**T.R.**

**GEBZE TECHNICAL UNIVERSITY**

**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES**

# SYSTEMS BIOLOGY APPROACH TO INVESTIGATE THE EFFECT OF NITROGEN AND PHOSPHATE METABOLISM ON ACTINORHODIN PRODUCTION BY STREPTOMYCES COELICOLOR USING GENOME SCALE METABOLIC MODELS

**HAMZA UMUT KARAKURT**

**A THESIS SUBMITTED FOR THE DEGREE OF**

**MASTER OF SCIENCE**

**DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS**

**GEBZE**

**2016**

**T.R.**

**GEBZE TECHNICAL UNIVERSITY**

**GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES**

# SYSTEMS BIOLOGY APPROACH TO INVESTIGATE THE EFFECT OF NITROGEN AND PHOSPHATE METABOLISM ON ACTINORHODIN PRODUCTION BY *STREPTOMYCES COELICOLOR* USING GENOME SCALE METABOLIC MODELS

**HAMZA UMUT KARAKURT**

**A THESIS SUBMITTED FOR THE DEGREE OF**

**MASTER OF SCIENCE**

**DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS**

THESIS SUPERVISOR

PROF. DR. SEDEF TUNCA GEDIK

II. THESIS SUPERVISOR

ASSIST. PROF. DR. TUNAHAN CAKIR

**GEBZE**

**2016**

**T.C.**
**GEBZE TEKNİK ÜNİVERSİTESİ**
**FEN BİLİMLERİ ENSTİTÜSÜ**

# *STREPTOMYCES COELICOLOR*'DA AZOT VE FOSFAT METABOLİZMASININ AKTİNORHODİN ÜRETİMİNE ETKİSİNİN GENOM-ÖLÇEKLİ METABOLİK MODELLER KULLANILARAK SİSTEM BİYOLOJİSİ UYGULAMALARI İLE İNCELENMESİ

**HAMZA UMUT KARAKURT**
**YÜKSEK LİSANS TEZİ**
**MOLEKÜLER BİYOLOJİ VE GENETİK ANABİLİM DALI**

TEZ DANIŞMANI
PROF. DR. SEDEF TUNCA GEDİK
II. TEZ DANIŞMANI
YRD. DOÇ. DR. TUNAHAN ÇAKIR

**GEBZE**
**2016**

| GEBZE TEKNİK ÜNİVERSİTESİ | YÜKSEK LİSANS JÜRİ ONAY FORMU |
|---|---|

GTÜ Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 22/06/2016 tarih ve 2016/41 sayılı kararıyla oluşturulan jüri tarafından 20/07/2016 tarihinde tez savunma sınavı yapılan Hamza Umut KARAKURT'ın tez çalışması Moleküler Biyoloji ve Genetik Anabilim Dalında YÜKSEK LİSANS tezi olarak kabul edilmiştir.

**JÜRİ**

| | |
|---|---|
| ÜYE (TEZ DANIŞMANI) | : Prof. Dr. Sedef TUNCA GEDIK |
| ÜYE | : Yrd. Doç. Dr. Emrah NİKEREL |
| ÜYE | : Yrd. Doç. Dr. Saliha İŞSEVER ÖZTÜRK |

**ONAY**

Gebze Teknik Üniversitesi Fen Bilimleri Enstitüsü Yönetim Kurulu'nun ...../...../..... tarih ve ...../..... sayılı kararı.

İMZA/MÜHÜR

# SUMMARY

*Streptomyces* species are soil-dwelling, gram positive, filamentous bacteria with a high G+C content genome. They have very complex regulatory systems and also metabolic switches that change the metabolism between primary and secondary metabolism. Medically and commercially important compounds are produced during the secondary metabolism, which is often observed when the culture has entered the stationary phase in a batch culture growth system. Complex regulatory network of *Streptomyces* species can be simplified with different approaches of systems biology (genome scale metabolic models, omics technologies etc.) and the results can suggest different ways of increasing production levels of valuable metabolites. In this study, a genome scale metabolic network of *S. coelicolor* was integrated with three different transcriptome data sets from the public Gene Expression Omnibus database: time dependent data of ΔphoP mutant, ΔargR mutant and wild type strain. The dynamic data spanned both primary and secondary phases of the metabolism. Statistical results of transcriptome data were used for reporter metabolite analysis and reporter pathway analysis, which identify the metabolites (or pathways) with a significant coordinated transcriptional change in response to gene deletion perturbation in phosphate and nitrogen metabolisms. Further, the production of actinorhodin, a pharmaceutically important compound, was modeled in the two deletion strains by calculating the metabolic fluxes subject to transcriptional level constraints on enzyme-coding genes. The metabolic switch from primary to secondary metabolism in *Streptomyces* was highlighted in terms of the activity of pathways and fluxes as a result of the computational analyses in this work, leading to a better understanding of the role of phosphate and nitrogen metabolisms in increasing production levels of valuable metabolites.

**Keywords: *Streptomyces, S. coelicolor,* Systems Biology, Transcriptome Data Analysis, Reporter metabolite analysis, Reporter Pathway Analysis, Metabolic fluxes.**

# ÖZET

*Streptomyces* türleri gram pozitif, filamentli ve genomları yüksek G+C oranlarına sahip toprak bakterileridir. Oldukça karmaşık regülasyon sistemlerine sahip olan bu bakterilerin, primer ve sekonder metabolizma arasındaki geçişi sağlayan metabolik anahtarlara sahip olduğu da bilinmektedir. Medikal ve ticari önemi olan metabolitler, kesikli büyüme sisteminde, kültürün durağan faza girdiği zaman başlayan sekonder metabolizma sırasında üretilirler. *Streptomyces* türlerinin karmaşık regülasyon ağları, sistem biyolojisinin farklı uygulamaları (genom ölçekli metabolik modeller, omik teknolojileri vb.) kullanılarak basitleştirilebilir ve sonuçlar değerli metabolitlerin üretim seviyelerini artırmak için farklı yollar ortaya koyabilir. Bu çalışmada, *S. coelicolor* genom-ölçekli metabolik ağ modeli, Gene Expression Omnibus veri bankasından elde edilmiş olan üç farklı transkriptom veri seti (Δ*argR* mutantı, Δ*phoP* mutantı ve yaban suşun zamana bağlı verileri) ile birlikte kullanılmıştır. Dinamik veriler, metabolizmanın hem primer hem de sekonder fazlarını içermektedir. Transkriptom verilerinin istatistiksel sonuçları, haberci metabolit ve haberci yolak analizleri için kullanılmıştır. Bu analizler, fosfat ve nitrojen metabolizmalarında düzensizlik yaratan mutasyonlara bağlı olarak belirgin koordineli transkripsiyonel değişimler gösteren metabolitlerin (ya da yolakların) belirlenmesi için kullanılmaktadır. Ayrıca, farmasötik önemi bulunan aktinorhodin üretimi, iki delesyon suşunda, enzim kodlayan genlerin transkripsiyonel düzeyde sınırlandığı metabolik akılar hesaplanarak modellenmiştir. Bu çalışmanın sonuçları, *Streptomyces*'lerde primer ve sekonder metabolizmalar arası geçişi, yolak ve akı aktiviteleriyle ilgili olarak bilgisayara dayalı analizlerle aydınlatmak suretiyle, değerli metabolitlerin üretim seviyelerini arttırmak için fosfat ve azot metabolizmalarının rolününün daha iyi anlaşılmasını sağlayacaktır.

**Anahtar Kelimeler**: ***Streptomyces coelicolor,*** **Transkriptom veri analizi, Haberci metabolit analizi, Haberci yolak analizi, Metabolik Akı.**

# ACKNOWLEDGEMENTS

First, I would like to express my deep and sincere gratitudes to my advisors, Prof. Dr. Sedef TUNCA GEDIK and Asst. Prof. Dr. Tunahan CAKIR for their support, mentorship and help. Thanks for patience and continuous support.

Also I want to thank Research assistant Emrah OZCAN for his help, supports and informations about computational techniques and to Isa YUCEL for his help on reporter pathway and metabolite analysis codes on Matlab.

I wish to express my sincere thanks to my mother for her support, patience and love.

I wish to express my sincere thanks to Prof. Kay Nieselt (Tubingen University, Germany) for her supportive information about the data sets I used and Mohammad Taqueer Alam (Cambridge University, UK) for his valuable support and information about genome scale metabolic model.

Also to my friends, I want express my thanks for their supports.

I dedicate this work to my father, Hasan Huseyin KARAKURT.

# TABLE of CONTENTS

# LIST of ABBREVIATIONS and ACRONYMS

| **Abbreviations and Acronyms** | **Explanations** |
|---|---|
| C: | : Carbon |
| N | : Nitrogen |
| P | : Phosphate |
| WT | : Wild-Type Strain |
| ACT | : Actinorhodin |
| ΔphoP | : *PhoP* Knock-Out Mutant Strain |
| ΔargR | : *ArgR* Knock-Out Mutant Strain |
| μ | : Growth Rate |
| h | : Hour |
| mmol | : Millimole |
| mRNA | : Messanger RNA |
| FBA | : Flux Balance Analysis |
| PCA | : Principle Component Analysis |
| KEGG | : Kyoto Encyclopedia of Genes and Genomes |
| COBRA | : The Constraint-based Reconstruction and Analysis |
| GEO | : Gene Expression Omnibus |
| GIMME | : Gene Activity Moderated by Metabolism and Expression |
| RMA | : Reporter Metabolite Analysis |
| RPA | : Reporter Pathway Analysis |
| BioCyc | : BioCyc Pathway/Genome Database Collection |

# LIST OF FIGURES

# LIST OF TABLES

# 1. INTRODUCTION

*Streptomyces* species are soil-dwelling bacteria that have high G+C content in their genome. They belong to one of the best studied and most diverse orders of bacterial taxonomy, *Actinomycetales* [Alam, 2011, Zhou et al., 2012]. They are the main producers of many medically and commercially important secondary metabolites like antibiotics, anti-cancer agents, anti-infectives. These gram-positive bacteria have very complex regulatory mechanisms, and systems-based elucidation of transition between primary and secondary metabolism can reveal new bioengineering approaches to increase the yield of production of pharmaceutically useful compounds [Bibb et al.,, 2005]. Their complex regulatory systems involve environmental processes and production of medically important compounds [Zhou et al., 2012]. Understanding their metabolic regulatory systems is crucial to optimize commercially important metabolites and to characterize their evolutionary interesting properties.

After the genome of first uni-cellular organism, *H. Influenzae*, was sequenced by Fleischmann et al in 1995, the number of genome sequences has rapidly increased due to the new-generation high-throughput sequencing technologies. The genome of *Streptomyces coelicolor* was sequenced in 2002 [Bentley et al., 2002]. This has opened a road to many computational biology methods. With the rapid evolution of "omics" fields like transcriptomics, proteomics and metabolomics, we can now analyze cellular systems at different conditions. The main challange is to process this huge amount of biological data to uncover the inherent knowledge.

Systems biology is a new branch of science, investigating cell as a system with the help of biological data and computational methods rather than focusing on individual cellular components. Due to the interactions between cellular components, focusing a part can hardly help to solve biological puzzle. Interaction of the related parts with other parts affects their function. It is necessary to measure many biomolecules in the cell simultaneously in order to understand the interactions between them. In the last decade, high quality technologies, called omics technologies, have enabled the researchers to measure hundreds of biomolecules (gene, transcript, protein, metabolite) in the same experiment. [Kitano, 2002].

Models of biological networks such as metabolic and signalling networks can be built to gain better insight about structures and functions of networks. Through the

genome sequences, genome scale metabolic models can be reconstructed. First genome scale model of *S. coelicolor* was constructed in 2005 [Borodina et al., 2005]. This model was used in several studies to investigate metabolism by integrating transcriptomic data and to show how the knock-out of specific genes reveals various bioengineering strategies. Other models that are more developed were reconstructed in 2010 [Alam et al., 2010] and in 2014 [Kim et al., 2014].

Among omics technologies, the most common one is transcriptomics. In transcriptomic studies, levels of all RNAs in a specific condition are identified by high-throughput technologies. Comparison can reveal the changes in transcriptome profile of a cell between conditions such as environmental changes, mutations and diseases.

In this thesis, I will use the genome-scale metabolic models of *S. coelicolor* and integrate them with specific transcriptome data of *PhoP* mutation, *ArgR* mutation and wild type dynamic response to investigate the metabolic properties. I also use these omics data in certain statistical analyses and methods that are commonly used in systems biology to reveal the metabolic properties of *S. coelicolor*. These systems biology methods are reporter metabolite and pathway analyses that reveal the most affected metabolites and pathways in certain conditions like mutations, and GIMME algorithm that creates condition specific genome scale models by changing the constraints of enzyme-coding genes of reactions via transcriptome levels.

# 2. BACKGROUND ASPECTS

## 2.1. *Streptomyces* and Related Actinomycetes

### 2.1.1. Physiology and Genomic Properties

Soil is a particularly complex and variable enviroment. *Streptomycetes* are among the most diverse soil bacteria. They are crucial in this environment because of their large variety of metabolic processes and biotransformation reactions including degradation of the remains of other organisms such as chitin and lignocellulose. This makes *Streptomycetes* one of the central organisms in carbon cycle. Unusual for bacteria, *Streptomycetes* species have a complex multicellular development and differentiation. This process involves branching, filamentous vegetative growth that gives rise to aerial hyphae [Bentley et al., 2002, Waksman et al., 1946].

*Streptomycetes* belong to one of the best studied and most diverse orders of bacterial taxonomy, Actinomycetales, of the phylum Actinobacteria. *Streptomyces* species are gram-positive bacteria with high content of G+C. They are amazingly biodiverse in terms of their genome size, ability to survive in extreme conditions, the pathogenicity and the varying capability to synthesize specific secondary metabolites [Hopwood et al., 2007]. Although they have remarkable diversity, these organisms are located in a single branch of the phylogenetic tree [Alam et al., 2010a]. Phylogenetic tree can be seen in Figure (2.1) [Zhou et al., 2011].

*Streptomyces* received much attention because of their medical and pharmaceutical properties, their importance in carbon recycling process, distinct developmental biology and complex regulatory networks. Members of genus *Streptomyces* have a complex life cycle that has been studied mostly on *S.coelicolor*. *Streptomyces* colonies show temporal and spatial control of gene expression, morphogenesis and metabolism. The lifecycle of *Streptomyces* is similar in many ways to lifecycle of filamentous fungi. They both grow as branching hyphae that form a vegetative mycelium [Shahab et al., 1994].

The complete genome sequence of *S.coelicolor* was published in 2002 [Bentley et al., 2002]. Genomes of 9 strains of *Streptomycetes* have been sequenced since. Each of the completed genome has its own unique biological importance and properties.

Despite the diverse physiological features, their genomes are relatively conserved. Their chromosomes are linear, G+C rich, and sizes vary from 6.5 Mbp to 11.9 Mbp. Since most of the bacteria have circular chromosomes, the linearity of chromosomes in *Streptomyces* is a unique property. Chromosomes of *Streptomycetes* are highly unstable and undergo rearrangements including DNA deletions and amplifications [Harrison et al., 2014].

Bacteria of the order actinomycetes are the most important producers of bioactive natural compounds including medically important compounds such as antibacterials, antifungals, antivirals and anticancer agents [Keulen et al., 2014].

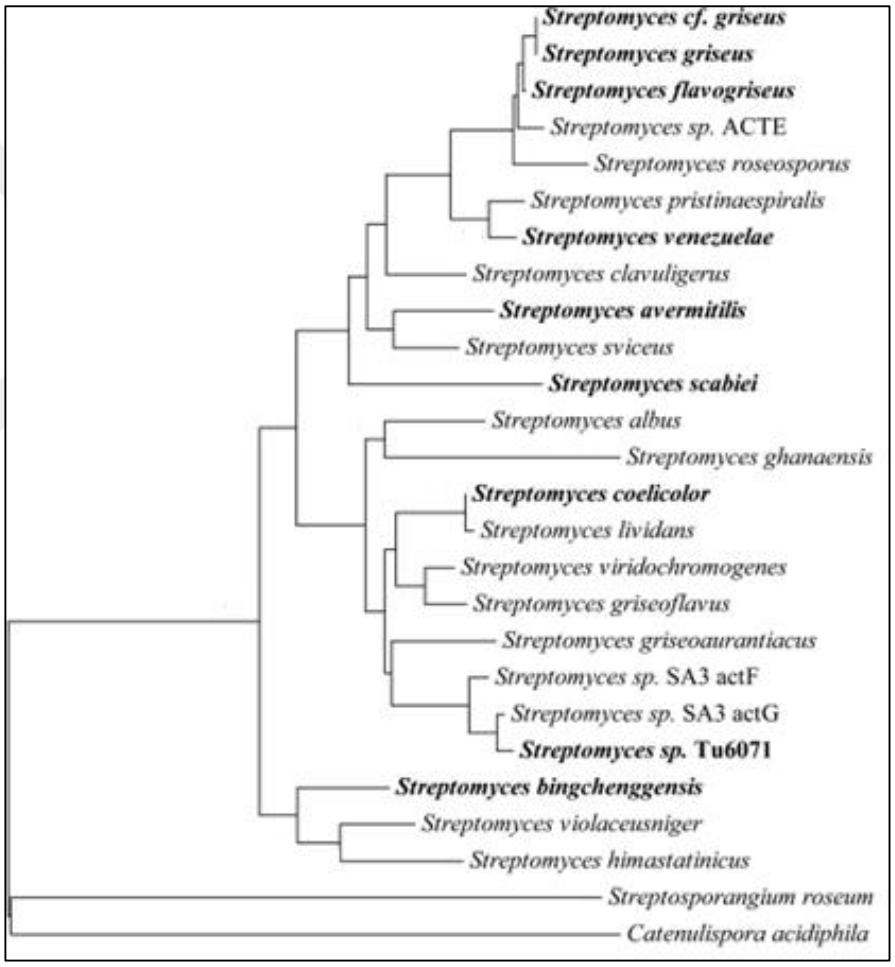


Figure 2.1: Phylogenetic tree of 16S ribosomal RNA of genome sequenced *Streptomyces.*

*S. coelicolor* is the genetically best known member of the genus. It has a single linear chromosome with a centrally located origin of replication. The genome has 8.667.507 bp and it is one of the largest complete genome for a bacteria.Almost all unconditionally essential genes such as genes that encode proteins involved in cell

division are located in the core of the genome.With 7.825 predicted genes, *S. coelicolor* genome has a great coding potential. 12.3% of the proteins (965) are predicted to have regulatory functions and the genome has about 65 sigma factors. Sigma factors are proteins that are needed only for the initiation of RNA synthesis. Different sigma factors are utilized under different environmental conditions. Its large amount of sigma factors shows the complex transcriptional regulatory mechanism of *S. coelicolor* [Bentley et al., 2002]. Figure (2.2) [Bentley, 2002] shows the circular representation of the S. coelicolor chromosome. Table (2.1) [Bentley, 2002] shows the properties of *S. coelicolor* genome.

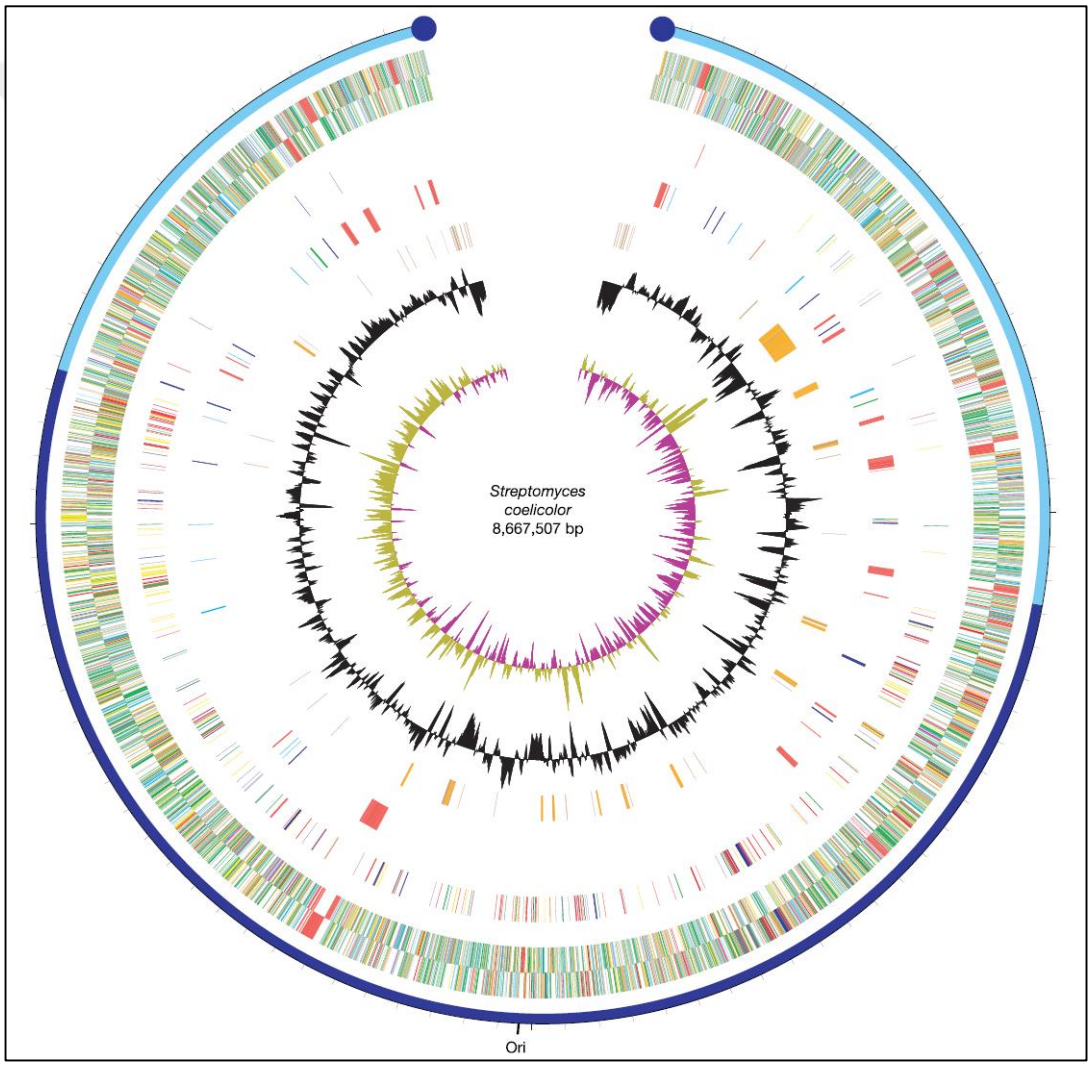


Figure 2.2: Circular representation of the *S. coelicolor* chromosome.

Table 2.1: General features of the chromosome.

| Component of Chromosome | Property |
|---|---|
| Total Size | 8.667.507 bp |
| Terminal Inverted Repeat | 21.653 bp |
| G+C Content | 72.12% |
| Coding sequences | 7.825 |
| Coding Density | 88.9% |
| Average Gene Length | 991 bp |
| Ribosomal RNAs | 6 x (16S–23S–5S) |
| Transfer RNAs | 63 |

### 2.1.2. Secondary Metabolism

The switch from primary metabolism to secondary metabolism due to nutrient limitation is found in many organisms. Understanding the switch has major importance in genetic studies in order to reveal the transciptional regulatory network and in order to optimize the production of commercially relevant secondary metabolites [Nieselt et al., 2010, Hopwood et al., 2007, Tarkka et al., 2008].

*Streptomycetes* are important sources of secondary metabolites with a range of biological activities that may find application as anti-infectives, anti-cancer agents or other pharmaceutically useful compounds. They have a complex secondary metabolism. The production of secondary metabolites generally concurs with the development of aerial hypae in surface grown cultures. In liquid cultures, it coincides with stationary phase and it is assumed to result from nutrient limitation. The genes that code the proteins needed for the production of individual secondary metabolites are arranged in clusters that vary in size from a few to 100 kb. Majority of these clusters involve pathway specific regulatory genes. Some of these genes are also needed for the formation of aerial hyphae [Bibb et al., 2005].

The metabolic versatility of *Streptomyces* species is also represented by the extensive range of classes of secondary metabolites. These metabolites include polyketides, pyrones, peptides, siderophores, γ-butyrolactones, butenolides, furans, terpenoids, fatty acids, oligopyrroles, and deoxysugars [Keulen et al., 2014].

In *Streptomycetes*, primary metabolism influences secondary metabolism by providing the metabolites, precursors and reducing agents. Primary metabolites that are mostly generated from carbon metabolism serve as building blocks for secondary metabolites. Some compounds, such as L-Valine for methylmalonyl-CoA and ethylmalonyl-CoA, are essential for both biomass and secondary metabolite biosynthesis. Consequently, secondary metabolism competes with biomass formation. Metabolic differentiation from primary metabolism to secondary metabolism decreases cellular growth. Connections and metabolic switches should be analyzed for optimal operation of secondary metabolism [Hwang et al., 2013]. The important secondary metabolites of *S. coelicolor* is shown in Figure (2.3) [Bentley et al., 2002]. In the figure, antibiotics (a), siderophores (b), pigments (c), lipids (d), other molecules (e) are grouped.

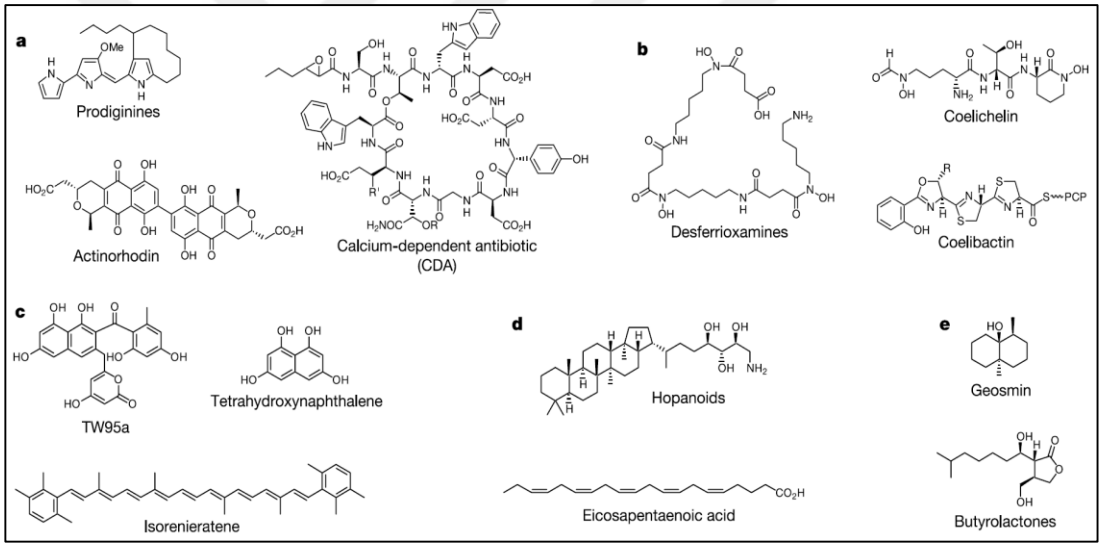


Figure 2.3: Secondary metabolites known or predicted to be produced by *S. coelicolor* A3(2), grouped according to their putative function.

## 2.1.2. Importance of Nitrogen and Phosphate in *Streptomyces coelicolor* Metabolism

Bacteria have evolved several regulatory mechanisms to respond to the nutrient exiguity. One of the universal regulatory mechanisms is the response to phosphate starvation. Phosphate is essential for metabolic, structural and regulatory activities in the cells. Reactions that involve phosphate are at the center of energy metabolism, biosynthesis of lipids, nucleic acids and cell wall components. Many microorganisms accumulate excess phosphate in the form of polyphosphate. Biologically synthesized

polyP is a linear polymer of phosphate with a chain length of up to 1000 residues or more. PolyP can serve as a phosphate source for the biosynthesis of nucleic acids and phospholipids under phosphate starvation conditions [Hirota et al., 2009].

In soil, where most of the *Streptomyces* species live, phosphate levels are low because phosphate forms complex salts with calcium and other cations. However, residuals of decaying plants and animals provide a significant phosphate source. Soil bacteria have adapted to changing amount of phosphate in the environment [Martin et al., 2012].

Increasing the concentration of phosphate downregulates the production of secondary metabolites in *Streptomyces* species [Masuma et al., 1986]. In *S. coelicolor,* regulation of phosphate metabolism is mediated by a two component system, PhoR-PhoP. PhoR is the membrane sensor protein kinase that senses phosphate limitation. PhoP is the response regulator that binds to the DNA and regulates the expression of genes that belong to the *pho* regulon. The most common members activated by the Pho regulon are: extracellular enzymes capable of obtaining inorganic phosphate (Pi) from organic phosphates, Pi-specific transporters, and enzymes involved in storage and saving of the nutrient. The Pst Pi-specific transporter is the most conserved member of the Pho regulon in all bacteria. Other Pi transporters commonly found in bacteria, such as the low affinity Pi transporter Pit, are regulated in a variable fashion manner in the different species.

PhoP also binds to glnR, the global regulator of nitrogen metabolism. PhoP regulates both primary and the secondary metabolism [Rodriguez-Garcia et al., 2009]. PhoP also regulates the promoter of amtB-glnK-glnD operon that encodes an ammonium transport and two putative nitrogen sensing/regulatory proteins.

The enzymes that catalyse the interconvention of inorganic nitrogen between different redox states lie at the heart of biogeochemical nitrogen cycle. Assimilation and utilization of nitrogen are important challanges for all kinds of bacteria. Mechanism of the regulation of nitrogen metabolism is different in every organism, but it is mainly regulated by a global regulator. In *Streptomyces*, the global regulator of nitrogen metabolism is GlnR [Pullan et al., 2011]. GlnR activates two glutamine synthetase coding genes, glnA and glnII. Also evidence was found that GlnR can regulate some ammonium assimilation genes and some genes that have not been identified yet [Tiffert et al., 2007]. Another key enzyme of nitrogen metabolism is the glutamine synthetase (GS). This enzyme has a conserved function in all organisms and

it has function in the assimilation of ammonia and the production of the amino acid glutamine through glutamate [Fink et al., 2002]. The common parts that are regulated in nitrogen and phosphate metabolism shows the complex structure of regulatory mechanisms and the regulation of phosphate metabolism over nitrogen metabolism.

Besides, some other proteins such as ArgR, an arginine biosynthesis regulator, was found to regulate nitrogen metabolism genes such as glnK, glnA, glnII [Perez-Redondo et al., 2012]. As a single amino acid regulatory system, ArgR gene can regulate not only the arginine metabolism but members of nitrogen regulation system that also are regulated by phosphate control system PhoR-PhoP directly or indirectly.

## 2.2 Systems Biology

For a long time, biologists have investigated the functions of parts of the cell through biochemistry of small and large molecules, structure of components of cell and principles of transcription and translation. Next step in this field of research is the systematic investigation of cells, organs and organisms. This approach has been termed “systems biology” [Klipp et al., 2005, Chuang et al., 2010].

“Systems biology” term has come out to describe the inter-disciplinary research in biology. This term was propelled in the mainstream about fifteen years ago in the same time with the completion of Human Genome Project and improved with the omics technologies such as transciptomics, proteomics and metabolomics [Likic et al., 2010].

Biological systems are complex systems. The typical behaviour of complex systems is the interaction of very large numbers of simple and identical elements to produce complex behaviours. In biological systems, the reality is slightly different. Biological systems have large numbers of functionally diverse, sometimes multifunctional, sets of elements. Molecular biology uncovered the biological facts such as genome sequences and protein properties, but this knowledge is not enough to predict the behaviour of biological systems [Kitano, 2002].

In systems biology research, understanding genes and proteins is still important, but understanding systems structure and dynamics is more important. Identifying all components of a biological system is like listing all parts of a machine. This kind of list provides a catalog of individual components, and it is not enough to understand the complexity and behaviour of system. Interactions of the parts need to be known

[Kitano, 2002]. Systems biology involves development of understanding of a biological system through mathematical and computational modelling of the interactions of components of the system. Computational and experimental biology have been seperate disciplines for many decades. Systems biology emerged as a new discipline and made collaboration of experimental and theoretical biology possible.

In systems biology, construction of complex systems from genes to cells by combining different databases is the dominant theme. Data itself is really important, but the way to utilize data is more important. The models that are built by using the data need to be as useful and accurate as possible. This cannot be achieved with data alone. The way the models are built is also critical as well as calculations, even the utilized softwares. This means that computational tool development in systems biology is also critical.

Systems biology profits from many fields such as bioinformatics, information technology and dynamic systems theory [Marcus et al., 2008]. Figure (2.4) [Palsson, 2006] shows the four principle steps in the implementation of systems biology. Noise in biological systems defines the cell to cell variation of gene expression levels. A typical DNA microarray requires more than 100.000 cells and measurements provide average level of expression of given genes. However the related gene can be expressed on different levels and this causes cell to cell variation or "noise". While at first glance noise may seem to be an undesired property of biological networks, it might be beneficial in some cases but in fact noise will increase functional heterogeneity in a population [Pilpel, 2011].

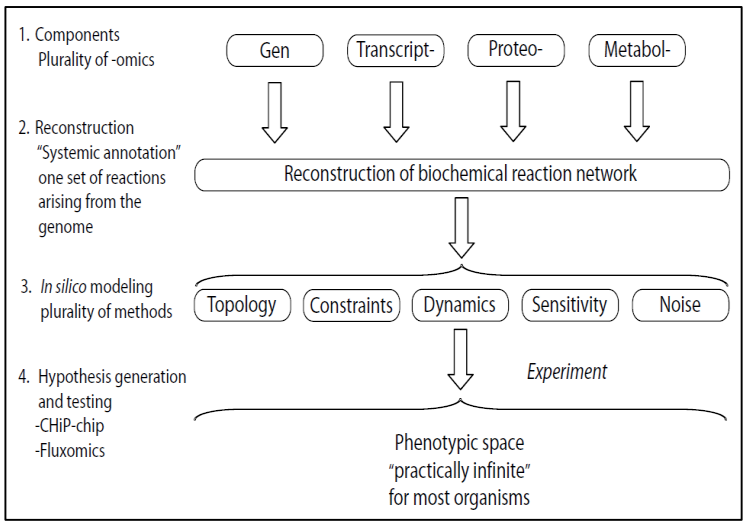


Figure 2.4: The four principal steps in the implementation of systems biology.

### 2.2.1. Omics Technologies

The word “ome” is defined by Oxford English Dictionary as “being used in cellular and molecular biology to form nouns with the sense ‘all constituents considered collectively’. The omics technologies are relatively new and the use of ome word dates back to early 1900s with the first usage of “biome”. Over the past few decades, usage of word “omics” increased with the new technologies that allow systems analysis of samples [Matthew-Bogyo et al., 2013].

Omics technologies adopt a view of the molecules in the cell, tissue or organism. Omics technologies preferentially aims global detection of genes (genomics), mRNA (transcriptomics), proteins (proteomics) and metabolites (metabolomics) in a specific biological sample in a non-targeted and non-biased manner. Basic aspect of omics technologies is that a complex system can be understood better when considered as a whole [Horgan et al., 2011]. The most common use of omics technologies are, genomics, transciptomics, proteomics and metabolomics. Figure (2.5) [Petit et al., 2012] shows the schematic representation of the five omics technologies.

