Elementary Mode Analysis of Ralstonia Eutropha H16 Metabolism for the Production of Useful Metabolites

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Gilsinia Lopez

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Prof. Friedrich Srienc

May 2013
Acknowledgements

This thesis would not have been possible without Your guidance. Throughout these 2 years, I have learnt that You will definitely come to my aid just a millisecond before I crash! You have sent me a few people to guard, guide and be with me. Thank You for

My family:

Father: My stronghold in the midst of chaos.
Mother: My path to righteousness and the guiding force of my life.
Consiniaa: The elixir of my life.

My Professors:

Prof Srienc: He has been a wonderful adviser, motivating and encouraging me with my research. Patience and abstraction are virtues that I must learn from him. I really, really enjoyed the many meetings we had.

Prof Sturtivant: He bought in joy through computer science. As a newbie in computer science, I initially felt lost among the huge undergraduate crowd. He believed in me and showed me how persistence and smart work can win any obstacle.

Prof Kazlauskas: A good friend, perfectionist and a nice mentor.

Prof Khodursky: Calm yet vigilant. A character to admire.

Prof Gralnick: The gel documentation system was very crucial for my project and he graciously gave me his lab key.

My Church:

Lawrence Newman Catholic Church: It is a wonderful community to worship and serve You. There is a peace that comes by just being there. It is and always will be a home away from home.
My Friends & Lab mates:

Dan Rouse: A font of optimism. Working with him was always a pleasure.

Pedro Pena: A very level headed person. He guided me with my experiments and sometimes in taking diplomatic decisions.

John Barrett: NaN error. A mystery as always! But he lets me call him Bhaiya (brother) so he is a nice person.

Chris Flynn: Though we have had very few conversations, they were very helpful in understanding my project.

Fr.Ivan Tou, Fr.Robert O’Donell, Neetha Eduthan, Caroline Lie, Azra Sultana, Buyu Chen, Diego Escalante, Florian Rossman, Peng Peng: Good companions through this wild journey.

Yours in Christ Jesus,

Lopez Gilsinia Gilroy
Abstract

*Ralstonia eutropha* H16 is a Gram-negative, facultative chemolithoautotrophic bacterium with the capability to synthesize many useful metabolites. One of the keys to the organism's lifestyle is its ability to use—alternatively or concomitantly—both organic compounds and molecular H$_2$ as sources of energy. It can fix CO$_2$ via the Calvin-Benson-Bassham (CBB) cycle and produces several useful metabolites like Poly(3-hydroxybutyric acid), isobutanol, 2,3-butanediol and ethanol. To quantitatively evaluate the capabilities of the metabolism we have set up models of the central metabolism that is based on known pathways of the organism. The lithoautotrophic metabolic model consists of 29 reversible reactions, 33 irreversible reactions, 59 internal metabolites and 11 external metabolites exchanged through the cell membrane. The heterotrophic model with fructose as a substrate has 31 reversible reactions, 47 irreversible reactions, 66 internal metabolites and 13 external metabolites. Elementary Mode Analysis identified 759 modes during lithoautotrophic growth and 135074 modes during heterotrophic growth. We have used the results from this analysis to predict key genetic alterations in the metabolism that would direct the metabolic flux towards the production of ethanol, isobutanol or polyhydroxybutanoate.
Table of Contents

List of tables .................................................................................................................. VII

List of figures .................................................................................................................. VIII

Chapter 1: Introduction ................................................................................................. 1
  1.1 Motivation .................................................................................................................. 1
  1.2 Elementary mode Analysis ....................................................................................... 2
  1.3 Ralstonia eutropha H16 ............................................................................................ 4
  1.4 Organization of the thesis ....................................................................................... 6

Chapter 2: Materials and Methods .............................................................................. 8
  2.1 Metabolic Models of R. eutropha H16 ..................................................................... 8
    2.1.1 Model construction ............................................................................................. 8
    2.1.2 Enumeration of Elementary modes ................................................................. 11
  2.2 Strains and media ................................................................................................... 11
    2.2.1 Ralstonia eutropha H16 .................................................................................... 11
    2.2.2 R. eutropha complex media ............................................................................ 11
    2.2.3 R. eutropha minimal media ............................................................................. 12
  2.3 Molecular Biology Techniques .............................................................................. 14
    2.3.1 PCR .................................................................................................................. 14
    2.3.2 Restriction digest ............................................................................................. 16
    2.3.3 Gel electrophoresis ......................................................................................... 16
    2.3.4 DNA purification ............................................................................................. 17
      2.3.4.1 Genomic DNA extraction ........................................................................... 17
      2.3.4.2 Plasmid DNA extraction .......................................................................... 18
      2.3.4.3 Linear DNA purification .......................................................................... 18
    2.3.5 In vitro DNA recombination ............................................................................ 18
    2.3.6 Plasmid design ................................................................................................ 19
    2.3.7 Primer Design .................................................................................................. 19
    2.3.8 Circular Polymerase Extension Cloning (CPEC) ........................................... 20
    2.3.9 S17-1 transformation ..................................................................................... 23
    2.3.10 Bi-parental mating ....................................................................................... 24
  2.4 Cell Growth ............................................................................................................ 25
    2.4.1 Ralstonia eutropha ........................................................................................... 25
    2.4.2 R. eutropha Minimal media growth .................................................................. 25
  2.5 Construction of the pexKan Suicide Plasmid ....................................................... 26
    2.5.1 Introduction to pex100 T suicide plasmid ..................................................... 26
    2.5.2 Methods for creation of pex 100 T suicide vector ....................................... 27
    2.5.3 Results ............................................................................................................ 29
Chapter 3: Metabolic models of *R. Eutropha* H16

3.1 Introduction
3.2 Lithoautotrophic (LA) model
3.3 Heterotrophic model
3.4 Results of elementary mode analysis
   3.4.1 Model Characterization & Comparison
   3.4.2 Validation
   3.4.3 Additional knockout target genes for model validation

Chapter 4: Automation of rational strain design

4.1 Introduction
4.2 Problem statement
4.3 Algorithm and software development
   4.3.1 System & Environment
   4.3.2 Requirements and software description
4.4 Results
   4.4.1 Ethanol Production
   4.4.2 Isobutanol Production
   4.4.3 PHB Production
   4.4.4 Predicted Molar Yields
   4.4.5 Identification of efficient secondary metabolite producing modes
4.5 Discussion
4.6 Future perspectives

Chapter 5: Estimation of elementary mode’s contributions to the overall metabolism from reaction entropies

5.1 Definition of the scientific problem
5.2 Significance
5.3 Methods
   5.3.1 Reconstruction of the input file
   5.3.2 Implementation
   5.3.3 Determination of flux from entropy
5.4 Pseudocode
   5.4.1 Function: Flux analysis (matrix)
   5.4.2 Function: newton_raphson()
5.5 Results & Discussion

Chapter 6: Construction of a ladh negative *R. Eutropha* strain, a fructose bisphosphatase negative *R. Eutropha* strain and a triose phosphate isomerase negative *R. Eutropha* strain

6.1 Introduction
6.2 Construction of suicide vector for knocking out lactaldehyde dehydrogenase, fructose 1,6 bisphosphatase and triose phosphate isomerase
6.3 Transfer of Plasmid to *R. eutropha*
6.4 SacB counter selection
6.4 Screening for the knockout phenotype following the counter selection
6.5 Results & Discussion
   6.5.1 Construction of the lactaldehyde dehydrogenase negative strain
   6.5.1.1 PCR for screening the constructed pexKanlac Plasmid
6.5.1.2 Verification of Properly Assembled pexKanlac Plasmid ........................................... 92
6.5.1.3 Screening for pexKanlac Integration into R. eutropha Genome ........................................... 93
6.5.1.4 Sac B counter selection for obtaining ladh negative strain: ........................................... 94
6.5.1.5 Sequencing of ladh negative strain ................................................................................. 95
6.5.2 Construction of the chromosomal fructose bisphosphatase negative strain ....................... 96
   6.5.2.1 Verification of Properly Assembled pexKancFBP Plasmid ........................................... 97
   6.5.2.2 Screening for pexKancFBP Integration into R. eutropha Genome ................................. 99
   6.5.2.3 Sac B counter selection for obtaining fbp negative strain: ........................................... 100
   6.5.2.4 Sequencing of fbp negative strain ............................................................................... 101
6.5.3 Construction of the Triose phosphate isomerase (TPI) negative strain ......................... 101
   6.5.3.1 PCR for screening Constructed pexKanTpi Plasmid: ........................................... 102
   6.5.3.2 Screening for pexKanTpi Integration into R. eutropha Genome ................................. 103
   6.5.3.3 Sac B counter selection for obtaining Tpi negative strain ........................................... 104

References ........................................................................................................................................ 105

Appendix A: List of reactions metatool input files ........................................................................ 113
   Lithoautotrophy ................................................................................................................. 113
   Heterotrophy ..................................................................................................................... 115

Appendix B: In depth description of rational strain design algorithm/software 118
   B.1 Requirements ............................................................................................................. 118
   B.2 Specifications ............................................................................................................ 119
   B.3 Wrapper Utility description (Code in DOS script) ....................................................... 119
   B.4 Minimal metabolic functionality (Coded in C++) ....................................................... 120
   B.5 Algorithm .................................................................................................................. 121
   B.6 operating instructions ............................................................................................... 123

Appendix C: List of primers: ...................................................................................................... 126
List of Tables

Table 2.1 The table describes the complex media recipe used to cultivate R. eutropha...

Table 2.2 Mineral Salts Media (MSM) recipe: The table describes MSM medium components. Multiple carbon sources like fructose/N-acetyl glucose amine can be used with the medium.

Table 2.3 Trace elements for Mineral Salts Medium:

Table 2.4 Template for PCR cycle:

Table 3.1 Metabolic Network map comprised of pathways and metabolities of R. eutropha during lithoautotrophy.

Table 3.2 Metabolic Network map comprised of pathways and metabolities of R. eutropha during heterotrophy with fructose.

Table 3.3 Characterization of the lithoautotrophic and heterotrophic models.

Table 3.4 All predicted single lethal deletion mutants of R. eutropha.

Table 4.1 Yields of ethanol, isobutanol or PHB optimized R. eutropha H16 strain.

Table 5.1 Description of the input file for determination of knockouts using entropy.
List of Figures

Fig 2.1 DNA ladder: The figure shows a 1 kb ladder obtained from Invitrogen ........ 17
Fig 2.2 A schematic diagram of the proposed CPEC mechanism for cloning an individual gene ........................................................................................................... 20
Fig 2.3: Map of plasmid pex100T with important features highlighted ....................... 26
Figure 2.4: 0.8 % agarose gel showing the amplified kanamycin region. Lane 1 – Kanamycin amplified at 70 °C from pgp564; Lane 2- Kanamycin amplified at 60 °C pgp564 ...................................................................................................................... 29
Fig 4.1 Implementation of multiple reaction deletion to rationally design the most efficient pathways for converting the substrate A into the product P. ......................... 55
Fig: 5.1 Plot of ΔS Vs the weighting factor .................................................................. 77
Fig 6.1 The map of the Ladh transcription unit .......................................................... 79
Fig 6.2 Diagrammatic representation of the pexKanLac integration into the R. Eutropha .................................................................................................................... 83
Fig 6.3 Map of primers Binding regions on the integrated pexKanLac plasmid .......... 84
Fig 6.4 Diagrammatic representation of the pexKancFBP integration into the R. Eutropha .................................................................................................................... 84
Fig 6.5 Diagrammatic representation of the pexKanTpi integration into the R. Eutropha .................................................................................................................... 85
Fig 6.6 Process flow diagram for the genomic integration of the suicide plasmid and the SacB counter selection ................................................................................. 87
Fig 6.8 Diagrammatic representation of the primers (LCTAD2R_ful_for- LCTAD2R_ful_rev) binding regions ......................................................................................... 88
Fig 6.9 Diagrammatic representation of the primers(FBP_For_veri - FBP_rev_veri ) binding regions .......................................................................................................... 89
Fig 6.10 Diagrammatic representation of the primers (Vec_Zn_for - Tpi_Rev) binding regions ................................................................................................................. 89
Fig 6.11 0.8 % gel displaying fragments used for invitro assembly of pexKanlac....... 90
Fig 6.12 PCR from E.coli (S17) colonies following transformation of the invitro assembled pexKanlac plasmid .................................................................................... 91
Fig 6.13 SmaI & Ast II digest of pexKanlac plasmid from E. coli plasmid extraction:92
Fig 6.14 0.8 % agarose gel verifying the genomic integration of pexKanLac plasmid into R. eutropha .............................................................................................................. 93
Fig 6.15 Map of the plasmid integrated into R. eutropha Genome: ............................ 94
Fig 6.16 PCR from R. eutropha colonies following growth on sucrose media: 10 colonies ..................................................................................................................... 94
Fig 6.17 0.8 % gel displaying fragments used for invitro assembly of pexKancFBP ... 96
Fig 6.18 PCR from E.coli (S17) colonies following transformation of the invitro assembled pexKancFBP plasmid ................................................................. 97
Fig 6.19 BglI digest of upstream-downstream fragment pexKancFBP plasmid: ....... 98
Fig 6.20 0.8 % agarose gel verifying the genomic integration of pexKancFBP plasmid into R eutropha ............................................................................................................ 99
Fig 6.21 Map of the penKanFBP plasmid integrated into R. eutropha Genome: .... 100
Fig 6.22 PCR from R. eutropha colonies following growth on sucrose ................. 100
Fig 6.23 0.8 % gel displaying fragments used for invitro assembly of pexKanlac:..... 101
Fig 6.24 PCR from E.coli (S17) colonies following transformation of the invitro assembled pexKanTpi plasmid................................................................. 102
Fig 6.25 0.8 % agarose gel verifying the genomic integration of pexKanTpi plasmid into R. eutropha.............................................................................. 103
Fig 6.26 Map of the penKanTpi plasmid integrated into R. eutropha Genome:....... 104
Fig 6.27 PCR from R. eutropha colonies following growth on sucrose media........ 104
Chapter 1: Introduction

1.1 Motivation

Biofuels are the common buzzwords found in the current environmental discussions. Our economies are oil dependent. Repeatedly history teaches us that supply shortfall leads to high oil prices and high oil prices lead to economic recessions. Economic growth requires oil, preferably cheap oil derived from non-feed biomass. Using greenhouse gas like carbon dioxide to produce biofuels would be equivalent to killing two birds with one stone. We have attempted to engineer the metabolic capabilities of *R. eutropha* H16 as a means to achieving this goal. In general, there are two main approaches to metabolic engineering. First, some enzyme(s) can be blocked completely or removed by knocking out the corresponding genes, or new enzyme(s) can be inserted. Thus, the topology of the reaction network is changed, and this can be supported by theoretical analysis. Second, enzyme activities can be changed gradually or regulatory interactions can be altered. This requires much more extensive data and sophisticated theoretical tools because complex kinetic information is involved. Elementary mode analysis has been employed to predict the metabolic flux of the organism. A model has been developed in this study to aid with identifying minimum number of enzymes to be knocked out for developing an organism optimized for the production of the desired metabolite.
1.2 Elementary mode Analysis

Elementary mode analysis is a useful metabolic pathway analysis tool to identify the structure of a metabolic network that links the cellular prototype to the corresponding genotype (Trinh et al., 2009). Elementary modes represent unique, non-decomposable, minimal sets of reactions within the metabolic network that operate under the steady-state stoichiometrically balanced reaction constraint. They support steady state operation of cellular metabolism that represents independent cellular physiological states. This analysis identifies the topology of the cellular metabolism based on only the stoichiometric structure and thermodynamic constraints of the reactions. Inputs to the elementary mode analysis are a set of irreversible and reversible reactions of the central metabolism along with the reactions required for the production of the desired secondary metabolite. Reactions that transform metabolites within the system can be considered internal reactions while reactions involving transport of metabolites in and out of the cell are considered to be exchange reactions. Some of the important terms relevant to this study are described in the box below.
**Elementary mode**: Minimal set of enzymes that could operate at steady state with all irreversible reactions proceeding in the appropriate direction.

**Irreversible reaction** A reaction in which the rate of the forward reaction is always (in all physiological conditions) so much higher than the rate of the reverse reaction that the latter is relatively negligible.

**Yield of a biotransformation** Ratio indicating how many moles of product are produced per mole of substrate used.

**External metabolites** Metabolites buffered by connection to reservoirs. They can be considered to be sources and sinks (nutrients and waste products, stored or excreted products, or precursors for further transformations).

**Internal metabolites** Metabolites only participating in reactions of the model. The formation of each is exactly balanced by its consumption (steady-state assumption).

**Linear combination (superposition)** Sum of several terms, each of which may be weighted by an appropriate coefficient.

**Box 1**: Glossary of important terms obtained from Schuster *et al.*, 2009.

For a better understanding of the concept and the principles behind the elementary mode analysis, the reader is directed to a concise review by Trinh *et al.*, 2009.
1.3 *Ralstonia eutropha* H16

*Ralstonia eutropha* H16 is an aerobic hydrogen oxidizing, facultative chemolithotrophic gram negative bacterium with the capability to synthesize many useful metabolites. It is the best studied producer of poly(3-hydroxybutyric acid) (Reinecke & Steinbüchel, 2009) and is known to produce a wide variety of other industrially important compounds like isobutanol (Lie et al., 2012; Lu et al., 2012), 2,3-butanediol, methyl citric acid (Brämer & Steinbüchel) and ethanol. The biomass generated as a byproduct can be used as animal feed (Repaske & Mayer, 1976). Also, it is a versatile organism which can be easily manipulated genetically for the creation of knock out mutations, for integration of genes into the chromosome and it can maintain plasmids with appropriate selection techniques.

The growth requirements for *R. eutropha H16* are fairly simple and are suited for large scale industrial production (Schlegel & Lafferty, 1971). One of the keys to the organism's lifestyle is its ability to use—alternatively or concomitantly—both organic compounds and molecular hydrogen as sources of energy. It can use a variety of organic substrates as electron donors and carbon sources (Rittenberg, 1969). During autotrophy, gaseous hydrogen or formate is used as the electron donor and oxygen as the electron acceptor to fix carbon dioxide. Fructose (Fleige et al., 2011) and N-acetyl-D-glucosamine are the only two sugars that can be metabolized. Apart from the sugars, it can also grow on tricarboxylic acid intermediates like pyruvate and lactate, fatty acids (Kelley & Srienc, 1999; Yu & Si, 2004), amino acids, organic acids (Wang et al., 2007). The growth rate is higher on organic substrates than under autotrophic conditions and undergoes versatile transition from heterotrophism with certain carbon
sources like fructose to autotrophism (Repaske, 1968; Stukus & Decico, 1970). Mixotrophy has also been documented. (DeCicco & Stukus, 1968; Rittenberg & Goodman, 1969).

The genome of *R. eutropha* H16 has mutli-replicon architecture with two chromosomes (Chr1, Chr2) and one megaplasmid (pHG1) (Fricke et al., 2009). Chromosome 1 encodes most key functions of DNA replication, transcription and translation, including the ribosomal proteins. Chromosome 2 harbors genes for central steps of the 2-keto-3-deoxy-6-phosphogluconate (KDPG) sugar and sugar acid catabolism path, the decomposition of aromatic compounds and the utilization of alternative nitrogen sources. *Cbb* genes are responsible for growth during lithoautotrophy and organo-autotrophy. One set of *cbb* genes (Gene locus: PHG416-428, Operon name: cbbₚ) maps on the megaplasmid pHG1, and the second (Gene locus: H16_B1384-1396, Operon name: cbbₗ) is located on chromosome 2. A regulatory gene *cbbR*, encoding a transcriptional activator, belongs to the chromosomal locus. A plasmid-borne copy of *cbbR* is present but defective, so the chromosomal copy controls the coordinate expression of both operon (Pohlmann et al., 2006).

An extensive protein repertoire aids the organism in adapting to alternative energy and carbon sources. For instance, phosphoglycerate mutase which catalyzes the interconversion 2- and 3- phosphoglycerate of is essential to lithoautotrophy (Reutz et al., 1982) but is dispensable when during growth on fructose because of the differences in the carbon utilization. To quantitatively evaluate the capabilities of the metabolism we have set up a model of the central metabolism that is based on known pathways of the organism. In this study, we describe the metabolic models for lithoautotrophy,
organo-autotrophy and heterotrophy. The metabolic profile of the organism varies with the substrates provided and we have attempted to build the model based on the existing literature, gene expression profile and experimental validations. The stoichiometric models described herein would serve as groundwork for network evaluation using elementary mode analysis or flux balance analysis.

