.47	1.84
Met+Asp+Phe+Leu	0.53	1.92
Met+Asp+Phe+Leu+Glu	0.67	1.49
Met+Asp+Phe+Leu+Glu+Asn	0.82	1.55
Met+Asp+Phe+Leu+Glu+Asn+Arg	0.97	-
Met+Asp+Phe+Leu+Glu+Asn+Arg+Thr	1.00	-

4.5 Defined medium optimization by mixture experiments method

The mixture experiment method was rearranged to obtain the best possible minimal medium composition for *Streptomyces lividans*. The standard simplex design is a boundary point design. With the exception of the overall centroid, all design points are on the boundaries of the simplex. To make predictions about properties of complete mixtures more runs are required in the interior of the simplex. Hence, the usual simplex design was augmented with four axial runs and five design points were done in replicate to check the accuracy of the model

and the lack of fit. Therefore, the total number of experiments supporting a quadratic model with four components would be 24. The experimental design and the results summarized in Table 4.4. The total amount of amino acids used in these experiments was 0.25 g/l.

Table 4.4: Experimental design and results for four-component mixture experiment

	Mole fraction				rHuIL-3 (mg/l)		
Run	Asp	Leu	Met	Phe	Experimental	Regression	Neural network
1	0.000	0.000	0.525	0.475	28.13	27.13	28.29
2	0.000	0.000	1.000	0.000	17.48	17.78	15.70
3	0.475	0.000	0.525	0.000	55.31	59.20	53.64
4	0.000	0.475	0.525	0.000	15.34	19.75	17.80
5	0.594	0.119	0.169	0.119	40.35	54.77	43.78
6	0.000	0.950	0.050	0.000	20.43	21.71	19.74
7	0.000	0.950	0.050	0.000	20.42	21.71	19.74
8	0.119	0.119	0.644	0.119	35.72	34.52	33.32
9	0.000	0.475	0.050	0.475	20.91	29.10	18.62
10	0.475	0.000	0.050	0.475	78.01	68.79	76.00
11	0.950	0.000	0.050	0.000	45.30	44.03	43.83
12	0.119	0.119	0.169	0.594	49.49	43.93	49.25
13	0.000	0.000	0.050	0.950	32.96	36.49	34.16
14	0.317	0.317	0.050	0.317	62.74	46.75	61.56
15	0.000	0.000	1.00	0.000	15.06	17.78	15.70
16	0.119	0.595	0.169	0.119	26.78	29.41	29.53
17	0.238	0.238	0.288	0.238	50.91	44.21	51.44
18	0.000	0.000	0.050	0.950	35.36	36.49	34.16
19	0.317	0.000	0.367	0.317	72.95	58.02	73.85

Continued on next page

Table 4.4 – continued from previous page

	Mole fraction				rHuIL-3 (mg/l)		
Run	Asp	Leu	Met	Phe	Experimental	Regression	Neural network
20	0.000	0.317	0.367	0.317	28.52	25.33	29.65
21	0.317	0.317	0.367	0.000	39.35	40.41	39.04
22	0.475	0.475	0.050	0.000	40.36	32.87	40.67
23	0.950	0.000	0.050	0.000	44.10	44.03	43.83
24	0.475	0.000	0.050	0.475	74.02	68.79	76.00

4.5.1 Multiple regression analysis

Based on the experimental values, statistical testing using Fisher's statistical test was carried out. The experimental design and statistical analysis of the results was performed with the aid of Design Expert version 6.0.4 statistical software (Stat-Ease Inc., Minneapolis, MN). The values of regression coefficients were calculated and the following equation which includes all terms regardless of their significance was derived for rHuIL-3:

$$\begin{aligned} rHuIL-3 = {} & 33.7 \times Asp + 18.8 \times Leu + 15.9 \times Met + 26.6 \times Phe \\ & +6.1 \times Asp \times Leu + 97.7 \times Asp \times Met + 151.4 \times Asp \times Phe \\ & +3.1 \times Leu \times Met + 24.7 \times Leu \times Phe + 59.8 \times Met \times Phe \end{aligned} \tag{4.1}$$

The model F-value was 4.04 implying that the model is significant. The P-values for and terms were less than 0.05 indicating that these terms are significant at a probability level of 95% and the remaining interaction terms were insignificant at that level. The insignificant terms of the model of Eq.4.1 were eliminated to give the following improved model:

$$\begin{aligned} rHuIL-3 = {} & 44.0 \times Asp + 21.7 \times Leu + 17.8 \times Met + 36.5 \times Phe \\ & +113.2 \times Asp \times Met + 114.1 \times Asp \times Phe \end{aligned} \tag{4.2}$$

The analysis of variance for the improved model is summarized in Table 4.5. An F-value of the improved model of 12.76, and a P-value of less than 0.0001 indicate that there is less than

0.05% chance that a model F-value this large could occur due to noise. P-values less that 0.05 indicate that the model terms are significant at a probability level of greater than 95%. In this case, linear component interactions, are significant model terms. The R^2 of the regression model was 0.78. Because the coefficient of Asp in the model was greater than the coefficients of Phe, Leu, and Met, one would conclude that Asp has higher impact on rHuIL-3 production by *Streptomyces lividans* than Phe, Leu, and Met. Furthermore, since the interaction between Asp and Met, and Asp and Phe has a positive coefficient and these coefficients are greater than the coefficients of the individual amino acids, the interaction terms have higher impact on product formation than the individual amino acids.

Table 4.5: ANOVA table for the improved model

Source of variation	SS	df	MS	F	$P > F$
Model	2.130	5	0.430	11.51	< 0.0001
Linear mixture	1.480	3	0.490	13.27	< 0.0001
Asp Met	0.280	1	0.280	7.45	0.0137
Asp Phe	0.400	1	0.400	10.88	0.0040
Residual	0.670	18	0.037		
Lack of Fit	0.540	13	0.041	1.56	0.3275
Pure error	0.130	5	0.026		
Total	2.800	23			

4.5.2 Artificial neural network

The architecture of the neural network developed in this work was an input layer of 4 neurons and an output layer of one neuron only corresponding to rHuIL-3 concentration. Only one hidden layer was used. Different numbers of neurons (4, 5, and 6 neurons) were tried in the hidden layer, checking each time if the developed neural network succeeded in reading

an error target. It was found that a network with one hidden layer of four neurons gave sufficiently accurate predictions. Both sigmoid and tangent hyperbolic functions were tried as transfer functions and it was found that because of its greater slope compared to sigmoid function, tangent hyperbolic function produced better results. The data set obtained from mixture experiments comprising 24 experimental runs was used as the training set to optimize the weights of the neural network. The R^2 of the neural network for the training data set was found to be 0.9987. The trained network was used to furnish predictions of rHuIL3 concentrations for the data points that were not used in the training network. This validation step is important to check generalization characteristics of the developed neural network. Another set of 8 experimental runs were carried out by choosing combinations of mole fractions of amino acids which were not taken earlier to validate the neural network model and verify the final regression model. The results of the verification experiments are given in Table 4.6. The R^2 of the neural network for the testing data was 0.9559. Fig.4.5 shows the network-predicted rHuIL3 concentrations for training and validation data sets plotted against the corresponding experimental data. The solid circles represent the network-trained outputs while the open circles denote the network-predicted outputs for input variables belonging to the validation set. The neural network model not only fits the training data very well but also provides predictions of the validation data that are very close to the experimental measurements. The rHuIL-3 concentrations calculated from the multiple regression equation (Eq. 4.2) is also shown in Fig. 4.5 for comparison. It is obvious that the neural network predictions are much closer to the line of perfect prediction than those of regression polynomial equation, confirming the usefulness of the neural networks as empirical models in response surface analysis.

4.5.3 Optimization of defined medium

Contour plot obtained from the regression model over the simplex region (Fig. 4.6) showed the existence of a local optimum within the experimentally explored range. The location of the optimum could be obtained by solving the optimization problem using sequential quadratic

Table 4.6: Verification of regression and neural network models

Run	Mole fraction				rHuIL-3 (mg/l)		
	Asp	Leu	Met	Phe	Experimental	Regression	Neural network
1	0.250	0.250	0.250	0.250	39.98	44.21	37.13
2	0.500	0.000	0.500	0.000	54.16	59.20	49.18
3	0.000	0.500	0.500	0.000	22.69	19.75	17.89
4	0.000	0.000	0.500	0.500	23.18	27.14	21.73
5	0.170	0.160	0.500	0.170	45.31	38.97	47.60
6	0.500	0.160	0.170	0.170	50.77	54.04	48.66
7	0.170	0.500	0.170	0.160	29.62	33.58	30.96
8	0.170	0.170	0.160	0.500	47.23	45.04	47.12

programming algorithm. The response surface and neural network models for rHuIL-3 production in terms of the amino acid composition were used to find the optimum level of amino acid by solving the following nonlinear optimization problem:

$$
\begin{aligned}
& Max[rHuIL-3 \quad production] \\
& s.t. \\
& Asp + Leu + Met + Phe = 1 \\
& 0 \leq Asp \leq 0.95 \\
& 0 \leq Leu \leq 0.95 \\
& 0.05 \leq Asp \leq 1 \\
& 0 \leq Phe \leq 0.95
\end{aligned}
\tag{4.3}
$$

This problem was solved using MATLAB (The MathWorks, 2007). MATLAB includes an optimization toolbox that implements various numerical optimization routines, including sequential quadratic programming algorithm to solve for the constrained optima. The MATLAB function *fmincon* solves the constrained problems and was employed in this research to find the optimal composition of amino acids in the medium. The maximum value of the objective

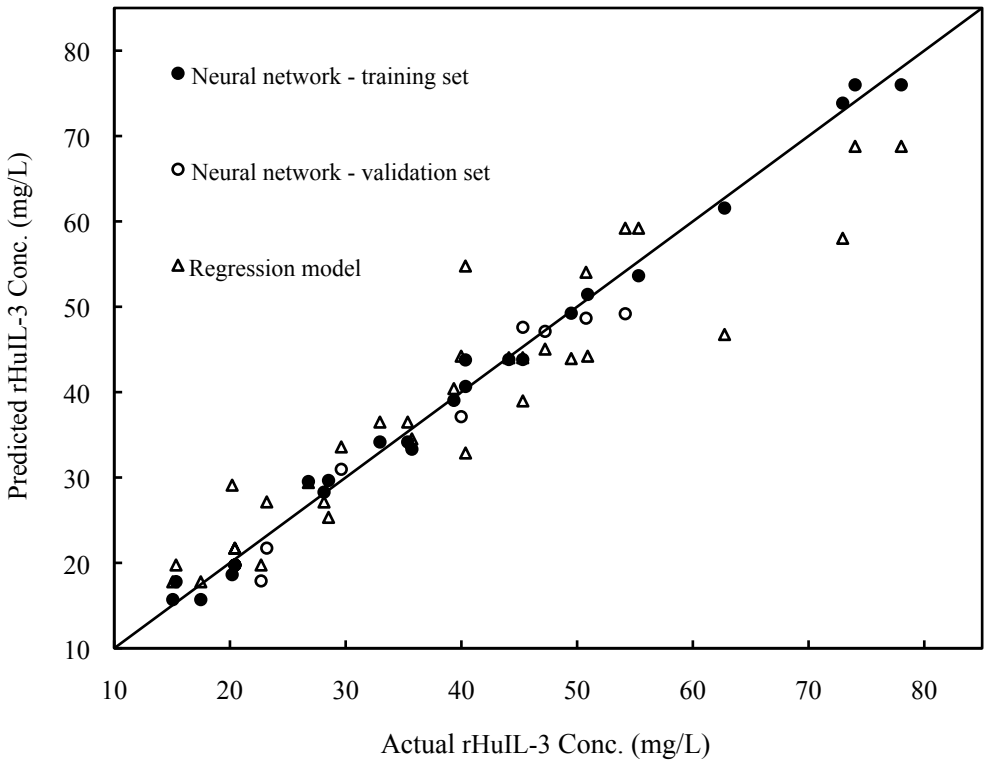

Figure 4.5: Predicted vs. actual rHuIL-3 concentration

function (rHuIl-3 production) using multiple regression model was found to be 67.95 mg/l, and according to the neural network model, the maximum achievable rHuIL-3 concentration for this fermentation was 76.00 mg/l. The combinations of mole fractions of the four amino acids giving the maximum concentration according to both models are listed in Table 4.7. To verify the results of the optimization, four shake flask fermentations were carried out using the optimal amino acid composition of the minimal medium and rHuIL-3 concentration was found to be 80.03 ± 8.24 mg/l. The maximum identified by the neural network model is 10.6% higher than the maximum identified by the multiple regression analysis. This difference indicates that solution obtained by multiple regression analysis with poor modeling capacity is not guaranteed to be optimum. Therefore, the ability of the model to approximate the true

response surface with a high degree of accuracy is of the key importance in optimization step.

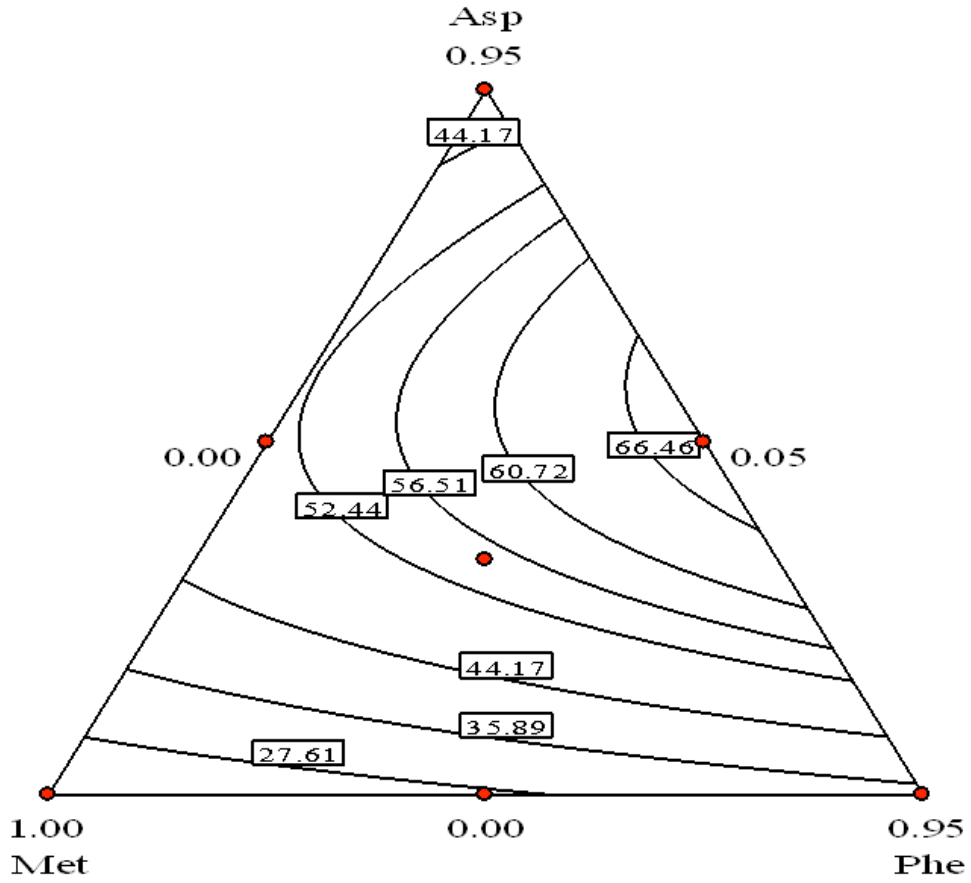

Figure 4.6: Contour plot of rHuIL-3 production (mg/l) over simplex region

Table 4.7: Maximum rHuIL-3 concentration and the optimum input sets

Model	rHuIL-3	Mole fraction			
	(mg/l)	Asp	Leu	Met	Phe
Multiple regression	67.95	0.53	0.00	0.05	0.42
Neural network	76.00	0.56	0.00	0.05	0.39

The recombinant host strain employed in this study has an absolute requirement for methionine. In cells methionine has multiple functions. Besides being an amino acid residue in proteins, it acts as an initiator of protein synthesis and is a methyl donor in several biochemical reactions. The genes leading to methinonine synthesis in *Streptomyces lividans* are not very well known, but the disruption of the methylenehydrofolate gene (metF) has been reported to lead to methionine dependence (Blanco J.and Coque J.J. and Martin J.F., 1998). The beneficial effect of aspartic acid in the medium on the growth of Streptomyces has been observed previously (Cheng Y.R. et al., 1995). The overproduction of recombinant protein is known to impose a metabolic burden on the host. It appears that aspartic acid provides a convenient entry into the Krebs cycle and acts as a supplementary energy source. The role of aspartic acid as an energy supplement is supported by our findings that the optimal level of aspartic acid is replaceable with a glutamic acid/aspartic acid mixture with nearly identical results. Phenylalanine provides a ready-made aromatic ring and is a known precursor for the synthesis of tyrosine. The important aspect of this study is not with regard to the specific acid requirements of *Streptomyces lividans* for the optimal production of a recombinant protein. Rather, this study presents a novel, statistics-based methodology for the rapid identification of the key amino acids and the determination of their optimal composition for producing a recombinant product. The methodology can be applied for any other organism.