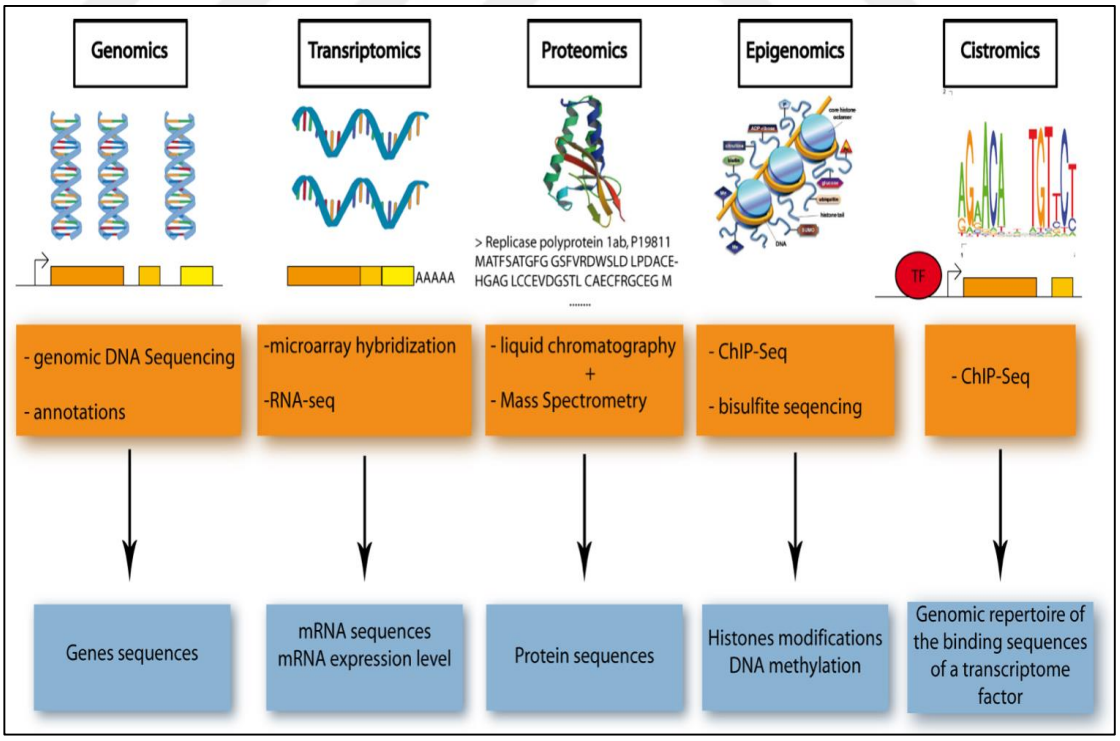


Figure 2.5: Schematic representation of the five omics technologies, focusing on the accessed element (blue) and the methods used (orange).

### 2.2.1.1. Transcriptomics

Transcriptome is the complete set of transcripts in a cell and their quantity in a specific condition. Understanding the transcriptomic behaviour of an organism gives very important insights about how cells change their gene expression levels in specific developmental stage, environmental or physiological condition. The key aim of the transcriptomics is to catalogue all gene transcripts [Wang et al., 2009]. In an organism, all cells have the same genome but they have different phenotypes. In single cell organisms like bacteria, expression levels of genes change simultaneously in different environmental or physiological conditions. The reason of this is the specific changes in transcriptome levels. Cell identity and function can be analyzed at molecular level by unique transcriptomic signatures. At organismal level, different tissues, conditions, mutations and diseases create different and unique transcriptomic profiles. With transcriptomic studies, discrepancy of different conditions such as disease, growth phase or growth medium can be compared in molecular level, and differences between analyses can reveal very important signs about the transcriptomic behaviour of organism [Stegle et al., 2015]. In transcriptome analysis, DNA microarrays, also called DNA chips, or RNA-seq techniques are the most common techniques. DNA microarray can be defined as an ordered collection of microspots that contains single defined species of nucleic acid. The technique is based on hybridisation. Sequence complementary causes hybridisation in a spot between two single stranded nucleic acids [Hwang et al., 2000, Gabig et al., 2001]. A brief overview of DNA microarray is shown in the Figure (2.6) [Hwang, 2000].

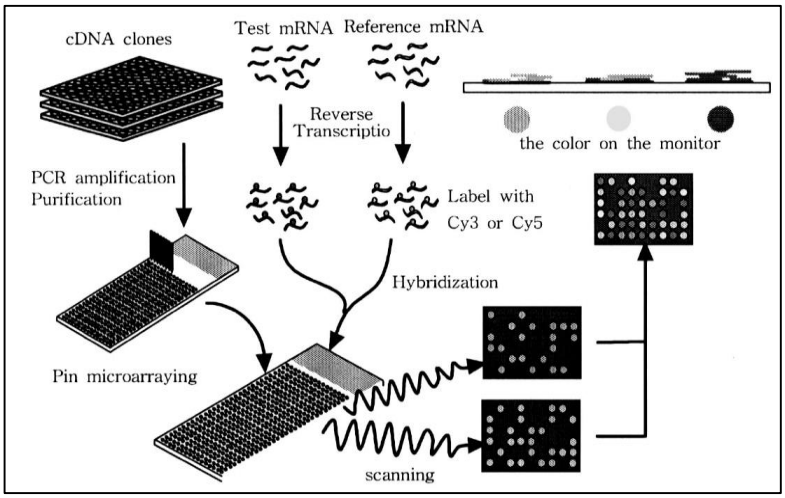


Figure 2.6: cDNA Microarray scheme.

RNA sequencing (RNA-seq) uses the capabilities of high-throughput sequencing methods to provide insight to the transcriptome of an organism. Compared to earlier techniques, RNA-seq provides higher scope and resolution of the dynamics of transcriptome [Kukurba et al., 2016]. A brief overview of RNA-seq is shown in Figure (2.7) [Wang et al., 2009].

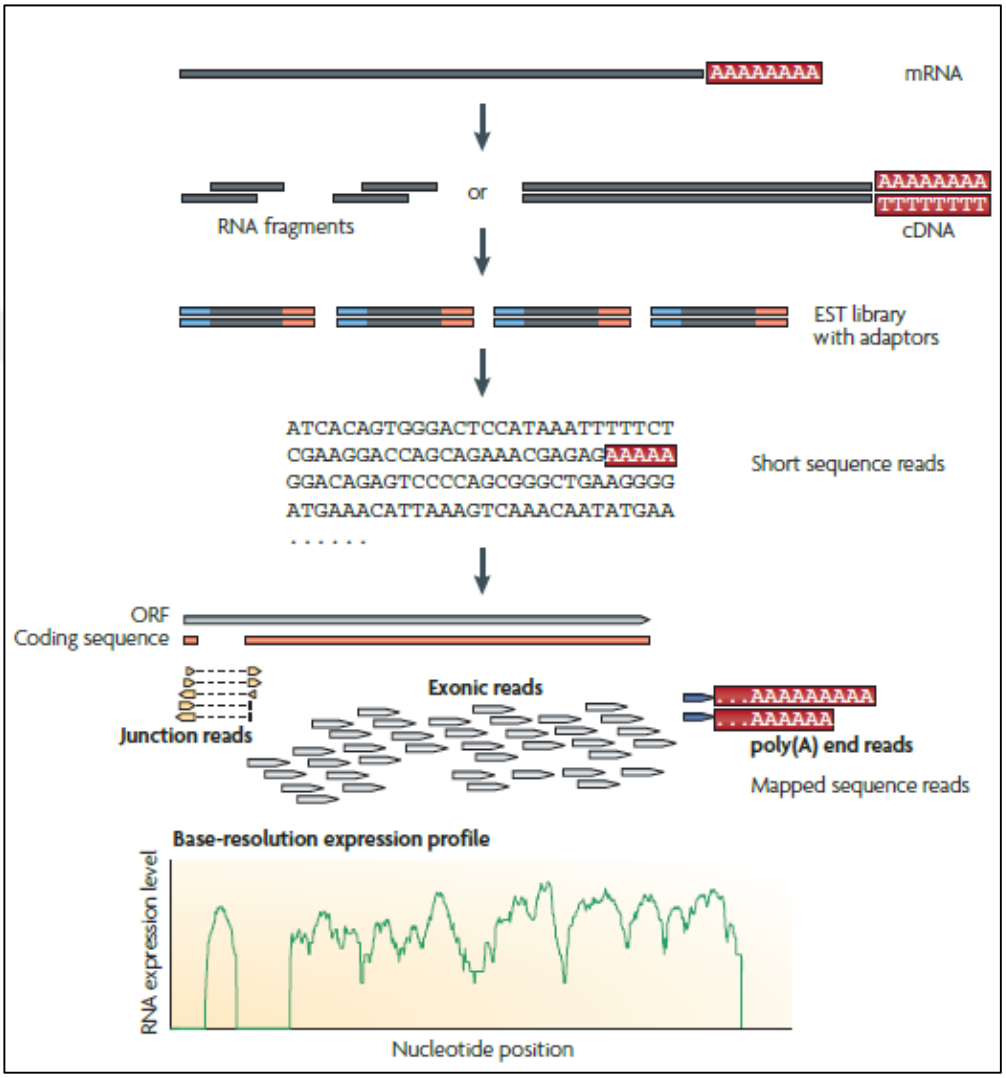


Figure 2.7: A typical RNA-seq experiment.

### 2.2.1.2. Other Omics Technologies

Genomics is a discipline that sequences and analyzes the function and structure of genomes and has DNA sequencing as its core technology. DNA sequencing technology has improved the cost and scale of genome characterization over 40 years [Koboldt et al., 2013]. Determining the genome sequences of organisms is an important achievement in the life sciences. Genomics is crucial to reconstruct genome

scale metabolic models by defining all enzyme-coding genes in a genome and building a model based on these genes and the corresponding proteins that are expressed from genome. Analysis of genomes also gives insight into genome structure [Miller, 2004]. Comparing genomes of different organisms can reveal the conserved sequences and enlighten the evolutionary roots of conserved sequences in different organisms.

Proteome is the expressed protein complements of a genome, and proteomics is the functional omics technology at the protein level. Proteomics is the large-scale study of proteins, and it studies protein properties such as expression levels and post-translational modifications on a large scale to obtain a systemic view of diseases, cellular processes and networks at the protein level. Whole genome and transcriptome analyses can be done, but the protein functions should also be revelaed. Also, post-transcriptional modifications affect the structure and function of proteins. Therefore, proteomic studies have potential to be used more than transcriptomic studies in future [Anderson et al., 1998, Blackstock et al., 1999].

Metabolomics is the systematic study of all endogenously produced metabolites and one of the youngest of omics sciences. This rapidly developing omics technology involves the total collection of small molecules present in cells, tissues or organ. Like other omics sciences, metabolomics will become increasingly useful in biological sciences [Ma et al., 2012]. The identities and concentrations of metabolites in an organism in a specific condition results in a complex interaction network among gene expression, protein expression and environmental conditions. Different from classical biochemical studies based on single metabolites, metabolomics involves a set of quantitative data of metabolites to gain overall insight of biological system [Kaddurah-Daouk et al., 2008]. Different from classical approaches on metabolites, new metabolomics technologies are not targeted, and they can identify dozens of metabolites at the same time [Villas-Boas et al., 2005, Villas-Boas et al., 2013].

There are also different kinds of omics technologies such as interactomics that studies the whole set of molecular interactions in a specific cell, tissue or organ, lipidomics that studies the pathways and networks of biological events that include lipids, fluxomics that studies the rates of metabolic reactions in a whole biological system, and phenomics that studies the physical and biochemical traits of organisms as they change in response to genetic mutation and environmental conditions.

### 2.2.2. Biological Networks

A network is a system that has units that are linked together, such as airport networks, social networks and species linked to the food chain. Every network has two typical elements: vertices and edges. A very simple network is shown in Figure (2.8).

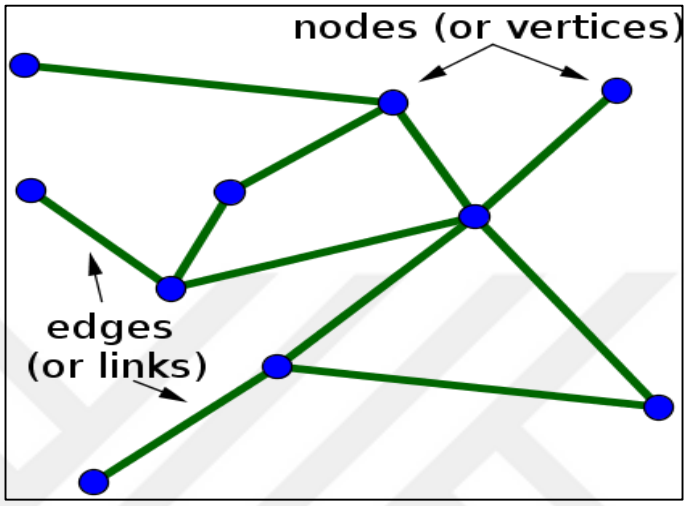


Figure 2.8: A very simple network.

A Biological network is a type of network that can be found in biological events such as protein-protein interactions. In biological systems, networks can be observed in different events such as food chain in ecology and biochemical nets in molecular biology. In fact, many interactions between cellular components are best represented by various complex networks [Wuchty et al., 2006]. A biological network has biological compounds as vertices and interactions as links. [Palsson, 2006]. Visualization of the interactions (links) between biological compounds (nodes) as a network graph can provide an integrated view of the system. Biological networks can be highly complex even in the simplest bacteria.

Modelling, analysis and visualization of biological networks are now important tasks in biological sciences. Several highly important biological networks are related to cellular molecules such as DNA, RNA, proteins and metabolites and interactions among them. Biological networks can be represented as dynamical, stoichiometric or logical models. To analyse biological networks experimentally, omics technologies are used and the data are evaluated to construct biological networks. Main types of biological networks are, transcriptional regulatory networks, signalling networks and metabolic networks [Bachmaier et al., 2013]. The main role of the network models is

to explain the emergence and behaviour of complex networks. The following subsections give details on these three major types of biological networks. Metabolic networks and related models will be discussed in a much more detailed way since the focus of this thesis study is metabolic networks (see section 2.2.3). Figure (2.9) [Bachmaier et al., 2013] shows the basic information flow in biological systems.

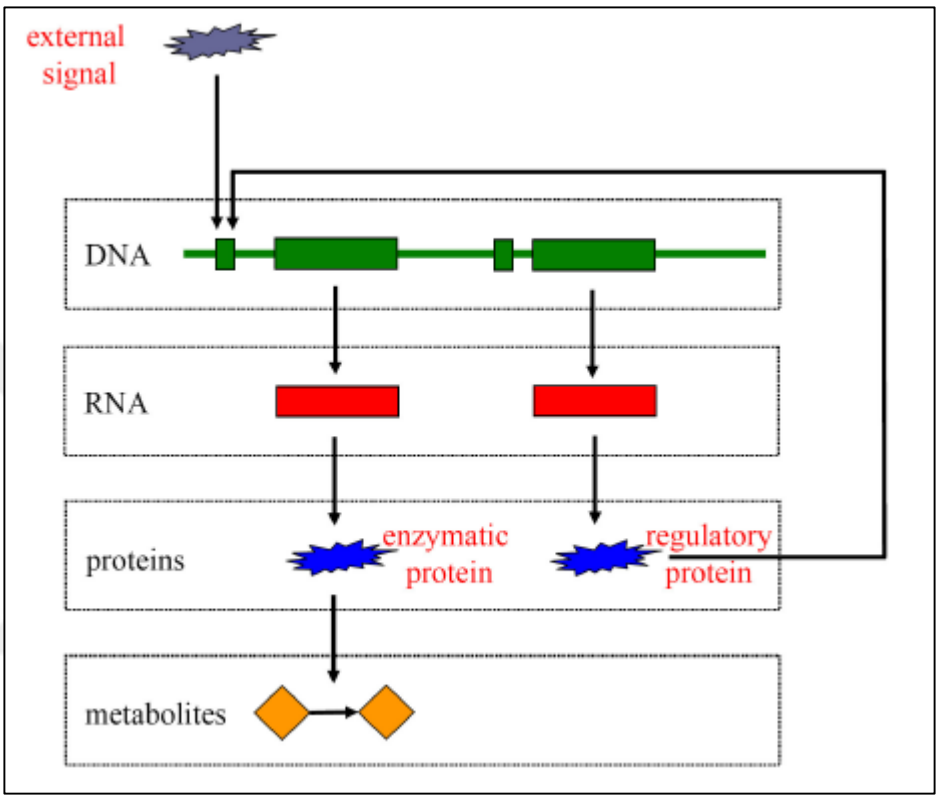


Figure 2.9: The standard pathway of information flow: DNA→RNA→protein.

### 2.2.2.1. Transcriptional Regulatory Networks

The expresssion of a gene of a genome is finely regulated sensitively in a cell. In a specific condition, only a subgroup of genes are expressed. In higher and multi-cellular organisms, all cells have the same set of genes but cells in different tissues have different functions and different transcriptome profile. In cellular organisms, different conditions require different set of proteins for the survival of cell. Changes in environmental conditions like nutrient limitation lead to expression of different genes. The control of gene expression is provided by transcriptional regulatory networks [Edwards et al., 2002]. Complex transcriptional regulatory systems evolved probably more than a billion year ago. Regulation of the gene expression by DNA binding proteins called transcription factors alters the binding rate of RNA polimerase

to the DNA [Levine et al., 2005, Davidson et al., 2005]. Transcription factors bind to the short DNA sequences called binding sites, which are mainly located in regions called promoters. Promoter is a region of the DNA that starts the transcription of a specific gene [Sharan et al., 2007]. Promoters are associated with a sigma factor. RNA polymerase binds to initiator sigma factor that directs the transcription process. Bacteria can contain more than one sigma factor. For example, *S.coelicolor* has 65 sigma factors. Different sigma factors recognize different promoters. Another mechanism of transcriptional regulation is via operons. An operon is a functional unit of a DNA cluster that is regulated by one promoter. Under a specific condition, the genes in the operon are expressed together and these genes have a functional similarity. Operons are mostly found in prokaryotes but also in some eukaryotes, like in *C.elegans* [Gruber et al., 2003, Palsson et al., 2002, Jacob et al., 1960, Alon, 2007]. A simple sketch of gene regulation systems is shown in Figure (2.10) [Sharan, 2007]

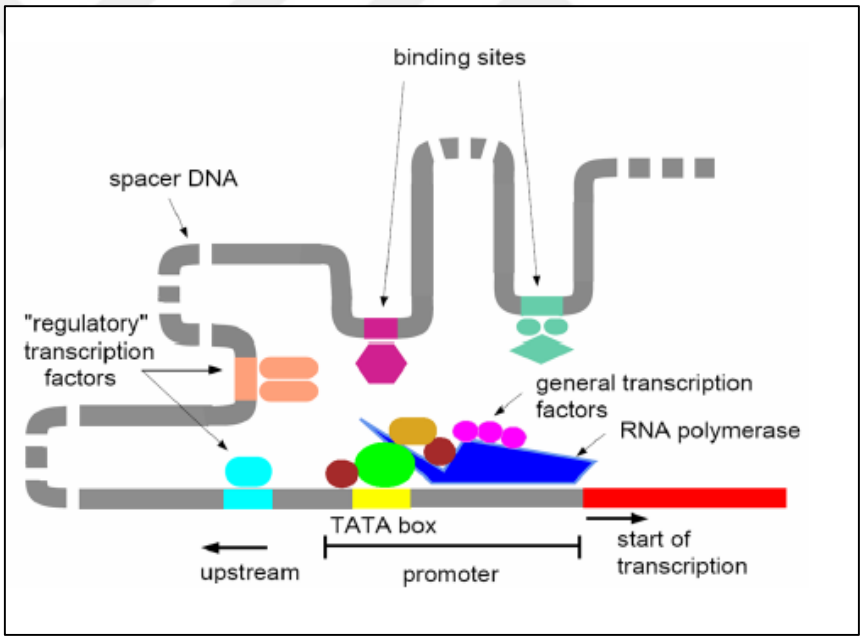


Figure 2.10: A sketch of gene regulation.

Transcriptional regulatory networks can be decomposed into a set of commonly occuring motifs. These motifs are conserved in evolution and can be found in all organisms [Alon, 2007, Edwards et al., 2002]. These motifs are shown in Figure (2.11) [Palsson, 2006].

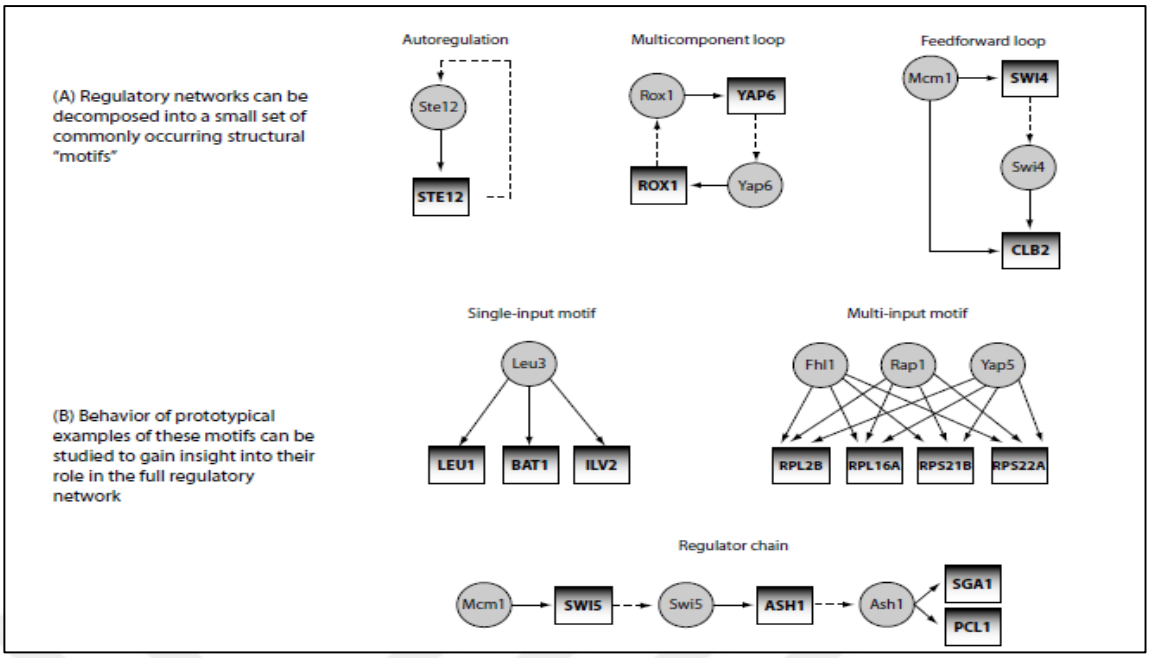


Figure 2.11: Motifs in transcriptional regulatory networks.

### 2.2.2.2 Signal Transduction Networks

A key issue in biology is the response to internal and external signals of a cell. Signal transduction is the process of communication within a cell to coordinate the transcription and metabolism to a change in environment [Bachmaier et al., 2013]. A signalling network is a network of reactions that stimulate the cell that responds to specific conditions. Cells needs to respond to changes in extracellular environment to survive. Physical changes like heat or nutrient limitations need to be responded quickly to regulate the metabolism [Donaldson, 2010]. The responses can be simple as in the case of opening of an ion channel or as complex as in the case of phosphorylation of a series of proteins. Signalling process mostly involves the binding of a signalling molecule (ligand) to a specific protein called receptor, followed by the phosphorylation of an internal protein, the passage of signal and the responses [Palsson, 2006].

### 2.2.2.3 Metabolic Networks

Function of cells involves complex networks of organized interactions and reactions. Due to this, a cell metabolism can be presented as collection of biochemical reactions [Edwards et al., 2002]. Metabolism is the set of chemical transformations within the cells of living organisms. The main task of the metabolism is the conversion

of nutrients to energy to run cellular processes and the conversion of building blocks to proteins, lipids, carbohydrates and nucleic acids. The reactions of the metabolism are catalyzed by enzymes that are encoded by the genome of the cell. Metabolism involves two kinds of reactions, anabolism and catabolism. Anabolism is the set of reactions that involves building the macromolecules from elementary metabolites by using energy, and the catabolism is the set of reactions that involves the conversion of macromolecules to elementary metabolites to obtain energy [Nelson et al., 2005, Smith et al., 2004]. Figure (2.12) [Palsson, 2006] shows the major parts of metabolism.

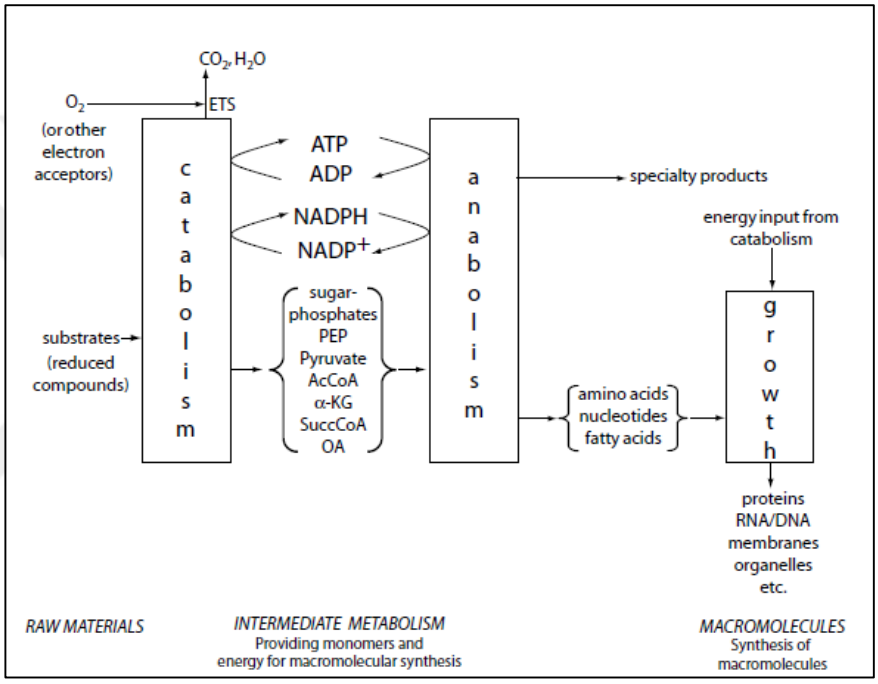


Figure 2.12: Major parts of metabolism.

Well-characterized metabolic networks can be shown by mathematical models at whole-genome scale. If kinetic parameters of the reactions are defined, a dynamical metabolic network can be constructed and numerical simulations can be carried out, but kinetic parameters of a complex metabolic network is hard to define. All enzymes have specific kinetic parameters and measuring these parameters experimentally is often hard. Purification of enzymes and other experimental procedures make determination of kinetic parameters hard. Although the data are abundant nowadays, the majority of kinetic parameters are still undefined. Nevertheles, with new computational methods, metabolism can be modelled without many parameters. The main concept of this type of modelling is the identification and mathematical definition of constraints [Terzer et al., 2009]. Reactions in metabolism are connected to each

other because, in a cell, substrate of a reaction is the product of another reaction. Using the connections between substrates and products in or out of a reaction, a metabolic network can be represented as a stoichiometric matrix. An example of a metabolic network in *E. coli* is shown in Figure (2.13) [Almaas et al., 2004]..

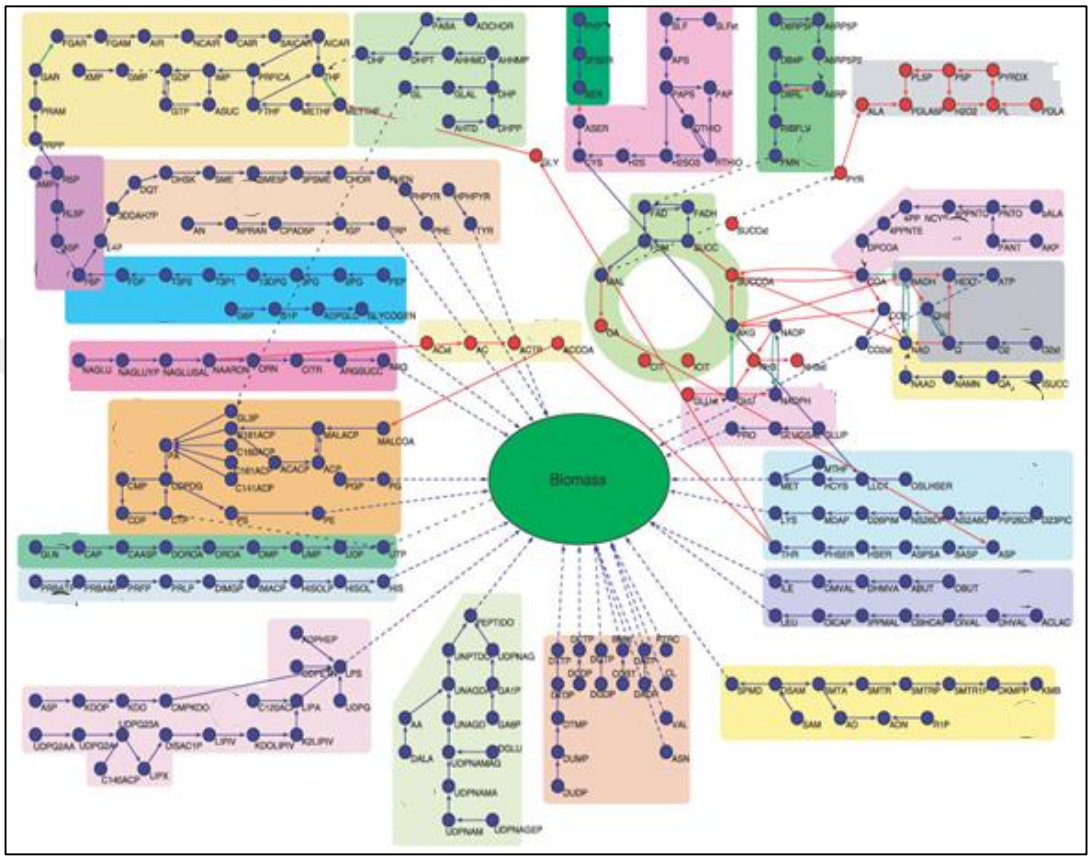


Figure 2.13: Metabolic network of *E. coli* on a glutamate-rich substrate.

## 2.2.3 Genome Scale Metabolic Models

Development of high quality techniques has led to a rapid increase in genome-scale metabolic models. Genome scale metabolic models are really useful tools to study the systems biology of metabolism. Number of this type of models rapidly increased and now many models are available for many organisms. Like other kinds of networks, genome scale metabolic models provide very effective ways of summarizing the information about the metabolism of related organisms. With the increase in the availability of whole genome sequence of variety of organisms, number of genome scale metabolic models have increased exponentially [Zommorodi et al., 2012, Patil et al., 2004].

There are many models available now for many organisms such as *Escherichia coli, Saccharomyces cerevisiae, Helicobacter pylori* and *S.coelicolor*. These models have been used in several studies to investigate metabolic properties of organisms in question, and in metabolic engineering studies that provide insights about mutant strains to increase the production of a specific metabolite [Patil et al., 2004, Edwards et al., 2000, Sohn et al., 2010, Sohn 2012, Bro et al., 2005].

Genome-scale metabolic models aim to cover all metabolic reactions and metabolites in a cell. Reconstruction of genome scale metabolic models involves few steps. First step to reconstruct a model is creating a draft model. Draft model is based on genome annotation of organisms and biochemical databases. Genome annotation is important to define the gene properties. With genome data, coding sequences and proteins and the biochemical reactions related to these genes can be identified. With the biochemical databases like gene ontology (GO), metabolic genes from genome can be retrieved. Metabolic reactions can be connected to the genes by EC numbers and by the use of biochemical databases like KEGG or Brenda, and all these data can be gathered in draft model. These data can contain many false-positives since, for example, proteins involved in DNA methylation have EC numbers too but they are not considered in metabolic reconstruction [Thiele et al., 2010, Palsson, 2006].

Genome scale metabolic models describe the metabolic networks as a set of stoichiometric equations that represent the biochemical equations. There are different models for different organisms but in fact they share one important requirement: a stoichiometric matrix. In a stoichiometric matrix, each column represents a reaction that can be biochemical or transport reaction and each row of the matrix represents a metabolite in the system. Stoichiometric matrix contains directionality which means metabolites can have negative or positive coefficients. Shortly, a stoichiometric matrix is a list of reactions in which metabolites participate [Terzer et al., 2009].

In genome scale models, maximum and minimum boundaries of the reactions are described in a vector. According to the physico-chemical laws, every reaction has boundaries. Reactions in a model can be classified as transport reactions, intracellular reactions, exchange reaction and biomass reaction [Liu et al., 2010]. Biomass reaction represents synthesis of biomass using certain amount of building blocks such as amino acids, carbohydrates, lipids and nucleic acids as precursors. With experimental studies, the content of biomass can be identified and corresponding equation can be added to the stoichiometric matrix.

After the reconstruction of a metabolic model, metabolism of the related organism can be investigated by computational approaches. As in engineering, a specific product can be maximized in a specific condition to see the result. By changing the boundaries of transport reactions, a growth medium can be simulated computationally. Basic steps of the reconstruction of a genome scale model is shown in Figure (2.14) [Palsson, 2002].

After the reconstruction of a genome scale metabolic model, mathematical optimization methods can be applied to this metabolic network computationally. Each optimization problem involves an objective function and a set of constraints. Objective function specifies the desired property of the system that should be maximized or minimized. Constraints are the boundaries that define the space of possible solutions to the optimization by the objective function. For this approach, there are several mathematical methods. Most common one is Flux Balance Analysis or FBA.

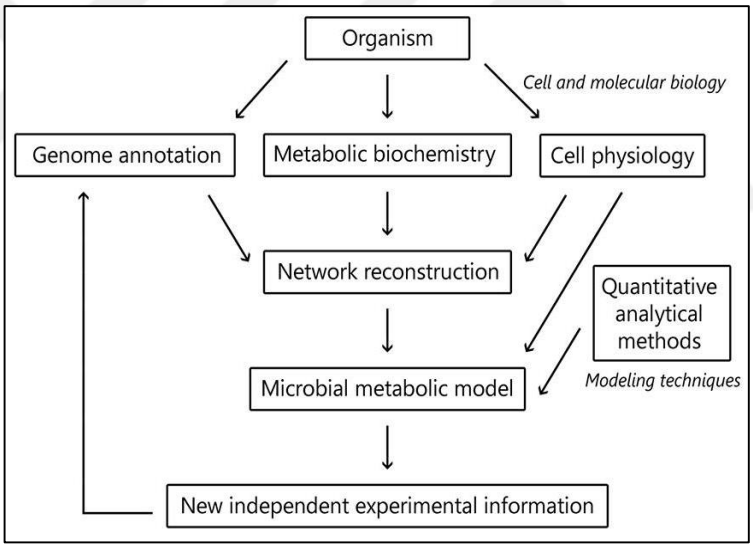


Figure 2.14: A schematic of the overall process of genome-scale metabolic network construction.

### 2.2.4 Flux Balance Analysis (FBA)

Flux balance analysis or FBA, is one of the most common computational approaches for studying the biochemical networks, especially genome scale metabolic networks. The metabolic networks are described in the model in the form of mass balance dynamics which represents the concentration change of a metabolite as being equal to the difference between reaction rates [Edwards et al., 2002].

Metabolic network of a cell is like a non-stop engine. For mathematical and computational analyses, FBA assumes that network will reach a steady state that is constrained by stoichiometric matrix. This state is called a pseudo-steady state. All fluxes in the system defines a bounded solution space. Once the solution space is described, capability of organism is defined and properties of the network can be studied [Kauffman et al., 2003, Lee et al., 2006].

