國立交通大學

生物醫學所

碩士論文

系統化的方法將植物二次代謝實現於微生物中:
用大腸桿菌生合成白藜蘆醇作為個案研究

A systematic method to implement a plant secondary metabolism in microbes: synthetically produce resveratrol in *Escherichia coli* as a case study

研究生：周致宏

指導教授：黃憲達　博士

中華民國九十八年七月

系統化的方法將植物二次代謝實現於微生物中: 用大腸桿菌生合成白藜蘆醇作為個案研究

A systemic method to implement a plant secondary metabolism in microbes: synthetically produce resveratrol in *Escherichia coli* as a case study

研 究 生：周致宏　　Student：Chih-Hung Chou

指導教授：黃憲達　　Advisor：Hsien-Da Huang

國 立 交 通 大 學

生 物 醫 學 所

碩 士 論 文

A Thesis

Submitted to Institute of Biomedicine Science

College of Biological Science and Technology

National Chiao Tung University

in partial Fulfillment of the Requirements

for the Degree of

Master

in

Biomedicine Science

July 2009

Hsinchu, Taiwan, Republic of China

中華民國九十八年七月

系統化的方法將植物二次代謝實現於微生物中:用大腸桿菌生合成白藜蘆醇作為個案研究

學生：周致宏　　指導教授：黃憲達

國立交通大學生物醫學研究所碩士班

摘　　要

合成生物學是一個近年迅速發展的跨學門研究，結合科學與工程的技術來設計和合成一個嶄新的生物功能系統。近年來發現植物二次代謝物對人類的健康有益，大宗為酚類化合物，像是異黃酮素、兒茶素和白藜蘆醇，皆是對人類有益的酚類化合物。基於上述兩點，本研究利用合成生物學的宗旨，在微生物中合成植物二次代謝物。研究一開始先收集並整合資料庫的資料，用有系統的方法來重建代謝途徑，包括植物的二次代謝。呈現重建的代謝途徑為新穎的 KEGG 連接地圖：將 KEGG 的地圖做連接並標上經過的所有代謝物和酵素。再者，使用跨物種分析所重建的代謝途徑，可了解此代謝途徑存在哪些物種上，提供每個酵素的基因，蛋白質和後轉譯修飾之資訊讓研究者了解是否可合成感興趣的代謝物在其它物種上。白藜蘆醇是一種抗氧化劑，能抑制發炎反應和腫瘤生長，有很大的經濟和研究價值。本研究選其生合成路徑在大腸桿菌中做個案研究。首先，系統化方法有詳細且視覺化的初步結果，發現酪胺酸到白藜蘆醇這段代謝途徑必須從植物轉殖到大腸桿菌中。第二，建立代謝網路並用基本通量路徑分析，以得知最高產率的生合成路徑，進一步刪除合成苯丙胺酸和色胺酸的代謝途徑後，發現可能可以提高白藜蘆醇生合成的效率。

此系統化的分析可以快速的得到非常多合成生物學和代謝工程需要的資訊，例如酵素從哪一種物種產生？其序列為何？它轉譯的蛋白質是否有後轉譯修飾？此系統化分析方法已經建置成網路工具 FMM(http://FMM.mbc.nctu.edu.tw/)。可應用在合成生物學上生產藥物和生質能源。此外，基本通量路徑分析可以評估最終產物的產率，改善傳統代謝工程隨機突變基因所消耗的金錢及時間。

A systematic method to implement a plant secondary metabolism in microbes: synthetically produce resveratrol in *Escherichia coli* as a case study

student：Chih-Hung Chou　　Advisors：Dr. Hsien-Da Huang

Institute of Biomedicine Science
National Chiao Tung University

ABSTRACT

Synthetic Biology, a multidisciplinary field, is growing rapidly. To understand biological systems clearly, scientists through mimicking and producing bio-orthogonal systems with new functions are two complementary pursuits. Recently, plant secondary metabolites are increasing importantly for human. The phenolic compounds, which are kinds of plant secondary metabolites, contain many important useful components, such as resveratrol, isoflavones, and catechin that are healthy for human. For the above reasons mentioned, this study for synthetic biology purposes is developed that synthesizes plant secondary metabolites in microbes. In the beginning of this study, the relative data was collected and integrated from databases. The metabolic pathway was reconstructed by systematic method containing plant secondary metabolism. Novel presentation for connecting different KEGG maps is newly provided. Furthermore, comparative analysis the reconstruction metabolic pathway will understand enzyme information about genes, proteins, and post-translational modifications (PTMs), which provide messages to find out whether a metabolite is synthesized by others species or not. Resveratrol is an antioxidant, which inhibits inflammation and tumor promotion and contains economic and scientific value. This study was chosen as a case. First, systematic method provides visualization primary results to find out the pathway synthesized from tyrosine to resveratrol having to be cloned from plant. Second, metabolic network is constructed and analyzed by elementary flux modes analysis (EFMs). The result shows that which pathway has maximum yield. Moreover, after deleting the phenylalanine and tryptophan biosynthesis pathway, the production of resveratrol may become efficient.

The systematic method is an efficient analysis to obtain information rapidly for synthetic biology and metabolic engineering, such as enzyme, sequence, and PTMs. The method has been developed to web tool: http://FMM.mbc.nctu.edu.tw/, which is an effective tool for applications in synthetic biology to produce both drugs and biofuels. Besides, traditional optimization strategies of random mutagenesis and selection will eventually have limited efficacy. However, EFMs can estimate the yield of end product for improving traditional methods to spend less time and cost.

誌　　謝

在交通大學碩士班兩年的日子裡，首先要非常感謝黃憲達教授提供這麼好的研究環境，加上老師的淵博學問，總是能夠在研究上提供很多很好的指引和建議，讓我能很快的從生物領域進入到生物資訊領域中。而實驗室的各位學長姊，也是幫助我能進入生物資訊領域做研究的一個推手，一開始是熙淵學長給了很多程式和資訊方面的知識，同時博凱學長、威霽學長、勝達學長、宗夷學長也都是很好討教的對象，在程式或研究機器上出了問題，都可以請教他們而很快的得到解決方式。而碩二時，實驗室又加入很多的新血，其中文綺學姊在我碩二期間幫助我非常的多，也是能完成此論文的一個關鍵。在碩一研究的一些資訊方法而碩二要將這些資訊方法真實實現在生物體中時，黃憲達老師跟文綺學姊真的是精通生物和資訊兩大領域，在不斷的研究與討論之下，讓我碩二作研究時能如此的順利，另外致閔和至昶也幫了很大的忙。而碩士班這兩年一直都坐我旁邊的恆毅，真的是非常的全能，無論在研究、伙食、玩樂、運動上都可以找他。昭昉也參與討論本論文題目，讓此題目更為明確。豐茂學長則給了很多報告和撰寫論文的建議。以上，感謝老師和各位實驗室成員的幫助。另外，在我碩一有許多籃球球友，碩二有更多的壘球球友，讓我除了研究之外，還有一個不錯的休閒運動。

最後要感謝我的父母親，感謝你們的支持和栽培，也要感謝我的哥哥和姊姊，你們在家中陪伴著父母親讓我無憂無慮的在新竹念書，在這裡由衷的感謝我最愛的家人。

Table of Contents

List of Figures

List of Tables

Chapter1 Introduction

1.1 Overview of schema

The plant secondary metabolites are useful because they are potential source as new natural drugs, antibiotics, insecticides, and herbicides[1]. Synthetic Biology, a rapidly growing multidisciplinary field, has two complementary goals- further elucidating biological systems through mimicry and producing bioorthogonal systems with new functions[2]. The diagram explains synthetic biology is shown in Figure 1.1. This study aims to implement a plant secondary metabolic pathway into microbes in order to accelerate and affect growth. Recently, because of the variety of biological database release and the bioinformatics tools development, we could use the useful data and tools to analysis and evaluate the primary result. Compare with the traditional methods about literature survey and experiment, a systematic method analysis will accelerate and primary understand whether the metabolic engineering experiment is working or not. Therefore, we also use the metabolic pathway analysis tools to model the pathway from central metabolism to resveratrol as a case study. The following will introduce the background about this research.

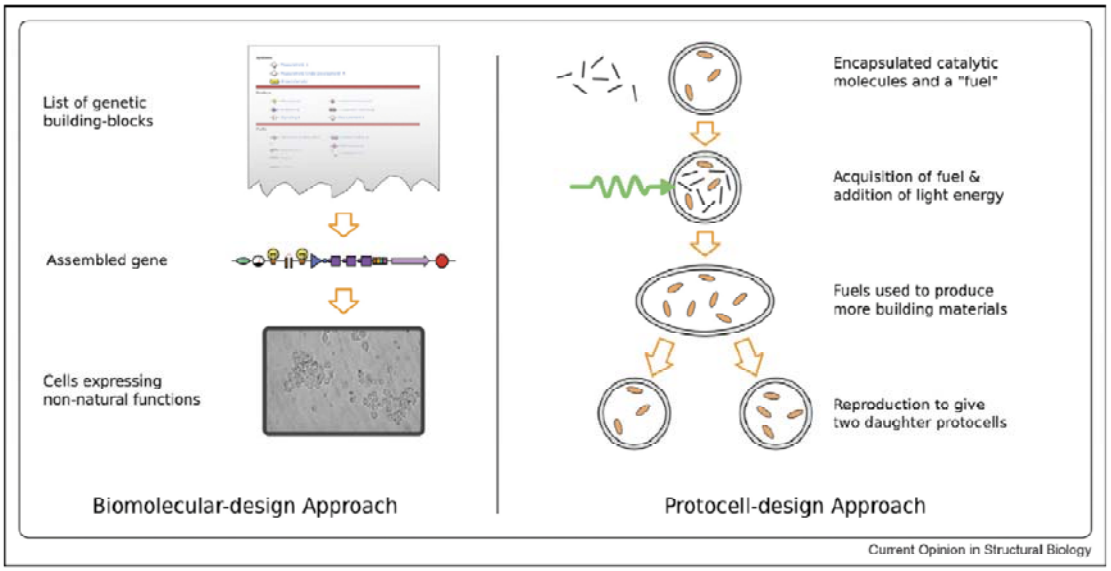


Figure 1.1 Using synthetic biology to building complex systems
(Channon, K. *et al.* 2008)
The left figure is BioBricks project (http://www.biobricks.org/). They aim to produce a wide range of standard 'parts', analogous to components in an electronic circuit, which can be added to a host 'chassis' to produce novel functions. The right figure is projects such as the Los Alamos 'protocell' aim to create an entirely new minimal reproducing machine.

1.1.1 The background of biological problems

1.1.1.1 The plant secondary metabolism

The metabolism is an important mechanism of physiologically in all organisms. All of the organisms have primary metabolism, which is a general pathway for synthesis the essential compounds and macromolecular to support growth and essential physiologically. For example, carbohydrates, proteins, fats, and nucleic acids are primary metabolites. In contrast to primary metabolism, the secondary metabolism occurs only in some organism such as plants, fungus, and some microbes (e.g. *Streptomyces coelicolor*). Secondary metabolites play a key role in plant survival because they can protect plants from herbivores and microbial infection, as attractants for pollinators and seed-dispersing animals, as allelopathic agents, UV protectants and signal molecules in the formation of nitrogen-fixing root nodules in legumes. They

are also used as dye, fibers, glues, oils, waxes, flavouring agents, drugs and perfumes[3]. For the above reasons given, we consider that the plant secondary metabolites have very important economic and scientific value. Figure 1.2 shows the classification of plant secondary metabolites.

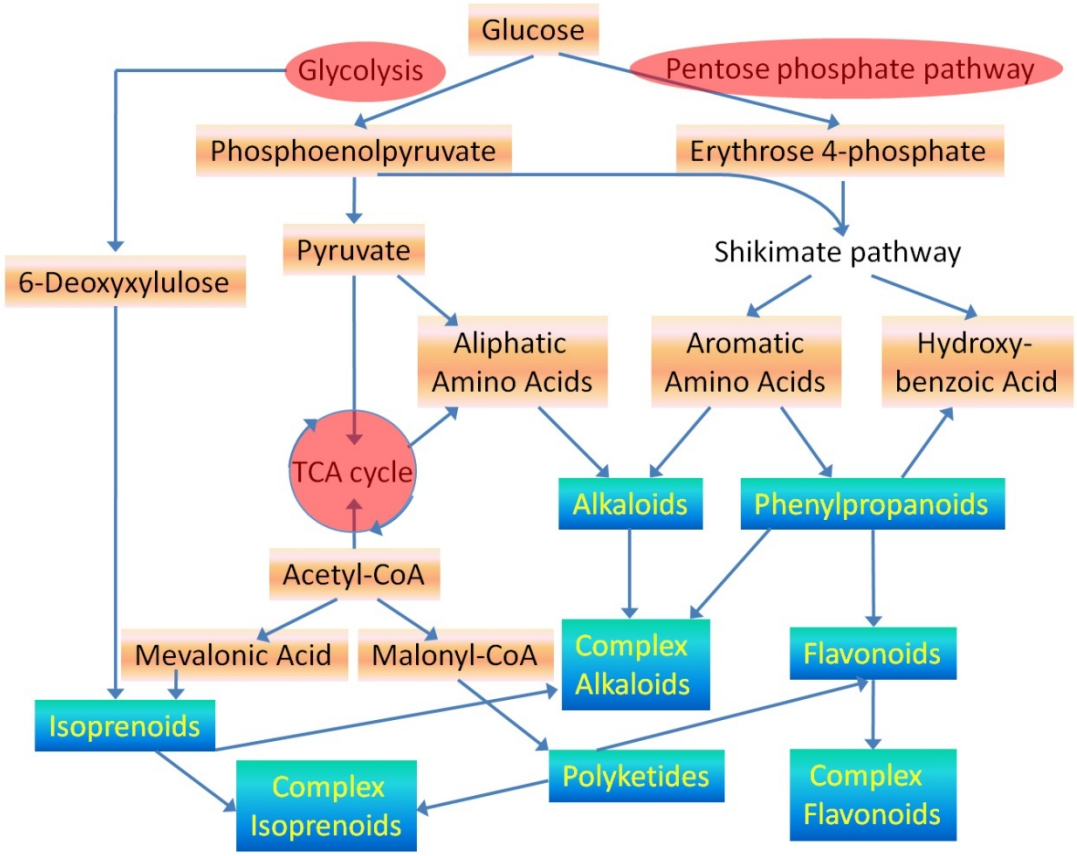


Figure 1.2 Overview of the plant metabolism involved secondary metabolism
There can be classified into eight kind of secondary metabolite involved isoprenoids, complex isoprenoids, alkaloids, complex alkaloids, polyketides, phenylpropanoids, flavonoids, and complex flavonoids show in blue box. Gold box show the primary metabolic. The central metabolic pathways are shown by red box.

1.1.1.2 Phenolic compounds

The definition of phenolic compounds is that they have one or more hydroxyl groups attached to an aromatic ring[4] (see Figure 1.3 (A)). Table 1.1 shows the main phenolic compounds in the natural world. Phenylpropanoids, which are classes of

plant secondary metabolites, are main phenolic compounds in plant (see Figure 1.2). The major role in nature is to protect plants. Many Chinese medicine research display the ingredient of Chinese herbal medicine are abundant of phenolic compounds[5]. Recently, their effects on human health are respected. The oldest medical is the usage of phenol as an antiseptic. However, the phenolic compound contain aromatic ring that can absorb UV-B radiation from the sum and thus prevent sunburns. The traditional sunscreens contain phenolic compound. In modern medicine, we used phenolic compound in following ways. Isoflavones (Figure 1.3 (B)) have estrogenic activity, which is an important mimicry hormone in women. Another important characteristic for phenol compound is that they have antioxidant properties. Catechin (Figure 1.3 (C)) is a powerful, water soluable polyphenol and antioxidant that exists in green tea. And another important antioxidant is resveratrol (Figure 1.3 (D)), which exists in red wine, grapes and peanuts.

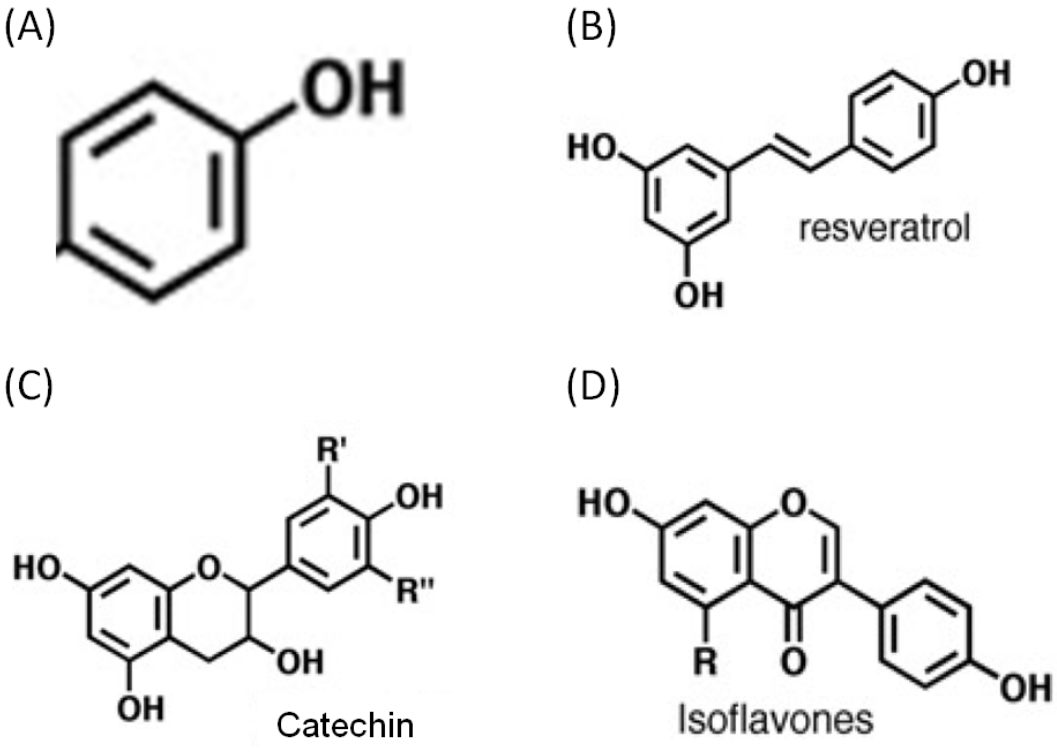
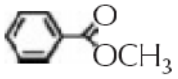
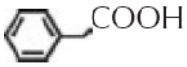
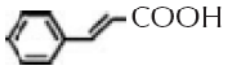
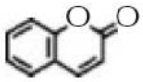
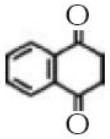
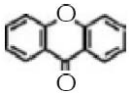
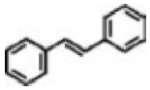
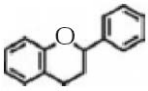


Figure 1.3 Phenolic compounds

(A) The example of phenolic, which have one or more hydroxyl groups attached to an aromatic ring. The (B) (C) (D) are chemical structure of isoflavones, catechin, and resveratrol, respectively.

Table 1.1 Structural skeletons of phenolic and polyphenolic compounds

Number of carbons	Skeleton	Classication	Example	Basic structure
7	C_6–C_1	Phenolic acids	Gallic acid	
8	C_6–C_2	Acetophenones	Gallacetophenone	
8	C_6–C_2	Phenylacetic acid	*p*-Hydroxyphenyl -acetic acid	
9	C_6–C_3	Hydroxycinnamic acids	*p*-Coumaric acid	
9	C_6–C_3	Coumarins	Esculetin	
10	C_6–C_4	Naphthoquinones	Juglone	
13	C_6–C_1–C_6	Xanthones	Mangiferin	
14	C_6–C_2–C_6	Stilbenes	Resveratol	
15	C_6–C_3–C_6	Flavonoids	Naringenin	

1.1.1.3 Synthetic biology

Synthetic biology is integrated by different field, including biology, chemistry, engineering, and some others field. The biologists have the ability to design the construction of the systems of synthetic biology and provide a direct and compelling method for testing the current understanding of natural biological systems. The chemists are able to create novel molecules and molecular systems, which promote the development of useful diagnostic assays, drugs, and biofuels. The re-writers mean the new genomes encoding natural biological systems. The engineers used the

technology to build up a system that works in genetic engineering and development of foundational technologies. Totally, the synthetic biology makes the design and construction of engineered biological systems easier[6]. Heinemann *et al.* think the synthetic biology is divided into two parts – design and fabrication. They consider that at the beginning of study synthetic biology, it has to possess the biological knowledge and then the computational tools would help the systems design. When the systems are designed by some knowledge and computational computing, the biological experiment such as cloning and DNA synthesis will implement a novel biological system [7] (see Figure 1.4).