4.6 Conclusion

Starvation trials used in the first step were an efficient approach to provide an initial screen concerning which amino acid affects rHuIL-3 production and cell growth. Mixture experiments and distance-based multivariate analysis used in the second step to rank the essential amino acids and finally another mixture experiment was used to find the optimal composition of the medium. The final composition of the chemically defined minimal medium to produce rHuIL-3 by *S. lividans* after the three steps is as follows: 0.25 g of an amino acid mixture comprising 42% Asp, 5% Met, 53% Phe, 10.6 g K_2HPO_4, 5.3 g KH_2PO_4, 2.5 g $(NH_4)_2SO_4$, 226 μl PPG

2025, 100 ml of 16% (w/v) glucose, 5 ml of 20% (w/v) $MgSO_4.7\,H_2O$, and 1 ml trace element solution. Evidently, the quantities of the components may be individually manipulated to achieve a desired nutrient limitation regarding growth yield.

The hybrid neural network-sequential quadratic programming algorithm approach is described in this work. This novel methodology serves as a viable alternative to the traditional multiple regression approach for modeling and optimization of fermentation media. The input space of a neural network model can be optimized using a sequential quadratic programming algorithm. It appears that neural networks provide a better fit to experimental data than multiple regression models. Starvation trials used in the first step were efficient to provide initial screening for those amino acids which affect rHuIL-3 production and cell growth.

Chapter 5

Metabolic flux analysis

5.1 Introduction

Human recombinant proteins, rather than extracts from plasma, are increasingly used today in pharmaceutical applications due to the lower immune response and reduced risk of transmission of disease. Naturally secreting bacteria such as *Streptomyces* spp. are promising hosts for the expression of products in which glycosylation is not crucial (Sathyamoorthy M. et al., 1996). Among the streptomycetes, *Streptomyces lividans* with its reduced restriction-modification barrier and low level of endogenous peptidase is particularly suitable for the expression of recombinant proteins (Brawner M.E., 1994).

The flux distribution in metabolic pathways determines the efficiency of the production process. Nutrients are partly used for biomass and product synthesis as well as energy supply. In addition, they are also being used for the synthesis of side products and energy dissipation in futile cycles. In order to optimize the capacity of microbial metabolism and increase yields the intracellular metabolite fluxes should be quantified. Metabolic flux analysis which only requires the measurement of extracellular metabolites has been proposed as a means to determine the flow through primary metabolic pathways, identify the energetic parameters and determine maximum theoretical yields (Naeimpoor F. and Mavituna F., 2000; Vallino J.J. and

Stephanopoulos G., 1990, 1993; Varma A. and Palsson B.O., 1994). With the measurement of only a limited number of extracellular metabolites, it is theoretically possible to measure and control metabolic conversions on the level of individual enzymatic reactions by metabolic flux analysis methods (van Gulik W.M. and Heijnen J.J., 1995). The analysis is based on network stoichiometry and conservation of mass. A great advantage of the stoichiometric approach is that it forms a set of linear algebraic equations at steady state, which enables linear optimization.

Although metabolic flux analysis has been applied for numerous microorganisms, there have been only two reports applying metabolic flux analysis to Streptomyces lividans. Daae and Isson (Daae E.B. and Ison A.P., 1999)developed a sensitivity analysis method based on stoichiometry only, and using their method they were able to analyze the impact of the perturbations in the measured fluxes on the unmeasured fluxes and also the impact of the changes in biomass composition on the calculated metabolic reactions. Avignone Rossa (Avignone Rossa C. et al., 2002) applied the metabolic flux analysis technique to carbon metabolism of S. lividans, producing two different antibiotics, the polyketide actinorhodin and the tripyrrole undecylprodigiosin in chemostat cultures and showed that the production of both antibiotics were negatively correlated with PP pathway. We applied the technique of metabolic flux analysis to carbon and nitrogen metabolism of S. lividans producing recombinant human interleukine-3 (rHuIL-3). The biochemical network developed was underdetermined and the system could be solved only by applying several constraints and using an optimization method. The linear programming approach with two different objective functions was utilized to obtain the intracellular metabolic flux distribution of *Streptomyces lividans 66* producing the recombinant human interleukin-3 (rHuIL-3). In previous studies the production of rHuIL-3 by *S. lividans* has been shown to be strictly growth associated (Yahya, 2003). Therefore it is convenient to optimize the growth rate in order to optimize recombinant protein productivity. The main objective was to develop a stoichiometry-based model to simulate the behavior of the *S. lividans*, to determine the maximum theoretical growth rate and to predict the effect of any change in some fluxes on the behavior of the cell. Also, another objective function, minimizing redox

potential production, was used for metabolic analysis. The topology of the system under two different objectives were compared.

5.2 Matrials and methods

Microorganism

The recombinant *Streptomyces lividans* NCIMB 11416/IL3 p002, was kindly provided by Cangene Corporation, Winnipeg, Manitoba, Ontario, Canada. This strain was derived from *Streptomyces lividans 66* as designated by the John Innes Institute and contains multiple copies of the plasmid encoding the rHuIL-3 gene with a thiostrepton-resistance marker. Cangene's patented production technology CANGENUS^{TM} was designed so that the product is constitutively over-expressed and secreted by the host in high levels relative to other naturally secreted components. The expressed recombinant protein is non-glycoslated, correctly folded, homogenous and bioactive.

5.2.1 Seed stock preparation

Streptomyces inocula were grown in 500 ml baffled flasks containing 12.5 mg/l thiostrepton in seed medium at a volume ratio of 5:1 for 24 hr, with temperature regulated at 30 - 32°C and agitation at 240 rpm on an Innova 4330 refrigerated incubator shaker (New Brunswick Scientific Co., NJ, USA). The mycelium was harvested after 24 hr by centrifugation at 5000 g for 5 min and resuspended in fresh sterile seed medium to 50% of the original suspension volume. 20% (w/v) sterile skim milk was added bringing the volume to the original culture volume. The mixture was thoroughly mixed and fractionated into 1 ml aliquots and stored at - 80°C until required for inoculum preparation.

5.2.2 Medium composition

The chemically defined medium used for growth studies had the following composition (per liter of de-ionized water): 10.6 g K_2HPO_4, 5.3 g KH_2PO_4, 2.5 g $(NH_4)_2SO_4$, 226 μl polypropylene glycol (PPG) 2025, 100 ml of 16% (w/v) glucose, 5 ml of 20% (w/v) $MgSO_4$, and 1 ml trace elemenat solution. The pH was adjusted to 6.7 before autoclaving for 20 min at 121°C. Glucose and magnesium sulfate solutions were autoclaved separately. Trace element solution consisted of (per liter of de-ionized water): 40 g $Na_2EDTA.2H_2O$, 11 g $CaCl_2.2H_2O$, 7 g $FeSO_4.7H_2O$, 2 g $MnCl_2.4H_2O$, 2 g $ZnSO_4.7H_2O$, 0.4 g $CuSO_4.5H_2O$, 0.4 g $CoCl_2.6H_2O$. The amino acid mixture used in this study was based on the minimal defined medium designed by Nowruzi et. al. (Nowruzi K. et al., 2008) with some modifications. The composition of amino acid mixture used in medium was (% by weight): aspartic acid 37, glutamic acid 41, methionine 5, and phenylalanine 17. A total amount of 1 g/l of the amino acid mixture was dissolved in deionized water and the resulting solution was sterilized by filtration through a 0.2 μm filter (Millipore) at room temperature.

5.2.3 Fermentation

100 μl of the thawed seed culture was used to inoculate 100 ml of the fermentation medium containing 10 mg/l of thiostrepton in a 500 ml baffled flask. The culture was grown under exactly the same conditions as for the seeds. This culture was used as inoculum for the fermenter.

Batch fermentation experiments were performed in either a 2 l Bioflo fermenter (New Brunswick Scientific Co., NJ, USA) with a working volume of 1.5 l or a 7 l Applikon bioreactor (Applikon Dependable instruments, Holland) with the working volume of 4.5 l using the medium described above. The bioreactor cultures were inoculated with a 5% (v/v) inoculum size. The culture temperature was kept at 32 ± 0.5°C. Air was supplied at a rate of 1 vvm and was sterilized through a hydrophobic 0.2 μm membrane filter (Millipore). The broth was agitated using three six bladed Rushton turbine impellers rotating at 200 - 400 rpm. The dissolved oxygen

was monitored and kept above 40% of saturation value using the variable agitation scheme. Carbon dioxide and oxygen concentrations were monitored on line at the inlet and outlet of the fermentor.

5.3 Theoretical considerations

5.3.1 Model construction

The metabolic model for the growth of *S. lividans* has been developed using the published information for this microorganism. The existence of Embden-Meyerhof pathway (EMP) and pentose phosphate pathway (PP) were reported by Salas et al. (Salas J.A. et al., 1984) for related species, namely *S.coelicolor, S. antibiotics, S. scabies, and S. reticuli.* Therefore, it was reasonable to assume that EMP and PP are functioning in *S. lividans* as well. Since several studies failed to produce any evidence for existence of Entener-Doudoroff (ED) pathway in any streptomycete, this pathway was not included in the metabolic model (Alves A.M.C.R. et al., 1994; Dekleva M.L. and Strohl W.R., 1988; Kieser T. et al., 2000). During growth on amino acids, the microorganism must transfer carbon from the malate/oxaloacetate pool to the phosphoenolpyruvate pool. This indicates that gluconeogenic pathway should be active as well (Roubos, 2002). It has been shown that pyruvate phosphate dikinase is responsible for the lower part of the gluconeogenic reaction in *Propionibacterium* and *Microbispora* when grown on carbon which enters the TCA cycle (Eisaki N. et al., 1999; Evans H.J. and Wood H.G., 1971). In this work pyruvate is assumed to be synthesized by an NAD-dependent malate dehydrogenase and pyruvate is subsequently converted to phosphoenolpyruvate by pyruvate dikinase. Dekleva and Strohl (1988) proved that a complete tricarboxylic acid (TCA) cycle exists in both *S.lividans* and *Streptomyces C5.* They, however, failed to detect glyoxylate pathway enzyme activity in *S. lividans, Streptomyces C5* and *S. aurofaciens.* The enzyme phosphoenolpyruvate carboxylase has been identified as the main anaplerotic enzyme in several streptomycetes (Bramwell H. et al., 1993; Dekleva M.L. and Strohl W.R., 1988). This enzyme is therefore

likely to be the case for the *S. lividans* as well. The main gluconeogenic step in *S. parvulus* was identified to be a reaction converting oxaloacetate to phosphoenolpyruvate (Inbar L. and Lapidot A., 1991). *Streptomyces* species contain both NADH dehydrogenase I and II and it is not clear whether both are functioning during exponential growth. In many microorganisms, the presence of NADH dehydrogenase II activity seems to be associated with rapid carbon conversion by fermentation as shown by *S. cerevisiae, C. glutamicum, B. subtillus* and *E. coli* (Roubos, 2002). One may assume that the lack of fermentation products by *S. lividans* is an indication that NADH dehydrogenase I is active and NADH dehydrogenase II is inactive. NADPH is needed for a number of anabolic reactions, but is normally required by a small number of enzymes, e.g. glucose-6-phosphate dehydrogenase, 6-phosphogluconate dehydrogenase, and isocitrate dehydrogenase. Both NADPH- and NADH-requiring 6-phosphogluconate dehydrogenase have been reported for various *Streptomyces* (Dekleva M.L. and Strohl W.R., 1988). Similar results have been found for glucose-6-phosphate dehydrogenase from various streptomycetes (Neuzil J et al., 1986, 1988). In this study it was assumed that NADPH needed for biosynthesis is mainly produced by the PP pathway. As in other organisms, adenosine triphosphate (ATP) is the main energy carrier for intracellular energy transfer, but several biochemical reactions are facilitated by pyrophosphate instead of ATP. Furthermore, it is known that polyphosphate can be used as an ATP substitute or as an energy source for the phosphorylation of AMP or NAD^+ (Blum E. et al., 1997; Kornberg A., 1995). It may be assumed that pyrophosphate hydrolysis is equivalent in energy to the formation of 1 ATP. Depending on the use of NADH dehydrogenase I or II the maximal ATP yield per oxygen, P/O ratio, is 2.5 or 1.5 (Nielsen J. et al., 2002).

A single reaction based on fixed biomass composition was employed for the formation of the biomass from monomers. It accounts for amino acid integration into various intracellular components ranging from protein to cell wall. The energy required for metabolic functions other than growth is also taken into account. These functions include, maintenance of transmembrane gradients, macromolecular turnover, cellular motility and maintaining cellular osmolarity. Since it has proven difficult to quantify each on of these functions individually, a

phenomenological measure known as maintenance coefficient has gained widespread use (Marr A.G. et al., 1963; Pirt S.J., 1965). Most maintenance requirements are energy related and may therefore be termed as ATP maintenance requirements. Maintenance represents a significant drain on metabolic resources at low growth rates. A schematic outline of the major metabolic pathways is depicted in Fig. 5.1. The complete set of reactions of the metabolic network is given in Appendix A.

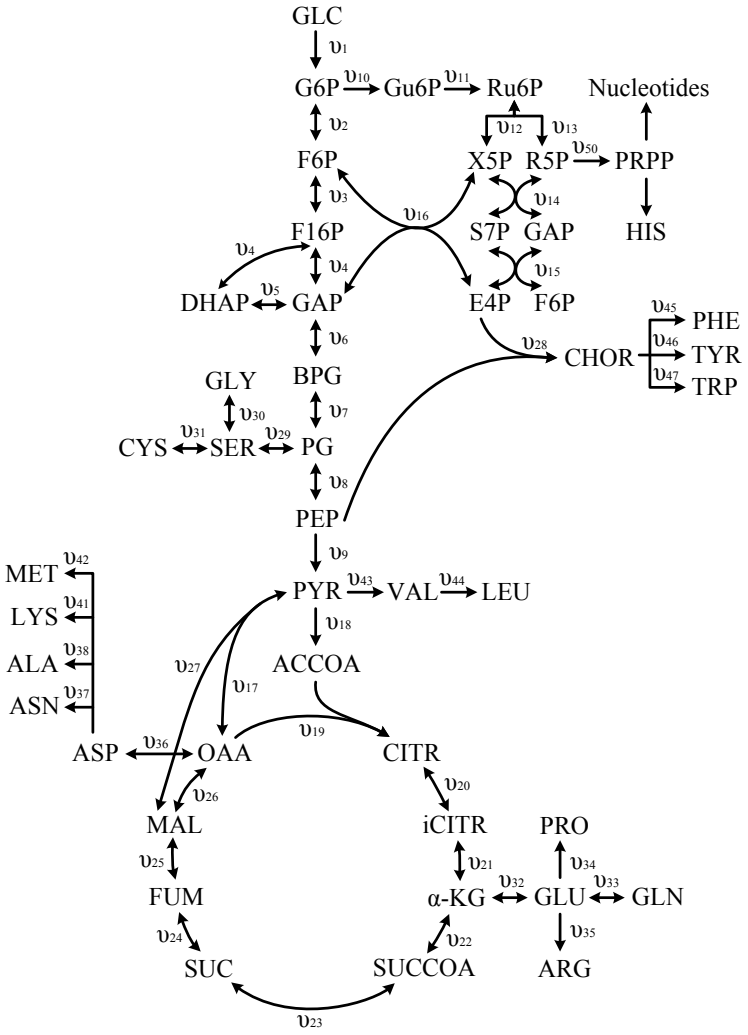

Figure 5.1: The general overview of the proposed metabolic pathways for *S. lividans*

Metabolic balancing

The transient material balances for all the components in a metabolic network containing n metabolites and m metabolic fluxes is given as follows:

$$\frac{d\boldsymbol{\psi}}{dt} = \mathbf{S} \cdot \boldsymbol{\nu} - \mathbf{R} \tag{5.1}$$

where $\boldsymbol{\psi}$ is an n-dimensional vector of metabolite mass per cell, $\boldsymbol{\nu}$ is the vector of m metabolites, $\mathbf{S}$ is an $m \times n$ stoichiometric matrix, m is the number of metabolites and n is the number of fluxes. $\mathbf{R}$ is the vector of the uptake and production rates of all the intracellular and extracellular metabolites, which can be calculated from experimental measurements. The element S_{ij} is the stoichiometric coefficient of metabolite i in reaction j. A negative value of S_{ij} indicates that metabolite i is consumed in reaction j, whereas a positive value of S_{ij} indicates that metabolite i is produced. Assuming pseudo-steady state conditions and negligible dilution effects for growth (Stephanopoulos G.N. et al., 1998), the transient mass balance can be simplified to consider only steady-state behavior as follows:

$$\mathbf{S} \cdot \boldsymbol{\nu} = \mathbf{R} \tag{5.2}$$

The mass balance equations for the fermentation can be written in the matrix form as follows:

$$\frac{d\boldsymbol{\psi}(t)}{dt} = \mathbf{R}\mathbf{X}(t) \tag{5.3}$$

To calculate the values of $\mathbf{R}$ for the extracellular metabolites from experimental data Eq. 5.3 is integrated as follows:

$$\begin{gathered} \int_0^t d\boldsymbol{\psi}(t) = \mathbf{R}\int_0^t \mathbf{X}(t)dt \Rightarrow \\ \boldsymbol{\psi}_t - \boldsymbol{\psi}_0 = \mathbf{R}(CH_t - CH_0) \end{gathered} \tag{5.4}$$

where CH is the cumulative volumetric cell-hour defined as (Dutton R.L. et al., 1998):

$$CH = \int_0^t X dt \tag{5.5}$$

As shown in Eq. 5.4 the slope relating the metabolite concentration to the cumulative volumetric cell-hours is the instantaneous value of $\mathbf{R}$. The use of the cumulative volumetric cell-hours concept allows for assessment of the productivity and the biological capacity for production on the same cumulative basis (Gao J. et al., 2007). The volumetric cell-hours can be calculated by a simple integration technique, using the logarithmic mean of the cell density between data points as follows (Dutton, 1998):

$$CH = \sum_{0}^{t} \left\{ \frac{X_{t+\Delta t} - X_t}{\ln(X_{t+\Delta t}/X_t)} \times (t_{t+\Delta t} - t_t) \right\} \tag{5.6}$$

