Improved microbial lipid production from *Rhodococcus opacus* PD630: Molecular and systems biology approach

A Thesis

Submitted in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

by

Bapi Mandal

Department of Biosciences and Bioengineering
Indian Institute of Technology Guwahati
Guwahati 781 039, Assam, India

March 2020

Dedicated to My Parents,
Wife and
Family Members

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING

DECLARATION

I, hereby declare that the research findings in this thesis titled "**Improved microbial lipid production from *Rhodococcus opacus* PD630: Molecular and systems biology approach**" is the result of research work carried out by me under the supervision of Prof. Venkata Dasu Veeranki and Prof. Kannan Pakshirajan, Department of Biosciences and Bioenginering, Indian Institute of Technology Guwahati, for the award of the degree of Doctor of Philosophy. This work has not been submitted elsewhere for any degree or membership of any institute or university to the best of my knowledge and belief.

As per the general norms of reporting research findings, due acknowledgements have been made, wherever the research findings of other researchers have been cited in this thesis.

Date: 29/01/2021 **Bapi Mandal**

Place: Guwahati Roll No: 146106022

INDIAN INSTITUTE OF TECHNOLOGY GUWAHATI

DEPARTMENT OF BIOSCIENCES AND BIOENGINEERING

CERTIFICATE

It is certified that the work described in this thesis entitled "**Improved microbial lipid production from *Rhodococcus opacus* PD630: Molecular and systems biology approach**" by Mr. Bapi Mandal for the award of degree of Doctor of Philosophy is an authentic record of the results obtained from the research work carried out under our supervision in the Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati, India. This work has not been submitted elsewhere for the award of any other degree.

Prof. Venkata Dasu Veeranki
(Thesis supervisor)

Prof. Kannan Pakshirajan
(Thesis supervisor)

ACKNOWLEDGEMENT

This thesis represents not only my work at the keyboard, it is a milestone in half a decade of work at Indian Institute of Guwahati, India and specifically within the Bioprocess Engineering Lab. My experience at Indian Institute of Technology Guwahati, India has been nothing short of amazing.

My deep gratitude goes first to my doctoral supervisor Professor Venkata Dasu Veeranki, who expertly guided me through my research and dissertation and shared the excitement of 5 years of discovery. His unwavering enthusiasm for science kept me constantly engaged with my research, and his personal generosity helped me make my time enjoyable in Indian Institute of Technology Guwahati. He continually and convincingly conveyed a spirit of adventure regarding research and discovery. Without his guidance and persistent help this dissertation would not have been possible.

I am also sincerely indebted to my doctoral another supervisor Professor Kannan Pakshirajan. His mentoring and encouragement have been especially valuable, and his early insights launched the greater part of the dissertation.

I would like to extend my sincere appreciation to Professor Alexandros for letting me fulfil my dream of being a part of an eminent institution like University College London, United Kingdom and carry out a major part of my research in collaboration under his guidance. He has been a tremendous mentor for me. I would like to thank him for encouraging my research and for allowing me to grow in the field of metabolic engineering. His advice on both research as well as on my career have been priceless.

I would also like to thank my committee members, Professor Vikas Kumar Dubey, Dr. Shankar Prasad Kanaujia, Dr. S Senthilkumar and Dr. Soumen Kumar Maiti for serving as my mentors even at hardship. I also want to thank them for letting all my annual presentations be superb learning experiences with all their brilliant comments and suggestions, thanks to you!

Dozens of people have helped and taught me immensely on this journey to this degree. This dissertation would not have been possible without the intellectual contribution of Dr. Ashish Prabhu, Dr. Nitin Bana, and Dr. Basavaraj. I would also like to thank my other lab mates and friends that include Sattu Jhakki, Bidkar, K K bhai, Abshar, Poloumi, Srirupa, Suraj, Subbi, Lalli, Ganesh, Pavan, Rakeshji, Mayur, Shreya, Tejas, Payel and Ansuman for making my experience in the Biochemical Engineering Lab and PhD exciting and fun.

Above ground, I am indebted to my parents, God and my sister Juhi whose value to me only grow with age. This dissertation would not have been possible without their warm love, continued patience, and endless support. They have cherished with me every great moment and supported me whenever I needed it.

Words cannot express how grateful I am to my wife Anwesha for all of the sacrifices that she has made on my behalf and especially I can't thank her enough for encouraging me throughout this experience. I would especially like to thank my father in law for all the assistance with my academics and health without which this dissertation would not have been meaningful. Additionally, a big thanks to my mother in law and Kirul for being great stress releasers in my life.

Finally, I am thankful to MHRD, GOI and Indian Institute of Technology Guwahati for the financial support that I otherwise would not have been able to develop my scientific discoveries.

Abstract

Biofuels are a better source of alternative energy in the state of fossil fuel exhaustion and energy crisis. Biofuels have been into scenario for quite a longer time now. However, catching less attention due to certain major drawbacks that would make them economically undesirable. After long years of research, scientists have discovered several potential microbial hosts being able to synthesize biofuel intracellularly. *Rhodococcus opacus* PD630 (*R. opacus* PD630) is one of them, which has caught attention in the field of industrial biodiesel production. Moreover, this microbe is also well known to degrade complex substances into simpler forms that had been used into waste management. However, it has not been explored well in managing the waste of dairy industry, which is a huge industrial interest in India.

The genes and their products used in biodiesel production are not well explored in *R. opacus* PD630. In the current study, *R. opacus* PD630 was used for simultaneous production of biodiesel and treatment of synthetic dairy wastewater (SDWW) with 80% of chemical oxygen demand (COD) removal as an outcome. Physiology of any organism is best known through its complete metabolism extended from its genome. Systems biology approach to see through its metabolic ability is a superior way to identify genomic problems when there is a lack of definite literature available. Taking into consideration, genomic scale metabolic model (GeM) is designed and constructed (iROPD630) to locate the functionality of genes involved in the complex blueprint, which eventually gives rise to designated phenotypes.

Biodiesel consists primarily of Fatty acid methyl esters (FAMEs) or Fatty acid ethyl Esters (FAEEs), which occurs by transesterification reaction between methanol/ethanol and free fatty-acids in the presence of sulphuric acid or sodium hydroxide acting as

catalyst. Methanol is a toxic substance and the handling alone makes it a little challenging apart from producing it in large volumes. Thus, ethanol is largely considered over methanol for biodiesel production. Though, ethanol has been achieved to be conveniently produced from microbes in a large volume, it adds an extra economical, time and manual burden in the biodiesel production. To have an alternative to this issue, we constructed a superior strain of *R. opacus* PD630 with an ability to synthesize FAEE intracellularly without an external supply of ethanol. To do this, a gene, pyruvate decarboxylase that is responsible for ethanol biosynthesis in *Zymomonas mobilis* ZM4 (*Z. mobilis*) was episomally overexpressed in *R. opacus* PD630 in order to achieve enough ethanol intracellularly to carry out FAEE biosynthesis in the cellular microenvironment. The cultivations led a FAEEs of 28.84% and 31.28% of the dry cell weight (DCW) in lactose and glucose respectively. The experimental data sets obtained from lactose were used to simulate the *i*ROPD630MD (obtained from *i*ROPD630) GeM for better understanding its physiology through metabolic flux analysis. With ethanol as the objective function, the model with carbon-constraint flux variability analysis correctly predicted the ethanol yield. This makes the GeM all the more viable and can be used to predict targets for enhancing ethanol or biodiesel yield.

All these collective approaches on exploring biofuels for the betterment of mankind shall further ease the dependence on fossil fuels and pave a new way for the rise of biofuels.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

GLOSSARY OF ACRONYMS

1G	First Generation
2G	Second Generation
3G	Third Generation
3D	3 Dimension
Acetyl-CoA	Acetyl coenzyme A
ACHR	Artificial Centred Hit and Run
ANN	Artificial Neural Network
ANOVA	Analysis of variance
ASTM	American Society for Testing and Materials
ATP	Adenosine Tri Phosphate
BBD	Box Behnken Design
BLAST	Basic Local Alignment Search tool
BOD	Biological Oxygen Demand
C/N ratio	Carbon to nitrogen ratio
CCD	Central Composite Design
CDS	Coding DNA sequences
CN	Cetane Number
COBRA	Constraint-Based Reconstruction and Analysis
COD	Chemical Oxygen Demand
CP	Cloud Point
DAG	Diacylglycerol
DCW	Dry Cell Weight
DF	Degrees of freedom
DNA	Deoxyribose Nucleic Acid
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
DU	Degree of Unsaturation
DWW	Dairy Wastewater

EBI	European Bioinformatics Institute
EC	Enzyme Commission
EN	European Committee for Standardization
EX	Exchange reactions
FAD	Flavin adenine dinucleotide
FAEE	Fatty Acid Ethyl Ester
FAME	Fatty Acid Methyl Ester
FAS	Fatty acid synthase
FASTA	FAST-All
FBA	Flux Balance Analysis
FDR	First Draft of Reconstruction
FVA	Flux Variability Analysis
GA	Genetic algorithm
GAM	Growth-associated ATP maintenance
GeM	Genome-scale Metabolic Model
GOLD	Genomes online database
GPR	Gene protein reaction
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HMM	Hidden Markov Model
HV	Highest Heating value
IV	Iodine Value
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LB	Luria Bertani
lb	Lower Bound
LP	Linear Programming
LPPs	Lipid Phosphate Phosphatases
MATLAB	Matrix laboratory
MC	Monte-Carlo
MCMC	Markov Chain Monte Carlo

MS	Mean square
MSE	Mean squared error
MSM	Minimal Salt media
NAD	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide + hydrogen+
NADPH	Nicotinamide adenine dinucleotide phosphate + hydrogen+
NCBI	National Centre for Biotechnology Information
NF	Nanofiltration
NS	Not specified
OD	Optical density
ORFs	Open reading Frames
OS	Osmotic stability
PAP2	Phosphatidic acid phosphatase type 2
PC	Principal component
PC1	First principal component
PC2	Second principal component
PCA	Principal component analysis
PCR	Polymerase Chain Reaction
pdc	Pyruvate decarboxylase
PGDB	Pathway/genome database
PHA	Polyhydroxyalkanoates
PP	Pour Point
PSO	Particle swarm optimization
RAST	Rapid Annotation using Subsystem Technology
RNA	Ribose Nucleic Acid
RPM	Revolution per minute
RSM	Response Surface Methodology
SBML	Systems Biology Markup Language
SDS PAGE	Sodium Do Decyl sulfate Poly Acrylamide Gel Electrophoresis
SDWW	Synthetic Dairy Wastewater

SS	Sum of squares
SV	Saponification Value
SVD	Singular value decomposition
TAG	Triacylglycerol
TIGR	The Institute for Genomic Research
TSB	Tryptic soy broth
ub	Upper Bound
UDP	Uridine diphosphate
UV	Ultraviolet
W/DCW	Weight/Dry cell weight
WT	Wild Type

LIST OF NOTATIONS

Variables

°C	Degree Celsius
b	Biases
c	Vector which was used to select linear combination of metabolic fluxes to be included at the objective function
C_0	Number of centre points
c1 and c2	Acceleration constants
g/g	Gram per gram
g/L	gram/Litre
K	Number of variables
KV/cm	Kilo volts per centimetre
L	Litre
mg/L	milligram/Litre
ml	millilitre
mm	millimetre
mM	millimolar
mmole	millimole
MW	The molecular mass of individual fatty acids
MW_i	Molecular weight of the i^{th} FAME component.
N	Total number of experiments/data points
N_D	Number of double bonds
nm	Nanometre
P_{FA}	Percentage of each fatty acid
p_g	Best solution for all particles
p_i	Best solution for this particle
r1 and r2	Elements from two random sequences in the range of (0, 1)
R^2	Coefficient of determination
rev min^{-1}	Revolution per minute

s -	Biomass constituent
S	Stoichiometric matrix
t	Generation of algorithm running
T_f	Flash Point
U	left singular vectors
v	Vector which defines the metabolic flux which includes both internal and external fluxes.
V^T	Rright singular vectors.
vvm	Volume of air/liquid per minute
w	Weights or weight of inertia
w/v	Weight by volume
W_C	Weighted-average number of carbon atoms in the fatty acids
W_{db}	Weighted-average number of double bonds
W_{MUFA}	Monounsaturated fatty acids in weight percentage
W_{PUFA}	Polyunsaturated fatty acids in weight percentage
W_{US}	Represents the total unsaturated FAME content (weight %)
X	Covariance matrix
x_i (t)	i^{th} particle in the t generation
x_i, x_j	Independent variables in coded values.
Y	Measured response (TAG %) variable
Y_a	Actual output,
Yp	Predicted output
$Y_{x/s}$	Yield coefficient

Greek Symbols

μ	Specific growth rate
μF	micro Farad
μg	microgram
μL	microliter
μM	micromolar
β_0, β_i, β_{ij}, β_{ii}	Constant and regression coefficients of the model
Ω	Ohm

Chapter 1

Introduction
and
Review of Literature

1.1 Introduction

As per the UN (United Nations) regulations of 2011, the world has been estimated to touch a population of 8 billion by 2025 and most of the population density shall limit itself to developing countries like India (KPMG 2007). With this fact that the most population will surround itself in developing counties calls to find better, proper, novel and pragmatic approaches to fight the need of this hour. Subsequently, with increasing population, the requirements upsurge and the focus should be directed towards sustainable development, out of which energy production is one of the most crucial factors (Mohebinejad 2011). To maintain a balanced economic growth as well as a decent standard of living, energy is a significant point for mankind. It has been predicted that by the year 2030, the energy demand shall rise by 50% than that of consumed today and of which 45% of this shall be consumed by China and India (International Energy Agency (IEA) 2007, Shahid and Jamal 2011).

The chief priority in the coming years will be providing safe, affordable and clean energy to the living population. To this end, there have been several barriers which the present world is fighting like dependence on high oil import, dependence and depletion of fossil fuels along with negative influences of fossil fuels on the environment. These are the most notable issues faced by mankind today which is compelling us to explore new options. To start with, the transportation sector alone depends on oil (Statistics , Mohebinejad 2011) by 95% and thus finding a replaceable alternative shall be a significant step forward. In this regard, biofuels have gained a significant amount of attention amidst the crisis of energy consumption and depletion. Biofuel is produced by many nations and it reliably proposes a better alternative with several advantages. However, the production of biofuel and its role in sustainable development is a serious

matter of debate as it would directly threaten food security, forests and its existing flora and fauna and encourages water depletion, monocropping and soil degradation (Mohebinejad 2011).

For the definite cause, the above listed issues have forced mankind to explore other sources of alternative energy which shall have a positive impact on the environment, use cheaper feedstocks, deliver clean and safe fuel along with reducing dependency on fossil fuels. Thus, in a developing country like India biofuel can contribute to boost economic growth and offer employments due to the presence of several transportation sector which play a major role and uses fossil fuel and contributes to 23% greenhouse gases related emissions (Lora, Palacio et al. 2011).

Keeping in view of the aforementioned issues developing a sustainable green energy using a robust organism with thorough metabolic knowledge and wide catabolic abilities is the urgent need of the hour.

1.2 Review of Literature

1.2.1 Energy consumption and fossil fuels

In this blind race of industrialization, oxygen/food once considered our energy needs, has dramatically lost its place to 'Fossil Fuels'. Fossil fuels are used throughout the world to power everything from cars to lights to industries. Oil is the world's primary fuel source or transportation, and this possesses a grave concern as between 50 and 60% of the oil produced is used for transport globally, which makes it the second largest emitter of carbon dioxide (CO_2).

In the United States, 28% of the energy is used to move people and goods from one place to another. Personal vehicles, on the other hand, consume more than 60% of the energy used for transportation. The combining effect of overpopulation and industrialization leading to increased needs have given rise to three notable issues, rise in energy prices, environmental health and energy security.

An irreparable environmental damage and climate resulting from surplus energy and transport fuel consumption owing to an expeditious rise in population have led to the research promotions on greenhouse compatible and sustainable biofuels. Biofuels have been in existence than the vehicles have, yet they have been kept on the outskirts by cheap gasoline and diesel. Our street travel, flights, and delivering represent almost a fourth of the world's ozone depleting substance outflows, and transportation today remains intensely reliant on petroleum derivatives. The thought behind biofuel is to supplant conventional fuels with those produced using plant material or different feedstocks that are inexhaustible. The need is so obvious from the fact that the population is expected to rise exponentially surpassing 9 billion by 2050 and so does the energy demand by 84% (Dutta, Daverey et al. 2014, Darda, Papalas et al. 2019).

A clear progression is evident in the development and production of first generation (1G), second generation (2G) and third generation (3G) biofuels. The success of 1G biofuels such as bioethanol and biodiesel were downgraded by their dependence on crucial edible food crops like sugarcane, corn, vegetable oil as feedstock. Thus, 2G biofuels feedstocks were derived from non-consumable or food crop waste biomass (generally lignocelluloses). The advantage of lower procurement costs of 2G feedstocks was unfortunately offset by the higher processing cost, which unfavorably impacted the economic viability of 2G biofuels. 3G algal based biofuels ruled out any competition

with food crops, the resources needed to produce them or subsequent costs of procurement and production (Chaudhary 2013, Castro, Rocha et al. 2016). In addition, the short life cycle of oil rich microalgae enabled multiple harvests; but their high energy requirements and yet unattained industrial level upscaling impacted their potential viability. Furthermore, collecting, dewatering and drying, which are the primary steps involved in biodiesel production from microalgae have proved quite expensive (Chaudhary 2013). These issues have shifted the focus to bacterial lipids such as triacylglycerol (TAG) as a source of biofuel in biotechnological and industrial applications (Castro, Rocha et al. 2016). There are specialized bacteria known to accumulate significant amounts of chemical energy, from organic molecules present in wastewater substantiated by the utility of lipid rich bacterial biomass in the direct production of biofuel (Muller, Sheik et al. 2014). Thus 10% FAME per DCW of sludge had resulted in a production cost of 3.23$ per gallon which at approximately 0.20 $/liter is significantly lower, than the present cost of petroleum-based biodiesel and alternatives (Alvarez and Steinbüchel 2002). Most bacterial communities synthesize polyhydroxyalkanoates (PHA) along with TAG, whereas a few like *Rhodococci* sp. accumulate a very high amount of TAG as their primary storage compound (Alvarez and Steinbüchel 2002).

TAG is the main derivative of energy rich portable biofuel. Bacteria has PHAs granules, which are used for nutrition storage. Similarly, TAGs are stored in the PHA granules by bacteria which are exploited later and refined to biodiesels. Biofuels and related products are mainly extracted from oleaginous organisms. Understanding the genetics of metabolism of the oleaginous organism *Rhodococcus* offers crucial insight into the production of biofuels (Holder, Ulrich et al. 2011). Only by developing a safe and

profitable alternative based on renewable and plentiful feedstock is it inevitable to overcome dependence on crude oil and rising environmental impairment. The production costs still remain one of the challenging issues, which can be reduced if the residues of industrial production processes are used as substrates for oleaginous microorganisms and converted them into valuable co-products. The production of chemicals and power integrated with conversion processes of biomass into biofuels, has been singled out as a way to achieve economically viable production chains, valorize residues and co-products, and reduce industrial waste (Almeida, Fávaro et al. 2012).

1.2.2 Industrial revolution and wastewater

The industrial revolution has recently sped up in developing countries like India and has contributed a lot in releasing industrial waste into the environment. The industrial pollutants present in the wastewater possesses a high amount of threat to the natural habitat. Of many industrial development, milk and dairy products have boomed to a significant level owing to the White Revolution in India and so is the dairy wastewater (DWW). Unpermitted discharges and/or accidental spills of milk and dairy products into a waterway have been known to impact water quality and cause fish kills. When milk is released into a stream or lake, it provides a nutrition source for bacteria, which uses oxygen in the water to live, grow and reproduce. Bacteria growth following a significant release or spill will often cause stress, or death, for fish and other aquatic life living in the impacted waterway. The ecological balance is disturbed by the wastewater discharged into the water bodies and deteriorates the water quality. The casein precipitation from the waste further degrades into highly odorous black sludge. The effluents from milk processing units contain suspended solids, soluble organics and trace

organics that release gases, responsible for taste and odour, impart colour and turbidity and promote eutrophication that majorly affects and disturbs aquatic flora-fauna (Edokpayi;, Odiyo; et al. 2017, Buchanan 2018). White Revolution has considered one of the major reformations India has witnessed.

1.2.2.1 Dairy wastewater

Industrialization is the spine for improvement of a nation. The contamination caused nowadays by these industries is a worldwide worry all through. The food sector has one of the highest water consumptions of all industrial activities and is considered one of the largest effluent producers per unit of production. The dairy sector is one of India's major food sectors, and India positions first among the greatest significant milk producing country. The wastewaters of the dairy industry are generated primarily from the milk processing plant's cleaning and washing operation. The dairy business is one of a significant wellspring of wastewater creating somewhere in the range of 3.739 and 11.217 million m^3 of waste for each year. The subsequent wastewater can contain sanitizers, salts, cleanser, base, and natural issue, contingent on the source (Tikariha and Sahu 2014).

Dairy wastewater with various nutrients, fats, lactose, sanitizing agents and detergents is known to have high biological oxygen demand (BOD) and chemical oxygen demand (COD). Eutrophication of receiving waters is caused by nutrients and detergents affect the aquatic life. The composition of wastewater from milk processing plants makes it exceptionally appropriate for biological wastewater treatment. During its processing and the production of a variety of milk products, about 1-2 m^3 of wastewater is created every day for each ton of milk in the dairy industry. Table 1.1 lists the volumes of wastewater

generated during the production processes in the dairy industry. It is evidenced that more than 90% of it organic contents hold production residues and milk (Quaiser and Gilbert 2016).

Table 1.1 Wastewater volumes for some dairy production processes

Process	**Volume of wastewater**
Production of saleable milk from 1 liter of milk as supplied	0.8 – 1.7 l/h
Production of milk powder from 1liter of milk as supplied	0.8 – 1.5 l/h
Production of 1 Kg of ice cream	2.7 – 4.0 l/h

The disposal of wastewater remains a primary concern and proper measures should be taken for the same. The disposal of wastewater into water bodies like rivers, ponds and land fields with improper treatment can produce serious health and hygiene problems (Tikariha and Sahu 2014).

Although the Indian Government has improvised strict laws of treatment and discharge of dairy wastewater, the proper disposal remains a major concern. Wastewater can be treated based on various methods, it includes primary, secondary and tertiary treatments. To name some of the processes listed under them are aerated lagoons, activate sludge, trickling filters, anaerobic sludge blanket, anaerobic filters etc. DWW is also post-treated with certain physico-chemical methods such as flocculation/coagulation with the aid of several organic and inorganic naturally occurring coagulants, reverse osmosis and nanofiltration (NF). Except for proteins and fats, most of the components in the dairy wastewater are biodegradable, proteins and fats take longer time to degrade. To this end, DWW consisting of a high percentage of organic matter is best treated with the aid of

association of robust microbes. The treatment with the microbes shall only be valuable if it gives rise to economically significant end products. The raw materials used as nutritional substrates can be by-products from industries and the aforementioned dairy contributes to a major part globally. However, this substrate has not been explored well enough for lipid accumulation by *R.opacus* PD630. In specific, bioreactor studies have not been conducted on lipid accumulation along with simultaneous dairy wastewater treatment (Kumar, Gupta et al. 2015).

By establishing a sustainable and competitive alternative based on renewable and abundant feedstock, biodiesels have the potential to reduce the dependence on crude oil and alleviate environmental adversity. In this context, various industrial wastes that have been reported to be rich substrates (Kumar, Gupta et al. 2015) for TAG production are shown in the Table 1.2. The dairy industry is one of the major food industries in most nations (Abdulsalam Tawfeeq Dawood., Arinjay Kumar. et al. 2011). Dairy processing industries employ water as a key processing medium and generate DWW from its operations. The amount of DWW generated is in the range of 0.2–10 liters (average = 2.5 liters) per liter of processed milk. DWW contains lactose which can be exploited as a carbon source (Tikariha and Sahu 2014). Owing to surplus dairy industries, large volumes of DWW generated become major pollutants, which are among the least explored carbon sources in India (Kushwaha, Srivastava et al. 2011).

Table 1.2:Lipid production by microorganisms using different substrates

Microorganism	**Substrate**	**TAG (%w/w of DCW)**	**References**
R. opacus PD630	Molasses	87.0	(Alvarez, Kalscheuer et al. 2000)
R. ruber	Acetate, citrate, valerate, pentadecane, hexadecane	26.0	(Alvarez, Pucci et al. 1997)
R. erythropolis 17	Pentadecane	56.0	(Alvarez, Luftmann et al. 2002)
R. opacus MR 22	Gluconate	48.0	(Alvarez and Steinbüchel 2002)
R. erythropolis	Gluconate, hexadecane	21.0	(Alvarez, Luftmann et al. 2002)
Nocardia asteroids	Glucose, pentadecane, hexadecane, gas-oil	12.2	(Alvarez, Luftmann et al. 2002)
R. opacus PD630	Carob waste	76.0	(Gouda, Omar et al. 2008)
R. opacus PD630	Sugar cane molasses	93.0	(Gouda, Omar et al. 2008)
Gordonia sp. DG	Orange waste	71.0	(Gouda, Omar et al. 2008)
R. opacus	Dairy wastewater	51.0	(Kumar, Gupta et al. 2015)
*R**. sp. YHY01	Glucose	61.7	(Bhatia, Kim et al. 2017)

R. opacus PD630 with its inherent genetic capability of importing lactose and degrading it through galactose makes it an excellent choice in comparison to other *Rhodococci* sp. for utilizing DWW for the production of TAG (Herrero and Alvarez 2016). Despite DWW's abundance and great potential to act as a suitable media for TAG production, it has been rarely used for lipid accumulation by *Rhodococci* sp. A combinatorial approach including utilization of by-products from other food industries would aid in achieving a better route to minimize waste and channel them in the production of superior and sustainable biofuel, which in turn would enhance the economic viability of biofuel.

1.2.3 *Rhodococcus opacus* PD630

One of the primary reserve materials in both prokaryotes and eukaryotes is TAG. They are water-insoluble, nonpolar fatty acid triesters of glycerol. Among the prokaryotic group, a Gram-positive bacterium named *R. opacus* PD630 is popular for accumulation TAG. A high density bioreactor level cultivation of *R. opacus* PD630 was achieved by Voss and Steinbeuchel (Voss, Steinbüchel et al. 2001) where they used sugar beet molasses and sucrose as chief carbon sources and obtained high concentrations of TAGs. This research was an example demonstrating that cheap raw materials like organic wastes and by-products from industrial wastes could act as a potential medium for lipid production. Furthermore, Gouda et.al (Gouda, Omar et al. 2008) used carob, orange waste, and sugarcane molasses to produce TAG from *R. opacus.* Therefore, using cheap raw materials for the production of TAG could be applied as a promising aspect in the biotechnological production of lipids.

The TAG accumulation is a stress response or synthesized in the stationary growth phase maintained by Rhodococcus and other similar members of the actinomycetes group. Although the nutritional stress is the primary inducer of TAG biosynthesis, other factors like the strain involved and carbon source it consumes also have an impact on the total content of TAGs accumulated (Abdulsalam Tawfeeq Dawood., Arinjay Kumar. et al. 2011).

Amongst other oleaginous microbes, *R. opacus* PD630 is known to be the most superior one which can convert 87% of its cellular dry weight to lipid. Thus, it is considered model organism and has the highest lipid storage compared to other oil accumulating microorganisms. The biosynthesis of lipids is a complex pathway that involves 22

reactions catalyzed by 457 enzymes. The TAG accumulation in *R. opacus* is balanced between the synthesis of TAG and degradation of the same (Chen, Ding et al. 2014). To this end, Actinomycetales, on the other hand, has several lipid synthesis pathways and take aid of multifunctional fatty acid synthase (FAS). In the TAG reaction cycle, there are prominently 20 distinct enzymes working, all starting from Acetyl-CoA and Acetyl-CoA functions as a key substrate in TAG biosynthesis. Basically, Acetyl-CoA appears as the final product in the in several catabolic reactions. Acetyl-CoA carboxylases work as the first step in the TAG biosynthesis and function as α/β complexes in *Rhodococcus* to giving rise to Malonyl-CoA which in turn participates into fatty acid synthesis pathway (Holder, Ulrich et al. 2011). Therefore, recognizing the chief enzymes essential for TAG accumulation and suppressing the non-essential ones shall drive the most flux towards TAG biosynthesis (Chen, Ding et al. 2014).

1.2.4 Improving TAG yields through system biology approach

A vast amount of research has been focussed on producing ecological and environment friendly fuel. Consequently, microbial oil production using oleaginous microorganisms has received further acclamations, however, it still receives several difficulties due to the usage of expensive feedstock as raw materials or low lipid yield (Herrero and Alvarez 2016).

TAGs, a biodiesel precursor, are the most frequent storage lipids in higher eukaryotes and also in eukaryotic micro-organisms such as yeast, molds, and algae (Alvarez, Alvarez et al. 2008). Also, the accumulation of large amounts of TAGs in bacteria has been frequently reported in strains belonging to the order Actinomycetales with a special mention in *R. opacus* PD630. It is one of the best-studied bacteria regarding biosynthesis

and accumulation of lipids with the ability to accumulate large amounts of TAGs of 76% - 87 % of the cellular dry weight under certain culture conditions (Alvarez, Mayer et al. 1996). The huge catabolic repertoire of this microorganism, in addition to its capability to adapt its metabolism to a wide range of nutritional conditions, make it a promising candidate for both biofuel production and bioremediation of polluted environments (Villalba, Hernández et al. 2013, Amara, Seghezzi et al. 2016). Although some significant advances in the basic knowledge of TAG metabolism in rhodococci have been made, in order to further explore the characteristics of this important industrial microorganism, a comprehensive understanding of its metabolism is essential.

Systems biology approach can be exploited to predict the way in which microorganisms consume and process their metabolic developments. Constraint based analysis of genome scale metabolic networks can be used to determine such processes (Alvarez, Mayer et al. 1996). With the help of experimental data, systems biology can conveniently build and validate *in-silico* models. In this regard, to effectively predict cellular behaviour GeMs were reconstructed to produce metabolic models (Alvarez, Mayer et al. 1996, Villalba, Hernández et al. 2013). To define model reconstruction, it is the cellular metabolic network wherein the genes, transcripts and proteins are working in association to give rise to desirable products (Alvarez, Mayer et al. 1996).

In this thesis, a genome-scale stoichiometric model of *R. opacus* PD630, was reconstructed based on its genome annotation and data from the literature and biochemical databases. The model accounts for nearly all the metabolic functions and pathways involved in TAGs, a biodiesel precursor production of *R. opacus* PD630 and can be helpful for further physiological studies and explorations.

1.2.4.1 Reconstruction of genome-scale metabolic model

Interpreting the basic functioning of biosynthetic circuits and microbial metabolic processes has been enhanced by the booming of genomics which may automatically derive the apparent metabolic construct of an organism, presented its genome. The metabolic network is based on the array of biochemical reactions catalysed by the enzymes found in an annotated genome sequence. This may contribute greatly in the case of industrially and medically noteworthy strains since their metabolism is a key factor that can be exploited to synthesize significant products and understand infection physiology. Bacteria has also been a common interest to researchers in terms of delivering usable substances of the microbial origin to the current world which includes biofuels, bioactive compounds and pharmaceuticals (Beloqui, de María et al. 2008, Zou, Liu et al. 2012). To this end, *in-silico* modelling and simulations have emerged out as a powerful tool to simulate chemical reaction fluxes within the whole microbial metabolism.

The constraint-based method is amongst the most conceivable approaches as they only require the metabolic reaction stoichiometry with mass balance around the substrates under pseudo-steady state assumption to build the tentative metabolic network of an organism (Varma and Palsson 1994, Oberhardt, Chavali et al. 2009, Oberhardt, Palsson et al. 2009, Orth, Thiele et al. 2010). Drafting the list of biomass assembly reactions, exchange fluxes and biochemical reactions with the environment are exploited to transform a cellular metabolic network into a model with the aid of computational approaches. The reconstructions of a particular strain are accounted for roles of all the functional genes which are incorporated in a standardized format, enabling *in-silico* model construction which can be further analysed with a variety of emerging

mathematical tools. The prime significance of constraint-based modelling is that it can automatically simulate to predict the resulting balance of all the chemical reactions and bridges the gap between theoretical information of the metabolic structure and observed metabolic phenotypes (Ramirez, Lynes et al. 2017, Mishra, Lee et al. 2018).

Though there has been the availability of various advanced tools for drafting a metabolic network, transforming them into a model that can clearly represent functional states of a given organism remains to be a hectic task. Most of the computational tools are not competent enough due to lacking or missing the metabolic steps for sustaining metabolic gap, stoichiometry, directionality and charge. Combining reliability and speed are difficult to incorporate into a model as they are highly time consuming and demand manual effort. To overcome these problems, Thiele and Palsson (Thiele and Palsson 2010) introduced a model that includes 4 stages and 94 steps required for a high quality and reliable metabolic model building. Speeding up the steps has clearly accelerated the overall process of network construction and constraint-based metabolic modelling.

1.2.4.2 Constraint-based metabolic modelling

Constraint-based modelling strategies, for example, Flux Balance Analysis (FBA), have been broadly used to unravel complex, data rich omics datasets to produce cellular metabolism behaviour patterns across the process. FBA has been effectively used to pick up understanding in a wide scope of uses, for example, the scope of substrate use, product yields and the development of metabolic engineering systems to improve bioprocess execution. Lack of knowledge in enzyme's kinetic parameters and concentration hinder beneficial application for several mathematical modelling methods. FBA does not necessitate to have a thorough details of concentration and kinetic parameters of the

enzymes and co-factors involved and therefore, it is an alternative approach to deal with accurate kinetics and concentration knowledge (Fell and Small 1986, Varma and Palsson 1994, Bonarius, Schmid et al. 1997, Pramanik, Keasling et al. 1997). Thus, by using a linear programming method along with allowable flux distribution an optimal metabolic flux distribution can be easily be determined (Chvátal 1983, Varma and Palsson 1994). An objective function, like microbial cellular growth, is essential to identify the optimal solution. Whether or not these flux distributions reveal the actual metabolic behaviour is still debatable. As for the microbial cells, the objective function is the maximal growth rate which would on the hypothesis would be agreed upon if the predicted and actual fluxes are comparable.

Hereby, the equation is mentioned which explains the mass balance constraints or stoichiometric:

$$S \bullet v = 0$$

In this equation, S represents the stoichiometric matrix and v is the vector that defines the metabolic flux which includes both internal and external fluxes. To identify the solution which would maximize an objective function, a metabolic flux distribution was expressed with the aid of a linear programming program. The formulated linear programming problem is mentioned below:

Minimize Z subject to:

$$S \bullet v = 0$$
$$\alpha i \leq vi \leq \beta i$$

where

$$Z = \sum_{i=0}^{n} ci.\mathrm{v}i = < c \bullet \mathrm{v} >$$

In this equation, αi and βi represent the lower and upper bound of the flux through reaction i, c represents vector which was used to select linear combination of metabolic fluxes to be included at the objective function (Varma and Palsson 1993). For this specific reaction, c was the cellular growth which was defined as the unit vector in the growth flux direction (Edwards, Ibarra et al. 2001). An acclaimed challenge related to enormous genome-scale metabolic systems is their underdetermined nature. Therefore, instead of a unique solution, FBA and similar approaches analyse range of flux values, non-unique solutions, that are in agreement with the observed physiological conditions. The more extensive the recorded flux values, the higher the vulnerability in the assurance of essential reaction properties, constraining the interpretability of and trust in the outcomes. To circumvent this issue and quantity the solution space one can perform the flux variability analysis (FVA). It enables the minimum and maximum permissible fluxes to be identified under the constraints imposed by each reaction in the reconstructed metabolic network.

There has been neither availability nor validation and analysis of the *Rhodococcus opacus* PD630 GeM despite its existence reported by (Holder, Ulrich et al. 2011). Herein we reconstructed and validated the GeM *i*ROPD630 right from the scratch to statistically analyse under different constraints. The GeM could as well be further used for strain improvement in the biofuel production process *in-silico* using various algorithms.

1.2.5 Sustainable *de novo* biosynthesis of biodiesel by oleaginous *Rhodococcus opacus* PD630 and its metabolic flux analysis

Biodiesel has numerous positive effects when the environment and ecology is concerned. However, it comes with certain drawbacks and limitations too, as follows: (a) as biodiesel totally depends on natural resources for its production like vegetable oil feedstocks, primarily rapeseeds from Europe, palm oil of south Asia and soybean from North America. Collectively, this makes biodiesel season restricted and geographically limited to areas with availability of required raw materials (Darda, Papalas et al. 2019). Though vegetable oils consist of a significant amount of TAGs, it cannot be taken into account due to oil's viscosity which will render problems when it comes to handling the material into huge quantity and to filter plugging at low temperature due to formation of crystals. Therefore, to obtain FAME and FAEE elements of biodiesel, the oil has to transesterified with the aid of short chain alcohols namely, methanol or ethanol. Furthermore, to obtain a transesterified final product it involves huge money and is energy consuming which eventually makes it economically undesirable and reduces possible energy yield as well (Chaudhary 2013). Although FAME and FAEE have comparable physiological properties in engine performances (Peterson, Reece et al. 1995), only FAME in more taken into consideration during commercial production due to its cost effectiveness. FAME uses methanol which is currently produced from natural gas and cheaper than ethanol. To this end, FAME based biodiesel cannot be fully considered as a renewable product due to usage to alcohol element which is derived from fossil origin. Alongside, methanol is an extremely toxic and hazardous material and it requires careful handling and precautions. In this regard, bioethanol can be used to replace ethanol in the case of FAEE which may aid in industrial production. However, producing bioethanol is itself an expensive affair in itself (Castro, Rocha et al. 2016).

The need for huge amount of land for the cultivation of vegetable feedstocks significantly limits the use of biodiesel. The rapeseed produces only 1300 l ha^{-1} biodiesel, as in this case only the seed oil is used, and the rest entire plant goes into waste. Difficulties come with the self-incompatibility of rapeseed and oilseed where their cultivation requires crop-rotation regime. Thus, in conclusion, in the future, vegetable feedstock-based biodiesel will not be able to meet and substitute more than 5-15% of petroleum-based diesel. In view of the aforementioned issues and to avoid usage of hazardous chemicals like methanol and sulfuric acid/sodium hydroxide here we report the construction of a genetically engineered, natural high yielding TAGs and fatty-acids strain, *R. opacus* PD630. Heterologous expression of the pdc gene, from the ethanol producing pathway of *Z. mobilis,* was performed in the above organism for *de novo* biosynthesis of FAEEs in the microenvironment of the cells. This novel approach might fulfil all the demands for the industrial production of biodiesel equivalents from recyclable resources as this organism is known to catabolize and thrive on variety of industrial wastes. In conclusion, this study come up with a platform to achieve more economical production process, and therefore a vital replacement of non-renewable fossil fuels by renewable biofuels in the future.

To better understand the physiology and metabolism of microdiesel producing *R. opacus* PD630, the reconstructed GeM of the organism, *i*ROPD630, is employed. Though FBA is undoubtedly the most used constraint-based approach for metabolic flux analysis (MFA), it fails to address the issue of yielding the same optimum in the complex underdetermined GeM network (Lee, Gianchandani et al. 2006, Orth, Fleming et al. 2010, Soh, Miskovic et al. 2012, Lularevic, Racher et al. 2019). FVA that, under a specific physiological constraint, measures a range of allowable flux values of a

particular reaction that can be used instead to study MFA (Mahadevan and Schilling 2003). The prediction however becomes uncertain with increase in size and underdetermine state of the GeM. With a network as big as GeM the flux values likely to be overestimated resulting in futile cycles having farfetched reaction flux. This allows the network to generate metabolic resources out of nothing (Price, Famili et al. 2002).

According to (Lularevic, Racher et al. 2019) a computationally effective carbon-constrain FVA (ccFVA) that depends on current experimental data or theories and can methodically decrease the flux variability to a significant physiological attribute can be used to refine the flux values obtained through FVA.

The strategy is based on the essential parity of carbon all through the system and guarantees that the recorded flux values are compatible with the measure of carbon atoms being consumed or discharged (Lularevic, Racher et al. 2019). This approach drastically reduces the solution space to help in finding the true optimum of the objective function.

1.3 Motivation and objectives

1.3.1 Definition of problem

With an economically rising nation like India, it faces challenges to meet food and fuel demands on a regular basis. This asks for the sustainable usage of all materials in the most desirable manner besides producing the required commodities through sustainable development.

In today's growing economy the responsibility is just not to produce sufficient sustainable green energy but also to make sure the commercial use of such energy products has no drastic consequence on the environment. To start with, the proposed

renewable resource should come with certain qualities namely: low production cost, easy production in huge volume and high energy content. If they fail to adhere to these four qualities, then the economic feasibility of the practice of such green fuels shall be debatable. Additionally, a check on the effect of usage of these sustainable resources on the environment is also important.

There have been several researches on choosing the best source for green fuel production and besides algae, several other organisms have come up to be superiorly producing biofuel. However, out of them the which provides maximum yield remains still a doubt. With 3G biofuels from microalgae using renewable feedstocks coming close to solving the ever-rising fossil fuel demands, their high energy requirements and yet unattained industrial level upscaling impacted their potential viability. *R. opacus* PD630, on the other hand, is known to thrive on cheap industrial wastes and wastewater to produce high amounts of bacterial lipids, TAG. This lipid is a biodiesel precursor that is transesterified to either FAMEs or FAEEs to be used as a potential biofuel.

With the recent submission of the organism's genome and lack of advanced genetic tools, there lies a potential gap that needs to be filled in the field of biofuel research involving *R. opacus* PD630. Also, the scarce information available for dairy wastewater being used by oleaginous organisms for biodiesel makes it a promising renewable feedstock and a required research domain. The present thesis work is aimed to address the gaps pertaining to the issues mentioned and tried to deliver on the solutions, however small, in that front. In this regard, the following objectives are formulated.

1.3.2 Objectives

1. Construction of a novel variant *Rhodococcus opacus* BM985 to achieve enhanced triacylglycerol-a biodiesel precursor, using synthetic dairy wastewater
2. Bioprocess development for simultaneous dairy wastewater treatment and biodiesel production using BM985
3. Reconstruction and validation of a genome scale metabolic model of *Rhodococcus opacus* PD630 for better understanding TAGs biosynthesis
4. Sustainable *de-novo* biosynthesis of biodiesel by oleaginous *Rhodococcus opacus* PD630 and its metabolic flux analysis

1.4 References

1. Abdulsalam Tawfeeq Dawood., Arinjay Kumar. and S. S. Sambi. (2011). "Study on Anaerobic Treatment of Synthetic Milk Wastewater under Variable Experimental Conditions " International Journal of Environmental Science and Development, **2**(1).
2. Almeida, J. R. M., L. C. L. Fávaro and B. F. Quirino (2012). "Biodiesel biorefinery: opportunities and challenges for microbial production of fuels and chemicals from glycerol waste." Biotechnology for Biofuels **5**(1): 1-16.
3. Alvarez, A. F., H. M. Alvarez, R. Kalscheuer, M. Wältermann and A. J. M. Steinbüchel (2008). "Cloning and characterization of a gene involved in triacylglycerol biosynthesis and identification of additional homologous genes in the oleaginous bacterium Rhodococcus opacus PD630." **154**(8): 2327-2335.
4. Alvarez, H., H. Luftmann, R. Silva, A. C Cesari, A. Viale, M. Wältermann and A. Steinbüchel (2002). Identification of phenyldecanoic acid as a constituent of triacylglycerols and wax ester produced by Rhodococcus opacus PD630.
5. Alvarez, H. and A. Steinbüchel (2002). "Triacylglycerols in prokaryotic microorganisms." Applied Microbiology and Biotechnology **60**(4): 367-376.
6. Alvarez, H. M., R. Kalscheuer and A. Steinbüchel (2000). "Accumulation and mobilization of storage lipids by Rhodococcus opacus PD630 and Rhodococcus ruber NCIMB 40126." Applied Microbiology and Biotechnology **54**(2): 218-223.
7. Alvarez, H. M., F. Mayer, D. Fabritius and A. Steinbüchel (1996). "Formation of intracytoplasmic lipid inclusions by Rhodococcus opacus strain PD630." Archives of Microbiology **165**.
8. Alvarez, H. M., O. H. Pucci and A. Steinbüchel (1997). "Lipid storage compounds in marine bacteria." Applied Microbiology and Biotechnology **47**(2): 132-139.
9. Amara, S., N. Seghezzi, H. Otani, C. Diaz-Salazar, J. Liu and L. D. Eltis (2016). "Characterization of key triacylglycerol biosynthesis processes in rhodococci." Scientific Reports **6**: 24985.
10. Beloqui, A., P. D. de María, P. N. Golyshin and M. J. C. o. i. m. Ferrer (2008). "Recent trends in industrial microbiology." **11**(3): 240-248.
11. Bhatia, S. K., J. Kim, H.-S. Song, H. J. Kim, J.-M. Jeon, G. Sathiyanarayanan, J.-J. Yoon, K. Park, Y.-G. Kim and Y.-H. Yang (2017). "Microbial biodiesel production from oil palm biomass hydrolysate using marine Rhodococcus sp. YHY01." Bioresource Technology **233**: 99-109.
12. Bonarius, H. P., G. Schmid and J. J. T. i. B. Tramper (1997). "Flux analysis of underdetermined metabolic networks: the quest for the missing constraints." **15**(8): 308-314.