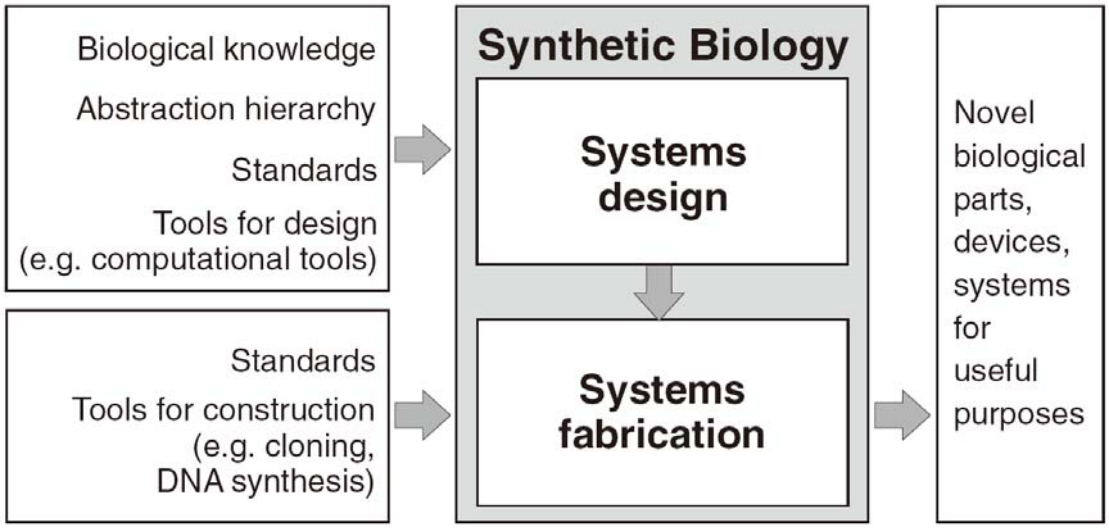


Figure 1.4 Synthetic biology — the systems design and system fabrication
(Heinemann, M. and S. Panke, 2006)

1.1.1.4 Suitability species for metabolic engineering

Microbes are well established as effective hosts for the biosynthesis of bio-molecules; consequently, engineered microbial biological systems represent critical frontier in synthetic biology[8]. Engineered microorganisms are employed in numerous applications, including food additives, pharmaceuticals, fuels, animal feed supplements, cosmetics and polymer materials[9]. For example, *Corynebacterium glutamicum* and *Escherichia coli* are used to produce lysine, methionine, valine and

threonine [9-13] which are essential intermediate precursors for antibiotics [14] and biofuels [15].

1.1.1.5 Protein post-translational modifications (PTMs)

If we choose microbes about *E. coli* or others prokaryote as a metabolic engineering host and the genes object to clone are eukaryote, we may suffer a problem in prokaryote because they do not exist PTMs. PTMs is an extremely important cellular control mechanism because it may alter proteins' physical and chemical properties folding, conformation distribution, stability, activity, and consequently, their functions[16]. In normal eukaryotic cell, they are more than 200 types of PTMs. Some important about acetylation, methylation, phosphorylation, ubiquitination, and GPI anchor [17] are shown in Figure 1.5.

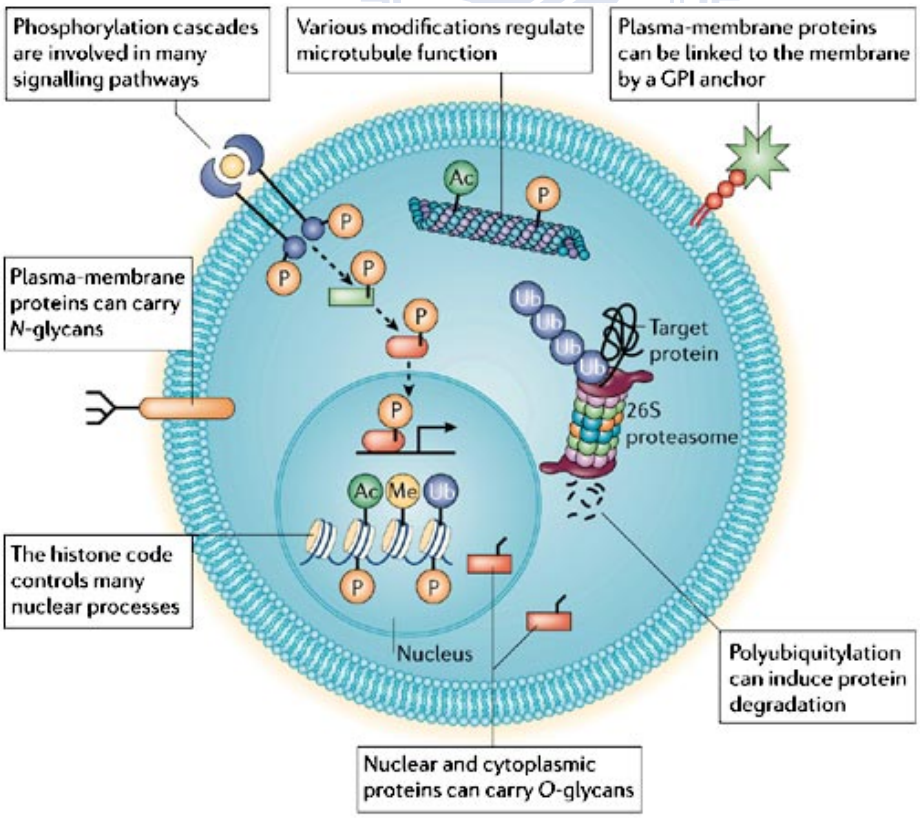


Figure 1.5 Cellular protein post-translational modifications
(Jensen, O.N. *et al.* 2006)
This figure shows some important post-translational modifications (PTMs) about acetylation, methylation, phosphorylation, ubiquitination, and GPI anchor. In normal cell, they are more than 200 types of PTMs.
Ac, acetyl group; GPI, glycosylphosphatidylinositol; Me, methyl group; P, phosphoryl group; Ub, ubiquitin.

1.1.2 Databases for metabolic pathway analysis

The general databases for metabolic pathways are *Kyoto Encyclopedia of Genes and Genomes* (KEGG) [18] and MetaCyc[19]. The KEGG database integrates genomic, chemical, metabolic pathways, and systematic functional information. MetaCyc is a non-redundant database and contains more than 1100 experimentally elucidated metabolic pathways from more than 1500 organisms; it provides a set of reference data for computationally predicting metabolic pathways. Several important databases of metabolic pathways exist with hundreds of metabolic pathways and thousands of biochemical reactions; even the metabolic pathways for a small organism constitutes a large network, such as Roche Applied Science Biochemical Pathways chart [20] and BRENDA[21].

1.2 Motivation

Plant secondary metabolites contain many useful compounds for human in plant organism. They have very important economic and scientific value. The extraction of compound from plant not only need amount of plant organism but also produce the final concentrate low. We think biosynthesis the efficient strain to produce the useful compound by microbe are very useful, so the analysis of the metabolic pathway cross specific is very important for this idea. Before that, the study used KEGG and Metacyc database to find the related pathways and enzymes. Besides, survey literature

to search interesting pathways. Although numerous servers can reconstruct metabolic pathways based on available metabolic maps, it is still inconvenient to reconstruct metabolic pathways from two interesting metabolites. For instance, determining the metabolic pathway from glucose to pyruvate from those databases is arduous unless the two metabolites are already known to be involved in the glycolysis pathway. Therefore, this study promotes a method to systems analysis the metabolic pathway. Then, so far the study of our interesting pathway, which is the resveratrol synthesis in microbe, is superficial. The genome scale of metabolic reconstruction is necessary.

1.3 Specific Aim

Figure 1.6 shows specific aim in this study.

1. Puzzle:

Biological researchers may encounter a similar problem, who are interested in two metabolites but the pathways are unknown of our knowledge. This study proposes a systematic method and model pathways to solve the problem.

2. Solving puzzle (systematic method):

(1). Using biochemical reaction matrix reconstructs a metabolic pathway from two interesting metabolites.

(2). The reconstruction pathways are presented visually by KEGG joint maps, which are building-up maps from KEGG pathway maps.

(3). Comparative analysis of all enzymes include in this reconstruction pathway, which can tell us whether the genes are exist in species or not.

(4). Providing detail information about genes, proteins, and post-translational modifications (PTMs).

3. Modeling:

In order to find out the pathways in detail, we can use a biosynthesis pathway as an example to our systematic methods analysis and model the pathway to understand their yield and efficiency.

4. Applications: If this study is successful in the above steps, researcher could apply it to synthetic metabolites from microbes or others organisms.

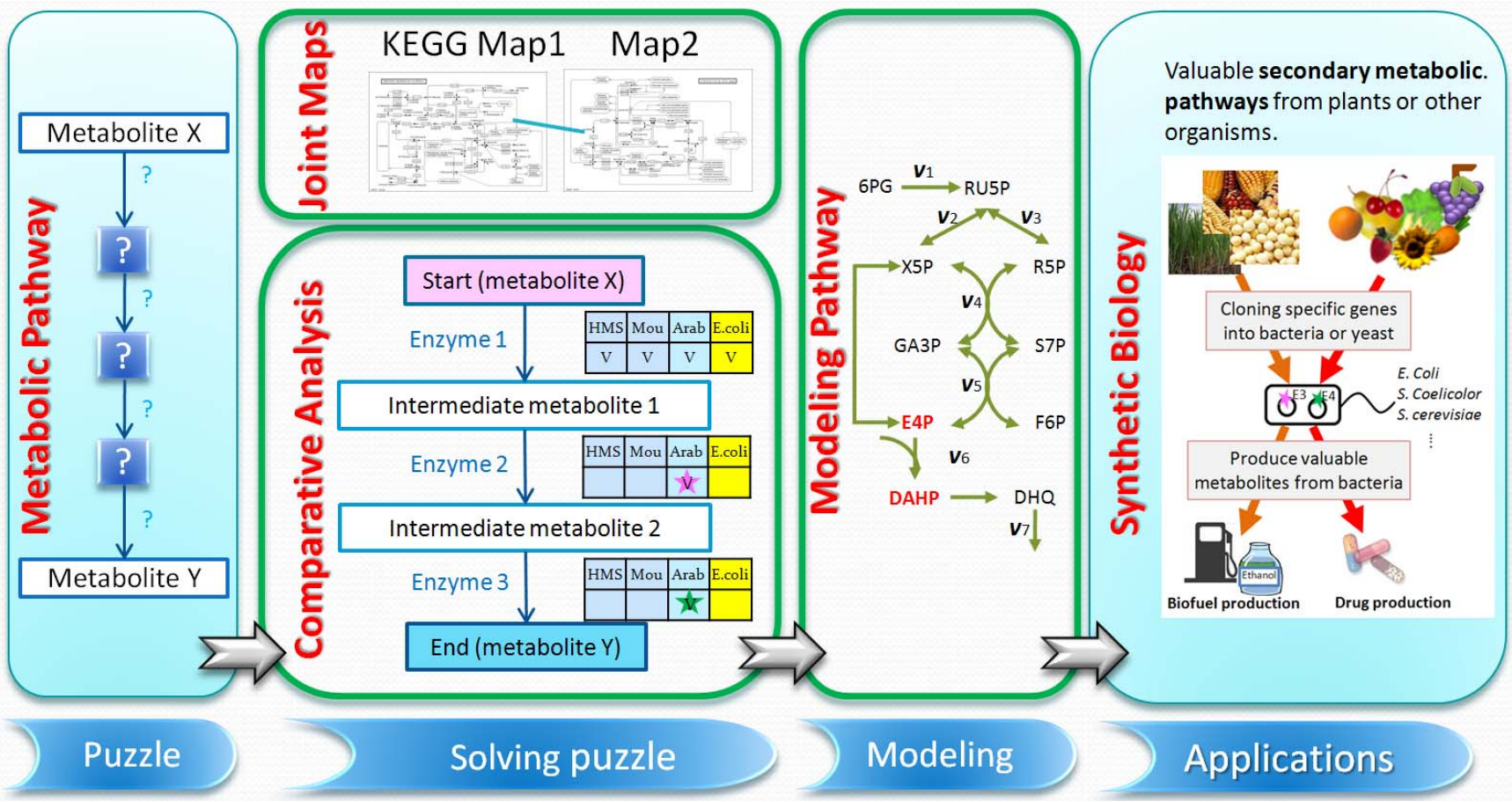


Figure 1.6 A conceptual diagram of goal

Four parts from left to right flow chart is our goal for synthetic biology.

Chapter2 Related Works

2.1 Related study of systematic method

2.1.1 *Kyoto Encyclopedia of Genes and Genomes* (KEGG) database

Kyoto Encyclopedia of Genes and Genomes (KEGG) is a useful database for biologists. It integrates genomic, chemical and systemic functional information. The databases contain three parts of category, which is systems information, genomic information, and chemical information (see Table 2.1).

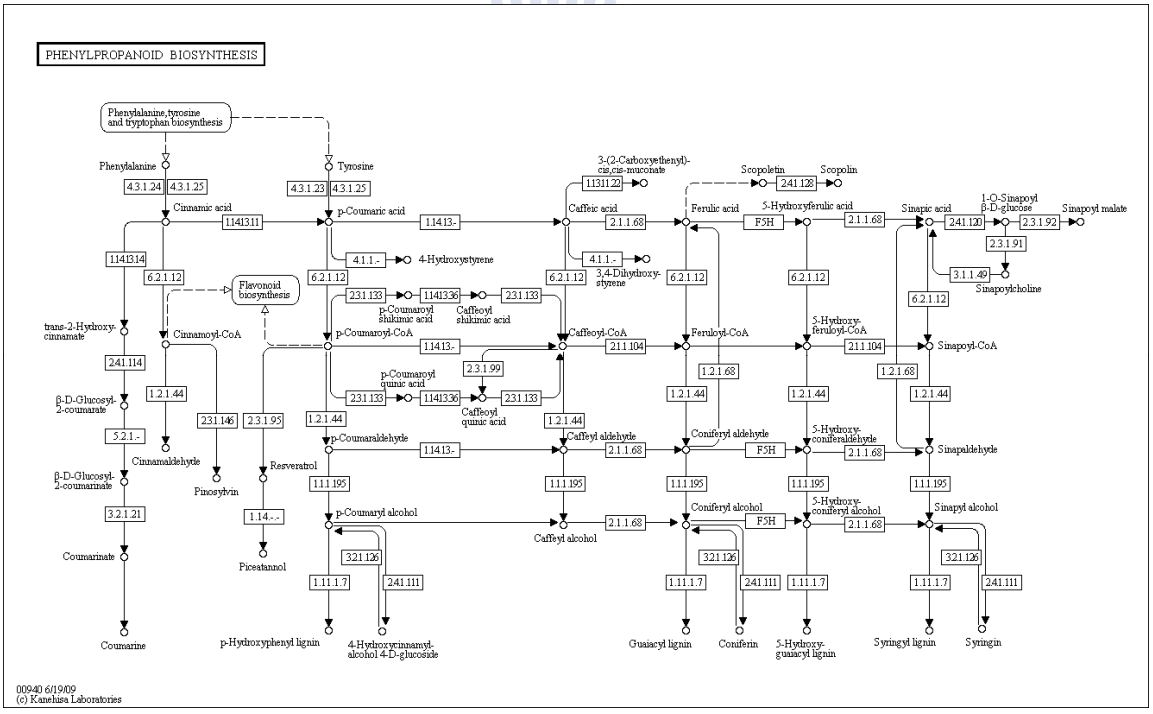
Table 2.1 KEGG databases (Kanehisa, M., et al. 2008)

Category	Database	Content
Systems information	KEGG PATHWAY	Pathway maps
	KEGG BRITE	Functional hierarchies
	KEGG MODULE	Pathway modules (released January 2008)
	KEGG DISEASE	Diseases (released January 2008)
Genomic information	KEGG ORTHOLOGY	KEGG orthology (KO) groups
(GENES)	KEGG GENOME	KEGG organisms
	KEGG GENES	Genes in high-quality genomes
	KEGG DGENES	Genes in draft genomes
	KEGG EGENES	Genes as EST contigs
	KEGG VGENOME	Viral genomes (to be fully integrated)
	KEGG VGENES	Genes in viral genomes (to be fully integrated)
	KEGG OGENES	Genes in organelle genomes (to be fully integrated)
	KEGG SSDB	Sequence similarities and best hit relations
Chemical information	KEGG COMPOUND	Metabolites and other chemical Compounds
(LIGAND)	KEGG DRUG	Drugs
	KEGG GLYCAN	Glycans
	KEGG ENZYME	Enzymes
	KEGG REACTION	Enzymatic reactions
	KEGG RPAIR	Reactant pairs and chemical transformations

2.1.2 Metacyc database

MetaCyc is a non-redundant, experimentally elucidated metabolic pathways database, which contain compounds, genes, protein and complexes, pathways, and reactions. MetaCyc emphasize that they contains only experimentally elucidated knowledge, it provides a uniquely high-quality resource for metabolic pathways and enzymes. KEGG database and Metacyc are powerful database for studying a metabolic pathway locally. For example, if researchers search resveratrol on these databases, they will output a map pathway of phenylpropanoid biosynthesis (see Figure 2.1). These two databases cannot search any two metabolites for global view all the pathways.

(A)



(B)

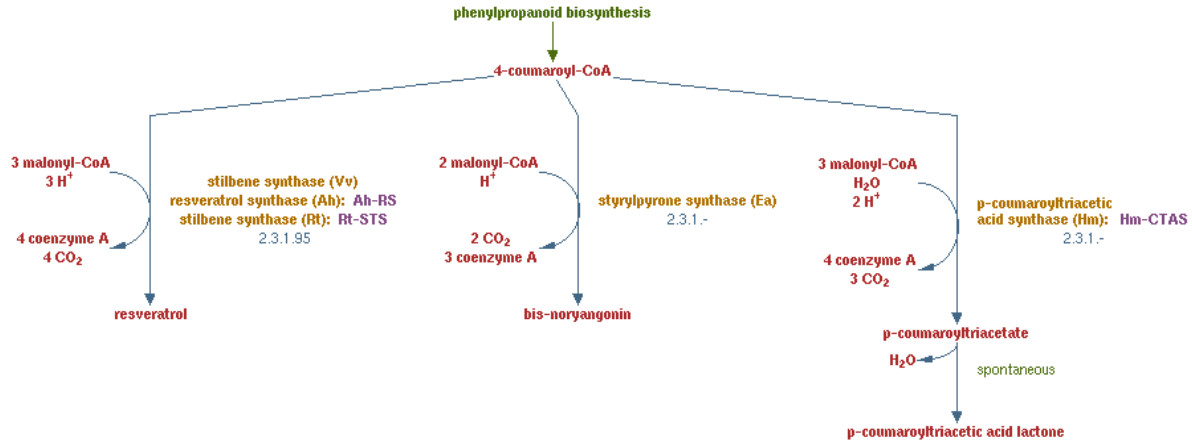


Figure 2.1 Compare with the phenylpropanoid biosynthesis of KEGG(A) and MetaCyc(B)

2.1.3 The Universal Protein Resource (UniProt)

The Universal Protein Resource (UniProt) contains protein sequence and annotation data, which is a comprehensive resource. The UniProt databases are composed of three parts of databae which is the UniProt Knowledgebase (UniProtKB), the UniProt Reference Clusters (UniRef), and the UniProt Archive (UniParc). The UniProt Metagenomic and Environmental Sequences (UniMES) database is a repository specifically developed for metagenomic and environmental data[22].

2.1.4 dbPTM database

Protein post translational modifications (PTMs) play a critical role in cellular control mechanism, such as phosphorylation for signal transduction, attachment of fatty acids for membrane anchoring and association. A database names dbPTM is collected and integrated this data. Furthermore, the databases have the function of computational annotation and prediction the PTM site [23] (see Figure 2.2).

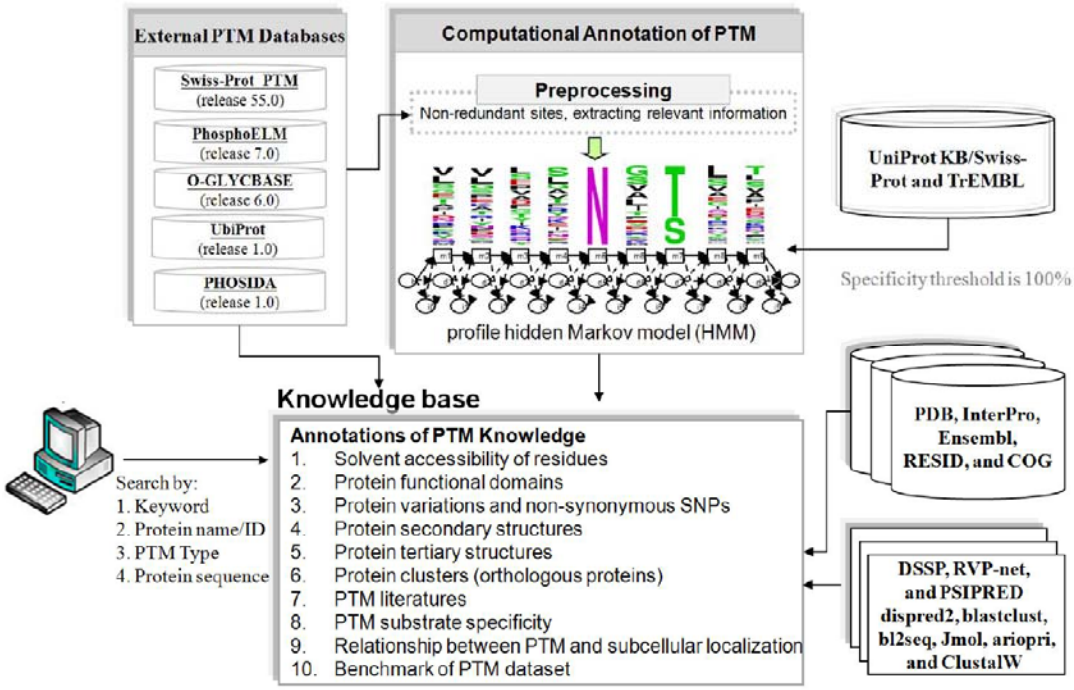


Figure 2.2 The main architecture of dbPTM
(Lee, T.Y., *et al.* 2009**)**

2.1.5 Other databases or tools

Li *et al*. adopted Metabolic Pathway Alignment and Scoring (M-PAS) (see Figure 2.3) for identifying and ranking conserved metabolic pathways based on a comprehensive and flexible similarity measuring method[21]. The Comparative Pathway Analyzer (CPA) calculates and displays the differences from the metabolic reaction contents of several organisms[24], but they only analyze one of the KEGG maps (see Figure 2.4). The *Roseobacter* systems biology database (ROSY) represents an integrated platform for studying the comparative genomics and systems biology of *Roseobacter*-related species[25]. However, even some of the resources for comparative analysis of metabolic pathways of various species, they focus on a few species only or do not provide a convenient interface for analysis.