Finally, the fluxes $\boldsymbol{\nu}$ can be calculated from the reaction rates $\mathbf{R}$ by using the matrix equation (Eq. 5.2). Since the number of unknown intracellular fluxes are greater than the number of metabolites i.e. $m < n$, Eq. 5.2 represents an underdetermined system of linear algebraic equations with a potentially infinite solutions. Consequently, the number of feasible flux distributions allowed by this equation is infinite. The flux distribution can be estimated by formulating a suitable objective function and using linear programming (Hillier F.S. and Lieberman G.J, 2001). The vector $\boldsymbol{\nu}$ can be partitioned into two vectors, one containing unknown internal fluxes and one containing unknown exchange fluxes (Gheshlaghi R. et al., 2007). Hence,

$$\mathbf{S^I} \cdot \boldsymbol{\nu}^{\mathbf{I}} + \mathbf{S^E} \cdot \boldsymbol{\nu}^{\mathbf{E}} = \mathbf{R} \tag{5.7}$$

Vector $\boldsymbol{\nu}^{\mathbf{I}}$ consists of n^I unknown internal fluxes and vector $\boldsymbol{\nu}^{\mathbf{E}}$ consists of n^E unknown exchange fluxes. $\mathbf{S^I}$ and $\mathbf{S^E}$ are the columns of $\mathbf{S}$ corresponding to $\boldsymbol{\nu}^I$ and $\boldsymbol{\nu}^E$, respectively. The linear programming is formulated as follows:

$$\begin{aligned} Min \quad & Z = \textstyle\sum c_j \nu_j \\ s.t. \quad & \mathbf{S^I} \cdot \boldsymbol{\nu}^{\mathbf{I}} + \mathbf{S^E} \cdot \boldsymbol{\nu}^{\mathbf{E}} = \mathbf{R} \\ & (1 - \epsilon_i) R_i \leq R_i \leq (1 + \epsilon_i) R_i \\ & 0 \leq \nu_j^I \leq \alpha_j \\ & \nu_{i,min}^E \leq \nu_i^E \leq \nu_{i,max}^E \end{aligned} \tag{5.8}$$

where c_j is a coefficient expressing to the importance of each flux in the objective function. Due to the characteristics of the Simplex algorithm, the elements of the solution vectors, $\boldsymbol{\nu}^{\mathbf{I}}$ and

$\boldsymbol{\nu}^{\mathbf{E}}$, will always be non-negative. Consequently, reversible reactions must be formulated as two separate reactions, in opposite directions. The third set of the constraints in Eq. 5.8 states this property for unknown internal fluxes. The first set of constraints is the steady state material balances, the second set is associated with any possible errors involved in the measured fluxes, and the last set of constraints implies upper and lower bounds on each exchange flux. In this model, the elements of $\mathbf{R}$ are defined to be positive if the metabolite was exiting the system and negative if it was entering the system. The mathematical model was programmed in the GAMS (General Algebraic Modelling System) environment which is used for optimization purposes (Rosenthal R.E., 2008).

The calculated fluxes are based on estimates from experimental measurements of exchange fluxes. It is therefore pertinent to calculate the sensitivity of the objective function Z with respect to perturbations in the i^{th} measured exchange flux. This could be done using the shadow price of the linear programming which is defined as follows (Pannel D.J., 1997):

$$\lambda_i = \frac{\partial Z}{\partial b_i} \tag{5.9}$$

The shadow prices were calculated from the mathematical duality of the primary linear optimization problem. Since the exchange fluxes vary significantly in their absolute value the relative or logarithmic sensitivity coefficient was used for comparison purposes (**?**). The logarithmic sensitivity of the objective function in response to changes in i^{th} measured flux can be expressed as follows (Gheshlaghi R. et al., 2007):

$$\Lambda(Z, b_i) = \frac{\partial Z/Z}{\partial b_i/b_i} = \frac{\partial \ln Z}{\partial \ln b_i} = \frac{b_i}{Z}\lambda_i \tag{5.10}$$

As shown in Eq. 5.10, to compute the logarithmic sensitivity coefficient, shadow price values, the measured exchange fluxes and the optimal value of the objective function are needed.

5.4 Experimental results and discussion

Three fermentations were performed under identical conditions and the averages of the three sets of observations were used as shown in Figs. 5.2, 5.3 and 5.4. The profiles were found

to be reproducible, since fermentation repeats produced almost identical results. During the first phase of fermentation, designated as growth Phase E1, immediately after inoculation the biomass increases in an exponential manner with $\mu = 0.078\,\mathrm{h}^{-1}$. Thereafter, the exponential growth rate decreases from $\mu = 0.078\,\mathrm{h}^{-1}$ to $\mu = 0.035\,\mathrm{h}^{-1}$, as observable in the biomass plot. The start of this growth phase (Phase E2) is estimated from logarithmic plot of biomass as shown in Fig. 5.5. The stationary phase is reached when glucose is completely exhausted. The elemental balances for carbon and nitrogen were calculated to verify the measurements. During Phase E1, aspartic acid and glutamic acid were used up completely and the concentration of phenylalanine, glucose and ammonia decrease by 50%, 9% and 4%, respectively. The complete depletion of aspartic acid and glutamic acid in fermentation medium corresponds to end of the Phase E1 and start of growth Phase E2, which indicates a switch in cellular metabolism.

5.4.1 Metabolic flux distribution

The extracellular metabolite fluxes were calculated from the metabolic concentration profiles for each phase. In order to verify the results of each simulation, the oxygen uptake rate and carbon dioxide evolution rates were set to be determined by the optimization program. Two different objective functions were examined:

1. Optimizing biomass production by maximizing the specific growth rate,
2. Optimizing with respect to redox metabolism by minimizing excess NAD(P)H production.

The mathematical form of the objective functions is shown in Table 5.1, and the results from these calculations and experimental measurements are summarized in Table 5.2 and discussed below.

The logarithmic sensitivities of the objective functions with respect to nutrients were determined for different periods of the fermentation and presented in Table 5.3.

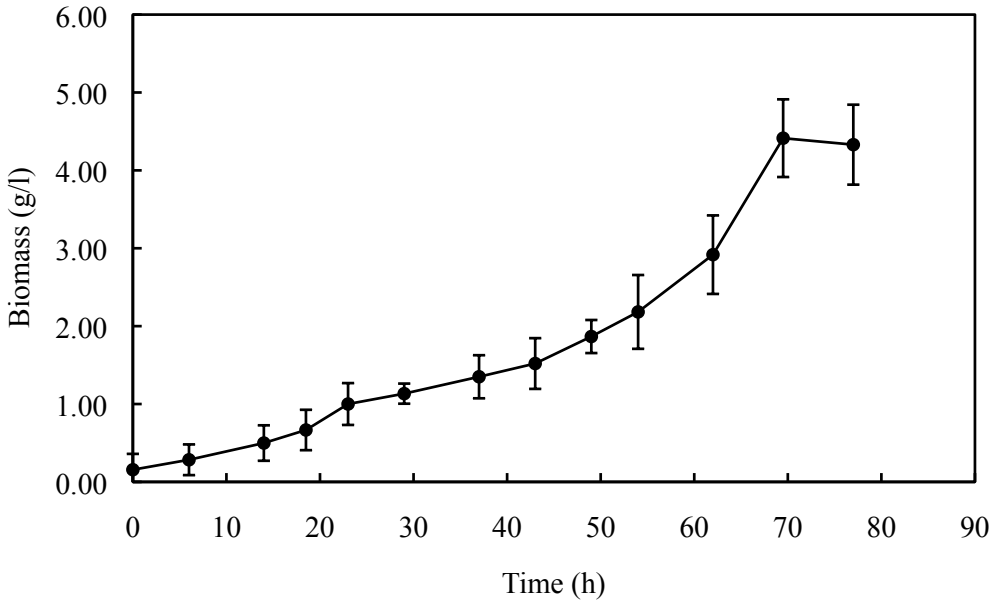

Figure 5.2: Biomass and rHuIL-3 concentration profiles during fermentation

Maximizing specific growth rate

Phase E1. Metabolic flux distribution at this period yielded a maximum specific growth rate of $\mu = 0.072\,\mathrm{h}^{-1}$ which deviates 8% from the measured value as given in Table 5.2. The specific oxygen uptake rate and specific carbon dioxide evolution rate was determined and compared to measured rates, showed 23% and 19% error, compared to the measured rates, respectively. The PP pathway was active and 58% of the glucose was diverted to this pathway. During high NADPH requirement of the cells, the flux through the PP pathway has been reported to be high (Jorgensen H. et al., 1995; Obanye A.I.C. et al., 1996). In most bacteria, glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase are NADP-dependent, therefore, every glucose-6-phosphate entering PP pathway would generate two NADPH. However, in *S. lividans* glucose-6-phosphate dehydrogenase is NADP-dependent but 6-phosphogluconate dehydrogenase has been found to be NAD-dependent (Alves A.M.C.R. et al., 1994; Dekleva M.L. and Strohl W.R., 1988). This means that only one NADPH is generated for every glucose-6-

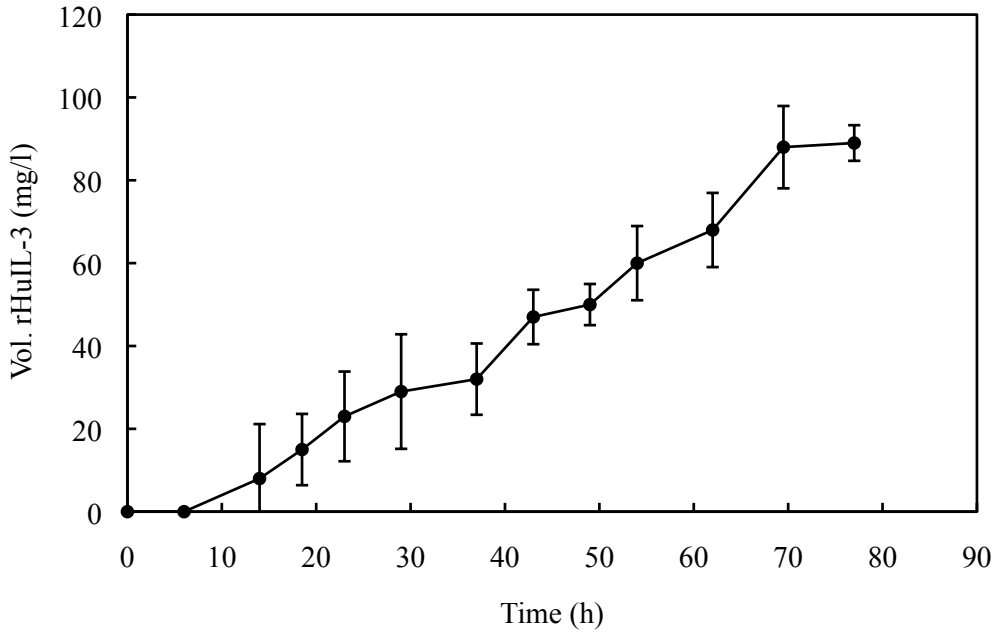

Figure 5.3: Biomass and rHuIL-3 concentration profiles during fermentation

phosphate entering PP pathway, therefore, this pathway needs to be more active in *S. lividans*, to meet the biosynthetic demand for NADPH. The net amount of redox potential produced was 5.81 mmol/g_{DW}hr. This includes the functioning of both NADH and NADPH. Since aspartic acid and glutamic acid directly enter the TCA cycle after deamination, the latter part of the gluconeogenic pathway is active and 30% of malate produced in TCA cycle is estimated to converted to pyruvate by malic enzyme. The flux distribution through the major pathways is shown in Figs. 5.6, 5.7 and 5.8. The total amount of ATP produced during this stage was 14.06 mmol/g_{DW}hr, which 28% was used for cellular maintenance such as transport, repair, heat production, and secondary metabolism.

Among the nutrients, aspartic acid, glutamic acid and ammonia were found to have positive shadow prices and phenylalanine and glucose were calculated to have zero shadow prices, indicating that aspartic acid, glutamic acid and ammonia can be used to improve the growth

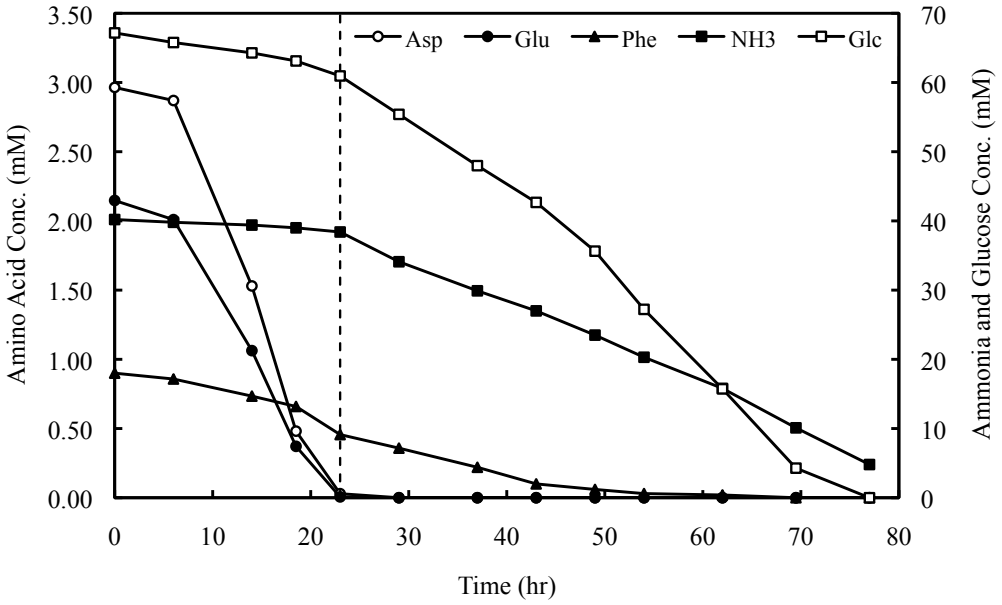

Figure 5.4: Amino acids and ammonia concentration profiles during the fermentation

during this period but phenylalanine and glucose have no effects on the growth during Phase E1. These results indicate that during Phase E1, cell growth was limited by nitrogen uptake, not by glutamic glucose. Moreover, all the remaining amino acids except histidine and phenylalanine, and all nucleotides other than IGP had positive shadow prices, indicating they can be used to improve the growth. Furthermore, examination of the computed logarithmic sensitivities led to some important observations. Although aspartic acid, glutamic acid and ammonia had the same shadow prices, their logarithmic sensitivities were different, so any changes in their uptake rate would have different effects on the specific growth rate. The highest value of the logarithmic sensitivity was 0.44 for aspartic acid, indicating that aspartic acid uptake was growth limiting during Phase E1. Quantitatively, this value means that should the uptake rate for aspartic acid increases by 10%, then the specific growth rate would increase by 4.4%.

Phase E2. The flux data corresponding to Phase E2 were used to evaluate the metabolic distribution of this phase. The maximum specific growth rate was determined to be $\mu = 0.028\,\mathrm{h}^{-1}$

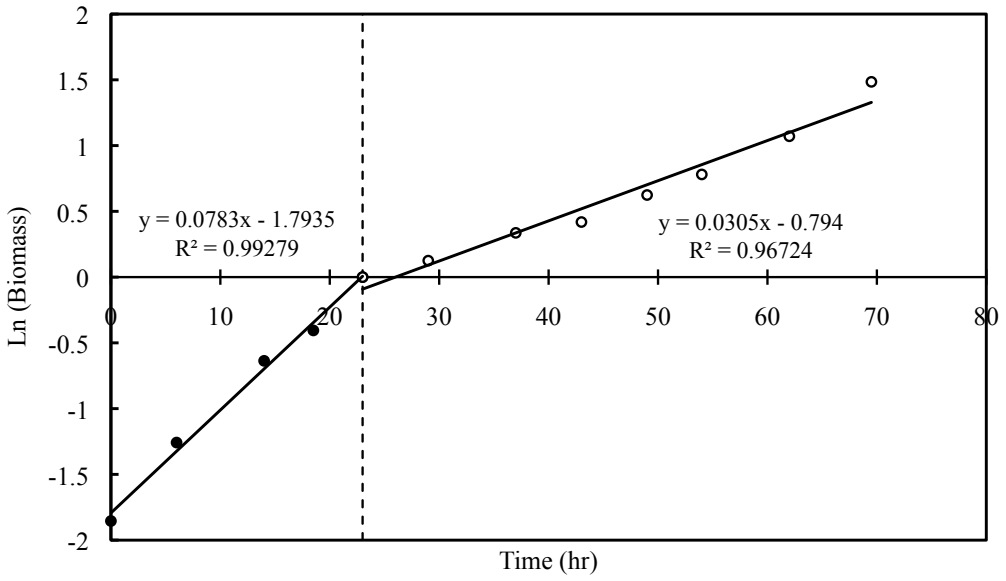

Figure 5.5: Natural logarithm of biomass v.s. time

which was 10% less than the measured value. However, both the specific oxygen uptake rate and carbon dioxide evolution rate were overestimated by 9% and 33% in comparison with experimental data. The carbon flux through the PP pathway decreased to 52% of total carbon flux due to decrease in biomass growth rate. The net amount of redox potential produced was $3.55\,\mathrm{mmol/g_{DW}h}$, which is 41% less than Phase E1. The total amount of ATP produced in this stage was $6.72\,\mathrm{mmol/g_{DW}h}$, which 31% was used for cellular maintenance. Unlike in Phase E1 the gluconeogenic pathway was not active in Phase E2, and both aspartic acid and glutamic acid were produced through the TCA cycle. Aspartic acid was produced by transamination of oxaloacetate and glutamic acid was produced by NADP-dependent glutamate dehydrogenase enzyme.