1.4 Organization of the thesis

An interest in development of biofuels and biopolymers has surged because of the urgency of achieving energy self-sufficiency. They are viewed as means of reducing dependency on fossil fuels and developing environmentally friendly and renewable energy. Biofuels produce fewer greenhouse gas (GHG) emissions than petroleum-based fuel. Their economic viability is still a critical issue to large-scale commercialization. The current study focusses on developing a metabolic model of *R. eutropha* which is capable of surviving on a gaseous mixture of Carbon di-oxide, oxygen and hydrogen and can produce isobutanol, ethanol and biopolymer (polyhydroxybutyrate). Chapter 2 describes the methods employed in manipulating *R. eutropha H16*. The metabolic model and its developments have been discussed in Chapter 3. This model serves as an input for elementary mode analysis. Elementary mode analysis is essential for rational strain design. Automation of the rational strain design has been described in Chapter 4. Metabolic flux distributions in living cells can be represented as non-negative linear combinations of elementary modes. Chapter 5 describes software for estimating the metabolic flux through each elementary mode family by using their reaction entropies. Success of any computational model depends on the experimental validation. Chapter 6
describes the development two model knockouts namely: (i) lactaldehyde dehydrogenase, an enzyme belonging the secondary metabolism and (ii) fructose bisphosphatase, an enzyme of the central metabolism (iii) triose phosphate isomerase.
Chapter 2: Materials and Methods

2.1 Metabolic Models of *R. eutropha* H16

2.1.1 Model construction

The model is based on the reactions described by Carlson & Srienc, 2003. Refinements were made with the metabolic information available from database at Kyoto Encyclopedia of Gene and Genomes KEGG (http://www.genome.jp/kegg/), biocyc project (http://www.biocyc.org/) and the Rhodocyc (http://rhodocyc.broadinstitute.org) project. The model comprises of the carbon fixation pathway, anaplerotic tricarboxylic acid cycle, anabolic gluconeogenesis, and energy generating Embden-Meyerhof with the alternate Entner–Doudoroff and oxidative phosphorylation reactions along with the reactions for the production of ethanol, isobutanol, poly-hydroxyvalerate, lactate, biomass and formate.

In general, lithoautotrophs fix carbon via the Calvin reductive pentose phosphate cycle (Calvin-Benson-Bassham) using energy from hydrogen/formate oxidation. In *R. eutropha* H16, one carbon compounds are assimilated via the Calvin cycle (Bowien & Schlegel, 1981). The carbons fixed are oxidized via the Embden-Meyerhof / Entner–Doudoroff pathway. The glycolytic pathway is incomplete, lacking a functional phosphofructokinase (E.C. 2.7.1.11) which catalyzes the reversible conversion of D-fructose 6-phosphate to D-fructose 1, 6-bisphosphate (Pohlmann *et al.*, 2006). The presence of fructose-1, 6-bisphosphatase (EC: 3.1.3.11) enables gluconeogenesis. Expression of enzymes belonging to the Entner–Doudoroff pathway is substrate dependent (Schlegel & Trüper, 1966).
The amphibolic TCA cycle is employed to generate energy via its link to the respiratory chain and to provide substrates for anabolic purposes. (Trüper, 1965; Glaeser & Schlegel, 1972). The NADPH generated via the TCA cycle was used predominantly for the synthesis of glutamate from alpha ketoglutarate when NH3 is abundant. In the absence of NH3, NADPH is used for synthesis of PHB (Katoh et al., 1999). It plays an important role in generating intermediates like 2-ketoglutarate, oxaloacetate and succinyl-CoA which are starting points for the synthesis of glutamate and aspartate. The organism utilizes a NADP dependent Isocitrate dehydrogenase (Trüper, 1965). Glutamate and aspartate are essential for biomass generation. Ralstonia can accumulate PHB at > 80 % of the cell dry weight (Steinbüchel, 1991). The stoichiometric formula for biomass without PHB was used in the model (Katoh et al., 1999 and Shi et al., 1997).

Like most fatty acid/acetate utilizing bacteria, the glyoxylate cycle is an indispensable constituent of the citric acid cycle in R. eutropha. Pyruvate has to be formed from oxaloacetate which is replenished via the glyoxylate cycle. The glyoxylate shunt has two unique enzymes Isocitrate lyase (ICL, EC 4.1.3.1) and malate synthase (MS, EC 4.1.3.2) which enables it to by-pass the decarboxylation step of the tricarboxylic acid cycle.

In R. eutropha, three reactions, (i) reversible carboxylation of phosphoenolpyruvate (PEP) to oxaloacetate (phosphoenolpyruvate carboxykinase), (ii)oxidative decarboxylation of malate to pyruvate (malic enzyme), (iii) conversion of pyruvate to PEP (PEP synthase) are responsible for the interconversion of C3 and C4
compounds (Schobert & Bowien, 1984; Trüper, 1965). Only NADP utilizing Malic enzyme (Ec 1.1.1.40) has been documented (Trüper, 1965; Schobert & Bowien, 1984; Brüland et al., 2010). Activity of NAD utilizing malic enzyme (Ec 1.1.1.38) could not be detected (Trüper, 1965). PEP carboxylase (C3 carboxylating enzyme) essential for PEP to OAA conversion and pyruvate carboxylase catalyzing reversible conversion of pyruvate to oxaloacetate are not functional (Schwartz et al., 2009; Brüland et al., 2010; IM & Friedrich, 1983). Hence, phosphoenolpyruvate carboxykinase is the only C3 carboxylating enzyme in Ralstonia eutropha.

*R. eutropha* H16 is capable of surviving under transient anoxia. The anaerobic growth proceeds at a slow rate in heterotrophic growth and is almost absent during autotrophy (Aragno & Schlegel, 1992). Fermentation enzymes are secreted when the oxygen supply is limited (Cramm, 2009). The models described below are designed to mimic the aerobic growth.

Among the oxidative phosphorylation reactions, ATPSYN: $1 \text{h}_2\text{O} + 1 \text{atp} = 1 \text{adp} + 1 \text{atp}_\text{base}$ was added to account for maintenance energy and ATP hydrolysis. A low activity of NADPH-NAD+ transhydrogenase which catalyzes the conversion reaction between NADH and NADPH has been detected in crude extracts (Lee et al., 1995; Katoh et al., 1999). For the purposes of the model, NADH and NADPH are considered to be equivalent. The efficiency of ATP production (P/O number) from NADH and FDH have been considered to be 2 and 1 respectively although there are some reports indicating P/O numbers can be as high 4 and 3 respectively when 20 micro molar iron citrate is used in the medium (Bongers, 1970; Ishaque & Aleem, 1970; Drozd & Jones, 1974; Jones et al., 1975).
2.1.2 Enumeration of Elementary modes

The elementary modes were calculated using CellNetAnalyzer, a toolbox for MATLAB ® (Mathworks Inc.) facilitating, in an interactive and visual manner (Klamt et al., 2007).

2.2 Strains and media

2.2.1 Ralstonia eutropha H16

The wild type Ralstonia eutropha H16 was used in the experiments. Both complex and minimal media were used during the course of the experiments.

2.2.2 R. eutropha complex media

The complex media used for the growth of R. eutropha was termed Doi media. The composition used is as follows:

Table 2.1 The table describes the complex media recipe used to cultivate R. eutropha

<table>
<thead>
<tr>
<th>Component</th>
<th>g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yeast Extract</td>
<td>10</td>
</tr>
<tr>
<td>Tryptose</td>
<td>10</td>
</tr>
<tr>
<td>Beef Extract</td>
<td>5</td>
</tr>
</tbody>
</table>

The media was made by dissolving all the components in distilled water and then brought to the final volume after the components were completely dissolved. The media was then autoclaved to sterilize.
2.2.3 *R. eutropha* minimal media

Minimal medium was used for multiple purposes while growing *R. eutropha*. In each case the only thing that differed was the carbon source. The minimal medium used was termed Mineral Salts Media (MSM) and was based off of one used by Ramsay *et al.* (Ramsay, 1990).
**Table 2.2** Mineral Salts Media (MSM) recipe: The table describes MSM medium components. Multiple carbon sources like fructose/N-acetyl glucose amine can be used with the medium.

<table>
<thead>
<tr>
<th>Component</th>
<th>g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Na₂HPO₄ * 7H₂O</td>
<td>6.7</td>
</tr>
<tr>
<td>KH₂PO₄</td>
<td>1.5</td>
</tr>
<tr>
<td>(NH₄)₂SO₄</td>
<td>4</td>
</tr>
<tr>
<td>MgSO₄ * 7H₂O</td>
<td>0.2</td>
</tr>
<tr>
<td>ferric ammonium citrate</td>
<td>6 mg</td>
</tr>
<tr>
<td>CaCl₂</td>
<td>7.5 mg</td>
</tr>
<tr>
<td>Trace elements (1000x)</td>
<td>1 mL</td>
</tr>
<tr>
<td>Carbon source</td>
<td>20</td>
</tr>
</tbody>
</table>

The medium was made by dissolving the phosphates in distilled water and autoclaving to sterilize the solution. The rest of the components were dissolved in distilled water, filtered to sterilize, and added after the autoclaved solution was allowed to cool. 1 mL of a filtered trace element solution was added at the same time as the filtered components. The recipe for the 1000X trace elements solution is shown in Table 2.3.
Table 2.3 Trace elements for Mineral Salts Medium: The stock solution of the trace elements has been described in this table.

<table>
<thead>
<tr>
<th>Component</th>
<th>g/L</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\text{H}_3\text{BO}_3$</td>
<td>0.3</td>
</tr>
<tr>
<td>$\text{CoCl}_2 \cdot 6\text{H}_2\text{O}$</td>
<td>0.2</td>
</tr>
<tr>
<td>$\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$</td>
<td>10 mg</td>
</tr>
<tr>
<td>$\text{MnCl}_2 \cdot 4\text{H}_2\text{O}$</td>
<td>30 mg</td>
</tr>
<tr>
<td>$\text{NiCl}_2 \cdot 6\text{H}_2\text{O}$</td>
<td>20 mg</td>
</tr>
<tr>
<td>$\text{NaMo}_4 \cdot 2\text{H}_2\text{O}$</td>
<td>30 mg</td>
</tr>
<tr>
<td>$\text{ZnSO}_4 \cdot 7\text{H}_2\text{O}$</td>
<td>0.1</td>
</tr>
</tbody>
</table>

The trace elements solution was made by dissolving all of the components in distilled water and filtering with a 0.2 µm filter to sterilize the solution.

2.3 Molecular Biology Techniques

2.3.1 PCR

PCR was routinely used to clone out regions of DNA from the genomes of microbes, or to detect the presence of certain genes in a culture. GoTaq® (Promega, Madison, WI) was used as the master mix for each reaction which contained everything needed for the PCR reaction in a 2X buffer except the primers and template DNA. All
reactions were carried out using a Bio-Rad (Hercules, CA) MJ Mini Personal Thermal Cycler. Although the annealing temperatures vary between PCR reactions depending on the melting temperatures of the primers used, the same program was used for each reaction with most steps left unchanged.

Table 2.4 Template for PCR cycle: Table shows the template for the thermal cycler settings used during all of the PCR reactions used in this thesis.

<table>
<thead>
<tr>
<th>Step</th>
<th>Name</th>
<th>Temp (°C)</th>
<th>Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Initial denaturing</td>
<td>94</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Primer anneal</td>
<td>Various</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>Extension</td>
<td>72</td>
<td>1 min/kb</td>
</tr>
<tr>
<td>4</td>
<td>Denature</td>
<td>94</td>
<td>0.5</td>
</tr>
<tr>
<td>5</td>
<td>Go to Step 2</td>
<td>28-32 times</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Primer anneal</td>
<td>Various</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>Final Extension</td>
<td>72</td>
<td>5</td>
</tr>
</tbody>
</table>

The initial denaturing step was run at 94°C for 5 minutes, the annealing steps were 1 minute, the extension was 72°C for 1 minute per 1000 bases to be cloned, the
repeated denaturing step was 94°C for 30 seconds, and the final extension period was run for 5 minutes.

2.3.2 Restriction digest

All enzymes were purchased from Invitrogen (Carlsbad, CA) or NEB (Ipswich, MA) and used according to their respective protocols. 100-1000 ng of DNA was incubated with a 1:10 dilution of the 10X reaction buffer supplied with the enzymes, and 1 µL of the respective enzyme. All digests were incubated at 37°C overnight unless otherwise stated.

2.3.3 Gel electrophoresis

PCR products and plasmid DNA were routinely analyzed by agarose gel electrophoresis on a 0.8% agarose gel run at 150 volts. Pictures of the gel were taken in a UV light box with a transilluminator to be further analyzed by computer. DNA fragments were regularly separated by electrophoresis and purified from the gel to be used in downstream applications such as homologous recombination. Figure 2-1 displays the DNA ladder used on all agarose gels run to measure the size of the DNA fragments analyzed by electrophoresis.
Fig 2.1 DNA ladder: The figure shows a 1 kb ladder obtained from Invitrogen

The above figure is from the manual supplied with the 1 Kb ladder purchased from Invitrogen (catalogue #15615-015). It displays the sizes of each band used for measurement during routine DNA fragment analysis. The same ladder was used for all experiments described throughout the thesis.

2.3.4 DNA purification

2.3.4.1 Genomic DNA extraction

Genomic DNA was routinely extracted from a culture to be used as the template for amplifying genes by PCR to be used in downstream applications such as homologous recombination, or used to screen for the insertion of genes into the genome
during experiments. Genomic DNA was purified from microbial cultures using Purelink™ Genomic DNA Mini Kits purchased from Invitrogen (catalogue # K1820-01) and used according to the manufacturer’s protocol.

2.3.4.2 Plasmid DNA extraction

Plasmid DNA was extracted from bacterial cultures to verify the presence of a certain plasmid or to harvest the DNA to be used for downstream applications such as restriction digests. Bacterial plasmid extractions were performed using the Wizard® Plus SV Miniprep kits from Promega (Madison, WI) according to the manufacturer’s protocol.

2.3.4.3 Linear DNA purification

Linear DNA fragments were purified for use in downstream applications such as homologous recombination in yeast from agarose gels or from a PCR reaction mixture. The fragments were purified from PCR mixtures and agarose gels using Qiaquick PCR Cleanup Kits and Qiaquick Gel Extraction Kit respectively according to the manufacturer’s protocol (Qiagen, Hilden, Germany).

2.3.5 Invitro DNA recombination

The plasmids described in Chapter 6 & 7 were created using invitro DNA assembly method. The procedure involves 5 stages: design of plasmids and primers, PCR amplification of desired fragments, Circular Polymerase Extension Cloning (CPEC) and transformation.
2.3.6 Plasmid design

In order for the recombined plasmid to replicate in S17 cells, an origin of replication and a selection marker are essential. After the plasmid has been constructed in S17 cells, it will be transferred to R. eutropha. Kanamycin was chosen as the resistance marker because R. eutropha is natively resistant to ampicillin.

2.3.7 Primer Design

The DNA pieces used in the recombination were combination of PCR products and linearized plasmid vector. The first step in the design stage was to create a map of the desired plasmid using Ape plasmid editor. Then the DNA fragments necessary to assemble the designed plasmid were determined, and primers were designed to clone the respective fragments from genomic DNA. 40 bases of homology to the adjacent DNA fragment on the final plasmid were added to the end of the primer sequence as an overhang to the original DNA template. For the pieces next to a digested plasmid vector, 40 terminal bases of homology were designed to match the vector sustaining the restriction site used to linearize the vector for easier screening downstream.

For the second step, the DNA fragments used in the recombination were amplified via PCR or digested by restriction enzymes and purified using the protocols described previously. Regardless of the protocol used to purify the DNA, the fragments were eluted from the spin columns using sterile nuclease free water at pH 7.
2.3.8 Circular Polymerase Extension Cloning (CPEC)

Polymerase extension was used for gene assembly with overlapping oligonucleotides or gene fragments. The method extending overlapping regions between the insert and vector fragments to form a complete circular plasmid and is therefore named “Circular Polymerase Extension Cloning” (Quan & Tian, 2009). A pictorial representation of the method is shown below.

Fig 2.2 A schematic diagram of the proposed CPEC mechanism for cloning an individual gene. The vector and the insert with overlapping regions at the ends are denatured and annealed (Step 1). The hybridized insert and vector extend using each other as a template until they complete a full circle and reach their own 5′-ends (Step 2). The final completely assembled plasmid has two nicks, one on each strand, at the positions marked by an arrowhead. They are used for transformation (Step 3) without further purification. (Adapted from Quan & Tian, 2009).
A three/two-piece *invitro* DNA assembly was performed. The DNA concentration was measured using Nanodrop 1000 Spectrophotometer. 100 ng of the linearized vector and equimolar amounts of other assembly pieces were added to a final volume of 25 µl assembly reaction mixture as follows:

<table>
<thead>
<tr>
<th>Components</th>
<th>Amount in µl</th>
</tr>
</thead>
<tbody>
<tr>
<td>linearized vector backbone (100 ng)</td>
<td>Varies</td>
</tr>
<tr>
<td>Upstream &amp; downstream pieces</td>
<td>Varies</td>
</tr>
<tr>
<td>5X HF Phusion Reaction Buffer</td>
<td>5</td>
</tr>
<tr>
<td>10 mM dNTPs</td>
<td>1</td>
</tr>
<tr>
<td>DMSO</td>
<td>0.75</td>
</tr>
<tr>
<td>2U/µl Phusion Polymerase</td>
<td>0.5</td>
</tr>
<tr>
<td>dH₂O</td>
<td>Varies</td>
</tr>
<tr>
<td>Final Volume</td>
<td>25</td>
</tr>
</tbody>
</table>
A PCR was performed with to extend the overlapping fragments. The program used is as follows:

<table>
<thead>
<tr>
<th>Step</th>
<th>Name</th>
<th>Temp (°C)</th>
<th>Time (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Initial denaturing</td>
<td>98</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>Primer anneal</td>
<td>55</td>
<td>0.5</td>
</tr>
<tr>
<td>3</td>
<td>Extension</td>
<td>72</td>
<td>0.25 min/kb</td>
</tr>
<tr>
<td>4</td>
<td>Denature</td>
<td>98</td>
<td>0.5</td>
</tr>
<tr>
<td>5</td>
<td>Go to Step 2</td>
<td></td>
<td>28-32 times</td>
</tr>
<tr>
<td>6</td>
<td>Primer anneal</td>
<td>55</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>Final Extension</td>
<td>72</td>
<td>10</td>
</tr>
</tbody>
</table>

5 µl of the assembly reaction was transformed into 100µl of competent S17 cells. The procedure for transformation is outlined.
2.3.9 S17-1 transformation

Mating strains of *E. coli* termed S17-1 were made chemically competent in the presence of a transformation and storage solution (TSS). TSS solution was made by mixing 15 g PEG 4000, 5 mL of 1 M MgCl₂, and adding LB media to 95 mL. The pH was adjusted to 6.5 and then autoclaved. 5 mL of DMSO were added after the solution was allowed to cool, and after mixing the solution aliquots were then stored at -20°C until needed.

Competent cells were prepared by growing S17-1 in complex medium (LB) overnight in 5mL culture tubes. 500 µL of the overnight culture were used to seed 50 mL of LB and incubated at 37°C to start the cells rapidly growing. The OD was monitored and when the cells reached an OD of 0.4-0.6 they were centrifuged for 10 min at 4000 x g at 4°C. The supernatant was poured off and the cells were carefully re-suspended in 2 mL of the cold TSS solution per 50mL of culture and incubated on ice for 5 minutes. The cell suspension was then dispensed into 100 µL aliquots, flash frozen in liquid nitrogen, and stored at -80°C for later use.

Typically 5 µl of the assembly reaction was added directly to the cell suspension and gently mixed by swirling the tip of the pipette around the culture. The cells were incubated on ice for 30 minutes, heat shocked at 42°C for 45 seconds, and then placed back on ice for 1 minute. 500 µL of 37°C LB media was added to the cell suspension and then incubated at 37°C for 1-4 hours. Serial dilutions were then plated on LB agar plates containing the appropriate antibiotic selection marker and incubated overnight at 37°C. Colonies were then screened to contain the appropriate plasmid and saved for downstream applications.
2.3.10 Bi-parental mating

Plasmid transfer to *R. eutropha* was achieved by conjugation with E. coli S17-1 as plasmid donor. This technique is called the bi-parental mating.

This method involves transfer of the plasmid of interest from one species of bacteria, called the donor strain, to a strain that you wish to accept the plasmid called the recipient strain through a physical linkage between the two bacteria. The process of bi-parental mating requires that the donor strain has the genes necessary for the linkage of the two bacteria and the transfer of the plasmid to the recipient strain called the tra genes. It is also required that the plasmid to be transferred has the P15A origin of replication (ori) sequence (origin of transfer). These DNA sequences are recognized by some of the tra genes and are necessary to facilitate the transfer of the plasmid to the recipient strain. The protocols for bi-parental mating in *R. eutropha* have been described by several authors (Hynes & Quandt, 1993; Srinivasan *et al.*, 2002).

The donor strain S17 was overnight in 5 mL of LB medium containing the kanamycin, and the recipient strain, *Ralstonia eutropha* in 5 mL of Doi media with no antibiotics. 500 µL of the donor strain was removed from the culture and centrifuged at 10,000 rpm in a microcentrifuge for 2 minutes. The supernatant was discarded and the residual antibiotic was removed by washing the cells twice with a 0.8% NaCl. The donor and recipient cells were suspended in 1mL of 0.8% NaCl and mixed together. The cells were centrifuged once more at 10,000 rpm for 2 minutes, the supernatant discarded, and then re-suspended in 1 mL of LB or Doi media ensuring a homogeneous mixture of the two bacterial strains. Then 100 µL was transferred to a Doi agar plate, spread in a circle about 2 inches in diameter, and allowed to incubate at 30°C, the
optimal temperature for growth of the recipient strain *R. eutropha*.