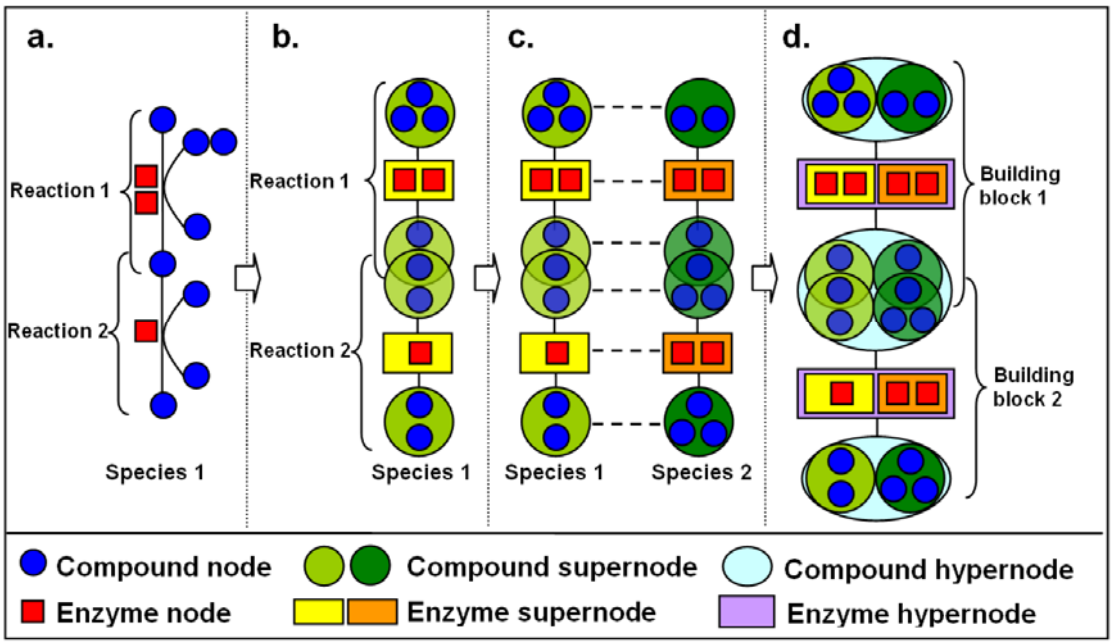


Figure 2.3 The schema of M-PAS methods for comparative analysis

(Li, Y., *et al.* 2008)

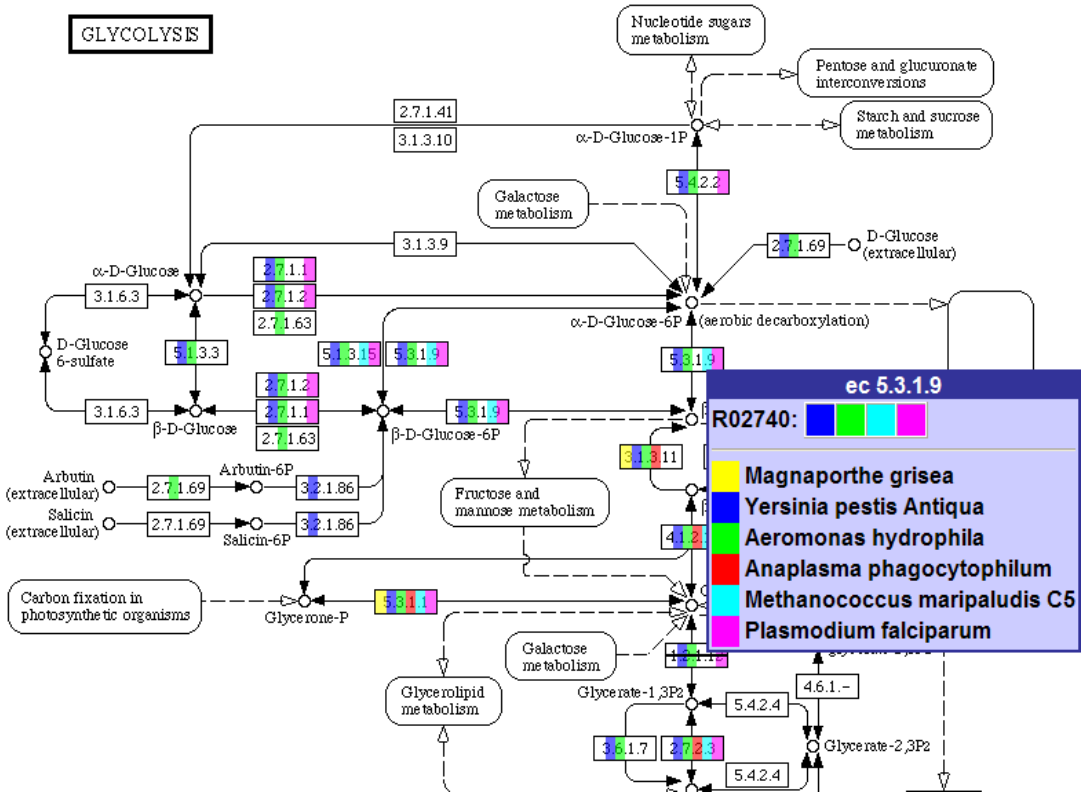


Figure 2.4 The CPA web tool display comparative analysis in glycolysis

(Oehm, S., *et al.* 2008)

2.2 Related study of modeling metabolic pathway

Katsuyama *et al.* successfully constructed an E.coli strain which can be produced resveratrol from glucose, but the yield of resveratrol is very low[26]. We think some powerful metabolic pathway tools can model this pathway. For example, flux balance analysis based on convex analysis by imposing an objective function to determine the metabolic flux vector. Several constraints such as uptake rates of substrate and/or secretion rates of product, thermodynamic constraints, metabolic regulation, and so on [27, 28] narrow down the range of solution (see Figure 2.5). The methods need a lot of experiment data, but the biosynthesis of plant secondary metabolites, which yield is very low to measure related data.

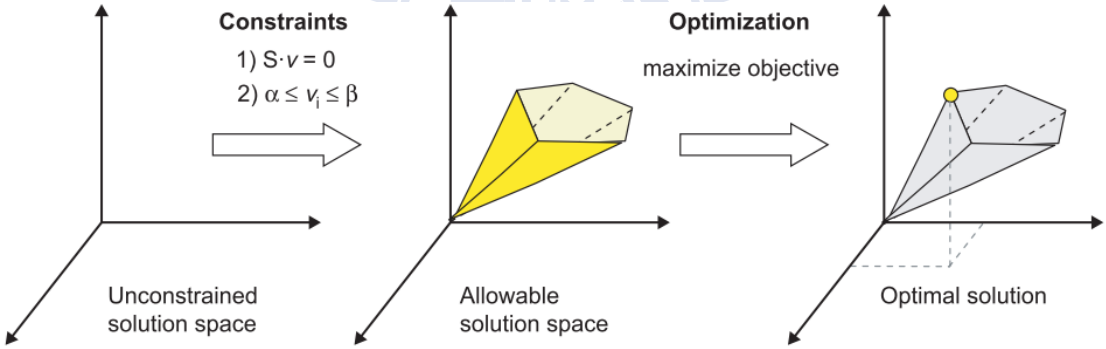


Figure 2.5 The concept of constraint-based modeling
(Reed, J.L. *et al.* 2003)
A three-dimensional flux space is given by metabolic network. The constraint of stoichiometric, thermodynamic and enzyme capacity was definite. Then, the possible solutions are confined to a region in the total flux space, termed the allowable solution space. Finally, linear optimization can be applied to identify a solution in the allowable solution space that maximizes or minimizes a defined objective, such as ATP, biomass, or end product production.

Chapter3 Materials and Methods

3.1 Systematic analysis of metabolic pathway

This study proposes a systematic flow to solve the problem from one metabolite to another amount different species. In part of *solving puzzle* (see Figure 1.6) displays the basic concept of systematic methods which demonstrates that it has numerous applications in synthetic biology and metabolic engineering.

Figure 3.1 presents the system flow of these systematic methods. First of all, the data were collected and integrated by three main databases, which are KEGG pathway and ligand database, UniProt Knowledgebase (UniProtKB), and dbPTM. Second, the map data, reaction data, and enzyme data are combined. Third, the reaction matrix is constructed by biochemical reaction. Fourth, the pathway is searched by KEGG map data and takes advantage of breadth first search (BFS) and KEGG module to search and select better pathways. Finally, the results are outputted by two form, one is the vertical map contain comparative analysis and another is KEGG joint map.

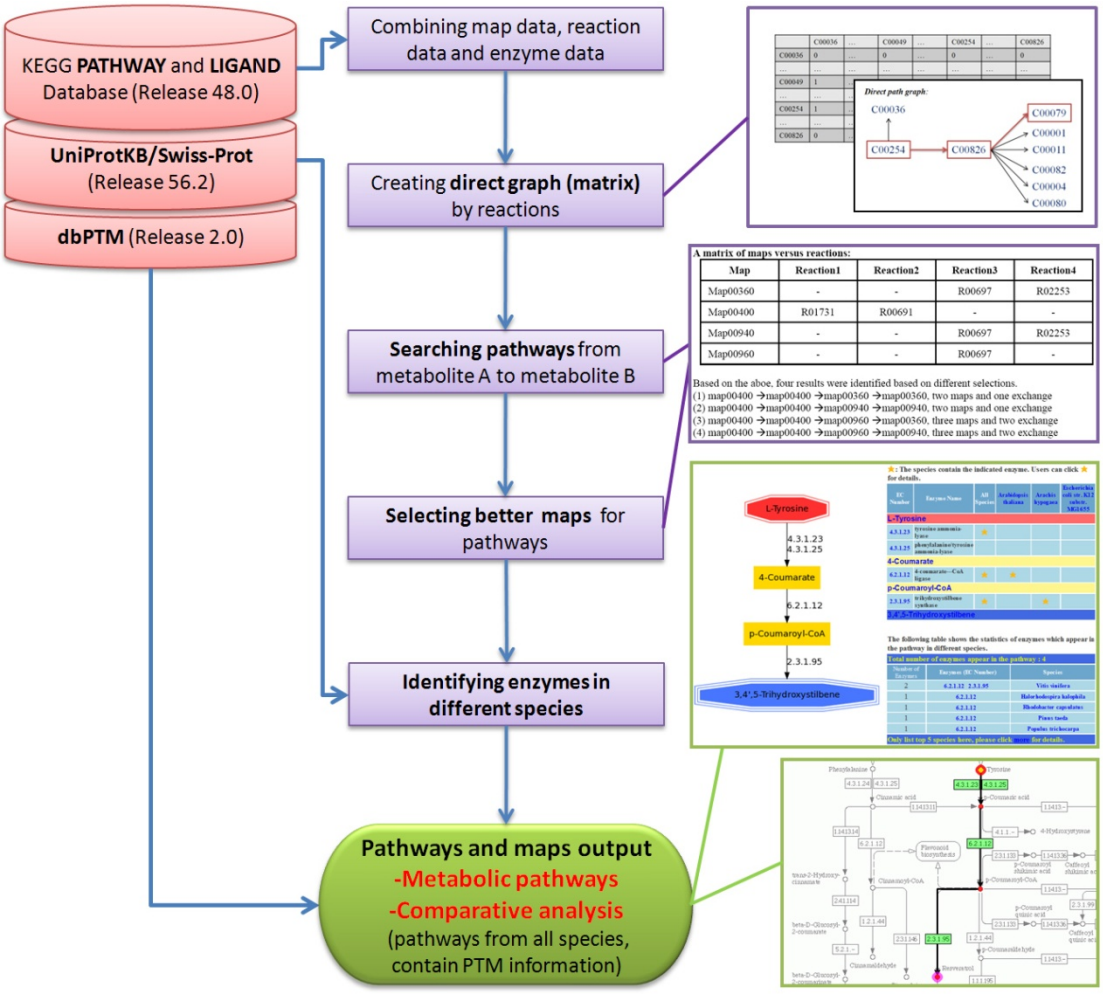


Figure 3.1 System flow of systematic analysis

3.1.1 Data collection and integration

Reaction definitions, species-specific reactions, reaction maps and enzyme lists were obtained from KEGG/LIGAND and KEGG/PATHWAY database Release 48.0, updated on Oct 1, 2008. Some information (including gene names, Enzyme Commission (EC) numbers, and species-specific enzymes) was supplied from UniProtKB/Swiss-Prot Release 14.0 and NCBI Taxonomy database. The current metabolites (e.g. ATP, NADH, H_2O, and CO_2 are lists in Appendix 1) are generally used as carriers for transferring electrons or certain functional groups (including phosphate group, amino group, one carbon unit, and methyl group); they are like the external metabolites that participate in many reactions and are not in pseudo steady

state in a sub-network. The connections through current metabolites should be avoided in calculating the path length from one metabolite to another. Hence, the connections through current metabolites and cofactors were deleted manually to make the path length analysis physiologically more meaningful. However, some current metabolites are primary metabolites. For example, asparate is not only a current metabolite involved in transamination mechanism (oxaloacetate +NH_3+NADPH = asparate +NADP+H_2O), but also a primary metabolite in biosynthesis of amino acids. When the current metabolites involved in the primary pathway as primary metabolites, such reactions and metabolites are extra collected in this systematic method. Therefore, biosynthesis of amino acids can be conducted in this systematic method. It is important to know the PTM information before cloning one gene from eukaryote to prokaryote. dbPTM [29] is a database that compiles information on protein post-translational modifications (PTMs), such as the catalytic sites, solvent accessibility of amino acid residues, protein secondary and tertiary structures, protein domains and protein variations. Protein post-translational modification (PTM) information was extracted from dbPTM.

3.1.2 Construction biochemical reaction matrix

Information on reactions and enzymes was obtained from KEGG pathway maps and the equations of each reaction were determined. Therefore, a reaction matrix was constructed based on maps, reactions and enzymes data. For instance, “2-oxoglutarate aminotransferase: Oxaloacetate + L-Arogenate <=> L-Aspartate + Prephenate” is converted into “R01731: C00036 + C00826 <=> C00049 + C00254” (where R01731 is KEGG reaction ID; C00036, C00826, C00049, and C00254 are KEGG compound ID). The direction of reaction R01731 is from C00254 to C00036 in the pathway map. Therefore, four reactions are described below:

C00254 => C00826

C00254 => C00036

C00049 => C00826

C00049 => C00036

Each equation corresponds to an edge in a graph; Table 3.1 depicts an example reaction matrix. The systematic method contains 16,884 compounds; and so the size of the reaction matrix is 16884*16884.

Example:

Reaction: R01731: L-Aspartate + Prephenate => Oxaloacetate + L-Arogenate

(C00049 + C00254 => C00036 + C00826)

Equation: C00254 => C00826

C00254 => C00036

C00049 => C00826

C00049 => C00036

Table 3.1 Construction of reaction matrix

	C00036	**…**	**C00049**	**…**	**C00254**	**…**	**C00826**
C00036	0	…	0	…	0	…	0
…	…	…	…	…	…	…	…
C00049	1	…	0	…	0	…	1
…	…	…	…	…	…	…	…
C00254	1	…	0	…	0	…	1
…	…	…	…	…	…	…	…
C00826	0	…	0	…	0	…	0

3.1.3 Searching pathway

3.1.3.1 Breadth first search (BFS)

Breadth first search is a graph search algorithm, begins at the root node and explores all the neighboring nodes. It can be applied to search our biochemical reaction matrix. Their algorithm was demonstrated in Figure 3.2. The brief procedure of BFS, contain four major steps, is listed as bellow:

1. Put the source node on the queue.

2. Pull a node from the beginning of the queue and examine it.

(1) If the searched element is found in this node, quit the search and return a result.

(2) Otherwise push all the (so-far-unexamined) successors (the direct child nodes) of this node into the end of queue, if they are any.

3. If the queue is empty, every node on the graph has been examined – quit the search and return “not found”.

4. Repeat from Step2.

It procedure assumes that the input graph $G = (V, E)$ is represented using adjacency lists. The pseudocode of BFS is implemented in C programming language shown in Figure 3.3. How we apply this algorithm to search our pathways is shown in Figure 3.4. When the direct path graph from compound C00254 to compound C00079 is identified by BFS, the search terminates when the C00079 node was found. It can search all possible pathways from our reaction matrix.

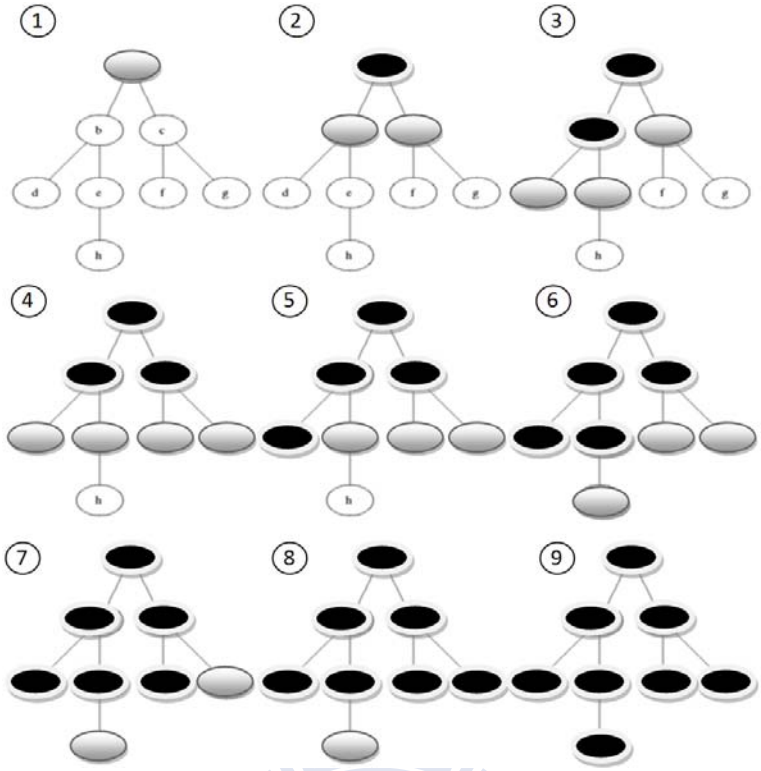


Figure 3.2 An example of breadth first search

A network contains eight metabolites. Breadth first search algorithm will search by night steps.

```
void BFS(VLink G[ ], int v) {
      int w;                              /*visit vertex v*/
      visited[v] = 1;                     /*mark v as visited : 1 */
      ADDQ(Q, v);
      while(!EMPTYQ(Q)) {
            v = DELQ(Q);                  /*Dequeue v*/
            w = FIRSTADJ(G, v);           /*Find first neighbor, return -1 if no neighbor*/
            while(w != -1) {
                  if(visited[w] == 0) {
                        VISIT(w);         /*visit vertex w*/
                        ADDQ(Q, w);       /*Enqueue current visited vertex w*/
                        visited[w] = 1;   /*mark w as visited*/
                  }
                  w = NEXTADJ(G, v); /*Find next neighbor, return -1 if no neighbor*/
            }
      }
}
```

Figure 3.3 Pseudocode of breadth-first search (BFS) algorithm

Question: C00254 $\xrightarrow{?}$ C00079

Reactions

(1) C00049 + C00254 => C00036 + C00826

(2) C00826 => C00079 + C00001 + C00011

(3) C00826 + C00003 => C00082 + C00011 + C00004 + C00080

Search steps by reaction matrix:

Step1	**C00254**						
Step2	C00254	C00036	C00826				
Step3	**C00036**	**C00826**					
Step4	C00826	C00079	C00001	C00082	C00011	C00004	C00080
Step5	**C00079**	**C00001**	**C00082**	**C00011**	**C00004**	**C00080**	

Direct path graph:

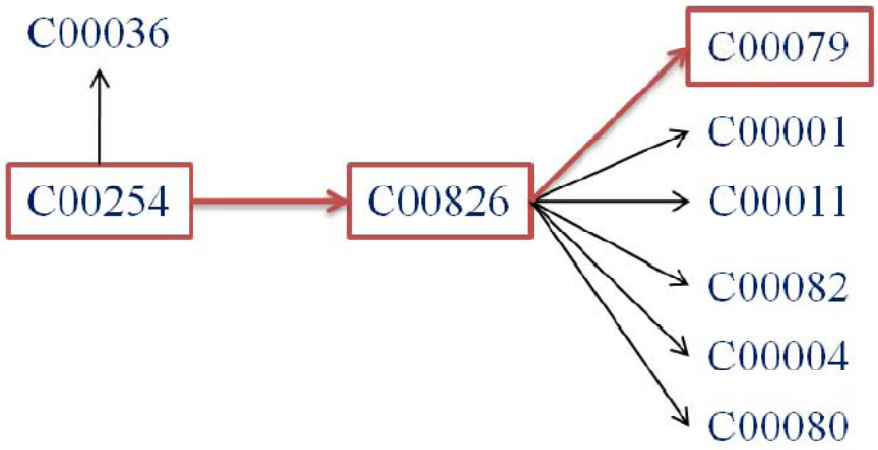


Figure 3.4 Construct of direct path graph

3.1.3.2 Implementing and modifying KEGG module

We modify KEGG MODULE by adding or deleting some metabolic pathways. KEGG MODULE is a new collection of pathway modules, molecular complexes, and other functional units, each represented as a list of KEGG Orthology (KO) identifiers, which were used in pathway reconstruction when the node we searched from BFS was a KEGG MODULE's metabolites. Table 3.2 shows the module, which contain the order of metabolite.

