Development of Genetic Modification Flux with Database for Estimating Metabolic Fluxes of Genetic Mutants

(遺伝子変異株の代謝流束を推定するためのデータベース構築と

Genetic Modification Flux ソフトウェアの開発)

Noorlin binti Mohd Ali

Preface

This thesis is submitted as a partial fulfillment for the degree of doctor of philosophy. The work is done under the supervision of Professor Dr. Hiroyuki Kurata in Systems and Synthetic Biology, Metabolic Engineering and Bioinformatics research group, at Kurata Sensei’s Laboratory, Graduate School of Computer Science and System Engineering, Department of Bioscience and Bioinformatics, Kyushu Institute of Technology.

List of Publication

The thesis is based on the following publication:

1. **Noorlin Mohd Ali**, Ryo Tsuboi, Yuta Matsumoto, Daisuke Koishi, Kentaro Inoue, Kazuhiro Maeda, and Hiroyuki Kurata: Web Application for Genetic Modification Flux with Database to Estimate Metabolic Fluxes of Genetic Mutants, *J. Biosci. Bioeng.*, (2015), http://dx.doi.org/10.1016/j.jbiosc.2015.12.001
   Received: 3 September 2015; accepted 7 December 2015

2. **Noorlin Mohd Ali**, Kazuhiro Maeda, and Hiroyuki Kurata: Database of calculable metabolic networks of genetically modified mutants, *3rd International Symposium on Applied Engineering and Sciences (SAES2015)*, November 23-24, 2015 Serdang Malaysia.

3. **Noorlin Mohd Ali**, Kentaro Inoue, and Hiroyuki Kurata: Database for Predicting Metabolic Flux Distribution within a Cell, *IPSJ SIG Technical Report, BIO-35(6)*, September 20, 2013, Hokkaido Japan.

4. **Noorlin Mohd Ali**, Kentaro Inoue, Soma Tabata, and Hiroyuki Kurata: The effect of implementing objective functions in analyzing the changes of enzyme activity profiles, *IPSJ SIG Technical Report, BIO-30(8)*, August 9, 2012, Fukuoka Japan.

Abstract

In understanding the complexity of a metabolic network structure, flux distribution is the key information to observe as it holds direct representation of cellular phenotype. To examine this, the study on genetically perturbed conditions (e.g. gene deletion/knockout) is one of the useful methods, which significantly contributes to metabolic engineering and biotechnology applications. Currently, metabolic flux analysis (MFA) is proven to be suitable mechanism for specific gene knockout studies, yet the method involves exhaustive computational effort since the calculation are derived by a stoichiometric model of major intracellular reactions and applying mass balances to the intracellular metabolites.

Metabolic Flux Analysis (MFA) is widely used to investigate the metabolic fluxes of a variety of cells. MFA is based on the stoichiometric matrix of metabolic reactions and their thermodynamic constraints. The matrix is derived from a metabolic network map, where the rows and columns represent metabolites, chemical/transport reactions, respectively. MFA is very effective in understanding the mechanism of how metabolic networks generate a variety of cellular functions and in rationally planning a gene deletion/amplification strategy for strain improvements.

Flux Balance Analysis (FBA) is used to predict the steady-state flux distribution of genetically modified cells under different culture conditions. Minimization of Metabolic Adjustment (MOMA) was developed to predict the flux distributions of gene deletion mutants. FBA and MOMA often lead to incorrect predictions in situations where the constraints associated with regulation of gene expression or

activity of the gene products are dominant, because they apply the Boolean logics or its related simple logics to gene regulations and enzyme activities. On the other hand, network-based pathway analyses, elementary modes (EMs) and extreme pathways emerge as alternative ways for constructing a mathematical model of metabolic networks with gene regulations. EM analysis was suggested to be convenient for integrating an enzyme activity profile into the flux distribution. Enzyme Control Fluxes (ECFs) uses the relative enzyme activity profile of a mutant to wild type to predict its flux distribution.

In facilitating the analysis of metabolic flux distributions, the support of computational approaches is significantly essential. In addition, the availability of real sample data particularly for further observation, a large number of knockout mutant data provides assistance in enhancing the process.

We had presented Genetic Modification Flux (GMF) that predicts the flux distribution of a broad range of genetically modified mutants. The feasibility of GMF to predict the flux distribution of genetic modification mutants is validated on various metabolic network models. The prediction using GMF shows higher prediction accuracy as compared to FBA and MOMA. To enhance the feasibility and usability of GMF, we developed two versions of simulator application with metabolic network database to predict flux distribution of genetically modified mutants. 112 data sets of *Escherichia coli* (*E.coli*), *Corynebacterium glutamicum* (*C.glutamicum*), *Saccharomyces cerevisiae* (*S.cerevisiae*), and *Chinese Hamster Ovary* (*CHO*) were registered as standard models.

Table of Contents

1 CHAPTER 1: INTRODUCTION AND BACKGROUND

1.1 Systems Biology

Systems Biology is a study to describe and understand the biological systems by integration of two major disciplines: quantitative sciences and experimental biology through systematic perturbation; monitoring the systems responses from multi-layered global information in deriving analytical models [1, 2]. In a common practice, to understand a whole system-level function, the subsystem and its component interactions are importantly to be identified. As such, the main focus of systems biology is to understand on the system structures and dynamics entirely; with the understanding on molecular level remained essential. There are four (4) main properties in understanding a biological system [3]:

(i) System structures

The study on component that formed the intracellular and extracellular structure of a biological network system; which included gene interactions and its associated biochemical pathways

(ii) System dynamics

The study on system responses under different conditions through metabolic analysis, sensitivity analysis or dynamic analysis; and identifying the mechanisms to achieve particular responses.

(iii) The control method

The study on methods to control a cell state, reduce malfunctions and identify prospective targets for diseases treatment.

(iv) The design method

The study on the approaches for desired properties of biological systems through design principles and simulations; as an alternative of exhaustive trial-and-error method.

Interestingly, the major challenge in biology systems is the nature of its multi-layered structures: genome (DNA), transcriptome (nRNA, sRNA, miRNA), proteome and interactome (proteins) and metabolome and fluxome (metabolites and fluxes). To date, this informative yet tedious process is supported by the breakthrough of multidisciplinary in quantitative sciences: mathematical or computational, genomics, measurement technologies and the integration of these disciplines; with the support of comprehensive database from existing knowledge.

1.2 Computational Systems Biology

The challenge to understand biological systems as systems able to achieve by combining computational, system analysis, updated technologies that support quantitative measurements, and high-throughput quantitative experimental data [4]. Figure 1 summarizes a basic cycle of systems biology research.

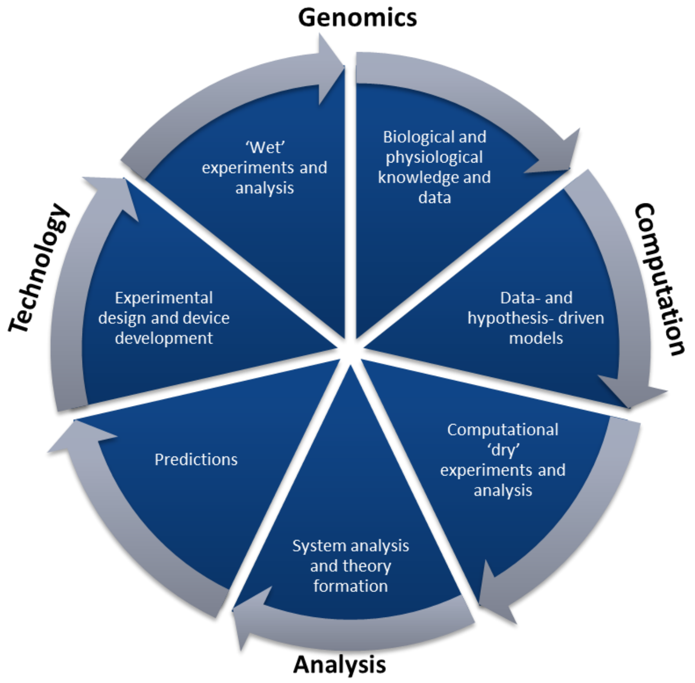

Genomics
'Wet' experiments and analysis
Biological and physiological knowledge and data
Computation
Data- and hypothesis- driven models
Computational 'dry' experiments and analysis
System analysis and theory formation
Analysis
Predictions
Technology
Experimental design and device development

Figure 1. The hypothesis-driven research cycle of systems biology

In realizing the objective for systems-level analysis, a comprehensive set of quantitative data is one of the essential components. It is necessarily to support simulation-based research where in-depth simulation with thorough exploratory and sufficient coverage is conducted before a validated hypothesis can be derived.

1.3 The Molecular Biology Database

The progression in systems biology is strengthening by the development of various molecular biology databases. Many specialized databases are developed as the main goal is to be more accessible to biologists. The early development of biological databases was towards sequence-based data e.g. nucleic-acid and amino-acid sequences, further the interest focuses on other types of molecular data, while the recent development emphasized to genetic disease and drugs. These included (i) nucleic acid sequence and structure, transcriptional regulation (GenBank, EMBL Nucleotide Sequence Database, DNA Bank of Japan); (ii) protein sequence and structure, motifs and domains, protein-protein interactions (GenProtEC, Protein Information Resource (PIR); (iii) metabolic and signaling pathways (Kyoto Encyclopedia of Genes and Genomes (KEGG), EcoCyc, ENZYME), metabolites, enzymes, protein modification ; (iv) viruses, bacteria, protozoa and fungi; (v) human genome, model organisms, comparative genomics (Human Gene Mutation Database (HMGD); (vi) genomic variation, diseases and drugs (EcoGene, GOBASE); (vii) plant databases and (viii) other molecular biology databases [2, 5-7]. Tables 1-5 list several examples of specialized molecular databases.

Table 1. Example of Primary Nucleotide Sequence database [5]

Database name	Database URL	Brief description
GenBank	http://www.ncbi.nlm.nih.gov/genbank/	All known nucleotide and protein sequences; International Nucleotide Sequence Database Collaboration
EMBL Nucleotide Sequence Database	http://www.ebi.ac.uk/	All known nucleotide and protein sequences; International Nucleotide Sequence Database Collaboration
DNA Data Bank of Japan (DDBJ)	http://www.ddbj.nig.ac.jp/	All known nucleotide and protein sequences; International Nucleotide Sequence Database Collaboration

Table 2. The example of Gene Expression database [5]

Database name	Database URL	Brief description
Gene Expression Database (GXD)	http://www.informatics.jax.org/	Mouse gene expression and genomics
Kidney Development Database	http://golgi.ana.ed.ac.uk/kidhome.html	Kidney development and gene expression
FlyBase	http://flybase.org/	A Database of Drosophila Genes & Genomes

Table 3. The example of Metabolic Pathways and Cellular Regulation database [5]

Database name	Database URL	Brief description
Kyoto Encyclopedia of Genes and Genomes (KEGG)	http://www.genome.jp/kegg/	Metabolic and regulatory pathways
EcoCyc	http://ecocyc.org/	*Escherichia coli* K-12 genome, gene products and metabolic pathways
ENZYME	http://enzyme.expasy.org/	Enzyme nomenclature

Table 4. The example of new online databases in the 2016 NAR Database issue [7]

Database name	Database URL	Brief description
AgingChart	http://www.agingchart.org/	Pathways of age-related processes
BreCAN-DB	http://brecandb.igib.res.in/	Breakpoint profiles of cancer genomes
MutationAligner	http://www.mutationaligner.org/	Mutation hotspots in protein domains in cancer

Table 5. The example of most recently published elsewhere databases in [7]

Database name	Database URL	Brief description
BiGG Models	http://bigg.ucsd.edu/	Biochemically, genetically and genomically structured metabolic network models
DGIdb	http://dgidb.genome.wustl.edu/	Drug-gene interaction database
iPPI-DB	http://www.ippidb.cdithem.fr/	Inhibitors of protein-protein interactions

In representing the qualitative data, a network model is required. The purpose of building a network model is for network dynamic analysis as well; however it is importantly to consider a model-based for experimental or simulation purposes, with high accuracy prediction performance, where the resources can be ideally distributed. The detailed description on reconstructing a genome scale metabolic network model is presented in **Section 1.9** in this chapter.

Another critical component of systems biology research is computer software support, which may varies by providing simulation software as a platform for modeling and analysis. The support should be open platform environment that commonly accepted in accordance to the emergence of online biological databases. Another concern of software support is to increase the development of common infrastructure that able to integrate the existing resources.

1.4 Metabolic Engineering

Metabolic engineering is the study to manipulate and modify metabolism with DNA recombination for the production of useful metabolites [8]. One of the novel aspects of metabolic engineering as compared to genetic engineering and other typical strain improvement technologies is the study on integrated metabolic pathways. In essence, this study includes the complete chains of biochemical reaction network, with associated issues of pathway synthesis and thermodynamics feasibility, and metabolic fluxes and their controls.

In examining a metabolic network and its pathway, gene expression levels, proteins and metabolites concentration provide some information. However, the interaction of these cellular phenotypes is manifested through metabolic fluxes. As such, fluxes are considered as the critical parameter to represent the fundamental basis of cellular phenotype and its corresponding pathways.

1.5 Systematically Perturbation of Biology Systems

Perturbation in biological systems is an approach to comprehend the complexity of cellular systems. This is performed by modifying the function of a biological system externally or internally; particularly done by genetic conditions (gene deletion, gene overexpression, undirected mutations) or environmental conditions (growth condition changes, temperature or hormone/drug stimuli) [2].

The responses from the modification process are monitored; further this hypothesis is validated to the experimental data set. Once validated, this will contribute as a new knowledge to systems biology. The study on perturbation is one of significant strategies to extract the information from complex structure of cellular system, this approach as well beneficial to describe gene relationships, identify drug responses, and determine the gene function (e.g. gene deletion) [9]. In general purposes, these quantitative observations provide valuable support for metabolic engineering and biotechnology applications.

The study on flux distribution under knockout condition becomes one of major interest, where the main purpose is to investigate the general and detailed responses of metabolic and regulatory network [10]. In the example of *E.coli* knockouts, the previous studies showed a significant contributions such as; discovering a novel hidden reaction in pentose phosphate pathway from double knockouts [11], monitoring the oxygen sensing and aerobic regulatory response by the combination of genetic and environmental perturbations [12-14], describing the regulations and dynamic of network pathway [15]. Table 6 summarized the reported publications of *E.coli* gene knockout studies. It is also recorded that the most studied knockouts were on the central carbon metabolism, global regulation and under substrate-rich conditions (e.g. batch) or substrate-limited conditions (e.g. continuous cultures) [10].

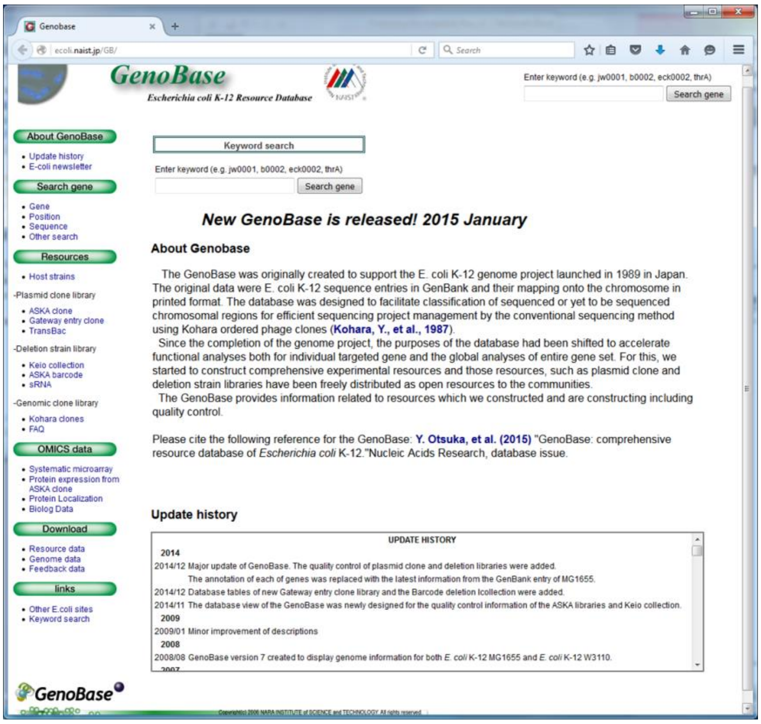
Table 6. The overview on *E.coli* knockout strains using 13C metabolic flux analysis studies [10]

***E.coli* knockout gene**	**Reference for ^{13}C metabolic flux analysis study**				
Central carbon metabolism	(micro-) aerobic		anaerobic		**Other growth conditions**
	Batch	Continuous	Batch	Continuous	
ptsG/crr	-	[16]	-	-	-
galM	-	[17]	-	-	-
glk	-	[17]	-	-	-
pgi	[18-23]	-	-	-	NH_4^+ limitation [24]; growth on galactose [22]
pgm	-	[17]	-	-	-
pfkA/pfkB	[23]	[17]	-	-	-
fbp	-	[17]	-	-	-
fbaA/fbaB	-	[17]	-	-	-
tpiA	[19]	-	-	-	-
gapAC	-	[17]	-	-	-
pgk	-	-	-	-	-
gpmA/gpmB	-	[17]	-	-	-
eno	[21]	-	-	-	-
pykA/pykF	[18, 23, 25]	[15, 17, 26, 27]	-	-	NH_4^+ limitation [27]
aceE/aceF	-	-	-	-	-
lpd	-	[28]	-	-	-
pflB/tdcE	-	-	-	-	-
zwf	[22, 23, 29]	[17, 24, 30, 31]	-	-	NH_4^+ limit [24]; growth on pyr [30] and ac [31]
pgl	-	[17]	-	-	-
gnd	-	[17, 30, 32]	-	-	Growth on pyruvate [30]
rpiA/rpiB	-	[17]	-	-	-
rpe	-	[17]	-	-	-
tktA/tktB	-	[17]	-	-	-
talA/talB	[11]	[17]	-	-	-
edd	-	-	-	-	-
eda	-	-	-	-	-
gltA	-	-	-	-	-
prpC	-	-	-	-	-
acnA/acnB	-	-	-	-	-
icd	-	-	-	-	-

sucA/sucB	-	[33]	-	-	-
sucC/sucD	-	[33]	-	-	-
sdhABCD	[22, 23]	-	-	-	Growth on galactose [23]
frdABCD	-	-	-	-	-
fumABC	[23]	-	-	-	-
mdh	[23]	-	-	-	-
aceA	-	-	-	-	-
aceB	-	-	-	-	-
ppc	[19, 25]	-	-	-	-
pck	[23]	[34]	-	-	-
maeA/maeB	[23]	-	-	-	-
ppsA	-	[17]	-	-	-
pta	[19, 25]	-	-	-	-
ackA	-	-	-	-	-
mgsA	-	[16]	-	-	-
Regulatory genes					
arcA	[22, 35]	[13, 36-39]	[35]	-	Nitrate as electron acceptor [35, 39]
arcB	[35]	[13]	-	-	-
cra	[35]	-	-	-	-
crp	[35]	[16]	-	-	-
cya	[35]	-	-	-	-
fnr	[35]	[37]	-	-	-
mlc	[35]	[16]	-	-	-
iclR	-	[36, 38]	-	-	-
fur	[22]	-	-	-	Growth on galactose [22]
pdhR	[22]	-	-	-	Growth on galactose [22]
ihfA	[22]	-	-	-	Growth on galactose [22]
ihfB	[22]	-	-	-	Growth on galactose [22]

1.5.1 GenoBase, the Single Knockout Mutant Database

The most related project of gene knockout database is GenoBase (http://ecoli.naist.jp/) [40]. The main purpose of GenoBase is to support the *E.coli* K-12 genome project launched in Japan in the year of 1989. This database was developed (1) to facilitate sequence classification towards efficient sequencing project management using Kohara-ordered phage clone and (2) to facilitate genome annotation. The main focus of GenoBase is to comprehensively construct the experimental resources and high-throughput data of large *E.coli* functional genomics. The resources of this database are recently included: (1) the plasmid clone libraries (i.e. ASKA ORFeome libraries) and (2) The single-gene deletion collection (i.e. Keio collection).



New GenoBase is released! 2015 January

About Genobase

The GenoBase was originally created to support the E. coli K-12 genome project launched in 1989 in Japan. The original data were E. coli K-12 sequence entries in GenBank and their mapping onto the chromosome in printed format. The database was designed to facilitate classification of sequenced or yet to be sequenced chromosomal regions for efficient sequencing project management by the conventional sequencing method using Kohara ordered phage clones (**Kohara, Y., et al., 1987**).

Since the completion of the genome project, the purposes of the database had been shifted to accelerate functional analyses both for individual targeted gene and the global analyses of entire gene set. For this, we started to construct comprehensive experimental resources and those resources, such as plasmid clone and deletion strain libraries have been freely distributed as open resources to the communities.

The GenoBase provides information related to resources which we constructed and are constructing including quality control.

Please cite the following reference for the GenoBase: **Y. Otsuka, et al. (2015)** "GenoBase: comprehensive resource database of *Escherichia coli* K-12."Nucleic Acids Research, database issue.

Update history

UPDATE HISTORY

2014

2014/12 Major update of GenoBase. The quality control of plasmid clone and deletion libraries were added.
The annotation of each of genes was replaced with the latest information from the GenBank entry of MG1655.

2014/12 Database tables of new Gateway entry clone library and the Barcode deletion Icollection were added.

2014/11 The database view of the GenoBase was newly designed for the quality control information of the ASKA libraries and Keio collection.

2009

2009/01 Minor improvement of descriptions

2008

2008/08 GenoBase version 7 created to display genome information for both *E. coli* K-12 MG1655 and *E. coli* K-12 W3110.

2007

Figure 2. The main page of GenoBase, the *E.coli* single gene knockout database

1.6 Computational Model

Enormous works have been done in implementing computational method to analyze metabolic flux under perturbed condition, which generally classified as network-based pathway and constraint-based flux analysis [1, 41, 42].

Constraint-based metabolic network analysis is also known as optimization-based analysis does not required detailed knowledge to predict feasible flux distributions. The main idea of this method is by imposing constraints (objective functions) and linear optimization techniques that likely represent cellular metabolism to desired growth condition.

Flux Balance Analysis (FBA) is one of optimization-based approaches that have been extensively used to predict metabolic fluxes by maximizing the growth rates. This method is based on convex analysis; by using an objective function, with subject to several constraints for example maximize cellular growth rates, substrate uptake rates, and/or product secretion rates, thermodynamic constraints, metabolic regulation or others.

FBA is able to perform estimation tasks with limited number of experimental data, yet the more fluxes is provided, the more accurate fluxes can be estimated.

The accuracy of FBA approach and objective functions has been proven in predicting fluxes. However, this accuracy is influenced by the used of suitable objective functions and valid cofactor assumptions. In addition, the use evolution-based objective function is questionable when genetically perturbed strains are unevolved. Furthermore, FBA is restricted to singular stoichiometric matrix model, which affected to estimate fluxes with recycled, bidirectional, and parallel types. It is notably that FBA determines only one optimal solution despite choices of optimal solution are available.

Minimization of Metabolic Adjustment (MOMA) is another optimization-based approach in predicting flux distributions. The basis of this approach is Euclidean distance, where MOMA proposes that mutant types should be very close to wild types, with minimal metabolic changes. This approach is implemented using quadratic programming (QP) optimization method. However the concept can be inconsistent with regulatory adaptation cost and flow linearity principles.

On the other hand, Regulatory On/Off Minimization (ROOM) is developed to overcome the inconsistency in MOMA by minimizing the total number of major flux changes from the wild type strains that satisfies FBA solution [43]. The assumptions underlying by ROOM is (i) the regulatory adaption cost is minimized by genetic regulatory changes that essential for flux changes are minimized by the cell, (ii) each regulatory changes is assigned by a fixed cost regardless its magnitude. Both MOMA and ROOM estimate the flux distribution that closed to wild type strains and not relied on to maximizing the growth rate.

RELATCH (RELATive CHAnge) is an approach that based on the relative optimality of relative flux changes. This approach uses experimental flux and gene expression data to estimate the flux distribution; suggests the assumptions that the perturbed strains will minimize the relative metabolic changes within a limited regulatory adaptation that further will increase the flux capacity of previously active pathways [44].

Another approach to predict the flux distribution is network-based pathway analysis. Metabolic Pathway Analysis has emerged as a main method in analyzing the structure and function of metabolic network. As compared to optimization-based flux analysis, metabolic pathway analysis is able to recognize a complete fluxes solution from a metabolic network without any cellular objective bias are provided. The associated techniques implies for metabolic pathway analysis are elementary mode analysis and extreme pathway analysis [1].