A clump of cells was removed from the mating plate and resuspended in a Doi media creating a homogeneous solution. Serial dilutions of the concentrated cell solution were plated on Doi media with gentamycin sulfate and kanamycin sulfate to preferentially allow the recipient strain harboring the plasmid of interest to grow while suppressing the growth of all other cells in the mixture. Alternatively for *R. eutropha*, Mineral Salt Medias (MSM) plates with kanamycin were made containing no carbon source and incubated at 30 °C in a hydrogen rich atmosphere for up to 5 days. Colonies that formed were picked and re-streaked for isolated colonies on another MSM plate and incubated in the same fashion to ensure no donor cells were able to contaminate the cells of interest. Single colonies were then picked from the secondary screening plates and screened for the plasmid of interest.

### 2.4 Cell Growth

#### 2.4.1 *Ralstonia eutropha*

*R. eutropha* was used in experiments. Cultures were maintained on Doi media plates containing 16 g/L agar. Cultures grown in tubes or shake flasks were cultured at 30°C in a Lab-Line incubator-shaker at 250 rpm.

#### 2.4.2 *R. eutropha* Minimal media growth

*R. eutropha* was grown heterotrophically on fructose and autotrophically with 80% CO2: 10% H2: 10% O2. For autotrophic growth, the MSM media described earlier was supplemented with 2% fructose and 1.6% w/vol of agarose. Heterotrophic growth requires 2% w/vol fructose. The cells were allowed to grow for 3-5 days after which
opaque colonies appeared due to the accumulation of PHB.

2.5 Construction of the pexKan Suicide Plasmid

2.5.1 Introduction to pex100 T suicide plasmid

Pex-100T, a suicide vector used in previous studies for chromosomal integration in *R. eutropha* (Ewering *et al*, 2006), was purchased from ATCC (ATCC™, Manassas, VA). Origin of replication on the plasmid is not recognized by *R. eutropha*. When challenged with antibiotics in the surrounding media, the organism is forced to integrate the plasmid into its genome in order to retain the antibiotic resistance gene located on the vector. The plasmid is the sacB gene from *Bacillus subtilis*. sacB codes for the peptide levansucrase, which catalyzes the reaction of Sucrose $\rightarrow \beta(2,6)$-linked fructan, otherwise known as levan, a polysaccharide toxic to most bacteria (Gay, Le Coq, Steinmetz, Ferrari, & Hoch, 1983). Below is a map of the plasmid as it was received from ATCC.

**Fig 2.3:** Map of plasmid pex100T with important features highlighted. The pEX100T suicide vector obtained from Invitrogen is shown here with its major components highlighted. ampR – ampicillin resistance marker; oriT – origin of replication; mcs – multiple cloning site.
Pex-100T is a suicide vector purchased from Invitrogen and used to create a suicide vector capable of integrating into the genome of R. eutropha. However, R. eutropha is natively resistant to ampicillin. Hence, kanamycin cassette of the pGP-564 (Thermo Scientific, Waltham, MA), a yeast – E. coli shuttle vector, was combined with the pex100T plasmid to gain all of the fragments necessary to assemble a suicide vector by using the CPEC, or circular polymerase extension cloning (Quan J. and Tian J, 2009). The CPEC method has been described earlier.

### 2.5.2 Methods for creation of pex 100 T suicide vector

The pex100 T vector was extracted from the E. coli colonies as described earlier in this chapter and digested with unique restriction enzyme SmaI (NEB, Ipswich, MA) and run on a 0.8% agarose gel to verify complete digestion of the circular plasmid. The Kanamycin region was amplified from the pGP564 plasmid using 60 base pair primers Kan R pex For - Kan R pex Rev.

**Kan R pex For**  
Forward primer for amplifying kanamycin fragment  
**Kan R pex Rev**  
Reverse primer for amplifying kanamycin fragment

The PCR program had an annealing temperature of 62 °C and extension time of 1.30 minutes.

The primers were designed to have a 40 base pair homology to the pex 100T plasmid. The vector and the fragment were run on a 0.8% agarose gel with 0.1% vol/vol crystal violet, excised and purified as described earlier. The purified fragments were added in equimolar quantities to the circular polymerase extension cloning
(CPEC) mix described earlier. After CPEC, 5 µl of the reaction mix was added to 100 µl of chemically competent S17 cells and selected on LB plates with ampicillin (100 mg/mL) and kanamycin (50 mg/mL).

Several colonies that were formed on the experimental plates were picked and screened with the primers used for amplifying the kanamycin resistance gene. The PCR products were then run on an agarose gel to check for the presence of the amplified fragment indicating the presence of the recombined plasmid in the colony. One of the colonies that displayed a positive PCR result was grown overnight in SD media and the plasmids were extracted. They were then digested with Bam HI to verify proper construction of the plasmid.
2.5.3 Results

The kanamycin amplified from the plasmid pGP564 using primers Kan R pex For and Kan R pex Rev. The gel below shows the 1099 base pair fragment.

![Image](image.png)

**Figure 2.4:** 0.8% agarose gel showing the amplified kanamycin region. Lane 1 – Kanamycin amplified at 70 °C from pgp564; Lane 2– Kanamycin amplified at 60 °C pgp564

The 1099 base pair band was purified and recombined with BamHI digested pex 100 T plasmid using CPEC to generate pexKan plasmid. The pexKan plasmid was transformed into E.coli DH5 α cells, selected on LB kanamycin media and the pexKan plasmid was extracted. The extracted plasmid was digested with the unique restriction enzyme SmaI.
Chapter 3: Metabolic Models of R. eutropha H16

3.1 Introduction

Stoichiometric modeling is based on the representation of the cell metabolism by use of metabolic network and pseudo steady state assumption. A stoichiometric model serves as a starting point for elementary mode analysis, metabolic flux analysis, flux spectrum and predictive modeling (Llaneras & Picó, 2008). Stoichiometric models describing the central metabolism of R. eutropha have been created based on the expression profiles obtained from the literature as described in the Chapter 2. The specifics of the lithoautotrophic and the heterotrophic models are described in the sections below.

3.2 Lithoautotrophic (LA) model

The presence of the enzymes of Calvin cycle during lithoautotrophy has been demonstrated by Friedrich et al. 1981 and Gottschalk et al., 1964. The LA model is characterized by the absence of the enzymes of Entner–Doudoroff pathway (Bowein & Schlegel, 1972). The Calvin cycle along with the pentose phosphate pathway is the main carbon utilizing pathway. Several studies have reported high expression levels of glucose-6-phosphate 1-dehydrogenase (Schlegel & Trüper, 1966; Gottschalk et al., 1964; Bowein & Schlegel, 1972), transketolase (Schäferjohann et al., 1993) and phosphoribulokinase (Abdelal & Schlegel, 1974). Molecular hydrogen induces the expression of Ribulose-1, 5-bisphosphate carboxylase oxygenase (IM & Friedrich, 1983). Energy is generated by the membrane bound and soluble hydrogenases which fix...
molecular hydrogen to Ubiquinol (UQH2) and NADH respectively (Schwartz et al., 2003). The soluble hydrogenases support rapid growth (Schink & Schlegel, 1978). An excess of NADH is produced during growth with 80 % H2 during lithoautotrophy. NAD is regenerated by 3 cytochrome oxidases, 3 quinol oxidases and cytochrome bd oxidases(a respiratory quinol) with oxygen as the terminal electron acceptor (Kohlmann et al., 2011) The membrane bound hydrogenases are oxygen tolerant and CO resistant (Pohlmann et al., 2006).

Apart from fructokinase, the glycolytic enzymes namely Glyceraldehyde 3 phosphate dehydrogenase (NAD utilizing enzyme), pyruvate kinase, phosphoglycerate kinase, fructose-1,6-/sedoheptulose-1,7-bisphosphatase,fructose-1,6-bisphosphate aldolase, phosphopyruvate hydratase (Enolase), triose phosphate isomerase, glucose-6-phosphate isomerase, phosphoglycerate mutase are also present (Friedrich et al 1980; Gottschalk et al., 1964; Schobert & Bowien, 1984; Schlegel & Trüper,1966; Kohlmann et al., 2011) in CO2 + H2+O2 grown Ralstonia eutropha. The TCA cycle is also fully functional in this growth model (Glaeser & Schlegel, 1972; Trüper, 1965; Brüland et al., 2010; Schwartz et al., 2009).

3.3 Heterotrophic model

Fructose is imported into the cell via ABC-type transporter (Gene locus: H16_B1498-1500), and catabolized via the Entner-Doudoroff (KDPG) pathway (Raberg et al., 2011). The calvin cycle enzymes are partially expressed but their activity is highly repressed (Gottschalk et al., 1964; Kuehn & McFadden, 1969; IM and Friedrich,1983; Friedrich et al., 1981; Bowien & Kussian,2002). Ribulose-1,5-
bisphosphate carboxylase oxygenase, is partially expressed but is inactivated up to 20% (Friedrich et al., 1981; Abdelal & Schlegel, 1974; Friedrich & Bowien, 1981). A recent study established the KDPG pathway as the sole means for carbon utilization during growth on fructose (Fleige et al., 2011). The study demonstrated the inability of the 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase negative mutants to grow on fructose. Presence of the fully functional Calvin cycle enzymes would permit growth in a KDPG aldolase negative strain which contradicts the experimental findings in the previous study. The reductive pentose phosphate cycle has been shown to be repressed (Abdelal & Schlegel, 1974). Since carbon dioxide fixation is an energy demanding process, repression of the calvin cycle enzymes during growth on fructose would make sense from an energetic standpoint (Bowien & Kussian, 2002). Partial depression of some enzymes of the pentose phosphate cycle like transketolase and phosphoribulokinase has been reported (Schäferjohann et al., 1993; Bowien et al., 1984; Abdelal & Schlegel, 1974a).

The enzymes of glycolysis/gluconeogenesis are shared by heterotrophic and autotrophic metabolism (Friedrich et al. 1980; Gottschalk et al., 1964; Schobert & Bowien, 1984; IM and Friedrich, 1983; Schlegel & Trüper, 1966; Schwartz et al., 2009). Like the lithoautotrophic model, TCA cycle is the energy yielding process (Glaeser & Schlegel, 1972; Trüper, 1965; Brüland et al., 2010).
A summary of the models are described below:

<table>
<thead>
<tr>
<th></th>
<th>CO₂ + H₂ + O₂</th>
<th>Fructose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of reactions</td>
<td>62</td>
<td>78</td>
</tr>
<tr>
<td>Reversible reactions</td>
<td>29</td>
<td>32</td>
</tr>
<tr>
<td>Irreversible reactions</td>
<td>33</td>
<td>46</td>
</tr>
<tr>
<td>External metabolites</td>
<td>11</td>
<td>13</td>
</tr>
<tr>
<td>Internal metabolites</td>
<td>59</td>
<td>66</td>
</tr>
</tbody>
</table>
Table 3.1: The table summarizes the list of reactions present in the Lithoautotrophic growth mode along with the corresponding enzyme name and positions on the genome.

<table>
<thead>
<tr>
<th>Reaction name</th>
<th>Description</th>
<th>Genes</th>
<th>Enzyme</th>
<th>Lithoautotrophy Reference</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEA</td>
<td>Malate → pyruvate</td>
<td>H16_A3153</td>
<td>malic Enzyme Ec 1.1.1.38</td>
<td>Truper,1965</td>
<td>low or no activity of NAD dependent enzyme</td>
</tr>
<tr>
<td>MEB</td>
<td>Malate → pyruvate</td>
<td>H16_A1002</td>
<td>malic Enzyme Ec 1.1.1.40 – nadh</td>
<td>Truper,1965 ; Schobert &amp; Bowien, 1984</td>
<td></td>
</tr>
<tr>
<td>PPCr</td>
<td>PEP = oxaloacetate</td>
<td>H16_A2921</td>
<td>Phosphoenolpyruvate carboxylase</td>
<td>Bruland et al., 2009</td>
<td>No expression of functional Ppc in R. eutropha H16</td>
</tr>
<tr>
<td>PCKr</td>
<td>Oxaloacetate → PEP</td>
<td>H16_A3711</td>
<td>phosphoenolpyruvate carboxykinase</td>
<td>Schobert &amp; Bowien, 1984</td>
<td></td>
</tr>
<tr>
<td>ICL/GLB1</td>
<td>Isocitrate → Glyoxylate + succinate</td>
<td>H16_A2211/H16_A2227</td>
<td>Isocitrate Lyase</td>
<td>Truper,1965; Schobert &amp; Bowien, 1984</td>
<td></td>
</tr>
<tr>
<td>MAL5/GLB2</td>
<td>Glyoxylate + acetylCoA → malate</td>
<td>H16_A2217</td>
<td>malate synthase</td>
<td>Truper,1965; Schobert &amp; Bowien, 1984</td>
<td></td>
</tr>
<tr>
<td>ALHD4</td>
<td>Butanal → 1-Butanol</td>
<td>H16_A0478/H16_A0679/ H16_A0861/H16_A0893/ H16_A1437/H16_A1562/ H16_A1702/H16_A1828/ H16_A2460/H16_A2474/ H16_A2586/H16_B0034/ H16_B0572/H16_B0663/ H16_B0713/H16_B0831/ H16_B0941/H16_B0942/ H16_B1417/H16_B2168/ H16_B2361/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHB9</td>
<td>3-Hydroxybutanoyl-CoA → PHB</td>
<td>H16_A1437/H16_A2003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHB4r</td>
<td>3-Hydroxybutanoyl-CoA = Acetoacetyl-CoA</td>
<td>H16_A1439/H16_A2002/ H16_A2171</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction name</td>
<td>Description</td>
<td>Genes</td>
<td>Enzyme</td>
<td>Lithoautotrophy</td>
<td>Reference</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------</td>
<td>-------</td>
<td>--------</td>
<td>-----------------</td>
<td>-----------</td>
</tr>
<tr>
<td>TCA7r</td>
<td>Malate --&gt; oxalaoacetate</td>
<td>H16_A2634/H16_B0334</td>
<td>Malate dehydrogenase (Nad)</td>
<td>Glaeser &amp; Schlegel, 1972; Truper, 1965</td>
<td>Expression of mdh is constitutive</td>
</tr>
<tr>
<td>TCA3</td>
<td>Isocitrate --&gt; 2-Oxoglutarate</td>
<td>H16_B1016</td>
<td>Isocitrate dehydrogenase - NAD</td>
<td>Truper, 1965</td>
<td>Low/none</td>
</tr>
<tr>
<td>TCA3</td>
<td>Isocitrate --&gt; 2-Oxoglutarate</td>
<td>H16_B1016</td>
<td>Isocitrate dehydrogenase - NADP</td>
<td>Truper, 1965</td>
<td></td>
</tr>
<tr>
<td>TCA6r</td>
<td>Fumarate = malate</td>
<td>H16_A2528/H16_B0103</td>
<td>Fumarase</td>
<td>Glaeser &amp; Schlegel, 1972; Truper, 1965</td>
<td>low but present</td>
</tr>
<tr>
<td>TCA2</td>
<td>Citrate --&gt; cis acotinate --&gt; Isocitrate</td>
<td>H16_A1907/H16_A2638/H16_B0568</td>
<td>Acotinase/aconitate hydratase</td>
<td>Glaeser &amp; Schlegel, 1972; Truper, 1965</td>
<td></td>
</tr>
<tr>
<td>TCA4br</td>
<td>Succinyl-CoA = succinate</td>
<td>H16_A0547&amp;H16_A0548</td>
<td>Succinyl coa kinase</td>
<td>Truper, 1965</td>
<td></td>
</tr>
<tr>
<td>TCA4a</td>
<td>2-Oxoglutarate --&gt; Succinyl-CoA</td>
<td>(H16_A2325)(H16_A1377/H16_A2323/H16_A3724/H16_B1098)(H16_A2324)</td>
<td>α ketoglutarate dehydrogenase</td>
<td>Glaeser &amp; Schlegel, 1972; Truper, 1965</td>
<td>Only NAD dependent enzyme is active</td>
</tr>
<tr>
<td>TCA1</td>
<td>Oxalaoacetate + Acetoacetyl-CoA --&gt; citrate</td>
<td>H16_A1229/H16_A2627/H16_B0357/H16_B0414/H16_B2211</td>
<td>Citrate synthase</td>
<td>Glaeser &amp; Schlegel, 1972; Truper, 1965</td>
<td></td>
</tr>
<tr>
<td>Reaction name</td>
<td>Description</td>
<td>Genes</td>
<td>Enzyme</td>
<td>Lithoautotrophy Reference</td>
<td>Comments</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------</td>
<td>--------------------------------</td>
<td>--------------------------------</td>
<td>---------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>TCA5r</td>
<td>Succinate $\rightarrow$ fumarate</td>
<td>H16_A2629&amp;H16_A2630&amp;H16_A2631&amp;H16_A2632&amp;H16_B0204</td>
<td>Succinate dehydrogenase</td>
<td>Glaeser &amp; Schlegel, 1972; Truper, 1965</td>
<td></td>
</tr>
<tr>
<td>LADH</td>
<td>pyruvate $\rightarrow$ Lactate</td>
<td>H16_A0666</td>
<td>lactaldehyde dehydrogenase</td>
<td>Kohlmann et al., 2011</td>
<td>No expression</td>
</tr>
<tr>
<td>GG6r</td>
<td>Glyceraldehyde 3-phosphate = 3-Phospho-D-glyceroyl phosphate</td>
<td>H16_A3146/H16_B1386</td>
<td>Glyceraldehyde 3 phosphate dehydrogenase(nad)</td>
<td>Friedrich et al 1980; Gottschalk et al., 1964</td>
<td></td>
</tr>
<tr>
<td>GG11</td>
<td>PEP $\rightarrow$ pyruvate</td>
<td>H16_A0567/H16_A3602/H16_B0961</td>
<td>pyruvate kinase</td>
<td>Gottschalk et al., 1964; Schobert &amp; Bowien, 1984</td>
<td></td>
</tr>
<tr>
<td>GG7r</td>
<td>3-Phospho-D-glycerate = 3-Phospho-D-glyceroyl phosphate</td>
<td>H16_A0566/H16_B1385/PHG417</td>
<td>phosphoglycerate kinase</td>
<td>Friedrich et al 1980; Gottschalk et al., 1964</td>
<td></td>
</tr>
<tr>
<td>FBP</td>
<td>Fructose 1,6-bisphosphate $\rightarrow$ Fructose 6-phosphate</td>
<td>H16_A0999/H16_B1390</td>
<td>fructose-1,6-/sedoheptulose-1,7-bisphosphatase</td>
<td>Friedrich et al 1980; Gottschalk et al., 1964</td>
<td></td>
</tr>
<tr>
<td>GG4r</td>
<td>Fructose 1,6-bisphosphate = Dihydroxyacetone phosphate + D-Glyceraldehyde 3-phosphate</td>
<td>H16_A0568/H16_B0278/H16_B1384</td>
<td>fructose-1,6-bisphosphate aldolase</td>
<td>Friedrich et al 1980; Gottschalk et al., 1964</td>
<td></td>
</tr>
<tr>
<td>GG9r</td>
<td>2-Phospho-D-glycerate = PEP</td>
<td>H16_A1188</td>
<td>phosphopyruvate hydratase(Enolase)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction name</td>
<td>Description</td>
<td>Genes</td>
<td>Enzyme</td>
<td>Lithoautotrophy Reference</td>
<td>Comments</td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------------</td>
<td>-------------------------------------</td>
<td>---------------------------------------------</td>
<td>-------------------------------------------------------------------------</td>
</tr>
<tr>
<td>GG5r</td>
<td>Dihydroxyacetone phosphate = D-Glyceraldehyde 3-phosphate</td>
<td>H16_A1047</td>
<td>triosephosphate isomerase</td>
<td>Gottschalk et al., 1964</td>
<td>Low when compared to heterotrophic growth mode</td>
</tr>
<tr>
<td>GG1r</td>
<td>Glucose 6-phosphate = Fructose 6-phosphate</td>
<td>H16_A1502/H16_B1502</td>
<td>glucose-6-phosphate isomerase</td>
<td>Gottschalk et al., 1964; Bowein &amp; Schlegel, 1972; Schlegel &amp; Truper, 1966</td>
<td>Indispensable for lithoautotrophic growth</td>
</tr>
<tr>
<td>GG8r</td>
<td>2-Phospho-D-glycerate = 3-Phospho-D-glycerate</td>
<td>H16_A0332/H16_A0493</td>
<td>phosphoglycerate mutase</td>
<td>Gottschalk et al., 1964; Reutz et al., 1982</td>
<td></td>
</tr>
<tr>
<td>SUCCD1</td>
<td>Succinate --&gt; fumarate</td>
<td>H16_A2629/H16_A2630/H16_A2631&amp;H16_A232&amp;H16_B0204</td>
<td>succinate dehydrogenase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NADPHOS</td>
<td>NADPH transhydrogenation</td>
<td>H16_A0143, H16_A1132</td>
<td>NAD + ATP = NADP + ADP + H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYTOOX</td>
<td></td>
<td>H16_A0342, H16_A0343 AND LOTS OF OTHER GENES</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPM4</td>
<td>NADH + Q --&gt; NAD</td>
<td>H16_A0251/H16_A1050/H16_A1051 /H16_A1052/H16_A1053/H16_A1054</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POX</td>
<td>Pyruvate --&gt; acetate</td>
<td>H16_A3123</td>
<td>cytochrome b561</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction name</td>
<td>Description</td>
<td>Genes</td>
<td>Enzyme</td>
<td>Lithoautotrophy Reference</td>
<td>Comments</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------</td>
<td>-------</td>
<td>--------</td>
<td>---------------------------</td>
<td>----------</td>
</tr>
<tr>
<td>CYTCOBDCYTCOBO3</td>
<td>QH₂ + O₂ → Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC1r</td>
<td>Nad/nadp transhydrogenation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUCCD2</td>
<td>FADH₂ → FAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPP1</td>
<td>Glucose 6-phosphate → 6-phospho-D-glucono-1,5-lactone</td>
<td>H16_A0316/H16_B1501/H16_B2566</td>
<td>glucose-6-phosphate 1-dehydrogenase</td>
<td>Schlegel &amp; Truper, 1966; Gottschalk et al., 1964; Bowein &amp; Schlegel, 1972</td>
<td>Low when compared to other modes</td>
</tr>
<tr>
<td>PPP5r</td>
<td>Ribose 5-phosphate + Xylose 5-phosphate = Glyceraldehyde 3-phosphate + Sedoheptulose 7-phosphate</td>
<td>H16_A3147/H16_B1388</td>
<td>transketolase</td>
<td>Schäferjohann et al., 1993</td>
<td>High levels are found</td>
</tr>
<tr>
<td>PPP7r</td>
<td>Xylose 5-phosphate + Erythrose 4-phosphate = Fructose 6-phosphate + D-Glyceraldehyde 3-phosphate</td>
<td>H16_A3147/H16_B1388</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entner Duroff enzymes</td>
<td></td>
<td></td>
<td></td>
<td>Bowein &amp; Schlegel, 1972</td>
<td>Absent</td>
</tr>
</tbody>
</table>
Table 3.2: The table summarizes the list of reactions present in the heterotrophic growth mode along with the corresponding enzyme name and positions on the genome.