Table 3.2 Pathway module

Module	Order of metabolites*
M00001	C00668;C05345;C05378;C00111;C00118;C00236;C00197;C00631;C00074;C00022
M00001	C00668;C05345;C05378;C00118;C00236;C00197;C00631;C00074;C00022
M00003	C00036;C00074;C00631;C00197;C00236;C00118;C00111;C05378;C05345
M00003	C00036;C00074;C00631;C00197;C00236;C00118;C05378;C05345
M00004	C01172;C01236;C00345;C00199;C00231;C05382;C00279;C05345
⋮	⋮
M00689	C00024;C00158;C00417;C00311;C00026;C00091;C00042

*The symbol of metabolites is KEGG compound ID (e.g. C00668 is glucose 6-phosphate).

3.1.4 Construction metabolic pathway map

3.1.4.1 Construction vertical map by Graph Visualization Software

The reconstruct pathway can directly be represented by Graph Visualization Software (*Graphviz*). *Graphviz* is a way of representing structural information as diagrams of abstract graphs and networks, which are open source graph visualization software; you can download them from http://www.graphviz.org/Download.php.

3.1.4.2 Construction KEGG joint map

KEGG pathways are fixed in KEGG maps. Many pathways may go across maps, so we joint it for vision. Our search pathways are not directly implemented to joint maps. Usually, found paths occurred not only in a single pathway map, but in a complicated fashion in several maps. Pathway maps that contain the most paths are selected and the one pathway map that has only one reaction is avoided. A matrix of maps versus reactions was employed to reconstruct metabolic pathway from different KEGG maps. We give an example to reconstruction of metabolic pathway from KEGG compound ID C00254 to C00811.

Pathway: C00254 => C00826 => C00079 => C00423 => C00811

Maps, reactions and enzymes involved in the pathway:

Path C00254 to C00826 involves in map00400 which contains reaction R01731 on it. One enzyme with EC number 2.6.1.57 is needed in reaction R01731.

Path C00254 to C00826 → *map00400:R01731:2.6.1.57*

Path C00826 to C00079 involves in map00400 which contains reaction R00691 on it. Two enzymes with EC number 4.2.1.51 and 1:4.2.1.91 are needed in reaction R00691.

Path C00826 to C00079 → *map00400:R00691:4.2.1.51, map00400:R00691:4.2.1.91*

Path C00079 to C00423 involves in 3 maps (map00360, map00940 and map00960) which contains reaction R00697 on them. Two enzymes with EC number 4.3.1.24 and 4.3.1.25 are needed in reaction R00697.

Path C00079 to C00423 → *map00360:R00697:4.3.1.24, map00360:R00697:4.3.1.25, map00940:R00697:4.3.1.24, map00940:R00697:4.3.1.25, map00960:R00697:4.3.1.24, map00960:R00697:4.3.1.25*

Path C00423 to C00811 involves in 2 maps (map00360 and map00940) which contains reaction R02253 on them. One enzyme with EC number 1.14.13.11 is needed in reaction R02253.

Path C00423 to C00811 →

map00360:R02253:1.14.13.11 ,map00940:R02253:1.14.13.11

Table 3.3 A matrix of maps versus reactions

Map	Reaction1	Reaction2	Reaction3	Reaction4
Map00360	-	-	R00697	R02253
Map00400	R01731	R00691	-	-
Map00940	-	-	R00697	R02253
Map00960	-	-	R00697	-

Based on the aboe, four results were identified based on different selections.

(1) map00400 →map00400 →map00360 →map00360, two maps and one exchange

(2) map00400 →map00400 →map00940 →map00940, two maps and one exchange

(3) map00400 →map00400 →map00960 →map00360, three maps and two exchange

(4) map00400 →map00400 →map00960 →map00940, three maps and two exchange

It is better to select less maps and less exchange times to reconstruct metabolic pathway. Therefore, based on the above example, "map00400 →map00400 →map00360 →map00360" is much better than "map00400 →map00400 →map00960 → map00360".

3.2 Metabolic pathway analysis by computational modeling

3.2.1 Construction a metabolic network

We used our systematic method for pathway reconstruct and then integrated KEGG, Biocyc, BiGG database to find detail information about reaction reversible or irrevsible, current metabolite contents, and enzyme name. The biomass formation data is referred Carlson R., 2004 [30] in doubling time equal to 200 minutes in *E.coli*.

3.2.2 Elementary flux modes analysis by *metatool*

Elementary flux modes analysis (EMS) is a useful metabolic pathway analysis tool to

identify the structure of a metabolic network. The concept of EMS provides a mathematical tool to define and comprehensively describe all metabolic routes that are both stoichiometrically and thermodynamically feasible for a group of enzymes[31]. These pathways consist of a minimal set of enzymes that can support steady state operation. We demonstrate the process in Figure 3.5. Metabolism could be characterized by EMS. EMS could application in strain improvement of metabolic engineering, such as identifying the key branch points of reaction, construct an efficient strain in minimal cell condition [32]. Metatool [33], which is one of the first programs dedicated to constraint-based modeling of metabolic networks, has been used in computing the nullspace matrix, elementary modes and other biochemical reaction networks. The newly metatool version is 5.0, which performance has been significantly increased and the environments of computing is the GNU octave or MATLAB[33].

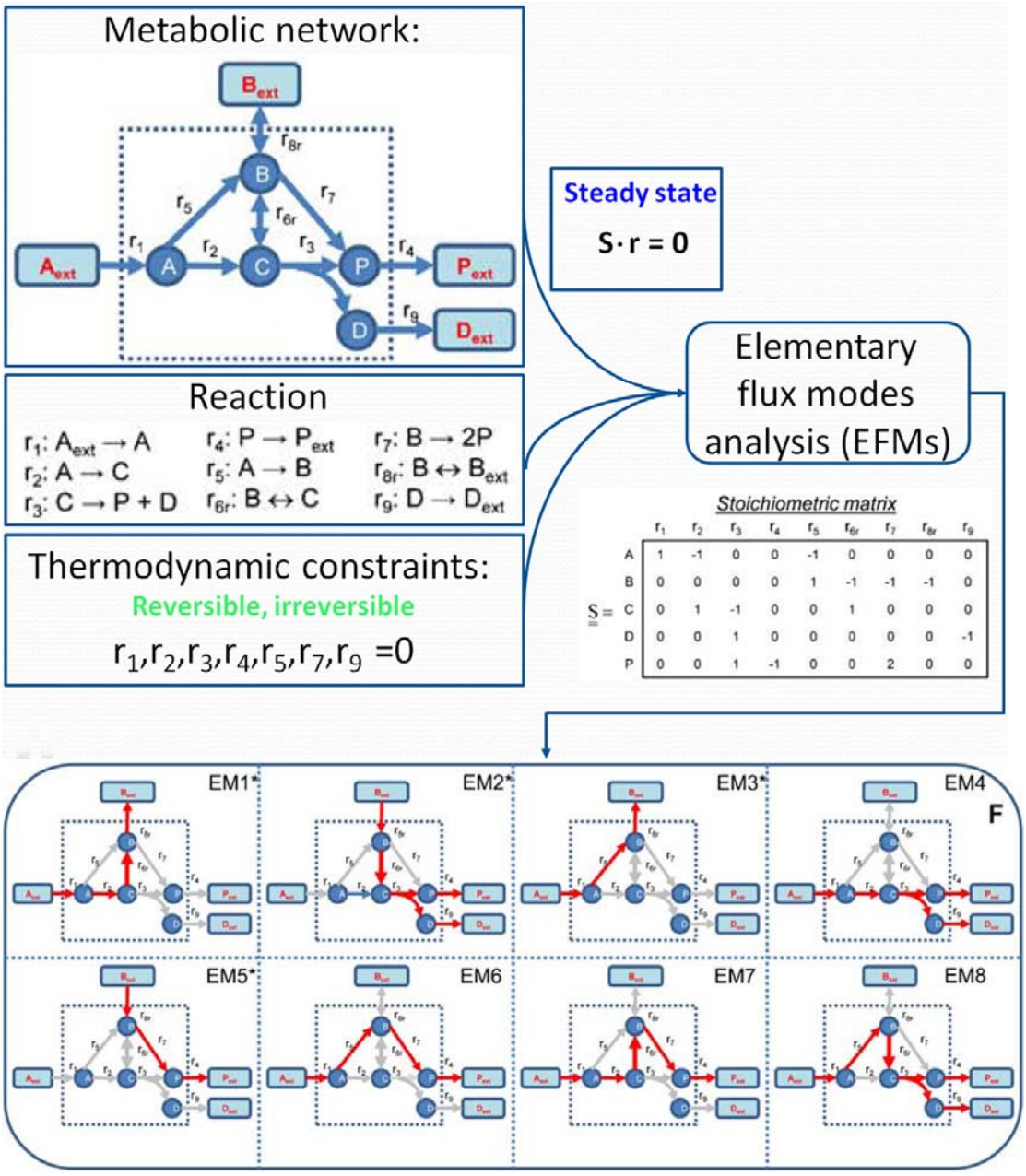


Figure 3.5 Elementary mode analysis

(Trinh, C.T. *et al.* 2009)

Construct a metabolic network system and definite internal and external metabolites. Input the reaction and a thermodynamic constraint data was counted in steady state. Eight possible of elementary mode will demonstrate and relative data can be analyzed further.

3.2.3 Calculation and analysis biomass and end product yield

The entire metabolic network can be calculated elementary flux modes by *metatool*. Therefore, we analyze all the elementary flux modes to understand production history. Biomass yield is calculated by carbon content. According to Carlson R., *et al.* experiment, when cell doubling time is 200 minutes the number of carbon are 2652. The uptake carbon source is glucose or glycerol. There number of carbon is six and three respectively. End product yield is calculated by molecular weight.

Chapter4 Results

4.1 The result of systematic method

4.1.1 The statistic of systematic methods

In our systematic methods, we collected and integrated data from different database. The data include species, reactions, enzymes, maps, modules, gene information, protein information, and PTMs are integrated to our database and their integration is statistic in Table 4.1.

Table 4.1 Statistic of systematic methods

Name	Number
Metabolites	16884
Species	1491
Reactions	4288
Enzymes	2336
Maps	149
Modules	689
Gene Information	667990
Protein Information	183462
PTMs	418386

4.1.2 Successfully reconstruct pathway between two metabolites

In materials and methods section, we introduce two approaches to search metabolic pathway from two metabolites. The result in Figure 4.1 (A) shows we can reconstruct the metabolic pathway clearly. All the blocks like red box region are one of the possible pathways. The pathway, which is searched by our module is displayed forward blocks. The orange box show all intermediates in the reconstruct pathway.

4.1.3 Creating KEGG joint map

KEGG maps tell us single pathway clearly. From their maps, we could know the metabolic network, which one metabolite is catalyzed by which enzyme. Here we connect and highlight the metabolic pathway in the KEGG map shows in Figure 4.1 (B). The result shows we provide a go across maps, global view, and highlight all the intermediate maps. The KEGG joint maps not only can rapidly find pathway but also can global view all the map.

4.1.4 Comparative analysis

Comparative analysis is useful in synthetic biology. For example, the reconstruction of metabolic pathways to produce valuable metabolites or secondary metabolites in bacteria or yeast is a promising biotechnological strategy. Comparative analysis provides an easy way to elucidate whose genes from which species should be cloned into those microorganisms. Firstly, the enzymes identified in the reconstructed pathway were processed to search for orthologous encoding genes from various species. Then, the presence or absence of the pathway in a particular species can be known. Furthermore, comparative tables of several organisms are provided. The species are ranked according to the numbers of enzymes involved in the reconstructed pathway. The results are presented in Figure 4.1 (C)-(F). Figure 4.1 (C) is a local view the metabolic pathway draw by *Graphviz.* Comparative is revealed in Figure 4.1 (D), in which the yellow star-shaped symbols indicated that the enzyme exist in this species. Figure 4.1 (E) is the statistics of enzymes which appear in the pathway in different species. A functional enzyme that needed PTM in eukaryote was no function when cloning into prokaryote. It is important to know the PTM information before cloning one gene from eukaryote to prokaryote. The result shows in Figure 4.1 (F). If an enzyme has PTMs, we can find out and know the enzyme if it’s appropriate for synthetic biology.

4.1.5 Glycolysis application in these systematic method

Our systematic method can be reconstruct the common pathway by our building module. Some important and common pathway such as glycolysis, pentose phosphate pathway, and citrate acid cycle are constructed by our building module. We use glycolysis as an example to demonstrate. Figure 4.2 shows all result of glycolysis analysis.

(A)

From Metabolite	To Metabolite	Pathway Size	Detailed Pathway
Phosphoenolpyruvate	3,4',5-Trihydroxystilbene	14 Metabolites	Vertical Map KEGG Map

KEGG Pathway Maps : Phenylalanine, tyrosine and tryptophan biosynthesis→Phenylpropanoid biosynthesis

Phosphoenolpyruvate –> 2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate –> 3-Dehydroquinate –> 3-Dehydroshikimate –> Shikimate –> Shikimate 3-phosphate –> 5-O-(1-Carboxyvinyl)-3-phosphoshikimate –> Chorismate –> Prephenate –> L-Arogenate –> L-Tyrosine –> 4-Coumarate –> p-Coumaroyl-CoA –> 3,4',5-Trihydroxystilbene

From Metabolite	To Metabolite	Pathway Size	Detailed Pathway
Phosphoenolpyruvate	3,4',5-Trihydroxystilbene	15 Metabolites	Vertical Map KEGG Map

KEGG Pathway Maps : Phenylalanine, tyrosine and tryptophan biosynthesis→Phenylpropanoid biosynthesis

Phosphoenolpyruvate –> 2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate –> 3-Dehydroquinate –> 3-Dehydroshikimate –> Shikimate –> Shikimate 3-phosphate –> 5-O-(1-Carboxyvinyl)-3-phosphoshikimate –> Chorismate –> Prephenate –> Phenylpyruvate –> L-Phenylalanine –> L-Tyrosine –> 4-Coumarate –> p-Coumaroyl-CoA –> 3,4',5-Trihydroxystilbene

From Metabolite	To Metabolite	Pathway Size	Detailed Pathway
Phosphoenolpyruvate	3,4',5-Trihydroxystilbene	14 Metabolites	Vertical Map KEGG Map

KEGG Pathway Maps : Phenylalanine, tyrosine and tryptophan biosynthesis→Phenylpropanoid biosynthesis

Phosphoenolpyruvate –> 2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate –> 3-Dehydroquinate –> 3-Dehydroshikimate –> Shikimate –> Shikimate 3-phosphate –> 5-O-(1-Carboxyvinyl)-3-phosphoshikimate –> Chorismate –> Prephenate –> 3-(4-Hydroxyphenyl)pyruvate –> L-Tyrosine –> 4-Coumarate –> p-Coumaroyl-CoA –> 3,4',5-Trihydroxystilbene

From Metabolite	To Metabolite	Pathway Size	Detailed Pathway
Phosphoenolpyruvate	3,4',5-Trihydroxystilbene	15 Metabolites	Vertical Map KEGG Map

KEGG Pathway Maps : Phenylalanine, tyrosine and tryptophan biosynthesis→Phenylpropanoid biosynthesis

Phosphoenolpyruvate –> 2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate –> 3-Dehydroquinate –> 3-Dehydroshikimate –> Shikimate –> Shikimate 3-phosphate –> 5-O-(1-Carboxyvinyl)-3-phosphoshikimate –> Chorismate –> Prephenate –> Phenylpyruvate –> L-Phenylalanine –> trans-Cinnamate –> 4-Coumarate –> p-Coumaroyl-CoA –> 3,4',5-Trihydroxystilbene

From Metabolite	To Metabolite	Pathway Size	Detailed Pathway
Phosphoenolpyruvate	3,4',5-Trihydroxystilbene	15 Metabolites	Vertical Map KEGG Map

KEGG Pathway Maps : Phenylalanine, tyrosine and tryptophan biosynthesis→Phenylpropanoid biosynthesis

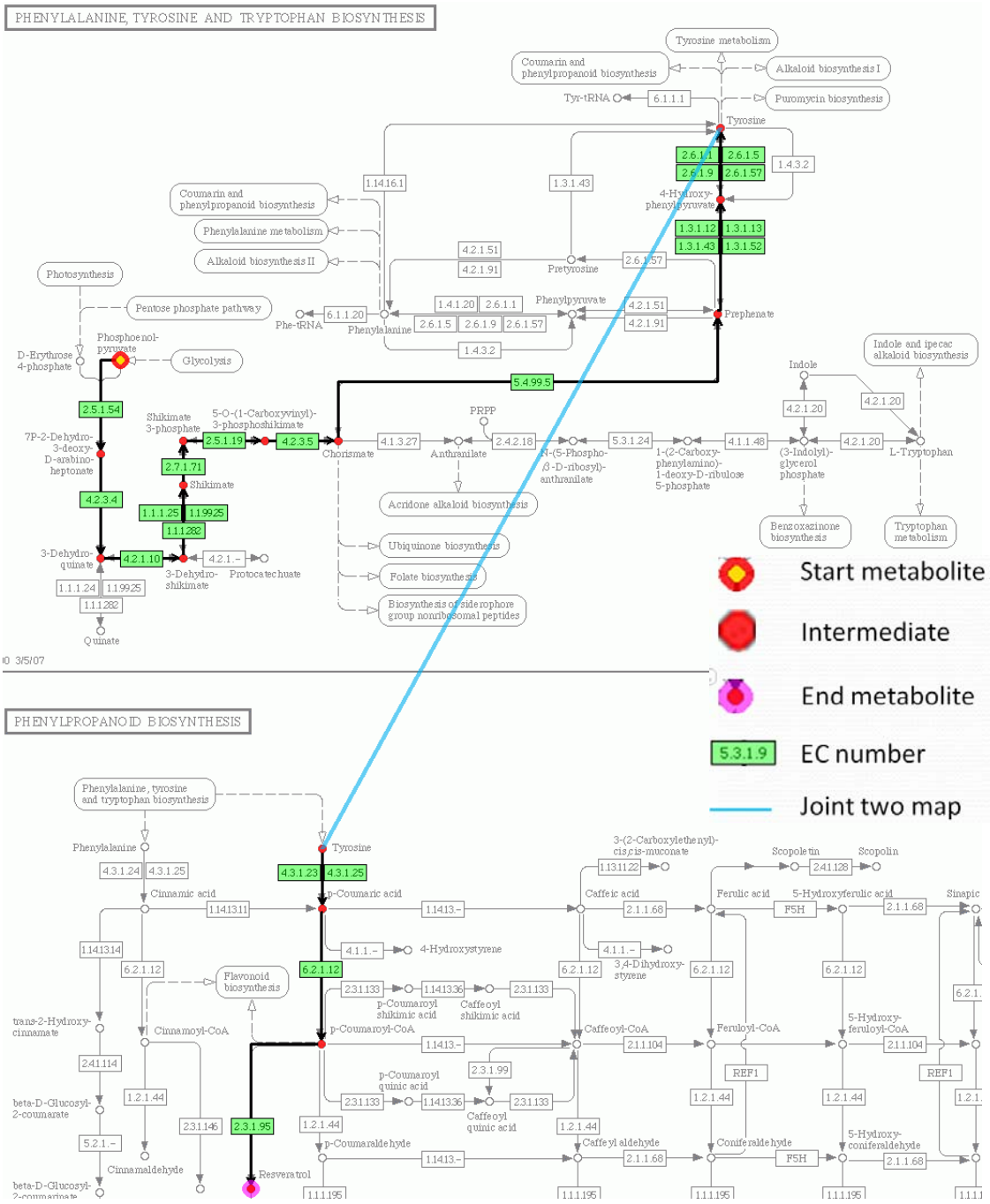
Phosphoenolpyruvate –> 2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate –> 3-Dehydroquinate –> 3-Dehydroshikimate –> Shikimate –> Shikimate 3-phosphate –> 5-O-(1-Carboxyvinyl)-3-phosphoshikimate –> Chorismate –> Prephenate –> L-Arogenate –> L-Phenylalanine –> L-Tyrosine –> 4-Coumarate –> p-Coumaroyl-CoA –> 3,4',5-Trihydroxystilbene