Similar to Phase E1 all amino acids except histidine and phenylalanine had a positive shadow price. Histidine had a negative shadow price and phenylalanine had a zero shadow price. Again glucose had a zero shadow price meaning that its uptake rate can not be used to increase the specific growth rate. Ammonia had a logarithmic sensitivity of 1, indicating that ammonia is

Table 5.1: Mathematical form of the objective functions used

Objective function	Mathematical form
Maximizing specific growth rate	ν_{64}
Minimizing redox potential	$\nu_6 + \nu_{10} + \nu_{11} + \nu_{18} + \nu_{21} + \nu_{22} + \nu_{26} + \nu_{27} + \nu_{29} + \nu_{31} + r_{32} + 2\,r_{39} + \nu_{44} + \nu_{46} + \nu_{52} + \nu_{57} + \nu_{59} + \nu_{60}$

Table 5.2: Experimental and simulation results obtained using different objective functions at different phases of fermentation

	Experimental		**Maximizde μ**		**Minimize NAD(P)H**	
	Phase E1	Phase E2	Phase E1	Phase E2	Phase E1	Phase E2
$\mu\ (h^{-1})$	0.078	0.031	0.072	0.028	0.070	0.027
$r_{O_2}\ (mmol/g_{DW}h)$	2.19	0.89	2.83	0.98	2.82	0.96
$r_{CO_2}\ (mmol/g_{DW}h)$	1.35	0.40	1.67	0.60	1.63	0.60

growth limiting in this stage.

Minimizing the production of excess redox potential

Phase E1. The objective of minimizing NAD(P)H production may be realistic, since it would not be beneficial for the cell to have an excess of oxidizing power. The results obtained with this objective function was almost the same as the objective previously discussed. Metabolic flux distribution at Phase E1 yielded a specific growth rate of $\mu = 0.070\,\mathrm{h}^{-1}$ which is 10% less than the measured value as given in Table 5.2. The specific oxygen uptake rate and specific carbon dioxide evolution rate was determined and showed 22% and 17% error, compared to the measured rates, respectively. The PP pathway was active and 56% of the glucose was

Table 5.3: Logarithmic sensitivities of the nutrients at different stages for objective functions

	Maximize μ		**Minimize redox potential**	
Nutrient	Phase E1	Phase E2	Phase E1	Phase E2
Asp	0.44	0	-0.07	0
Glc	0	0	1.22	0
Glu	0.32	0	0.03	0
NH3	0.24	1	-0.2	-0.17
Phe	0	0	0.02	1.17

diverted to this pathway which is 2% less than employing the objective function to maximize the specific growth rate. The optimal redox potential produced was 5.78 $\mathrm{mmol/g_{DW}h}$ which is 0.5% less than the redox potential produced with the specific growth rate as the objective. The total amount of ATP produced in this stage was 12.58 $\mathrm{mmol/g_{DW}h}$ and 25% of the produced amount was used for cellular maintenance purposes. Malic enzyme was active converting malate into pyruvate.

Aspartic acid and ammonia had negative shadow prices, while glucose, glutamic acid and phenylalanine had positive shadow prices. This means that higher uptake rates of glucose, glutamic acid and phenylalanine favours the minimization of the excess redox potential. However, the logarithmic sensitivities of the amino acids were very small compared to glucose and ammonia. Glucose had the highest logarithmic sensitivity of 1.21.

Phase E2. The flux data corresponding to this phase were used to evaluate the metabolic distribution of this phase. The specific growth rate was determined to be $\mu = 0.027\,\mathrm{h^{-1}}$ which was 10% less than the measured value. However, both the specific oxygen uptake rate and carbon dioxide evolution rate were overestimated by 7% and 33% in comparison to the observed experimental data. The carbon flux through the PP pathway was 51% of total carbon flux and was 1% less than the pervious objective. Unlike the Phase E1 results, malic enzyme was

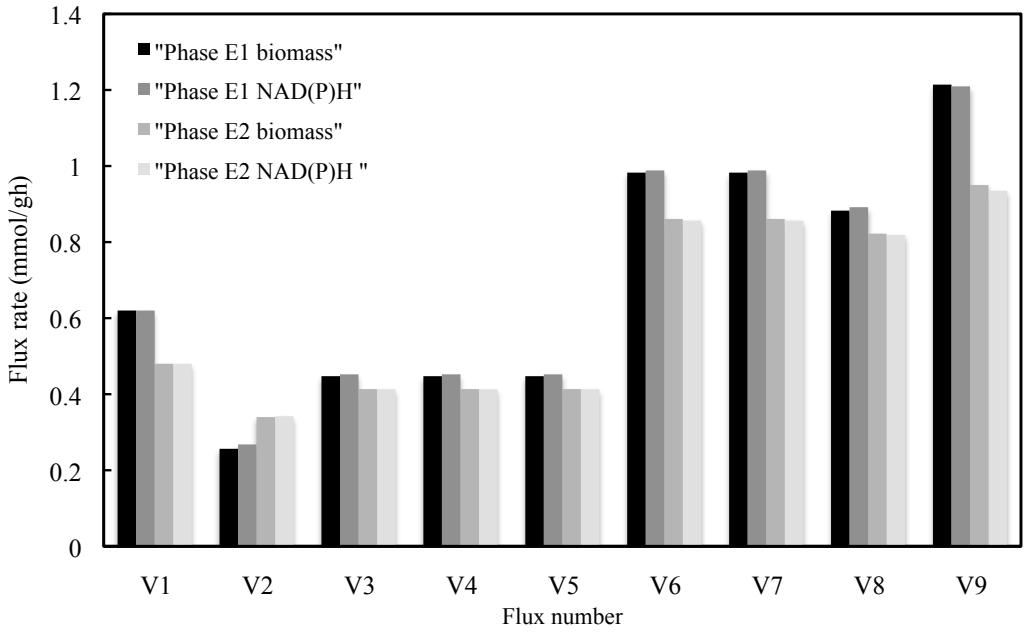

Figure 5.6: Flux distribution in EMP pathway during different stages for the objective functions

not active in Phase E2. The optimal redox potential produced was 3.42 mmol/g_{DW}h which is 40% less than the Phase E1 and 4% less than the amount of redox potential produced when tmaximizing the specific growth rate was the objective. The net amount of ATP produced in this phase was 6.61 mmol/g_{DW}h and 30% of this was consumed for cellular maintenance.

5.5 Discussion

All the major pathways were active throughout the fermentation and the flux distribution was almost the same for the two different objective functions. However, fluxes through the pathways were different during different phases of the fermentation. The relative flux size through the PP pathway depends on the medium composition and the specific growth rate. The results of the simulations showed that the flux through PP pathway increased with the specific growth rate from 51% to 58% during the fermentation. According to Cochrane et. al.

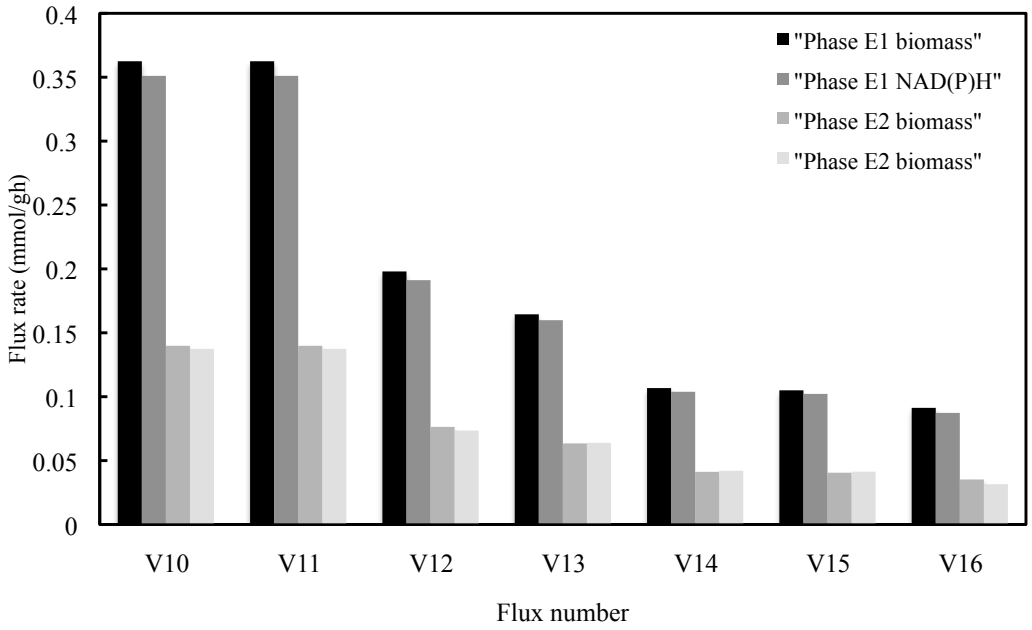

Figure 5.7: Flux distribution in PP pathway during different stages for the objective functions

(Cochrane V.W. et al., 1953), the PP pathway was the primary glucose catabolic route in *S. reticuli* and *S. coelicolor*. Furthermore, it has been reported that during spore germination of *S. antibioticus* the PP pathway for glucose catabolism increases relative to the EMP pathway, and it became the primary metabolic route (Salas J.A. et al., 1984). All this evidence supports the importance of the PP pathway in glucose catabolism and the assumption that NADPH needed for biosynthesis is mainly produced by the PP pathway.

During the growth Phase E1, both aspartic acid and glutamic acid are available for the cell and both enter the TCA cycle directly. This causes malic acid to be active during the first stage of the growth. But this enzyme is inactive during the second stage since aspartic acid and glutamic acid are no more available for the cell.

The energy requirement for maintenance was expressed in terms of ATP hydrolysis. The results showed that the maintenance requirements accounted for up to 31% of the total production depending on the objective function and the fermentation stage.

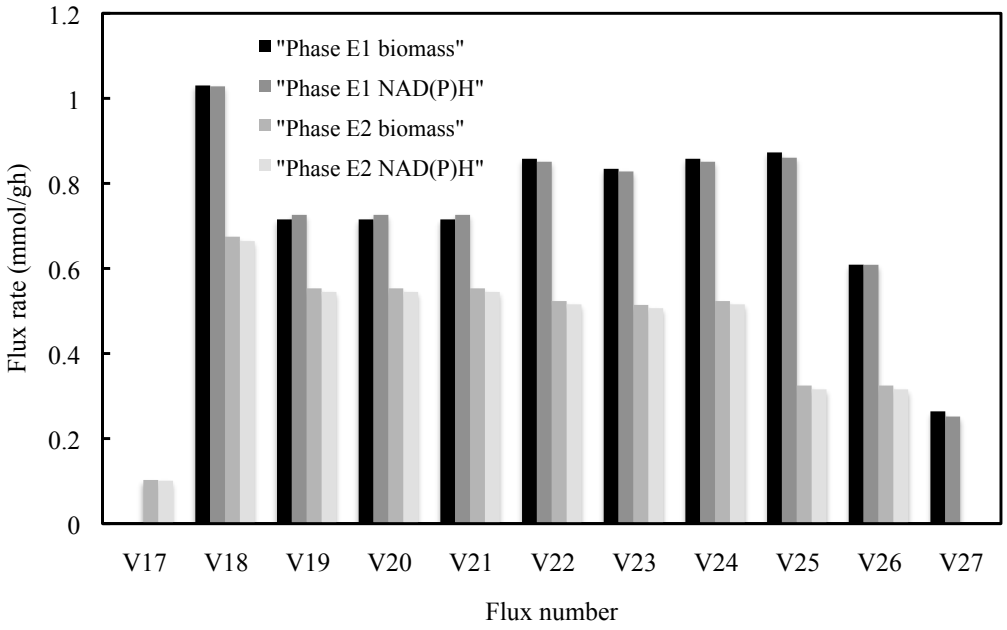

Figure 5.8: Flux distribution in TCA pathway during different stages for the objective functions

As depicted in Fig. 5.5 the microorganism displayed a diauxic growth pattern that was caused by a shift in metabolic pathways in the middle of the growth cycle. The specific growth rate increased rapidly during the first 23 h of the fermentation when the amino acids served as both the main carbon and nitrogen sources. After the amino acids were exhausted, ammonia became the main nitrogen source and the growth rate dropped during this time. The specific growth rate gradually decreased and reached zero at 70 h as the cell concentration reached its maximum value. One could conclude that the low concentration of organic nitrogen sources was the main reason for the slow growth after 23 h, while the depletion of glucose was the reason for cessation of the growth at 70 h. After this time specific growth rate became slightly negative due to cell lysis.

Since monomeric cellular compositions for streptomycetes are currently unavailable, the biomass composition and demands assumed for *S. lividans* were taken from a typical *E. coli* cell

(Daae E.B. and Ison A.P., 1999; Ingraham J.L. et al., 1983). The process of determining the monomeric composition is very laborious and expensive, hence, before committing cost and time to such study one should consider the implications of using an assumed composition taken from *E. coli* cell. This study assessed the implications of this assumption assuming a uniform distribution for each of the biomass components. The range assumed for each biomass component was $\pm 20\%$ of the reported values for *E. coli*. The percentage change in each calculated flux due to changes in each of the biomass components are presented in Table 4. The percent change in each flux was calculated by running the simulation program 15 times for each phase with of the objective functions by randomly choosing a value for each biomass component from an assumed uniform distribution. Then, the calculated fluxes were averaged and a 95% confidence interval was calculated for each flux. The reported values for percent change in each flux in Table 4 are the ratio of the 95% confidence interval to average value of the fluxes calculated by the simulation program.

The changes in flux vary from 0.22% in metabolites such as BPG and PG to 7.91% in MTHF. Changes in ATP production was 0.35% to 0.93% and changes in objective functions, specific growth rate and redox potential were 1.64% to 1.69%, 0.29% to 0.57%, respectively. The highest percentage changes are for the smaller fluxes, therefore, they do not have any significant effect on the objective functions. Therefore, it may be concluded that possible changes in the biomass composition will have little or no impact on the primary metabolic fluxes. This concluded that using a typical *E. coli* cellular monomeric composition to represent *Streptomyces lividans* biomass composition will not significantly affect the calculated fluxes.