<table>
<thead>
<tr>
<th>Reaction name</th>
<th>Description</th>
<th>Genes</th>
<th>Enzyme</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEA</td>
<td>Malate → pyruvate</td>
<td>H16_A1002</td>
<td>malic Enzyme</td>
<td>Low or no expression of Nad Dependent Enzyme</td>
</tr>
<tr>
<td>MBB</td>
<td>Malate → pyruvate</td>
<td>H16_A1002</td>
<td>malic Enzyme</td>
<td>Low or no expression of Nad Dependent Enzyme</td>
</tr>
<tr>
<td>PEP</td>
<td>PiP → oxaloacetate</td>
<td>H16_A2921</td>
<td>Phosphoenolpyruvate carboxylase</td>
<td>No expression of functional Ppc in R. eutropha H16</td>
</tr>
<tr>
<td>PCK</td>
<td>Oxaloacetate → PEP</td>
<td>H16_A3711</td>
<td>phosphoenolpyruvate carboxykinase</td>
<td>Phosphoenolpyruvate carboxykinase</td>
</tr>
<tr>
<td>ALDH4</td>
<td>Butanal → 1-Butanol</td>
<td>H16_A2217</td>
<td>Isoconate Lyase</td>
<td>Isoconate Lyase</td>
</tr>
<tr>
<td>MALSG, GLB2</td>
<td>Glyoxylate + acetyl CoA → malate</td>
<td>H16_A2217</td>
<td>Isoconate Lyase</td>
<td>Isoconate Lyase</td>
</tr>
<tr>
<td>ALDH4</td>
<td>3-Hydroxybutyryl-CoA → PHB</td>
<td>H16_A1457</td>
<td>malate synthase</td>
<td>Malate synthase</td>
</tr>
</tbody>
</table>

Heterotrophy reference
Truper, 1965
Bruland et al., 2009; Schobert & Bowien, 1984
<table>
<thead>
<tr>
<th>Reaction name</th>
<th>Description</th>
<th>Genes</th>
<th>Enzyme</th>
<th>Heterotrophy reference</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>PHB4r</td>
<td>3-Hydroxybutanoyl-CoA = Acetoacetyl-CoA</td>
<td>H16_A1439/H16_A2002/H16_A2171</td>
<td>Malate dehydrogenase (Nad)</td>
<td>Glaeser &amp; Schlegel, 1972; Truper, 1965; Bruland et al., 2009</td>
<td>Shows constitutive expression of mdh</td>
</tr>
<tr>
<td>TCA7r</td>
<td>Malate → oxaloacetate</td>
<td>H16_A2634/H16_B0334</td>
<td>Isocitrate dehydrogenase - NAD</td>
<td>Truper, 1965</td>
<td>Low or no expression of the NAD Dependent Enzyme</td>
</tr>
<tr>
<td>TCA3</td>
<td>Isocitrate → 2-Oxoglutarate</td>
<td>H16_B1016</td>
<td>Isocitrate dehydrogenase - NAD</td>
<td>Truper, 1965</td>
<td></td>
</tr>
<tr>
<td>TCA3</td>
<td>Isocitrate → 2-Oxoglutarate</td>
<td>H16_B1016</td>
<td>Isocitrate dehydrogenase - NAD</td>
<td>Truper, 1965</td>
<td></td>
</tr>
<tr>
<td>MDH2</td>
<td>Malate → oxaloacetate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CTLr</td>
<td>Citrate → oxaloacetate + acetate</td>
<td>H16_A2635/H16_B0353/H16_B0680/H16_B2113</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCA6r</td>
<td>Fumarate = malate</td>
<td>H16_A2528/H16_B0103</td>
<td>Fumarase</td>
<td>Glaeser &amp; Schlegel, 1972; Truper, 1965</td>
<td></td>
</tr>
<tr>
<td>TCA2</td>
<td>Citrate → cis aconitate - Insoluble</td>
<td>H16_A1907/H16_A2638/H16_B0568</td>
<td>Acotinase/aconitate hydratase</td>
<td>Glaeser &amp; Schlegel, 1972; Truper, 1965</td>
<td></td>
</tr>
<tr>
<td>Reaction name</td>
<td>Description</td>
<td>Genes</td>
<td>Enzyme</td>
<td>Heterotrophy reference</td>
<td>Comment</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------------------</td>
<td>-----------------------------------------------------------------------</td>
<td>-------------------------------</td>
<td>------------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>TCA4br</td>
<td>Succinyl-CoA = succinate</td>
<td>H16_A0547&amp;H16_A0548</td>
<td>Succinyl coa kinase</td>
<td>Truper,1965</td>
<td></td>
</tr>
<tr>
<td>TCA4a</td>
<td>2-Oxoglutarate→ Succinyl-CoA</td>
<td>(H16_A2325)(H16_A1377/H16_.A2323/H16_A3724/H16_B1098)(H16_A2324)</td>
<td>α ketoglutarate dehydrogenase</td>
<td>Glaeser &amp; Schlegel,1972; Truper,1965</td>
<td>Presence of a NAD dependent enzyme</td>
</tr>
<tr>
<td>TCA1</td>
<td>Oxaloacetate + Acetoacetyl-CoA → citrate</td>
<td>H16_A1229/H16_A2627/H16_B0 357/H16_B0414/H16_B2211</td>
<td>Citrate synthase</td>
<td>Glaeser &amp; Schlegel,1972; Truper,1965</td>
<td></td>
</tr>
<tr>
<td>TCA5r</td>
<td>Succinate → fumarate</td>
<td>4</td>
<td>Succinate dehydrogenase</td>
<td>Glaeser &amp; Schlegel,1972; Truper,1965</td>
<td></td>
</tr>
<tr>
<td>HEXf</td>
<td>Fructose → Fructose 6 phosphate</td>
<td>H16_B1503</td>
<td>fructokinase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LADH</td>
<td>pyruvate → Lactate</td>
<td>H16_A0666</td>
<td>Lactaldehyde dehydrogenase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG6r</td>
<td>Glyceraldehyde 3-phosphate = 3-Phospho-D-glyceroyl phosphate</td>
<td>H16_A3146/H16_B1386</td>
<td>Glyceraldhey 3 phosphate dehydrogenase(nad)</td>
<td>Friedrich et al 1980 ; Gottschalk et al., 1964</td>
<td></td>
</tr>
<tr>
<td>GG11</td>
<td>PEP → pyruvate</td>
<td>H16_A0567/H16_A3602/H16_B0 961</td>
<td>pyruvate kinase</td>
<td>Gottschalk et al., 1964 ; Schobert &amp; Bowien, 1984</td>
<td></td>
</tr>
<tr>
<td>Reaction name</td>
<td>Description</td>
<td>Genes</td>
<td>Enzyme</td>
<td>Heterotrophy reference</td>
<td>Comment</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------</td>
<td>-------</td>
<td>--------</td>
<td>------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>GG7r</td>
<td>3-Phospho-D-glycerate = 3-Phospho-D-glyceroyl phosphate</td>
<td>H16_A0566/H16_B1385/PHG417</td>
<td>phosphoglycerate kinase</td>
<td>Friedrich et al 1980; Gottschalk et al., 1964</td>
<td></td>
</tr>
<tr>
<td>FBP</td>
<td>Fructose 1,6-bisphosphate -- Fructose 6-phosphate</td>
<td>H16_A0999/H16_B1390</td>
<td>fructose-1,6-bisphosphate aldolase</td>
<td>Friedrich et al 1980; Gottschalk et al., 1964</td>
<td></td>
</tr>
<tr>
<td>GG4r</td>
<td>Fructose 1,6-bisphosphate = Dihydroxyacetone phosphate + D-Glyceraldehyde 3-phosphate</td>
<td>H16_A0568/H16_B0278/H16_B1384</td>
<td>fructose-1,6-bisphosphate aldolase</td>
<td>Friedrich et al 1980; Gottschalk et al., 1964</td>
<td></td>
</tr>
<tr>
<td>GG9r</td>
<td>2-Phospho-D-glycerate = PEP</td>
<td>H16_A1188</td>
<td>phosphopyruvate hydratase/Enolase</td>
<td>IM and Friedrich, 1983</td>
<td></td>
</tr>
<tr>
<td>GG5r</td>
<td>Dihydroxyacetone phosphate = D-Glyceraldehyde 3-phosphate</td>
<td>H16_A1047</td>
<td>triosephosphate isomerase</td>
<td>Gottschalk et al., 1964</td>
<td></td>
</tr>
<tr>
<td>GG1r</td>
<td>Glucose 6-phosphate = Fructose 6-phosphate</td>
<td>H16_A1502/H16_B1502</td>
<td>glucose-6-phosphate isomerase</td>
<td>Gottschalk et al., 1964; Schlegel &amp; truper, 1966</td>
<td></td>
</tr>
<tr>
<td>GG8r</td>
<td>2-Phospho-D-glycerate = 3-Phospho-D-glycerate</td>
<td>H16_A0332/H16_A0493</td>
<td>phosphoglycerate mutase</td>
<td>Gottschalk et al., 1964; Reutz et al., 1982</td>
<td>Dispensable</td>
</tr>
<tr>
<td>LACDHl</td>
<td>Lac --pyruvate</td>
<td>H16_B0460/H16_B1817</td>
<td>lactate cytochrome c reductase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction name</td>
<td>Description</td>
<td>Genes</td>
<td>Enzyme</td>
<td>Heterotrophy reference</td>
<td>Comment</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------</td>
<td>-------</td>
<td>--------</td>
<td>------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>FDH</td>
<td>Formate → CO₂</td>
<td>H16_A2629&amp;H16_A2630&amp;H16_A2631&amp;H16_A232&amp;H16_B0204</td>
<td>succinate dehydrogenase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUCCD1</td>
<td>Succinate → fumarate</td>
<td>H16_A0850&amp;H16_A0851&amp;H16_A0852&amp;H16_A1264&amp;H16_A1265&amp;H16_A1266&amp;H16_A3128&amp;H16_A3130&amp;H16_A3131&amp;H16_B1714&amp;H16_B1715</td>
<td>NAD(P) transhydrogenase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FC3</td>
<td>Nad /nadp transhydrogenation</td>
<td>H16_A0143 , H16_A1132</td>
<td>NAD + ATP =NADP +ADP + H</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NADPHOS</td>
<td>NADPH transhydrogenation</td>
<td>H16_A0342,H16_A0343 and lots of other genes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYTOOX</td>
<td></td>
<td>H16_A0251/H16_A1050/H16_A1051/H16_A1052/H16_A1053/H16_A1054/H16_A1055/H16_A1056/H16_A1057/H16_A1058/H16_A1059/H16_A1060/H16_A1061/H16_A1062/H16_A1063</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OPM4</td>
<td>NADH + Q → NAD</td>
<td>H16_A1062/H16_A1063</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>POX</td>
<td>Pyruvate → acetate</td>
<td>H16_A3123</td>
<td>cytochrome b₅₆₁</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYTCORD/CY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCOBO3</td>
<td>QH₂ + O₂ → Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction name</td>
<td>Description</td>
<td>Genes</td>
<td>Enzyme</td>
<td>Heterotrophy reference</td>
<td>Comment</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------</td>
<td>-------</td>
<td>--------</td>
<td>------------------------</td>
<td>---------</td>
</tr>
<tr>
<td>FC1r</td>
<td>Nad/nadp trans hydrogenation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUCCD2</td>
<td>FADH2 → FAD</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPP1</td>
<td>Glucose 6-phosphate → 6-phospho-D-glucono-1,5-lactone</td>
<td>H16_A0316/H16_B1501/H16_B2566</td>
<td>glucose-6-phosphate 1-dehydrogenase</td>
<td>Bowien et al, 1974; Schlegel &amp; Truper, 1966; Gottschalk et al., 1964</td>
<td>Present but inhibited when hydrogen is added to air</td>
</tr>
<tr>
<td>PPP5r</td>
<td>Ribose 5-phosphate + Xyulose 5-phosphate = Glyceraldehyde 3-phosphate + Sedoheptulose 7-phosphate</td>
<td>H16_A3147/H16_B1388</td>
<td>transketolase</td>
<td>Schäferjohann et al., 1993</td>
<td>partial depression of the gene expression</td>
</tr>
<tr>
<td>PPP7r</td>
<td>Xyulose 5-phosphate + Erythrose 4-phosphate = Fructose 6-phosphate + D-Glyceraldehyde 3-phosphate</td>
<td>H16_A3147/H16_B1388</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entner Duroff enzymes</td>
<td></td>
<td></td>
<td></td>
<td>IM and Friedrich, IM and Friedrich, 1983</td>
<td>Essential for growth on Fructose</td>
</tr>
</tbody>
</table>
3.4 Results of elementary mode analysis

3.4.1 Model Characterization & Comparison

The lithoautotrophic model generates 759 elementary modes. 27 % of these modes produce biomass. The most optimal biosynthetic mode has a yield of 0.01 g of biomass per g of carbon dioxide consumed. 199 (26.5%) elementary modes consume hydrogen and oxygen without the net generation of biomass or other any compounds. These modes were identified by setting Tra1 Tra2 Tra19 Tra20 Biomass Tra6r Tra15 to zero. 5 futile cycles were identified by setting the only ATP, ADP and water to be external metabolites. Futile cycles are indicators of allosteric regulations of the key enzymes involved (Qian & Beard, 2006). They are also biochemically relevant because they generate additional heat and warmth in the organism. (Dandekar & Sauerborn, 2002). Only one futile cycle which converts 3 phosphoglycerate to 1, 3 bisphosphoglycerate operates anaerobically.

During growth on fructose, 135074 elementary modes are predicted. Alternatively, 0.03 g of biomass could be obtained per g of fructose. In the absence of oxygen, no biomass can be generated due to the lack of electron acceptor. Fructose can however be metabolized anaerobically and converted to lactate, ethanol or isobutanol. 15551(11.5%) elementary modes completely oxidize fructose to CO2. Of the 15551 modes, 2544 mode generate ATP for cell maintenance. 19 futile cycles are present when fructose is used as the carbon source.

In addition to synthesis of biomass, cells expend energy at a constant rate for cell maintenance (Pirt, 1965; Du et al., 2001). Cell maintenance includes functions like
motility, maintaining internal ion /concentration gradients, futile cycles, turnover of macromolecular components like RNA, protein, and cell-wall polymers (Turner et al., 1989; Tempest and Neijssel, 1984). *R. eutropha* has been shown to excrete intermediary metabolites under conditions of maintenance energy insufficiency (Park & Lee, 1995). Lithoautotrophic model has 2 energy generating modes which do not produce biomass. Membrane bound hydrogenase generates 1 ATP while oxidation with soluble hydrogenase via NAD produces 2 ATP molecule for every hydrogen molecule consumed. With fructose, 4948 elementary modes can contribute to the maintenance energy. The most efficient modes produce 26 molecules of ATP for every fructose molecule consumed. Fructose is degraded via the KDPG pathway to give glyceraldehyde 3 phosphate and pyruvate, which are then completed oxidized to CO2. In all these modes, an extra ATP molecule is generated by converting pyruvate to acetate which combines with oxaloacetate to give citrate.
Fig 3.1: Metabolic Network map comprised of the pathways and metabolites for R. eutropha during lithoautotrophy. The corresponding reaction stoichiometries are listed in the appendix. Reversible reactions are indicated by double headed arrows. All the other reactions proceed in the designation of the arrows.
Fig 3.2: Metabolic Network map comprised of the pathways and metabolites for R. eutropha during heterotrophy with fructose as a substrate. The corresponding reaction stoichiometries are listed in the appendix. Reversible reactions are indicated by double headed arrows. All the other reactions proceed in the direction of the arrows.
**Table 3.3** Characterization of the lithoautotrophic and heterotrophic models. The table describes the number of viable elementary modes under various conditions. “Required Reactions” refers to the reactions that must occur for the elementary modes to be selected. “Reactions to be eliminated” refers to the reactions that must not occur for the elementary modes to be selected.

<table>
<thead>
<tr>
<th></th>
<th>Required Reactions</th>
<th>Reactions to be eliminated</th>
<th>CO$_2$ + H$_2$ + O$_2$</th>
<th>Fructose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total number of elementary modes</td>
<td></td>
<td></td>
<td>759</td>
<td>135074</td>
</tr>
<tr>
<td>Biomass producing mode</td>
<td>Biomass</td>
<td></td>
<td>210 (27.66%)</td>
<td>93327</td>
</tr>
<tr>
<td>Growth with maintenance energy</td>
<td>Biomass, ATPSYN</td>
<td></td>
<td>0</td>
<td>8001</td>
</tr>
<tr>
<td>Growth without maintenance energy</td>
<td>Biomass, Atpsyn</td>
<td></td>
<td>210 (27.66%)</td>
<td>85326</td>
</tr>
<tr>
<td>Anaerobic mode</td>
<td>Biomass</td>
<td>Tra13</td>
<td>1</td>
<td>18</td>
</tr>
<tr>
<td>Anaerobic growth</td>
<td>Biomass</td>
<td>Tra13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ethanol producing modes</td>
<td>Tra1</td>
<td></td>
<td>42</td>
<td>15396</td>
</tr>
<tr>
<td>Growth with Ethanol</td>
<td>Tra1, Biomass</td>
<td></td>
<td>6</td>
<td>8599</td>
</tr>
<tr>
<td>Isobutanol producing modes</td>
<td>Tra19</td>
<td></td>
<td>114</td>
<td>12317</td>
</tr>
<tr>
<td>Growth with Isobutanol</td>
<td>Tra19, Biomass</td>
<td></td>
<td>6</td>
<td>7319</td>
</tr>
<tr>
<td>PHB producing modes</td>
<td>Tra2</td>
<td></td>
<td>114</td>
<td>8717</td>
</tr>
<tr>
<td>Growth with PHB</td>
<td>Tra2, Biomass</td>
<td></td>
<td>6</td>
<td>5746</td>
</tr>
<tr>
<td>Lactate producing modes</td>
<td>Tra6</td>
<td></td>
<td>0</td>
<td>20078</td>
</tr>
<tr>
<td>Growth with Lactate</td>
<td>Tra6, Biomass</td>
<td></td>
<td>0</td>
<td>13204</td>
</tr>
</tbody>
</table>
3.4.2 Validation

The models can be validated by comparing model predictions with viability phenotypes observed in single knockout strains under various growth conditions. The lithoautotrophic model developed here confirms an earlier study (Reutz et al., 1982) where it has been demonstrated that phosphoglycerate mutase negative mutant cannot survive on CO$_2$ + H$_2$ + O$_2$. This can be attributed to the absence of the enzymes of the KDPG pathway (Bowein & Schlegel, 1972; IM & Friedrich, 1983). The other reactions of glycolysis such as those responsible for conversion of Fructose 1,6-phosphate to dihydroxyacetone phosphate and glyceraldehyde-3-phosphate (enzyme: fructose-bisphosphate aldolase), glyceraldehyde-3-phosphate to 1,3-bisphosphoglycerate (enzyme: glyceraldehyde-3-phosphate dehydrogenase (NADP$^+$) (phosphorylating)) and glyceraldehyde-3-phosphate to dihydroxyacetone phosphate (enzyme: triose-phosphate isomerase) have also been identified as essential reactions by the model. All of the above-named enzymes with the exception of triose phosphate isomerase have at least three isoenzymes and have been shown to be expressed in lithoautotrophic cultures (Schwartz et al., 2009). The study by Schwartz et al., 2009 predicts that the enzyme fructose bisphosphatase which catalyzes the FBP: Fructose 1, 6-bisphosphate $\rightarrow$ Fructose 6-phosphate conversion is essential during lithoautotrophy. Fructose bisphosphatase is encoded by three enzymes cbbFc and cbbFp being present in two highly homologous cbb operons located on the megaplasmid and on chromosome 2 and third enzyme FBP on chromosome 1 (Yoo & Bowien, 1997). The enzymes cbbFc and cbbFp are bifunctional: they also act as sedoheptulose-1, 7-
bisphosphatase. When the Fructose bisphosphatase activity is completely eliminated, the reaction GG13r: sedoheptulose 1, 7 bisphosphate $\rightarrow$ Sedoheptulose 7-phosphate is also affected because it is catalyzed by the bifunctional enzyme sedoheptulose-1, 7-bisphosphatase/ fructose bisphosphatase. Eliminating the FBP and GG13r reactions together produces a mutant, which is unable to grow lithoautotrophically. *R. eutropha H16* is capable of utilizing alternative electron acceptors like nitrate and nitrite for growth under anoxic conditions. In the absence of oxygen and nitrates, both the lithoautotrophic and the heterotrophic models predict no growth as expected (Schwartz *et al.*, 2003).