13. Buchanan, R. A. (2018). "History of technology." The Industrial Revolution (1750–1900) Retrieved 27th September, 2019, from https://www.britannica.com/technology/history-of-technology.

14. Castro, A. R., I. Rocha, M. M. Alves and M. A. Pereira (2016). "Rhodococcus opacus B4: a promising bacterium for production of biofuels and biobased chemicals." AMB Express **6**(1): 35.

15. Chaudhary, N. S. (2013). Microalgae in Biofuel Production-Current Status and Future Prospects. Biofuels Production, John Wiley & Sons, Inc.**:** 167-209.

16. Chen, Y., Y. Ding, L. Yang, J. Yu, G. Liu, X. Wang, S. Zhang, D. Yu, L. Song, H. Zhang, C. Zhang, L. Huo, C. Huo, Y. Wang, Y. Du, H. Zhang, P. Zhang, H. Na, S. Xu, Y. Zhu, Z. Xie, T. He, Y. Zhang, G. Wang, Z. Fan, F. Yang, H. Liu, X. Wang, X. Zhang, M. Q. Zhang, Y. Li, A. Steinbüchel, T. Fujimoto, S. Cichello, J. Yu and P. Liu (2014). "Integrated omics study delineates the dynamics of lipid droplets in Rhodococcus opacus PD630." Nucleic Acids Research **42**(2): 1052-1064.

17. Chvátal, V. J. N. Y. (1983). "Linear Programming WH Freeman and Company." 13-26.

18. Darda, S., T. Papalas and A. Zabaniotou (2019). "Biofuels journey in Europe: Currently the way to low carbon economy sustainability is still a challenge." Journal of Cleaner Production **208**: 575-588.

19. Dutta, K., A. Daverey and J.-G. Lin (2014). "Evolution retrospective for alternative fuels: First to fourth generation." Renewable Energy **69**: 114-122.

20. Edokpayi;, J. N., J. O. Odiyo; and O. S. Durowoju (2017). Impact of Wastewater on Surface Water Quality in Developing Countries: A Case Study of South Africa. Water Quality. H. Tutu. South Africa, IntechOpen.

21. Edwards, J. S., R. U. Ibarra and B. O. J. N. b. Palsson (2001). "*In-silico* predictions of Escherichia coli metabolic capabilities are consistent with experimental data." **19**(2): 125.

22. Fell, D. A. and J. R. J. B. J. Small (1986). "Fat synthesis in adipose tissue. An examination of stoichiometric constraints." **238**(3): 781.

23. Gouda, M. K., S. H. Omar and L. M. Aouad (2008). "Single cell oil production by Gordonia sp. DG using agro-industrial wastes." World Journal of Microbiology and Biotechnology **24**(9): 1703.

24. Gouda, M. K., S. H. Omar, L. M. J. W. J. o. M. Aouad and Biotechnology (2008). "Single cell oil production by Gordonia sp. DG using agro-industrial wastes." **24**(9): 1703.

25. Herrero, O. M. and H. M. Alvarez (2016). "Whey as a renewable source for lipid production by Rhodococcus strains: Physiology and genomics of lactose and galactose utilization." European Journal of Lipid Science and Technology **118**(2): 262-272.

26. Holder, J. W., J. C. Ulrich, A. C. DeBono, P. A. Godfrey, C. A. Desjardins, J. Zucker, Q. Zeng, A. L. Leach, I. Ghiviriga and C. J. P. g. Dancel (2011). "Comparative and functional genomics of Rhodococcus opacus PD630 for biofuels development." **7**(9): e1002219.

27. Holder, J. W., J. C. Ulrich, A. C. DeBono, P. A. Godfrey, C. A. Desjardins, J. Zucker, Q. Zeng, A. L. B. Leach, I. Ghiviriga, C. Dancel, T. Abeel, D. Gevers, C. D. Kodira, B. Desany, J. P. Affourtit, B. W. Birren and A. J. Sinskey (2011). "Comparative and Functional Genomics of Rhodococcus opacus PD630 for Biofuels Development." PLoS Genetics **7**(9): e1002219.

28. International Energy Agency (IEA). (2007). "World Energy Outlook 2007." Retrieved 2nd February, 2012, from http://www.iea.org/textbase/nppdf/free/2007/weo_2007.pdf.

29. KPMG, I. (2007). The Indian sugar industry sector roadmap 2017, KPMG.

30. Kumar, S., N. Gupta and K. Pakshirajan (2015). "Simultaneous lipid production and dairy wastewater treatment using Rhodococcus opacus in a batch bioreactor for potential biodiesel application." Journal of Environmental Chemical Engineering **3**(3): 1630-1636.

31. Kushwaha, J. P., V. C. Srivastava and I. D. Mall (2011). "An Overview of Various Technologies for the Treatment of Dairy Wastewaters." Critical Reviews in Food Science and Nutrition **51**(5): 442-452.

32. Lee, J. M., E. P. Gianchandani and J. A. Papin (2006). "Flux balance analysis in the era of metabolomics." Briefings in Bioinformatics **7**(2): 140-150.

33. Lora, E. E. S., J. C. E. Palacio, M. H. Rocha, M. L. G. Renó, O. J. Venturini and O. A. J. E. Del Olmo (2011). "Issues to consider, existing tools and constraints in biofuels sustainability assessments." **36**(4): 2097-2110.

34. Lularevic, M., A. J. Racher, C. Jaques and A. Kiparissides (2019). "Improving the accuracy of flux balance analysis through the implementation of carbon availability constraints for intracellular reactions." **116**(9): 2339-2352.

35. Mahadevan, R. and C. J. M. e. Schilling (2003). "The effects of alternate optimal solutions in constraint-based genome-scale metabolic models." **5**(4): 264-276.

36. Mishra, P., N.-R. Lee, M. Lakshmanan, M. Kim, B.-G. Kim and D.-Y. J. B. s. b. Lee (2018). "Genome-scale model-driven strain design for dicarboxylic acid production in Yarrowia lipolytica." **12**(2): 12.

37. Mohebinejad, S. (2011). Biofuel Production in Developing Countries: A key to Sustainable Development? . MA Master, University of Vienna.

38. Muller, E. E. L., A. R. Sheik and P. Wilmes (2014). "Lipid-based biofuel production from wastewater." Current Opinion in Biotechnology **30**: 9-16.

39. Oberhardt, M. A., A. K. Chavali and J. A. Papin (2009). Flux balance analysis: interrogating genome-scale metabolic networks. Systems Biology, Springer**:** 61-80.

40. Oberhardt, M. A., B. Ø. Palsson and J. A. J. M. s. b. Papin (2009). "Applications of genome-scale metabolic reconstructions." **5**(1).

41. Orth, J., R. Fleming and B. Palsson (2010). "Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide."

42. Orth, J. D., I. Thiele and B. Ø. J. N. b. Palsson (2010). "What is flux balance analysis?" **28**(3): 245.

43. Peterson, C., D. Reece, B. Hammond, J. Thompson and S. Beck (1995). Commercialization of Idaho biodiesel (HySEE) from ethanol and waste vegetable oil. Annual ASAE Meeting. Paper.

44. Pramanik, J., J. J. B. Keasling and bioengineering (1997). "Stoichiometric model of Escherichia coli metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements." **56**(4): 398-421.

45. Price, N. D., I. Famili, D. A. Beard and B. Ø. Palsson (2002). "Extreme Pathways and Kirchhoff's Second Law." Biophysical Journal **83**(5): 2879-2882.

46. Quaiser, J. and E. Gilbert (2016). Wastewater treatment in the dairy processing industry - recovering energy using anaerobic technology.

47. Ramirez, A. K., M. D. Lynes, F. Shamsi, R. Xue, Y.-H. Tseng, C. R. Kahn, S. Kasif and J. M. J. C. r. Dreyfuss (2017). "Integrating extracellular flux measurements and genome-scale modeling reveals differences between Brown and white adipocytes." **21**(11): 3040-3048.

48. Shahid, E. M. and J. Jamal (2011). "Production of biodiesel: A technical review." Renewable and Sustainable Energy Reviews **15**(9): 4732-3745.

49. Soh, K. C., L. Miskovic and V. Hatzimanikatis (2012). "From network models to network responses: integration of thermodynamic and kinetic properties of yeast genome-scale metabolic networks." FEMS Yeast Research **12**(2): 129-143.

50. Statistics, I. Key World Energy Statistics. Paris. International Energy Agency, 2014. 82 p.

51. Thiele, I. and B. Ø. J. N. p. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." **5**(1): 93.

52. Tikariha, A. and O. Sahu (2014). "Study of Characteristics and Treatments of Dairy Industry Waste Water." Journal of Applied & Environmental Microbiology **2**(1): 16-22.

53. Varma, A. and B. O. J. A. E. M. Palsson (1994). "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110." **60**(10): 3724-3731.

54. Varma, A. and B. O. J. B. t. Palsson (1994). "Metabolic flux balancing: basic concepts, scientific and practical use." **12**(10): 994.

55. Varma, A. and B. O. J. J. o. T. B. Palsson (1993). "Metabolic capabilities of Escherichia coli II. Optimal growth patterns." **165**(4): 503-522.

56. Villalba, M. S., M. A. Hernández, R. A. Silva and H. M. J. J. o. M. B. Alvarez (2013). "Genome sequences of triacylglycerol metabolism in Rhodococcus as a platform for comparative genomics." **2**(2).

57. Voss, I., A. J. A. m. Steinbüchel and biotechnology (2001). "High cell density cultivation of Rhodococcus opacus for lipid production at a pilot-plant scale." **55**(5): 547-555.

58. Zou, W., L. Liu, J. Zhang, H. Yang, M. Zhou, Q. Hua and J. J. J. o. b. Chen (2012). "Reconstruction and analysis of a genome-scale metabolic model of the vitamin C producing industrial strain Ketogulonicigenium vulgare WSH-001." **161**(1): 42-48.

Chapter 2

Construction of a novel variant *Rhodococcus opacus* BM985 to achieve enhanced triacylglycerol-a biodiesel precursor, using synthetic dairy wastewater

2.1 Outline

Biofuel from bacterial lipids like TAG is known to alleviate difficulties encountered in the production of first (1G), second (2G) and third generation (3G) biofuels. *Rhodococcus opacus* converts glucose into TAG, a biodiesel precursor, by using intermediates in the Kennedy pathway and specific (phosphatidic acid phosphatase type 2) *PAP2* enzymes. The elucidation of genes involved, their functions and operational parameters is incomplete. Consequently, we have constructed a *PAP2* over-expressing improved variant *Rhodococcus opacus* BM985 (BM985) from wild type (WT) *R. opacus* PD630, for maximizing biodiesel production using synthetic dairy wastewater (SDWW).

2.2 Materials and methods

Genetic engineering approach for enhanced lipid production

2.2.1 *Rhodococcus opacus* PD630 and Vector pNit-QT1

Rhodococcus opacus PD630 (DSM 44193) was obtained from the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ, Germany). The *E. coli-Rhodococcus* shuttle vector pNit-QT1 was graciously donated by Dr. Tomohiro Tamura (National Institute of Advanced Industrial Science and Technology, Japan) and used to generate the clones (see Figure 2.1).

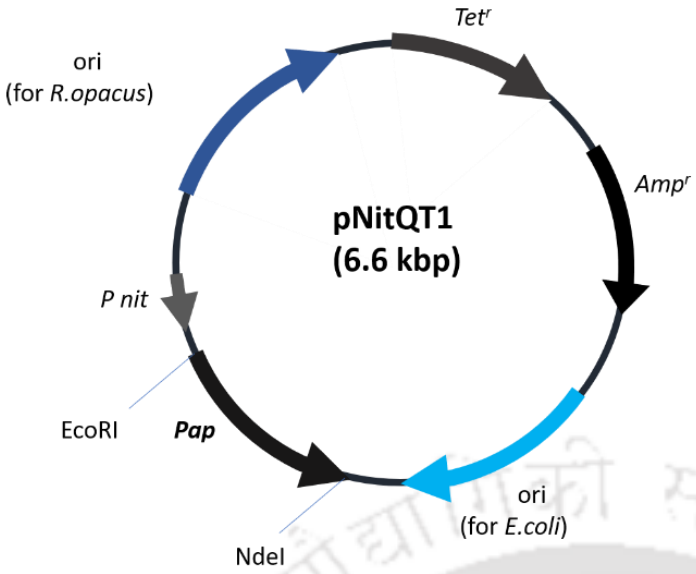

Figure 2.1:Shuttle vector pNit-QT1 with the gene of interest Pap

2.2.2 Amplification and cloning of *PAP2* gene

Genomic DNA of the WT was isolated by using GeneJET Genomic DNA Purification Kit (Thermo Fisher Scientific, Waltham, MA, USA) according to the manufacturer's protocol. *PAP2* gene was PCR amplified (with 32 cycles) using 0.025 units/µL Phusion high fidelity DNA polymerase (NEB); 1X polymerase buffer (supplemented with Mg^{2+}), 0.5 µM P1 and P2 primers (Invitrogen, Grand Island, NY) and 200 µM dNTPs (NEB). The sequence of gene's primers with incorporated restriction sites used for amplification and PCR conditions are listed in Table 2.1 and Table 2.2 respectively. Both, the PCR products *PAP2* gene and the vector pNit-QT1 were subjected to electrophoresis and extracted using QIA quick gel extraction kit (Qiagen). *PAP2* gene insertion into the *E. coli-Rhodococcus* shuttle vector pNit-QT1 using restriction digestion (*NdeI* and *EcoRI*) was followed by ligation. All the restriction endonucleases were purchased from NEB (MA). The vector with *PAP2* gene insert was transformed into competent *Escherichia coli (E.coli)* TOP10 cells (prepared by $CaCl_2$ method) by heat shock procedure. In this process, the cells are treated with high concentrated, ice cold calcium chloride. The

calcium ion which is positively charged attracts both DNA backbone (negatively charged) and the cell's lipopolysaccharide (negatively charged) to each other and on heat shock the cells take the DNA inside. (Schäfer, Tauch et al. 1994). The screened colonies of *E. coli* TOP10 with plasmid were grown on 100 µg/ml ampicillin LB plate over-night. Plasmid DNA was extracted using QIA prep miniprep kit (Qiagen, Valencia, CA) followed by confirmation of the clone by sequencing.

Table 2.1:Primer sequences and gene used in this study. They are aligned in 5' – 3' directions

Primer	Target gene	Sequence (5`-3`)	Cohesive end
*PAP2*_FP	*PAP2*	TAGCATATGATGCCCCACACCTCTGCCGCT	*NdeI*
*PAP2*_RP	*PAP2*	GTACGAATTCTCAGCCTCCCACTCGGTCGA	*EcoRI*

Table 2.2:PCR programme and conditions for amplifying PAP2 gene from bacterial genome

Step	Temperature	Time
Initial denaturation	94 °C	1 min
Denaturation	94 °C	30 sec
Primers annealing	65 °C	1 min
Extension	72 °C	50 sec
Final extension	72 °C	2 min
Hold	4 °C	∞

2.2.3 Transformation and Over-Expression of *PAP2* gene in WT/*R. opacus* PD630

The confirmed plasmid DNA with *PAP2* gene insert was introduced and overexpressed into *R. opacus* PD630 cells by electroporation on Electro Cell Manipulator 620 (BTX Inc., San Diego, CA, USA) using steps described by Na et. al (Na, Nagayasu et al. 2005). Electrocompetent cells were obtained by culturing of 0.1 mL of *R. opacus* PD630 (overnight TSB pre-culture) inoculum on 10 mL of Tryptic soy broth (TSB), supplemented with 0.5% (w/v) glycine in a 50-mL screw-capped vial at 28 °C for 24 hrs. Harvested cells were washed twice with ice-cold HS buffer containing 7 mM HEPES and

252 mM sucrose (pH 7.0) concentrated 10-fold in ice-cold HS buffer. About 400 µL of competent cells were pre-incubated at 40 ^{o}C for 10 min and mixed with plasmid DNA (final concentration 0.1–1 µg/mL) immediately before the electroporation. Electro-cuvettes (BTX Inc.) were used to perform electroporation with gaps of 2 mm, 6.5 kV/cm, 725 Ω, and 50 µF. Pulsed cells were immediately diluted with 4 mL of TSB and regenerated at 28 ^{o}C for 24 hrs before they were plated on TSB agar plates supplemented with 10-15 µg/mL of tetracycline (Berni, Dorileo et al. 2013). The clones of *R. opacus* PD630 overexpressing *PAP2* gene were generated and the better TAG yielding clone was selected.

A loopful of culture was inoculated into 100 mL of LB media with 10-15 ug/mL tetracycline for 3 days in shaking at 200 rev min^{-1} at 30 °C. The vector pNit-QT1 has a constitutive promoter which stays induced at all the time. The culture was then harvested at 13,000 rpm, 10 mins. The cells were collected and analyzed for the over-expression of *PAP2* gene. The SDS PAGE was run to ensure the over-expression of *PAP2* gene. PAP2 is native to WT, a control (without the plasmid) and BM895 (pNit-QT1 with *PAP2* gene) was expressed for 3 days and evaluated on SDS PAGE to observe an improved *PAP2* gene expression in BM985.

2.2.4 Growing WT and BM985 for comparing their lipid contents

WT and BM985 cultures were maintained on Luria-Bertani (LB) agar containing 10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl. Initial experiments were carried out in 250 mL Erlenmeyer flask containing 50 mL of production medium, mineral salt medium (MSM) containing 9 g/L $Na_2HPO_4 \cdot 12H_2O$, 1.5 g/L KH_2PO_4, 0.1 g/L NH_4SO_4, 0.2 g/L $MgSO_4 \cdot 7H_2O$, 0.02 g/L $CaCl_2 \cdot 2H_2O$, 0.05 g/L $NaHCO_3$ and 1.2 mg/L ferric ammonium citrate.

A loopful of cells from a single colony grown on LB agar at 30 ºC for 3 days was inoculated in 50 mL of LB medium for preparing the seed culture. The culture was then incubated on a rotary shaker (200 rev min^{-1}) at 30 ºC for 24 hrs until it reached an optical density of approximately 0.6 at 660 nm. The production medium having different concentrations of SDWW and MSM in the ratio 2:1 was inoculated with seed culture to reach an initial OD_{660} of about 0.03 and incubated on a rotatory shaker (200 rev min^{-1}) at 30 ºC. The sampling was done every 6 to 12 hrs for a total of 72 hrs batch time to estimate the growth (dry cell weight, DCW of biomass) and lipid profile (TAG%, W/DWC) of cells.

SDWW was prepared as reported in (Leal, Freire et al. 2006), using commercially available dried milk powder (2 g/L) and fat (0.2 g/L). The major constituents of milk powder (Amulya™, Amul, India) and fat (Ghee, Britannia, India) were used are listed respectively in Table 2.3 and Table 2.4. The synthetic wastewater had an initial pH 6.0 and an average COD content of 2,800 mg/L.

Table 2.3:Composition of the commercially available milk powder (Amulya™, Amul, India)

Constituents	Amount per 100 g
Total fat	20 g
Saturated fat	12.4 g
Cholesterol	52 mg
Sodium	88 mg
Total carbohydrates	50 g
Sucrose	18 g
Protein	20 g
Calcium	1 g

Table 2.4:Composition of the commercially available Fat (local name: ghee, Britania, India)

Constituents	Amount per 100 g
Milk fat	99.7 g
Saturated fatty acids	63.0 g
Polyunsaturated fatty acids	1.7 g
Monounsaturated fatty acids	24.5 g
Trans fatty acids	3.0 g
Cholesterol	0.4 g
Vitamin A	700 µg

2.3 Results and Discussions

The construction and confirmation of a high yielding *R. opacus* PD630 strain, BM985 was performed prior to comparing the lipid yield with that of wild strain.

2.3.1 Construction of BM985

The clone BM985 was confirmed through double digestion (with *NdeI* and *EcoRI*), releasing two bands at 5.8 kbp and 700 kbp corresponding to the size of plasmid pNit-QT1 and *PAP2* gene respectively as shown in Figure 2.2. Expression of the gene leading to a stable protein was confirmed through spotting a size of the *PAP2* protein (24 kDa) on the sodium dodecyl sulfate (SDS) gel illustrated in Figure 2.3.

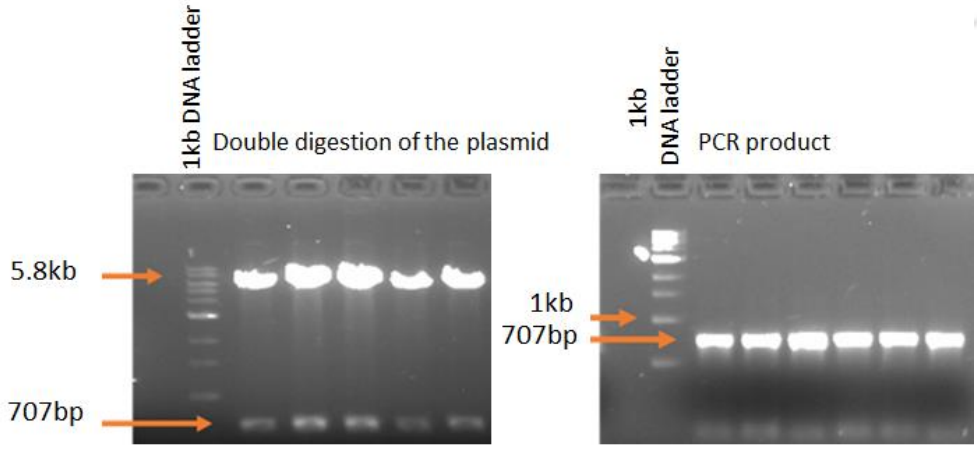

Figure 2.2:Confirmation of BM985 having the expression plasmid pNit-QT1 with PAP2 gene on agarose gel through PCR with isolated plasmid as template and double digestion

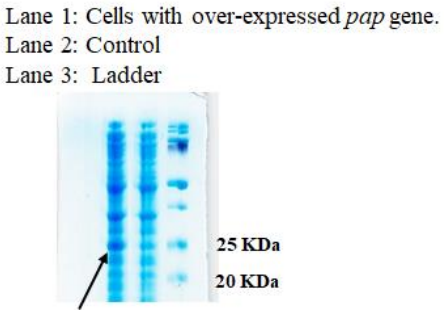

Figure 2.3:Confirmation of BM985 having the expression plasmid pNit-QT1 with PAP2 gene on SDS gel

2.3.2 Comparing the potential of the improved variant BM985 with WT in terms of TAG

Of many transformed cells, the clone with the highest TAG content was chosen for further experiments. The improved variant BM985 and WT were grown in synthetic dairy wastewater for the production of TAG. The result is shown in the Figure 2.4. The result clearly shows the BM985 to be better than its WT strain in terms of TAG contents. The process could be further optimized to enhance the TAG yield in BM985.

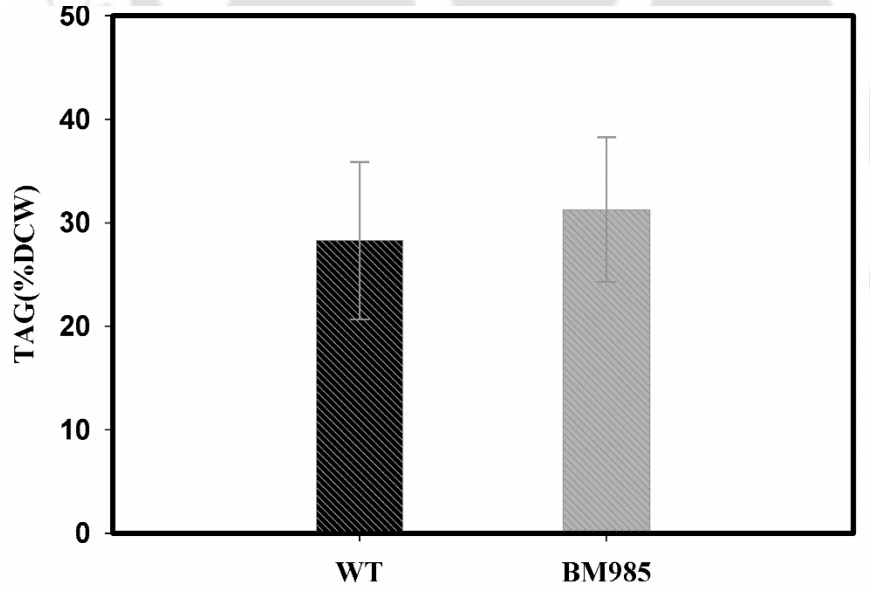

Figure 2.4:Comparing the TAG content in BM985 and WT. Representing TAG as (BM985) and (WT)

Kennedy pathway as shown in the Figure 2.5 is the primary source of TAG production in *Rhodococci* and other Actinobacteria (Amara, Seghezzi et al. 2016). The phosphatidic acid formed in this pathway is dephosphorylated to diacylglycerol (DAG) and catalyzed by *PAP2*. There are very few reports on the role of *PAP2* enzymes in *Rhodococcal* TAG accumulation (Hernández, Comba et al. 2015). *PAP2* enzyme family is currently grouped under Lipid Phosphate Phosphatases (LPPs) and is reported to be predominantly associated with lipid metabolism and its signaling pathways (Pyne, Long et al. 2005). Unlike *PAP1*, *PAP2* enzyme category is found in both eukaryotic and prokaryotic cells (Pfam 01569) (Carman and Han 2006). The Kennedy pathway through a series of enzymatic conversions converts glucose to the biodiesel precursor TAG; with major flux directed towards biomass through phospholipids, which is a bacterial survival tool.

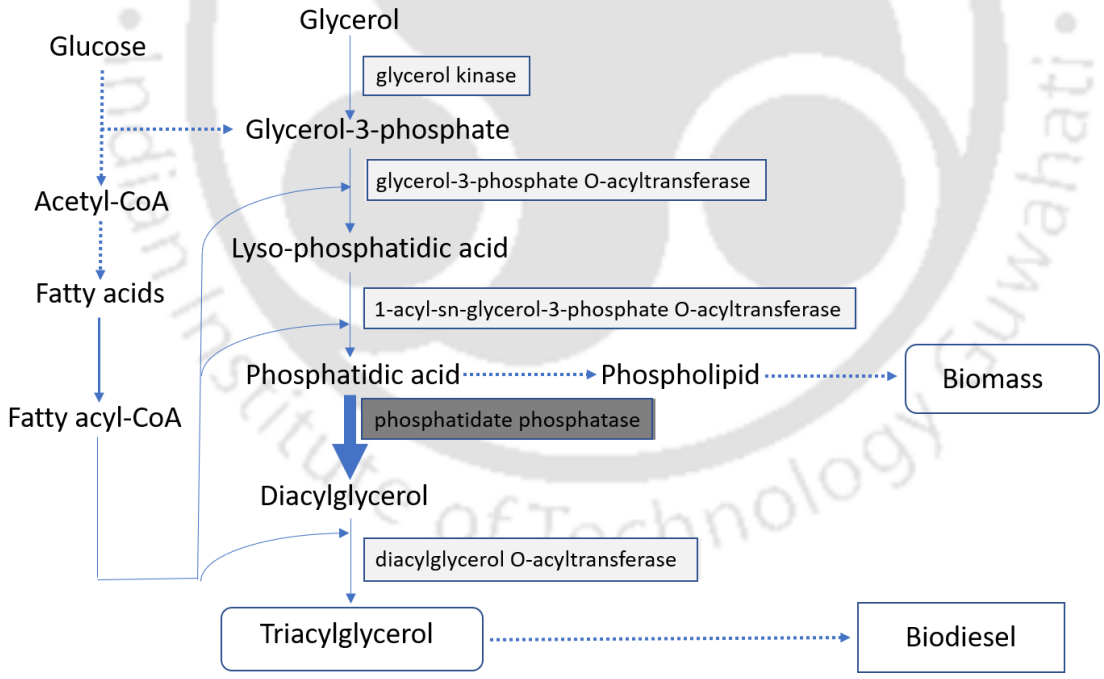

Figure 2.5:Kennedy pathway illustrating the formation of TAG with thin arrows representing the normal expression of genes and thick arrows for overexpression. The structure of the expression vector pNit-QT1 is shown beside the metabolic pathway

Our strategy was to channel the maximum possible flux towards TAG synthesis while maintaining the phospholipids in a balanced proportion for growth. It was thus only

appropriate to choose the phosphatidate phosphatase (*PAP*) gene for an enhanced TAG synthesis with minimum required flux towards biomass. Therefore, functional analysis of *PAP2* enzyme in *Rhodococcal* cells would contribute to a clear understanding of its regulation and consequently be central to designing a novel strategy for improving TAG production.

2.4 References

1. Amara, S., N. Seghezzi, H. Otani, C. Diaz-Salazar, J. Liu and L. D. Eltis (2016). "Characterization of key triacylglycerol biosynthesis processes in rhodococci." Scientific Reports **6**: 24985.
2. Berni, M. D., I. L. Dorileo, J. M. Prado, T. Forster-Carneiro and M. A. A. Meireles (2013). Advances in Biofuel Production. Biofuels Production, John Wiley & Sons, Inc.**:** 11-58.
3. Carman, G. M. and G.-S. Han (2006). "Roles of phosphatidate phosphatase enzymes in lipid metabolism." Trends in biochemical sciences **31**(12): 694-699.
4. Hernández, M., S. Comba, A. Arabolaza, H. Gramajo and H. Alvarez (2015). "Overexpression of a phosphatidic acid phosphatase type 2 leads to an increase in triacylglycerol production in oleaginous Rhodococcus strains." Applied Microbiology and Biotechnology **99**(5): 2191-2207.
5. Leal, M. C. M. R., D. M. G. Freire, M. C. Cammarota and G. L. Sant'Anna (2006). "Effect of enzymatic hydrolysis on anaerobic treatment of dairy wastewater." Process Biochemistry **41**(5): 1173-1178.
6. Na, K.-s., K. Nagayasu, A. Kuroda, N. Takiguchi, T. Ikeda, H. Ohtake and J. Kato (2005). "Development of a genetic transformation system for benzene-tolerant Rhodococcus opacus strains." Journal of Bioscience and Bioengineering **99**(4): 408-414.
7. Pyne, S., J. S. Long, N. T. Ktistakis and N. J. Pyne (2005). "Lipid phosphate phosphatases and lipid phosphate signalling." Biochemical Society Transactions **33**(6): 1370-1374.
8. Schäfer, A., A. Tauch, W. Jäger, J. Kalinowski, G. Thierbach and A. Pühler (1994). "Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum." Gene **145**(1): 69-73.

Chapter 3

Bioprocess development for simultaneous dairy wastewater treatment and biodiesel production using BM985

3.1 Outline

PAP2 over-expressing improved variant *Rhodococcus opacus* BM985 (BM985) was optimized and evaluated for its parameters for maximizing biodiesel production using SDWW as the primary medium. Post application of optimization tools, we witnessed an impressive spike in TAG from 34.5% to 68% TAG/DCW content in BM985 cells at reactor level – a 1.98-fold increment compared to WT together with a lower BM985 biomass yield (0.76 g/g) but significant biomass concentration (3.7 g/L DCW). BM985 was effective in treating synthetic dairy wastewater with a COD removal of ~88% as compared to 65% by wild-type (WT). The fatty acid compositions obtained and thus the biodiesel properties such as cetane number, iodine number, osmotic stability etc. from BM985 strongly support the potential of BM985 as an alternative source for biofuel production and for DWW treatment.

3.2 Materials and methods

3.2.1 Parameters optimization

3.2.1.1 Optimization of SDWW: MSM for culturing *Rhodococci* cells

WT and BM985 cultures were maintained on LB agar containing 10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl. Initial experiments were carried out in 250 mL Erlenmeyer flask containing 50 mL of production medium, mineral salt medium (MSM) containing 9 g/L $Na_2HPO_4 \cdot 12H_2O$, 1.5 g/L KH_2PO_4, 0.1 g/L NH_4SO_4, 0.2 g/L $MgSO_4 \cdot 7H_2O$, 0.02 g/L $CaCl_2 \cdot 2H_2O$, 0.05 g/L $NaHCO_3$ and 1.2 mg/L ferric ammonium citrate.

A loopful of cells from a single colony grown on LB agar at 30 ^oC for 3 days was inoculated in 50 mL of LB medium for preparing the seed culture. The culture was then incubated on a rotary shaker (200 rev min^{-1}) at 30 ^oC for 24 hrs until it reached an optical density of approximately 0.6 at 660nm. The production medium having different

concentrations of SDWW and MSM in the ratio 1:1, 2:1 and 3:1 was inoculated with seed culture to reach an initial OD_{660} of about 0.03 and incubated on a rotatory shaker (200 rev min^{-1}) at 30 °C. The sampling was done every 6 to 12 hrs for the total of 72 hrs batch time to estimate the growth (dry cell weight, DCW of biomass) and lipid profile (TAG%, W/DWC) of cells as described in section 3.2.2.

SDWW was prepared as reported in (Leal, Freire et al. 2006), using commercially available dried milk powder (2 g/L) and fat (0.2 g/L). The major constituents of milk powder (Amulya™, Amul, India) and fat (Ghee, Britannia, India) were used are listed respectively in Table 3.1 and Table 3.2. The synthetic wastewater had an initial pH 6.0 and an average COD content of 2,800 mg/L.

Table 3.1: Composition of the commercially available milk powder (Amulya™, Amul, India)

Constituents	Amount per 100 g
Total fat	20 g
Saturated fat	12.4 g
Cholesterol	52 mg
Sodium	88 mg
Total carbohydrates	50 g
Sucrose	18 g
Protein	20 g
Calcium	1 g

Table 3.2: Composition of the commercially available Fat (local name: ghee, Britania, India)

Constituents	Amount per 100 g
Milk fat	99.7 g
Saturated fatty acids	63.0 g
Polyunsaturated fatty acids	1.7 g
Monounsaturated fatty acids	24.5 g
Trans fatty acids	3.0 g
Cholesterol	0.4 g
Vitamin A	700 µg

3.2.1.2 Characterization of both WT and BM985 strains under different cultivation conditions

Using optimized SDWW: MSM ratio as a medium, effects of pH, temperature, nitrogen, and inoculum size on bacterial growth were accessed initially through shake-flask experiments. The strain was grown under varied pH (2, 4, 6, 7, 9 and 11), varied temperature (25, 28, 30, 35, and 37 °C), different nitrogen sources (ammonium chloride, ammonium sulfate, sodium nitrate, urea, and glycine) with equimolar concentration of nitrogen (1 g/L) and under different inoculum sizes (1, 3, 5, 10, 20 and 30%). The effect of pH was evaluated under the optimized value of temperature, inoculum size, and nitrogen source.

Since acetate has been reported to improve lipid biosynthesis in some strains of microalgae and contribute to ease of degrading organic matter (Silva, Prete et al. 2016, Stemmler, Massimi et al. 2016), we also evaluated the effect of different concentrations of sodium acetate (0.5, 1 and 2 g/L) on growth and lipid production. In addition, phosphate, magnesium, and nitrate concentrations were reported to act as the drivers for lipid induction in rhodococci (Janßen, Ibrahim et al. 2013). We sought to optimize these parameters and evaluate their combined effects on TAG production.

3.2.1.3 Further optimization of the screened parameters

Different optimization tools and methods described below were used to optimize the above-mentioned parameters responsible for enhancing TAG production in both WT and BM985. This optimization was performed under the optimized values of temperature, pH, inoculum size and nitrogen source.

3.2.1.3.1 Box-Behnken Design for optimization of process parameters

Box-Behnken Design's (BBD) primary advantage is in addressing the issue of where the experimental boundaries should be to avoid treatment combinations that are extreme.

During the initial studies we carried out preliminary shake flask studies to determine the levels for experimental design. This led to the selection of BBD as Central Composite Design (CCD) with axial point that would result in a greater number of experiments. Minitab software was used to design the experiment according to the design matrix. Analysis of Variance (ANOVA) with first order regression model was used for screening the significant components and the significance of the model coefficients was assessed using Student's t-test. Four significant parameters (Sodium nitrate, Sodium acetate, Mg^{2+} and PO^{2-}) were screened using Plackett–Burman experiment and later optimized with Box-Behnken Design (BBD) to assess their combined effect. The four significant variables were designated as X_1, X_2, X_3, X_4, and TAG percentage was designated as Y (predicted response). The values of the screened variables used in the three coded levels (-1, 0, +1) are summarized in Table 3.3. WT and BM985 strains were cultivated in defined mediums with varying concentrations of the above-mentioned substrates. The experiments were performed in triplicates. The experimental design consisted of fitting a second-order response surface using a multi-dimensional cube that defined the region of interest. The total number of experiments (N) performed was calculated from the equation **(1).**

$$N = 2K(K-1) + C_0 \qquad \textbf{(1)}$$

Where, K = number of variables and C_0 = number of centre points

The second order polynomial for predicting the optimal levels of the parameters is shown in equation (**2**).

$$Y = \beta_0 + \sum \beta_i X_i + \sum \beta_{ij} x_i x_j + \sum \beta_{ii} x_i^2 \qquad \textbf{(2)}$$

Where, Y is the measured response (TAG %) variable; β_0, β_i, β_{ij}, β_{ii} are constant and regression coefficients of the model, and x_i, x_j represent the independent variables in coded values.

Table 3.3:Coded values of independent variables for Box-Benkhen design

Independent variable	Coded value			
	Symbol code	-1	0	1
Sodium Nitrate (g/L)	X_1	0.1	0.55	1
Sodium Acetate (g/L)	X_2	0.5	1	2
KH_2PO_4(g/L)	X_3	3	6.5	10
$MgCl_2$ (g/L)	X_4	0.1	1.05	2

3.2.1.3.2 Artificial neural network (ANN)

We used an artificial neural network (ANN) based on feed-forward back-propagation method. The network architecture consisted of four input layers viz., Sodium nitrate, Sodium acetate, Mg^{2+} and PO_4^{2-}, eight hidden layers and one output layer representing the percentage of TAG production. In this work, a Neural Network Toolbox of MATLAB (2016b) mathematical software was used to predict the TAG production. We used particle swarm optimization (PSO) and genetic algorithm (GA) algorithms to further optimize the process parameters. In the feed-forward training system, the data flowed in a channel from input to output layer via hidden layer. These four input layers are interconnected by the parameters known as weights (w) and biases (b) (K.M. Desai, S.A. Survase et al. 2008, M. Hhayet and C. Cojocaru 2012). At the beginning of the process, the weighted input including the biases were summed up and transferred to nonlinear transfer function (tansig, equation 3) situated between the input and hidden layer. The output of the hidden layers was transferred as input for linear transfer function purelin (equation 4) and the output of this function would subsequently be transferred to the output layer.

$$tansig = \frac{1+\exp(-sum)}{1-\exp(-sum)} \quad \textbf{(3)}$$

$$Purelin = sum \quad \textbf{(4)}$$

The total experimental data was divided into three different sets for training [9 datasets], validation [3 datasets] and testing [3 datasets]. The neural network works based on an iterative method that calculated the error (the difference between actual output and predicted output), used an error function and minimized the error by adjusting the weights. We used mean squared error (MSE, equation **5**) function in the present study.

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Y_a - Y_p)^2 \quad \textbf{(5)}$$

Where Y_a is the actual output, Yp is the predicted output and N is the number of data points. The MSE is minimized by appropriate adjustment of the weights and biases. During training step the weight and biases get iterated by Levenberg–Marquardt algorithm convergence to achieve the optimal value (Y. Yasin, F. Bin et al. 2014, Prabhu and Jayadeep 2017, Prabhu, Purkayastha et al. 2017).

3.2.1.3.3 Particle Swarm Optimization (PSO) Algorithm

The global optimal output from ANN model was used to optimize the parameters using PSO algorithm. The velocity and position of each particle was evaluated using t (equation **6),** subsequently, the particle is self-directed to a new position identified by the velocity matrix (equation **7**).

$$v_i(t+1) = wv_i(t) + c_1r_1\big(p_i(t) - x_i(t)\big) + c_2r_2\left(p_g(t) - x_i(t)\right) \quad \textbf{(6)}$$

$$x_i(t+1) = x_i + v_i(t+1) \quad \textbf{(7)}$$

In this equation t is the generation of algorithm running, w is weight of inertia, c_1 and c_2 are acceleration constants, r_1 and r_2 are elements from two random sequences in the range of (0, 1), x_i (t) is the i^{th} particle in the t generation; p_i is the best solution for this particle; p_g is the best solution for all particles. “ANN-PSO” function of MATLAB was used to

optimize all the models (Eberhart and Kennedy 1995, Huang, Mei et al. 2007, Zhang, Zhang et al. 2007, Khaouane, Si-Moussa et al. 2012).

3.2.1.3.4 Genetic algorithm (GA)

The genetic algorithm (GA) is a global optimization procedure. In this method, the convergence is independent of the initial value. Post-development of ANN the input space may be further optimized by GA. Initialization of the solution for the population will take place followed by fitness computation which in turn is dependent on objective function; it is in this step that the better individual will be selected. The creation of new sets of the individual is based on the selected individual that had undergone crossing over and mutation. This process is repeated until an appropriate result is achieved (K.M. Desai, S.A. Survase et al. 2008, Prabhu A, Chityala et al. 2017).

3.2.2 Model validation

To determine the consistency of the results predicted by the model with the actual results, validation experiments were carried out with optimized values from the initial screening experiments as well as from BBD and Hybrid ANN-PSO-GA optimization in 250 mL Erlenmeyer baffled flasks in triplicates at a temperature of 28 °C and at a speed of 200 rev min^{-1}.

3.2.2.1 Increased TAG production under controlled and optimized conditions in bioreactor

The optimized values from the shake-flask validation experiments were taken for the scaleup experiment at reactor level. Stirred Tank Bioreactor (Applikon Dependable Instruments, ez Control, The Netherlands) was utilized in this investigation. The organisms (WT and BM985) were grown in optimized medium in a 3-L bioreactor (1-L working volume). The bioreactor was aerated through a sintered stainless-steel sparger at

a constant airflow rate of 1 vvm. The rotational speed of impeller in the reactor was kept constant at 300 rev min^{-1} throughout the cultivation. The pH and the temperature in the reactor were maintained at 7±0.1 and 30±1 °C, respectively. The inoculum and other medium constituents were as per the values obtained and validated from ANN-GA optimization algorithm. At the end of the cultivation cycle on 4^{th} day, the reactor was harvested by removal of biomass and analysed for DCW, TAG content, residual sugar, chemical oxygen demand.

3.2.3 Analysis of growth and substrates utilization

At each sampling point, the biomass, utilization of substrates and lipid production were evaluated. A known sample volume was centrifuged at 8000 g for 10 minutes at 4 °C; the supernatant was collected for extracellular substrate analysis (lactose, nitrate, and phosphate) and the pellet was used for biomass and neutral lipid analysis. The biomass was evaluated using a UV-Visible spectrophotometer (Cary 100, Varian, Australia), measuring the absorbance at 660 nm. TAG production was evaluated using Nile red assay as documented by Fei et al (Fei, O'Brien et al. 2016).

3.2.4 Chemical oxygen demand (COD) estimation

Samples were centrifuged at 8,000 rpm for 15 min to separate the bacterial biomass and SDWW. The supernatants were appropriately diluted and COD analysis was carried using standard methods (Clesceri, Greenberg et al. 1998).

3.2.5 Analysis of fatty acid methyl esters (FAMEs)

The fatty acid composition of the cells was analysed by gas chromatography (GC) using methods reported (Brandl, Gross et al. 1988, Timm, Byrom et al. 1990) with some modifications (Kumar, Muthuraj et al. 2014). Briefly, fatty acid in the cells was first converted to fatty acid methyl esters (FAMEs) in a two-step *in situ* transesterification

method for analysis. Step 1: Lyophilized biomass of BM985 was subjected to alkali treatment by adding 1.5 mL of 2% (w/v) NaOH in methanol to 5-15 mg bacterial cells and incubated for 20 min at 200 rev min^{-1} in a shaking water bath at 90 °C. Step 2: The mixture was subjected to acid treatment by adding 1.8 mL 5% (v/v) H_2SO_4 in methanol and incubated for 10 min at 150 rev min^{-1}, 90 °C. Equal ratios of water and hexane were added to the mixture before centrifugation to obtain FAMEs in the hexane layer. The Hexane layer with FAMEs was filtered through a 0.2 µm filter for analysis using GC against a FAME mix C14-C22 (Supelco, USA) as the standard. The compositions estimated are represented as the total lipid of the biomass in % (w/w, DCW). Biodiesel

properties based on fatty acid compositions are calculated empirically using the equations suggested in different reports (Francisco, Neves et al. 2010). Properties such as viscosity (η), cetane number (CN), flash point (T_f), cloud point (CP), pour point (PP)were calculated based on the equations described by Su et al. (Su, Liu et al. 2011) and the values for saponification value (SV), iodine value (IV), degree of unsaturation (DU) and highest heating value (HV)were evaluated using the equations obtained from Francisco et al.(Francisco, Neves et al. 2010).

$$\eta = 0.235W_C - 0.468W_{db} \quad (8)$$

$$CN = 3.930W_C - 15.936W_{db} \quad (9)$$

$$T_f = 23.362W_C + 4.854W_{db} \quad (10)$$

$$CP = 18.134W_C - 0.790W_{US} \quad (11)$$

$$PP = 18.880W_C - 1.00W_{US} \quad (12)$$

$$SV = \sum(560 \times P_{FA})/MW \quad (13)$$

$$IV = \sum(254 \times P_{FA} \times N_D)/MW \quad (14)$$

$$DU = (W_{MUFA}) + (2 \times W_{PUFA}) \quad (15)$$

$$HV = 46.19 - \left(\frac{1794}{MW_i}\right) - 0.21 \times N_D \quad (16)$$

Where W_C is the weighted-average number of carbon atoms in the fatty acids, W_{db} is the weighted-average number of double bonds, W_{US} represents the total unsaturated FAME content (weight %),P_{FA} is the percentage of each fatty acid, MW is the molecular mass of individual fatty acids, N_D is the number of double bonds, W_{MUFA} is the monounsaturated fatty acids in weight percentage, W_{PUFA} is polyunsaturated fatty acids in weight percentage, MW_i represents the molecular weight of the i^{th} FAME component.

