# RECONSTRUCTION AND VALIDATION OF GENOME-SCALE METABOLIC MODEL FOR *GLUCONACETOBACTER DIAZOTROPHICUS*; A PLANT GROWTH PROMOTING BACTERIUM

by

Ülker Alkaya

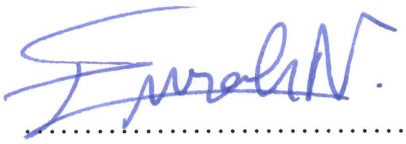
Submitted to Graduate School of Natural and Applied Sciences
in Partial Fulfillment of the Requirements
for the Degree of Master of Science in
Biotechnology

Yeditepe University

2019

RECONSTRUCTION AND VALIDATION OF GENOME-SCALE METABOLIC MODEL FOR *GLUCONACETOBACTER DIAZOTROPHICUS*; A PLANT GROWTH PROMOTING BACTERIUM

APPROVED BY:

Asisst. Prof. Dr. Emrah Nikerel ..... 
(Thesis Supervisor)
(Yeditepe University)

Assoc. Prof. Dr. Tunahan Çakır .....
(Gebze Technical University)

Assist. Prof. Dr. Bahar Soğutmaz Özdemir .....
(Yeditepe University)

DATE OF APPROVAL: ..../..../2019

*To my mum and dad ...*

# ACKNOWLEDGEMENTS

I would like to thank my supervisor Assit. Prof. Dr. Emrah Nikerel, for his endless guidance, patience and support he has given me for both my personal and school life during both my research in graduate school and bachelor's degree.

I would like to thank Assist. Prof. Dr. Bahar Soğutmaz Özdemir for giving me all of her support and guidance from the first day I have met her.

I am thankful to my friends Ebru Yılmaz, Kaya İşleyen, Filiz Erçelik, Gizem Banger, Gülşah Akçadağ, Kerem Kaya and Burcu Şirin for being there for me during my long thesis journey. They had to endure all of my talks on physiology of various animals such as pandas. I am grateful to Dr. Işıl Bilican for the guidance and mental support she has provided me when I most needed them.

I would like to thank my cats; Kedikun, Pıtırcık, Albus and Çipil. They always manage to find a way to comfort me and make me forget about my daily struggles whether by cuddles or play time.

Finally I would like to thank my family members Zarif Alkaya, Ali Alkaya, Çiğdem Alkaya, Hasibe Özkılıç, Zeynep Gencer and Jeroen Blikmans for their love, understanding, patience, support they continuously provide me to this day and forward. I would also like to thank my late father Temel Alkaya, I wish he could have seen what I have accomplished. If it was not for my parents love, trust, patience I could have never been where I am today. They have always had my back no matter what I have done or chosen for my life. There are not enough words to describe my gratitude for them. I know that I do not say it out loud much, but I love all of you with my whole heart.

# ABSTRACT


## RECONSTRUCTION AND VALIDATION OF GENOME-SCALE METABOLIC MODEL FOR *GLUCONACETOBACTER DIAZOTROPHICUS*; A PLANT GROWTH PROMOTING BACTERIUM

*Gluconacetobacter diazotrophicus* is an acid tolerant, gram negative and obligate aerobe bacterium whose genome is fully sequenced. It has important agricultural value such as nitrogen fixation, producing plant promoting enhancers, increasing solubility of zinc and phosphorus and increasing plant's resistance against certain bacteria or fungi. It also produces industrially and pharmaceutically valuable carboxyl acid gluconic acid which is used as antiseptic and chelating agents. Unlike other diazotrophs, *G. diazotrophicus* is an endophyte bacterium living within the plant without causing any harm. Its natural host is sugarcane: a monocot plant with high sugar content. There are studies about plant growth promoting and nitrogen fixation characteristics but the underlying metabolic pathways are not fully understood. Adopting a system biology approach; reconstructing the genome-scale metabolic model (GSMM) for the *Gluconacetobacter diazotrophicus* and applying Flux Balance Analysis (FBA), can not only help us to find the carbon flux of its metabolism but also to elucidate how it works and responds to the environmental factors. Consequently, the analysis of the GSMM can be also used for determining the target pathways for metabolic engineering considering utilitarian exploitation of the organism. In the light of the above, a genome-scale metabolic model for *G. diazotrophicus* consisting of 1754 reactions, 1629 metabolites and 848 genes was constructed.

# ÖZET


## BİTKİ BÜYÜMESİNİ TEŞVİK EDEN BAKTERİ (PGPB) OLAN *GLUCONACETOBACTER DIAZOTROPHICUS´*UN GENOM-ÖLÇEKLİ METABOLİK MODELİNİ OLUŞTURMA VE DOĞRULAMA

*Gluconacetobacter diazotrophicus*; gram negatif, aside toleranslı, zorunlu aerob ve tüm genomu dizilenmiş bir bakteridir. Azot fiksleme, bitki büyümesine yardımcı etkenler üretme, çinko ve fosfor çözünürlüğünü sağlama ve bitkilerin patojenlere karşı dirençlerini arttırma gibi önemli zirai özellikleri vardır. Ayrıca inşaat ve ilaç endüstrisinde antiseptik ve şelat oluşturucu olarak da kullanılan bir karboksil asit olan glukonik asit üretmektedir. *G. diazotrophicus* endofit bir bakteridir; yani konak bitkisinin içinde, ona zarar vermeden yaşamaktadır. Yüksek şeker içeren ve monokot bir bitki olan şeker kamışı, bu bakterinin doğal konağıdır. Azot bağlayıcı ve bitkilerin büyümesine yardımcı karakterlerini inceleyen birçok çalışma olmasına rağmen, bunlardan sorumlu metabolik yolaklar hakkında bilgi oldukça yetersiz veya başlangıç aşamasındadır. Sistem biyolojisi yaklaşımını kullanarak *G. diazotrophicus* bakterisinin genom ölçekli metabolik modelinin oluşturulması, modele akı denge analizinin uygulanması, hem bakteri metabolizmasındaki karbon akısının incelenmesine hem de bu metabolizmanın genel çalışma prensiplerinin ve çevresel faktörlerden nasıl etkilendiğinin anlaşılmasına yardım edecektir. Ayrıca böyle bir metabolik model, metabolizma mühendisliği için organizmanın endüstriyel açıdan faydalı olmasını sağlayacak hedef yolakların bulunması için de yararlı olacaktır. Bu bilgiler ışığında bu çalışmada, genom dizisinden yola çıkarak 1754 reaksiyon, 1629 metabolit ve 848 genden oluşan bir metabolik model oluşturulmuştur.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF SYMBOLS/ABBREVIATIONS

| | |
|---|---|
| 6PG | 6-Phosphogluconate |
| ADH | Alcohol dehydrogenase |
| ADP | Adenosine diphosphate |
| ALDH | Aldehyde dehydrogenase |
| ATP | Adenosine triphosphate |
| BLAST | Basic local alignment search tool program |
| BNF | Biological nitrogen fixation |
| bp | Base pair |
| $C_2H_2$ | Acetylene |
| $C_2H_4$ | Ethylene |
| Ca | Calcium |
| Da | Dalton |
| DCA | C4-dicarboxylates |
| DNA | Deoxyribonucleic acid |
| E.C. | Enzyme commission |
| ETC | Electron transport chain |
| fasta | Read data file format |
| FBA | Flux balance analysis |
| Fe | Iron |
| FVA | Flux variability analysis |
| g | Gas state |
| GA | Gluconic acid |
| GaDH | Gluconate dehydrogenase |
| GA1 | Giberellin 1 |
| GA3 | Giberellin 3 |
| GD | *Gluconacetobacter diazotrophicus* |
| GPR | Gene-protein-reaction |
| GSMM | Genome-scale metabolic model |

| | |
|---|---|
| IAA | Indole-3-acetic acid |
| ISR | Inducing systemic resistance |
| kb | Kilobase |
| kPa | KiloPascal |
| LsdA | Levansucrase |
| Mb | Megabase |
| Mg | Magnesium |
| Mo | Molybdenum |
| N | Nitrogen |
| NAD-GDH | NAD glucose dehydrogenase |
| O | Oxygen |
| P | Phosphorus |
| PGPB | Plant growth promoting bacteria |
| PGPR | Plant growth promoting rhizobacteria |
| PPP | Pentose phosphate pathway |
| PQQ | Pyrroloquinoline quinone |
| PQQ-GDH | Pyrroloquinoline quinone glucose dehydrogenase |
| q | Specific rate of production or consumption |
| RAST | Rapid annotation using subsystem technology |
| RNA | Ribonucleic acid |
| rxn | Reaction |
| S | Stoichiometric matrix |
| SBML | System biology markup language |
| TCA | Tricarboxylic acid |
| UDPG | Uridine diphosphate glucose |
| Z | Objective funtion |
| Zn | Zinc |

# 1. INTRODUCTION

Bacteria are prokaryotic organisms responsible for several metabolic functions which are beneficial to soil and plant health such as nutrient cycling, decomposition of organic matter, plant growth promotion and increased resistance to pathogens.

Nitrogen (N) is essential for life as amino acids, proteins and nucleotides contain N in their structure. Nitrogen is found abundantly in nature in the form of $N_2$(g), yet it is not readily available for most organisms to use. The amount of accessible nitrogen is one of the important crop yield-limiting factors. Thus in agriculture, to prevent production loss, application of fertilizers and rotational crop practices are applied. However the production of N-containing fertilizers and the excess usage of these have negative impacts on the environment. Biological Nitrogen Fixation (BNF) can be a strategical move to make agriculture more productive by providing the plants the needed nitrogen using environmentally benign methods.

Some of the N-fixing bacteria have symbiotic relationships with plants. While plants provide a habitat with essential nutrients to the bacteria, they receive ammonia which is a product of BNF in return. The most common example of the symbiotic relationship between a plant and nitrogen-fixing bacterium is the root nodules of legumes such as beans, lentils and peas, hosting various plant growth promoting rhizobacteria (i.e. *Rhizobia*). Fortunately, such relationship is not only limited to legumes. *Gluconacetobacter diazotrophicus*, an endophyte which was initially isolated from monocot plant sugarcane, not only has the ability to fix nitrogen but also has many other plant growth promoting properties such as production of plant hormones (i.e. auxins, giberellins), increasing the solubility of zinc and phosphate by acidification, showing antagonistic effects to plant pathogens. The molecular research which includes sequencing of the whole genome of bacterium *G. diazotrophicus* determined many candidate genes for understanding the relationship between the endophyte and its host sugarcane. This relationship can be a model for diazotroph-monocot interactions.

Although the plant growth promoting properties and nitrogen fixing characteristics of *Gluconacetobacter diazotrophicus* have been studied, the responsible pathways are yet to

be correctly defined. Taking a systems biology approach in the post genomic era may lead us better than classical methods for a rational understanding of the metabolism, increasing the yield and designing optimal media for growth. An important part that was missing for systems biology approach was the genome-scale metabolic model for *G. diazotrophicus*.

This study is mainly computational; reconstruction and analysis of a mathematical model. The experimental data were taken from the published literature. The main outcome of this study is the "genome-scale metabolic model" of *Gluconacetobacter diazotrophicus*, the plant growth promoting, N-fixing bacterium. The computational analysis was performed by Flux Balance Analysis (FBA) method, which allowed us to study the metabolic capabilities of the bacterium. This newly gathered knowledge will help us to adopt engineering strategies to design novel strains with increased capabilities to produce industrially, agriculturally or medically important products such as phytohormones or gluconic acid with minimal economical cost.

# 2. LITERATURE REVIEW

The world's population is over 7 billion people and expected to reach 8.6 billion in 2030 and 9.8 billion in 2050 [1]. One of the biggest challenges of today's world is to be able to grow enough food to sustain population with the limited and diminishing arable land. Most of the current agricultural strategies consist of using chemical fertilizers to achieve higher yields since most of the time the essential elements are the limiting factors for growth. This strategy has negative impacts on the environment and nutrient cycles. Developing biotechnological tools which can be used for sustainable agriculture is a significant social and industrial endeavor [2].

## 2.1. NUTRIENT CYCLES

According to the elemental analyzes performed on various living organisms from plants to animals, there are eleven essential elements common to all living matter. From these, six of them is found abundantly in biomass compared to others. These six elements are carbon (C), hydrogen (H), oxygen (O), nitrogen (N), phosphorus (P) and sulfur (S) and required for building blocks of all the biological macromolecules; carbohydrates, proteins, lipids and nucleic acids [3].

Earth can be assumed as a closed ecosystem where intake or loss of matter is negligible compared to the earth's mass, thus limiting the amount of chemical elements available which makes recycling of essential elements a fundamental process for life on earth [4]. Recycling of elements consists of both biotic and abiotic components of earth, hence are called biogeochemical cycles [5]. Two of the major biogeochemical cycles are carbon and nitrogen cycles.

Carbon is elemental base for life since it is found in all organic compounds making the carbon cycle one of the most important biogeochemical cycle. Carbon cycle consists of four processes; photosynthesis (carbon fixation), respiration, combustion and decomposition (Figure 2.1). During photosynthesis the carbon dioxide gas ($CO_2$) is taken by the primary producers (autothrophs) such as plants, and synthesized into carbohydrates via the enzyme

ribulose bisphosphate carboxylase/oxygenase (rubisco), storing the light energy from the sun into the chemical bonds. The plants incorporate the fixed carbon into their biomass and become carbon source for consumers. The plants or animals which die without being consumed by others, are decomposed by bacteria and fungi releasing the carbon to either to the air or soil. Some of the carbon reservoir is found in the soil as fossils and fossil fuels. The carbon also can return to the atmosphere as $CO_2$ gas via either respiration process of all living things, or combustion process where organic materials such as wood of fossil fuel burn with an oxidant [4].

Another major element for life is nitrogen as it is the building block for proteins and nucleic acids. Nitrogen is found abundantly as nitrogen gas ($N_2$) in air making up 79 per cent of the atmosphere. Nitrogen cannot be used in this form by most of the organism but diazotrophs [4, 6]. Similar to carbon cycle, consumers obtain this element from the plants or other animals. On the other hand, most plants obtain nitrogen from soil in the form of nitrite, nitrate or ammonium (Figure2.2) [7, 8].

Nitrogen is fixed and becomes available via two natural and one artificial pathways. Atmospheric disposition where lightning is responsible for the 15 per cent of nitrogen entering the ecosystem [6, 9]. The energy in the lightning breaks the bonds between the two nitrogen atoms and form nitrogen oxides. These nitrogen oxides dissolves in the rain and enter the soil to be available for plants to use. Biological nitrogen fixation is another natural pathway where bacteria convert the $N_2$ gas into ammonia ($NH_3$) and responsible for 60 per cent of the fixed nitrogen [4, 9]. The third, artificial nitrogen fixation method is the Haber-Bosch process where atmospheric nitrogen gas is converted to $NH_3$ using hydrogen gas and a metal catalyst [6, 7, 9]. Via this process, around 500 million tonnes of artificial fertilizers are produced each year [8]. Unfortunately, plants can only consume thirty to fifty percentage of the applied nitrogen based chemical fertilizers. The rest is wasted, either converted and released to the air in the form of nitrogen oxide, a greenhouse gas, by the soil bacteria or leached into the water supplies. These wasted part of the nitrogen fertilizers are danger to environment as they lead to both climate change and water pollution [8]. BNF is alternative nitrogen source for plants, which is cleaner and sustainable compared to fertilizers.

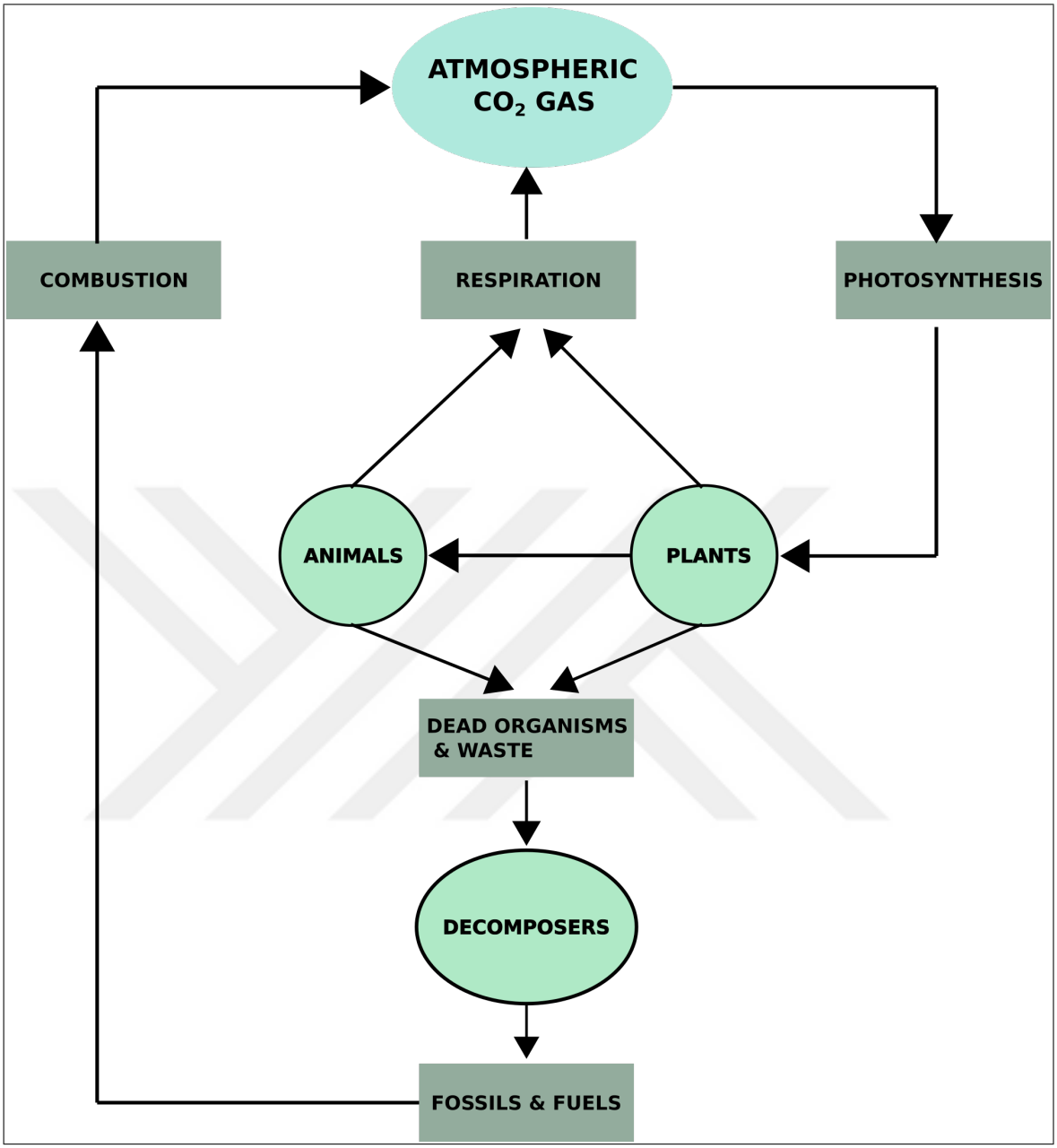



Figure 2.1. Simplified carbon cycle diagram.

Overall reactions for nitrogen fixing processes are presented below:

Lightning : $N_2 + O_2 \rightarrow 2NO, NO + \frac{3}{2}O_2 \rightarrow 2NO_2$

BNF : $N_2 + 8e + 8H^+ + 16ATP + 8H_2O \rightarrow 2NH_3 + 16ADP + H_2 + 16Pi$

Haber-Bosch : $N_2 + 3H_2 \rightarrow 2NH_3$

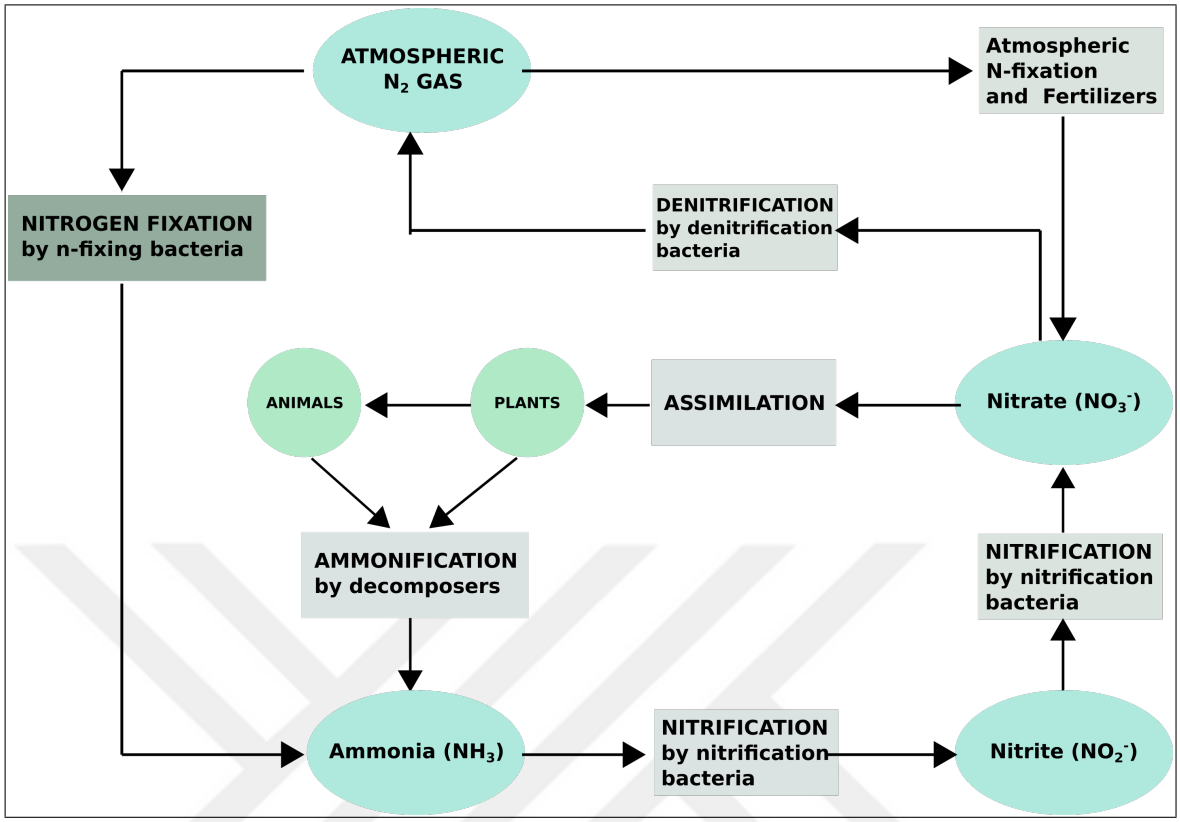



Figure 2.2. Simplified nitrogen cycle diagram.

Although the plants can use the ammonium ($NH_4^+$), the protonated form of ammonia directly, significant amount of the ammonium is used as energy source by the nitrification bacteria and convert it to nitrite ($NO_2^-$) and nitrate ($NO_3^-$) forms. Plants also use both nitrite and nitrate and assimilated into amino acids and nucleic acids. Consumers can assimilate the organic nitrogen by feeding on other animals or plants. Decomposers can also disintegrate the organic compounds containing nitrogen into ammonium via a process known as ammonification [4].

Possible Ammonification Reactions :

- $C$-$NH_2$(protein) $+ H_2O \rightarrow C - OH + NH_3, \quad NH_3 + H_2O \rightarrow NH_4^+ + OH^-$

- $CO(NH_2)_2$(urea) $+ H_2O \rightarrow HCO_3^- + NH_3, \quad NH_3 + H_2O \rightarrow NH_4^+ + OH^-$

Dinitrifcation bacteria, do not use the oxygen from the air but instead use the oxygen found in the nitrate compounds. This conversion enables the release of nitrogen gas back to the atmosphere [4].

This thesis focuses on the biological nitrogen fixation part of the nitrogen cycle.