To our knowledge, single knockout strains which are lethal when grown on fructose (with the exception of the edp2 negative strain described in section 3.3) have not been created. The next section predicts some of the possible reactions that may be knocked out to test the validity of this model.
3.4.3 Additional knockout target genes for model validation:

Minimal cut set lists a set of reactions whose failure causes lethality. They are the set of smallest failure modes in a network that render the correct functioning of a cellular system impossible. (Klamt & Gilles, 2003). Using the CellNetAnalyzer we have determined the single gene knockouts which would prevent the formation of biomass. They are listed in the table 2.3.

**Table 3.4** All predicted single lethal deletion mutants of R. eutropha. The table describes the reactions that can be knocked out to test for viability under lithoautotrophic and heterotrophic growth conditions. Single knockout strains generated with the below reactions will be unable to produce biomass and hence will serve to validate the model. The enzymes for these reactions can be obtained from table 3.1 and table 3.2.

<table>
<thead>
<tr>
<th>CO2 + H2 + O2</th>
<th>Fructose</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCA3</td>
<td>✓</td>
</tr>
<tr>
<td>TCA2</td>
<td>✓</td>
</tr>
<tr>
<td>TCA1</td>
<td>✓</td>
</tr>
<tr>
<td>HEXf</td>
<td></td>
</tr>
<tr>
<td>GG7r</td>
<td>✓</td>
</tr>
<tr>
<td>GG1r</td>
<td></td>
</tr>
<tr>
<td>GG9r</td>
<td>✓</td>
</tr>
<tr>
<td>GG5r</td>
<td>✓</td>
</tr>
<tr>
<td>GG1r</td>
<td>✓</td>
</tr>
<tr>
<td>GG8r</td>
<td>✓</td>
</tr>
<tr>
<td>C TCO</td>
<td>✓</td>
</tr>
<tr>
<td>Gene</td>
<td>CO2 + H2 + O2</td>
</tr>
<tr>
<td>----------</td>
<td>---------------</td>
</tr>
<tr>
<td>PPP5r</td>
<td>✓</td>
</tr>
<tr>
<td>PPP7r</td>
<td>✓</td>
</tr>
<tr>
<td>PPP3r</td>
<td>✓</td>
</tr>
<tr>
<td>PPP4r</td>
<td>✓</td>
</tr>
<tr>
<td>GG12</td>
<td>✓</td>
</tr>
<tr>
<td>cal1</td>
<td>✓</td>
</tr>
<tr>
<td>Glut1r</td>
<td>✓</td>
</tr>
<tr>
<td>GLUM1r</td>
<td>✓</td>
</tr>
<tr>
<td>Biomass</td>
<td>✓</td>
</tr>
<tr>
<td>TRA13</td>
<td>✓</td>
</tr>
<tr>
<td>TRA15</td>
<td>✓</td>
</tr>
<tr>
<td>TRA6r</td>
<td>✓</td>
</tr>
<tr>
<td>TRA5r</td>
<td>✓</td>
</tr>
<tr>
<td>SH</td>
<td></td>
</tr>
<tr>
<td>PPP1</td>
<td></td>
</tr>
<tr>
<td>PGL</td>
<td></td>
</tr>
<tr>
<td>EDP2</td>
<td></td>
</tr>
<tr>
<td>EDP1</td>
<td></td>
</tr>
<tr>
<td>TRA11</td>
<td></td>
</tr>
<tr>
<td>TRA6</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 4: Automation of rational strain design

And God said; Let us make man in our image, after our likeness: and let them have dominion over the fish of the sea, and over the fowl of the air, and over the cattle, and over all the earth, and over every creeping thing that creepeth upon the earth. Genesis 1: 26

4.1 Introduction

Elementary modes are unique, non-decomposable, minimal sets of reactions within the metabolic network that operate under the pseudo steady-state constraint. They allow for the identification of pathway efficiency and the prediction of the effect of pathway elimination on cellular metabolism. The molar yield information is used to design an efficient strain by identifying target reactions to be eliminated for improved cell performance.

Trinh et al., 2008, have demonstrated strain optimization by elementary mode analysis. The optimization was achieved via multiple rounds of reaction deletion. The target for knockout was identified based on the following 3 rules:

a. The effect of elimination of individual reactions on the number of EMs was evaluated

b. The maximum yields of the desired product were determined

c. Reaction with the smallest fraction of remaining elementary modes that still support maximum yields of desired products and biomass was selected.

The figure below obtained from Trinh et al., 2009 summaries the concept. The objective is to increase the conversion of A to P. The molar yield of P can be determined from the set of existing elementary modes consisting of 6 elementary modes. The strain could be optimized by eliminating all the elementary modes with no
net production of D. Based on the analysis depicted in Fig 3.1a, where the number of remaining elementary modes are sorted in an increasing order, Reaction R2 is selected for the first round of deletion. On progressing iteratively, we can also eliminate reactions R3 and R8r.

![Diagram of reaction deletion](image)

**Fig 4.1** Implementation of multiple reaction deletion to rationally design the most efficient pathways for converting the substrate A into the product P. a–d Evaluation of sequential reaction deletion. E Summary of the effect of multiple reaction deletion on the total number of elementary modes as well as the yield range of the desired product P. f Alternative sets of reaction deletion to achieve the most efficient pathways to convert the substrate A into the product P. The figure has been taken from Trinh et al. (2009). Please refer the text for additional details.

In a living cell, eliminating these reactions would force the cell to utilize only the remaining elementary modes. Such a cell will have the most efficient conversion of
A to product P. It would desirable to incorporate multiple constraints like coupling growth and product synthesis or retaining biomass synthesis modes while optimizing for product production.

In a larger network, with many reactions the sequential reaction deletion strategy would be tedious and monotonous. Automating the approach described above would drastically bring down the development time and would increase the human productivity.

### 4.2 Problem statement

Given a set of elementary modes along with the substrate, production and biomass reactions, identify minimum of target knockouts which would

(i) Retain at least one elementary mode with maximum product yield

(ii) Retain at least one biomass producing elementary mode. Ideally we seek to retain biomass producing mode which is coupled to product production.

(iii) Eliminate elementary modes which do not produce product of interest

### 4.3 Algorithm and software development

Initially a macro was designed in Visual Basic. However, the system tends to crash when the number of reactions and the number of elementary modes are huge. Hence effective buffer utilization strategy had to be formulated. C++ was used to code the software.
The program has been written to handle 5 different optimizations. Each of these optimizations differs in the types of constraints used:

1 - The KO will be found retaining maximum product and bio mass yield with respect to one substrate

2 - The KO will be found retaining maximum product, bio mass and product/biomass yield with respect to 1 substrate

3 - The KO will be found retaining maximum product and bio mass yield with respect to 2 substrates

4 - The KO will be found retaining maximum product, bio mass and product/biomass yield

5 - The KO will be found retaining maximum product yield

6 - to run all the of the above

Based on the user input (integer), the script will check for the presence of necessary files and will compile the source code.

- It will execute the C++ executable
- It will keep a track of all actions in a log file
- It will move the log files to log folder. It will create a log folder if absent
- The list of KO’s will be moved to results folders. It will create a results folder if absent
- The input file will be moved to input folders. It will create a input folder if absent
- The Effects of KO files (eg. KO_2subs_max_prod_bio1.txt) will be moved to KO_effects folder. If the folder is absent it will be created
• The executable will be deleted

4.3.1 System & Environment

CGYWIN was used to stimulate Linux-like environment for Windows and the GCC compiler was used. A batch file was written in MS DOS to provide a user-friendly interface.

4.3.2 Requirements and software description

The code has been developed for a Windows based PC with gcc compilers. The software is to be used in conjunction with other software like CellNetAnalyzer (Steffen et al., 2007) or Metatool (Pfeiffer et al., 1999) which enumerate the elementary modes. A wrapper utility has been designed in DOS to increase the user friendliness. The elementary modes can be saved in a matrix form of tab delimited text files and are used as input for the software designed.

Apart the text file containing the elementary modes, the application requires the number of total reactions required to generate the file, the names of the substrate, product and biomass producing reactions as input. The application attempts to couple biomass production to product synthesis. If it is unable to do so, it retains maximum product producing mode along with atleast one biomass producing mode. The output includes a text file under the results folder which contains a list of the knockouts that might be implemented to obtain a strain which can effectively convert the substrate to product. The results file contains information about the minimum and the maximum yield of the substrate and the product after each reaction is knocked out. Many a times,
multiple reactions can have the same effect on the number of remaining elementary modes. The software arranges these reactions in alphabetical order and picks out the first reaction. One reaction may be preferred over the other based on the presence of isozymes or the ease of knocking the reaction. The alternate reactions and the effect of the reactions on each elementary mode set can be determined from files generated in an ancillary folder labeled KO.

4.4 Results

The lithoautotrophic model developed earlier was used as the input file and the program was run to optimize *R. eutropha* for the production of ethanol, isobutanol and PHB. The results of the program along with manually curated gene details are shown below. After the reactions are knocked out, 3 elementary modes remain. One of the elementary modes produces the product at the maximum yield, the second elementary mode is responsible for biomass coupled product production and the last elementary mode is involved with ATP hydrolysis.
4.4.1 Ethanol Production

1. PPP6R (H16_A2346)
2. GG11 (H16_A0567/H16_A3602/H16_B0961)
3. CAL3 (H16_A3146/H16_B1386)
4. TCA6r (H16_A2528/H16_B0103)
5. TRA20 (succinate excretion)
6. ADHB (H16_A1439/H16_A2002/H16_A2171)
7. GG10 (H16_A2038)
8. PHB8 (H16_A1437/H16_A2003)

4.4.2 Isobutanol Production

1. PPP6R (H16_A2346)
2. GG11 (H16_A0567/H16_A3602/H16_B0961)
3. CAL3 (H16_A3146/H16_B1386)
4. TCA6r (H16_A2528/H16_B0103)
5. TRA20
6. ALCDETR
   (H16_A0757/H16_A3330/H16_B0517/H16_B1433/H16_B1699/H16_B1745/H16_B1834/H16_B2470)
7. GG10 (H16_A2038)
8. PHB8 (H16_A1437/H16_A2003)
4.4.3 PHB Production

1. PPP6R (H16_A2346)
2. GG11 (H16_A0567/H16_A3602/H16_B0961)
3. CAL3 (H16_A3146/H16_B1386)
4. TCA6r (H16_A2528/H16_B0103)
5. TRA20 (succinate excretion)
6. ADHB (Knock one out : H16_A1439/H16_A2002/H16_A2171)
7. GG10 (H16_A2038)
8. ALCDETR
   (H16_A0757/H16_A3330/H16_B0517/H16_B1433/H16_B1699/H16_B1745/H16_B1834/H16_B2470)

4.4.4 Predicted Molar Yields

The program also determines theoretical molar yields expected for the products. The expected yields after knocking out the reactions for each of the three products been documented in the table.
Table 4.1  Yields of ethanol, isobutanol or PHB optimized R. eutropha H16 strain as predicted by the software. The table shows the expected yield of products and biomass after the genes described in the text have been knocked out from R. eutropha H16 strain.

<table>
<thead>
<tr>
<th>Product</th>
<th>Biomass Yield (in gms) with respect to CO₂</th>
<th>Product Yield (in moles) with respect to CO₂ when Biomass is produced</th>
<th>Product</th>
<th>Biomass Yield (in gms) with respect to H₂</th>
<th>Product</th>
<th>Biomass Yield (in gms) with respect to H₂ when Biomass is produced</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethanol</td>
<td>0.5</td>
<td>0.00472596</td>
<td>0.0758099</td>
<td>0.105263</td>
<td>0.00116801</td>
<td>0.0182548</td>
</tr>
<tr>
<td>Isobutanol</td>
<td>0.25</td>
<td>0.00472596</td>
<td>0.065829</td>
<td>0.0526316</td>
<td>0.00116801</td>
<td>0.0155569</td>
</tr>
<tr>
<td>PHB</td>
<td>0.25</td>
<td>0.00472596</td>
<td>0.065829</td>
<td>0.0625</td>
<td>0.00116801</td>
<td>0.0163185</td>
</tr>
</tbody>
</table>

4.4.5 Identification of efficient secondary metabolite producing modes:

Similar yields of the product (with respect to one substrate) can be produced by multiple elementary modes which have different energetic consequences. For instance, maximum theoretical yield of isobutanol i.e. 0.25 moles of isobutanol per moles of fructose can be attained via 12 different elementary mode families. Two elementary modes with the yield of 0.25 for isobutanol are shown in the table 4.2. The model developed here can help identify the most efficient families.
Table 4.2: Comparison of two different elementary mode families with similar isobutanol yield per mole of CO$_2$.

**Family 1**

$$0.88 \text{ O}_2 + 1 \text{ CO}_2 + 4.75 \text{ H}_2 \rightarrow \text{0.25 ISOBUTANOL}$$

**Family 2**

$$2.63 \text{ O}_2 + 1 \text{ CO}_2 + 8.25 \text{ H}_2 \rightarrow \text{0.25 ISOBUTANOL}$$
The most efficient stoichiometries with respect to the carbon source for various products have been documented below

**Isobutanol production:**

Lithoautotrophy: \(0.88 \text{ O}_2 + 1 \text{ CO}_2 + 4.75 \text{ H}_2 \rightarrow 0.25 \text{ ISOBUTANOL}\)

Heterotrophy: 1 FRUCTOSE \(\rightarrow 1 \text{ ISOBUTANOL} + 2 \text{ CO}_2\)

As shown by (Li et al., 2012), production of isobutanol during lithoautotrophy requires 19.011 moles of H2 per mole of isobutanol.

**Ethanol production:**

Lithoautotrophy: \(0.88 \text{ O}_2 + 1 \text{ CO}_2 + 4.75 \text{ H}_2 \rightarrow 0.5 \text{ ETOH}\)

Heterotrophy: 1 FRUCTOSE \(\rightarrow 2 \text{ ETOH} + 2 \text{ CO}_2\)

**PHB production:**

Lithoautotrophy: \(0.88 \text{ O}_2 + 1 \text{ CO}_2 + 4 \text{ H}_2 \rightarrow 0.25 \text{ PHB}\)

Heterotrophy: 1.5 \(\text{ O}_2 + 1 \text{ FRUCTOSE} \rightarrow 1 \text{ PHB} + 2 \text{ CO}_2 + 6 \text{ ATP\_BASE}\)

**Lactate production:**

Heterotrophy: 1 FRUCTOSE \(\rightarrow 2 \text{ LACTATE}\)
**Biomass production:**

Lithoautotrophy: \(0.86 \text{ O}_2 + 0.29 \text{ NH}_3 + 1 \text{ CO}_2 + 4.05 \text{ H}_2 \rightarrow 0.01 \text{ BIOMASS}\)

Heterotrophy: \(0.49 \text{ O}_2 + 1.26 \text{ NH}_3 + 1 \text{ FRUCTOSE} \rightarrow 1.22 \text{ CO}_2 + 0.19 \text{ ACETATE} + 0.03 \text{ BIOMASS}\)

(Or)

\(0.55 \text{ O}_2 + 1.33 \text{ NH}_3 + 1 \text{ FRUCTOSE} \rightarrow 1.32 \text{ CO}_2 + 0.03 \text{ BIOMASS}\)

Several families give an \(\text{Y}_{bio}/\text{fructose}\) of 0.03. The choice of the best family depends on the desired product. Families which co-produce a secondary product give better yield when compared to families that produce biomass only.

The Stoichiometry for production of secondary metabolites associated with biomass production during growth on fructose or \(\text{CO}_2 + \text{H}_2 + \text{O}_2\) is listed in the table below:
Table 4.3: Stoichiometry for production of secondary metabolites associated with biomass production during growth on fructose or CO₂ + H₂ + O₂

<table>
<thead>
<tr>
<th>Secondary metabolite</th>
<th>CO₂ + H₂ + O₂</th>
<th>Fructose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ethanol + maximum Biomass</td>
<td>0.87 O₂ + 0.25 NH₃ + 1 CO₂ + 4.16 H₂ ----&gt; 0.08 ETOH + 0.01 BIOMASS</td>
<td>0.65 O₂ + 1.21 NH₃ + 1 FRUCTOSE ----&gt; 0.14 ETOH + 1.49 CO₂ + 0.03 BIOMASS</td>
</tr>
<tr>
<td>Isobutanol + maximum Biomass</td>
<td>0.87 O₂ + 0.21 NH₃ + 1 CO₂ + 4.24 H₂ ----&gt; 0.07 ISOBUTANOL + 0.01 BIOMASS</td>
<td>0.8 O₂ + 1.21 NH₃ + 1 FRUCTOSE ----&gt; 0.05 ISOBUTANOL + 1.59 CO₂ + 0.03 BIOMASS</td>
</tr>
<tr>
<td>PHB + maximum Biomass</td>
<td>0.87 O₂ + 0.21 NH₃ + 1 CO₂ + 4.04 H₂ ----&gt; 0.07 PHB + 0.01 BIOMASS</td>
<td>0.86 O₂ + 1.21 NH₃ + 1 FRUCTOSE ----&gt; 0.05 PHB + 1.58 CO₂ + 0.03 BIOMASS</td>
</tr>
<tr>
<td>Lactate + maximum Biomass</td>
<td>-</td>
<td>0.65 O₂ + 1.21 NH₃ + 1 FRUCTOSE ----&gt; 1.35 CO₂ + 0.14 LACTATE + 0.03 BIOMASS</td>
</tr>
</tbody>
</table>
4.5 Discussion

*R. eutropha* produces several secondary like lactate, acetate, formate, isobutanol and ethanol among many others. Three metabolites considered for the optimizations are isobutanol, ethanol and PHB. The optimizations were done to retain the maximum yield of the product with respect to Carbon dioxide and hydrogen. The program predicts eight candidate reactions for optimizations. Implementing these knockouts would generate a strain, which is highly efficient in the production of one of these metabolites. Seven knockouts are common to the optimization of *R. eutropha* for the production of these three metabolites. TRA20, the succinate excretion reaction is one of the suggested knockouts. This reaction is carried out by H16_A0693 (gltp2), a generic dicarboxylate transporter. Eliminating the reaction might have unintended consequences. Knocking out the reactions GLB2 and TCA4A would achieve the same result as knocking out the TRA20 reaction.

Eight elementary modes remain after implementing the above knockouts. These include 2 product only producing modes, 2 biomass only producing modes, 2 product and biomass producing modes and 2 modes generating ATP via the hydrogenases. The ATP generating modes are modes which take in hydrogen via membrane bound hydrogenase (reaction: HYDQ) or soluble hydrogenase (reaction: SH). The Hydrogen is converted to NADH and then finally to ATP. Elimination of these modes requires elimination of hydrogenases, which will destroy biomass production. In addition, these modes are essential because they are the only modes, which contribute towards energy for cell maintenance.
We have two biomass producing modes which do not produce the products of interest. No further knockouts can eliminate these non-ethanol/non-isobutanol/non-PHB producing elementary modes without affecting the maximum product yield and non-lethality constraints. However, using nitrogen as the limiting factor during product generation stage silences these elementary modes. This model presents a case where product production cannot be coupled to biomass production.

4.6 Future perspectives

The running time and the memory usage can be minimized by determining the linear reactions and combining them to form a single vertex. The depth first search algorithm can be used in linear time to determine the SCC with Tarjan's algorithm (Tarjan, 1971).
Chapter 5: Estimation of elementary mode’s contributions to the overall metabolism from reaction entropies

5.1 Definition of the scientific problem

With the developing of high throughput sequencing technologies, vast number of genomes has been sequenced and the functions of many genes have been annotated. The next step would be to predict the metabolic flux in an organism and develop strategies rational strain development.