3.3 Results and Discussion

3.3.1 Screening Nitrogen source for optimal growth

In the current experiment, among the five nitrogen sources tested, both WT and BM985 strains exhibited enhanced cell growth when cultured in the presence of sodium nitrate (see Figure 3.1). The medium with sodium nitrate resulted in the highest biomass concentration of 2.5 g/L and 2.7 g/L DCW in both BM985 and WT respectively as shown in Figure 3.1. At the end of batch, the biomass yield ($Y_{x/s}$ = g of DCW/g of carbon being consumed) recorded for WT and BM985 was 0.96 g/g and 0.89 g/g respectively. The order of nitrogen sources for biomass by DCW were: sodium nitrate > ammonium chloride > glycine > ammonium sulphate > urea (see Figure 3.1). Media with ammonium sulphate, urea and glycine resulted in suboptimal growth in both strains.

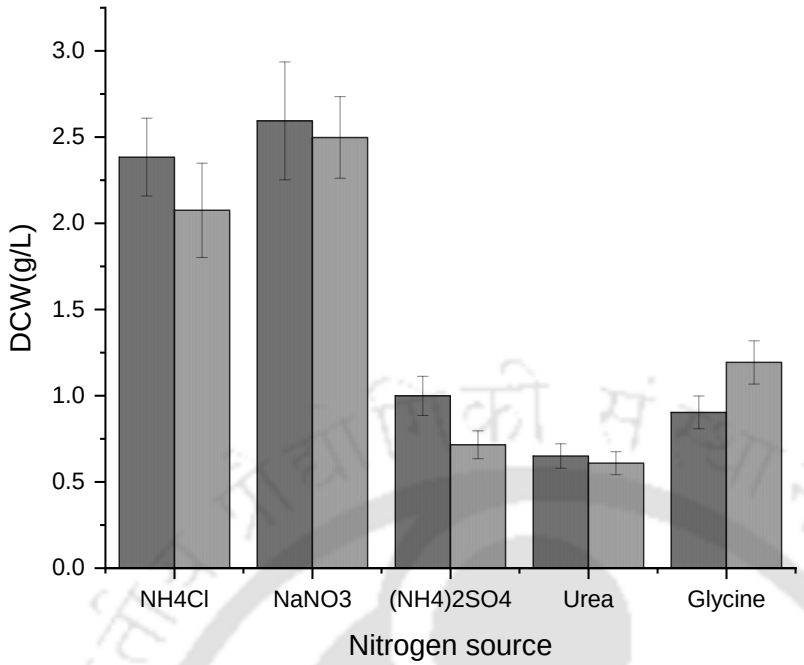

Figure 3.1: Effect of different nitrogen source on biomass. Representing TAG as (BM985) and (WT)

3.3.2 Effect of inoculum size on optimal growth and lipid production

The effect of inoculum size on growth and TAG production in both strains is represented in Figure 3.2. A higher biomass concentration and yield were obtained for the lower inoculum size compared to higher inoculum sizes. With 5% inoculum size, maximum DCW and $Y_{x/s}$ in both BM985 (1.2 g/L and 0.4 g/g) and WT (1.1 g/L and 0.43 g/g) was attained at the end of 3rd day and the observed DCW amount for both strains decreased with inoculum sizes above and below 5%. The lipid production and growth are mutually coupled to each other as observed. However, the inoculum size of 5 % achieved both higher biomass concentration and an enhanced lipid production than the high inoculum size. BM985 and WT achieved 28 % and 26 % TAG of DCW respectively at 5% inoculum size. Comparatively, TAG production was more in BM985 than WT for all inoculum sizes from 1 % to 30 %. Like biomass, TAG production also decreased above and below 5% inoculum size.

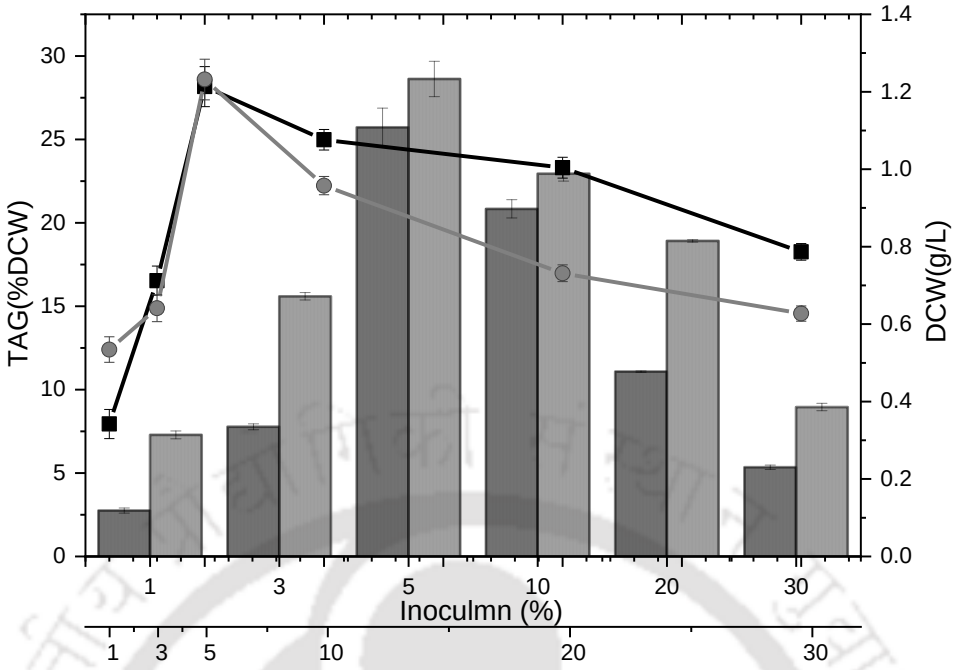

Figure 3.2: Effect of inoculum size on TAG and biomass profile. Representing TAG as (BM985) and (WT) while biomass concentration as (BM985) and (WT)

3.3.3 Effect of pH, temperature and sodium acetate

The maximum biomass, as well as TAG production, were observed at neutral pH (pH 7) for both WT and BM985 as shown in Figure 3.3. At pH 7, BM985 produced a higher amount of TAG (30% with DCW of 2.5 g/L) compared to WT (25 % with DCW of 2 g/L) (see Figure 3.3). At both extremes of pH range, both strains showed suboptimal growth and TAG production.

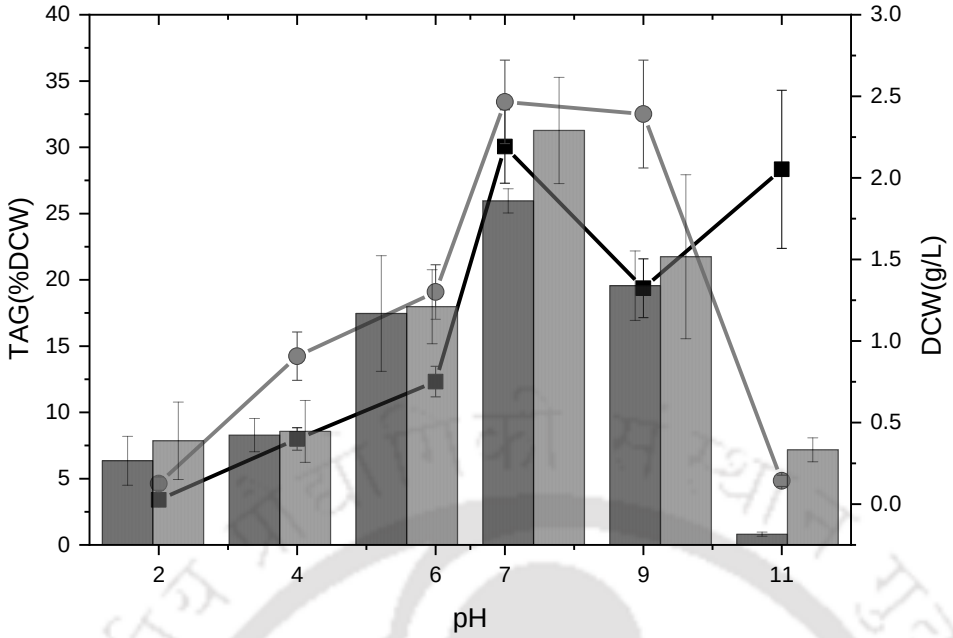

Figure 3.3: Effect of different pH range on TAG and biomass profile. Representing TAG as (BM98) and (WT) while biomass concentration as (BM985) and (WT)

Maintaining an optimal temperature is significant in achieving a suitable TAG yield. The effect of temperature on biomass and TAG production in both strains are represented in Figure 3.4. Even though BM985 achieved a maximum biomass of 2.2 g/L at 30 ^{o}C, the highest TAG yield (36% of DCW) was achieved at 28 ^{o}C. In contrast, WT achieved a maximum of biomass (2.5 g/L) and TAG yield (22% of DCW) at 30 ^{o}C. At tested temperatures lower than 28 ^{o}C and higher than 30 ^{o}C, both strains showed suboptimal growth and TAG yield.

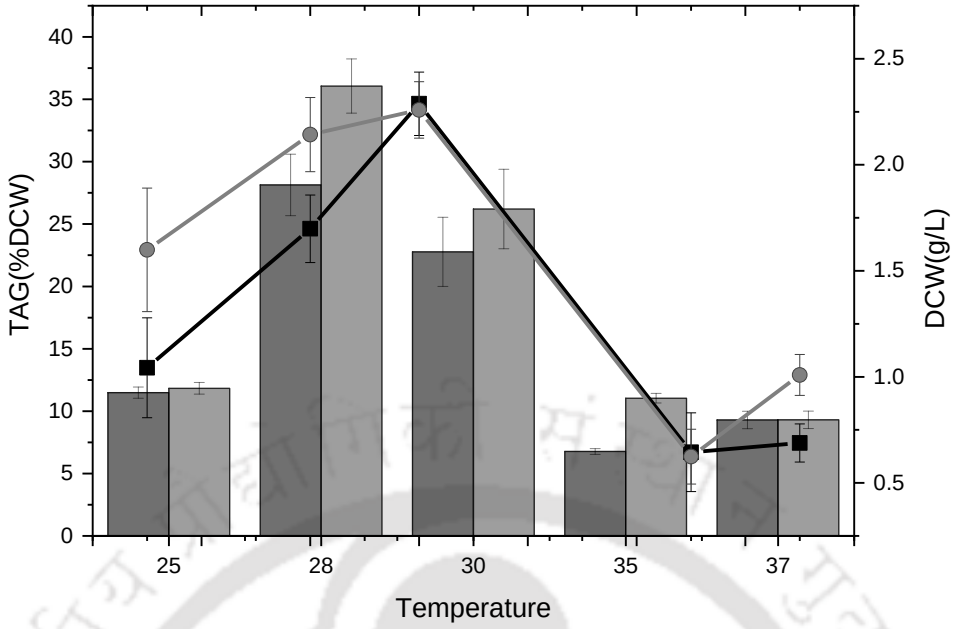

Figure 3.4: Effect of different temperature range on TAG and biomass profile. Representing TAG as (BM98) and (WT) while biomass concentration as (BM985) and (WT).

The effect of sodium acetate on biomass concentration and TAG production in both strains are represented in Figure 3.5. At 1 g/L sodium acetate concentration higher biomass with maximum TAG were observed in both BM985 (2.8 g/L DCW, 37 % TAG) and WT (3.4 g/L, 27 % TAG). The highest biomass concentration in BM985 (3.1 g/L DCW) was observed with 0.5 g/L sodium acetate, whereas maximum TAG in BM985 (37 % of DCW) was observed with 1 g/L sodium acetate. However, WT showed no difference in TAG yield (27 % of DCW) between 0.5 g/L and 1 g/L sodium acetate despite an increase in biomass from 2.7 g/L to 3.4 g/L DCW, respectively. Both BM985 and WT strains showed significantly lower growth and TAG yield at 2 g/L of sodium acetate owing to cell toxicity under higher concentration of sodium acetate. Therefore, optimal growth and TAG yield were observed in both strains at 1 g/L sodium acetate, whereas other concentrations of sodium acetate (0.5 g/L and 2 g/L) were suboptimal. The profile for biomass yield shows nearly the same trends as biomass concentration with $Y_{x/s}$ for

WT (0.85 g/g) being higher than BM985 (0.61 g/g) at 1 g/L sodium acetate and $Y_{x/s}$ for WT (0.71 g/g) being lower than BM985 (0.78 g/g) at 0.5 g/L sodium acetate.

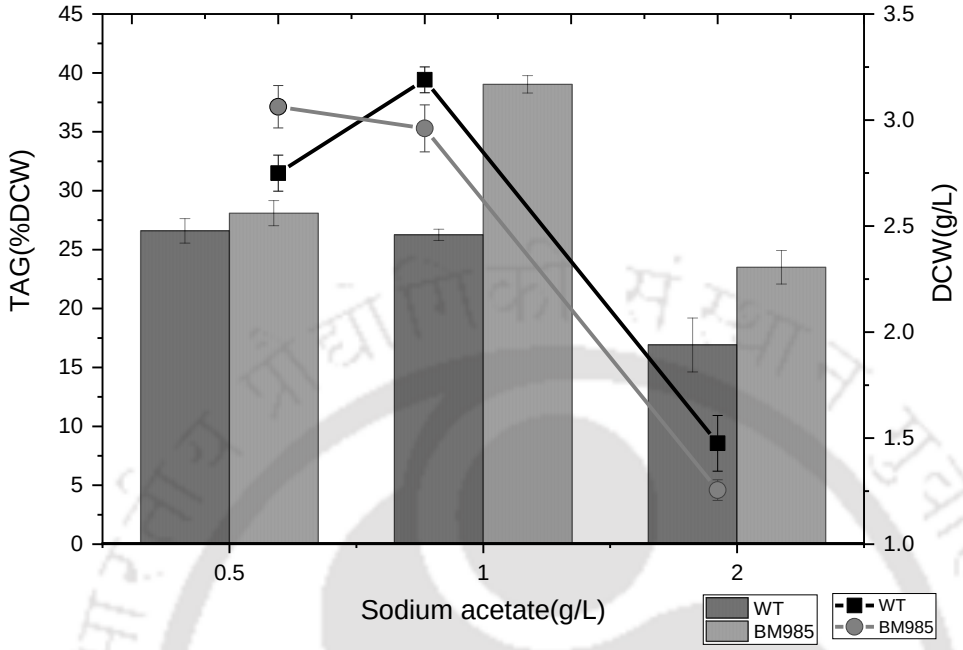

Figure 3.5: Effect of different sodium acetate concentration on TAG and biomass profile.

3.3.4 Screening variables for optimal growth

The importance of optimization of individual medium constituents for lipid production was reflected by the observed wide differences in lipid concentration. The four significant substrates: sodium nitrate, sodium acetate, Mg^{2+}, and PO^{2-} were considered cardinal to optimization studies (M. Hhayet and C. Cojocaru 2012). Screening sodium acetate and Mg^{2+} as the suitable nutrients leading to improved cellular growth further supported our findings on sodium acetate. The effect of Mg^{2+} on *Rhodococcus* growth and lipid production was evaluated (data not shown) by supplementing the media with $MgSO_4$ (0.1 g/L, 0.5 g/L, and 1 g/L), which showed maximum growth at 0.95 g/L $MgSO_4$, whereas 0.1 g/L $MgSO_4$ was insufficient for the cell growth and metabolism. Addition of Mg^{2+} ions led to increased biomass production and lipid content (~1.7-fold).

3.3.5 Optimization of screened variables for maximization of TAG

To analyze the effect of screened variables on TAG response, we adopted a three-level Box-Behnken design. Based on multiple regression analysis, the second-order polynomial response was fitted using the following equation, according to the methods reported previously (Jungo, Marison et al. 2007). The concentration of TAG was observed to be varying between 1.81 % to 25 % of DCW.

$$Y_{TAG} = 13.12 - 3.03X_1 + 2.15X_2 - 1.67X_3 + 1.71X_4 + 3.95X_1^2 - 6.36X_2^2 + 2.35X_3^2 - 7.05X_4^2 - 1.30X_1X_2 + 1.154X_1X_3 - 0.72X_1X_4 - 0.44X_2X_3 + 0.45X_2X_4 - 0.32X_3X_4$$

Where, X_1 = sodium Nitrate, X_2 = sodium Acetate, X_3 = KH_2PO_4 and X_4 = $MgCl_2$

The results from ANOVA of the quadratic model are listed in Table 3.4 and the estimated model coefficients and their significance (by student t-test) are listed in Table 3.5.

Table 3.4: ANOVA for quadratic model

Source	DF	SS	MS	F	P
Model	14	861.48	61.53	8.83	0
Residual (Error)	13	90.61	6.97		
Lack-of-Fit	10	79.37	7.94	2.12	0.29
Pure Error	3	11.23	3.74		
Total	27	952.09			

R^2 = 90.48% R^2(pred) = 52.98% R^2(adj) = 80.23%
DF= degrees of freedom, SS =sum of squares, MS= mean square

TAG production was explored by the tested variables with the help of 3D surface plots built on two independent variables (along x and y-axis) to that the response (%TAG W/DCW, z-axis), as shown in Figure 3.6. The plots showed the bell-shaped response of TAG production with increasing sodium acetate concentration with a peak at approximately 1.2 g/L of sodium acetate illustrating the importance and ease of acetate

assimilation in the lipid biosynthesis pathway. A similar pattern was observed for $MgCl_2$, where the optimum $MgCl_2$ (1.25 g/L) showed maximum TAG production. The concentrations above or below the optimum range of sodium acetate (1.2 g/L) and $MgCl_2$ (1.25 g/L) resulted in a sharp decrease in TAG production. In the contrary, a lower concentration of sodium nitrate and KH_2PO_4 were optimal for maximum TAG production. These findings supported the phenomenon of nitrogen limitation as a prerequisite for initiating lipid biosynthesis.

Table 3.5 Model coefficient estimated by multiple linear regressions

		SE Coef	**T**	**P**
Constant	13.117	0.9129	14.369	0
X1	-3.0281	0.6223	-4.866	0
X2	2.1545	0.6223	3.462	0.004
X3	-1.6696	0.6223	-2.683	0.019
X4	1.7055	0.6223	2.741	0.017
X1*X1	3.9548	1.6438	2.406	0.032
X2*X2	-6.356	1.6438	-3.867	0.002
X3*X3	2.3493	1.6438	1.429	0.177
X4*X4	-7.0512	1.6438	-4.29	0.001
X1*X2	-1.3035	0.66	-1.975	0.07
X1*X3	1.145	0.66	1.735	0.106
X1*X4	-0.7152	0.66	-1.084	0.298
X2*X3	-0.4422	0.66	-0.67	0.515
X2*X4	0.452	0.66	0.685	0.505
X3*X4	-0.3157	0.66	-0.478	0.64

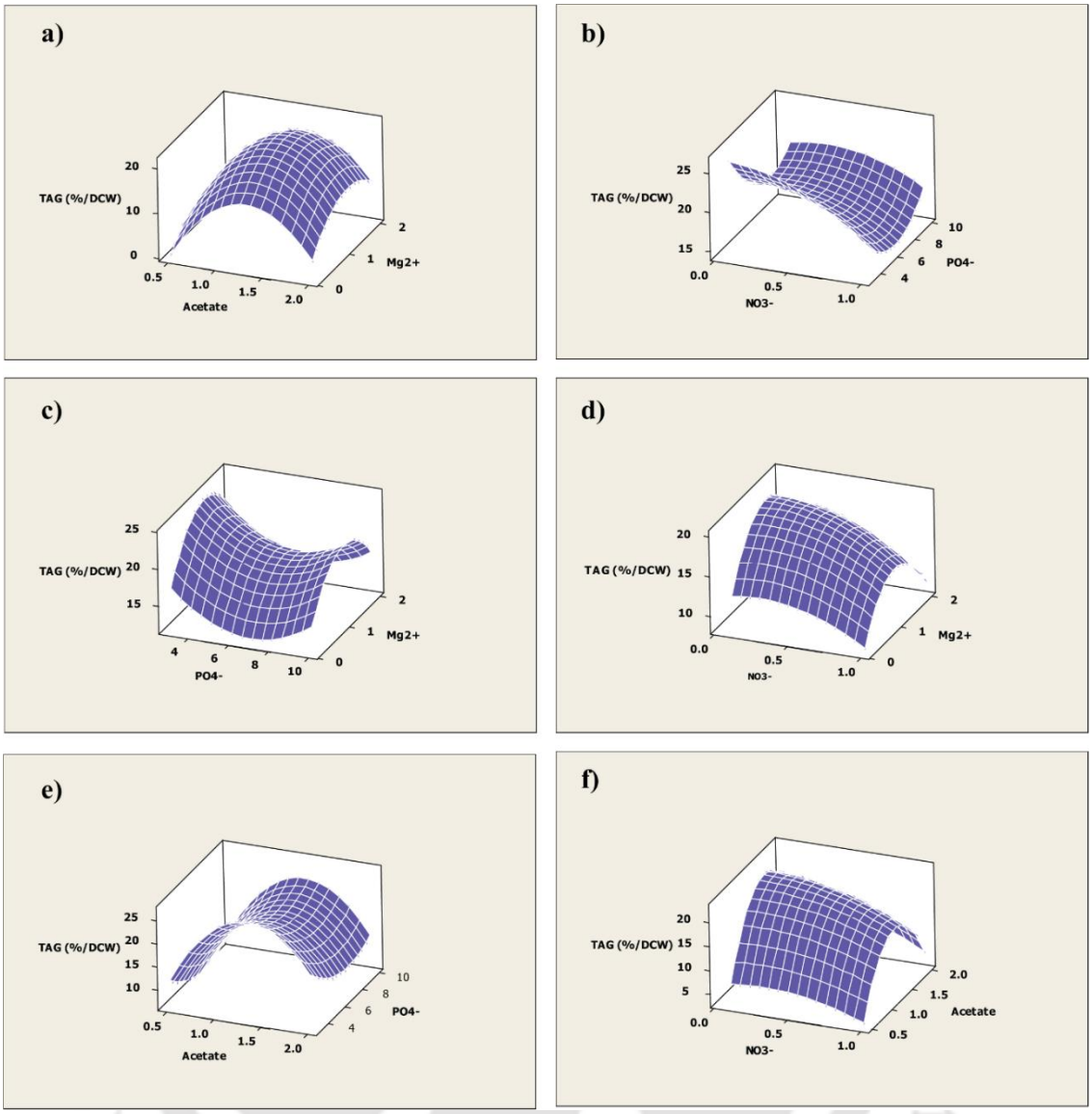

Figure 3.6:Three-dimensional response surface plot for TAG production (z-axis) showing the interactive effects of (a) Sodium acetate and $MgCl_2$(b) $NaNO_3$ and KH_2PO_4 (c) KH_2PO_4 and $MgCl_2$ (d) $NaNO_3$ and $MgCl_2$ (e) Sodium acetate and KH_2PO_4 and (f) $NaNO_3$ and Sodium acetate

3.3.6 Hybrid ANN-PSO-GA modeling

The non-linear problems of the RSM was addressed by the feed forward ANN with backpropagation. The experimental data from RSM-BBD were then used as input in developing and training the ANN model. Using the Levenberg–Marquardt back-propagation ANN was trained up to 1000 epochs and the optimum value was reached by

5 epochs. The specific experimental runs' output, TAG% of DCW were used in ANN model for training, validation and testing with corresponding R^2 and MSE. The correlation between model output versus prediction output (R^2 = 0.93, see Figure 3.7) showed the model explained ~93% of the total variability.

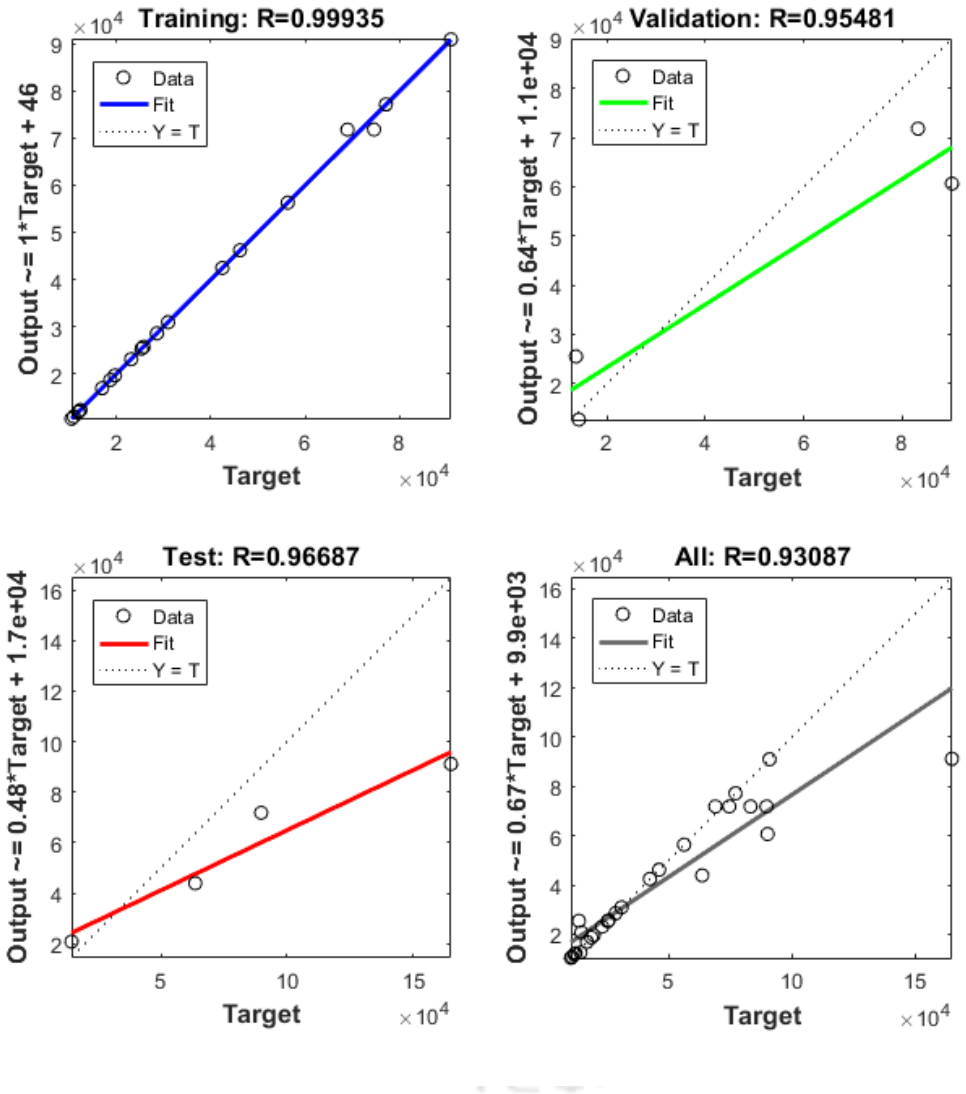

Figure 3.7: The regression plot for prediction performance of ANN models for the TAG Production

In GA, the ANN-PSO trained values were used to optimize input space with the primary aim of maximizing TAG yield. After 171 iterations, the optimized value of 23.99 % TAG

of DCW (see Figure 3.8) was achieved by maintaining the variables as follows: sodium nitrate = 0.1 g/L, sodium acetate = 1.073 g/L, KH_2PO_4 = 3 g/L and $MgCl_2$ = 0.115 g/L.

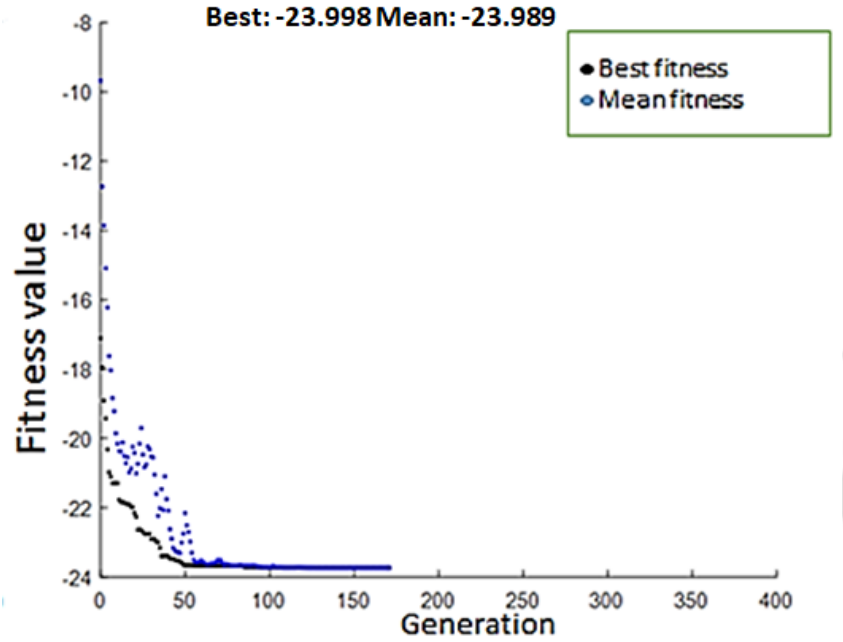

Figure 3.8: Representative plots generated from the optimization by GA using MATLAB (2016b). Best and average fitness values with successive generations showed gradual convergence to the optimum value for TAG production

3.3.7 Verification of the model

The validation of the model was conducted based on the optimization results acquired from RSM-BBD and ANN-GA which were further compared with the final TAG value in Table 3.6. The experimental TAG (35·28 ± 1·78 % of DCW) obtained by ANN was higher compared to that obtained by RSM (28·12 ± 1·92 % of DCW). The high R^2 value (0·931) of ANN model than the RSM-BBD regression model (0.905), implies the ANN model to be more accurate for the present study.

Table 3.6: Comparison of two optimization model and their validation

Variable	**Optimum concentration (g/L)**				**TAG (% of DCW)**		
	Sodium nitrate	Sodium acetate	KH_2PO_4	$MgCl_2$	Predicted	Observed	R^2
RSM (BBD)	0.201	1.24	4.65	1.25	22.58	28.12±1.92	90.48
ANN-hybrid model	0.102	1.07	3.00	0.12	23.99	35.28±1.78	93.08

3.3.8 Reactor Studies

Based on the fermentation conditions (inoculum size = 5%, sodium nitrate = 0.1 g/L, sodium acetate = 1.073 g/L, KH_2PO_4 = 3 g/L and $MgCl_2$ = 0.115 g/L) obtained from a shake flask, scale-up fermentation was conducted. A combinatorial approach to dairy wastewater treatment with lipid accumulation was further carried out in a 1L bioreactor to evaluate the potential of *R. opacus* as a potential renewable fuel feedstock. The results from batch studies shown in Figure 3.9 illustrated that a controlled reactor environment showed increased biomass and lipid accumulation in both BM985 and WT compared to controlled shake-flask experiments. The biomass concentration and lipid production in bioreactor were: WT (3.92 g/L DCW with 34% TAG) and BM985 (3.73 g/L DCW with 68% TAG), which is clearly more than that in shake-flasks batch (see Figure 3.9). Along with increased biomass and lipid production in BM985 in batch studies, about 88% of COD removal in SDWW was achieved when compared to flask experiments (80% COD removal) under optimized conditions. The COD removal levels of BM985 (88%) were 1.29-fold higher than WT (65%). The biomass yield shows some interesting trends with $Y_{x/s}$ from the reactor for WT (0.88 g/g) being very much higher than BM985 $Y_{x/s}$ (0.76 g/g). Shake flask studies reveal nonsignificant difference $Y_{x/s}$ (0.63 g/g) between WT and BM985.

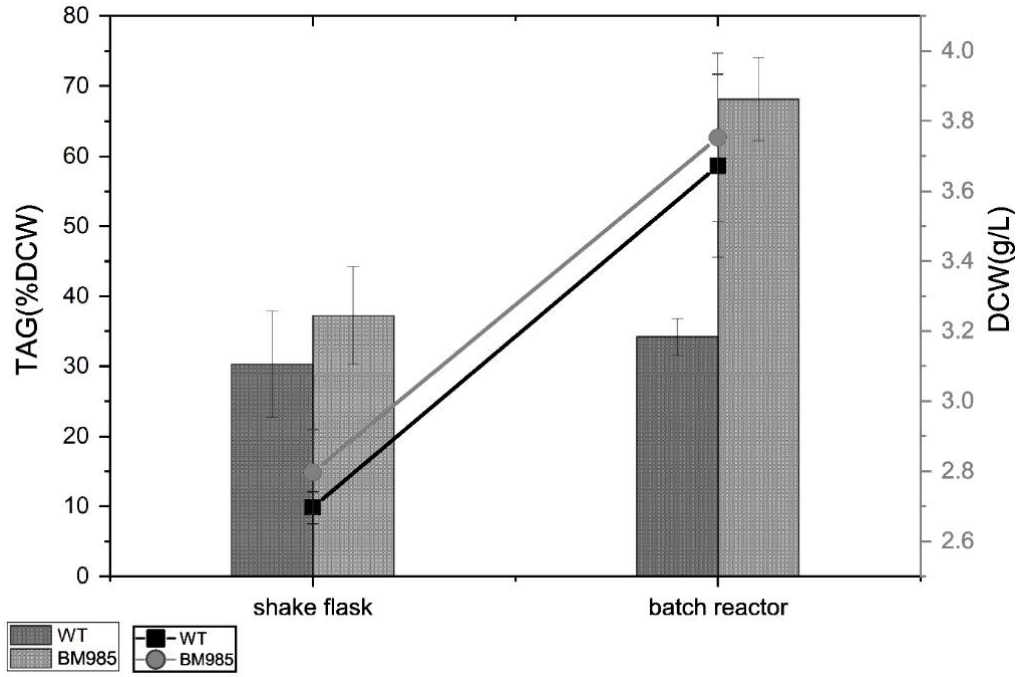

Figure 3.9: Effect of controlled parameters on DCW and TAG for WT and BM985

3.3.9 Composition of fatty acids

In our study, the composition of fatty acids obtained with lactose as present in SDWW and acetate with ghee (lipid) as substrates are shown in Table 3.7. Fatty acid composition of different microbes along with BM985 is shown in

Table 3.8, indicating it to be more like *R. opacus* B4. The principal fatty acid occurring in the stored lipids is the saturated palmitic acid (see Table 3.7) which amounted to approximately 34% of the total fatty acid resulting in nearly three-fold higher saturated species (74.41%) than that of unsaturated (25.59%) ones. The evaluated biodiesel properties, as shown in Table 3.9, based on empirical equations suggest it is of commercial biodiesel grade.

Table 3.7: Fatty acids composition of BM985 biomass in SDWW

FAMEs	**Relative proportion of fatty acids (%, w/w)**					
	Glucose	**Acetate**	**Gluconic acid**	**Acetic acid**	**Gluconate**	**SDWW and Acetate**
C12:0	-	-	Trace (<0.5)	Trace (<0.5)	0.8	-
C13:0	-	-	-	-	-	-
C14:0	2.0	2.04	1.9	1.9	4.3	0.95
C15:0	-	0.72	6.4	11.4	6.3	-
C16:0	38.5	30.4	36.4	34.8	25.7	34.13
C16:1	5.0	17.5	4.6	4.0	9.5	6.05
C17:0	22.9	0.03	11.4	16.8	12.3	18.73
C17:1	15.0	0.97	10.6	12.5	15.4	9.07
C18:0	2.8	10.27	9.6	5.7	3.5	13.45
C18:1	13.7	38.04	19.1	12.9	22.0	10.67
C20:0	-	-	-	-	Trace (<0.2)	7.82
Sigma-Aldrich	(Castro, Rocha et al. 2016)	(Bhatia, Kim et al. 2017)	(Alvarez, Mayer et al. 1996)		(Alvarez and Steinbüchel 2002)	This study

Table 3.8: FAMEs profile of various microbes

FAMEs	***Streptomyces coelicolor* %(w/w)**	***Rhodococcus* sp. *YHY01* %(w/w)**	***Rhodococcus opacus B4* %(w/w)**	***Nannochloropsis salina* %(w/w)**	**BM985 %(w/w)**
C12:0	5.03	-		5.0	-
C14:0	-	0.97	1.6	-	0.95
C15:0	-	0.17	-	-	-
C15:1	1.43	-	-	-	-
C16:0	-	38.16	35.6	37.5	34.13
C16:1	20.07	12.83	4.5	23.3	6.05
C17:0		0.11	22.4	0.4	18.73
C17:1	1.42	0.30	16.5		9.07
C18:0	-	16.59	3.8	0.9	13.45
C18:1	-	30.84	-	-	10.67
C20:0	-	-	-	-	7.82
C20:4	-	-	-	3.3	-
C20:5	-	-	1.6	-	-
Sigma-Aldrich	(Bhatia, Yi et al. 2015)	(Bhatia, Kim et al. 2017)	(Castro, Rocha et al. 2016)	(Bhatia, Bhatia et al. 2017)	This study

Table 3.9: Specifications for commercial biodiesel

Properties	This study	Microalgal[b]	Biodiesel blend stock (B100)	
			ASTM D6751	EN 14214
Density, (kg/m^3)	858.146	0.876	NS	860 - 900
Kinematic Viscosity, (mm^2/s)	4.251	4.35	1.9 – 6.0	3.5 – 5.0
Cetane number (CN)	55.233	52.6	47 (min.)	51 (min.)
Flash point (T_f), (°C) (minimum)	126.101	160	93 (min.)	101 (min.)
Cloud point (CP), (°C)	-5.956	-	Report[a]	Country specific[a]
Pour point (PP)	-8.614	-	NS	NS
Iodine Value (IV), (g I_2/100 g)	24.217	-	NS	120 (max.)
Saponification value (SV), (mg KOH/g)	209.526	-	164-220	NS
Cold flow plugging point (CFPP), (°C)	273.150	-	NS	Variable
Vapor pressure, (Pa)	7.40E^{-8}	-	NS	NS
Heat Capacity, (J/gK)	2.215	-	NS	NS
Heat of vaporization, (KJ/mol)	231.138	-	NS	NS
Higher heating value, (MJ/kg)	-	38.25	-	-
Osmotic stability, (h)	infinity	1.2	-	-

a = Low temperature properties are not strictly specified but should be agreed upon by the fuel supplier.
b = (Bhatia, Kim et al. 2017)
NS = Not specified

Over the years, diverse forms of PAP like enzymes have evolved and appeared in rhodococci with specialized functions in lipid metabolism, primarily to adapt glycerolipid biosynthesis under changing physiological and environmental conditions. These evolved PAP proteins might have a role in forming separate pools of DAG in rhodococci for the specific biosynthesis of TAG (Hernández, Comba et al. 2015). We sought to assess the potential of the *PAP2* overexpressing BM985 to thrive, enhance TAG production and simultaneously treat the dairy wastewater. The TAG accumulation was maximized on the 3rd day of the batch experiment. SDWW with Mineral Salt Media (MSM) at 3:1 ratio resulted in both an enhanced biomass and lipid biosynthesis by BM985 (data not shown) compared to minimal culture titers observed when cells were grown in absence of MSM. These findings clearly illustrated mineral salts were essential for the efficient utilization of organics present in SDWW to produce lipids and improved biomass.

Unfavourable changes in physiological, environmental and nutritional conditions, inhibit cell growth and impact lipid accumulation. While nitrogen deficiency is a key trigger for lipid accumulation in *R. opacus* PD630, it simultaneously restricts overall biomass development (Xia, Zhang et al. 2011). However, as lipid biosynthesis is directly coupled to the biomass, it may be assumed that those BM985 cells achieving higher biomass would automatically lead to improved lipid content. Recent reports (Blasco R, Martínez-Luque M et al. 2001, Pacheco, Ciapina et al. 2010) support our finding that sodium nitrate as a suitable nitrogen source for growth of *Rhodococcus*. Alternative nitrogen sources like ammonium sulphate, urea, and glycine a produce only suboptimal growth, possibly due to uptake of ammonia coupled with the release of H^+ ions, resulting in lower biomass driven by a faster stationary phase (Isleten-Hosoglu, Gultepe et al. 2012).

Inoculum size has a significant effect on the optimal growth rate, lipid production and synthesis of metabolites during fermentation (Ding, Tian et al. 2009, Carrau, Medina et al. 2010). High inoculum size results in short fermentation times to start simple batch fermentation at high cell density. However, conditions for growth and metabolism at high cell densities are less favourable due to hindered access to nutrients, space limitations, and cell interactions (Jarzębski, Malinowski et al. 1989). The difference in biomass could also be partly described by the differences in total initial concentrations of the nutrients in the cultures (95% vs 70% nutrients for 5% vs 30% inoculum size) as the nutrients get exhausted, it results in cessation of growth. This is supported by the conclusion that culture medium with a lower inoculum size could get hold a larger (~6 % larger) concentration of all nutrients contrary to the medium with larger inoculum (Skandamis, Stopforth et al. 2007, Carrau, Medina et al. 2010). However, this does not hold correct for the inoculum sizes below 5% that resulted in both suboptimal growth and TAG. Significantly lower inoculum size might fall below the threshold number of cells

required for utilizing the high concentration of nutrients and achieving the normal growth and thus lower biomass concentration and TAG.

In another characteristics the capability of *R. opacus* to grow in alkaline pH makes it a suitable strain for contamination free cultivation wherein the unnecessary contaminants can be reduced by maintaining an increased pH (Muthuraj, Kumar et al. 2014). For optimal growth and TAG production, though BM985 displayed metabolic flexibility regarding a wide pH range, its temperature requirements were centered at 28 ^{o}C; with TAG production dropping at 30 ^{o}C, with minor change in biomass. Thus, the temperature of 28 ^{o}C is a specific and unique characteristic of BM985.

Sodium acetate as carbon source has been reported to have maximum lipid induction in rhodococci (Thanapimmetha, Suwaleerat et al. 2016) due to their simple structure containing only two carbon atoms that can be easily assimilated for conversion to acetyl-CoA (Qiao and Wang 2009, Estévez-Landazábal, Barajas-Solano et al. 2013). Increase in the concentration of acetyl CoA increases the flux towards lipid biosynthesis pathway which explains the lipid induction property of sodium acetate (Muthuraj, Chandra et al. 2015). Similarly, magnesium is an important cofactor for many enzymes (acetyl-coA carboxylase) and has significant roles in most of the central metabolic pathways including lipid biosynthesis pathway (Janßen, Ibrahim et al. 2013).

All the above-mentioned parameters were optimized, and Fisher *F* test of the model was highly significant ($P<0.001$) with coefficient of determination (R^2) of 0.905. Therefore, 90.5% of the variation in the model could be well explained. In addition, the “lack of fit” analyses which evaluates the failure of the model was not significant ($P=0.29$). To yield maximum TAG production (output), the Generic PSO algorithm was used for identifying the optimal set of input conditions. In this study, the particle size between 30-100 was

investigated and observed no differences in examining the optimal output. Therefore, the number of particles and the learning factor c_1 and c_2 used were 30, 1.5, and 2.5, respectively. Once the number of generations reached 50, the inertia weight started with a large value of 0.9 and linearly decreased to 0.4. Further optimization was carried out with ANN. We carried out 1000 epoch; however, the optimum values were achieved at 9 epochs, having network topology of 4 inputs, 8 hidden layers and 1 output layer.

The values of GA specific parameters used in the optimization step were as follows: population size = 20, crossover probability = 0.8, mutation probability=0.01, No. of generation = 100. The global optimum values were obtained heuristically by training the GA multiple times with different initial parameters. The results of the various optimization tools clearly reveal that a higher C/N ratio is achieved by maintaining magnesium ion concentration, phosphate concentration and sodium acetate in defined media and simultaneously decreasing sodium nitrate concentration are the most substantial factors for enhancing lipid synthesis. In batch-culture with controlled experimental conditions, increased SDWW: $NaNO_3$ (higher C/N ratio) in the production media, led to enhanced fatty acids biosynthesis. These findings supported that higher carbon to nitrogen ratio (C/N) was required for increased TAG production.

The outcome of the reactor studies illustrated that the optimal conditions (pH: 7 and temperature: 28 ^{o}C) were more specific for BM985 than for WT, as the former was observed to produce twice the amount of TAG compared to later, despite similar biomass levels. This supported the well-known phenomenon that organisms' direct maximum energy flux to thrive and metabolize under optimized conditions, whereas under unfavourable conditions energy flux is directed towards cellular growth maintenance (6). Also, the lower biomass yield in case of BM985 under optimum conditions proves that

comparatively high DCW gets produced per carbon consumed than that in shake flasks. The COD removal by BM985 (88%) was considerably higher compared to 60% COD removal by *R. opacus* stated previously (Carrera, Vicent et al. 2004). Thus, the saturation of nitrogen and phosphorus in the nutrient medium was attributed to be the primary cause for lower COD removal. These results strongly endorse the *PAP2* overexpressing variant BM985 as a future potential candidate for biofuel production from dairy waste, authenticated by its ability to concomitantly treat dairy wastewater and enhance TAG.