Table 5.4: Percentage change in the calculated fluxes due to changes in each biomass component

	Maximize μ		**Minimize redox potential**	
Flux	Phase E1 (%)	Phase E2 (%)	Phase E1 (%)	Phase E2 (%)
ν_1	0.00	0.00	0.00	0.00
ν_2	5.46	2.37	5.09	2.30
ν_3	1.03	0.47	0.98	0.46
ν_4	1.03	0.47	0.98	0.46
ν_5	1.03	0.47	0.98	0.46
ν_6	0.50	0.22	0.49	0.23
ν_7	0.50	0.22	0.49	0.23
ν_8	0.62	0.26	0.59	0.27
ν_9	0.80	0.37	0.78	0.37
ν_{10}	3.87	2.16	3.87	2.13
ν_{11}	3.87	2.16	3.87	2.13
ν_{12}	4.80	2.46	4.80	2.43
ν_{13}	2.82	1.76	2.82	1.73
ν_{14}	4.47	2.37	4.47	2.32
ν_{15}	4.54	2.39	4.54	2.34
ν_{16}	5.20	2.56	5.20	2.60
ν_{17}	-	0.92	-	0.92
ν_{18}	0.58	0.36	0.56	0.41
ν_{19}	1.90	1.03	1.82	1.05
ν_{20}	1.90	1.03	1.82	1.05
ν_{21}	1.90	1.03	1.82	1.05
ν_{22}	1.54	1.06	1.51	1.08

Continued on next page

Table 5.4 – continued from previous page

	Maximize μ		**Minimize redox potential**	
Flux	Phase E1 (%)	Phase E2 (%)	Phase E1 (%)	Phase E2 (%)
ν_{23}	1.60	1.10	1.57	1.12
ν_{24}	1.54	1.06	1.51	1.08
ν_{25}	1.66	1.00	1.64	1.02
ν_{26}	2.54	1.00	2.47	1.02
ν_{27}	0.93	-	0.95	-
ν_{28}	3.97	3.97	3.97	4.32
ν_{29}	2.43	2.43	2.43	2.42
ν_{30}	2.95	2.95	2.95	2.97
ν_{31}	5.99	5.99	5.99	5.94
ν_{32}	-	0.53	-	0.52
ν_{33}	2.31	2.31	2.31	2.36
ν_{34}	7.06	7.06	7.06	7.01
ν_{35}	4.69	4.69	4.69	4.70
ν_{36}	3.75	1.20	3.32	1.20
ν_{37}	5.38	5.38	5.38	5.33
ν_{38}	5.79	5.79	5.79	5.79
ν_{39}	4.32	4.32	4.32	4.30
ν_{40}	6.60	6.60	6.60	6.63
ν_{41}	6.16	6.16	6.16	6.15
ν_{42}	7.63	7.63	7.63	7.63
ν_{43}	4.15	4.15	4.15	4.15
ν_{44}	6.48	6.48	6.48	6.45
ν_{46}	5.04	5.04	5.04	5.00
ν_{47}	6.83	6.83	6.83	6.79

Continued on next page

Table 5.4 – continued from previous page

	Maximize μ		**Minimize redox potential**	
Flux	Phase E1 (%)	Phase E2 (%)	Phase E1 (%)	Phase E2 (%)
ν_{48}	7.76	7.76	7.76	7.72
ν_{49}	6.18	-	5.61	-
ν_{50}	2.02	2.02	2.02	2.03
ν_{51}	3.71	3.71	3.71	3.75
ν_{52}	3.90	3.90	3.90	3.90
ν_{53}	7.18	7.18	7.18	7.16
ν_{54}	6.71	6.71	6.71	6.69
ν_{55}	7.76	7.76	7.76	7.72
ν_{56}	2.92	2.92	2.92	2.94
ν_{57}	6.37	6.37	6.37	6.40
ν_{58}	6.18	6.18	6.18	6.18
ν_{59}	2.60	2.60	2.60	2.64
ν_{60}	7.91	7.91	7.91	7.90
ν_{61}	0.97	0.44	0.95	0.47
ν_{62}	1.46	1.05	1.44	1.07
ν_{63}	6.40	5.02	6.07	5.03
ν_{65}	0.71	0.40	0.70	0.42
ν_{66}	1.05	0.58	1.02	0.61
μ	1.69	1.69	1.69	1.64
NAD(P)H	0.57	0.29	0.55	0.33
ATP	0.93	0.45	0.90	0.35

5.6 Conclusion

A flux based approach was performed for analysis of the metabolic network of *S. lividans*. The metabolic network included both degradative and biosynthetic reactions. The network consisted of 91 biochemical reactions and 62 intracellular metabolites. It included glycolysis (EMP), pentose phosphate (PP) pathways, tricarboxylic acid cycle (TCA), anaplerotic reactions, the metabolism of amino acids, biosynthesis of nucleotides and biosynthesis of biomass. The proposed model along with the experimental measurements was used to construct the stoichiometric matrix corresponding to the process with unknown fluxes. Linear programming with experimental constraints was used to determine the the optimized specific growth rate and redox potential at different stages of fermentation. The bioreaction network presented in this study is a general pathway which can be valid for any bacterial cell with some minor modifications. This model can be used to find the theoretical metabolic distribution and clarify the metabolic behavior of the microorganisms along with experimental data. In this study, two objectives were considered, optimizing biomass production by maximizing specific growth rate, and optimizing redox metabolism by minimizing excess NAD(P)H production. The model developed in this work was able to predict the specific growth rate very accurately with a maximum error of 10%. Moreover, the oxygen uptake rate and carbon dioxide evolution rate were evaluated with maximum error of 27% and 35%, respectively. It should be noted that the off-gas analyzer was not able to display the levels of oxygen and carbon dioxide in outlet gas accurately, consequently the experimentally measured values of these two metabolites were less accurate.

Sensitivity analysis revealed that amino acids uptake was the growth limiting fluxes during the Phase E1 of the fermentation. During the second growth phase the uptake rate of ammonia had a significant effect on the specific growth rate. Glucose was not growth limiting within the experimental range. Logarithmic sensitivity analysis revealed that during the Phase E1 a 10% increase in aspartic acid uptake rate will increase biomass by 4.4% and during the Phase E2 a 10% increase in ammonia uptake rate will increase the biomass uptake rate by 10%.

The sensitivity analysis of the specific growth rate and redox potential with respect to the biomass components showed that biomass building blocks would not affect the specific growth rate and redox potential production as well as the calculated flux pattern significantly. Hence, it was concluded that using *E. coli* biomass composition to represent *S. lividans* biochemical demands is an acceptable compromise.

Chapter 6

Conclusions and recommendations

6.1 Conclusions

In this thesis two approaches, the first based on statistical methods and the second on metabolic flux analysis, are presented for maximizing the recombinant protein (rHuIL-3) production rate. An emerging recombinant protein production platform, *Streptomyces lividans 66*, was used as the host organism in a model bacterial fermentation. This section summarizes the findings of the thesis.

Medium formulation had a profound impact on rHuIL-3 expression and post-secretion stability. Complex medium offered higher growth rates and crude rHuIL-3 concentrations at the expense of elevated protease activity. Defined medium supported growth and retained product authenticity and repressed protease activity.

As a first step, starvation trials were used to provide an initial screening of amino acids that affect cell growth and rHuIL-3 production. Using this approach the number of potentially significant amino acids were narrowed down to eight, namely: Arg, Asn, Asp, Glu, Leu, Met, Phe, Thr.

Screening mixture design in combination with distance-based multivariate analysis was efficient

in specifying and ranking the most significant amino acids with regard to cell growth and product formation. Based on this analysis the ranking of the amino acids was as follows: Met, Asp, Phe, Leu, Glu, Asn, Arg, and Thr. Among these amino acids Met, Asp, Phe, and Leu were proven to be essential for cell growth and product formation.

An augmented simplex mixture design with four components (Asp, Leu, Met, and Phe) was used to explain the nature of the response surface. Analysis of variance (ANOVA) revealed that the obtained model was significant and there was no statically significant lack of fit. According to this analysis, Asp $\times$ Met and Asp $\times$ Phe interactions had the greatest effect on the product formation. An artificial neural network with one input layer of four neurons corresponding to the four amino acids and one hidden layer of four neurons and one output layer of one neuron corresponding to rHuIL-3 was trained to predict the concentration of the product. A tangent hyperbolic transfer function produced better predictions compared to a sigmoid transfer function. The artificial neural network provided a better fit to experimental data than a multiple regression model.

A sequential quadratic algorithm was used to optimize both the multiple regression and the artificial neural network models in terms of product formation. The optimum amino acid levels were found to be Asp 56%, Met 5%, and Phe 39%. The maximum theoretical rHuIL-3 concentration was 76.00 mg/L at the optimum levels of the amino acids. An experimental maximum rHuIL-3 concentration of 80.03 ± 8.24 mg/L was achieved using the optimum medium composition supported the applied methodology.

Based on the aforementioned results, one may conclude that mixture designs, distance-based multivariate analysis and artificial neural networks are powerful tools in designing and optimizing a defined medium for recombinant protein production by *S. lividans*.

A metabolic network was developed for *S. lividans* and used as a predictive tool. The biochemical network was underdetermined therefore the solution space was too wide. To reduce the solution space, experimental flux measurements of the key metabolites were considered as inputs to the model. This approach was found to be effective in order to achieve a realistic

model. The proposed model was able to predict biomass concentration accurately throughout the fermentation.

The growth period was divided into two distinct phases, Phase E1 and E2, based on the slope of the logarithmic plot of biomass. During Phase E1, Asp and Glu acid were used up completely and the concentration of Phe, glucose and ammonia decreased by 50%, 9% and 4%, respectively. The specific growth rate during this phase was μ=0.078 h^{-1}. During Phase E2 the specific growth rate decreased to μ=0.035 h^{-1}. Growth stopped when glucose was completely exhausted.

Two different objective functions were considered: (1) optimizing biomass production by maximizing the specific growth rate and (2) optimizing with respect to redox metabolism by minimizing excess NAD(P)H production. The flux distribution obtained from the two objectives was almost identical. The PP pathway was active during Phase E1 for both objectives and 56-58% of glucose flux was diverted to this pathway. The glucose flux through the PP pathway during Phase E2 reduced slightly to 51-52% of the total.

Sensitivity analysis indicated that, in the concentration range chosen for nutrients, the nitrogen sources (amino acids in Phase E1 and ammonium nitrogen in Phase E2) were growth limiting rather than glucose. Sensitivity of the biomass formation reaction with respect to biomass components revealed that the specific growth rate was not affected by the biomass composition. Also, this analysis showed that the internal fluxes were not affected by biomass composition as well.

6.2 Recommendations

Although two methodologies have been applied successfully to identify and optimize key metabolites that enhance recombinant protein production, the following recommendations for future studies are proposed.

S. lividans was proven to be a promising host, but this study was limited to recombinant

rHuIL-3 expression only. The proposed statistical technique along with the artificial neural network model should be examined for *S. lividans* expressing other recombinant proteins. It should be also applied for optimizing medium and recombinant protein production with new emerging hosts, for example *Pichia pastoris*. Various *Pichia* species have been widely used on a laboratory scale for protein expression by recombinant DNA methodology.

The metabolic flux model should be reduced to elementary flux modes, which are the simplest metabolic paths connecting the substrate to the products. Correlation analysis should be performed to formulate a biomass model for predicting biomass and recombinant protein concentrations as a function of the extracellular metabolite concentration. Then the biomass model could be combined with the elementary flux modes to get an integrated model capable of predicting concentration values of substrates, biomass, and product by utilizing only starting concentrations as input. The metabolic model could be modified by introducing integer variables in the optimization program in order to evaluate different sets of the solutions that have the identical objective value and satisfy the constraints.

Throughout this study the steady-state analysis of the network stoichiometry was considered which produced algebraic equations as the constraints of the optimization problem. An alternative approach would be investigating the system under unsteady-state conditions and using differential equations as the constraints of the optimization problem. This approach would allow investigating the benefits and optimization of a fed-batch culture for producing rHuIL-3. An important consideration to assess is whether to initiate fed-batch operation during the Phase E1 of growth using amino acid mixtures or in Phase E2 using inorganic nitrogen. Presumably, the maintenance of the Phase E1 for extended time periods would result in faster growth, hence higher productivity. On the other hand, the sustained maintenance of Phase E2 would be economically more attractive, since the medium costs would be significantly less.

All experiments in this study were run at a constant temperature 32°C and neutral pH. It is known, that the bacterium, *S. lividans* contains multiple copies of the plasmid encoding rHuIL-3. Temperature and pH shifts have been reported to be beneficial for recombinant protein

production in several bacteria by increasing the plasmid copy number. The temperature or pH sensitivity of rHuIL-3 production by *Streptomyces* is unknown. Therefore, the effects of temperature and pH shifts on growth and product formation need further investigations. The basal growth medium for growth of *Streptomyces* contains animal-based nutrients including peptone. The use of complex medium resulted in faster growth than the amino acid mixtures. However, a shown in this thesis, peptides in this medium led to the production of proteases thereby compromising the authenticity of the product. The trend in the pharmaceutical industry is the replacement of animal-based fermentation medium components. Therefore, an investigation of the effect of the plant-based peptide medium on the productivity and the authenticity of the recombinant protein is recommended.

Bibliography

Alves A.M.C.R., Euverink G.J.W., Hector H.J., Hessels G.I., van der Vlag J., Vrijbloed J.W., Hondmann D., Visser J., and Dijkhuizen L. Enzymes of glucose and methanol metabolism in the actinomycete *Amycolatopsis methanolica*. *J. Bacteriol.*, 176:6827–6835, 1994. 65, 72

Anderson M.J. DSTLM *forward*: a FORTRAN computer program to calculate a distance-based multivariate analysis for a linear model using forward selection. http://www.stat.auckland.ac.nz/∼ mja/Programs.htm (2003). 23, 48

Avignone Rossa C., White J., Kuiper A., Postma P.W., and Bibb M. Teixeira de Mattos M.J. Carbon flux distribution in antibiotic-producing chemostat cultures of *Streptomyces lividans*. *Metab. Eng.*, 4:138–150, 2002. 62

Ayadi D.Z., Chouayekh H., Mhiri S., Zerria K., Fathallah D. M., and Bejar S. Expression by *Streptomyces lividans* of the rat α integrin CD11b A-domain as a secreted and soluble recombinant protein. *J. Biomed. Biotechnol.*, 2007:6 pages, 2007. 6

Bailey. Towards a science of metabolic engineering. *Science*, 252:1668–1647, 1991. 2

Baishan F., Hongwen C., Xiaolan X., Ning W., and Zongding H. Using genetic algorithms coupling neural networks in a study of xylitol production: medium optimization. *Process Biochem.*, 38:979–985, 2003. 43

Baughman D.R. and Liu Y.A. *Neural networks in bioprocessing and chemical engineering*. Academic press limited, California, 1995. 25, 26

Beynon R.J. and Bond J.S. *Proteolytic enzymes: A practical approach.* Oxford University Press, 2000. 39

Bidlingmeyer B.A., Cohen S.A., and Tarvin T.L. Rapid analysis of amino acids using pre-column derivatization. *J. Chromatogr.*, 336:93–104, 1984. 40

Blanco J.and Coque J.J. and Martin J.F. The folate branch of the methionine biosynthesis pathway in *Streptomyces lividans*: Disruption of the 5,10-methylenetetrahydrofolate reductase gene leads to methionine auxotrophy. *J. Bacteriol.*, 180:1586–1591, 1998. 59

Blum E., Py B., Carpousis A.J., and Higgins C.F. Polyphosphate kinase is a component of the *Escherichia coli* RNA degradosome. *Mol. Microbiol.*, 26(2):387–398, 1997. 66

Bonarius B.P.J., Hatzimanikatis V., and Meesters K.P.H. Metabolic flux analysis of hybridoma cells in different culture media using mass balances. *Biotechnol. Bioeng.*, 50:299–318, 1996. 30

Bramwell H., Hunter H.G., and Coggins J.R. Phosphoenolpyruvate carboxylase from *Streptomyces coelicolor A3(2)*: purification of the enzyme, cloning of the *ppc* gene and over-expression of the protein in a *Streptomycete*. *Biochem. J.*, 293:131–136, 1993. 65

Brawner M.E. Advances in heterologous gene expression by *Streptomyces*. *Curr. Opin. Biotechnol.*, 5:475–481, 1994. 61

Cannizzaro C., Christensen B., Nielsen J., and von Stockar. Metabolic network analysis on *Phaffia rhodozyma* yeast using ^{13}clabeled glucose and gas chromatography-mass spectrometry. *Metab. Eng.*, 6:340–351, 2004. 3

Cawthray G.R. An improved reversed-phase liquid chromatographic method for the analysis of low-molecular mass organic acids in plant root exudates. *J. Chromatogr. A*, 1011:233–240, 2003. 41

Chary C.V.K., Rambhav S., Venkateswerlu G., and Ramachandran L.K. Possible precursor for thiostrepton in *Streptomyces azureus* - culture medium production optimization. *Indian J. of Microbiol.*, 15:414–417, 1989. 42

Cheng Y.R., Fang A., and Demain A.L. Effect of amino acids on rapamycin biosynthesis by *Streptomyces hygroscopicus*. *Appl. Microbiol. Biotechnol.*, 43:1096–1098, 1995. 59

Clarke K.R. Nonparametric multivariate analyses of changes in community structure. *Aus. J. Ecol.*, 18:117–143, 1993. 22

Cochrane V.W., Peck H.D., Jr., and Harrison A. The metabolism of species of *Streptomyces*. VII. the hexosemonophosphate shunt and associated reactions. *J. Bacteriol.*, 66:17–23, 1953. 79

Cornell J.A. *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*. John Wiley & Sons, Inc, New York, 2002. 17, 19, 21, 43

Daae E.B. and Ison A.P. Classification and sensitivity analysis of proposed primary metabolic reaction network for *Streptomyces lividans*. *Metab. Eng.*, 1:153–165, 1999. 62, 81

Dagley S. and Nicholson D.E. *An introduction to metabolic pathways*. John Wiely & Sons, Inc, New York, 1970. 13

Dekleva M.L. and Strohl W.R. Biosynthesis of E-rhodomycinone from glucose by *Streptomyces* c5 and comparison with intermediary metabolism of other polyketide-producing streptomycetes. *Can. J. Microbiol*, 34:1235–12404, 1988. 65, 66, 72

Demain A.L. Microbial biotechnology. *Trends Biotecnnol*, 18:26–31, 2000. 2

Demain A.L. and Solomon. *Biology of industrial microorganisms*. The Benjamin/Cummings Publishing Company, INC, San Francisco, 1985. 7, 8

Diez B., Mellado E., Rodriguez M., Fouces R., and Barredo J.L. Recombinant microorganisms for industrial production of antibiotics. *Biotechnol. Bioeng.*, 55:216–226, 1997. 2

Roshni Lynne Dutton. *Growth and productivity of a recombinant Chinese Hamster Ovary cell line in batch culture.* PhD Thesis. Department of Chemical engineering, University of Waterloo, 200 University Ave. West, Waterloo, Ontario, N2L 3G1, Canada, 1998. 69

Dutton R.L., Scharer J.M., and Moo-Young M. Descriptive parameter evaluation in mammalian cell culture. *Cytotechnology*, 26:139–152, 1998. 68

Edwards J.S. and Palsson B.O. Robustness analysis of the *Escherichia coli* metabolic network. *Biotechnol. Prog.*, 16:927–939, 2000. 30

Eisaki N., Tatsumi H., Murakami S., and Horiuchi T. Pyruvate phosphate dikinase from a thermophilic actinomyces *Microbispora rosea* subsp. *aerata*: purifcation, characterization and molecular cloning of the gene. *Biochim. Biophys. Acta*, 1431:363–373, 1999. 65

Elkamel A., Abdul Wahab S., Bouhamra W., and Alper E. Measurment and prediction of ozone levels around a heavily industrialized area: A neural network approach. *Adv. Enviro. Res.*, 5:47–59, 2001. 26