To quantitatively analyze the cellular phenotypes, Metabolic Flux Analysis (MFA) becomes an emerged alternative technology and one of central importance to metabolic engineering [8, 45]. The formation of MFA is based on mass balances of internal metabolites at the steady state assumption. MFA is derived by a stoichiometric matrix that describes the cellular metabolism, which is formed based on a metabolic network model [46, 47].

1.7 The Theory to Analyzing Metabolic Network

1.7.1 Stoichiometric model

A stoichiometric model column is based on the transportation reaction; represented by non-zero values that identify the metabolites involve in the reaction and the stoichiometric coefficients correspond to each metabolite. The rows contain with zero represents the non-participation of corresponding metabolites. The matrix also denotes the directionality, where substrate and product metabolites are having negative (-) and positive (+) coefficients respectively. A standard stoichiometric matrix denotes as S and defined as:

$$\begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1n} \\ S_{21} & S_{22} & \cdots & S_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ S_{m1} & S_{m2} & \cdots & S_{mn} \end{bmatrix}$$

A stoichiometric matrix of $m \times n$ represents a metabolic network with *m* internal metabolites and *n* reactions. To describe the mass conservation of metabolites in a system, the general equation is defined as:

$$\frac{d}{dt} C_1 = \sum_{i=1}^{n} S_{li} v_i \quad \text{for } l = 1, \ldots., \text{m} \tag{1}$$

where C_1 denotes the concentrations of the *l*-th metabolite a network.

The element of *l*-th row and *i*-th column of S represents the amount of *l*-th metabolite consumed or produced by *i*-th reaction. The flux values of all *n* reactions is represented as flux vector denoted as $v_{n\times 1}$. A metabolic network may contain irreversible reaction (s), where the flux must be non-negative. With the consideration of thermodynamic, additional constraints need to be added as:

$$v_i \geq 0, \tag{2}$$

where $i \in \text{irrev}$ are the indices of the irreversible reactions.

1.7.2 Elementary mode analysis

Elementary Mode (EM) analysis is one of mathematical-related approaches to represent fundamental ‘interaction’ routes in biochemical networks [46]. It is often defined as a minimum set of sub-networks (associated enzymes) that enabled a metabolic system to operate at a steady state, through all irreversible reactions [48, 49]. It is used to recognize a metabolic network structure by involving all possible pathways for a group of enzymes that cannot further decomposed. At the steady state, the mass balance equation is given as:

$$\mathrm{S} \cdot \mathrm{v} = 0, \tag{3}$$

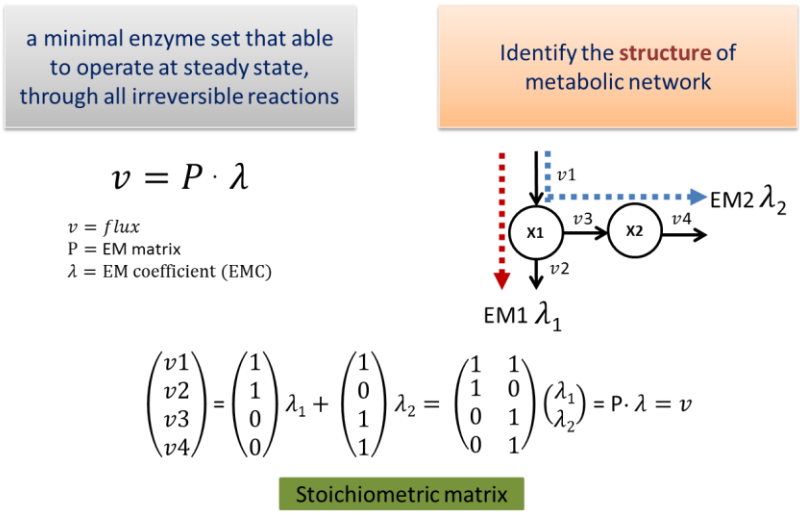
where $\mathrm{v} = (v_1, v_2, \ldots, v_n)^t$ is the vector of reaction flux rate and n is the number of reactions. The set of vectors are determined from all possible solutions of the equation in (1). Some reactions are irreversible and additional constraints on positive flux values are required as in equation (2). From equation (1), EM needs to fulfill the constraints in (2) and non-decomposability constraints.

To represent the EM matrix **P**, it is determined using the stoichiometric matrix and the flux vector as:

$$\mathrm{v} = \mathrm{P} \cdot \lambda, \tag{4}$$

where $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{ne})^t$ is the vector of EM coefficient and ne is the number of EMs.

An example to transform a simple network is summarized in Figure 3.



a minimal enzyme set that able to operate at steady state, through all irreversible reactions

Identify the structure of metabolic network

$$v = P \cdot \lambda$$

$v = flux$
P = EM matrix
λ = EM coefficient (EMC)

$$\begin{pmatrix} v1 \\ v2 \\ v3 \\ v4 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \end{pmatrix} \lambda_1 + \begin{pmatrix} 1 \\ 0 \\ 1 \\ 1 \end{pmatrix} \lambda_2 = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} = \mathrm{P} \cdot \lambda = v$$

Stoichiometric matrix

Figure 3. Example of transforming a simple metabolic network to EM. The pathway through X1 is considered as EM1, and the pathway from X1 to X2 is considered as EM2. The fluxes involved in EM1 are represented as ‘1’ while ‘0’ for non-involvement. From this network, it will be then transform into the equation that is represented in stoichiometric matrix

Most of metabolism models are classified as underdetermined [50], where the number of determined EM is more than the fluxes data. This situation occurs since only a few constraints are available. The solution to overcome this problem is by providing more constraints until an optimized coefficient is achieved. To add more constraints, implementing objective function is one of the solutions. The use of an objective function is as an optimizer element that maximizes the targeted cell growth, energy or metabolite synthesis [51].

EM analysis enables us to identify unique pathway from a complex metabolic network and to calculate all solutions from a flux space. Therefore, EM analysis is considered as powerful tool to recognize the structure a metabolic network. In addition this tool is also potentially effective for integrating transcriptome or proteome data into metabolic network, which further provides the mechanism of how phenotypic or metabolic flux distributions change with respect to environmental and genetic perturbations [52].

1.7.3 Extreme pathway analysis

Extreme pathway analysis is closely related concept to elementary modes, yet all reactions are controlled by the flux direction [1, 48, 49]. In extreme pathway analysis, the reversible reactions are separated into two irreversible reactions; i.e. forward and backward directions, as compared to elementary modes that allows for reversible reactions. The solution set derived by extreme pathway is a subset of elementary modes and it is systematically independent. Therefore, extreme pathway analysis is implies based on the additional conditions of (1) network reconfiguration, and (2) systematic independence. Figure 4 illustrates the difference between elementary modes and extreme pathway analysis.

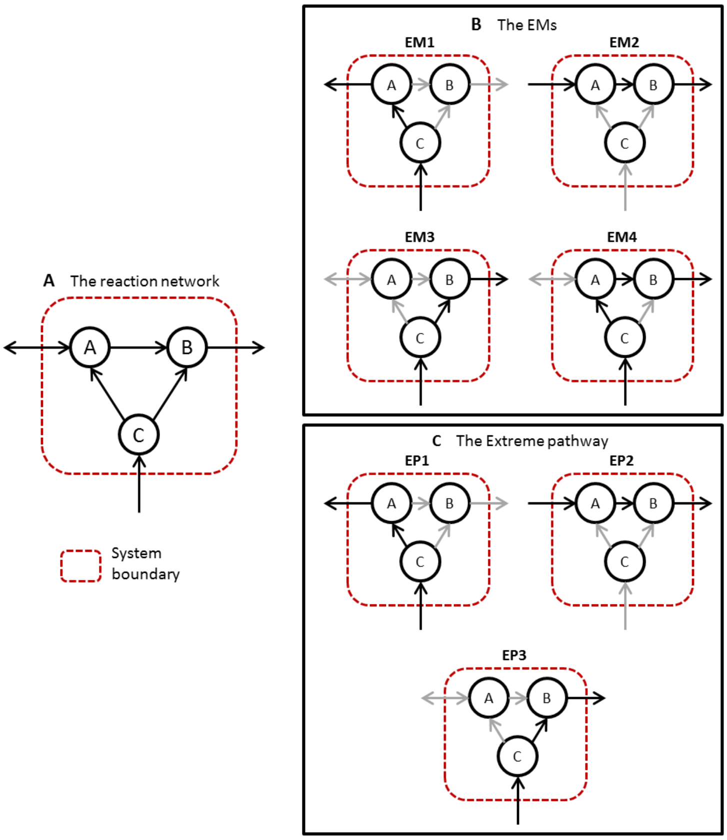


Figure 4. The difference between elementary modes (EMs) and extreme pathway analysis.

1.7.4 Application Programming Interface (API) for EM analysis

Calculating EM requires highly effort and resources, especially when large metabolic network is involved. A number of APIs to calculate elementary mode are publicly available, with some earlier versions such as METATOOL [53], GEPASI/COPASI [54], and FluxAnalyzer [55]. The APIs is mainly developed using C language, yet FluxAnalyzer developed on MATLAB environment (The Mathworks, Inc., USA) with a user friendly interface and advances features to analyze metabolic network.

Recent enhancements in APIs development had made the ability to calculating larger metabolic networks with other advances analyses. The upgraded version of METATOOL has incorporated the Null Space algorithm with an efficient rank test to check the mode elementarily, available in either C language or MATLAB. The newer version of FluxAnalyzer, CellNetAnalyzer (CNA) is further improved by implementing binary approach that able to decrease 96% of memory consumption. CNA also provides signal transduction pathways analysis.

Table 7: A list of available EM analysis API

API name	Tool	URL Reference
CellNetAnalyzer (CNA)	Matlab	http://www2.mpi-magdeburg.mpg.de/projects/cna/cna.html
ScrumPy	Python	http://mudshark.brookes.ac.uk/ScrumPy
Gepasi	C/C++ Ms Windows Program	http://www.gepasi.org/
efmtool	Java (integrated into Matlab)	http://www.csb.ethz.ch/tools/efmtool
Metatool	C	http://pinguin.biologie.uni-jena.de/bioinformatik/networks/

1.8 Application Programs for Estimating Metabolic Fluxes, Gene Knockout Study

User friendly computer applications in MFA are exist with different functions to improve the analysis tasks. OpenFLUX is a software application for small and large scale ^{13}C metabolic flux analysis [56]. The application is developed based on the new Elementary Metabolite Unit (EMU) framework which comprises two main modules

(1) to automate metabolic models construction or to modify user-entered reaction data and (2) to calculate fluxes from experimental data, with statistical flux analysis option. OpenFLUX implemented gradient-based minimization search function (FMINCON) in MATLAB Optimization Toolbox to estimate flux parameter and execute the sensitivity analysis.

Various works in estimating flux distributions in **perturbed conditions** were proposed. The works that applied different quantitative and mechanistic of mathematical and computational methods such as differential equation based models [57], cybernetic models [58], and combination of regulatory and metabolic models [59]. However, the works continuity towards well-developed software/computational applications tool to improve the analysis and quantitative understanding is not yet implemented [10].

1.9 Computational Metabolic Network Model

In understanding a network structure and for further analysis, a metabolic network model is required. The process of building a model should be started by considering the purpose of a model; either for the comprehensive understanding of system behavior or prediction of complex simulation purposes. This consideration is important as it will define the model scope and level of abstraction [4]. Each model intuitively develops for a purpose based on the requirement [1]:

(a) Good data fitting

The objective of this model is to describe each data point individually using a general mathematical function, which applies to dynamic modeling. A good data fitting model will have a well definition between the parameters and data curves.

(b) Good prediction

If the main requirement of a model is to obtain good prediction accuracy, a model to build is supposed able to emphasis general relationships among major quantities. This is important for future interpretations when new data set are tested to the model.

(c) Biological comprehensive

The main objective of comprehensive or mechanistic model is should be able to describe the actuality. In biological practice, this kind of model will focus on certain part of cells only up to the traceable level, with supporting simplified assumptions

(d) Key principles

A key principles model should only highlight the fundamental properties that represent a biological process, thus it is needs to be very simple. This kind of model is appropriate for experimental model systems.

The process of reconstructing a genome-scale metabolic model generally involves the steps of: (1) create a draft model; (2) reconstruct a detailed model; (3) convert into a mathematical format; (4) identify and filling the gaps; and (5) simulation and visualization [60]. The output of this process is known as genome-scale models (GEMs). GEMs are defined as a structured knowledge-based; which constructed on a combination of genome sequence and detailed biochemical information. This model is used to perform computational and quantitative queries to answer various questions on the capabilities of an organism. The process used for reconstructing the metabolic model is generally described in Figure 5.

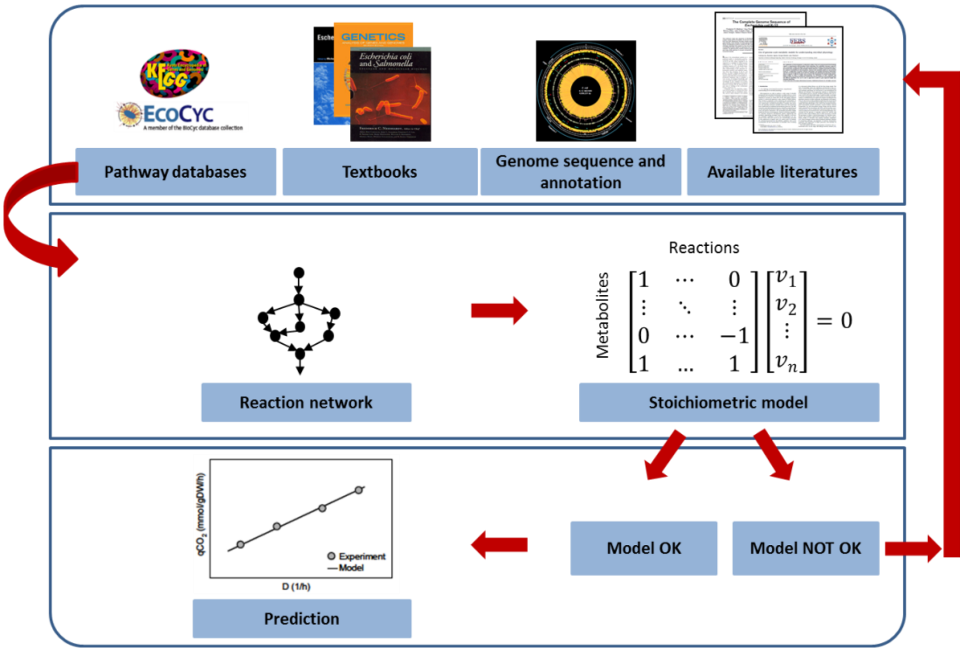

KEGG
EcoCyc
A member of the BioCyc database collection
GENETICS
Escherichia coli and Salmonella
Pathway databases
Textbooks
Genome sequence and annotation
Available literatures
Reactions
Metabolites
$\begin{bmatrix}1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & -1\\ 1 & \cdots & 1\end{bmatrix}\begin{bmatrix}v_1\\ v_2\\ \vdots\\ v_n\end{bmatrix} = 0$
Reaction network
Stoichiometric model
qCO$_2$ (mmol/gDW/h)
D (1/h)
Experiment
Model
Prediction
Model OK
Model NOT OK

Figure 5. The process of reconstructing a metabolic network

The first draft of metabolic pathways is starts by identifying the coding sequence and functional annotation of particular genes. This process will only include the gene encoded for building the enzymes or membrane transporters that will be used in the model. The functional annotation of enzymes needs to be translated into biochemical reactions that will build a chain of complete metabolic network model. To accomplish this task, the information are available from the genome sequence annotation [61], biochemical pathway databases [62, 63] , related textbooks and publications.

To identify enzyme-catalyzed reactions, EC numbers is beneficial to directly match between EC numbers and reactions in various databases. The identified genes and its given EC numbers is compared and matched to biochemical reaction databases, e.g.

KEGG, Biocyc, or to the registered metabolic network models e.g. BiGG database. The important information that needs to carefully check is: (1) metabolites and co-factors; (2) each metabolites chemical formula; (3) metabolites identifiers; (4) reaction stoichiometry and directionality (reversible or irreversible); (5) gene and reaction localization; (6) reaction identifier; and (7) metabolic subsystems.

An essential yet challenging phase in reconstructing a GEM model is to add the reactions that are not concluded in the genome annotation, such as: (1) spontaneous reactions; (2) extracellular transport reactions; (3) intracellular transport reactions; and (4) exchange reactions. Adding some new reactions will minimize the dead-end metabolites and improve the network connectivity.

The biomass reaction is another set of required reaction that needs to consider while building the model. Within an equation, this reaction described all biomass components, information on energy requirement (e.g. ATP molecules), maintenance (e.g. turgor pressure) and their contributions to the cell growth. This information normally derived from the literature studies or experimentally determined.

In the third phase, the metabolic network model that comprises of reactions list is converted in a constraint-based mathematical format (e.g. stoichiometric model) to analyze its structural properties. Since the model is a representing living cell, constraints need to be applied for better approximation of flux solution space. To set the boundary of cellular functions, there are four types of constraints: physiochemical, topological, environmental conditions and regulatory constraints. For the scope of

GEM, the constraint normally used is physiochemical and/or environmental conditions: flux balance ($S \cdot \mathrm{v} = 0$), energy balance ($\Delta E = 0$), enzyme or transporter capacity ($v_i \leq v_{max}$) and thermodynamics ($0 \leq v_{min}$).

From this stage, it will further need to be verified and evaluated. The first process of verification is by checking the model consistency, identify metabolic gaps and examine the catabolized process of different substrates into different metabolites. In this step, new reaction may need to be included to fill the metabolic gaps and the problem of dead-end metabolites.

The remaining step is to test the model for prediction by comparing to experimental data, in which the prediction will be the basis if the model needs for further refining until a desired model is achieved.

1.10 The Purpose of Study

1.10.1 The arising problems

Exploring knockout fluxes is potentially significant, however due to the lack of coverage in different experimental conditions and methodology has leads to the difficulties for further analysis and generalizing the results [10].

It is interestingly to note that, the cellular responses of both conditions are significantly different. From the observation on *E.coli*, the data set of *zwf* knockout strains grown under continuous conditions [30] is compared to batch condition [29]; it is recorded that the acetate flux grown under continuous conditions was 29 and citrate synthase flux was 87, while the acetate flux was 44, and citrate synthase flux was 51 in a batch conditions.

In addition, varies flux distributions were reported for the same knockout strains and growth condition. As for the example, fumarate synthase flux was 71 [30] to 109 [17] for *gnd* knockout of dilution rates 0.2h^{-1} under continuous conditions. There is also major difference on flux distribution of *pgi*, *pyk* and *ppc* genes of *pykF* knockout at dilution rates of 0.1h^{-1} and 0.2h^{-1} under continuous conditions [10].

By considering the potential and current situation, it would be valuable to provide analytical platform to help biologist to access, analyze and interpret the information.

Based on our laboratory research progression, we designed Genetic Modification Flux (GMF) to predict flux distribution of a broad range of genetically modified mutants with under-expressed/over-expressed genes [42, 52, 64] in previous work.

The feasibility of GMF to predict the flux distribution of genetic modification mutants is validated on various metabolic network models of *E.coli*, *S.cerevisae*, and *C.glutamicum*, *Bacillus Subtilis* (*B.subtilis*), and *CHO* [52, 64, 65]. The performance of GMF is compared to FBA and MOMA. The prediction using GMF shows higher prediction accuracy as compared to FBA and MOMA when tested on experimental data set of *E.coli* gene deletion mutants [64]. The applicability of

GMF in estimating the flux distribution is also proven on over- and under-expressed mutants; which is a promising strategy for enhanced production of genetically strains. The detail of GMF algorithms is described in Section 1.12 of this chapter.

Despite the usefulness of ECF and GMF, there have been no user-friendly applications programs are developed as reported by [10]. Use of them had required handling computer programs, which often hampers the general and broad use.

Furthermore, the analysis requires real experimental data; particularly for further observation a large number of knockout mutant data becomes necessary. The current experimental data are not presented in simulation-ready format. The large-scale metabolic network models are available in many public databases; however refinement processes are required to limit the boundary of a network. Reconstructing a metabolic network for computer simulation purposes normally contain blocked reaction problem, due to dead end metabolites and missing metabolites and/or reactions.

1.10.2 The research target

With consideration of the stated problems from the current situations in both progression: (1) the study on flux distribution under knockout condition in general, and (2) the progression research in our laboratory; we aim to develop Genetic Modification Flux (GMF), a user-friendly web application together with the database of metabolic networks that helps users accessing metabolic network data

[10]. In achieving the above, we initiate a metabolic network database by collecting a variety of experimental data of different microorganisms.

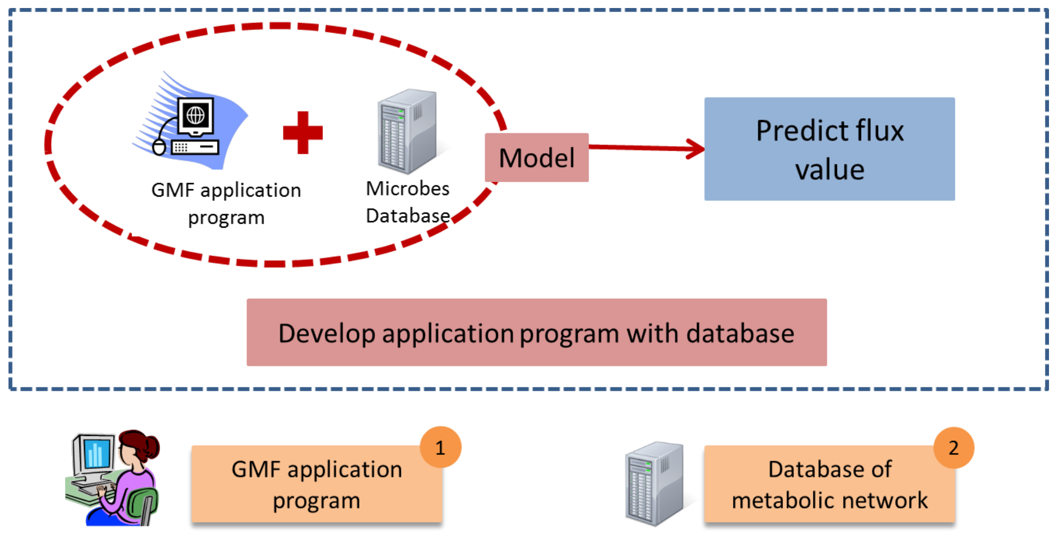


Figure 6. The diagram representing the research direction

1.11 Genetic Modification of Flux (GMF)

GMF is an EM-based method, integrates enzyme activity profiles i.e. gene expression or enzyme activity data to predict the flux distributions. This algorithm is consists of two other algorithms: (1) modified Control Effective Flux (mCEF) and (2) Enzyme Control Flux (ECF).

1.11.1 modified Control Effective Flux (mCEF)

mCEF is an algorithm derived from the Control Effective Flux (CEF), which estimates the relative expression ratios of metabolic genes of a mutant to wild type from changes in target gene expression.

1.11.2 Control Effective Flux (CEF)

The main function of CEF algorithm is to estimate the changes in transcriptional regulations when the substrates changes. This estimation is based on a metabolic network topology with specified biological reactions [48, 66]. For each cellular objective, $\varepsilon_{j,CELLOBJ}$, the efficiency of the *j*-th EM is defined as the ratio of EM output (reaction that involving the objectives) to the necessary investment to form each EM (the total of absolute elements in EM):

$$\varepsilon_{j,CELLOBJ} = \frac{P_{CELLOBJ,j}}{\sum_i |P_{i,j}|} \qquad (5)$$

where $P_{i,j}$ is the normalized element of the *i*-th reaction in the *j*-th EM and $CELLOBJ$ is the reaction number of specified biological function (biomass production and ATP generation). CEF of the *i*-th reaction, which is associated to the flux of *i*-th reaction, is indicated by the total weight of the *i*-th elements from all EMs based on the efficiency $\varepsilon_{j,CELLOBJ}$:

$$cef_i = \sum_{CELLOBJ} \frac{1}{P_{CELLOBJ}^{max}} \frac{\sum_j (\varepsilon_{j,CELLOBJ} \cdot |P_{i,j}|)}{\sum_j \varepsilon_{j,CELLOBJ}} \qquad (6)$$

where $P_{CELLOBJ}^{max}$ is the maximum element in the row of biological functions.