Flux balance analysis can be applied to metabolic networks from relatively small networks such as a production network of specific metabolite, to whole genome scale metabolic models. To apply FBA, some parameters must be defined. The parameters can be limited to a specific value to mimic a condition such as in an *E.coli model*, where oxygen uptake rate (transport reaction) was defined as 0 to mimic anaerobic environment [Orth et al., 2010].

Required parameters for FBA are similar to those required in the reconstruction of a genome scale metabolic model. These steps are, system definition, mass balance, defining fluxes and optimization. In the system definition step, all reactions must be defined. These steps are exemplified over a toy model of a network shown in Figure (2.15) [Kauffman et al., 2003].

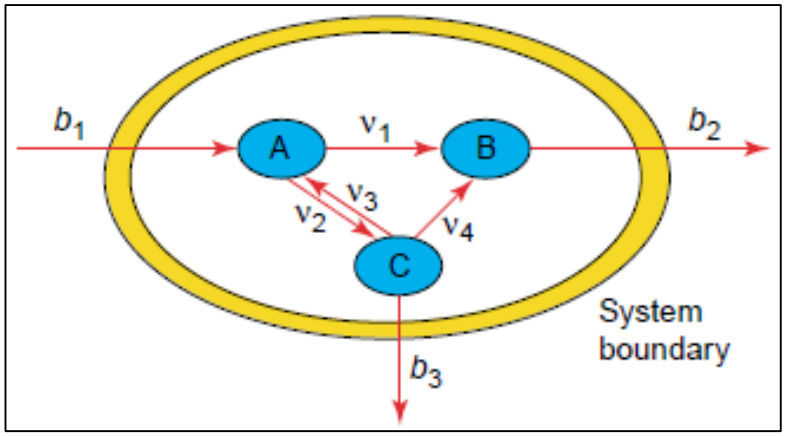


Figure 2.15: A model system that contains three metabolites (A, B and C), three internal reactions and three exchange reactions.

Once the system is defined, a dynamic mass balance is derived for all metabolites in the network. Mass balance is defined by using the flux through each reaction and the stoichiometry of that reaction. Mass balance leads to a set of differential equations. These equations can represented using a stoichiometric matrix 'S' and a vector of

fluxes ‘v’. Differential equations and matriceses (S and v) of the system in Figure (2.15) are shown in Figure (2.16) [Kauffman et al., 2003].

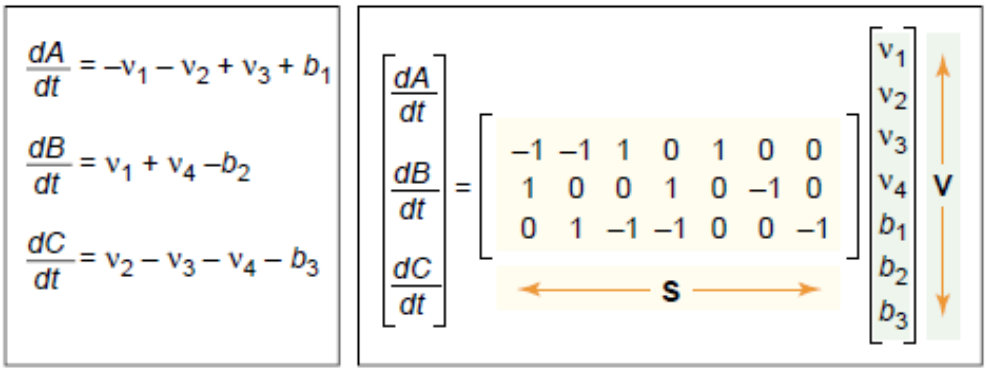


Figure 2.16: Mass balance equations of the system.

Usually, the exact fluxes in metabolic network (the matrix v) is not well defined but their boundaries are known. Direction of the reactions plays an important role to define the boundaries. Flux of reversible reactions can be positive or negative but irreversible reactions can only have one direction. Besides, some fluxes can be determined experimentally to decrease the degrees of freedom. For instance, the rate of the reaction called “non-growth associated ATP maintenance reaction” is usually set at its minimum boundary. Mathematical identification of fluxes is shown in Equation 2.1 and 2.2. Irreversible reactions and reversible reactions are constrained to take values between [0, ∞ ) and (-∞ , ∞) respectively. In equation, $v^{lb}$ is the minimum value (lower bound) of flux and $v^{ub}$ is the maximum value (upper bound) of flux.

$$v_i^{lb} \leq v_i \leq v_i^{ub} \tag{2.1}$$

After these main steps, the parameters can be used for FBA. In a defined system, we have an *m x n* size stoichiometric matrix, S, that represents a metabolic network that consists of m metabolites and n reactions. For a pseudo-steady state we can assume the change in the metabolite concentrations over time is zero [Sun et al., 2013]. FBA aims to solve the problem by linear programming.

$$S_{m \times n} \times v_{n \times 1} = 0_{m \times 1} \tag{2.2}$$

After these steps, the last step of FBA is the optimization step. Typically, metabolic networks have more fluxes in the system than metabolites. Therefore, the system of equations is underdetermined. For a given objective function (Z), an optimal

set of fluxes can be obtained subject to the mass balance. By changing the boundaries of exchange fluxes, specific growth conditions can be mimicked using FBA. Figure (2.17) [Orth et al 2010] represents conceptual basis of FBA and constraint-based modelling.

Generally, maximization of growth rate is used as objective function but, for analysing the production of specific metabolite production, other objective functions can be used too. By optimization, the phenotype of the organism can be defined. FBA has some alternative usages such as robustness analysis of a specific reaction.

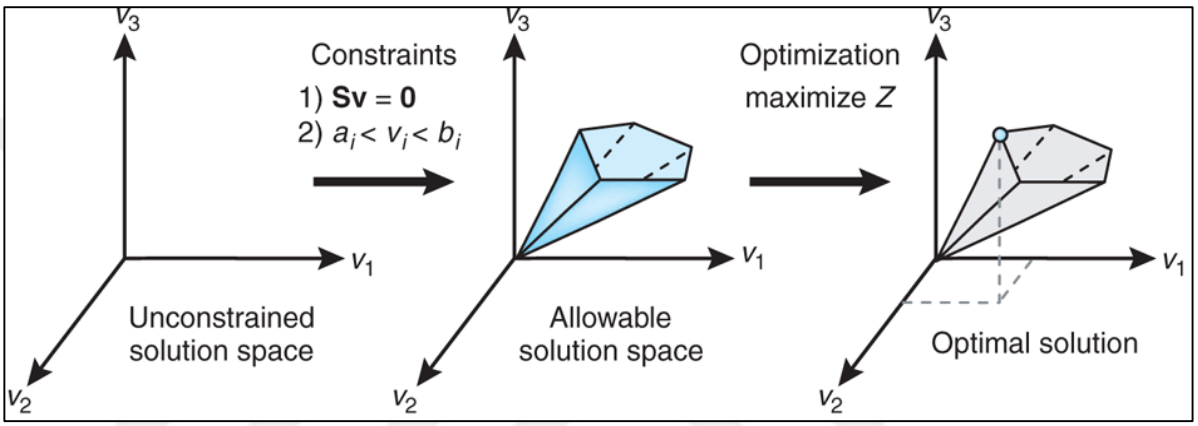


Figure 2.17: The conceptual basis of constraint-based modeling and FBA.

## 2.2.5 Reporter Metabolite and Pathway Analysis

Basically, reporter metabolite analysis is a tool that integrates omics datasets to a metabolic network model to identify the most significantly affected metabolites between two conditions in which omics datasets were generated. These conditions or perturbations include gene deletion, insertion or environmental changes such as nutrient change.

"Reporter metabolites" algorithm was developed by Nielsen and Patil in 2005 [Patil et al., 2005] and it is a hypothesis-driven algorithm and it is represented in Figure (2.18) [Sertbaş 2013]. In this algorithm, each metabolite in the metabolic network is scored based on the normalized transcriptional response of its neighbouring enzymes. Equation 2.3 shows the formula to convert p-values to Z-scores of each gene. Each pi can subsequently be converted to a Z score of the enzyme node ($Z_{ni}$) by using the inverse normal cumulative distribution function ($\theta^{-1}$).

$$Z_{ni} = \theta^{-1}(1 - p_i) \quad (2.3)$$

Equation 2.4. shows the formula to calculate the Z-score of each metabolite. In case of differential data, the normalized transcriptional response was calculated as size independent aggregated Z scores of the *k* neighboring enzymes.

$$Z_{metabolite} = (1 \div \sqrt{k}) \times \sum_i^n Z_i \quad (2.4)$$

These scores can be corrected by auto-scaling for the background distribution as shown in Equation 2.5.

$$Z_{corrected} = \frac{(Z_{metabolite} - \mu_k)}{\sigma_k} \quad (2.5)$$

In the last step, Z scores are converterd to p-values and the metabolites that have p-values lower than the selected cut-off are named as reporter metabolites. These metabolites are the metabolites perturbed through the transcriptional change in the associated reactions.

Reporter pathway analysis (RPA), similar to the reporter metabolite analysis, is an algorithm that integrates the omics data sets to a metabolic network to reveal the most significant changed pathways in metabolism.

Studying the transcriptome data set in pathway scale reveals more useful information about the changes in conditions that are compared in omics data. In old approach, RPA is done via the p-values of reactions to calculate the p-values of pathways. In this study, metabolite-based RPA that was introduced by Çakır et. (2015) is used. In this approach p-values of each metabolite is calculated by using reporter metabolite analysis and p-values of pathways are calculated via metabolite p-values. The equations that are used in metabolite based reporter pathway analysis algorithm are shown in equation 2.6.

$$Z^m_{pathway} = (1 \div \sqrt{n}) \sum^n_{metabolite=1} Z_{metabolite} \quad (2.6)$$

In these algorithms, Z-score is used instead of the more common t-score. These scores are both used in hypothesis testing. Generally, when the sample size is below 30 and the population of sample standart deviation is unknown, t-score is used.

Technically, z-scores are a conversion of individual scores into a standard form. The conversion is based on the knowledge about the population's standard deviation and mean. A z-score indicates you how many standard deviations from the mean the result is. Like z-scores, t-scores are also a conversion of individual scores into a standard form. However, t-scores are used when the conversion is made without knowledge of the population standard deviation and mean. Basically, t-score is more error-prone than z-score in hypothesis testing of large populations.

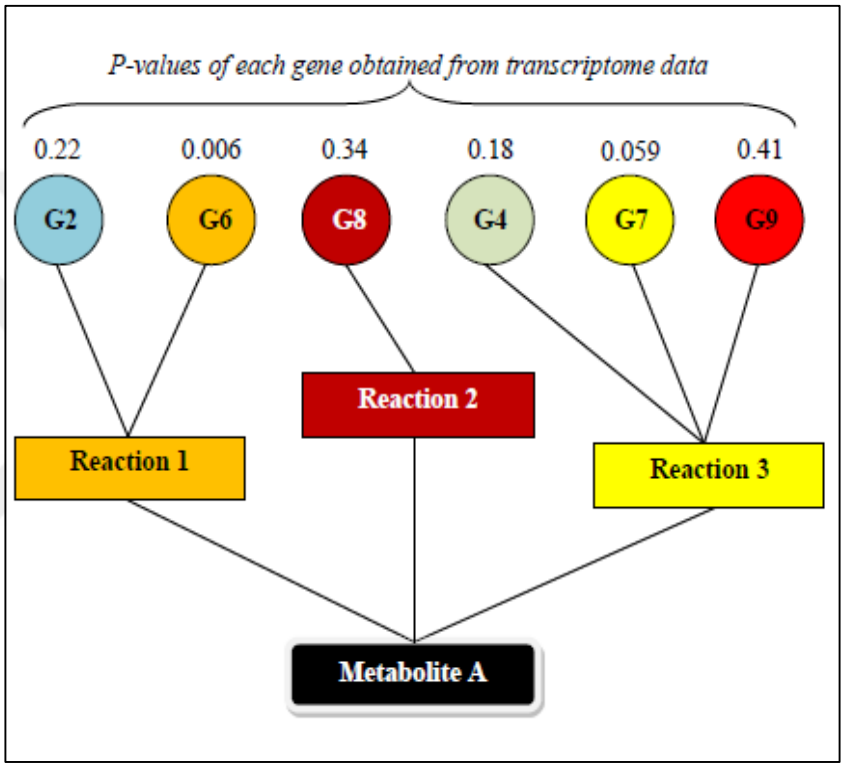


Figure 2.18: Outline of reporter metabolite analysis.

## 2.2.6. Statistical Significance Tests

Experiments usually have a goal of testing a theory or hypothesis. If model is suitable, hypothesis of interest can be characterized by values taken by one or more parameters. A statistical test is used to find evidence about hypothesis [Bowman, 1987]. Shortly, significance is a measure of strength of data associated with null hypothesis. Null hypothesis ($H_0$) is a term that is used in statistics to measure the connection with the hypothesis of "no difference" [Upton et al., 2014].

Statistical significance is gained when a p-value is less than a significance cut-off level. The p-value is the probability of randomness. P-value concept is commonly

used in natural and social sciences. The first step of statistical testing is developing a null hypothesis. Null hypothesis is chosen as the opposite of the research hypothesis [Johnson, 1999].

Statistical significance tests can be either parametric or non-parametric. Parameter is a numerical quantity that characterizes the distribution or random variable or population [Upton et al., 2014]. Parametric tests use some statistical parameters such as mean, variance or standart deviation. To use parametric tests, data set must show a normal distribution. Mean ($\bar{x}$) is the average of all numbers and shown in Equation 2.7. The mean of a population sized *N* is the sum of its elements divided by *N*. This is commonly referred as arithmetic mean and it is the most widely used parameter of a data [Upton et al., 2014]. Standart deviation (σ), is a measure that is used to quantify the amount of variation or dispersion of a set of data values and show in Equation 2.9. The standard deviation of a population size N is given by the square root of the squared deviations from the mean divided by N [Upton et al., 2014,Bowman 1987, Lieber 1990]. The square of the standard deviation is called variance and shown in Equation 2.8.

It should be noted that significance tests can be one-tailed or two-tailed depending on the hypothesis. For instance, if we are interested in both higher and lower values than cut-off, this is a two-tailed test. If we are only interested in one of the higher or lower values than cut-off, this is a one-tailed test [Mould, 1998].

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{x_1 + x_2 + \cdots + x_N}{N} \tag{2.7}$$

$$s^2 = \frac{\sum (X - \bar{X})^2}{N-1} \tag{2.8}$$

$$\sigma = \sqrt{\frac{\sum (x - \bar{x})^2}{N}} \tag{2.9}$$

Most common parametric test is Student's t-test. t-test can be applied to data sets from two different conditions exhibiting normal distrubution. In biology, these conditions can be changes in environment or knockout of a gene.

### 2.2.6.1 Student t-Test

Student t-test is used for statistical analysis to determine the significance of change between two data sets. It uses mean and standard deviation to make a comparison. The formula for t-test is given below. After calculation of the t-value, from the t-table, according to the degrees of freedom ($n_1 + n_2 - 2$), p-value is calculated and if the p-value is lower than the cut-off value we choose, null hypothesis is rejected and we can say there is a significant change [Mould, 1998]. Equation 2.10 shows the calculation formula of t score.

$$t = \frac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \tag{2.10}$$

## 2.2.7. Principle Component Analysis

Principal Component Analysis (PCA) is the general name for a technique which uses sophisticated underlying mathematical principles to transform a number of possibly correlated variables into a smaller number of variables called principal components. For example, a data set that has n number of features is transformed into a data set that has k dimensions($k < n$). Also because PCA reduces the multi-variable data, it is used in visualization [Richardson, 2009, Abdi et al 2010].

PCA is used in all forms of analysis, from neuroscience to computer graphics, because of its simplicity. Some of the other common applications include, de-noising signals, blind source separation, and data compression. The main ideas behind PCA are drawing some lower dimensional space and represent each data point by its projection [Shlens, 2003]. Using mathematical projection, the original data set, which may have involved many variables, can often be interpreted in just a few variables (the principal components).

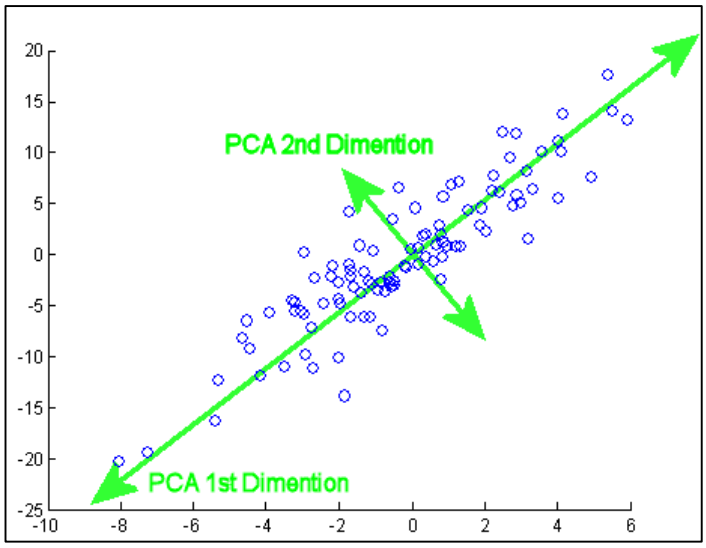

Figure 2.19: Principle Component Analysis Example.

# 3. MATERIALS AND METHODS

## 3.1. *Streptomyces coelicolor* Genome Scale Model

First *S. coelicolor* genome scale metabolic model (iIB711) was developed by Borodina et al., in 2005 [Borodina et al., 2005]. This model has 972 reactions and 496 metabolites in its stoichiometric matrix and has 711 genes associated with reactions. 819 of the reactions are biochemical conversions and 152 are transport reactions. Model was validated by making predictions about the growth of *S. coelicolor*. In silico analyses of the model were compared with the chemostat data from Melzoch et al., 1997.

Another model (iMA789) for *S. coelicolor* was developed by Alam et al., in 2010 [Alam et al., 2010]. The model has 1015 reactions and 759 metabolites and 789 genes associated with 666 enzymatic reactions. 747 reactions are metabolite biosynthesis and degradation reactions, 152 are transport reactions and 116 are additional inputs and outputs of the system. The model was validated by using the chemostat data from Melzoch et al., 1997 just like the model of Borodina et al., in 2005. The model has been used to investigate the metabolic switch from primary metabolism to secondary metabolism.

The most recent metabolic network model (iMK1208) was developed by Kim et al., in 2014 [Kim et al., 2014]. In this model, properties of old models were improved, some reactions, metabolites and genes were added. The model has 1643 reactions, 1436 metabolites and 1208 genes associated with the reactions. In construction of the model, genome annotations from StrepDB, BioCyc and KEGG were retrieved and these data banks were used for Gene-Protein-Reaction relationships. For the data of transport reactions, TransportDB database was used. When necessary, BLASTP was used against the proteins of *E. coli*. This model was also compared with the chemostat data of Melzoch et al., (1997). It was also used with an algorithm called “flux scanning based on enforced objective flux” (FSEOF) to design a strain that produces actinorhodin more than wild-type strain. The designed strain was validated experimentally and the model was proven as to be accurate.

The *S. coelicolor* metabolic network that was downloaded from BioCyc *S. coelicolor* data base was only used for reporter metabolite and reporter pathway

analysis in this study. Comparison of properties of the genome-scale networks are shown in Table 3.1.

Table 3.1: Comparison of *S. coelicolor* genome scale models.

| Model Name | iIB711 | iMA789 | iMK1208 | BioCyc Network |
|---|---|---|---|---|
| Genes Included | 711 | 789 | 1208 | 1103 |
| Reactions | 972 | 1015 | 1643 | 1705 |
| Metabolites | 496 | 759 | 1436 | 1301 |

## 3.2. Transcriptomic Data Sets

Transcriptomic studies were performed by using time dependent transcriptomic analysis data set of the *ArgR* knockout mutant and *PhoP* knockout mutant.

*ArgR* knockout mutant and its wild-type strain transcriptomic study was performed by Botas et al., (2012) and this data set has 30 samples for 5 time points ($32^{nd}$-$42^{nd}$-$49^{th}$-$56^{th}$-$66^{th}$ hours) for both wild type and ΔargR strains. The data have 3 biological replicates. Analyses have been made with Agilent-017905 DNA chip and all data were collected from the same growth medium (MG–maltose glutamate medium). The data is stored in Gene Expression Omnibus(GEO) database under the code, GSE58666.

*PhoP* knockout mutant data set was generated by Nieselt et al., in 2012. This data set is sub-data set of time series study of Nieselt et al., [2010]. *PhoP* dataset has 36 samples from hour 23 to hour 60, one hour resolution from $23^{rd}$ to $36^{th}$ h and $41^{st}$ to $48^{th}$ h, half hour resolution between 36-41 h, two hour resolution between 48-60 h, sample missing for 34 h, with no replicates. The data set is stored in Gene Expression Omnibus (GEO) data base under the code, GSE31068. For the comparison with the wild type data, time series data set of Nieselt et al., (2010) is used. This data set is stored in GEO database under the code, GSE18489. Total number of samples in the time series data set is 32. For 32 samples, there are no replicates, with one hour resolution between 20-44 h, two hour resolution between 44-60 h, with sample missing for 25 h. All of the experiments have been done in same growth media in the same fermentors under the same environmental conditions and same starting volumes. With same experimental conditions, *PhoP* mutant data set and data set of wild type strain

can be analyzed together. For the analyses, mutual time points of both *PhoP* and wild type transcriptomics data sets were taken. Total number of time points is 28.

## 3.3. Softwares and Toolboxes for Analyses

For FBA and other analyses on iMK1208, the following softwares were used: Matlab 2011r software, Systems Biology Markup Language (SBML) Toolbox for Matlab [Keating et al., 2006], Constraint Based Reconstruction and Analysis (COBRA) Toolbox [Becker et al., 2007] and Reconstruction, Analyses, and Visualization of Metabolic Networks (RAVEN) Toolbox for Matlab [Agren et al., 2013]. For Reporter Metabolite Analyses, RAVEN toolbox is used. The Reporter Pathway Analyses were performed by using in-house developed Matlab codes. Statistical analyses has been done by Statistical Toolbox of Matlab. For data and output editing, Microsoft Office Excel 2013 and Notepad++ softwares were also used as subsidiary softwares.

The reaction that was maximized was selected as objective function via "changeObjective" command of COBRA Toolbox for Matlab. Optimization of genome-scale metabolic models was done via "OptimizeCbModel" command of COBRA Toolbox. Exchange reaction fluxes were changed via "changeRxnBounds" command in the toolbox.

## 3.4. Data Processing

The transcriptome datasets, GSE58666, GSE31068 and GSE18489, were preprocessed before statistical analyses. Since GSE58666 dataset has repeating genes, the first step to preprocess the data is making the data unique for each unique gene in the array. For this, the average transcription value of genes that repeat more than one was used. After the preprocessing of the expression data, data were saved as two variables, WildType Expression data and Δ*argR* Expression data. These data were used in t-test between wild type and mutant conditions. Wild type data of GSE58666 was also used in t-test between primary and secondary metabolism.

The other datasets, GSE31068 and GSE18489, are the subsets of each other but some of the data are missing for both of them. To preprocess both of the data, first the mutual time steps of datasets were found and extracted. The data sets were saved in

Matlab 2011r as Wild Type Expression Data and Δ*pho*P mutant Expression Data. Just like GSE58666, these datasets were used in t-test between conditions and primary-secondary metabolism.

Data extraction was in Matlab 2011 software and extraction results were checked in Microsoft Office Excel 2013.

Also, datasets were used in PCA and Cluster Analysis to show that the experiments that have similar conditions show similar behaviours.

## 3.5. Statistical Significance Test

After data processing step, expression profiles between mutant and wild type strains were compared via Student t-test. Also, wild type expression profiles of both data sets on different time steps such as 20-21 hour and 58-60 hour time steps were compared for secondary metabolism analysis. After t-test, for each unique gene in our platform data, a p-value was obtained as a measure of significance. These p-values were used in reporter metabolite and pathway analyses.

All t-tests was done by "ttest2" command of Matlab 2011r software and saved in both Matlab m files and text files.

## 3.6. Reporter Feature Analyses

The p-values that were obtained as a result of statistical significance analyses were used in reporter metabolite and reporter pathway analysis. For reporter metabolite analysis, both genome-scale metabolic model of *S. coelicolor* [Kim et al., 2014] and the genome-scale metabolic network which was downloaded from BioCyc website were used. Reporter metabolite analysis was done via Reporter Metabolite Analysis Tool of RAVEN Toolbox.

The Metabolite Centric Reporter Pathway Analysis was done in Matlab 2011r by using in-house codes developed by Isa Yuksel (2016). This analysis was only performed by using the BioCyc network since it has much higher pathway coverage. The results were saved in Microsoft Excel and in text files.

## 3.7. Creating Condition Specific Models

Gene Inactivity Moderated by Metabolism and Expression (GIMME) algorithm was developed by Becker et al., in 2008 to create a condition specific genome-scale metabolic model by integrating a generic genome-scale metabolic model and transcriptome data [Becker et al., 2008]. The COBRA Toolbox includes a tool script called "Create Tissue Specific Model" for applications of this algorithm. Basically, GIMME uses a logical expression pattern that shows the genes below threshold and deletes the related reactions from model. In the next step, algorithm identifies the Required Metabolic Functionalities (RMFs) such as reactions needed to produce biomass. Typically the threshold that was used to identify the genes that are under or over-expressed is selected as the half of the average expression value of all genes. There is not any official usage to define the threshold. As a new method to identify unexpressed genes, fold change and p-value comparison was used between conditions. In this study, if a gene has a p-value lower than threshold and its fold change is below 0.01, related gene is considered as not expressed and the flux of related reactions is set to zero in GIMME algorithm.

Both of the threshold methods were used to obtain a Δ*arg*R mutant model, a Δ*pho*P mutant model and secondary metabolism models from corresponding transcriptome data sets by using the GIMME approach.

# 4. RESULTS

## 4.1. Model Validation

### 4.1.1. Comparison with Experimental Data

The *S. coelicolor* genome-scale metabolic model was compared with the experimental chemostat data of Melzoch et al., (1997). In the experimental study, in specific glucose and oxygen consumption rates, growth values are reported. In the present study, by using "COBRA Toolbox", biomass production rates of iMA789 (Alam et. al., 2010) and iMK1208 (Kim et. al., 2014) models were optimized in the same glucose and oxygen consumption rates (Table 4.1 and Figure 4.1). As it can be seen from the validation results, iMK1208 predicts the specific growth rate in different culture media conditions accurately. The model of Kim et. al (2014) shows closer results to the experimental results than the model of Alam et. al (2010). The prediction accuracy is mostly more than 80%. These results validate our choice of iMK1208 in this thesis study.

Table 4.1: Comparison of experimentally observed data from chemostat data and predicted specific growth rates.

| Glucose consumption rate (m.mol/g h) | $O_2$ consumption rate(m.mol /g h) | Experimental Results μ /h (Melzoch et al., 1997) | Predicted Specific Growth Rate μ /h [iMA789 (Alam 2010 et al.,)] | Predicted Specific Growth Rate μ /h [iMK1208 (Kim et al., 2014)] |
|---|---|---|---|---|
| 0.5 | 1.80 | 0.035 | 0.0272 | 0.0278 |
| 0.6 | 2.00 | 0.045 | 0.0396 | 0.0405 |
| 0.8 | 2.40 | 0.060 | 0.0539 | 0.0582 |
| 0.9 | 2.50 | 0.072 | 0.0657 | 0.0683 |
| 1.10 | 3.10 | 0.092 | 0.0862 | 0.0886 |
| 1.85 | 6.60 | 0.115 | 0.1088 | 0.1113 |
| 2.10 | 7.20 | 0.128 | 0.1385 | 0.1417 |

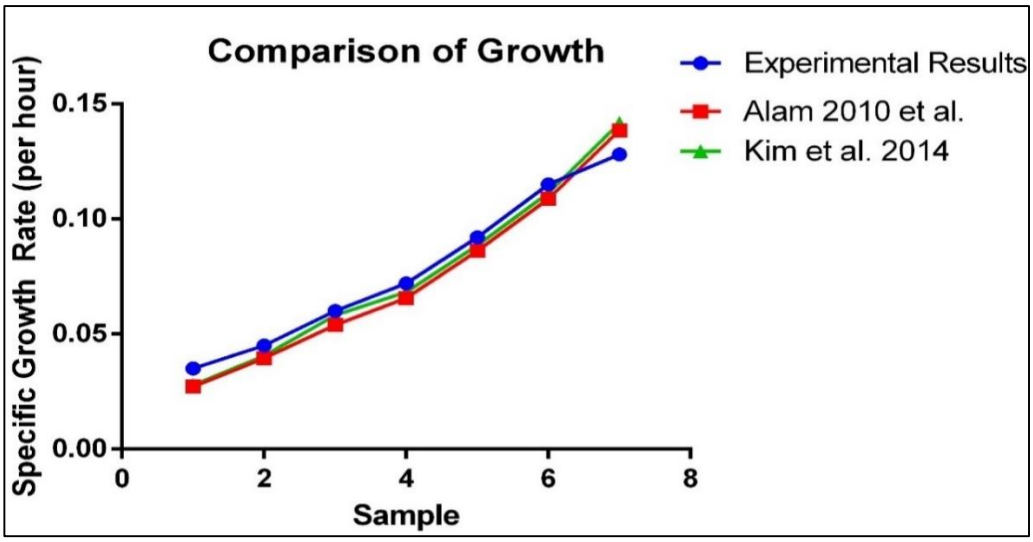


Figure 4.1: Model Validation.

Model simulated growth rates in different carbon sources were compared. The model predicts non zero growth rate for 46 different carbon sources and 11 nitrogen sources. According to the experimental study of Melzoch et al., (1997), *S. coelicolor* didn't grow when L-aspartate and L-glutamate was used as carbon sources, but in model, iMK1208, these two carbon sources can form biomass. 57 out of 64 different carbon and nitrogen sources form biomass based on iMK1208 simulation results.

For simulations by using different carbon sources, "biomass formation reaction" in the model was selected as the objective function to be maximized, "exchange reaction flux" of carbon sources were set to 1 mmol/g.h and "glucose exhange reaction flux" was set to zero. According to results in Table 4.2 and Table 4.3, *S. coelicolor* genome scale model (Kim et al 2014) can simulate growth on different carbon and nitrogen sources, and these results are compatible with experimental results. As a result, genome-scale metabolic model iMK1208 (Kim et al., 2014) can use all carbon and nitrogen sources in agreement with experimental data except the use of glutamate and aspartate as carbon sources. Deficiency of carbon metabolism regulatory systems in the model could be the main reason of this small incompatibility with the experimental data.

Table 4.2: Growth on Different Carbon Sources.

| C source | | | C source | | |
|---|---|---|---|---|---|
| Substrate | Exchange Rxn | Growth | Substrate | Exchange Rxn | Growth |
| acetate | EX_ac(e) | + | L-histidine | EX_his-L(e) | + |
| cellobiose | EX_celb(e) | + | L-isoleucine | EX_ile-L(e) | + |
| citrate | EX_cit(e) | + | L-lactate | EX_lac-L(e) | + |
| D-arabinotol | EX_abt-D(e) | + | L-leucine | EX_leu-L(e) | + |
| D-fructose | EX_fru(e) | + | L-lysine | EX_lys-L(e) | + |
| D-galactose | EX_gal(e) | + | L-malate | EX_mal-L(e) | + |
| D-gluconate | EX_glcn(e) | + | L-methionine | EX_met-L(e) | - |
| D-glucose | EX_glc(e) | + | L-phenylalanine | EX_phe-L(e) | - |
| D-mannitol | EX_mnl(e) | + | L-proline | EX_pro-L(e) | + |
| D-mannose | EX_man(e) | + | L-rhamnose | EX_rmn(e) | + |
| D-ribose | EX_rib-D(e) | + | L-serine | EX_ser-L(e) | + |
| D-tartrate | EX_tartr-D(e) | + | L-threonine | EX_thr-L(e) | + |
| D-xylose | EX_xyl-D(e) | + | L-tryptophan | EX_trp-L(e) | + |
| glycerol | EX_glyc(e) | + | L-valine | EX_val-L(e) | + |
| glycine | EX_gly(e) | + | maltose | EX_malt(e) | + |
| lactose | EX_lcts(e) | + | melibiose | EX_melib(e) | + |
| L-alanine | EX_ala-L(e) | + | propionate | EX_ppa(e) | + |
| L-arabinose | EX_arab-L(e) | + | pyruvate | EX_pyr(e) | + |
| L-arginine | EX_arg-L(e) | - | salicin | EX_salcn(e) | + |
| L-asparagine | EX_asn-L(e) | + | succinate | EX_succ(e) | + |
| L-aspartate | EX_asp-L(e) | + (*in vivo*: -) | sucrose | EX_suc(e) | - |

For simulation of biomass formation on different nitrogen sources, NH3 exchange reaction flux, was set to zero. As can be seen in Table 4.2 the model can simulate biomass production on different nitrogen sources.

Table 4.3: Growth on Different Nitrogen Sources

| N-source | | |
|---|---|---|
| Substrate | Exchange Reaction | Growth |
| L-asparagine | EX_asn-L(e) | + |
| L-aspartate | EX_asp-L(e) | + |
| L-glutamate | EX_glu-L(e) | + |
| L-glutamine | EX_gln-L(e) | + |
| L-isoleucine | EX_ile-L(e) | + |
| L-leucine | EX_leu-L(e) | + |
| L-lysine | EX_lys-L(e) | + |
| L-methionine | EX_met-L(e) | + (iIB711: -) |
| L-phenylalanine | EX_phe-L(e) | + (iIB711: -) |
| L-proline | EX_pro-L(e) | + |
| L-valine | EX_val-L(e) | + |

As can be seen from the results of model validation, *S. coelicolor* genome-scale metabolic model, iMK1208, is suitable for growth simulation on different conditions.

*S. coelicolor* genome-scale metabolic model (iMK1208), contains a 1436 x 1859 stoichiometric matrix (S). The model S matrix can be visualized as a graph (Figure 4.2). This graph represents all non-zero entries in stoichiometric matrix with a dot. This graph simply shows the connection of reactions and metabolites. As can be seen, most of the reactions involves a few metabolites and only few of them involves many metabolites. In this graph, topology of metabolic network can be seen. X axis shows the reactions and the Y axis shows the metabolites. The metabolites that are involved in many reactions are called currency metabolites like $NAD^+$, ATP and $H^+$, and these kinds of metabolites are shown close to the top of the graph with many dots. The graph looks triangular because the reactions that are listed towards the end of the reactions list do not involve the currency metabolites. The continuous lines inside the upper triangular part shows reactions series.

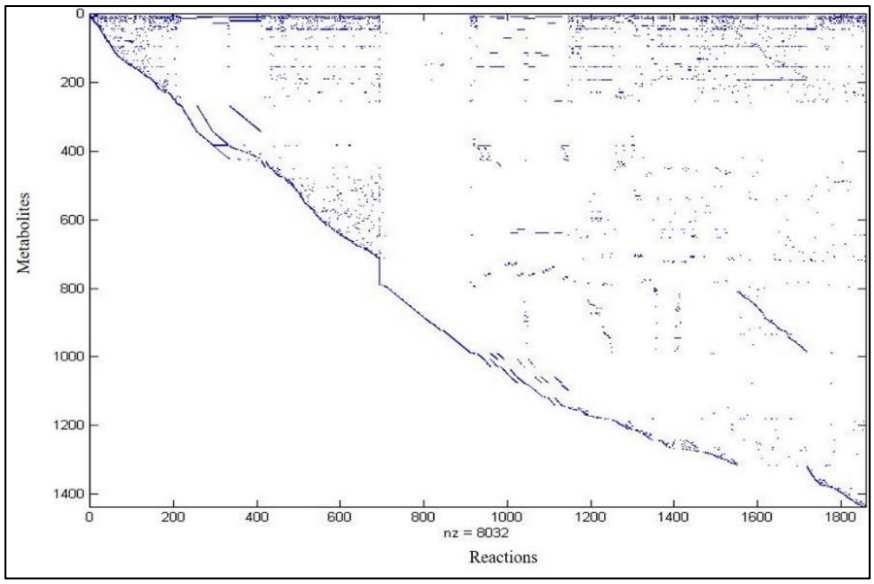


Figure 4.2: The (1436 x 1859) S matrix of Model. All non-zero entries are marked with a dot.