Eriksson L., Johansson E., and Wikstrom C. Mixture design design generation, PLS analysis, and model usage. *Chemometr. Intell. Lab.*, 43:1–24, 1998. 21

Evans H.J. and Wood H.G. Purification and properties of pyruvate phosphate dikinase from propionic acid bacteria. *Biochemistry*, 10(5):721–729, 1971. 65

Gao J., Gorenflo V.M., Scharer J.M., and Budman M. Dynamic metabolic modeling for a MAb bioprocess. *Biotechnol. Prog.*, 23:168–181, 2007. 69

Garson G. D. Interpreting neural-network connection wights. *AI Expert*, pages 47–51, 1991. 26

Gheshlaghi R., Scharer J.M., Moo-Young M., and Douglas P.L. Medium optimization for hen egg white lysozyme production by recombinant *Aspergilus niger* using statistical methods. *Biotechnol. Bioeng.*, 90:754–760, 2005. 43

Gheshlaghi R., Scharer J.M., Moo-Young M., and Douglas P.L. Metabolic flux analysis for optimizing the specific growth rate of recombinant *Aspergilus niger*. *Bioprocess Biosyst Eng*, 30:397–418, 2007. 69, 70

Glorfeld L.W. A methodology for simplification and interpretation of backpropagation-based neural network models. *Expert Syst. Appl.*, 10:37–54, 1996. 27

Goh A.T.C. Back-propagation neural networks for modeling complex systems. *Artif. Intell. Eng.*, 9:143–151, 1995. 26

Gottschalk G. *Bacterial metabolism*. Springer-Verlag, Berlin, 1979. 17

Gouveia E.R., Baptista Neto A., Badino Jr. A.C., and Hokka C.O. Optimization of medium composition for clavulanic acid production by *Streptomyces clavuligerus*. *Biotechnol. Lett.*, 23:157–161, 2001. 43

Heijnen J.J., van Gulik W.M., Shimizu H., and Stephanoloulos. Metabolic flux control analysis of branch points: an improved approach to obtain flux control coefficients from large perturbation data. *Metab. Eng.*, 6:391–400, 2004. 3

Hilborn R. and Stearns S.C. On inference in ecology and evolutionary biology: The problem of multiple causes. *Acta Biotheror.*, 31:145–164, 1982. 22

Hillier F.S. and Lieberman G.J. *Introduction to operations research*. McGraw-Hill, New York, 2001. 69

Hochreiter S. and Wagner R. *Bioinformatics Research and Development*. Springer-Verlag, Berlin, Germany, 2007. 30

Ihmels J., Levy R, and Barkai N. Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*. *Nat. Biotechnol.*, 22(1):86–92, 2004. 3

Inbar L. and Lapidot A. ^{13}C nuclear magnetic resonance and gas chromatography-mass spectrometry studies of carbon metabolism in the actinomycin D producer *Streptomyces parvulus* by use of ^{13}C-labeled precursors. *J. Bacteriol.*, 173:7790–7801, 1991. 66

Ingraham J.L., Maaloe O., and Neidhardt F.C. *Growth of the bacterial cell.* Sinauer Associates Inc., Massachusetts, 1983. 81

Jorgensen H., Nielsen J., Villadsen J., and Mollgaard H. Metabolic flux distribution in *Penicillium chrysogenum* during fed-batch cultivation. *Biotechnol. Bioeng.*, 46:117–131, 1995. 72

Kalil S.J, Rodrigues M.I., and Maugeri F. Response surface analysis and simulation as a tool for bioprocess design and optimization. *Process Biochem.*, 35(6):539–550, 2000. 2

Kennedy M. and Krouse D. Strategies for improving fermentation medium performance: a review. *J. Ind. Microbiol. Biotech.*, 23:456–475, 1999. 42

Kieser T., Bibb M.J. Buttner M.J., Chater K.F., and Hopwood D.A. *Practical Streptomyces genetics.* The John Innes Foundation, Colney, Norwich NR4 7UH, England, 2000. 13, 65

Klamt S., Schuster S., and Gilles E.D. Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. *Biotechnol. Bioeng.*, 77(7):734–751, 2002. 29

Kornberg A. Inorganic polyphosphate: Toward making a forgetten polymer unforgettable. *J. Bacteriol.*, 177(3):491–496, 1995. 66

Laemmli U.K. Cleavage of structural proteins during the assembly of head of bacteriophage T4. *Nature*, 227:680–685, 1970. 37

Lee S.U. and Papoutsakis E.T. *Metabolic engineering.* Marcel Dekker, Inc., New York, 1999. 28, 31

Legendre P. and Anderson M.J. Distance-based multivariate analysis: Testing multispecies responses in multifactorial ecological experiments. *Ecol. Monogr.*, 69:1–24, 1999. 22, 23

Mantel N. and Valand R.S. A technique of nonparametric multivariate analysis. *Biometrics*, 26:547–558, 1970. 22

Marr A.G., Nilson E.H., and Clark D. J. The maintenance requirements of *Escherichia coli*. *Ann. N.Y. Acad. Sci.*, 102:536–548, 1963. 67

McArdle B.H. and Anderson M.J. Fitting multivariate models to community data: A comment on distance-based redundancy analysis. *Ecology*, 82:290–297, 2001. 23, 48

Monot F., Martin J.R., Petitdemange H., and Gay R. Aceton and butanol production by clostridium acetobutylicum in synthetic medium. *Appl. Environ. Microbiol.*, 44:1318–1324, 1982. 42

Montgomery D.C. and Runger G.C. *Applied statistics and probability for engineers*. John Wiley & Sons, Inc, New York, 2006. 17

Myers R.H. and Montgomery D.C. *Response surface methodology: process and product optimization using designed experiments*. John Wiley & Sons, Inc, New York, 2003. 19, 43

Naeimpoor F. and Mavituna F. Metabolic flux analysis in *Streptomyces coelicolor* under various nutrient limitations. *Metab. Eng.*, 2:140–148, 2000. 61

Nagata Y. and Chu K.H. Optimization of a fermentation medium using neural networks and genetic algorithms. *Biotechnol. Lett.*, 25:1837–1842, 2003. 43

Neidhardt F.C., Ingraham J.L., and Schaechter M. *Physiology of the bacterial cell. A molecular approach*. Sinaure Associate, Sunderland, 1990. 17

Neuzil J, Novotna J, Behal V, and Hostalek Z. Inhibition studies of glucose-6-phosphate dehydrogenase from tetracycline-producing *Streptomyces aureofaciens*. *Biotechnol. Appl. Bioc.*, 8:375–378, 1986. 66

Neuzil J, Novotna J, Erban V, Behal V, and Hostalek Z. Glucose-6-phosphate dehydrogenase from a tetracycline producing strain of *Streptomyces aureofaciens*: some properties and regulatory aspects of the enzyme. *Biochem. Int.*, 17:187–196, 1988. 66

Nicola N.A. *Guidebook to cytokines and their receptors*. Sambrook and Tooze Publishing Partnership, New York, 1994. 9

Nielsen J. Metabolic engineering. *Appl. Microbiol. Biotechnol.*, 55:263–283, 2001. 27

Nielsen J., Villadsen J., and Liden G. *Bioreaction Engineering Principles*. Plenum publishing corporation, New York, 2002. 66

Nowruzi K., Elkamel A., Scharer J.M., Cossar D., and Moo-Young M. Develoment of a minimal defined medium for recombinant human interleukin-3 production by *Streptomyces lividans 66*. *Biotechnol. Bioeng.*, 99:214–222, 2008. 64

Obanye A.I.C., Hobbs G., Gardner D.C.J., and Oliver S.G. Correlation between carbon flux through the pentose phosphate pathway and production of the antibiotic methylenomycin in *Streptomyces coelicolor* a3(2). *Microbiology*, 142:133–137, 1996. 72

Ooijkaas L.P., Wilkinson E.C., Tramper J., and Buitelaar R.M. Medium optimization for spore production of coniothyrium minitans using statistically-based experimental designs. *Biotechnol. Bioeng.*, 64:92–100, 1999. 43

Pallson B. O. What lies beyond bioinformatics? *Nature Biotechnol.*, 15:3–4, 1997. 3

Pannel D.J. *Introduction to practical linear programming*. John Wiley & Sons, Inc., New York, 1997. 70

Parekh S. and Strobel R. J. Improvement of microbial strains and fermentation processes. *Biochem. Eng. J.*, 54:287–301, 2000. 2

Pirt S.J. The maintenance energy of bacteria in growing cultures. *Proc. R. Soc. B.*, 163: 224–231, 1965. 67

Rosenthal R.E. *GAMS - A user's guide*. GAMS Development Corporation, Washington, DC, USA, 2008. 70

Hans Roubos. *Bioprocess modelling and optimization: Fed-batch Clavulanic acid production by Steptomyces clavuligerus*. PhD in Control system engineering, Delft University of Technology, Postbus 5, 2600 AA Delft, The Netherlands, 2002. 65, 66

Salas J.A., Quiros L.M., and Hardisson C. Pathways of glucose catabolism during germination of *Streptomyces* spores. *Microbiol. Lett.*, 22:229–233, 1984. 65, 79

Sathyamoorthy M., Stemke D., and Speedie M.K. Native and heterologous secretion by *Streptomyces lividans*. *Appl. Microbiol. Biotechnol.*, 46:347–352, 1996. 61

Savinell J.M and Palsson B.O. Network analysis of intermediary metabolism using linear optimization. I. development of mathematical formalism. *J. Theor. Biol.*, 154:421–454, 1992. 30

Schuster S.and Klamt S., Weckwerth W., Moldenhauer F., and Pfeiffer T. Use of network analysis of metabolic systems in bioengineering. *Bioproc. Biosyst. Eng.*, 24:363–372, 2002. 27

Stephanopoulos G.N. and Vallino J.J. Network rigidity and metabolic engineering overproduction. *Science*, 252:1675–1681, 1991. 2

Stephanopoulos G.N., Aristidou A.A., and Nielsen J. *Metabolic Engineering Principles and Methadologies*. Academic Press, San Diego, 1998. 15, 27, 29, 30, 68

Techapun C., Sinsuwongwat S., Watanabe M., Sasaki K., and Poosaran N. Production of cellulase-free xylanase by a thermotolerant *Streptomyces sp.* grown on agricultural waste and media optimization using mixture design and Plackett & Burman experimental design methods. *Biotechnol. Lett.*, 24:1437–1442, 2002. 43

Torres N.V. and Voit E.O. *Pathway analysis and optimization in metabolic engineering*. Cambridge university press, UK, 2002. 2

Vallino J.J. and Stephanopoulos G. *Flux determination in cellular bioreaction network: Applications to lysine fermentations*. CRC Press, Boca Rotan, FL, 1990. 2, 61

Vallino J.J. and Stephanopoulos G. Metabolic flux distribution in *Corynebacterium glutamicum* during growth and lysine overproduction. *Biotechnol. Bioeng.*, 41:633–646, 1993. 62

van Gulik W.M. and Heijnen J.J. A metabolic network stoichiometry analysis of microbial growth and product formation. *Biotechnol. Bioeng.*, 48:681–698, 1995. 3, 62

Varma A. and Palsson B.O. Biochemical production capabilities of *Escherichia. coli. Biotechnol. Bioeng.*, 42:59–73, 1993a. 3

Varma A. and Palsson B.O. Metabolic capabilities of *E. Coli*. I. Synthesis of biosynthetic precursors and cofactors. *J. Theor. Biol.*, 165:477–502, 1993b. 3

Varma A. and Palsson B.O. Metabolic capabilities of *E. Coli*. II. Optimal growth patterns. *J. Theor. Biol.*, 165:503–522, 1993c. 3

Varma A. and Palsson B.O. Metabolic flux balancing: Basic concepts, scientific and practical use. *Bio/Technol.*, 12:994–998, 1994. 2, 29, 62

Voet D. and Voet J.G. *Biochemistry*. John Wiley & Sons, INC, New York, 2end edition, 1995. 11

Wittmann C., Kim H.M., and Heinzle E. Metabolic network analysis of lysine producing *Corynebacterium glutamicum* at a miniaturized scale. *Biotechnol. Bioeng.*, 87(1):1–6, 2004. 29

Ahmad Ramil Mohammad Yahya. *Bioprocessing strategies in improving recombinant human interleukin-3 production in Streptomyces lividans fermentation*. PhD Thesis. Department of Chemical engineering, University of Waterloo, 200 University Ave. West, Waterloo, Ontario, N2L 3G1, Canada, 2003. 39, 62

Yee L. and Blanch H.W. Defined medium optimization for growth of recombinant *Escherichia coli* X90. *Biotechnol. Bioeng.*, 41:221–230, 1992. 42

Yun S., Yahya A.R.M., Cossar D., Anderson W.A., Scharer J.M., and Moo-Young M. Temperature downshift increases recombinant cytokine titer in *Streptomyces lividans* fermentation. *Biotechnol. Lett.*, 23:1903–1905, 2001. 32

Zhang J. and Greasham R. Chemically defined media for commercial fermentations. *Appl. Microb. Biotechnol.*, 51:407–421, 1999. 42

Zubay G.L. *Biochemistry*. Macmillan, New York, 1998. 13

APPENDICES

Appendix 1

Component List

αKG	α-ketoglutarate
ACCOA	acetyl coenzyme A
AICAR	5-aminoimidazole-4-carboxamide ribonucleotide
ADP	adenosine-5-diphosphate
ALA	alanine
AMP	adenosine-5-monophosphate
ARG	argenine
ASN	asparagine
ASP	aspartic acid
ATP	adenosine-5-triphosphate
BPG	1,3-bisphosphoglycerate
CHOR	chorismate
CITR	citrate
CMP	cytidine monophosphate
CO_2	carbon dioxide
CYS	cysteine
DAHP	dihydroxyacetone phosphate
E4P	erythrose-4-phosphate
F16P	fructose-1,6-bisphosphate
F6P	fructose-6-phosphate
FAD	flavine adenine dinucleotide (oxidized)
FADH2	flavine adenine dinucleotide (reduced)
FTHF	formyl tetrahydrofolate
FUM	fumurate

G6P	glucose-6-phosphate
GAP	3-phosphoglyceraldehyde
GLC	glucose-6-phosphate
GLN	glutamine
GLU	glutamic acid
GLY	glycine
GMP	guanosine-5-monophosphate
Gu6P	gluconate-6-phosphate
HIS	histidine
ICIT	isocitrate
IGP	imidazole glycerol phosphate
ILE	isoleucine
IMP	inosine monophosphate
LEU	leucine
LYS	lysine
MAL	malate
MET	methionine
MTHF	methylene tetrahydrofolate
NAD	nicotinamide adenine dinucleotide (oxidized)
NADH	nicotinamide adenine dinucleotide (reduced)
NADP	nicotinamide adenine dinucleotide phosphate (oxidized)
NADPH	nicotinamide adenine dinucleotide phosphate (reduced)
NH_3	ammonia
O_2	oxygen
OAA	oxaloacetate
PEP	phosphoenolpyruvate
PG	3-phosphoglycerate

PHE	phenylalanine
PRO	proline
PRPP	5-phosphoribosyl pyrophosphate
PYR	pyruvate
R5P	ribulose-5-phosphate
Ru5P	ribose-5-phosphate
S7P	sedoheptulose-7-phosphate
SER	serine
SUC	succinate
SUCCOA	succinate coenzyme A
THF	tetrahydrofolate
THR	thronine
TMP	thymidine monophosphate
TRP	tryptophane
TYR	tyrosine
UMP	uridine monophosphate
VAL	valine
X5P	xylulose-5-phosphate

Biochemical reactions in the metabolic network

Glycosis

1. GLC + ATP → G6P + ADP
2. G6P ↔ F6P
3. F6P + ATP ↔ F16DP + ADP
4. F16DP + H_2O ↔ GAP + DHAP
5. DHAP ↔ GAP
6. GAP + NAD^+ ↔ BPG + NADH
7. BPG + ADP ↔ GP + ATP
8. GP ↔ PEP
9. PEP + ADP→ PYR + ATP

Pentose Phosphate Pathway

10. G6P + $NADP^+$ → Gu6P + NADPH
11. Gu6P + $NADP^+$ → Ru5P + NADPH + CO_2
12. Ru5P ↔ X5P
13. Ru5P ↔ R5P
14. R5P + X5P ↔ GAP + S7P
15. GAP + S7P ↔ E4P + F6P

16. E4P + X5P $\leftrightarrow$ GAP + F6P

Tricarboxylic Cycle Pathway

17. PYR + ATP + CO_2 $\leftrightarrow$ OAA + ADP
18. PYR + NAD^+ $\rightarrow$ ACCOA + NADH + CO_2
19. ACCOA + OAA $\rightarrow$ CITR
20. CITR $\leftrightarrow$ iCITR
21. iCITR + NAD^+ $\leftrightarrow$ αKG + NADH + CO_2
22. αKG + CoA + NAD^+ $\leftrightarrow$ SUCCOA + NADH + CO_2
23. SUCCOA + ADP $\leftrightarrow$ SUC + ATP + CoA
24. SUC + FAD $\leftrightarrow$ FUM + $FADH_2$
25. FUM $\leftrightarrow$ MAL
26. MAL + NAD^+ $\leftrightarrow$ OAA + NADH
27. MAL + NAD^+ $\leftrightarrow$ PYR + CO_2 + NADH