From Metabolite	To Metabolite	Pathway Size	Detailed Pathway
Phosphoenolpyruvate	3,4',5-Trihydroxystilbene	15 Metabolites	Vertical Map KEGG Map

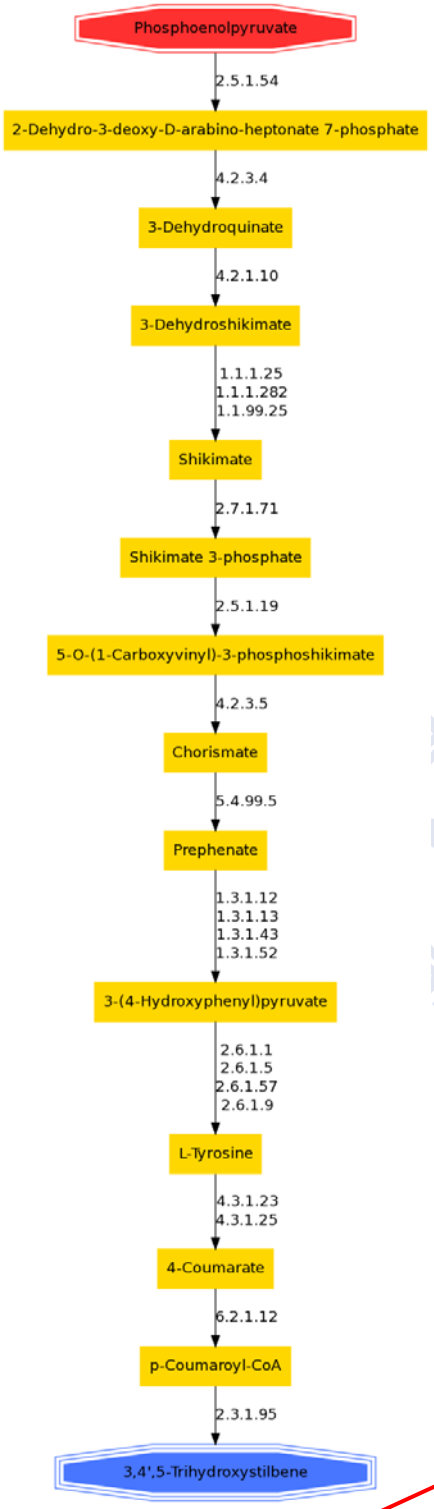
KEGG Pathway Maps : Phenylalanine, tyrosine and tryptophan biosynthesis→Phenylpropanoid biosynthesis

Phosphoenolpyruvate –> 2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate –> 3-Dehydroquinate –> 3-Dehydroshikimate –> Shikimate –> Shikimate 3-phosphate –> 5-O-(1-Carboxyvinyl)-3-phosphoshikimate –> Chorismate –> Prephenate –> L-Arogenate –> L-Phenylalanine –> trans-Cinnamate –> 4-Coumarate –> p-Coumaroyl-CoA –> 3,4',5-Trihydroxystilbene

(B)


PHENYLALANINE, TYROSINE AND TRYPTOPHAN BIOSYNTHESIS
Tyrosine metabolism
Coumarin and phenylpropanoid biosynthesis
Alkaloid biosynthesis I
Puromycin biosynthesis
Tyr-tRNA
Tyrosine
Phenylalanine metabolism
Alkaloid biosynthesis II
Photosynthesis
Pentose phosphate pathway
Glycolysis
Phe-tRNA
Phenylalanine
Phenylpyruvate
Prephenate
Pretyrosine
4-Hydroxyphenylpyruvate
Phosphoenolpyruvate
D-Erythrose 4-phosphate
Shikimate 3-phosphate
5-O-(1-Carboxyvinyl)-3-phosphoshikimate
Chorismate
Anthranilate
PRPP
N-(5-Phospho-β-D-ribosyl)-anthranilate
1-(2-Carboxyphenylamino)-1-deoxy-D-ribulose 5-phosphate
(3-Indolyl)-glycerol phosphate
Indole
L-Tryptophan
Indole and ipecac alkaloid biosynthesis
7P-2-Dehydro-3-deoxy-D-arabino-heptonate
Shikimate
3-Dehydroquinate
3-Dehydroshikimate
Protocatechuate
Quinate
Acridone alkaloid biosynthesis
Ubiquinone biosynthesis
Folate biosynthesis
Biosynthesis of siderophore group nonribosomal peptides
Benzoxazinone biosynthesis
Tryptophan metabolism
Start metabolite
Intermediate
End metabolite
EC number
Joint two map
PHENYLPROPANOID BIOSYNTHESIS
Phenylalanine, tyrosine and tryptophan biosynthesis
Cinnamic acid
p-Coumaric acid
Caffeic acid
Ferulic acid
5-Hydroxyferulic acid
Sinapic
4-Hydroxystyrene
3,4-Dihydroxystyrene
Scopoletin
Scopolin
Flavonoid biosynthesis
p-Coumaroyl shikimic acid
Caffeoyl shikimic acid
p-Coumaroyl quinic acid
Caffeoyl quinic acid
Cinnamoyl-CoA
p-Coumaroyl-CoA
Caffeoyl-CoA
Feruloyl-CoA
5-Hydroxy-feruloyl-CoA
trans-2-Hydroxy-cinnamate
beta-D-Glucosyl-2-coumarate
Cinnamaldehyde
p-Coumaraldehyde
Caffeyl aldehyde
Coniferaldehyde
5-Hydroxy-coniferaldehyde
Resveratrol
REF1
F5H

(C)



Vitis vinifera contain the most enzymes in the pathway.

(D)

★: The species contain the indicated enzyme. Users can click ★ for details.

EC Number	Enzyme Name	All Species	Arabidopsis thaliana	Arachis hypogaea	Vitis vinifera	Saccharomyces cerevisiae	Escherichia coli str. K12 substr MG1655	Streptomyces coelicolor
Phosphoenolpyruvate								
2.5.1.54	3-deoxy-7-phosphoheptulonate synthase	★	★		★	★	★	★
2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate								
4.2.3.4	3-dehydroquinate synthase	★	★		★	★	★	★
3-Dehydroquinate								
4.2.1.10	3-dehydroquinate dehydratase	★	★		★	★	★	★
3-Dehydroshikimate								
1.1.1.25	shikimate dehydrogenase	★	★		★	★	★	
1.1.1.282	quinate/shikimate dehydrogenase	★					★	
1.1.99.25	quinate dehydrogenase (pyrroloquinoline-quinone)	★						
Shikimate								
2.7.1.71	shikimate kinase	★	★		★	★	★	★
Shikimate 3-phosphate								
2.5.1.19	3-phosphoshikimate 1-carboxyvinyltransferase	★	★		★	★	★	★
5-O-(1-Carboxyvinyl)-3-phosphoshikimate								
4.2.3.5	chorismate synthase	★	★		★	★	★	★
Chorismate								
5.4.99.5	chorismate mutase	★	★		★	★	★	
Prephenate								
1.3.1.12	prephenate dehydrogenase	★					★	
1.3.1.13	prephenate dehydrogenase (NADP+)	★				★		
1.3.1.43	arogenate dehydrogenase	★						
1.3.1.52	2-methyl-branched-chain-enoyl-CoA reductase							
3-(4-Hydroxyphenyl)pyruvate								
2.6.1.1	aspartate transaminase	★	★		★	★	★	
2.6.1.5	tyrosine transaminase	★	★					
2.6.1.57	aromatic-amino-acid transaminase	★				★	★	
2.6.1.9	histidinol-phosphate transaminase	★	★		★	★	★	★
L-Tyrosine								
4.3.1.23	tyrosine ammonia-lyase	★						
4.3.1.25	phenylalanine/tyrosine ammonia-lyase							
4-Coumarate								
6.2.1.12	4-coumarate---CoA ligase	★	★		★			
p-Coumaroyl-CoA								
2.3.1.95	trihydroxystilbene synthase	★		★	★			
3,4',5-Trihydroxystilbene								

There are no enzymes from tyrosine to resveratrol in *E. coli*.

(E)

The following table shows the statistics of enzymes which appear in the pathway in different species.

Total number of enzymes appear in the pathway : 22		
Number of Enzymes	Enzymes (EC Number)	Species
12	2.5.1.54 4.2.3.4 4.2.1.10 1.1.1.25 2.7.1.71 2.5.1.19 4.2.3.5 5.4.99.5 2.6.1.1 2.6.1.9 6.2.1.12 2.3.1.95	Vitis vinifera
11	2.5.1.54 4.2.3.4 4.2.1.10 1.1.1.25 2.7.1.71 2.5.1.19 4.2.3.5 5.4.99.5 1.3.1.12 2.6.1.1 2.6.1.9	Haemophilus influenzae
11	2.5.1.54 4.2.3.4 4.2.1.10 1.1.1.25 2.7.1.71 2.5.1.19 4.2.3.5 5.4.99.5 1.3.1.12 2.6.1.1 2.6.1.9	Synechococcus elongatus PCC 7942
11	2.5.1.54 4.2.3.4 4.2.1.10 1.1.1.25 2.7.1.71 2.5.1.19 4.2.3.5 5.4.99.5 1.3.1.12 2.6.1.1 2.6.1.9	Synechocystis sp. PCC 6803
11	2.5.1.54 4.2.3.4 4.2.1.10 1.1.1.25 2.7.1.71 2.5.1.19 4.2.3.5 5.4.99.5 1.3.1.12 2.6.1.1 2.6.1.9	Bacillus subtilis
Only list top 5 species here, please click more for details.		

(F)

Enzyme Information

EC Number	Enzyme Name	Enzyme Classification	Reaction
2.5.1.54	3-deoxy-7-phosphoheptulonate synthase	Transferases	phosphoenolpyruvate + D-erythrose 4-phosphate + H2O = 3-deoxy-D-arabino-hept-2-ulosonate 7-phosphate + phosphate

KEGG Gene Information

Species Name	KEGG Gene ID (Gene Name)
Arabidopsis thaliana	AT1G22410
Arabidopsis thaliana	AT4G33510 (DHS2)
Arabidopsis thaliana	AT4G39980 (DHS1)
Cyanidioschyzon merolae 10D	CMQ120C
Cyanidioschyzon merolae 10D	CMT165C
Saccharomyces cerevisiae S288C	YBR249C (ARO4)
Saccharomyces cerevisiae S288C	YDR035W (ARO3)
Ashbya gossypii	AGOS_ABL102C
Ashbya gossypii	AGOS_AFL047W
Schizosaccharomyces pombe 972h-	SPAC24H6.10c
Schizosaccharomyces pombe 972h-	SPAP8A3.07c

UniProt Protein Information

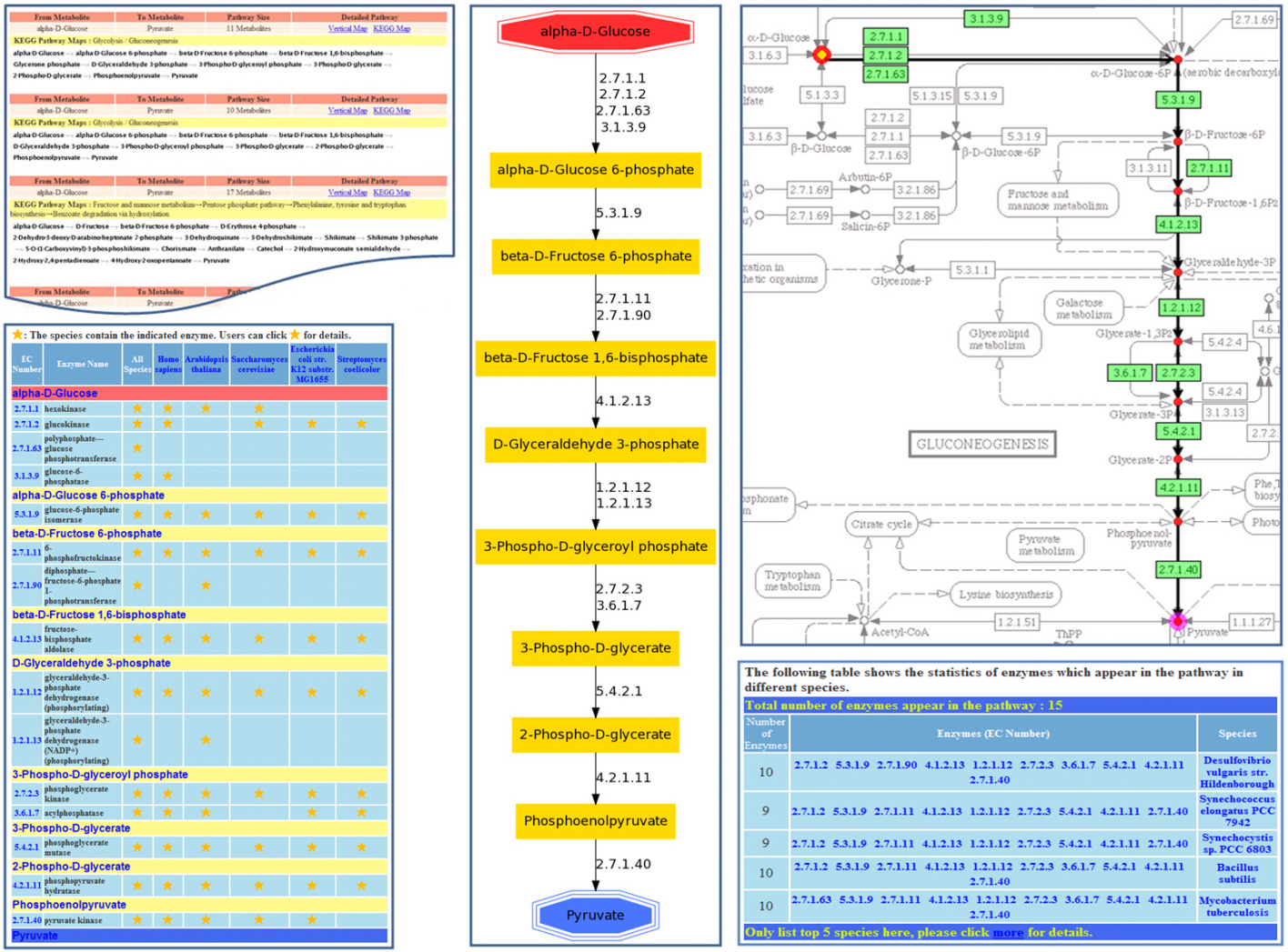
Species Name	UniProt Protein ID
Pseudomonas fluorescens	Q51789
Pseudomonas chlororaphis	Q51517
Pantoea agglomerans	Q02285
Pantoea agglomerans	O54459
Salmonella typhi	P0A1B6
Salmonella typhi	Q8Z6I9
Salmonella typhimurium	P0A1B5
Salmonella typhimurium	Q8ZPS4
Shigella flexneri	P59736
Haemophilus influenzae	P44303
Bacillus subtilis	P39912

Posttranslational Modification Information (Adapted from dbPTM)

Species Name	dbPTM	UniProt	Modification Site	Modification Type
Solanum tuberosum	AROF_SOLTU	P21357	75	Blocked amino end (Thr).
Saccharomyces cerevisiae	AROF_YEAST	P14843	26	Phosphothreonine.
Bacillus subtilis	AROG_BACSU	P39912	2	Phosphoserine.
Saccharomyces cerevisiae	AROG_YEAST	P32449	2	Phosphoserine.
Saccharomyces cerevisiae	AROG_YEAST	P32449	4	Phosphoserine.
Saccharomyces cerevisiae	AROG_YEAST	P32449	44	Phosphothreonine.
Saccharomyces cerevisiae	AROG_YEAST	P32449	47	Phosphoserine.

Figure 4.1 The results of systematic method analysis

(A) Our search algorithm shows the possible routes from PEP to resveratrol. (B) The KEGG joint map base on Fig 5(A) red circle. (C) The *Graphviz* draws map base on Fig 5(A) red circle. (D) Comparative analysis result including different enzyme. (E) The statistics of enzymes which appear in the pathway in different species. (F) The detail information about gene, protein, and post-translational modification.



From Metabolite | To Metabolite | Pathway Size | Detailed Pathway

alpha-D-Glucose | Pyruvate | 11 Metabolites | Vertical Map KEGG Map

KEGG Pathway Maps : Glycolysis / Gluconeogenesis

alpha-D-Glucose → alpha-D-Glucose 6-phosphate → beta-D-Fructose 6-phosphate → beta-D-Fructose 1,6-bisphosphate → Glycerone phosphate → D-Glyceraldehyde 3-phosphate → 3-Phospho-D-glyceroyl phosphate → 3-Phospho-D-glycerate → 2-Phospho-D-glycerate → Phosphoenolpyruvate → Pyruvate

From Metabolite | To Metabolite | Pathway Size | Detailed Pathway

alpha-D-Glucose | Pyruvate | 10 Metabolites | Vertical Map KEGG Map

KEGG Pathway Maps : Glycolysis / Gluconeogenesis

alpha-D-Glucose → alpha-D-Glucose 6-phosphate → beta-D-Fructose 6-phosphate → beta-D-Fructose 1,6-bisphosphate → D-Glyceraldehyde 3-phosphate → 3-Phospho-D-glyceroyl phosphate → 3-Phospho-D-glycerate → 2-Phospho-D-glycerate → Phosphoenolpyruvate → Pyruvate

From Metabolite | To Metabolite | Pathway Size | Detailed Pathway

alpha-D-Glucose | Pyruvate | 17 Metabolites | Vertical Map KEGG Map

KEGG Pathway Maps : Fructose and mannose metabolism→Pentose phosphate pathway→Phenylalanine, tyrosine and tryptophan biosynthesis→Benzoate degradation via hydroxylation

alpha-D-Glucose → D-Fructose → beta-D-Fructose 6-phosphate → D-Erythrose 4-phosphate → 2-Dehydro-3-deoxy-D-arabino-heptonate 7-phosphate → 3-Dehydroquinate → 3-Dehydroshikimate → Shikimate → Shikimate 3-phosphate → 5-O-(1-Carboxyvinyl)-3-phosphoshikimate → Chorismate → Anthranilate → Catechol → 2-Hydroxymuconate semialdehyde → 2-Hydroxy-2,4-pentadienoate → 4-Hydroxy-2-oxopentanoate → Pyruvate

From Metabolite | To Metabolite | Pathway Size | Detailed Pathway

alpha-D-Glucose | Pyruvate

★: The species contain the indicated enzyme. Users can click ★ for details.

EC Number	Enzyme Name	All Species	Homo sapiens	Arabidopsis thaliana	Saccharomyces cerevisiae	Escherichia coli str. K12 substr. MG1655	Streptomyces coelicolor
alpha-D-Glucose							
2.7.1.1	hexokinase	★	★	★	★		
2.7.1.2	glucokinase	★	★		★	★	★
2.7.1.63	polyphosphate---glucose phosphotransferase	★					
3.1.3.9	glucose-6-phosphatase	★	★				
alpha-D-Glucose 6-phosphate							
5.3.1.9	glucose-6-phosphate isomerase	★	★	★	★	★	★
beta-D-Fructose 6-phosphate							
2.7.1.11	6-phosphofructokinase	★	★	★	★	★	★
2.7.1.90	diphosphate---fructose-6-phosphate 1-phosphotransferase	★		★			
beta-D-Fructose 1,6-bisphosphate							
4.1.2.13	fructose-bisphosphate aldolase	★	★	★	★	★	★
D-Glyceraldehyde 3-phosphate							
1.2.1.12	glyceraldehyde-3-phosphate dehydrogenase (phosphorylating)	★	★	★	★	★	★
1.2.1.13	glyceraldehyde-3-phosphate dehydrogenase (NADP+) (phosphorylating)	★		★			
3-Phospho-D-glyceroyl phosphate							
2.7.2.3	phosphoglycerate kinase	★	★	★	★	★	★
3.6.1.7	acylphosphatase	★	★	★		★	★
3-Phospho-D-glycerate							
5.4.2.1	phosphoglycerate mutase	★	★	★	★	★	★
2-Phospho-D-glycerate							
4.2.1.11	phosphopyruvate hydratase	★	★	★	★	★	★
Phosphoenolpyruvate							
2.7.1.40	pyruvate kinase	★	★	★	★	★	
Pyruvate							

The following table shows the statistics of enzymes which appear in the pathway in different species.

Total number of enzymes appear in the pathway : 15

Number of Enzymes	Enzymes (EC Number)	Species
10	2.7.1.2 5.3.1.9 2.7.1.90 4.1.2.13 1.2.1.12 2.7.2.3 3.6.1.7 5.4.2.1 4.2.1.11 2.7.1.40	Desulfovibrio vulgaris str. Hildenborough
9	2.7.1.2 5.3.1.9 2.7.1.11 4.1.2.13 1.2.1.12 2.7.2.3 5.4.2.1 4.2.1.11 2.7.1.40	Synechococcus elongatus PCC 7942
9	2.7.1.2 5.3.1.9 2.7.1.11 4.1.2.13 1.2.1.12 2.7.2.3 5.4.2.1 4.2.1.11 2.7.1.40	Synechocystis sp. PCC 6803
10	2.7.1.2 5.3.1.9 2.7.1.11 4.1.2.13 1.2.1.12 2.7.2.3 3.6.1.7 5.4.2.1 4.2.1.11 2.7.1.40	Bacillus subtilis
10	2.7.1.63 5.3.1.9 2.7.1.11 4.1.2.13 1.2.1.12 2.7.2.3 3.6.1.7 5.4.2.1 4.2.1.11 2.7.1.40	Mycobacterium tuberculosis

Only list top 5 species here, please click more for details.