The efficiency and commercial viability of biodiesels are ultimately dictated by constituent fatty acid and their esters. Consequently, their chain length, heat of combustion, exhaust emission, viscosity, degree of unsaturation oxidative stability and cold flow are the vital indices (Refaat 2009). The fatty acid composition of bacterial cells and the chemical properties of biodiesel obtained are significantly affected by the substrate used. With glucose and acetate as substrates, fatty acids are reported to vary from 14 to 18 carbon atoms (Alvarez, Mayer et al. 1996, Castro, Rocha et al. 2016, Bhatia, Kim et al. 2017). This is a favourable fatty acid profile for biodiesel productions. This biodiesel constitutes most of the straight chain fatty acids of which C-16 and C-18 contribute the major portions. In turn, the properties such as heat capacity and cetane number of biodiesels depend on the straight chain length of fatty acids (Dunn 2009, Knothe 2010). The combustion behaviour of the biodiesel is determined by the CN, which is the ignition delay time (the time between the injection and ignition). The CN of BM985 biodiesel satisfies the standard value fixed by European (EN14214) American (ASTM), and it is better compared to algal biodiesel (see Table 3.9). Like CN, iodine number is also considered to determine the quality of biodiesel and iodine number for FAMEs of BM985 was found to be less than maximum limit which demonstrates the highest quality of biodiesel. The osmotic stability (OS) value depends on the C18:2 and C18:3 contents

and this biodiesel is free of these fatty acids. It is related to stability, performance quality and effects the combustion behaviour. The unstable fuel owing to low OS may result in an increase in viscosity, formation of gums and other deposits that may plug the engine. Issues such as the fuel's oxidative stability and cold flow are the attributes of polyunsaturated and saturated fatty esters respectively (Pinto, Guarieiro et al. 2005, Demirbas 2008, Bhatia, Kim et al. 2017). The evaluated biodiesel properties, as shown in Table 3.9, based on empirical equations suggest it is of commercial biodiesel grade. Differential fatty acid composition accumulated from SDWW may be attributed to the utilization of ghee (lipids) as carbon source besides lactose and acetate by *Rhodococcal* cells.

An improved variant BM985 comparatively yielded more TAG than WT. BM985 grown in SDWW: MSM of 3:1 ratio yielded higher biomass concentration (~3.7 g/L of DCW) and TAG (68% of DCW) with optimal nitrogen source, pH, temperature and inoculum size and essential MSM salts. With significant COD removal (~88%) from DWW and fuel grade lipid production, BM985 justifies our preliminary objective of constructing a novel bacterial variant to achieve biodiesel production and dairy wastewater treatment simultaneously.

3.4 Conclusion

A superior mutant BM985 comparatively yielded more TAG than WT. BM985 grown in SDWW: MSM of 1:2 ratio yielded higher biomass (~37% DCW) and TAG (68% of DCW) with optimal nitrogen source, pH, temperature and inoculum size besides MSM salts. With significant COD removal (~88%) from DWW and fuel grade lipid production,

BM985 may be suggested to exploit simultaneously for biodiesel production and wastewater treatment.

3.5 References

1. Alvarez, H. M., F. Mayer, D. Fabritius and A. Steinbüchel (1996). "Formation of intracytoplasmic lipid inclusions by Rhodococcus opacus strain PD630." Archives of Microbiology **165**.
2. Alvarez, H. M., F. Mayer, D. Fabritius and A. Steinbüchel (1996). "Formation of intracytoplasmic lipid inclusions by Rhodococcus opacus strain PD630." Archives of Microbiology **165**(6): 377-386.
3. Alvarez, H. M. and A. Steinbüchel (2002). "Triacylglycerols in prokaryotic microorganisms." Appl Microbiol Biotechnol **60**.
4. Bhatia, S. K., R. K. Bhatia and Y.-H. Yang (2017). "An overview of microdiesel — A sustainable future source of renewable energy." Renewable and Sustainable Energy Reviews **79**: 1078-1090.
5. Bhatia, S. K., J. Kim, H.-S. Song, H. J. Kim, J.-M. Jeon, G. Sathiyanarayanan, J.-J. Yoon, K. Park, Y.-G. Kim and Y.-H. Yang (2017). "Microbial biodiesel production from oil palm biomass hydrolysate using marine Rhodococcus sp. YHY01." Bioresource Technology **233**: 99-109.
6. Bhatia, S. K., D.-H. Yi, Y.-H. Kim, H.-J. Kim, H.-M. Seo, J.-H. Lee, J.-H. Kim, J.-M. Jeon, K.-S. Jang, Y.-G. Kim and Y.-H. Yang (2015). "Development of semi-synthetic microbial consortia of Streptomyces coelicolor for increased production of biodiesel (fatty acid methyl esters)." Fuel **159**: 189-196.
7. Blasco R, Martínez-Luque M, Madrid MP, Castillo F and M.-V. C (2001). "Rhodococcus sp. RB1 grows in the presence of high nitrate and nitrite concentrations and assimilates nitrate in moderately saline environments." Archives Microbiology **175**(6): 435-440.
8. Brandl, H., R. A. Gross, R. W. Lenz and R. C. Fuller (1988). "Pseudomonas oleovorans as a Source of Poly(β-Hydroxyalkanoates) for Potential Applications as Biodegradable Polyesters." Applied and Environmental Microbiology **54**(8): 1977-1982.
9. Carrau, F., K. Medina, L. Fariña, E. Boido and E. Dellacassa (2010). "Effect of Saccharomyces cerevisiae inoculum size on wine fermentation aroma compounds and its relation with assimilable nitrogen content." International Journal of Food Microbiology **143**(1–2): 81-85.
10. Carrera, J., T. Vicent and J. Lafuente (2004). "Effect of influent COD/N ratio on biological nitrogen removal (BNR) from high-strength ammonium industrial wastewater." Process Biochemistry **39**(12): 2035-2041.
11. Castro, A. R., I. Rocha, M. M. Alves and M. A. Pereira (2016). "Rhodococcus opacus B4: a promising bacterium for production of biofuels and biobased chemicals." AMB Express **6**(1): 35.

12. Clesceri, L. S., A. E. Greenberg and A. D. Eaton (1998). Standard Methods for the Examination of Water and Wastewater, 20th Edition, APHA American Public Health Association.

13. Demirbas, A. (2008). "Relationships derived from physical properties of vegetable oil and biodiesel fuels." Fuel **87**(8): 1743-1748.

14. Ding, M.-Z., H.-C. Tian, J.-S. Cheng and Y.-J. Yuan (2009). "Inoculum size-dependent interactive regulation of metabolism and stress response of Saccharomyces cerevisiae revealed by comparative metabolomics." Journal of Biotechnology **144**(4): 279-286.

15. Dunn, R. O. (2009). "Effects of minor constituents on cold flow properties and performance of biodiesel." Progress in Energy and Combustion Science **35**(6): 481-489.

16. Eberhart, R. and J. Kennedy (1995). A new optimizer using particle swarm theory. Micro Machine and Human Science, 1995. MHS '95., Proceedings of the Sixth International Symposium on.

17. Estévez-Landazábal, L.-L., A.-F. Barajas-Solano, C. Barajas-Ferreira and V. Kafarov (2013). "IMPROVEMENT OF LIPID PRODUCTIVITY ON Chlorella vulgaris USING WASTE GLYCEROL AND SODIUM ACETATE." CT&F - Ciencia, Tecnología y Futuro **5**: 113-126.

18. Fei, Q., M. O'Brien, R. Nelson, X. Chen, A. Lowell and N. Dowe (2016). "Enhanced lipid production by Rhodosporidium toruloides using different fed-batch feeding strategies with lignocellulosic hydrolysate as the sole carbon source." Biotechnology for Biofuels **9**(1): 130.

19. Francisco, É. C., D. B. Neves, E. Jacob-Lopes and T. T. Franco (2010). "Microalgae as feedstock for biodiesel production: Carbon dioxide sequestration, lipid production and biofuel quality." Journal of Chemical Technology & Biotechnology **85**(3): 395-403.

20. Hernández, M., S. Comba, A. Arabolaza, H. Gramajo and H. Alvarez (2015). "Overexpression of a phosphatidic acid phosphatase type 2 leads to an increase in triacylglycerol production in oleaginous Rhodococcus strains." Applied Microbiology and Biotechnology **99**(5): 2191-2207.

21. Huang, J., L.-H. Mei and J. Xia (2007). "Application of artificial neural network coupling particle swarm optimization algorithm to biocatalytic production of GABA." Biotechnology and Bioengineering **96**(5): 924-931.

22. Isleten-Hosoglu, M., I. Gultepe and M. Elibol (2012). "Optimization of carbon and nitrogen sources for biomass and lipid production by Chlorella saccharophila under heterotrophic conditions and development of Nile red fluorescence based method for quantification of its neutral lipid content." Biochemical Engineering Journal **61**(Supplement C): 11-19.

23. Janßen, H. J., M. H. Ibrahim, D. Bröker and A. Steinbüchel (2013). "Optimization of macroelement concentrations, pH and osmolarity for triacylglycerol accumulation in Rhodococcus opacus strain PD630." AMB Express **3**.

24. Jarzębski, A. B., J. J. Malinowski and G. Goma (1989). "Modeling of ethanol fermentation at high yeast concentrations." Biotechnology and Bioengineering **34**(9): 1225-1230.

25. Jungo, C., I. Marison and U. von Stockar (2007). "Mixed feeds of glycerol and methanol can improve the performance of Pichia pastoris cultures: A quantitative study based on concentration gradients in transient continuous cultures." Journal of Biotechnology **128**(4): 824-837.

26. K.M. Desai, S.A. Survase, P.S. Saudagar, S.S. Lele and R. S. Singhal (2008). "Comparison of artificial neural network (ANN) and response surface methodology (RSM) in fermentation media optimization: case study of fermentative production of scleroglucan." Biochem Eng J **41**: 266-273.

27. Khaouane, L., C. Si-Moussa, S. Hanini and O. Benkortbi (2012). "Optimization of culture conditions for the production of Pleuromutilin from Pleurotus Mutilus using a hybrid method based on central composite design, neural network, and particle swarm optimization." Biotechnology and Bioprocess Engineering **17**(5): 1048-1054.

28. Knothe, G. (2010). "Biodiesel and renewable diesel: a comparison." Prog Energy Combust Sci **36**.

29. Kumar, V., M. Muthuraj, B. Palabhanvi, A. K. Ghoshal and D. Das (2014). "Evaluation and optimization of two stage sequential in situ transesterification process for fatty acid methyl ester quantification from microalgae." Renewable Energy **68**(Supplement C): 560-569.

30. Leal, M. C. M. R., D. M. G. Freire, M. C. Cammarota and G. L. Sant'Anna (2006). "Effect of enzymatic hydrolysis on anaerobic treatment of dairy wastewater." Process Biochemistry **41**(5): 1173-1178.

31. M. Hhayet and C. Cojocaru (2012). "Artificial neural network modeling and optimization of desalination by air gap membrane distillation." Sep Purif Technol **86**: 171-182.

32. Muthuraj, M., N. Chandra, B. Palabhanvi, V. Kumar and D. Das (2015). "Process Engineering for High-Cell-Density Cultivation of Lipid Rich Microalgal Biomass of Chlorella sp. FC2 IITG." BioEnergy Research **8**(2): 726-739.

33. Muthuraj, M., V. Kumar, B. Palabhanvi and D. Das (2014). "Evaluation of indigenous microalgal isolate Chlorella sp. FC2 IITG as a cell factory for biodiesel production and scale up in outdoor conditions." Journal of Industrial Microbiology & Biotechnology **41**(3): 499-511.

34. Pacheco, G. J., E. M. P. Ciapina, E. d. B. Gomes and N. P. Junior (2010). "Biosurfactant Production by Rhodococcus Erythropolis and its Application to Oil Removal." Brazilian Journal of Microbiology **41**(3): 685-693.

35. Pinto, A. C., L. L. N. Guarieiro, M. J. C. Rezende, N. M. Ribeiro, E. A. Torres, W. A. Lopes, P. A. d. P. Pereira and J. B. d. Andrade (2005). "Biodiesel: an overview." Journal of the Brazilian Chemical Society **16**: 1313-1330.

36. Prabhu A, A., S. Chityala, Y. Garg and V. Venkata Dasu (2017). "Reverse micellar extraction of papain with cationic detergent based system: An

optimization approach." Preparative Biochemistry and Biotechnology **47**(3): 236-244.

37. Prabhu, A. A. and A. Jayadeep (2017). "Optimization of enzyme-assisted improvement of polyphenols and free radical scavenging activity in red rice bran: A statistical and neural network-based approach." Preparative Biochemistry and Biotechnology **47**(4): 397-405.

38. Prabhu, A. A., A. Purkayastha, B. Mandal, J. P. Kumar, B. B. Mandal and V. D. Veeranki (2017). "A novel reverse micellar purification strategy for histidine tagged human interferon gamma (hIFN-γ) protein from Pichia pastoris." International Journal of Biological Macromolecules.

39. Qiao, H. and G. Wang (2009). "Effect of carbon source on growth and lipid accumulation in Chlorella sorokiniana GXNN01." Chinese Journal of Oceanology and Limnology **27**(4): 762.

40. Refaat, A. A. (2009). "Correlation between the chemical structure of biodiesel and its physical properties." International Journal of Environmental Science & Technology **6**(4): 677-694.

41. Silva, H. R., C. E. C. Prete, F. Zambrano, V. H. de Mello, C. A. Tischer and D. S. Andrade (2016). "Combining glucose and sodium acetate improves the growth of Neochloris oleoabundans under mixotrophic conditions." AMB Express **6**: 10.

42. Skandamis, P. N., J. D. Stopforth, P. A. Kendall, K. E. Belk, J. A. Scanga, G. C. Smith and J. N. Sofos (2007). "Modeling the effect of inoculum size and acid adaptation on growth/no growth interface of Escherichia coli O157:H7." International Journal of Food Microbiology **120**(3): 237-249.

43. Stemmler, K., R. Massimi and A. E. Kirkwood (2016). "Growth and fatty acid characterization of microalgae isolated from municipal waste-treatment systems and the potential role of algal-associated bacteria in feedstock production." PeerJ **4**: e1780.

44. Su, Y.-C., Y. A. Liu, C. A. Diaz Tovar and R. Gani (2011). "Selection of Prediction Methods for Thermophysical Properties for Process Modeling and Product Design of Biodiesel Manufacturing." Industrial & Engineering Chemistry Research **50**(11): 6809-6836.

45. Thanapimmetha, A., T. Suwaleerat, M. Saisriyoot, Y. Chisti and P. Srinophakun (2016). "Production of carotenoids and lipids by Rhodococcus opacus PD630 in batch and fed-batch culture." Bioprocess and Biosystems Engineering: 1-11.

46. Timm, A., D. Byrom and A. Steinbüchel (1990). "Formation of blends of various poly(3-hydroxyalkanoic acids) by a recombinant strain of Pseudomonas oleovorans." Applied Microbiology and Biotechnology **33**(3): 296-301.

47. Xia, C., J. Zhang, W. Zhang and B. Hu (2011). "A new cultivation method for microbial oil production: cell pelletization and lipid accumulation by Mucor circinelloides." Biotechnology for Biofuels **4**(1): 15.

48. Y. Yasin, F. Bin, H. Ahmad, M. Ghaffari-Moghadamc and M. Khajeh (2014). "Application of a hybrid artificial neural network–genetic algorithm approach to optimize the lead ions removal from aqueous solutions using intercalated

tartrate-Mg–Al layered double hydroxides." Environmental Nanotechnology, Monitoring & Management **1-2**: 2-7.

49. Zhang, J.-R., J. Zhang, T.-M. Lok and M. R. Lyu (2007). "A hybrid particle swarm optimization–back-propagation algorithm for feedforward neural network training." Applied Mathematics and Computation **185**(2): 1026-1037.

Chapter 4

Reconstruction and validation of a Genome-Scale Metabolic model of *Rhodococcus opacus* PD630 for better understanding TAGs biosynthesis

4.1 Outline

GeMs have been successfully used to allocate functionality to annotated genome sequences by providing a blueprint of how the cell's biochemical components interact molecularly to affect the evolving cell phenotype. In combination with FBA, GeMs can be utilized to explore the metabolic potential of any organism for which the stoichiometry of the participating biochemical reactions is well-established. FBA assumes that the metabolic network is in a pseudo-steady state and consequently mass balance constraints can be imposed to reflect any given set of physiological metabolite uptake and secretion rates. This results in an underdetermined problem which, is routinely formulated as a linear programming problem that can estimate the distribution of carbon, energy, and metabolic assets over the entire network. Herein, we present and validate against literature and experimental data a metabolic network reconstruction for *R. opacus* PD630 (*i*ROPD630). The *i*ROPD630 model was used to investigate growth under two different carbon sources. An Artificial Centering Hit and Run (ACHR) algorithm was used to generate a representative number of sampled flux distributions for each condition. Principal component analysis (PCA) was performed to identify key metabolic differences between the two conditions. The *i*ROPD630 GeM can be used further to understand *R. opacus* PD630 metabolism and to identify targets for metabolic engineering to improve the production of TAG.

4.2 Materials and methods

The systematic steps involve in the reconstruction of metabolic model of *R. opacus* PD630 and analysis thereof are illustrated in the Figure 4.1.

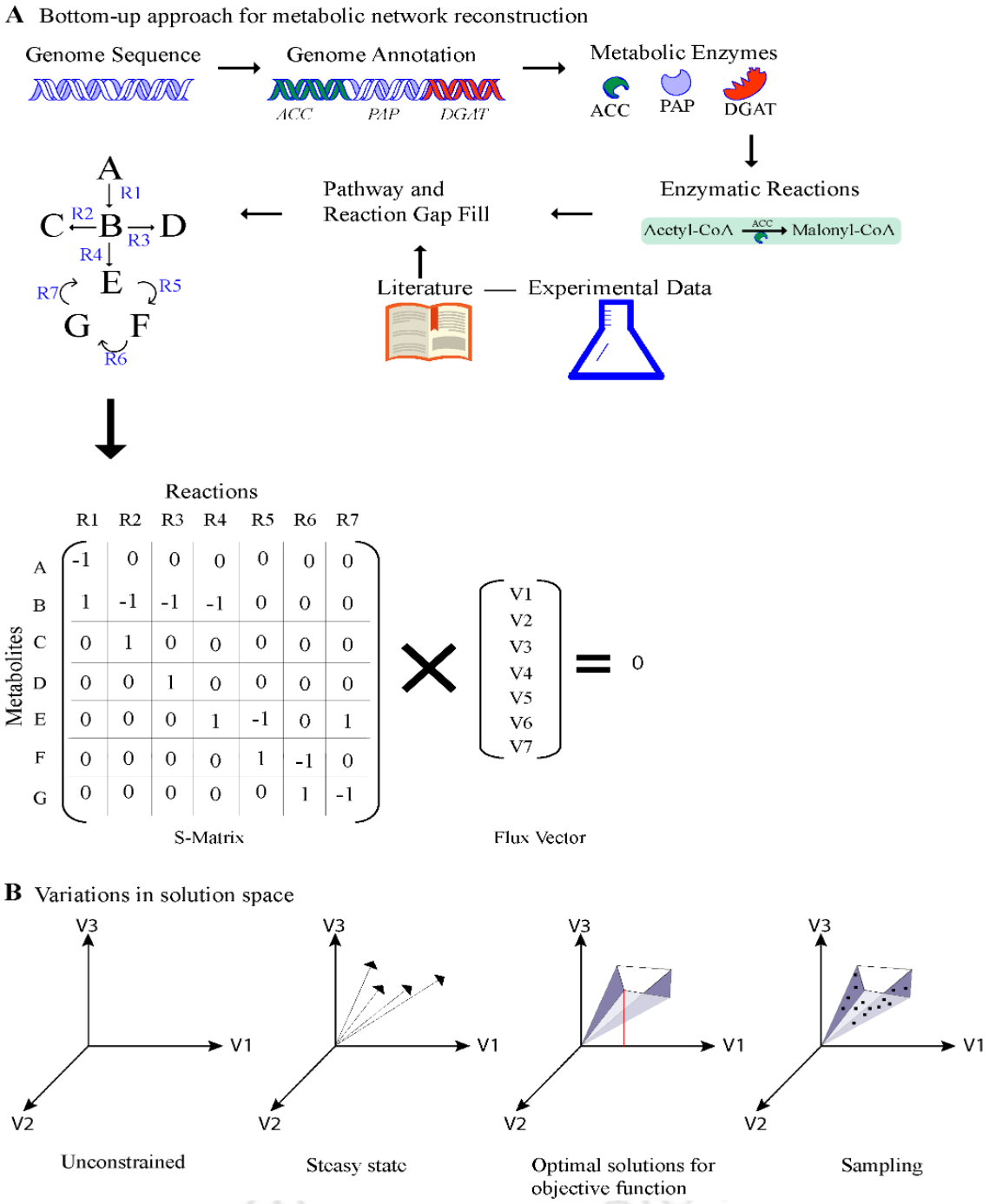

Figure 4.1: Steps involved in the bottom-up approach for GEM reconstruction of Rhodococcus opacus PD630 *A and its analysis through constraint-based approach B. The steps in the A illustrates how genome sequence can be used to arrive at a metabolic mathematical model*

4.2.1 Materials

4.2.1.1 Genome sequence databases

Genome sequencing has widely been accepted as one of the better tools for preparing a metabolic model of an organism. However, for the analysis, a file in FASTA format is required along with a set of coding sequences that need to be studied. At the time of constructing a genome scale model, certain aspects must be well taken care of, such as placements of open reading frames (ORFs), genome assembly, structural and functional annotation. This information can be attained from a handful of reliable databases available, like GOLD, TIGR, NCBI and EBI (Thiele and Palsson 2010). The whole genome sequence of *R. opacus* PD630 was taken from NCBI for the reconstruction of GeM. At times when the databases fail to provide information on role of certain ORFs putative functions extracted from orthology-based applications like BLAST or Inparanoid may be considered (Baart, Langenhof et al. 2010, Baart and Martens 2012). With the help of these applications, one can conveniently find the putative function of the gene from the similarity score and can further consider adding it to the reaction model. A similar strategy was employed to include around 400 missing enzymes and thus reactions from the published metabolic model of the closely related genus *Rhodococcus jostii* RHA1 (Tajparast and Frigon 2015). Additionally, membrane transport reactions are equally significant and should be included along with the inclusion genes or ORFs. As stated, the first step of the model is to define the functional annotation meaning determining the specific functions of each ORF. Part of these ORFs shall also code for enzymes that are involved in the regulation of metabolic pathways. The reaction stoichiometry along with information on reversibility of the reaction is obtained for such enzymes, which is then next incorporated. Finally, the reaction of biomass formation is included which is a sum of several biochemical reactions and not connected to just a single enzyme.

4.2.1.2 Online reconstruction tools for the first draft generation

A combination of RAST and Model SEED pipelines are quite popular in rapidly producing high quality evaluation of gene functions. For bacterial and archaeal genome, they provide a completely automated and model reconstruction service which includes auto-filling of metabolic gaps and FBA of the model (Aziz, Bartels et al. 2008). Every model network consists of all the reactions that are associated with an organism's genome and set of reactions which are devoid of enzymatic catalysis as well (Aziz, Bartels et al. 2008, Henry, DeJongh et al. 2010). Metabolic network construction happens to be quite a tedious process and existence of such automated tools makes it a little easier for model drafting. It took ~48 hours to reconstruct a metabolic model from an assembled genome sequence.

Above-mentioned automated application is well known for time saving to a significant level; however, a manual inspection is still required to ensure that all the possible errors and variations have been eliminated. Determination of exact reaction stoichiometry is required once the genome has been effectively mined. Reaction stoichiometry includes knowledge of co-factors, reaction reversibility/irreversibility and elemental balances of each reaction. Regarding co-factors, whether both NADH and NADPH were included as the exact candidate was not known, Gibb's free energy was used to calculate the reversibility/irreversibility of the reaction which led to an apt thermodynamic model constraint in the direction of reactions. Defining compartments and localization of enzymes in these compartments are key aspects in metabolic network construction. This was conveniently achieved by software tools like SignalP and TargetP (Thiele and Palsson 2010). A compartment specific naming is essential in order to distinguish between same metabolites positioned in different compartments. For an example, M[e0] metabolite can be named for the metabolite M present extracellularly and M[c0] can be

named for the same metabolite located in the cytosol. Biomass formation is the association of macromolecules (DNA, RNA, proteins, carbohydrates and lipids) which is the last step in the metabolic network construction. The macromolecules taking part in the biomass formation are a result of their respective precursors. The reaction stoichiometry vastly depends on composition of biomass and biopolymers. For an instance, the reaction stoichiometry for protein can be directly derived from its amino acid sequence which demonstrates the importance of determination of biomass composition; one such component is ATP which must be included as per unit biomass (Baart and Martens 2012).

4.2.2 Methods

4.2.2.1 Construction of a draft metabolic model

Most application tools listed allows CDS or coding sequences, which in this case obtained from NCBI, to be uploaded and in return generate a SBML- formatted metabolic model. One such example is RAST which can generate a metabolic model draft by selecting "build metabolic model" option starting the genome annotation process. All such application tools generate a SBML metabolic model draft which further needs to be manually examined before one considers turning that into a functional model (Fondi and Liò 2015).

Subsystem and non-subsystem based assertions are the two classes in RAST server to implement asserted gene functions. The subsystem-based assertion system is based upon recognizing the functional variants whereas the non-subsystem integrates evidence from other common tools. RAST uses the subsystem based assertion system which is more reliable and accurate as which makes RAST a superiorly dependable and be exploited as a better start point to build a comprehensive annotation (Aziz, Bartels et al. 2008).

SEED is basically a community-based genome annotation and analysis environment on which we have applied our approach. SEED is capable to implement a sub system approach which is implementation of cross-organismal methodology. The subsystem is defined as a set that consists the related gene products and another set that encodes the relations between functions and those specific genes forms the products in particular organisms. One very important role of this subsystem in SEED is the enzymes, which have active functions in the metabolic pathways demonstrated by the subsystems. Furthermore, SEED entails subsystems that shelter the central and intermediary metabolism in many various diverse organisms. These functions provided by SEED are in the form of organism-independent style. The superior functions of SEED like determining related functional roles of gene products, grouping them all together, associating specific genes with their designated functions across many genomes and determining functional variants make this approach a better path to construct our metabolic model (DeJongh, Formsma et al. 2007).

4.2.2.2 Preliminary model reconstruction through Model SEED pipeline

To generate the initial draft models for *R. opacus* PD630 Model SEED pipeline was taken into account. The RAST sever was used to import the genome sequence of *R. opacus* into the SEED system which further would perform gene and genome sequence annotations in approximately 15-24 hours. The pipeline system continues by constructing a primary model of each organism using RAST annotations. This primary model analysis has certain functions like, reaction networks along with GPR associations, Gibbs free energy values, biomass reaction which includes cell wall components, lipids and cofactors. Every preliminary reaction consists of all the reactions catalysed by respective enzymes coded in the definite organism's genome and other spontaneous reaction pathways which do not include enzymes to form the products. A mapping between biochemical pathways,

genome annotations and specific role of genes assigned during RAST annotations are significant in generating GPR associations of each particular reaction pathway in the model network system. This is mostly used in a situation to distinguish products formed from multiple genes giving rise to a protein complex which further catalyses a reaction pathway and where products from genes may act independently to catalyse the same reaction pathways. However, these GPR associations are well defined and accurate but a manual curation is always recommended (DeJongh, Formsma et al. 2007, Henry, DeJongh et al. 2010).

4.2.2.3 Automatic completion of model gaps of the model

At most times there are gaps generated at the primary level of building a model and consequently it inhibits the synthesis of one or several biomass components in the reaction pathway. The "auto-completion" process integrated into the SEED pipeline exists an optimization algorithm that is capable to identify the minimal set of reactions that have to be added to fill the gaps (Kumar, Dasika et al. 2007, Suthers, Dasika et al. 2009). This process in the SEED guarantees to produce the metabolic functions which are anticipated to be missing from the genome annotations and makes the metabolic model efficient of simulating cell growth. The reactions are all selected from a database that provides mass and charge balanced reactions in aqueous conditions and at a neutral pH. This database will collect all the reactions from KEGG pathway and shall list into a single, non-redundant set (Du, Yuan et al. 2014, Kanehisa, Sato et al. 2018). The number of metabolic genes and thus their corresponding functions likely to have missed from genome annotation was listed by the autocompletion process with the assumption that every SEED model is entitled to simulate cell growth (Henry, DeJongh et al. 2010). MATLAB is one community-driven software tool that plays a very essential role in model curation. One of the most prominent toolbox available for model curation in the COBRA

Toolbox available in MATLAB (Heirendt, Arreckx et al. 2019). Cobra toolbox is primarily used for refining and interpreting constraint-based network models with the help of its given built-in functions (Schellenberger, Que et al. 2011, Heirendt, Arreckx et al. 2019). One such function 'fastGapFill' from cobra toolbox and auto-completion from SEED has been employed to curate the first rough draft.

4.2.2.4 Model evaluation and addition of specific exchange reactions

It should be treated essential, at this stage, to check probable sources of errors before starting of modelling procedures because without this the reconstructed model may be inadequate and lack metabolic genes and/or functions. Unlike, draft model which is created from the preliminary reconstruction are "analysis ready" as they are capable of simulating biomass production and nutrients, our model failed responding to the auto-completion procedure tried on ModelSEED and fastGapFill function present in cobra box.

In order to classify the reactions in SEED as active, inactive or essential a flux variability analysis named (FVA) (Mahadevan and Schilling 2003) was used which comes under the "model analysis" step in the ModelSEED pipeline. The essential and active are classified as the ones which must carry flux for basic growth of the organism and must carry flux but are not essential for growth, respectively. The inactive classification of flux does not carry any flux and are inactive gaps as well in the metabolic model. This classification needs further manual curation. To match the quality and accuracy of most published models, further manual curation of the model created by ModelSEED is required (Henry, DeJongh et al. 2010).

In case of potentially missing reactions the following steps were implemented

1. In view of the described alternative metabolic reactions for a specific pathway, the databases and scientific pieces of literature pertaining to *Rhodococcus opacus* PD630 were examined.
2. Comparative genomics of closely related microorganism, *Rhodococcus jostii* RHA1 was used. At this stage the SEED model viewer can help much because multiple reconstructions can be concurrently shown onto the same map. To do this, other models of closely related organisms from the initial page may be selected and then clicked on the biosynthetic pathway that is required to be examined in detail. Prospective errors during the metabolic reconstruction of the other strain may be suggested by the presence or absence of any reactions in any of the two organisms. The sequence of the enzyme encoding for that reaction may be retrieved to validate this indication in one of the organisms possessing it and a BLAST search in the genome of the organism missing it may also be performed. A strong indication for gap filling the corresponding metabolic step is the presence of a homologous sequence in the probed genome. For retrieving information on the metabolic features of closely related microorganisms databases such as KEGG (Karlsen, Schulz et al. 2018) and MetaCyc (Karp, Riley et al. 2000) were explored.

4.2.2.5 Refining the model through manual curation

- In the current study, we had compared our results with ModelSEED pipeline (fig 1) with the one published in metabolic reconstruction protocol, we observed that the pipeline replicated 73 times out of 82. Moreover, most of the initial 42 steps in the automated reconstruction steps were automated (Henry, DeJongh et al. 2010). The step that requires the most attention and is important for any model to be viable is the ‘manual model curation’ step. Before employing any techniques

for model curation, one needs to understand the basic nature of any organism to produce biomass as a survival instinct. Similarly, the GeM was constructed and curated with the main purpose of fulfilling and providing the basic needs such as DNA, RNA, amino acids, phospholipids, peptidoglycan, cofactors and growth/non-growth associated ATP requirements of *R. opacus* PD630 biomass by fermentation of glucose/lactose. The biomass reaction for *R. opacus* PD630 was taken from the published model of related organism *Rhodococcus jostii* RHA1 as they share 95% similarity. Based on the above prerequisite and need we employed three main model curation techniques (Karp, Weaver et al. 2018):

- Manual inference from biochemical requirement, (e.g., a need to synthesize an amino acid)
- Manual inference from text or EC number in the genome annotation
- Hidden Markov Model (HMM) searches against the genome sequence

The current model for manual gap filling process was focussed on the principle of ability of genome to self-reproduce itself within basic nutritional context and that:

(a) The metabolic model should be able to attain a basic biomass synthesis objective in terms of demonstrating the metabolic reactions concerning amino acids, nucleotides lipids etc.

(b) The model should be able to demonstrate such metabolic activity while carrying out the basic aerobic fermentation along with successfully maintaining the potential across the membrane.

(c) The model should produce forward reactions favouring the redox and phosphorylation processes which excludes reduction of $NADP^+$ by NADH and formation of ATP without being associated with an energy yielding reaction.

Briefly, the metabolic model was built using the reactions catalysed by identified genes as a guide, starting with glucose and moving forward through the pathways and reactions to reach the objective metabolites. We conducted curation to fix the model's gaps by inserting metabolites that had been missed in the automated metabolite-import process required for reaction instantiation.

Our manual curation was primarily based on refining the connectivity and annotation of the metabolic components in order to identify varying compounds' names and to fill the reaction gaps in the network. Failing to synthesize the essential biomass precursor or unstable products leads to incorrect reaction during the metabolic simulations. To this end, it can be reverted by deriving new annotations from well-designed metabolic databases, the connectivity of peripheral pathways, in contrast is progressively improved. Locating network gaps is to allow the synthesis of biomass precursors which is widely examined by KEGG PATHWAY, which is a quick and accurate application to locate network gaps. The second objective of the same can be achieved by referring to the biochemical literature and likewise reproducing the known metabolic functions. Similarly, one can consider scanning through the metabolic pathway of an organism visually to look for the missing reactions. This mostly helps in finding the synthetic pathways for the components of biomass. These pathways are found to be usually linear and can be further checked against literature (Quek and Nielsen 2008).

The metabolic model was fundamentally built considering glucose as the first energy source and then gradually going down into pathways with various linear and branched reactions to identify the target metabolite. The model was further repaired by adding compounds to the reactions so as to heal the network gaps which was bounced by the automatic metabolite import process.

Before applying FVA following the failure to respond to auto-completion process and Fastgapfill in SEED and cobra toolbox respectively, the reconstructed model was compared manually with the most related and similar organism *Rhodococcus jostii* RHA1 to check for any missing reactions. 360 genes found exclusively in the published *Rhodococcus jostii* RHA1 model was excluded in the RAST and SEED generated model in light of the fact that either the genes doled out to these qualities in the SEED were conflicting with the reactions mapped to them in the published models or the genes were not explicit enough to take into account mapping to express reactions. Each of the missing genes was search against *R. opacus* PD630 genome sequence using blastn for any match.

4.2.2.6 Assign in-silico medium composition

The lower bounds (lb) and upper bound (ub) of exchange reactions can help in defining the growth medium constituents. These two bounds play a crucial role in determining maximum utilization rate of every compound that must be incorporated into the model through exchange reactions. Usually, lb values are used for defining the uptake rate of a given compound in the corresponding exchange reaction. To this end, inorganic ions and water are present in a non-limiting concentration and thus the lb for their corresponding exchange reaction is set to very high values.

4.2.2.7 Optimize model for biomass production

Biomass formation is an integral part during construction of metabolic model and should be given utmost importance. The reactions that contribute for biomass can be considered in two ways, first by considering set of reactions that denote macromolecules into biomass and second, set of building blocks that constitute the biomass. As precursor fractions are converted to mmol $gDCW^{-1}h^{-1}$, conventionally, the biomass reaction is expressed in h^{-1}. The mole fraction of each precursor necessary to produce 1 g dry weight of cells is

determined by the biomass assembly reaction (Thiele and Palsson 2010). Therefore, at this point, a detailed information on biomass composition should be available. Missing pieces of information may be derived from the biomass composition from closely related strains if available experimental data is not enough. We have borrowed the biomass reaction together with growth-associated ATP maintenance (GAM) from the *Rhodococcus jostii* RHA1.

Linear programming can be considered to examine the flux distribution for optimization which alters the output of the specific set of reaction as objective function. By setting biomass production as the model objection function, one is enabled to test the compounds that are involved in the biomass assembly and whether those can be synthesized (see 4.3.3). In the first case, the flux out of the biomass (f) assembly will be greater than 0; conversely, in case one (or more) biomass constituent(s) cannot be produced, f will be equal to 0.

4.2.2.8 Model validation against experimental data sets

Genome-scale metabolic models such as the one described in this work can be used to calculate experimentally verifiable phenotypic predictions. One of the key scientific uses of these models is to enable systematic improvement of our current knowledge of metabolic networks by comparing model predictions to experimental data and identify potential improvements to the model, which then in turn can be used to design new experiments.

4.2.2.8.1 *In- silico* growth conditions

COBRA toolbox (Heirendt, Arreckx et al. 2019) and FBA (Orth, Thiele et al. 2010) were used to determine the growth rate and metabolite production. The predictions of the model were all compared with data sets from both experimental and reported literature data

(Kurosawa, Boccazzi et al. 2010, Mandal, Prabhu et al. 2019). To this end, the capacity of the strain to grow in minimal condition was achieved by *in-silico* simulations. Unlimited uptake of certain elements was given to simulate the media composition. Such elements were: H^+, H_2O, Mg^{2+}, Mn^{2+}, Co^{2+}, Cu^{2+}, Ca^{2+}, Cl^-, Fe^{2+}, K^+, Na^+, Mo^{7+} and Zn^{2+}. The lower bounds were set to −100 mmol $gDCW^{-1}h^{-1}$ for the respective exchange reactions and default lower bounds were set to −10 mmol $gDCW^{-1}h^{-1}$. In the default category, carbon, nitrogen, phosphorus and sulphur constituents were glucose, ammonium, phosphate and sulphate respectively. In order to identify the growth promoting nutrients, all the default nutrients were removed one at a time with a lower bound set to 0 mmol $gDCW^{-1}h^{-1}$and subsequently FBA was used to determine the growth rates. The aerobic condition for the microbe was simulated by limiting oxygen intake and the rate was fixed to 10 mmol $gDCW^{-1}h^{-1}$.

The exchange reactions (conventionally regarded as "EX") defines the composition of the growth medium and environmental conditions at the time of simulation. This defines the concentration range of compounds that are incorporated into the metabolic model and metabolites those are synthesized, to form biomass or other cellular substituents. To execute this, the draft model included minimal set of exchange reactions.

4.2.2.8.2 Gene expression and model validation

To examine the expression of *PAP2* gene *Rhodococcus opacus* BM985 from (Mandal, Prabhu et al. 2019) was grown in 50 mL lactose minimal salt medium having 10 g/L lactose, 9 g/L $Na_2HPO_4{\cdot}12H_2O$, 1.5 g/L KH_2PO_4, 0.2 g/L $MgSO_4{\cdot}7H_2O$, 0.1 g/L NH_4SO_4, 0.05 g/L $NaHCO_3$, 0.02 g/L $CaCl_2{\cdot}2H_2O$, and 1.2 mg/L ferric ammonium citrate with each concentration being converted to mmol $gDCW^{-1}h^{-1}$ for model simulation.

Real time PCR for *PAP2* gene with internal control for *Rhodococcus opacus* PD630 was performed in triplicate as according to (DeLorenzo and Moon 2018). The fold change in *PAP2* gene expression form BM985 is compared with *PAP2* gene from *R. opacus* PD630. The model was evaluated for the flux change using flux balance analysis (FBA) through the reaction R02239 involving *PAP2* with TAG as objective function for both BM985 and *R. opacus* PD630.

4.2.2.9 Flux balance analysis

A stoichiometric matrix is at the core of FBA. It is a mathematical representation of collection of mass-balanced biochemical reactions that completely explains the established topology of a cell's metabolic network (Varma and Palsson 1994).

Hereby, the equation is mentioned which explains the mass balance constraints or stoichiometric:

$S \bullet v = 0$

In this equation, S represents the stoichiometric matrix and v is the vector which defines the metabolic fluxes which includes both internal and external fluxes. To identify the solution which would maximize an objective function, a metabolic flux was expressed with the aid of a linear programming program. The formulated linear programming problem is mentioned below:

Minimize Z subject to:

$$S \bullet v = 0$$

$$\alpha i \leq vi \leq \beta i$$

where

$$Z = \sum_{i=0}^{n} ci.vi = < c \bullet v >$$

In this equation c represents the vector which was used to select linear combination of metabolic fluxes to be included at the objective function (Varma and Palsson 1993). For this specific example, c is a vector of zeros with a one in the position of the growth rate flux which was defined as the unit vector in the growth flux direction (Edwards, Ibarra et al. 2001). An acclaimed challenge related to enormous genome-scale metabolic systems is their underdetermined nature. Therefore, instead of a unique solution, FBA and similar approaches return a non-unique solutions which however, is in agreement with the imposed constraints. The more extensive the non-unique solutions, the higher the vulnerability in the assurance of essential reaction properties. This in turn constrains the interpretability of and trust in the outcomes. To circumvent this issue and quantity the solution space one can perform the FVA. It enables the minimum and maximum permissible fluxes to be identified under the constraints imposed by each reaction in the reconstructed metabolic network (Mahadevan and Schilling 2003).

The task that remained was to connect these extra added reactions to rest of the network and check for the model's viability. FBA is a linear optimization to interpret whether the known metabolic functions be mimicked *in-silico* (Domach, Leung et al. 2000) and for such job done, a superior mathematical modelling and computer simulation is required. In this regard, for our current study, we had to connect the extra added reactions to the network model and interpret its feasibility. The newly reconstructed network model was to analyse and predict the flux distribution. The competences of a reconstructed metabolic model are based on synthetic mass-balance and reaction capacity which can be analysed using FBA (Edwards 1999). FBA therefore cannot provide any solution for the incomplete set of constraints and provides a solution space for all the possible outcomes of steady state flux distribution that placate the applied constraints. The capacity constraints are considered for the transport reactions where these are used to set the rate

for uptake reactions and gives the rate of reversibility of every metabolic reaction. Therefore, these capacity constraints were in the network model for every flux value. The transport fluxes are considered zero for the metabolites not found in the media and capacity constraints were varied on rate of uptake on carbon and oxygen to construct phenotype phase planes. Furthermore, the metabolic bi-products were permissible to leave the system and transportation of inorganic phosphate, sulfate, sodium, ammonia, CO_2, ammonia were always unconstraint. The reversibility of the reaction could also be defined by the aid of capacity constraint, like, the reversible fluxes were for the internal fluxes were set to zero.

4.2.2.10 Flux variability analysis (FVA)

The focus of the current study is the characterization of alternate solutions within the *R. opacus PD630* genome-scale network consisting of 1879 fluxes and 1404 metabolites. In this LP-based approach the accentuation is with respect to deciding the minimum and maximum estimations of all the flux values that will fulfil the imperatives and take into account the equivalent ideal objective value. It should be noted that this methodology doesn't recognize all conceivable interchange ideal solutions, yet rather the scope of transition changeability that is conceivable inside some random solutions (Mahadevan and Schilling 2003). The methodology starts with estimating the objective function's wild-type value by solving the LP problem according to the equation. From this solution the scope of variability that can occur in each flux in the system because of substitute optimal solutions can be determined through a progression of LP problems in which the estimation of the original objective is set and every response in the system is maximized and then minimized to decide the plausible scope of flux values for every reaction (Hay and Schwender 2011). To illustrate the basic concepts of the approach described in, we analyse the flux variability for aerobic growth on glucose, and lactose in the *R. opacus*

PD630 model. Comparable examination has been utilized to define flux bounds in order to further constrain the flux space for limited reaction sets to be identified.

4.2.2.10.1 Sampling

Artificial hit and run (ACHR) algorithm was used for sample selection of 10 x 10^6 unique flux distributions from the allowable solution space with 50 step size (Schellenberger and Palsson 2009). . CHRR has better runtime and convergence when compared to ACHR, however due to restricted availability of MATLAB in a high-speed computing system ACHR was chosen for this study. The input for the ACHR function with 50 step size were 50000 points per file and 500000 points.Warm-up points of size four times the model's reactions count for each condition were generated as required by the ACHR sampler.The experimental data sets were used to constraint the model as per FVA results prior to sampling.

It has as of late been understood that randomized sampling of solutions, got through FVA, is a successful method in the solution space to describe its substance and dispense with candidates for non-unique solutions. Inspecting of the candidate solutions can be utilized to determine the most instructive estimations to be taken (Savinell and Palsson 1992, Savinell and Palsson 1992). Sampling by the Monte-Carlo based Artificial Centred Hit and Run (ACHR) was used to simulate and distribute arbitrary experimental noise across the network. The measurement noise ratio to calculated noise was quantified statistically, thus rejecting experiments with poor design for progressively instructive ones (Schellenberger and Palsson 2009).

4.2.2.11 Principal component analysis

GeM with a greater number of reactions than metabolites, there exists an infinite number of solutions under a specific set of constraints owing to its underdetermined nature. Thus,

sampling warns the distribution of a reaction fluxes to be probabilistic and limits basic investigation of the information. Having many reactions (= variables) in GeM makes it difficult to compare different conditions. To alleviate the problem, we employ PCA to reduce the dimensionality of the solution space in a K-dimensional data set. PCA is a commonly used statistical method banking on the concepts of covariance, standard deviation and singular value decomposition (SVD). Maximum variance from original K-dimensional space is used to define the K new dimensions through PCA (Jolliffe 2002, Ringnér 2008).