Amino acid biosynthesis

28. 2 PEP + E4P + NADPH + ATP $\rightarrow$ CHOR + ADP
29. PG + GLU + NAD^+ $\leftrightarrow$ SER + αKG + NADH
30. SER + FTHF $\leftrightarrow$ Gly + MTHF
31. SER + NAD^+ $\leftrightarrow$ CYS + NADH
32. αKG + NADPH + NH_3 $\leftrightarrow$ GLU + $NADP^+$

33. $\mathrm{GLU} + \mathrm{NH_3} + \mathrm{ATP} \leftrightarrow \mathrm{GLN} + \mathrm{ADP}$

34. $\mathrm{GLU} + 2\,\mathrm{NADPH} + \mathrm{ATP} \rightarrow \mathrm{PRO} + 2\,\mathrm{NADP^+} + \mathrm{ADP}$

35. $2\,\mathrm{GLU} + \mathrm{ASP} + \mathrm{NADPH} + 2\,\mathrm{ATP} \rightarrow \mathrm{ARG} + \alpha\mathrm{KG} + \mathrm{FUM} + \mathrm{NADP^+} + 2\,\mathrm{ADP}$

36. $\mathrm{OAA} + \mathrm{GLU} \leftrightarrow \mathrm{ASP} + \alpha\mathrm{KG}$

37. $\mathrm{ASP} + \mathrm{NH_3} + \mathrm{ATP} \rightarrow \mathrm{ASN} + \mathrm{ADP}$

38. $\mathrm{ASP} \rightarrow \mathrm{ALA} + \mathrm{CO_2}$

39. $\mathrm{ASP} + \mathrm{NADPH} + \mathrm{NADH} + 2\,\mathrm{ATP} \leftrightarrow \mathrm{THR} + \mathrm{NADP^+} + \mathrm{NAD^+} + \mathrm{ADP}$

40. $\mathrm{THR} + \mathrm{GLU} + \mathrm{NADPH} \rightarrow \mathrm{ILE} + \alpha\mathrm{KG} + \mathrm{NH_3} + \mathrm{NADP^+}$

41. $\mathrm{ASP} + \mathrm{GLU} + \mathrm{PYR} + \mathrm{SUCCOA} + 3\,\mathrm{NADPH} + \mathrm{ATP} \rightarrow \mathrm{LYS} + \alpha\mathrm{KG} + \mathrm{SUC} + \mathrm{CO_2} + 3\,\mathrm{NADP^+} + \mathrm{ADP}$

42. $\mathrm{ASP} + 2\,\mathrm{NADPH} + 2\ \mathrm{ATP} \rightarrow \mathrm{MET} + 2\,\mathrm{NADP^+} + 2\,\mathrm{ATP}$

43. $\mathrm{PYR} + \mathrm{GLU} + \mathrm{NADPH} \rightarrow \mathrm{VAL} + \alpha\mathrm{KG} + \mathrm{NADP^+}$

44. $\mathrm{VAL} + \mathrm{ACCOA} + \mathrm{NAD^+} \rightarrow \mathrm{LEU} + \mathrm{CO_2} + \mathrm{NADH}$

45. $\mathrm{CHOR} + \mathrm{GLU} \rightarrow \mathrm{PHE} + \alpha\mathrm{KG} + \mathrm{CO_2}$

46. $\mathrm{CHOR} + \mathrm{GLU} + \mathrm{NAD^+} \rightarrow \mathrm{TYR} + \alpha\mathrm{KG} + \mathrm{CO_2} + \mathrm{NADH}$

47. $\mathrm{CHOR} + \mathrm{GLN} + \mathrm{PRPP} + \mathrm{SER} \rightarrow \mathrm{TRP} + \mathrm{PYR} + \mathrm{GLU} + \mathrm{GAP} + \mathrm{CO_2}$

48. $\mathrm{IGP} + 2\,\mathrm{NADH} \rightarrow \mathrm{HIS} + 2\,\mathrm{NAD^+}$

49. $\mathrm{ASP} \leftrightarrow \mathrm{FUM} + \mathrm{NH_3}$

Nucleotide biosynthesis

50. $\mathrm{R5P} + \mathrm{ATP} \rightarrow \mathrm{PRPP} + \mathrm{ADP}$

51. PRPP + ASP + 2 GLN + GLY + CO_2 + 4 ATP → AICAR + 2 GLU + 4 ADP

52. PRPP + ASP + NH_3 + NAD^+ + 2 ATP → UMP + NADH + 2 ADP

53. UMP + MTHF → TMP + THF

54. UMP + GLN + ATP → CMP + GLU + ADP

55. PRPP + GLN + ATP → IGP + AICAR + GLU + ADP

56. AICAR + ASP + FTHF + ATP → IMP + FUM + THF

57. IMP + GLN + NAD^+ + ATP → GMP + GLU + NADH + ADP

58. IMP + ASP + ATP → AMP + FUM + ADP

59. ACCOA + THF + NADPH + ATP → FTHF + CO_2 + NADH + $FADH_2$

60. MTHF + NAD^+ ↔ FTHF + NADH

Oxidative Phosphorylation

61. NADH + 0.5 O_2 + $(\frac{p}{o})$ ADP → NAD + $(\frac{p}{o})$ ATP

62. $FADH_2$ + 0.5 O_2 + $(\frac{p}{o})$ ADP → FAD + $(\frac{p}{o})$ ATP

63. ATP → ADP

Biomass

64. 2.4998 ACCOA + 0.1897 AMP + 4.2856 ATP + 0.1514 CMP + 0.0709 F6P + 0154 G6P + 0.129 GAP + 0.0485 MTHF + 5.1875 NADPH + 0.0276 OAA + 0.0511 PEP + 0.0235 PG + 0.0276 PYR + 0.0235 R5P + 0.0235 S7P + 0.0247 TMP + 0.1360 UMP + 0.5432 ALA + 0.281 ARG + 0.229 ASN + 0.229 ASP + 0.087 CYS + 0.25 GLN + 0.5478 GLU + 0.585 GLY + 0.09 HIS + 0.276 ILE + 0.428 LEU + 0.326 LYS + 0.146 MET + 0.176 PHE

$+\ 0.21\,\mathrm{PRO} + 0.334\,\mathrm{SER} + 0.241\,\mathrm{THR} + 0.054\,\mathrm{TRP} + 0.131\,\mathrm{TYR} + 0.402\,\mathrm{VAL} + \rightarrow$
$\mathrm{Biomass} + 0.2109\,\alpha\mathrm{KG} + 0.0235\,\mathrm{NADH} + 0.2856\,\mathrm{ADP}$

65. $O_2 \rightarrow$

66. $\rightarrow CO_2$

Appendix 2

Flux distribution through the metabolic network (mmol/g_{DW}h)

	Maximize μ		**Minimize redox potential**	
Flux	Phase E1	Phase E2	Phase E1	Phase E2
ν_1	0.6200	0.4800	0.6200	0.4800
ν_2	0.2564	0.2280	0.2668	0.2300
ν_3	0.4474	0.3762	0.4523	0.3756
ν_4	0.4474	0.3762	0.4523	0.3756
ν_5	0.4474	0.3762	0.4523	0.3756
ν_6	0.9826	0.8235	0.9900	0.8189
ν_7	0.9826	0.8235	0.9900	0.8189
ν_8	0.8827	0.7850	0.8929	0.7810
ν_9	1.2141	1.0245	1.2148	1.0095
ν_{10}	0.3625	0.2515	0.3521	0.2495
ν_{11}	0.3625	0.2515	0.3521	0.2495
ν_{12}	0.1980	0.1508	0.1924	0.1483
ν_{13}	0.1645	0.1007	0.1598	0.1013
ν_{14}	0.1067	0.0784	0.1037	0.0794
ν_{15}	0.1050	0.0777	0.1020	0.0787
ν_{16}	0.0913	0.0724	0.0887	0.0689
ν_{17}	0.0000	0.1026	0.0000	0.1009
ν_{18}	1.0303	0.6375	1.0314	0.6268
ν_{19}	0.7154	0.5160	0.7255	0.5074
ν_{20}	0.7154	0.5160	0.7255	0.5074
ν_{21}	0.7154	0.5160	0.7255	0.5074

Continued on next page

– continued from previous page

	Maximize μ		**Minimize redox potential**	
Flux	Phase E1	Phase E2	Phase E1	Phase E2
ν_{22}	0.8579	0.4861	0.8503	0.4781
ν_{23}	0.8343	0.4770	0.8273	0.4691
ν_{24}	0.8579	0.4861	0.8503	0.4781
ν_{25}	0.8729	0.5109	0.8599	0.5024
ν_{26}	0.6089	0.5109	0.6083	0.5024
ν_{27}	0.2640	0.0000	0.2516	0.0000
ν_{28}	0.0137	0.0053	0.0133	0.0098
ν_{29}	0.0983	0.0379	0.0955	0.0373
ν_{30}	0.0657	0.0253	0.0638	0.0249
ν_{31}	0.0061	0.0023	0.0059	0.0023
ν_{32}	0.0000	0.2234	0.0000	0.2242
ν_{33}	0.1008	0.0389	0.0980	0.0382
ν_{34}	0.0146	0.0056	0.0141	0.0055
ν_{35}	0.0215	0.0083	0.0209	0.0081
ν_{36}	-0.1085	0.0967	-0.1191	0.0951
ν_{37}	0.0159	0.0061	0.0154	0.0060
ν_{38}	0.0386	0.0149	0.0375	0.0146
ν_{39}	0.0353	0.0136	0.0343	0.0134
ν_{40}	0.0187	0.0072	0.0182	0.0071
ν_{41}	0.0236	0.0091	0.0230	0.0090
ν_{42}	0.0107	0.0041	0.0104	0.0041
ν_{43}	0.0636	0.0245	0.0618	0.0241
ν_{44}	0.0353	0.0136	0.0343	0.0134
ν_{46}	0.0097	0.0037	0.0094	0.0037

Continued on next page

– continued from previous page

	Maximize μ		**Minimize redox potential**	
Flux	Phase E1	Phase E2	Phase E1	Phase E2
ν_{47}	0.0040	0.0016	0.0039	0.0015
ν_{48}	0.0064	0.0024	0.0062	0.0024
ν_{49}	-0.0492	0.0000	-0.0527	0.0000
ν_{50}	0.0560	0.0216	0.0544	0.0213
ν_{51}	0.0229	0.0088	0.0223	0.0087
ν_{52}	0.0228	0.0088	0.0221	0.0086
ν_{53}	0.0018	0.0007	0.0017	0.0007
ν_{54}	0.0110	0.0043	0.0107	0.0042
ν_{55}	0.0064	0.0024	0.0062	0.0024
ν_{56}	0.0293	0.0113	0.0284	0.0111
ν_{57}	0.0159	0.0061	0.0154	0.0060
ν_{58}	0.0134	0.0052	0.0130	0.0051
ν_{59}	0.0912	0.0352	0.0886	0.0346
ν_{60}	-0.0267	-0.0103	-0.0259	-0.0101
ν_{61}	4.7049	1.4343	4.6959	1.4149
ν_{62}	0.9491	0.5213	0.9388	0.5126
ν_{63}	3.7606	1.6645	3.8498	1.6329
ν_{65}	1.6640	0.6041	1.6284	0.5987
ν_{66}	2.8270	0.9778	2.8174	0.9638
μ	0.0723	0.0279	0.0703	0.0274
NAD(P)H	5.8062	3.4168	5.7834	3.5486
ATP	8.3516	3.9915	8.3451	4.1677

Random biomass components chosen from uniform distribution for Runs 1 to 8

Comp.	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8
ACCOA	2.1312	2.7249	2.0732	2.4790	2.7304	2.9510	2.6372	2.5263
AKG	0.2515	0.2362	0.1984	0.2179	0.2484	0.2483	0.1715	0.2083
ALA	0.5884	0.5038	0.6482	0.5091	0.5405	0.5827	0.4542	0.4750
AMP	0.2031	0.1519	0.1549	0.2085	0.2094	0.1860	0.1559	0.1905
ARG	0.2941	0.3224	0.2280	0.3184	0.2931	0.3005	0.3048	0.2830
ASN	0.2575	0.1950	0.2144	0.2150	0.1961	0.2238	0.2333	0.2570
ASP	0.1855	0.2259	0.2499	0.2443	0.2288	0.2605	0.2352	0.2023
ATP	3.3213	3.3127	4.2357	4.2627	4.1714	3.5874	3.8123	4.3921
CMP	0.1604	0.1773	0.1387	0.1327	0.1764	0.1284	0.1292	0.1545
CYS	0.0801	0.0805	0.0963	0.0770	0.0706	0.0739	0.0797	0.0924
F6P	0.0732	0.0644	0.0834	0.0691	0.0714	0.0820	0.0830	0.0581
G6P	0.0149	0.0155	0.0145	0.0169	0.0168	0.0124	0.0176	0.0134
GAP	0.1415	0.1069	0.1223	0.1490	0.1199	0.1334	0.1297	0.1194
GLN	0.2591	0.2383	0.2265	0.2537	0.2742	0.2442	0.2919	0.2237
GLU	0.5972	0.4910	0.6147	0.6315	0.5695	0.4772	0.5233	0.6562
GLY	0.5622	0.4912	0.6143	0.6265	0.6378	0.6339	0.5633	0.6575
GMP	0.2496	0.2224	0.2702	0.2387	0.2056	0.1858	0.2671	0.2216
HIS	0.1017	0.0735	0.0897	0.1047	0.0793	0.0857	0.0795	0.1051
ILE	0.3273	0.2322	0.2502	0.2294	0.2743	0.3239	0.2521	0.2235
LEU	0.4910	0.4726	0.4370	0.5656	0.4240	0.5020	0.4001	0.5783
LYS	0.3780	0.3710	0.2851	0.3067	0.2790	0.3160	0.3725	0.2681
MET	0.1281	0.1469	0.1294	0.1536	0.1417	0.1487	0.1735	0.1240
MTHF	0.0567	0.0427	0.0527	0.0442	0.0487	0.0441	0.0519	0.0541
NADH	0.0255	0.0207	0.0215	0.0245	0.0281	0.0269	0.0214	0.0209

Continued on next page

– continued from previous page

Comp.	Run 1	Run 2	Run 3	Run 4	Run 5	Run 6	Run 7	Run 8
NADPH	4.8559	4.6025	5.6287	4.5261	4.1823	5.1129	5.6672	6.0738
OAA	0.0269	0.0233	0.0307	0.0253	0.0283	0.0311	0.0240	0.0287
PEP	0.0526	0.0508	0.0554	0.0547	0.0525	0.0487	0.0496	0.0440
PG	0.0213	0.0240	0.0215	0.0196	0.0282	0.0201	0.0243	0.0263
PHE	0.1622	0.1827	0.1584	0.1574	0.1441	0.1637	0.1456	0.2062
PRO	0.1719	0.1747	0.2055	0.1934	0.1811	0.2217	0.2246	0.2294
PYR	0.0270	0.0312	0.0270	0.0240	0.0233	0.0239	0.0296	0.0274
R5P	0.0204	0.0260	0.0257	0.0199	0.0242	0.0245	0.0198	0.0240
S7P	0.0196	0.0254	0.0257	0.0240	0.0272	0.0248	0.0257	0.0200
SER	0.2707	0.3542	0.2939	0.2725	0.3314	0.3049	0.2764	0.3468
THR	0.2095	0.2209	0.2149	0.2714	0.2360	0.2061	0.1954	0.2815
TMP	0.0277	0.0280	0.0296	0.0245	0.0256	0.0206	0.0241	0.0217
TRP	0.0611	0.0609	0.0623	0.0434	0.0525	0.0614	0.0593	0.0646
TYR	0.1470	0.1536	0.1347	0.1358	0.1241	0.1100	0.1291	0.1336
UMP	0.1518	0.1402	0.1480	0.1308	0.1603	0.1460	0.1402	0.1113
VAL	0.3281	0.4027	0.4757	0.4256	0.3995	0.4689	0.4253	0.3668

Random biomass components chosen from uniform distribution for Runs 9 to 15

Comp.	Run 9	Run 10	Run 11	Run 12	Run 13	Run 14	Run 15
ACCOA	2.3828	2.9643	2.9353	2.7947	2.2937	2.8625	2.6062
AKG	0.1776	0.2102	0.2244	0.2315	0.2241	0.1832	0.2165
ALA	0.6127	0.4547	0.5625	0.5658	0.4711	0.5161	0.5346
AMP	0.1816	0.2001	0.1696	0.2263	0.1595	0.2020	0.1856
ARG	0.2879	0.3348	0.3261	0.3114	0.2709	0.2859	0.2972
ASN	0.2033	0.2040	0.1921	0.2306	0.2371	0.2134	0.2195
ASP	0.2590	0.2521	0.2242	0.2361	0.2139	0.2441	0.2330
ATP	3.4324	4.8170	3.6617	3.8676	3.2790	4.6507	3.9146
CMP	0.1545	0.1402	0.1655	0.1780	0.1255	0.1747	0.1526
CYS	0.0977	0.1001	0.0826	0.0898	0.0754	0.0795	0.0840
F6P	0.0810	0.0712	0.0834	0.0670	0.0674	0.0687	0.0731
G6P	0.0178	0.0130	0.0170	0.0180	0.0167	0.0122	0.0155
GAP	0.1119	0.1254	0.1049	0.1167	0.1399	0.1492	0.1264
GLN	0.2258	0.2235	0.2857	0.2107	0.2547	0.2274	0.2457
GLU	0.4746	0.4490	0.5924	0.6440	0.4454	0.4522	0.5442
GLY	0.5914	0.6508	0.5967	0.6167	0.4989	0.5499	0.5922
GMP	0.2106	0.2056	0.1877	0.1911	0.2077	0.2090	0.2195
HIS	0.0793	0.0883	0.0820	0.0769	0.1068	0.0764	0.0878
ILE	0.2302	0.2373	0.2828	0.2625	0.2518	0.2463	0.2588
LEU	0.5361	0.5709	0.4913	0.4153	0.4306	0.5257	0.4886
LYS	0.3830	0.3391	0.3223	0.3252	0.2860	0.3482	0.3272
MET	0.1261	0.1739	0.1546	0.1648	0.1750	0.1286	0.1478
MTHF	0.0510	0.0575	0.0528	0.0484	0.0507	0.0399	0.0497
NADH	0.0221	0.0226	0.0216	0.0224	0.0234	0.0211	0.0231