Figure 4.2 The results of glycolysis pathway

These figures indicate that our systematic analysis can apply to glycolysis pathway.

4.2 Web Interface

The systematic method is to develop a web server named FMM. FMM is an effective tool for applications in synthetic biology to produce both drugs and biofuels. This novel and innovative resource is now freely available at http://FMM.mbc.nctu.edu.tw/ and published in *Nucleic acids research*[34]. Figure 4.3 presents the user-friendly web interface of the FMM. In "Start FMM", two metabolites of interest (keyword or KEGG compound ID) can be inputted and several species can be selected from four categories (animals, plants, fungi and prokaryotic). The metabolic pathways corresponding to the two input metabolites can be reconstructed, as given in Figure 4.3 **(**A). As to the comparative analysis of metabolic pathways, Figure 4.3 (B) includes one species in the "major species" column (of microorganisms commonly used in synthetic biology) and several in the "comparative species" column (of several organisms in common use in the laboratory).

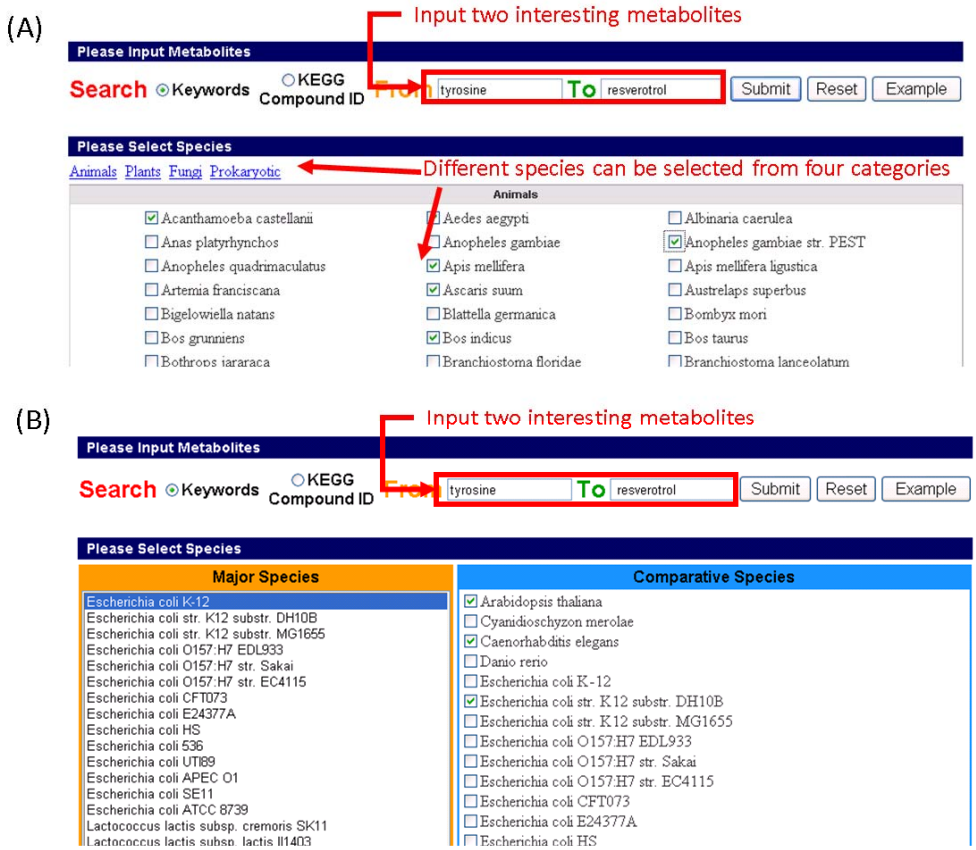


Figure 4.3 The input interface of web server
(A) The input interface of “Start FMM” (B) The input interface of “Comparative Analysis”

4.3 Case study: Synthetically produce resveratrol in *E. coil*

4.3.1 Background

According to our introduction, we are interested in plant secondary metabolites including isoflavones, catechins, and resveratrol. When we deeply studied on their biosynthesis pathways, we found out their biosynthesis from *p*-courmonyl-CoA through one enzyme for resveratrol, three enzymes for isoflavones, and five enzymes for catechins [35] (see Figure 4.4). We think the biosynthesis of resveratrol may easier than others two secondary metabolites because of their pathway length. We query the price from sigma aldrich (http://www.sigmaaldrich.com/sigma-aldrich/home.html) and discovered the price was much higher than other precursor. The query result is shown in Table 4.2. The property of resveratrol is that resveratrol is an antioxidant, positively regulates many physiological and cellular processes in animals by inhibiting lipid peroxidation, reducing cholesterol levels, downregulating platelet aggregation and exhibiting anti-inflammatory activity[36]. Resveratrol was added to the food starting in early adulthood and caused a dose-dependent increase of median and maximum lifespan[37]. Related study of biosynthesis resveratrol in microbe, Beekwilder *et al.* reported the yield of resveratrol is three-fold higher than yeast[38]. On the basis of the above mentioned, we want to synthesize resveratrol in *E. coli*.

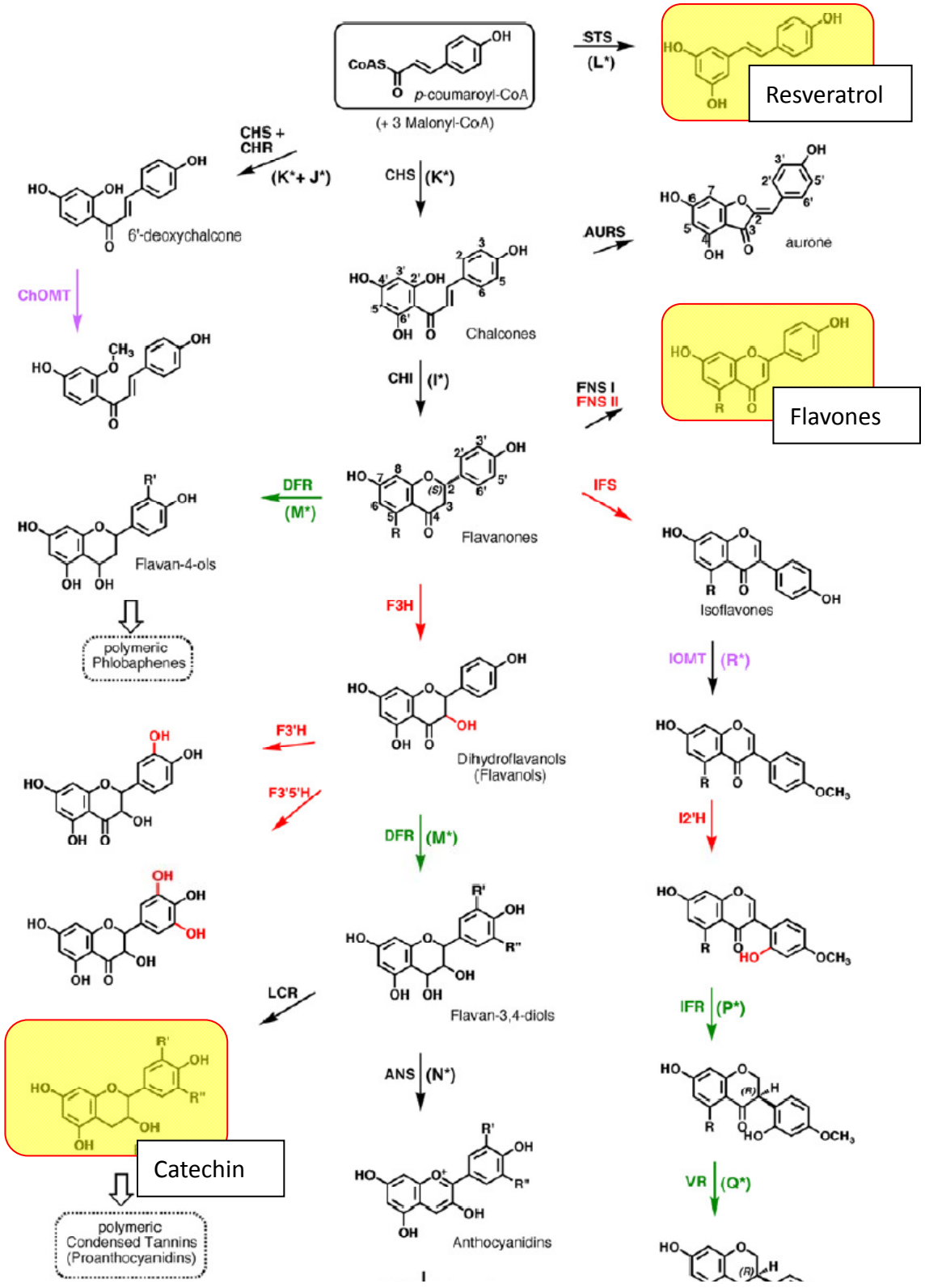


Figure 4.4 Resveratrol, isoflavones, and catechin biosynthesis pathway

(Ferrer, J.-L. *et al.*)

The biosynthesis from *p*-courmonyl-CoA is through one enzyme for resveratrol, three enzymes for isoflavones, and five enzymes for catechin.

Table 4.2 Compare the price with different metabolites

Sigma Product Number	Name	Price (USD)
G8270	D-Glucose	0.01/g
P8740	L-Phenylalanine	0.429/g
T4321	L-Tyrosine	0.429/g
R5010	**Resveratrol**	**937/g**

4.3.2 Systematic method analysis

On our knowledge of biology, the glycolysis is a central metabolism in almost all of the organisms. It can convert glucose into pyruvate. Our tool wants to know whether a biosynthesis pathway is fitting in metabolic engineering species or not. We will search the phosphoenolpyruvate (PEP) because it is the end of glycolysis and the intermediate of resveratrol biosynthesis. PEP (C00074) and resveratrol (C03582) were inputted into our tool. We select two parts of species which one is the resveratrol specific organisms (*Arabidopsis thaliana*, *Arachis hypogaea*, *Vitis vinifera*) and another is the metabolic engineering organisms (*Saccharomyces cerevisiae*, *Escherichia coli str. K12 substr. MG1655*, *Streptomyces coelicolor*). The results are similar to Figure 4.1. It is display we can reconstruct the metabolic pathways from PEP to resveratrol. Comparative analysis is shown in Figure 4.1 (D) reveal that if we want to synthesize resveratrol in *E. coli*, we have to clone three gene including tyrosine ammonia-lyase or phenylalanine/tyrosine ammonia-lyase, 4-coumarate-CoA ligase, and trihydroxystilbene synthase from others species. *E. coli* was lack enzymes from tyrosine to resveratrol but *A. thaliana*, *A .hypogaea*, and *V. vinifera* contain related enzymes. And then the detail analysis the enzyme of this three species, there are no post-translational modification. Also, we have their gene and protein sequence, so we may clone it from these specific to *E. coli*.

4.3.3 Computational modeling

4.3.3.1 Metabolic network from carbon source to resveratrol

The metabolite network and reaction data are integrated from KEGG, MetaCyc, BiGG [28] database. We built up a metabolic network for modeling the resveratrol production. The network is shown in Figure 4.5, which contains glycolysis, pentose phosphate pathway, citrate acid cycle, aromatic amino acid, resveratrol biosynthesis, and biomass formation. The first control point for synthesis resveratrol is erythrose-4-phosphate (E4P) and PEP to produce 3-deoxy-D-arabinoheptulosonate 7-phosphate (DAHP) by the enzyme DAHP synthase. The production of DAHP is close to theoretical yield by elementary mode analysis with metabolic engineering experiment. The second control point is chorismate to anthranilate, which is a branch point to synthesis L-tryptophan. The third control point is prephenate to phenylpyruvate, which is another branch point to synthesis L-phenylalanine. The result of elementary flux modes analysis shows that the highly yield of resveratrol from glucose (about 0.38-0.40 (g resveratrol/g sugars)) is not contains the second and third reaction (see Figure 4.6). The second and third reaction will delete for elementary mode analysis. Serine and 5-Phospho-alpha-D-ribose 1-diphosphate (PRPP) in this network are only for L-tryptophan biosynthesis, so we will delete the reaction in knock out analysis. Figure 4.6 shows an example of the maximum resveratrol yield elementary mode, which net reaction is Net reaction: 29 GLU_ext + 69 OXY_ext = 75 CO_2_ext + 9 Resveratrol_ext (GLU_ext: external glucose; OXY_ext: external oxygen; CO_2_ext: external CO_2). Trinh *et al.* construct a most efficient biomass producing E. coli by delete the reaction from fumarate to succinate[39]. We also delete this reaction in our metabolic network.

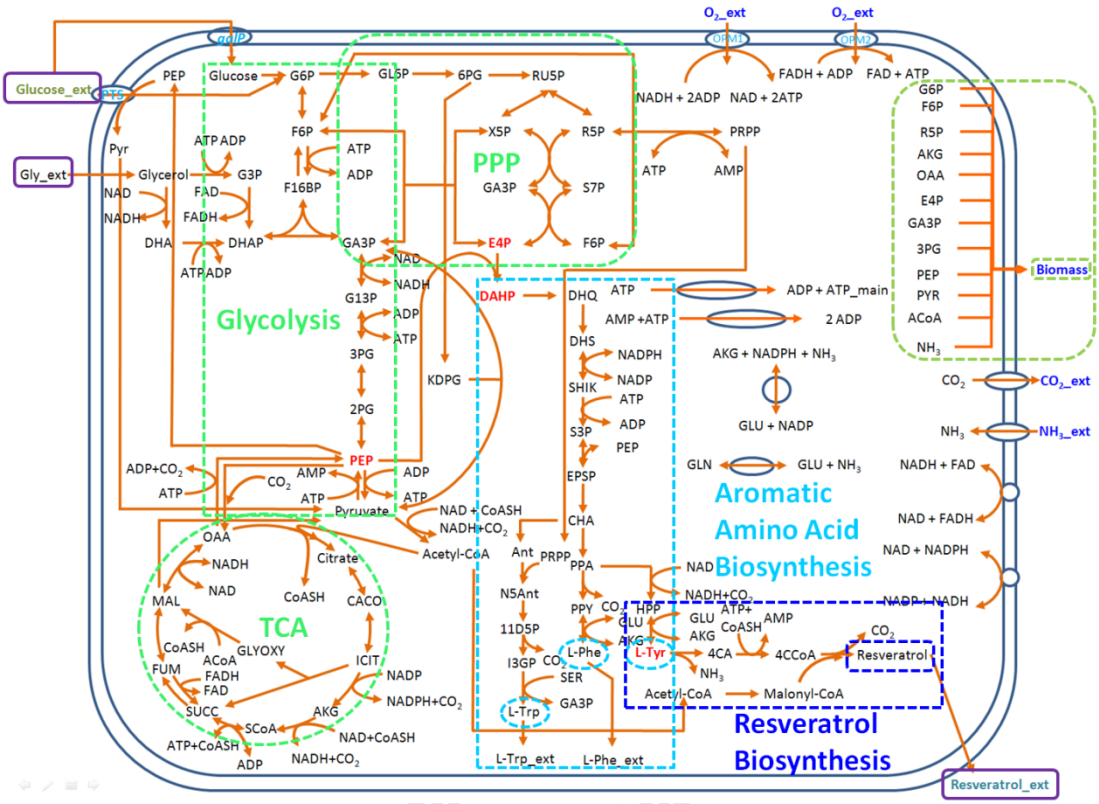


Figure 4.5 Modeling pathway from carbon source to resveratrol

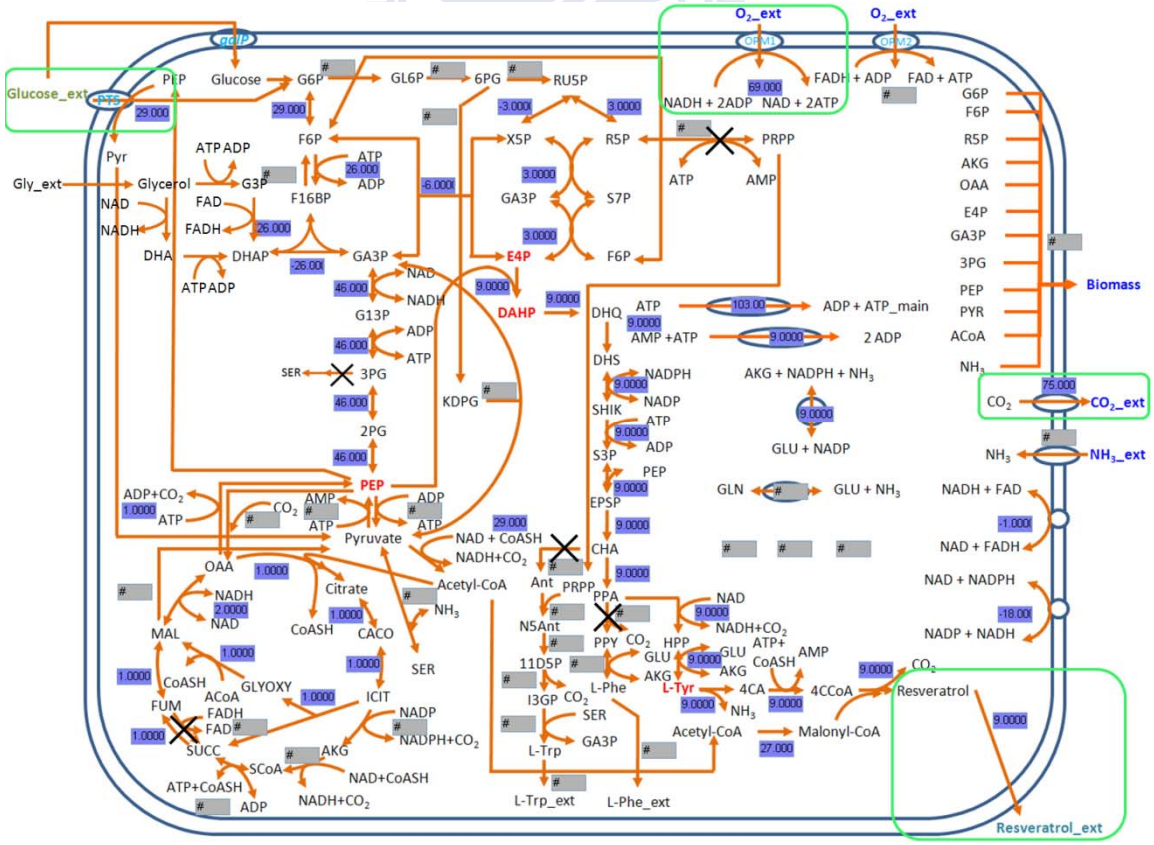


Figure 4.6 An example of the maximum resveratrol yield elementary mode

4.3.3.2 Glucose as carbon source

Based on the metabolic network considered, we identified 8885 EMs that *E.coli* can use to metabolize glucose. We find out 1360 can make resveratrol, 2795 can synthesize biomass, and 840 can coproduce resveratrol and biomass (see Table 4.3). The maximum yield of resveratrol is 0.40 (g resveratrol/g sugar). On the knock out condition analysis, the totally elementary mode is show fell 80%, and the yield of resveratrol, biomass, and resveratrol and biomass are reduced to 37%, 21%, 36% respectively, but the yield of resveratrol is maintained in 0.4 (g resveratrol/g sugar). The wild type profile of resveratrol yield is similar to knock out, and the biomass, too (see Figure 4.7). In resveratrol and biomass co-produce modes, knock out condition is negative correlation to wild type (see Figure 4.8). It means the resveratrol production in wild type is not correlated to biomass as knock out condition. We consider that the knock out condition may be an efficient strain to synthesis resveratrol, but resveratrol yield may be more effective by biomass yield.