## 2.2. BIOLOGICAL NITROGEN FIXATION (BNF)

Biological Nitrogen Fixation is defined as the capability of prokaryotes to reduce dinitrogen ($N_2$) to ammonia ($NH_3$) which helps sustaining the soil's nitrogen supply [10, 11, 8]. BNF is catalyzed by the enzyme called "Nitrogenase" (EC 1.18.6.1). Nitrogenase enzyme which is a Molybdenum-dependent system is consist of two proteins. One of the protein contains the ATP binding sites, called dinitrogenase reductase or Fe protein. The other protein which possess the substrate binding sites is named dinitrogenase or MoFe protein [8]. Both of these proteins are very sensitive to oxygen and reactive oxygen species (ROS) and can be deactivated irreversibly by both, yet the nitrogen fixation is a very energy demanding process requiring aerobic respiration [8, 12].

$N_2$ needs six electrons to be reduced to $NH_3$ by the nitrogenase complex, moreover two extra electrons are needed as at least one mole of $H_2$ is produced during biological nitrogen fixation. The electrons needed by this process is provided by the reduction of ferrodoxin. The energy needed to break the $N_2$ bonds are provided by the dinitrogenase reductase part of nitrogenase complex, which hydrolysis 2 ATP per electron hence 16 ATP in total. The oxygen is consumed while oxidizing the reduced ferrodoxin [9].

Although there is no ideal method exists for quantification the nitrogenase enzyme activity, there are two common methods; $^{15}N_2$ enrichment and acetylene ($C_2H_2$) - ethylene ($C_2H_4$) assay[13]. As a direct measurement method of $N_2$ fixation during the $^{15}N_2$ enrichment, the samples are analyzed for $^{15}N$ using a mass spectrometer after a incubation period with $^{15}N_2$. On the other hand, for the acetylene ($C_2H_2$) - ethylene ($C_2H_4$) assay which is an indirect measurement method for $N_2$ fixation , first the samples are incubated with the acetylene gas. Then using gas chromatography, the produced ethylene gas is measured. The biochemical basis if this assay method is, acetylene acts as an exogenous electron acceptor instead of N2

and reduced to ethylene in the reaction catalyzed by the nitrogenase enzyme [13].

## 2.3. PLANT GROWTH PROMOTING BACTERIA (PGPB)

Apart from the biogeochemical cycles, large number of bacteria also have important metabolic roles in plant growth and are called “Plant Growth Promoting Bacteria (PGPB)” [14, 15]. PGPB include both free-living bacteria and endophytes. The free living bacteria generally found in the soil around plant roots which is termed as ”rhizosphere” by Lorenz Hiltner in 1904. On the other hand, endophytes are able to enter and live within the plant tissue without any pathogenic activity [14, 16, 17].

PGPB can enhance plant growth either using direct or indirect metabolic mechanisms. Direct metabolic mechanisms include solubilizing minerals such as phosphate, producing phytohormones (cytokinins, giberellins, auxins and ethylene) or increasing plant’s own production of these hormones and most importantly fixing nitrogen which is very significant plant nutrient. On the other hand indirect mechanisms are producing antibiotics and lytic enzymes against pathogenic microorganisms, helping other beneficial bacteria to improve their effects on plants, production of siderophores which are chaletors for iron and other heavy metals, and inducing systemic resistance (ISR) for plants to help their immune system against pathogens [14, 17, 15]

## 2.4. GLUCONACETOBACTER DIAZOTROPHICUS

In 1988, Cavalcante and Dobereiner discovered a new species of nitrogen fixing bacterium on roots of sugarcane, which is resistant to acidic conditions. Initially they suggested that the bacterium species was called ”*Saccharobacter nitrocaptans*” [18]. Later Gillis and colleagues named the bacterium as ”*Acetobacter diazotrophicus*” according to its phenotypic, genomic and chemo-taxonomic properties which are the similarities between the biocomposition of the organisms [19]. In 1997, after 16S ribosomal RNA analysis, it was reclassified as *Gluconacetobacter diazotrophicus* [11, 20].

*Gluconacetobacter diazotrophicus* is an acid tolerant, gram negative and obligate aerobe

bacteria. It has a rod shape and rounded ends with the size of 0.7-0.9 $\mu$m by 1-2 $\mu$m. The bacteria cells can grow on high sucrose content and low acidity. The optimum sugar concentration is 10 per cent and pH is 5.5 for growth although its recorded that they can live at pH as low as 3. It is an endophyte, located at the internal tissues of its host without causing any harmful effect on the plant (Figure 2.3) [18, 19, 21, 22, 23].

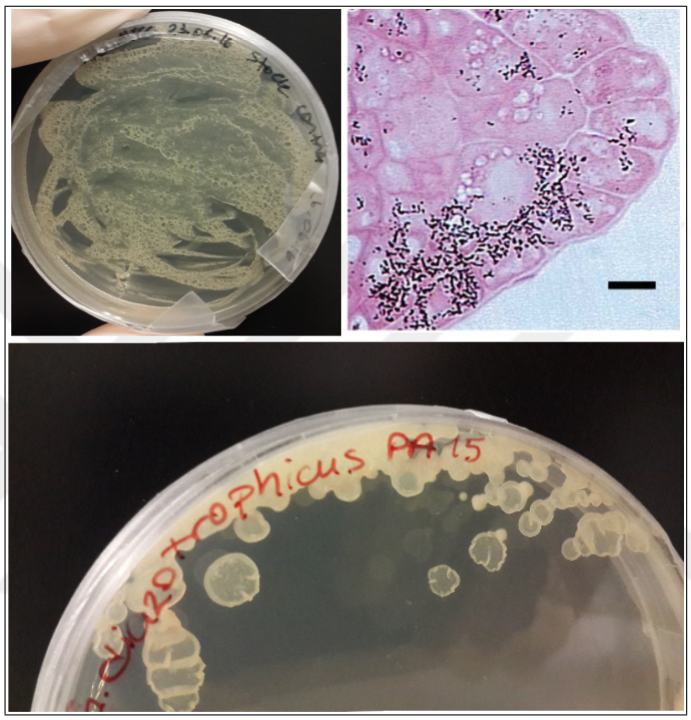


Figure 2.3. *Gluconacetobacter diazotrophicus* colonies on agar and inside root tissue of *Arabidopsis thaliana* [24].

The whole genome of *G. diazotrophicus* strain Pal5 (ATCC 49037) has been sequenced by two different institutes. First group to sequence and publish the genome is RioGene FAPRJ, Brazil. Second group is from US DOE JGI, USA. The second genome sequence has not been published as a paper but the sequence can be found on NCBI with the assembly name ASM2132v1. There is a large difference between the two sequenced genomes [25]. RioGene determined *G. diazotrophicus* to have a single circular chromosome (3.9 Mb) and two plasmids (38 kb and 16 kb). 3938 coding sequences has been annotated [26]. JGI sequence has one plasmid (27 kb) and one chromosome (3.9 Mb). Giongo et al (2010), reconstructed an optical map of the *G. diazotrophicus* and compared two DNA sequence against it. Optical map is a physical restriction map of a genome which gives structural

information about a genome thus can be used as validation tool in sequencing projects.[25] Giongo and colleagues have found that RioGene sequence has several rearrangement positions and regions which did not align with the optical map. On the other hand they detected only small number of inversions on JGI DNA sequence. JGI DNA sequence aligned more accurately with the optical map representation [25].

The most important characteristic of *Gluconacetobacter diazotrophicus* is the ability known as "Biological Nitrogen Fixation (BNF)". *G. diazotrophicus* can provide its host with sufficient amount of fixed nitrogen without nodule formation. For example, sugarcane obtains nearly 150 kg N $ha^{-1}$ per year from *G. diazotrophicus* [18]. The bacteria has also the ability to secrete part of the fixed nitrogen in suspension cultures hence can be used in nitrogen production as a biological work horse [27]. *G. diazotrophicus* also has advantage over other N-fixing organisms (i.e. *Azosprillium vinelendii*, *Rhizobium etli*). It does not have the nitrate reductase enzyme which reduces nitrate to nitrite. It can also fix nitrogen even in the presence of soil nitrogen since both nitrite and excess nitrogen in soil can have negative effect on the nitrogenase enzyme. Although the oxygen shows inhibition on nitrogenase, the *G. diazotrophicus* nitrogenase enzyme can fix nitrogen under a wide range of atmospheric oxygen pressure (5 to 60 kPa $pO_2$) [28, 29]. *Gluconacetobacter diazotrophicus* can fix nitrogen on semi-solid media but when it is grown on liquid culture, a started dose of 1mM $(NH_4)_2SO_4$ is needed [18].

*Gluconacetobacter diazotrophicus* has important characteristics for agricultural use other than Biological Nitrogen Fixation (BNF) such as producing plant growth promoting agents (i.e. phytohormones; Indole-3-acetic acid (IAA) and gibberellins (GA1 and GA3)). Furthermore, *G. diazotrophicus* increases the solubility of the zinc and phosphorus which also increase the plant growth [11, 26]. Additional studies showed that it produces bacteriocin against *Xanthomonas albilineans*; a sugarcane pathogen and has antifungal properties against *Fusarium sp.* and *Helminthosporium carbonum* [26].

The *G. diazotrophicus* host species is not only limited to sugarcane but also includes banana, coffee, tea, wetland rice, pineapple and coconut plants. There are several studies where artificial inoculates were successfully applied in high sugar content plants such as tomato, wheat, corn and *Arabidopsis thaliana*; a model plant for dicots. Both natural and artificial

inoculation studies showed that *G. diazotrophicus* enables an increase in plant growth and crop yields [11, 30, 31].

*G. diazotrophicus* has application areas other than agriculture such as the production of gluconic acid (GA) and the enzyme levansucrase.

### 2.4.1. PGPB Mechanisms

As a non-nodular endophytic bacterium, *Gluconacetobacter diazotrophicus* is a interesting organism which has several direct and indirect PGPB mechanisms other than biological nitrogen fixation (Figure 2.4).

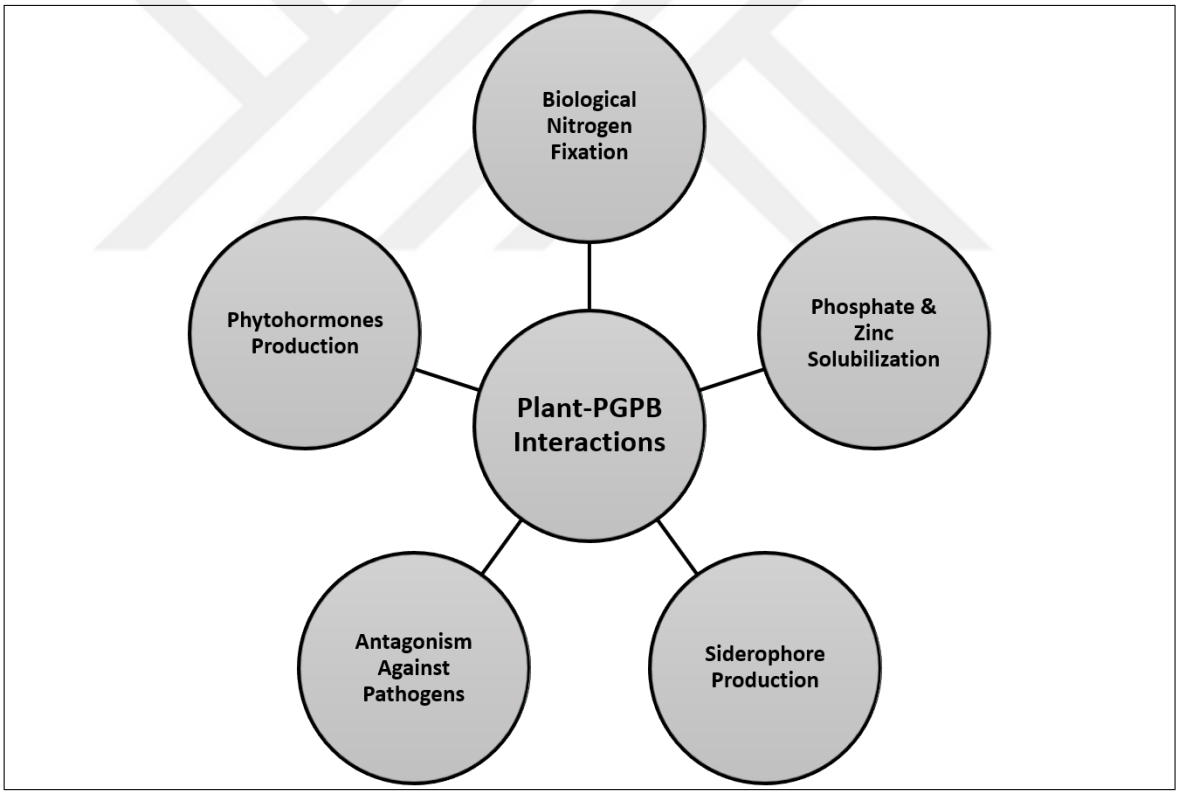



Figure 2.4. Plant - PGPB interaction.

#### *2.4.1.1. Phytohormone Production*

Saravanan et al. showed that both wild type and nif-mutants (lacking the nif gene responsible for the nitrogenase enzyme) could still promote plant growth in nitrogen rich soils which was

an indication of phytohormone production [32].

In one study, 18 different strains of *G. diazotrophicus* were isolated from sugarcane grown in Mexico. Then, all 18 strains were grown in defined cultre medium and analysed for indoleacetic acid (IAA), a type of auxin. It was determined that all strains were able to produce IAA into the medium from 0.14 to 2.42 $\mu$g per liter [33]. In another study, *G. diazotrophicus* was found to produce different amounts of IAA according to the sucrose content of the medium. It was measured that, when the bacteria were grown in a ten percent sucrose medium the IAA production was the highest rate of 32 ng per ml [34]. Cytochrome C biogenesis genes are involved in the production of IAA, and based on the presence of the intermediate metabolites, the bacterium uses the IpyA pathway for IAA production [35, 36]. *G. diazotrophicus* also has been studied for giberellins synthesis and found to produce GA1 and GA3 types in chemically defined medium [34, 32]. GA1 secretion and GA3 secretion were 1.6 ng per ml and 11.9 ng per ml respectively [11].

Polyamines such as putrescine, spermidine and spermine are phytohormone-like molecules found in most of the organisms. They were demonstrated to enhance plant tolerance to several abiotic stresses such as temperature, salt, water and pollutions. During stress times, bacterial polyamine production can help the plant to cope with the environmental stress [37].

#### *2.4.1.2. Phosphate and Zinc Solubility*

Besides nitrogen, phosphorus (P) is one the significant nutrient limiting factor in agriculture. Although it has been found in both organic and inorganic forms in the soil, not all forms are readily available to plants and bacteria. Even though soluble phosphorus compounds are readily given to the agricultural crops, they become insoluble quickly, become inaccessible for plants and pollutes the soil [38]. *G. diazotrophicus* has been shown to increase P solubilization via organic acid production and decreasing the pH of the environment. For example, strawberry plants inoculated with GD strain, had increased P content and reversed P deficiency [38, 39].

*G. diazotrophicus* can also increase the solubility of the trace element Zinc (Zn) via the acidification of the environment [40]. It was also demonstrated via radiotracers that maize

plants inoculated with *G. diazotrophicus* had higher uptake of Zinc than plants that were not inoculated [41].

#### *2.4.1.3. Siderophore Production*

Iron (Fe) is a micronutrient required for many biochemical reactions as cofactors. Under aerobic conditions, the iron ion is found in $Fe^{+3}$ form which bacteria and plants cannot use readily. Bacteria are known to produce small (<1000 Da) molecules, which can bind $Fe^{+3}$ ions. These molecules are known as siderophores. While binding to siderophores, Fe ions are in the form of $Fe^{+2}$ which can pass through the cell walls. Once they are inside the cell, the Fe ions and siderophores are separated. Plants are known to use these chelating molecules produced by the bacteria. These molecules not only help to prevent iron deficiency and enhance plant growth but also help against infection by pathogenic bacteria since the pathogens are deprived from the iron ions they require to grow [15, 42]. *G. diazotrophicus* has been investigated to produce such small molecules to increase the plants uptake of Fe ions [32, 43].

#### *2.4.1.4. Antagonism Against Plant Pathogens*

*Gluconacetobacter diazotrophicus* has been characterized to have antagonistic properties against plant pathogens. The bacterium not only produces elicitor molecules which increases the defense response of the plant but also secretes bacteriocins that kills other microorganisms [44]. Plant pathogens which *G. diazotrophicus* has been identified to show antagonisms are; *Xanthomonas albilineans* a sugar cane pathogen, *Colletotrichum falcatum* (red dot fungal pathogen of sugarcane) and *Gossypium hirsutum* (cotton) pathogen root-know nematode (*Meloidogyne incofnita*) [32, 45, 46].

Gluconic acid which is key molecule of the carbon metabolism of the *Gluconacetobacter diazotrophicus* has antimicrobial activity and also responsible for the antimicrobial characteristics of *Gluconacetobacter diazotrophicus* PAl5 strain. Although gram-positive and negative bacteria show sensitivity to gluconic acid, eukaryotes are resistant [47].

### 2.4.2. Metabolism

#### *2.4.2.1. Carbon Metabolism*

Glycolysis is an anaerobic process where one molecule of glucose is reduced to two molecules of pyruvate. The whole pathway generates 4 molecules of ATP, yet it consumes 2 ATP. Thus the net ATP production is 2 ATP molecules [9]. Gluconeogenesis seems to be the opposite pathway of the glycolysis (although it is not) where glucose is produced from the non-carbohydrate precursors [9].

*G. diazotrophicus* has been demonstrated to have gluconeogenesis pathway since key and related enzymes for the gluconeogenesis pathway has been detected in the crude extracts. Presence of this pathway can explain why *G. diazotrophicus* can grow on glycerol and other C2, C3 substrates [48].

*Gluconacetobacter diazotrophicus* has also pentose phosphate pathway and an uncommon aldose oxidation pathway which reduces glucose to gluconic acid in exterior of the periplasm via ”pyrroloquinoline quinone (PQQ) linked glucose dehydrogenase (PQQ-GDH) enzyme (Figure 2.5)[32, 49, 50]. *G. diazotrophicus* lacks both the Entner-Doudoroff and Embden-Meyerhof pathway as the bacterium does not posses the key enzymes 6-phosphogluconate (6PG) dehydratase, KDPG aldolase and phosphofructokinase [48, 49].

*Gluconacetobacter diazotrophicus* posseses full TCA cycle along a respiratory chain-linked malate dehydrogenase which explains the ability to fully oxidize ethanol to acetate [48]. *G. diazotrophicus* can oxidize ethanol to acetic acid in the periplasm without excreting the acetaldehyde into the medium and two membrane-bound enzymes catalyze this process. These enzymes are the alcohol dehydrogenase (ADH) and the aldehyde dehydrogenase (ALDH) [52].

The bacterium grows well on sucrose, glucose, fructose, gluconate, mannitol, sorbitol and glycerol but not on C4-dicarboxylates (DCA) such as succinate, fumarate and malate. The enzymes needed to metabolize DCA exists in the crude extract of bacterium thus this indicate that *G. diazotrophicus* has no transport mechanism for DCA [23, 48, 53]. Although the optimum growth for the bacteria has been reported on ten percent sucrose, there is

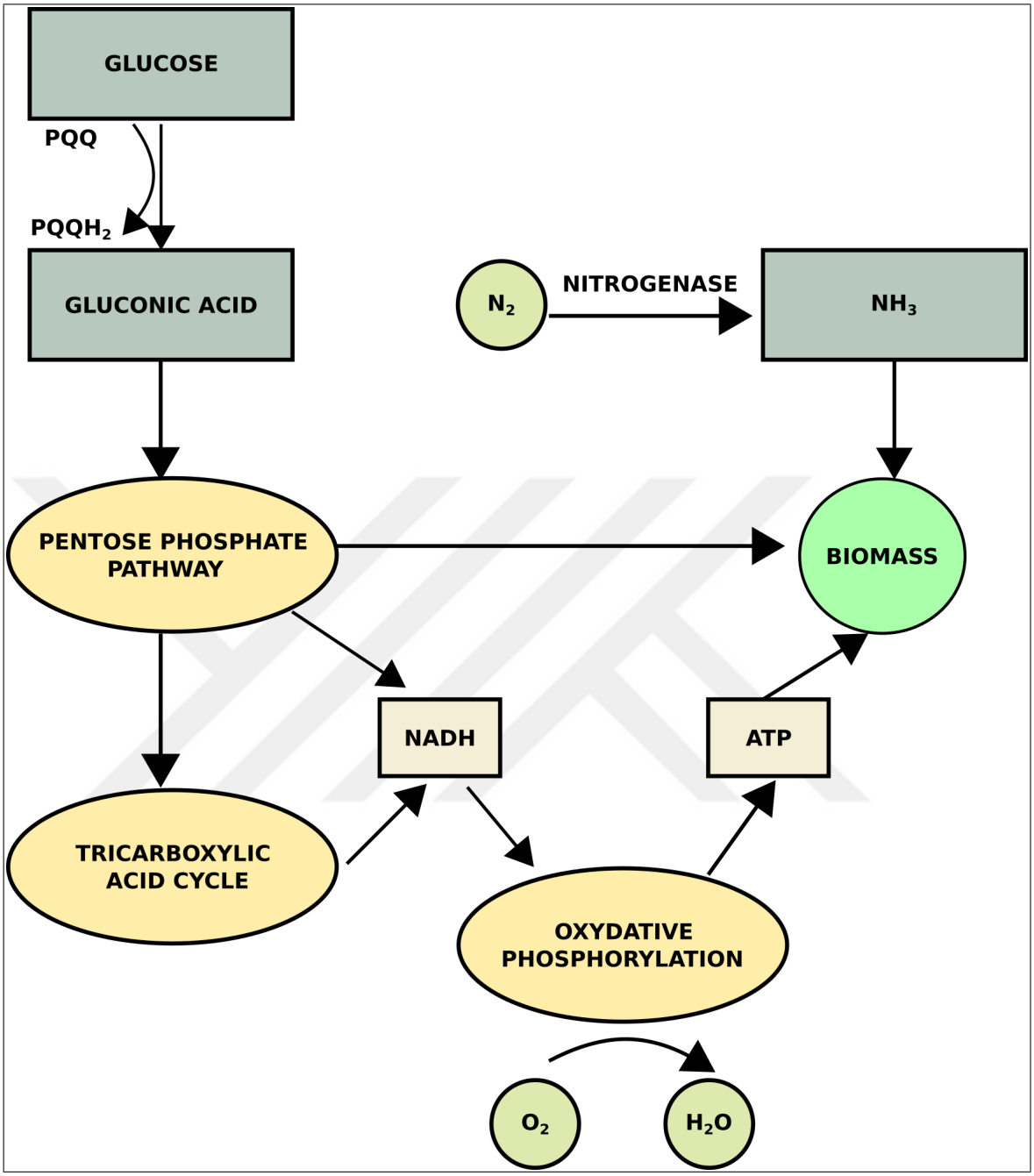



Figure 2.5. Simplified carbon metabolism (adapted from [51])

also no transport mechanism for sucrose. It is metabolized externally via the secretion of levansucrase [11, 48].

*Gluconacetobacter diazotrophicus* can directly oxidize glucose, gluconate and ketogluconate through respiratory chain-linked enzymes. Glucose, as with the other aldoses, is reduced to gluconic acid externally via the enzyme PQQ-GDH which

has a significant role in glucose metabolism [23, 49, 54]. Gluconate either fully metabolized through the hexose monophosphate pathway or reduced further to ketogluconic acid via the periplasmic enzyme gluconate dehydrogenase (GaDH) externally [48, 54, 55].

Oxidation via the PQQ-GDH enzyme is the main metabolic pathway for aldoses especially under carbon limiting and BNF conditions. During $N_2$ fixation, the metabolism requires abundant amounts of energy so that the synthesis of PQQ-GDH is accelerated [56]. But under glucose excess condition, an alternative pathway is simultaneously used. Excess glucose is taken intercellulary and oxidized via NAD-GDH enzyme [54, 55].