Projected linear combination of the original variables, coefficients of these linear combinations and the transformed data points in a new space are called principal components (PCs), loadings and scores respectively scores (Sarıyar, Perk et al. 2006, Worley, Halouska et al. 2013). Initially the Monte-Carlo (MC) based ACHR generated samples with each row and a column representing a sampled flux distribution and a reaction respectively were combined into one data matrix followed by its assessing through PCA. Any systematic errors were avoided by shuffling the rows of the reconstructed data matrix.

With PCA being the maximum variance projection method, the variables need to undergo pre-processing to avoid any impact on their analysis as different variables tend to have different numerical ranges. During the pre-processing method the mean centred data that is obtained through subtracting the mean from the corresponding flux values of each column is made unit variance by dividing them by their respective standard deviation (Bro and Smilde 2003). The data was then used to calculate the covariance matrix before undergoing singular value decomposition (SVD): $X = USV^T$, where X is covariance matrix, U having the left singular vectors (in columns), the diagonal of S being the eigenvalues, and V^T being the right singular vectors (in rows). The scores were calculated

using this V^T or loadings by multiplying original mean centred data matrix with right singular vectors (Worley and Powers 2013).

PCA is selected in light of the fact that it effectively figures out how a perturbation proliferates through a linear network and to draw comparisons between different fluxes under glucose and lactose.

4.3 Results and Discussion

The metabolic reconstruction of *R. opacus* PD630 was carried out in a series of successive refinements followed by model validation and analysis. The refinement steps include modelling process, manual curation, curation of biomass reaction, model characteristics and explained in detail below.

4.3.1 Modelling process

In the current study, the metabolic network model of *R. opacus* PD630 was reconstructed, validated and a computational prediction study was performed. To derive a metabolic network, initially the genome of *R. opacu*s PD630 was integrated with literature to create the initial draft of the model. The process was initiated by considering enzymes and predicting their function probability encoded to the organism's genome (Galagan, Calvo et al. 2003). All the predicted enzymes were then collected and integrated into experimental pathways which were acquired from the pathway/genome database (PGDB) generated through ModelSEED (Latendresse, Krummenacker et al. 2012, Aite, Chevallier et al. 2018). Subsystem based category distribution of the *R. opacu*s PD630 metabolism as generated from RAST and ModelSEED pipeline is shown in the Figure 4.2.With the aid of this automated approach we could successfully curate the model using the literature specific to Rhodococcus and to recognize gene ontology terms to proteins, enzymes determined through experiments, characterize and differentiate isozymes from

their respective enzyme complexes and catalogue growth of the organism (Holder, Ulrich et al. 2011).

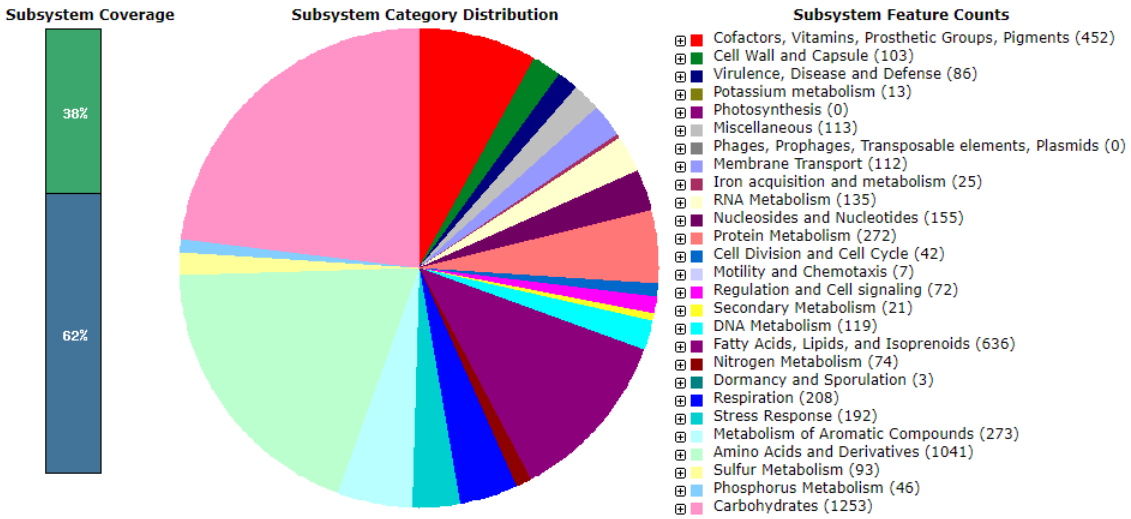

Figure 4.2: Pie chart illustrating the subsystem statistics of the FDR for R. opacus PD630 with carbohydrate concerning reactions sharing the highest portion followed by amino acids and derivatives and fatty acids, lipids and isoprenoids reactions.

4.3.2 Model analysis

Despite having used the tools for automatically generating FDR, a thorough manual curation of the model is needed to optimize the reconstruction's prediction (Henry, DeJongh et al. 2010). While the automatically created first draft GeM has been often seen to produce biomass, this ability may not be inferred from the accuracy of a GeM in predicting experimental results (Heinken, Sahoo et al. 2013). However, in the current study the model was even incapable to produce the biomass without constraints and thus a thorough manual intervention, re-visiting the networks structure was necessary. The manual curation can be done in several ways (i) examining the thermodynamic probability of a reaction, (ii) examining mass and charge balance (iii) filling gaps in the model (Karlsen, Schulz et al. 2018).

4.3.3 Manual curation

The validity of the model was ensured prior to simulating for growth. Maintaining accuracy of all the reactions in the metabolic model is essential as it influences the energy flux which in turn determines the overall flux capabilities of the network. To this end, the direction of all the reactions was determined and these stipulations were further added on to applications like KEGG and MetaCyc to further correct and expand the network built in RAST and ModelSEED. On a general note, Thiele's and Palsson's (Thiele and Palsson 2010) approach was followed to determine the directionality of a specific reaction. The direction of a particular reaction is typically based on two main rules: (a) NADP+ should not be reduced by NADH without presence of an energy coupling. (b) Formation of ATP should occur only at substrate level phosphorylation in association with ATP synthase. These two simple rules ensure biochemically feasible reactions through such oxidative pathways. One of the dominant roles of ferredoxin-mediated $NADP^+$ reduction in *R. opacus* PD630's metabolism provides it with an utmost importance to carefully create routes to NADPH.

The reconstructed model of *R. opacus* PD630 was compared to the nearest related bacterial strain, *Rhodococcus jostii* RHA1, for which a GeM was published (Tajparast and Frigon 2015) in view of coming across any missing reactions. 360 genes found exclusively in the published *Rhodococcus jostii* RHA1 model was not included in the SEED *R. opacus* PD630 model because either the functions assigned to these genes in the SEED were inconsistent with the reactions mapped to them in the published models or the functions were not specific enough to allow for mapping to explicit reactions. Care was taken to avoid incorporating reactions for which the genome lacks the corresponding genes. Each of the missing genes was searched against *R. opacus* PD630 genome sequence using blastn for any match. For each sequence hit, the matched gene/enzyme

was confirmed for its presence in *R. opacus* PD630 from the literature and KEGG. 270 out of 360 genes were found to share the similar sequence from blastn results and 170 genes/reactions of those matched genes were added in *R. opacus* PD630 after having them verified from literature. In addition to the reactions required in the synthesis of biomass constituents, 104 reactions without any gene associations were incorporated to the GeM based on literature

There are not many details available on metabolic reactions of *R.opacus*; to this regard, a minimal set of biomass metabolites, waste products and set of nutrients were obtained from experimental data and literature (Kurosawa, Boccazzi et al. 2010, Holder, Ulrich et al. 2011, Hernández, Comba et al. 2015, Herrero and Alvarez 2016, Thanapimmetha, Suwaleerat et al. 2016, Suwaleerat, Thanapimmetha et al. 2018). The experimental sets result in a set of 15 simple nutrients (glucose, lactose, nitrate, inorganic phosphate, oxygen, vitamin and sulphate) with no extracellular fermentative waste products and secretion. The literature yielded ten core biomass metabolites: phosphorylated nucleotides, amino acids, phospholipids, isoprenoids, NAD(P), sugar nucleotides, crosslinked-peptidoglycan, carbohydrates, corynomycolic acid and small molecules. These metabolites defined the inputs and outputs of the *R. opacus* PD630 metabolic model.

4.3.4 Curation of growth rate data

In order to validate our growth rate data, we found several experimental conditions in literature that included information on glucose levels for our organism reporting doubling times and glucose uptake rates (Kurosawa, Boccazzi et al. 2010, Suwaleerat, Thanapimmetha et al. 2018). We developed the initial metabolic model iteratively with a manually curated training set on limited and augmented media with experimentally

observed viability phenotypes. Only when consistent with published experimental evidence, these changes were reviewed and accepted manually.

4.3.5 Model validation

The curated model was validated by comparing the specific growth rate (μ) with that of the literature and experiment. Model with biomass as the objective function yielded a specific growth rate of 0.089 h^{-1}, which is comparable with 0.114 h^{-1} obtained by (Kurosawa, Boccazzi et al. 2010) under the similar constraints. In addition, the model was also validated by comparing specific growth rate of the growing *R. opacus* PD630 under similar conditions as our another work in (Mandal, Prabhu et al. 2019). However, the model was simulated with lactose instead of the synthetic wastewater as carbon source. The simulated μ = 0.179 h^{-1} is in the acceptable range as the experimental value of 0.248 h^{-1} considering the anomaly in the carbon source. FVA has played a vital role in model's behaviour in accordance with the literature.

4.3.5.1 qPCR

In comparison, inferring the gene knockout targets *in-silico* is much easier than deducing gene overexpression targets. Determining targets for gene overexpression requires precisely controlling fluxes to specific values, which in itself is difficult to achieve experimentally due to the complexities involved. However, there has been a report for identifying overexpression targets coupled with additional manipulations using optForce through FBA and FVA (Ranganathan, Suthers et al. 2010). Nevertheless, identifying solely overexpression targets independently through FBA has not been possible yet. We tried to hypothesize and establish a link between the change in flux through a reaction and an expression level of corresponding gene catalysing it when overexpressed. The molecular species such DNA, RNA and protein are linked tightly by the central dogma

of biology. Despite this principle there is no concrete link defining the relationship between mRNA and protein concentration due to the presence of various factors governing the underlying relationship, such as complicated post-translational modifications, varied *in-vivo* half-lives for different proteins, experimental error (McManus, Cheng et al. 2015, Liu, Beyer et al. 2016). Only under the assumption that increased expression translates to increased translation (mRNA to protein) and that increased enzyme concentration is the original bottleneck for that reaction, an increase in the gene expression level compared to its control would result in an increased flux and thus the end product/metabolite concentration of that particular reaction.

The expression level of *PAP2* gene in both *R. opacus* PD630 and BM985 from (Mandal, Prabhu et al. 2019) was compared to further validate the curated model. The model was simulated with TAG values from the literature for both *R. opacus* PD630 and BM985 and compared the corresponding flux change through the *PAP2* catalysed reaction, R02239. Though the real-time PCR (see Figure 4.3) revealed the expression value to be around 2.2 times more in case of the overexpressed *PAP2* gene in BM985, the flux change through R02239 was slight on a lower side. Only under the oversimplifying assumption that increased expression leads to increased translation and further assuming that enzyme concentration was the rate limiting factor, overexpression of a single gene lead to increased flux through the respective reaction. The P-value for the RT-PCR is 0.0015, which is highly significant. The simulated flux through reaction increased 1.46 times in case of the overexpressed *PAP2* gene. This gives a room for further curation and better constraints for the model to mimic the in-vivo conditions.

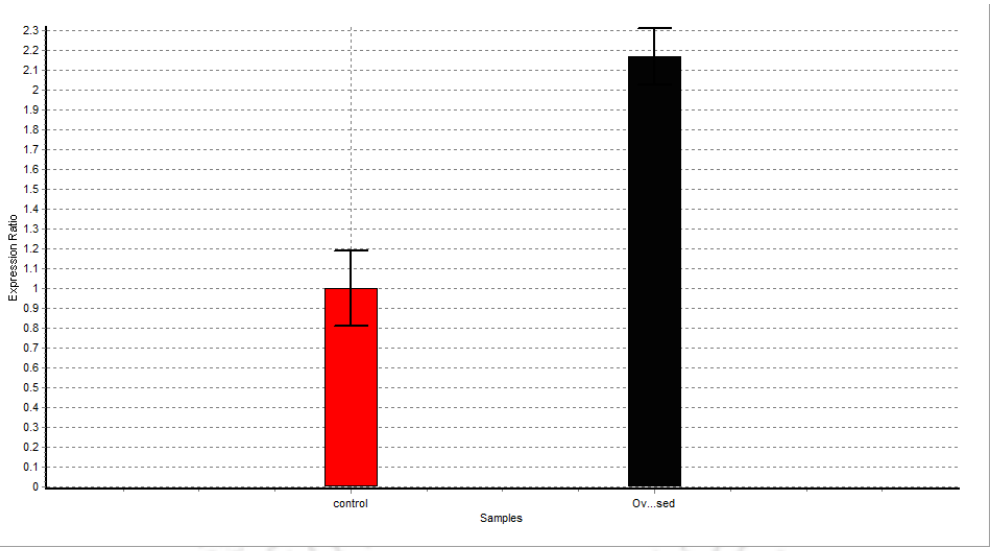

Figure 4.3: Comparison of the expression level of PAP2 gene from Rhodococcus opacus PD630 (control-red bar) with that of Rhodococcus opacus BM985 (black bar). The run was performed in triplicates.

4.3.6 Model characteristics

The final version of the *R opacus* PD630 GSM consists of 1472 genes mapped to 1879 different reactions. An aggregate of 52 reactions with obscure gene affiliations was added to the GeM. This rundown excludes biomass drains, auto-catalytic and membrane transporter responses that were included the premise that they were required. An interaction map and a communication guide of the metabolites reveal global functionality and highlights respectively of the GeM (see Figure 4.4). For the biosynthesis of significant biomass components, only a very small set of reactions are fundamental. There are around 270 critical reactions, despite the fact that this number differs relying upon the input constraints imposed. The input constraints dictate the availability of input nutrients, and therefore the biosynthetic pathways that must be activated for growth. Nodes from the critical reaction array will, in general, be grouped at the centre point of the interaction map, indicating a higher degree of connectivity to these metabolites.

Although the GeM includes a large number of reactions, only 1050 reactions are viewed as active (i.e., have non-zero flux). These reactions, less the critical ones, are parts of

pathways that are redundant for growth. Primer evaluation of the redundant reactions uncovered that they are usually part of parallel or cyclic pathways. Right off the bat, an enormous number of these reactions include transhydrogenation, whereby at least two of them can be clustered into a pathway that delivers a net move of redox equivalent starting with one cofactor then onto the next (e.g., NAD^+, $NADP^+$, FAD, ferredoxin).

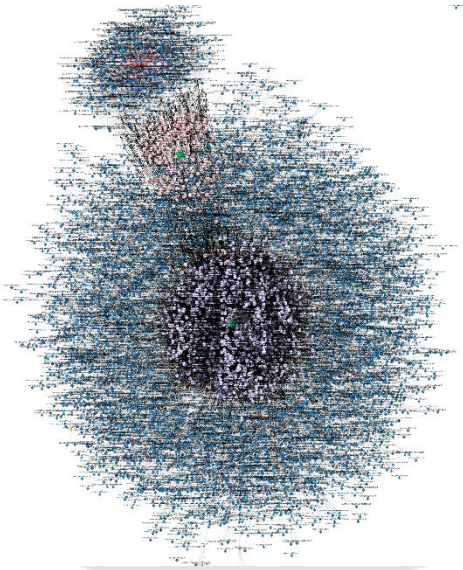

Figure 4.4: Cytoscape generated Rhodococcus opacus PD630 GEM network describing its complexity. Cytosol lies at the centre surrounded with M[c0] metabolites. M[e0] are at the periphery of the network near extracellular

4.3.7 Sampling of the solution space

FBA is an application used to predict and analyse metabolic reaction pathways from genome scale metabolic network models. It represents a metabolic network by acquiring the stoichiometries of reactions from stoichiometric matrix, S. The set of reactions contains lower-bound (lb) and upper-bound (ub) constraints which define all the probable fluxes in the network at a steady state. Formulating a linear programming problem (LP) allows to identify whether a constructed model supports growth under a specific nutrient condition or not.

There are often redundancies in biological systems that contribute fundamentally to their robustness. In this, the redundancies in the underdetermined metabolic system that lead to alternate optimal flux dispersions and subsequently robustness with regards to optimal growth and flux variability are investigated. An approach is proposed to describe and analyse flux variability in alternative, optimal solutions. Using the *R. opacus* PD630 genome-scale metabolic model, this approach to characterizing the alternative optima is demonstrated. The alternative optimal growth solutions for *R. opacus* PD630 on glucose, and lactose had been recognized utilizing the proposed methodology. The outcomes here exhibited that the quantity of pathways with optimal biomass yield and henceforth the quantity of fluxes that differ is subject to the substrate that was being taken up according to the claim reported by (Mahadevan and Schilling 2003).

Monte Carlo based sampling techniques allow to sample the allowable flux space (determined by network properties and bounds imposed on model) in order to generate flux probabilities to get a clearer idea as to what the underlying network behaviour is. The predominant algorithm used to generate such flux probability data has been the Markov Chain Monte Carlo (MCMC) sampler. Dissimilar to the elimination algorithm usable for littler systems, the ACHR algorithm creates a substantial arrangement point at each emphasis. In accordance with probabilistic rules, an initial valid point repeatedly moves within the allowable solution space. The sample becomes the trail of valid points generated. A main downside of this algorithm is that there is no guarantee that in a finite time these points will occupy the whole space evenly. Such action is referred to as "slow mixing".

The artificial centring hit-and-run (ACHR) is an improvement to the MCMC algorithm that is efficient in solving the issues arising from using MCMC algorithm (Schellenberger and Palsson 2009). ACHR is implemented primarily with uniformly selected directions

on the unit sphere. It generates points that approach the optimal probability distribution asymptotically (Kaufman and Smith 1998).

4.3.8 Principal components analysis

PCA allows to reduce dimensionality of complex datasets in order to make it more interpretable. The method aims to identify and order the dimension which capture the most variability within the data set. Analysis of flux probability samples for example underlie thousands of dimensions. Each reaction in the metabolic network represents one dimension in the multi-dimensional space. When comparing the flux samples of different conditions (e.g. network exposed to different carbon sources) PCA may help to identify the main metabolic pathway differences. The new variables called the principal components are a linear combination of the original variables where the coefficients used to project the original data into the new space are called loadings. Generally speaking, the larger the absolute coefficient, i.e. loading value of a variable (reaction), the more it contributes to the overall variance. In other words, when analysing a data matrix containing two sets of flux samples, reactions with vastly different fluxes between the analysed conditions will contribute a lot to the overall variance, i.e. will have a large absolute loading value.

In order to validate the herein presented metabolic reconstruction of *R. opacus* PD630, the model was used to simulated growth on two different carbon sources, glucose and lactose.

Flux sampling of the feasible solution spaces and subsequent PCA resulted in a separation of the solution spaces on the first principal component (PC1) (see Figure 4.5). As mentioned before, PCA aims to reduce dimensionality in an effort to highlight the most variability within a given dataset. With that in mind it can be said that the most variability

captured by the reconstructed network is due to the different carbon sources, i.e. different pathway utilization. Figure 4.5 shows two coloured sample clouds (each dot represents a solution vector/sample), red and blue. The colours indicate the model under glucose and lactose constraints respectively. The result clearly shows the distinct model behaviour with scores being recognizably stretched under different carbon sources proving the fact that model and thus, organism behaves and chooses different pathways to attain the objective function under different constraints.

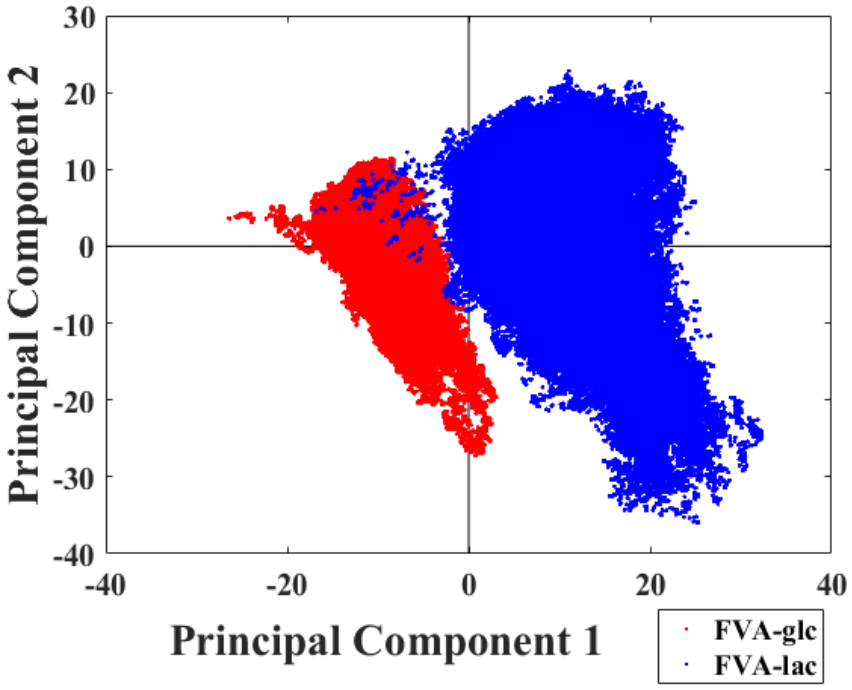

Figure 4.5: PCA score plot for the model under glucose (FVA-glu) and lactose (FVA-lac)

The first principal component explains 8.6 % of the total variance in the data. Figure 4.6 shows the percentage contribution by PC. As PC1 separates the two distinct conditions and also represents the direction with maximal amount of variance among samples, it can be used to deduce the reaction networks which are differently used between the two conditions. According to the analysis of the normalized absolute eigenvectors (loadings), the first 100 reactions that are half the size of the largest normalized absolute loadings

(1.00) are considered significant. This is well illustrated in the Figure 0.1 showing the cut-off for the loadings.

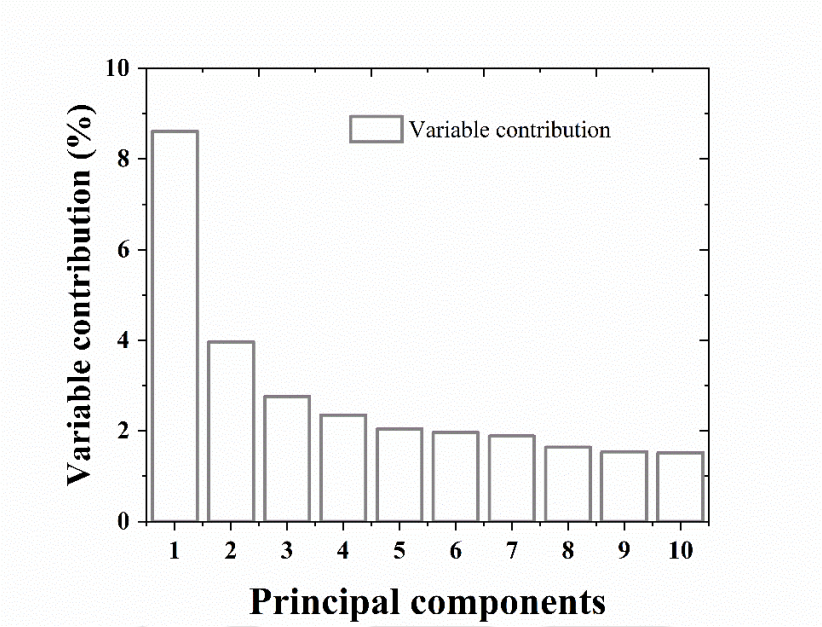

Figure 4.6: Variable contributions in percentage by the first ten components

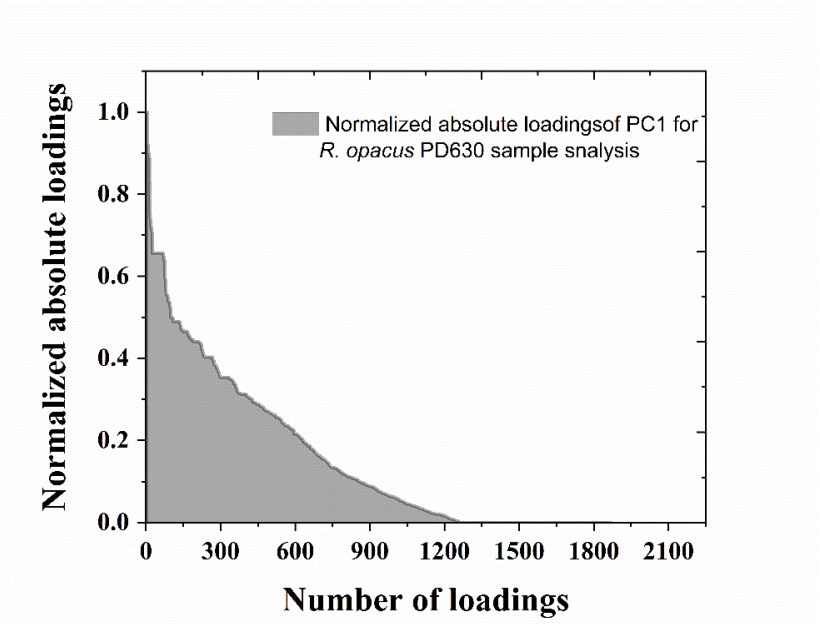

Figure 0.1: Normalized absolute loadings of PC 1

Figure 0.2 gives a small overview of the most influential reactions on the separation of the two conditions. Comparing the reactions, falling above the threshold normalized

absolute loadings (0.5), with the ones in Figure 0.2 there are 23 reactions (Table 0.1) influencing the PCA-score plot outcome for glucose and lactose.

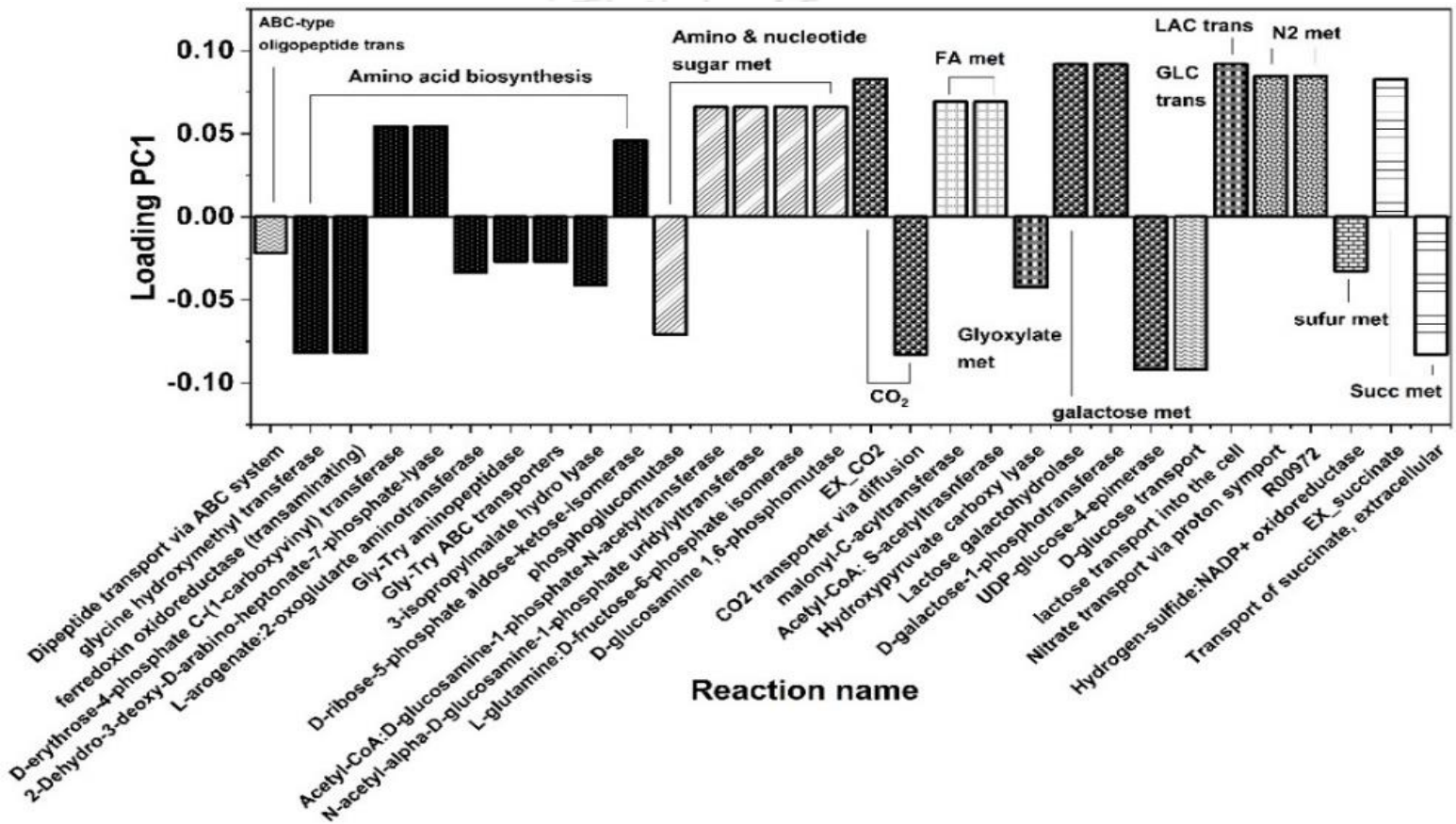

Figure 0.2: Small list of most influential loadings for principal component 1 responsible for the separation of the two tested conditions.

Table 0.1: List of reactions responsible for bringing out the metabolic behavioural change under glucose and lactose.

S.No.	Reactions	Abs flux glucose*	Abs flux lactose*	Median glucose	Median lactose
1	Lactose galactohydrolase	0	0.629946	0	9.935523
2	rxn05598 (lactose transport into the cell)	0	0.629946	0	9.935523
3	ATP:D-galactose-1-phosphotransferase	0	0.629946	0	9.935523
4	UDP-glucose-4-epimerase	0.002542	0.630947	0.000598	-9.93457
5	rxn05573 (D-glucose transport in via proton symport)	0.519803	0	9.935570	0
6	EX_CO2[e0]	19.86715	32.52339	30.93376	40.38788
7	CO2 transporter via diffusion	19.86715	32.52339	-30.9338	-40.3879
8	Nitrate transport via proton symport	3.59446	5.237561	3.399785	5.482229
9	R00792 (ferrocytochrome-b:nitrate oxidoreductase)	3.59446	5.237561	3.399785	5.482229
10	R09490 (oxidative phosphorylation)	3.593707	5.227654	-3.39733	-5.47797
11	EX_succinate[e0]	6.548708	11.33627	0.333326	7.110977
12	Transport of succinate, extracellular	6.548708	11.33627	-0.33333	-7.11098
13	5,10-methylenetetrahydrofolate: glycine hydroxymethyl transferase	13.41404	17.16206	-8.44059	-12.6363
14	L-glutamate: ferredoxin oxidoreductase (transaminating)	41.31904	65.02246	-30.6082	-45.6471
15	Acyl-[acyl-carrier-protein]: malonyl-(acyl-carrier-protein] C-acyltransferase	0.415379	0.957017	0.216591	0.480522
16	Acetyl-CoA: [acyl-carrier-protein] S-acetyltrasnferase	0.415379	0.957017	0.216591	0.480522
17	R00959 (phosphoglucomutase)	32.33182	31.79335	4.565615	-5.41659
18	Acetyl-CoA:D-glucosamine-1-phosphate-N-acetyltransferase	0.098528	0.180773	0.044399	0.061872
19	UTP: N-acetyl-alpha-D-glucosamine-1-phosphate uridylyltransferase	0.098528	0.180773	0.044399	0.061872
20	R00768 (L-glutamine:D-fructose-6-phosphate isomerase)	0.098528	0.180773	0.044399	0.061872
21	R02060 (D-glucosamine 1,6-phosphomutase)	0.098528	0.180773	0.044399	0.061872
22	Phosphoenoylpyruvate:D-erythrose-4-phosphate C-(1-carboxyvinyl) transferase	1.476628	2.080353	0.726003	1.11079
23	2-Dehydro-3-deoxy-D-arabino-heptonate-7-phosphate-lyase (cyclizing)	1.476628	2.080353	0.726003	1.11079

- **Unit of flux = mM/gDCWh**

The first five reactions, Lactose galactohydrolase, rxn05598 (lactose transport into the cell), ATP:D-galactose-1-phosphotransferase, UDP-glucose-4-epimerase, and rxn05573 (D-glucose transport in via proton symport) contribute equally to the variance of the flux samples with Lactose galactohydrolase, lactose transport into the cell, and ATP:D-galactose-1-phosphotransferase having the positive covariance and the remaining two with negative covariance. The effect thereafter starts decreasing gradually as evidenced from the amount of spread across each reaction has in the Figure 0.3 along the direction of PC1 with rxn05535 (ATP phosphohydrolase) having the least effect.

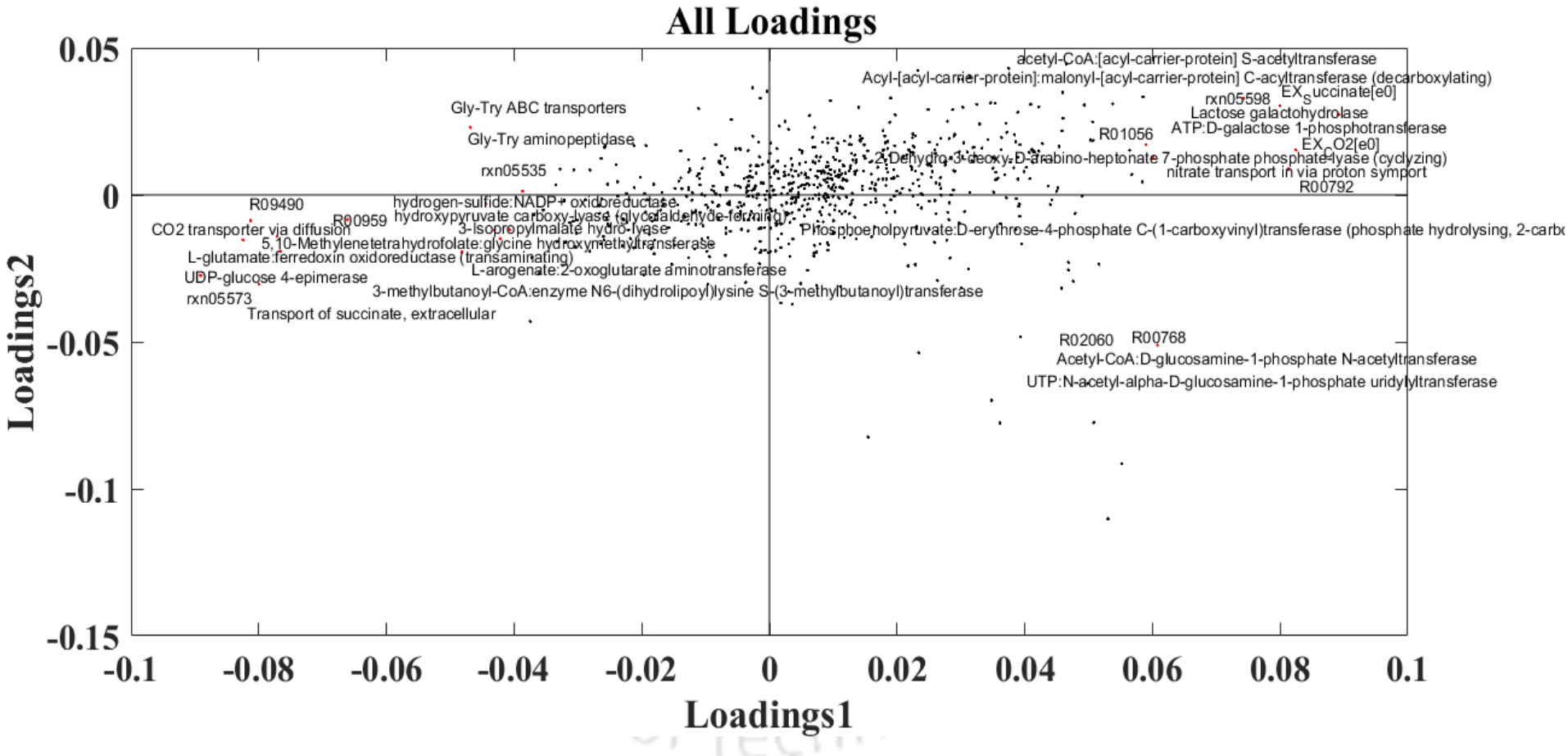

Figure 0.3:Loadings scatter plot. Loading values of PC1 and PC2 are plotted against each other. Most influential loadings on the far ends of PC1 (positive and negative) are labelled with the reaction they correspond to.

Clearly one can infer from the PC1 loading plot (see

Figure 0.2) the existence of many reactions, which can be grouped together, responsible for bringing out the effect on the model under the mentioned constraints. In comparison, the assembly of the number and type of reactions can be visualized easily from the Figure 0.3 for the PCA-loading scatter plot. In addition, the variables (reactions) which are projected onto the PC1 (Loadings1-axis) in the direction having the maximum variance followed by PC2 (loadings 2-axis) with the second most maximum variance, can as well be estimated from this plot. Most samples are near the origin, whereas the samples having the highest variance for the displayed variables are labelled and extended out in their respective directions along the PC1. Evidently one can conclude having the maximum positive covariance in case of Lactose galactohydrolase, lactose transport into the cell, and ATP:D-galactose-1-phosphotransferase with the corresponding negative covariance for UDP-glucose-4-epimerase, and rxn05573 (D-glucose transport in via proton symport). These are most contributed to the first PC indicator and thus to the total variability of the basic set. The highest and least contribution to the description of the PC1 indicator was due to lactose galactohydrolase and rxn05535 respectively as per their distribution along the PC1 direction. These observations are also supported by their respective absolute normalized loadings value falling above or below the 0.5 threshold mark. Observations that are similar will fall close to each other, i.e. they cluster together, and could be catalogued together. Clustering the reactions with highest contribution can be evidenced from the fact that the reactions, lactose galactohydrolase, lactose transport into the cell, $EX\_CO_2[e0]$ and ATP:D-galactose-1-phosphotransferase, are a part of galactose metabolism pathway involved in lactose degradation using oxygen. Oxygen is utilized more for the condition grown under lactose for its degradation through glucose and galactose. This makes sense for the reaction $EX\_CO_2[e0]$ to be included in the same cluster as more oxygen utilization means more carbon dioxide secretion. Similarly, other

catalogued reactions R09490 (oxidative phosphorylation), EX_succinate[e0] and Transport of succinate, extracellular too have an influence on the cells grown under different conditions. These reactions are linked to the requirement of additional oxygen for lactose metabolism, where EX_succinate[e0] and Transport of succinate, extracellular influence the uptake rate of succinate for the aerobic Kreb's cycle. The succinate inside the cells is converted to fumarate through R09490 (oxidative phosphorylation), where oxygen is involved in the transfer of electron from reduced menaquinone to an oxidized one. The reactions combined or independently have a significant impact on the organism depending on the carbon source under which the organism is grown. The PCA here directed us in estimating the patterns in the data and to guide the date by highlighting their distinctions and connections.

4.4 Conclusion

In order to study the complete metabolic characteristics of an organism at system level, reconstruction and simulation of GeM have been playing a strategic role in the field of strain improvement. Feeding additional biochemical information through literature and experiments renders the GeM more powerful and functional. Developing a universal GeM of a particular organism is highly impossible owing to the fact that neither the organism behaves consistently under a specific condition nor any information is ever enough for incorporating into GeM. Herein, we performed the reconstruction of a GeM of *R. opacus* PD630, *i*ROPD630, for the production of a neutral lipid TAG with glucose as the initial point. The model's behaviour is in good agreement with data sets from both literature and experiments. The model distinctively illustrates the metabolic behaviour of the bacterium under two different carbon sources, glucose and lactose, highlighting the principal reactions responsible for the flux change. This validated functional model could be employed further to computationally estimate the candidate target points to enhance the

production of TAG and thus biofuel. This *i*ROPD360 can be used as a template for further curation and polishing to yield a more compelling model as per the requirement.

4.5 Reference

1. Aite, M., M. Chevallier, C. Frioux, C. Trottier, J. Got, M. P. Cortés, S. N. Mendoza, G. Carrier, O. Dameron and N. J. P. c. b. Guillaudeux (2018). "Traceability, reproducibility and wiki-exploration for "à-la-carte" reconstructions of genome-scale metabolic models." **14**(5): e1006146.
2. Aziz, R. K., D. Bartels, A. A. Best, M. DeJongh, T. Disz, R. A. Edwards, K. Formsma, S. Gerdes, E. M. Glass and M. J. B. g. Kubal (2008). "The RAST Server: rapid annotations using subsystems technology." **9**(1): 75.
3. Baart, G. J., M. Langenhof, B. van de Waterbeemd, H.-J. Hamstra, B. Zomer, L. A. van der Pol, E. Beuvery, J. Tramper and D. E. J. M. Martens (2010). "Expression of phosphofructokinase in Neisseria meningitidis." **156**(Pt 2): 530.
4. Baart, G. J. and D. E. Martens (2012). Genome-scale metabolic models: reconstruction and analysis. Neisseria meningitidis, Springer**:** 107-126.
5. Bro, R. and A. K. J. J. o. C. Smilde (2003). "Centering and scaling in component analysis." **17**(1): 16-33.
6. DeJongh, M., K. Formsma, P. Boillot, J. Gould, M. Rycenga and A. J. B. b. Best (2007). "Toward the automated generation of genome-scale metabolic networks in the SEED." **8**(1): 139.
7. DeLorenzo, D. M. and T. S. J. S. r. Moon (2018). "Selection of stable reference genes for RT-qPCR in Rhodococcus opacus PD630." **8**(1): 6019.
8. Domach, M., S. Leung, R. Cahn, G. Cocks, M. J. B. Shuler and bioengineering (2000). "Computer model for glucose-limited growth of a single cell of Escherichia coli B/r-A." **67**(6): 827-840.
9. Du, J., Z. Yuan, Z. Ma, J. Song, X. Xie and Y. J. M. b. Chen (2014). "KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model." **10**(9): 2441-2447.
10. Edwards, J. S. (1999). Functional genomics and the computational analysis of bacterial metabolism, University of California, San Diego.
11. Edwards, J. S., R. U. Ibarra and B. O. J. N. b. Palsson (2001). "*In-silico* predictions of Escherichia coli metabolic capabilities are consistent with experimental data." **19**(2): 125.
12. Fondi, M. and P. Liò (2015). Genome-scale metabolic network reconstruction. Bacterial Pangenomics, Springer**:** 233-256.
13. Galagan, J. E., S. E. Calvo, K. A. Borkovich, E. U. Selker, N. D. Read, D. Jaffe, W. FitzHugh, L.-J. Ma, S. Smirnov and S. J. N. Purcell (2003). "The genome sequence of the filamentous fungus Neurospora crassa." **422**(6934): 859.

14. Hay, J. and J. Schwender (2011). "Metabolic network reconstruction and flux variability analysis of storage synthesis in developing oilseed rape (Brassica napus L.) embryos." **67**(3): 526-541.

15. Heinken, A., S. Sahoo, R. M. Fleming and I. J. G. m. Thiele (2013). "Systems-level characterization of a host-microbe metabolic symbiosis in the mammalian gut." **4**(1): 28-40.

16. Heirendt, L., S. Arreckx, T. Pfau, S. N. Mendoza, A. Richelle, A. Heinken, H. S. Haraldsdottir, J. Wachowiak, S. M. Keating and V. J. N. p. Vlasov (2019). "Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v. 3.0." **14**(3): 639.

17. Henry, C. S., M. DeJongh, A. A. Best, P. M. Frybarger, B. Linsay and R. L. J. N. b. Stevens (2010). "High-throughput generation, optimization and analysis of genome-scale metabolic models." **28**(9): 977.

18. Hernández, M., S. Comba, A. Arabolaza, H. Gramajo and H. Alvarez (2015). "Overexpression of a phosphatidic acid phosphatase type 2 leads to an increase in triacylglycerol production in oleaginous Rhodococcus strains." Applied Microbiology and Biotechnology **99**(5): 2191-2207.

19. Herrero, O. M. and H. M. Alvarez (2016). "Whey as a renewable source for lipid production by Rhodococcus strains: Physiology and genomics of lactose and galactose utilization." European Journal of Lipid Science and Technology **118**(2): 262-272.

20. Holder, J. W., J. C. Ulrich, A. C. DeBono, P. A. Godfrey, C. A. Desjardins, J. Zucker, Q. Zeng, A. L. Leach, I. Ghiviriga and C. J. P. g. Dancel (2011). "Comparative and functional genomics of Rhodococcus opacus PD630 for biofuels development." **7**(9): e1002219.

21. Jolliffe, I. T. J. P. c. a. (2002). "Principal components in regression analysis." 167-198.

22. Kanehisa, M., Y. Sato, M. Furumichi, K. Morishima and M. J. N. a. r. Tanabe (2018). "New approach for understanding genome variations in KEGG." **47**(D1): D590-D595.

23. Karlsen, E., C. Schulz and E. J. B. b. Almaas (2018). "Automated generation of genome-scale metabolic draft reconstructions based on KEGG." **19**(1): 467.