Table 4.3 Elementary mode analysis of glucose uptake

EM type (Only glucose growth)	**WT**	**KO**
Total EMs	8885	1828
Resveratrol-production EMs	1360	507
Growth EMs	2795	588
Growth &Resveratrol-production EMs	840	303
Resveratrol yield (g resveratrol/g sugars)	0.0004-0.40	0.003-0.40
Biomass yield (g biomass/g sugars)	0.0050-0.83	0.05-0.83

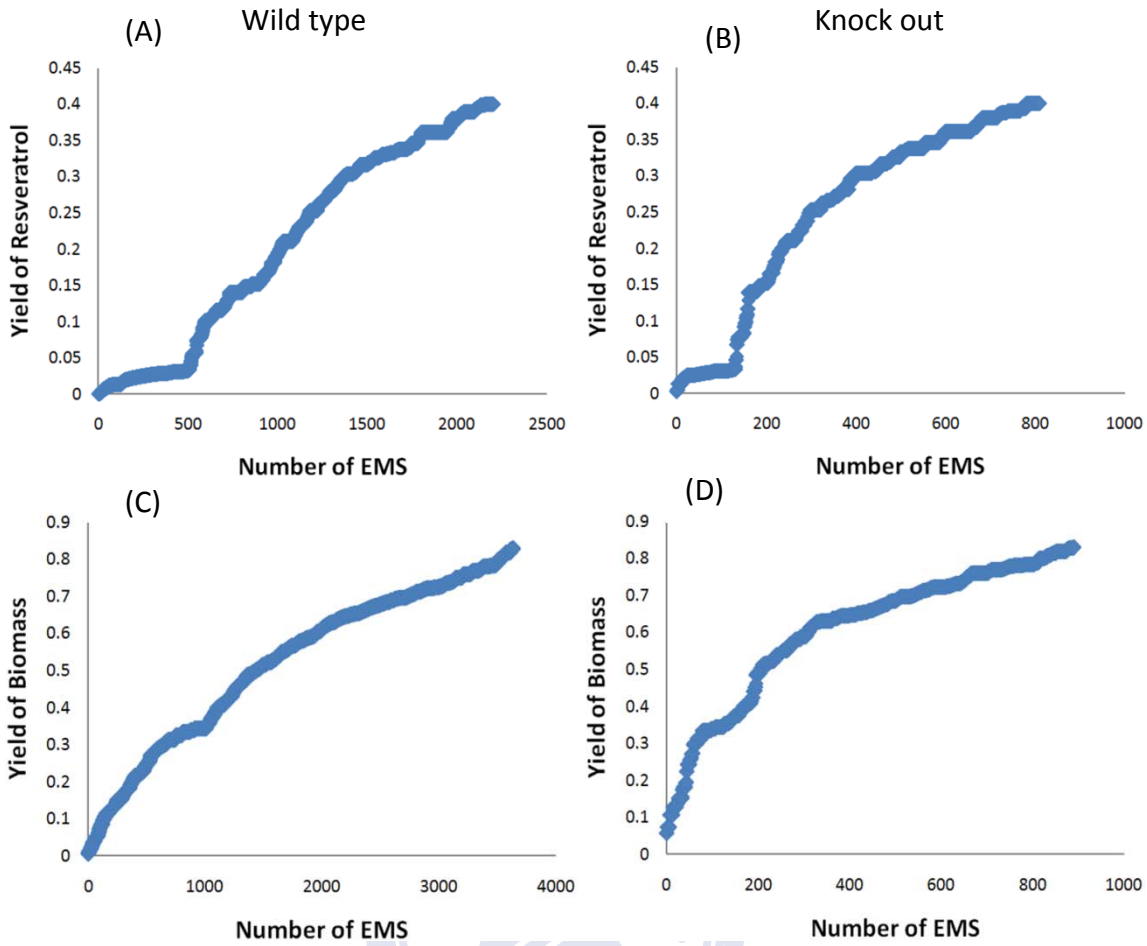


Figure 4.7 Elementary flux modes analysis of glucose uptake to produce resveratrol

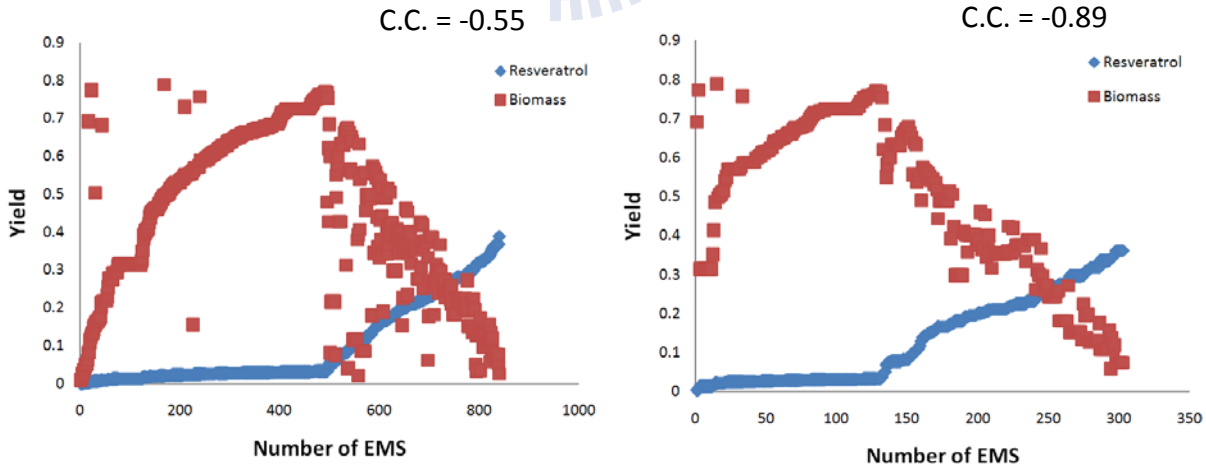


Figure 4.8 The relationship of resveratrol and biomass yield in glucose uptake

C.C. is correlation coefficient

4.3.3.3 Coultilization glucose and glycerol

Glycerol is a by-product of biodiesel production (via the transesterification of vegetable oils or animal fats). Nowadays, we regard crude glycerol as a 'waste stream' with a disposal cost, and their cost will low[40]. Karla Martínez et.al, coutilization of glucose and glycerol enhances the production of aromatic compounds precursor DAHP in an *E. coli* strain[41]. DAHP is a precursor of resveratrol; we think we can try to synthesize resveratrol by coutilization glucose and glycerol. The result is shown in Table 4.4 that reveal the maximum yield of resveratrol maintained in 0.4 (g resveratrol/g sugar), but the maximum yield of biomass increased from 0.83 to 0.95 (g biomass/g sugar). The present of knock out condition reduce mode are similar to only glucose uptake. The wild type profile of resveratrol yield is also similar to knock out, and the biomass, too (see Figure 4.9). In resveratrol and biomass co-produce modes, knock out condition is also negative correlation than wild type (see Figure 4.10). It is also means the resveratrol production in wild type is not correlated to biomass than knock out condition. We consider that coutilization of glucose and glycerol may synthesize more biomass to support cell growth, indirectly increase resveratrol production, than yield.

Table 4.4 Elementary mode analysis of coutilization glucose and glycerol

EM type (Glucose and glycerol growth)	WT	KO
Total EMs	36955	7269
Resveratrol-production EMs	5644	2290
Growth EMs	15027	2697
Growth &Resveratrol-production EMs	4397	1517
Resveratrol yield (g resveratrol/g sugars)	0.0002-0.40	0.0005-0.40
Biomass yield (g biomass/g sugars)	0.0050-0.95	0.04-0.95

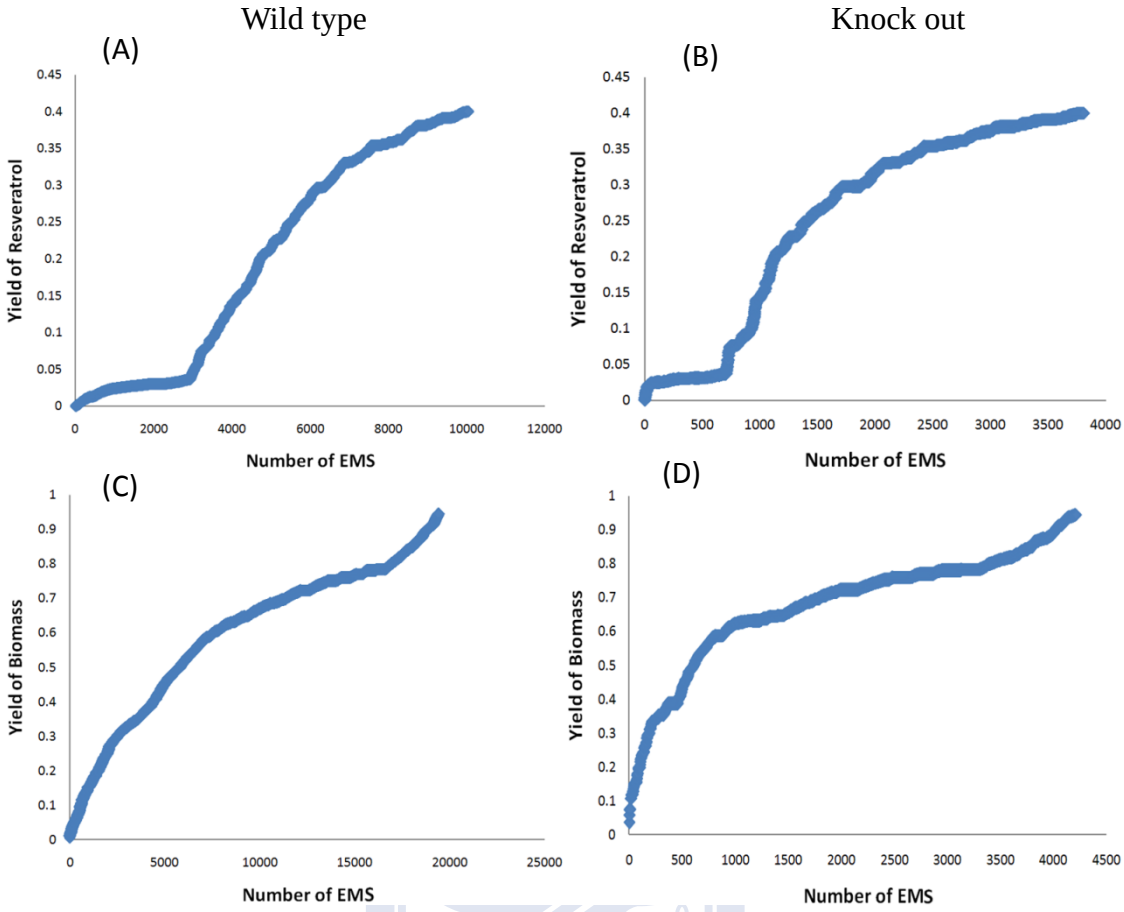


Figure 4.9 Elementary flux modes analysis of coutilization glucose and glycerol to produce resveratrol

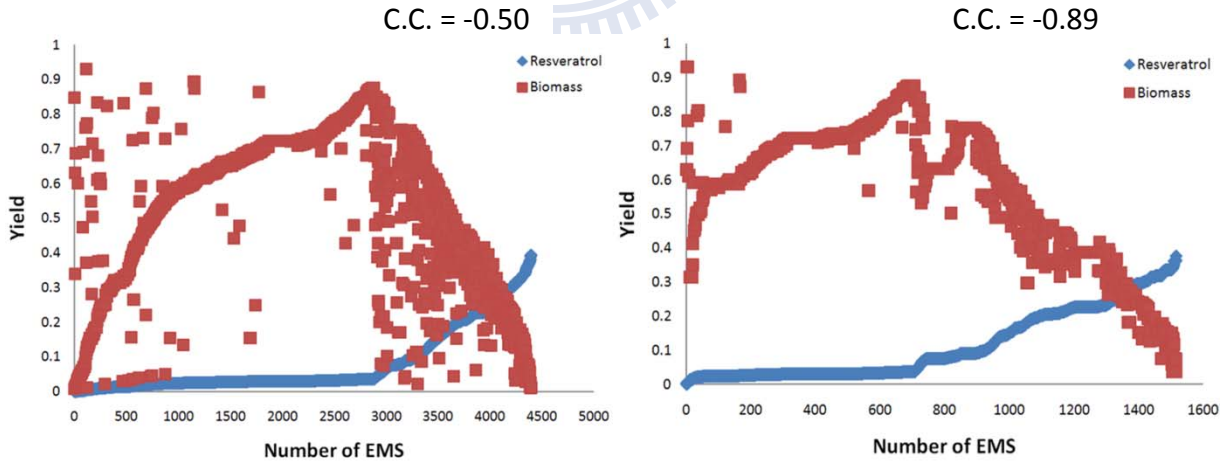


Figure 4.10 The relationship of resveratrol and biomass yield in coutilization of glucose and glycerol

C.C. is correlation coefficient

Chapter5 Discussion

5.1 Systematic method search possible find non-biological pathway

Base on breadth first search, the system will find non-biological pathway, because it is applied to find the connection between only two compounds. Therefore, some disjoint map or unique pathway will be displayed in our result, we provide the KEGG joint map for biological research to view they want and put the possible pathway in the top of the result by module. We think there is a powerful tool to analyze any two metabolites and comparative analysis.

5.2 Phenylalanine, tyrosine, and tryptophan biosynthesis contain many feedback inhibition

The pathway of phenylalanine, tyrosine, and tryptophan biosynthesis contain some feedback pathway[42]. The elementary mode analysis are not considered the feedback inhibit effect. All of the tools for metabolic pathway analysis are also not considered the effect. Therefore, metabolic engineering has been modified *E. coli* to remove these feedback resistant [41, 43, 44], such as *aroG*, *tyrA*, and *aroB* have removed feedback resistant (see Figure 5.1). We assume that based on these modify E. coli strain, we model the pathway from carbon source to resveratrol.

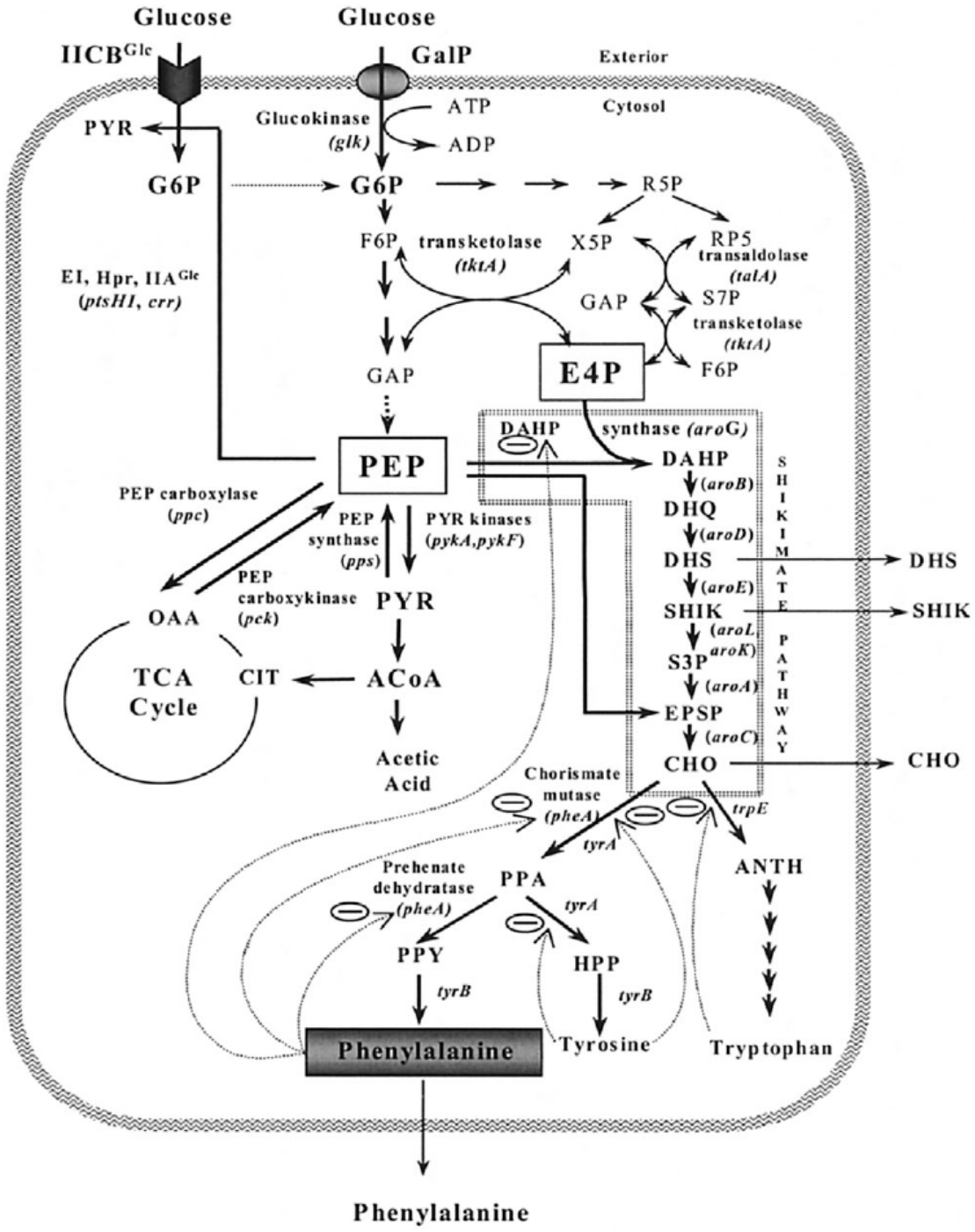


Figure 5.1 Feedback resistance in aromatic amino acid biosynthesis

(Baez-Viveros, J.L., *et al* 2004)

Metabolites symbols: G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; GAP, glyceraldehyde 3-phosphate; R5P, ribulose 5-phosphate; R5P, ribose 5-phosphate; X5P, xylulose 5-phosphate;S7P, sedoheptulose 7-phosphate; PYR, pyruvate; PEP, phosphoenolpyruvate; E4P, erythrose 4-phosphate; DAHP, 3-deoxy-D-arabino-heptulosonate 7-phosphate; DHQ, 3-dehydroquinate; DHS, 3-dehydroshikimate; SHIK, shikimate; EPSP, 5-enolpyruvylshikimate 3-phosphate;

CHO, chorismate; PPA, prephenate; PPY, phenylpyruvate; Phe, phenylalanine; Tyr, tyrosine; Trp, tryptophan; OAA, oxaloacetate; CIT, citrate; ACoA, acetyl coenzyme A. Protein and gene symbols: PTS, phosphotransferase transport system; GalP, galactose permease; glk, glucokinase; ptsHI, crr, PTS general proteins and enzyme IIA^{Glc}, respectively; talA, transaldolase; tktA, transketolase; aroG, DAHP synthase; aroB, DHQ synthase; aroD, DHQ dehydratase; aroE, SHIK dehydrogenase; aroL, aroK, SHIK kinase; aroA, EPSP synthase; aroC, CHO synthase; pheA, chorismate mutase-prephenate dehydratase.

5.3 Yield of DAHP in *E. coli*

DAHP is an important precursor for aromatic amino acid. Martinez *et al.* use glucose and glycerol as carbon source to enhance the production of aromatic compounds in an E. coli strain. The result of DAHP production by elementary mode analysis shows in Table 5.1. The result of biomass production is similar to Martinez *et al.*, which the coutilization of glucose and glycerol can promote biomass production.

Table 5.1 Elementary mode analysis of DAHP yield

Carbon souece	**Glucose**	**Glucose&Glycerol**
Total EMs	1661	5792
DAHP-production EMs	422	1235
Growth EMs	588	2697
Growth &Resveratrol-production EMs	221	1094
DAHP yield (g DAHP/g sugars)	0.004-0.995	0.0006-1.00
Biomass yield (g biomass/g sugars)	0.01-0.83	0.01-0.95

Chapter6 Conclusion and future work

6.1 Conclusion

The concept of our goal is illustrated in Figure 1.6, which demonstrates that metabolic pathways can be easily reconstructed among various species using the systematic method; moreover, modeling metabolic pathway can detail understand the yield of our interesting. The results suggest that systematic method is helpful in metabolic engineering by reconstructing metabolic pathways for producing some valuable metabolites or secondary metabolites in bacteria or yeast, whose approach has potential in drug production. Moreover, it can be used to compare the metabolic pathways with numerous species and connect metabolic pathways of different KEGG maps, suggesting that FMM is not only an effective tool for synthetic biology, such as in the production of drugs and biofuels, but also a useful resource for investigation in metabolism.

6.2 Future work

The systematic method has been published in *Nucleic acids research*. We have to update the data which we collected from databases, such as KEGG, Uni-ProtKB. The MetaCyc database contains a lot of data which has been provided by experiment, so we have to try to integrate their data. Before that, based on KEGG map, we did not integrate MetaCyc data. Figure 6.1 shows the summary of my study (blue box). In the future, we may try to prove the problem by experiments, because we have obtained three key enzymes for synthesis resveratrol (see Table 6.1). Finally, the following step and result will implement the synthetic biology.