Inactivity of PQQ-GDH in mild alkaline pH may be the reason why *Gluconacetobacter diazotrophicus* does not grow under such conditions [57].

#### *2.4.2.2. Nitrogen Metabolism*

*G. diazotrophicus* has BNF capacity of 400 - 417 nmoles of $C_2H_4$ $hr^{-}1$ $mg^{-}1$ cell protein as nitrogenase potential is measured by the acetylene ($C_2H_2$) - ethylene ($C_2H_4$) assay [11].

BNF bacteria has developed different strategies to cope with the energy demand and protect their enzyme from the oxygen gas. For example, rhizobacteria assemble nodule type of structure around the roots of host plants that is an oxygen free environment while plants provide the required energy [58, 59] In the case of endophytic bacterium *Gluconacetobacter diazotrophicus*, the direct oxidation of glucose might be an evolutionary strategy to protect their enzyme from the $O_2$ gas [23, 48]. *G. diazotrophicus* has one of the highest reported respiratory capability for an aerobic bacteria [12]. The oxygen requirement for energy production during the BNF conditions, utilizes the oxygen from the surroundings, hence devising an environment which is friendly for the activity of nitrogenase enzyme [12]. Stephan et al. also demonstrated that nitrogenase had high oxygen tolerance in the presence of glucose and gluconate [23]. It has been also shown that *Gluconacetobacter diazotrophicus* can alter its redox metabolism during nitrogen fixing and decreases the ROS generation via upregulating the antioxidant gene levels [59]. According to another study, *Gluconacetobacter diazotrophicus* colonies can fix nitrogen under a range of different oxygen pressure (5 to 60 kPa) and sustain the activity of their nitrogenase enzyme in

response to changes in the oxygen pressure [29].

Nitrate reductase is not detectable in crude extracts of *Gluconacetobacter diazotrophicus* and the fixed nitrogen is assimilated via the glutamine synthetase enzyme [48, 32].

### 2.4.3. Industrial Products

#### *2.4.3.1. Gluconic Acid*

One of the important products that *G. diazotrophicus* produces is gluconic acid ($C_6H_{12}O_7$), or gluconate. The bacteria secrete an enzyme called Pyrroloquinoline Quinone-Linked Glucose Dehydrogenase EC 1.1.5.2 which oxidizes glucose into gluconic acid (GA) [11]. Gluconic acid is a carboxyl acid which is used as for its chelating and antiseptic properties. Chelating agents are molecules which can bind to metal ions with more than one bonds. As a chelating agent, gluconic acid can bind to Ca, Mg, Fe [47].

Gluconic acid is used in numerous industrial sectors such as pharmaceuticals, textile, food and even construction. Microbial production of GA is an alternative methods to chemical, catalytic or electrochemical processes [60]. GA is a biodegradable, water soluble, weak, harmless, odorless, non-corosive and non-toxic acid [61].

In food industry, GA and its derivatives are used as acidity regulators (E574-E580) since in both Europe and USA, it is assessed to be safe to use. They give a bitterness in taste of foodstuff. They also have properties of raising, hardening and sequestering the food they are added to [60, 61].

In pharmaceutical industry, gluconate derivatives are used as mineral supplements in treatments of hypocelcaemia, anaemia and hypomagnesaemia [60].

Amids its other uses, GA has chelating ability of alkaline solutions hence it has been used in both cleaning and construction industries. As an additive to cement, it increases the resistance, hardness and stability to some of the extreme weather conditions [60, 62] .

#### *2.4.3.2. Levansucrase*

The natural host of *Gluconacetobacter diazotrophicus* is sugarcane which is rich in plant sucrose. Since *G. diazotrophicus* does not have a transport for sucrose, it secretes a constitutively expressed enzyme levansucrase (LsdA, EC 2.4.1.10) to break down sucrose [63]. The GA levansucrase has two different enzyme activities. First, it hydrolysis the sucrose into glucose and fructose so that GA cells can transport these molecules into the cytoplasm. Second, it has a role in the polymerization of fructosyls and levan production [48, 64]. The LsdA enzyme has an adequate kestose-production which are used as natural low-calorie sweeteners reported to boost human and animal gastrointestinal health as they promote the growth of healthy gut bacteria [65]. The activity of this extracellular enzyme has been measured as 1600 nmol glucose formed /min. mg protein by Alvarez et al. [48].

There are also several other medical and pharmaceutical uses of levan, a product of levansucrase enzyme. It has biodegradable and film-forming abilities. It can be used as filler, radioprotector, coating and binder properties in drug market. It can also be a plasma substitute. [66]

## 2.5. SYSTEMS BIOLOGY

System biology is an interdisciplinary science where the objective is to understand the systemic properties of cells or organisms and eventually predict the behavior of the phenotype when environmental conditions fluctuate [67, 68]. System biology uses and combines the powers of both experimental (omics technologies) and computational methods (mathematical models) for these predictions [69, 68]. Mathematical models are required because biological systems are too complex due to the large number of components which interact in a nonlinear fashion [67].

One of the goals in life sciences is to elucidate the relationship between the genotype and the phenotype. Before the emergence of systems biology, the reductionist view (bottom up approach) was popular in the fields of molecular biology, genetics and biochemistry where the functions and properties of individual parts such as a molecule, a gene or a reaction were studied and identified [70]. Although this approach was useful to gain information about the

components, it was inadequate when it comes to the emergent properties which are the collective features of the biological systems (i.e. biofilm production by microorganism), and the interactions of components within themselves or with their environment. On the other hand, system biology adopts a top-down approach and takes on a holistic view when it comes to complex biological problems [67, 69, 71].

One of the most studied and modeled subject in systems biology is metabolic networks. Although still incomplete, a remarkable number of reactions and how they interact stoichiometrically in the metabolism are known. Information on the enzymes catalyzing these reactions and genes encoding these enzymes are also available. Using the particular information Genome-scale Metabolic Models (GSMM) are constructed to analyze the metabolic network of target organisms [67, 71, 72].

### 2.5.1. Fluxome

The fluxome is defined as the complete collection of metabolic fluxes; where a flux is the rate of a given reaction typically at a steady state. The fluxome is the dynamic snap shot of the phenotype's response to the environment based on the metabolome of the cell [69].

GSMMs aim to capture the fluxome of the modeled organism with respect to the environmental conditions that is being simulated.

### 2.5.2. Genome-scale Metabolic Modeling

Metabolism is the total collection of enzyme-catalyzed reactions in an organism required for maintaining the living state. Enzymes encoded by the genome of the organism are responsible for catalyzing these metabolic reactions. Metabolic models are mechanistic models based on the relationship between the genome and the metabolism [70, 73]. Genome-scale metabolic models are the set of reactions needed for growth and maintaining life, reconstructed via the information on the genome and bibliome which is the experimental literature on the target organism [70, 74, 75, 76, 77].

While GSMMs lack the enzyme kinetics, they can still used for analyzing the metabolic

capabilities of organisms via the well established stoichiometry of the reactions using constraint based modeling approaches [72, 74]. Constraint based approaches are used to limit the solution space of the models. The mass conservation principle, energy balance and thermodynamic constraints are universal constraints applied to all systems whether its biological or not. In order to further limit the solution space, organism's characteristics and behaviors can be applied as constraints on the models [78, 79].

When a mass-balance equation is written for each intracellular metabolite ($x$);

$$\frac{dx}{dt} = \sum v_{\text{produced}} - \sum v_{\text{consumed}} \tag{2.1}$$

where $\frac{dx}{dt}$ is the time derivative of the given metabolite concentration, $v_{produced}$ is the rate of formation and $v_{consumed}$ is the rate consumption for that particular metabolite. Using a matrix notation, the equation 2.1 can be written as;

$$\frac{d\boldsymbol{x}}{dt} = \mathbf{S} \cdot \mathbf{v} \tag{2.2}$$

In equation 2.2, **S** denotes the stoichiometric matrix; where the stoichiometric coefficients for each metabolite is given per corresponding reaction and **v** stands for the rate of reaction matrix. If the metabolite is consumed during the reaction, the stoichiometric coefficient in the **S** matrix is negative, while if the metabolite is the product of that reaction the coefficient has positive value.

Genome-scale models assume that the cells are in a pseudo stead-state because the turn over time of metabolites are much smaller compared to the time-dependent behavior of the intracellular of metabolite concentrations [80]. In this special case, the change in concentration over time is equal to zero which yields the algebraic relation:

$$\mathbf{S} \cdot \mathbf{v} = 0 \tag{2.3}$$

Reconstruction of a genome-scale model has been methodized using a 96 steps protocol by Thiele and Palsson, it can be summarized in five steps [77, 81]:

1. Draft model reconstruction (usually done via automated tools such as Model SEED)
2. Gap-filling
3. Manual curation of the draft model into a detailed model
4. Transformation of the model into a mathematical one
5. Simulation of the model

The reconstruction processes is an iterative process in which if there is inconsistencies between the model and the experimental data, the model should be re-evaluated and the same steps should be taken from the start [74, 81].

The first step in generating a genome-scale metabolic model is functional annotation of the genome to identify the metabolic reactions and determine their stoichiometry. The reversibility of the reactions depending on the thermodynamic constraints (either calculated using the Gibbs Free Energy or taken from the literature) can be added to the model as well. The latter steps include assigning the reactions' proper localization, creating a lumped biomass reaction that will reflect the biomass composition of the cell, and evaluation of the energy requirements of the cell [74, 81].

The gene-reaction relationships and cofactor requirements can be mined from databases such as KEGG, BioCyc. When uncertain about whether NADH or NADPH is used as a cofactor for a reaction, it is better to add both reactions to the model [74]

Definition of compartments in a model and assigning reactions to their proper localization is also another important step of the metabolic reconstruction. Transport reactions between the compartments should be added to the model. The identical metabolites of reactions belonging to different compartments should be named accordingly. For example, one way of naming metabolite X is, as X_c if it belongs to reaction in cytosol while if it belongs to a reaction localized on periplasm it should be named as X_p [74, 81].

The biomass reaction is a lump, artificial reaction formed from the precursor metabolites (i.e. amino acids, fatty acids, sugars etc) of macro molecules (i.e proteins, DNA, lipids

etc.) constituting the biomass of the cell [74, 81]. The proportions of the biomass elements cannot be derived from the annotated genome but should be either measured experimentally or literature data of a similar organism can be used [82].

Another artifical reaction that should be added to the model is the non-growth associated maintenance (NGAM) reaction which is an irreversible reaction. Non-growth associated maintenance energy is the energy needed for survival of the cell without any growth. NGAM accounts all the ATP expenses, that can not be taken stoichiometricaly into account, e.g. via Futile cycles, signaling processes etc. In this reaction ATP is converted into ADP and orthophosphate [73, 74].

After mining the set of reactions from both the annotated genome and the bibliome, the stoichiometric matrix **S**(m x n) is constructed where m is the number of metabolites and n is the number of reactions. The rows on the **S** matrix denote the metabolites while the columns denote reactions [74, 83, 84]. For every matrix element, the $\mathbf{S}ij$, shows the stoichiometry of that metabolite in the given reaction. If the $\mathbf{S}ij$ is zero, it means that metabolite is not present in the reaction [83].

Once the model is converted into $\boldsymbol{S}$ matrix, using a constrain based optimization such as Flux Balance Analysis (FBA), the model can be analyzed. All the possible flux values for each reactions can be calculated [74].

Apart from the reactions and their stoichiometric information, genome-scale metabolic models can also include available enzyme commission numbers (E.C.) and gene-protein-reaction (GPR) associations. GPR association is a description of which genes are responsible for encoding the enzyme associated with the particular reaction [81, 85]. Although genome-scale metabolic models lack the regulatory information about the cells, GPR association allows integration of -omics data (i.e. transcriptomics) to increase the predictive power of the constraint based modeling approaches [70, 83].

Although all biological models including the genome-scale metabolic models, are simpler interpretations of reality evaluating only part of the real life constraints, their predictions are very useful[78]. Genome-scale metabolic models and constraint based model analysis have been used first and foremost to elucidate metabolism and to predict its theoretical

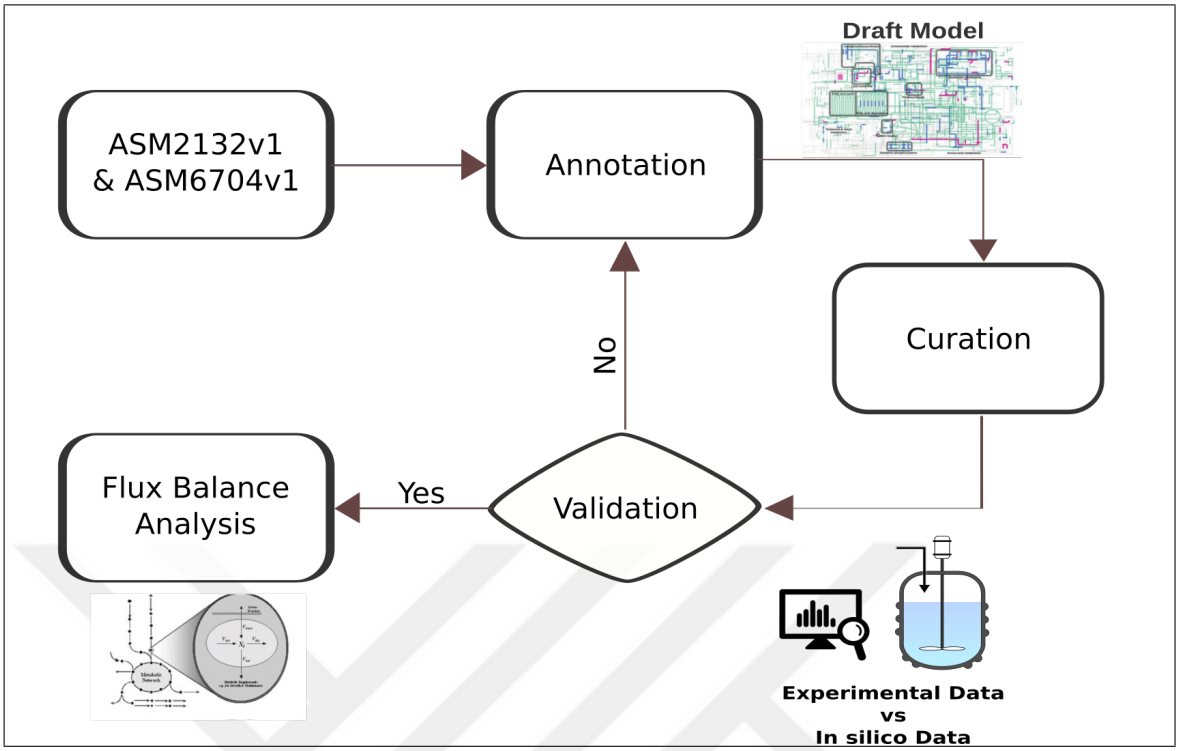



Figure 2.6. Workflow for reconstruction of a GSMM

capabilities [70, 74, 86, 87]. In vitro experiments are both time and money consuming, thus *in silico* hypothesis testing using the GSMMs are optimal to reduce time and cost while gathering as much data as possible on external and internal effects on the metabolism [76, 86]. Even the incorrect predictions of the metabolic models can help to find missing metabolic information or right the wrong ones. The inconsistencies between the model prediction and the experiments can be used to construct targeted experiments to fix such inaccurate metabolic knowledge [70].

GSMMs are also used in finding metabolic engineering targets [70, 87, 88]. For example using the gene-reaction relationships in the model, new gene-knock out strategies can be developed for acquiring new or improved phenotypes [74, 86]. Optimum media design for targeted cellular bioprocess is also possible using GSMMs [74].

GSMMs are optimal tools to calculate yield of a particular product single or more substrates. Yield is defined as the maximum amount of product can be produced per unit substrate[89].

### 2.5.3. Flux Balance Analysis (FBA)

Flux Balance Analysis is a constraint based optimization method which is used to analyze genome-scale metabolic models [74, 84]. FBA can evaluate reaction fluxes without any regulatory information and kinetic enzyme parameters assuming a steady-state [69, 79, 85].

To apply FBA, a metabolic network reconstruction is required in the form of a stoichiometric matrix, **S**(m x n). The **S** matrix is multiplied by a column vector, **v**, which is the unknown flux vector to establish the linear equation system. Since the assumption that the system is in a steady state exits, the product of this matrix has to be equal to zero [83, 84, 90].

$$\mathbf{S} \cdot \mathbf{v} = 0 \tag{2.4}$$

In FBA, the system is generally under-determined, as number of reactions is much larger than the number of metabolites in the metabolic models [69]. A bounded solution space can be found as FBA limits it via the constraints set on the equation system. These constraints are thermodynamics constraints, mass balance, upper/lower limits of fluxes and organism specific ones; such as constraining oxygen uptake while being an anaerobe. Constraining the solution space gives us a more accurate representation of the target organism [69, 74, 79, 83, 90]. The single solution calculated via the FBA is not necessarily unique, there can be numerous flux distributions where the objective function can reach optimum value [67, 74].

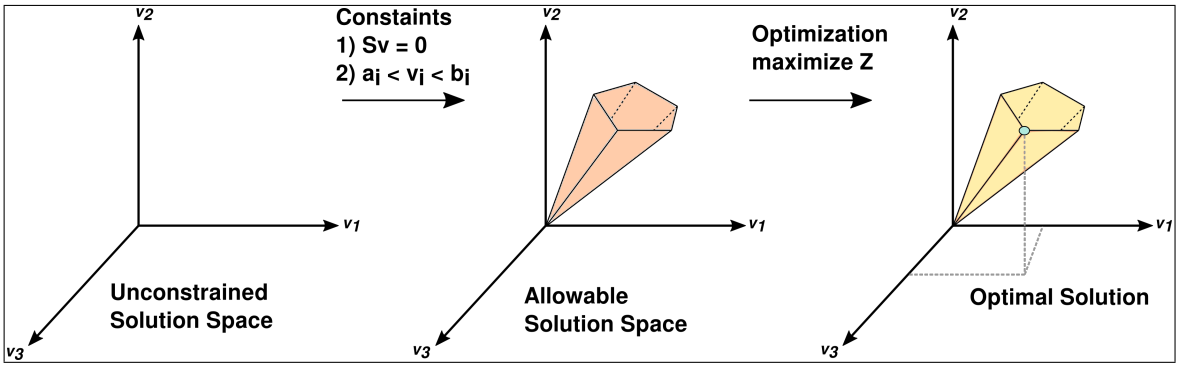



Figure 2.7. The basis of constrain-based analysis (adapted from [84])

FBA uses linear optimization to solve the system to find the representation of the metabolic

phenotype under certain conditions [77, 86]. Linear optimization demands an objective function, which will be used to optimize the equation system [69, 74]. Although there are no single objective function to describe the phenotype under all conditions, some of them can be used to make biological meaningful predictions (i.e. maximum growth or ATP production) [67]. In genome-scale metabolic models, objective function is generally biomass reaction representing the growth rate since it is assumed that organisms usually maximize their growth [67, 73]. Different type of objective functions can be used as well. For example; maximization of ATP synthesis or a product synthesis can all be used as objective functions [74, 83].

Since FBA maximizes or minimizes the objective function, Z, the dot product of vectors **c** and **v**, where **c** represents the weight of each reaction contributing to the objective. Since only one reaction is needed to be maximized or minimized, all except one value is zero on the vector **c** [83, 84]

$$Z = \mathbf{c} \cdot \mathbf{v} \quad \text{where} \quad \mathbf{LB}_i \leq \mathbf{v}_i \leq \mathbf{UB}_i \tag{2.5}$$

# 3. THE OBJECTIVE OF THE STUDY

*Gluconacetobacter diazotrophicus* is plant growth promoting bacterium with interesting characteristics. Albeit the plant promoting properties and nitrogen fixing aspect of the *Gluconacetobacter diazotrophicus* haven been the subjects of various studies, the bacterium has not been studied at the systems biology perspective. In order to approach the metabolism of *Gluconacetobacter diazotrophicus* from a systems biology perspective, a key tool is missing; the genome-scale metabolic model which can also serve as a blue print of biochemical reactions for the bacterium. The model would be an effective instrument to answer questions about the organism such as;

- How the bacterium reacts to changes in its environment?
- How does it interact with the plants or other microorganisms?
- How can the genome and pathways of the organism can be exploited for utilitarian desires for targeted metabolic engineering? (for example, improving production of phytohormones, nitrogen fixation or gluconic acid)

The aim of this study is to reconstruct the genome-scale metabolic model of the plant growth promoting bacterium *Gluconacetobacter diazotrophicus*. The GSMM was validated and analyzed using Flux Balance Analysis. Validation was performed by the comparison of simulation results and the experimental data from literature. The mathematical model was used to investigate the metabolic capabilities of the bacterium; determining the carbon and nitrogen flow and flux distributions under constraint conditions.

# 4. MATERIALS AND METHODS

## 4.1. MODEL RECONSTRUCTION

The genome sequences of the *Gluconacetobacter diazotrophicus* have been downloaded from the NBCI with the assembly names ASM2132v1 and ASM6704v1. The ASM2132v1 has been submitted by the US DOE Joint Genome Institute (JGI-PGF) and has one chromosome and one plasmid. On the other hand, the ASM6704v1 has been submitted and published by FAPERJ - RioGene and has one chromosome and two plasmids.

Table 4.1. Sequence information of *Gluconacetobacter diazotrophicus*

| **Assembly Name** | **ASM2132v1** | **ASM6704v1** |
|---|---|---|
| Total Sequence Length (bp) | 3,914,947 | 3,999,591 |
| Total Ungapped Length (bp) | 3,914,947 | 3,999,591 |
| Number of Chromosomes | 1 | 1 |
| Number of Plasmids | 1 | 2 |
| G-C content | 66.3 | 66.3 |

Table 4.2. RAST options chosen during genome annotation step

| **UPLOAD OPTIONS** | |
|---|---|
| Scheme | Classic RAST |
| Gene Caller | RAST |
| FIGfam Version | Release 70 |
| Fix errors | Yes |
| Fix Frameshift | Yes |
| Build Metabolic Model | Yes |
| Backfill gaps | Yes |
| Turn on debug | Yes |
| Verbose level | 5 |
| Disable replication | Yes |

## 4.2. MANUAL CURATION

Using the pathway maps on the Model SEED database, E.C. numbers and subsystem information were added for the available reactions.

Biomass formula was derived from the draft model's biomass reaction.

### 4.2.1. Elemental Balances

In order to find if there were any gaps or mistakes in elemental balances in the model, using the COBRA toolbox's checkBalance function was used.

### 4.2.2. Gap Filling

Online gap-filling algorithm of Model Seed was run. The algorithm uses a single, non-redundant set of reactions from KEGG database and 13 published GSMMs to add reactions to ensure that the model is able to simulate growth[91].

Manual curations were done according to the annotations on KEGG pathways, reference metabolic models such as *Gluconacetobacter xylinus* & *Escherichia coli* and literature.

### 4.2.3. FBA Validation

Fermentation and enzyme assay data are taken from the literature for validation of the model. When needed, specific consumption or formations rates are calculated from the given data and/or unit conversions were performed. The model was simulated with either biomass as the objective function, or any other reaction with an experimental value. The computational results compared with the experimental data.