24. Karp, P. D., M. Riley, M. Saier, I. T. Paulsen, S. M. Paley and A. J. N. a. r. Pellegrini-Toole (2000). "The ecocyc and metacyc databases." **28**(1): 56-59.

25. Karp, P. D., D. Weaver and M. J. B. s. b. Latendresse (2018). "How accurate is automated gap filling of metabolic models?" **12**(1): 73.

26. Kaufman, D. E. and R. L. J. O. R. Smith (1998). "Direction choice for accelerated convergence in hit-and-run sampling." **46**(1): 84-95.

27. Kumar, V. S., M. S. Dasika and C. D. J. B. b. Maranas (2007). "Optimization based automated curation of metabolic reconstructions." **8**(1): 212.

28. Kurosawa, K., P. Boccazzi, N. M. de Almeida and A. J. J. J. o. B. Sinskey (2010). "High-cell-density batch fermentation of Rhodococcus opacus PD630 using a high glucose concentration for triacylglycerol production." **147**(3-4): 212-218.

29. Latendresse, M., M. Krummenacker, M. Trupp and P. D. J. B. Karp (2012). "Construction and completion of flux balance models from pathway databases." **28**(3): 388-396.

30. Liu, Y., A. Beyer and R. J. C. Aebersold (2016). "On the dependency of cellular protein levels on mRNA abundance." **165**(3): 535-550.

31. Mahadevan, R. and C. J. M. e. Schilling (2003). "The effects of alternate optimal solutions in constraint-based genome-scale metabolic models." **5**(4): 264-276.

32. Mandal, B., A. Prabhu, K. Pakshirajan and V. Veeranki Dasu (2019). "Construction and parameters modulation of a novel variant Rhodococcus opacus BM985 to achieve enhanced triacylglycerol-a biodiesel precursor, using synthetic dairy wastewater." Process Biochemistry **84**: 9-21.

33. McManus, J., Z. Cheng and C. J. M. B. Vogel (2015). "Next-generation analysis of gene expression regulation–comparing the roles of synthesis and degradation." **11**(10): 2680-2689.

34. Orth, J. D., I. Thiele and B. Ø. J. N. b. Palsson (2010). "What is flux balance analysis?" **28**(3): 245.

35. Quek, L.-E. and L. K. Nielsen (2008). On the reconstruction of the Mus musculus genome-scale metabolic network model. Genome Informatics 2008: Genome Informatics Series Vol. 21, World Scientific**:** 89-100.

36. Ranganathan, S., P. F. Suthers and C. D. J. P. c. b. Maranas (2010). "OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions." **6**(4): e1000744.

37. Ringnér, M. J. N. b. (2008). "What is principal component analysis?" **26**(3): 303.

38. Sarıyar, B., S. Perk, U. Akman and A. J. J. o. t. b. Hortaçsu (2006). "Monte Carlo sampling and principal component analysis of flux distributions yield topological and modular information on metabolic networks." **242**(2): 389-400.

39. Savinell, J. M. and B. O. J. J. o. t. b. Palsson (1992). "Optimal selection of metabolic fluxes for in vivo measurement. I. Development of mathematical methods." **155**(2): 201-214.

40. Savinell, J. M. and B. O. J. J. o. t. b. Palsson (1992). "Optimal selection of metabolic fluxes for in vivo measurement. II. Application to Escherichia coli and hybridoma cell metabolism." **155**(2): 215-242.

41. Schellenberger, J. and B. Ø. J. J. o. b. c. Palsson (2009). "Use of randomized sampling for analysis of metabolic networks." **284**(9): 5457-5461.

42. Schellenberger, J., R. Que, R. M. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis and S. J. N. p. Rahmanian (2011). "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2. 0." **6**(9): 1290.

43. Suthers, P. F., M. S. Dasika, V. S. Kumar, G. Denisov, J. I. Glass and C. D. J. P. C. B. Maranas (2009). "A genome-scale metabolic reconstruction of Mycoplasma genitalium, iPS189." **5**(2): e1000285.

44. Suwaleerat, T., A. Thanapimmetha, M. Srisaiyoot, Y. Chisti, P. J. J. o. C. T. Srinophakun and Biotechnology (2018). "Enhanced production of carotenoids and lipids by Rhodococcus opacus PD630." **93**(8): 2160-2169.

45. Tajparast, M. and D. J. B. s. b. Frigon (2015). "Genome-scale metabolic model of Rhodococcus jostii RHA1 (i MT1174) to study the accumulation of storage compounds during nitrogen-limited condition." **9**(1): 43.

46. Thanapimmetha, A., T. Suwaleerat, M. Saisriyoot, Y. Chisti and P. Srinophakun (2016). "Production of carotenoids and lipids by Rhodococcus opacus PD630 in batch and fed-batch culture." Bioprocess and Biosystems Engineering: 1-11.

47. Thiele, I. and B. Ø. J. N. p. Palsson (2010). "A protocol for generating a high-quality genome-scale metabolic reconstruction." **5**(1): 93.

48. Varma, A. and B. O. J. A. E. M. Palsson (1994). "Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110." **60**(10): 3724-3731.

49. Varma, A. and B. O. J. J. o. T. B. Palsson (1993). "Metabolic capabilities of Escherichia coli II. Optimal growth patterns." **165**(4): 503-522.

50. Worley, B., S. Halouska and R. J. A. b. Powers (2013). "Utilities for quantifying separation in PCA/PLS-DA scores plots." **433**(2): 102-104.

51. Worley, B. and R. J. C. M. Powers (2013). "Multivariate analysis in metabolomics." **1**(1): 92-107.

Chapter 5

Sustainable *de-novo* biosynthesis of biodiesel by oleaginous *Rhodococcus opacus* PD630 and its metabolic flux analysis

5.1 Outline

Biodiesel, a promising substitute for fossil fuel consists mainly of FAMEs or FAEEs produced through the transesterification reaction of methanol (or ethanol) and TAGs from plants and algae in the presence of sulfuric acid or sodium hydroxide as catalyst. To avoid usage of hazardous chemicals like methanol and sulfuric acid/sodium hydroxide here we report the construction of a genetically engineered natural high yielding TAGs and fatty-acids strain *R. opacus* PD630 for *de-novo* biosynthesis of FAEEs in the microenvironment of the cells by the heterologous expression of pyruvate decarboxylase gene from the ethanol producing pathway of *Z. mobilis*. The cultivations yielded a FAEEs of 0.88 g/L and 1.22 g/L in lactose and glucose respectively. This novel approach might pave the way for industrial production of biodiesel equivalents from renewable resources as this organism is known to catabolize and thrive on a variety of industrial wastes.

5.2 Introduction

Population explosion and their ever-increasing energy demand have taken a toll on the rapidly depleting non-renewable energy sources like fossil fuels. The direct emission and exposure of fossil fuels contribute to increase in greenhouse gases and consequently global warming. In this regard, a quick alternative to this situation is needed which not only shall meet the fuel demands but as well reduce the emission of harmful greenhouse gases into the environment (Mandal, Prabhu et al. 2019). Biodiesel, a promising substitute for the fossil fuel which is produced from oils form plant e.g. oil seed crops can potentially replace fossil fuels and serve the need of this hour. The biodiesel obtained from plant sources are almost carbon neutral as the carbon dioxide released during burning of the fuel is absorbed by the crop itself while farming (Mathews 2008).

Biodiesel consists of FAMEs and FAEEs produced by transesterification of plant oils and animal fats which is a high energy consuming process. Moreover, during the transesterification process methanol or ethanol is used to generate FAMEs or FAEEs respectively. The main disadvantage of this homogenous catalysts in transesterification is the undesirable production of both soap and glycerol, which as well increases the production costs through energy requirement for their removal. The transesterification using ethanol at 75^{o}C industrially is carried out as compared to 60^{o}C in case of methanol, which further increases the energy demand for this process. In order to overcome the above problems, triglycerides can also be transesterified with lipases, at a temperature not exceeding 50^{o}C industrially which is analogous to the reaction occurring inside a cell with constant supply of enzymes catalysing at a temperature of 50^{o}C. Enzymes lower this 'energy hill' of activation energy by providing an alternative reaction pathway of lower activation energy so that the reactions can occur more easily, e.g. at lower temperatures. As a result, some metabolic processes occur rapidly at body temperature, which is relatively cool in terms of chemical reactions (Schuchardt, 1998). Also, there aren't any steps to remove the unwanted soap and glycerol. This strategy makes this *in-situ* transesterification more energy efficient.

A superior yield of biodiesel is produced when methanol is used in the transesterification process instead of ethanol, however due to methanol's toxic properties it is not considered over ethanol. Alongside, methanol is primarily obtained from natural gases and thus the fuel produced from the aid of methanol cannot be called as fully renewable (Kalscheuer, Stölting et al. 2006).

Subsequently, in order to have a better fuel resource, microdiesel has potentially achieved superiority over crop aided biodiesels. Microdiesel is produced into the cellular environment of microbes like bacteria (Duan, Zhu et al. 2011). Although the yield can go

significantly higher in case of microalgae, if genetically modified, in compared to bacteria, the latter contributes to much easier and economical feasible process (Duan, Zhu et al. 2011, Mandal, Prabhu et al. 2019).

An engineered *E. coli* strain was constructed using pdc and adhB of *Z. mobilis* for ethanol production, wax ester and TAGs forming gene from *Acinetobacter baylyi* for intracytoplasmid lipid storage (Kalscheuer, Stölting et al. 2006). In another relevant study, (Duan, Zhu et al. 2011) optimized the effect of culture conditions such as effect of initial culture conditions, temperature, initial feeding condition and the induction time point on microdiesel production in engineered *E. coli.* A maximum production of 922 mg/L of FAEEs was evidenced to be achieved in the optimized conditions. Although, they attained low yield, according to their work microbe can also produce microdiesel in absence of oily feed rather using sugars derived from lignocellulosic biomass (Duan, Zhu et al. 2011). In (Duan, Zhu et al. 2011) study ethyl oleate was found to be in greater extent with little amount of ethyl palmitate and ethyl palmitoleate. 1.28 g/L FAEE was achieved in his study that accounted for the 26% of the cellular dry biomass. In another study (Elbahloul and Steinbüchel 2010) used an engineered *E. coli* strain to produce microdiesel in a pilot scale and achieved a yield in between 25-26%. Alongside, the group also reported inferior microdiesel yield when the microorganism starts to depend on the lipid produced by them. Although the bacteria could produce microdiesel, the yield remained considerably low. There could be several reasons for *E. coli* to fail as a suitable host for biodiesel synthesis, to name the most relevant one is the integration of multiple foreign genes which induces stress on the regular cellular mechanism of *E. coli* leading to inferior production of biodiesel. This drawback might be overcome by using an organism with native ability to accumulate wax esters and TAGs so that integrating a gene for ethanol

biosynthesis can suffice the issue with lesser yield (Kalscheuer, Stölting et al. 2006, Duan, Zhu et al. 2011).

The economics of Fermentation largely depend upon several factors. To name the most important ones would be (a) investment costs of construction and maintenance (b) fermentation plant and (c) raw materials along with upstream and downstream processing expenses. The downstream process becomes significantly expensive for the transesterification process due to utilization of expensive materials like methanol/ethanol and acid/base. To this end, we report a novel method which would reduce the cost of downstream processing significantly and make biodiesel economically viable. During the growth of the microbe, it is the biomass that primarily effects the formation of FAEE; therefore, in order to attain a decent amount of FAEEs suitable substrates should be used like glucose and lactose as carbon sources (Mathews 2008, Elbahloul and Steinbüchel 2010).

Lipases have evolved to be an industrially significant enzyme especially in the arena of biofuel. One of its very evident usefulness is hydrolysis of vegetable oil to fatty acids and that it can esterify both fatty acids and glycerides. Lipases have a wide variety of substrates, like esters, thioesters, amides to name a few. However, the higher costs involved in both upstream and downstream processing make it economically futile. To this end, a better alternative will be to have a microorganism which may hydrolyse the lipid feed into TAG in its cellular environment. Likewise, lipases present intracellularly in *R. opacus* PD630 can hydrolyse lipid feed into TAG, which a precursor of biodiesel. (Shah, Sharma et al. 2004, Chatterjee, Yadav et al. 2014).

R. opacus PD630, an aerobic and Gram-positive bacterium, capable of degrading hydrogen and accumulation of intracytoplasmic TAGs, a vital step to produce biodiesel. In this study, *R. opacus* PD630 was used to produce microdiesel considering its ability to

produce intracellular fatty acids by consuming simple sugar like glucose under nitrogen limiting condition (Mandal, Prabhu et al. 2019). Moreover, ethanol, a vital component for production of FAEEs is not produced by wild *R. opacus* PD630 due to absence of pdc enzyme. However, pyruvate decarboxylase helps in conversion of 2-hydroxy-ethyl-thpp to acetaldehyde which further gets converted to ethanol as shown in Figure 5.1 (KEGG). Recombinant *R. opacus* PD630 with an ability to produce ethanol was constructed by inserting a codon optimized pdc gene from *Z. mobilis*, a notable bacterium for bioethanol production. The pdc gene is responsible for synthesis of pyruvate decarboxylase enzyme which enables recombinant *R. opacus* PD630 to produce ethanol intracellularly. In the current study, recombinant *R, opacus* PD630 was chosen to degrade lipids into TAG which eventually was converted to FAEE with the aid of intracellular lipase and in presence of ethanol. The microbe has been evidenced to synthesize TAG in a nitrogen limiting condition following a de novo fatty acid biosynthesis and β-oxidation pathway for the synthesis of storage lipids.

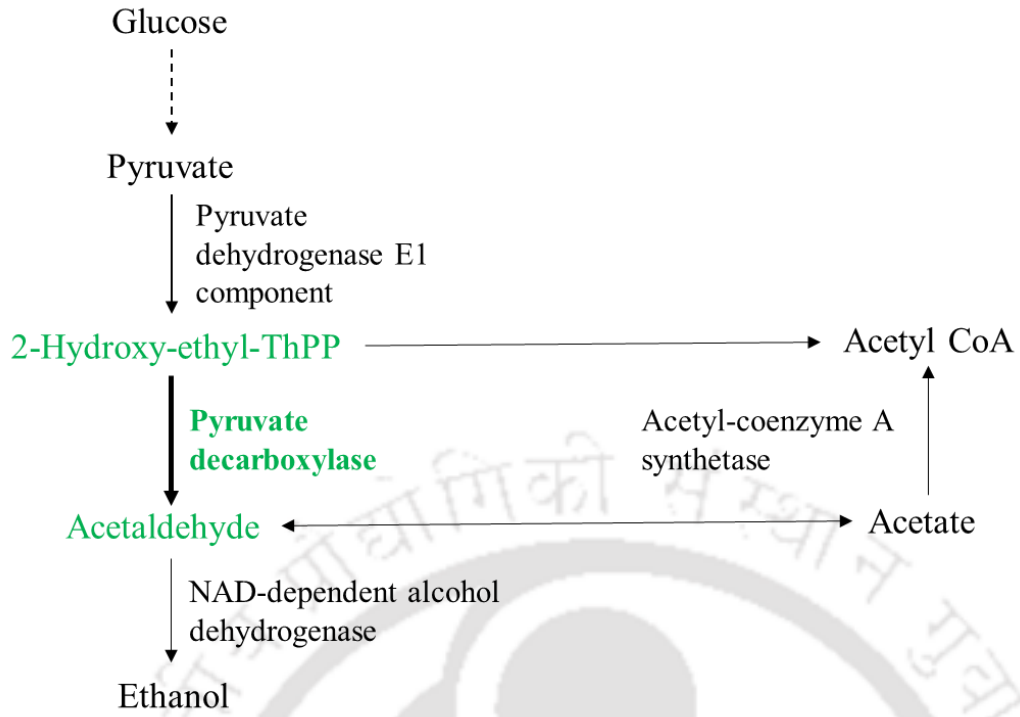

Figure 5.1: Schematic representation of the pathway leading to ethanol through pdc catalysed reaction

5.3 Materials and methods

5.3.1 Strain and Construction of plasmid

R. opacus PD630 was used in this study. *E. coli-Rhodococcus* shuttle vector pNit-QT1 was used to construct plasmid with pdc gene from *Z. mobilis*. This plasmid was transformed into *E. coli* to attain high copy number. Later, the plasmid was isolated from *E. coli* (TOP10) and transformed into *R. opacus* PD630 for expression of the gene (production of ethanol).

Recombinant *E. coli* and *R. opacus* PD630 were grown and maintained on LB agar medium containing 5 g/L of NaCl and yeast extract each along with 10g/L tryptone. Ampicillin and tetracycline were added to the agar medium for specific selection of strains.

5.3.2 Amplification and cloning of pdc gene

GeneJET Genomic Purification kit (Thermo Fisher Scientific, Waltham, MA, USA) was used to isolate Genomic DNA of the *Z. mobilis*. For PCR amplification of pdc gene, 0.025 units/µL DNA polymerase (NEB), 0.5 µM specific primers i.e. P1 and P2, 200 µM dNTPs and 1X polymerase buffer supplemented with Mg^{2+} ion. Agarose gel electrophoresis was run for 20 minutes to further gel purify pdc gene and pNit-QT1 vector using QIAquick gel extraction kit. Cloning of vector and gene was performed with the aid of restriction digestion followed by ligation. NdeI and EcoRI restriction enzymes were used to perform digestion on the vector and insertion of pdc gene followed by ligation. The competent *E. coli* TOP10 cells were prepared by $CaCl_2$ method. The recombinant vector was then transformed into freshly prepared competent *E. coli* TOP10 cells by provision of heat shock. To screen and confirm the transformed cells; *E. coli* TOP10 cells were grown on LB agar plates with ampicillin (100µg/mL) for 12 hrs. QIA prep miniprep kit was used to extract plasmid DNA from transformed colonies and sequencing was done to confirm the clone (Schäfer, Tauch et al. 1994, Mandal, Prabhu et al. 2019).

5.3.3 Transformation and over-expression of pdc gene in *R. opacus* PD630

Once the pdc gene was confirmed with the help of sequencing results, further electroporation was done as according to (Mandal, Prabhu et al. 2019) using Electro Cell Manipulator 620 (BTX Inc., San Diego, CA, USA) to introduce the gene into *R. opacus* PD630. Subsequently, 0.1 mL of *R. opacus* PD630 inoculum (overnight TSB pre-culture) was cultured in 10 mL of tryptic soy broth along with 0.5 %(w/v) glycine (in a 50mL screw-capped vial) for 24hrs at 28 ^{o}C. Ice cold HS buffer was prepared using 7mM HEPES and 252 mM sucrose of pH 7.0, then concentrated 10-fold in ice-cold HS buffer. This buffer was used to wash harvested cells. Before performing electroporation, 400 µL of competent cells were preincubated at 40 ^{o}C for 10 min and then immediately mixed

with plasmid DNA with final concentration 0.1-1 µg/mL. Gaps of 2 mm, 6.5 KV/cm, 725 Ω, and 50 µF were maintained while performing electroporation using Electro-cuvettes (BTX Inc.). The cells were then regenerated for 24 hrs at 28 ^{o}C after diluting them with 4 mL of TSB. Afterwards, they were spread on TSB agar plates containing TSB agar with tetracyclin (15 µg/mL). The TAG yielding clone was selected from the colonies of *R. opacus* PD630 overexpressing pdc gene. Inoculum of loopful culture was added into LB media with 15 µg/mL tetracycline and a final volume of 100 mL at 30 ^{o}C and 200 rev min-1 for 72 hrs. The vector pNit-QT1 has a constitutive promoter which stays induced at all the time. The cells were collected by harvesting the culture at 13000 rpm for 10 mins. pdc is native to *Z. mobilis*, a control (without the plasmid) and recombinant *R. opacus* PD630 (pNit-QT1 with pdc gene) was expressed for 3 days and evaluated.

5.3.4 Shake flask cultivation

Experiments were carried out in 250 mL Erlenmeyer flask containing 50 mL of production medium, Glucose and mineral salt medium (MSM) containing 9 g/L $Na_2HPO_4 \cdot 12H_2O$, 1.5 g/L KH_2PO_4, 0.2 g/L $MgSO_4 \cdot 7H_2O$, 0.1 g/L NH_4SO_4, 0.05 g/L $NaHCO_3$ 0.02 g/L $CaCl_2 \cdot 2H_2O$, and 1.2 mg/L $C_6H_8FeNO_7$. The flasks were stirred at 200 rpm and maintained at 30 ^{o}C till the end of the batch.

5.3.5 Thin-layer chromatography

Thin-layer chromatography (TLC) analysis was performed as described (Kalscheuer & Steinbu¨chel, 2006) using the solvent system methanol and chloroform. The biomass was treated with combinations of methanol and chloroform mixture and chloroform. The solvent system hexane/diethyl ether/acetic acid (90:7.5:1, by vol.) was used as the mobile phase. Lipids were visualized by spraying with 40 % (v/v) sulfuric acid and charring.

Ethyl oleate was purchased from Sigma-Aldrich Chemie and used as the reference substance for FAEEs.

5.3.6 Bioreactor cultivation

Scale up experiment in bioreactor was done considering all the optimized values obtained from shake-flask validation experiments. Stirred Tank Bioreactor (Applikon Dependable Instruments, ez Control, The Netherlands) was utilized in this experiment to grow *Z. mobilis* and *R. opacus* PD630. 3 liter bioreactor was used for experiment containing 1 liter of production medium as mentioned above. Aeration was done at a constant air flow rate of 1 vvm through a stainless sintered sparger in the bioreactor. The rotational speed of the impeller, temperature and pH were set at 300 rev min^{-1}, 30 ± 1 °C and 7 ± 0.1 respectively. After 96 hrs of cultivation cycle, the culture was harvested and then DCW, sugar consumption, TAG content and amount of FAEE was analysed.

In both flask and bioreactor cultivation 1 % (v/v) inoculum was used from overnight seed culture grown on LB medium. Sample was collected in 3 hrs interval for analysis.

5.3.7 *In-silico* analysis of the microdiesel producing *R opacus* PD630

The metabolism of any organism can be easily understood and analyse using the validated GeM **iROPD630**.

5.3.7.1 Metabolic flux analysis of iROPD630MD

In this study the GeM *i*ROPD630 reconstructed in the previous chapter (chapter 4) was used for metabolic flux analysis. The native *i*ROPD630 does not include the reactions for ethanol production. As the pdc gene was heterologously expressed in this study, the reaction catalysed by pyruvate decarboxylase was added externally to the *i*ROPD630 GeM to form *i*ROPD630MD. This addition was also used to test the viability of the model

using FBA and FVA. This constraint-based approach was performed to analyse the cell's metabolism as described in the sections 4.2.2.9 and 4.2.2.10.

5.3.7.2 Carbon constraints-based flux balance analysis (ccFBA)

ccFBA aims to solve the issues of FBA and FVA solutions with extremely high flux values well beyond the physiological range by obliging the allowable flux through intracellular reactions based on the measure of carbon absorbed by the cell under the physiological conditions studied. ccFBA demands that all carbon absorbed, and their specific chemical formulae be identified, and that the FBA model be properly constrained. Therefore, it is necessary to provide the availability of experimental values in the form of absorption and secretion rates for major carbon sources. In this study ccFBA algorithm was run according to (Lularevic, Racher et al. 2019) under lactose and glucose to compare the behaviour of the metabolic model.

5.4 Results and Discussion

5.4.1 Thin-layer chromatography

Thin-layer chromatography (TLC) was performed on biomass obtained from shake-flask batch cultures to confirm and analyse the presence of FAEE in the different solvent systems. The results obtained can be seen in the Figure 5.2.

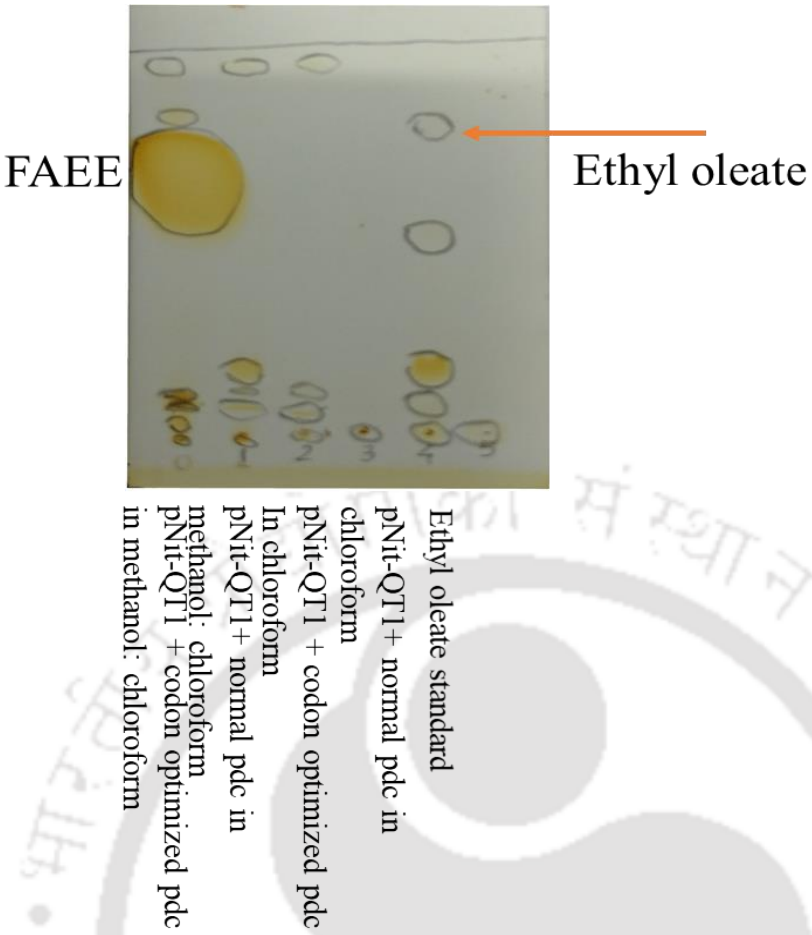

Figure 5.2:Thin layer chromatography of the samples extracted with different solvents from transformed cells with codon optimized and unoptimized pdc gene

In TLC normal non codon-optimized and codon optimized pdc gene products were tested after being treated with combinations of chloroform and methanol. All the results were compared with Ethyl oleate standard. The Figure 5.2 clearly shows that FAEE was found to be forming in methanol: chloroform solvent system where codon optimized pdc gene was used. This may be due to the codon optimization has increased the protein expression i.e pyruvate decarboxylase formation. And this enzyme produced ethanol which in further reaction with TAGs, helped in formation of FAEE. The chloroform failed to extract any traceable amount of FAEE to be detected in TLC both for codon-optimized and non-codon-optimized pdc gene. The non-codon-optimized construct shows no results both in the presence of methanol: chloroform and chloroform.

5.4.2 FAEEs from batch fermentation in bioreactor

Formation of ethanol & FAEEs at different interval from 0 to 100 hrs is plotted in graph as shown in Figure 5.3. To check biomass OD is taken at 660 nm and carbon source utilization is also checked. In the batch fermentation where glucose used as sole carbon source, 4.62 g/L ethanol is observed to be maximum at 20 hrs and maximum 1.22 g/L FAEE is found after 60 hrs at around 65 hrs. At 60 hours Glucose got completely depleted and after 60 hrs growth also slowed down. So, the experiment can be run for more time with greater amount of substrate because the substrate depletion is the sole reason for the lack of further growth and FAEE production. Also, the substrate amount must not exceed minimum substrate inhibitory concentration.

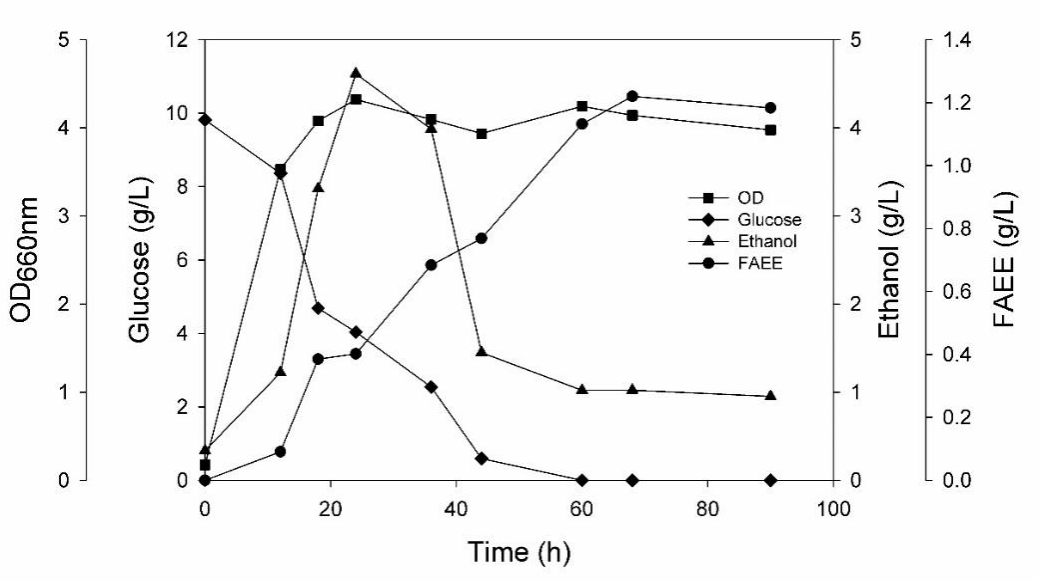

Figure 5.3:Growth profile of the organism showing glucose consumption rate with ethanol and FAEE yield

When the experiment was done using lactose as sole carbon source (see Figure 5.4), both ethanol and FAEE yield found to be less than that of glucose. Maximum 2.92 g/L ethanol was found to be at 20 hrs and 0.88 g/L FAEE (maximum) at 90 hrs. Even if growth is found to be increasing at 90 hrs, the lactose is not utilized much after 65 hrs. From this observation we can conclude that the lactose batch should continue up to 120 hrs as

lactose is not completely depleted after 90 hrs growth was increasing at that stage as mentioned earlier.

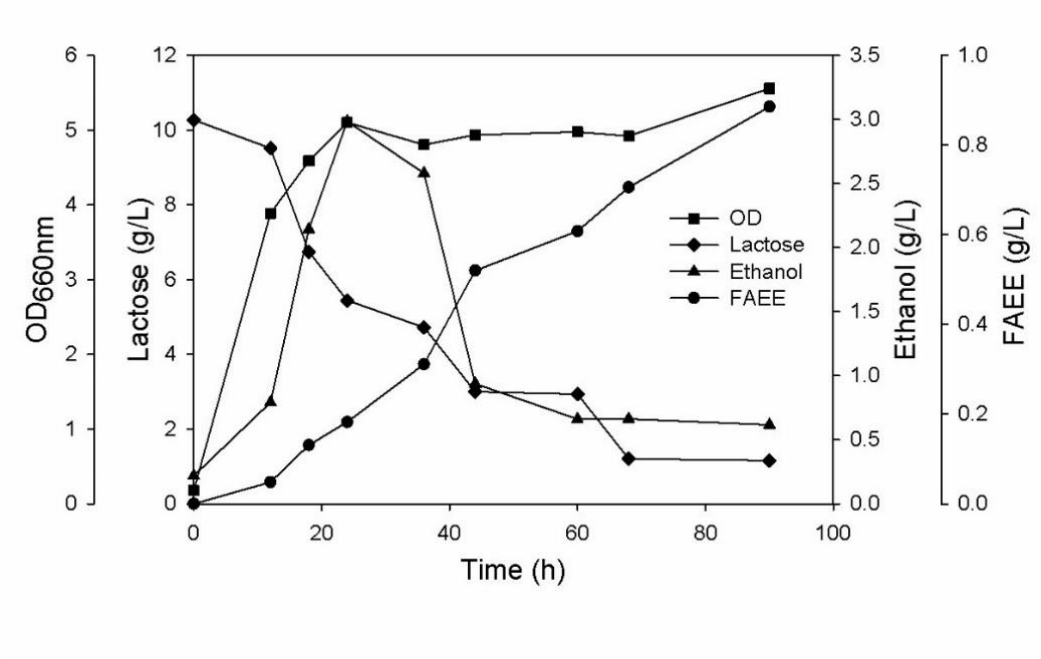

Figure 5.4:Growth profile of the organism showing lactose consumption rate with ethanol and FAEE yield

In both cases concentration of ethanol found to be decreasing after 20 hrs must be due to formation of FAEEs where ethanol combines with fatty acids to form FAEEs. The increasing concentration seen after 20 hrs support this discussion as well. It is reported that high concentration of ethanol inhibits microbial growth, but in this study, ethanol is seen to be get utilized after reaching a maximum value of around 5 g/L and 3g/L respectively. So, the inhibitory effect of ethanol can be completely avoided even if the experiment is performed in fed batch process or continuous process.

5.4.3 Metabolic flux analysis of the *i*ROPD360MD

Inclusion of the reaction R00755 catalysed by pyruvate decarboxylase fills the gap between pyruvate and acetaldehyde. The already present enzyme ethanol: NAD^+ oxidoreductase can now act upon acetaldehyde to convert it into ethanol through a reaction R00754. The flow of carbon and thus flux from lactose through this reaction is

analysed and compared between ccFVA and unconstrained FVA. For the organism to use further in the dairy wastewater treatment the analysis is restricted to lactose to estimate the model's behaviour.

The experimentally measured uptake and secretion rates obtained in this study are used to simulate and constraints the model for its analysis. All the units are normalized to millimole g cell dry weight^{-1} hour^{-1} (mmol gDCW^{-1}h^{-1}). The objective function is set for ethanol secretion reaction. This will further help in comparing it with the experimental ethanol yield and validating the model to check for its viability. The simulation yielded varying results as shown in the Table 5.1. The table compares the potential of ccFVA with that of the normal FBA and FVA.

Table 5.1: FBA and comparison between iROPD360 and iROPD360MD with ethanol as the objective function

	***i*ROPD630**	***i*ROPD630MD**			
	FBA/FVA	FBA	absolute (FVA)	ccFVA absolute	This study
Ethanol (mmol gDCW^{-1}h^{-1})	0	36.96	11.70	5.85	5.61

The table shows the range of permissible flux through the reaction R00754 (ethanol:NAD$^+$ oxidoreductase) leading to ethanol. The *i*ROPD630 does not have any ethanol producing reaction and so the FBA yields 0 mmol gDCW^{-1}h^{-1}. The FVA for *i*ROPD630MD resulted in a higher value (36.96 and 11.70 mmol gDCW^{-1}h^{-1}) of ethanol than the actual experimental yield (5.61 mmol gDCW^{-1}h^{-1}). This large flux values for FBA and FVA through R00754 is due to the general rule of common FBA of assigning undetermined intracellular fluxes with large bounds constraining unidirectional and bidirectional reactions between [0, 1000] mmol gDCW^{-1}h^{-1} and [-1000, 1000] mmol gDCW^{-1}h^{-1} respectively (Bordel, Agren et al. 2010, Chaudhary, Tøndel et al. 2016, Opdam, Richelle et al. 2017). Though the significant reduction in the solution space of

around 85.34% can be observed in the case of FVA the value at large remains farfetched from the experimental value. This wide flux ranges restrict our scope for any valid inferences from drawing for the metabolic phenotype.

The issue of wide flux range is circumvent using the ccFVA strategy. One of the factors that influences the total wide flux ranges is the total number of carbons consumed by the cells. This factor is controlled within a physiologically realistic range by the ccFVA (Maranas and Zomorrodi 2016, Lularevic, Racher et al. 2019). Employing this approach resulted in a 96.59% solution space reduction yielding a comparable ethanol flux with that of the experimental value (see table). Having fluxes with no unrealistic range, in a complex GeM network, also helps in mitigating the effect of futile cycles/internal loops that produces metabolic resources out of thin air (Price, Reed et al. 2004).

ccFVA, with Biomass as an objective function, was performed on both *i*ROPD630 and *i*ROPD630MD to compare and analyse the flux distribution under lactose. Simulated specific growth rate remains relatively same for both the models despite the requirement for carbon and energy in ethanol production in *i*ROPD630MD. Constant biomass yield could be fulfilled by the reaction R00006 (pyruvate: pyruvate acetaldehyde transferase). This reaction transfers more flux (17.62 mmol $gDCW^{-1}h^{-1}$) than that of *i*ROPD630 (16.64 mmol $gDCW^{-1}h^{-1}$) from pyruvate to acetolactate. This acetolactate then gets converted to acetyl-CoA, which is the key intermediate in the central carbon metabolism. Acetyl-CoA eventually gets oxidized and converted to biomass constituents through tricarboxylic acid (TCA) cycle, glyoxylate cycle and fatty acids biosynthesis pathways (Kremer, van Teeseling et al. 2019). Oxidation of acetyl-CoA also generates NADPH and ATP that aid in the formation of free fatty acids (Yan, Cheng et al. 2013, Kremer, van Teeseling et al. 2019) and thus FAEEs.

Incorporating pdc gene helps in binding the gap between pyruvate and ethanol through acetaldehyde (see Figure 5.1). More of acetaldehyde needed to be produced by cells through other means or pathways in order to compensate for the constant biomass despite producing ethanol. This is achieved by the reaction R06973 (Malonate semialdehyde decarboxylase) producing more of acetaldehyde from 3-oxopropanoate. A flux of 4.78 mmol $gDCW^{-1}h^{-1}$ more flows through this reaction in *i*ROPD630MD than in *i*ROPD630 keeping an ample pool of acetaldehyde available for ethanol production. Both R00006 and R06973 resulted in excess carbon dioxide production which is evidenced through a 0.938 mmol $gDCW^{-1}h^{-1}$ more flux through a carbon dioxide exchange reaction ($EX\_CO_2[e0]$) in *i*ROPD630MD as compared to *i*ROPD630.

5.5 Conclusion

The transesterification reaction products namely FAMEs and FAEEs produced from either methanol and ethanol respectively are the constituents of biodiesel. The transesterification reaction is usually energy driven making the entire process laborious and economically less desirable. To this end, in the current study, we have developed a one-step transesterification strategy carried out in the cellular microenvironment. The overexpression of codon optimized pyruvate decarboxylase gene from *Z. mobilis* was observed to be successfully expressing itself and superiorly functioning in the reaction pathway as compared to non-codon optimized gene. Also, the extraction in it was evidenced that, with the episomal integration of pyruvate decarboxylase gene the FAEE yield reached a 28.84% and 31.28% of DCW in the presence of lactose and glucose as medium of nutrition, respectively. Therefore, this one step transesterification strategy shall prove itself beneficial in the fermentation industry arena with larger culture volumes and superior biodiesel yields.

FBA through carbon constraints turns out to be better analysis technique as compared to normal FBA. FBA on *i*ROPD630MD gives insight into the cell's physiology during ethanol production. The comparison between GeM could offer the strategy to target for nodes or reactions leading towards more ethanol and fatty acids production.

5.6 References

1. Bordel, S., R. Agren and J. Nielsen (2010). "Sampling the Solution Space in Genome-Scale Metabolic Networks Reveals Transcriptional Regulation in Key Enzymes." PLOS Computational Biology **6**(7): e1000859.
2. Chatterjee, S., D. Yadav, L. Barbora, P. Mahanta and P. J. J. o. C. Goswami (2014). "Silk-Cocoon matrix immobilized lipase catalyzed transesterification of sunflower oil for production of biodiesel." **2014**.
3. Chaudhary, N., K. Tøndel, R. Bhatnagar, V. A. P. Martins dos Santos and J. Puchałka (2016). "Characterizing the optimal flux space of genome-scale metabolic reconstructions through modified latin-hypercube sampling." Molecular BioSystems **12**(3): 994-1005.
4. Duan, Y., Z. Zhu, K. Cai, X. Tan and X. J. P. O. Lu (2011). "De novo biosynthesis of biodiesel by Escherichia coli in optimized fed-batch cultivation." **6**(5).
5. Elbahloul, Y. and A. J. A. E. M. Steinbüchel (2010). "Pilot-scale production of fatty acid ethyl esters by an engineered Escherichia coli strain harboring the p (Microdiesel) plasmid." **76**(13): 4560-4565.
6. Kalscheuer, R., T. Stölting and A. J. M. Steinbüchel (2006). "Microdiesel: Escherichia coli engineered for fuel production." **152**(9): 2529-2536.
7. Kremer, K., M. C. F. van Teeseling, L. Schada von Borzyskowski, I. Bernhardsgrütter, R. J. M. van Spanning, A. J. Gates, M. N. P. Remus-Emsermann, M. Thanbichler and T. J. Erb (2019). "Dynamic Metabolic Rewiring Enables Efficient Acetyl Coenzyme A Assimilation in Paracoccus denitrificans." **10**(4): e00805-00819.
8. Lularevic, M., A. J. Racher, C. Jaques and A. Kiparissides (2019). "Improving the accuracy of flux balance analysis through the implementation of carbon availability constraints for intracellular reactions." **116**(9): 2339-2352.
9. Mandal, B., A. Prabhu, K. Pakshirajan and V. Veeranki Dasu (2019). "Construction and parameters modulation of a novel variant Rhodococcus opacus BM985 to achieve enhanced triacylglycerol-a biodiesel precursor, using synthetic dairy wastewater." Process Biochemistry **84**: 9-21.
10. Maranas, C. D. and A. R. Zomorrodi (2016). Optimization methods in metabolic networks, John Wiley & Sons.
11. Mathews, J. A. J. E. p. (2008). "Carbon-negative biofuels." **36**(3): 940-945.

12. Opdam, S., A. Richelle, B. Kellman, S. Li, D. C. Zielinski and N. E. Lewis (2017). "A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models." Cell Systems **4**(3): 318-329.e316.

13. Price, N. D., J. L. Reed and B. Ø. J. N. R. M. Palsson (2004). "Genome-scale models of microbial cells: evaluating the consequences of constraints." **2**(11): 886-897.

14. Schäfer, A., A. Tauch, W. Jäger, J. Kalinowski, G. Thierbach and A. Pühler (1994). "Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum." Gene **145**(1): 69-73.

15. Shah, S., S. Sharma, M. J. E. Gupta and Fuels (2004). "Biodiesel preparation by lipase-catalyzed transesterification of Jatropha oil." **18**(1): 154-159.

16. Yan, J., R. Cheng, X. Lin, S. You, K. Li, H. Rong, Y. J. A. m. Ma and biotechnology (2013). "Overexpression of acetyl-CoA synthetase increased the biomass and fatty acid proportion in microalga Schizochytrium." **97**(5): 1933-1939.

Chapter 6

Conclusion and Future aspects

With significant carbon oxygen demand (COD) removal from synthetic dairy wastewater (SDWW) and fuel grade lipid production BM985, an improved variant of *R. opacus* PD630, justifies our preliminary objective of constructing a novel variant to achieve biodiesel production and DWW treatment simultaneously. BM985 could be further employed in the biorefinery industry to obtain high-value and low-cost product using dairy wastewater.

GEMs with constrains based modelling (CBM) methods, have been successfully applied to understand the lipogenesis and associated metabolic fluxes in *R. opacus* PD630. It has been emphasized that the regulatory hotspots are promising engineering targets for improving the lipid production. The GEMs are also promising tools for guiding metabolic engineering, however, there is a need to adapt computational strain design algorithms (CSOMs) to consider the metabolic switch. Finally, most of the current systems biology studies on the oleaginous microorganisms only take advantage of one of the two systems biology approaches: top-down, data-driven or bottom-up, model-driven. Integrative use of both approaches on *i*ROPD630 will bridge the knowledge gap between the two approaches, and better guide the metabolic engineering in the oleaginous microorganisms.

Microdiesel from *R. opacus* PD630 instead of *E. coli* could be an attractive alternative bioenergy source that could contribute to the energy market as a substitute for diesel in addition to conventional biodiesel. The dependence on feeding a source of lipid to cells synthesizing these FAEEs makes *E. coli* a less attractive choice for the microdiesel production. This was overcome by engineering free fatty acids producing *R. opacus* PD630 heterologously to produce ethanol resulting in *de-novo* FAEEs production. The metabolism of the engineered *R. opacus* PD630 was significantly understood using ccFVA through *i*ROPD630MD instead of a normal FBA. FAEEs yield depends on the

availability of both ethanol and free fatty acids. Free fatty acids get broken down during β-oxidation in cytosol to lower moieties, thereby decreasing its concentration. Metabolic engineering strategy through a systems biology approach could be employed on *i*ROPD630MD to target nodes/pathways blocking β-oxidation cycle or to target reactions to enhance ethanol production.

LIST OF PUBLICATIONS

Bapi Mandal, Ashish Prabhu, Kannan Pakshirajan and Veeranki Venkat Dasu (2019). "Construction and parameters modulation of a novel variant Rhodococcus opacus BM985 to achieve enhanced triacylglycerol-a biodiesel precursor, using synthetic dairy wastewater." Process Biochemistry **84**: 9-21.

Ashish A Prabhu, **Bapi Mandal** and Veeranki Venkata Dasu (2017), "Medium optimization for high yield production of extracellular Human Interferon Gamma from Pichia Pastoris: A Statistical optimization and Neural network based approach", Korean Journal of Chemical Engineering, 34(4), 1109–1121.

Ashish A Prabhu, Anwesha Purkayastha**, Bapi Mandal**, Jadi Praveen Kumar, Biman B Mandal and V. Venkata Dasu. "A Novel reverse micellar purification strategy for histidine tagged human interferon gamma (hIFN-γ) protein: A two stage mathematical optimization approach." Int J Biol Macromol. 2018 Feb;107(Pt B):2512-2524.