In genome-scale metabolic model of *S. coelicolor*, there are very few metabolites that participate in relatively more reactions, while most of the metabolites participate only in a few reactions (Figure 4.3). The few highly connected metabolites are "global" players, similar to hubs in protein-protein-interaction networks, while the low connected metabolites are "local" players, many of which only occur in linear pathways. The global players are the currency metabolites that are mentioned before. Connectivity of a metabolite shows the number of vertices (reactions) that an edge (metabolite) has.

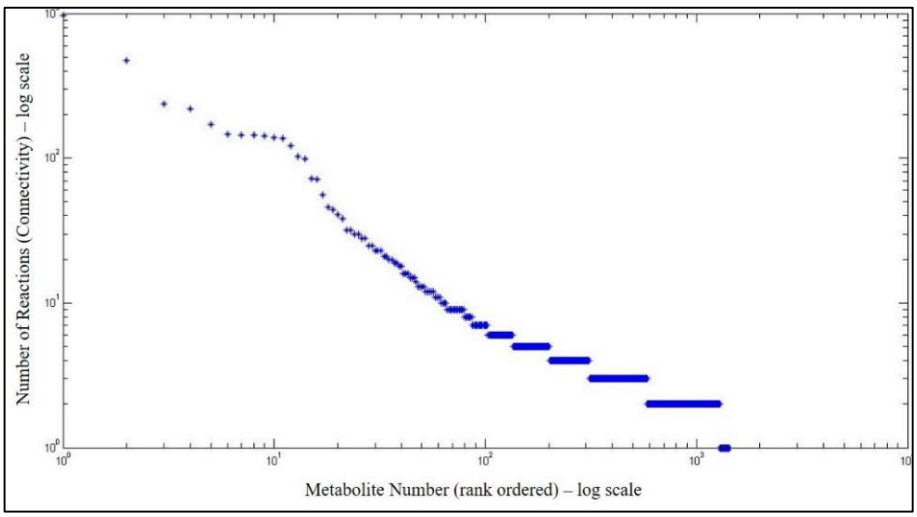


Figure 4.3: Connectivity graph of the model metabolites.

### 4.1.2. Biomass Formation on Different Carbon and Nitrogen Sources

After the validation of formation of biomass on different carbon sources, next step to validate the model is comparing the amount of biomass on different carbon sources. For this analysis, total amount of carbon atom value was selected as 6 that means if a carbon source has 3 carbon atoms, its exchange flux value will be set to 2 mmol/g h, while a carbon source with 6 carbons such as glucose has 1 mmol/g.h exchange flux rate. This arrangement on uptake rate values was done to make the results of the simulations comparable.

For analyses that simulate the growth on different amino acids as nitrogen sources, $NH_4$ uptake flux was set to 0 and biomass formation was simulated via FBA in the same glucose and oxygen uptake rates. Results were compared with experimental values of different *Streptomyces* species.

In previous studies, which include experimental and computational results of biomass production rates, genome-scale models were validated on different carbon sources. Because of the lack of enough experimental data about growth of *S. coelicolor* on different carbon sources, experimental data of other *Streptomyces* species (*S. lividans, S. kanamyceticus, S. hygroscopicus, S.pristinaespiralis, Salbidoflavus, S.thermoviolaceus)* were used for comparison in this study. When the FBA results were compared qualitatively with the experimental results of 10 different studies [Borodina et al 2005, Basak et al 1973, Ilic 2010 et al , Voelker et al 2011, Tawfik et al 1991, Romano et al 1957, Hodgson et al 1982, Narayana et al 2008, Gesheva et al 2004, James et al 1991, Melzoch et al 1997], *S. coelicolor* genome-scale model predicted the growth on different carbon sources sufficient to validate the model (Table 4.4).

According to the simulation of growth of *S. coelicolor* on different carbon sources, propionate gives the highest biomass production rate while glycine gives the lowest value (Figure 4.4). Figure 4.5 shows the effect of amino acids on biomass production when they are used as carbon source. Maximum biomass was reached with isoleucine while minimum biomass was obtained when glycine was used as carbon source. Generally, aliphatic amino acids like leucine, isoleucine, proline and valine gave the highest biomass values.

Table 4.4: The Articles of Experimental Evidences About Carbon and Nitrogen Metabolism.

| Article | *Streptomyces* strain | Type |
|---|---|---|
| Borodina et al., (2005) | *Streptomyces coelicolor* | Carbon Source |
| Basak et al., (1973) | *Streptomyces kanamyceticus* | Carbon Source |
| Basak et al., (1973) | *Streptomyces kanamyceticus* | Sugar Alcohols |
| Basak et al., (1973) | *Streptomyces kanamyceticus* | Amino Acids |
| Ilic et al., (2010) | *Streptomyces hygroscopicus* | Carbon Source |
| Voelker et al., (2001) | *Streptomyces pristinaespiralis* | Nitrogen Source |
| Tawfik et al., (1991) | *Streptomyces pristinaespiralis* | Carbon Source |
| Romano et al., (1957) | *Streptomyces fradiae* | Carbon Source |
| Hodgson et al., (1982) | *Streptomyces coelicolor J802* | Carbon Source |
| Madden et al., (1996) | *Streptomyces lividans* | Nitrogen Source |
| Narayana et al., (2008) | *Streptomyces albidoflavus* | Carbon Source |
| Narayana et al., (2008) | *Streptomyces albidoflavus* | Nitrogen Source |
| Gesheva et al., (2004) | *Streptomyces hygroscopicus* | Carbon Source |
| James et al., (1991) | *Streptomyces thermoviolaceus* | Carbon Source |

Total biomass was also calculated when the organic acids were used as carbon source (Figure 4.6). Although the results are close to each other, it is possible to

conclude that the highest biomass was obtained with gluconate and the lowest with citrate.

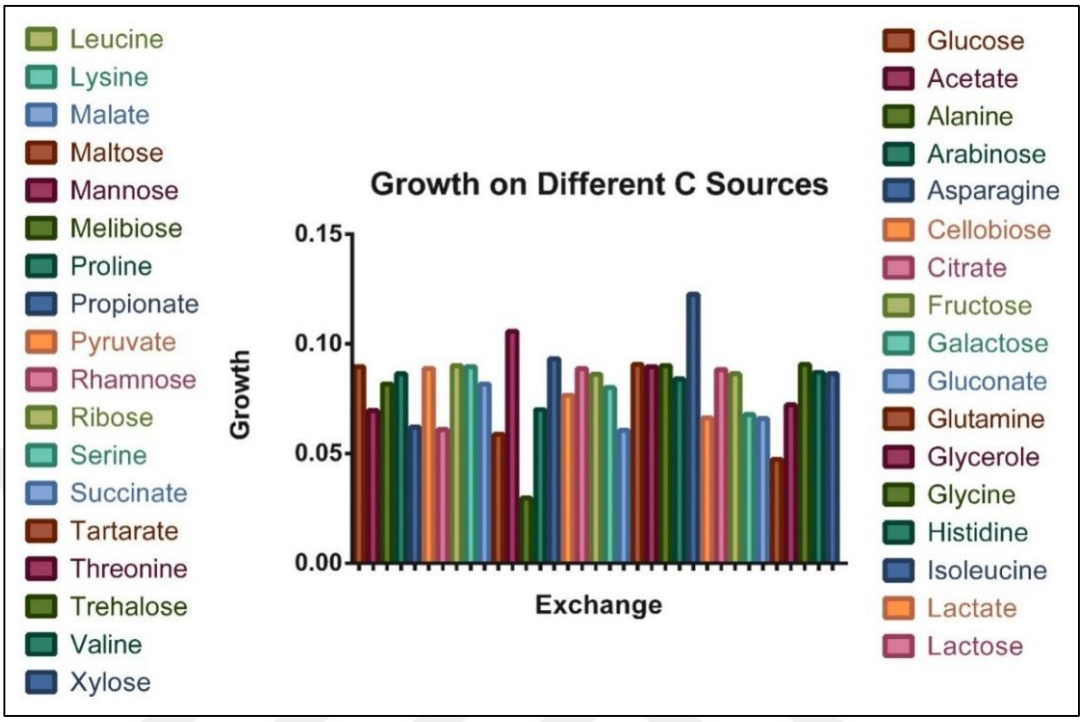


Figure 4.4: Biomass Formation on Different Carbon Sources obtained via FBA simulation.

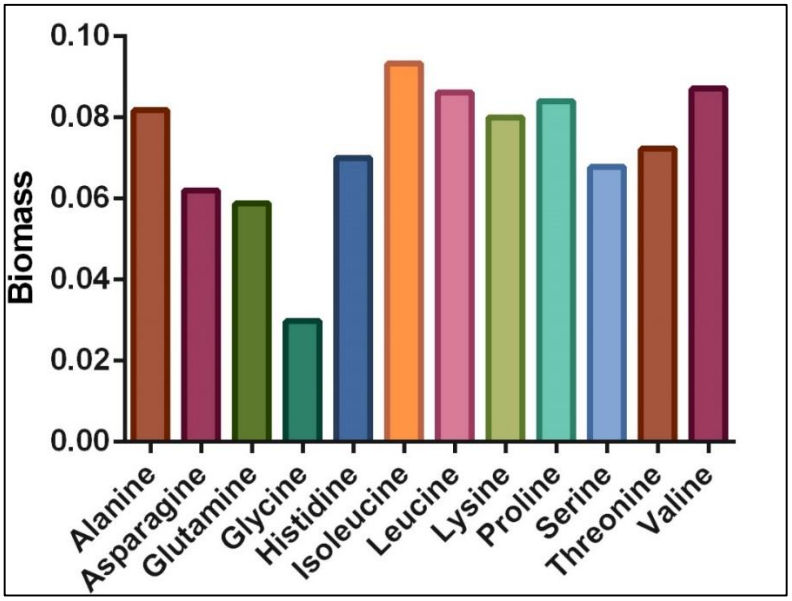


Figure 4.5: Effects of Amino Acids as Carbon Sources on Growth rate, simulated by FBA.

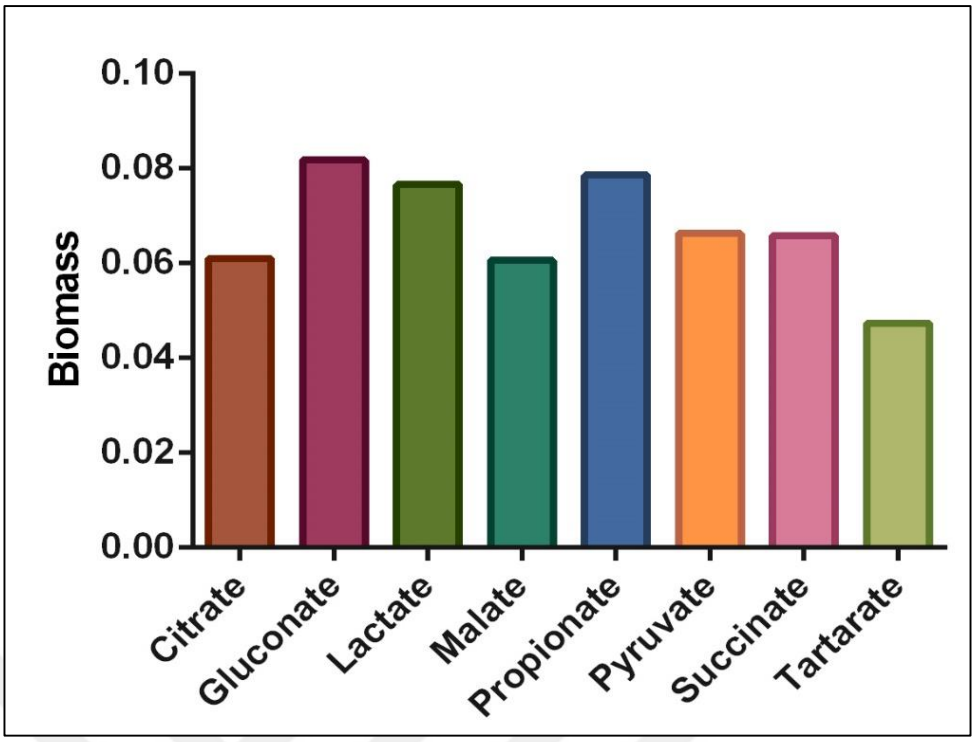


Figure 4.6: Effects of Organic Acids as Carbon Sources on Growth rate, simulated by FBA.

Biomass rates were similar to each other when different pentoses were used as carbon source (Figure 4.7). Different monosaccharides (Figure 4.8) and disaccharides (Figure 4.9) didn't affect the biomass since FBA results are similar to each other.

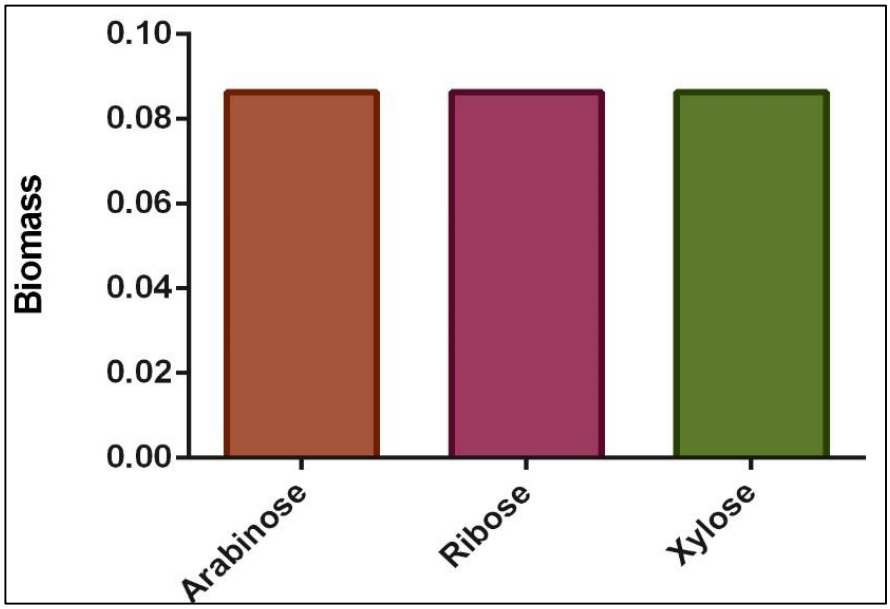


Figure 4.7: Effects of Pentoses as Carbon Sources on Growth rate, simulated by FBA.

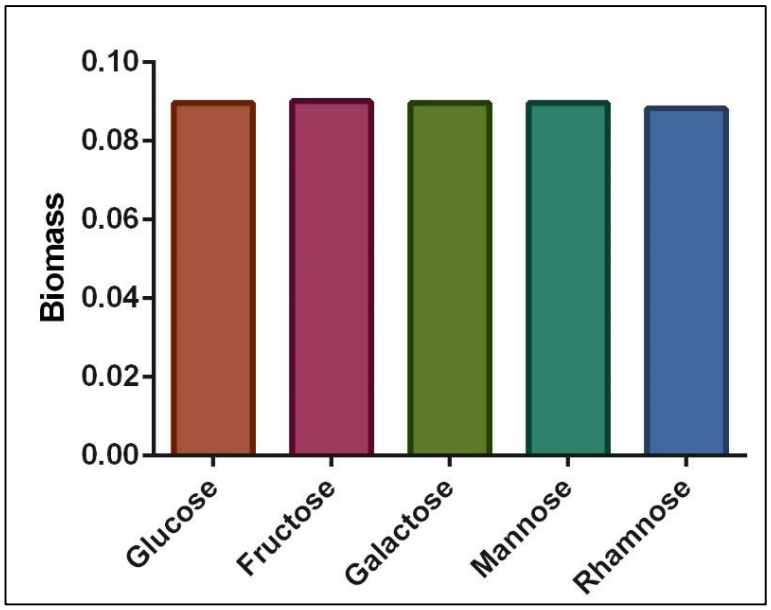


Figure 4.8: Effects of Monosaccharides as Carbon Sources on Growth rate, simulated by FBA.

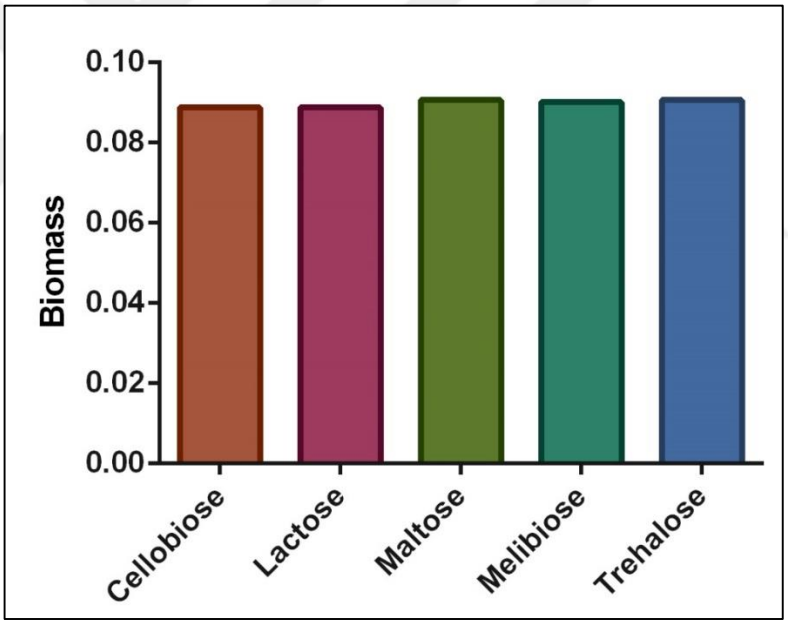


Figure 4.9: Effects of Disaccharides as Carbon Sources on Growth rate, simulated by FBA.

The ability of the model to simulate the growth on different carbon sources is promising. More than 60% of the simulation results are compatible with experimental ones. It can be concluded that computational results can be used to predict the experimental data.

In Figure 4.10, effects of amino acids as nitrogen sources on biomass production rate were shown. Isoleucine has the lowest value of all. Glutamate has the highest

value. This result suggests that glutamate supports *S. coelicolor* growth better than the other amino acids when it is used as nitrogen source.

Effects of ammonia, sulfate and phosphate on biomass production were also simulated by FBA in this study (Figure 4.11). For this simulation, all exchange flux rates except the one that was simulated was set to a fixed value and FBA was applied to model. The highest biomass production was obtained with sulfate while the highest biomass production was reached with ammonia.

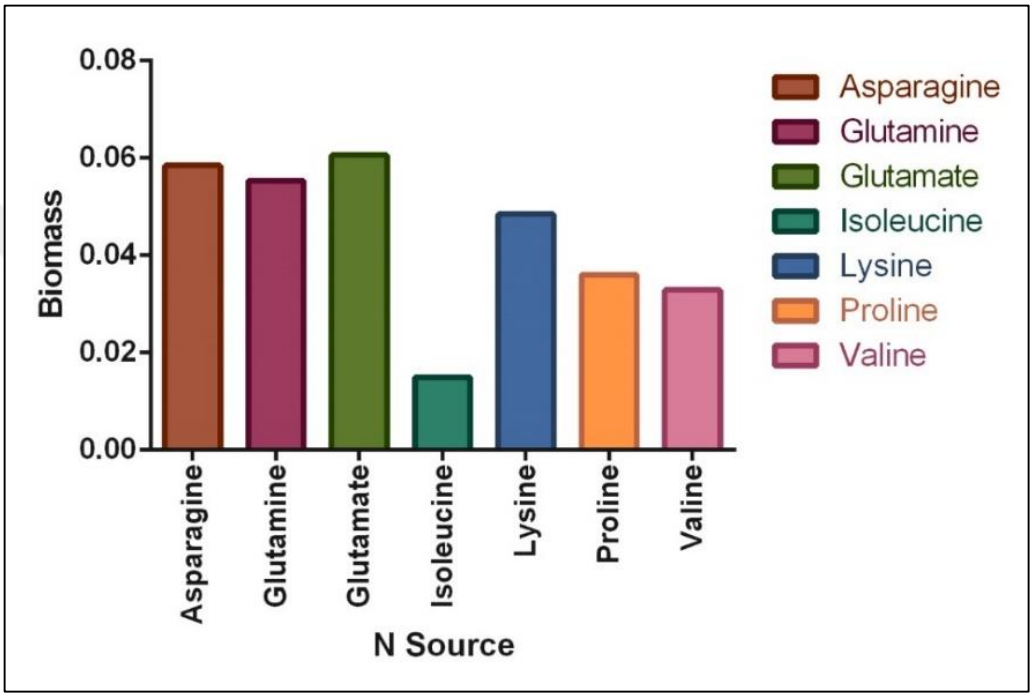


Figure 4.10: Effects of Amino Acids as Nitrogen Sources on Biomass Formation.

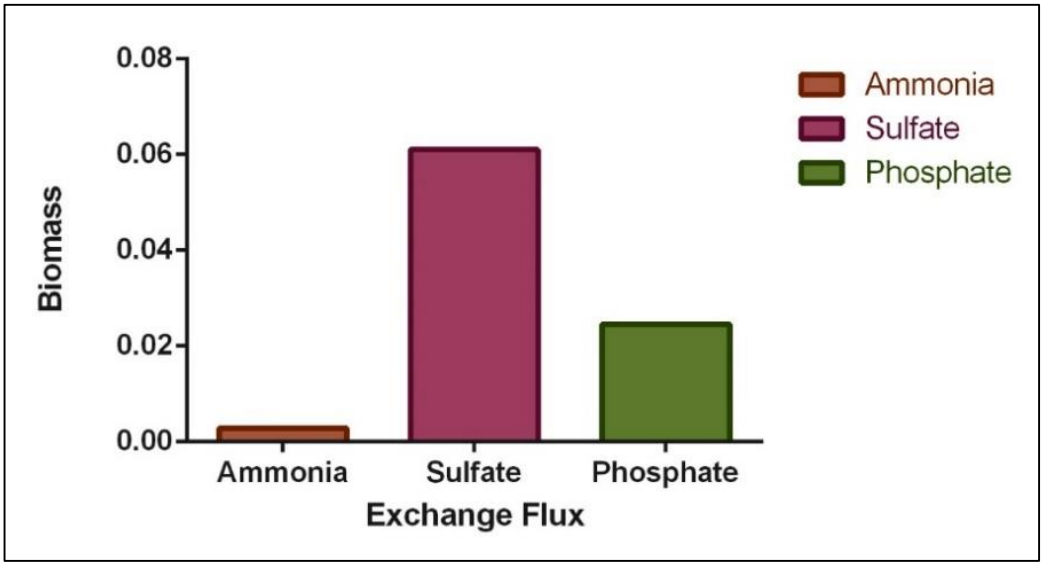


Figure 4.11. Effects of Ammonia, Sulfate and Phosphate on Biomass Formation.

### 4.1.3. Actinorhodin Production on Different Carbon Sources

Actinorhodin (ACT) is a benzoisochromanequinone antibiotic produced by *S. coelicolor*. It is the most common secondary metabolite that is used in studies of *S. coelicolor* metabolism. As mentioned before, secondary metabolism competes with biomass formation. Due to this, when actinorhodin production is maximized in model on a specific condition, biomass production is zero because all carbon is used for actinorhodin production. To obtain results closer to in vivo, a new prediction approach was used in this study. The model, iMK1208, has a non-growth associated ATP maintenance reaction that expresses the minimum ATP production rate, which is 3.65 mmole/g/hr, that cell needs to survive. Each different carbon source has a unique minimum amount of uptake flux value to provide the minimum ATP value. To simulate the actinorhodin production on different carbon sources, the biomass production was determined according to the minimum ATP value. Shortly, the model used the given carbon source for survival (based on non-growth associated ATP mantenance reaction), and then used the residual carbon source to produce ACT. Similar to the simulations on biomass formation, the amount of carbon atoms was set as six.

To compare the results, different antibiotic production rates of different *Streptomyces* species on different carbon sources were used. Although most of the *Streptomyces* species produce different types of antibiotics since they belong to a single branch of phylogenetic tree, their secondary metabolism shows similar behaviour. Figure 4.12 shows the effect of different carbon sources on actinorhodin production. Glutamine has the highest actinorhodin production of all, while glycine is still the lowest one. Asparagine, on the other hand has the second lowest value. According to these results, *S. coelicolor* produces the highest level of actinorhodin while it uses glutamine as carbon source compared to other carbon sources.

In order to compare the results easily and to have better projection, Figure 4. 12 was divided into different graphics (Figure 4.13, Figure 4.14, Figure 4.15, Figure 4.16, Figure 4.17).

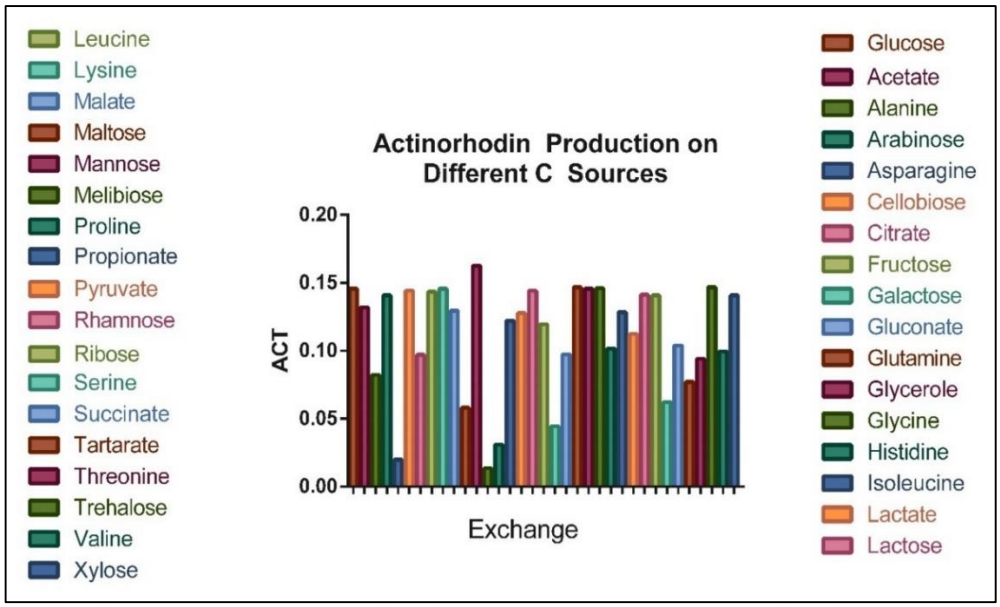


Figure 4.12. Effects of Different Carbon Sources on Actinorhodin Production.

In Figure 4.13 the effects of amino acids as carbon source on actinorhodin production was shown. Isoleucine has the highest value and glycine is the lowest one just like the biomass production results in Figure 4.5. Asparagine is the most remarkable amino acid in the results. It gave relatively high biomass production rate but has really low actinorhodin production rate. This simulation suggests that, compared to other amino acids, *S. coelicolor* produces the highest level of actinorhodin when it uses isoleucine and lysine as carbon sources.

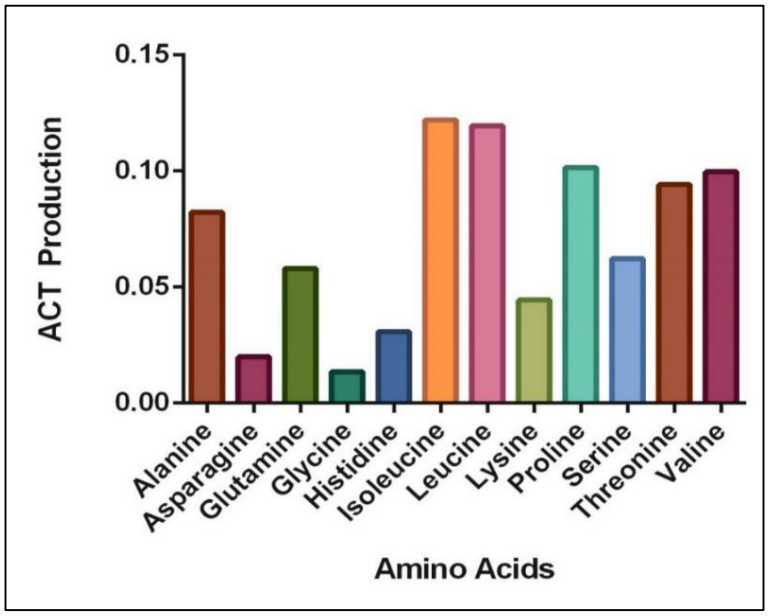


Figure 4.13. Effects of Amino Acids on Actinorhodin Production.

In Figure 4.14, effects of organic acids on ACT production were shown. Gluconate, lactate and propionate resulted in the highest values while citrate gave the lowest. This simulation result suggests that *S. coelicolor* produces the highest level of actinorhodin when it uses gluconate, lactate or propionate as carbon sources when compared to other organic acids.

Different monosaccharides, disaccharrides and pentoses did not affect the ACT production (Figure 4.15, Figure 4.16, Figure 4.17).

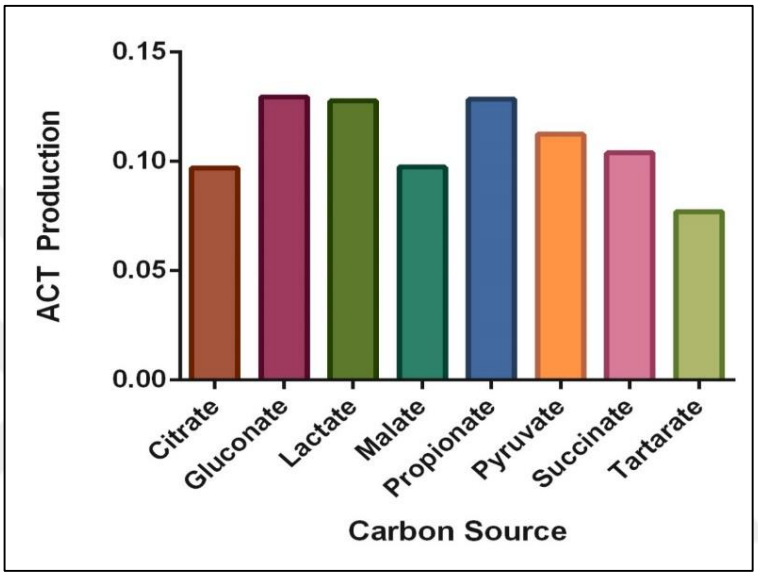


Figure 4.14: Effects of Organic Acids on Actinorhodin Production.

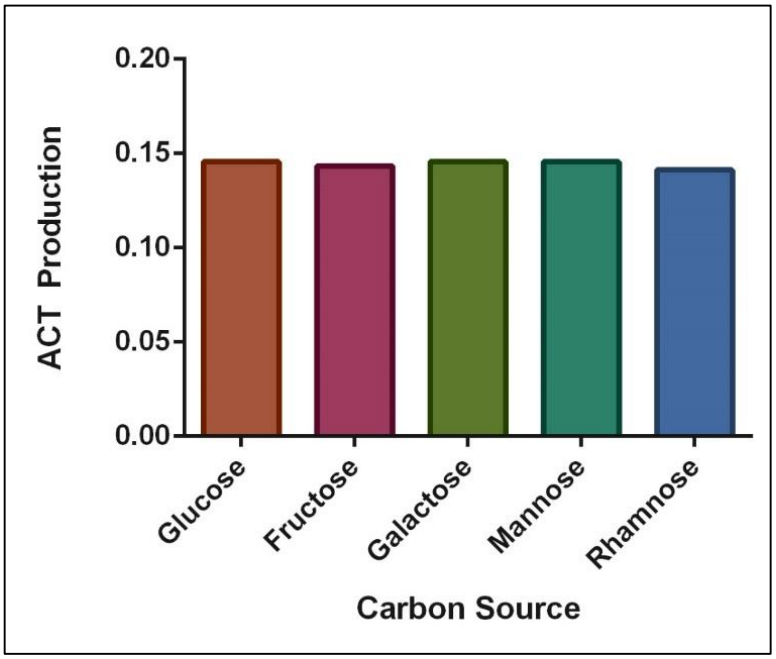


Figure 4.15: Effects of Monosaccharides on Actinorhodin Production.

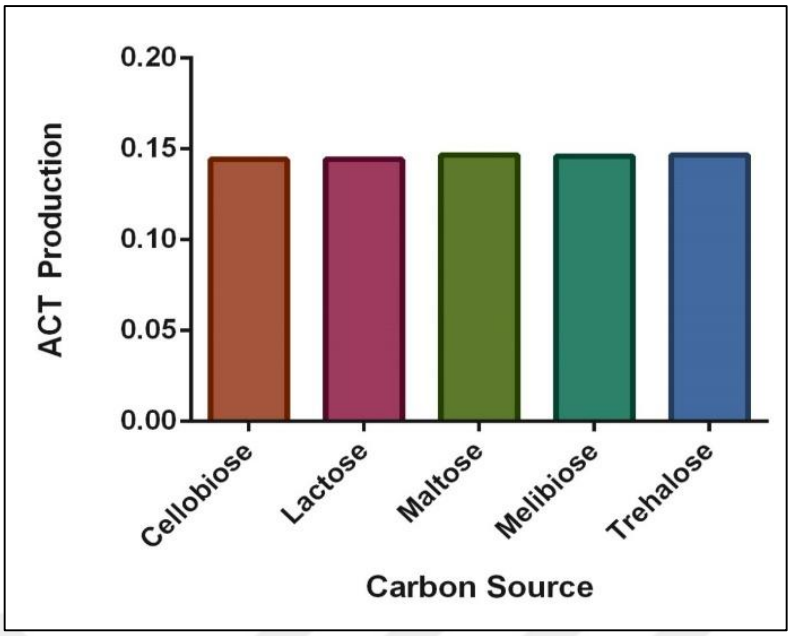


Figure 4.16: Effects of Disaccharides on Actinorhodin Production.

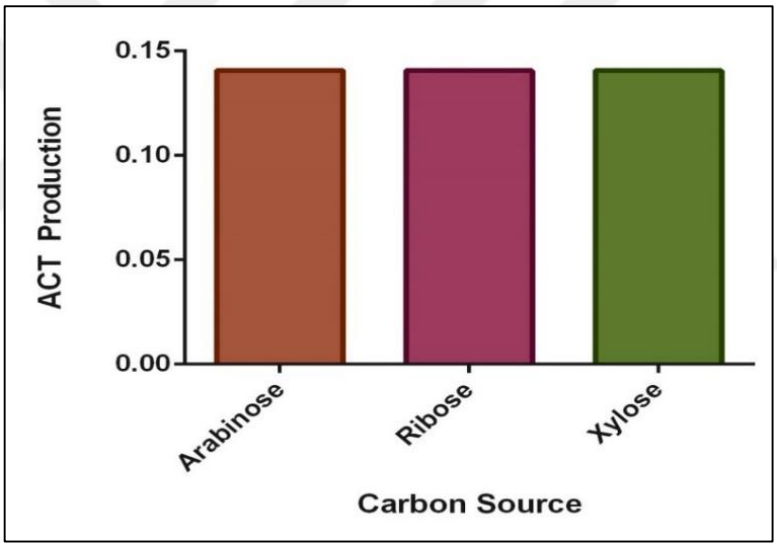


Figure 4.17: Effects of Pentoses on Actinorhodin Production.