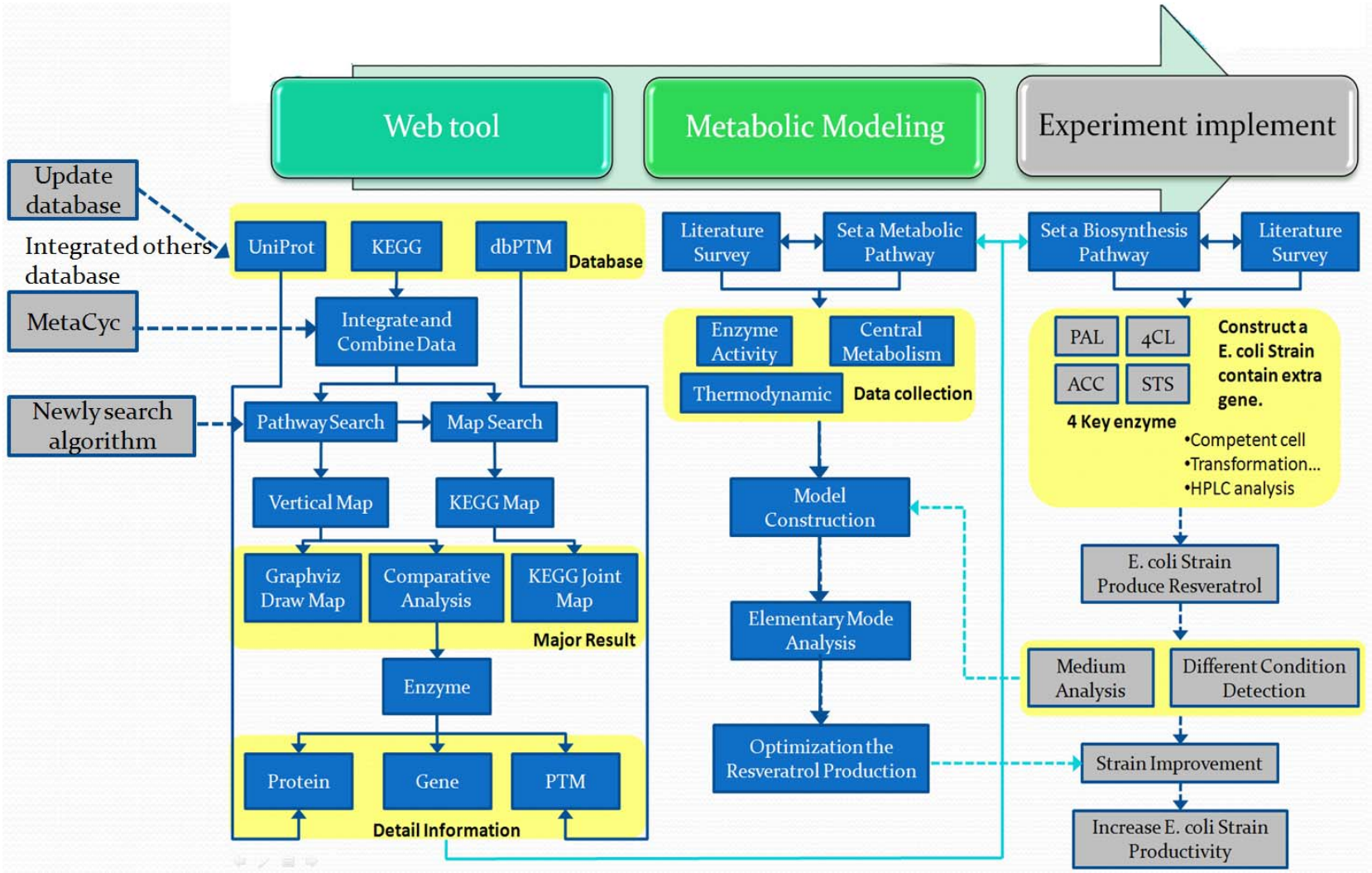


Figure 6.1 Summary and future work

Table 6.1 The plasmid for resveratrol biosynthesis in *E. coli*

Gene name	Vector	Origin	Anti	Restriction E	Gene Sequence	Promoter
PAL	pET16b based	pBR322	Amp	Nde1 : BamH1	X13094.1	T7
4CL	pBluescript SK-	pUC	Amp	EcoR1 : Xho1	D49366.1	
ACC (dtsR1,accBC)	pRSFDuet-1	ColeE1	Kan	BamH1:HindIII/Nde1:Xho1	unpublished	T7
PMT	pACYCDuet-1	CDF	Cam	EcoR1 : HindIII	Contain res site	T7
STS					GeneSynthesis	

Symbols: PAL: Phenylalanine ammonia lyase; 4CL: Cinnamate 4-hydroxylase; ACC: Acetyl-CoA carboxylase; PMT: Pinosylvin methyltransferase ; STS: Stilbene synthase

Reference

1. Dewick, P.M., *Medicinal natural products: a biosynthetic approach*. 2nd ed. 2002, Chichester: John Wiley & Sons, Ltd.
2. Channon, K., E.H. Bromley, and D.N. Woolfson, *Synthetic biology through biomolecular design and engineering.* Curr Opin Struct Biol, 2008. **18**(4): p. 491-8.
3. Crozier, A., M.N. Clifford, and H. Ashihara, *Plant Secondary Metabolites*. 1st ed. 2006, Oxford: Blackwell Publishing Ltd.
4. Vermeris, W. and R. Nicholson, *Phenolic Compound Biochemistry*. 2006, Dordrecht: Springer.
5. Zhu, Y.Z., et al., *Antioxidants in Chinese herbal medicines: a biochemical perspective.* Nat Prod Rep, 2004. **21**(4): p. 478-89.
6. Endy, D., *Foundations for engineering biology.* Nature, 2005. **438**(7067): p. 449-53.
7. Heinemann, M. and S. Panke, *Synthetic biology--putting engineering into biology.* Bioinformatics, 2006. **22**(22): p. 2790-9.
8. Prather, K.L. and C.H. Martin, *De novo biosynthetic pathways: rational design of microbial chemical factories.* Curr Opin Biotechnol, 2008. **19**(5): p. 468-74.
9. Park, J.H. and S.Y. Lee, *Towards systems metabolic engineering of microorganisms for amino acid production.* Curr Opin Biotechnol, 2008. **19**(5): p. 454-60.
10. Lee, J.H., et al., *Global analyses of transcriptomes and proteomes of a parent strain and an L-threonine-overproducing mutant strain.* J Bacteriol, 2003. **185**(18): p. 5442-51.
11. Ohnishi, J., et al., *A novel gnd mutation leading to increased L-lysine production in Corynebacterium glutamicum.* FEMS Microbiol Lett, 2005. **242**(2): p. 265-74.
12. Park, J.H., et al., *Metabolic engineering of Escherichia coli for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation.* Proc Natl Acad Sci U S A, 2007. **104**(19): p. 7797-802.
13. Ruckert, C., A. Puhler, and J. Kalinowski, *Genome-wide analysis of the L-methionine biosynthetic pathway in Corynebacterium glutamicum by targeted gene deletion and homologous complementation.* J Biotechnol, 2003. **104**(1-3): p. 213-28.
14. Garg, R.P., et al., *Investigations of valanimycin biosynthesis: elucidation of the*

role of seryl-tRNA. Proc Natl Acad Sci U S A, 2008. **105**(18): p. 6543-7.

15. Atsumi, S., T. Hanai, and J.C. Liao, *Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels.* Nature, 2008. **451**(7174): p. 86-9.
16. Diella, F., et al., *Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins.* BMC Bioinformatics, 2004. **5**: p. 79.
17. Jensen, O.N., *Interpreting the protein language using proteomics.* Nat Rev Mol Cell Biol, 2006. **7**(6): p. 391-403.
18. Kanehisa, M., et al., *KEGG for linking genomes to life and the environment.* Nucleic Acids Res, 2008. **36**(Database issue): p. D480-4.
19. Caspi, R., et al., *The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases.* Nucleic Acids Res, 2008. **36**(Database issue): p. D623-31.
20. Chang, A., et al., *BRENDA, AMENDA and FRENDA the enzyme information system: new content and tools in 2009.* Nucleic Acids Res, 2009. **37**(Database issue): p. D588-92.
21. Li, Y., et al., *Metabolic pathway alignment between species using a comprehensive and flexible similarity measure.* BMC Syst Biol, 2008. **2**: p. 111.
22. Boutet, E., et al., *UniProtKB/Swiss-Prot.* Methods Mol Biol, 2007. **406**: p. 89-112.
23. Lee, T.Y., et al., *A comprehensive resource for integrating and displaying protein post-translational modifications.* BMC Res Notes, 2009. **2**: p. 111.
24. Oehm, S., et al., *Comparative Pathway Analyzer--a web server for comparative analysis, clustering and visualization of metabolic networks in multiple organisms.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W433-7.
25. Pommerenke, C., et al., *ROSY--a flexible and universal database and bioinformatics tool platform for Roseobacter related species.* In Silico Biol, 2008. **8**(2): p. 177-86.
26. Katsuyama, Y., N. Funa, and S. Horinouchi, *Precursor-directed biosynthesis of stilbene methyl ethers in Escherichia coli.* Biotechnol J, 2007. **2**(10): p. 1286-93.
27. Price, N.D., J.L. Reed, and B.O. Palsson, *Genome-scale models of microbial cells: evaluating the consequences of constraints.* Nat Rev Microbiol, 2004. **2**(11): p. 886-97.
28. Reed, J.L., et al., *An expanded genome-scale model of Escherichia coli K-12*

(iJR904 GSM/GPR). Genome Biol, 2003. **4**(9): p. R54.
29. Lee, T.Y., et al., *dbPTM: an information repository of protein post-translational modification.* Nucleic Acids Res, 2006. **34**(Database issue): p. D622-7.
30. Carlson, R. and F. Srienc, *Fundamental Escherichia coli biochemical pathways for biomass and energy production: identification of reactions.* Biotechnol Bioeng, 2004. **85**(1): p. 1-19.
31. Schuster, S., D.A. Fell, and T. Dandekar, *A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks.* Nat Biotechnol, 2000. **18**(3): p. 326-32.
32. Trinh, C.T., A. Wlaschin, and F. Srienc, *Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism.* Appl Microbiol Biotechnol, 2009. **81**(5): p. 813-26.
33. von Kamp, A. and S. Schuster, *Metatool 5.0: fast and flexible elementary modes analysis.* Bioinformatics, 2006. **22**(15): p. 1930-1.
34. Chou, C.H., et al., *FMM: a web server for metabolic pathway reconstruction and comparative analysis.* Nucleic Acids Res, 2009.
35. Ferrer, J.L., et al., *Structure and function of enzymes involved in the biosynthesis of phenylpropanoids.* Plant Physiol Biochem, 2008. **46**(3): p. 356-70.
36. Kundu, J.K. and Y.J. Surh, *Molecular basis of chemoprevention by resveratrol: NF-kappaB and AP-1 as potential targets.* Mutat Res, 2004. **555**(1-2): p. 65-80.
37. Valenzano, D.R., et al., *Resveratrol prolongs lifespan and retards the onset of age-related markers in a short-lived vertebrate.* Curr Biol, 2006. **16**(3): p. 296-300.
38. Beekwilder, J., et al., *Production of resveratrol in recombinant microorganisms.* Appl Environ Microbiol, 2006. **72**(8): p. 5670-2.
39. Trinh, C.T., et al., *Design, construction and performance of the most efficient biomass producing E. coli bacterium.* Metab Eng, 2006. **8**(6): p. 628-38.
40. Yazdani, S.S. and R. Gonzalez, *Anaerobic fermentation of glycerol: a path to economic viability for the biofuels industry.* Curr Opin Biotechnol, 2007. **18**(3): p. 213-9.
41. Martinez, K., et al., *Coutilization of glucose and glycerol enhances the production of aromatic compounds in an Escherichia coli strain lacking the phosphoenolpyruvate: carbohydrate phosphotransferase system.* Microb Cell Fact, 2008. **7**: p. 1.
42. Baez-Viveros, J.L., et al., *Metabolic engineering and protein directed*

evolution increase the yield of L-phenylalanine synthesized from glucose in Escherichia coli. Biotechnol Bioeng, 2004. **87**(4): p. 516-24.

43. Lutke-Eversloh, T. and G. Stephanopoulos, *L-tyrosine production by deregulated strains of Escherichia coli.* Appl Microbiol Biotechnol, 2007. **75**(1): p. 103-10.

44. Olson, M.M., et al., *Production of tyrosine from sucrose or glucose achieved by rapid genetic changes to phenylalanine-producing Escherichia coli strains.* Appl Microbiol Biotechnol, 2007. **74**(5): p. 1031-40.

Appendix

Appendix 1

The connections through current metabolites and cofactors were deleted to make the path length analysis physiologically more meaningful, that deleted data are given in the following table.

Compound ID	Number of reaction	Compound name
C00001	2572	H2O;Water
C00080	1658	H+
C00007	1016	Oxygen;O2
C00006	819	NADP+;NADP;Nicotinamide adenine dinucleotide phosphate
C00005	817	NADPH;TPNH
C00003	737	NAD+;NAD;Nicotinamide adenine dinucleotide
C00004	727	NADH;DPNH
C00002	492	ATP;Adenosine 5'-triphosphate
C00011	460	CO2;Carbon dioxide
C00010	453	CoA;Coenzyme A;CoA-SH
C00009	418	Orthophosphate;Phosphate;Phosphoric acid
C00008	348	ADP;Adenosine 5'-diphosphate
C00013	327	Pyrophosphate;Pyrophosphoric acid;Diphosphate;PPi
C00014	310	NH3;Ammonia
C00019	280	S-Adenosyl-L-methionine;S-Adenosylmethionine
C00021	268	S-Adenosyl-L-homocysteine;S-Adenosylhomocysteine
C00015	259	UDP;Uridine 5'-diphosphate
C00028	201	Acceptor;A
C00030	200	Reduced acceptor;AH2
C00027	181	H2O2;Hydrogen peroxide;Oxydol
C00020	173	AMP;Adenosine 5'-monophosphate;Adenylic acid
C00055	64	CMP;Cytidine-5'-monophosphate;Cytidylic acid
C00016	49	FAD;Flavin adenine dinucleotide
C00363	48	dTDP;Deoxythymidine 5'-diphosphate
C01352	46	FADH2
C00035	41	GDP;Guanosine 5'-diphosphate;Guanosine diphosphate

C00138	35	Reduced ferredoxin
C00139	35	Oxidized ferredoxin
C00162	34	Fatty acid
C00017	16	Protein
C00151	14	L-Amino acid;L-2-Amino acid
C00237	14	CO;Carbon monoxide
C05359	13	e-;electron
C03024	12	Reduced flavoprotein
C03161	12	Oxidized flavoprotein
C00662	11	Reduced adrenal ferredoxin;Reduced adrenodoxin
C00667	11	Oxidized adrenal ferredoxin;Oxidized adrenodoxin
C00075	24	UTP;Uridine 5'-triphosphate;Uridine triphosphate
C00044	31	GTP;Guanosine 5'-triphosphate
C00105	28	UMP;Uridylic acid;Uridine monophosphate
C00144	15	GMP;Guanosine 5'-phosphate;Guanosine monophosphate
C00063	30	CTP;Cytidine 5'-triphosphate;Cytidine triphosphate

Appendix 2

Input file for the program METATOOL of resveratrol biosynthesis from carbon source in *E. coli*.

Identifiers: ENZREV, reversible enzymes; ENZIRREV, irreversible enzymes; METINT, internal metabolites; METEXT, external metabolites; CAT, catalyzed reactions.

-ENZREV
R2r R5r R6r R7_1r R7_2r R7_3r R8r R11r R12r R13r R14r R15r R23_1r R23_2r
R26r R28r R29r
R83r R84r
SHK3Dr PSCVTr PHETA1r TYRTAr
GLUDr GLSr
PRPPsr

-ENZIRREV
R1 R3 R4 R9 RR9 R10_1 R10_2 R10_3
glpF glpK GLYD3 GLYD4 GLYD5
R21 R22 R24 R25 R27 R30
R40 R41 R42 GLB1 GLB2
R50 R51
R70
R80 R81 R82 R85
R93 R97
PGCD PSERT PSP_L SERD_L
DAHPsase DHQS DHQTi SHKK CHORS ANS ANPRT PRAIi IGPS TRPS1
CHORM PPNDH PPND
RSP 4CCoAL CoAT STS Resext
TRPext Pheext

-METINT
ATP ADP GLU_6_P FRU_6_P FRU_BIS_P DHAP GA_3P NAD NADH 3PGP 3PG
2PG AMP
Glycerol G3P DHA
RIBULOSE_5_P XYL_5_P RIBOSE_5_P SED_7_P ERYTH_4_P PYR PEP
CITRATE GL6P 6PG NADP NADPH

OXALO MALATE CoASH ACETYL_CoA FADH FAD CACO
AKG ISOCIT SUCC FUMARATE GLYOXY
SUCC_CoA NH3 CO2
KDPG
3PHP PSER
DAHP DHQ DHS SHIK S3P EPSP CHA GLN ANT GLU PRPP N5ANT 11D5P
I3GP SER TRP PPA PPY Phe HPP TYR
4CA 4CCoA Malonyl_CoA Resveratrol

-METEXT
ATP_main GLU_ext CO2_ext
NH3_ext BIOMASS OXY_ext
Resveratrol_ext
TRP_ext Phe_ext
Gly_ext

-CAT
#Glycolysis
R1 : GLU_ext + PEP = GLU_6_P + PYR
R2r : GLU_6_P = FRU_6_P
R3 : FRU_6_P + ATP = FRU_BIS_P + ADP
R4 : FRU_BIS_P = FRU_6_P
R5r : FRU_BIS_P = DHAP + GA_3P
R6r : GA_3P = DHAP
R7_1r : GA_3P + NAD = 3PGP + NADH
R7_2r : 3PGP + ADP = 3PG + ATP
R7_3r : 3PG = 2PG
R8r : 2PG = PEP
R9 : PEP + ADP = PYR + ATP
RR9 : PYR + ATP = PEP + AMP

#Glycerol uptake
glpF : Gly_ext = Glycerol
glpK : Glycerol + ATP = G3P + ADP
GLYD3 : Glycerol + NAD = DHA + NADH
GLYD4 : DHA + ATP = DHAP + ADP
GLYD5 : G3P + FAD = DHAP + FADH

#Pentose Phosphate Pathway
R10_1 : GLU_6_P + NADP = GL6P + NADPH
R10_2 : GL6P = 6PG
R10_3 : 6PG + NADP = RIBULOSE_5_P + CO2 + NADPH
R11r : RIBULOSE_5_P = XYL_5_P
R12r : RIBULOSE_5_P = RIBOSE_5_P
R13r : RIBOSE_5_P + XYL_5_P = SED_7_P + GA_3P
R14r : GA_3P + SED_7_P = ERYTH_4_P + FRU_6_P
R15r : ERYTH_4_P + XYL_5_P = GA_3P + FRU_6_P
PRPPsr : RIBOSE_5_P + ATP = PRPP + AMP

#TCA cycle
R21 : PYR + NAD + CoASH = ACETYL_CoA + CO2 + NADH
R22 : OXALO + ACETYL_CoA = CITRATE + CoASH
R23_1r : CITRATE = CACO
R23_2r : CACO = ISOCIT
R24 : ISOCIT + NADP = AKG + NADPH + CO2
R25 : AKG + NAD + CoASH = NADH + SUCC_CoA + CO2
R26r : SUCC_CoA + ADP = SUCC + ATP + CoASH
R27 : SUCC + FAD = FUMARATE + FADH
R28r : FUMARATE = MALATE
R29r : MALATE + NAD = OXALO + NADH
R30 : FUMARATE + FADH = SUCC + FAD

#Anapleurotic reactions
R40 : PEP + CO2 = OXALO
R41 : MALATE + NAD = PYR + NADH + CO2
R42 : OXALO + ATP = PEP + ADP + CO2
GLB1 : ISOCIT = GLYOXY + SUCC
GLB2 : GLYOXY + ACETYL_CoA = MALATE + CoASH

#Entner-Doudrorff pathway
R50 : 6PG = KDPG
R51 : KDPG = PYR + GA_3P

#Biomass production (see Table II for doubling time dependent reaction coefficients (doubling time = 200))
R70 : 4 GLU_6_P + 46 RIBOSE_5_P + 31 ERYTH_4_P + 156 PEP + 237 PYR + 72 ACETYL_CoA + 86 AKG + 139 OXALO + 2921 ATP + 856 NADH + 731 NH3 = BIOMASS + 72 CoASH + 2921 ADP + 856 NAD + 35 CO2

#Oxidative phosphorylation/maintenance energy:
R80 : NADH + 2 ADP + OXY_ext = NAD + 2 ATP
R81 : FADH + ADP + OXY_ext = FAD + ATP
R82 : ATP = ADP + ATP_main
R83r : NADH + FAD = NAD + FADH

#Futile cycle
R84r : NAD + NADPH = NADP + NADH
R85 : AMP + ATP = 2 ADP
GLUDr : AKG + NADPH + NH3 = GLU + NADP
GLSr : GLN = GLU + NH3

#Membrane transport reactions
R93 : NH3_ext = NH3
R97 : CO2 = CO2_ext

#Serine biosynthesis
PGCD : 3PG + NAD = 3PHP + NADH
PSERT : 3PHP + GLU = PSER + AKG
PSP_L : PSER = SER
SERD_L : SER = NH3 + PYR

#Shikimate pathway
DAHPsase : ERYTH_4_P + PEP = DAHP
DHQS : DAHP = DHQ
DHQTi : DHQ = DHS
SHK3Dr : DHS + NADPH = SHIK + NADP
SHKK : SHIK + ATP = S3P + ADP
PSCVTr : PEP + S3P = EPSP
CHORS : EPSP = CHA
ANS : CHA + GLN = ANT + GLU + PYR
ANPRT : ANT + PRPP = N5ANT

PRAIi : N5ANT = 11D5P
IGPS : 11D5P = I3GP + CO2
TRPS1 : I3GP + SER = GA_3P + TRP
TRPext : TRP = TRP_ext
CHORM : CHA = PPA
PPNDH : PPA = PPY + CO2
PHETA1r : PPY + GLU = Phe + AKG
Pheext : Phe = Phe_ext
PPND : PPA + NAD = HPP + NADH + CO2
TYRTAr : HPP + GLU = TYR + AKG

#Resveratrol biosynthesis
RSP : TYR = 4CA + NH3
4CCoAL : ATP + 4CA + CoASH = AMP + 4CCoA
CoAT : ACETYL_CoA + ATP = Malonyl_CoA + ADP
STS : 3 Malonyl_CoA + 4CCoA = 4 CoASH + Resveratrol + 4 CO2
Resext : Resveratrol = Resveratrol_ext