The transcript ratio principle for *i*-th reaction under different substrate conditions, S_1 and S_2, is given by:

$$\Theta_i(S_1, S_2) = \frac{cef_i(S_2)}{cef_i(S_1)} \qquad (7)$$

For genetic mutants that over-, under-expressed, or lack of metabolic gene, the original CEF algorithm is modified [64] by the efficiency of the *j*-th EM for such a genetic mutant is defined by:

$$\varepsilon^{m}_{j,CELLOBJ} = \frac{P_{CELLOBJ,j} \cdot EA_j}{\Sigma_i(|P_{i,j}| \cdot \eta_i)} \tag{8}$$

$$\eta_i = \begin{cases} EAP_i \text{ (if reaction } i \text{ is modified)} \\ 1 \text{ (if reaction } i \text{ is not modified)} \end{cases}$$

where EAP_i is the enzyme activity parameter (i.e. relative gene expression or enzyme activity) responsible for the i-th reaction of a mutant to wild type. EAP_i is set as 0 if the gene of i-th reaction is deleted; it is set as more than 1 ($EAP_i > 1$) for over-expressed and less than 1 ($EAP_i < 1$) if it is under-expressed condition. η_i is the correcting factor to compute the investment for genetic mutants. EA_j is the correcting factor which includes the change in the modified reaction into each EM's output, by:

$$EA_j = \prod_{i=1}^{n} ge_{i,j} \tag{9}$$

$$ge_{i,j} = \begin{cases} EAP_i & \text{if } P_{i,j} \neq 0 \\ 1 & \text{if } P_{i,j} = 0 \end{cases}$$

where $ge_{i,j}$ is the parameter indicating the gene expression state of the i-th reaction in the j-th EM. The state is computed by the numerator in Equation (8), where it will increase or decrease, if a gene within an EM is over-expressed or under-expressed respectively. As $EAP_i = 0$, the containing EM is ignored$\left(\varepsilon^{m}_{j,CELLOBJ} = 0\right)$, which is consistent with EM analysis of gene deletion mutants. For $EAP_i = 1$, in which the gene expressions do not affected by any changes, the Equation (8) is consistent with Equation (5). Both equations are the efficiency of genetic mutants, yet Equation (8) is extended of Equation (5).

The mCEF for the mutant is defined by:

$$mCEF_i(mut) = \sum_{CELLOBJ} \frac{1}{P^{max}_{CELLOBJ}} \frac{\sum_j \left(\varepsilon^m_{j,CELLOBJ} \cdot |P_{i,j}| \cdot \eta_i \right)}{\sum_j \varepsilon^m_{j,CELLOBJ}} \quad (10)$$

where η_i indicates the weight of associated elements for each EM.

The calculation of mCEF for wild type is resembles from the original CEF:

$$mCEF_i(w) = \sum_{CELLOBJ} \frac{1}{P^{max}_{CELLOBJ}} \frac{\sum_j \left(\varepsilon^m_{j,CELLOBJ} \cdot |P_{i,j}| \right)}{\sum_j \varepsilon_{j,CELLOBJ}} \quad (11)$$

Therefore the relative change in a gene expression profile of a mutant type to wild type is derived from the Equation (7), which is:

$$\Theta_i(w, mut) = \frac{mCEF_i(mut)}{mCEF_i(w)} \quad (12)$$

1.11.3 Enzyme Control Flux (ECF)

ECF is an EM-based algorithm, to estimate the correlation between enzyme activity profiles and its associated flux distribution based on the EMs [51]. ECF is very effective in the case that an enzyme activity profile is provided. The principle of ECF defines that the changes in enzyme activities for both wild type and mutant type are correlated to the changes in the EMCs. The principle is presented by the power-law formula. The feasibility of ECF in estimating flux distribution of mutants by integrating the enzyme activity profiles were validated in *E.coli* and *B. subtilis* model [51].

The estimation process is performed by calculating the EMCs of wild type

$\lambda^{wt} = (\lambda_1^{wt}, \lambda_2^{wt}, \ldots, \lambda_m^{wt})^t$ using quadratic programming [67, 68] from the flux distribution of wild type by:

$$\min \textstyle\sum_j \left(\lambda_j^{wt}\right)^2$$

$$\text{subject to } \mathrm{P} \cdot \lambda^{wt} = v \qquad (13)$$

$$\lambda_j^{wt} \geq 0$$

Further, the EMCs of a mutant are defined by:

$$\lambda_j^{mut} = \beta \cdot \lambda_j^{wt} \textstyle\prod_{i=1}^{n} a_{i,j} \qquad (14)$$

$$a_{i,j} = \begin{cases} a_i & \text{if } P_{i,j} \neq 0 \\ 1 & \text{if } P_{i,j} = 0 \end{cases}$$

where $\lambda^{mut} = (\lambda_1^{mut}, \lambda_2^{mut}, \ldots, \lambda_m^{mut})^t$, $a_{i,j}$ is the relative enzyme activity for the i-th reaction in the j-th EM of a mutant type to wild type, $a_{i,j}$is the enzyme activity ratio for the i-th reaction of the mutant type to wild type. β is the factor used to normalize λ^{mut}, therefore the substrate uptake flux is the same as wild type. The flux distribution of the mutant type is given by:

$$v^{mut} = P \cdot \lambda^{mut} \qquad (15)$$

1.11.4 Genetic Modification of Flux (GMF)

GMF predicts the flux distribution of genetically modified mutants; gene knockout mutants, over-expressed or under-expressed genes using the topological structures of metabolic networks [64]. The flow algorithm of GMF is illustrated in Figure 7:

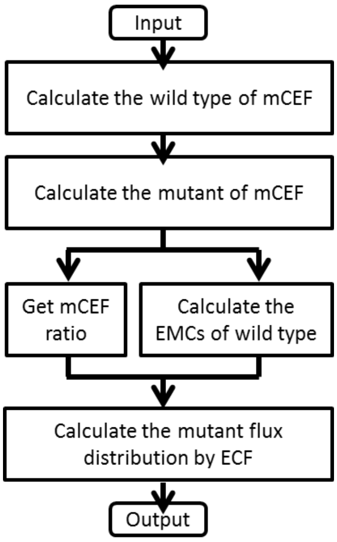

Input
Calculate the wild type of mCEF
Calculate the mutant of mCEF
Get mCEF ratio
Calculate the EMCs of wild type
Calculate the mutant flux distribution by ECF
Output

Figure 7. The flow of GMF algorithm

By the assumption that there is linear correlation between a gene expression and its associated enzyme activity profiles, the EMCs of a mutant can be estimated based on the flux distribution of wild type using quadratic programming as in Equation (13). This is supported by the existence of quantitative correlation between mRNA expression and protein levels in some studies [2, 15]. Since the enzyme activity ratios is possible to be substituted using the CEF ratios, the EMCs for the mutant is derived by the Equation (14):

$$\lambda_j^{mut} = \beta \cdot \lambda_j^{wt} \prod_{i=1}^{n} \theta_i(wt, mut) \qquad (16)$$

Therefore, the predicted flux distribution is given as:

$$v^{mut} = P \cdot \lambda^{mut} \qquad (17)$$

1.11.5 Objective functions

The EM coefficients (EMCs) must be estimated by using an objective function to calculate flux distributions. Estimation of the EMCs is an underdetermined problem [50, 69], because the number of EMs is much more than the experimental flux data. GMF is implemented with four types of objective functions; Linear Programming (LP), Quadratic Programming (QP) [67], Linear Programming based on alpha spectrum (MeanLP) [51], Maximum Entropy Principle (MEP) [52].

Table 8: The objective functions in GMF application

Method	Description
LP	To maximize biomass or specific metabolite formation.
QP	To optimize emc by defining minimal norm of emc
MeanLP	To optimize emc by calculating the mean (average) from maximizing and minimizing each emc
MEP	To optimize emc by derivation of Shannon’s theory and Lagrange Multipliers (LM).

1.12 The Thesis Organization

The thesis is organized in the following structure:

Chapter 1 reviews a brief introduction and background of the study that covers: Systems Biology, Computational Systems Biology, The Molecular Biology Database, Metabolic Engineering, Systematically Perturbation of Biology System, Computational Model, the theory in analyzing a metabolic network, the process in reconstructing a metabolic network model and the direction of this research works.

Chapter 2 describes the materials and methods used and implemented to achieve the targeted objectives. The process on reconstructing metabolic network models, preparing the metabolic network input files and the GMF algorithm are described in this chapter.

Chapter 4 discusses the results and outcomes obtained from the study, which presents the details on database collection and the GMF prediction performance.

Chapter 5 concludes this study and discusses the advantages, contributions and the gap and potential for the future research.

2 CHAPTER 2: MATERIALS AND METHODS

2.1 Systems overview

Figure 8 shows a workflow of the web application of GMF. Metabolic reaction network files written in the Microsoft Excel format are registered in the database attached to the GMF web application. These files can be freely downloaded. Users either select a registered or uploaded user's own data file. The application reads the selected or uploaded file and generates its associated stoichiometric matrix with the format available for the efmtool [70]. Users can select one algorithm out of the three: GMF, mCEF or ECF to predict the flux distribution of genetic mutants. To perform GMF and ECF, they select one of the four objective functions and specify a ratio type of gene or enzyme. The calculated result is displayed and can be downloaded.

In addition, we have developed the stand-alone version of GMF application that functions on the MATLAB (The MathWorks). The main workflow of the GMF stand-alone version is the relatively same as the web version.

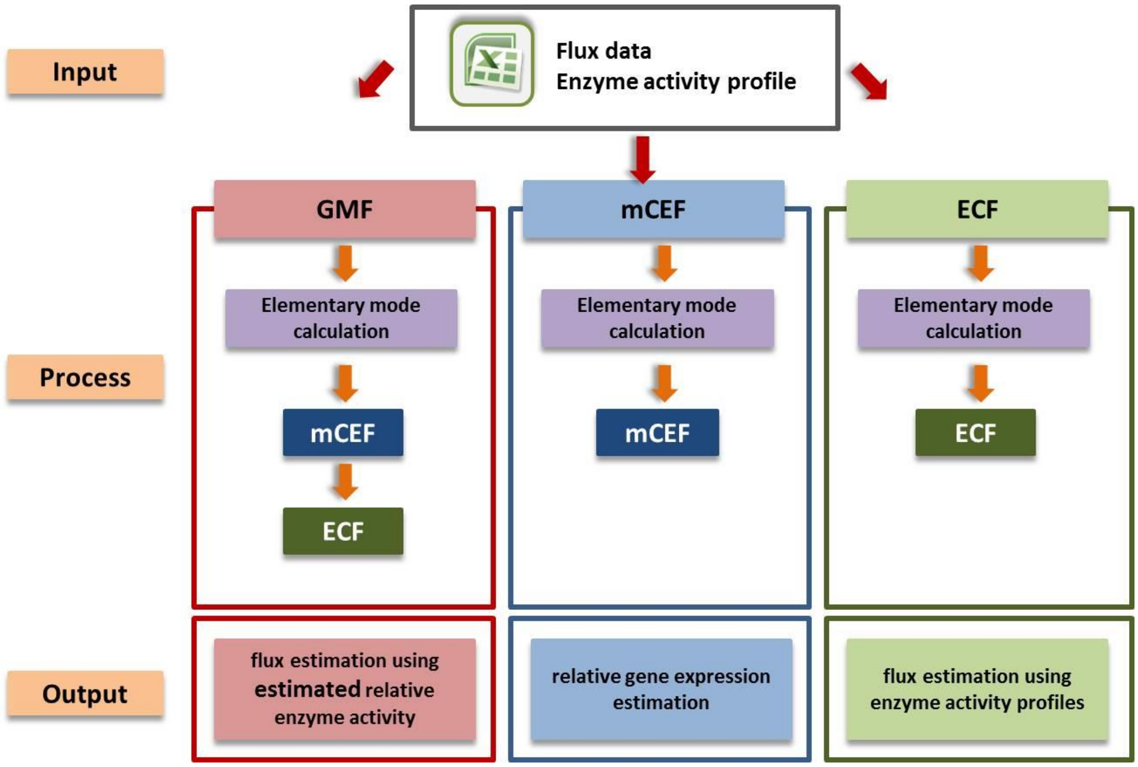

Input
Flux data
Enzyme activity profile
Process
GMF
Elementary mode calculation
mCEF
ECF
mCEF
Elementary mode calculation
mCEF
ECF
Elementary mode calculation
ECF
Output
flux estimation using estimated relative enzyme activity
relative gene expression estimation
flux estimation using enzyme activity profiles

Figure 8. The main workflow of GMF application. A metabolic network file written in the Microsoft Excel format is put in the application. To perform GMF and ECF, users need to select an objective function out of Maximum Entropy Principle (MEP), Linear Programming based on alpha spectrum (MeanLP), Linear Programming (LP), and Quadratic Programming (QP).

2.2 The Gene Knockout Database

2.2.1 Preparing the metabolic input files for GMF

To prepare the input files which represented in a metabolic network model begin with the process to reconstruct a metabolic network. As presented in Chapter 1 (Computational Metabolic Network Model), the metabolic network model employed in GMF is constructed based on the described phases.

Beforehand the reconstructing task begins; the main purpose of building the model is defined. Based on the priority to create a knockout gene experimental database, our focus is mainly on central carbon metabolic pathways, since these pathways are considered as the bottle neck of metabolic systems. In addition, with the consideration that the model will be used as computer-executed model and able to estimate various types of experimental conditions (e.g. batch or continuous conditions), we defined a small scale metabolic model, that purposely for prediction task; where experimental flux data are available [71] and its significant applications [10].

Building a computer-executed metabolic model that will be used as an input file for computer simulation application, the key value of a computer application should be considered is the system usability. From the time-consumption point of view, analyzing a metabolic network depends on the network size; the larger metabolic network will need a longer time for analysis, and produce higher number of feasible solutions. With the limitations of high-end machine and its memory capacity (i.e. super computers) to conduct the simulation task and time-consuming (i.e. user does not prefer to wait longer) [72], building a model that focused on central carbon metabolism would be most appropriate.

The reconstructed model is designed based on a comprehensive literature from varies sources such as online pathway databases, biochemistry textbooks, functional annotation genome sequence and information extraction from published journals. A series of academic discussion was also conducted among the experts (i.e. Professors, postdoctoral personnel) in completing the reconstruction phase.

We defined the functional annotation of genes based on gene catalog from KEGG metabolic pathways databases (http://www.genome.jp/kegg/). The information is further organized in central carbon metabolic pathways: glycolysis, pentose phosphate, entner-doudoroff, pyruvate metabolism, and TCA cycle. The details for Open Reading Frame (ORFs) name, gene name, enzyme name, EC numbers and KEGG metabolic chart were used to reconstruct the metabolic network.

To create a particular reaction list, the reaction stoichiometry was referred from several online databases, such as: KEGG, Biocyc (http://biocyc.org/) and BiGG database (http://bigg.ucsd.edu/). In many databases, the information such as cofactors utilization is not yet been completely clarified, as for example either a reaction only require to include NADH or NADPH as a cofactor; or might involve both cofactors. In such cases, two reactions were included in the reconstructed metabolic network.

In relation to the gene and reaction localization, all reactions were localized in cytosol as most of central carbon metabolism takes place in this compartment. The information directionality of reactions (reversible or irreversible) was extracted from pathway databases or registered metabolic network sample models.

The outcome of reconstruction process is a set of associated biochemical reactions that might be used in constructing the stoichiometric models metabolism using metabolite balancing [8, 73]. This model basically depends on mass balance principle on metabolic intermediates and allow for steady state behavior. Further, based on the information on reactions stoichiometry, localization, and reversibility, the biomass composition needs to be defined. Table 9 shows the biomass composition of *E.coli* model [74].

The *E.coli* reconstructed model comprises 48 reactions that are most frequently encountered pathways: glycolysis (11 reactions), pentose phosphate (7 reactions), Entner-Doudoroff (ED) (1 reaction), TCA Cycle (8 reactions), pyruvate metabolism (2 reactions), anaplerotic reactions (5 reactions), energy/redox metabolism (5 reactions), transport reactions (3 reactions) and exchange reactions (5 reactions).

Table 9. Biomass composition for *E.coli* metabolic model from [74]

Metabolite	**mmole g DW**	**Metabolite**	**mmole g DW**
Fructose 6 phosphate	0.1	Nicotinamide adenine dinucleotide phosphate-reduced	18.2
3 Phosphoglycerate	1.5	Glyceraldehyde 3 phosphate	0.1
Acetyl coenzyme A	3.7	Nicotinamide adenine dinucleotide	3.5
Glucose 6 phosphate	0.2	Pyruvate	2.8
Adenosine triphosphate	41.3	Phosphate	41.3
Phosphoenolpyruvate	0.5	Coenzyme A	3.7
alpha Ketoglutarate	1.1	Adenosine diphosphate	41.3
Erythrose 4 phosphate	0.4	Nicotinamide adenine dinucleotide - reduced	3.5
Ribose 5 phosphate	0.9	Nicotinamide adenine dinucleotide phosphate	18.2
Oxaloacetate	1.8	Carbon dioxide	1.68

2.2.2 The reconstructed metabolic network for *E.coli*

Figure 9 shows the employed metabolic network model for *E.coli* experimental data and the associated enzymes and metabolites are described in **Tables 10-11**. The characteristics and reaction distribution to its associate pathways of this model are summarized in **Tables 12-13** respectively.

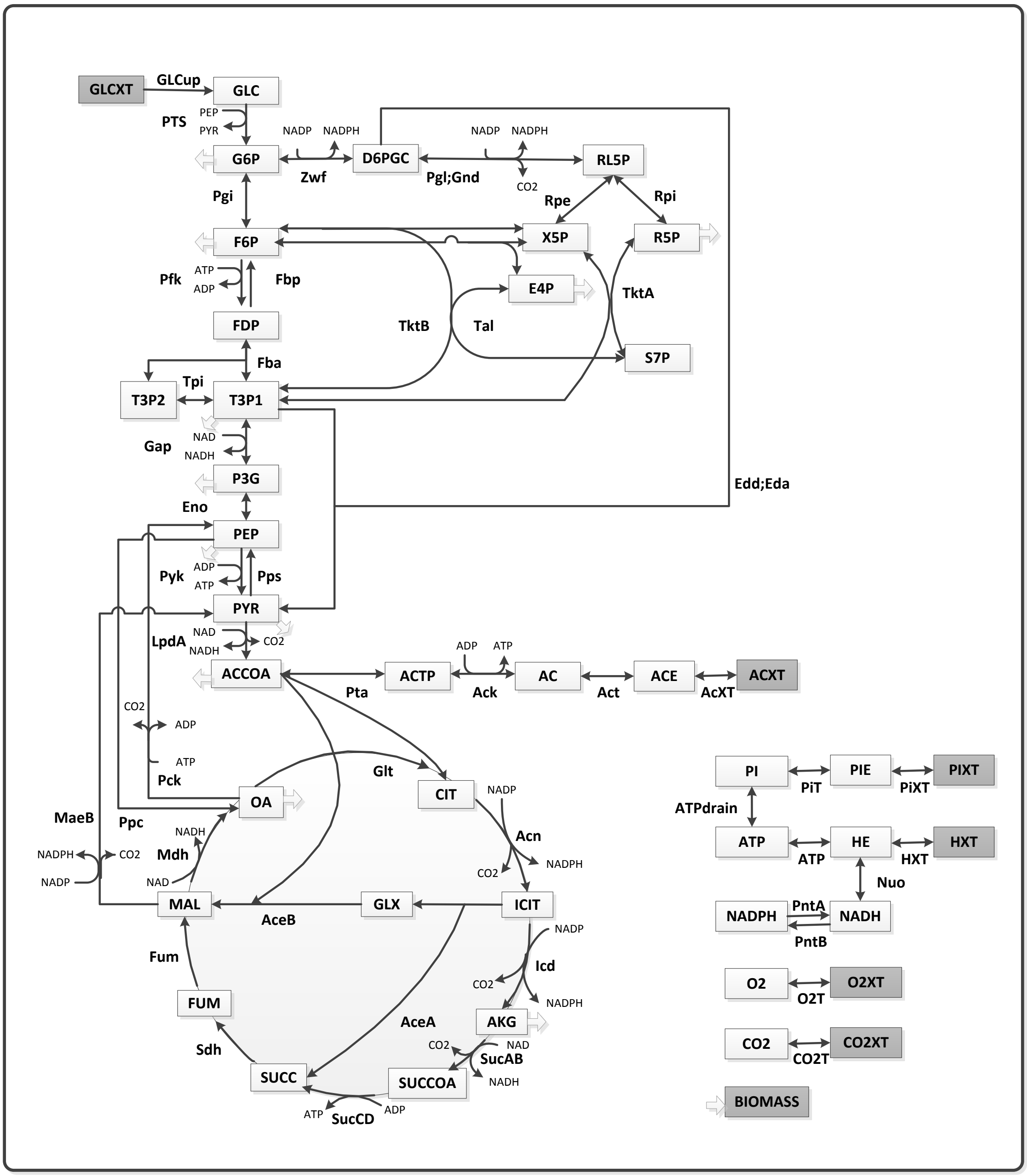


Figure 9: The employed *Escherichia coli* metabolic network map

Table 10: The employed *Escherichia coli* metabolic model reactions

Pathway	**Enzyme catalyzing**	**Gene**	**Reaction**
Glycolysis	PTS	*pts*	PEP + GLC --> G6P + PYR
	Pgi	*pgi*	G6P <--> F6P
	Pfk	*pfkA,B*	F6P + ATP --> ADP + FDP
	Fbp	*fbp*	FDP --> F6P + PI
	Fba	*fba*	FDP <--> T3P2 + T3P1
	Tpi	*tpi*	T3P2 <--> T3P1
	GAPDH	*gapA,C*	PI + T3P1 + NAD <--> P3G + NADH
	Eno	*Eno*	P3G <--> PEP
	Pyk	*pykF,A*	PEP + ADP --> ATP + PYR
	Pdh	*lpdA*	COA + NAD + PYR --> ACCOA + CO2 + NADH
	Pps	*ppsA*	ATP + PYR --> PI + PEP
Pentose Phosphate	G6PDH	*zwf*	G6P + NADP --> NADPH + D6PGC
	6PGDH	*pgl; gnd*	D6PGC + NADP --> RL5P + CO2 + NADPH
	Rpi	*rpiAB*	RL5P <--> R5P
	Rpe	*rpe*	RL5P <--> X5P
	Tkt1	*tktA*	X5P + R5P <--> S7P + T3P1
	Tal	*tal*	S7P + T3P1 <--> F6P + E4P
	Tkt2	*tktB*	X5P + E4P <--> F6P + T3P1
Entner-Doudoroff	KDPG	*edd;eda*	D6PGC --> T3P1 + PYR
Pyruvate Metabolism	Pta	*pta*	ACCOA + PI <--> ACTP + COA
	Ack	*ackA*	ACTP + ADP <--> ATP + AC
TCA cycle	CS	*gltA*	ACCOA + OA <--> COA + CIT
	Acn	*acn*	CIT <--> ICIT
	ICDH	*icd*	ICIT + NADP <--> AKG + CO2 + NADPH
	aKGDH	*sucAB*	AKG + COA + NAD --> CO2 + SUCCOA + NADH
	SCS	*sucCD*	PI + ADP + SUCCOA --> ATP + SUCC + COA
	SDH	*sdhABCD*	SUCC --> FUM

	Fum	*fumABC*	FUM <--> MAL
	MDH	*mdh*	MAL + NAD <--> OA + NADH
Anapleurotic Reactions	Ppc	*ppc*	PEP + CO2 --> PI + OA
	Pck	*pckA*	ATP + OA --> PEP + CO2 + ADP
	Mez	*maeB*	MAL + NADP --> CO2 + NADPH + PYR
	Icl	*aceA*	ICIT --> SUCC + GLX
	MS	*aceB*	ACCOA + GLX --> MAL + COA
Energy/Redox Metabolism	ATP	*atp*	PI + 4 HE + ADP <--> ATP + 3 H
	ATPDr	*atpdrain*	ATP --> PI + ADP
	NUO	*nuo*	NADH + 3 H + O2 --> 4 HE + NAD
	PNTA	*pntA*	NADPH + NAD --> NADH + NADP
	PNTB	*pntB*	NADH + NADP --> NADPH + NAD
Transport Reactions	ACt	*act*	AC + H --> ACE + HE
	PIt	*pit*	PIE + HE <--> PI + H
	CO2t	*co2t*	CO2 --> CO2XT
Exchange Reactions	GLCUP	*glcup*	GLCXT --> GLC
	ACxt	*acxt*	ACE <--> ACXT
	Hxt	*hxt*	HE <--> HXT
	PIxt	*pixt*	PIXT <--> PIE
	O2xt	*o2xt*	O2XT <--> O2
Biomass Reaction	Growth	*growth*	0.1 F6P + 1.5 P3G + 3.7 ACCOA + 0.2 G6P + 41.3 ATP + 0.5 PEP + 1.1 AKG + 0.4 E4P + 18.2 NADPH + 1.8 OA + 0.9 R5P + 0.1 T3P1 + 3.5 NAD + 2.8 PYR --> 41.3 PI + 3.7 COA + 41.3 ADP + 3.5 NADH + 18.2 NADP + 1.677 CO2 + 1 BIOMASS

The metabolic reactions are based on Figure 9.