## 4.2. Double Gene Deletion Analyses

Double Gene Deletion analysis is a tool that knocks outs the genes in a genome scale metabolic model in pairs to analyse the effects on objective function. The *S. coelicolor* genome-scale metabolic model involves 1209 genes. With double gene deletion analysis tool of COBRA Toolbox, effects of knock out of double genes on growth (Figure 4.18) and actinorhodin production (Figure 4.19) can be revealed.

The results of these analyses are given in 1209 x 1209 matrixes in Matlab 2011r software. There is not any increase on biomass production rate with any gene pair

knock out. An essential gene for biomass production such as SCO0386, asparagine synthetase, show biomass production “0” when it was knocked out with all other genes. Analysis results show that 228 gene pairs affect biomass production negatively. All genes that show biomass production rate as zero when knocked out when it paired with itself on result matrix, also shows no growth with any other genes that knocked out together. In results 27 of these genes effects biomass production but does not make it “0”. Glycine dehydrogenase (SCO1378) causes a small decrease in biomass production when paired with any of the genes. Other gene deletions are lethal. The gene SCO2149, the iron-sulfur protein, decreases biomass production rate and when it is knocked out with gene SCO1378, glycine dehydrogenase, it causes more decrease on biomass production rate. The gene SCO1464, ribulose phosphate 3 epimerase, has the same effect when paired with SCO1378.

Same analysis was done for actinorhodin production on *S. coelicolor* (Figure 4.19) and 34 genes were found to affect actinorhodin production negatively. There is not any increase with any gene pair knock-out. The gene SCO0617, phosphoketolase, and SCO1446, a hypothetical protein, decrease actinorhodin production when it is knocked out but when they are knocked out together. However when SCO1946, phosphoglycerate kinase, it causes more decrease on actinorhodin production.

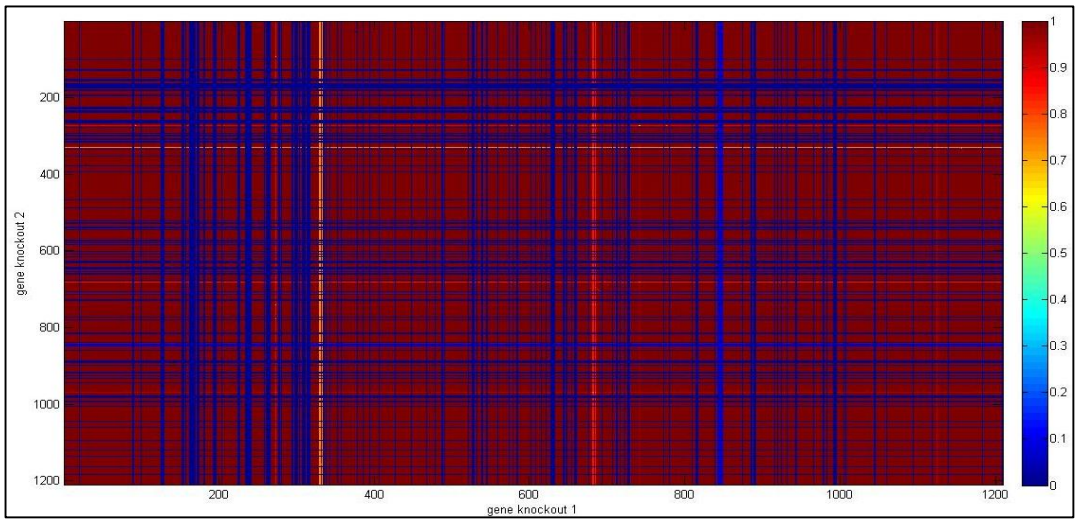


Figure 4.18: Double Gene Deletion Analysis. (Objective Function : Biomass Formation) (Red color shows same flux value as wild type model and blue shows 0 as production. Color scales are shown beside).

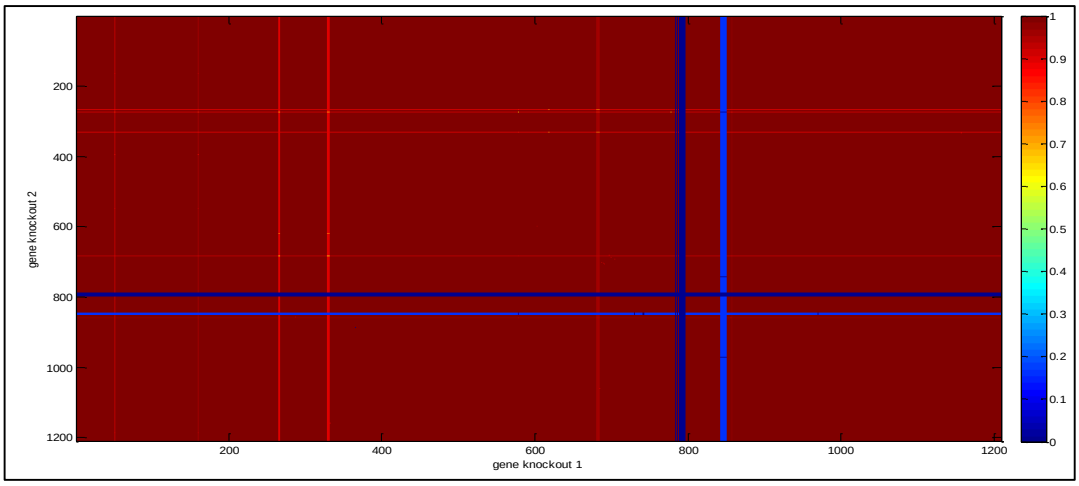


Figure 4.19: Double Gene Deletion Analysis. (Objective Function : Actinorhodin Production) (Red color shows same flux value as wild type model and blue shows 0 as production. Color scales are shown beside).

## 4.3. Principle Component Analyses and Cluster Analyses of Transcriptome Data Sets

The transcriptome data sets (GSE58666, GSE31068, GSE18489) were checked with PCA and cluster analysis to detect the resolution of datasets and detect the outliers. PCA and cluster analyses show that datasets are well defined, and similar data form groups together. PCA and cluster analyses were used to show the suitability of transcriptome data sets. There are two types of PCA results in Figure 4.20, Figure 4.21, Figure 4.22 and Figure 4.23. Time-dependent wild-type data were shown in Figure 4.20 and Figure 4.21.

In Figure 4.20 all data of GSE18489 are shown in PCA graph. Close time points are neighbour in the graph. This shows that the data set is proper and close time points are grouped together and there is not any outliers that must be removed.

In Figure 4.21 only wild-type data of GSE58666 are shown in PCA graph. The graph shows that the data set is really good to use. The data has 5 time points and they are all grouped together and there is not any outliers.

Comparison of wild-type and mutant transcriptome data were shown in Figure 4.22 and Figure 4.23. In Figure 4.22 all data of GSE31068 (wild type and Δ*phoP* transcriptome data) and in Figure 4.23 all data of GSE58666 (wild type and Δ*argR* transcriptome data) were shown in PCA graph. As can be seen in the graph, *ΔphoP* data and wild-type data are grouped together as in the case of *ΔargR* data and wild-type data.

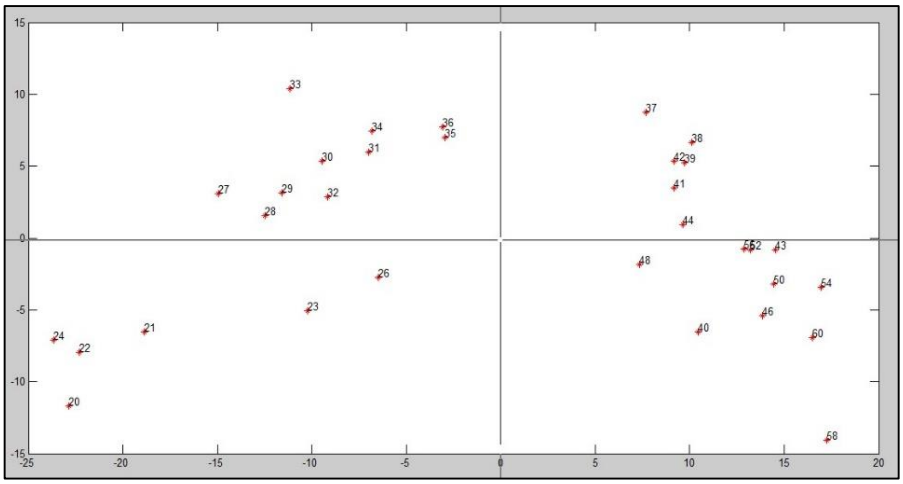

Figure 4.20: PCA of Time-Course Transcriptome Analysis of *S. coelicolor* (GSE18489) (Red Dots Show a certain sub-data) (Numbers are time points).

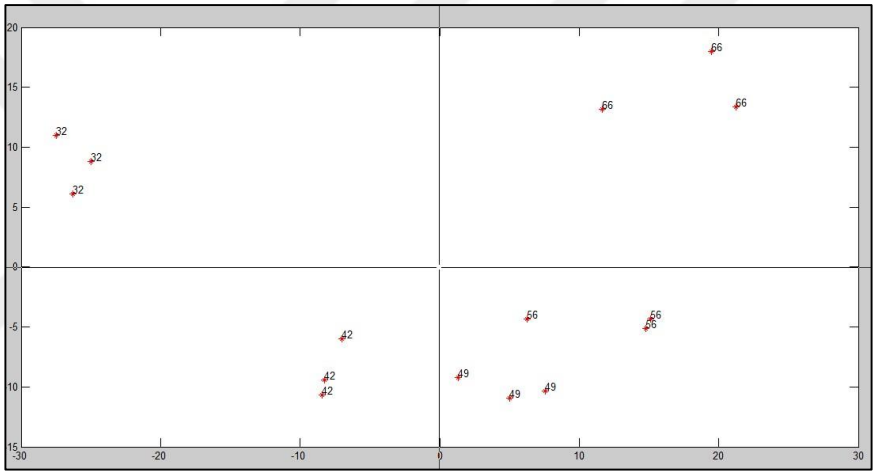

Figure 4.21: PCA of ΔargR mutant of *S. coelicolor* (GSE58666) (Only Wild-Type Data) (Red Dots Show a certain sub-data) (Numbers are time points).

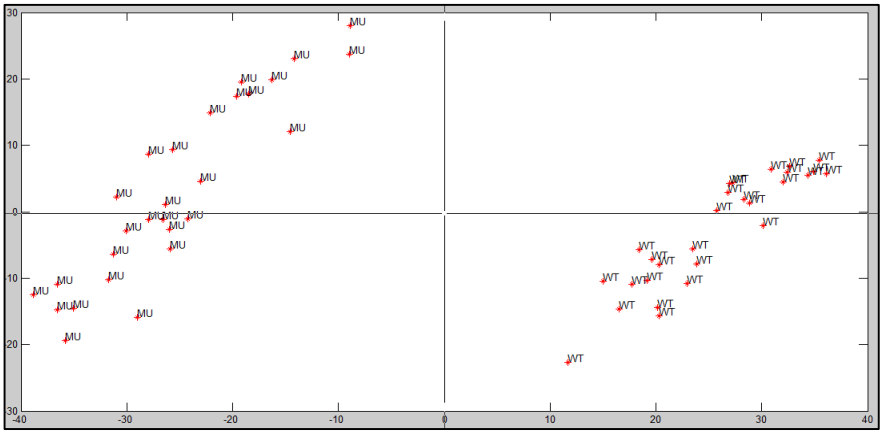

Figure 4.22: PCA of Wild-type vs. Δ*pho*P data set (GSE 31068) (WT = Wild-Type MU = Mutant Strain).

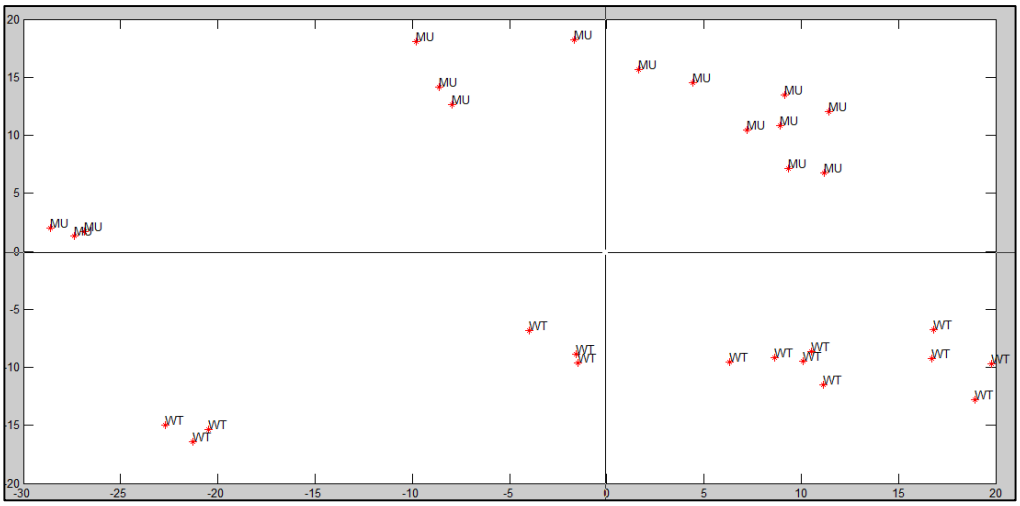


Figure 4.23: PCA of Wild-Type vs. Δ*arg*R data set (GSE58666) WT = Wild-Type MU = Mutant Strain.

Cluster analyses show the grouping of datasets. Data collected at the same hour as replicates are grouped together. In Figure 4.24 and Figure 4.25, only wild-type transcriptome data are used. In Figure 4.26 and Figure 4.27, wild-type and mutant strain data were used. Cluster graphs show similar results to PCA graphs. In Figure 4.24, only wild type data of GSE58666 was shown in a cluster graph. Replicate data points are clustered together just like in PCA graph. All of the data that belong to the same time point create a cluster together.

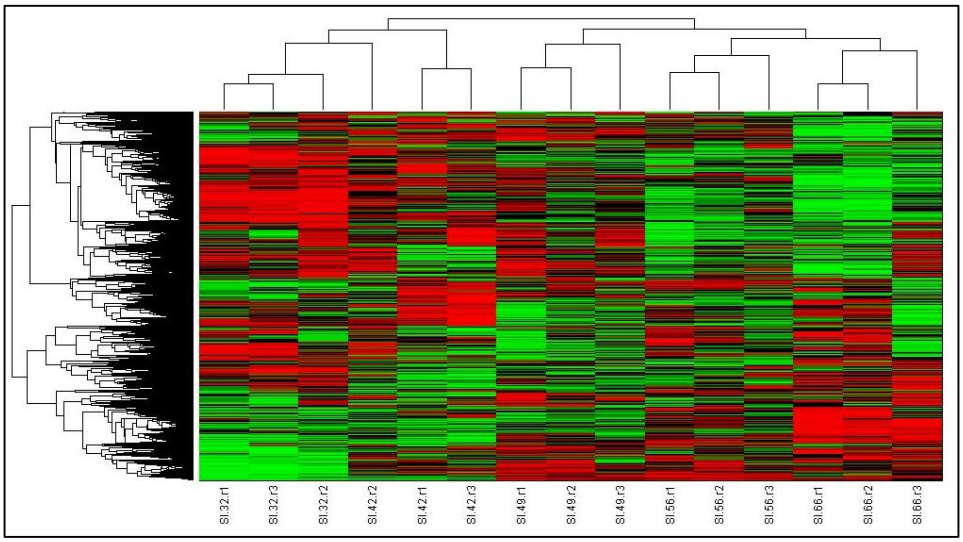


Figure 4.24: Cluster Analysis of Wild Type Expression Data Part of GSE58666 (SI = Wild Type strain of *S.coelicolor*, numbers = time points, r = replicate*)*( Red and green shows the expression profiles of genes to create gene clusters. Because of there are too many genes to show in cluster graph, gene names are not included in cluster analyses).

In Figure 4.25, wild type expression profile GSE18489 was shown in cluster graph. Just like the PCA graph of this data, same time points are clustered together. Just like in Figure 4.24, gene names are not shown in cluster graph.

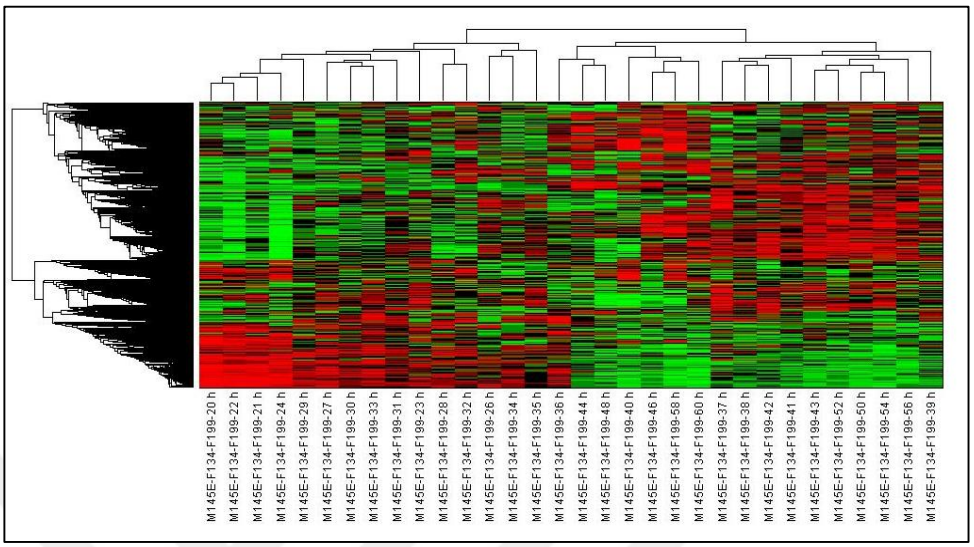


Figure 4.25: Cluster Analysis of Wild Type Expression Data of GSE18489 (M145E = Wild Type strain of *S.coelicolor*, F134-F199 = Fermentor and sample number) ( Red and green shows the expression profiles of genes to create gene clusters. Because of there are too many genes to show in cluster graph, gene names are not included in cluster analyses).

In Figure 4.26 GSE58666 data (*ΔargR* and wild-type expression profiles) was shown in cluster graph. Cluster graph shows that most of the data of wild type and mutant are clustered in itself. Only the 32$^{nd}$ hour data of wild type strain is clustered with 32$^{nd}$ hour data of *ΔargR* strain but in that cluster, the wild-type data are clustered in the same smaller cluster. This graph suggests that wild type and mutant strain data are grouped together and they can be used to compare the expression profiles.

In Figure 4.27 the data of GSE18489 and GSE31068 were shown in cluster graph. As mentioned before, wild type expression data and Δ*phoP* strain expression data are sub-data sets and to use in cluster analyses, mutual time points of each data are extracted first. As can be seen in the graph, wild type and Δ*phoP* data sets are clustered together and they can be used to compare the wild-type and mutant strain.

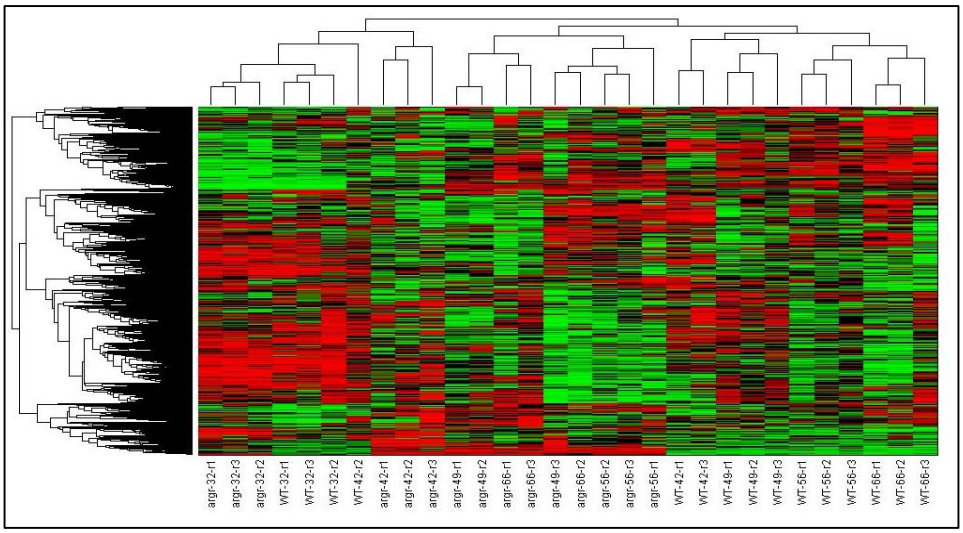


Figure 4.26: Cluster Analysis of All Expression Data of GSE58666 (WT = Wild Type strain of *S.coelicolor* , argr = Δ*argR* strain)( Red and green shows the expression profiles of genes to create gene clusters. Because there are too many genes to show in cluster graph, gene names are not included in cluster analyses).

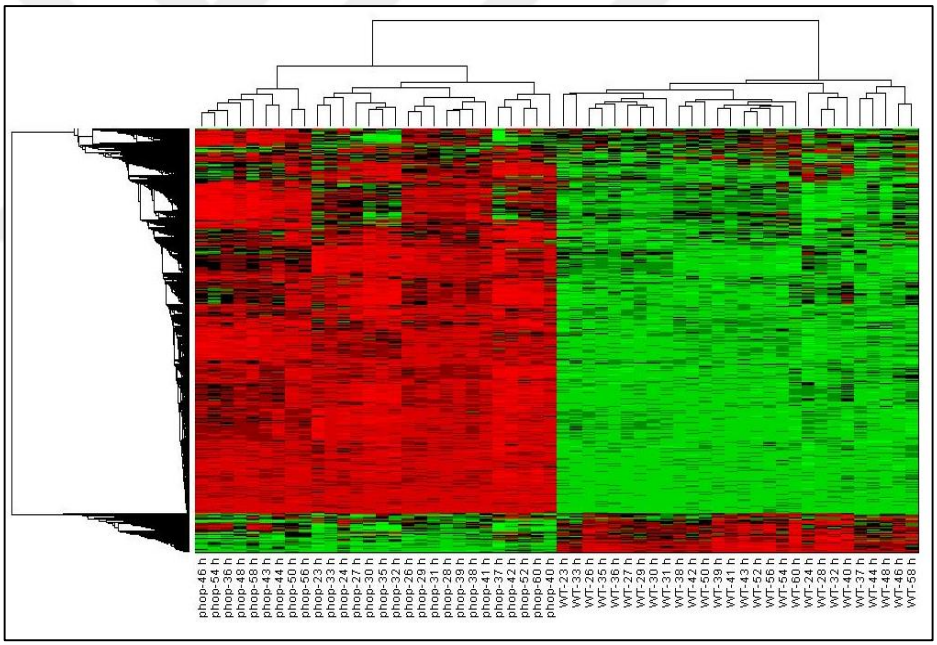


Figure 4.27: Cluster Analysis of All Expression Data of Wild Type and Δ*pho*P mutant (GSE18489 – GSE31068) (WT = Wild Type strain of *S.coelicolor* , phop = Δ*phoP* strain)( Red and green show the expression profiles of genes to create gene clusters. Because there are too many genes to show in cluster graph, gene names are not included in cluster analyses).

Despite their similar objectives, PCA and Cluster Analyses used different parameters and different algorithms to group data. PCA uses eigen values to show closeness of data sets and cluster analysis uses correlation of data sets.

## 4.4. Reporter Metabolite Analysis Results

To reveal the significantly changed genes belonging to primary and/or secondary metabolism under different conditions, t-test was used. To find the significantly changed genes in secondary metabolism, $32^{th}$ and $66^{th}$ hours of wild-type data of Δ*arg*R mutant (GSE58666) and $20^{th}$-$21^{th}$ and $58^{th}$-$60^{th}$ hours of Δ*pho*P time series data (GSE18489) were used. To reveal the alterations of genes between wild-type and mutant strains, all data of the GSE58666 and GSE31068 were used. Wild-type data were used as control. The p-values obtained from student t-test result were used in Reporter Metabolite Analysis mentioned in Materials and Methods section. Results of reporter metabolite analysis were obtained by using two alternative *S. coelicolor* metabolic networks: genome-scale metabolic model, iMK1208 and the network downloaded from BioCyc (Table 4.5 - Table 4.19).

Reporter metabolite analysis gave the most perturbed metabolites that were affected in a certain condition such as deletion of *argR* or *phoP* genes and/or during transition between primary and secondary metabolism. Tables 4.5, Table 4.7, Table 4.9, Table 4.11 and Table 4.13 show the analysis results by using the network of Genome-Scale Metabolic Model, Table 4.6, Table 4.8, Table 4.10, Table 4.12 show the results obtained by using Gene-Metabolite network downloaded from BioCyc database. Tables also show the p-values (lower than 0.05) of reporter metabolites and the numbers of connected genes to the corresponding metabolites.

Table 4.5 and Table 4.6 show the most affected metabolites in the lack of *argR* gene. In the absence of *arg*R gene the most three affected metabolites are Ring 1,2-epoxyphenylacetyl-CoA, phenylacetyl-CoA and taurine (Table 4.5). The metabolite that is significantly perturbed with the highest numbers of connected genes is acetate. This result was obtained by using the genome-scale metabolic model of *S.coelicolor* by Kim et al (2014).

Table 4.5: Reporter Metabolite Analysis Results of Wild Type vs. Δ*arg*R mutant expression profile comparison (Genome-Scale Metabolic Model) (metNGenes=Number of Connected Genes).

| Metabolite Name | P-Value | metNGenes |
|---|---|---|
| Ring 1,2-epoxyphenylacetyl-CoA | 1.38E-05 | 6 |
| Phenylacetyl-CoA | 6.69E-05 | 7 |
| Taurine | $4 \times 10^{-4}$ | 3 |
| Fe3+ | $8 \times 10^{-4}$ | 6 |
| L-Phenylalanine | $9 \times 10^{-4}$ | 8 |
| Fe3+ | 0.001 | 3 |
| (R)-Glycerate | 0.0014 | 9 |
| Taurine | 0.0096 | 5 |
| 2-Hydroxy-3-oxopropanoate | 0.01829 | 3 |
| Acetate | 0.01835 | 21 |
| Cu2+ | 0.01880 | 7 |
| D-Glucose 6-phosphate | 0.02023 | 16 |
| UDPgalactose | 0.02231 | 5 |
| dTDP-4-dehydro-6-deoxy-D-glucose | 0.02385 | 3 |
| Cu+ | 0.02635 | 5 |
| MoaD Protein with bound AMP | 0.02804 | 3 |
| Thiamin | 0.02872 | 3 |
| Thiamin | 0.02872 | 3 |
| L-Glutamine | 0.03487 | 3 |
| Superoxide anion | 0.03555 | 3 |
| 10-Formyltetrahydrofolate | 0.03677 | 5 |
| Hypoxanthine | 0.03868 | 12 |
| 2,3-Dihydrodipicolinate | 0.03902 | 5 |
| 2-Deoxy-D-ribose 1-phosphate | 0.04001 | 3 |
| Cytidine | 0.04041 | 3 |
| Uridine | 0.04042 | 3 |

Table 4.6 shows the results of RMA Wild Type vs. Δ*argR* mutant expression profile comparison. The most three affected metabolites are L-phenylalanine, 3-

hydroxypropanoyl-CoA and chitin. The two metabolites that have the highest number of connected genes are $H_2O$ and urate. This result is obtained via using Gene-Metabolite network downloaded from BioCyc database.

Table 4.6: Reporter Metabolite Analysis Results of Wild Type vs. ΔargR mutant expression profile comparison (BioCyc Gene-Metabolite Network).

| Metabolite | P-values | MetNGenes |
|---|---|---|
| L-phenylalanine | 0.0002 | 5 |
| 3-hydroxypropanoyl-CoA | 0.0007 | 14 |
| chitin | 0.0012 | 14 |
| n+1_a_chitodextrin | 0.0012 | 14 |
| acryloyl-CoA | 0.0037 | 18 |
| xanthine | 0.0078 | 19 |
| n_H2O | 0.0086 | 43 |
| glycolate | 0.0103 | 5 |
| D-glycerate | 0.0111 | 7 |
| a_carboxylate | 0.0192 | 15 |
| urate | 0.0232 | 16 |
| (2S,4S)-4-hydroxy-2,3,4,5-tetrahydrodipicolinate | 0.0244 | 5 |
| dTDP-4-dehydro-6-deoxy-alpha-D-glucopyranose | 0.0262 | 3 |
| 3-aminopropanal | 0.0304 | 4 |
| D-glyceraldehyde | 0.0304 | 4 |
| a_fatty_aldehyde | 0.0304 | 4 |
| octanal | 0.0304 | 4 |
| UDP-alpha-D-galactose | 0.0344 | 4 |
| an_aldehyde | 0.0353 | 9 |
| superoxide | 0.0386 | 3 |
| (R)-lipoate | 0.0412 | 3 |
| a_[lipoyl-carrier_protein]_N6-dihydrolipoyl-L-lysine | 0.0412 | 3 |
| dihydrolipoate | 0.0412 | 3 |
| dimethylallyl_diphosphate | 0.0443 | 5 |
| a_[pyruvate_dehydrogenase_E2_protein]_N6-lipoyl-L-lysine | 0.0455 | 8 |
| 3-oxoadipate | 0.0475 | 3 |

Table 4.7, Table 4.8, Table 4.9 and 4.10 show the most affected metabolites in the lack of *phoP* gene. Table 4.11, Table 4.12, Table 4.13 and Table 4.14 are the results

of reporter metabolite analysis of metabolic switch between primary and secondary metabolism via GSE18489 and GSE58666.

In Table 4.7, comparison of the expression profile of RMA results of Wild Type vs. Δ*phoP* mutant at 23rd and 24th hour was shown. 10 metabolites were significantly affected from the deletion of *pho*P gene. The three most affected metabolites are L-aspartate, 3-Methyl-2-oxobutanoate and glycolate. The significant metabolite with the highest number of connected genes is glycerol. This result was obtained by using the genome-scale metabolic model of *S. coelicolor* by Kim et al (2014).

Table 4.7: Reporter Metabolite Analysis Results of Wild Type vs. ΔphoP mutant 23rd and 24th hour expression profile comparison (Genome-Scale Metabolic Model).

| Metabolite Name | P-Value | metNGenes |
|---|---|---|
| L-Aspartate | 0.0083 | 6 |
| 3-Methyl-2-oxobutanoate | 0.0120 | 18 |
| Glycolate | 0.0192 | 4 |
| 5-Oxo-4,5-dihydrofuran-2-acetate | 0.0229 | 2 |
| L-Glutamate | 0.0231 | 5 |
| (S)-3-Methyl-2-oxopentanoate | 0.0255 | 15 |
| Glycerol | 0.0353 | 19 |
| 2-Oxoadipate | 0.0411 | 6 |
| Glycolate | 0.0428 | 2 |

In Table 4.8 the results of RMA of wild type vs. *ΔphoP* mutant expression profile comparison were shown. The three most affected metabolites in the deletion of *pho*P gene are D-Alanyl-D-alanine, 3-Methyl-2-oxobutanoate and succinyl-CoA. The metabolite that is significantly altered and has the highest number of connected genes is coenzyme A. The result is obtained by the genome scale meabolic model.

Table 4.8: Reporter Metabolite Analysis Results of Wild Type vs. Δ*pho*P mutant expression profile comparison (Genome-Scale Metabolic Model).

| Metabolite Name | P-Value | metNGenes |
|---|---|---|
| D-Alanyl-D-alanine | 0.0036 | 14 |
| 3-Methyl-2-oxobutanoate | 0.0168 | 18 |
| Succinyl-CoA | 0.0171 | 29 |
| N2-Formyl-N1-(5-phospho-D-ribosyl)glycinamide | 0.0186 | 4 |
| 2-Oxoadipate | 0.0228 | 6 |
| Guanine | 0.0231 | 4 |
| 2-(Formamido)-N1-(5-phospho-D-ribosyl)acetamidine | 0.0235 | 4 |
| CMP | 0.0283 | 8 |
| Coenzyme A | 0.0322 | 102 |
| (R)-Pantothenate | 0.0360 | 4 |
| (S)-3-Methyl-2-oxopentanoate | 0.0367 | 15 |
| dATP | 0.0401 | 3 |
| 7,9,12-Octaketide intermediate 1-[acp] | 0.0407 | 3 |
| Farnesyl diphosphate | 0.0431 | 6 |
| 2-Methylbutanoyl-CoA | 0.0448 | 17 |
| Isovaleryl-CoA | 0.0448 | 17 |
| D-Alanyl-D-alanine | 0.0460 | 19 |
| Precorrin 8 | 0.0462 | 3 |

Table 4.9 shows the results of RMA of Wild Type vs. Δ*pho*P mutant by using 23$^{rd}$ and 24$^{th}$ hour expression profile obtained by using Gene-Metabolite Network downloaded from BioCyc database. The two most significantly perturbed metabolites are NAD and deoxynucleotides. NAD is also the significantly affected metabolite that has the highest number of connected genes. This shows the importance of NAD in the metabolism of phosphate. Phosphoenolpyruvate (PEP) shows the effect of *phoP* mutation on central carbon metabolism.

Table 4.9: Reporter Metabolite Analysis Results of Wild Type vs. Δ*pho*P mutant 23rd and 24th hour expression profile comparison (BioCyc Gene-Metabolite Network)

| Metabolite | P-values | MetNGenes |
|---|---|---|
| NAD+ | 0.0058 | 119 |
| (deoxynucleotides)(m) | 0.0098 | 5 |
| (deoxynucleotides)(n+m) | 0.0098 | 5 |
| beta-nicotinamide_D-ribonucleotide | 0.0098 | 5 |
| [dihydrolipoyllysine-residue(2-methylpropanoyl)transferase]lipoyllysine | 0.0151 | 7 |
| [dihydrolipoyllysine-residue_(2-methylpropanoyl)transferase]_ | 0.0151 | 7 |
| S-(2-methylpropanoyl)dihydrolipoyllysine | 0.0151 | 7 |
| (deoxynucleotides)(n) | 0.0294 | 19 |
| (R)-lipoate | 0.0294 | 3 |
| a_[lipoyl-carrier_protein]_N6-dihydrolipoyl-L-lysine | 0.0294 | 3 |
| dihydrolipoate | 0.0312 | 3 |
| L-glutamate_(outside) | 0.0316 | 4 |
| glycerol | 0.0331 | 3 |
| NADH | 0.0388 | 111 |
| phosphoenolpyruvate | 0.0432 | 14 |
| L-arginine | 0.0449 | 5 |
| malonyl-CoA | 0.0449 | 12 |
| a_beta-D_glucoside | 0.0449 | 10 |
| a_non_glucosylated_D-glucose_acceptor | 0.0449 | 10 |
| a_[pyruvate_dehydrogenase_E2_protein]_N6-dihydrolipoyl-L-lysine | 0.0480 | 6 |
| acetone_cyanohydrin | 0.0480 | 8 |
| linamarin | 0.0480 | 8 |

Table 4.10 shows the results of RMA Wild Type vs. Δ*phoP* mutant by using Gene-Metabolite Network that was downloaded from BioCyc database. The three most affected metabolites in the Δ*phoP* strain are (R)-lipoate, a_[lipoyl-carrier_protein]_N6-dihydrolipoyl-L-lysine and dihydrolipoate. The metabolite that has the highest number of connected genes is diphosphate.