Manuscripts under pipeline

Bapi Mandal, Maximillian Lularevic, Alexandros Kiparissides and Veeranki Venkata Dasu, 'Reconstruction and validation of a genome scale metabolic model of *R. opacus* PD630 for better understanding TAGs biosynthesis', Plos Computational Biology (submitted).

Bapi Mandal, Ansuman Sahoo, Kannan Pakshirajan and Veeranki Venkata Dasu,' Sustainable *de novo* biosynthesis of biodiesel by oleaginous *Rhodococcus opacus* PD630 and its metabolic flux analysis'. (under preparation)

Published in National/International Conference Proceedings

Bapi Mandal, Ansuman Sahoo and Veeranki Venkata Dasu, "Sustainable de novo biosynthesis of biodiesel by oleaginous *Rhodococcus opacus* PD630". NHBT 2019, CSIR-NIIST, India.

Bapi Mandal, Kannan Pakshirajan and Veeranki Venkata Dasu, "Construction of a novel *Rhodococcus opacus* BM985 to achieve an enhanced biodiesel grade fuel using synthetic dairy wastewater". RABEB 2018, IIT BHU, India.

Bapi Mandal, Kannan Pakshirajan and V. Veeranki Dasu," Construction of a novel mutant *Rhodococcus opacus* BM985 and its optimum design methodology to achieve an enhanced biodiesel grade fuel using synthetic dairy wastewater'. BPI 2018, IIT Delhi, India

Bapi Mandal, Ashish A Prabhu, Kannan Pakshirajan and V. Venkata Dasu, "Advancing *Rhodococcus opacus* PD630 To Produce Biofuel: A Functional Genomics and Metabolic Engineering Approach". 57th International Conference of the Association of Microbiologists of India.

Ashish A Prabhu, Sushma Chityala, Nitin Kumar, **Bapi Mandal**, Rajat Pandey and V. Venkata Dasu (2015). Bioprocess Development of Therapeutic Proteins and Industrial Enzymes. Research conclave"15, IIT Guwahati, India

Ashish A Prabhu, **Bapi Mandal** and Venkata Dasu V (2015). Development of chemically defined medium for high level expression of extracellular Human Interferon Gamma from Pichia Pastoris GS115. International Conference on Recent advances in Biosciences and Applications 22 | P a g e of Engineering in production of Biopharmaceuticals, 14-16 Dec. KL University, Andhra Pradesh, India.

Ashish A Prabhu, Sushma Chityala, **Bapi Mandal**, (2016). Development of chemically defined medium for high level expression of extracellular Human Interferon Gamma from Pichia Pastoris GS115. Research conclave"16, IIT Guwahati, India

APPENDIX

Reactions and Subsystem of *i*ROPD630

Subsystems/KEGG Map	Reactions (KEGG ID)
Folate biosynthesis	R03067
Glutathione metabolism	R00497
Valine, leucine and isoleucine degradation	R07599
Purine metabolism	R01132
Pyrimidine metabolism	R00512
Unknown	U207
Peptidoglycan biosynthesis	R05032
Glutathione metabolism	R00894
Unknown	7-methyl-trans-oct-2-enoyl-ACP:NAD+ oxidoreductase (A-specific)
Pyruvate metabolism	R01257
Pyrimidine metabolism	R02326
Unknown	NO2t2r
Glycolysis / Gluconeogenesis/ Citrate cycle (TCA cycle)/ Pyruvate metabolism	R03270
	R01324
Unknown	R06514
Fatty acid elongation in mitochondria/ Fatty acid metabolism	R04737
Unknown	7-methyl-octanoyl-ACP:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
	R05346
Citrate cycle (TCA cycle)	R00268
Purine metabolism	R01863
Valine, leucine and isoleucine degradation	R05064
Glycine, serine and threonine metabolism	R02565
Unknown	MEAMP1(met-ala)
Pyrimidine metabolism	R01665
Glycolysis / Gluconeogenesis/ Citrate cycle (TCA cycle)/ Pyruvate metabolism	R00726
Propanoate metabolism	R00353
Purine metabolism/ Alanine and aspartate metabolism	R01083
Urea cycle and metabolism of amino groups	R00239

Benzoate degradation via hydroxylation/ Carbazole degradation	R00816
Urea cycle and metabolism of amino groups	R00259
Pyrimidine metabolism	R01664
Unknown	6-methyl-octanoyl-ACP:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
Glutamate metabolism/ Tyrosine metabolism/ Butanoate metabolism	R00714
	R04150
Unknown	U205
Pentose and glucuronate interconversions	R01639
Inositol metabolism/ Alanine and aspartate metabolism/ beta-Alanine metabolism/ Propanoate metabolism	R00705
	R02145
Histidine metabolism	R03243
Porphyrin and chlorophyll metabolism	R04972
Unknown	XYLabc
Folate biosynthesis	R03504
	R07219
Alanine and aspartate metabolism	R00357
Unknown	13-methyl-tetra-decanoyl-ACP:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
Unknown	THD2
Biosynthesis of steroids/ Terpenoid biosynthesis	R02061
Unknown	R01209
Unknown	15-methyl-trans-hexa-dec-2-enoyl-ACP:NAD+ oxidoreductase (A-specific)
Riboflavin metabolism	R07281
Unknown	11-methyl-dodecanoyl-ACP:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
Urea cycle and metabolism of amino groups	R02282
Arginine and proline metabolism	R02922
Unknown	TRANS-RXNBWI-115637.ce.maizeexp.GLY_GLY
Phenylalanine metabolism/ Styrene degradation	R02540
Unknown	R01100
Nicotinate and nicotinamide metabolism	R04292
Methane metabolism	R06983
	R01700

Unknown	R00220
Propanoate metabolism	R00930
	R00004
Unknown	DIPEPabc7
Unknown	R00922
Pyrimidine metabolism	R00965
Thiamine metabolism	R03223
Urea cycle and metabolism of amino groups/ Glutathione metabolism	R00670
Tyrosine metabolism	R04134
Tyrosine metabolism	R01647
One carbon pool by folate/ Methane metabolism	R01224
Unknown	D-Lysine:2-oxoglutarate aminotransferase
Unknown	DIPEPabc3
Pyrimidine metabolism	R00974
Purine metabolism	R01768
Unknown	stearyl-ACP:[acyl-carrier-protein] transferase
Unknown	13-methyl-tetra-decanoyl-ACP:[acyl-carrier-protein] transferase
Phenylalanine, tyrosine and tryptophan biosynthesis/ Benzoxazinone biosynthesis	R02340
Biosynthesis of unsaturated fatty acids	R03814
Unknown	Gly-Leu aminopeptidase
Glycine, serine and threonine metabolism/ Lysine biosynthesis	R00480
Vitamin B6 metabolism	R00277
Unknown	6-methyl-trans-oct-2-enoyl-ACP:NAD+ oxidoreductase (A-specific)
Purine metabolism	R00191
Fatty acid biosynthesis	R04952
Histidine metabolism/ Nitrogen metabolism	R01168
Fatty acid elongation in mitochondria/ Fatty acid metabolism	R04747
Phenylpropanoid biosynthesis	R07441
Alanine and aspartate metabolism/ Nitrogen metabolism	R00578
Unknown	2-MEBUCOA-FAD-RXN.c
Citrate cycle (TCA cycle)	R03316
Phenylalanine, tyrosine and tryptophan biosynthesis	R00733
Fluorobenzoate degradation	R08113

Valine, leucine and isoleucine biosynthesis/ Pyruvate metabolism	R01213
Phenylalanine, tyrosine and tryptophan biosynthesis	R02412
Nicotinate and nicotinamide metabolism	R03346
Purine metabolism/ Sulfur metabolism	R00509
Fatty acid metabolism	R03777
Valine, leucine and isoleucine degradation	R07602
Lysine biosynthesis	R02735
Purine metabolism	R00336
Unknown	MEAMP1(ala-glu)
Unknown	NIRBD-RXN.c
Unknown	NACtpp
	R04340
Porphyrin and chlorophyll metabolism	R00084
Taurine and hypotaurine metabolism/ Pyruvate metabolism	R00315
Fatty acid biosynthesis	R04957
Fructose and mannose metabolism	R01059
Glycine, serine and threonine metabolism/ Glycerolipid metabolism/ Glyoxylate and dicarboxylate metabolism	R01514
Unknown	rbs_51
2,4-Dichlorobenzoate degradation	R02561
Pantothenate and CoA biosynthesis	R00130
	R03314
Unknown	12-methyl-trans-tetra-dec-2-enoyl-ACP:NAD+ oxidoreductase (A-specific)
Unknown	3.6.3.20-RXN.ce
Unknown	NADH-DEHYDROG-A-RXN.c.metaexp.CPD-9956_UBIQUINONE-8
Glycolysis / Gluconeogenesis/ Carbon fixation in photosynthetic organisms	R01061
1- and 2-Methylnaphthalene degradation	R06913
Unknown	13-methyl-3-oxo-tetra-decanoyl-ACP:NADP+ oxidoreductase
Porphyrin and chlorophyll metabolism	R03165
Unknown	PSD161

Peptidoglycan biosynthesis	R05629
Purine metabolism	R00127
Phenylalanine, tyrosine and tryptophan biosynthesis	R03083
Butanoate metabolism	R03937
Fatty acid biosynthesis	R04724
Porphyrin and chlorophyll metabolism	R05220
Glycine, serine and threonine metabolism/ Cyanoamino acid metabolism/ One carbon pool by folate/ Methane metabolism	R00945
Fructose and mannose metabolism	R02071
Urea cycle and metabolism of amino groups/ Alanine and aspartate metabolism/ Arginine and proline metabolism	R01086
Glycine, serine and threonine metabolism	R01511
Histidine metabolism	R03013
Tyrosine metabolism/ Styrene degradation	R02519
Fatty acid metabolism	R03857
Purine metabolism	R01227
Unknown	R06513
Unknown	FLDR
Biosynthesis of steroids	R05688
Fatty acid metabolism/ Synthesis and degradation of ketone bodies/ Valine, leucine and isoleucine degradation/ Lysine degradation/ Tryptophan metabolism/ Pyruvate metabolism/ Benzoate degradation via CoA ligation/ Propanoate metabolism/ Butanoate metabolism	R00238
Pyrimidine metabolism	R01870
Folate biosynthesis	R04639
Bile acid biosynthesis	R01459
Unknown	NADH8
Geraniol degradation	R08086
Benzoate degradation via hydroxylation/ Benzoate degradation via CoA ligation	R00813
Unknown	NADK4

Valine, leucine and isoleucine degradation	R04203
beta-Alanine metabolism/ Propanoate metabolism	R03158
Phenylalanine, tyrosine and tryptophan biosynthesis	R00986
Unknown	ACOAD20
Glycolysis / Gluconeogenesis/ Citrate cycle (TCA cycle)/ Pyruvate metabolism	R00431
Purine metabolism	R01968
Unknown	R00844
Unknown	3-Hydroxyoctodecanoyl-ACP:NADP+ oxidoreductase
Glycine, serine and threonine metabolism	R00366
Phenylalanine metabolism	R04376
Arginine and proline metabolism	R00552
Riboflavin metabolism	R00425
Valine, leucine and isoleucine degradation	R02662
Glycolysis / Gluconeogenesis	R03321
Glycerophospholipid metabolism	R02817
Unknown	LIPAtex
Phenylalanine metabolism	R01372
	R02232
Glycine, serine and threonine metabolism	R00610
Folate biosynthesis	R05046
Propanoate metabolism	R00931
Glycerophospholipid metabolism	R01030
Purine metabolism	R02422
Methionine metabolism	R02025
Arginine and proline metabolism	R00671
Fatty acid biosynthesis	R04963
Unknown	8-methyl-trans-dec-2-enoyl-ACP:NAD+ oxidoreductase (A-specific)
Fatty acid elongation in mitochondria/ Fatty acid metabolism	R04744
Glyoxylate and dicarboxylate metabolism/ One carbon pool by folate	R01220
Fatty acid elongation in mitochondria/ Fatty acid metabolism	R04170

Glyoxylate and dicarboxylate metabolism	R01745
Toluene and xylene degradation	R05298
Unknown	Gly-Phe aminopeptidase
Unknown	13-methyl-trans-tetra-dec-2-enoyl-ACP:NAD+ oxidoreductase (A-specific)
	R05705
Unknown	R03959
Unknown	isotetradecanoyl-glycerol-3-phosphate O-acyltransferase
Fatty acid biosynthesis	R04726
Purine metabolism	R02103
Glyoxylate and dicarboxylate metabolism	R00479
Biosynthesis of steroids	R05612
Pyrimidine metabolism	R01993
Methionine metabolism	R04405
Unknown	myristoyl-glycerol-3-phosphate O-acyltransferase
Glycine, serine and threonine metabolism	R02529
Unknown	3.6.3.27-RXN.ce
Citrate cycle (TCA cycle)	R02570
Glycine, serine and threonine metabolism/ Lysine biosynthesis	R02291
1,4-Dichlorobenzene degradation	R04258
Citrate cycle (TCA cycle)/ Pyruvate metabolism/ Glyoxylate and dicarboxylate metabolism/ Carbon fixation in photosynthetic organisms/ Reductive carboxylate cycle (CO2 fixation)	R00342
Glycolysis / Gluconeogenesis/ Pyruvate metabolism	R00703
Unknown	5-methyl-3-oxo-hexanoyl-ACP:NADP+ oxidoreductase
Pyrimidine metabolism	R00962
Pyrimidine metabolism	R00517
Urea cycle and metabolism of amino groups	R02283
Cysteine metabolism	R00782
Pyrimidine metabolism	R00963
Propanoate metabolism	R00409
Valine, leucine and isoleucine degradation	R07601
Riboflavin metabolism	R00549
Glycolysis / Gluconeogenesis/ Inositol metabolism/ Fructose and	R01015

mannose metabolism/ Carbon fixation in photosynthetic organisms	
Inositol metabolism/ Inositol phosphate metabolism	R01183
Porphyrin and chlorophyll metabolism	R00036
Unknown	MEAMP1(gly-asn)
	R04733
Pantothenate and CoA biosynthesis	R03035
	R04339
Histidine metabolism	R04037
Pentose phosphate pathway/ Pentose and glucuronate interconversions	R05605
Porphyrin and chlorophyll metabolism	R06530
Nucleotide sugars metabolism/ Streptomycin biosynthesis/ Polyketide sugar unit biosynthesis	R02328
Porphyrin and chlorophyll metabolism	R05221
Pyruvate metabolism	R00203
Valine, leucine and isoleucine degradation/ Valine, leucine and isoleucine biosynthesis	R02199
Synthesis and degradation of ketone bodies/ Valine, leucine and isoleucine degradation/ Butanoate metabolism	R01978
Peptidoglycan biosynthesis	R02786
Glycine, serine and threonine metabolism	R02566
Fatty acid elongation in mitochondria/ Fatty acid metabolism	R04738
Unknown	PDE2_2
Unknown	4-methyl-3-oxo-hexanoyl-ACP:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
Riboflavin metabolism	R04457
Pentose and glucuronate interconversions	R01432
Unknown	RXN-12559.c
Pyruvate metabolism/ Reductive carboxylate cycle (CO2 fixation)	R00199
Glycine, serine and threonine metabolism/ Valine, leucine and isoleucine biosynthesis	R00996

1- and 2-Methylnaphthalene degradation	R06914
Phenylalanine, tyrosine and tryptophan biosynthesis	R01715
Pentose and glucuronate interconversions/ Galactose metabolism/ Starch and sucrose metabolism/ Nucleotide sugars metabolism	R00289
Benzoate degradation via hydroxylation/ Carbazole degradation	R00817
Fatty acid biosynthesis	R04961
Fructose and mannose metabolism	R03014
Urea cycle and metabolism of amino groups	R03180
Unknown	TRANS-RXNBWI-115637.ce.maizeexp.ILE_ILE
Glutamate metabolism/ Arginine and proline metabolism	R00707
Purine metabolism	R01245
Purine metabolism	R02143
	R05805
	R05172
Purine metabolism	R00190
Purine metabolism	R01234
Purine metabolism/ Glutamate metabolism	R01072
Biosynthesis of steroids	R05637
Purine metabolism	R00330
Porphyrin and chlorophyll metabolism	R05225
Unknown	DASYN161
Histidine metabolism	R04640
Folate biosynthesis	R05048
Unknown	isoheptadecanoyl-CDPdiacylglycerol-serine O-phosphatidyltransferase
Fatty acid biosynthesis	R04543
Unknown	PSSA161
Aminosugars metabolism	R02059
Propanoate metabolism	R03157
Citrate cycle (TCA cycle)/ Propanoate metabolism/ Reductive carboxylate cycle (CO2 fixation)	R00405
Unknown	8-methyl-decanoyl-ACP:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
Glutamate metabolism/ Arginine and proline	R00243

metabolism/ D-Glutamine and D-glutamate metabolism/ Nitrogen metabolism	
Unknown	ETHSabc
beta-Alanine metabolism/ Propanoate metabolism	R03045
Propanoate metabolism	R00740
Unknown	3.6.3.36-RXN.ce
Histidine metabolism	R02288
Glutamate metabolism/ Butanoate metabolism	R01648
Unknown	Gly-Leu ABC transporters
Methionine metabolism	R00177
	R01740
Glyoxylate and dicarboxylate metabolism	R01394
Unknown	HEMEAS
	R00081
Tryptophan metabolism/ Indole and ipecac alkaloid biosynthesis	R00685
Unknown	TRANS-RXNBWI-115637.ce.maizeexp.SER_SER
Glycolysis / Gluconeogenesis	R01600
	R00424
Unknown	7-methyl-3-oxo-octanoyl-ACP:NADP+ oxidoreductase
Unknown	MDDEP4pp
Unknown	ALDD31
Fatty acid biosynthesis/ Biosynthesis of unsaturated fatty acids	R07763
Valine, leucine and isoleucine biosynthesis	R01215
Thiamine metabolism	R03471
Phenylalanine metabolism	R06782
Unknown	MEAMP1(gly-pro-L)
Phenylalanine, tyrosine and tryptophan biosynthesis	R01373
Pyrimidine metabolism	R00511
Urea cycle and metabolism of amino groups	R01990
Glyoxylate and dicarboxylate metabolism	R01747
Nicotinate and nicotinamide metabolism	R02323
Unknown	RXN0-6978.c
Glycine, serine and threonine metabolism	R01771
Galactose metabolism	R03033
Unknown	R01068

Tetracycline biosynthesis/ Biosynthesis of type II polyketide products	R05463
Biosynthesis of steroids	R05634
Purine metabolism/ Alanine and aspartate metabolism	R01135
Unknown	9-methyl-decanoyl-ACP:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
Purine metabolism	R01126
Starch and sucrose metabolism	R02778
Inositol phosphate metabolism/ Phosphatidylinositol signaling system	R01187
Phenylalanine, tyrosine and tryptophan biosynthesis	R01714
Unknown	ACOAD5
Inositol metabolism	R05378
Toluene and xylene degradation	R05295
Unknown	ACGApts
Unknown	R08682
Galactose metabolism/ Nucleotide sugars metabolism	R00291
Glyoxylate and dicarboxylate metabolism/ One carbon pool by folate	R01655
Tyrosine metabolism/ Phenylalanine, tyrosine and tryptophan biosynthesis/ Novobiocin biosynthesis/ Alkaloid biosynthesis I	R00734
Fatty acid biosynthesis	R04429
Unknown	Gly-Try aminopeptidase
Porphyrin and chlorophyll metabolism	R05222
Urea cycle and metabolism of amino groups	R02649
Galactose metabolism	R03387
Fatty acid metabolism	R04754
Purine metabolism	R02090
Lysine biosynthesis/ Peptidoglycan biosynthesis	R02788
Unknown	R02424
Taurine and hypotaurine metabolism/ Pyruvate metabolism	R00230
Tyrosine metabolism	R02382
Purine metabolism	R02147

Pyrimidine metabolism	R01569
Unknown	6-methyl-3-oxo-octanoyl-ACP:NADP+ oxidoreductase
1,4-Dichlorobenzene degradation	R05441
Unknown	TRANS-RXN-206.ce
Glutamate metabolism/ Tyrosine metabolism/ Butanoate metabolism	R00713
Glutamate metabolism/ Nitrogen metabolism	R00253
1,4-Dichlorobenzene degradation	R05406
Fluorobenzoate degradation	R08114
Styrene degradation	R05404
Folate biosynthesis	R03503
Fatty acid biosynthesis	R04534
Fructose and mannose metabolism	R00866
Histidine metabolism	R04558
	R08502
Purine metabolism	R01229
Pentose phosphate pathway	R02036
One carbon pool by folate	R02301
Phenylalanine, tyrosine and tryptophan biosynthesis	R00674
Purine metabolism	R04591
Unknown	12-methyl-3-oxo-tetra-decanoyl-ACP:NADP+ oxidoreductase
Fatty acid metabolism	R02487
Valine, leucine and isoleucine biosynthesis	R04001
Valine, leucine and isoleucine biosynthesis	R04426
Butanoate metabolism	R01171
Propanoate metabolism	R00921
Purine metabolism	R02142
Histidine metabolism	R01163
Unknown	PGSA161
Pyrimidine metabolism	R00966
Phenylalanine, tyrosine and tryptophan biosynthesis	R01731
Nitrogen metabolism	R01221
Purine metabolism	R01561
Nicotinate and nicotinamide metabolism	R03348
Glyoxylate and dicarboxylate metabolism	R00013
Unknown	14-methyl-hexa-decanoyl-ACP:[acyl-carrier-protein] transferase
Fatty acid biosynthesis	R01626

Unknown	Gly-Cys aminopeptidase
Unknown	14-methyl-3-oxo-hexa-decanoyl-ACP:NADP+ oxidoreductase
Fatty acid biosynthesis	R04355
Porphyrin and chlorophyll metabolism	R06558
Benzoate degradation via hydroxylation/ Benzoate degradation via CoA ligation	R05621
Pyruvate metabolism	R00196
Peptidoglycan biosynthesis	R05030
Phenylpropanoid biosynthesis	R07442
Unknown	ACOAD7
Unknown	5-methyl-hexanoyl-ACP:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
Fructose and mannose metabolism	R02568
Methionine metabolism/ Sulfur metabolism	R01776
Unknown	MALATE-DEHYDROGENASE-ACCEPTOR-RXN.c.metaexp.CPD-9956_UBIQUINONE-8
Urea cycle and metabolism of amino groups/ Glutamate metabolism/ Arginine and proline metabolism/ Nitrogen metabolism	R00149
Phenylalanine metabolism	R00699
Fatty acid biosynthesis/ Biosynthesis of unsaturated fatty acids	R07764
	R00188
Unknown	R00842
Valine, leucine and isoleucine degradation	R07603
Pyrimidine metabolism	R02331
Fatty acid metabolism	R01175
Glycine, serine and threonine metabolism	R00611
Arginine and proline metabolism	R01251
Phenylalanine, tyrosine and tryptophan biosynthesis	R01627
Unknown	INSTtex
Unknown	ACOAD3
Nicotinate and nicotinamide metabolism	R01268
	R07140
Unknown	TRANS-RXN-199.ce
Pyrimidine metabolism	R00573
Glutathione metabolism	R00899

	R01515
Purine metabolism	R00722
Unknown	11-methyl-3-oxo-dodecanoyl-ACP:NADP+ oxidoreductase
One carbon pool by folate	R04326
Purine metabolism	R01857
Unknown	ACOAt
Fatty acid elongation in mitochondria/ Fatty acid metabolism	R04743
Glycerophospholipid metabolism	R00848
Tyrosine metabolism	R04895
Histidine metabolism	R02285
Lysine biosynthesis	R00451
Valine, leucine and isoleucine degradation	R04224
Valine, leucine and isoleucine biosynthesis/ C5-Branched dibasic acid metabolism	R00994
Unknown	SUCFUMt
Purine metabolism	R01054
Fatty acid elongation in mitochondria/ Fatty acid metabolism	R03778
Unknown	15-methyl-3-oxo-hexa-decanoyl-ACP:NADP+ oxidoreductase
Unknown	4-methyl-hexanoyl-ACP:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
	R05684
Unknown	R00760
Benzoate degradation via hydroxylation/ 2,4-Dichlorobenzoate degradation	R01631
Caprolactam degradation	R06941
Fatty acid biosynthesis	R04958
Unknown	5.3.1.17-RXN.c
Purine metabolism	R05554
Unknown	MANpts
Tryptophan metabolism	R04917
Pyruvate metabolism	R01446
Biotin metabolism	R03231
Purine metabolism	R04144
Thiamine metabolism	R07462
Valine, leucine and isoleucine degradation	R04204
Glutamate metabolism/ D-Glutamine and D-glutamate	R00256

metabolism/ Nitrogen metabolism	
Phenylalanine, tyrosine and tryptophan biosynthesis	R03509
Unknown	11-methyl-trans-dodec-2-enoyl-ACP:NAD+ oxidoreductase (A-specific)
Glycolysis / Gluconeogenesis	R00754
Valine, leucine and isoleucine degradation/ Propanoate metabolism	R02765
Glycine, serine and threonine metabolism	R00582
Unknown	U220
Unknown	isoheptadecanoyl-phosphatidate cytidylyltransferase
Urea cycle and metabolism of amino groups	R00669
Fatty acid elongation in mitochondria/ Fatty acid metabolism	R03991
	R07237
1- and 2-Methylnaphthalene degradation	R06906
Riboflavin metabolism	R00066
	R08551
Purine metabolism	R02017
Alanine and aspartate metabolism/ D-Alanine metabolism	R00401
Phenylalanine, tyrosine and tryptophan biosynthesis	R00985
Glutamate metabolism/ Arginine and proline metabolism/ D-Glutamine and D-glutamate metabolism/ Nitrogen metabolism	R00248
Purine metabolism/ Sulfur metabolism	R00529
Valine, leucine and isoleucine degradation	R04095
Riboflavin metabolism	R03459
Unknown	3-Oxooctodecanoyl-ACP:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
Thiamine metabolism	R04509
Pyrimidine metabolism/ Alanine and aspartate metabolism	R01397
Fructose and mannose metabolism	R00875
Unknown	L-LACTDEHYDROGFMN-RXN.c.chlamyexp.CPD-9956_UBIQUINONE-8
Fatty acid metabolism	R01280
Unknown	R04260

Fatty acid elongation in mitochondria/ Fatty acid metabolism	R04746
Taurine and hypotaurine metabolism/ Reductive carboxylate cycle (CO2 fixation)	R00396
Pentose phosphate pathway	R02035
Glutathione metabolism	R00274
Aminophosphonate metabolism	rxn11649
Valine, leucine and isoleucine biosynthesis	R03968
Butanoate metabolism	R01357
Pyrimidine metabolism	R00571
	R04096
Peptidoglycan biosynthesis	R05662
Fatty acid biosynthesis	R04566
Glyoxylate and dicarboxylate metabolism	R00475
Unknown	RXN0-6973.c
Pyrimidine metabolism/ Glutamate metabolism	R00575
Tyrosine metabolism/ Alkaloid biosynthesis I	R00736
Phenylpropanoid biosynthesis	R05700
Pyruvate metabolism/ Carbon fixation in photosynthetic organisms	R00214
Fatty acid biosynthesis	R01624
Purine metabolism	R00429
Citrate cycle (TCA cycle)/ Reductive carboxylate cycle (CO2 fixation)	R01082
Lipopolysaccharide biosynthesis	R03351
Glycolysis / Gluconeogenesis/ Citrate cycle (TCA cycle)/ Valine, leucine and isoleucine biosynthesis/ Pyruvate metabolism/ Butanoate metabolism	R00014
	R07598
Folate biosynthesis	R03066
Glycine, serine and threonine metabolism	R01466
Phenylalanine, tyrosine and tryptophan biosynthesis	R01073
Tyrosine metabolism	R04132
Purine metabolism	R00089

Pyrimidine metabolism	R00513
Tyrosine metabolism	R02521
Unknown	NADH18pp
Nicotinate and nicotinamide metabolism	R02294
Fatty acid elongation in mitochondria/ Fatty acid metabolism	R04742
Glycine, serine and threonine metabolism/ Lysine biosynthesis	R01775
Benzoate degradation via CoA ligation	R05601
Urea cycle and metabolism of amino groups/ Purine metabolism/ Atrazine degradation	R00131
Pentose phosphate pathway	R01066
Pyrimidine metabolism	R00516
Unknown	9-methyl-3-oxo-decanoyl-ACP:NADP+ oxidoreductase
Thiamine metabolism	R02133
Phenylalanine metabolism	R06788
2,4-Dichlorobenzoate degradation	R02562
Unknown	R00470
Porphyrin and chlorophyll metabolism	R03220
Pyruvate metabolism	R00317
Valine, leucine and isoleucine degradation	R04138
Unknown	R00772
Glyoxylate and dicarboxylate metabolism/ One carbon pool by folate	R00944
Taurine and hypotaurine metabolism	R05651
Synthesis and degradation of ketone bodies/ Valine, leucine and isoleucine degradation/ Butanoate metabolism	R00410
Biosynthesis of steroids	R05633
Porphyrin and chlorophyll metabolism	R05223
Unknown	THYMtex
Histidine metabolism	R04035
Unknown	reac_398
Glycolysis / Gluconeogenesis	R00658
Histidine metabolism	R02914
Unknown	Gly-Phe ABC transporters

Porphyrin and chlorophyll metabolism	R03222
Glycerolipid metabolism/ Glycerophospholipid metabolism	R00851
Unknown	10-methyl-trans-dodec-2-enoyl-ACP:NAD+ oxidoreductase (A-specific)
Unknown	3.6.3.30-RXN.ce
Glycine, serine and threonine metabolism/ Glycerophospholipid metabolism	R03424
Unknown	pentso3abcpp
Pyrimidine metabolism	R07307
Pyrimidine metabolism	R01876
Unknown	GLYBt2r
Valine, leucine and isoleucine biosynthesis	R01652
Nicotinate and nicotinamide metabolism	R02295
Unknown	TRANS-RXN-139.ce
Glutamate metabolism/ Alanine and aspartate metabolism/ Carbon fixation in photosynthetic organisms	R00355
Fatty acid metabolism	R03990
Aminosugars metabolism	R00660
Unknown	CHACH
Purine metabolism	R04559
	R07475
Citrate cycle (TCA cycle)	R01899
Galactose metabolism	R01103
Purine metabolism	R04208
Phenylalanine metabolism/ Benzoate degradation via hydroxylation/ Biphenyl degradation/ Toluene and xylene degradation/ 1,4-Dichlorobenzene degradation/ Fluorene degradation/ Carbazole degradation/ Ethylbenzene degradation/ Styrene degradation	R00750
Glutathione metabolism	R00494
Purine metabolism/ One carbon pool by folate	R04560
Pentose phosphate pathway/ Glutathione metabolism	R01528
Synthesis and degradation of ketone bodies/ Butanoate metabolism	R01361

Sulfur metabolism	R02021
Methionine metabolism	R01287
Purine metabolism	R02719
Pyrimidine metabolism	R02098
Glyoxylate and dicarboxylate metabolism/ Methane metabolism	R00519
	R07238
Tyrosine metabolism	R06897
Pyrimidine metabolism	R02485
Unknown	12-methyl-tetra-decanoyl-ACP:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
Propanoate metabolism	R04424
Pyrimidine metabolism	R02093
Folate biosynthesis	R05553
One carbon pool by folate/ Methane metabolism	R07168
Fatty acid elongation in mitochondria/ Fatty acid metabolism	R04748
Valine, leucine and isoleucine degradation	R00927
Pyrimidine metabolism	R00968
Fructose and mannose metabolism	R00890
Unknown	LCYSTabc
Folate biosynthesis	R01716
Unknown	SUCCtex
Valine, leucine and isoleucine degradation/ Valine, leucine and isoleucine biosynthesis/ Pantothenate and CoA biosynthesis	R01214
Pyrimidine metabolism	R02296
Phenylalanine, tyrosine and tryptophan biosynthesis	R03084
Unknown	9-methyl-trans-dec-2-enoyl-ACP:NAD+ oxidoreductase (A-specific)
Tryptophan metabolism	R04916
Biosynthesis of steroids	R05613
Unknown	15-methyl-hexa-decanoyl-ACP:[acyl-carrier-protein] transferase
Purine metabolism	R01547
Urea cycle and metabolism of amino groups	R01989
Glycerolipid metabolism	R00847
Unknown	RXN-13418.c
Unknown	RXN0-6377.c.metaexp.NADP_NADPH
Fatty acid biosynthesis	R04536

Phenylpropanoid biosynthesis	R05701
Unknown	isoheptadecanoyl-glycerol-3-phosphate O-acyltransferase
Nicotinate and nicotinamide metabolism	R00112
Phenylalanine metabolism/ Nitrogen metabolism/ Phenylpropanoid biosynthesis/ Alkaloid biosynthesis II	R00697
Benzoate degradation via hydroxylation	R03470
Nicotinate and nicotinamide metabolism	R01724
Fatty acid elongation in mitochondria/ Fatty acid metabolism	R01177
Purine metabolism	R01130
Glycolysis / Gluconeogenesis/ Carbon fixation in photosynthetic organisms	R01063
Pentose phosphate pathway/ Methane metabolism/ Carbon fixation in photosynthetic organisms	R01621
Benzoate degradation via hydroxylation	R03307
Urea cycle and metabolism of amino groups	R01986
Nicotinate and nicotinamide metabolism	R02324
Unknown	NO3t2
Unknown	RXN0-6977.c
Cysteine metabolism	R02618
Unknown	X00002
Tryptophan metabolism	R08164
Phenylalanine, tyrosine and tryptophan biosynthesis	R01826
Tyrosine metabolism	R02698
Methionine metabolism/ One carbon pool by folate	R00946
Biosynthesis of steroids	R05636
Glycine, serine and threonine metabolism	R01513
Galactose metabolism/ Nucleotide sugars metabolism	R00502
Purine metabolism	R04463
Unknown	R02782
Unknown	isoheptadecanoyl-phosphatidylserine decarboxylase
Unknown	R00771

Unknown	MEAMP1(gly-glu)
Alanine and aspartate metabolism/ beta-Alanine metabolism	R00489
1- and 2-Methylnaphthalene degradation	R06917
Purine metabolism	R02557
Glutamate metabolism/ Butanoate metabolism	R00261
Valine, leucine and isoleucine degradation/ Valine, leucine and isoleucine biosynthesis	R01090
Fatty acid biosynthesis	R04969
Ubiquinone biosynthesis	R07263
Urea cycle and metabolism of amino groups/ beta-Alanine metabolism	R02549
Unknown	TRANS-RXNBWI-115572.ce.maizeexp.HYPOXANTHINE_HYPOXANTHINE
Taurine and hypotaurine metabolism	R05320
Fatty acid biosynthesis	R04966
Unknown	SULAabc
Unknown	DIPEPabc14
Purine metabolism	R02088
Nicotinate and nicotinamide metabolism	R00104
Citrate cycle (TCA cycle)/ Glyoxylate and dicarboxylate metabolism/ Reductive carboxylate cycle (CO2 fixation)	R01900
Aminosugars metabolism	R05332
Unknown	14-methyl-trans-hexa-dec-2-enoyl-ACP:NAD+ oxidoreductase (A-specific)
Riboflavin metabolism/ Porphyrin and chlorophyll metabolism	R04148
Tyrosine metabolism	R02656
Purine metabolism	R01969
Glycine, serine and threonine metabolism/ Methionine metabolism/ Nitrogen metabolism	R01001
Galactose metabolism	R03634
Synthesis and degradation of ketone bodies/ Valine, leucine and isoleucine degradation/ Butanoate metabolism	R01360

Naphthalene and anthracene degradation	R05643
Fatty acid elongation in mitochondria/ Fatty acid metabolism	R04741
Unknown	MEPROPCOA-FAD-RXN.c
Arginine and proline metabolism	R01566
Pyrimidine metabolism	R01878
Unknown	R02047
Purine metabolism	R02019
Methionine metabolism	R00192
Nucleotide sugars metabolism/ Streptomycin biosynthesis/ Polyketide sugar unit biosynthesis	R02777
Unknown	10-methyl-dodecanoyl-ACP:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
Unknown	R01641
Unknown	R01698
Butanoate metabolism/ Reductive carboxylate cycle (CO2 fixation)	R00408
Unknown	R03050
Glycerolipid metabolism	R01036
Unknown	TSULabc
Histidine metabolism	R03457
Pentose phosphate pathway	R01737
Unknown	(S)-Dihydroorotate:(acceptor) oxidoreductase
Urea cycle and metabolism of amino groups/ Alanine and aspartate metabolism/ Arginine and proline metabolism	R01954
Pantothenate and CoA biosynthesis	R01226
Inositol metabolism	R08503
Porphyrin and chlorophyll metabolism	R05177
beta-Alanine metabolism	R00924
Pyrimidine metabolism	R02018
Propanoate metabolism	R01353
Unknown	Gly-Cys ABC transporters
Unknown	isohexadecanoyl-Diacylglycerol kinase
One carbon pool by folate/ Folate biosynthesis	R02236
Unknown	AGt3
Propanoate metabolism	R04425
Unknown	4-methyl-trans-hex-2-enoyl-ACP:NAD+ oxidoreductase (A-specific)

Fatty acid biosynthesis	R04968
Fatty acid elongation in mitochondria/ Fatty acid metabolism	R04740
	R02102
Glycine, serine and threonine metabolism	R01022
Nicotinate and nicotinamide metabolism	R00481
Unknown	3-methylbutanoyl-ACP:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
Pentose phosphate pathway/ Glutathione metabolism	R02736
Urea cycle and metabolism of amino groups	R03313
Phenylalanine metabolism/ Benzoate degradation via hydroxylation/ Biphenyl degradation/ Toluene and xylene degradation/ 1,4-Dichlorobenzene degradation/ Fluorene degradation/ Carbazole degradation/ Ethylbenzene degradation/ Styrene degradation	R02601
Purine metabolism	R02297
Styrene degradation	R05551
Glutamate metabolism/ Nitrogen metabolism	R00114
Pyrimidine metabolism	R00158
Pyrimidine metabolism	R00964
Valine, leucine and isoleucine degradation	R07600
Unknown	TRANS-RXNAVI-26568.ce
Biosynthesis of steroids/ Terpenoid biosynthesis	R02003
Citrate cycle (TCA cycle)/ Glyoxylate and dicarboxylate metabolism/ Reductive carboxylate cycle (CO2 fixation)	R01325
Tyrosine metabolism	R04131
D-Glutamine and D-glutamate metabolism/ Peptidoglycan biosynthesis	R02783
Peptidoglycan biosynthesis	R05630
Purine metabolism	R01676
Unknown	R503-RXN.c
Porphyrin and chlorophyll metabolism	R03197
Galactose metabolism	R01064

Pyrimidine metabolism	R00570
Ubiquinone biosynthesis	R04030
Unknown	DIPEPabc10
Ubiquinone biosynthesis/ Biosynthesis of siderophore group nonribosomal peptides	R01717
Unknown	TRANS-RXNBWI-115637.ce.maizeexp.VAL_VAL
Pentose phosphate pathway/ Purine metabolism	R01057
Peptidoglycan biosynthesis	R05627
Purine metabolism	R02748
Galactose metabolism	R01092
Unknown	ACOAD4
Ubiquinone biosynthesis	R04031
Purine metabolism	R02423
Unknown	MSO3abc
	R06064
Fatty acid metabolism	R01279
Glutamate metabolism/ D-Glutamine and D-glutamate metabolism	R00260
Fatty acid biosynthesis	R04955
beta-Alanine metabolism	R00904
Unknown	3.6.3.21-RXN.ce.metaexp.GLN_GLN
Fatty acid metabolism	R04751
Arginine and proline metabolism	R01248
Alanine and aspartate metabolism/ Cyanoamino acid metabolism/ Nitrogen metabolism	R00485
Citrate cycle (TCA cycle)/ Glyoxylate and dicarboxylate metabolism	R00351
Pentose and glucuronate interconversions/ Methane metabolism	R05339
	R05171
Purine metabolism/ One carbon pool by folate	R04325
Nicotinate and nicotinamide metabolism/ Nitrogen metabolism	R00189
Unknown	DIPEPabc12
	R04564
Benzoate degradation via CoA ligation	R01238
Unknown	DIPEPabc8

Alanine and aspartate metabolism/ Nitrogen metabolism	R00490
Unknown	8-methyl-3-oxo-decanoyl-ACP:NADP+ oxidoreductase
Purine metabolism	R02556
Fructose and mannose metabolism	R02263
Unknown	ADNK4
	R00321
	R05683
Unknown	4-methyl-3-oxo-hexanoyl-ACP:NADP+ oxidoreductase
Biotin metabolism	R03210
Pyrimidine metabolism	R02484
Unknown	THD2pp
Purine metabolism	R00332
Citrate cycle (TCA cycle)	R00621
Folate biosynthesis	R00428
Arginine and proline metabolism	R00667
Unknown	MANt2
Urea cycle and metabolism of amino groups	R03443
Tryptophan metabolism	R03096
beta-Alanine metabolism/ Pantothenate and CoA biosynthesis	R02473
Arginine and proline metabolism	R00245
	R01573
Sulfur metabolism	R00858
Fructose and mannose metabolism	R01818
Urea cycle and metabolism of amino groups/ Atrazine degradation	R00005
Phenylalanine, tyrosine and tryptophan biosynthesis	R03508
Urea cycle and metabolism of amino groups	R01416
	R01178
Unknown	isohexadecanoyl-1-acylglycerol-3-phosphate O-acyltransferase
Unknown	trans-Octodec-2-enoyl-ACP:NAD+ oxidoreductase (A-specific)
Phenylalanine, tyrosine and tryptophan biosynthesis	R02722
Pantothenate and CoA biosynthesis	R03269
Unknown	ACOAD19
Pyrimidine metabolism	R02100

Purine metabolism	R01560
Glyoxylate and dicarboxylate metabolism	R01393
Tryptophan metabolism	R00678
Pyrimidine metabolism	R02094
Unknown	10-methyl-3-oxo-dodecanoyl-ACP:NADP+ oxidoreductase
	R03854
Vitamin B6 metabolism	R00278
Riboflavin metabolism	R03458
Fatty acid elongation in mitochondria/ Fatty acid metabolism	R04745
	R06974
Benzoate degradation via hydroxylation	R02991
Unknown	MAL31
Pyrimidine metabolism	R00970
Urea cycle and metabolism of amino groups/ Arginine and proline metabolism	R01398
Valine, leucine and isoleucine degradation	R03172
Phenylalanine, tyrosine and tryptophan biosynthesis	R00732
Unknown	5-methyl-trans-hex-2-enoyl-ACP:NAD+ oxidoreductase (A-specific)
Toluene and xylene degradation	R04089
Unknown	TRANS-RXNBWI-115637.ce.maizeexp.PRO_PRO
Glycine, serine and threonine metabolism	R04173
Folate biosynthesis	R02237
One carbon pool by folate/ Folate biosynthesis	R00939
Unknown	4.2.3.14-RXN.c
Benzoate degradation via hydroxylation/ Benzoate degradation via CoA ligation	R05622
Folate biosynthesis	R00942
Glycerolipid metabolism	R01752
Pyrimidine metabolism	R02016
Lysine biosynthesis	R02734
Unknown	CO2t
Pyrimidine metabolism	R00156
Fatty acid biosynthesis	R04960
Pyrimidine metabolism/ One carbon pool by folate	R02101
Pyruvate metabolism/ Glyoxylate and dicarboxylate metabolism	R00472

Purine metabolism	R01677
Phenylalanine metabolism/ Phenylalanine, tyrosine and tryptophan biosynthesis	R00694
Urea cycle and metabolism of amino groups	R00774
Pyruvate metabolism	R02530
Unknown	R00416
Unknown	rbs_360
Purine metabolism	R01137
1- and 2-Methylnaphthalene degradation	R06923
Ubiquinone biosynthesis	R04988
Unknown	RXNIM-6209
Unknown	stearoyl-cardiolipin synthase
Biotin metabolism	R05145
Ubiquinone biosynthesis	R08165
Unknown	isoheptadecanoyl-1-acylglycerol-3-phosphate O-acyltransferase
Benzoate degradation via hydroxylation/ Toluene and xylene degradation	R01294
Unknown	8-methyl-3-hydroxy-decanoyl-ACP hydro-lyase
Lipopolysaccharide biosynthesis	R05646
Unknown	3-deoxy-D-manno-octulosonic acid transferase
Ubiquinone biosynthesis	R04990
Ubiquinone biosynthesis	R04986
Unknown	DHNAOT
Phenylalanine, tyrosine and tryptophan biosynthesis	R03460
Porphyrin and chlorophyll metabolism	R00310
Unknown	R03051
Unknown	HEPK1
Riboflavin metabolism	R07280
Unknown	12-methyl-3-hydroxy-tetra-decanoyl-ACP hydro-lyase
Unknown	14-methyl-3-hydroxy-hexa-decanoyl-ACP hydro-lyase
Ubiquinone biosynthesis	R06146
Unknown	GLCTR1
Unknown	10-methyl-3-hydroxy-dodecanoyl-ACP hydro-lyase
Unknown	NCAIR synthetase and NCAIR mutase
Lipopolysaccharide biosynthesis	R05074
Unknown	PQQ recharge using O2
Unknown	2-OCTAPRENYL-6-METHOXYPHENOL-HYDROX-RXN
Unknown	Acetyl-CoA:(acceptor) oxidoreductase
Lipopolysaccharide biosynthesis	R04658