Continued on next page

– continued from previous page

Comp.	Run 9	Run 10	Run 11	Run 12	Run 13	Run 14	Run 15
NADPH	4.2977	6.0988	4.6573	5.9970	4.8734	5.6971	5.1622
OAA	0.0275	0.0316	0.0282	0.0280	0.0268	0.0240	0.0274
PEP	0.0536	0.0582	0.0416	0.0590	0.0473	0.0468	0.0511
PG	0.0253	0.0251	0.0217	0.0245	0.0221	0.0212	0.0232
PHE	0.1910	0.1590	0.1520	0.1864	0.1757	0.1686	0.1681
PRO	0.2344	0.1807	0.1831	0.1727	0.2144	0.2285	0.2011
PYR	0.0275	0.0326	0.0321	0.0227	0.0303	0.0315	0.0279
R5P	0.0256	0.0263	0.0229	0.0188	0.0227	0.0232	0.0232
S7P	0.0279	0.0268	0.0199	0.0237	0.0229	0.0260	0.0243
SER	0.3156	0.3044	0.3951	0.3458	0.2697	0.2721	0.3110
THR	0.2254	0.2673	0.1942	0.1937	0.2826	0.2026	0.2287
TMP	0.0256	0.0221	0.0219	0.0248	0.0290	0.0209	0.0247
TRP	0.0594	0.0512	0.0494	0.0439	0.0537	0.0557	0.0556
TYR	0.1511	0.1456	0.1229	0.1521	0.1146	0.1257	0.1343
UMP	0.1482	0.1382	0.1229	0.1313	0.1430	0.1148	0.1376
VAL	0.3247	0.3442	0.4368	0.3644	0.3310	0.3878	0.3915

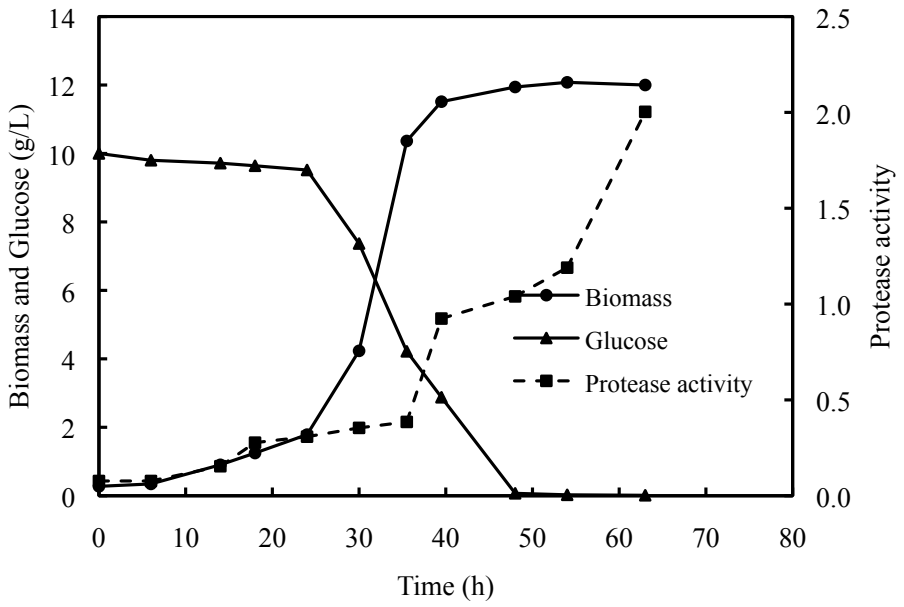

Figure 1: Protease activity and Biomass and rHuIL-3 concentration profiles during fermentation with casein peptone

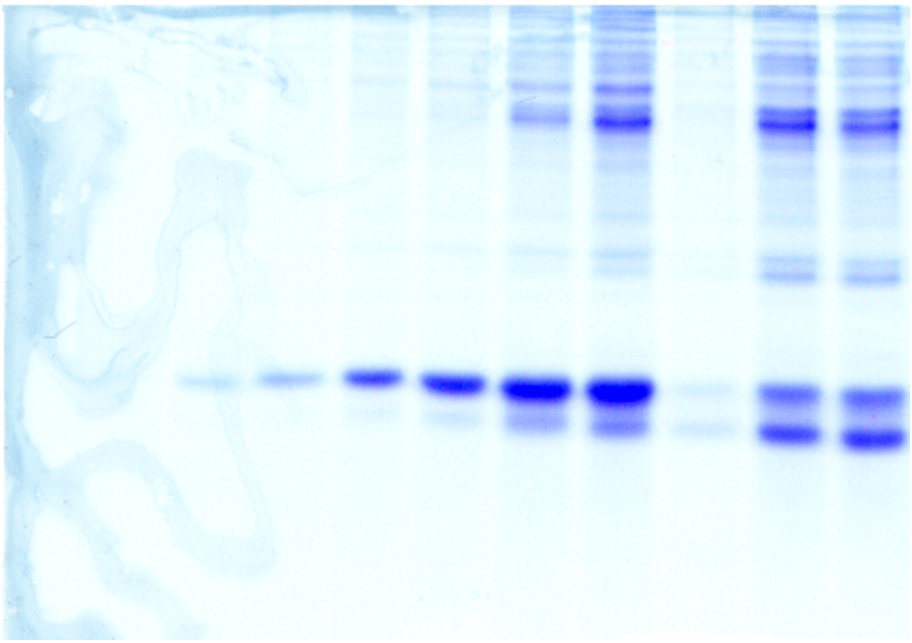

Figure 2: SDS-PAGE of supernatant samples from fermentor with casein peptone

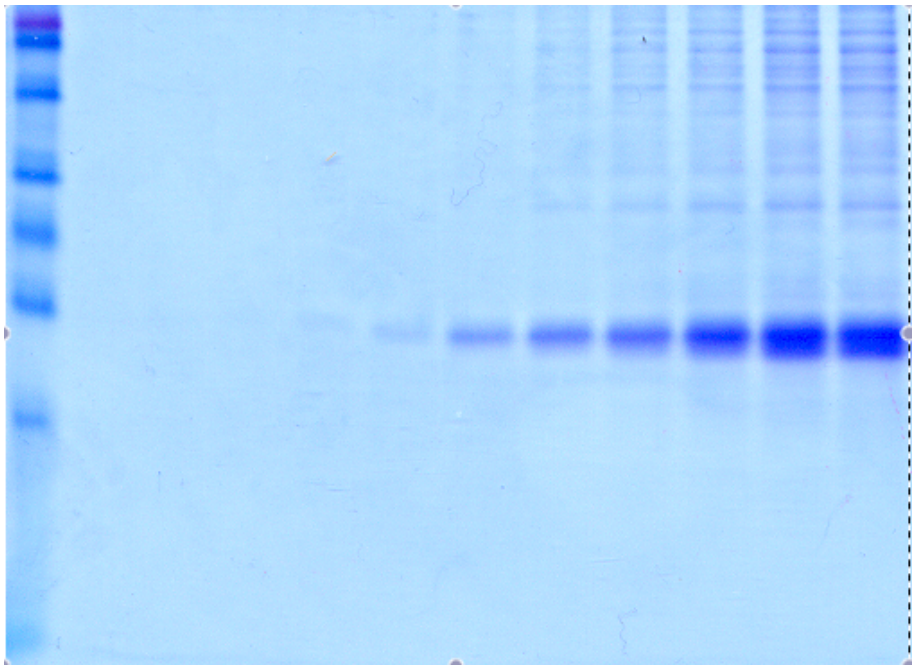

Figure 3: SDS-PAGE of supernatant samples from fermentor with defined medium

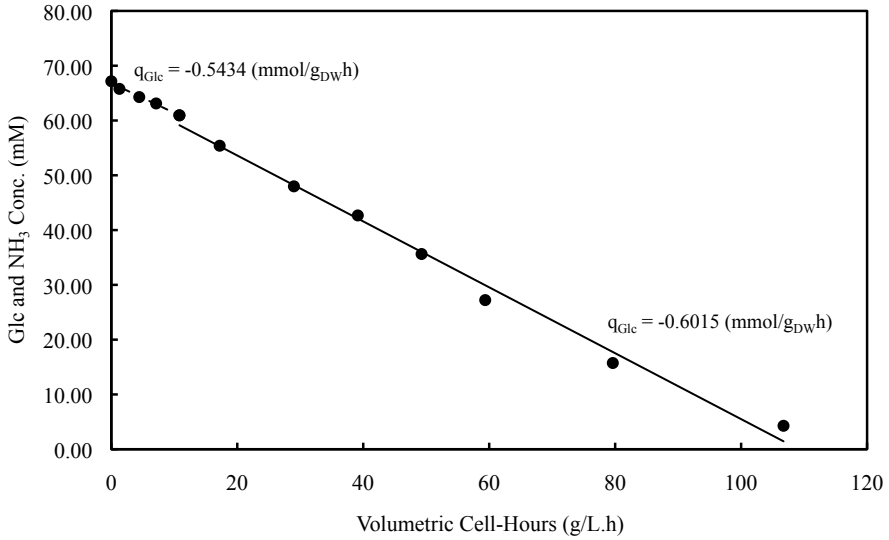

Figure 4: Dependence of glucose uptake on volumetric cell hours

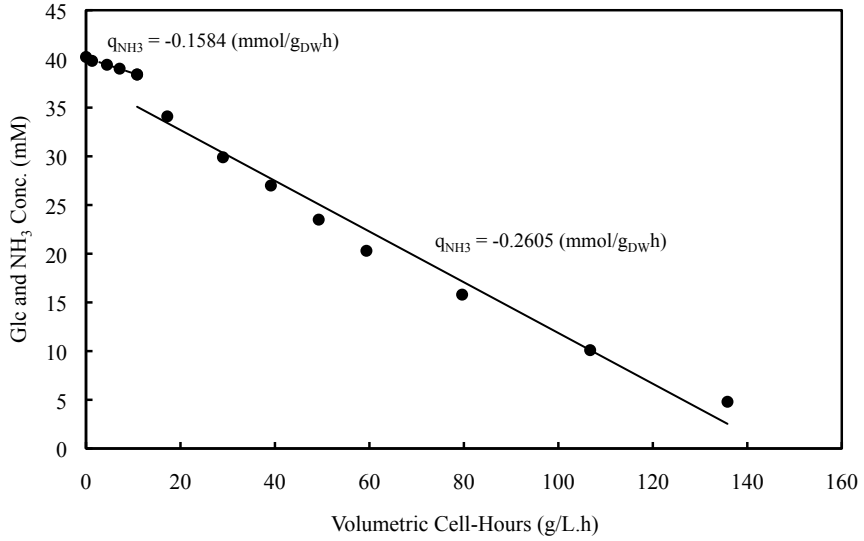

Figure 5: Dependence of ammonia uptake on volumetric cell hours

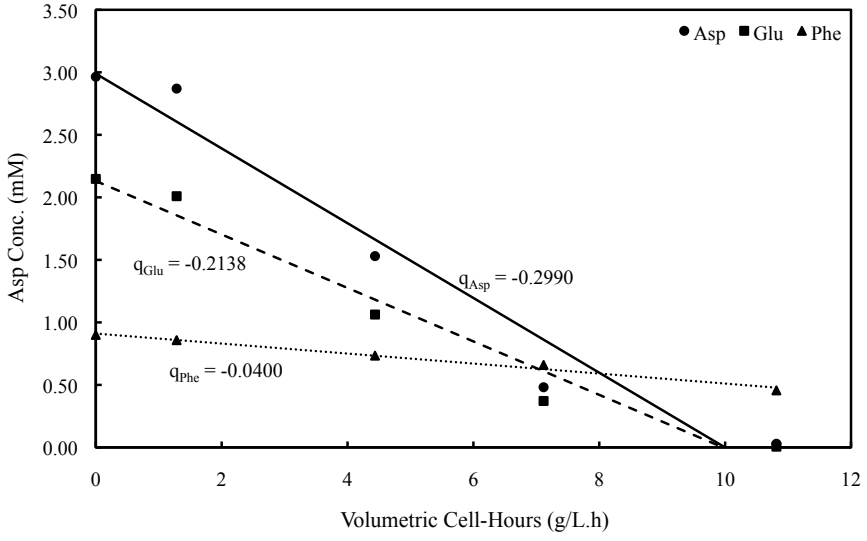

Figure 6: Dependence of amino acid uptake on volumetric cell hours

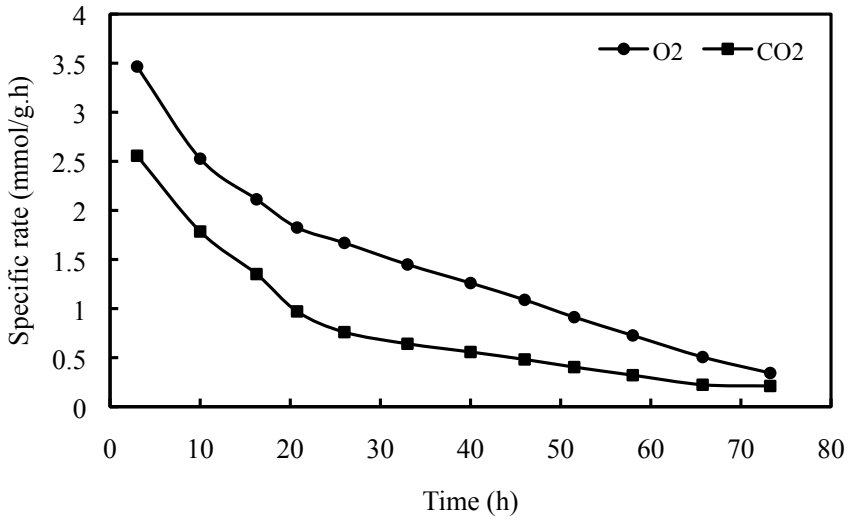

Figure 7: O_2 and CO_2 concentration profiles during fermentation

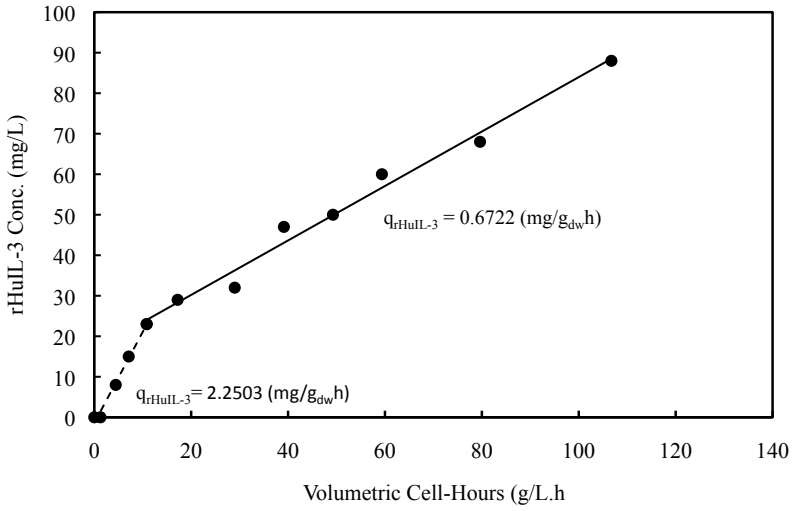

Figure 8: Dependence of rHuIL-3 production on volumetric cell hours

Table 5: Extracellular concentration of measured metabolites

	(g/l)						
Time (h)	Biomass	Asp	Glu	Phe	Glucose	NH_3	IL-3 (mg/l)
0	0.16±0.20	0.39	0.32	0.15	12.10	0.68	0
6	0.28±0.19	0.38	0.30	0.14	11.85	0.68	0
14	0.53±0.23	0.20	0.16	0.12	11.58	0.67	8±13
19	0.67±0.26	0.06	0.05	0.11	11.37	0.66	15±9
23	1.00±0.27	0.00	0.00	0.08	10.98	0.65	23±11
29	1.13±0.13	0.00	0.00	0.06	9.98	0.58	29±14
37	1.35±0.28	0.00	0.00	0.04	8.64	0.51	32±9
43	1.52±0.33	0.00	0.00	0.02	7.69	0.46	47±7
49	1.87±0.21	0.00	0.00	0.01	6.42	0.40	50±5
54	2.18±0.47	0.00	0.00	0.00	4.90	0.35	60±9
62	2.92±0.50	0.00	0.00	0.00	2.84	0.27	68±9
69	4.41±0.50	0.00	0.00	0.00	0.77	0.17	88±10
77	4.33±0.51	0.00	0.00	0.00	0.00	0.08	89±4