Table 4.10: Reporter Metabolite Analysis Results of Wild Type vs. Δ*pho*P mutant expression profile comparison (BioCyc Gene-Metabolite Network).

| Metabolite | P-values | MetNGenes |
|---|---|---|
| (R)-lipoate | 0.0061 | 3 |
| a_[lipoyl-carrier_protein]_N6-dihydrolipoyl-L-lysine | 0.0061 | 3 |
| dihydrolipoate | 0.0061 | 3 |
| N2-formyl-N1-(5-phospho-beta-D-ribosyl)glycinamide | 0.0183 | 4 |
| [dihydrolipoyllysine-residue(2-methylpropanoyl)transferase]lipoyllysine | 0.0196 | 7 |
| [dihydrolipoyllysine-residue_(2-methylpropanoyl)transferase]_ | 0.0196 | 7 |
| S-(2-methylpropanoyl)dihydrolipoyllysine | 0.0196 | 7 |
| GMP | 0.0212 | 4 |
| dimethylallyl_diphosphate | 0.0221 | 5 |
| 2-(formamido)-N1-(5-phospho-beta-D-ribosyl)acetamidine | 0.0231 | 4 |
| coenzyme_A | 0.0238 | 75 |
| a_[pyruvate_dehydrogenase_E2_protein]_N6-dihydrolipoyl-L-lysine | 0.0246 | 6 |
| GDP | 0.0269 | 16 |
| geranylgeranyl_diphosphate | 0.0334 | 6 |
| succinate | 0.0364 | 25 |
| succinyl-CoA | 0.0369 | 17 |
| a_thioredoxin | 0.0419 | 3 |
| succinate_semialdehyde | 0.0430 | 4 |
| guanine | 0.0431 | 3 |
| diphosphate | 0.0434 | 126 |
| (deoxynucleotides)(n) | 0.0436 | 19 |

In Table 4.11, the results of RMA of $32^{nd}$ and $66^{th}$ hour expression profile of wild type data of GSE58666 were compared to reveal metabolic changes during transition between primary and secondary metabolism. This result is obtained by genome scale metabolic network of Kim et al (2014). The most significantly affected metabolite is hydrogen peroxide. The most remarkable metabolites in Table 4.11 are acyl-carrier proteins.

Table 4.11: Reporter Metabolite Analysis Results of Wild Type data of GSE58666 when 32nd and 66th hour expression profile are compared (Genome Scale Metabolic Model).

| Metabolite Name | P-Value | metNGenes |
|---|---|---|
| Hydrogen peroxide | 0.0002 | 25 |
| Ferrioxamine B | 0.0012 | 4 |
| Ferrioxamine B | 0.0012 | 4 |
| Ferricytochrome c | 0.0032 | 11 |
| Ferrocytochrome c | 0.0032 | 11 |
| Superoxide anion | 0.0039 | 3 |
| GMP | 0.0048 | 7 |
| 3-Methylbut-2-enoyl-CoA | 0.0061 | 3 |
| Pyridoxal 5-phosphate | 0.0096 | 8 |
| 3-(Imidazol-4-yl)-2-oxopropyl phosphate | 0.0133 | 4 |
| Dihydroxyacetone phosphate | 0.0146 | 15 |
| 3-Oxoisodecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoisoundecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoisododecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoisotridecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoisotetradecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoisopentadecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoisohexadecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoisoheptadecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoisooctadecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoisoheptanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoisooctanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoisononanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| Orotate | 0.0163 | 4 |
| (R)-Pantothenate | 0.0163 | 3 |
| L-Citrulline | 0.0191 | 4 |
| L-Histidinol phosphate | 0.0212 | 4 |
| dCTP | 0.0221 | 3 |
| 3-Oxoante-isoundecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoante-isotridecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoante-isopentadecanoyl-[acyl-carrier protein] | 0.0261 | 10 |

Contination of the Table.

| | | |
|---|---|---|
| 3-Oxoante-isoheptadecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoante-isononanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-oxo-cis-dodec-5-enoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-oxo-cis-myristol-7-eoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-oxo-cis-palm-9-eoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-oxo-cis-vacc-11-enoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxododecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxodecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoheptanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoheptadecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxohexanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxopentanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxotridecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxoundecanoyl-[acyl-carrier protein] | 0.0261 | 10 |
| 3-Oxo-glutaryl-[acyl-carrier protein] methyl ester | 0.0261 | 10 |
| 3-Oxo-pimeloyl-[acyl-carrier protein] methyl ester | 0.0261 | 10 |
| Ferricoelichelin | 0.0300 | 7 |
| Ferricoelichelin | 0.0300 | 7 |
| Pyridoxine 5-phosphate | 0.0315 | 3 |
| L-alanine-D-glutamate | 0.0345 | 5 |
| Propanoyl-CoA | 0.0366 | 22 |
| L-Glutamine | 0.0366 | 3 |
| CDP | 0.0368 | 9 |
| dCDP | 0.0376 | 6 |
| 2-C-methyl-D-erythritol 2,4-cyclodiphosphate | 0.0377 | 11 |
| D-Fructose 1,6-bisphosphate | 0.0381 | 3 |
| 3-Oxoisohexanoyl-[acyl-carrier protein] | 0.0391 | 8 |
| L-alanine-D-glutamate-meso-2,6-diaminoheptanedioate | 0.0396 | 11 |
| dCMP | 0.0431 | 14 |
| 3-Oxododecanoyl-[acyl-carrier protein] (redQ) | 0.0441 | 8 |
| 3-Oxodecanoyl-[acyl-carrier protein] (redQ) | 0.0441 | 8 |
| 3-Oxohexanoyl-[acyl-carrier protein] (redQ) | 0.0441 | 8 |

Contination of the Table.

| | | |
|---|---|---|
| 3-Oxooctanoyl-[acyl-carrier protein] (redQ) | 0.0441 | 8 |
| UTP | 0.0460 | 7 |
| Butanoyl-CoA | 0.0474 | 18 |
| Reduced glutathione | 0.0483 | 4 |
| L-methionine-R-sulfoxide | 0.0483 | 5 |

Table 4.12 shows the result of RMA of 32$^{nd}$ and 66$^{th}$ hour expression profile of Wild Type data of GSE58666 obtained by using Gene-Metabolite network that was downloaded from BioCyc database. The three most significantly changed metabolites are hydrogen peroxide, superoxide and 1,4 alpha-D-glucan. The perturbed metabolite with the highest number of connected genes is propanoyl-CoA.

Table 4.12: Reporter Metabolite Analysis Results of Wild Type data of GSE58666 when 32$^{nd}$ and 66$^{th}$ hours expression profile are compared (BioCyc Gene-Metabolite Network).

| Metabolite | P-values | MetNGenes |
|---|---|---|
| hydrogen_peroxide | 0.0001 | 20 |
| superoxide | 0.0036 | 3 |
| a_1,4-alpha-D-glucan | 0.0041 | 8 |
| acryloyl-CoA | 0.0044 | 18 |
| L-citrulline | 0.0059 | 5 |
| a_beta-lactam | 0.0069 | 6 |
| a_substituted_beta-amino_acid | 0.0069 | 6 |
| 2-oxo-3-phenylpropanoate | 0.0074 | 4 |
| L-phenylalanine | 0.0088 | 5 |
| imidazole_acetol-phosphate | 0.0111 | 4 |
| fructose_1,6-bisphosphate | 0.0113 | 7 |
| L-arabinose | 0.0116 | 3 |
| propanoyl-CoA | 0.0131 | 17 |
| L-histidinol-phosphate | 0.0180 | 4 |
| a_large-branched_glucan | 0.0185 | 7 |
| 3-methylcrotonyl-CoA | 0.0203 | 6 |
| isovaleryl-CoA | 0.0203 | 6 |
| a_linear_malto-oligosaccharide | 0.0231 | 6 |
| a_maltooligosyl-trehalose | 0.0231 | 6 |
| 5-deoxy-dihydrokalafungin | 0.0305 | 3 |
| acrylamide | 0.0335 | 5 |
| acrylate | 0.0335 | 5 |

Contination of the Table.

| | | |
|---|---|---|
| alpha,alpha-trehalose | 0.0376 | 8 |
| a_long-linear_glucan | 0.0380 | 7 |
| RedQ | 0.0418 | 3 |
| RedQ-dodecanoyl_thioester | 0.0418 | 3 |
| alpha-ribazole_5-phosphate | 0.0437 | 3 |
| a_[pyruvate_dehydrogenase_E2_protein]_ N6-dihydrolipoyl-L-lysine | 0.0445 | 6 |
| dUTP | 0.0468 | 3 |
| maltose | 0.0469 | 12 |
| acetaldehyde | 0.0472 | 11 |
| a_ubiquinone[membrane] | 0.0476 | 3 |

Table 4.13 shows results of RMA of $20^{th}$-$21^{st}$ and $58^{th}$-$60^{th}$ hour expression profile comparison of Wild Type data of GSE18489, obtained by using genome scale metabolic model by Kim et al (2014). Results show the most affected metabolites during the switch between primary and secondary metabolism. The three most affected metabolite are D-glucosamine 6-phosphate, L-glutamate and L-aspartate and the metabolite that significantly changed and has the highest number of genes is hydrogen peroxide.

Table 4.13: Reporter Metabolite Analysis Results of Wild Type data of GSE18489, $20^{th}$-$21^{st}$ and $58^{th}$-$60^{th}$ expression profile comparison (Genome Scale Metabolic Model).

| Metabolite Name | Pvalues | metNGenes |
|---|---|---|
| D-Glucosamine 6-phosphate | 0.0003 | 5 |
| L-Glutamate | 0.0009 | 5 |
| L-Aspartate | 0.0016 | 6 |
| Oxidized thioredoxin | 0.0030 | 17 |
| Reduced thioredoxin | 0.0030 | 17 |
| dUDP | 0.0060 | 9 |
| dCDP | 0.0072 | 10 |
| dGDP | 0.0084 | 10 |
| L-Methionine Sulfoxide | 0.0145 | 6 |
| Allantoin | 0.0155 | 5 |
| 3-Phosphoadenylyl sulfate | 0.0172 | 7 |
| dADP | 0.0243 | 11 |
| UDP | 0.0248 | 15 |

Contination of the Table.

| | | |
|---|---|---|
| (S)-2-Aceto-2-hydroxybutanoate | 0.0277 | 6 |
| (S)-2-Acetolactate | 0.0277 | 6 |
| Trehalose | 0.0338 | 6 |
| Trehalose | 0.0343 | 9 |
| GDP | 0.0381 | 19 |
| D-Fructose 6-phosphate | 0.0415 | 16 |
| Hydrogen peroxide | 0.0469 | 26 |
| L-Proline | 0.0489 | 9 |

Table 4.14 shows the RMA results of the same data set used in Table 4.13 by using Gene-Metabolite network downloaded from BioCyc database. The most affected metabolite is L-glutamate taken from outside media. Table 4.14 also includes some secondary metabolite precursors as expected. The metabolite that has the highest number of connected genes is acetyl-CoA which can be thought as an important molecule affecting transition to secondary metabolism.

Table 4.14: Reporter Metabolite Analysis Results of Wild Type data of GSE18489, $20^{th}$-$21^{st}$ and $58^{th}$-$60^{th}$ expression profile comparison (BioCyc Gene-Metabolite Network).

| Metabolite | P-values | MetNGenes |
|---|---|---|
| L-glutamate_(outside) | 0.0032 | 4 |
| malonyl-CoA | 0.0039 | 12 |
| D-glucosamine_6-phosphate | 0.0061 | 5 |
| alpha-D-glucopyranose | 0.0093 | 3 |
| glycine_betaine | 0.0103 | 3 |
| 3-hydroxypropanoyl-CoA | 0.0140 | 14 |
| a_reduced_thioredoxin | 0.0167 | 7 |
| a_2,3,4-saturated_fatty_acyl_CoA | 0.0192 | 16 |
| D-mannopyranose_6-phosphate | 0.0239 | 5 |
| betaine_aldehyde | 0.0260 | 4 |
| (S)-2-aceto-2-hydroxybutanoate | 0.0279 | 7 |
| (S)-2-acetolactate | 0.0279 | 7 |
| D-gluconate_6-phosphate | 0.0308 | 6 |
| L-histidinol-phosphate | 0.0319 | 4 |
| 2-oxo-3-phenylpropanoate | 0.0344 | 4 |
| geranylgeranyl_diphosphate | 0.0352 | 6 |
| 2-methyl-3-hydroxybutyryl-CoA | 0.0368 | 11 |

Contination of the Table.

| | | |
|---|---|---|
| dADP | 0.0380 | 4 |
| dCDP | 0.0380 | 4 |
| dGDP | 0.0380 | 4 |
| dUDP | 0.0380 | 4 |
| acetyl-CoA | 0.0421 | 51 |
| an_oxidized_thioredoxin | 0.0463 | 5 |
| cytidine | 0.0476 | 3 |

## 4.4. Reporter Pathway Analysis Results

Reporter Pathway Analysis (Çakır et al., 2015) was applied to the Gene-Pathway Network obtained from BioCyc database (Table 4.15, Table 4.16, Table 4.17, Table 4.18 and Table 4.19). Same p-values that was previously used in RMAs were used for the analysis. All tables include the p-values and connected gene numbers of each pathway.

Table 4.15 shows the comparison of RPA results of Wild Type vs. ΔargR mutant expression profile. *argR* gene deletion affects 11 pathways. The most significantly perturbed pathway is germicidin biosynthesis pathway.

Table 4.16 shows comparison of RPA result of 23$^{rd}$ and 24$^{th}$ hour expression profile of Wild Type vs. ΔphoP mutant. In the lack of *phoP* gene, the most affected pathways are nucleotide degradation pathways. Some secondary metabolite synthesis pathways or pathways supplying precursors for them are also affected by *phoP* deletion. Also valine, isoleucine, threonine and leucine pathways are affected significantly. Moreover, some pathways that are connected with carbon metabolism is also significantly affected. The pathway that has the highest number of connected genes is streptorubin B biosynthesis. This pathway is a part of secondary metabolism showing that *phoP* is an important gene for secondary metabolism.

Table 4.15: Reporter Pathway Analysis Results of Wild Type vs. Δ*arg*R mutant expression profile comparison (MetNGenes = Connected Number of Genes)

| Pathway | P-values | PathNGenes |
|---|---|---|
| germicidin_biosynthesis | 1.29E-12 | 16 |
| urate_biosynthesis/inosine_5-phosphate_degradation | 0.0023 | 11 |
| methylenomycin_furan_biosynthesis | 0.0108 | 6 |
| adenosine_nucleotides_degradation_I | 0.0136 | 14 |
| (deoxy)ribose_phosphate_degradation | 0.0138 | 19 |
| guanosine_nucleotides_degradation_I | 0.0165 | 10 |
| adenosine_nucleotides_degradation_II | 0.0256 | 13 |
| trans,_trans-farnesyl_diphosphate_biosynthesis | 0.0328 | 5 |
| guanosine_nucleotides_degradation_II | 0.0339 | 10 |
| acrylonitrile_degradation_I | 0.0395 | 4 |
| aldoxime_degradation | 0.0457 | 4 |

Table 4.16: Reporter Pathway Analysis Results of Wild Type vs. Δ*pho*P mutant 23$^{rd}$ and 24$^{th}$ hour expression profile comparison

| Pathway | P-values | PathNGenes |
|---|---|---|
| guanosine_nucleotides_degradation_II | 0.0002 | 10 |
| adenosine_nucleotides_degradation_II | 0.0005 | 13 |
| guanosine_nucleotides_degradation_I | 0.0006 | 10 |
| ethanol_degradation_I | 0.0010 | 7 |
| adenosine_nucleotides_degradation_I | 0.0020 | 14 |
| L-lactaldehyde_degradation_(anaerobic) | 0.0021 | 4 |
| pyruvate_fermentation_to_propanoate_I | 0.0028 | 18 |
| streptorubin_B_biosynthesis | 0.0028 | 44 |
| glyoxylate_cycle | 0.0033 | 13 |
| ethanol_degradation_II | 0.0042 | 12 |
| actinorhodin_biosynthesis | 0.0048 | 28 |
| valine_degradation_III | 0.0053 | 25 |
| aerobic_respiration_I_(cytochrome_c) | 0.0057 | 7 |
| urate_biosynthesis/inosine_5-phosphate_degradation | 0.0057 | 11 |
| pyruvate_fermentation_to_ethanol_I | 0.0072 | 9 |
| nitrate_reduction_I_(denitrification) | 0.0077 | 3 |
| phenylacetate_degradation_I_(aerobic) | 0.0084 | 17 |
| isoleucine_degradation_III | 0.0085 | 23 |
| respiration_(anaerobic) | 0.0087 | 28 |
| superpathway_of_glyoxylate_cycle_and fatty_acid_degradation | 0.0088 | 15 |

Contination of the Table.

| | | |
|---|---|---|
| L-threonine_degradation_II | 0.0102 | 8 |
| pyruvate_decarboxylation_to_acetyl_CoA | 0.0126 | 10 |
| cardiolipin_biosynthesis_I | 0.0130 | 8 |
| 3-oxoadipate_degradation | 0.0177 | 6 |
| propanoyl_CoA_degradation_I | 0.0191 | 9 |
| trans,_trans-farnesyl_diphosphate_biosynthesis | 0.0214 | 5 |
| TCA_cycle_IV_(2-oxoglutarate_decarboxylase) | 0.0228 | 21 |
| phosphoPantothenate_biosynthesis_I | 0.0233 | 16 |
| benzoyl-CoA_degradation_II_(anaerobic) | 0.0285 | 12 |
| anaerobic_aromatic_compound_degradation_ (Thauera_aromatica) | 0.0321 | 5 |
| ppGpp_biosynthesis | 0.0335 | 11 |
| mixed_acid_fermentation | 0.0337 | 33 |
| 2-oxopentenoate_degradation | 0.0377 | 8 |
| streptomyces_coelicolor_butyrolactone_biosynthesis | 0.0381 | 16 |
| pyruvate_fermentation_to_lactate | 0.0382 | 5 |
| lipoate_salvage_I | 0.0387 | 7 |
| formaldehyde_oxidation_IV_(thiol-independent) | 0.0391 | 6 |
| L-leucine_biosynthesis | 0.0405 | 16 |
| guanosine_ribonucleotides_de_novo_biosynthesis | 0.0409 | 15 |
| UDP-alpha-D-glucuronate_biosynthesis_(from_UDP-glucose) | 0.0424 | 6 |
| formate_oxidation_to_CO2 | 0.0426 | 4 |
| leucine_degradation_IV | 0.0442 | 23 |
| fatty_acid_biosynthesis_--_initial_steps_(branched) | 0.0444 | 7 |
| fatty_acid_biosynthesis_--_initial_steps_(straight) | 0.0444 | 7 |
| trans-lycopene_biosynthesis_II_(plants) | 0.0450 | 4 |
| formaldehyde_assimilation_I_(serine_pathway) | 0.0498 | 28 |

Table 4.17 shows the comparison of RPA result of wild type vs. Δ*phoP* mutant expression profile. The most affected pathway is farnesyl diphosphate biosynthesis when *phoP* is deleted. Significantly changed pathway that has highest number of connected genes is salvage pathways of purine and pyrimidine nucleotides which shows *phoP* plays important role in nucleotide metabolism

Table 4.17: Reporter Pathway Analysis Results of Wild Type vs. Δ*pho*P mutant expression profile comparison

| Pathway | P-values | PathNGenes |
|---|---|---|
| trans,_trans-farnesyl_diphosphate_biosynthesis | 0.0004 | 5 |
| anaerobic_aromatic_compound_degradation | 0.0007 | 5 |
| 3-oxoadipate_degradation | 0.0014 | 6 |
| acetate_conversion_to_acetyl-CoA | 0.0019 | 6 |
| pyruvate_fermentation_to_propanoate_I | 0.0025 | 18 |
| aerobic_respiration_I_(cytochrome_c) | 0.0028 | 7 |
| phenylacetate_degradation_I_(aerobic) | 0.0039 | 17 |
| L-threonine_degradation_II | 0.0050 | 8 |
| guanosine_ribonucleotides_de_novo_biosynthesis | 0.0051 | 15 |
| ppGpp_biosynthesis | 0.0055 | 11 |
| ethanol_degradation_II | 0.0060 | 12 |
| phosphoPantothenate_biosynthesis_I | 0.0061 | 16 |
| lipoate_salvage_I | 0.0078 | 7 |
| superpathway_and_fatty_acid_degradation | 0.0079 | 15 |
| anthranilate_degradation_III_(anaerobic) | 0.0087 | 6 |
| trans-lycopene_biosynthesis_II_(plants) | 0.0087 | 4 |
| adenosylcobalamin_biosynthesis_from_cobyrinate_a,c-diamide_I | 0.0103 | 23 |
| actinorhodin_biosynthesis | 0.0118 | 28 |
| germicidin_biosynthesis | 0.0124 | 16 |
| salvage_pathways_of_pyrimidine_ribonucleotides | 0.0142 | 18 |
| ethanol_degradation_I | 0.0147 | 7 |
| guanosine_nucleotides_degradation_II | 0.0154 | 10 |
| thiosulfate_assimilation | 0.0157 | 11 |
| TCA_cycle_IV_(2-oxoglutarate_decarboxylase) | 0.0177 | 21 |
| coenzyme_A_biosynthesis_I | 0.0196 | 12 |
| 5-aminoimidazole_ribonucleotide_biosynthesis_II | 0.0206 | 16 |
| guanosine_deoxyribonucleotides_de_novo_biosynthesis_I | 0.0221 | 8 |
| guanosine_deoxyribonucleotides_de_novo_biosynthesis_II | 0.0221 | 9 |
| guanosine_nucleotides_degradation_I | 0.0248 | 10 |
| enterobactin_biosynthesis | 0.0269 | 7 |
| ethanol_fermentation_to_acetate | 0.0282 | 20 |
| sulfate_reduction_I_(assimilatory) | 0.0288 | 15 |
| acetate_formation_from_acetyl-CoA_I | 0.0300 | 7 |
| acetyl-CoA_dissimilation | 0.0300 | 7 |
| isoleucine_degradation_III | 0.0317 | 23 |
| L-methionine_biosynthesis_I | 0.0365 | 14 |
| mixed_acid_fermentation | 0.0365 | 33 |

Contination of the Table.

| | | |
|---|---|---|
| APS_pathway_of_sulfate_reduction | 0.0381 | 15 |
| 5-aminoimidazole_ribonucleotide_biosynthesis_I | 0.0386 | 17 |
| L-selenocysteine_biosynthesis_I_(bacteria) | 0.0399 | 9 |
| adenosylcobalamin_biosynthesis_from_cobyrinate_a,c-diamide_II | 0.0399 | 25 |
| octane_oxidation | 0.0408 | 18 |
| glyoxylate_cycle | 0.0408 | 13 |
| pantothenate_biosynthesis_II | 0.0409 | 17 |
| sulfate_activation_for_sulfonation | 0.0410 | 7 |
| urate_biosynthesis/inosine_5-phosphate_degradation | 0.0428 | 11 |
| adenosine_nucleotides_degradation_II | 0.0440 | 13 |
| salvage_pathways_of_purine_and_pyrimidine_nucleotides | 0.0445 | 31 |
| UDP-N-acetyl-D-glucosamine_biosynthesis_I | 0.0446 | 13 |

To reveal pathway level changes during transition of primary metabolism to secondary metabolism, RPA results of 32$^{nd}$ and 66$^{th}$ hour expression profile of Wild Type data of GSE58666 were compared (Table 4.18). The most significantly changed pathways are aerobic respiration, urate degradation and superoxide radical degradation. The most significantly changed pathway is aerobic respiration I (cytochrome c). Mixed acid fermentation pathway is also interesting. Succinate, 2-oxoglutarate, acetate, ethanol and lactate biosynthesis are part of mixed acid fermentation pathway. L-histidine biosynthesis is another significantly changed pathway and it has 22 connected genes to it. It shows the importance of L-histidine on secondary metabolism phase.

Table 4.18: Reporter Pathway Analysis Results of Wild Type data of GSE58666, 32$^{nd}$ and 66$^{th}$ expression profile comparison.

| Pathway | P-values | PathNGenes |
|---|---|---|
| aerobic_respiration_I_(cytochrome_c) | 0.0003 | 7 |
| urate_degradation_to_allantoin_I | 0.0006 | 9 |
| superoxide_radicals_degradation | 0.0016 | 5 |
| superpathway_of _fatty_acid_degradation | 0.0031 | 15 |
| pyruvate_fermentation_to_propanoate_I | 0.0036 | 18 |
| mixed_acid_fermentation | 0.0049 | 33 |
| superpathway_of_phenylalanine,_tyrosine and_tryptophan_biosynthesis | 0.0094 | 10 |

Contination of the Table.

| | | |
|---|---|---|
| pyridoxal_5-phosphate_salvage_I | 0.0178 | 10 |
| adenosylcobalamin_biosynthesis | 0.0193 | 25 |
| formaldehyde_assimilation_II_(RuMP_Cycle) | 0.0211 | 13 |
| starch_degradation_I | 0.0212 | 5 |
| L-histidine_biosynthesis | 0.0213 | 22 |
| pyrimidine_ribonucleotides_de_novo_biosynthesis | 0.0276 | 26 |
| NAD_biosynthesis_I_(from_aspartate) | 0.0281 | 20 |
| adenosylcobalamin_salvage_from_cobinamide_I | 0.0284 | 19 |
| octaprenyl_diphosphate_biosynthesis | 0.0363 | 4 |
| UMP_biosynthesis | 0.0396 | 18 |
| nitrate_reduction_I_(denitrification) | 0.0414 | 3 |
| phenylethylamine_degradation_I | 0.0449 | 9 |
| biotin_biosynthesis_III | 0.0465 | 18 |

Table 4.19 shows comparison of the result of RPA of $20^{th}$-$21^{st}$ and $58^{th}$-$60^{th}$ hour expression profile of wild type data of GSE18489. The most significantly changed pathway is butyrolactone biosynthesis pathway. The significantly changed pathways with the highest number of genes are streptorubin B biosynthesis and calcium dependent antibiotic (CDA) biosynthesis pathway. Albaflavenone biosynthesis pathway is another significantly changed pathway which starts with farnesyl diphosphate, a compound that was also found as significantly changed in the lack of *argR* and *phoP* genes. Albaflavenone is a bacterial compound that is found in *S. coelicolor* and it is an anti-microbial agent. This time, this compound is revealed as significantly changed instead of its precursor, farnesyl diphosphate.

Table 4.19: Reporter Pathway Analysis Results of Wild Type data of GSE18489, $20^{th}$-$21^{st}$ and $58^{th}$-$60^{th}$ expression profile comparison (MetNGenes = Number of Connected Genes).

| Pathway | P-values | PathNGenes |
|---|---|---|
| streptomyces_coelicolor_butyrolactone_biosynthesis | 0.0001 | 16 |
| L-arabinose_degradation_IIIa | 0.0056 | 12 |
| guanosine_deoxyribonucleotides_de_novo_biosynthesis_II | 0.0085 | 9 |
| D-mannose_degradation | 0.0088 | 2 |
| guanosine_deoxyribonucleotides_de_novo_biosynthesis_I | 0.0102 | 8 |
| methylenomycin_biosynthesis | 0.0147 | 3 |
| urate_degradation_to_allantoin_I | 0.0148 | 9 |

Contination of the Table.

| | | |
|---|---|---|
| streptorubin_B_biosynthesis | 0.0172 | 44 |
| N-acetylglucosamine_degradation_I | 0.0191 | 7 |
| albaflavenone_biosynthesis | 0.0292 | 11 |
| germicidin_biosynthesis | 0.0368 | 16 |
| starch_degradation_I | 0.0383 | 5 |
| calcium-dependent_antibiotic_biosynthesis | 0.0393 | 46 |
| GDP-mannose_biosynthesis | 0.0413 | 8 |
| heme_biosynthesis_II_(anaerobic) | 0.0491 | 13 |
| trans-lycopene_biosynthesis_II_(plants) | 0.0499 | 4 |

## 4.5. Results of Creation of Condition Specific Models

GIMME algorithm was used for transcriptome data sets via both threshold and fold change comparison methods. The models specific for each mutant were compared and the most suitable ones according to their reaction and metabolite numbers were selected for FBA. The fluxes of mutant models were compared to the fluxes of wild type model and the reactions with the highest differences were selected (Table 4.20 and Table 4.21). Both Table 4.20 and Table 4.21 include reactions, flux differences and the subsystems or pathways with which the reaction is associated.

After obtaining the mutation specific genome-scale models, FBA was applied to these models and the flux rates of internal reactions were compared. Table 4.20, which includes reaction names, flux differences and the subsystem that the reaction belongs, shows the reactions with the highest flux differences between wild-type strain and Δ*arg*R strain. Although the biomass production of mutant and wild type strains were calculated as 0.0699, their internal fluxes were different. The highest flux differences are in alternate carbon metabolism, oxidative phosphorylation and transport from membrane reactions.

Table 4.20: Flux Differences between Wild Type Model and Δ*ArgR* Mutant (Threshold Method) (Table does not include differences lower than 0.01).

| RXNS | Difference | SubSystem |
|---|---|---|
| ALCD19 | 1997.6285 | Alternate Carbon Metabolism |
| IDOND2 | 1000.0000 | Alternate Carbon Metabolism |
| NADTRHD | 1000.0000 | Oxidative Phosphorylation |
| ALAt2r | 1000.0000 | Transport, Membrane |
| CSNt2r | 1000.0000 | Transport, Membrane |
| G3PD2 | 3.4282 | Alternate Carbon Metabolism |
| PYK | 3.2952 | Glycolysis/Gluconeogenesis |
| NADH17 | 3.1358 | Oxidative Phosphorylation |
| PEPCK | 2.9232 | Anaplerotic Reactions |
| NDPK1 | 2.9229 | Nucleotide Salvage Pathway |
| HCO3E | 2.5696 | Unassigned |
| PC | 2.5690 | Pyruvate Metabolism |
| ATPS4r | 2.3453 | Oxidative Phosphorylation |
| MDH | 0.7275 | Citric Acid Cycle |
| PDH | 0.7226 | Glycolysis/Gluconeogenesis |
| HEX1 | 0.5000 | Glycolysis/Gluconeogenesis |
| ICL | 0.4348 | Anaplerotic Reactions |
| MALS | 0.4348 | Anaplerotic Reactions |
| ENO | 0.2927 | Glycolysis/Gluconeogenesis |
| GAPD | 0.2927 | Glycolysis/Gluconeogenesis |
| FBA | 0.2927 | Glycolysis/Gluconeogenesis |
| PFK | 0.2927 | Glycolysis/Gluconeogenesis |
| SUCD3 | 0.2927 | Oxidative Phosphorylation |
| FUM | 0.2927 | Citric Acid Cycle |
| ACONTa | 0.2927 | Citric Acid Cycle |
| ACONTb | 0.2927 | Citric Acid Cycle |
| CS | 0.2927 | Citric Acid Cycle |
| TPI | 0.2926 | Glycolysis/Gluconeogenesis |
| RPI | 0.2926 | Pentose Phosphate Pathway |
| PPA | 0.1480 | Anaplerotic Reactions |
| MYCTR | 0.0483 | Cofactor and Prosthetic Group Biosynthesis |
| TRDR | 0.0481 | Oxidative Phosphorylation |
| PAPSR | 0.0448 | Cysteine Metabolism |
| ALCD2x | 0.0173 | Pyruvate Metabolism |
| AKGDH | 0.0168 | Citric Acid Cycle |
| CYSTGL | 0.0050 | Cysteine Metabolism |
| SHSL2 | 0.0050 | Methionine Metabolism |

Contination of the Table.

| | | |
|---|---|---|
| THRD_L | 0.0033 | Valine, Leucine, and Isoleucine Metabolism |
| HSK | 0.0033 | Threonine and Lysine Metabolism |
| THRS | 0.0033 | Threonine and Lysine Metabolism |
| EX_h(e) | 0.0023 | Exchange |
| POX3 | 0.0023 | Oxidative Phosphorylation |
| ACKr | 0.0023 | Pyruvate Metabolism |
| FE3abc | 0.0018 | Inorganic Ion Transport and Metabolism |
| EARai70y | 0.0014 | Cell Envelope Biosynthesis |
| APTNAT | 0.0013 | Threonine and Lysine Metabolism |
| OXPTNDH | 0.0013 | Threonine and Lysine Metabolism |
| DAPAT | 0.0013 | Threonine and Lysine Metabolism |
| GLCOAS | 0.0013 | Threonine and Lysine Metabolism |
| GLUTCOADHc | 0.0013 | Threonine and Lysine Metabolism |
| PPRDNDH | 0.0013 | Threonine and Lysine Metabolism |
| LYSDC | 0.0013 | Threonine and Lysine Metabolism |
| DAPDC | 0.0013 | Threonine and Lysine Metabolism |
| DAPE | 0.0013 | Threonine and Lysine Metabolism |
| DHDPRy | 0.0013 | Threonine and Lysine Metabolism |
| DHDPS | 0.0013 | Threonine and Lysine Metabolism |
| SDPDS | 0.0013 | Threonine and Lysine Metabolism |
| THDPS | 0.0013 | Threonine and Lysine Metabolism |
| ASPK | 0.0013 | Threonine and Lysine Metabolism |
| H2Ot | 0.0013 | Transport, Membrane |
| RNDR3 | 0.0012 | Nucleotide Salvage Pathway |
| RNDR4 | 0.0012 | Nucleotide Salvage Pathway |

Table 4.21, which includes reaction names, flux differences and the subsystem that reaction belongs, shows the flux differences of internal reactions of wild type genome scale model and Δ*phoP* strain. Biomass production of mutant strain and that of wild type strain were the same. Table contains 42 reactions as the highest flux differences. The reactions with the highest flux differences belong to the transport from membrane, anaplerotic reactions and pyruvate metabolism.