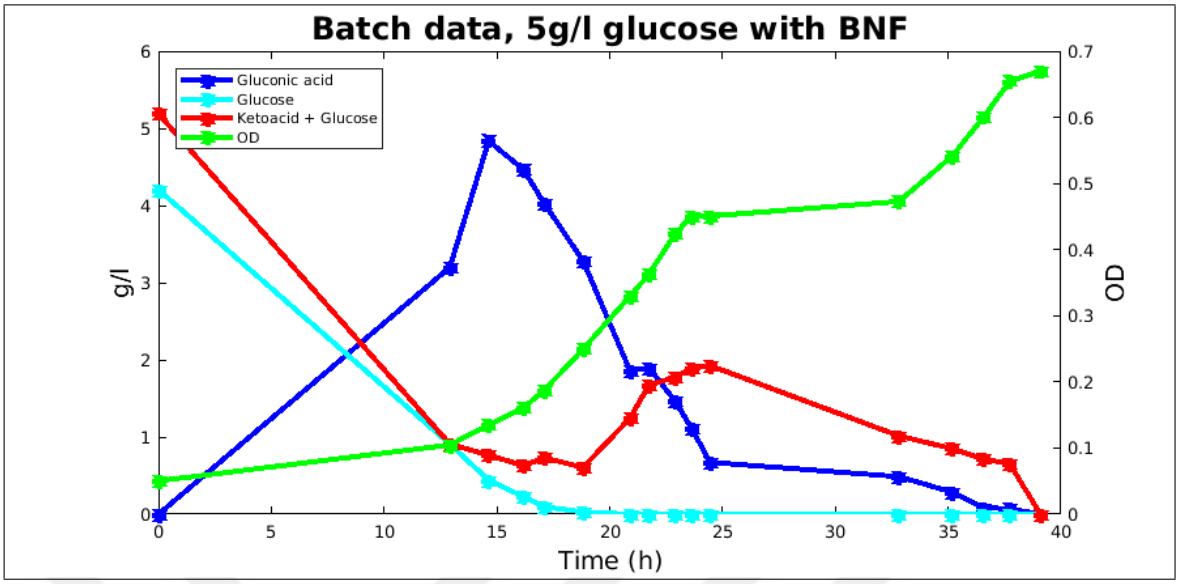



Figure 4.1. *Gluconacetobacter diazotrophicus* batch fermentation data with 5 g/l glucose and nitrogen fixation [51]

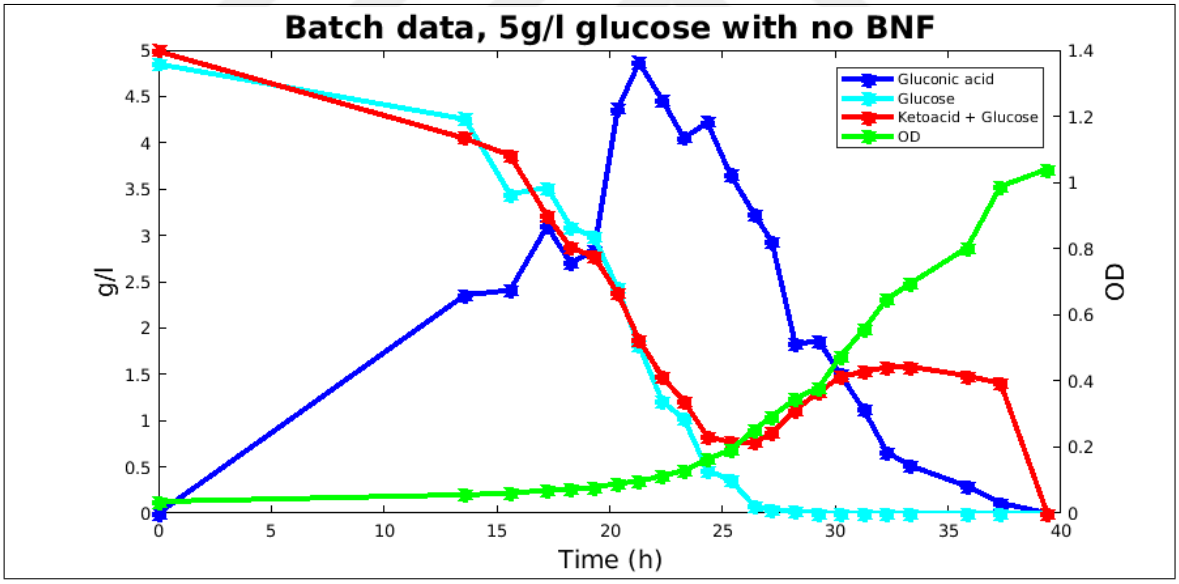



Figure 4.2. *Gluconacetobacter diazotrophicus* batch fermentation data with 5 g/l glucose and without nitrogen fixation [51]

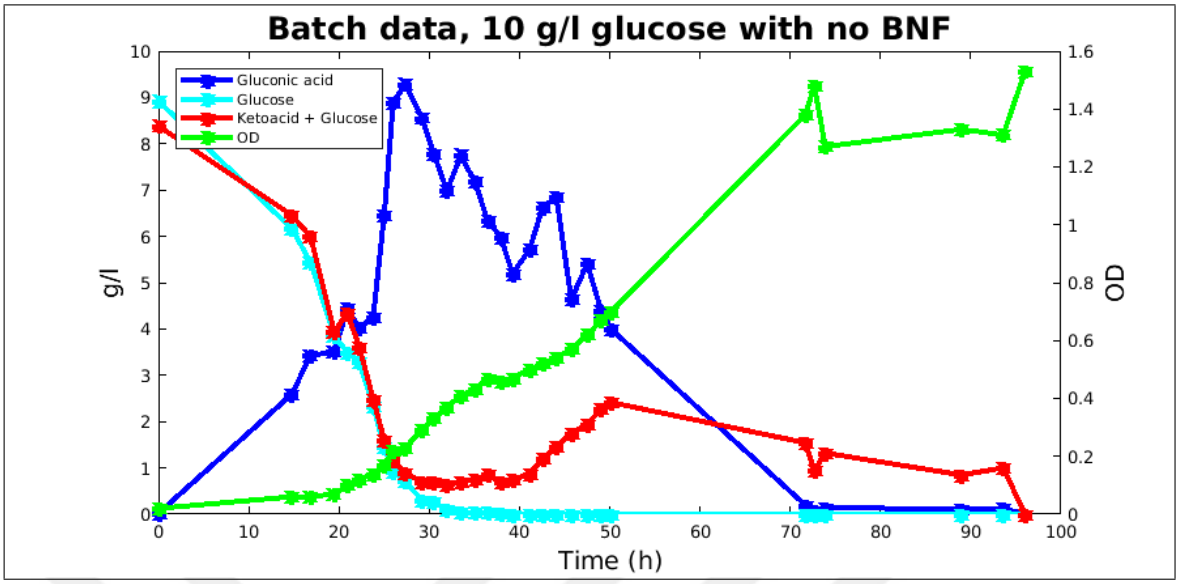



Figure 4.3. *Gluconacetobacter diazotrophicus* batch fermentation data with 10 g/l glucose and without nitrogen fixation [51]

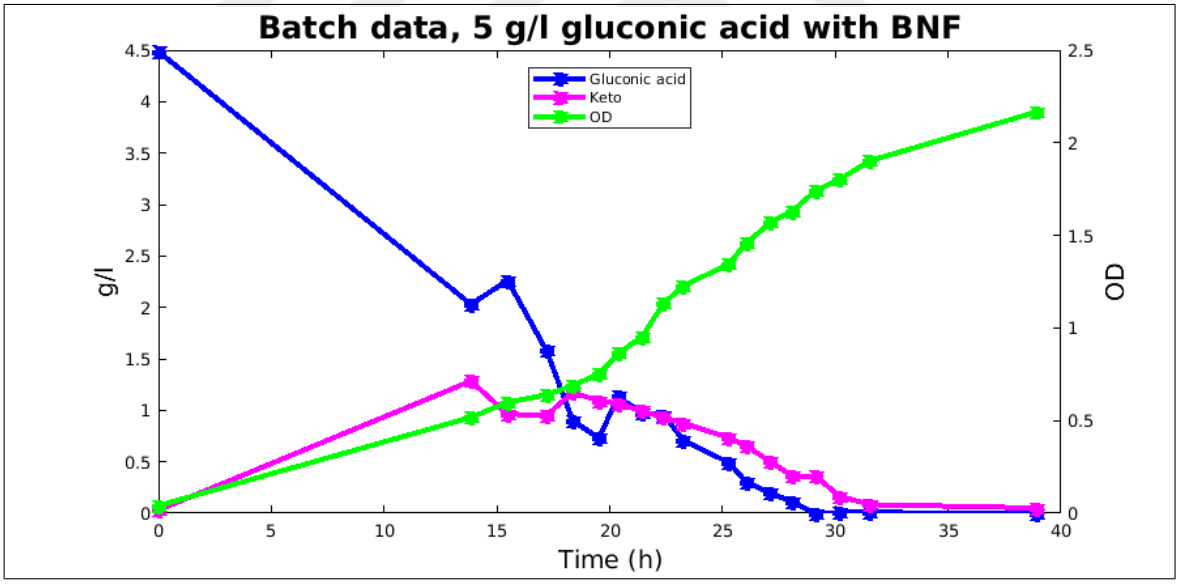



Figure 4.4. *Gluconacetobacter diazotrophicus* batch fermentation data with gluconic acid and nitrogen fixation [51]

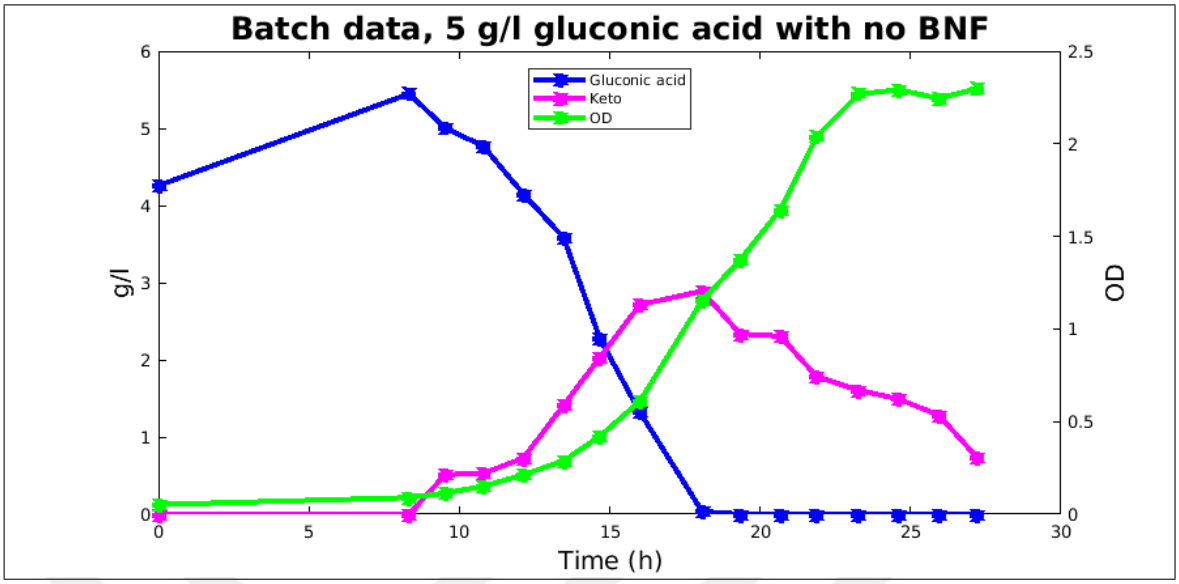



Figure 4.5. *Gluconacetobacter diazotrophicus* batch fermentation data with gluconic acid and without nitrogen fixation [51]

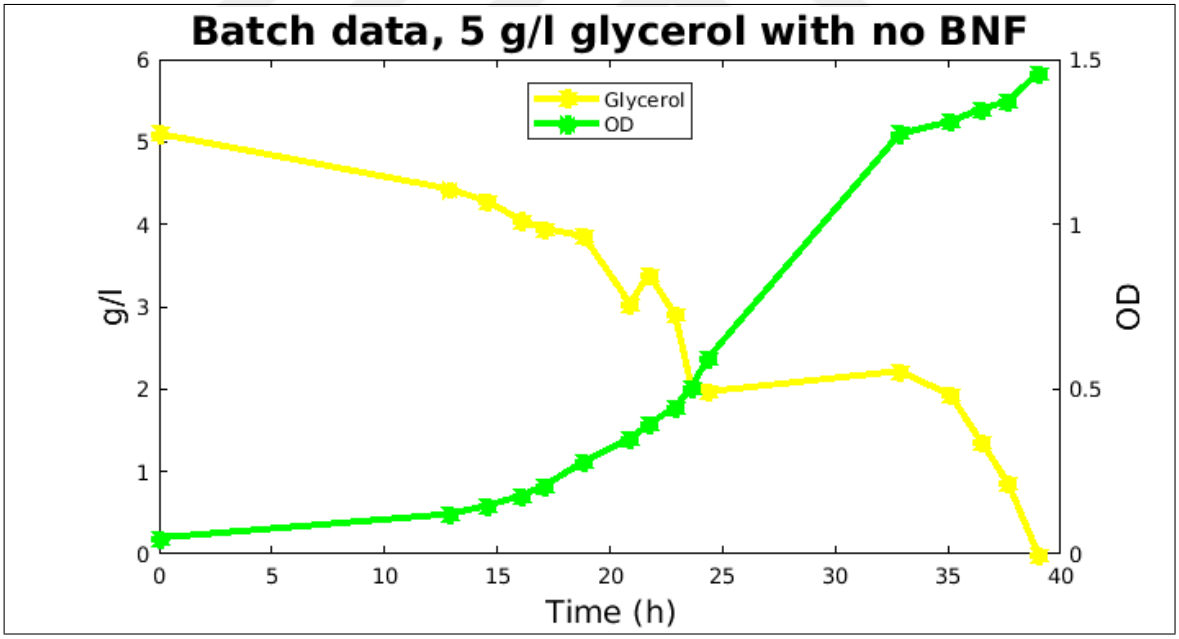



Figure 4.6. *Gluconacetobacter diazotrophicus* batch fermentation data with glycerol and without nitrogen fixation [51]

## 4.3. DATABASES AND ONLINE TOOLS USED

### 4.3.1. RAST server (Rapid Annotation using Subsystem Technology)

The SEED was built in 2004 to be a platform which could be used to store precise genome annotations and to develop new annotations. The SEED also provides information about compilations of functionally related protein families and their FIGfams (protein families) [91]. In 2008, the RAST server was established to annotate sequenced genomes of microorganisms [92, 93]. After a new genome in fasta format uploaded to the RAST server and the genes are called, their annotations are generated by comparing them to the FIGfam collection in the SEED database [91].

RAST annotations are also used a base for automated metabolic model reconstruction in the modelSEED and KBASE [92, 93].

### 4.3.2. Model SEED

The Model SEED is a online resource that automates the first 72 steps of 96 stepped gnome-scale metabolic model reconstruction protocol written by Thiele et al.. It is built on the basis of the SEED annotation database. Model SEED is able to reconstruct a draft metabolic model with gene-protein-reaction (GPR) associations, a biomass reaction, and information on the thermodynamic reversibility of the reactions [94].

### 4.3.3. KEGG (Kyoto Encyclopedia of Genes and Genomes)

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a reference knowledge base where genomic information is linked to higher order functional information [95]. There are also reference pathway maps for core metabolic pathways. A user can check organism specific pathway maps with KEGG annotation [95].

### 4.3.4. BRENDA (BRaunschweig ENzyme DAtabase)

BRENDA (BRaunschweig ENzyme DAtabase) is a public database for enzymes and enzyme-ligand associated information. The relevant information is directly extracted from the primary literature. The enzyme information is based on the EC classification system belonging to the International Union of Biochemistry and Molecular Biology (IUBMB). There are over 4 million manually annotated experimental on 84.000 enzyme found in the BRENDA. The database also integrated with other type of external databases for additional information such as KEGG, UniProt etc [96].

### 4.3.5. EcoCyc

EcoCyc is a free to use database which provides experimental data for the bacterial model *Escherichia coli K-12*. Information on gene products, functions and regulations, metabolic pathways, enzymes and related cofactors can be extracted from the database which is updated regularly. There are also SmartTable tools which enable users to browse the content easier. *E. coli* metabolic models can be directly simulated and modified directly from the EcoCyc [97].

### 4.3.6. National Center for Biotechnology Information (NCBI)

National Center for Biotechnology Information (NCBI) was established as a division of the National Library of Medicine (NLM) at the National Institutes of Health (NIH), USA in 1988. NCBI not only hosts several databases on molecular biology, biochemistry and genetics, but also provide several bioinformatics tools such as BLAST [98]. One of the notable databases belonging to NCBI is GenBank, where genome sequences of organism can be found. Data on genes, proteins, single nucleotide polymorphisms, taxonomy, can all be reached using NBCI.

## 4.4. SOFTWARE USED

### 4.4.1. MATLAB

MATLAB is a software and a matrix-based scripting language developed for scientists and engineers by MathWorks. MATLAB is designed primarly for numeric operations yet with the optional toolboxes, symbolic computations are also possible. MATLAB enables users to make matrix manipulations, function and data plotting, algorithm developments and model creations [99].

MATLAB R2018a version was used in this thesis for model analysis with the help of addition third party toolboxes and solvers.

### 4.4.2. GLPK (GNU Linear Programming Kit)

GLPK (GNU Linear Programming Kit) is a solver organized as a callable library, written in ANSI c and intended for linear programming (LP) and mixed integer programming (MIP) problems [100].

### 4.4.3. COBRA (COnstraint-Based Reconstruction and Analysis) Toolbox

COnstraint-Based Reconstruction and Analysis (COBRA) Toolbox is compilation of several algorithms required for constrain based analysis of metabolic models on MATLAB platform. COBRA toolbox V.3 consists of updated methods for model reconstruction, analysis, strain and experimental design and network integration for various data such as proteomics, transcriptomics and metabolomics [101].

### 4.4.4. BLAST (Basic Local Alignment Search Tool)

BLAST (Basic Local Alignment Search Tool) is one of the popular bioinformatics software for sequence comparison. Both the online and stand alone version is hosted by the NCBI. There are different versions of BLAST and algorithms for different types of biological

sequences. Blastn is used for nucleotide sequences and Blastp is used for protein sequence alignments. BLAST takes a heuristic approach and tries to find local and short similarities between two sequences. It uses substitution score matrices to measure the similarities between the two nucleotides or amino acids [102].

## 4.5. CALCULATIONS OF Q RATES

The special formation or consumption rates were calculated using the fermentation data found from literature using the formula 4.1 where c denotes the metabolite concentration; t, time and x, the biomass. $\Delta$c is the difference between the metabolite concentrations at $t_0$ and $t_1$, while $\Delta$t denotes the difference between $t_0$ and $t_1$.

$$q = \frac{\Delta c}{\Delta t \cdot (\frac{x_1+x_2}{2})} \tag{4.1}$$

The enzyme activity data were converted into the q rates via the unit conversions. For example if enzyme activity is defined as mol per gram protein per unit time, it was converted into mmol per gram dry weight per unit time assuming that biomass is 60 per cent protein.

# 5. RESULTS AND DISCUSSION

The genome of ASM2132v1 assembly was uploaded to the RAST (Rapid Annotation using Subsystem Technology) server which is an automated annotation service for bacteria and archaea [91, 92, 93]. For annotation, the options on Table 4.2 were chosen. Once the annotation was complete, Model SEED server was used to create the draft model using the complete media option. The SBML file was downloaded. The same steps were taken for the assembly ASM6704v1 for a second draft model. After both reconstructions were completed, the reactions which were unique to the model ASM6704v1 were added to the model ASM2132v1 in order to reconstruct a super model.

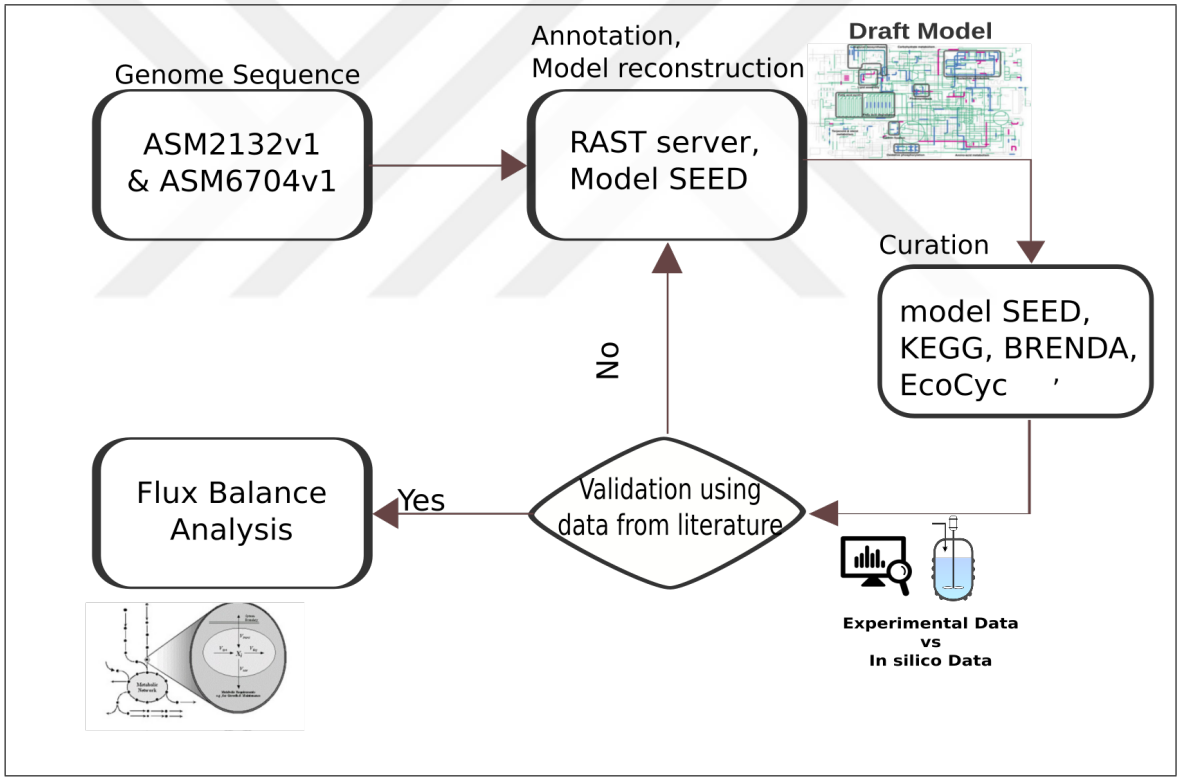



Figure 5.1. Workflow for the model reconstruction

Initially there were 1378 reactions and 1469 metabolites in model named iAU848. To check if there were any gaps in the initial draft model, the gap-fill algorithm of the Model SEED was run for the first time and there were 57 added reactions.

In order to check the carbon and nitrogen balances in the draft model, the function

”checkBalance” which is included in the COBRA Toolbox was used. This function checks if each reaction is elementally balances using the metabolic formulas and stoichiometric data found in the model. Exchange reactions and three more reactions corresponding to RNA,DNA and protein biosynthesis were identified to be elementally unbalanced. Exchange reactions were expected to be unbalanced as they represent one way reactions such as secretion of a metabolite to the environment or addition of metabolites in the environment of cells representing the *in silico* media in which cells grow. The three extra reactions were removed from the model as they did not have any reaction stoichiometry or formulas for the metabolites they consisted of.

After the first simulation of the draft model with the objective function biomass, exchange reactions were set to zero one by one to find out which metabolites were not produced by the model.

The model was taking up various dipeptides from media. After constraining these dipeptide exchange reactions to zero, the model stopped simulating the growth. This indicated that there were some gaps in the amino acid biosynthesis pathways. For all 20 amino acids, first of all new exchange reactions were added to the model. Then all the dipeptide exchange reactions were constrained to zero. Finally, to find out which amino acids were not being synthesized, the amino acid exchange reactions were set to zero one at a time and checked if the model still simulated growth. For five amino acids, gaps were found in the biosynthesis pathways, these amino acids and reactions that were added to ensure model growth can be seen on Table 5.1.