A study by Wlaschin et al., 2006 attempts to predict the metabolic flux in the organism using the stoichiometry of the participating reactions and the entropy of the metabolites. The technique presented in this study is based on Elementary mode analysis which can predict the metabolic flux vector without requiring any knowledge of any fixed flux rates or imposing any objective function for cellular metabolism (Trinh et al., 2009). The flux through any system has been so far determined by metabolic flux analysis (MFA). MFA is dependent on having enough measurable metabolite fluxes to solve \( \mathbf{R} = \alpha \mathbf{M} \) where \( \mathbf{R} \) is a column vector of net metabolic fluxes, \( \alpha \) is the stoichiometric matrix of all reactions and metabolites in the network and \( \mathbf{M} \) is the column velocity vector comprised of fluxes through each reaction. The system is often underdetermined and requires an objective function. Determining the most probable objective function for any cell requires the appropriate experimental data along with reaction set. There may not exist an universal objective function and several studies have been conducted to determine the best match (Andrea L. et al, 2006). The authors in this paper attempted to eliminate the need for an objective function by resorting to alternative techniques using EMA.
In contrast to MFA, in EMA the overall fluxes through metabolic pathways can be described as weighted linear combinations of reactions.

5.2 Significance

This study by Wlaschin et al., 2006, provides the first experimental evidence that metabolic fluxes can be estimated with the thermodynamic properties of the elementary modes. The metabolism of a cell can be viewed as a weighted sum of the elementary modes (Wlaschin et al., 2006). Based on the results of this study, Srienc and Unrean in 2010 demonstrated that the usage probabilities of individual elementary modes are distributed according to Boltzmann’s distribution law. Evolution of metabolic networks appears to be biased towards entropy maximization for the whole system using preferentially efficient pathways with small reaction entropies. The weights (or usage probability) of the individual elementary modes can determined using the equation

\[ w_i = \exp\left(-\frac{\Delta S_i}{b}\right) \]

where \( S \) is the entropy generated by the elementary modes and \( b \) is a constant obtained by solving the nonlinear equation

\[ F = \sum \exp\left(-\frac{\Delta S_i}{b}\right) \text{ for } i = 1 \text{ to total number of elementary modes}; \]

\[ F = 1 \]

And \( F = 1 \)
5.3 Methods

5.3.1 Reconstruction of the input file:

- Construct the model
  - The metabolic model of *E. coli* which was published by Carlson and Srienc, 2004 along with the modifications described by Carlson et al., 2005 has been used. The model was updated to include the production and secretion of 3 Hydroxy butanoate.
- Run the elementary mode analysis using the Metatool 5.0
  - The model produced 4374 elementary modes
- Simplify the model by constraints
  - Consider anaerobic, production (absence of biomass synthesis), no lactate and no succinate producing state. The authors found 40 elementary modes
- Identify unique families of elementary modes to reduce the number of unknowns
  - The 40 Elementary modes belong to 5 families
- Solve equation 1 and 2 to find the value of b using newton raphson method.
  - The value of b was found to be variable as described in the results section below.

5.3.2 Implementation

The model was reconstructed as described in the paper and the elementary modes were computed using the CellNetAnalyzer software. Software was developed in C++ to automate the process. This software will help automate the process and thereby enable the user to handle data of higher complexity.
5.3.3 Determination of flux from entropy

The code takes in the elementary mode matrix (obtained from CellNetAnalyzer) and list of transport reactions along with Delta G and Delta H values in the following format. The reactions and constraints used to obtain the elementary modes can be found at the end of the document.

Table 5.1: Description of the input file for determination of knockouts using entropy: The Delta G and Delta H values can be determined as described by Sandler et al., 1991 at standard conditions. For testing purposes, these values from Wlaschin's thesis were used. The energy values are in terms of moles of metabolites. The definition of biomass used in the table below has been derived from Roels, 1983. Please note that the first line describing the columns has been given here for illustration purposes. It must not be included in the actual input file.

<table>
<thead>
<tr>
<th>Reaction Name</th>
<th>Metabolite</th>
<th>Reactant(0)/Product</th>
<th>Delta G (kJ/mol)</th>
<th>Delta H (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>R1</td>
<td>GLU_ext</td>
<td>0</td>
<td>2872</td>
<td>2807</td>
</tr>
<tr>
<td>R90r</td>
<td>ETOH_ext</td>
<td>1</td>
<td>1319</td>
<td>1369</td>
</tr>
<tr>
<td>R91r</td>
<td>ACETATE_ext</td>
<td>1</td>
<td>894</td>
<td>876</td>
</tr>
<tr>
<td>R96</td>
<td>FORMATE_ext</td>
<td>1</td>
<td>140.5</td>
<td>127.5</td>
</tr>
<tr>
<td>R95</td>
<td>SUCC_ext</td>
<td>1</td>
<td>1550.1</td>
<td>1526.6</td>
</tr>
<tr>
<td>R94</td>
<td>LACTATE_ext</td>
<td>1</td>
<td>1322.6</td>
<td>1308.4</td>
</tr>
<tr>
<td>R99</td>
<td>3HB_ext</td>
<td>1</td>
<td>1984.4</td>
<td>1971.4</td>
</tr>
<tr>
<td>R98</td>
<td>PHB_ext</td>
<td>1</td>
<td>2078.5</td>
<td>2077.3</td>
</tr>
<tr>
<td>r70</td>
<td>BIOMASS</td>
<td>1</td>
<td>529.1</td>
<td>523.4</td>
</tr>
</tbody>
</table>
5.4 Pseudocode

5.4.1 Function: Flux analysis (matrix)

- Read the input from CellNetAnalyzer into two different data structures. Have one datastructure to store the reaction names. Use the other one to store the stoichiometric details
- Read the transport reactions file from the user
- Using both files generate a new matrix with only the external reactions and generate the stoichiometry after appropriate normalization
- Use the sort and unique function in C++ to identify the families.
- Use the delta G and delta H value to find out the delta S

- To find the value of the constant b use the following strategies to fine tune the value.

For (i=0; i< 1000; i++)

- Determine the value of b using 0.01 as the seed value. Call the newton_raphson function described later in this document.
- If newton_raphson returns a difference of > 0.01 add 0.01 to the seed value and call newton_raphson
- Else if newton_raphson returns a difference of > 0.001 add 0.001 to the seed value and call newton_raphson
Else If newton_raphson returns a difference of > 0.0001 add 0.0001 to the seed value if the difference in the previous iteration is less than the current else subtract 0.0001 and call newton_raphson

- Else If newton_raphson returns a difference of 0 return the value of b
- Based on the value of b determine the flux values using the formula for f 1 and print it.

5.4.2 Function: newton_raphson ()

- for (itercount=0;itercount<=1000;++itercount)
  - find the value of f over all the families using f1 = f1 + (exp(-(S)/b));
  - if F is between 0.995 and 1.05 return the value of b and set difference to zero
  - Else f1 = f1-1;
  - Find the value of f2 as f2 = f2 + (exp(-(S)/b) * (1/b));
  - Determine b1 as b1 = b - (f1/f2);
  - Find the difference between b and b1
  - If the difference is less than tolerance limit return the value of b and difference.

5.5 Results & Discussion

The software was tested with the input file used by Wlaschin et al., 2006. The software assigned the elementary modes to the 7 different families. Wlaschin et al., 2006 predicted 5 families. The difference lies in the precision used to estimate the stoichiometry. The 7 families could be reduced to 5, if a precision of 2 was used. The
entropy values correspond to the entropy values reported in the study. The value of the constant predicted by the software differs from the values predicted by the paper. This could be attributed to the fact that the authors have used experimental data to arrive at the value of "b". A plot of the $\Delta S$ Vs the weighting factor is shown in figure 5.1. The table 5.2 compares the result generated by the software with those described by Wlaschin et al, 2006.

The success of the software depends on the correctness of the input file. Incorrect value of the Gibbs free energy or enthalpy in the input file would jeopardize the value of the constant ‘b’ calculated by the program. When no solution exists, the program terminates with an error message. The advantage of the software lies in the magnitude of data that can be supplied processed.

However, there room for further improvement, if some machine learning algorithm could be used to find the seed (initial) value of "b".
Table 5.2 2 compares the result generated by the software with those described by Wlaschin et al., 2006.

<table>
<thead>
<tr>
<th>Family</th>
<th>Stoichiometry (current study)</th>
<th>Stoichiometry (Wlaschin et al., 2006)</th>
<th>Entropy (current study) kJ/K</th>
<th>Entropy (Wlaschin et al., 2006) kJ/K</th>
<th>Flux through each family (current study) Sum = 1.047</th>
<th>Flux through each family (Wlaschin et al., 2006) Sum= flux</th>
<th>Value of B according to Wlaschin et al., 2006 Exp (- S/b)= flux</th>
<th>Actual Fluxes according to Wlaschin et al., 2006</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family 1</td>
<td>1 GLU_EXT ----&gt; 1 ETOH_EXT + 1 ACETATE_EXT + 2 FORMATE_EXT</td>
<td>1 GLU_EXT ----&gt; 1 ETOH_EXT + 1 ACETATE_EXT + 2 FORMATE_EXT</td>
<td>0.24</td>
<td>0.23</td>
<td>0.18</td>
<td>0.281</td>
<td>0.184339</td>
<td>0.547</td>
</tr>
<tr>
<td>Family 2</td>
<td>1 GLU_EXT ----&gt; 1.71429 ETOH_EXT + 1.71429 FORMATE_EXT</td>
<td>1 GLU_EXT ----&gt; 1.71 ETOH_EXT + 1.71 FORMATE_EXT</td>
<td>0.43</td>
<td>0.42</td>
<td>0.04</td>
<td>0.075</td>
<td>0.163304</td>
<td>0.069</td>
</tr>
<tr>
<td>Family 3</td>
<td>1 GLU_EXT ----&gt; 2 ETOH_EXT</td>
<td>1 GLU_EXT ----&gt; 2 ETOH_EXT</td>
<td>0.55</td>
<td>0.54</td>
<td>0.02</td>
<td>0.036</td>
<td>0.163647</td>
<td>0.046</td>
</tr>
<tr>
<td>Family 4a</td>
<td>1 GLU_EXT ----&gt; 0.666667 ETOH_EXT + 2 FORMATE_EXT + 0.666667 PHB_EXT</td>
<td>1 GLU_EXT ----&gt; 0.67 ETOH_EXT + 2 FORMATE_EXT + 0.67 PHB_EXT</td>
<td>0.24</td>
<td>0.24</td>
<td>0.18</td>
<td>0.277</td>
<td>0.1838</td>
<td>0.148</td>
</tr>
<tr>
<td>Family 4b</td>
<td>1 GLU_EXT ----&gt; 0.66665 ETOH_EXT + 2 FORMATE_EXT + 0.66665 PHB_EXT</td>
<td></td>
<td>0.24</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Family 5a</td>
<td>1 GLU_EXT ----&gt; 0.666667 ETOH_EXT + 2 FORMATE_EXT + 0.666667 3HB_EXT</td>
<td>1 GLU_EXT ----&gt; 0.67 ETOH_EXT + 2 FORMATE_EXT + 0.67 3HB_EXT</td>
<td>0.21</td>
<td>0.21</td>
<td>0.22</td>
<td>0.33</td>
<td>0.189417</td>
<td>0.148</td>
</tr>
<tr>
<td>Family 5b</td>
<td>1 GLU_EXT ----&gt; 0.66665 ETOH_EXT + 2 FORMATE_EXT + 0.66665 3HB_EXT</td>
<td></td>
<td>0.21</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig: 5.1 Plot of ΔS Vs the weighting factor. (a) The plot is based on the data present in Wlaschin et al., 2006 (b) the plot is based on the data generated by the software developed. In both the cases the slope gives the value of “b”
Chapter 6 Construction of a Ladh negative R. eutropha strain, a fructose bisphosphatase negative R. eutropha strain and a triose phosphate isomerase negative R. eutropha strain

6.1 Introduction

The validation of the model developed earlier requires the creation of gene knockouts. In order to develop an efficient protocol for knocking out genes in R. eutropha, an attempt was made to knockout the Lactaldehyde dehydrogenase (ladh) gene. The ladh gene was selected because (i) it is part of the secondary metabolism and (ii) generating LADH negative mutant will help in producing R. eutropha optimized for ethanol production on fructose. Since ladh is a part of the secondary metabolism, knocking this gene out would not produce a lethal mutant and would have no effect on the growth under aerobic conditions.

The ladh gene is responsible for the production of S lactate from lactaldehyde. The ladh gene is a part of a single gene operon and is located upstream from the H16_A1918 gene. Being in the downstream position it presents fewer challenges in terms of open reading frame destruction, elimination of upstream enhancers, promoters or other regions of regulation. The gene is 1542 base pairs long and lies between 716,645 <-> 718,186 on chromosome 1.
The structure of the transcription unit is

**Fig 6.1** The map of the Ladh transcription unit. The figure depicts the ladh transcription unit. It was obtained from biocyc.org

*R. eutropha* has three copies of Fructose bisphosphatase (cbbFc, cbbFp, FBP) being present in two highly homologous cbb operons located on the megaplasmid and on chromosome 2 and one on chromosome 1. CbbFc and cbbFp have the same orientation as the adjacent upstream and downstream cbb genes, comprised 1065 base pairs, with 97% amino acid identity (Yoo & Bowien, 1997). They are bifunctional: they also have sedoheptulose-1,7-bisphosphatase activity. An attempt was made to knock out the FBP gene on chromosome 1. The fact that the upstream and the downstream regions of the FBP are different from cbbFc & cbbFp has been exploited in designing a vector for targeting the chromosomal Fructose bisphosphatase (FBP). The FBP forms an independent transcription unit unlike the cbbFp & cbbFc. The gene lies between 1,090,943 < 1,091,959 on chromosome 1. Fructose bisphosphatase is redundant and its absence would not make much difference metabolic flux on the organism which makes it an ideal candidate for mastering the knocking out protocol.

Triose phosphate isomerase catalyzes the interconversion of glyceraldehyde-3-phosphate <=> dihydroxyacetone phosphate. It is encoded by the gene tpiA present on chromosome. It represents one of the few reactions catalyzed by a single enzyme (with no isozymes). The reaction is a part of the glycolysis and is known to be essential
during lithoautotrophy. The gene for Triose phosphate isomerase (H16_A1047) lies between positions 1,145,822 -> 1,146,559 on chromosome 1.

6.2 Construction of suicide vector for knocking out lactaldehyde dehydrogenase, fructose 1,6 bisphosphatase and triose phosphate isomerase:

Three suicide vectors were constructed to knockout the lactaldehyde dehydrogenase gene (pexKanLac) chromosomal fructose 1, 5 bisphosphatase (pexKancFBP) and triose phosphate isomerase (pexKanTpi).

Lactaldehyde dehydrogenase gene is 1542 base pairs long. In order to minimize any effect the deletion of gene sequence would have on neighboring genes, 494 base pairs of the 1542 base pairs ladh gene was eliminated. The upstream region included the 3’ region of ladh and 5’ end of H16_A0667 gene while the downstream region included the 5’ part of ladh and 3’ end of H16_A0665 gene. Both H16_A0667 and H16_A0665 have an open reading frame but their products are unknown. Care was taken not to disrupt the ORF of these genes. The fragments of the upstream and downstream regions that extended into these ORF were selected as multiples of three. The primers used to amplify the upstream and downstream regions with an annealing temperature of 60 and extension time of 1 minute is

Vec_up_for_lac2dr : Forward primer for upstream region of pexKanLac
Up_Rev_lac2dr : Reverse primer for upstream region of pexKanLac
Down_for_lac2dr : Forward primer for downstream region of pexKanLac
Vec_down_rev_lac2dr : Reverse primer for downstream region of pexKanLac

The FBP gene is 1017 base pairs long. Since it is a single gene transcription unit, the entire gene sequence was eliminated. Two fragments

(i) 716 base pairs upstream of the FBP open reading frame which comprises of a part of aminopeptidase N

(ii) 504 base pairs downstream of the fructose bisphosphatase ORF including a part of tRNA-thr

were amplified using the following primers at a Tm of 62°C and extension time of 1 minute.

FBP_Up_For_Vec : Forward primer for upstream region of pexKancFBP

FBP_Dn_Rev_Vec : Reverse primer for upstream region of pexKancFBP

FBP_Down_For : Forward primer for downstream region of pexKancFBP

FBP_Up_Rev : Reverse primer for downstream region of pexKancFBP

The TPI gene is 738 basepairs long and is present as a single transcription unit. The Upstream region putative Zn-dependent oxidoreductase (578 bps) and downstream region pre-protein translocase subunit SecG (617 bps) were amplified from the Genomic DNA using the following primers:

GG5R_VEC_FOR Forward primer for upstream region of pexKanTpi

GG5R_ZN_REV Reverse primer for upstream region of pexKanTpi

GG5R_TPI_FOR Forward primer for downstream region of pexKanTpi
GG5R_TPI_REV  Reverse primer for downstream region of pexKanTpi

The primers were designed to have a 30 basepair overlap with the vector / A tm of 55 °c was used with 2 minutes extension time.

The pexKan plasmid was digested with unique restriction enzyme SmaI (NEB, Ipswich, MA). All the fragments and the backbone were gel purified with 0.8 % agarose and 0.1 % crystal violet. Three separate CPEC reactions described in Chapter 2 were carried out to assemble the pexKanLac, pexKancFBP and pexKanTpi plasmids. Colonies formed on the plate were analyzed with primers Vec_up_for_lac2dr & Kan_rev for the pexKanLac plasmid and with FBP_Up_F for_Vec and Kan_rev for pexKancFBP plasmid. The pexKanTpi plasmid was verified with the primers GG5R_VEC_FOR & GG5R_TPI_REV designed to amplify the upstream and downstream regions.

Few positive colonies were selected and inoculated into 5 mL of LB media supplemented with Kanamycin and ampicillin. Plasmids were extracted as described in Chapter 2. The pexKanLac plasmid was screened by digestion with AAT II & SmaI (NEB, Ipswich, MA) while the pexKancFBP was restricted with BamHI (NEB, Ipswich, MA). The pexKanTpi plasmid was extracted and digested with BsaBI (NEB, Ipswich, MA). One of the BsaBI sites is unique to the inserted fragment. The presence of the restriction site at the right position would verify the presence and orientation of the upstream and downstream fragments in the vector.
6.3 Transfer of Plasmid to R. eutropha

The plasmids from mating strain of E. coli (S17-1) were transferred into R. eutropha by conjugation using the mating strategy described previously. Colonies from the LB agar plates supplemented with Gentamycin sulfate (10mg/mL), Kanamycin sulfate (50mg/mL) and ampicillin (100mg/mL) were selected and genomic DNA was extracted for the PCR screened. For the R. eutropha strains carrying the pexKanLac, the LCTAD2R_ful_for-Kan_rev primers were used with a melting temperature of 65 °C and extension time of 3 minutes. LCTAD2R_ful_for binds outside the upstream region and hence can be used to confirm the presence of the plasmid at the right location in the genome.

Two kinds of recombination events may occur. Depending on the position of occurrence the possibilities are

Recombination at the downstream region
Recombination at the upstream region

Fig 6.2 Diagrammatic representation of the pexKanLac integration into the R. Eutropha: Integration of the plasmid can occur at the upstream or downstream homologous regions. Based on position of recombination two different orientations of the plasmid are possible as shown in the map above.

A second PCR was done with the pexKanLac strain to verify if the plasmid integrated at the right position. The Sac_b_pex_for- LCTAD2R_ful_rev primers were used and a map of the primer binding position is shown below. Annealing temperature
was set to 63 °C and extension time was set to 3 minutes. Since the expected product size is around 6.0 Kb the Phusion PCR was used.

The primer binding regions are shown in the diagram below.

**Fig 6.3** Map of primers Binding regions on the integrated pexKanLac plasmid: Sac_b_for and lac_full_rev primers bind on the plasmid and below the downstream regions as shown in the figure above.

The *R. eutropha* strains carrying the pexKancFBP were screened with the FBP_for_veri- Kan_rev primers at a Tm of 63 °C and an extension time of 3.0 minutes.

Two different cases of recombination possible with the pexKancFBP plasmid are

**Recombination at the upstream region**

**Recombination at the downstream region**

**Fig 6.4** Diagrammatic representation of the pexKancFBP integration into the *R. Eutropha*: Integration of the plasmid can occur at the upstream or downstream homologous regions. Based on position of recombination two different orientations of the plasmid are possible as shown in the map above.

The *R. eutropha* strains with pexKanTpi were tested with primers Zn_for & GG5r_rev_veri. Both of the primers are 20 basepairs long. Zn_for binds a few base
before the upstream region while GG5r_rev_veri binds within the downstream region. The PCR was done with an annealing temperature of 65 °C and an extension time of 4 minutes. Depending on the recombination event, two differently sized fragments can be amplified as shown below:

Recombination at the Upstream region

Recombination at the downstream region

**Fig 6.5** Diagrammatic representation of the pexKanTpi integration into the R. Eutropha: Integration of the plasmid can occur at the upstream or downstream homologous regions. Based on position of recombination two different orientations of the plasmid are possible as shown in the map above

A second PCR was done with the pexKanTpi integrated R. eutropha to confirm the orientation of the plasmid inside the host genome. The Secg_Rev & GG5r_for_veri primers were used and a diagrammatic representation is shown below:

Recombination at the Upstream region

Recombination at the downstream region

**Fig 6.6** Map of primers Binding regions on the integrated pexKanTpi plasmid Secg_Rev binds a few base pairs after the downstream region while GG5r_for_veri binds within the upstream region.