Table 4.21: Flux Differences between WT Model and *PhoP* Knock-Out Mutant (Threshold Method)

| RXNS | Difference | SubSystem |
|---|---|---|
| CSNt2r | 1000.0000 | Transport, Membrane |
| PPA | 3.1887 | Anaplerotic Reactions |
| PTAr | 3.0401 | Pyruvate Metabolism |
| ADK1 | 3.0401 | Nucleotide Salvage Pathway |
| ACS | 3.0401 | Pyruvate Metabolism |
| NADH17 | 2.9650 | Oxidative Phosphorylation |
| ALCD19y | 2.3714 | Alternate Carbon Metabolism |
| ATPS4r | 2.1325 | Oxidative Phosphorylation |
| PGI | 1.3614 | Glycolysis/Gluconeogenesis |
| NADTRHD | 0.7375 | Oxidative Phosphorylation |
| HEX1 | 0.5000 | Glycolysis/Gluconeogenesis |
| SUCD3 | 0.4629 | Oxidative Phosphorylation |
| CS | 0.4629 | Citric Acid Cycle |
| FUM | 0.4629 | Citric Acid Cycle |
| GAPD | 0.4629 | Glycolysis/Gluconeogenesis |
| TPI | 0.4629 | Glycolysis/Gluconeogenesis |
| ACONTa | 0.4629 | Citric Acid Cycle |
| ACONTb | 0.4629 | Citric Acid Cycle |
| ENO | 0.4629 | Glycolysis/Gluconeogenesis |
| PYK | 0.4625 | Glycolysis/Gluconeogenesis |
| MDH | 0.4625 | Citric Acid Cycle |
| PDH | 0.4555 | Glycolysis/Gluconeogenesis |
| FBA | 0.4538 | Glycolysis/Gluconeogenesis |
| RPI | 0.4538 | Pentose Phosphate Pathway |
| PFK | 0.4538 | Glycolysis/Gluconeogenesis |
| AKGDH2 | 0.4451 | Citric Acid Cycle |
| ICDHyr | 0.4451 | Citric Acid Cycle |
| ICL | 0.0178 | Anaplerotic Reactions |
| GLYK | 0.0091 | Alternate Carbon Metabolism |
| GLXCL | 0.0091 | Glyoxylate Metabolism |
| TRSARr | 0.0091 | Alternate Carbon Metabolism |
| ACOAD1f | 0.0035 | Membrane Lipid Metabolism |
| OIVD2 | 0.0035 | Valine, Leucine, and Isoleucine Metabolism |
| IBMi | 0.0035 | Butanoate Metabolism |
| ACLS | 0.0035 | Valine, Leucine, and Isoleucine Metabolism |
| DHAD1 | 0.0035 | Valine, Leucine, and Isoleucine Metabolism |

Contination of the Table.

| | | |
|---|---|---|
| KARA1i | 0.0035 | Valine, Leucine, and Isoleucine Metabolism |
| DMPPS | 0.0017 | Cofactor and Prosthetic Group Biosynthesis |
| EARai70y | 0.0014 | Cell Envelope Biosynthesis |
| EARi60y | 0.0010 | Cell Envelope Biosynthesis |
| MCTP1A | 0.0008 | Murein Biosynthesis |
| MDDEP3 | 0.0008 | Murein Recycling |

# 5. DISCUSSION

As mentioned in Results section, *S. coelicolor* genome-scale metabolic model reconstructed by Kim et al., in 2014 is validated and compared to experimental results in this thesis study. Also transcriptome data of Δ*pho*P and Δ*arg*R from GEO database was used to analyze the transcriptomic profiles of nitrogen and phosphate regulatory systems in *S. coelicolor*. Transcriptome data and the genome-scale model were computationally integrated for reporter features analyses that reveal the most significantly altered features such as metabolites in specific conditions. Reporter metabolite analyses were done via both the genome-scale metabolic model of Kim et al., and the genome-scale network obtained from BioCyc *S. coelicolor* database. For reporter pathway analyses, only network from BioCyc database was used since its pathway coverage is much larger. Also, the p-values obtained after the statistical analyses of transcriptome data were used in the creation of condition specific metabolic models via GIMME algorithm. FBA was applied to the condition specific models, and the calculated internal reaction fluxes were compared to the values obtained from the simulation of wild-type genome-scale metabolic model.

The experimental results of growth rates and secondary metabolite production rates on different carbon sources both for *S. coelicolor* and other *Streptomyces* species were found to be mostly compatible with the simulation results of *S. coelicolor*. Since more than 60% of the simulation results are compatible with the experimental results, one can assume that the simulation results without experimental evidences can be used as simulation evidences. As mentioned before, the number of consumed carbon atoms from a carbon source was arranged to be the same in all simulations by modifying the uptake rates accordingly. Glycine has the lowest growth rate in all different carbon sources and propionate has the highest value of all. The results are classified into groups to have a better visualization to compare them. In all amino acids as carbon sources, isoleucine has the highest growth rate. Glycerol is a sugar alcohol, and it shows the highest growth value between sugar alcohols and the highest value of secondary metabolite production in all carbon sources. Our model shows different growth and secondary metabolite production profiles for different carbon sources. The qualitative result can be different from experimental result because our model involves only metabolic reactions but not regulatory systems. The compounds such as amino acids and essential elements that are used in the functioning of regulatory systems are

not included in model, so the effect of usage of these building blocks is not covered in our study. The model simulations tell us that the maximal biomass and secondary metabolite production does not show similar behaviour under conditions that includes same carbon source. Since different carbon sources joins the central carbon metabolism as different derivatives, their effect on growth is different. As mentioned in Results section, although the genome-scale model shows growth on glutamate or aspartate including growth medium there is not any growth in these two carbon sources in the literature. The main reason of this is that these two amino acids are linked to central carbon metabolism with only one reaction. Glutamate joins to citric acid cycle (TCA) via α-ketoglutarate that is a keto-acid produced by deamination of glutamate. Aspartate joins to TCA via oxaloacetate. Because these biochemical reactions are available in *S. coelicolor* genome-scale model, our model use these two amino acids as carbon sources ignoring the regulatory interactions which control the expression of the genes controlling these reactions.

Mostly amino acids can be used as nitrogen sources as well. In the model, as mentioned before, 7 amino acids can be used as nitrogen sources. In the simulation results, glutamate and asparagine show the highest biomass production when they are used as nitrogen sources, and isoleucine, that showed a high biomass production when used as a carbon source, shows the lowest biomass production when used as a nitrogen source. Glutamate shows important functions in the metabolism because its derivative, α-ketoglutarate, is the main α-keto acid that is involved in the deamination of amino acids. Its presence in the system affects the nitrogen metabolism more than the other amino acids. Asparagine, on the other hand, is synthesised from aspartate with a reaction that involves glutamate. It means asparagine metabolism affects glutamine metabolism as well. Glutamine also shows high biomass production value. Glutamine is synthesised from glutamate with a single reaction. Simulation results on nitrogen source show that glutamate metabolism affects the nitrogen utilization.

As mentioned before, our genome-scale metabolic model involves 1208 genes that represent the enzymes that catalyze the reactions in the model. Each gene has different influence on biomass and actinorhodin production. In double gene deletion analysis results, the genes were computationally knocked out in pairs. This simulation gives realistic results since there are many gene knock outs that do not prevent the cell from biomass production. A wrong model could involve genes any of whose deletions can block the production of biomass and actinorhodin as secondary metabolite.

The main applications of this thesis are reporter feature analyses and condition specific model generations. Reporter metabolite analyses were applied to metabolic network of Kim et al.,(2014) and to the metabolic network that was downloaded from BioCyc *S. coelicolor* database to compare the effect of metabolic network reconstruction on the resulting metabolite list.

In experimental studies, evidences suggested that *argR* knock out affects nucleotide and nucleic acid biosynthesis, arginine biosynthesis, glutamine synthetase (glnA) activity and fatty acid biosynthesis [Perez-Redondo 2012].

In the results of WildType strain and Δ*argR* mutant strain comparison based on the reporter metabolite analyses via genome-scale metabolic model by Kim et al., (2014), the most affected metabolite under this specific condition is ring 1,2-epoxyphenylacetyl-Coa. This metabolite is connected to six genes that contribute to its p-value. It is involved in phenylalanine metabolism just like the second most affected metabolite, phenylacetyl-CoA. These two metabolites suggest that the mutation of *argR* mostly affects the phenylalanine metabolism. We can see the L-phenylalanine in the list as the most affected 5th metabolite in the mutation condition. In our Table, Table 4.5, there are 26 metabolites that have p-values lower than 0.05. The metabolites that are connected to highest number of genes are acetate, D-glucose-6-phosphate(G6P), and hypoxanthine. Acetate is a carbon source that cell can utilize, G6P is glucose sugar phosphorylated on sixth carbon. These two metabolites suggest that carbon metabolism on *S. coelicolor* is affected because of the lack of *argR* gene. This gene is a regulator on arginine metabolism, so it affects the amino acid and arginine degradation/biosynthesis metabolism basically. The transcriptome analysis result showed us this gene also affects the metabolism of these two carbon sources. Hypoxantine is a derivative of purine. It is basically constituent of nucleic acids, and it is a necessary additive in certain cell and bacteria as nitrogen source. So, the lack of this nitrogen regulatory gene affects this nitrogen source and also purine metabolism. In other results, some certain essential elements are found to be affected in the studied condition. These elements are $Fe^{+3}$ and $Cu^{+2}$. This result suggests that any perturbation on nitrogen regulatory system causes the degeneracy on essential elements metabolism. The other remarkable metabolites that are on the reporter list are glycerate and taurine. Glycerate or 3-phosphoglyceric acid is a three carbon molecule that is a metabolic intermediate in both glycolysis and Calvin cycle. In *S. coelicolor*, sucrose biosysthesis takes shape from the intermediates of Calvin cycle. This suggests that the

lack of *argR* affects the sucrose biosynthesis in *S. coelicolor*. Taurine is a derivative of cysteine. Taurine metabolism involves amino acids like L-Cysteine and also includes acetate, acetyl-CoA and pyruvate. The result mentioned before coincides with the change in this pathway since acetate is found to be a metabolite affected from a high number of genes.

The results of reporter metabolite analysis of Δ*argR* via metabolic network downloaded from BioCyc database shows a similar result that tells us the L-phenylalanine as the most affected metabolite from the mutation, with 5 edges that represent the genes that affect the level of this metabolite. Different from the genome-scale metabolic model results, there are different metabolites in the Table 4.6. The main reason of this difference is that the network of BioCyc database involves more and different metabolites like some hypothetical metabolites such as "a carboxylate" and "a fatty aldehyde". In the Table, the most remarkable metabolites are 3-hydroxypropanoyl-CoA, acryloyl-CoA, xantine and urate (uric acid). 3-hyrdoxypropanoyl-CoA is a metabolic compound that is involved in β-alanine biosynthesis. It is produced by a reaction that involves acryloyl-CoA and consumed by a reaction that involves 3-hydroxypropanote. Basically, it is a compound that is involved in β-alanine biosynthesis. So, it means that lack of Δ*argR* gene affects the alanine biosynthesis with the other amino acids that were found in the results of genome-scale metabolic model network. Metabolites that involve Coenzyme A in their structure are affected in both results. Xantine is purine base that is product of the pathway of purine degradation. This result coincides with the result of genome-scale metabolic model. Urate or uric acid is a heterocyclic compound of carbon, nitrogen, oxygen and hydrogen. It is the product of degradation of metabolic breakdown of purine nucleotides. As mentioned before, both reporter metabolite analysis results of genome-scale metabolic network and BioCyc network suggested that purine base degradation is one of the most affected pathways.

Table 4.7, 4.8, 4.9 and 4.10 show the results of reporter metabolite analysis of Δ*phoP* mutant from two different perspectives. Transcriptome data set for only 23rd and 24th hours and for all hours that are included in the data were separately analyzed to allow discovery of different metabolic mechanisms altered by the deletion. Both of the analysis results are shown in the tables. As mentioned before, *phoP* is the part of the regulation system of phosphate metabolism in *S. coelicolor.*

In experimental studies, the knock out of *phoP* was reported to affect the oxidative phosphorylation, cell wall biosynthesis, *bldA, bldC* and *bldM* activation that are part of *bldD* regulon required for morphological development and antibiotic production [Hengst 2010], nucleotide biosynthesis, and glycogen metabolism [Nicholas 2012]. As shown in different studies, *phoP* affects cell envelope biosynthesis, nitrogen and phosphate metabolism, glyceraldehyde-3-phosphate dehydrogenase activity [Rodriguez-Garcia 2007], super-oxide dismutase, some putative membrane proteins, glutamate binding protein, ATP synthase, citrate synthase, and enolase [Thomas 2011].

When using only the 23rd and 24th hour time-point transcriptome data, reporter metabolite analysis results of the genome-scale metabolic model suggests that the most affected metabolite is L-aspartate and the metabolites that have the highest number of connected genes are 3-Methyl-2-oxobutanoate, (S)-3-Methyl-2-oxopentanoate and glycerol. Both 3-Methyl-2-oxobutanoate and (S)-3-Methyl-2-oxopentanoate are compounds that are involved in valine, leucine and isoleucine degradation and biosynthesis, and in secondary metabolism biosynthesis. Glycerol is a sugar alcohol and it has one of the highest growth rate and the highest actinorhodin production rate based on the in silico results reported in previous sections. As mentioned, the lack of *phoP* gene decreases the production of secondary metabolites in *S. coelicolor*. Our result suggests that this decrease in antibiotic production can be because of the significant change on pathways that involve glycerol. Glycolate or glycolic acid is another perturbed metabolite in Δ*phoP* mutant. It is the smallest α-hydroxy acid. It is involved in the biosynthesis of antibiotics pathway and this result coincides with experimental results. The same results via using the BioCyc network can be seen in Table 4.9. In this Table, the most central metabolites are $NAD^+$ and NADH due to their very high number of connectivity in the network. In the mutant strain, oxidative phosphorylation is known to be affected, and these two compounds are the major parts of oxidative phosphorylation. The number of metabolites that have p-values lower than 0.05 is 21. In the results, deoxynucleotides are predicted to be mostly perturbed. Phosphoenolpyruvate (PEP) is another metabolite that was significantly affected in the mutant strain. PEP participates in glycolysis and it has a high energy phosphate bond. PEP is transformed to pyruvate by a single reaction catalyzed by pyruvate kinase that produces pyruvate. It is the final and one of the most important reactions in glycolysis. The compounds that are produced in glycolysis are then used in oxidative

phosphorylation, so this perturbation on PEP indirectly affects oxidative phosphorylation. β-nicotinamide-D-ribonucleotide is another significantly perturbed metabolite. This compound is involved in nicotinate and nicotinamide metabolism. The compounds, [dihydrolipoyllysine-residue (2-methylpropanoyl) transferase] lipoyllysine and [dihydrolipoyllysine-residue (2-methylpropanoyl) transferase] S(2-methylpropanoyl) dihydrolipoyllysine are two compounds that are involved in valine, leucine and isoleucine degradation. In Δ*phoP* mutant, metabolism of these amino acids can be affected because of the affected metabolism of phosphate.

Same reporter metabolite analyses were applied to all of the wild-type and Δ*phoP* transcriptome data rather than considering only two time points. Different time steps in the data were considered as replicates in this analysis. Reporter metabolite analysis was repeated both for the genome-scale metabolic model network of Kim et al., (2014) and the BioCyc Gene-Metabolite network. In Table 4.8, reporter metabolite analysis results of all data of Δ*phoP* can be seen. First, Coenzyme A is the metabolite that is affected from highest number of genes. Coenzyme A is the major component of oxidative phosphaorylation and as reported in the experimental results, oxidative phosphorylation is affected from the knock out of *phoP* gene. The metabolites that include Coenzyme A such as succinyl-CoA, 2-Methylbutanoyl-CoA and Isovaleryl-CoA are also predicted to be affected as well in our analysis with a high number of connected genes. Among the other metabolites that are significantly perturbed with high number of metabolites are D-Alanyl-D-Alanine, 3-Methyl-2-oxobutanoate. (S)-3-Methyl-2-oxopentanoate metabolite is also in Table 4.7 that shows the results of reporter metabolite analysis of Δ*phoP* mutant based on only 23rd and 24th hour points. This shows us the consistency of the results. D-alanyl-D-alanine is a component of bacterial peptidoglycan that forms the cell wall. In experimental results, cell wall biosynthesis is shown to be affected in *phoP* knock out mutant. The computational results suggest that, by affecting D-Alanyl-D-Alanine, *phoP* knock out indirectly perturbs the cell wall biosynthesis. CMP or cytidine monophosphate is a nucleotide that is used as monomer in RNA. It takes part in cardiolipin biosynthesis in *S. coelicolor*. Cardiolipin is a key player in bacterial adaptation to osmotic stress [Romantsov 2009]. This results implies that there is a change in the osmotic stress on *S. coelicolor* growing with a lack of *phoP* gene.

The result of reporter metabolite analysis of the same dataset via using BioCyc Gene-Metabolite network can be seen in Table 4.10. Similar to the results discussed

above, Coenzyme A is a significantly perturbed metabolite with one of the highest number of connected genes. Also succinyl-CoA and and succinate, that forms succinyl-CoA with Coenzyme A, are significantly affected metabolites. Diphosphate has the highest number of connected genes. As mentioned before, *phoP* is a part of phosphate regulatory system, so disphosphate metabolite is affected because of the perturbation of regulatory system. Pyruvate dehydrogenase E2 protein N6-dihydrolipoyl L lysine is a compound that is involved in pyruvate decarboxylation to acetyl CoA and lipoate salvage pathway. Dihydroxylipoate is another remarkable metabolite in the Table, and it is the reduced form of lipoate. It contains two sulfur atoms and it is a cofactor of pyruvate dehydrogenase complex. Dihydrolipoyllysine-residue(2-methylpropanoyl)transferase lipoyllysine compound is involved in valine, leucine and isoleucine degradation, and this coincides with other reporter metabolite analysis results. Dimethylallyl diphosphate is a metabolite in the identified list with 5 connected genes. It involves many different pathways such as biosynthesis of secondary metabolites, biosynthesis of terpenoid. It shows indirect effect of *phoP* knock out on secondary metabolism.

To investigate the metabolic differences of primary and secondary metabolism, data set of wild type transcriptome data of $32^{nd}$ and $66^{th}$ hours of Δ*argR* data set and GSE18489 data set were used. As in the case of the other reporter metabolite analyses, both genome-scale metabolic model network and BioCyc *S. coelicolor* Gene-Metabolite Network were used. Results are shown in Tables 4.11, 4.12, 4.13, 4.14.

In the Table 4.11, the results of reporter metabolite analysis of wild type strain based on $32^{nd}$ and $66^{th}$ hour expression profile, the most remarkable metabolite is hydrogen peroxide. Hydrogen peroxide is a chemical compound that is formulated as $H_2O_2$, and it is the simplest peroxide. Despite its toxic effect, hydrogen peroxide is a secondary messanger in some bacteria. It also shows the increasing activity in oxygen levels. Ferroxamine b is a compound of iron metabolism in *S. coelicolor*. It is a siderophore that takes iron from environment for metabolic processes. These results indicate changes in iron metabolism in secondary metabolism and changes in many different types of acyl-carrier proteins. In the secondary metabolism phase, these acyl-carrier proteins are involved in secondary metabolite production, so these proteins that are connected to 10 genes show the major change between primary and secondary metabolism. Ferricytochrome c is a major part of oxidation chain with 11 genes connected. Pyridoxal 5"-phosphate, another significantly perturbed metabolite, is the

active form of vitamin $B_6$ and it is a cofactor of variety of enzymes. This means that metabolic changes in that cofactor affects different types of enzymes and these enzymes are important in transition between primary and secondary metabolism. Dihydroxyacetone phosphate (DAP) is another compound that is predicted to be significantly affected. As it is known, secondary metabolism is triggered with nutrient limitation like phosphate limitation. Therefore, this means *S. coelicolor* still can metabolize the carbon sources but primary metabolite production is shifted to secondary metabolite production. This change in central carbon metabolism shows its effect on DAP metabolism that takes place in glycolysis. Propanoyl-CoA, Coenzyme A derivative of propanoyl, and Butanoyl-CoA, Coenzyme A derivative of butanoyl, are also significantly perturbed metabolites between primary and secondary metabolism and they are involved in several biochemical pathways such as beta-oxidation of fatty acids and isoleucine and valine metabolism. In secondary metabolism phase, fatty acid biosynthesis and degradation is known to be one of the major changes in *S. coelicolor*. Thus, these compounds that are involved in fatty acid oxidation pathway changed significantly. L-alanine-D-glutamate-meso-2,6-diaminoheptanedioate is involved in the peptidoglycan biosynthesis, so it is a part of cell wall biosynthesis. The perturbation on this metabolite shows us the secondary metabolism effect on cell wall biosynthesis.

The same analysis was done by using BioCyc Gene-Metabolite network and the results are shown in Table 4.12. Hydrogen peroxide is still the most significantly changed metabolite in results. Acryloyl-CoA, with its 18 genes connected, is one of the most significantly altered metabolites based on these results, and it is involved in Propanoyl-CoA degradation and β-Alanine biosynthesis. As mentioned in the results shown in Table 4.11, propanoyl-CoA is changed in transition between primary and secondary metabolism and, in the results of BioCyc network, the compounds that are produced in the reactions of propanoyl-CoA is remarkably changed. Propanoyl-CoA can also be seen in the list of the results obtained by using the genome-scale metabolic model. In the results, maltose appears as a significantly perturbed metabolite. Just because the growth medium in this experiment contains maltose as carbon source, it shows the depletion on maltose in the medium with time.

To reveal the metabolic changes in secondary metabolism by using another data set and, hence, allow a validation of the results, 20th-21st and 58th-60th hour

transcriptome data of GSE18489 were used. Both genome-scale metabolic model and BioCyc network were used as input networks as done in other RMAs.

First, the results of reporter metabolite analysis based on the genome-scale metabolic model can be seen in Table 4.13. The metabolites that are remarkably perturbed are oxidized and reduced thioredoxins. Thioredoxin is an important compound and it can be found in all organisms. It plays roles in important processes like redox signalling. By an enzyme called thioredoxin reductase, electrons are transferred between pyridine nucleotide and disulfide componds. Basically, thioredoxins allow electrons to pass from NADPH to FAD and then to the active site disulfide bond. According to the computational results, both oxidized and reduced forms of thioredoxin are significantly changed in secondary metabolism. This prediction implies that some enzymes that use thioredoxins, FAD and NADPH are significantly changed during the transition between primary and secondary metabolism. Cytidine diphosphate, deoxycytidine diphosphate, deoxyuracil diphosphate, dioxyguanosine diphosphate, dioxyadenine diphosphate, uracil diphosphate are nucleotides that are significantly perturbed during secondary metabolism based on the computational results. Probably, during the switch between primary and secondary metabolism, the lack of phosphate in the growth medium causes the degradation of these nucleotides to supply the needed phosphate for metabolism. Hydrogen peroxide is also on the list as a significantly affected metabolite, but its p-value is higher. That is just because the genome-scale metabolic network and BioCyc network have different connections between genes and metabolites. L-Methionine sulfoxide is a derivative of methionine that is formed by the oxidation of sulfur in the methionine. This reaction is NADPH dependent so the thioredoxin is involved in this reaction. With the change in nucleotide metabolism and thioredoxin metabolism, 3"-Phosphoadenylyl sulfate, a compound that is involved in sulfur metabolism, purine metabolism and glycosaminoglycan metabolism, changes in cell wall metabolism in addition to the sulfur metabolism are also significantly affected. Fructose-6-Phosphate(F6P) is also significantly changed and it is one of the major metabolites in glycolysis. It implies changes in the central carbon metabolism. As mentioned before, primary and secondary metabolism competes, and secondary metabolism uses the precursors from primary metabolism. This change in F6P can be caused because of the changes of cofactors in the secondary metabolism.

As the last part of the reporter metabolite analyses, 20th-21st and 58th-60th hour transcriptome data of GSE18489 was analyzed by using the BioCyC network. The most significantly perturbed metabolite is L-glutamate(outside). The growth medium contains glutamate as phosphate source and this implies significant associated alterations during the transition. This coincides with the phosphate depletion effect of switch between primary and secondary metabolism. With 51 connected genes, Acetyl-CoA is one of the most marked metabolites affected during the transition. With other compounds that involve Coenzyme A such as malonyl-CoA, 3-hydroxypropanoyl-CoA and 2-methyl-3-hydroxybutyryl-CoA, Coenzyme A metabolism is found to be significantly changed. Coenzyme A is also involved in oxidative phosphorylation and fatty acid biosynthesis, and functions as acyl group carrier. With its significant perturbation, one can assume that during transition, acyl group carriage, oxidative phosphorylation and metabolites that need acyl carriage such as antibiotics are also perturbed.. A_2,3,4-saturated_fatty_acyl_CoA metabolite represents a typical fatty acid that has 14-24 carbon atoms. This putative compound is involved in fatty acid oxidation, acyl-CoA hydrolysis and diacylglycerol and triacylglycerol biosynthesis. As in the case of the other results in the list, acyl-CoA compound is also significantly affected during the transition, and this result shows the close relationship between secondary metabolite biosynthesis and fatty acid synthesis. Oxidized thioredoxin is in the list of results based on BioCyc network as well.

The other reporter feature analysis is reporter pathway analysis. For this analysis, the Gene-Pathway Network downloaded from BioCyc *S. coelicolor* database was used. These results are shown in Table 4.15, 4.16, 4.17, 4.18 and 4.19.

The first reporter pathway analysis performed is based on the comparison of Wild Type strain and Δ*argR* strain. Results are shown in Table 4.15. The most significantly altered pathway is germicidin biosynthesis pathway, with 16 genes connected to it. Germicidins are a group of natural products arising from Streptomyces species that act as autoregulatory inhibitor of spore germination. Significant change in this pathway shows us the spore germination and aerial hypae formation during secondary metabolism. As mentioned in the results of reporter metabolite analysis, adenosine, guanosine, deoxyribose phosphate degradation pathways are identified as significantly changed pathways. In addition to this, urate biosynthesis/inosine 5''-phosphate degradation pathway, methylenomycin furan biosynthesis, trans,-trans-farnesyl diphosphate biosynthesis, acrylonitrile degradation I and aldoxime

degradation are also identified. Ureate is produced from purines. Change in urate biosynthesis can be caused by a change in the metabolism of nucleotide degradation. This pathway also includes inosine-5' phosphate degradation pathway. Methylenomycin furan (MMF) is a type of compound that includes 5 different types in *S. coelicolor*. The pathway starts with glycerol degradation. This pathway belongs to the super class autoinducer biosynthesis. The MMFs are signalling molecules and they are needed to induce the production of methylenomycin. Methylenomycins are cyclopentanoid antibiotics produced by *S. coelicolor* that is effective against both Gram-negative and Gram-positive bacteria. With the lack of *argR* gene, these signalling molecules and surely the production of methylenomycins changed significantly. Farnesyl diphosphate is an intermediate in both the mevalonate and non-mevalonate pathways used by organisms in the biosynthesis of terpenes, terpenoids, and sterols. According to the analysis results, this pathway significantly changed because of the deleted gene. In *S. coelicolor*, from farnesyl diphosphate, albaflavenones, an anti-microbial agent is produced. In acrylonitrile degradation pathway, acrylate is produced from acrylonitrile by two step reactions, and acrylate is transformed to ammonium with a single reaction that uses $H_2O$. These results can indicate the changes in ammonia metabolism and this coincides with the lack of arginine pathway regulatory gene, *argR.* In aldoxime degradation pathway, nicotinade is produced in two reaction steps, and from nicotinade pyridine nucleotides are produced. As mentioned in the results of reporter metabolite analysis, nucleotide biosynthesis and degradation pathways changed in Δ*argR* strain.

As shown on the Table 4.16, reporter pathway analysis was also applied to 23rd and 24th hour expression profiles of Δ*phoP* strain. Results show 48 pathways whose p-values are lower than 0.05. The first three pathways that significantly changed are guanosine nucleotides degradation I-II and adenosine nucleotides degradation. According to the experimental results, nucleotide biosynthesis pathways are repressed with the lack of *phoP*. In the pathway analysis result identified in this study, this effect appears as a significant change in degradation pathways. Also, the significantly perturbed pathway with the highest number of connected genes is streptorubin B biosynthesis. Streptorubin antibiotic production is controlled by *red* gene cluster which also controls the production of undecylprodigiosin. The knock out of *phoP* gene affects the *red* gene cluster, and thereby the streptorubin B biosynthesis pathway. Another secondary metabolite biosynthesis pathway that is found to be significantly affected in

the bioinformatics analysis is actinorhodin biosynthesis pathway. In the experimental results, in Δ*phoP* strain actinorhodin production is reduced [Nicholas 2002]. 28 genes are connected to actinorhodin biosynthesis pathway. With this change in secondary metabolite production, ethanol degradation, anaerobic L-lactaldehyde degradation, glyoxylate cycle, valine degradation pathway III, guanosine pentaphosphate (ppGpp) biosynthesis, formaldehyde assimilation I (serine_pathway), leucine degradation and biosynthesis pathways are significantly altered. With the lack of phosphate metabolism regulator, ppGpp biosynthesis is a response to the stringent response to amino acid starvation. The changes in the amino acid metabolism such as valine and leucine degradation pathway and serine pathway coincide with these results. Lactaldehyde is the oxidized form of lactic acid. Glyoxylate cycle is an anabolic pathway that converts Acetyl-CoA to succinate for carbohydrate biosynthesis. Mixed acid fermentation pathway, with its 0.03 p-value and 33 connected genes, shows the metabolic effects of *phoP* knock-out on succinate, 2-oxoglutarate, acetate, ethanol and lactate biosynthesis. Besides, ethanol degradation pathway is one of the most significantly affected pathways of all, and together with the change on mixed acid fermentation pathways, pathways connected with ethanol are important pathways that are affected in Δ*phoP* strain. Farnesyl diphosphate biosynthesis pathway can be seen in this Table too.

Same analysis was applied to the whole data of GSE31068. Results are shown in Table 4.17. In this analysis, farnesyl diphosphate biosynthesis pathway is the most significantly changed pathway. Significantly changed pathway that has highest number of connected genes is salvage pathways of purine and pyrimidine nucleotides. Salvage pathways are used to recover bases and nucleotides that are formed during degradation of RNA and DNA. The salvaged bases and nucleosides can then be converted back into nucleotides. It shows the effect of *phoP* knock out on nucleotide metabolism. Phosphate group is one of the major monomers of nucleotides, and perturbations on phosphate regulatory systems causes significant changes on nucleotide metabolism. Just like the analysis presented before (Table 4.16), pyruvate fermentation to propanoate, L-threonine degradation, ethanol degradation, actinorhodin and germicidin biosynthesis, guanosine and adenosine nucleotide degradation pathways, glyoxalate cycle, isoleucine degradation and mixed acid fermentation showed significant change in Δ*phoP* strain. PhosphoPantothenate consumed biosynthesis pathway is a part of coenzyme A biosynthesis pathway. It coincides with the significant change in the pathways that are related to coenzyme A.

Adenosyl cobalamine biosynthesis from cobyrinate a,c-diamine pathway is a part of vitamin biosynthesis pathways. Also 2-oxoglutarate decarboxylase step of TCA cycle is affected from the lack of *phoP* gene.

To reveal the changed osecondary metabolism of *S. coelicolor* on the pathway level, just like in reporter metabolite analysis, 32$^{nd}$ and 66$^{th}$ hours of wild type data of GSE58666 and 23-24$^{th}$ hours and 58-60$^{th}$ hours of wild type data of GSE18489 were used, and reporter pathway analysis was applied.

Results of reporter pathway analysis of GSE58666 data are shown in Table 4.18. The most significantly perturbed pathways are identified as aerobic respiration, urate degradation and superoxide radicals degradation. Their connected gene numbers are 7,9 and 5. The first result shows the effects of secondary metabolism on cytochrome c level. Urate degradation, as in the other results, produces allantoin. Urate is produced in the purine nucleobases degradation pathways. As mentioned before, mostly secondary metabolism is related to nutrient limitation such as phosphate limitation. This result can show metabolic effect of phosphate limitation in growth medium. Because of the lack of phosphate, *S. coelicolor* can start or accelerate the degradation of purine nucleotides to obtain the phosphate that is required for metabolism, and this can significantly alter the urate metabolism. Superoxide radicals degradation pathway is basically a pathway that produces oxygen from 4 superoxide molecules. Superoxide molecules are also related to iron metabolism. With two oxygen molecules and $Fe^{2+}$, a superoxide molecule and $Fe^{+3}$ are produced. For the transition between primary and secondary metabolism, the most affected pathways with high number of genes show the metabolic effect of genetic regulation of secondary metabolism. The most remarkable pathways are mixed acid fermentation, superpathway of phenylalanine, tyrosine and tryptophan biosynthesis, adenosyl cobalamine biosynthesis, pyrimidine ribonucleotides de novo biosynthesis, NAD biosynthesis, UMP biosynthesis and biotin biosynthesis. As mentioned before, succinate, 2-oxoglutarate, acetate, ethanol and lactate biosynthesis are part of mixed acid fermentation pathway. Significant change in this pathway points to the alterations in the carbon metabolism. With the lack of phosphate but not any perturbations on carbon source uptake, *S. coelicolor* changed its carbon utilization pathways to mixed organic acid fermentation pathways. The superpathway of phenylalanine, tyrosine and tryptophan biosynthesis is significantly changed with 10 genes connected to it. These three amino acid biosynthesis pathways are sub pathways of proteinogenic amino acid biosynthesis. According to experimental

results [Nieselt 2010], protein biosynthesis decreased in secondary metabolism of *S. coelicolor*. Also, tryptophan biosynthesis, although it has no correlation to secondary metabolite production, it is directly related to growth. At a higher growth rate, more tryptophan would be synthesized. It could be related to stringent response [Hood 1992]. In secondary metabolism phase, growth rate decreases, and tryptophan biosynthesis would decrease too. As is known through the experimental results, proline is directly related to secondary metabolism because of antibiotic undecylprodigiosin. Pyridoxal phosphate salvage pathway is another significantly changed pathway during transition between primary and secondary metabolism. As the active form of vitamin $B_6$, it acts as coenzyme in all transamination reactions and in some decarboxylation and deamination reactions. These results imply alterations in the metabolism of pyridoxal phosphate, and the reactions in which it acts as coenzyme would be also affected in secondary metabolism. L-histidine biosynthesis is another significantly changed pathway, and it has 22 connected genes to it. It appears as one of the important amino acids in transitions. This pathway has many reaction steps in it and many internal metabolites and genes. As shown in the results before, nucleotide metabolism is perturbed during the transition, and pyrimidine ribonucleotides de novo biosynthesis is shown in Table 4.18 as a part of nucleotide metabolism. NAD biosynthesis is another important and significantly changed pathway that affected in secondary metabolism. $NAD^+$ is not only involved in oxidative phosphorylation, it is also an important coenzyme in many reactions. A pathway in Table 4.18 is the biosynthesis of NAD from L-aspartate. In this pathway, there are many reactions that involve 2-iminosuccinate, nicotinade-D-ribonucleotide, diphosphate and glutamate transformation to glutamine. As mentioned before, reporter pathway analysis revealed a significant change in ribonucleotide biosynthesis, and now a significant change is predicted in a pathway that involves a ribonucleotide in. This indicates that the main change in $NAD^+$ biosynthesis pathway is caused because of ribonucleotide biosynthesis pathway. UMP (uridine monophosphate) biosynthesis is another significantly changed pathway and it has 18 genes connected to it. UMP involves reactions that are part of peptidoglycan, pyrimidine nucleobases and ribonucleobases salvage pathways and tRNA process. This reveals the UMP is an important metabolite, and its biosynthesis is an important pathway in secondary metabolism.

Another reporter pathway analysis that revealed the metabolic changes during transition between primary and secondary metabolism is RPA of 23-24th hours and

58-60th hours of wild type data of GSE18489. It is shown in Table 4.19 and it shows 16 significantly changed pathways. The most significantly changed pathway is butyrolactone biosynthesis pathway. This pathway starts with glycerone phosphate and ends with three different butyrolactone, SCB1, SCB2 and SCB3. They are classified as secondary metabolites and as expected, during transition this pathway that involves secondary metabolites is significantly altered. The second most significantly changed pathway is L-arabinose degradation III pathway. It represents the utilization of L-arabinose outside. This result came up because of the change in growth medium during the experiment. The significantly changed pathways with the highest number of genes are streptorubin B biosynthesis and calcium dependent antibiotic (CDA) biosynthesis pathway. These two secondary metabolites that are commercially important are produced during secondary metabolism. We expected this result because, in about 35 hours, *S. coelicolor* cells changed their metabolism from primary to secondary metabolism. Albaflavenone biosynthesis pathway starts with farnesyl diphosphate, a compound that was mentioned before as significantly perturbed in the lack of *argR* and *phoP* genes. Albaflavenone is a bacterial compound that is found in *S. coelicolor*, and it is an anti-microbial agent. This time, this compound itself was revealed as significantly changed instead of its precursor, farnesyl diphosphate. It shows the connection of lack of *argR* and *phoP* genes and secondary metabolism. Germicidin biosynthesis pathway involves the synthesis of germicidin B, germicidin A and isogermicidin A, and as mentioned before, germicidins are autoregulatory inhibitor of spore germination. As is known, spore germination is connected with secondary metabolism via aerial hypae formation. Urate degradation is another pathway that is found to be significantly affected during transition in the computational analysis, and it is also found to be significantly changed in mutations that was analyzed before. Because of the change in the metabolism of nucleotides, urate degradation pathway is also affected during the secondary metabolism. The most different pathway that did not appear in the results before is heme biosynthesis (anaerobic) pathway. Heme molecule is an important compound in iron metabolism, and it can show us the effect of secondary metabolism on iron metabolism.