D-Glutamine and D-glutamate metabolism/ Peptidoglycan biosynthesis	R03193
	R04548
Unknown	13-methyl-3-hydroxy-tetra-decanoyl-ACP hydro-lyase
Glycine, serine and threonine metabolism/ Porphyrin and chlorophyll metabolism	R00830
Ubiquinone biosynthesis	R05615
Lipopolysaccharide biosynthesis	R05644
	R01576
Unknown	11-methyl-3-hydroxy-dodecanoyl-ACP hydro-lyase
Pyrimidine metabolism	R01869
Unknown	GALM2pp
Lipopolysaccharide biosynthesis	R05176
Riboflavin metabolism/ Porphyrin and chlorophyll metabolism	R04594
D-Alanine metabolism/ Peptidoglycan biosynthesis	R01150
Lysine biosynthesis	R04199
Lipopolysaccharide biosynthesis	R04606
Lipopolysaccharide biosynthesis	R05647
	R01530
Unknown	MECDPDH2
Unknown	stearoyl-1-acylglycerol-3-phosphate O-acyltransferase
Unknown	CYRDAS
Unknown	FAS140COA
Unknown	R_PYDXS
Unknown	isohexadecanoyl-CDPdiacylglycerol:sn-glycerol-3-phosphate 3-phosphatidyltransferase
Lysine biosynthesis	R02292
Unknown	6-methyl-3-hydroxy-octanoyl-ACP hydro-lyase
Lysine biosynthesis/ Peptidoglycan biosynthesis	R04617
Unknown	HEPT3
Lipopolysaccharide biosynthesis	R04549
Unknown	GLCTR2
Unknown	GLCtex
Unknown	UACGAMtex
Fatty acid elongation in mitochondria/ Fatty acid metabolism	R04749
Unknown	RXN0-5061.c
Unknown	RXN0-5118.c

Propanoate metabolism	R06987
Lipopolysaccharide biosynthesis	R04587
Folate biosynthesis	R04620
Unknown	PGPP161pp
Unknown	R5Ptex
Unknown	7-methyl-3-hydroxy-octanoyl-ACP hydro-lyase
Unknown	myristoyl-cardiolipin synthase
Unknown	5-methyl-3-hydroxy-hexanoyl-ACP hydro-lyase
Unknown	15-methyl-3-hydroxy-hexa-decanoyl-ACP hydro-lyase
Unknown	HEPK2
Histidine metabolism	R01071
Unknown	DHBSH
Unknown	HEMEOSm
Valine, leucine and isoleucine degradation	R02085
Lipopolysaccharide biosynthesis	R04657
Unknown	9-methyl-3-hydroxy-decanoyl-ACP hydro-lyase
Lipopolysaccharide biosynthesis	R05645
Unknown	4-methyl-3-hydroxy-hexanoyl-ACP hydro-lyase
Unknown	PDXA-RXN
Unknown	EX_H+[e0]
Unknown	EX_Fumarate[e0]
Unknown	EX_Nitrite[e0]
Unknown	EX_Succinate[e0]
Unknown	EX_Xylose[e0]
Unknown	EX_Glycine[e0]
Unknown	EX_L-alanylglycine[e0]
Unknown	EX_ala-L-Thr-L[e0]
Unknown	EX_PPi[e0]
Unknown	EX_ala-L-glu-L[e0]
Unknown	EX_D-Fructose[e0]
Unknown	EX_Glycerol-3-phosphate[e0]
Unknown	EX_H2O[e0]
Unknown	EX_L-Leucine[e0]
Unknown	EX_D-Mannitol[e0]
Unknown	EX_Cys-Gly[e0]
Unknown	EX_Gly-Met[e0]
Unknown	EX_Phosphate[e0]
Unknown	EX_L-Methionine[e0]
Unknown	EX_fe3[e0]
Unknown	EX_ala-L-asp-L[e0]
Unknown	EX_D-Glucose[e0]
Unknown	EX_Sucrose[e0]

Unknown	EX_L-Isoleucine[e0]
Unknown	EX_ethanesulfonate[e0]
Unknown	EX_Taurine[e0]
Unknown	EX_Gly-Phe[e0]
Unknown	EX_L-Serine[e0]
Unknown	EX_N-Acetyl-D-glucosamine[e0]
Unknown	EX_Gly-Cys[e0]
Unknown	EX_XAN[e0]
Unknown	EX_Gly-Gln[e0]
Unknown	EX_Citrate[e0]
Unknown	EX_Sorbitol[e0]
Unknown	EX_Cytosine[e0]
Unknown	EX_D-Galactonate[e0]
Unknown	EX_D-Mannose[e0]
Unknown	EX_Thiamin[e0]
Unknown	EX_Gly-Tyr[e0]
Unknown	EX_L-Proline[e0]
Unknown	EX_BET[e0]
Unknown	EX_Co2+[e0]
Unknown	EX_Cl-[e0]
Unknown	EX_Ala-Leu[e0]
Unknown	EX_Aminoethanol[e0]
Unknown	EX_L-Cysteate[e0]
Unknown	EX_gly-asp-L[e0]
Unknown	EX_Nitrate[e0]
Unknown	EX_TRHL[e0]
Unknown	EX_HYXN[e0]
Unknown	EX_Sulfoacetate[e0]
Unknown	EX_gly-pro-L[e0]
Unknown	EX_H2S2O3[e0]
Unknown	EX_Gly-Leu[e0]
Unknown	EX_O2[e0]
Unknown	EX_gly-glu-L[e0]
Unknown	EX_L-Valine[e0]
Unknown	EX_methanesulfonate[e0]
Unknown	EX_Ala-Gln[e0]
Unknown	EX_L-Glutamine[e0]
Unknown	EX_met-L-ala-L[e0]
Unknown	EX_gly-asn-L[e0]
Unknown	EX_Ala-His[e0]
Unknown	EX_L-Malate[e0]
Unknown	EX_L-Inositol[e0]
Unknown	EX_CO2[e0]
Unknown	ATPM
Purine metabolism	R00720

Purine metabolism	R01855
Purine metabolism	R04209
Arginine and proline metabolism	R00557
Phenylalanine metabolism/ Phenylalanine, tyrosine and tryptophan biosynthesis	R00688
Benzoate degradation via hydroxylation	R00829
Toluene and xylene degradation	R03566
Toluene and xylene degradation	R03608
2,4-Dichlorobenzoate degradation	R05274
	R03981
Isoflavonoid biosynthesis	R07778
Glyoxylate and dicarboxylate metabolism/ One carbon pool by folate	R00943
Glycine, serine and threonine metabolism/ Glyoxylate and dicarboxylate metabolism	R01388
1,2-Dichloroethane degradation	R05284
Propanoate metabolism	R00919
3-Chloroacrylic acid degradation	R05367
3-Chloroacrylic acid degradation	R05368
Butanoate metabolism	R03027
Methane metabolism	R01142
Atrazine degradation	R05558
Atrazine degradation	R05559
Porphyrin and chlorophyll metabolism	R06895
Caprolactam degradation	R05231
Purine metabolism	R00776
Glycine, serine and threonine metabolism	R00371
gamma-Hexachlorocyclohexane degradation	R06834
Novobiocin biosynthesis	R06776
gamma-Hexachlorocyclohexane degradation	R06857
Fluorobenzoate degradation	R08118
Tryptophan metabolism	R00988
Glycerophospholipid metabolism	R01801

Glycerophospholipid metabolism	R02027
Naphthalene and anthracene degradation	R03543
Benzoate degradation via CoA ligation	R05311
Methane metabolism	R00527
Glycine, serine and threonine metabolism	R01465
Arginine and proline metabolism	R01563
Benzoate degradation via hydroxylation	R02990
1,4-Dichlorobenzene degradation	R04259
Toluene and xylene degradation	R05442
Toluene and xylene degradation	R05443
Propanoate metabolism	R01359
Taurine and hypotaurine metabolism	R05652
2,4-Dichlorobenzoate degradation	R02877
Biotin metabolism	R03209
Phenylalanine, tyrosine and tryptophan biosynthesis	R01872
Biosynthesis of steroids	R02063
Ascorbate and aldarate metabolism	R02754
Ascorbate and aldarate metabolism	R03277
Unknown	3-demethylubiquinone-9 3-o-methyltransferase
Tyrosine metabolism	R02695
Benzoate degradation via hydroxylation	R02762
Tryptophan metabolism	R03889
Pyruvate metabolism	R01736
Biphenyl degradation	R03462
Biphenyl degradation	R05245
Toluene and xylene degradation	R05353
Naphthalene and anthracene degradation	R07708
Styrene degradation	R07854
Biotin metabolism	R03182
Porphyrin and chlorophyll metabolism	R06529
Porphyrin and chlorophyll metabolism	R07302
Sulfur metabolism	R00533
Unknown	R02508

Caprolactam degradation	R02231
Glycolysis / Gluconeogenesis/ Reductive carboxylate cycle (CO2 fixation)	R00235
Glycolysis / Gluconeogenesis	R01602
Citrate cycle (TCA cycle)	R00362
Pentose and glucuronate interconversions	R05338
Fructose and mannose metabolism	R00885
Galactose metabolism/ Nucleotide sugars metabolism	R00955
Fatty acid biosynthesis	R04428
Ubiquinone biosynthesis	R06858
Ubiquinone biosynthesis	R06859
Urea cycle and metabolism of amino groups	R01151
Urea cycle and metabolism of amino groups	R01157
Urea cycle and metabolism of amino groups	R04025
Urea cycle and metabolism of amino groups	R05050
Purine metabolism	R00430
Purine metabolism	R00434
Purine metabolism	R01138
Purine metabolism	R01230
Purine metabolism/ Glutamate metabolism	R01231
Purine metabolism	R01856
Purine metabolism	R01858
Purine metabolism	R02106
Purine metabolism	R03409
Pyrimidine metabolism	R00568
Pyrimidine metabolism	R01570
Pyrimidine metabolism	R02325
Pyrimidine metabolism/ One carbon pool by folate	R06613
Alanine and aspartate metabolism/ beta-Alanine metabolism/ Propanoate metabolism	R00908
Glycerophospholipid metabolism/ Phosphatidylinositol signaling system	R01799
Glycine, serine and threonine metabolism	R06977

Glycine, serine and threonine metabolism	R06978
Glycine, serine and threonine metabolism	R06979
Glycine, serine and threonine metabolism	R08050
Methionine metabolism	R00999
Methionine metabolism	R03217
	R03248
Cysteine metabolism/ Sulfur metabolism	R00897
Cysteine metabolism	R04859
Valine, leucine and isoleucine degradation/ Propanoate metabolism	R00833
Valine, leucine and isoleucine degradation/ Valine, leucine and isoleucine biosynthesis	R01088
	R05061
Valine, leucine and isoleucine biosynthesis	R02196
Valine, leucine and isoleucine biosynthesis/ C5-Branched dibasic acid metabolism	R03896
Valine, leucine and isoleucine biosynthesis/ C5-Branched dibasic acid metabolism	R03898
Valine, leucine and isoleucine biosynthesis	R04440
Lysine biosynthesis	R00457
Arginine and proline metabolism	R03187
Arginine and proline metabolism	R03291
Arginine and proline metabolism	R04444
Arginine and proline metabolism	R05051
Tyrosine metabolism/ Styrene degradation	R01364
Tyrosine metabolism	R01645
Tyrosine metabolism	R03868
Tyrosine metabolism	R04304
Phenylalanine metabolism/ Styrene degradation	R02536
Phenylalanine metabolism	R02539
Phenylalanine metabolism	R02603
Phenylalanine metabolism	R02613
Phenylalanine metabolism	R06783

Phenylalanine metabolism	R06789
gamma-Hexachlorocyclohexane degradation/ 1,2-Dichloroethane degradation	R05287
gamma-Hexachlorocyclohexane degradation	R06835
gamma-Hexachlorocyclohexane degradation	R06838
Benzoate degradation via hydroxylation/ 2,4-Dichlorobenzoate degradation	R01298
Benzoate degradation via hydroxylation	R02672
Benzoate degradation via hydroxylation/ Fluorene degradation	R03966
Benzoate degradation via hydroxylation	R04061
Geraniol degradation	R08099
Fluorobenzoate degradation	R08100
Fluorobenzoate degradation	R08108
Fluorobenzoate degradation	R08106
Fluorobenzoate degradation	R08110
Fluorobenzoate degradation	R08111
Fluorobenzoate degradation	R08115
Fluorobenzoate degradation	R08119
Fluorobenzoate degradation	R08120
Tryptophan metabolism	R00679
Tryptophan metabolism	R02173
Tryptophan metabolism	R02678
Tryptophan metabolism	R03936
Phenylalanine, tyrosine and tryptophan biosynthesis/ Novobiocin biosynthesis	R01728
beta-Alanine metabolism	R03139
Taurine and hypotaurine metabolism	R01682
Taurine and hypotaurine metabolism	R02466
Selenoamino acid metabolism	R03601
Aminosugars metabolism	R03191
Inositol metabolism	R05661
	R07324
Pyruvate metabolism	R00236
Pyruvate metabolism	R00316

Pyruvate metabolism	R03145
Toluene and xylene degradation	R01419
	R01487
Toluene and xylene degradation	R01764
Toluene and xylene degradation	R05282
Toluene and xylene degradation	R05289
Toluene and xylene degradation	R05347
Toluene and xylene degradation	R05348
Toluene and xylene degradation	R05663
Toluene and xylene degradation	R05664
Toluene and xylene degradation	R07666
1- and 2-Methylnaphthalene degradation	R06915
Naphthalene and anthracene degradation/ Fluorene degradation	R00818
1,4-Dichlorobenzene degradation	R03891
1,4-Dichlorobenzene degradation	R03893
1,4-Dichlorobenzene degradation	R05390
1,4-Dichlorobenzene degradation	R05511
Glyoxylate and dicarboxylate metabolism/ Nitrogen metabolism	R00524
Glyoxylate and dicarboxylate metabolism	R00717
Glyoxylate and dicarboxylate metabolism	R02545
Glyoxylate and dicarboxylate metabolism	R06180
Benzoate degradation via CoA ligation	R05586
3-Chloroacrylic acid degradation	R05237
3-Chloroacrylic acid degradation	R05238
Styrene degradation	R05509
1,2-Dichloroethane degradation	R05286
Propanoate metabolism	R00926
Propanoate metabolism	R01354

Styrene degradation	R05379
Carotenoid biosynthesis - General	R07852
Butanoate metabolism	R04000
One carbon pool by folate	R00936
Thiamine metabolism	R00615
Thiamine metabolism	R02135
Nicotinate and nicotinamide metabolism	R00103
Fatty acid biosynthesis/ Biosynthesis of unsaturated fatty acids	R07765
	R02364
Porphyrin and chlorophyll metabolism	R05219
Unknown	rxn06885
Porphyrin and chlorophyll metabolism	R07412
Caprolactam degradation	R06943
	R03758
	R04443
Histidine metabolism	R04875
Porphyrin and chlorophyll metabolism	R02272
Porphyrin and chlorophyll metabolism	R03948
Porphyrin and chlorophyll metabolism	R05149
Porphyrin and chlorophyll metabolism	R05150
Porphyrin and chlorophyll metabolism	R05180
Porphyrin and chlorophyll metabolism	R05181
	R05217
Terpenoid biosynthesis	R06447
Limonene and pinene degradation	R05784
Caprolactam degradation	R06942
Urea cycle and metabolism of amino groups	R01152
Lysine biosynthesis	R04475
Phenylpropanoid biosynthesis	R04998
Histidine metabolism	R00254
Tyrosine metabolism	R01085
gamma-Hexachlorocyclohexane degradation	R06849

gamma-Hexachlorocyclohexane degradation	R06840
gamma-Hexachlorocyclohexane degradation	R06850
Benzoate degradation via hydroxylation	R03892
Fluorobenzoate degradation	R08105
	R04293
	R01650
	R01410
Glycerophospholipid metabolism	R02029
Toluene and xylene degradation	R05290
Toluene and xylene degradation	R05292
Toluene and xylene degradation	R05314
	R05309
Benzoate degradation via hydroxylation	R02589
Toluene and xylene degradation	R05389
Toluene and xylene degradation	R05428
Toluene and xylene degradation	R05665
Benzoate degradation via hydroxylation	R01427
Naphthalene and anthracene degradation	R05534
Naphthalene and anthracene degradation	R05536
	R05535
1,4-Dichlorobenzene degradation	R07784
Fluorene degradation	R05538
Fluorene degradation	R05540
Fluorene degradation	R05542
Fluorene degradation	R05544
Unknown	rxn06929
Benzoate degradation via hydroxylation/ 1,4-Dichlorobenzene degradation/ Carbazole degradation	R02604
Benzoate degradation via CoA ligation	R05305
Styrene degradation	R05362

Porphyrin and chlorophyll metabolism	R03166
Caprolactam degradation	R05099
	R04496
2,4-Dichlorobenzoate degradation	R05699
Unknown	GLYCtex
Unknown	GLYCTO2
Unknown	GLYCTO3
Unknown	RXN0-6987.c
Unknown	RXN0-6979.c
Unknown	RXN0-6975.c
Unknown	RXN0-6974.c
Unknown	MEAMP1(ala-thr)
	R05169
	R05170
	R05576
Unknown	FAS1_6_c
Unknown	FAS1_7_c
gamma-Hexachlorocyclohexane degradation	R06833
gamma-Hexachlorocyclohexane degradation	R06839
Unknown	3.6.3.23-RXN.ce.metaexp.ALA-GLY_ALA-GLY
Unknown	3.6.3.23-RXN.ce.metaexp.CYS-GLY_CYS-GLY
Unknown	3.6.3.23-RXN.ce.metaexp.CPD-13393_CPD-13393
Unknown	3.6.3.23-RXN.ce.metaexp.CPD-13404_CPD-13404
Unknown	3.6.3.23-RXN.ce.metaexp.CPD-13394_CPD-13394
Unknown	3.6.3.23-RXN.ce.metaexp.CPD-13398_CPD-13398
Unknown	3.6.3.23-RXN.ce.metaexp.CPD-13406_CPD-13406
Unknown	3.6.3.23-RXN.ce.metaexp.CPD-13403_CPD-13403
Unknown	3.6.3.23-RXN.ce.metaexp.CPD-13401_CPD-13401
Unknown	R01289
Glycerophospholipid metabolism	R02038
Unknown	R_R02222_c
	R02223
Unknown	3.6.3.22-RXN.ce.metaexp.LEU_LEU
Unknown	3.6.3.22-RXN.ce.metaexp.MET_MET
Unknown	3.6.3.22-RXN.ce.metaexp.ILE_ILE
Unknown	3.6.3.22-RXN.ce.metaexp.VAL_VAL
Histidine metabolism	R03012
	R01158
Atrazine degradation	R05562
	R05563

Tetracycline biosynthesis/ Pyruvate metabolism/ Propanoate metabolism	R00742
Alanine and aspartate metabolism/ beta-Alanine metabolism/ Propanoate metabolism	R00907
Ascorbate and aldarate metabolism	R08056
Ascorbate and aldarate metabolism	R02752
Unknown	TRANS-RXN-1.cp
Unknown	FRD2
Cysteine metabolism	R03105
Nicotinate and nicotinamide metabolism	R03005
Nicotinate and nicotinamide metabolism	R00137
	R04133
Tyrosine metabolism	R04380
One carbon pool by folate	R02300
	R04124
Unknown	BALAtex
Unknown	BSORx
Unknown	FACOAL181
Unknown	RXN-7904.c.metaexp.CPD-7836_TETRADECANOYL-COA
Porphyrin and chlorophyll metabolism	R03947
Unknown	pd001
Unknown	pd002
Unknown	pd003
Unknown	pd004
Unknown	pd005
Unknown	pd013
Unknown	pd016
Unknown	pd018
Unknown	pd019
Unknown	pd020
Unknown	pd023
Unknown	pd024
Unknown	pd025
Unknown	pd038
Unknown	pd039
Unknown	pd040
Unknown	pd041
Unknown	pd042
Unknown	pd044

Unknown	pd046
Unknown	pd047
Unknown	pd048
Unknown	pd049
Unknown	pd050
Unknown	pd051
Unknown	pd052
Unknown	pd054
Unknown	pd055
Unknown	pd056
Unknown	pd057
Unknown	pd059
Unknown	pd060
Unknown	pd061
Unknown	pd062
Unknown	pd064
Unknown	pd066
Unknown	pd067
Unknown	pd068
Unknown	pd069
Unknown	pd070
Unknown	pd071
Unknown	pd073
Unknown	pd074
Unknown	pd075
Unknown	pd076
Unknown	pd077
Unknown	pd078
Unknown	pd079
Unknown	pd080
Unknown	pd081
Unknown	pd082
Unknown	pd083
Unknown	pd084
Unknown	pd085
Unknown	pd088
Unknown	pd089
Unknown	pd091
Unknown	pd095
Unknown	pd096
Toluene and xylene degradation	R03560
Unknown	pd106
Toluene and xylene degradation	R07663
Unknown	pd111

Unknown	pd110
Unknown	pd112
Unknown	pd113
Unknown	pd125
Unknown	pd138
Unknown	pd139
Unknown	pd140
Unknown	pd141
Unknown	pd142
Unknown	pd150
Unknown	pd157
Unknown	pd170
Unknown	pd171
Unknown	pd172
Unknown	pd173
Unknown	pd174
Unknown	pd205
Unknown	pd242
Tryptophan metabolism	R04323
Unknown	pd257
Unknown	pd258
Unknown	pd260
Unknown	pd261
Unknown	pd282
Unknown	pd283
Unknown	pd284
Unknown	pd285
Unknown	pd287
Unknown	pd288
Unknown	pd289
Unknown	pd290
Unknown	pd297
Unknown	pd301
Unknown	pd303
Unknown	pd304
Unknown	pd305
Unknown	pd306
Unknown	pd307
Unknown	pd308
Unknown	pd309
Unknown	pd310
Unknown	pd311
Unknown	pd312
Unknown	pd313
Unknown	pd315

Unknown	pd316
Unknown	pd317
Unknown	pd319
Unknown	pd325
Unknown	pd328
Unknown	pd329
Unknown	pd361
Unknown	pd362
Unknown	pd363
Unknown	R09883
Unknown	R09859
Unknown	R09818
	R04597
Unknown	R09882
Unknown	pd400
Unknown	pd404
Unknown	RETINYL-PALMITATE-ESTERASE-RXN.c
Unknown	pd411
Unknown	pd417
Unknown	pd422
Unknown	pd423
Unknown	pd424
Unknown	R09705
Benzoate degradation via hydroxylation	R04277
	R01999
Unknown	pd433
Fatty acid biosynthesis	R04544
Unknown	R11353
	R00614
Unknown	R00895
	R01686
	R04861
Ubiquinone biosynthesis	R07262
Unknown	RXN-1521.c
Flavonoid biosynthesis/ Isoflavonoid biosynthesis	R07712
Arginine and proline metabolism	R01884
Pentose phosphate pathway	R01519
Toluene and xylene degradation	R05138
1- and 2-Methylnaphthalene degradation	R06918
1- and 2-Methylnaphthalene degradation	R06919

1- and 2-Methylnaphthalene degradation	R06920
1- and 2-Methylnaphthalene degradation	R06921
1- and 2-Methylnaphthalene degradation	R06922
1- and 2-Methylnaphthalene degradation	R06924
1- and 2-Methylnaphthalene degradation	R06925
Taurine and hypotaurine metabolism	R01681
Unknown	R00006
beta-Alanine metabolism	R01914
	R03816
Tryptophan metabolism	R01937
Tryptophan metabolism	R01933
Toluene and xylene degradation	R04136
Glycine, serine and threonine metabolism	R03759
Unknown	rxn06903
1,4-Dichlorobenzene degradation	R05408
1,4-Dichlorobenzene degradation	R05403
1,4-Dichlorobenzene degradation	R05402
Benzoate degradation via hydroxylation/ Fluorene degradation	R02602
Naphthalene and anthracene degradation	R00812
Unknown	alpha-D-Aldosyl-beta-D-fructoside:aldose1-beta-D-fructosyltransferase
Unknown	pd483
Caprolactam degradation	R06944
Atrazine degradation	R05560
Biphenyl degradation	R05241
Biphenyl degradation	R05261
Biphenyl degradation	R05359
2,4-Dichlorobenzoate degradation	R03932
	R04098
1,1,1-Trichloro-2,2-bis(4-chlorophenyl)ethane (DDT) degradation	R05363
3-Chloroacrylic acid degradation	R06973
gamma-Hexachlorocyclohexane degradation	R06856

1,4-Dichlorobenzene degradation	R05510
Toluene and xylene degradation	R05374
Toluene and xylene degradation	R05864
Fluorene degradation	R05410
Fluorene degradation	R05422
Fluorene degradation	R05412
Terpenoid biosynthesis	R05555
Fluorobenzoate degradation	R08104
Carotenoid biosynthesis - General	R08102
Fluorobenzoate degradation	R08103
gamma-Hexachlorocyclohexane degradation	R06831
gamma-Hexachlorocyclohexane degradation	R06832
gamma-Hexachlorocyclohexane degradation	R06848
1,4-Dichlorobenzene degradation	R07780
Benzoate degradation via hydroxylation	R02988
Benzoate degradation via hydroxylation	rxn00255
Pyruvate metabolism/ Pyruvate metabolism	rxn00146
Toluene and xylene degradation	R02550
Naphthalene and anthracene degradation	R00815
Urea cycle and metabolism of amino groups	R01154
Phenylalanine metabolism	R06785
Unknown	pd506
Carotenoid biosynthesis - General	R07656
Atrazine degradation	R05561
Vitamin B6 metabolism	R01909
Arginine and proline metabolism	R03293
Unknown	pd518
Unknown	pd519
Unknown	pd521
Unknown	pd522
Phenylalanine metabolism	R02252
Styrene degradation	R05450

Cyanoamino acid metabolism	R00486
Unknown	R09097
Benzoate degradation via CoA ligation	R01033
Biotin metabolism	R01074
Lipopolysaccharide biosynthesis	R05146
Unknown	R09692
Carotenoid biosynthesis - General	R04787
Carotenoid biosynthesis - General	R04798
Carotenoid biosynthesis - General	R04800
Carotenoid biosynthesis - General	R05341
Carotenoid biosynthesis - General	R03824
Peptidoglycan biosynthesis	R04573
Butanoate metabolism	R03534
Glycerophospholipid metabolism	R02563
Unknown	rxn33678
Unknown	N00008 Transport
Unknown	N00011 Transport
Unknown	4-Chlorobiphenyl transport
Unknown	N00017 tansport
Unknown	2,4-Dichlorobiphenyl tansport
Unknown	N00029 tansport
Unknown	N00033 tansport
Unknown	2,2,5,5-TCB tansport
Unknown	N00038 tansport
Unknown	N00044 tansport
Unknown	p-Chlorophenol tansport
Unknown	rxn12443
Unknown	o-Xylene tansport
Unknown	m-Xylene tansport
Unknown	p-Xylene tansport
Unknown	Tosylate tansport
Unknown	Atrazine tansport
Unknown	Cholesterol tansport
Unknown	PCP tansport
Unknown	1-Methylnaphthalene tansport
Unknown	rxn05488
Unknown	Vanillyl alcohol tansport
Unknown	rxn12571
Unknown	Dibenzothiophene tansport

Unknown	Fluorene tansport
Unknown	Dutch liquid tansport
Unknown	Trichlorobenzene A tansport
Unknown	2-Hydroxy-3-methylbenzalpyruvate tansport
Unknown	Protein Synthesis
Unknown	DNA Synthesis
Unknown	RNA Synthesis
Unknown	Phospholipid Synthesis
Unknown	Peptideglycan Synthesis
Unknown	Small molecules Synthesis
Unknown	Carbohydrates Synthesis
Unknown	Corynomycolic acid Synthesis
Urea cycle and metabolism of amino groups	R00566
Valine, leucine and isoleucine biosynthesis	R07399
Unknown	R09517
Tryptophan metabolism	R02665
Cysteine metabolism	R07633
Fatty acid biosynthesis	R04568
Galactose metabolism	R05549
Galactose metabolism	R01101
Galactose metabolism	R03635
Unknown	RXN0-6976.c
Benzoate degradation via hydroxylation	R04399
3-Chloroacrylic acid degradation	R05372
Unknown	Naphthyl-2-methyl-succinic acid transport
Unknown	Geranial transport
Unknown	EX_Geranial[e0]
Unknown	EX_Naphthyl-2-methyl-succinic-acid[e0]
Unknown	EX_2-Fluorobenzoate[e0]
Unknown	pd169
Unknown	EX_cpd01368[e0]
Unknown	EX_N00008[e0]
Unknown	EX_N00011[e0]
Unknown	EX_cpd03990[e0]
Unknown	EX_N00017[e0]
Unknown	EX_cpd09945[e0]
Unknown	EX_N00029[e0]
Unknown	EX_N00033[e0]
Unknown	EX_cpd09898[e0]
Unknown	EX_N00038[e0]
Unknown	EX_N00044[e0]
Unknown	EX_cpd01437[e0]

Unknown	EX_cpd01034[e0]
Unknown	EX_cpd04450[e0]
Unknown	EX_cpd04446[e0]
Unknown	EX_cpd04136[e0]
Unknown	EX_cpd04078[e0]
Unknown	EX_cpd03959[e0]
Unknown	EX_cpd00160[e0]
Unknown	EX_cpd01685[e0]
Unknown	EX_cpd09785[e0]
Unknown	EX_cpd00029[e0]
Unknown	EX_cpd03758[e0]
Unknown	EX_cpd00248[e0]
Unknown	EX_cpd21358[e0]
Unknown	EX_cpd04851[e0]
Unknown	EX_cpd04132[e0]
Unknown	EX_cpd03998[e0]
Unknown	EX_cpd01157[e0]
Unknown	EX_cpd09789[e0]
Unknown	EX_cpd08133[e0]
Unknown	Fluorobenzene transport
Unknown	3-Fluorobenzoate transport
Unknown	EX_cpd01579[e0]
D-Glutamine and D-glutamate metabolism	R00579
Unknown	rxn13291
Unknown	EX_cpd00507[e0]
Unknown	1,2-Anthracenediol transport
Unknown	EX_cpd14923[e0]
Unknown	EX_cpd02864[e0]
Unknown	cis-2-Chlorodienelactone transport
Unknown	EX_cpd02568[e0]
Unknown	EX_cpd00432[e0]
Unknown	EX_cpd04013[e0]
Unknown	EX_cpd04014[e0]
Unknown	Bromoxynil transport
Unknown	rxn11398
Unknown	trans-1,3-Dichloropropene transport
Unknown	cis-1,3-Dichloropropene transport
Unknown	Creatinine transport
Unknown	EX_cpd00585[e0]
Citrate cycle (TCA cycle)/ Alanine and aspartate metabolism/ Pyruvate metabolism	R00344
Valine, leucine and isoleucine degradation/ Propanoate metabolism	R01859

Unknown	EX_cpd01727[e0]
Unknown	rxn10979
Unknown	rxn05598
Unknown	EX_cpd00208[e0]
Unknown	(1-Hydroxycyclohexan-1-yl)acetyl-CoA transport
Unknown	EX_cpd02648[e0]
Unknown	EX_cpd27379[e0]
Unknown	(L-rhamnose transport
Unknown	EX_cpd04358[e0]
Unknown	(Styrene transport
Unknown	EX_cpd02304[e0]
Unknown	(2-Pyrone-4,6-dicarboxylate transport
Unknown	EX_cpd12238[e0]
Unknown	(Reduced flavoprotein transport
Unknown	EX_cpd25440[e0]
Unknown	(2,6-dihydroxybenzoate transport
Unknown	EX_cpd01133[e0]
Unknown	rxn18552
Unknown	EX_cpd00170[e0]
Unknown	Gluconolactone transport
Unknown	EX_cpd11847[e0]
Unknown	RCl transport
Unknown	EX_cpd14919[e0]
Unknown	Phenylboronic acid transport
Unknown	4-Hydroxybenzoylformate transport
Unknown	EX_cpd12847[e0]
Unknown	Apo-[carboxylase] transport
Unknown	EX_cpd11855[e0]
Unknown	RH transport
Unknown	EX_cpd15671[e0]
Unknown	1-isoheptadecanoyl-sn-glycerol 3-phosphate transport
Vitamin B6 metabolism	R01580
Fructose and mannose metabolism	R00868
Unknown	rxn13853
Unknown	EX_cpd03638[e0]
Unknown	(+)-Limonene transport
Urea cycle and metabolism of amino groups	R01987
Unknown	rxn09658
Unknown	FRUpts
Unknown	rxn09681
Unknown	rxn05170
Unknown	rxn05648
Unknown	EX_cpd00264[e0]
Unknown	rxn09660

Unknown	EX_cpd02259[e0]
Unknown	R09083
Unknown	rxn07979
Unknown	EX_cpd00179[e0]
Phenylalanine, tyrosine and tryptophan biosynthesis	R02413
Unknown	2-Fluorobenzoate transport
Unknown	3,4,5,6-Tetrachlorocyclohexene transport
Unknown	EX_cpd04308[e0]
Unknown	Biomass
Valine, leucine and isoleucine biosynthesis	R05070
Valine, leucine and isoleucine biosynthesis	R05068
Valine, leucine and isoleucine biosynthesis	R04673
Valine, leucine and isoleucine biosynthesis	R05069
Unknown	TRANS-RXN-2.cp
Unknown	rxn08469
Unknown	rxn08471
C5-Branched dibasic acid metabolism	R00932
Unknown	R00935
Purine metabolism	R01770
Unknown	R00801
Nicotinate and nicotinamide metabolism	R01273
Unknown	R01286
Glycine, serine and threonine metabolism/ Lysine biosynthesis	R01773
Methionine metabolism/ Sulfur metabolism	R01777
Caprolactam degradation	R03751
	R02065
	R04218
Unknown	rll_338
Unknown	rxn05211
Unknown	rxn05319
Sulfur metabolism	R00859
Unknown	R09490
Unknown	rxn34011
Unknown	DM_cpd15137[c0]
Unknown	DM_cpd00854[c0]
Unknown	DM_cpd03597[c0]
Unknown	DM_cpd08261[c0]
Unknown	DM_cpd00591[c0]

Unknown	DM_cpd01490[c0]
Unknown	DM_cpd00739[c0]
Unknown	DM_cpd15674[c0]
Unknown	DM_cpd03492[c0]
Unknown	DM_cpd12347[c0]
Unknown	DM_cpd12264[c0]
Unknown	DM_cpd12660[c0]
Unknown	DM_cpd04019[c0]
Unknown	DM_cpd12844[c0]
Unknown	DM_cpd04129[c0]
Unknown	DM_cpd04366[c0]
Unknown	DM_cpd21432[c0]
Unknown	DM_cpd22474[c0]
Unknown	DM_cpd01892[c0]
Unknown	DM_cpd01420[c0]
Unknown	DM_cpd00116[c0]
Unknown	DM_cpd15432[c0]
Unknown	DM_cpd15793[c0]
Unknown	DM_cpd15794[c0]
Unknown	DM_cpd02061[c0]
Unknown	sink_cpd12526[c0]
Unknown	sink_cpd02522[c0]
Unknown	sink_cpd06227[c0]
Unknown	sink_cpd00552[c0]
Unknown	sink_cpd02701[c0]
Unknown	sink_cpd01599[c0]
Unknown	sink_cpd15533[c0]
Unknown	sink_cpd00151[c0]
Unknown	sink_cpd00166[c0]
Unknown	sink_cpd15696[c0]
Unknown	sink_cpd03495[c0]
Unknown	sink_cpd00109[c0]
Unknown	sink_cpd11451[c0]
Unknown	R02687
Unknown	Triacylglycerol
Unknown	TAG metabolism
	R01998
	R00390
Biotin metabolism	R01078
Lipoic acid metabolism	R07767
Lipoic acid metabolism	R07768
Biphenyl degradation	R02606
Fatty acid biosynthesis	R04386
Fatty acid biosynthesis	R04385

Selenoamino acid metabolism	R04942
Unknown	R09365
Selenoamino acid metabolism	R04770
Unknown	R09371
Cysteine metabolism/ Sulfur metabolism	R00586
Glycerolipid metabolism/ Glycerophospholipid metabolism	R02241
Fatty acid biosynthesis	R04953
Fatty acid biosynthesis	R04533
	R03790
Unknown	EX_cpd01357[e0]
Unknown	Mandelate transport
Fatty acid metabolism/ Lysine degradation/ Tryptophan metabolism/ Benzoate degradation via CoA ligation/ Butanoate metabolism	R03026
Benzoate degradation via CoA ligation/ Butanoate metabolism	R01976
C21-Steroid hormone metabolism	R04851
Porphyrin and chlorophyll metabolism	R07411
Naphthalene and anthracene degradation	R07697
Naphthalene and anthracene degradation	R07696
	R07695
Fluorene degradation	R07804
	R07803
Unknown	R04254
Unknown	R01332
Unknown	pd143
Unknown	pd183
Unknown	rxn13669
Unknown	rxn13673
Fatty acid biosynthesis	R08163
Unknown	rxn08551
Unknown	rxn08088
Unknown	rxn10954
Fatty acid biosynthesis/ Biosynthesis of unsaturated fatty acids	R03370
Fatty acid biosynthesis	R08161

Unknown	pd256
Fatty acid metabolism	R02281
Fatty acid metabolism	R01347
Fatty acid biosynthesis/ Biosynthesis of unsaturated fatty acids	R07762
Thiamine metabolism	R07464
Glutathione metabolism	R03916
2,4-Dichlorobenzoate degradation	R01301
Fluorobenzoate degradation	R08122
Fluorobenzoate degradation	R08123
Urea cycle and metabolism of amino groups	R00559
Toluene and xylene degradation	R03559
Unknown	rxn10097
Unknown	rxn10098
Unknown	rxn10099
Unknown	rxn10100
Unknown	rxn10101
Unknown	rxn10102
Unknown	rxn10103
Unknown	rxn10109
Histidine metabolism/ One carbon pool by folate	R02287
One carbon pool by folate	R02302
Lysine biosynthesis	R04365
Valine, leucine and isoleucine biosynthesis	R01434
Vitamin B6 metabolism	R01710
Selenoamino acid metabolism	R04946
	R03964
	R07311
Unknown	R11161
Unknown	R11162
Unknown	R11164
Biphenyl degradation	R05239
Porphyrin and chlorophyll metabolism	R02864
Unknown	rxn13675
Unknown	rxn05557
Nitrogen metabolism	R05712
	R05318
Porphyrin and chlorophyll metabolism	R02480
Unknown	cytochrome-C3

Purine metabolism	R06975
Purine metabolism	R01127
Pyrimidine metabolism	R01080
Pentose phosphate pathway	R01051
Carbon fixation in photosynthetic organisms	R01829
Carbon fixation in photosynthetic organisms	R01845
	R00311
Porphyrin and chlorophyll metabolism	R00078
	R00092
Pantothenate and CoA biosynthesis	R04230
Pantothenate and CoA biosynthesis	R04231
Pantothenate and CoA biosynthesis	R03018
Pantothenate and CoA biosynthesis	R02472
Unknown	R05071
Lipoic acid metabolism	R07769
Lipoic acid metabolism	R07766
Unknown	apo-ACP transport
Unknown	EX_cpd12370[e0]
Pantothenate and CoA biosynthesis	R01625
Vitamin B6 metabolism	R05840
Unknown	rxn09997
Thiamine metabolism	R00619
Thiamine metabolism	R04448
Unknown	UNK2
	R00798
Unknown	R09492
Unknown	ATPSYN-RXN.ce
Nitrogen metabolism	R00790
Cysteine metabolism	R02433
	R08212
Glycine, serine and threonine metabolism	R08211
Vitamin B6 metabolism	R02494
Vitamin B6 metabolism	R01711
Vitamin B6 metabolism	R00174
Unknown	CYTBD
Unknown	ETC
Unknown	rxn05573
Starch and sucrose metabolism	R01557

Starch and sucrose metabolism	R02110
Unknown	R02737
Unknown	R00299
Glycolysis / Gluconeogenesis/ Pentose phosphate pathway	R02739
Nitrogen metabolism	R00792
Methionine metabolism	R01288
	R05717
Unknown	EX_cpd00048[e0]
Unknown	rxn05153
Unknown	rxn08688
Porphyrin and chlorophyll metabolism	R05227
Unknown	R11361
Naphthalene and anthracene degradation	R07692
Naphthalene and anthracene degradation	R07701
Glutathione metabolism	R04951
Glycine, serine and threonine metabolism/ Glycerophospholipid metabolism	R01800
Glycine, serine and threonine metabolism/ Glycerophospholipid metabolism	R02055
Glycerophospholipid metabolism	R00749
Unknown	R11261
Unknown	R11262
Unknown	R09994
Unknown	R09945
Unknown	rxn10951
Toluene and xylene degradation	R07664
Toluene and xylene degradation	R07665
Unknown	R01830
Riboflavin metabolism	R00161
Pentose and glucuronate interconversions	R05606
Riboflavin metabolism	R00160
Glutamate metabolism/ Aminosugars metabolism	R00768
Aminosugars metabolism	R02060
Glycerolipid metabolism/ Glycerophospholipid metabolism	R02239

	R01322
Starch and sucrose metabolism	R00028
	R00354
Pyruvate metabolism/ Carbon fixation in photosynthetic organisms/ Reductive carboxylate cycle (CO2 fixation)	R00345
Glycerolipid metabolism	R05333
Glycine, serine and threonine metabolism/ Glycerophospholipid metabolism	R02052
Glycerophospholipid metabolism	R01013
Glycerophospholipid metabolism	R02756
Inositol phosphate metabolism	R03332
Streptomycin biosynthesis/ Inositol phosphate metabolism/ Phosphatidylinositol signaling system	R01185
Glycerolipid metabolism	R02757
Glycerophospholipid metabolism/ Phosphatidylinositol signaling system	R01802
Unknown	R01351
Glycerolipid metabolism/ Glycerophospholipid metabolism/ Phosphatidylinositol signaling system	R02240
Glycine, serine and threonine metabolism/ Glycerophospholipid metabolism	R01021
Glycerophospholipid metabolism	R01890
Glycerophospholipid metabolism	R01321
Glycerophospholipid metabolism	R01310
Unknown	R02250
Pyruvate metabolism	R00319
Pyruvate metabolism	R00228
Geraniol degradation	R08091
Geraniol degradation	R08090
Geraniol degradation	R03493

Geraniol degradation	R03494
Geraniol degradation	R08098
Geraniol degradation	R08097
Geraniol degradation	R08095
Geraniol degradation	R08094
Geraniol degradation	R08093
Geraniol degradation	R08092
Glycolysis / Gluconeogenesis/ Purine metabolism/ Pyruvate metabolism/ Carbon fixation in photosynthetic organisms	R00200
Glycolysis / Gluconeogenesis/ Citrate cycle (TCA cycle)/ Pyruvate metabolism/ Butanoate metabolism/ Reductive carboxylate cycle (CO2 fixation)	R01196
Glycolysis / Gluconeogenesis/ Citrate cycle (TCA cycle)/ Pyruvate metabolism	R02569
Unknown	R07618
Glycolysis / Gluconeogenesis/ Galactose metabolism/ Starch and sucrose metabolism/ Streptomycin biosynthesis	R00959
Unknown	R04779
Glycolysis / Gluconeogenesis/ Carbon fixation in photosynthetic organisms	R01512
Glycolysis / Gluconeogenesis	R01518
Pyruvate metabolism	R00753
Glycolysis / Gluconeogenesis/ Pyruvate metabolism	R00710
Pyruvate metabolism/ Carbon fixation in photosynthetic organisms	R00216
Glutathione metabolism/ Reductive carboxylate cycle (CO2 fixation)	R00267
Pentose phosphate pathway/ Purine metabolism	R01049
Pentose phosphate pathway/ Carbon fixation in photosynthetic organisms	R01056
Pentose phosphate pathway/ Pentose and glucuronate interconversions/ Carbon	R01529

fixation in photosynthetic organisms	
Unknown	Gluconate trasport
Unknown	EX_cpd00222[e0]
Unknown	R01827

Table i: List of useful functions in the COBRA Toolbox for reconstruction purposes. * For details on syntax, please refer to the COBRA Toolbox and doi: 10.1038/nprot.2009.203

Action	**COBRA Toolbox command***	**Comment**
Add reaction	model = addReaction(model,'newRxn1','A + 2 B -> C')	Can be used to add reactions such as: A + 2 B --> C
Add demand reaction	[model,rxnNames] = addDemandReaction(model,metaboliteNameList)	Can be used to add reactions such as: A -->
Add sink reaction	[model] = AddSinkReactions(model,metabolites,bounds)	Can be used to add reactions such as: A <==>
Remove reaction	modelOut = removeRxns(model,rxnRemoveList)	
Write model in file	writeCbModelRecon(model,format,fileName)	Format of the file can be sbml, plain text, or xls
Open sbml model	model = readCbModel(fileName)	
Open reconstruction from xls	model = xls2model(RxnFileName,MetFileName)	
Change gene association	model = changeGeneAssociation(model,rxnName,grRule,geneNameList,systNameList)	

Change reaction bounds	model = changeRxnBounds(model,rxnNameList,value,boundType)	
Change objective function	model = changeObjective(model, rxnNameList, objectiveCoeff)	Multiple reactions can be set as objective function. They have to be satisfied in the specified ratio as define by objectiveCoeff (default = 1).
Print constraint reactions	PrintConstraints(model,MinInf, MaxInf)	The function identifies all reactions whose upper bounds (or lower bounds) are less (or greater) than the maximum value (MaxInf) (or minimum value – MinInf)
Load reconstruction into Matlab	model = xls2model(RxnFileName,MetFileName);	