Following the screening process, a few colonies of *R. eutropha* carrying either the pexKancFBP or pexKanLac or pexKanTpi were selected and grown on LB media
containing 10 mg/mL Gentamycin Sulfate and 100 mg/mL Kanamycin sulfate to an OD600 of 0.4.

6.4 SacB counter selection

The cells were centrifuged at 4,000 rpm for 5 minutes and washed twice with 0.8 % NaCl. The cells suspended in 200 µl of 0.8 % NaCl, serially diluted and plated on Doi media plates containing 5 % sucrose and 10mg/mL of Gentamycin sulfate. This was done to remove all the parts (except the homologous regions) of the plasmid. Two kinds of populations, namely the revertents and knockouts result as shown in the fig 6.6
**Fig 6.7** Process flow diagram for the genomic integration of the suicide plasmid and the SacB counter selection. The diagram shows the suicide vector being integrated into the R. eutropha genome followed by its removal after growth on 5 % sucrose.
6.4 Screening for the knockout phenotype following the counter selection

Few colonies were picked from the doi sucrose plates and PCR screened. To identify the ladh negative phenotype the LCTAD2R_ful_for-LCTAD2R_ful_rev primers were used

These primers bind outside the upstream and downstream regions respectively. An annealing temperature of 65 °C was used and an extension time of 3 minutes was used. A map depicting the primer binding region and the expected fragment sizes is shown below:

Revertent Phenotype                  Knockout Phenotype

Fig 6.8 Diagrammatic representation of the primers (LCTAD2R_ful_for-LCTAD2R_ful_rev) binding regions: As the figure shows the primers LCTAD2R_ful_for-LCTAD2R_ful_rev binds outside the upstream and downstream regions and can be used to differentiate between the knockouts and Revertents

A similar procedure was repeated to identify the FBP negative mutant. The FBP_For_veri - FBP_rev_veri primers were used at a Tm of 70 °C and an extension time of 3.00 minutes. A map of the primer binding region along with the expected fragment sizes for the wild type and revertents is shown in fig 6.8
Fig 6.9 Diagrammatic representation of the primers (FBP_For_veri - FBP_rev_veri) binding regions: As the figure shows the primers FBP_For_veri - FBP_rev_veri binds outside the upstream and downstream regions and can be used to differentiate between the knockouts and Revertents.

Primers Vec_Zn_for - Tpi_Rev were used to screen for GG5r negative mutant. As the figure below shows these primers can differentiate between the reverent and knockout phenotype.

Fig 6.10 Diagrammatic representation of the primers (Vec_Zn_for - Tpi_Rev) binding regions: As the figure amplification with the Vec_Zn_for and Tpi_Rev primers can be used to differentiate between the knockouts and Revertents.
6.5 Results & Discussion

The results for the lactaldehyde dehydrogenase mutants are discussed first followed by the results for fructose bisphosphatase.

6.5.1 Construction of the lactaldehyde dehydrogenase negative strain

**Fig 6.11** 0.8 % gel displaying fragments used for invitro assembly of pexKanlac: Gels contained 0.8% agarose and ethidium bromide for visualization of DNA under UV light. (a) Cloned fragments (Lanes 1 & 2) from Ralstonia genomic DNA using Vec_up_for_lac2dr- Up_Rev_lac2dr and Down_for_lac2dr - Vec_down_rev_lac2dr. (b) Sample of pex-100T digested with unique cutter BamHI.

The upstream region and the downstream regions have bands at the right sizes as expected. The plasmid appears linear and at the correct size of 6846 base pairs.
6.5.1.1 PCR for screening the constructed pexKanlac Plasmid:

Fig 6.12 PCR from E.coli (S17) colonies following transformation of the invitro assembled pexKanlac plasmid: Plasmids were extracted from the three colonies picked from the LB amp kan selection plates after the transformation of pexKanLac plasmid and used as the DNA templates for screening S17 colonies to verify the presence of the recombined plasmid. Primers used: Vec_up_for_lac2dr- Vec_down_rev_lac2dr

Two of the three colonies showed bands at the right position which corresponds to 1220 base pairs. This indicates that the upstream and the downstream regions have reconstructed as expected. Ability of the cells to replicate in the presence of ampicillin and kanamycin indicates the presence of the pexKan plasmid.
6.5.1.2 Verification of Properly Assembled pexKanlac Plasmid

![Image of gel showing bands for 3652, 2194/1959, and 100 base pairs.]

Fig 6.13 SmaI & AAt II digest of pexKanlac plasmid from E. coli plasmid extraction: Plasmid DNA extracted from 3 colonies on LB amp kan plates and restricted with SmaI and AAtII is being shown on a 0.8 % agarose gel

The expected fragment sizes are 3692, 2194, 1959, 100 base pairs. All the 3 colonies show bands at the desired positions. Colony 3 did not show a band in the diagnostic PCR. This could be due to insufficient amounts of template or presence of some PCR inhibitors. The 0.8 % gel could not resolve 2194 and 1959 base pair bands. Also the 100 base pair fragment was insufficient to be visualized.
6.5.1.3 Screening for pexKanlac Integration into *R. eutropha* Genome

**Fig 6.14** 0.8% agarose gel verifying the genomic integration of pexKanLac plasmid into *R. eutropha* (a) Five *R. eutropha* colonies (Lanes 1-5) were initially screened with primers LCTAD2R_ful_for-KanRev to verify the integration of the plasmid. Lane 6 shows PCR with pexKanLac plasmid as template. Lane 7 shows PCR with Wild type *R. eutropha* Genomic DNA as template. Lane 8 is the negative control with no DNA template (b) The first three positive colonies (Lanes 1-3) were subjected to a second and longer PCR with primers Sac_b_pex_for- LCTAD2R_ful_rev

Of the 5 colonies tested for the presence of the pexKanLac plasmid, 4 colonies gave a positive result. LCTAD2R_ful_for binds outside the upstream region and the presence of a band verifies the integration at the right location in the genome. They pointed to a recombination event between the upstream region of the plasmid and the upstream region in the genome. To verify the findings further, the second and a longer PCR was done with the first 3 positive colonies. Colonies 1 and 3 run on lane 1 and 3 of fig 6.15 b show a band at the right position. These colonies were selected for sacB counter selection. Based on these PCR we can confirm that the plasmid position in the genome is as shown in the figure below:
Fig 6.15 Map of the plasmid integrated into R. eutropha Genome: The upstream and the downstream regions of the *R. eutropha* genome are shaded in blue while the regions from the pexKan plasmid are shaded in orange.

6.5.1.4 Sac B counter selection for obtaining *ladh* negative strain:

![Image of PCR results](image)

**Fig 6.16** PCR from *R. eutropha* colonies following growth on sucrose media: 10 colonies (Lanes 1-10) picked from the LB plates with 5% sucrose were used as template for PCR screening the colonies using primers LCTAD2R_ful_for-LCTAD2R_ful_rev. Lane W shows PCR with Wild type *R. eutropha* Genomic DNA as template. Lane N is the negative control with no DNA template.

Colonies 1 and 3 which showed positive results for the presence of the plasmid were grown in LB media supplemented with Gentamycin sulfate and kanamycin sulfate and used for sac B counter selection. 10 colonies were screened with primers,
LCTAD2R_ful_for- LCTAD2R_ful_rev which amplifies regions outside the upstream and downstream regions. Among the 10 colonies screened, 8 colonies showed knockout phenotype with band at the expected position. The ratio of the knockout vs. revertents was not 50:50 as expected.

6.5.1.5 Sequencing of ladh negative strain

Two of the colonies which showed positive knockout traits were grown on minimal salt plates supplemented with gentamycin sulfate in gas chamber with 80:10:10 ratio of CO2: H2: O2. Genomic DNA was extracted from these cultures and, LCTAD2R_ful_for- LCTAD2R_ful_rev primers were used to amplify the fragment sent for sequencing. Sequencing results confirmed the knockout phenotype.
6.5.2 Construction of the chromosomal fructose bisphosphatase negative strain

**Fig 6.17** 0.8 % gel displaying fragments used for invitro assembly of pexKancFBP: Gels contained 0.8% agarose and ethidium bromide for visualization of DNA under UV light. (a) Upstream and Downstream regions amplified from Ralstonia genomic DNA using FBP_Up_For_Vec- FBP_Up_Rev and FBP_Down_For- FBP_Dn_ Rev_Vec . (b) Sample of pex-100T digested with unique cutter BamHI. The upstream region and the downstream regions have bands at the right sizes as expected. The plasmid appears linear and at the correct size of 6846 base pairs.
6.5.2.1 Verification of Properly Assembled pexKancFBP Plasmid:

PCR verification

**Fig 6.18** PCR from E.coli (S17) colonies following transformation of the invtro assembled pexKancFBP plasmid: Seven colonies picked from the LB amp kan selection plates after the transformation of pexKancFBP plasmid were used as the DNA template for screening S17 colonies to verify the presence of the recombined plasmid in one of them. Primers used: FBP_Up_For_Vec-FBP_Dn_Rev_Vec.

Five of the seven colonies showed bands at the right position which corresponds to 1004 base pairs. This indicates that the upstream and the downstream regions have reconstructed as expected. Ability of the cells to replicate in the presence of ampicillin and kanamycin indicates the presence of the pexKan plasmid
Verification by restriction

Fig 6.19 BglII digest of upstream-downstream fragment pexKancFBP plasmid: Purified fragment of Upstream-Downstream region amplified by using FBP_Up_For_Vec-FBP_Dn_Rev_Vec after restriction with BglII is being shown on a 0.8 % agarose gel.

The restriction digest of upstream and downstream regions purified from two colony PCR show bands at 706 base pairs and 299 base pairs as expected.
6.5.2.2 Screening for pexKancFBP Integration into *R. eutropha* Genome

![0.8% agarose gel verifying the genomic integration of pexKancFBP plasmid into *R. eutropha*: Five *R. eutropha* colonies (Lanes 1-5) were screened with primers FBP_for_veri–KanRev. Lane 6 shows PCR with Wild type *R. eutropha* Genomic DNA as template. Lane 7 shows PCR with pexKancFBP plasmid as template. Lane 8 is the negative control with no DNA template.

Of the 5 colonies tested for the presence of the pexKancFBP plasmid, all the colonies gave a positive result. FBP_for_veri binds outside the upstream region and the presence of a band verifies the integration at the right location in the genome. They pointed to a recombination event between the upstream region of the plasmid and the upstream region in the genome. Based on these PCR we can confirm that the plasmid position in the genome is as shown in the figure below:
Fig 6.21 Map of the penKanFBP plasmid integrated into R. eutropha Genome: The upstream and the downstream regions of the R. eutropha genome are shaded in blue while the regions from the pexKan plasmid are shaded in orange.

6.5.2.3 Sac B counter selection for obtaining fbp negative strain:

Fig 6.22 PCR from R. eutropha colonies following growth on sucrose media : 8 colonies (Lanes 1-8) picked from the LB plates with 5 % sucrose were used as template for PCR screening the colonies using primers FBP_for_veri- FBP_rev_veri. Lane G shows PCR with Wild type R. eutropha Genomic DNA as template. Lane N is the negative control with no DNA template.

Colonies 1 and 7 which showed positive results for the presence of the plasmid were grown in LB media supplemented with Gentamycin sulfate and kanamycin sulfate and used for sac B counter selection.10 colonies were screened with primers FBP_for_veri- FBP_rev_veri which amplify regions outside the upstream and downstream regions. Among the 8 colonies screened, 1 colony showed knockout phenotype with band at the expected position.
6.5.2.4 Sequencing of fbp negative strain

The single colony which showed positive knockout traits were grown on minimal salt plates supplemented with gentamycin sulfate in gas chamber with 80:10:10 ratio of CO2: H2: O2. Genomic DNA was extracted from these cultures and, FBP_for_veri- FBP_rev_veri primers were used to amplify the fragment sent for sequencing. Sequencing results confirmed the knockout phenotype.

6.5.3 Construction of the Triose phosphate isomerase (TPI) negative strain

![Fig 6.23](image)

**Fig 6.23** 0.8 % gel displaying fragments used for invitro assembly of pexKanlac: Gels contained 0.8% agarose and ethidium bromide for visualization of DNA under UV light. (a) Lane 1: Downstream region amplified using GG5R_TPI_FOR and GG5R_TPI_REV from *R. eutropha* genomic DNA. Lane 2: Upstream region amplified using GG5R_VEC_FOR and GG5R_VEC_Rev from *R. eutropha* genomic DNA (b) Bsa BI restriction digest of plasmids extracted from two different colonies.
The upstream region and the downstream regions have bands at the right sizes as expected. The restriction with BsaBI produced two fragments of sizes 5612 and 2337 as expected. This confirms the presence of the upstream and downstream fragments in the right orientation within the pexkanTpi plasmid.

6.5.3.1 PCR for screening Constructed pexKanTpi Plasmid:

![Image of gel electrophoresis with bands at 1100 and 1600 base pairs]

**Fig 6.24** PCR from E.coli (S17) colonies following transformation of the invitro assembled pexKanTpi plasmid: Three colonies picked from the LB amp kan selection plates after the transformation of pexKanTpi plasmid and used as the DNA template for screening S17 colonies (lanes 1-3) to verify the presence of the recombined plasmid. Lane 4 shows the fragment amplified using genomic DNA from *R.eutropha* as the template. Primers used: GG5R_VEC_FOR & GG5R_TPI_REV

All the three colonies showed bands at the right position which corresponds to 1100 base pairs. This indicates that the upstream and downstream regions have reconstructed as expected. Ability of the cells to replicate in the presence of ampicillin and kanamycin indicates the presence of the pexKan plasmid.
6.5.3.2 Screening for pexKanTpi Integration into *R. eutropha* Genome

![Agarose Gel](image)

**Fig 6.25** 0.8% agarose gel verifying the genomic integration of pexKanTpi plasmid into *R. eutropha* (a) Five *R. eutropha* colonies (Lanes 1-5) were initially screened with primers Secg_Rev & Zn_for to verify the integration of the plasmid. Lane 6 & 7 show PCR with Wild type *R. eutropha* Genomic DNA as template. Lane 8 is the negative control with no DNA template. (b) Five *R. eutropha* colonies (Lanes 1-5) were then screened with primers Zn_for & GG5r_rev_veri to confirm the integration of the plasmid. Lane 6 & 7 show PCR with Wild type *R. eutropha* Genomic DNA as template. Lane 8 is the negative control with no DNA template.

All the 5 colonies tested for the presence of the pexKanTpi plasmid gave a positive result with primers Secg_Rev & Zn_for. Zn_for binds outside the upstream region and the presence of a band verifies the integration at the right location in the genome. They pointed to a recombination event between the downstream region of the plasmid and the downstream region in the genome. To verify the findings further, the second PCR was done with the positive colonies. Colonies 1 to 3 run on lanes 1, 2, 3 of fig 6.27b show a band at the right position. These colonies were selected for sacB counter selection.

Based on these PCR we can confirm that the plasmid position in the genome is as shown in the figure 6.26.
Fig 6.26 Map of the penKanTpi plasmid integrated into R. eutropha Genome: The upstream and the downstream regions of the *R. eutropha* genome are shaded in blue while the regions from the pexKan plasmid are shaded in orange. All the five colonies tested had a recombination event at the downstream region.

6.5.3.3 Sac B counter selection for obtaining Tpi negative strain:

Fig 6.27 PCR from *R. eutropha* colonies following growth on sucrose media: 13 colonies (Lanes 1-13) picked from the LB plates with 5% sucrose were used as template for PCR screening the colonies using primers Vec_zn_for and Tpi_rev primers with a Tm of 60 °C. These primers should produce a band of size 1100 bps with the mutant and 1600 with the wild type. Lane 14 shows PCR with pexKanTpi plasmid DNA as template. Lane 15 shows PCR with Wild type *R. eutropha* Genomic DNA as template. Lane 16 is the negative control with no DNA template.

Though several rounds of screening were done, we were unable to identify a positive knockout. LB–sucrose media was replaced with minimal media + lactic acid + sucrose or minimal media + citric acid + sucrose. However, the mutant could not be isolated.
References


Dandekar, T., & Sauerborn, T. - *Comparative genome analysis and pathway reconstruction* - Future Medicine. doi:- 10.1517/14622416.3.2.245


Appendix A: List of reactions metatool input files

Lithoautotrophy

-ENZREV
PCKr PHB4r TCA7r TCA6r TCA4br TCA5r GG6r GG7r GG4r GG9r GG5r GG1r GG8r FC1r PPP5r PPP7r PPP6r PPP3r PPP4r ALCDet1r ADH Glut1r GLUM1r gg13r TRA13 TRA15 TRA6r TRA5r cal3

-ENZIREV
MEB GLB1 GLB2 PHB9 TCA3 TCA2 TCA4a TCA1 GG11 FBP SUCCD1 OPM4 CYTCo SUCCD2 GG10 GG12 PHB8 ATPSYN cal1 Biomass FC2 TRA1 TRA2 TRA19 TRA20 CAL4a HYDGq SH ALS ILV1 ILV2 KDC ADHB

-METINT
glx eth akg xu5p r5p dhap h2 succ icit fum 3pg nh3 f6p 2pg pbhb glum rgt adp oaa succoa rl5p co2 fdp pyr glut s7p atp ackoa cit GA3P mal 3PGP aacoa acac e4p acal pep o2 3HB uq uqh2 fad coa fadh2 amp nadh nadph nadp nad g6p s17bp AcLAC DHXisoVAL KetoisoVAL IsoBUTANAL IsoBUTANOL

-METEXT
nh3_e biomass pbhb_e eth_e succ_e co2_e h2_e atp_base o2_e IsoBUTANOL_e

-CAT
MEB : 1 mal + 1 nadp = 1 co2 + 1 pyr + 1 nadph
PCKr : 1 oaa + 1 atp = 1 adp + 1 co2 + 1 pep
GLB1 : 1 icit = 1 glx + 1 succ
GLB2 : 1 glx + 1 accoa = 1 mal + 1 coa
PHB9 : 1 3HB = 1 pbhb + 1 coa
PHB4r : 1 3HB + 1 nadp = 1 aacoa + 1 nadph
TCA7r : 1 mal + 1 nad = 1 oaa + 1 nadh
TCA3 : 1 icit + 1 nadp = 1 akg + 1 co2 + 1 nadph
TCA6r : 1 fum = 1 mal
TCA2 : 1 cit = 1 icit
TCA4br : 1 succ + 1 atp + 1 coa = 1 adp + 1 succoa
TCA4a : 1 akg + 1 coa + 1 nad = 1 succoa + 1 co2 + 1 nadh
TCA1 : 1 oaa + 1 accoa = 1 cit + 1 coa
TCA5r : 1 succ + 1 fad = 1 fum + 1 fadh2
GG6r : 1 GA3P + 1 nad = 1 3PGP + 1 nadh
GG11 : 1 adp + 1 pep = 1 pyr + 1 atp
GG7r : 1 3pg + 1 atp = 1 adp + 1 3PGP
FBP : 1 fdp = 1 f6p
GG4r : 1 fdp = 1 dhap + 1 GA3P
GG9r : 1 2pg = 1 pep
GG5r : 1 dhap = 1 GA3P
GG1r : 1 g6p = 1 f6p
GG8r : 1 2pg = 1 3pg
SUCCD1 : 1 succ + 1 uq = 1 fum + 1 uqh2
OPM4 : 1 adp + 1 uq + 1 nadh = 1 adp + 1 uq + 1 nad
CYTCO : 1 adp + 0.5 o2 + 1 uq + 2 = 1 atp + 1 uqh2 + 1 nad
FC1r : 1 atph + 1 nad = 1 atph + 1 nadp
SUCCD2 : 1 uq + 1 fadh2 = 1 uq + 2 + 1 fad
PPP5r : 1 xu5p + 1 r5p = 1 s7p + 1 GA3P
PPP7r : 1 xu5p + 1 e4p = 1 f6p + 1 GA3P
PPP6r : 1 s7p + 1 GA3P = 1 f6p + 1 e4p
PPP3r : 1 r5p = 1 xu5p
PPP4r : 1 r5p = 1 r5p
GG10 : 1 pyr + 1 atp = 1 pep + 1 amp
ALCDetr : 1 eth + 1 nad = 1 acal + 1 nadh
ADH : 1 acal + 1 coa + 1 nad = 1 acoa + 1 nadh
GG12 : 1 pyr + 1 coa + 1 nad = 1 co2 + 1 acoa + 1 nadh
PHB8 : 2 acoa = 1 aacoa + 1 coa
ATPSYN : 1 atp = 1 adp + 1 atp_base
cal1 : 1 r5p + 1 co2 + 1 atp = 2 3pg + 1 adp
Glut1r : 1 akg + 1 nh3 + 1 nadph = 1 glut + 1 nadp
GLUM1r : 1 nh3 + 1 glut + 1 atp = 1 glum + 1 adp
Biomass : 3.91 r5p + 8.062 3pg + 0.269 f6p + 1.71 glum + 13.55 oaa + 32.13 pyr + 56.54 glut + 81.03 atp + 4.6 acoa + 0.7644 GA3P + 1.799 e4p + 3.09 pep + 11.09 nadph + 4.7 nad + 0.807 g6p = 51.399 akg + 81.03 atp + 51.399 co2 + 4.6 coa + 4.7 nadh + 11.09 nadp + 1 biomass
gg13r : 1 s17bp = 1 s7p
FC2 : 1 atp + 1 amp = 2 adp
TRA1 : 1 eth = 1 eth_e
TRA13 : 1 o2_e = 1 o2
TRA15 : 1 nh3_e = 1 nh3
TRA2 : 1 pbhb = 1 pbhb_e
TRA19 : 1 IsoBUTANOL = 1 IsoBUTANOL_e
TRA20 : 1 succ = 1 succ_e
CAL4a : 1 dhap + 1 e4p = 1 s17bp
TRA6r : 1 co2_e = 1 co2
TRA5r : 1 h2_e = 1 h2
HYDGq : 1 h2 + 1 uq = 1 uqh2
SH : 1 h2 + 1 nad = 1 nadh
cal3 : 1 GA3P + 1 nadp = 1 3PGP + 1 nadph
ALS : 2 pyr = 1 co2 + 1 AcLAC
ILV1 : 1 nadph + 1 AcLAC = 1 nadp + 1 DHXisoVAL
ILV2 : 1 DHXisoVAL = 1 KetoisoVAL
KDC : 1 KetoisoVAL = 1 co2 + 1 IsoBUTANAL
ADHB : 1 nadh + 1 IsoBUTANAL = 1 nad + 1 IsoBUTANOL
**Heterotrophy**