“-->” represents irreversible reaction; “<-->” represents reversible reaction.

Table 11: The employed *E.coli* metabolites

Abbreviations	**Full name**	**Abbreviations**	**Full name**
AC	Acetate	ICIT	Isocitrate
ACCOA	Acetyl coenzyme A	KetoPGluc	2 Keto 3 desoxy 6 phospho gluconate
ACE	Medium Acetate	MAL	Malate
ACTP	Acetyl phosphate	NAD	Nicotinamide adenine dinucleotide
ACXT	External Acetate	NADH	Nicotinamide adenine dinucleotide - reduced
ADP	Adenosine diphosphate	NADP	Nicotinamide adenine dinucleotide phosphate
AKG	Alpha Ketoglutarate	NADPH	Nicotinamide adenine dinucleotide phosphate-reduced
ATP	Adenosine triphosphate	O2	Oxygen
BIOMASS	Biomass	O2XT	External Oxygen
CIT	Citrate	OA	Oxaloacetate
CO2	Carbon dioxide	P3G	3 Phosphoglycerate
CO2XT	External Carbon Dioxide	PEP	Phosphoenolpyruvate
COA	Coenzyme A	PI	Phosphate
D6PGC	6 Phospho D gluconate	PIE	Medium Phosphate
E4P	Erythrose 4 phosphate	PIXT	External Phosphate
F6P	Fructose 6 phosphate	PYR	Pyurvate
FDP	2 6 bisphosphate	R5P	Ribose 5 phosphate
FUM	Fumarate	RL5P	Ribulose 5 phosphate
G6P	Glucose 6 phosphate	S7P	Sedoheptulose 7 phosphate
GLC	Glucose	SUCC	Succinate
GLCXT	External glucose	SUCCOA	Succinyl coenzyme A
GLX	Glyoxylate	T3P1	Glyceraldehyde 3 phosphate
H	Proton	T3P2	Dihydroxyacetate phosphate
HE	Medium Proton	X5P	Xylulose 5 phosphate
HXT	External Proton		

The metabolites are based on Figure 9

Table 12. Network characteristic of reconstructed metabolic network of *E.coli*

Metabolites (total)		49
Cytosolic metabolites	39	
Transportation metabolites	3	
Extracellular metabolites	7	
Reactions (total)		48
Cytosolic reactions	39	
Exchange fluxes	5	
Transportation reactions	3	
Growth reactions	1	

Table 13. Distribution of reactions for *E.coli* reconstructed metabolic network

Reactions (total)		48
Glycolysis pathway reactions	11	
Pentose Phosphate reactions	7	
Entner-Doudoroff reactions	1	
Pyruvate metabolism reactions	2	
TCA cycle reactions	8	
Anapleurotic Reactions	5	
Energy/Redox Metabolism	5	
Transport Reactions	3	
Exchange Reactions	5	

The mass balance equations are based on the metabolic network in Figure 9:

GLC: $v_{GLCup} - v_{PTS} = 0$ (18)

G6P: $v_{PTS} - v_{Pgi} - v_{Zwf} - 0.2 \times v_{Biomass} = 0$ (19)

F6P: $v_{Pgi} + v_{TktB} + v_{Tal} - v_{Pfk} - 0.1 \times v_{Biomass} = 0$ (20)

FDP: $v_{Pfk} - v_{Fba} = 0$ (21)

T3P1: $v_{Fba} + v_{TktB} + v_{TktA} - v_{Gap} - 0.1 \times v_{Biomass} = 0$ (22)

T3P2: $v_{Fba} - v_{Tpi} = 0$ (23)

P3G: $v_{Gap} + v_{Eno} - 1.5 \times v_{Biomass} = 0$ (24)

PEP: $v_{Eno} + v_{Pck} + v_{Pps} - v_{Pyk} - v_{Ppc} - 0.5 \times v_{Biomass} = 0$ (25)

PYR: $v_{Pyk} + v_{maeB} + v_{Eda} - v_{LpdA} - v_{Pps} - 2.8 \times v_{Biomass} = 0$ (26)

ACCOA: $v_{LpdA} - v_{Glt} - v_{Pta} - 3.7 \times v_{Biomass} = 0$ (27)

CIT: $v_{Glt} + v_{Pta} - v_{Acn} = 0$ (28)

ICIT: $v_{Acn} - v_{Icd} - v_{AceA} = 0$ (29)

AKG: $v_{Icd} - v_{SucAB} - v_{Biomass} = 0$ (30)

SUCCOA: $v_{SucAB} - v_{SucCD} = 0$ (31)

SUCC: $v_{SucCD} - v_{AceA} - v_{Sdh} = 0$ (32)

FUM: $v_{Sdh} - v_{Fum} = 0$ (33)

MAL: $v_{Fum} + v_{AceB} - v_{Mdh} - v_{MaeB} = 0$ (34)

OA: $v_{Mdh} + v_{Ppc} + v_{Pck} - v_{Pyk} - v_{Glt} - 1.8 \times v_{Biomass} = 0$ (35)

ACTP: $v_{Pta} - v_{Ack} = 0$ (36)

AC: $v_{Ack} - v_{Act} = 0$ (37)

ACE: $v_{Act} - v_{AcXT} = 0$ (38)

D6PGC: $v_{Zwf} - v_{Gnd} - V_{Edd} = 0$ (39)

RL5P: $v_{Gnd} - v_{Rpe} - v_{Rpi} = 0$ (40)

X5P: $v_{Rpe} + v_{Tal} - v_{TktA} - v_{TktB} - v_{Tal} - 0.9v_{Biomass} = 0$ (41)

R5P: $v_{Rpi} - v_{TktA} - v_{Biomass} = 0$ (42)

S7P: $v_{TktA} - v_{Tal} = 0$ (43)

E4P: $v_{Tal} - v_{TktB} - 0.4 \times v_{Biomass} = 0$ (44)

CO2: $v_{CO2T} - v_{LpdA} - v_{Icd} - v_{SucAB} - v_{Gnd} - v_{MaeB} - v_{Pck} - 1.68 \times v_{Biomass} = 0$ (45)

O2: $v_{O2T} - v_{Nuo} = 0$ (46)

NADPH: $v_{PntA} - v_{PntB} - v_{Acn} - v_{Icd} - v_{Zwf} - v_{Gnd} - 18.2 \times v_{Biomass} = 0$ (47)

NADH: $v_{PntB} - v_{Gap} - v_{LpdA} - v_{SucAB} - v_{Mdh} - v_{MaeB} - 3.5 \times v_{Biomass} = 0$ (48)

NAD: $v_{Gap} + v_{LpdA} + v_{SucAB} + v_{Mdh} + v_{PntA} - v_{Nuo} - v_{PntB} - 3.5 \times v_{Biomass} = 0$ (49)

NADP: $v_{Zwf} + v_{Gnd} + v_{Icd} + v_{MaeB} + v_{PntB} - v_{PntA} - 18.2 \times v_{Biomass} = 0$ (50)

COA: $v_{LpdA} + v_{SucAB} - v_{Pta} - v_{Glt} - v_{SucCD} - v_{AceB} - 3.7 \times v_{Biomass} = 0$ (51)

ATP: $v_{Pfk} + v_{ATPDrain} + v_{Pck} - v_{ATP} - v_{Pyk} - v_{SucCD} - v_{Ack} - 41.3 \times v_{Biomass} = 0$ (52)

ADP: $v_{Pyk} + v_{ATP} + v_{succCD} + v_{Ack} - v_{Pfk} - v_{Pck} - v_{ATPDrain} - 41.3 \times v_{Biomass} = 0$ (53)

PI: $v_{Gap} + v_{Pta} + v_{succCD} + v_{ATP} - v_{Fbp} - v_{Ppc} - v_{Pps} - v_{PiT} - v_{ATPDrain} - 41.3 \times v_{Biomass} = 0$ (54)

PIE: $v_{PiT} - v_{PiXT} = 0$ (55)

H: $v_{Act} + v_{Nuo} - v_{PiT} - v_{ATP} = 0$ (56)

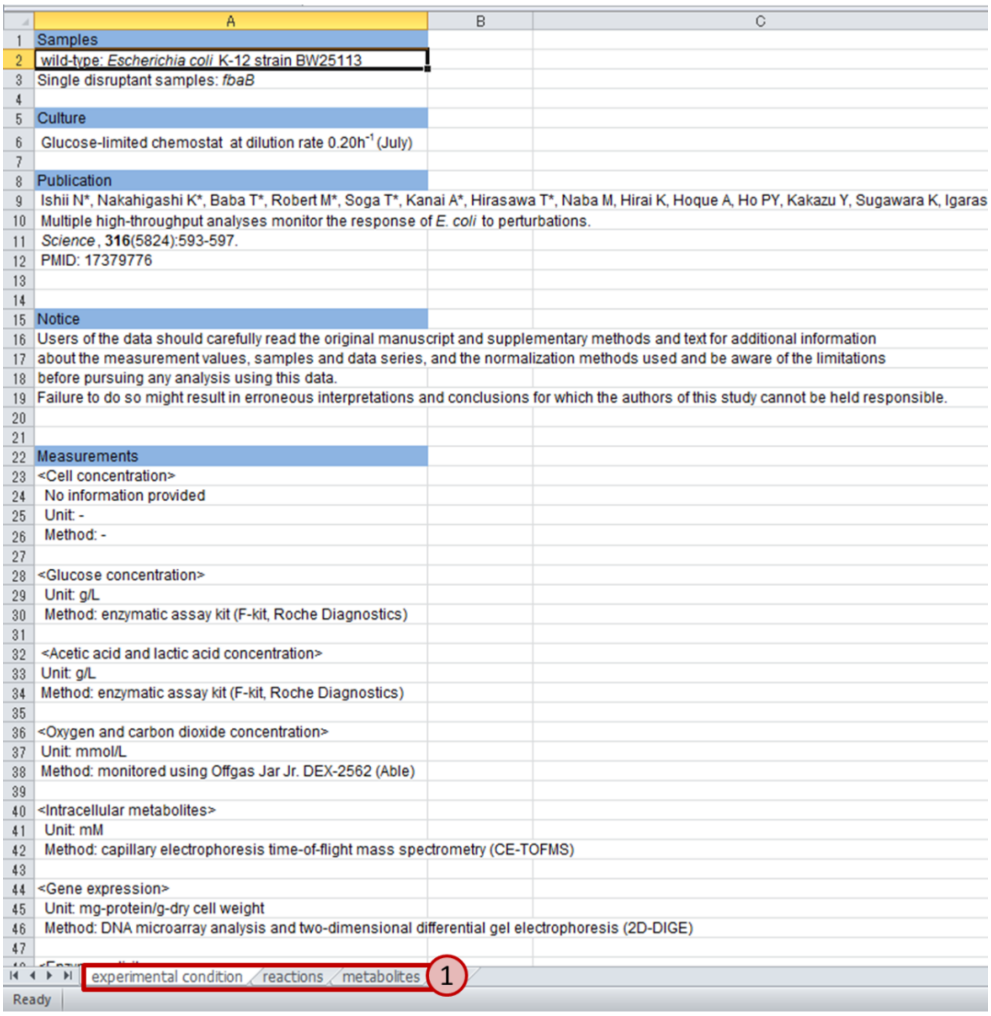
HE: $v_{HEXT} + v_{PiT} + v_{ATP} - v_{Act} - v_{Nuo} = 0$ (57)

2.2.3 Preparing the metabolic network files

The GMF application is equipped with a database as the input file. All data files are written in Microsoft Excel format, with three sheets; (1) experimental condition, (2) reactions, and (3) metabolites as shown in **Figures 10-12** respectively. Sheet 1 (experimental condition), which contains the experiment information, including the author and title of original publication from where the data were extracted. Sheet 2 (reactions) provides the metabolic reactions and their associated flux distributions, gene expressions, and enzyme activities with their experimental values for the reference (e.g., wild type) and target (e.g., mutant type) cells. Enzyme activity or gene expression distributions are used for the ECF algorithm. The Sheet 3 (metabolites) lists the corresponding metabolites with their experimental concentration values for both the reference and target cells. The internal or external index is added. Details on the input file setting are described in **Tables 14-16**.

The reconstructed metabolic network model (as described in Section 2.2.2) will be put in '*reactions*' sheet, in column A – D, which represents its corresponding model, i.e.

enzyme name (column A), gene name (column B), reaction formula list (column C), and directionality information (column D) respectively.

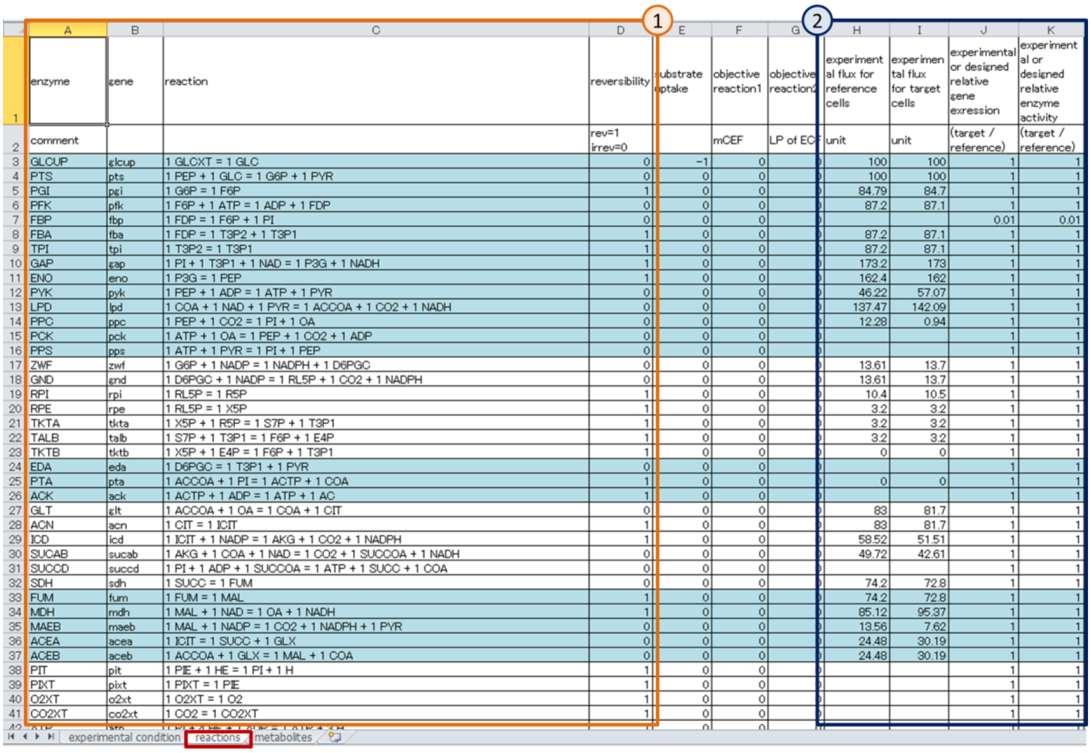


	A	B	C
1	Samples		
2	wild-type: *Escherichia coli* K-12 strain BW25113		
3	Single disruptant samples: *fbaB*		
4			
5	Culture		
6	Glucose-limited chemostat at dilution rate 0.20h^{-1} (July)		
7			
8	Publication		
9	Ishii N*, Nakahigashi K*, Baba T*, Robert M*, Soga T*, Kanai A*, Hirasawa T*, Naba M, Hirai K, Hoque A, Ho PY, Kakazu Y, Sugawara K, Igaras		
10	Multiple high-throughput analyses monitor the response of *E. coli* to perturbations.		
11	*Science*, **316**(5824):593-597.		
12	PMID: 17379776		
13			
14			
15	Notice		
16	Users of the data should carefully read the original manuscript and supplementary methods and text for additional information		
17	about the measurement values, samples and data series, and the normalization methods used and be aware of the limitations		
18	before pursuing any analysis using this data.		
19	Failure to do so might result in erroneous interpretations and conclusions for which the authors of this study cannot be held responsible.		
20			
21			
22	Measurements		
23	<Cell concentration>		
24	No information provided		
25	Unit: -		
26	Method: -		
27			
28	<Glucose concentration>		
29	Unit: g/L		
30	Method: enzymatic assay kit (F-kit, Roche Diagnostics)		
31			
32	<Acetic acid and lactic acid concentration>		
33	Unit: g/L		
34	Method: enzymatic assay kit (F-kit, Roche Diagnostics)		
35			
36	<Oxygen and carbon dioxide concentration>		
37	Unit: mmol/L		
38	Method: monitored using Offgas Jar Jr. DEX-2562 (Able)		
39			
40	<Intracellular metabolites>		
41	Unit: mM		
42	Method: capillary electrophoresis time-of-flight mass spectrometry (CE-TOFMS)		
43			
44	<Gene expression>		
45	Unit: mg-protein/g-dry cell weight		
46	Method: DNA microarray analysis and two-dimensional differential gel electrophoresis (2D-DIGE)		
47			

Figure 10. The example of a metabolic network file. All data files is organized into three sheets (1); Sheet 1: *experimental condition*, Sheet 2: *reactions* and Sheet 3: *metabolites.*

Table 14. The descriptions on the setting of input file in ‘experimental condition’ sheet

Column	Column name	Description
A1-A3	Samples	The strain sample
A5-A6	Culture	The experimental culture condition
A8-A11	Publication	The details of original publication: A9: The author (s) A10: The publication title A11: The publication journal
A13-A17	Notice	The notice for user to use the data
A19-A52	Measurements	The corresponding measurements of the experimental (a) Cell concentration (b) Glucose concentration (c) Acetic and lactic acid concentration (d) Oxygen and carbon dioxide concentration (e) Intracellular metabolites (f) Gene expression (g) Enzyme activity (h) Flux
A53	Remarks	Experimental additional information (if related)



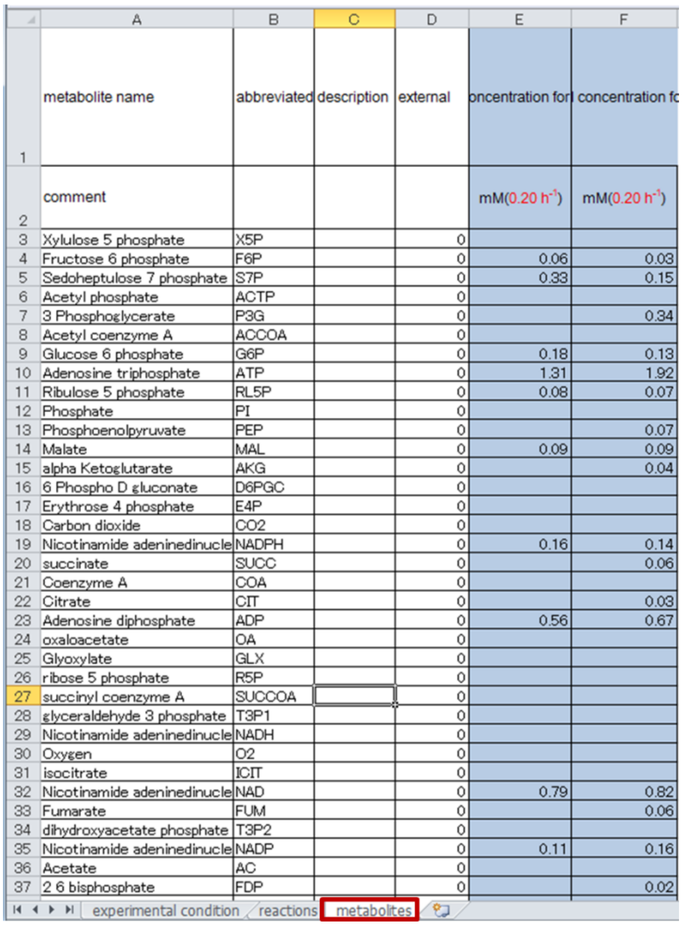
	A	B	C	D	E	F	G	H	I	J	K
1	enzyme	gene	reaction	reversibility	ubstrate ptake	objective reaction1	objective reaction2	experimental flux for reference cells	experimental flux for target cells	experimental or designed relative gene exression	experimental or designed relative enzyme activity
2	comment			rev=1 irrev=0		mCEF	LP of EC	unit	unit	(target / reference)	(target / reference)
3	GLCUP	glcup	1 GLCXT = 1 GLC	0	-1	0	0	100	100	1	1
4	PTS	pts	1 PEP + 1 GLC = 1 G6P + 1 PYR	0	0	0	0	100	100	1	1
5	PGI	pgi	1 G6P = 1 F6P	1	0	0	0	84.79	84.7	1	1
6	PFK	pfk	1 F6P + 1 ATP = 1 ADP + 1 FDP	0	0	0	0	87.2	87.1	1	1
7	FBP	fbp	1 FDP = 1 F6P + 1 PI	0	0	0	0			0.01	0.01
8	FBA	fba	1 FDP = 1 T3P2 + 1 T3P1	1	0	0	0	87.2	87.1	1	1
9	TPI	tpi	1 T3P2 = 1 T3P1	1	0	0	0	87.2	87.1	1	1
10	GAP	gap	1 PI + 1 T3P1 + 1 NAD = 1 P3G + 1 NADH	1	0	0	0	173.2	173	1	1
11	ENO	eno	1 P3G = 1 PEP	1	0	0	0	162.4	162	1	1
12	PYK	pyk	1 PEP + 1 ADP = 1 ATP + 1 PYR	0	0	0	0	46.22	57.07	1	1
13	LPD	lpd	1 COA + 1 NAD + 1 PYR = 1 ACCOA + 1 CO2 + 1 NADH	0	0	0	0	137.47	142.09	1	1
14	PPC	ppc	1 PEP + 1 CO2 = 1 PI + 1 OA	0	0	0	0	12.28	0.94	1	1
15	PCK	pck	1 ATP + 1 OA = 1 PEP + 1 CO2 + 1 ADP	0	0	0	0			1	1
16	PPS	pps	1 ATP + 1 PYR = 1 PI + 1 PEP	0	0	0	0			1	1
17	ZWF	zwf	1 G6P + 1 NADP = 1 NADPH + 1 D6PGC	0	0	0	0	13.61	13.7	1	1
18	GND	gnd	1 D6PGC + 1 NADP = 1 RL5P + 1 CO2 + 1 NADPH	0	0	0	0	13.61	13.7	1	1
19	RPI	rpi	1 RL5P = 1 R5P	1	0	0	0	10.4	10.5	1	1
20	RPE	rpe	1 RL5P = 1 X5P	1	0	0	0	3.2	3.2	1	1
21	TKTA	tkta	1 X5P + 1 R5P = 1 S7P + 1 T3P1	1	0	0	0	3.2	3.2	1	1
22	TALB	talb	1 S7P + 1 T3P1 = 1 F6P + 1 E4P	1	0	0	0	3.2	3.2	1	1
23	TKTB	tktb	1 X5P + 1 E4P = 1 F6P + 1 T3P1	1	0	0	0	0	0	1	1
24	EDA	eda	1 D6PGC = 1 T3P1 + 1 PYR	0	0	0	0			1	1
25	PTA	pta	1 ACCOA + 1 PI = 1 ACTP + 1 COA	1	0	0	0	0	0	1	1
26	ACK	ack	1 ACTP + 1 ADP = 1 ATP + 1 AC	1	0	0	0			1	1
27	GLT	glt	1 ACCOA + 1 OA = 1 COA + 1 CIT	0	0	0	0	83	81.7	1	1
28	ACN	acn	1 CIT = 1 ICIT	1	0	0	0	83	81.7	1	1
29	ICD	icd	1 ICIT + 1 NADP = 1 AKG + 1 CO2 + 1 NADPH	1	0	0	0	58.52	51.51	1	1
30	SUCAB	sucab	1 AKG + 1 COA + 1 NAD = 1 CO2 + 1 SUCCOA + 1 NADH	0	0	0	0	49.72	42.61	1	1
31	SUCCD	succd	1 PI + 1 ADP + 1 SUCCOA = 1 ATP + 1 SUCC + 1 COA	0	0	0	0			1	1
32	SDH	sdh	1 SUCC = 1 FUM	0	0	0	0	74.2	72.8	1	1
33	FUM	fum	1 FUM = 1 MAL	1	0	0	0	74.2	72.8	1	1
34	MDH	mdh	1 MAL + 1 NAD = 1 OA + 1 NADH	1	0	0	0	85.12	95.37	1	1
35	MAEB	maeb	1 MAL + 1 NADP = 1 CO2 + 1 NADPH + 1 PYR	0	0	0	0	13.56	7.62	1	1
36	ACEA	acea	1 ICIT = 1 SUCC + 1 GLX	0	0	0	0	24.48	30.19	1	1
37	ACEB	aceb	1 ACCOA + 1 GLX = 1 MAL + 1 COA	0	0	0	0	24.48	30.19	1	1
38	PIT	pit	1 PIE + 1 HE = 1 PI + 1 H	1	0	0	0			1	1
39	PIXT	pixt	1 PIXT = 1 PIE	1	0	0	0			1	1
40	O2XT	o2xt	1 O2XT = 1 O2	1	0	0	0			1	1
41	CO2XT	co2xt	1 CO2 = 1 CO2XT	1	0	0	0			1	1

Figure 11. The '*reactions*' sheet template: The input file is designed by (1) enzyme, gene, reaction list, and reversibility type to represent the metabolic network, (2) experimental data retrieved from publications.