Table 5.1. Amino acids that could not be produced by the initial draft model

| **Amino acid name** | **Added reactions** |
|---|---|
| Tyrosine | rxn01256 |
| Phenylalanine | rxn01256 |
| Asparagine | rxn00416 |
| Histidine | rxn02160 |
| Lysine | rxn01644, rxn02928, rxn02929, rxn03087 |

From the literature, the *Gluconacetobacter diazotrophicus* is known to produce three different plant hormones; Indole-3-acetic acid (IAA), giberellin 1 (GA1) and giberellin 3 (GA3). The draft genome lack the reactions to produce any plant hormones. According to

the literature the IAA was produced from the IpyA pathway so the missing reactions were added from the literature information about that pathway [35, 36]. Giberrellin biosynthesis reactions were not detailed for bacteria, hence reaction for precursors and a super reaction for both GA1 and GA3 productions were added to the model (Table 5.2). Three exchange reactions were added to the model to simulate hormone secretion.

Table 5.2. Hormone biosynthesis reactions added to the model

| **Plant hormone name** | **Added reactions** |
|---|---|
| Indole-3-acetic acid | tryptpyramin, rxn00483, rxn00477, rxn01937, rxn11964 |
| Giberellin 1 | rxn23454, rxn23455, rxn03452, rxn01490, ga1 |
| Giberellin 3 | rxn23454, rxn23455, rxn03452, rxn01490, ga3 |

External sucrose hydrolyses reaction was added to the model and the levan production reaction was changed into an external reaction instead of cytoplasmic one since the bacteria do not posses a sucrose transport pathway [11, 48]. Gluconic acid production reaction was also changed into an external one. Reactions which represent the possible external catabolism of gluconic acid into 2-ketogluconic acid,5-ketogluconic acid and 2,5-diketogluconic acid were also added to the model. Transport and exchange reactions were added for gluconic acid, 2-ketogluconic acid,5-ketogluconic acid and 2,5-diketogluconic acid. Glucose transport was initially missing from the model was added.

There were three metabolites that were absent in the model but needed for the biomass formation; spermidine, myristic acid, lauric acid. Reactions ”rxn02061, rxn15947 and rxn15949” were added manually for spermidine production. For the other two metabolite missing reactions belonging the fatty acid biosynthesis, elongation and degredation pathways were added.

The model was not simulating growth in the absence of metabolite trehalose, which turned out to be a problem with reactions related to Uridine diphosphate glucose (UDPG). The *Escherichia coli* genome scale model iAF1260 was used as a reference for those reactions [103]. The absentee reactions that were present in the *E. coli* model were manually added.

Ethanol, acetate and levan exchange reactions were added.

Simulations were run with different type of carbon sources. There were no growth on

sorbitol and galactose contrary to literature [48]. Missing reactions were added.

Lery et al., did a proteomics analysis on the *Gluconacetobacter diazotrophicus*. From this study's supplementary files, enzymes (and their corresponding reactions) that were validated to be present in the *G. diazotrophicus* metabolism were added to the model using modelSEED's reaction database [104].

Using KEGG's organism specific pathway maps, the reactions that were absent on modelSEED annotation but present on the KEGG *G. diazotrophicus* annotations were added to the model from all available pathways.

When the model was simulated with minimizing the objective function "oxygen exchange", the FBA resulted in infeasible solution which indicated missing reactions from oxidative phosphorylation. For those reactions, a genome-scale metabolic model for *Gluconacetobacter xylinus* and iAF1260 were used references and 11 reactions were added [103, 105].

Apart from the annotated reactions from modelSEED and KEGG, all the enzymes for added reactions were blasted against the *Gluconacetobacter diazotrophicus*' proteome using protein BLAST with the DELTA-BLAST algorithm. The BLAST results can be seen on appendix Table 7.1. The logic behind using the proteome rather than the genome of the bacteria that proteins are more conserved and genes for the target proteins may vary because of the codon bias

After the gap-fill step and manual curation, the genome-scale metabolic model of *Gluconacetobacter diazotrophicus* consists of 1754 reactions, 1629 metabolites and 848 genes.

*G. diazotrophicus* has limited genetic diversity as the environment within the sugarcane stems which is the natural habitat of this bacterium is relatively constant [106]. It has been observed that PAL 3 and PAL 5 strains have same observable characteristics, for instance they can grow on the same carbon sources and under same pH range, they both fix nitrogen even in the presence of $NO_3^-$ [23]. Due to this reason, the validation of the model was done using the literature data on both strains.

## 5.1. BIOMASS COMPOSITION

Using the components of the biomass function taken from the model SEED for a generic gram negative bacterium, a general biomass formula was generated. After close inspection of the biomass function, the metabolites which are the product of the three removed reactions correspoding for RNA,DNA and protein synthesis were discarded as they do not linked to any other metabolites or reactions . The biomass formula per one carbon was found to be $CH_{1.60}O_{0.37}N_{0.22}P_{0.021}S_{0.006}$.

The stoichiometric ratios of the elements in the biomass formula is in agreement with the ones in the literature [50].

## 5.2. VALIDATION

Metabolic model validation requires experimental data acquired from either from batch fermentation or preferably chemostate. For this thesis, the validations were performed using the batch data found in the literature [51]. The batch fermentation data contains specific comsumption/production rates for glucose, gluconic acid, glycerol and 2-ketogluconic acid (Figure 4.1-4.6). The data did not contain any information about the $qCO_2$ or $qO_2$ rates, these values were calculated from carbon balance and degrees of reduction balance. The calculated experimental q rates for carbon sources,oxygen, products and byproducts for different phases of the growth are given in tables 5.3-5.8.

In order to validate the model, the specific consumption rates of carbon sources were given to the model as constrains, the production rates for products were also used as constrains. The $CO_2$ and $O_2$ q rates were constrained using the calculated values as upper or lower boundaries, but there were instances were the calculated values did not reflect the biological characteristics of the bacteria. In these cases, the upper and lower boundaries were set to the default values in the model. The validation simulations were performed using biomass reaction as the objective function.

According to the data taken from the Luna (2003), when the bacteria are grown on the media containing glucose as carbon source, they demonstrate a three-phase growth pattern [51]. In

the first phase the glucose is converted into gluconic acid which results in the drop of pH. During early stage of this phase, the specific growth rate is low; 0.065 1/h until the bacteria convert the most of the glucose into gluconic acid. During the second phase, bacteria starts metabolizing the gluconic acid with a concomitant accumulation of keto-acids outside the cell. At this stage the specific growth rate is doubled. This phase has the highest specific growth rate compared to other phases, as most of the carbon source (in this case gluconic acid) is used for biomass growth. At the third phase, when the gluconic acid concentration drops low, the cells start using the keto-acids as carbon sources. The specific growth rate also decreases. This three phase growth pattern can be observed both with and without nitrogen fixing conditions (Tables 5.3-5.5) [51].

As reported by Luna (2003), when the carbon source is only gluconic acid, there are two phases for bacterial growth. During the first phase gluconic acid is used for growth while small percentage the gluconic acid is converted to the 2-ketogluconic acid. After the gluconic acid concentration reduces to a certain point, the consumption of 2-ketogluconic acid starts. The specific growth rate is halved during the second phase as the total carbon source concentration is lower than the first phase. This pattern is again observable under both with BNF and without BNF conditions (Table 5.6-5.7) [51].

Luna (2003) also demonstrated that G. diazotrophicus can grow on glycerol as carbon source, during which the growth rate is slow (Table 5.8). During this fermentation, gluconic acid or ketoacid were not produced hence the pH of the environment dropped slowly [51].

Table 5.3. Validation of experimental data; 5 g/l glucose with BNF

| **Condition: 5 g/l Glucose with BNF** | | **qGlucose (mmol/gDW/h)** | **q2-Ketogluconic Acid (mmol/gDW/h)** | **qGluconic Acid (mmol/gDW/h)** | **$qCO_2$ (mmol/gDW/h)** | **$qO_2$ (mmol/gDW/h)** | **qBiomass (1/h)** |
|---|---|---|---|---|---|---|---|
| **1st Phase Glucose Consumption** | **Experimental Value** | -28.471 | -3.875 | 29.355 | -7.891 | -13.572 | 0.065 |
| | **In Silico Value** | -28.471 | -3.875 | 29.355 | 9.039 | -15.61 | 0.228 |
| **2nd Phase Gluconic Acid Consumption** | **Experimental Value** | -1.097 | 6.848 | -16.851 | 103.105 | -189.359 | 0.115 |
| | **In Silico Value** | -1.097 | 6.848 | -16.851 | 27.028 | -6.655 | 1.013 |
| **3rd Phase 2-Ketogluconic Acid Consumption** | **Experimental Value** | 0 | -2.666 | -0.921 | 4.458 | -7.995 | 0.027 |
| | **In Silico Value** | 0 | -2.666 | -0.921 | 14.473 | -7.995 | 0.181 |

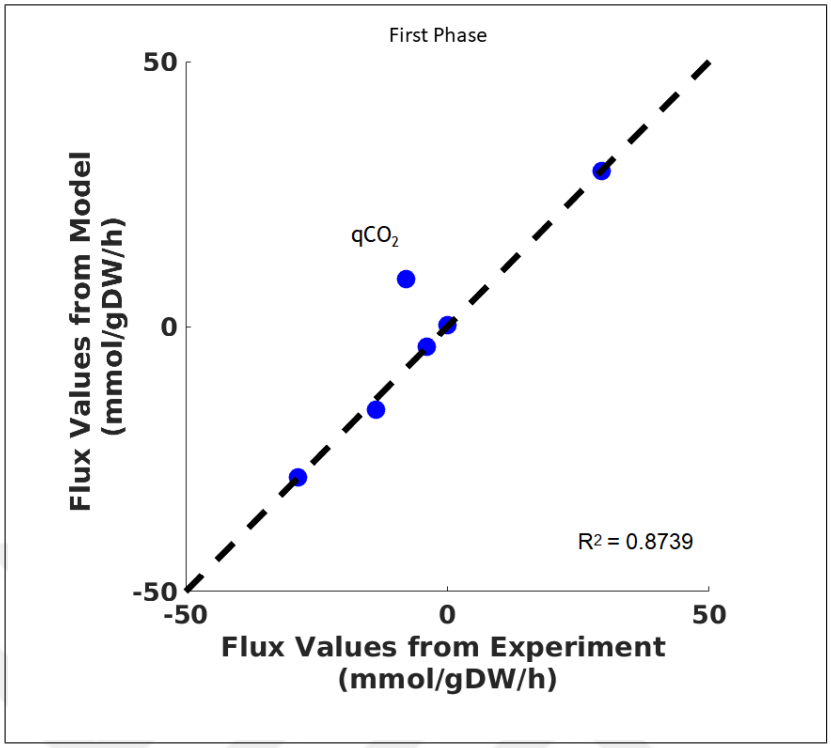



Figure 5.2. Comparison of experimental and *in silico* data for 5 g/l glucose and BNF during first phase of growth

When the simulations were performed with constraints taken from the table 5.3 using biomass reaction as objective function and the biological nitrogen fixation was allowed. For the first phase, $qCO_2$ and $qO_2$ values could not be set to the experimental values as the simulation could not be solved. Both exchange reactions were set to default constrains (-1000,1000). According to the *in silico* values, the $CO_2$ is produced not consumed which agrees with the biological characteristics of this species, as it is not a carbon-fixer. The $qO_2$ *in silico* value is close to the experimental value and also coherent with the biology of the bacterium as it is an obligate aerobe. According to the simulations, the specific growth rate is almost 3.5 times higher than the experimental value. This difference might be due to the fact that, during the experiment, other byproducts were produced but they were not measured and taken into the account. Since the simulations are performed using the data from the experiment, the lack of constraints on byproducts result in using the most of the carbon from the substrates into the biomass rather than other biological products.

For the second phase simulation, again the specific rates for glucose, gluconic acid and 2-ketogluconic acid were set as constraints. The specific rates for $CO_2$ and $O_2$ were set

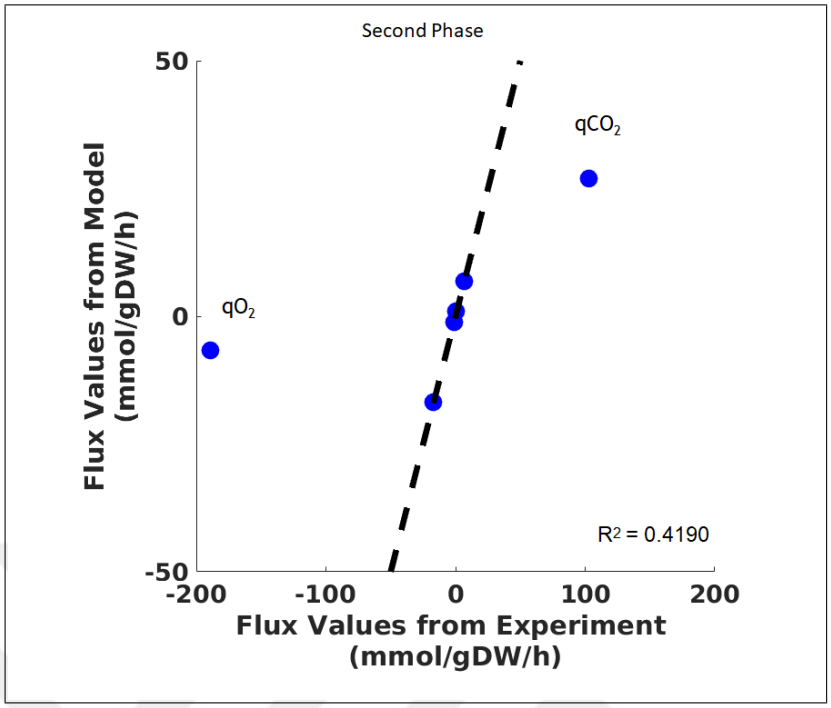



Figure 5.3. Comparison of experimental and *in silico* data for 5 g/l glucose and BNF during second phase of growth

as upper bound and lower bound respectively. The *in silico* values for specific rates of biomass, $CO_2$ and $O_2$ were divergent from the experimental values. Again like the first phase simulation, the growth rate was higher than the experimental value, but the specific rates of $CO_2$ and $O_2$ were lower. For the third phase simulation, only specific rate of $CO_2$ could not be constrained to the experimental value. The $CO_2$ specific rate was estimated to be higher than the experimental value and like the other two phases, the growth rate was higher as well. The inconsistencies between the experimental and in silico values might be due to the fact that experimental data were missing measurements for products other than they targeted. Also, the experiment was a batch fermentation where the concentration of metabolites changes instantaneously while the genome scale metabolic model assumes that the concentrations do not change with respect to time.

Although there were inconsisties between the simulation and the batch fermentation,for all phases, the pattern of the growth rate observed during the fermentation could be simulated with the model as well.

For simulations performed using the data from Table 5.4 and 5.5, the biological nitrogen

Table 5.4. Validation of experimental data; 5 g/l glucose without BNF

| **Condition: 5 g/l Glucose without BNF** | | **qGlucose (mmol/gDW/h)** | **q2-Ketogluconic Acid (mmol/gDW/h)** | **qGluconic Acid (mmol/gDW/h)** | **qCO2 (mmol/gDW/h)** | **qO2 (mmol/gDW/h)** | **qBiomass (1/h)** |
|---|---|---|---|---|---|---|---|
| **1st Phase Glucose Consumption** | **Experimental Value** | -26.778 | -0.601 | 39.364 | -77.246 | 115.127 | 0.043 |
| | **In Silico Value** | -26.778 | -0.601 | 27.000 | 2.077 | -14.958 | 0.228 |
| **2nd Phase Gluconic Acid Consumption** | **Experimental Value** | -3.390 | 2.470 | -7.879 | 63.284 | -118.689 | 0.108 |
| | **In Silico Value** | -3.390 | 2.470 | -7.879 | 52.306 | -50.791 | 0.013 |
| **3rd Phase 2-Ketogluconic Acid Consumption** | **Experimental Value** | 0.000 | -5.151 | -0.998 | 3.082 | -5.167 | 0.073 |
| | **In Silico Value** | 0.000 | -5.151 | -0.998 | 7.000 | -1.320 | 0.009 |

Table 5.5. Validation of experimental data; 10 g/l glucose without BNF

| **Condition: 10 g/l Glucose without BNF** | | **qGlucose (mmol/gDW/h)** | **q2-Ketogluconic Acid (mmol/gDW/h)** | **qGluconic Acid (mmol/gDW/h)** | **$qCO_2$ (mmol/gDW/h)** | **$qO_2$ (mmol/gDW/h)** | **qBiomass (1/h)** |
|---|---|---|---|---|---|---|---|
| **1st Phase Glucose Consumption** | **Experimental Value** | -28.836 | 2.316 | 29.903 | -8.843 | -12.217 | 0.061 |
| | **In Silico Value** | -28.836 | 2.316 | 26.391 | 0.500 | -15.986 | 0.007 |
| **2nd Phase Gluconic Acid Consumption** | **Experimental Value** | -0.821 | 2.388 | -5.577 | 36.636 | -67.694 | 0.044 |
| | **In Silico Value** | -0.821 | 2.388 | -5.577 | 23.860 | -23.441 | 0.005 |
| **3rd Phase 2-Ketogluconic Acid Consumption** | **Experimental Value** | 0.000 | -0.536 | -0.877 | 4.616 | -8.355 | 0.016 |
| | **In Silico Value** | 0.000 | -0.536 | -0.877 | 4.616 | -3.635 | 0.002 |

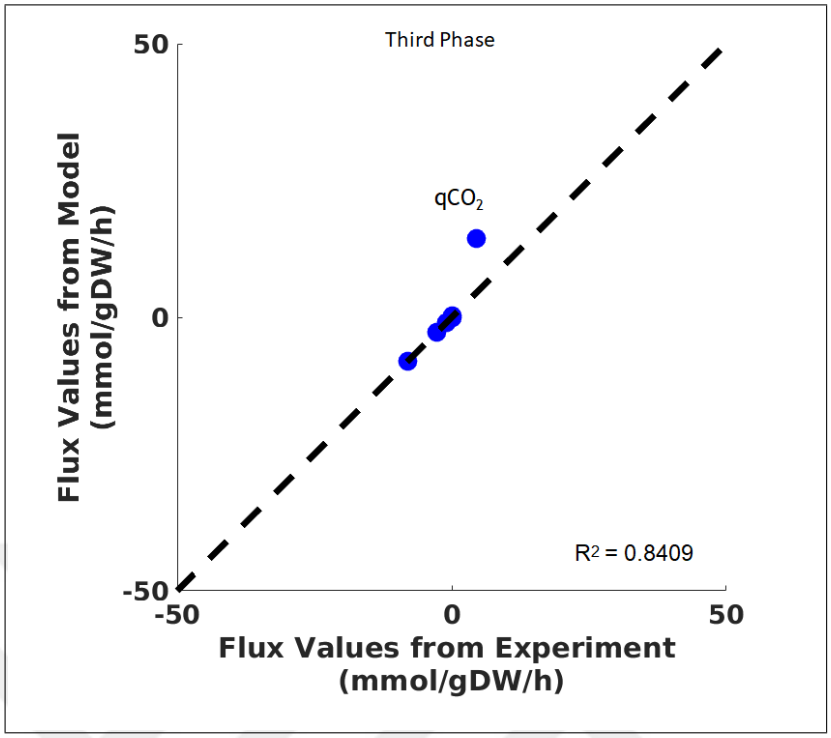



Figure 5.4. Comparison of experimental and *in silico* data for 5 g/l glucose and BNF during third phase of growth

fixation was not simulated. Instead, since the media that batch fermentation grew consisted of $NH_3$ as nitrogen source, the exchange $NH_4$ reaction used. The specific rate of the $NH_4$ exchange reaction were set to experimental growth rate value, as it was assumed that, the nitrogen source was only used for the growth and the rate it is taken should be equal to the growth rate. Although it is expected that the growth rate should be higher than the BNF conditions as the model does not use ATP or sources for nitrogen fixation. It is calculated to be very low compared to experimental values. The reason behind this case may be that the model nitrogen source was constrained very strictly and the nitrogen was not enough to grow for such high rates.

When the G. diazotrophicus is grown with gluconic acid as the carbon source, it shows a two phase growth pattern (Table 5.6-5.7). During the first phase, just like the second phase of the previous experiment, the gluconic acid is consumed while a percentage of it is converted into 2-ketogluconic acid. During the second stage, both the remaining gluconic acid and the 2-ketogluconic acid are used as carbon sources. The cells grow with a faster rate during the first phase. This pattern is observed with both BNF and without BNF conditions [51].

Table 5.6. Validation of experimental data; 5 g/l gluconic acid with BNF

| **Condition: 5 g/l Gluconic Acid with BNF** | | **q2-Ketogluconic Acid (mmol/gDW/h)** | **qGluconic Acid (mmol/gDW/h)** | **$qCO_2$ (mmol/gDW/h)** | **$qO_2$ (mmol/gDW/h)** | **qBiomass (1/h)** |
|---|---|---|---|---|---|---|
| **1st Phase Gluconic Acid Consumption** | **Experimental Value** | 1.936 | -6.013 | 32.203 | -58.394 | 0.097 |
| | **In Silico Value** | 1.936 | -6.013 | 24.000 | -22.712 | 0.012 |
| **2nd Phase 2-Ketogluconic Acid Consumption** | **Experimental Value** | -0.434 | -0.343 | 0.049 | 0.245 | 0.050 |
| | **In Silico Value** | -0.434 | -0.343 | 2.174 | -0.385 | 0.064 |

Table 5.7. Validation of experimental data; 5 g/l gluconic acid without BNF

| **Condition: 5 g/l Gluconic Acid without BNF** | | **q2-Ketogluconic Acid (mmol/gDW/h)** | **qGluconic Acid (mmol/gDW/h)** | **$qCO_2$ (mmol/gDW/h)** | **$qO_2$ (mmol/gDW/h)** | **qBiomass (1/h)** |
|---|---|---|---|---|---|---|
| **1st Phase Gluconic Acid Consumption** | **Experimental Value** | 3.017 | -4.358 | 22.104 | -39.850 | 0.101 |
| | **In Silico Value** | 3.017 | -4.358 | 7.589 | -8.384 | 0.012 |
| **2nd Phase 2-Ketogluconic Acid Consumption** | **Experimental Value** | -1.552 | -0.026 | -2.747 | 5.520 | 0.073 |
| | **In Silico Value** | -1.552 | -0.027 | 9.138 | -7.543 | 0.009 |

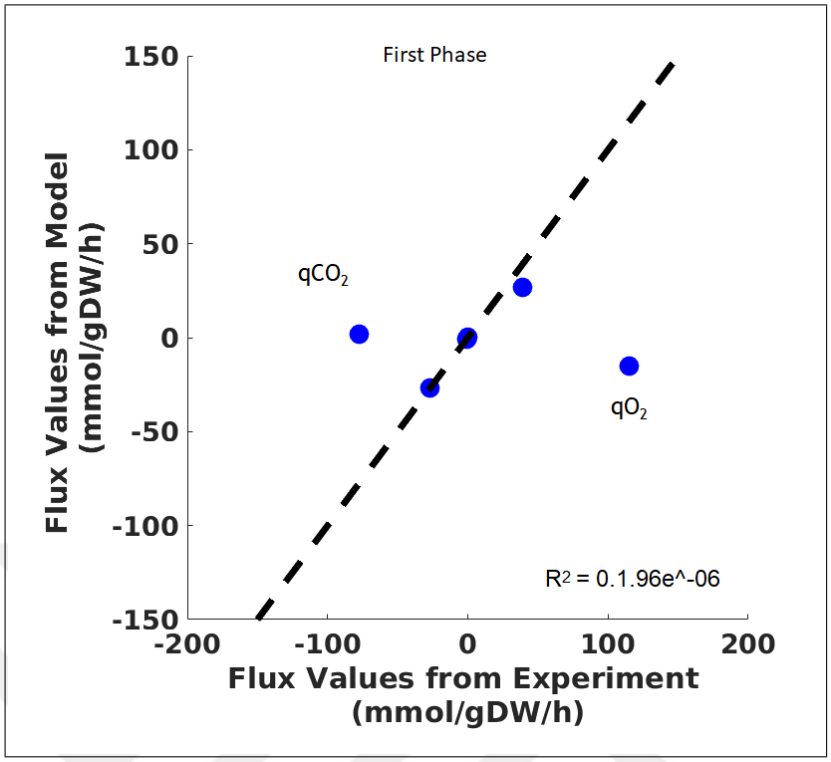



Figure 5.5. Comparison of experimental and *in silico* data for 5 g/l glucose and without BNF during the first phase of growth

Albeit the model simulated the growth patterns for BNF and without BNF conditions with 5g/l gluconic acid, the growth rates were distinctively lower than the experimental values especially when BNF was not active. The differences between the experimental and *in silico* values might be due the fact that at the points of the batch fermentation where the data is measured and calculated from might not be suitable to simulate using GSMMs.