- **ENZREV**
  PCKr PHB4r TCA7r CITLr TCA6r TCA4br TCA5r LADHr GG6r GG7r GG9r GG5r GG1r GG8r FC1r PPP5r PPP7r PPP6r PPP3r PPP4r MGS1r ALCDetr ADHr PTAr ACKr Glut1r GLUM1r gg13r TRA13 TRA15 GG4r cal3

- **ENZIRREV**
  MEB GLB1 GLB2 PHB9 TCA3 TCA2 TCA4a TCA1 HEXf GG11 FBP LACDHq SUCCD1 OPM4 POX CYTCO SUCCD2 PPP1 PGL EDP2 EDP1 ALHD1 GG10 MGS2 LGTTHL APPS2 ACS GG12 PHB8 ATPSYN Biomass FC2 TRA11 TRA8 TRA1 TRA2 TRA19 TRA20 TRA7 CAL4a TRA6 ALS ILV1 ILV2 KDC ADHB

- **METINT**
  glx eth akg mg xu5p r5p dhap h2 succ icit fum 3pg nh3 f6p 2pg pbhb glum rgt adp oaa succoa rl5p co2 fdp pyr glut lac s7p atp accoa cit GA3P mal 3PGP aacoa acac e4p acal pep ac ltg o2 3HB uq uqh2 fad coa fadh2 amp nadh nadph nadp nad fru d6pgc 6PG g6p kdpg r3hbn s17bp actp AcLAC DHXisoVAL KetoisoVAL IsoBUTANAL IsoBUTANOL

- **METEXT**
  nh3_e lac_e biomass pbhb_e eth_e succ_e co2_e h2_e atp_base o2_e fru_e AC_E IsoBUTANOL_e

- **CAT**
  MEB : 1 mal + 1 nadp = 1 co2 + 1 pyr + 1 nadph
  PCKr : 1 oaa + 1 atp = 1 adp + 1 co2 + 1 pep
  GLB1 : 1 icit = 1 glx + 1 succ
  GLB2 : 1 glx + 1 accoa = 1 mal + 1 coa
  PHB9 : 1 3HB = 1 pbhb + 1 coa
  PHB4r : 1 3HB + 1 nadp = 1 aacoa + 1 nadph
  TCA7r : 1 mal + 1 nad = 1 oaa + 1 nadh
  TCA3 : 1 icit + 1 nadp = 1 akg + 1 co2 + 1 nadph
  CITLr : 1 cit = 1 oaa + 1 ac
  TCA6r : 1 fum = 1 mal
  TCA2 : 1 cit = 1 icit
  TCA4br : 1 succ + 1 atp + 1 coa = 1 adp + 1 succoa
  TCA4a : 1 akg + 1 coa + 1 nad = 1 succoa + 1 co2 + 1 nadh
  TCA1 : 1 oaa + 1 accoa = 1 cit + 1 coa
  TCA5r : 1 succ + 1 fad = 1 fum + 1 fadh2
  HEXf : 1 atp + 1 fru = 1 f6p + 1 adp
  LADHr : 1 pyr + 1 nadh = 1 lac + 1 nad
  GG6r : 1 GA3P + 1 nad = 1 3PGP + 1 nadh
  GG11 : 1 adp + 1 pep = 1 pyr + 1 atp
GG7r : 1 3pg + 1 atp = 1 adp + 1 3PGP
FBP : 1 fdp = 1 f6p
GG9r : 1 2pg = 1 pep
GG5r : 1 dhap = 1 GA3P
GG1r : 1 g6p = 1 f6p
GG8r : 1 2pg = 1 3pg
LACDHq : 1 lac + 1 uq = 1 pyr + 1 uqh2
SUCCD1 : 1 succ + 1 uq = 1 fum + 1 uqh2
OPM4 : 1 adp + 1 uq + 1 nadh = 1 atp + 1 uqh2 + 1 nad
POX : 1 pyr + 1 uq = 1 co2 + 1 ac + 1 uqh2
CYTCO : 1 adp + 0.5 o2 + 1 uqh2 = 1 atp + 1 uq
FC1r : 1 nadph + 1 nad = 1 nadh + 1 nadp
SUCCD2 : 1 uq + 1 fadh2 = 1 uqh2 + 1 fad
PPP1 : 1 nadp + 1 g6p = 1 nadph + 1 6PG
PPP5r : 1 xu5p + 1 r5p = 1 s7p + 1 GA3P
PPP7r : 1 xu5p + 1 e4p = 1 f6p + 1 GA3P
PPP6r : 1 s7p + 1 GA3P = 1 f6p + 1 e4p
PGL : 1 6PG = 1 d6pgc
EDP2 : 1 kdpg = 1 pyr + 1 GA3P
EDP1 : 1 d6pgc = 1 kdpg
PPP3r : 1 rl5p = 1 xu5p
PPP4r : 1 r5p = 1 rl5p
ALHD1 : 1 acal + 1 nad = 1 ac + 1 nadh
GG10 : 1 pyr + 1 atp = 1 pep + 1 amp
MGS2 : 1 ltg = 1 rgt + 1 lac
MGS1r : 1 dhap = 1 mg
LGTTHL : 1 mg + 1 rgt = 1 ltg
ALCDetr : 1 eth + 1 nad = 1 acal + 1 nadh
ADHr : 1 acal + 1 coa + 1 nad = 1 accoa + 1 nadh
PTAr : 1 accoa = 1 coa + 1 actp
ACKr : 1 adp + 1 actp = 1 atp + 1 ac
APPS2 : 1 actp = 1 ac
ACS : 1 atp + 1 ac + 1 coa = 1 accoa + 1 amp
GG12 : 1 pyr + 1 coa + 1 nad = 1 co2 + 1 accoa + 1 nadh
PHB8 : 2 accoa = 1 aacoa + 1 coa
ATPSYN : 1 atp = 1 adp + 1 atp_base
Glut1r : 1 akg + 1 nh3 + 1 nadph = 1 glut + 1 nadp
GLUM1r : 1 nh3 + 1 glut + 1 atp = 1 glum + 1 adp
Biomass : 3.91 r5p + 8.062 3pg + 0.269 f6p + 1.71 glum + 13.55 oaa + 32.13 pyr +
56.54 glut + 81.03 atp + 4.6 accoa + 0.7644 GA3P + 1.799 e4p + 3.09 pep + 11.09
nadph + 4.7 nad + 0.807 g6p = 51.399 akg + 81.03 adp + 51.399 co2 + 4.6 coa + 4.7
nadh + 11.09 nadp + 1 biomass
gg13r : 1 s17bp = 1 s7p
FC2 : 1 atp + 1 amp = 2 adp
TRA11 : 1 atp + 1 fru_e = 1 adp + 1 fru
TRA8 : 1 lac = 1 lac_e
TRA1 : 1 eth = 1 eth_e
TRA13 : 1 o2_e = 1 o2
TRA15 : 1 nh3_e = 1 nh3
TRA2 : 1 pbhb = 1 pbhb_e
TRA19 : 1 IsoBUTANOL = 1 IsoBUTANOL_e
TRA20 : 1 succ = 1 succ_e
TRA7 : 1 ac = 1 AC_E
CAL4a : 1 dhap + 1 e4p = 1 s17bp
TRA6 : 1 co2 = 1 co2_e
GG4r : 1 fdp = 1 dhap + 1 GA3P
cal3 : 1 GA3P + 1 nadp = 1 3PGP + 1 nadph
ALS : 2 pyr = 1 co2 + 1 AcLAC
ILV1 : 1 nadph + 1 AcLAC = 1 nadp + 1 DHXisoVAL
ILV2 : 1 DHXisoVAL = 1 KetoisoVAL
KDC : 1 KetoisoVAL = 1 co2 + 1 IsoBUTANAL
ADHB : 1 nadh + 1 IsoBUTANAL = 1 nad + 1 IsoBUTANOL
Appendix B: In Depth description of rational strain design algorithm/software

B.1 Requirements

The code has been developed for a Windows based PC with gcc compilers. The software is to be used in conjunction with other software like CellNetAnalyzer (Steffen et al., 2007) or Metatool (Pfeiffer et al., 1999) which enumerate the elementary modes. A wrapper utility has been designed in DOS to increase the user friendliness. The elementary modes can be saved in a matrix form of tab delimited text files and are used as input for the software designed.

Apart the text file containing the elementary modes, the application requires the number of total reactions required to generate the file, the names of the substrate, product and biomass producing reactions as input. The application attempts to couple biomass production to product synthesis. If it is unable to do so, it retains maximum product producing mode along with at least one biomass producing mode. The output includes a text file under the results folder which contains a list of the knockouts that might be implemented to obtain a strain which can effectively convert the substrate to product. The results file contains information about the minimum and the maximum yield of the substrate and the product after each reaction is knocked out. Many a times, multiple reactions can have the same effect on the number of remaining elementary modes. The software arranges these reactions in alphabetical order and picks out the first reaction. One reaction may be preferred over the other based on the presence of isozymes or the
ease of knocking the reaction. The alternate reactions and the effect of the reactions on each elementary mode set can be determined from files generated in an ancillary folder labeled KO.

B.2 Specifications

The program reads in an elementary mode matrix is stored in a tab delimited text format. This file can generally be obtained as an output of CellNetAnalyzer, software to calculate elementary modes.

The wrapper utility has been described first followed by the algorithm for the minimal metabolic functionality.

B.3 Wrapper Utility description (Code in DOS script)

The program has been written to handle 5 different optimizations. Each of these optimizations differs in the types of constraints used:

1 - The KO will be found retaining maximum product and bio mass yield with respect to one substrate

2 - The KO will be found retaining maximum product, bio mass and product/biomass yield with respect to 1 substrate

3 - The KO will be found retaining maximum product and bio mass yield with respect to 2 substrates

4 - The KO will be found retaining maximum product, bio mass and product/biomass yield

5 - The KO will be found retaining maximum product yield
6 - to run all the of the above

Based on the user input (integer), the script will check for the presence of necessary files and will compile the source code.

- It will execute the C++ executable
- It will keep a track of all actions in a log file
- It will move the log files to log folder. It will create a log folder if absent
- The list of KO’s will be moved to results folders. It will create a results folder if absent
- The input file will be moved to input folders. It will create a input folder if absent
- The Effects of KO files (eg. KO_2subs_max_prod_bio1.txt) will be moved to KO_effects folder. If the folder is absent it will be created
- The executable will be deleted

B.4 Minimal metabolic functionality (Coded in C++)

All the functions referred to in this section have been described in the following section. Each of the 5 optimizations has been written as separate program files. As a general case, a description of the third optimization has been presented below:

The first row should refer to the reaction names and the rows following it to the elementary modes.
B.5 Algorithm

1. Read the first row into a vector called the “reaction” vector. The user will provide the number of columns in each row (m)

2. Read the remaining rows into a matrix called “Data”. The matrix has been implemented as a row of vectors

3. Find out the column number for the substrates, product and biomass using the function col_find described below.

4. Initialise ko_count =1. This refers to the first knockout. It will be updated when the program progresses to next knockouts when we call the function “KO_filename”. When ko_count =1, KO_r (the variable which stores the reaction to be knocked) will be empty.

5. While (ko_count==1) or (KO_r is not empty) )
   // this loop will terminate when the program is unable to find a reaction which will bring down the number of elementary modes while retaining the yields
   {
   a. Find out the effect of knocking out each reaction by using the function reaction_eff using the Data matrix as input
   b. Print out the effect to “KO_2subs_max_prod_bio1.txt” using the function print_table . The 1 in green refers to the KO number and it will be updated using the function KO_filename .

   }
c. Determine the reaction to be knocked out using the function “KO_rcn”.

d. If the program is able to find a reaction (to be knocked out) which will bring down the number of Elementary modes while retaining the product yields, then it will create another data matrix “Data1” containing only those elementary modes (rows) which will survive when it knocks the reaction. Then it will clear the original data matrix “Data”. The size of “Data” matrix becomes zero at this point. This deletion is to minimize the memory usage.

e. Find out the effect of knocking out each reaction by using the function reaction_eff using the “Data1” matrix as input.

f. Print out the effect to “KO_2subs_max_prod_bio2.txt” using the function print_table. The “2” in green refers to the KO number and it will be updated using the function KO_filename.

g. Determine the reaction to be knocked out using the function “KO_rcn”.

h. If program is able to find a reaction (to be knocked out) which will bring down the number of Elementary modes while retaining the product yields, then it will update data matrix “Data” containing only those elementary modes (rows) which will survive when the program knocks the reaction. Then delete the data matrix “Data1”. The size of “Data1” matrix becomes zero at this point. This deletion is to minimize the memory usage.

i. Find out the effect of knocking out each reaction by using the function reaction_eff using the “Data” matrix as input.
j. Determine the reaction to be knocked out using the function “KO_rcn”.

6. Print out the List of KO’s generated when the “KO_rcn” function was run.

**B.6 operating instructions**

- Copy the flux analysis folder
- Press Shift + right click → Open command window (this is for a windows 7 system). If this does not work, go to start → run → type cmd → enter → type cd <path to the folder without the angular braces here> → enter
- Have a copy of the input file inside the flux analyzer folder
• Type JMJ in the window that pops up. JMJ is the name of the program.

• The next screen will be:

```
This program can handle 6 cases
ECHO is off.
1 - The RO will be found using maximum product and bio mass yield with respect to one substrate.
2 - The RO will be found using maximum product, bio mass and product/bio mass yield with respect to 1 substrate.
3 - The RO will be found using maximum product and bio mass yield with respect to 2 substrates.
4 - The RO will be found using maximum product, bio mass and product/bio mass yield with respect to 2 substrates.
5 - The RO will be found using maximum product yield with respect to 2 substrate.
6 - to run all of the above
Enter the number corresponding to the optimization you want
```

• Type the number corresponding to the desired optimization strategy.

• System will check for the source files and report error if any files are missing.

• It will also check for the presence of the C++ compiler and report error, if any.

• When prompted type the:
  
  • Input file name

  • # of reactions

  • Substrate rcn name
- Product rcn name
- Biomass rcn name

- When the program has run completely, Go to the flux analyzer folder
  - Will have 4 new folders
  - Input folder has all input files
  - Results folder: List of KO for the optimization strategy selected
  - Logs: complete list of all activities with the date, time, file name and error/success message
  - KO_effect: Additional details about the KO. This folder will have files which will details of the effect of knocking out each reaction in an elementary mode matrix
**Appendix C: List of Primers:**

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence(5' - 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG5r_Zn_for</td>
<td>GGTGGGCGACCGCGTCTG</td>
</tr>
<tr>
<td>GG5r_secg_re</td>
<td>TGCCGGCGCCCGATGC</td>
</tr>
<tr>
<td>GG5r_zn_rev</td>
<td>TCATGGAAAGAAGCTGGCAGCAGGGGATCCAGGGGATCCCCAGGTCGTCATGACAAATGCAACTGCGGGATCCAGGTC</td>
</tr>
<tr>
<td>GG5r_tpi_for</td>
<td>GAGGGGCACTTCGTGAGACAAAAGCCGGACAATGCGGCGACGTCGAGCATTAGGATCCGGCG</td>
</tr>
<tr>
<td>GG5r_tpi_rev</td>
<td>GTTAAATTGTTTGAACACTGGCAGGAGCATTAGGATCCGGCG</td>
</tr>
<tr>
<td>FBP_up_for_v</td>
<td>CTCATTACCTGTGATCCGGAATCCAGCCAGCCAGGATCCAGGTCGTAACACTGGCAGGCAGGCTTGGTTGCAGGTC</td>
</tr>
<tr>
<td>FBP_dn_rev_vec</td>
<td>GAGCACCCGCCGCTTCCCAACC</td>
</tr>
<tr>
<td>FBP_down_for</td>
<td>CGTGGAGATCGTGTGTGTGGTTGGTCAGGAGCAACAGGCTGAGGATCCAGGTCGAGCAGGCTTGGAGAGCAGGCTG</td>
</tr>
<tr>
<td>FBP_up_rev</td>
<td>TGCCGGCGGAGGCCGCAACAGGCTGAGGATCCAGGTCGAGCAGGCTGAGGATCCAGGTCGAGCAGGCTGAGGATCCG</td>
</tr>
<tr>
<td>GG5r_for_veri</td>
<td>GGCGGCAACGTTGATCCAGGTCGAGCAGGCTGAGGATCCAGGTCGAGCAGGCTGAGGATCCAGGTCGAGGATCCG</td>
</tr>
<tr>
<td>GG5r_rev_veri</td>
<td>CACCGAAACAGGCTGAGCAGGCTGAGGATCCAGGTCGAGCAGGCTGAGGATCCAGGTCGAGGATCCAGGTCGAGGATCCG</td>
</tr>
<tr>
<td>Kan_for</td>
<td>AGTGCCAAGGCTCATTACCTGCTTTATCCCTACCCGGGCCGATCCACAGGCTGAGGATCCAGGTCGAGGATCCGAGGATC</td>
</tr>
<tr>
<td>Kan_rev</td>
<td>CCATGATTACGCAAGCTCTAGGGATACAGGGAACACAGGATCCAGGTCGAGGATCCAGGTCGAGGATCCAGGTCGAGGATC</td>
</tr>
</tbody>
</table>

126
LCTAD2R_ful_for  GTGTGATGGCCTCGGGTGC
LCTAD2R_ful_rev AGGATGGCCTCGGCATTGCG
Vec_up_for_lac2d CCAAGCTCATTACCCTGTATTCCCCCTCCGGGGATCCATG
Up_Rev_lac2dr CAACAACGAGACAAACTC
Down_for_lac2dr CCCTGGCCGCGCGCTGACCACGTCGTGCTGCAGGAAGCCAG
Vec_down_rev AACACTGCCAGGCACTTACCGACCCTAGGCTCAGGCCCGCCGGCTTC
_vec2dr TTCAGCTGCAGCAGCTTGT
Sac_b_pex_for CGCGCGTTCGCTGTGATGACG
Sac_b_pex_rev TCGCGCGGCTTTGTTACTGAT
FBP_for_veri GTGGTCAGGGCCAGCGACG
FBP_rev_veri CAACCTGGCGCTCGGCTACC
Kan_R_pex_For AGTGCCAAGTCATTACCCTGTATTCCCCCTCCGGGGATCCCGATC
Kan_R_pex_Rev ATAGGGGATCGATCCTCCCGGATTACCCTGTATTCCCTA
Gg5r_vec_for AAGCTCATTACCCTGTATTCCCCCTACCGGGGGATCCCGATC
Gg5r_zn_rev TCATGGAAAAACAGTTCGCGGCGCATTTGTCGGGCTTTTGTCTC
Gg5r_tpi_for GAGGGGCACTTCGTAGACAAAAGCCGGCAATGCCGGCG
AAGCTTGGCGTAATCATGG
Gg5r_tpi_for AGTGCCAAGTCATTACCCTGTATTCCCGGGATCCCGATC
CTAATGCTCTGCCAGTGTT
Kan_R_pex_Rev ATAGGGGATCGATCCTCCCGGATTACCCTGTATTCCCTA
GAGCTTTGGCGTAAATCATGG
Gg5r_vec_for AAGCTCATTACCCTGTATTCCCCCTACCGGGGGATCCCGATC
Gg5r_zn_rev TCATGGAAAAACAGTTCGCGGCGCATTTGTCGGGCTTTTGTCTC
ACGAAGTGCCGCCCTC
Gg5r_tpi_for GAGGGGCACTTCGTAGACAAAAGCCGGCAATGCCGGCG
AAGCTTGGCGTAATCATGG
127
Gg5r_tpi_rev GTTAATTGGTTGTAACACTGGCAGAGCATTAGGATCCGCGGCGCCGATGCATCAGCG