Table 15. The descriptions on the setting of input file in '***reactions***' sheet

Column	Column name	Description
A	Enzyme	Enzyme name corresponding to a metabolic reaction
B	Gene	Gene name corresponding to a metabolic reaction
C	Reaction	Reaction formula
D	Reversibility	Reversibility of a reaction; 1 (reversible), 0 (irreversible).
E	Substrate uptake	The reaction corresponding to substrate uptake is indicated by -1; the others are set to 0.
F	Objective reaction1	The objective reaction for mCEF is indicated by -1; the others are set to 0.
G	Objective reaction2	The objective reaction for LP (in ECF) is indicated by -1; the others are set to 0.

H	Experimental flux for reference cells	Experimental flux value for reference cells.
I	Experimental flux for target cells	Experimental flux value for target cells
J	Experimental or designed relative gene expression (target/reference)	The relative gene expression ratio of the target cells to the reference ones. It is given by an experimental value or a designed value.
K	Experimental or designed relative enzyme activity (target/reference)	The relative enzyme activity ratio of the target cells to the reference ones. It is given by an experimental value or a designed value.
L	Experimental gene expression for reference cells	The experimental gene expression values of reference cells
M	Experimental gene expression for target cells	The experimental gene expression values of target cells
N	Experimental enzyme activity for reference cells	Experimental enzyme activity values of reference cells.
O	Experimental enzyme activity for target cells	Experimental enzyme activity values of target cells
P	Predicted relative gene expression	The predicted relative ratio of target gene expression to the reference expression
Q	Predicted flux for reference cells	The predicted flux value for reference cells
R	Predicted flux for target cells	The predicted flux value for target cells
S	Predicted EMC for reference cells	The predicted elementary mode coefficient for reference cells
T	Predicted EMC for target cells	The predicted elementary mode coefficient for target cells



	A	B	C	D	E	F
1	metabolite name	abbreviated	description	external	oncentration for	concentration fc
2	comment				mM(0.20 h^{-1})	mM(0.20 h^{-1})
3	Xylulose 5 phosphate	X5P		0		
4	Fructose 6 phosphate	F6P		0	0.06	0.03
5	Sedoheptulose 7 phosphate	S7P		0	0.33	0.15
6	Acetyl phosphate	ACTP		0		
7	3 Phosphoglycerate	P3G		0		0.34
8	Acetyl coenzyme A	ACCOA		0		
9	Glucose 6 phosphate	G6P		0	0.18	0.13
10	Adenosine triphosphate	ATP		0	1.31	1.92
11	Ribulose 5 phosphate	RL5P		0	0.08	0.07
12	Phosphate	PI		0		
13	Phosphoenolpyruvate	PEP		0		0.07
14	Malate	MAL		0	0.09	0.09
15	alpha Ketoglutarate	AKG		0		0.04
16	6 Phospho D gluconate	D6PGC		0		
17	Erythrose 4 phosphate	E4P		0		
18	Carbon dioxide	CO2		0		
19	Nicotinamide adeninedinucle	NADPH		0	0.16	0.14
20	succinate	SUCC		0		0.06
21	Coenzyme A	COA		0		
22	Citrate	CIT		0		0.03
23	Adenosine diphosphate	ADP		0	0.56	0.67
24	oxaloacetate	OA		0		
25	Glyoxylate	GLX		0		
26	ribose 5 phosphate	R5P		0		
27	succinyl coenzyme A	SUCCOA		0		
28	glyceraldehyde 3 phosphate	T3P1		0		
29	Nicotinamide adeninedinucle	NADH		0		
30	Oxygen	O2		0		
31	isocitrate	ICIT		0		
32	Nicotinamide adeninedinucle	NAD		0	0.79	0.82
33	Fumarate	FUM		0		0.06
34	dihydroxyacetate phosphate	T3P2		0		
35	Nicotinamide adeninedinucle	NADP		0	0.11	0.16
36	Acetate	AC		0		
37	2 6 bisphosphate	FDP		0		0.02

Figure 12. The '*metabolites*' sheet template: Metabolite sheet keeps the experimental metabolites concentration value of reference cell (wild type) and target cell (mutant type).

Table 16. The descriptions on the setting of input file in 'metabolites' sheet

Column	Column name	Description
A	Metabolite name	The metabolite name corresponding to the metabolic reaction
B	Abbreviated name	The metabolite abbreviated name
C	Description	Description of metabolites
D	External	The metabolites status; 1 (internal), 0 (external).
E	Experimental concentration for reference cells	The experimental metabolite concentration for reference cells
F	Experimental concentration for target cells	The experimental metabolite concentration for target cells

2.3 The simulation algorithms

Figure 13 represents the basic system flow of GMF. As mentioned before, the input for GMF is the metabolic network file. In general, GMF consists of five (5) main modules: (1) data preprocessing, (2) EM calculation, (3) GMF, (4) mCEF, and (5) ECF.

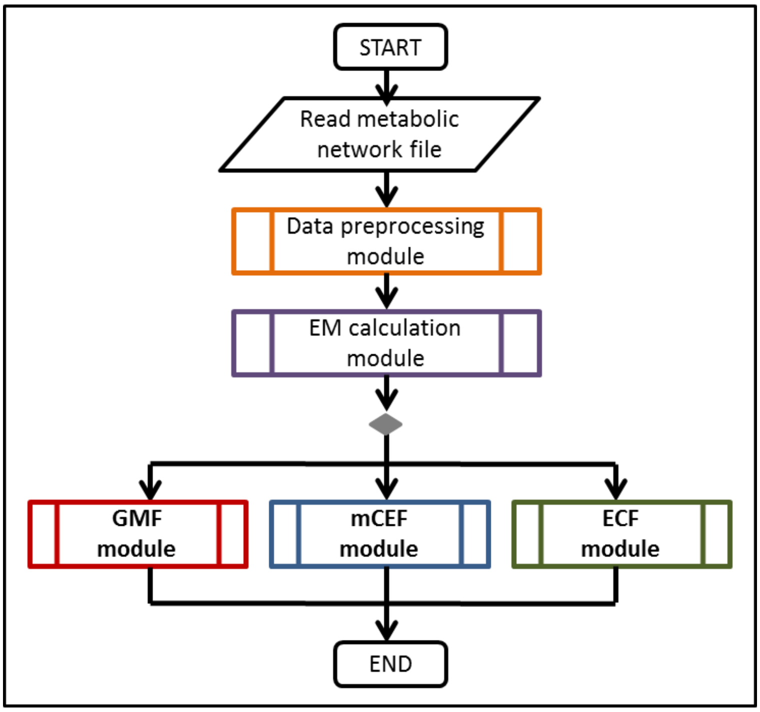


Figure 13. The main system flow of GMF, which consists of five main modules: (1) Data preprocessing, (2) EM calculation, (3) GMF calculation, (4) mCEF calculation, and (5) ECF calculation

In data preprocessing module, the system will extract the particular data required for the calculation from input file to the memory. The data needed are extracted from reaction and metabolite sheets. The system program will convert the format of reaction formula listed in the input file in the format that accepted by efmtool.

Once converted, the system will calculate the EM (EM calculation module). We implemented the calculation of EM by invoking the function `CalculateFluxModes (reactionFormula)` as provided by efmtool. efmtool will analyze the metabolic network from the reaction formulas as listed in the input file. From the called function, we extract the information of EM that will be used for calculation.

GMF integrated Elementary Flux Mode Tool (efmtool) to produce stoichiometric matrix and calculate EM. efmtool is developed in Java programming language, and integrated into MATLAB. The implementation of bit pattern tress algorithm resulted efmtool is currently the most efficient method for computing EM in large networks [75].

After these processes completed, the system program will proceed to the estimation process based on the algorithm selected by the user. As mentioned previously, if user selects GMF or ECF algorithm, the EM calculated by efmtool will be optimized by the given objective functions. The GMF application implemented four types of objective functions; Linear Programming (LP), Quadratic Programming (QP) [67], Linear Programming based on alpha spectrum (MeanLP) [51], Maximum Entropy Principle (MEP) [52].

The implementation is performed in Matlab. The nonlinear optimization (MEP) is using the function fmincon, while for the other objective functions (QP, LP, MeanLP) the present programs are improved to feed bigger metabolic networks.

GMF is consists of two algorithms: modified Control Effective Flux (mCEF) and

Enzyme Control Flux (ECF). GMF predicts the flux distribution of genetically modified mutants [64]. The mCEF algorithm, which is derived from the Control Effective Flux (CEF), estimates the relative expression ratios of metabolic genes of a mutant to wild type from changes in target gene expression. ECF estimates the flux distributions of genetically modified mutants by integrating their enzyme activity profiles into EMs. ECF is very effective in the case that an enzyme activity profile is provided.

2.4 The implementation

The Hypertext Preprocessor (PHP) is used as the GUI of GMF web application. All the programs for simulation and visualization are written in MATLAB R2014a and run on a Linux server. The efmtool program, an open source application computer interface, is employed to calculate EMs. The GMF web application is available at: http://kurata22.bio.kyutech.ac.jp/gmf/pub/top.php. The recommended web browser to use the application is Mozilla Firefox or Google Chrome. The user manuals and application programs of stand-alone version is shown in **Appendix A**.

3 CHAPTER 3: RESULT AND DISCUSSION

3.1 The gene knockout database

We have collected 112 metabolic network models that contain key metabolism processes and able to be calculated by the application [17, 18, 24, 26-30, 32-34, 76-81]. Details are described in **Tables 17-21**.

Table 17: The number of files according to microorganism in the database

Microorganism	**Number of files**
Escherichia coli	104
Corynebacterium glutamicum	4
Saccharomyces cerevisiae	3
Chinese Hamster Ovary	1
Total	**112**

Table 18: List of *E.coli* wild type data file

Dilution rate				**Total**
0.10h^{-1}	**0.40h^{-1}**	**0.50h^{-1}**	**0.70h^{-1}**	
1	1	1	1	4
Total				**4**

Table 19: List of *E.coli* genetic deletion mutant files

Pathway	Gene deletion	Dilution rate													Total
		0.10h^{-1}	**0.20h^{-1}** **June**	**0.20h^{-1}** **July**	**0.20h^{-1}** **Sept**	**0.20h^{-1}** **Oct**	**0.20h^{-1}**	**0.22h^{-1}**	**0.40h^{-1}**	**0.60h^{-1}**	**0.66h^{-1}**	**0.50h-^{1}** **(5H)**	**0.50h-^{1}** **(6H)**	**0.50h-^{1}** **(7H)**	
Glycolysis	1. *fbaB*		1*	1*	1*	1*									4
	2. *fbp*		1*	1*	1*	1*									4
	3. *gapC*		1*	1*	1*	1*									4
	4. *gpmA*		1*	1*	1*	1*									4
	5. *gpmB*		1*	1*	1*	1*									4
	6. *pfkA*		1*	1*	1*	1*									4
	7. *pfkB*		1*	1*	1*	1*									4
	8. *pgi*	1*	1*	1*	1*	1*									5
	9. *pykA*		1*	1*	1*	1*									4
	10. *pykF*	1*	1*	1*	1*	1*			1**			1**	1**	1**	9
	11. *ppsA*		1*	1*	1*	1*									4
	12. *lpdA*							1*							1
Pentose Phosphate	13. *gnd*		1*	1*	1*	1*	2*								6
	14. *pgl*		1*	1*	1*	1*									4
	15. *rpe*		1*	1*	1*	1*									4
	16. *rpiA*		1*	1*	1*	1*									4
	17. *rpiB*		1*	1*	1*	1*									4
	18. *tktA*		1*	1*	1*	1*									4
	19. *tktB*		1*	1*	1*	1*									4

	20. *talA*		1*	1*	1*	1*									4
	21. *talB*		1*	1*	1*	1*									4
	22. *sucA*						1*								1
	23. *zwf*	1*	1*	1*	1*	1*	1*			1*	1*				8
Anapleurotic Reactions	24. *ppc*						1*								1
	25. *pck*	1*													1
														Total	**100**

"*" represents continuous culture; "**" represents batch culture

Table 20: List of *C.glutamicum* genetic deletion mutant files

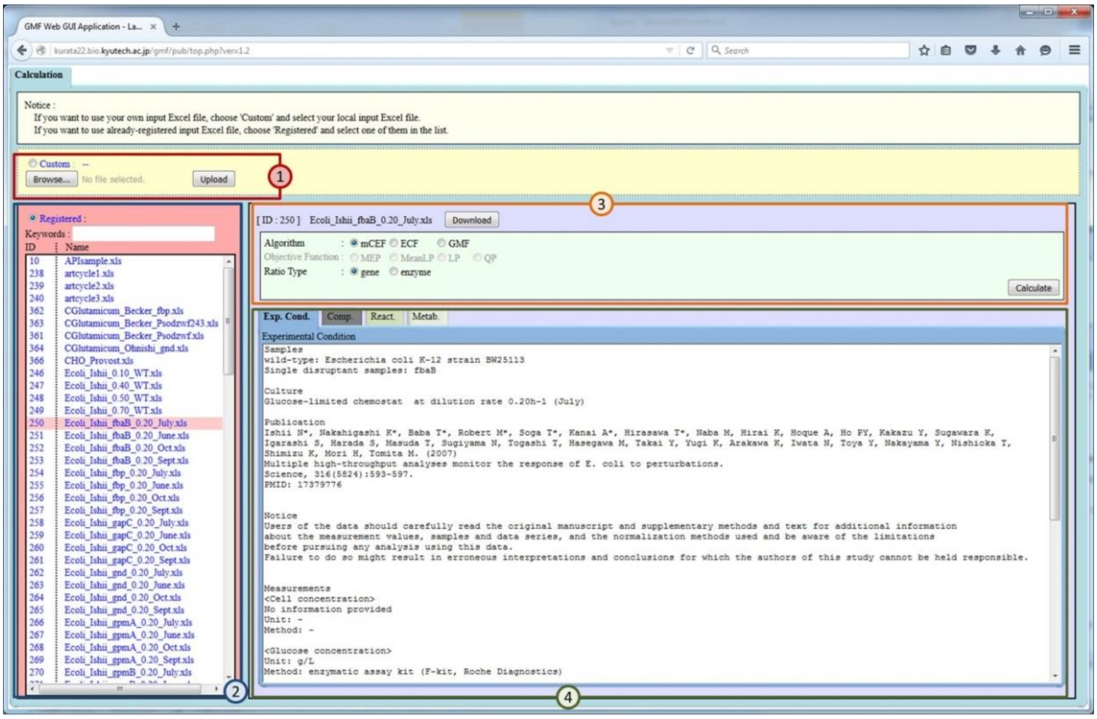
Pathway	**Gene deletion**	**Total**
Glycolysis	1. *fbp*	1
	2. *gnd*	1
	3. *zwf*	2
	Total	**4**

Table 21: List of *S.cerevisiae* wild type files

Dilution rate			**Total**
$0.15h^{-1}$	**$0.30h^{-1}$**	**$0.40h^{-1}$**	
1	1	1	3
		Total	**3**

3.2 The GMF application (web and standalone version)

Figure 14 shows the main page of the GMF web application. Users can (1) upload their own file or (2) select a file out of the registered files. (3) Users select an algorithm to estimate a flux distribution and an objective function out of the four functions (MEP, QP, LP, or MeanLP). When ECF is used, they can select a ratio type of gene expression and enzyme activity profiles. (4) Once an input file is selected, details in the metabolic network are displayed. 112 metabolic network files were registered with their associated experimental data.


GMF Web GUI Application - La...
kurata22.bio.kyutech.ac.jp/gmf/pub/top.php?ver=1.2
Search
Calculation
Browse...
Upload
Download
Calculate
Exp. Cond.
Comp.
React.
Metab.
1
2
3
4

Notice :
If you want to use your own input Excel file, choose 'Custom' and select your local input Excel file.
If you want to use already-registered input Excel file, choose 'Registered' and select one of them in the list.

○ Custom : --
No file selected.

● Registered :
Keywords :

ID	Name
10	APIsample.xls
238	artcycle1.xls
239	artcycle2.xls
240	artcycle3.xls
362	CGlutamicum_Becker_fbp.xls
363	CGlutamicum_Becker_Psodzwf243.xls
361	CGlutamicum_Becker_Psodzwf.xls
364	CGlutamicum_Ohnishi_gnd.xls
366	CHO_Provost.xls
246	Ecoli_Ishii_0.10_WT.xls
247	Ecoli_Ishii_0.40_WT.xls
248	Ecoli_Ishii_0.50_WT.xls
249	Ecoli_Ishii_0.70_WT.xls
250	Ecoli_Ishii_fbaB_0.20_July.xls
251	Ecoli_Ishii_fbaB_0.20_June.xls
252	Ecoli_Ishii_fbaB_0.20_Oct.xls
253	Ecoli_Ishii_fbaB_0.20_Sept.xls
254	Ecoli_Ishii_fbp_0.20_July.xls
255	Ecoli_Ishii_fbp_0.20_June.xls
256	Ecoli_Ishii_fbp_0.20_Oct.xls
257	Ecoli_Ishii_fbp_0.20_Sept.xls
258	Ecoli_Ishii_gapC_0.20_July.xls
259	Ecoli_Ishii_gapC_0.20_June.xls
260	Ecoli_Ishii_gapC_0.20_Oct.xls
261	Ecoli_Ishii_gapC_0.20_Sept.xls
262	Ecoli_Ishii_gnd_0.20_July.xls
263	Ecoli_Ishii_gnd_0.20_June.xls
264	Ecoli_Ishii_gnd_0.20_Oct.xls
265	Ecoli_Ishii_gnd_0.20_Sept.xls
266	Ecoli_Ishii_gpmA_0.20_July.xls
267	Ecoli_Ishii_gpmA_0.20_June.xls
268	Ecoli_Ishii_gpmA_0.20_Oct.xls
269	Ecoli_Ishii_gpmA_0.20_Sept.xls
270	Ecoli_Ishii_gpmB_0.20_July.xls

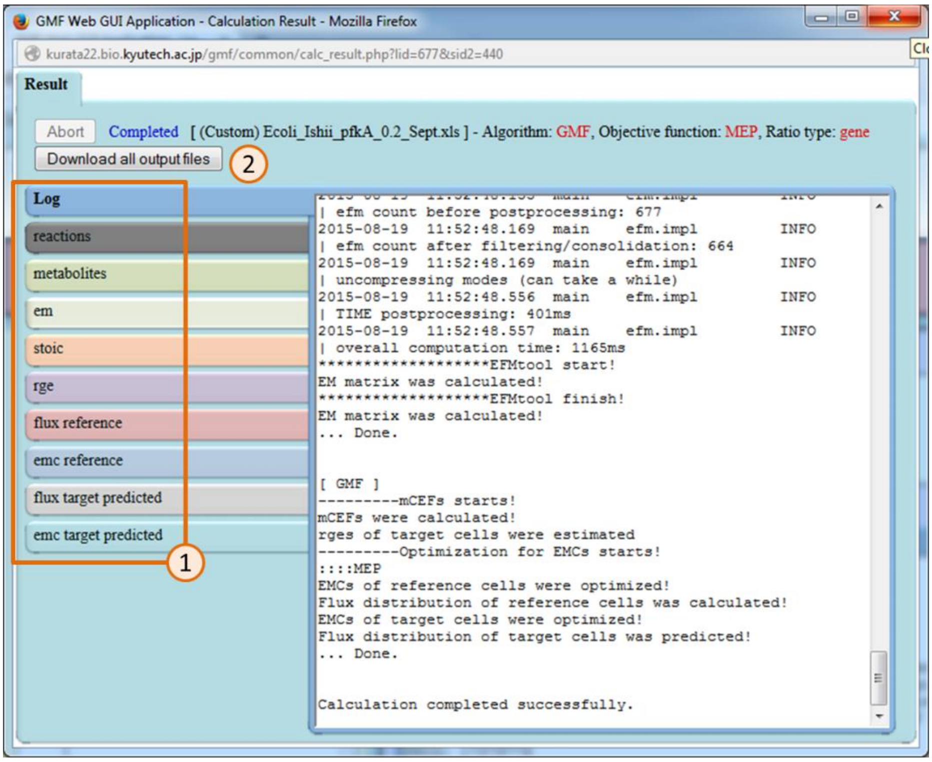
[ID : 250] Ecoli_Ishii_fbaB_0.20_July.xls

Algorithm : ● mCEF ○ ECF ○ GMF
Objective Function : ○ MEP ○ MeanLP ○ LP ○ QP
Ratio Type : ● gene ○ enzyme

Experimental Condition

```
Samples
wild-type: Escherichia coli K-12 strain BW25113
Single disruptant samples: fbaB

Culture
Glucose-limited chemostat  at dilution rate 0.20h-1 (July)

Publication
Ishii N*, Nakahigashi K*, Baba T*, Robert M*, Soga T*, Kanai A*, Hirasawa T*, Naba M, Hirai K, Hoque A, Ho PY, Kakazu Y, Sugawara K,
Igarashi S, Harada S, Masuda T, Sugiyama N, Togashi T, Hasegawa M, Takai Y, Yugi K, Arakawa K, Iwata N, Toya Y, Nakayama Y, Nishioka T,
Shimizu K, Mori H, Tomita M. (2007)
Multiple high-throughput analyses monitor the response of E. coli to perturbations.
Science, 316(5824):593-597.
PMID: 17379776


Notice
Users of the data should carefully read the original manuscript and supplementary methods and text for additional information
about the measurement values, samples and data series, and the normalization methods used and be aware of the limitations
before pursuing any analysis using this data.
Failure to do so might result in erroneous interpretations and conclusions for which the authors of this study cannot be held responsible.


Measurements
<Cell concentration>
No information provided
Unit: -
Method: -

<Glucose concentration>
Unit: g/L
Method: enzymatic assay kit (F-kit, Roche Diagnostics)
```

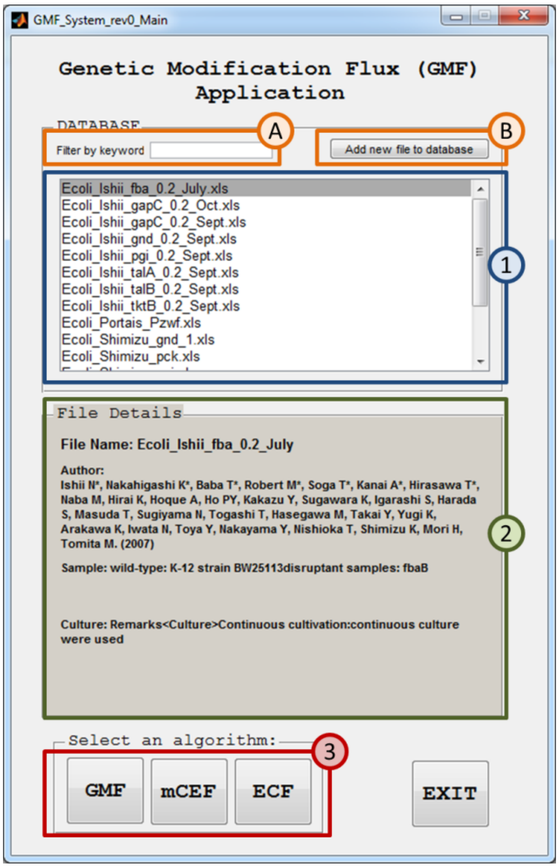
Figure 14. The main page of GMF web application. Users can (1) upload their own file or (2) select a file out of the registered files. (3) Users select an algorithm to estimate a flux distribution or gene expression profile, a ratio type of gene expression or enzyme activity profiles, and an objective function. (4) Once an input file is selected, details in the metabolic network are displayed.