Table 5.8. Validation of experimental data; 5 g/l glycerol without BNF

| **Condition: 5 g/l Glycerol without BNF** | **qGlycerol (mmol/gDW/h)** | **$qCO_2$ (mmol/gDW/h)** | **$qO_2$ (mmol/gDW/h)** | **qBiomass (1/h)** |
|---|---|---|---|---|
| **Experimental Value** | -4.149 | 22.978 | -45.956 | 0.048 |
| **In Silico Value** | -4.149 | 12.229 | -14.283 | 0.006 |

When the batch fermentation was performed with glycerol for the carbon source, it was measured that the growth rate is lower than the other two carbon source glucose and gluconic acid [51]. The simulation growth rate was again lower than the experimental value. This might be due to the fact that the rate of exchange reaction for the nitrogen source ($NH_4$) was

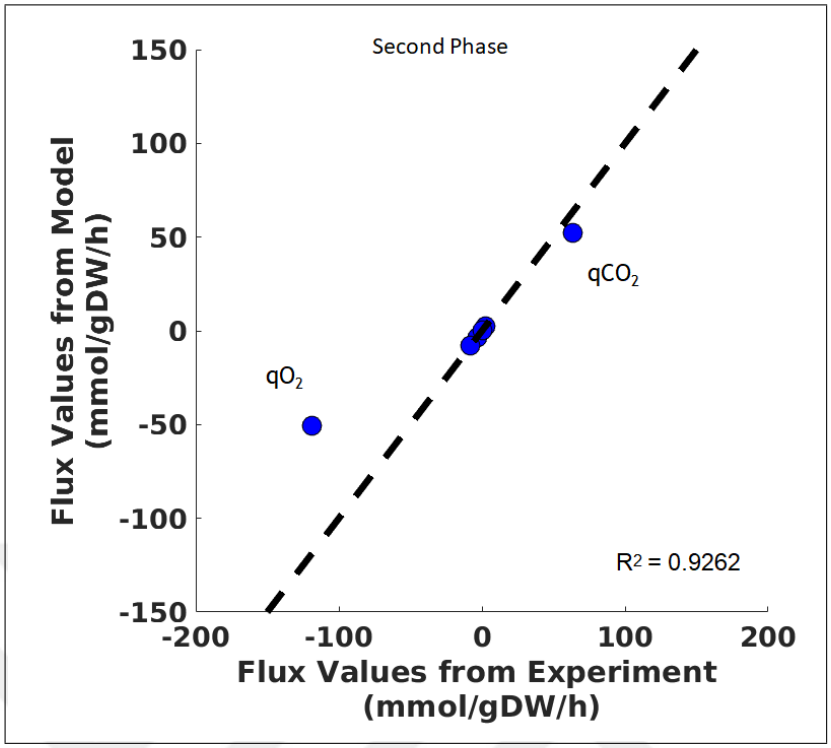



Figure 5.6. Comparison of experimental and *in silico* data for 5 g/l glucose and without BNF during second phase of growth

set low than the experimental value. Since there are no data for specific consumption rate of N-source during the batch experiment, the constraint for the $NH_4$ was assumed to be equal to the batch fermentation growth rate.

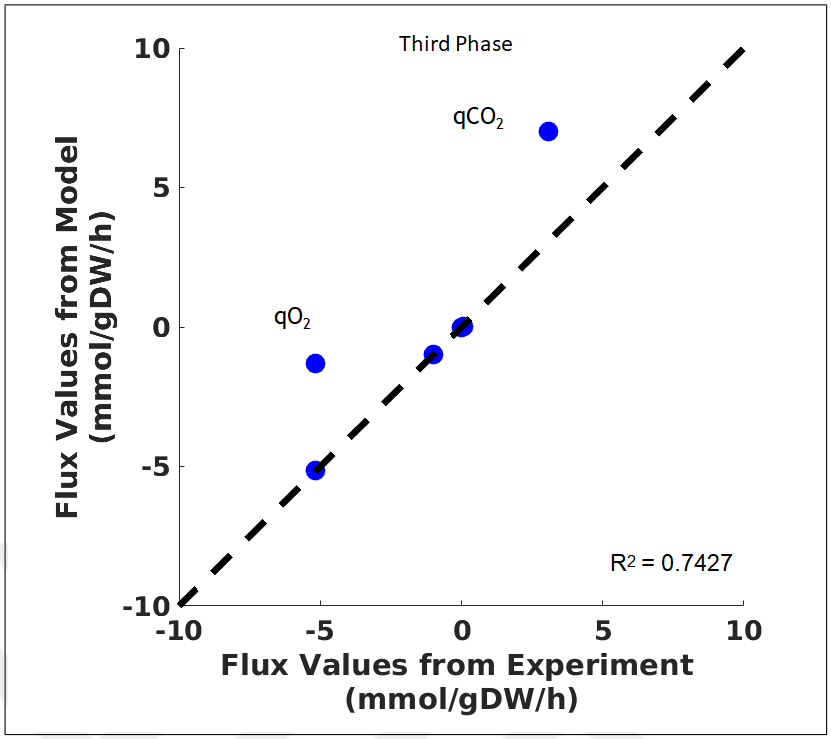



Figure 5.7. Comparison of experimental and *in silico* data for 5 g/l glucose and without BNF during third phase of growth

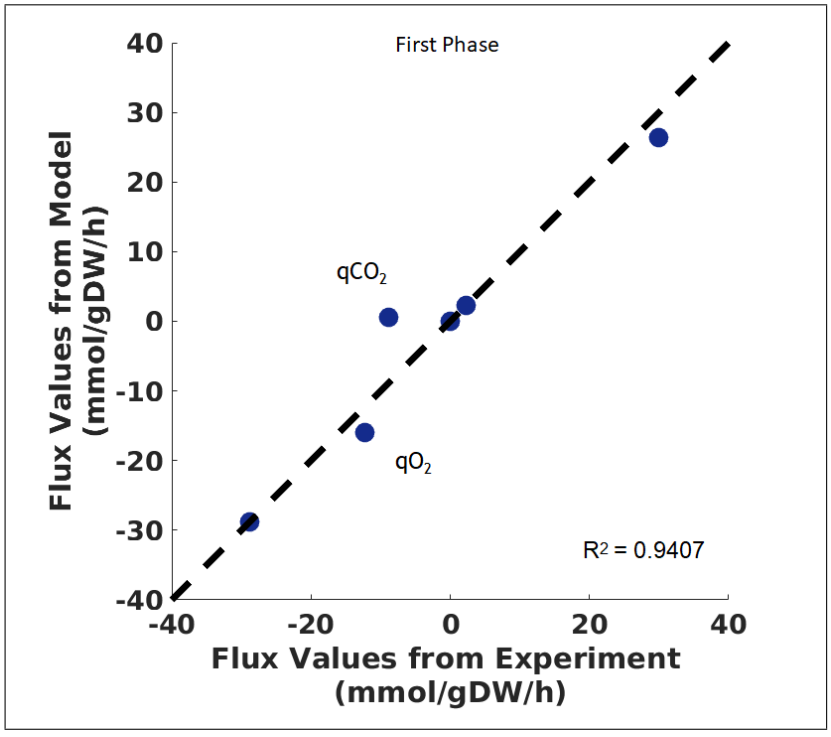



Figure 5.8. Comparison of experimental and *in silico* data for 10 g/l glucose and without BNF during first phase of growth

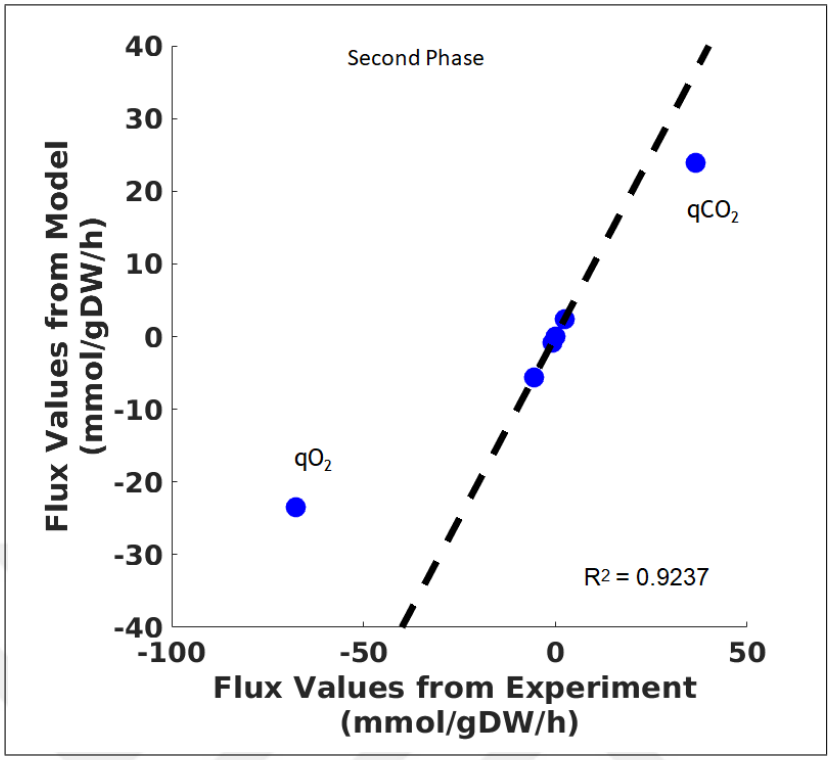



Figure 5.9. Comparison of experimental and *in silico* data for 10 g/l glucose and without BNF during second phase of growth

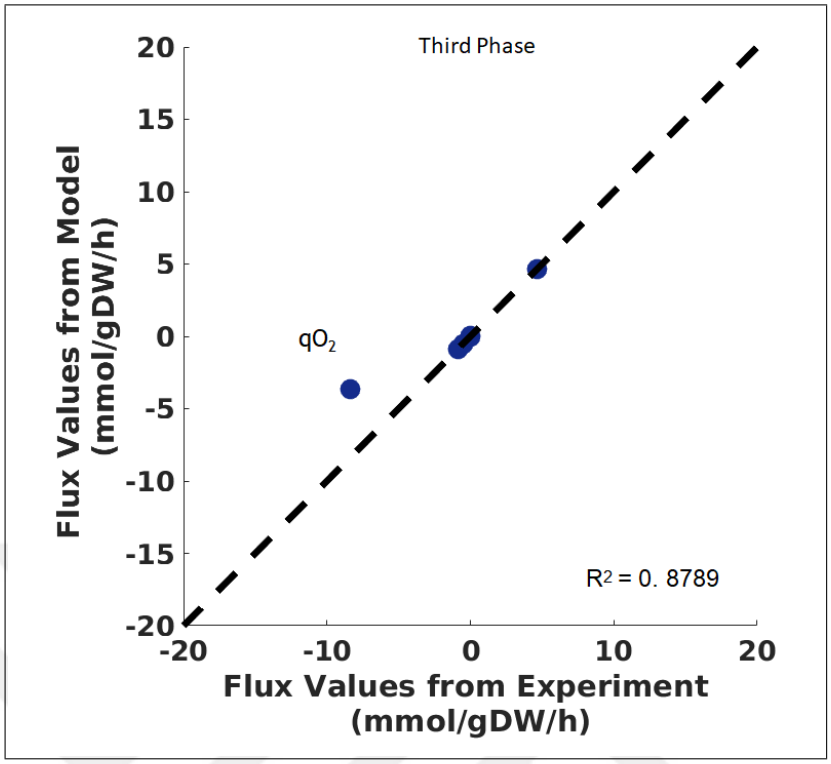



Figure 5.10. Comparison of experimental and *in silico* data for 10 g/l glucose and without BNF during third phase of growth

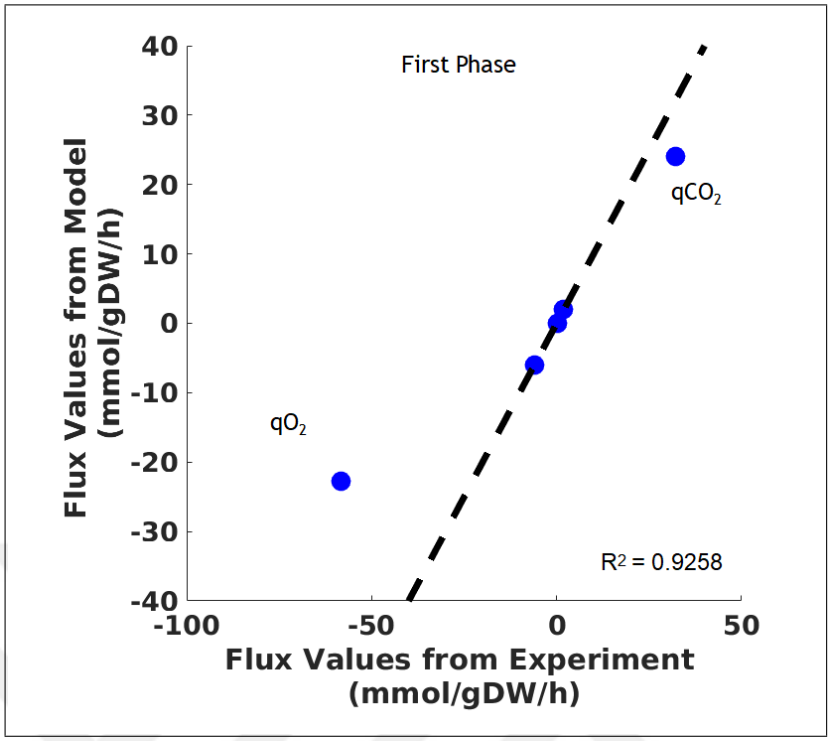



Figure 5.11. Comparison of experimental and *in silico* data for 5 g/l gluconic acid and BNF during first phase of growth

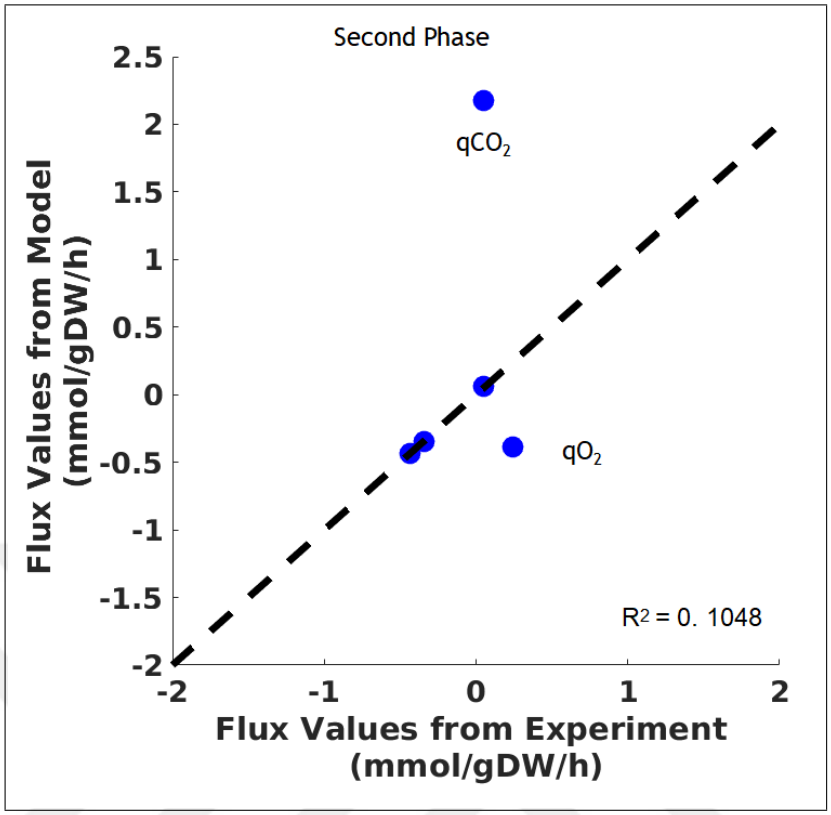



Figure 5.12. Comparison of experimental and *in silico* data for 5 g/l gluconic acid and BNF during second phase of growth

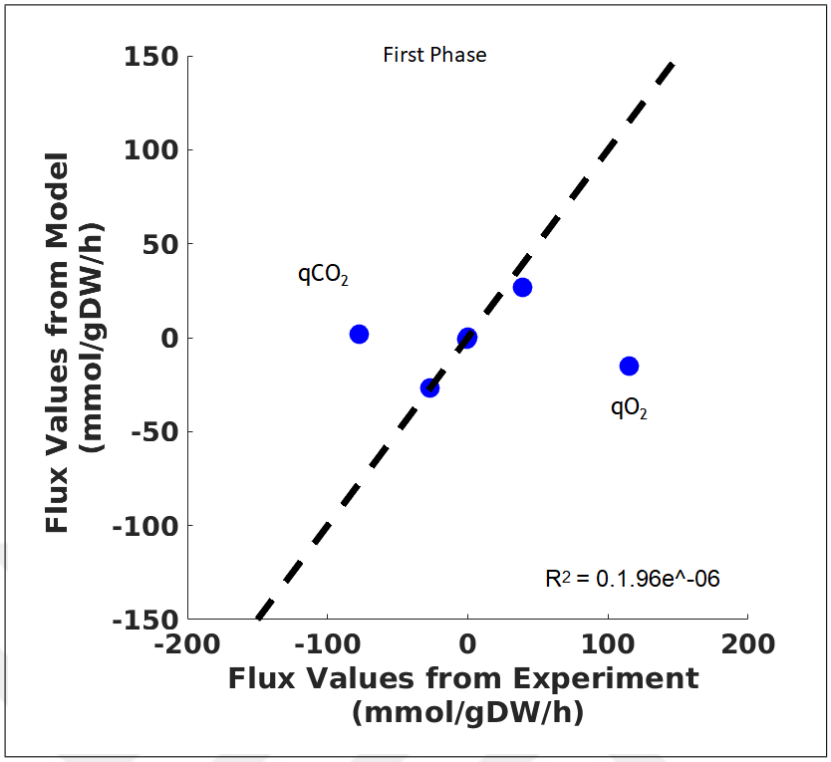



Figure 5.13. Comparison of experimental and *in silico* data for 5 g/l gluconic acid and without BNF during first phase of growth

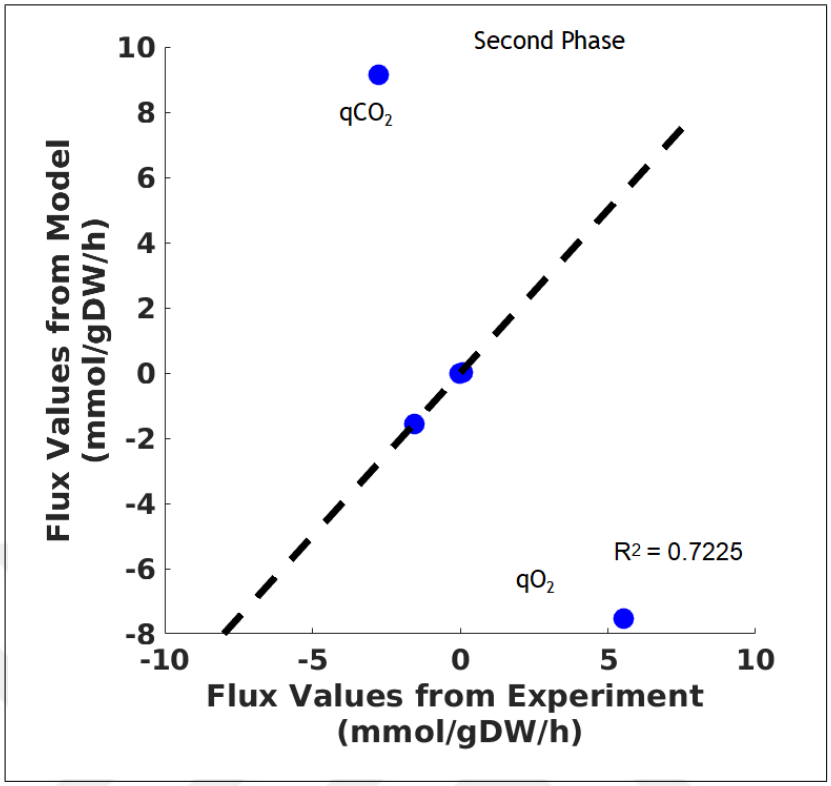



Figure 5.14. Comparison of experimental and *in silico* data for 5 g/l gluconic acid and without BNF during second phase of growth

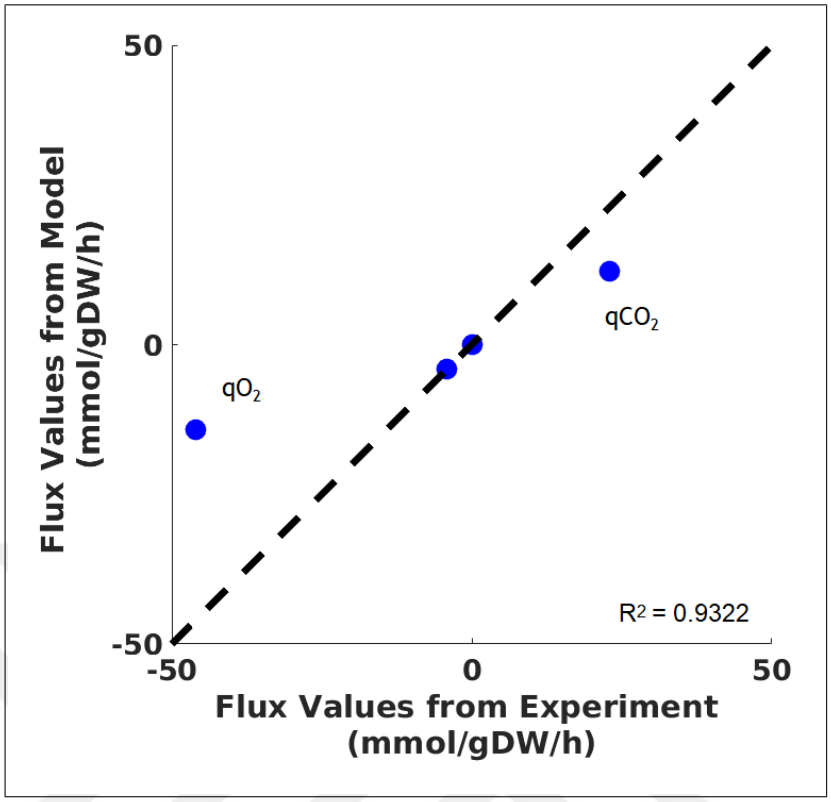



Figure 5.15. Comparison of experimental and in silico data for 5 g/l glycerin without BNF

## 5.3. CARBON AND NITROGEN METABOLISMS ANALYSIS

The model is able to simulate growth using different carbon sources such as glucose, sucrose, fructose, gluconic acid etc. which agrees with the literature [48]. The growth rates and calculated yields are shown on Table 5.9 for various carbon sources. As expected, when nitrogen source is supplied to the model rather than using nitrogen fixation as sole nitrogen source, the yields are higher. Since nitrogen fixation requires 16 mol of ATP per two moles of NH3 fixed, when there is an external nitrogen source, the cell can use the 16 ATP on growth rather than BNF [9].