To reveal the effects of *argR* and *phoP* genes, the GIMME algorithm was used to create the genome-scale metabolic models specific to deletion strains. Models were obtained by using the threshold method as mentioned before. Both of the models showed the same growth rate in same conditions but they used different internal fluxes

to reach this specific growth rate. The internal fluxes of mutant models were compared to the wild-type genome-scale metabolic model and internal fluxes are compared according to their flux rates. In Tables 4.20 and 4.21 the most different flux rates are shown.

Table 4.20 shows the results of FBA of Δ*argR* genome-scale model and the flux rates of reactions that involves the model and comparison of flux rates to wild type metabolic model. The 63 of the most changed flux rates are shown in the table. Reactions that has the most changed flux rates are alcohol dehydrogenase, L-idonate 5-dehydrogenase, NAD transhydrogenase, L-alanine reversible transport via proton symport and cytosine transport in via proton symport. The subsystems of these reactions are alternate carbon metabolism, oxidative phosphorylation and membrane transport. The other 58 reactions are belong to the alternate carbon metabolism, glycolysis, oxidative phosphorylation, anaplerotic reactions, nucleotide salvage pathways, pyruvate metabolism, TCA, pentose phosphate pathway, cofactor and prosthetic group biosynthesis, cysteine metabolism, methionine metabolism, valine,leucine and isoleucine metabolism, threonine and lysine metabolism, inorganic ion transport and cell envelope biosynthesis subsystems. As mentioned before, according to the experimental results, Δ*argR* affects the nucleotides and nucleic acid synthesis, arginine biosynthesis, polyketide biosynthesis and actinorhodin biosynthesis. FBA with objective function as actinorhodin production was used, and the FBA result was zero. It shows the effect of Δ*argR* on actinorhodin production. According to the results, comparison of flux rates of wild type strain and Δ*argR*, carbon metabolism, nucleotide metabolism, amino acid metabolism, pentose phosphate pathway, TCA, oxidative phosphorylation and cell envelope biosynthesis are remarkably changed as well. The results show how *S. coelicolor* changes the metabolic pathways in the lack of *argR* mutant. According to the reporter pathway analysis results, germicidin biosythesis is the most significantly changed pathway, and it coincides with the result of cell envelope biosynthesis. Germicidins are autoregulators of spore germination and during germination cell envelope structure changes. Changes in the internal fluxes of nucleotide metabolism match up with the results of reporter pathway analysis. The biggest differences, as mentioned before, are in the central carbon metabolism. It appears that *S. coelicolor* changed its central carbon metabolism and oxidative phosphorylation by increasing the flux of the

reaction that is catalyzed by alcohol dehydrogenase to produce glycerol. Flux rates of nucleotide salvage pathway are also changed in the Δ*argR*.

Table 4.21 shows the results of FBA of Δ*phoP* genome-scale metabolic model and the flux rates of reactions that are involved in the model and comparison of flux rates to wild type metabolic model. The table contains 42 reactions as the highest flux differences. The 6 reactions with highest differences are belong to membrane transport, anaplerotic reactions, pyruvate metabolism, nucleotide salvage pathways and oxidative phosphorylation. The other 36 reactions are belong to the alternate carbon metabolism, oxidative phosphorylation, glycolysis, TCA, anaplerotic reactions, glyoxylate metabolism, membrane lipid metabolism, valine, leucine and isoleucine metabolism, butanoate metabolism and cell envelope biosynthesis.

To gain an overall insight about the effects of mutations and switch of primary metabolism to secondary metabolism and to observe new metabolic and genetic engineering insights, the results were examined all together. As known, both Δ*argR* and Δ*phoP* strains produce secondary metabolites less than wild-type strain. Mutual metabolic effects of these gene knock outs can show a new target to improve secondary metabolite production. As mentioned before, both of the mutants show different metabolic and pathway level effects on *S. coelicolor* and, according to reporter metabolite analyses via BioCyc Gene-Metabolite network results, three metabolites are shared in the results. These three metabolites are, (R)-lipoate, dihydrolipoate and lipoyl carrier protein $N^6$ dihydrolipoyl-L-lysine. The common effect of mutations is a decrease in actinorhodin biosynthesis, and these three metabolites can reveal the direct effect of perturbations on secondary metabolite biosynthesis. (R)-lipoate and lipoyl carrier protein $N^6$ dihydrolipoyl-L-lysine are the substrate and product of lipoate salvage pathway. Dihyrdolipoate, on the other hand, is a compound that lipoate is synthesized from. As mentioned before, these metabolites are common in the two mutant strains that show decreased secondary metabolite production. Lipoate is an organosulfur compound that is used in many enzymes as cofactor. These results indicate that the reactions that use lipoate as cofactors affect the secondary metabolite biosynthesis pathways. To compare further the effects on secondary metabolism, the Δ*argR* strain reporter metabolite results and secondary metabolism analysis of wild type transcriptome sub-data of Δ*argR* mutant were compared. In this comparison, three different metabolites show significant change. As is known, in secondary metabolism phase, actinorhodin production increases and in the mutant straint there is

a decrease in actinorhodin production. The three metabolites that show significant change in both of the conditions are, L-phenylalanine, acryloyl-CoA and superoxide. Analysis results do not show the direction of change such as decrease or increase in metabolite levels but this situation shows us the aromatic amino acid, L-phenylalanine, can be important in secondary metabolite production. Acryloyl-CoA, by the way, is a compound that takes part in β-alanine biosynthesis. Significant change of this metabolite can also show the indirect effect of β-alanine in secondary metabolite production.

To gain more insights, the results or RMAs and RPAs were compared. As is known, in Δ*argR* and Δ*phoP* strains, actinorhodin production is decreased and in secondary metabolism phase actinorhodin production is increased. When results are compared, the common significant changed metabolites and pathways can reveal the common properties of actinorhodin production and enable to find new genetic or metabolic engineering targets to increase the production. When RMAs of Δ*argR* and Δ*phoP* strains compared, the results that obtained via BioCyc network reveals three metabolites in common. These are , R-lipoate, dihydrolipoate and a_[lipoyl-carrier_protein]_N6-dihydrolipoyl-L-lysine. These three metabolites involves in lipoate salvage pathway and this situation indicated the possible metabolic engineering targets for increased actinorhodin production.

Reporter pathway analysis results can reveal better targets because this results shows the effects of genetic perturbations on pathway level. When RMAs of Δ*argR* and Δ*phoP* strains compared, six pathway are common in both of the analyses. These are urate biosynthesis, adenosine nucleotide degredation pathways I and II, guanosine degredation pathways I and II and farnesyl diphosphate biosynthesis pathway. In these two strains, as mentioned before, ACT production is decreased and these six pathways are the common significantly changed pathways. This result indicated that adenoside and guanosine nucleotide degredation pathways can be important in actinorhodin biosynthesis pathways. As mentioned this could be the effect of nutrient limitation and boosts cells to degrade the nucleotides to replace the limited nutrient.

Another comparison of RPA results is comparison of Δ*phoP* and transition between primary and secondary metabolism. This comparison shows four common pathway that significantly changed. In secondary metabolism, actinorhodin production increases as mentioned before and in Δ*phoP* strain actinorhodin production is decreased. Because of p-values are scalar, RPA results are scalar too and do not show

the increase and decrease. They only shows significance. Secondary metabolism and Δ*phoP* have a common property, change in actinorhodin production so their RPA results can be compared and reveals potential targets to increase actinorhodin production. The four common significantly changed are, guanosine deoxyribonucleotides de novo biosynthesis I and II, germicidin biosynthesis and trans-lycopene biosynthesis pathways. Just like results before, nucleotide biosynthesis pathways are significantly changed and this indicates that nucleotide biosynthesis and degredation pathways are suitable metabolic engineering targets to develop new strains for increased actinorhodin production. Germicidin biosynthesis pathway is another common significantly changed pathway. As mentioned, germicidins are autoregulatory inhibitors of spore germination. This pathway and spore germination mechanisms could be suitable targets to develop new strains. The last common significantly changed pathway is lycopene biosynthesis pathway. Trans lycopene is a bright red carotenoid pigment and found in many photosynthetic organisms. This pathway as noted as plant pathway in BioCyc *S. coelicolor* data base but in genome scale network it shown as connected to some genes in *S. coelicolor*. Probably this is because the lack of information. Trans lycopenes are members of secondary metabolites. More experimental evidences about the genes that connected to this pathway could reveal new possible targets to increase actinorhodin production.

As a last comparison, FBA results of mutant specific genome-scale models obtained via GIMME algorithm are compared. The subsystems that changed in the models are mostly common but few of them are different. These subsystems are methionine metabolism and inorganic ion transport for Δ*argR* strain and glyoxylate metabolism, membrane lipid metabolism, butanoate metabolism and murein metabolism. Common subsystems can identify the pathways that changed in the mutant strain, and a mutual property can reveal the connected pathway to actinorhodin production as well. Pathway analysis results were done before, but the comparison of deletion strain specific models can reveal better insights. As in the reporter pathway results, nucleotide salvage pathways are both changed in the models and it shows the metabolic effect of mutations on nucleotide metabolism. Also change in alternate carbon metabolism indicates that *S. coelicolor* changed its metabolic pathway to utilize the carbon that is required for metabolic processes. TCA and metabolism of some certain amino acids like valine,leucine and isoleucine had high flux differences in both of the mutant models.

Consequently, the secondary metabolism of *S. coelicolor* shows complex regulatory and metabolic networks and it is controlled directly or indirectly by many metabolites and pathways. Although their regulatory tasks are different, *phoP* and *argR* shows some common metabolic effects. These common effects indicate the common results such as reduced actinorhodin production. Connections between these effects can reveal the complex regulatory systems of both primary and secondary metabolism of *S. coelicolor*.

The results of reporter feature analyses show different range of results due to the different properties of genome scale metabolic model network and BioCyc networks. Also some errors can be caused because of errors on the transcriptome data during experimental procedure and computational errors on statistical test.

# 6. CONCLUSION

In this study, the phosphate and nitrogen regulatory mechanisms and secondary metabolism of *S. coelicolor* were investigated to reveal the key metabolites and pathways of *phoP* and *argR* knock out perturbations and transition between primary and secondary metabolism. *S. coelicolor* is very important producer of many medically and commercially important compound and the identification of key systems of their metabolism can enable to develop high production rate strains.

To study their complex regulatory systems, several approaches of systems biology such as genome-scale metabolic models, omics technologies, statistical approaches, reporter feature analyses and creation of condition specific models were used. Study of metabolic networks as a system reveals the interactions of genes, metabolites and pathways. As is known, complex biological systems shows their specific behaviours with interactions between parts of system. Because of this systematic behaviour of systems and robustness of biological systems, their behaviours on specific conditions and perturbations cannot be predicted by focusing only individual parts.

According to results of this study, *S. coelicolor* has common points that take part on nitrogen metabolism regulation, phophate metabolism regulation and transition between primary and secondary metabolism. These common metabolites and pathways are suitable targets for metabolic and genetic engineering methods to develop better strains. Mostly, secondary metabolism phase and secondary metabolite production are known to be connected with nutrient limitation but any perturbation on nutrient uptake and regulation systems causes the same effect too. This effect can be seen on the significant change of degradation of some metabolites such as nucleotides. Lack of regulatory systems and proteins in genome-scale metabolic models can cause some errors and wrong predictions. In genetic engineering, mostly regulatory proteins are selected as targets to knock-out or over-expression. In this study we invetigated *S. coelicolor* at metabolic network level and results are all on metabolite and pathway level.

As a conclusion, detailed investigation of the complex regulatory and metabolic network of S. coelicolor also gives important clues on the properties

important for medical and commercial approaches. Yet, it enlightens their evolutionary properties as well.

# REFERENCES


[1] Abdi H., Williams L.J., (2010) "Principal component analysis", 2010 John Wiley & Sons, Incorporated, 2, 433 - 459.

[2] Agren R., Liu L., Shoaie S., Vongsangnak W., Nookaew I., et al., (2013) "The RAVEN Toolbox and Its Use for Generating a Genome-scale Metabolic Model for Penicillium chrysogenum", PLoS Computational Biolology, 9(3), 1 – 16.

[3] Alam et al., (2010), "Metabolic modeling and analysis of the metabolic switch in Streptomyces coelicolor. BioMedCentral Genomics, 11(202), 1 - 9.

[4] Alam M., Merlo M., Takano E., Breitling R., et al., (2010) "Genome-based phylogenetic analysis of Streptomyces and its relatives", Molecular Phylogenetics and Evolution, 54(2010), 763 - 772.

[5] Alam M.T., (2010), "Metabolic modeling of *Streptomyces* and its relatives: A constraints-based approach", PhD Thesis, Nijmegen University.

[6] Almaas E., Kovacs B., Vicsek T., Oltvai Z.N., Barabasi A., (2004), "Global organization of metabolic fluxes in the bacterium Escherichia coli". Nature, 427, 834 – 843.

[7] Alon U., (2007), "Introduction to Systems Biology: Design Principles of Biological Circuits", 1st Edition, Chapman & Hall/CRC.

[8] Anderson N. L., Anderson N. G., (1998), "Proteome and proteomics: New technologies, new concepts, and new words". Electrophoresis 1998, 19, 1853-1861.

[9] Bachmaier C., (2013), "Handbook of Graph Drawing and Visualization", 1st Edition, CRC Press.

[10] Basak K. Majumdar S.K., (1973), "Utilization of Carbon and Nitrogen Sources by Streptomyces kanamyceticus for Kanamycin Production", Antimicromal Agents and Chemotherapy, 4(1), 6-10.

[11] Becker S.A., Feist A.M., Mo M.L., Hannum G., Palsson B., Herrgard M., (2007), "Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox", Nature Protocols, 2(3), 727 - 738

[12] Becker S.A., Palsson B.O., (2008), "Context-Specific Metabolic Networks Are Consistent with Experiments", PLoS Computational Biology, 4(5), 1 - 10

[13] Bibb M., (2005), "Regulation of secondary metabolism in streptomycetes", Elsevier Current Opinion in Microbiology, 8, 208 – 215.

[14] Blackstock W., Weir M.P., (1999) "Proteomics: quantitative and physical mapping of cellular proteins", Elsevier TibTech, 17, 121 - 127.

[15] Borodina I., Krabben P., Nielsen J., (2005), "Genome-scale analysis of Streptomyces coelicolor A3(2) metabolism", Cold Spring Harbor Laboratory Press Genome Research, 15(5), 820 – 829..

[16] Bowman A.W., Robinson D.R., (1987), "Introduction to Statistics", 5th Edition, CRC Press.

[17] Boygo, M., Rudd P.M., (2013), "New technologies and their impact on 'omics' research", Elsevier Current Opinion in Chemical Biology 2013, 17(2), Pages 1–3.

[18] Bro C., Regenberg B., Förster J., Nielsen J., (2006), "In silico aided metabolic engineering of Saccharomyces cerevisiae for improved bioethanol production", Elsevier Metabolic Engineering, 8(2006), 102 - 111.

[19] Chuang H., Hofree M, Ideker T., (2010), "A Decade of Systems Biology", Annual Review of Cell and Developmental Biology, 26(44), 721 – 744.

[20] Cohen G., Yanko M., Mislovati M., Argaman A., Schreiber R., Av-gay Y., Aharonowitz Y., (1993), "Thioredoxin-Thioredoxin Reductase System of *Streptomyces clavuligerus*: Sequences, Expression, and Organization of the Genes", Journal of Bacteriology, 175(16), 5159-5167.

[21] Corre C., Haynes S.W., Malet N., Song L., Challis G.L., (2010), "A butenolide intermediate in methylenomycin furan biosynthesis is implied by incorporation of stereospecifically 13C-labelled glycerols", The Royal Society of Chemistry Journal, 46(23), 4079 – 4081.

[22] Çakır T., (2015), "Reporter pathway analysis from transcriptome data: Metabolitecentric versus Reaction-centric approach", Scientific Reports, 5, 14563, doi: 10.1038/srep14563.

[23] Davidson E., (2005), "Gene regulatory networks", Proceedings of National Academy of Sciences, 102(14), 4935.

[24] Donalson R., Talcott C., Knapp M., Calder M., (2010),"Understanding Signalling Networks as Collections of Signal Transduction Pathways",International Conference on Computational Methods in Systems Biology, Trento- Italy, September 29 – October 1

[25] Dworkin M., Falkow S., Rosenberg, Schleifer K, Stackebranda., "The Prokaryotes Third Edition: A Handbook on the Biology of Bacteria", 3rd Edition, Springer, 2006.

[26] Edwards J. , Covert M., Palsson B., (2002), "Metabolic modelling of microbes: the flux-balance approach", Journal of Environmental Microbiology, 4(3), 133–140.

[27] Edwards J. , Palsson B., (2000), “Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* K-12 gene deletions”, BMC Bioinformatics, 2000, 1(1), 1471 – 2105.

[28] Fink D. , Weiβschuh N., Reuter J., Wohlleben W., Engels A., (2002), “Two transcriptional regulators GlnR and GlnRII are involved in regulation of nitrogen metabolism in Streptomyces coelicolor A3(2).” Journal of Molecular Microbiology, 2(46), 331–347.

[29] Fleischmann R.D., Adams M. D., White O., Clayton A. R., Kirkness E.F. , Kerlavage A. R ., Bult C. J. , Tomb J. F. , Dougherty B. A., Merrick J. M., (1995), “Whole-genome random sequencing and assembly of Haemophilus influenzae rd.” Science, 269(5223), 496–512.

[30] Gabig M., Wegrzyn G., (2001) “An introduction to DNA chips: principles, technology, applications and analysis”, Acta Biochimica Polonica, 48(3), 615 – 622.

[31] Gesheva V., Ivanova V., Gesheva R., (2005), “Effects of nutrients on the production of AK-111-81 macrolide antibiotic by Streptomyces hygroscopicus”, Elsevier Journal of Microbiological Research, 160(2005), 243 - 248.

[32] Graves P.R., Haystead T.A.J., (2002), “Molecular Biologist’s Guide to Proteomics”, Microbiology and Molecular Biology Reviews, 66(1), 39–63.

[33] Harrison J., Studholme D.J., “Recently published Streptomyces genome sequences”, Microbial Biotechnology by John Wiley & Sons Limited, 7(1), 373–380

[34] Hengst C.D., Tran N.T., Bibb M.J., Chandra G., Leskiw B.K., Buttner M.J., (2010) “Genes essential for morphological development and antibiotic production in Streptomyces coelicolor are targets of BldD during vegetative growth”, Molecular Microbiology, 78(2), 361–379.

[35] Hirota R., Kuroda A., Kato J., Ohtake H., (2010), “Bacterial phosphate metabolism and its application to phosphorus recovery and industrial bioprocesses”, Journal of Bioscience and Bioengineering, 109(5), 423–432.

[36] Hodgson D., “Glucose Repression of Carbon Source Uptake and Metabolism in Streptomyces coelicolor A3(2) and its Perturbation in Mutants Resistant to 2-Deox yglucose”, Journal of General Microbidogy, 128, Pages: 2417-2430.

[37] Hood D.W., Heidstra R,., Swoboda U.K., Hodgson D.A., (1992), “Molecular genetic analysis of proline and tryptophan biosynthesis in Streptomyces coelicolor A3(2): interaction between primary and secondary metabolism - a review”, Gene, 115, 5-12

[38] Hopwood D.A., (2007) “Streptomyces in Nature and Medicine: The Antibiotic Markers”, 1st Edition, Oxford University Press.

[39] Horgan RP, Kenny LC., (2011), "'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics.", The Obstetrician & Gynaecologist, 13, 189–195.

[40] Hwang K., Kim H., Charusanti P., Palsson B., Lee S., (2013), "Systems biology and biotechnology of Streptomyces species for the production of secondary metabolites.", Elsevier Biotechnology Advances, 32(5), 255–268.

[41] Hwang S.Y., Lim G., (2000), "DNA Chip Technologies", Biotechnology and Bioprocess Engineering, 5(3), 159-163.

[42] Ilic S., Konstantinovic S., Veljkovic V., Dragisa S., Gordana D., (2010), "The impact of different carbon and nitrogen sources on antibiotic production by Streptomyces hygroscopicus CH-7" , Current Research, Technology and Education Topics in Applied Microbiology and Microbial Biotechnology, Formatex 2010, 1337 – 1342.

[43] James P.D.A. , Edwards C., (1991), "Effect of Carbon Source and Oxidative Metabolism on Secondary Metabolism in *Streptomyces thermoviolaceus"*, Current Microbiology, 23(1991), 227-232.

[44] Johnson D., (1999), "The Insignificance of Statistical Significance Testing", Journal of Wildlife Management, 63(3), 763-772.

[45] Bentley S.D. K. F. Chater, A.-M. Cerden o-Tarraga, G. L. Challis, N. R. Thomson, K. D. James, D. E. Harris, M. A. Quail, H. Kieser, D. Harper, A. Bateman, S. Brown, G. Chandra, C. W. Chen, M. Collins, A. Cronin, A. Fraser, A. Goble, J. Hidalgo, T. Hornsby, S. Howarth, C.-H. Huang, T. Kieser, L. Larke, L. Murphy, K. Oliver, S. O'Neil, E. Rabbinowitsch, M.-A. Rajandream, K. Rutherford, S. Rutter, K. Seeger, D. Saunders, S. Sharp, R. Squares, S. Squares, K. Taylor, T. Warren, A. Wietzorrek, J. Woodward, B. G. Barrell, J. Parkhill, D. A. Hopwood, (2002), "Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2)", Nature, 417, 141 – 147.

[46] Kaddurah-Daouk R., Kristal B.S., Weinshilboum R.M., (2008), "Biochemical Approach to Drug Response and Disease.", Annual Review of Pharmacology and Toxicology, 48, 653 - 683.

[47] Kauffman K., Prakash P., Edwards J., (2003), "Advances in flux balance analysis", Elsevier Current Opinion in Biotechnology, 14, 491–496.

[48] Keating S.M., Bornstein B. J., Finney A., Hucka M., (2006) "SBMLToolbox: an SBML toolbox for MATLAB users", Bioinformatics Application Notes, 22(10), 1275–1277.

[49] Keulen G., Dyson P., (2014), "Production of Specialized Metabolites by *Streptomyces coelicolor* A3(2).", Advances in Applied Microbiology, 89(3), 218 – 256.

[50] Kim M., Yi J.S., Kim J., Kim J., Kim M.W., Kim B., (2014), "Reconstruction of a high-quality metabolic model enables the identification of gene overexpression targets for enhanced antibiotic production in Streptomyces coelicolor A3(2).", Biotechnology Journal, 9(, 1185–1194.

[51] Kitano, H., (2010) "Computational systems biology", Nature, 420, 206 - 210.

[52] Kitano, H., (2010a) "Systems Biology: A Brief Overview", Science, 295, 1662 - 1664.

[53] Klipp E., Herwig R., Kowald A., Wierling C., Lehrach H., (2005), "Systems Biology In Practice", 4th Edition, Wiley-Vch Verlag GmbH & Company.

[54] Koboldt D.C., Steinberg K.M., Larson D.E., Wilson R.K., Mardis E.R., (2013), "The Next-Generation Sequencing Revolution and Its Impact on Genomics" , Cell, 155, 27 - 38.

[55] Kukurba R.K., Montgomery S.B., (2016), "RNA Sequencing and Analysis", Cold Spring Harbor Protocols, 10, 951 – 970.

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

[57] Levine M., Davidson E.H., (2005), "Gene regulatory networks for development", Proceedings of National Academy of Sciences, 102(14), 4936 - 4942.

[58] Lieber R.L., (1990), "Statistical Significance and Statistical Power in Hypothesis Testing", Journal of Orthopaedic Research, Raven Press, 8, 304 – 309.

[59] Likic V.A., McConville M.J., Lithgow T., Bacic A., (2010), "Systems Biology: The Next Frontier for Bioinformatics", Advenced in Bioinformatics, 2010, 1 - 10.

[60] Liu L, Agren R., Bordel S., Nielsen J., (2010), "Use of genome-scale metabolic models for understanding microbial physiology", Federation of European Biochemical Societies Letters, Elsevier, 584(2010), 2556 - 2564.

[61] Ma Y., Zhang P., Yang Y., Wang F., Qin H., (2012), "Metabolomics in the fields of oncology: a review of recent research", Molecular Biology Reports, 39, 7505–7511

[62] Madden T., Ward J.M., Ison A.P., (1996), "Organic acid excretion by *Streptomyces lividans* TK24 during growth on defined carbon and nitrogen sources", Microbiology, 142, 3181 – 3185

[63] Marcus F.B. (2008), "Bioinformatics and Systems Biology", 1st Edition. Springer-Verlag Berlin Heidelberg.

[64] Martin J.F., Santos-Beneit F., Rodriguez-Garcia A., Sola-Landa A., (2012), ”Transcriptomic studies of phosphate control of primary and secondary metabolism in *Streptomyces coelicolor*”, Applied Microbiol Biotechnology, 95, 61–75.

[65] Masuma R., Tanaka Y., Tanaka H., Omura S., (1986), “Production Of Nanaomycin And Other Antibiotics By Phosphate-Depressed Fermentation Using Phosphate-Trapping Agents”, The Journal Of Antibiotics, 29(11), 1557 - 1564

[66] Miller W., Makova K.D., Nekrutenko A., Hardison R.C., (2004), “Comparative Genomics”, Annual Reviews of Genomics and Human Genetics, 5, 15–56.

[67] Mo S., Sydor P.K., Corre C., (2008), “Elucidation of the *Streptomyces coelicolor* Pathway to 2-Undecylpyrrole, a Key Intermediate in Undecylprodiginine and Streptorubin B Biosynthesis”, Elsevier Journal of Chemistry & Biology, 15(1), 137–148.

[68] Mould R., (1998), “Introductory Medical Statistics 3rd Edition”, 4th Edition. IOP Publishing Limited,

[69] Narayana K.J.P., Vijayalakshmi M., (2008), “Optimization of Antimicrobial Metabolites Production by Streptomyces albidoflavus”, Research Journal of Pharmacology, 2(1), 4-7.

[70] Nicholas E. E. A., Laing E., Bucca G., Kierzek A. M., Smith C. P., (2012), “Diverse control of metabolism and other cellular processes in Streptomyces coelicolor by the PhoP transcription factor: genome-wide identification of in vivo targets.”, Nucleic Acid Research, 40(19), 9543 – 9556.

[71] Nieselt K., Battke F., Herbig A., Bruheim P., Wentzel A., Jakobsen O.M., Sletta H., Alam M.T., Merlo M.E., Moore J., Omara W.A.M., Morissey E.R., Juarez-Hermosillo M.A., Rodriguez-Garcia A., Nentwich M., Thomas L., Iqbal M., Legaie R., Gaze H.W., Challis L.G., Jansen R.C., Dijkhuizen L., Rand D.A., Wild D.L., Bonin M., Reuther J., Wohlleben W., Smith M.C.M., Burroughs N.J., Martin J.F., Hodgson D.A., Takano E., Breitling R., Ellingsen T.E., Wellington E.M.G., (2010), “The dynamic architecture of the metabolic switch in Streptomyces coelicolor”. BMC Genomics, 10(1), 1 – 9.

[72] Orth J., Thiele I., Palsson B., (2010), “What is flux balance analysis?”, Nature Biotechnology Journal, 28(3), 245–248.

[73] Palsson B., “Systems Biology: Properties of Reconstructed Networks”, 1st Edition Cambridge University Press.

[74] Patil K.R., Akesson M., Nielsen J., (2004), “Use of genome-scale microbial models for metabolic engineering”, Current Opinion in Biotechnology, 15, 64–69.

[75] Patil K.R., Nielsen J., (2005), "Uncovering transcriptional regulation of metabolism by using metabolic network topology", Proceedings of National Academy of Sciences, 102(8), 2685 – 2689.

[76] Perez-Redondo R, Rodriguez-Garcia A, Botas A, Santamarta I, Martin JF, Liras P., (2012), "*ArgR* of Streptomyces coelicolor Is a Versatile Regulator". PLOS One, 7(3), 1 – 9.

[77] Petit C., (2013), "Evodevomıcs", BioSciences Master Reviews, 10(01), 1 – 8.

[78] Pilpel Y., (2011), "Noise in Biological Systems: Pros, Cons, and Mechanisms of Control", Methods in Molecular Biology, 759, 407 - 425

[79] Pullan S.T., Chandra G., Bibb M.J., Merrick M., (2011), "Genome-wide analysis of the role of GlnR in *Streptomyces venezuelae* provides new insights into global nitrogen regulation in actinomycetes". BioMedCentral Genomics, 12(175), 1 – 14.

[80] Richardson M., (2009), "Principle Component Analysis", Lecture Notes, Federal University of Campina Grande, Brazil.

[81] Rodriguez-Garcia A., Sola-Landa A., Apel K., Santos-Beneit F., Martin J.F., (2009), "Phosphate control over nitrogen metabolism in *Streptomyces coelicolor*: direct and indirect negative control of glnR, glnA, glnII and amtB expression by the response regulator PhoP", Nucleic Acids Research, 37(10), 3230 – 3242.

[82] Romantsov C., Guan Z., Wood J., (2009), "Cardiolipin and the osmotic stress responses of bacteria", Biochimica et Biophysica Acta, 1788(2009), 2092–2100.

[83] Sertbaş M., "Investıgatıon Of Effects Of Neurological Diseases On Human Brain Metabolism: Computatıonal Systems Biology Approach", B.S. Thesis, Chemical Engineering, Hacettepe University, 2009

[84] Shalab N., Kamaruddin K., Platt J., Butler P.R., Oliver S.G., Hobbs G., (1994), "Cell Physiology And Antibiotic Production Of *Streptomyces coelicolor* Grown On Solid Medium", Biotechnology Letters, 16(10), 1015 – 1020.

[85] Sharan R., (2007), "Analysis of Biological Networks: Transcriptional Networks - Promoter Sequence Analysis", Lecture Notes 11, January 2007, The Israel Young Academy, Tel Aviv University.

[86] Shlens J., (2003), "A Tutorıal On Prıncıpal Component Analysis: Derivation, Discussion And Singular Value Decomposition", Princeton University, Lecture Notes.

[87] Smith E., Morowitz H.J., (2004), “Universality in intermediary metabolism”, Proceedings of National Academy of Sciences, 101(36), 13168 – 13173.

[88] Sohn S.B., Graf A. B., Kim T. Y., Gasser B., Maurer M., Ferrer P., Mattanovich D., Lee S. Y., (2010), “Genome-scale metabolic model of methylotrophic yeast *Pichia pastoris* and its use for in silico analysis of heterologous protein production”, Biotechnology Journal, 5(2010), 705–715.

[89] Sohn S.B., Kim T.Y., Lee J. H., Lee S. Y., (2011), “Genome-scale metabolic model of the fission yeast Schizosaccharomyces pombe and the reconciliation of insilico/in vivo mutant growth”. BioMedCentral Systems Biology, 6(49), 1 – 12.

[90] Stegle O., Teichmann S.A., Marioni J.C., (2014), “Computational and analytical challenges in single-cell transcriptomics”, Nature Reviews Genetics, Advence Online Publication doi:10.1038/nrg3833.

[91] Sun Y., Fleming R. M. T., Thiele I., Saunders M. A., (2013), “Robust flux balance analysis of multiscale biochemical reaction networks.”, BioMedCentral Bioinformatics, 14(240).

[92] Tarkka M., Hampp R., (2008), “Secondary Metabolites of Soil Streptomycetes in Biotic Interactions”, Secondary Metabolites in Soil Ecology, Chapter 6, 1st Edition, Springer.

[93] Tawfik K.A., Ramadan E. M., (1991), “Factors Affecting the Biological Activity of Streptomyces aureofaciens MY18 and Str. Roseviolaceus”,Journal of King Abdulaziz University, 3, 5-19.

[94] Terzer M., Maynard N.D., Covert M.W., Stelling J., (2009), “Genome-scale metabolic networks”, 2009 John Wiley & Sons, Incorporated., 1(2009), 285 – 297.

[95] Thiele I., Palsson B., (2010), “A protocol for generating a high-quality genome-scale metabolic reconstruction.”, Nature Protocols, 5(1), 93 – 121.

[96] Thomas L., Hodgson D.A., Wentzel A., Nieselt K., Ellingsen T.E., Moore J., Morrissey E.R., Legaie R., The STREAM Consortium, Wohlleben W., Rodríguez-García A., Martín J., Burrough N.J., Wellington E. and Smith M.C.M., (2012), “Metabolic Switches and Adaptations Deduced from the Proteomes of Streptomyces coelicolor Wild Type and *phoP* Mutant Grown in Batch Culture”, Mollecular and Cellular Proteomics, 11(2), 1 - 17.

[97] Tiffert Y., Supra P., Wurm R., Wohlleben W., Wagner R., Reuther J., (2008), “The Streptomyces coelicolor GlnR regulon: identification of new GlnR targets and evidence for a central role of GlnR in nitrogen metabolism in actinomycetes”, Molecular Microbiology, 67(4), 861–880.

[98] Upton G., Cook I., (2014), “A Dictionary of Statistics”, 1st Edition, Oxford University Press.

[99] Villas-Boas S.G., “Analytical Techniques & Applications of Metabolomics in Systems Medicine and Systems Biotechnology”, (2013), Computational and Structural Biotechnology Journal, 4(5), 1.

[100] Villas-Boas S.G., Rasmussen S., Lane G.A., (2005), “Metabolomics or metabolite profiles?”, Trends in Biotechnology, 23(8), 385 – 386.

[101] Waksman S.A., (1946), “Certain Aspects Of The Physiology Of Actinomycetes”, Jouraal Series Paper, Department of Microbiology, New Jersey Agricultural Experiment Station, Rutgers University, New Brunswick, N.J., 1946.

[102] Wang Z., Gerstein M., Snyder M., (2009), “RNA-Seq: a revolutionary tool fortranscriptomics”, Nature Reviews, Genetics, 10(1), 57 – 63.

[103] Wuchty S., Ravasz E., Barabasi A., (2006), “The Architecture of Biological Networks”, Department of Physics, 225 Nieuwland Science Hall, University of Notre Dame

[104] Yuksel I., “Reporter Pathway Analysis algoritmasının gelistirilmesi ve yonteminin kok hucre mikrodizin verilerine uygulanması” (2016), Bachelors Degree Thesis, Gebze Technical University.

[105] Zhou Z., Gu J., Du Y., Li Y., Wang Y, (2011) “The -omics Era- Toward a Systems-Level Understanding of Streptomyces, Current Genomics, 12(2011) , 404 - 416

[106] Zhou Z., Gu J., Li Y., Wang Y., (2012) “Genome plasticity and systems evolution in Streptomyces”. BioMedCentral Bioinformatics, 13(10), 1 – 17.

[107] Zomorrodi A.R., Suther P.F., Ranganathan S., Maranas C.D. ., “Mathematical optimization applications in metabolic networks”, Elsevier Metabolic Engineering, 14 (2012), 672–686.

# BIOGRAPHY

Hamza Umut Karakurt was born in Golcuk/KOCAELI in 1990. He graduated from Ege University Biochemistry Department with Biotechnology major in 2013, same year he starter Molecular Biology and Genetics Msc. Degree education in Gebze Technical University (former Gebze Institute of High Technology). He is also a Msc. student in Bioinformatics and Systems Biology Department in the same university.