GMF Web GUI Application - Calculation Result - Mozilla Firefox
kurata22.bio.kyutech.ac.jp/gmf/common/calc_result.php?lid=677&sid2=440
Result
Abort
Completed [(Custom) Ecoli_Ishii_pfkA_0.2_Sept.xls] - Algorithm: GMF, Objective function: MEP, Ratio type: gene
Download all output files
2
Log
reactions
metabolites
em
stoic
rge
flux reference
emc reference
flux target predicted
emc target predicted
1

```
| efm count before postprocessing: 677
2015-08-19  11:52:48.169  main    efm.impl        INFO
| efm count after filtering/consolidation: 664
2015-08-19  11:52:48.169  main    efm.impl        INFO
| uncompressing modes (can take a while)
2015-08-19  11:52:48.556  main    efm.impl        INFO
| TIME postprocessing: 401ms
2015-08-19  11:52:48.557  main    efm.impl        INFO
| overall computation time: 1165ms
******************EFMtool start!
EM matrix was calculated!
******************EFMtool finish!
EM matrix was calculated!
... Done.


[ GMF ]
---------mCEFs starts!
mCEFs were calculated!
rges of target cells were estimated
---------Optimization for EMCs starts!
::::MEP
EMCs of reference cells were optimized!
Flux distribution of reference cells was calculated!
EMCs of target cells were optimized!
Flux distribution of target cells was predicted!
... Done.


Calculation completed successfully.
```

Figure 15. The sample of GMF calculation result page in the web version. Users can click (1) on a particular tab to get the desired output or (2) download all output files, by clicking on ‘Download all output files’.



Ecoli_Ishii_fba_0.2_July.xls
Ecoli_Ishii_gapC_0.2_Oct.xls
Ecoli_Ishii_gapC_0.2_Sept.xls
Ecoli_Ishii_gnd_0.2_Sept.xls
Ecoli_Ishii_pgi_0.2_Sept.xls
Ecoli_Ishii_talA_0.2_Sept.xls
Ecoli_Ishii_talB_0.2_Sept.xls
Ecoli_Ishii_tktB_0.2_Sept.xls
Ecoli_Portais_Pzwf.xls
Ecoli_Shimizu_gnd_1.xls
Ecoli_Shimizu_pck.xls

File Details

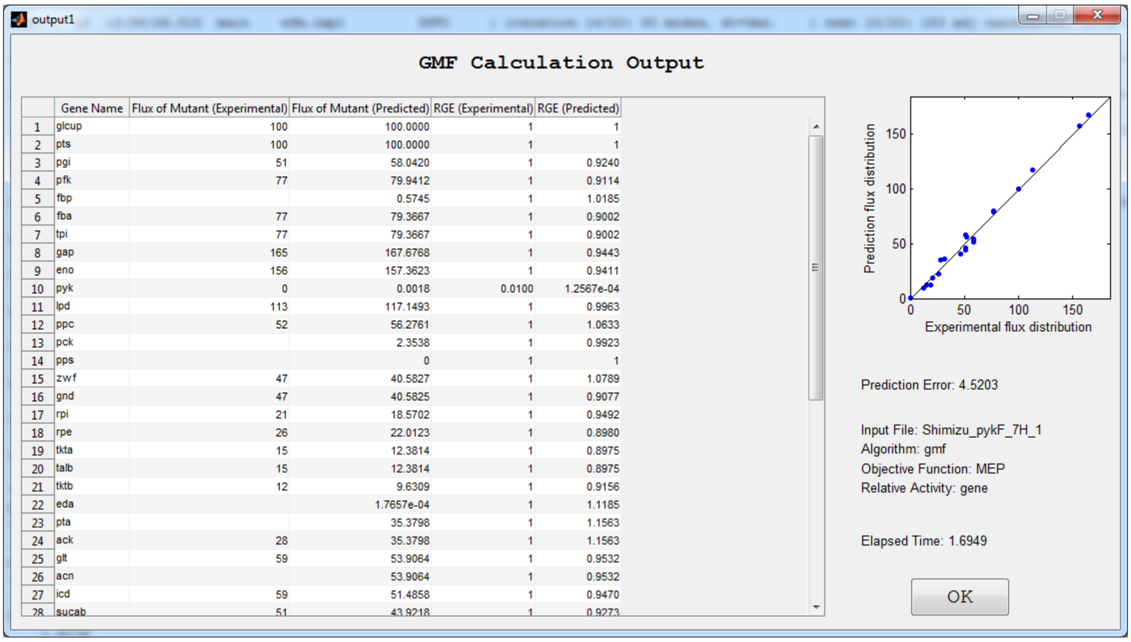
File Name: Ecoli_Ishii_fba_0.2_July

Author:
Ishii N*, Nakahigashi K*, Baba T*, Robert M*, Soga T*, Kanai A*, Hirasawa T*, Naba M, Hirai K, Hoque A, Ho PY, Kakazu Y, Sugawara K, Igarashi S, Harada S, Masuda T, Sugiyama N, Togashi T, Hasegawa M, Takai Y, Yugi K, Arakawa K, Iwata N, Toya Y, Nakayama Y, Nishioka T, Shimizu K, Mori H, Tomita M. (2007)

Sample: wild-type: K-12 strain BW25113disruptant samples: fbaB

Culture: Remarks<Culture>Continuous cultivation:continuous culture were used

Figure 16. The main graphical user interface (GUI) of GMF stand-alone application. The data files are available in (1). Once a particular file is clicked, the details are shown as in (2). User need to select an algorithm as marked in (3) to start the calculation.



GMF Calculation Output

	Gene Name	Flux of Mutant (Experimental)	Flux of Mutant (Predicted)	RGE (Experimental)	RGE (Predicted)
1	glcup	100	100.0000	1	1
2	pts	100	100.0000	1	1
3	pgi	51	58.0420	1	0.9240
4	pfk	77	79.9412	1	0.9114
5	fbp		0.5745	1	1.0185
6	fba	77	79.3667	1	0.9002
7	tpi	77	79.3667	1	0.9002
8	gap	165	167.6768	1	0.9443
9	eno	156	157.3623	1	0.9411
10	pyk	0	0.0018	0.0100	1.2567e-04
11	lpd	113	117.1493	1	0.9963
12	ppc	52	56.2761	1	1.0633
13	pck		2.3538	1	0.9923
14	pps		0	1	1
15	zwf	47	40.5827	1	1.0789
16	gnd	47	40.5825	1	0.9077
17	rpi	21	18.5702	1	0.9492
18	rpe	26	22.0123	1	0.8980
19	tkta	15	12.3814	1	0.8975
20	talb	15	12.3814	1	0.8975
21	tktb	12	9.6309	1	0.9156
22	eda		1.7657e-04	1	1.1185
23	pta		35.3798	1	1.1563
24	ack	28	35.3798	1	1.1563
25	glt	59	53.9064	1	0.9532
26	acn		53.9064	1	0.9532
27	icd	59	51.4858	1	0.9470
28	sucab	51	43.9218	1	0.9273

Prediction Error: 4.5203

Input File: Shimizu_pykF_7H_1
Algorithm: gmf
Objective Function: MEP
Relative Activity: gene

Elapsed Time: 1.6949

Figure 17. The sample of GMF calculation result page using stand-alone version. The information related to estimation result is displayed in the center table of the page. Users can refer the information of selected file name, algorithm to perform the estimation in the right hand side of the page.

3.3 Feasibility of application programs

To validate the feasibility of the application programs, we tested them with registered models. The prediction accuracy by GMF or ECF was evaluated by:

$$\text{Prediction error (PE)} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(v_{prediction_i} - v_{experimental_i}\right)^2} \qquad (58)$$

where $v_{prediction_i}$ is the ith flux predicted, $v_{experimental_i}$ is the ith experimental flux, and n the number of reactions.

We picked up *E.coli* gene deletion mutants: *gapC*, *talB* [17], *pck* [34], *pykF* [18], and *zwf* [30], and *E.coli* over-expression mutants: *zwf* [29], and estimated the flux distributions by GMF and ECF with MEP, QP, LP, or MeanLP objective functions. **Table 22** shows the effect of an objective function on the prediction error of the genetic mutants by GMF and ECF. The MEP predicted their flux distributions more accurately than other objective functions.

Table 22. Effect of objective functions on the prediction errors using GMF and ECF.

E.coli mutant	Sample	Growth rate	Prediction error by GMF				Prediction error by ECF			
			MEP	QP	LP	Mean LP	MEP	QP	LP	Mean LP
Gene deletion	*gapC*	$0.20h^{-1}$ (Oct)	11.52	13.63	46.58	13.03	10.08	11.83	41.93	11.47
	talB	$0.20h^{-1}$ (Sept)	5.27	6.61	44.49	5.55	5.13	6.53	45.02	5.51
	pck	$0.10h^{-1}$	9.48	19.33	46.57	25.36	8.86	16.94	44.51	20.39
	pykF	$0.50h^{-1}$ (7hrs)	4.52	7.89	39.71	6.45	5.98	6.75	38.05	5.41
	zwf	$0.20h^{-1}$	7.00	8.11	17.81	7.08	4.32	6.69	26.19	5.15
Over expression	*zwf*	$0.66h^{-1}$	4.72	7.58	43.43	9.28	5.13	8.53	45.31	10.25

Thus, we used MEP to compare the GMF- and ECF- predicted flux distributions of the genetic mutants with the experimental flux distributions, as shown in **Figures 18** and **19**. The predicted flux distributions were consistent with the experimental data. The estimated flux distributions using GMF are shown in **Tables 24-29**. To statistically validate the prediction errors, we performed linear regression analysis between the GMF- or ECF- predicted flux distributions and experimental data, as shown in **Table 23.** The coefficients of determination (R^2) ranged between 0.940 and 0.993 for GMF and 0.940 to 0.997 for ECF, respectively. The Pearson correlations (r) for GMF and ECF were from 0.935 to 0.994 and 0.950 to 0.997, respectively. Both methods provided significant correlation between the predicted gene expression and experimental data.

Table 23. The coefficients of determination (R^2) and Pearson correlation (r) of prediction accuracy by GMF and ECF. MEP is used

E.coli **Mutant condition**	Sample	Growth rate	**GMF**		**ECF**	
			R^2	**r**	**R^2**	**r**
gene deletion	*gapC*	0.20h^{-1} (Oct)	0.942	0.983	0.955	0.986
	talB	0.20h^{-1} (Sept)	0.989	0.935	0.990	0.950
	pck	0.10h^{-1}	0.940	0.971	0.940	0.971
	pykF	0.50h^{-1} (7hrs)	0.977	0.989	0.981	0.991
	zwf	0.20h^{-1}	0.985	0.994	0.997	0.997
Over expression	*zwf*	0.66h^{-1}	0.993	0.985	0.992	0.984

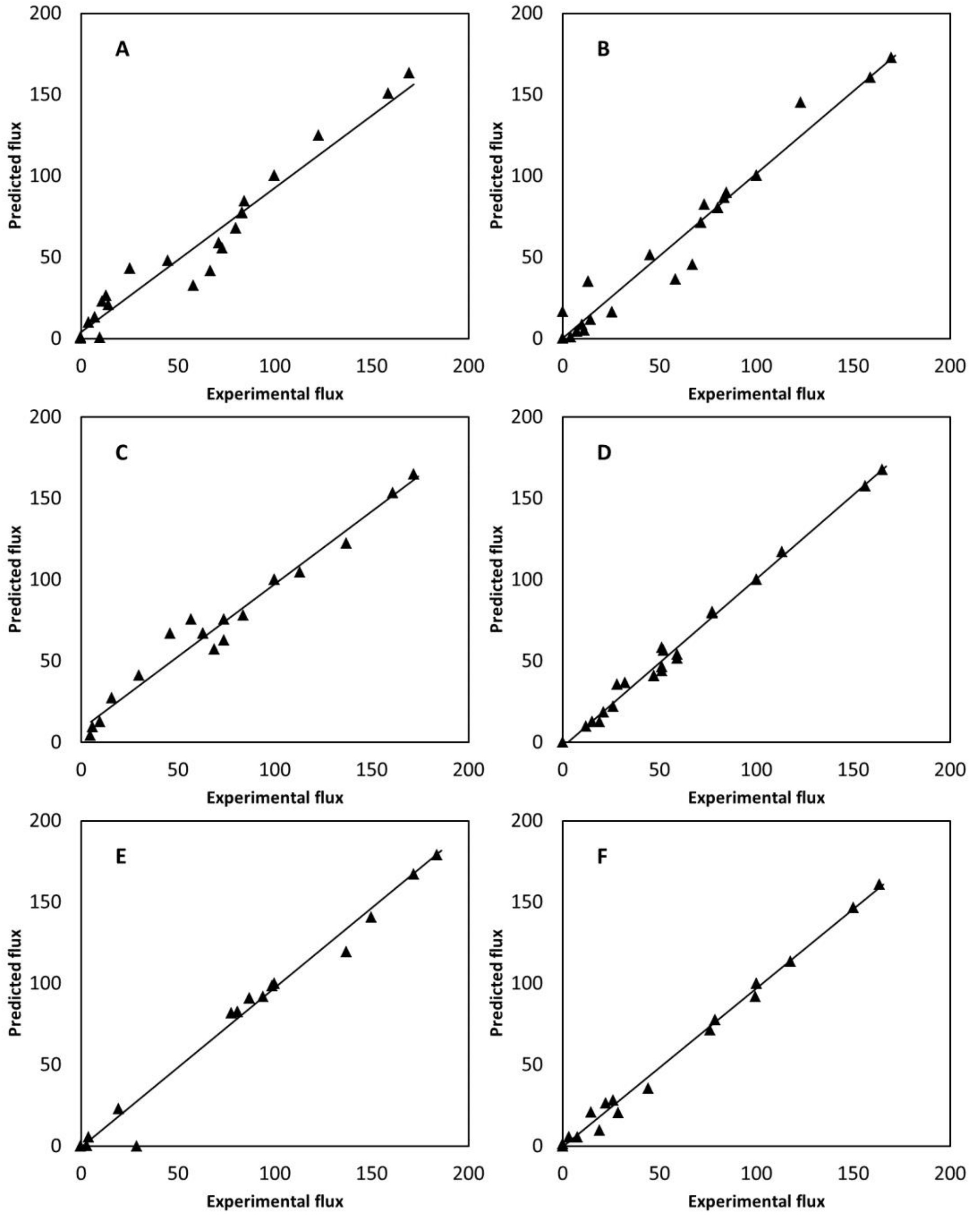


Figure 18. Comparison between the predicted and experimental flux distributions for the *E.coli* mutants of (A) *gapC* gene deletion, (B) *talB* gene deletion, (C) *pck* gene deletion, (D) *pykF* gene deletion, (E) *zwf* gene deletion, and (F) *zwf* overexpression. GMF is tested with the MEP objective function.

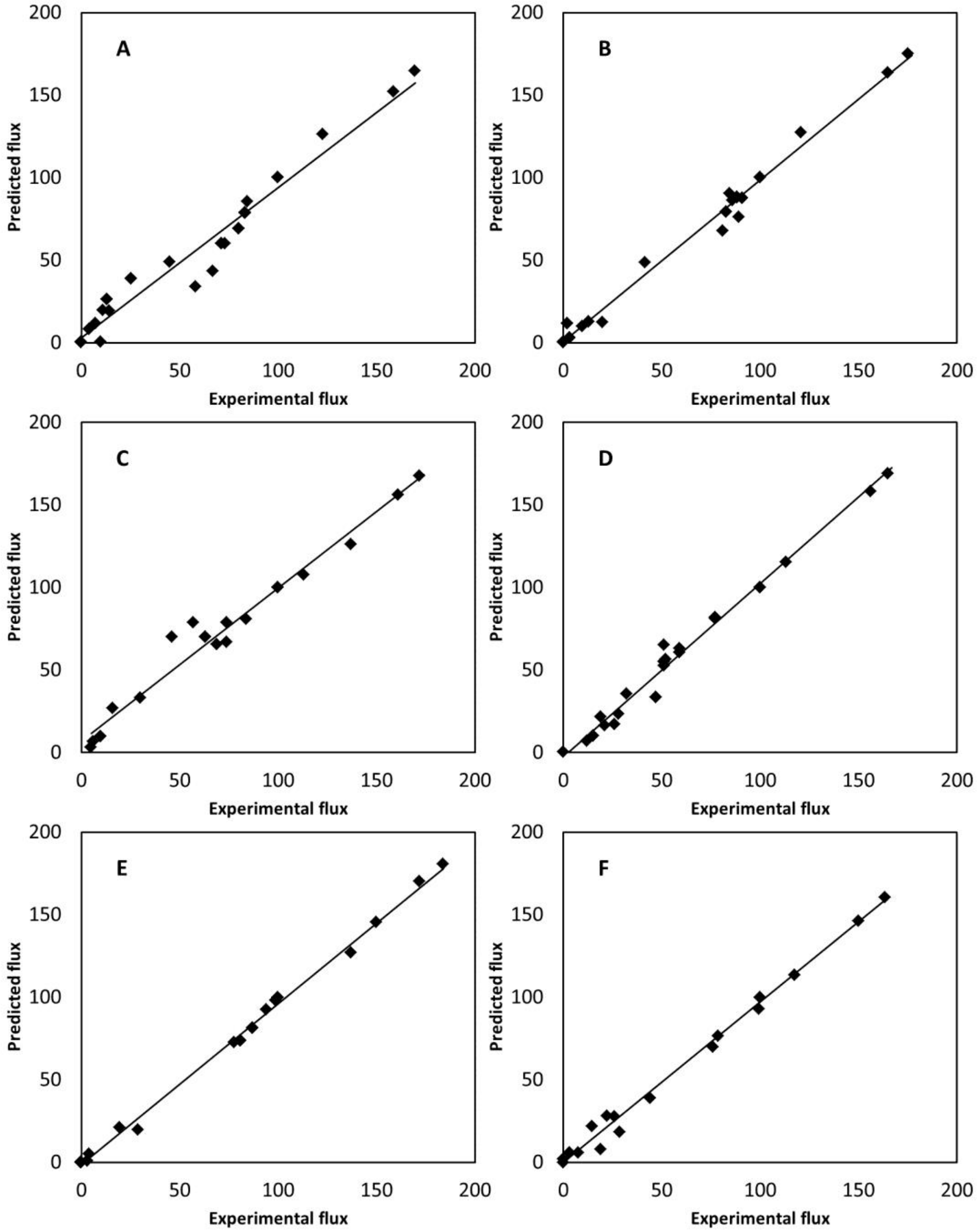


Figure 19. Comparison between the predicted and experimental flux distributions for the *E.coli* mutants of (A) *gapC* gene deletion, (B) *talB* gene deletion, (C) *pck* gene deletion, (D) *pykF* gene deletion, (E) *zwf* gene deletion, and (F) *zwf* overexpression. ECF is tested with the MEP objective function.

Table 24. Prediction result of *E.coli gapC* 0.20h^{-1} (Oct) gene deletion using GMF. Experimental fluxes (exp fluxes) are from [17]

Gene	Exp fluxes	Predicted fluxes			
		MEP	QP	LP	MeanLP
glcup	100.00	100.00	100.00	100.00	100.00
pts	100.00	100.00	100.00	100.00	100.00
pgi	73.15	55.45	49.19	-1.30	51.187
pfk	83.38	77.28	78.96	95.40	79.66
fba	83.38	77.07	69.90	-1.89	70.96
tpi	83.38	77.07	69.90	-1.89	70.96
gap	169.48	162.87	155.92	88.83	157.13
eno	158.68	150.34	143.59	80.65	144.92
pyk	44.94	47.64	45.58	7.24	44.28
lpd	122.84	124.77	132.35	196.66	130.76
ppc	9.85	0.42	5.85	75.61	7.35
zwf	25.25	42.88	49.17	100.21	47.19
gnd	25.25	42.88	41.34	5.94	39.82
rpi	14.22	20.42	19.81	5.98	19.24
rpe	11.03	22.46	21.53	-0.04	20.58
tktA	7.12	12.90	12.41	1.07	11.92
talB	7.12	12.90	12.41	1.07	11.92
tktB	3.92	9.56	9.12	-1.11	8.66
pta	0	0.01	0	51.36	4.50
glt	80.08	67.63	70.02	69.52	67.24
acn	80.08	67.63	70.02	69.52	67.24
icd	66.93	41.42	38.10	13.93	38.33
sucAB	58.23	32.23	29.06	7.93	29.38
sdh	71.38	58.44	60.97	63.51	58.30
fum	71.38	58.44	60.97	63.51	58.30
mdh	84.54	84.14	90.92	108.66	85.31
maeB	0.00	0.51	1.97	10.44	1.90
aceA	13.16	26.21	31.91	55.58	28.91
aceB	13.16	26.21	31.91	55.58	28.91

Table 25. Prediction result of *E.coli talB* 0.20h^{-1} (Sept) gene deletion using GMF. Experimental fluxes (exp fluxes) are from [17].

Gene	Exp fluxes	Predicted fluxes			
		MEP	QP	LP	MeanLP
glcup	100.00	100.00	100.00	100.00	100.00
pts	100.00	100.00	100.00	100.00	100.00
pgi	85.80	90.97	84.90	16.81	91.35
pfk	88.10	89.81	87.79	95.65	90.28
fba	88.10	89.69	87.42	15.34	89.80
tpi	88.10	89.69	87.42	15.34	89.80
gap	175.10	176.84	174.14	105.96	177.00
eno	165.00	165.50	162.43	97.68	165.71
pyk	41.48	50.92	48.27	8.68	51.34
lpd	120.68	130.00	127.74	183.69	131.21
ppc	19.82	11.94	12.68	69.98	13.28
zwf	12.70	7.52	13.54	82.09	7.15
gnd	12.70	7.52	13.54	4.69	7.09
rpi	9.70	8.05	10.24	5.61	7.89
rpe	3.00	-0.53	3.30	-0.92	-0.79
tkta	3.00	1.25	3.21	0.64	1.11
talb	3.00	1.25	3.21	0.64	1.11
tktb	0.00	-1.78	0.09	-1.56	-1.90
pta	0.00	0.00	0.00	61.68	0.23
glt	90.90	90.69	85.12	58.78	91.00
acn	90.90	90.69	85.12	58.78	91.00
icd	89.03	79.34	71.40	15.97	78.89
sucab	80.83	71.03	62.81	9.90	70.60
sdh	82.70	82.38	76.53	52.71	82.72
fum	82.70	82.38	76.53	52.71	82.72
mdh	84.58	93.49	88.92	82.47	93.95
maeB	0.00	0.24	1.33	13.06	0.89
aceA	1.88	11.35	13.72	42.81	12.12
aceB	1.88	11.35	13.72	42.81	12.12

Table 26. Prediction result of *E.coli pck* 0.10h^{-1} gene deletion using GMF. Experimental fluxes (exp fluxes) are from [34].

Gene	Exp fluxes	Predicted fluxes			
		MEP	QP	LP	MeanLP
glcup	100.00	100.00	100.00	100.00	100.00
pts	100.00	100.00	100.00	100.00	100.00
pgi	69.00	65.31	42.51	8.53	28.64
fba	84.00	80.89	51.85	7.09	38.54
gap	172.00	167.63	140.86	95.86	127.73
eno	161.00	155.93	131.16	85.95	118.19
pyk	137.00	126.16	105.74	55.48	88.73
lpd	113.00	107.55	125.15	143.12	126.20
ppc	16.00	26.77	22.72	27.45	26.29
pck		0.90	0.55	0.28	0.01
gnd	30.00	33.13	22.11	6.09	22.80
tktA	10.00	9.74	6.29	0.93	6.54
talB	10.00	9.74	6.29	0.93	6.54
tktB	6.00	6.62	3.70	-1.71	4.00
glt	74.00	78.68	74.09	33.85	51.20
icd	57.00	78.68	74.09	19.78	51.05
sucab	46.00	70.10	66.98	12.52	44.06
sdh	63.00	70.10	66.98	26.58	44.20
fum	63.00	70.10	66.98	26.58	44.20
mdh	74.00	66.86	63.56	18.57	36.36
maeB	5.00	3.24	3.42	22.08	7.98

Table 27. Prediction result of *E.coli pykF* 0.50h^{-1} (7H) gene deletion using GMF. Experimental fluxes (exp fluxes) are from [18].