Table 5.9. Yields and specific growth rate with different carbon sources

| **Carbon Source qS=-10 mmol/gDW/h** | **Yield (BNF)** | **Yield (without BNF)** | **$\mu$ (BNF)** | **$\mu$ (without BNF)** |
|---|---|---|---|---|
| **Glucose** | 0.48 | 0.59 | 0.87 | 1.06 |
| **Sucrose** | 0.53 | 0.69 | 1.81 | 2.33 |
| **Fructose** | 0.52 | 0.71 | 0.94 | 1.27 |
| **Gluconic Acid** | 0.44 | 0.54 | 0.86 | 1.06 |
| **Ribose** | 0.52 | 0.71 | 0.79 | 1.06 |
| **Glycerol** | 0.51 | 0.69 | 0.47 | 0.64 |
| **Mannose** | 0.52 | 0.71 | 0.94 | 1.27 |
| **Xylose** | 0.52 | 0.71 | 0.79 | 1.06 |

The natural host of the *G. diazotrophicus* is sugarcane. Inside the sugarcane tissue where the bacteria live, the abundant carbon source is sucrose. The optimal condition for the bacterial growth is 10 per cent sucrose [11]. When the grow rates for different the carbon sources are compared, the highest growth rate is calculated for sucrose. The *G. diazotrophicus* hydrolyses the sucrose into fructose and glucose, and can metabolize both monomer for growth. Although when the yields are compared for different carbon sources, the sucrose yield is only higher than the glucose yield. This could be the due the fact that although the bacteria can grow on sucrose fast, cannot use it efficiently, but since inside the host there is no carbon limitation, this inefficiency is not disadvantageous.

The model can simulate the production of three plant hormones, which the *G. diazotrophicus* is known to produce [11, 34, 36]. When the simulation is performed using the biomass reaction as the objective function, the model does not produce any of the hormones, which suggest that the pathway responsible for the hormone production are not required for the

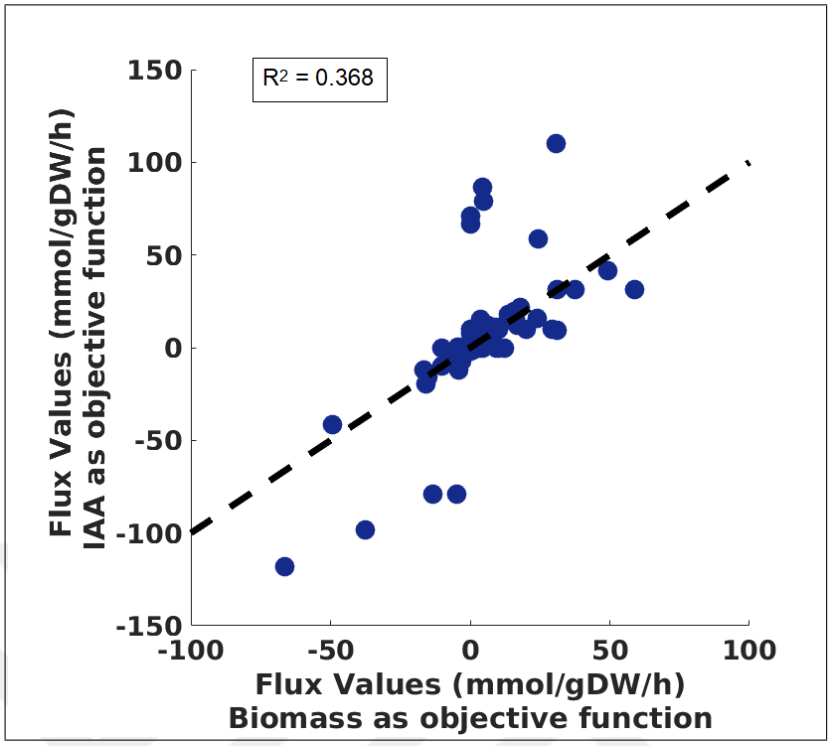



Figure 5.16. Metabolic comparison of phytohormone IAA and biomass productions

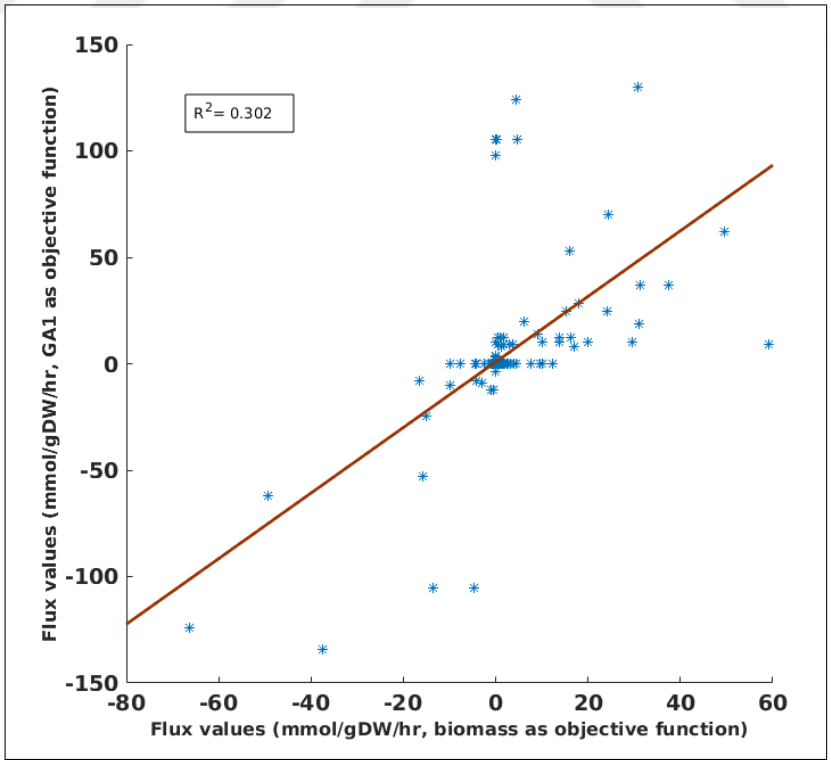



Figure 5.17. Metabolic comparison of phytohormone GA1 and biomass productions

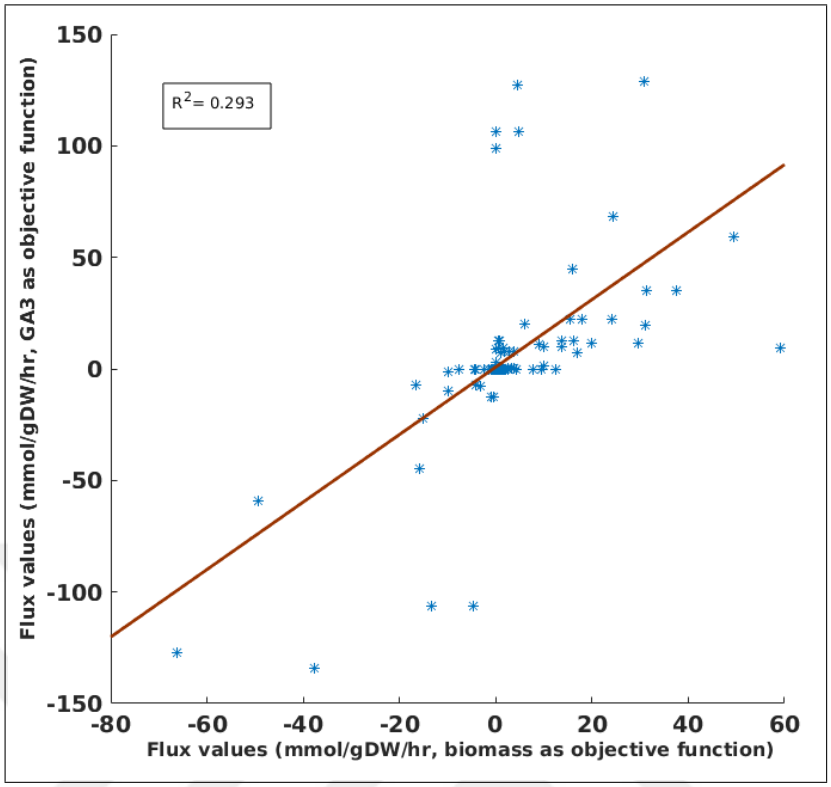



Figure 5.18. Metabolic comparison of phytohormone GA3 and biomass productions

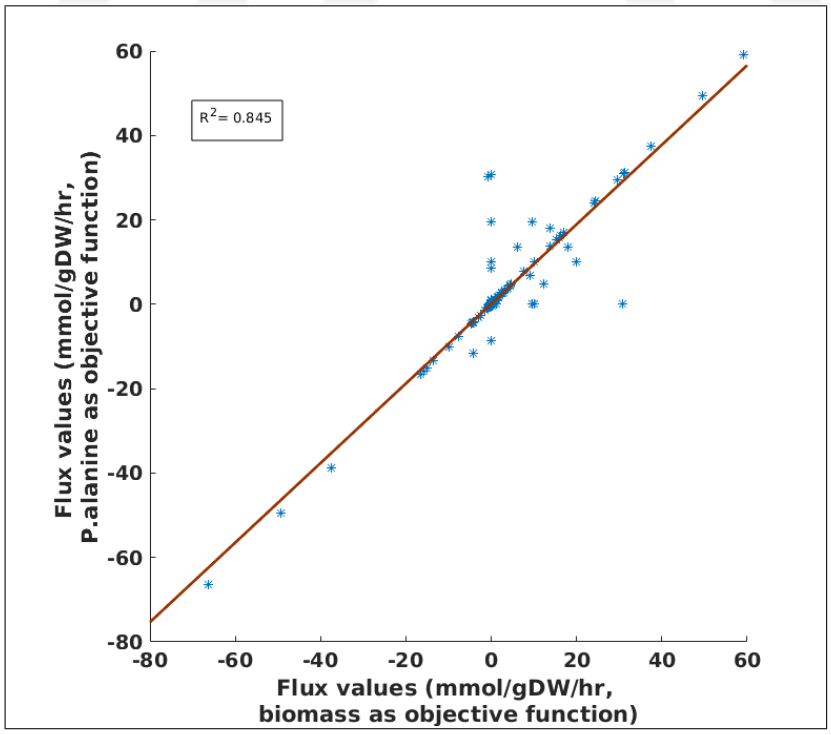



Figure 5.19. Metabolic comparison of amino acid phenylalanine and biomass productions

biomass. As it is indicated by the scatter plots Figures 5.16-5.18, the metabolic pathways required for the biomass production and the hormone production are not the same. In order to see the difference between a product of a pathway which is related to biomass, an amino acid production was simulated and plotted against the biomass production simulation. Figure 5.19 clearly suggests that if a product is linked to the pathways related to the growth, the simulations would have almost identical trends.

In order to investigate the affect of the maintenance ATP, which is the energy required to sustain a cell's living state without any growth, mATP values are plotted against the bacterial yield. By theory, the more maintenance ATP is required, the slower the growth will be. In figure 5.20, Yield vs mATP graph can be seen for one of the most curated metabolic models iND750 [107]. As expected the yield decreases with increasing mATP. In this model, the yield decrease can be observed instantaneously whereas in the model iUA848, the maintenance ATP does not affect the yield till the mATP value reaches to a certain value which is much larger than the *Saccharomyces cerevisiae*, then follows the decrease trend and yield reaches zero at larger mATP values. This could be explained that *G. diazotrophicus* has different energy metabolism where it starts oxidizing glucose outside of the cytoplasm and does not have a full glycolysis patyway thus the metabolism uses PPP mainly for growth [48, 51].

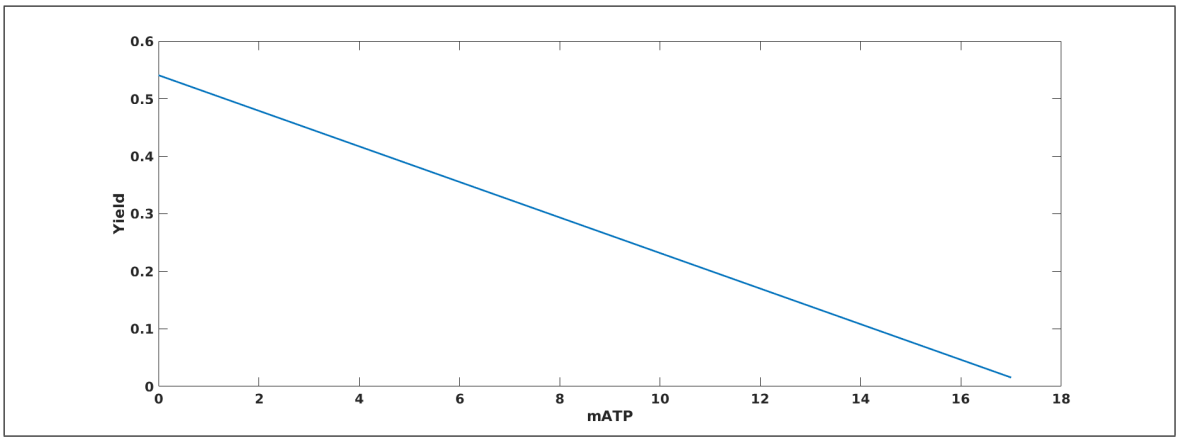



Figure 5.20. Yield vs maintenance ATP values for *Saccharomyces cerevisiae* using model iND750

Since the bacterium *G. diazotrophicus* does not posses a full glycolysis pathway, the

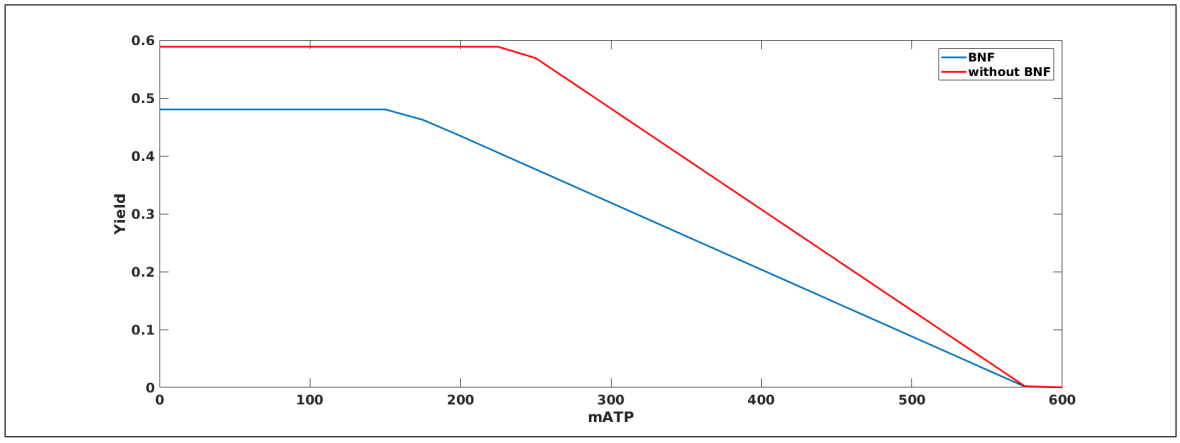



Figure 5.21. Yield vs maintenance ATP values for *G. diazotrophicus* using model iUA848 under both BNF and without BNF conditions

model is constructed to simulate this biological characteristics. Although the bacterium does not use the glycolysis, it only lacks one enzyme from that pathway which is the phosphofructokinase [48]. In order to observe what happens to the central carbon metabolism of the bacterium when it would have a full glycolysis route, the missing enzyme reaction was added to the model. To compare both wildtype metabolism (missing phosphofructokinase) and glycolysis metabolism, both conditions were simulated. Then the reactions which have values different than zero and related to the pathways glycolysis, PPP, TCA and oxydative phosphorylation were selected. For both conditions, scatter graph was plotted Figure 5.22. As is seen from the figure 5.8, compared to wildtype metabolism, the full glycolysis in silico strain, posses more active pathways where as the number of central carbon pathways is lower for the wildtype strain. In wildtype metabolism, cell uses the PPP, to produce pyruvate directly or by generating precursors of pyruvate.

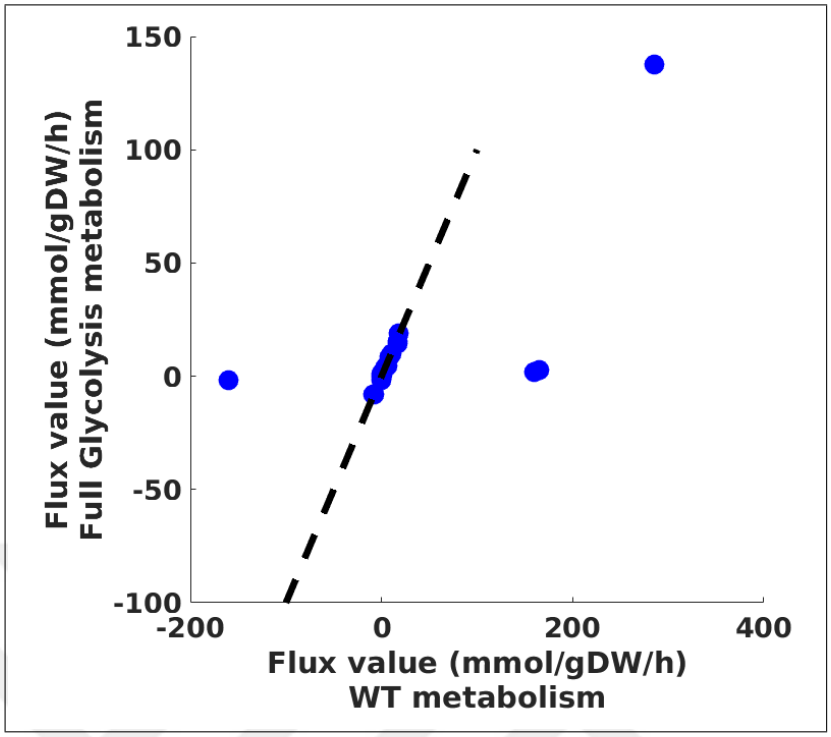



Figure 5.22. Metabolic comparison of wildtype *G. diazotrophicus* and full glycolysis in silico strain using model iUA848.

# 6. CONCLUSIONS

*Gluconacetobacter diazotrophicus* is an acid tolerant, gram negative and obligate aerobe plant growth promoting bacterium with interesting characteristics. To be able to elucidate the bacterium's metabolism a key systems biology tool was missing from the literature, which was the focus of this thesis; the genome-scale metabolic model iUA848. The genome-scale metabolic model was reconstructed using annotated genome sequence, online tools and databases for metabolic reconstruction. The metabolic model consists of 1754 reactions, 1629 metabolites and 848 genes. The validation of the model was performed via comparing literature data and *in silico* generated data. A systems biology approach was taken and the model was analyzed using FBA. *Gluconacetobacter diazotrophicus* is an intriguing organism because it lacks a complete glycolysis pathway and investigating the carbon flux of the metabolism was essential. Using this metabolic model, environmental responses of the organism also investigated under certain conditions, such as with or without nitrogen fixation. Furthermore, the model can be used for targeted metabolic engineering as a guide to exploit the organism's metabolic capabilities. For future work, the model can be improved in way that the model can be integrated with plant metabolic models in order to investigate the interaction between the diazotrophic organisms and plants.

# REFERENCES


1. United Nations Department of Economic and Social Affairs, Population Division World Population Prospects: The 2017 revision, Volume 2. 2017.

2. Carvalho TL, Pires E, Saraiva R, Vargas L, Bomfim ACJ, Ballesteros H, Baldani JI, Ferreira PC, Hemerly AS. Nitrogen fixation in grasses-gluconacetobacter activates genes in sugarcane. *InBMC proceedings, 2014 The 5th Congress of the Brazilian Biotechnology Society (SBBIOTEC) on;* 2014: BioMed Central.

3. Da Silva JF, Williams RJP. *The Biological chemistry of the elements: the inorganic chemistry of life*. New York: Oxford University Press; 2001.

4. Reece JB, Urry LA, Cain ML, Wasserman SA, Minorsky PV, Jackson RB. *Campbell biology*. Boston: Pearson; 2014.

5. Finzi AC, Austin AT, Cleland EE, Frey SD, Houlton BZ, Wallenstein MD. Responses and feedbacks of coupled biogeochemical cycles to climate change: examples from terrestrial ecosystems. *Frontiers in Ecology and the Environment*. 2011;9(1):61–67.

6. Galloway JN, Dentener FJ, Capone DG, Boyer EW, Howarth RW, Seitzinger SP, Asner GP, Cleveland CC, Green P, Holland EA. Nitrogen cycles: past, present, and future. *Biogeochemistry*. 2004;70(2):153–226.

7. Pérez-Montaño F, Alías-Villegas C, Bellogín R, Del Cerro P, Espuny M, Jiménez-Guerrero I, López-Baena FJ, Ollero F, Cubo T. Plant growth promotion in cereal and leguminous agricultural important plants: from microorganism capacities to crop production. *Microbiological Research*. 2014;169(5-6):325–336.

8. Dent D. Non-nodular endophytic bacterial symbiosis and the nitrogen fixation of Gluconacetobacter diazotrophicus. *Symbiosis*. 2018:1–29.

9. Berg J, Tymoczko J, Stryer L, Stryer L. *Biochemistry, Ed 5th*. New York: WH

Freeman; 2002.

10. Mylona P, Pawlowski K, Bisseling T. Symbiotic nitrogen fixation. *The Plant Cell*. 1995;7(7):869–885.

11. Eskin N, Vessey K, Tian L. Research progress and perspectives of nitrogen fixing bacterium, Gluconacetobacter diazotrophicus, in monocot plants. *International Journal of Agronomy*. 2014;1:1-13.

12. Flores-Encarnacion M, Contreras-Zentella M, Soto-Urzua L, Aguilar G, Baca B, Escamilla J. The respiratory system and diazotrophic activity of Acetobacter diazotrophicus PAL5. *Journal of Bacteriology*. 1999;181(22):6987–6995.

13. Hardy R, Burns RC, Holsten RD. Applications of the acetylene-ethylene assay for measurement of nitrogen fixation. *Soil Biology and Biochemistry*. 1973;5(1):47–81.

14. Glick BR. Plant growth-promoting bacteria: mechanisms and applications. *Scientifica*. 2012;1:1-15.

15. Turan M, Kıtır N, Alkaya Ü, Günes A, Tüfenkçi Ş, Yıldırım E, Nikerel E. Making soil more accessible to plants: the case of plant growth promoting rhizobacteria. *Plant Growth*. 2016;1:61–69.

16. Barea J. Future challenges and perspectives for applying microbial biotechnology in sustainable agriculture based on a better understanding of plant-microbiome interactions. *Journal of Soil Science and Plant Nutrition*. 2015;15(2):261–282.

17. Antoun H, Kloepper J. Plant growth promoting rhizobacteria. *Encyclopedia of Genetics*. 2013:5:1477–1480.

18. Cavalcante VA, Dobereiner J. A new acid-tolerant nitrogen-fixing bacterium associated with sugarcane. *Plant and Soil*. 1988;108(1):23–31.

19. Gillis M, Kersters K, Hoste B, Janssens D, Kroppenstedt R, Stephan M, Teixeira

KdS, Dobereiner J, De Ley J. Acetobacter diazotrophicus sp. nov., a nitrogen-fixing acetic acid bacterium associated with sugarcane. *International Journal of Systematic and Evolutionary Microbiology*. 1989;39(3):361–364.

20. Yamada Y, Yukphan P, Vu HTL, Muramatsu Y, Ochaikul D, Nakagawa Y. Subdivision of the genus Gluconacetobacter Yamada, Hoshino and Ishikawa 1998: the proposal of Komagatabacter gen. nov., for strains accommodated to the Gluconacetobacter xylinus group in the $\alpha$-Proteobacteria. *Annals of Microbiology*. 2012;62(2):849–859.

21. Muthukumarasamy R, Revathi G, Seshadri S, Lakshminarasimhan C. Gluconacetobacter diazotrophicus (syn. Acetobacter diazotrophicus), a promising diazotrophic endophyte in tropics. *Current Science*. 2002;25:137–145.

22. Chawla N, Phour M, Suneja S, Sangwaan S, Goyal S. Gluconacetobacter diazotrophicus: An overview. *Research in Environment and Life Sciences*. 2014; 7(1):1–10.

23. Stephan M, Oliveira M, Teixeira K, Martinez-Drets G, Döbereiner J. Physiology and dinitrogen fixation of Acetobacter diazotrophicus. *FEMS Microbiology Letters*. 1991;77(1):67–72.

24. Cocking EC, Stone PJ, Davey MR. Intracellular colonization of roots of Arabidopsis and crop plants by Gluconacetobacter diazotrophicus. *In Vitro Cellular & Developmental Biology-Plant*. 2006;42(1):74–82.

25. Giongo A, Tyler HL, Zipperer UN, Triplett EW. Two genome sequences of the same bacterial strain, Gluconacetobacter diazotrophicus PAl 5, suggest a new standard in genome sequence submission. *Standards in Genomic Sciences*. 2010;2(3):309–317.

26. Bertalan M, Albano R, de Pádua V, Rouws L, Rojas C, Hemerly A, Teixeira K, Schwab S, Araujo J, Oliveira A. Complete genome sequence of the sugarcane nitrogen-fixing endophyte Gluconacetobacter diazotrophicus Pal5. *BMC Genomics*.

2009;10(1):450.

27. Cojho EH, Reis VM, Schenberg ACG, Döbereiner J. Interactions of Acetobacter diazotrophicus with an amylolytic yeast in nitrogen-free batch culture. *FEMS Microbiology Letters*. 1993;106(3):341–346.

28. Pedraza RO. Recent advances in nitrogen-fixing acetic acid bacteria. *International Journal of Food Microbiology*. 2008;125(1):25–35.

29. Pan B, Vessey JK. Response of the Endophytic Diazotroph Gluconacetobacter diazotrophicus on Solid Media to Changes in Atmospheric Partial O2 Pressure. *Applied Environmental Microbiology* 2001;67(10):4694–4700.

30. Cocking EC. OBPC Symposium: maize 2004 & beyond—intracellular colonization of cereals and other crop plants by nitrogen-fixing bacteria for reduced inputs of synthetic nitrogen fertilizers. *In Vitro Cellular & Developmental Biology-Plant*. 2005;41(4):369–373.

31. Luna MF, Aprea J, Crespo JM, Boiardi JL. Colonization and yield promotion of tomato by Gluconacetobacter diazotrophicus. *Applied Soil Ecology*. 2012;61:225–229.

32. Saravanan V, Madhaiyan M, Osborne J, Thangaraju M, Sa T. Ecological occurrence of Gluconacetobacter diazotrophicus and nitrogen-fixing Acetobacteraceae members: their possible role in plant growth promotion. *Microbial Ecology*. 2008; 55(1):130–140.

33. Fuentes-Ramirez L, Jimenez-Salgado T, Abarca-Ocampo I, Caballero-Mellado J. Acetobacter diazotrophicus, an indoleacetic acid producing bacterium isolated from sugarcane cultivars of Mexico. *Plant and Soil*. 1993;154(2):145–150.

34. Bastián F, Cohen A, Piccoli P, Luna V, Bottini R, Baraldi R. Production of indole-3-acetic acid and gibberellins A1 and A3 by Acetobacter diazotrophicus

and Herbaspirillum seropedicae in chemically-defined culture media. *Plant Growth Regulation*. 1998;24(1):7–11.

35. Lee S, Flores-Encarnacion M, Contreras-Zentella M, Garcia-Flores L, Escamilla J, Kennedy C. Indole-3-acetic acid biosynthesis is deficient in Gluconacetobacter diazotrophicus strains with mutations in cytochrome c biogenesis genes. *Journal of Bacteriology*. 2004;186(16):5384–5391.

36. Rodrigues EP, Soares CdP, Galvao PG, Imada EL, Simoes-Araujo JL, Rouws LF, de Oliveira AL, Vidal MS, Baldani JI. Identification of Genes Involved in Indole-3-Acetic Acid Biosynthesis by Gluconacetobacter diazotrophicus PAL5 Strain Using Transposon Mutagenesis. *Frontiers in Microbiology*. 2016;7:1572-1583.

37. Frank AC. The Genomes of Endophytic Bacteria. *Endophytes of Forest Trees*. 2018: 1:141–176.

38. Crespo J, Boiardi J, Luna M. Mineral phosphate solubilization activity of Gluconacetobacter diazotrophicus under P-limitation and plant root environment. *Agricultural Sciences*. 2011;2(01):16-22.

39. Delaporte-Quintana P, Grillo-Puertas M, Lovaisa NC, Teixeira KR, Rapisarda VA, Pedraza RO. Contribution of Gluconacetobacter diazotrophicus to phosphorus nutrition in strawberry plants. *Plant and Soil*. 2017;419(1-2):335–347.

40. Saravanan V, Madhaiyan M, Thangaraju M. Solubilization of zinc compounds by the diazotrophic, plant growth promoting bacterium Gluconacetobacter diazotrophicus. *Chemosphere*. 2007;66(9):1794–1798.

41. Sarathambal C, Thangaraju M, Paulraj C, Gomathy M. Assessing the Zinc solubilization ability of Gluconacetobacter diazotrophicus in maize rhizosphere using labelled 65 Zn compounds. *Indian Journal of Microbiology*. 2010;50(1):103–109.

42. Souza Rd, Ambrosini A, Passaglia LM. Plant growth-promoting bacteria as inoculants in agricultural soils. *Genetics and Molecular Biology*. 2015;38(4):401–419.

43. Saravanan VS. An investigation on solubilization of zinc and certain other insoluble compounds by Gluconacetobacter diazotrophicus [dissertation]. Caimbatore: Tamil Nadu Agricultural University; 2004.

44. Pinon D, Casas M, Blanch M, Fontaniella B, Blanco Y, Vicente C, Solas MT, Legaz ME. Gluconacetobacter diazotrophicus, a sugar cane endosymbiont, produces a bacteriocin against Xanthomonas albilineans, a sugar cane pathogen. *Research in Microbiology*. 2002;153(6):345–351.

45. Arencibia AD, Vinagre F, Estevez Y, Bernal A, Perez J, Cavalcanti J, Santana I, Hemerly AS. Gluconoacetobacter diazotrophicus elicitate a sugarcane defense response against a pathogenic bacteria Xanthomonas albilineans. *Plant Signaling & Behavior*. 2006;1(5):265–273.

46. Chawla N, Anand R, Narula N. Colonization behaviour of Gluconoacetobacter diazotrophicus in root-knot nematode (Meloidogyne incognita) infected and healthy cotton plants. *International Journal of Microbial Resources & Technology*. 2013; 2:24–35.

47. Nieto-Peñalver CG, Savino MJ, Bertini EV, Sánchez LA, de Figueroa LI. Gluconic acid produced by Gluconacetobacter diazotrophicus Pal5 possesses antimicrobial properties. *Research in Microbiology*. 2014;165(7):549–558.

48. Alvarez B, Martínez-Drets G. Metabolic characterization of Acetobacter diazotrophicus. *Canadian Journal of Microbiology*. 1995;41(10):918–924.

49. Attwood MM, Van Dijken JP, Pronk JT. Glucose metabolism and gluconic acid production by Acetobacter diazotrophicus. *Journal of Fermentation and Bioengineering*. 1991;72(2):101–105.

50. Luna M, Mignone C, Boiardi J. The carbon source influences the energetic efficiency of the respiratory chain of N 2-fixing Acetobacter diazotrophicus. *Applied Microbiology and Biotechnology*. 2000;54(4):564–569.

51. Luna MF. Caracteristicas fisiologicas de Gluconacetobacter diazotrophicus [dissertation]. La Plata: Universidad Nacional de La Plata; 2003.

52. Gómez-Manzo S, Escamilla JE, González-Valdez A, López-Velázquez G, Vanoye-Carlo A, Marcial-Quino J, de la Mora-de la Mora I, Garcia-Torres I, Enríquez-Flores S, Contreras-Zentella ML. The oxidative fermentation of ethanol in Gluconacetobacter diazotrophicus is a two-step pathway catalyzed by a single enzyme: alcohol-aldehyde dehydrogenase (ADHa). *International Journal of Molecular Sciences*. 2015;16(1):1293–1311.

53. Ureta A, Alvarez B, Ramon A, Vera M, Martinez-Drets G. Indentification of Acetobacter diazotrophicus, Herbaspirillum seropedicae and Herbaspirillum rubrisubalbicans using biochemical and genetic criteria. *Plant and Soil*. 1995; 172(2):271–277.

54. Luna MF, Bernardelli CE, Galar ML, Boiardi JL. Glucose metabolism in batch and continuous cultures of Gluconacetobacter diazotrophicus PAL 3. *Current Microbiology*. 2006;52(3):163–168.

55. Luna MF, Bernardelli CE, Mignone CF, Boiardi JL. Energy generation by extracellular aldose oxidation in N2-fixing Gluconacetobacter diazotrophicus. *Applied and Environmental Microbiology*. 2002;68(4):2054–2056.

56. Galar M, Boiardi J. Evidence for a membrane-bound pyrroloquinoline quinone-linked glucose dehydrogenase in Acetobacter diazotrophicus. *Applied Microbiology and Biotechnology*. 1995;43(4):713–716.

57. Luna MF, Boiardi JL. Growth yields and glucose metabolism of N2-fixing Gluconacetobacter diazotrophicus at different culture pH values. *World Journal of*

*Microbiology and Biotechnology*. 2008;24(4):587–590.

58. Fischer HM. Genetic regulation of nitrogen fixation in rhizobia. *Microbiological Reviews*. 1994;58(3):352–386.

59. Alquéres SM, Oliveira JHM, Nogueira EM, Guedes HV, Oliveira PL, Câmara F, Baldani JI, Martins OB. Antioxidant pathways are up-regulated during biological nitrogen fixation to prevent ROS-induced nitrogenase inhibition in Gluconacetobacter diazotrophicus. *Archives of Microbiology*. 2010;192(10):835–841.

60. Canete-Rodriguez AM, Santos-Duenas IM, Jimenez-Hornero JE, Ehrenreich A, Liebl W, Garcia-Garcia I. Gluconic acid: Properties, production methods and applications—An excellent opportunity for agro-industrial by-products and waste bio-valorization. *Process Biochemistry*. 2016;51(12):1891–1903.

61. Ramachandran S, Fontanille P, Pandey A, Larroche C. Gluconic acid: Properties, applications and microbial production. *Food Technology & Biotechnology*. 2006; 44(2):185–195.

62. Singh OV, Kumar R. Biotechnological production of gluconic acid: future implications. *Applied Microbiology and Biotechnology*. 2007;75(4):713–722.

63. Arrieta JG, Sotolongo M, Menéndez C, Alfonso D, Trujillo LE, Soto M, Ramírez R, Hernández L. A type II protein secretory pathway required for levansucrase secretion by Gluconacetobacter diazotrophicus. *Journal of Bacteriology*. 2004;186(15):5031–5039.

64. Velázquez-Hernández ML, Baizabal-Aguirre VM, Cruz-Vázquez F, Trejo-Contreras MJ, Fuentes-Ramírez LE, Bravo-Patiño A, Cajero-Juárez M, Chávez-Moctezuma MP, Valdez-Alarcón JJ. Gluconacetobacter diazotrophicus levansucrase is involved in tolerance to NaCl, sucrose and desiccation, and in biofilm formation. *Archives of Microbiology*. 2011;193(2):137–149.

65. Hernández L, Ramírez R, Hormaza J, Madrazo J, Arrieta J. Increased levansucrase production by a genetically modified Acetobacter diazotrophicus strain in shaking batch cultures. *Letters in Applied Microbiology*. 1999;28(1):41–44.

66. Tomulescu C, Stoica R, Sevcenco C, Căşărică A, Moscovici M, Vamanu A. Levan-a mini review. *Scientific Bulletin Series Fermentation Biotechnologies*. 2016;20:309–317.

67. Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. *Molecular Systems Biology*. 2007; 3(1):119–133.

68. Radde NE, Hütt MT. The Physics behind Systems Biology. *EPJ Nonlinear Biomedical Physics*. 2016;4(1):7–25.

69. Aon M, Cortassa S. Systems biology of the fluxome. *Processes*. 2015;3(3):607–618.

70. Bordbar A, Monk JM, King ZA, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. *Nature Reviews Genetics*. 2014; 15(2):107–120.

71. Price ND, Reed JL, Palsson BØ. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology*. 2004;2(11):886-897.

72. Chen PW, Theisen MK, Liao JC. Metabolic systems modeling for cell factories improvement. *Current Opinion in Biotechnology*. 2017;46:114–119.

73. Durot M, Bourguignon PY, Schachter V. Genome-scale models of bacterial metabolism: reconstruction and applications. *FEMS Microbiology Reviews*. 2008; 33(1):164–190.

74. Baart GJ, Martens DE. Genome-scale metabolic models: reconstruction and analysis. *Neisseria Meningitidis*. 2012;107–126.

75. O'Brien EJ, Monk JM, Palsson BO. Using genome-scale models to predict biological capabilities. *Cell*. 2015;161(5):971–987.

76. Banos DT, Trébulle P, Elati M. Integrating transcriptional activity in genome-scale models of metabolism. *BMC Systems Biology*. 2017;11(7):134–143.

77. Liu L, Agren R, Bordel S, Nielsen J. Use of genome-scale metabolic models for understanding microbial physiology. *FEBS Letters*. 2010;584(12):2556–2564.

78. Stalidzans E, Seiman A, Peebo K, Komasilovs V, Pentjuss A. Model-based metabolism design: constraints for kinetic and stoichiometric models. *Biochemical Society Transactions*. 2018;46(2):261–267.

79. Kauffman KJ, Prakash P, Edwards JS. Advances in flux balance analysis. *Current Opinion in Biotechnology*. 2003;14(5):491–496.

80. Nikerel E, van Winden WA, Verheijen PJ, Heijnen JJ. Model reduction and a priori kinetic parameter identifiability analysis using metabolome time series for metabolic reaction networks with linlog kinetics. *Metabolic Engineering*. 2009;11(1):20–30.

81. Thiele I, Palsson BØ. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*. 2010;5(1):93–121.

82. Beck AE, Hunt KA, Carlson RP. Measuring cellular biomass composition for computational biology applications. *Processes*. 2018;6(5):38–64.

83. Blazier AS, Papin JA. Integration of expression data in genome-scale metabolic network reconstructions. *Frontiers in Physiology*. 2012;3:299–306.

84. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nature Biotechnology*. 2010;28(3):245–248.

85. Maia P, Rocha M, Rocha I. In silico constraint-based strain optimization methods: the quest for optimal cell factories. *Microbiology and Molecular Biology Reviews*.

2016;80(1):45–67.

86. Banga JR. Optimization in computational systems biology. *BMC Systems Bbiology*. 2008;2(1):47–67.

87. McCloskey D, Palsson BØ, Feist AM. Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli. *Molecular Systems Biology*. 2013;9(1):661–675.

88. Kim B, Kim WJ, Kim DI, Lee SY. Applications of genome-scale metabolic network model in metabolic engineering. *Journal of Industrial Microbiology & Biotechnology*. 2015;42(3):339–348.

89. Feist AM, Palsson BO. The biomass objective function. *Current Opinion in Microbiology*. 2010;13(3):344–349.

90. Gianchandani EP, Lee JM, Papin JA. Flux balance analysis in the era of metabolomics. *Briefings in Bioinformatics*. 2006;7(2):140–150.

91. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research*. 2013; 42(D1):D206–D214.

92. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics*. 2008;9(1):75–89.

93. Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, Olson R, Overbeek R, Parrello B, Pusch GD. RASTtk: a modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Scientific Reports*. 2015;5:8365–8370.

94. Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B, Stevens RL. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature Biotechnology*. 2010;28(9):977–982.

95. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*. 2016;45(D1):D353–D361.

96. Jeske L, Placzek S, Schomburg I, Chang A, Schomburg D. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Research*. 2018;47(D1):D542–D549.

97. Keseler IM, Mackie A, Santos-Zavaleta A, Billington R, Bonavides-Martínez C, Caspi R, Fulcher C, Gama-Castro S, Kothari A, Krummenacker M. The EcoCyc database: reflecting new knowledge about Escherichia coli K-12. *Nucleic Acids Research*. 2016;45(D1):D543–D550.

98. Geer LY, Marchler-Bauer A, Geer RC, Han L, He J, He S, Liu C, Shi W, Bryant SH. The NCBI biosystems database. *Nucleic Acids Research*. 2009;38(suppl_1):D492–D496.

99. What is Matlab?: MathWorks; [cited 2019 22 June]. Available from: https://uk.mathworks.com/discovery/what-is-matlab.html.

100. Linear Programming Kit Version 2.7: GNU; [cited 2019 22 June]. Avaliable from: https://www.gnu.org/software/glpk/glpk.html.

101. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, Haraldsdóttir HS, Wachowiak J, Keating SM, Vlasov V. Creation and analysis of biochemical constraint-based models: the COBRA Toolbox v3. 0. *arXiv Preprint arXiv:1710.04038*. 2017;1–78.

102. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search

tool. *Journal of Molecular Biology*. 1990;215(3):403–410.

103. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BØ. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*. 2007;3(1):1–18.

104. Lery LM, Coelho A, von Kruger WM, Gonçalves MS, Santos MF, Valente RH, Santos EO, Rocha SL, Perales J, Domont GB. Protein expression profile of Gluconacetobacter diazotrophicus PAL5, a sugarcane endophytic plant growth-promoting bacterium. *Proteomics*. 2008;8(8):1631–1644.

105. Zhang H, Ye C, Xu N, Chen C, Chen X, Yuan F, Xu Y, Yang J, Sun D. Reconstruction of a genome-scale metabolic network of Komagataeibacter nataicola RZS01 for cellulose production. *Scientific Reports*. 2017;7(1):7911-7919.

106. Caballero-Mellado J, Martinez-Romero E. Limited genetic diversity in the endophytic sugarcane bacterium Acetobacter diazotrophicus. *Applied and Environmental Microbiology*. 1994;60(5):1532–1537.

107. Duarte NC, Herrgård MJ, Palsson BØ. Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. *Genome Research*. 2004;14(7):1298–1309.

# APPENDIX A: MODEL iUA484 RELATED TABLE

Table A.1. Blastp results of added reactions

| ADDED RXN ID | PROTEIN COVERAGE (%) | E-VALUE | PER. IDENT (%) |
|---|---|---|---|
| rxn01256 | 24 | 6E-28 | 35.48 |
| rxn00416 | 54 | 6.00E-08 | 17.05 |
| rxn01644 | 100 | 1.00E-144 | 46.08 |
| rxn02928 | 99 | 7E-100 | 40.07 |
| rxn02929 | 99 | 7E-100 | 40.07 |
| rxn03087 | 96 | 3.00E-174 | 40.15 |
| rxn02160 | 54 | 2E-124 | 48.99 |
| rxn23454 | 89 | 2E-76 | 17.84 |
| rxn23455 | 89 | 2E-76 | 17.84 |
| rxn03452 | 89 | 2E-76 | 17.84 |
| rxn01490 | 89 | 2E-76 | 17.84 |
| rxn00701 | 26 | 1E-13 | 20.20 |
| rxn00212 | 37 | 3E-10 | 11.32 |
| rxn09028 | 87 | 1E-41 | 14.25 |
| rxn09029 | * | * | * |
| rxn00213 | 100 | 3E-69 | 40.00 |
| rxn00483 | 98 | 5E-107 | 31.60 |
| rxn00477 | 89 | 2E-147 | 25.19 |
| R10180 | 91 | 9E-67 | 17.53 |
| rxn01937 | 96 | 0 | 68.29 |
| rxn11964 | 51 | 0 | 29.74 |
| rxn05329 | 99 , 99, 97 | 1E-84, 1E-84, 4E-59 | 45.79, 45.79, 45.58 |
| rxn05330 | 99 , 99, 97 | 1E-84, 1E-84, 4E-59 | 45.79, 45.79, 45.58 |
| rxn05331 | 99 , 99, 97 | 1E-84, 1E-84, 4E-59 | 45.79, 45.79, 45.58 |
| rxn05332 | 99 , 99, 97 | 1E-84, 1E-84, 4E-59 | 45.79, 45.79, 45.58 |

*Continued on next page*

Table A.1. *Blastp results of added reactions (continued)*

| ADDED RXN ID | PROTEIN COVERAGE | E-VALUE | PER. IDENT |
|---|---|---|---|
| rxn05333 | 99 , 99, 97 | 1E-84, 1E-84, 4E-59 | 45.79, 45.79, 45.58 |
| rxn05334 | 99 , 99, 97 | 1E-84, 1E-84, 4E-59 | 45.79, 45.79, 45.58 |
| rxn05335 | 99 , 99, 97 | 1E-84, 1E-84, 4E-59 | 45.79, 45.79, 45.58 |
| rxn06556 | 84 | 1E-17 | 14.80 |
| rxn08434 | 91 | 2E-39 | 33.67 |
| rxn03249 | 28, 27 | 4E-59, 3E-65 | 28.50, 29.80 |
| rxn03246 | 28, 27 | 4E-59, 3E-65 | 28.50, 29.80 |
| rxn03244 | 28, 27 | 4E-59, 3E-65 | 28.50, 29.80 |
| rxn03242 | 28, 27 | 4E-59, 3E-65 | 28.50, 29.80 |
| rxn06777 | 28, 27 | 4E-59, 3E-65 | 28.50, 29.80 |
| rxn03239 | 28, 27 | 4E-59, 3E-65 | 28.50, 29.80 |
| rxn03250 | 98 | 1E-78 | 32.55 |
| rxn00943 | 62 | 3E-16 | 18.18 |
| rxn04793 | 98, 86 | 6E-92, 1E-24 | 21.54, 18.34 |
| rxn02678 | 98, 86 | 6E-92, 1E-24 | 21.54, 18.34 |
| rxn03252 | 98, 86 | 6E-92, 1E-24 | 21.54, 18.34 |
| rxn02719 | 98, 86 | 6E-92, 1E-24 | 21.54, 18.34 |
| rxn02802 | 98, 86 | 6E-92, 1E-24 | 21.54, 18.34 |
| rxn00945 | 98, 86 | 6E-92, 1E-24 | 21.54, 18.34 |
| rxn00947 | 96 | 6.00E-88 | 17.85 |
| rxn01409 | 4 | 0.53 | 32.14 |
| rxn01451 | 28 | 4.00E-59 | 28.50 |
| rxn00946 | 82 | 9.00E-16 | 14.47 |
| rxn02803 | 86 | 2.00E-19 | 17.10 |
| rxn02720 | 41 | 6.00E-12 | 15.06 |
| rxn03253 | 55 | 6.00E-19 | 12.68 |
| rxn02679 | 41 | 6.00E-12 | 15.06 |
| rxn03251 | 41 | 6.00E-12 | 15.06 |

*Continued on next page*

Table A.1. *Blastp results of added reactions (continued)*

| ADDED RXN ID | PROTEIN COVERAGE | E-VALUE | PER. IDENT |
|---|---|---|---|
| rxn00872 | 82 | 9E-16 | 14.47 |
| R10619 | 97 | 3E-130 | 37.07 |
| rxn00808 | 42 | 0.066 | 25.00 |
| rxn00701 | 26 | 1E-13 | 20.00 |
| rxn02061 | 98 | 4E-106 | 46.15 |
| rxn03886 | 99 | 2E-86 | 25.10 |
| rxn00500 | 96 | 5E-77 | 40.33 |
| rxn10112 | 90 | 5E-48 | 37.43 |
| rxn30498 | * | * | * |
| rxn13848 | 90 | 4E-48 | 37.43 |
| rxn13930 | 59 | 8E-15 | 17.36 |
| rxn06106 | * | * | * |
| rxn08973 | * | * | * |
| rxn08978 | 100 | 7E-99 | 30.58 |
| rxn30496 | * | * | * |
| rxn13835 | 98 | 1E-107 | 37.61 |
| rxn00104 | 99 | 0 | 79.68 |

* No enzyme/Gene information for the given reaction

** Multiple values means, the reaction has more than one enzyme catalyzing it