Gene	Exp fluxes	Predicted fluxes			
		MEP	QP	LP	MeanLP
glcup	100.00	100.00	100.00	100.00	100.00
pts	100.00	100.00	100.00	100.00	100.00
pgi	51.00	58.04	60.10	9.39	53.41
pfk	77.00	79.94	81.70	94.68	79.12
fba	77.00	79.37	79.91	8.53	77.62
tpi	77.00	79.37	79.91	8.53	77.62
gap	165.00	167.68	167.96	100.09	165.78
eno	156.00	157.36	157.42	92.64	155.33
pyk	0.00	0.00	0.00	0.00	0.00
lpd	113.00	117.15	122.53	181.51	118.51
ppc	52.00	56.28	60.79	84.65	56.97
zwf	47.00	40.58	38.49	89.62	45.20
gnd	47.00	40.58	38.49	4.92	45.02
rpi	21.00	18.57	17.98	5.28	20.11
rpe	26.00	22.01	20.51	-0.36	24.91
tktA	15.00	12.38	11.66	0.81	13.85
talB	15.00	12.38	11.66	0.81	13.85
tktB	12.00	9.63	8.85	-1.17	11.06
ack	28.00	35.38	37.50	83.05	37.50
glt	59.00	53.91	50.35	45.14	49.07
icd	59.00	51.49	41.67	10.19	42.89
sucAB	51.00	43.92	33.94	4.73	35.23
sdh	51.00	46.34	42.62	39.67	41.41
fum	51.00	46.34	42.62	39.67	41.41
mdh	19.00	12.36	9.09	63.92	9.76
maeB	32.00	36.40	42.21	10.71	37.83

Table 28. Prediction result of *E.coli zwf* 0.20h^{-1} gene deletion using GMF. Experimental fluxes (exp fluxes) are from [30].

Gene	Exp fluxes	Predicted fluxes			
		MEP	QP	LP	MeanLP
glcup	100.00	100.00	100.00	100.00	100.00
pts	100.00	100.00	100.00	100.00	100.00
pgi	98.90	98.33	98.27	87.56	98.44
fba	94.20	92.70	91.85	82.40	91.93
gap	184.00	180.85	178.57	171.35	178.65
eno	172.00	170.39	166.85	161.59	166.93
pyk	150.00	145.64	139.48	125.17	139.51
lpd	137.00	127.17	118.47	161.74	118.55
ppc	19.40	21.26	23.46	52.68	23.53
pck	0.00	0.00	0.00	19.51	0.02
zwf	0.00	0.27	0.17	11.14	0.00
gnd	0.00	0.27	0.13	0.39	0.00
rpi	4.10	5.20	5.77	4.90	5.73
rpe	-4.10	-4.93	-5.64	-4.51	-5.73
tkta	-0.94	-1.07	-1.26	-0.95	-1.30
talb	-0.94	-1.07	-1.26	-0.95	-1.30
tktb	-3.20	-3.86	-4.38	-3.55	-4.43
eda	0.00	0.00	0.04	10.75	0.00
ack	29.00	19.72	5.89	46.82	0.06
glt	87.00	81.65	83.68	61.14	89.53
acn	87.00	81.65	83.68	61.14	89.53
icd	87.00	81.65	83.67	31.43	89.47
sucab	80.80	73.98	75.08	24.28	80.88
succd	80.80	73.98	75.08	24.28	80.88
sdh	80.80	73.98	75.08	53.99	80.94
fum	80.80	73.98	75.08	53.99	80.94
mdh	77.80	72.93	74.27	39.68	80.08
maeb	3.00	1.05	0.82	44.03	0.92

Table 29. Prediction result of *E.coli zwf* 0.66h^{-1} gene deletion (overexpression) using GMF. Experimental fluxes (exp fluxes) are from [29]

Gene	Exp fluxes	Predicted fluxes			
		MEP	QP	LP	MeanLP
glcup	100.00	100.00	100.00	100.00	100.00
pts	100.00	100.00	100.00	100.00	100.00
pgi	75.90	71.53	58.48	-1.02	56.15
pfk	78.50	77.51	69.85	26.08	67.12
gap	163.50	161.03	152.07	85.31	147.73
eno	149.90	146.51	138.46	74.25	133.93
pyk	117.40	113.49	106.01	50.07	100.80
lpd	99.40	91.97	97.34	142.07	98.05
ppc	26.00	28.18	27.92	38.24	28.63
zwf	22.20	26.53	39.71	99.54	42.01
gnd	14.60	21.04	24.88	7.52	22.32
tkta	3.20	5.40	6.78	1.28	5.91
talb	3.20	5.40	6.78	1.28	5.91
tktb	0.00	1.53	3.15	-1.67	2.23
ack	44.00	35.49	39.16	51.44	38.96
glt	28.60	20.65	24.28	41.82	23.59
mdh	18.80	9.90	12.70	34.60	11.62
maeb	0.00	0.10	1.91	20.62	3.31

4 CHAPTER 4: CONCLUSION

Molecular biology encompasses uncovered coherent biological facts, which significantly essential in upholding life. The progression of systems biology approach is rapidly emerges, from the definition of single components i.e. cells, tissues, organs, and organisms towards its specific interactions. Thousands of genome sequences from humans, plants, animals and disease tissues are now made available; the recent systems biology application is now giving extra focuses on the needs to produce quantitative interpretation that demonstrate the potential contribution for disease and drug discovery. In accomplishing as such, understanding system-level becomes the primary goal of systems biology.

The specific interactions of systems biology components are manifested through metabolic network. In understanding as such, the integration of heterogeneous biological data becomes a major concern. This concern is promisingly solved by the combination of experimental and computational approaches, i.e. computational biology.

In examining a metabolic network and its pathway, the study on genetic perturbed condition such as genetic knockout is one of significant strategies to comprehend the complexity of cellular systems. Due to its significant contribution to support the metabolic engineering and biotechnology applications, various methods have been proposed, which implements either optimization-based or pathway-based analysis. FBA, MOMA, ROOM are some of the methods that include constraints and/or linear optimization techniques to analyze metabolic fluxes. Alternatively, MFA, EM, Extreme Pathway, CEF, mCEF, ECF, GMF are the example of pathway-based method that able to recognize a complete fluxes solution from a metabolic network without any cellular objective bias are provided.

Over the multi omics level of cellular systems, fluxomics provides essential information. To facilitate in-depth analysis and generalization, a comprehensive, systematic and standardize flux knockout data set with different experimental conditions and methodology would be useful. The variety is important since for the knockout study, different culturing conditions has affected on flux results.

In this work, we developed the web application of GMF to estimate the flux distribution of genetic mutants with overexpressed or deleted genes. The originality of GMF is derived based on EM, and the former study showed that the performance of GMF is outperformed as compared to FBA and MOMA. This application implements the GMF and ECF with four types of objective functions: MEP, QP, LP, and MeanLP. As an alternative, GMF is also developed in stand-alone version.

To assist the analysis process, a database was attached that registers metabolic network files with a variety of experimental data. To the date, we have collected 112 data set; which included *E.coli* (104), *C.glutamicum* (4), *S.cerevisiae* (3) and *CHO* (1). In representing the data, the experimental data of fluxes, enzyme activity profiles and metabolite concentration are collected in re-arranged in a consistent and standardized data files. The information on experimental condition and method was recorded as well.

The metabolic network models presented in GMF were reconstructed and designed based on central carbon metabolism. We focus on this metabolism system since it contains ‘busy’ pathways with high-traffic of energy, cofactors and precursors that would be high priority for understanding in metabolic engineering purposes. To demonstrate the feasibility of the application programs, we tested the registered models. Based on the measured prediction accuracy, the predicted flux distributions were consistent to the experimental data. The MEP predicted their flux distributions more accurately than other objective functions.

4.1 The contribution and advantages

GMF provides the real-time or simultaneous analysis platform with original experimental data of flux, enzyme profiles, and metabolite concentration. This simulator application can be readily extended by adding latest simulation tools and be a user-friendly application that contributes to advances in metabolic flux analysis.

A part of the simulator, the database provides a variation of experimental data files, represented in a metabolic network model and simulation-ready format. The availability of real sample data contributes a valuable reference platform in facilitating the analysis for systems biology tasks; particularly for further observation where a large number of knockout mutant data becomes necessary.

Furthermore, the proposed metabolic network used in representing the data set can be the basis as predictive model in analysis tasks.

To date, the use of GMF and ECF algorithms had required an expensive MATLAB license and its associated command line operation, but the new web application solved such problems. Users can use the GMF and ECF through the web without any license and command line operation.

Metabolic network data are often written in the SBML format [82], where each reaction is decomposed into multiple classified components. This format has an advantage in the exact definition of each component, while it requires lots of memory due to redundancy of XML tags and sometimes hampers human readability. On the other hand, the GMF web application presents one metabolic reaction in one cell in an ordinary Microsoft Excel format, enhancing the human readability and usability.

4.2 The future works

The works that have been done in building GMF to publicly accessible is still having room of improvements. With the consideration based on the current works, further improvement should be planned in future:

(1) Towards automatic reconstructing the metabolic network

The reconstruction of metabolic network in GMF now is done manually. This reconstructing process should be improved towards computerized process by extracting relevant information from available online genome and pathway databases. As mentioned previously, the processes in reconstructing a computer executable model will having challenges (blocked reaction problem, missing gap), however by implementing suitable algorithm these problems will solve.

(2) Towards a large, standardized flux data set

We have produced a standardized knockouts data set of central carbon metabolism for *E.coli*, *C.glutamicum*, *S.cerevisae* and *CHO*. These data set are arranged in a consistent metabolic network (according to particular organism model), each files comprises the experimental data on: fluxes, enzyme activity profiles (gene expression and enzyme activity), mRNA, and metabolite concentrations. The condition and method conducted during the experimental process were also recorded in the files.

The current data set will become more valuable by increasing the current number of data files, as more variety files will facilitates comprehensive analysis. Another valuable data set should include the data set of multiple perturbation experiment i.e. double, triple or more knockouts [10].

The next practical value of data set should include the regulators of central metabolism. By including such information, more comprehensive models are needed to reconstruct in future. The regulatory network layer is essential to substitute the modeling principle that applied stand-in concept (e.g. objective function) or other heuristics, to fundamental mechanistic models. It is also important to have a data collection of gene set that related to the aerobic/anaerobic responses, stress response and carbon sources catabolism from other sources than glucose, i.e. xylose, glycerol, and acetate.

(3) Towards improved quantitative analysis

The current development of GMF is only able to perform estimation process within a singular input files. It would be more favorable if two or more files (e.g. the same gene knockout type of different growth rates) are estimated concurrently. This would assist for efficient analysis tasks.

In addition, to improve the interpretation of metabolism, it could be ideal to apply many existing and proven theoretical frameworks, for example graph theory of metabolic robustness, flux coupling, transcriptional versus metabolic limited fluxes classifications, modularized network analysis or other relevant principles.

References

1. Edda Klipp, W.L., Christoph Wierling, Axel Kowald, Hans Lehrach, and Ralf Herwig, *Systems Biology A Textbook*2009, Federal Republic of Germany: Wiley-Blackwell.
2. Ideker, T., T. Galitski, and L. Hood, *A New Approach to Decoding Life: Systems Biology.* Annual Review of Genomics Human Genetics, 2001. **2**: p. 343-72.
3. Kitano, H., *Systems biology: a brief overview.* Science, 2002. **295**(5560): p. 1662-4.
4. Kitano, H., *Computational systems biology.* Nature, 2002. **420**(6912): p. 206-10.
5. Baxevanis, A.D., *The molecular biology database collection: an online compilation of relevant database resources.* Nucleic Acids Research, 2000. **28**(1): p. 1-7.
6. Galperin, M.Y., D.J. Rigden, and X.M. Fernandez-Suarez, *The 2015 Nucleic Acids Research Database Issue and molecular biology database collection.* Nucleic Acids Research, 2015. **43**(Database issue): p. D1-5.
7. Rigden, D.J., X.M. Fernandez-Suarez, and M.Y. Galperin, *The 2016 database issue of Nucleic Acids Research and an updated molecular biology database collection.* Nucleic Acids Research, 2016. **44**(D1): p. D1-6.
8. Stephanopoulos, G., A.A. Aristidou, and J.H. Nielsen, *Chapter 1: The Essence of Metabolic Engineering*. Metabolic Engineering : Principles and Methodologies1998, San Diego: Academic Press.
9. Nielsen, J., *Metabolic engineering: techniques for analysis of targets for genetic manipulations.* Biotechnology and Bioengineering, 1998. **58**(2-3): p. 125-32.
10. Long, C.P. and M.R. Antoniewicz, *Metabolic Flux Analysis of Escherichia coli Knockouts: Lessons from the Keio Collection and Future Outlook.* Current Opinion in Biotechnology, 2014. **28**: p. 127-33.
11. Nakahigashi, K., et al., *Systematic phenome analysis of Escherichia coli multiple-knockout mutants reveals hidden reactions in central carbon metabolism.* Molecular Systems Biology, 2009. **5**: p. 306.
12. Nizam, S.A. and K. Shimizu, *Effects of arcA and arcB genes knockout on the metabolism in Escherichia coli under anaerobic and microaerobic conditions.* Biochemical Engineering Journal, 2008. **42**(3): p. 229-236.
13. Nizam, S.A., et al., *Effects of arcA and arcB genes knockout on the metabolism in Escherichia coli under aerobic condition.* Biochemical Engineering Journal, 2009. **44**(2-3): p. 240-250.
14. Portnoy, V.A., et al., *Deletion of Genes Encoding Cytochrome Oxidases and Quinol Monooxygenase Blocks the Aerobic-Anaerobic Shift in Escherichia coli K-12 MG1655.* Applied and Environmental Microbiology, 2010. **76**(19): p. 6529-6540.

15. Siddiquee, K.A., M.J. Arauzo-Bravo, and K. Shimizu, *Effect of a pyruvate kinase (pykF-gene) knockout mutation on the control of gene expression and metabolic fluxes in Escherichia coli.* FEMS Microbiology Letters, 2004. **235**(1): p. 25-33.

16. Yao, R., et al., *Catabolic regulation analysis of Escherichia coli and its crp, mlc, mgsA, pgi and ptsG mutants.* Microbial Cell Factories, 2011. **10**: p. 67.

17. Ishii, N., et al., *Multiple high-throughput analyses monitor the response of E. coli to perturbations.* Science, 2007. **316**(5824): p. 593-7.

18. Toya, Y., et al., *^{13}C-Metabolic Flux Analysis for Batch Culture of Escherichia coli and Its pyk and pgi Gene Knockout Mutants Based on Mass Isotopomer Distribution of Intracellular Metabolites.* Biotechnology Progress, 2010. **26**(4): p. 975-992.

19. Fong, S.S., et al., *Latent pathway activation and increased pathway capacity enable Escherichia coli adaptation to loss of key metabolic enzymes.* Journal of Biological Chemistry, 2006. **281**(12): p. 8024-33.

20. Canonaco, F., et al., *Metabolic flux response to phosphoglucose isomerase knock-out in Escherichia coli and impact of overexpression of the soluble transhydrogenase UdhA.* FEMS Microbiology Letters, 2001. **204**(2): p. 247-52.

21. Usui, Y., et al., *Investigating the effects of perturbations to pgi and eno gene expression on central carbon metabolism in Escherichia coli using (13)C metabolic flux analysis.* Microbial Cell Factories, 2012. **11**: p. 87.

22. Haverkorn van Rijsewijk, B.R., et al., *Large-scale 13C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in Escherichia coli.* Molecular Systems Biology, 2011. **7**: p. 477.

23. Fischer, E. and U. Sauer, *Metabolic flux profiling of Escherichia coli mutants in central carbon metabolism using GC-MS.* European Journal of Biochemistry, 2003. **270**(5): p. 880-91.

24. Hua, Q., et al., *Responses of the central metabolism in Escherichia coli to phosphoglucose isomerase and glucose-6-phosphate dehydrogenase knockouts.* Journal of Bacteriology, 2003. **185**(24): p. 7053-7067.

25. Zhu, J. and K. Shimizu, *Effect of a single-gene knockout on the metabolic regulation in Escherichia coli for D-lactate production under microaerobic condition.* Metabolic Engineering, 2005. **7**(2): p. 104-15.

26. Al Zaid Siddiquee, K., M.J. Arauzo-Bravo, and K. Shimizu, *Metabolic flux analysis of pykF gene knockout Escherichia coli based on ^{13}C-labeling experiments together with measurements of enzyme activities and intracellular metabolite concentrations.* Applied Microbiology and Biotechnology, 2004. **63**(4): p. 407-17.

27. Emmerling, M., et al., *Metabolic flux responses to pyruvate kinase knockout in*

Escherichia coli. Journal of Bacteriology, 2002. **184**(1): p. 152-164.

28. Li, M., et al., *Effect of lpdA gene knockout on the metabolism in Escherichia coli based on enzyme activities, intracellular metabolite concentrations and metabolic flux analysis by ^{13}C-labeling experiments.* Journal of Biotechnology, 2006. **122**(2): p. 254-266.

29. Nicolas, C., et al., *Response of the central metabolism of Escherichia coli to modified expression of the gene encoding the glucose-6-phosphate dehydrogenase.* FEBS Letters, 2007. **581**(20): p. 3771-6.

30. Zhao, J., et al., *Global metabolic response of Escherichia coli to gnd or zwf gene-knockout, based on ^{13}C-labeling experiments and the measurement of enzyme activities.* Applied Microbiology and Biotechnology, 2004. **64**(1): p. 91-8.

31. Zhao, J., et al., *Effect of zwf gene knockout on the metabolism of Escherichia coli grown on glucose or acetate.* Metaboling Engineering, 2004. **6**(2): p. 164-174.

32. Jiao, Z., et al., *Analysis of metabolic and physiological responses to gnd knockout in Escherichia coli by using C-13 tracer experiment and enzyme activity measurement.* FEMS Microbiology Letters, 2003. **220**(2): p. 295-301.

33. Li, M., et al., *Effect of sucA or sucC gene knockout on the metabolism in Escherichia coli based on gene expressions, enzyme activities, intracellular metabolite concentrations and metabolic fluxes by C-13-labeling experiments.* Biochemical Engineering Journal, 2006. **30**(3): p. 286-296.

34. Chen Yang, Q.H., Tomoya Baba, Hirotada Mori, Kazuyuki Shimizu, *Analysis of Escherichia coli Anaprerotic Metabolism and its Regulation Mechanisms From the Metabolic Responses to Altered Dilution Rates and Phosphoenolpyruvate Carboxylinase Knockout.* Biotechnology and Bioengineering, 2003. **84**(2): p. 129-144.

35. Perrenoud, A. and U. Sauer, *Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in Escherichia coli.* Journal of Bacteriology, 2005. **187**(9): p. 3171-9.

36. Waegeman, H., et al., *Effect of iclR and arcA knockouts on biomass formation and metabolic fluxes in Escherichia coli K12 and its implications on understanding the metabolism of Escherichia coli BL21 (DE3).* BMC Microbiology, 2011. **11**: p. 70.

37. Zhu, J., et al., *Effect of the global redox sensing/regulation networks on Escherichia coli and metabolic flux distribution based on C-13 labeling experiments.* Metabolic Engineering, 2006. **8**(6): p. 619-27.

38. Waegeman, H., et al., *Effect of iclR and arcA deletions on physiology and metabolic fluxes in Escherichia coli BL21 (DE3).* Biotechnology Letters, 2012. **34**(2): p. 329-37.

39. Toya, Y., et al., *Metabolic regulation analysis of wild-type and arcA mutant*

Escherichia coli under nitrate conditions using different levels of omics data. Molecular Biosystems, 2012. **8**(10): p. 2593-604.

40. Otsuka, Y., et al., *GenoBase: comprehensive resource database of Escherichia coli K-12.* Nucleic Acids Research, 2015. **43**(Database issue): p. D606-17.

41. Kim, H.U., T.Y. Kim, and S.Y. Lee, *Metabolic flux analysis and metabolic engineering of microorganisms.* Molecular Biosystems, 2008. **4**(2): p. 113-20.

42. Badsha, M.B., R. Tsuboi, and H. Kurata, *Complementary elementary modes for fast and efficient analysis of metabolic networks.* Biochemical Engineering Journal, 2014. **90**: p. 121-130.

43. Shlomi, T., O. Berkman, and E. Ruppin, *Regulatory on/off minimization of metabolic flux changes after genetic perturbations.* Proceedings National Academy Sciences of U S A, 2005. **102**(21): p. 7695-700.

44. Kim, J. and J.L. Reed, *RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations.* Genome Biology, 2012. **13**(9): p. R78.

45. Matsuoka, Y. and K. Shimizu, *Chapter 15: Metabolic Flux Analysis for Escherichia coli by Flux Balance Analysis.* Metabolic Flux Analysis : Methods and Protocols, ed. J.O. Krömer, L.K. Nielsen, and L.M. Blank2014, New York: Springer New York.

46. Schuster, S., T. Dandekar, and D.A. Fell, *Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering.* Trends Biotechnology, 1999. **17**(2): p. 53-60.

47. Schilling, C.H., D. Letscher, and B.O. Palsson, *Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective.* Journal of Theoretical Biology, 2000. **203**(3): p. 229-248.

48. Stelling, J., et al., *Metabolic network structure determines key aspects of functionality and regulation.* Nature, 2002. **420**(6912): p. 190-3.

49. Trinch C. T., W.A., and Srienc Friedrich, *Elementary Mode Analysis: A Useful Metabolic Pathway Analysis Tool for Characterizing Cellular Metabolism.* Applied Microbiology and Biotechnology, 2009. **81**(5): p. 813-826.

50. Toya, Y., et al., *Metabolic flux analysis and visualization.* J Proteome Res, 2011. **10**(8): p. 3313-23.

51. Kurata, H., et al., *Integration of enzyme activities into metabolic flux distributions by elementary mode analysis.* BMC Systems Biology, 2007. **1**.

52. Zhao, Q. and H. Kurata, *Use of maximum entropy principle with Lagrange multipliers extends the feasibility of elementary mode analysis.* Journal of

Bioscience and Bioengineering, 2010. **110**(2): p. 254-261.

53. Thomas Pfeiffer, J.C.N., Ferdinand Moldenhauer and Stefan Schuster. *METATOOL*. 2000.

54. Hoops, S., et al., *COPASI--a COmplex PAthway SImulator*. Bioinformatics, 2006. **22**(24): p. 3067-74.

55. Klamt., S.a.v.K.A. *CellNetAnalyzer*. 2013 January, 2013; Available from: http://www2.mpi-magdeburg.mpg.de/projects/cna/cna.html.

56. Quek, L.E., et al., *OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis.* Microbial Cell Factories, 2009. **8**: p. 25.

57. Kadir, T.A., et al., *Modeling and simulation of the main metabolism in Escherichia coli and its several single-gene knockout mutants with experimental verification.* Microbial Cell Factories, 2010. **9**: p. 88.

58. Young, J.D., et al., *Integrating cybernetic modeling with pathway analysis provides a dynamic, systems-level description of metabolic control.* Biotechnology and Bioengineering, 2008. **100**(3): p. 542-59.

59. Yeang, C.H. and M. Vingron, *A joint model of regulatory and metabolic networks.* BMC Bioinformatics, 2006. **7**: p. 332.

60. Liu, L.M., et al., *Use of genome-scale metabolic models for understanding microbial physiology.* FEBS Letters, 2010. **584**(12): p. 2556-2564.

61. Rudd, J.Z.a.K.E. *EcoGene 3.0 Nucleic Acids Research, 41 (D1): D613-D624.* 2013; Available from: http://www.ecogene.org/.

62. Kanehisa, M.a.G., S, *KEGG: Kyoto Encyclopedia of Genes and Genomes.* Nucleic Acids Research, 2012. **28**: p. 27-30.

63. Caspi, R.A., T. Billington. R, Dreher. K, Foerster. H, Fulcher. CA, Holland. T.A, Keseler. I.M, Kothari. A, Kubo. A, Krummenacker. M, Latendresse. M, Mueller. LA, Ong. Q, Paley. S, Subhraveti. P, Weaver. DS, Weerasinghe. D, Zhang P, and Karp, P.D., *The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases.* Nucleic Acids Research, 2014. **42**(1): p. D459-D471.

64. Zhao, Q. and H. Kurata, *Genetic modification of flux for flux prediction of mutants.* Bioinformatics, 2009. **25**(13): p. 1702-8.

65. Zhao, Q. and H. Kurata, *Maximum entropy decomposition of flux distribution at steady state to elementary modes.* Journal of Bioscience and Bioengineering, 2009. **107**(1): p. 84-9.

66. Cakir, T., B. Kirdar, and K.O. Ulgen, *Metabolic pathway analysis of yeast strengthens the bridge between transcriptomics and metabolic networks.*

Biotechnology and Bioengineering, 2004. **86**(3): p. 251-60.

67. Schwartz, J.M. and M. Kanehisa, *A quadratic programming approach for decomposing steady-state metabolic flux distributions onto elementary modes.* Bioinformatics, 2005. **21 Suppl 2**: p. ii204-5.

68. Schwartz, J.M. and M. Kanehisa, *Quantitative elementary mode analysis of metabolic pathways: the example of yeast glycolysis.* BMC Bioinformatics, 2006. **7**: p. 186.

69. Shimizu, K., *Toward systematic metabolic engineering based on the analysis of metabolic regulation by the integration of different levels of information.* Biochemical Engineering Journal, 2009. **46**(3): p. 235-251.

70. Terzer, M. and J. Stelling, *Large-scale computation of elementary flux modes with bit pattern trees.* Bioinformatics, 2008. **24**(19): p. 2229-2235.

71. Orth, J., R. Fleming, and B. Palsson *Reconstruction and Use of Microbial Metabolic Networks: the Core Escherichia coli Metabolic Model as an Educational Guide.* EcoSal Plus, 2010. **4**, DOI: doi:10.1128/ecosalplus.10.2.1.

72. *Jakob Nielsen's Designing Web Usability: The Practice of Simplicity.* Library Journal, 2000. **125**(7): p. 6.

73. Nielsen, J., *Metabolic engineering.* Applied Microbiology and Biotechnology, 2001. **55**(3): p. 263-83.

74. Wiback, S.J., R. Mahadevan, and B.O. Palsson, *Using metabolic flux data to further constrain the metabolic solution space and predict internal flux patterns: The Escherichia coli spectrum.* Biotechnology and Bioengineering, 2004. **86**(3): p. 317-331.

75. Klamt, S. *Manual CellNetAnalyzer*. 2014 [cited 2014; Available from: http://www2.mpi-magdeburg.mpg.de/projects/cna/manual_cellnetanalyzer.pdf.

76. Peng, L., M. Arauzo-Bravo, and K. Shimizu, *Metabolic flux analysis for a ppc mutant Escherichia coli based on ^{13}C-labelling experiments together with enzyme activity assays and intracellular metabolite measurements.* FEMS Microbiology Letters, 2004. **235**(1): p. 17-23.

77. Becker, J., et al., *Amplified expression of fructose 1,6-bisphosphatase in Corynebacterium glutamicum increases in vivo flux through the pentose phosphate pathway and lysine production on different carbon sources.* Applied and Environmental Microbiology, 2005. **71**(12): p. 8587-8596.

78. Becker, J., et al., *Metabolic flux engineering of L-lysine production in Corynebacterium glutamicum—over expression and modification of G6P dehydrogenase* Journal of Biotechnology, 2007. **132**(2): p. 99-109.

79. Frick, O. and C. Wittmann, *Characterization of the metabolic shift between oxidative and fermentative growth in Saccharomyces cerevisiae by comparative C-13 flux analysis.* Microbial Cell Factories, 2005. **4**.

80. Ohnishi, J., et al., *A novel gnd mutation leading to increased L-lysine production in Corynebacterium glutamicum.* FEMS Microbiology Letters, 2005. **242**(2): p. 265-274.

81. Provost, A. and G. Bastin, *Dynamic metabolic modelling under the balanced growth condition.* Journal of Process Control, 2004. **14**(7): p. 717-728.

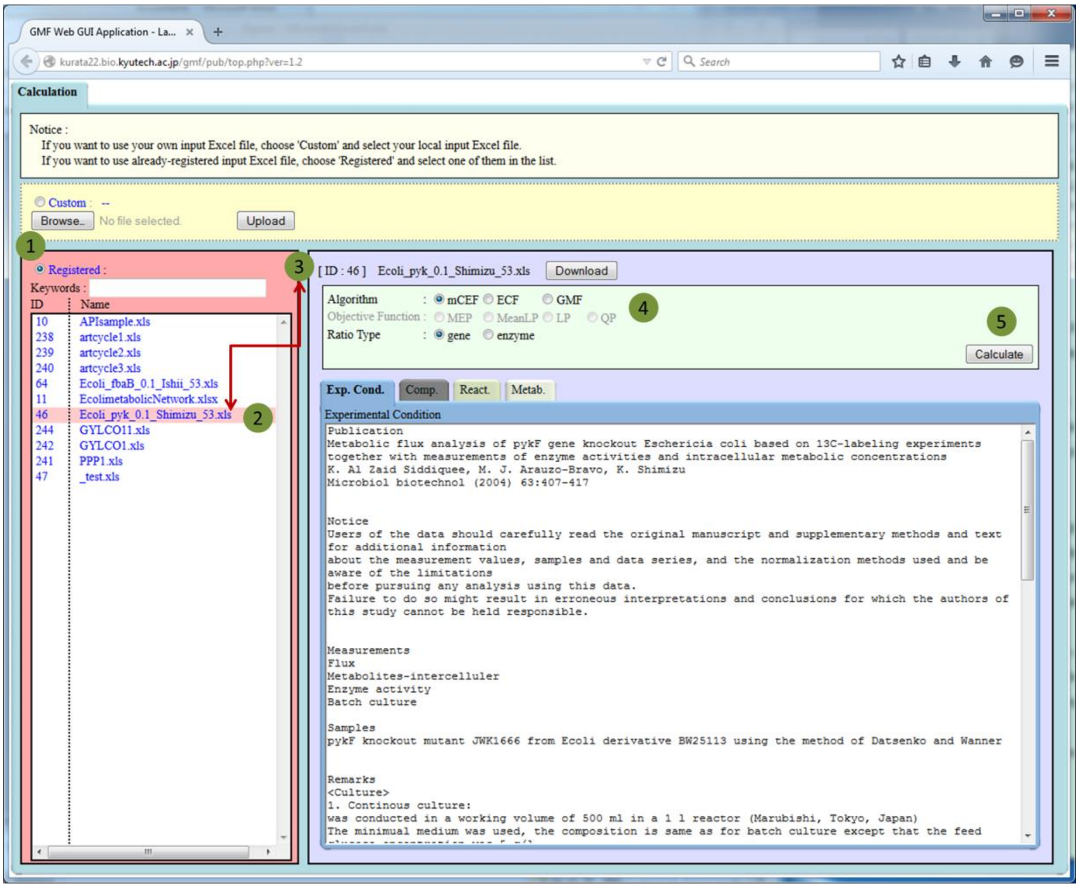
82. Hucka, M., et al., *The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models.* Bioinformatics, 2003. **19**(4): p. 524-31.

Appendix A: The Instruction of application programs (User Manual)

Genetic Modification Flux (GMF) web application

1. Selecting a file

   (a) GMF for a registered file


GMF Web GUI Application - La...
kurata22.bio.kyutech.ac.jp/gmf/pub/top.php?ver=1.2
Search
Calculation
Browse..
Upload
Download
Calculate
Exp. Cond.
Comp.
React.
Metab.
1
2
3
4
5

Notice :
If you want to use your own input Excel file, choose 'Custom' and select your local input Excel file.
If you want to use already-registered input Excel file, choose 'Registered' and select one of them in the list.

○ Custom : --
No file selected.

● Registered :
Keywords :

ID	Name
10	APIsample.xls
238	artcycle1.xls
239	artcycle2.xls
240	artcycle3.xls
64	Ecoli_fbaB_0.1_Ishii_53.xls
11	EcolimetabolicNetwork.xlsx
46	Ecoli_pyk_0.1_Shimizu_53.xls
244	GYLCO11.xls
242	GYLCO1.xls
241	PPP1.xls
47	_test.xls

[ID : 46]   Ecoli_pyk_0.1_Shimizu_53.xls

Algorithm : ● mCEF ○ ECF ○ GMF
Objective Function : ○ MEP ○ MeanLP ○ LP ○ QP
Ratio Type : ● gene ○ enzyme

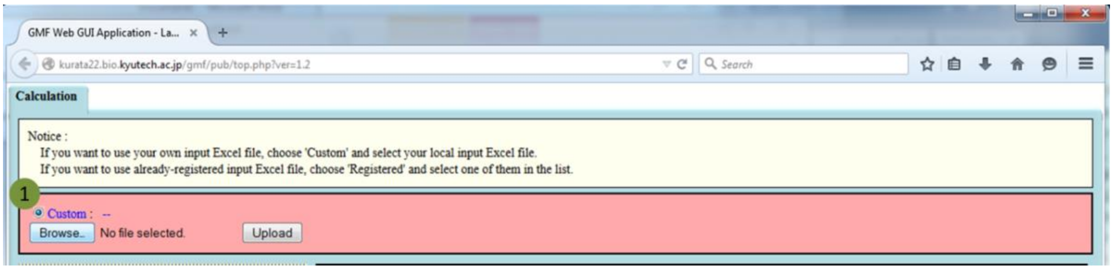
Experimental Condition

```
Publication
Metabolic flux analysis of pykF gene knockout Eschericia coli based on 13C-labeling experiments
together with measurements of enzyme activities and intracellular metabolic concentrations
K. Al Zaid Siddiquee, M. J. Arauzo-Bravo, K. Shimizu
Microbiol biotechnol (2004) 63:407-417


Notice
Users of the data should carefully read the original manuscript and supplementary methods and text
for additional information
about the measurement values, samples and data series, and the normalization methods used and be
aware of the limitations
before pursuing any analysis using this data.
Failure to do so might result in erroneous interpretations and conclusions for which the authors of
this study cannot be held responsible.


Measurements
Flux
Metabolites-intercelluler
Enzyme activity
Batch culture

Samples
pykF knockout mutant JWK1666 from Ecoli derivative BW25113 using the method of Datsenko and Wanner


Remarks
<Culture>
1. Continous culture:
was conducted in a working volume of 500 ml in a 1 l reactor (Marubishi, Tokyo, Japan)
The minimual medium was used, the composition is same as for batch culture except that the feed
```

Fig. B1: The GMF main page to calculate a registered network in the database

1. Click on the ‘Registered’ radio button.
2. Click on a file name from the list.
3. Once a file is selected, the file name will be displayed in 3.
4. Click on the desired algorithm and ratio type.
5. Click the ‘Calculate’ button.

(b) GMF for user’s own file

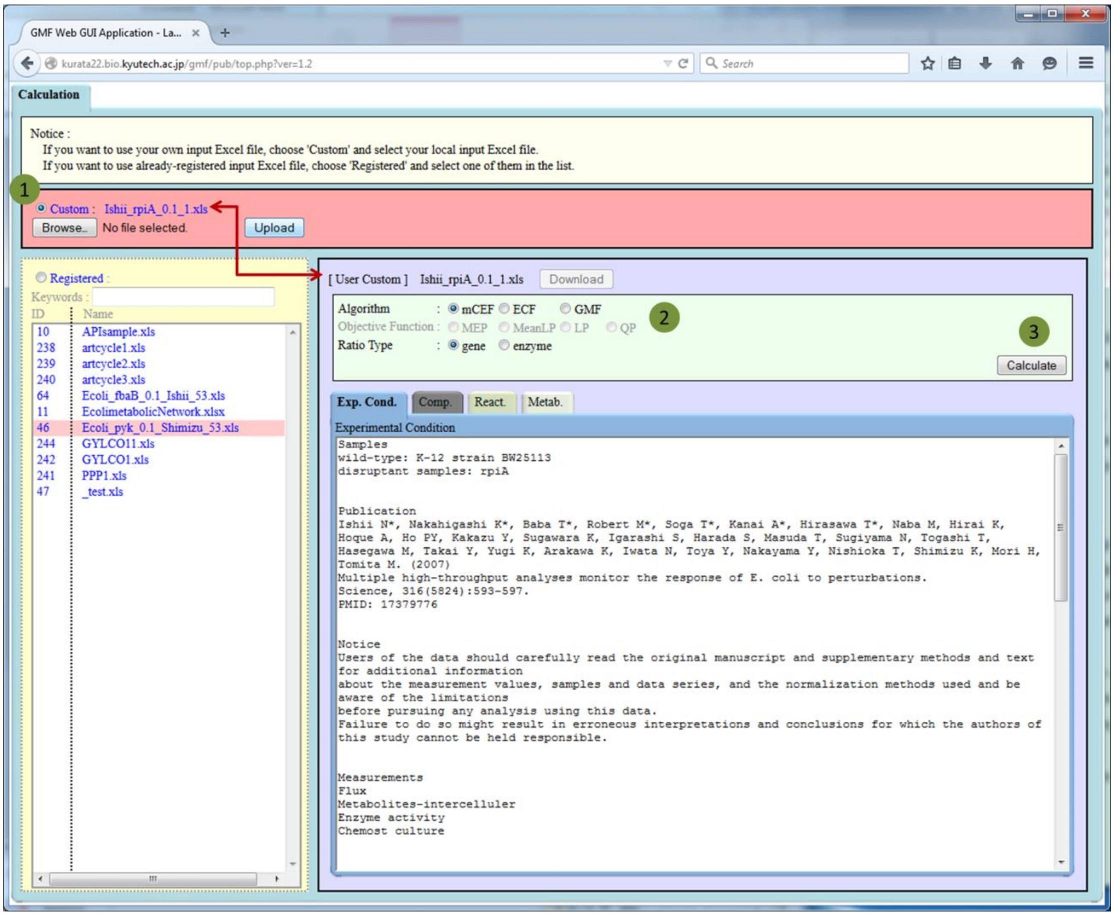

GMF Web GUI Application - La...
kurata22.bio.kyutech.ac.jp/gmf/pub/top.php?ver=1.2
Search
Calculation
1
Browse...
Upload

Notice :
If you want to use your own input Excel file, choose 'Custom' and select your local input Excel file.
If you want to use already-registered input Excel file, choose 'Registered' and select one of them in the list.

● Custom : --
No file selected.

Fig. B2: The GMF main page to calculate user’s own file (before a file is selected)

1. Click on the ‘Custom’ radio button. Click on the ‘Browse’ button and make a file selection from the user’s local drive. Click on the ‘Upload’ button.


GMF Web GUI Application - La...
kurata22.bio.kyutech.ac.jp/gmf/pub/top.php?ver=1.2
Search
Calculation
1
Browse...
Upload
Download
2
3
Calculate
Exp. Cond.
Comp.
React.
Metab.

Notice :
If you want to use your own input Excel file, choose 'Custom' and select your local input Excel file.
If you want to use already-registered input Excel file, choose 'Registered' and select one of them in the list.

● Custom : Ishii_rpiA_0.1_1.xls
No file selected.

○ Registered :
Keywords :

ID	Name
10	APIsample.xls
238	artcycle1.xls
239	artcycle2.xls
240	artcycle3.xls
64	Ecoli_fbaB_0.1_Ishii_53.xls
11	EcolimetabolicNetwork.xlsx
46	Ecoli_pyk_0.1_Shimizu_53.xls
244	GYLCO11.xls
242	GYLCO1.xls
241	PPP1.xls
47	_test.xls

[User Custom] Ishii_rpiA_0.1_1.xls

Algorithm : ● mCEF ○ ECF ○ GMF
Objective Function : ○ MEP ○ MeanLP ○ LP ○ QP
Ratio Type : ● gene ○ enzyme

Experimental Condition

```
Samples
wild-type: K-12 strain BW25113
disruptant samples: rpiA


Publication
Ishii N*, Nakahigashi K*, Baba T*, Robert M*, Soga T*, Kanai A*, Hirasawa T*, Naba M, Hirai K,
Hoque A, Ho PY, Kakazu Y, Sugawara K, Igarashi S, Harada S, Masuda T, Sugiyama N, Togashi T,
Hasegawa M, Takai Y, Yugi K, Arakawa K, Iwata N, Toya Y, Nakayama Y, Nishioka T, Shimizu K, Mori H,
Tomita M. (2007)
Multiple high-throughput analyses monitor the response of E. coli to perturbations.
Science, 316(5824):593-597.
PMID: 17379776


Notice
Users of the data should carefully read the original manuscript and supplementary methods and text
for additional information
about the measurement values, samples and data series, and the normalization methods used and be
aware of the limitations
before pursuing any analysis using this data.
Failure to do so might result in erroneous interpretations and conclusions for which the authors of
this study cannot be held responsible.


Measurements
Flux
Metabolites-intercelluler
Enzyme activity
Chemost culture
```

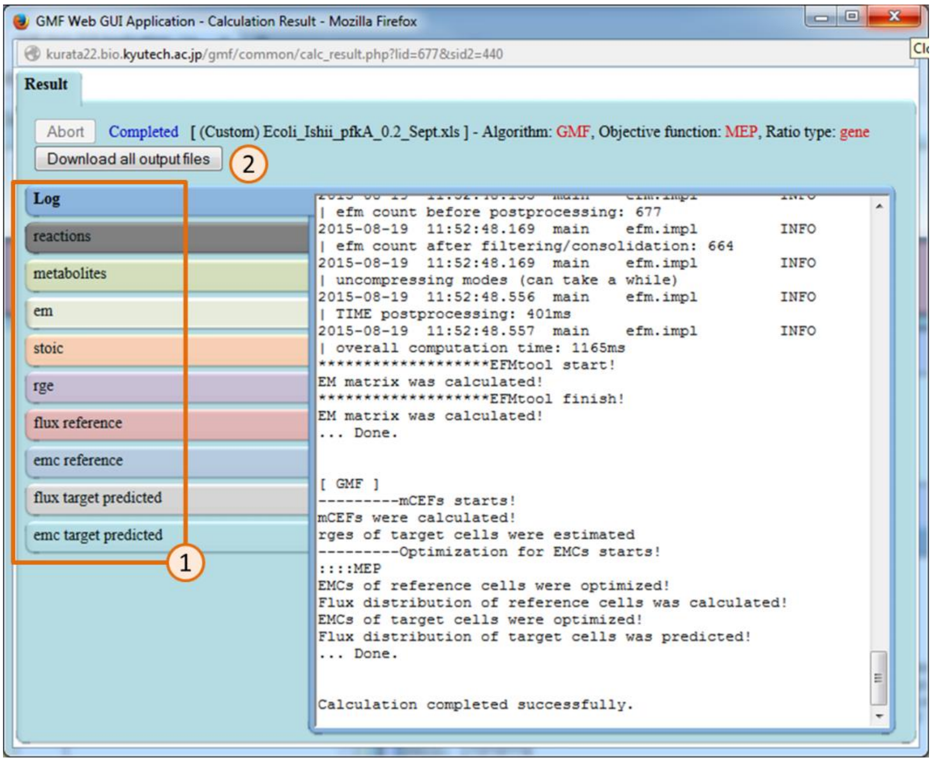
Fig. B3: The GMF main page to calculate user’s file (after a file is uploaded)

2. Click on the desired algorithm and ratio type.
3. Click on the ‘Calculate’ button. The calculation completion time depends on the metabolic network size, algorithms, and objective functions. The calculation completion time is shown as follows.

File name: Ecoli_Ishii_pfkA_0.2_Sept, Reaction number: 48

Calculation completion time (sec)								
GMF				ECF				mCEF
MEP	QP	LP	MeanLP	MEP	QP	LP	MeanLP	
42	37	42	157	42	37	41	157	37

2. Retrieving the Calculated Results


GMF Web GUI Application - Calculation Result - Mozilla Firefox
kurata22.bio.kyutech.ac.jp/gmf/common/calc_result.php?lid=677&sid2=440
Result
Abort
Download all output files
2
Log
reactions
metabolites
em
stoic
rge
flux reference
emc reference
flux target predicted
emc target predicted
1

Completed [(Custom) Ecoli_Ishii_pfkA_0.2_Sept.xls] - Algorithm: GMF, Objective function: MEP, Ratio type: gene

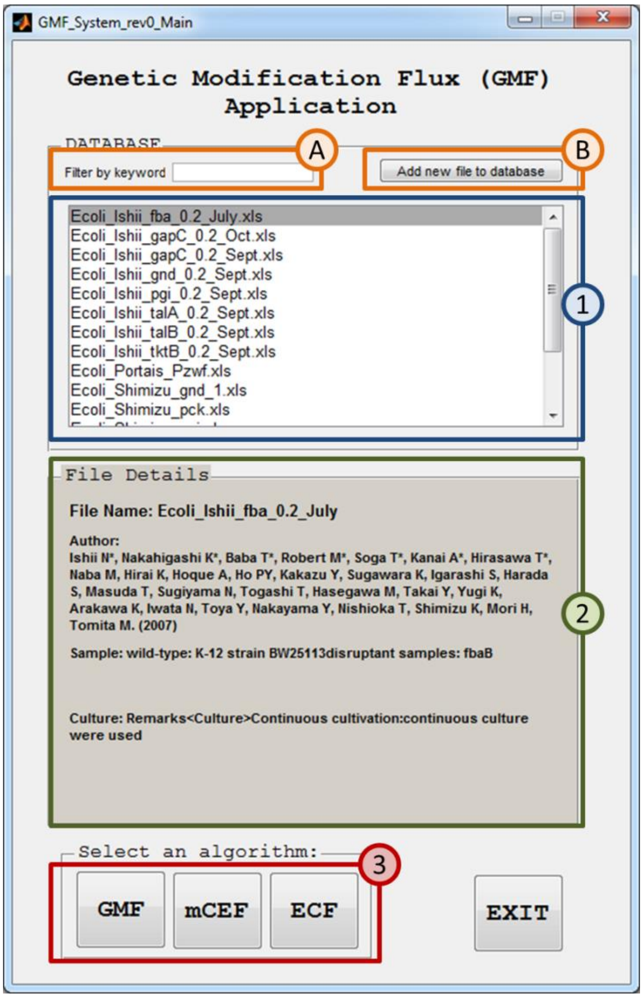
```
| efm count before postprocessing: 677
2015-08-19  11:52:48.169  main    efm.impl        INFO
| efm count after filtering/consolidation: 664
2015-08-19  11:52:48.169  main    efm.impl        INFO
| uncompressing modes (can take a while)
2015-08-19  11:52:48.556  main    efm.impl        INFO
| TIME postprocessing: 401ms
2015-08-19  11:52:48.557  main    efm.impl        INFO
| overall computation time: 1165ms
******************EFMtool start!
EM matrix was calculated!
******************EFMtool finish!
EM matrix was calculated!
... Done.


[ GMF ]
---------mCEFs starts!
mCEFs were calculated!
rges of target cells were estimated
---------Optimization for EMCs starts!
::::MEP
EMCs of reference cells were optimized!
Flux distribution of reference cells was calculated!
EMCs of target cells were optimized!
Flux distribution of target cells was predicted!
... Done.


Calculation completed successfully.
```

Fig. B4: The GMF calculation result page

1. Click on a particular tab to get the desired output.
2. To download all output files, click on ‘Download all output files’.

Matlab Stand-Alone Version

1. Main graphical user interface (GUI)



Ecoli_Ishii_fba_0.2_July.xls
Ecoli_Ishii_gapC_0.2_Oct.xls
Ecoli_Ishii_gapC_0.2_Sept.xls
Ecoli_Ishii_gnd_0.2_Sept.xls
Ecoli_Ishii_pgi_0.2_Sept.xls
Ecoli_Ishii_talA_0.2_Sept.xls
Ecoli_Ishii_talB_0.2_Sept.xls
Ecoli_Ishii_tktB_0.2_Sept.xls
Ecoli_Portais_Pzwf.xls
Ecoli_Shimizu_gnd_1.xls
Ecoli_Shimizu_pck.xls

File Details

File Name: Ecoli_Ishii_fba_0.2_July

Author:
Ishii N*, Nakahigashi K*, Baba T*, Robert M*, Soga T*, Kanai A*, Hirasawa T*, Naba M, Hirai K, Hoque A, Ho PY, Kakazu Y, Sugawara K, Igarashi S, Harada S, Masuda T, Sugiyama N, Togashi T, Hasegawa M, Takai Y, Yugi K, Arakawa K, Iwata N, Toya Y, Nakayama Y, Nishioka T, Shimizu K, Mori H, Tomita M. (2007)

Sample: wild-type: K-12 strain BW25113disruptant samples: fbaB

Culture: Remarks<Culture>Continuous cultivation:continuous culture were used

Fig. B5: The main graphical user interface (GUI) of GMF stand-alone program.

1. The data collection
   (i) Select a file by clicking on a file name from the list.
2. The details of selected file, i.e. Author (s), Sample and Culture are displayed
3. Calculation algorithms selection. Click on the desired algorithm to perform the calculation
4. Additional functions: A: Search a file by using keyword; B: Add a new file to the database

2. Selection of the GMF algorithm

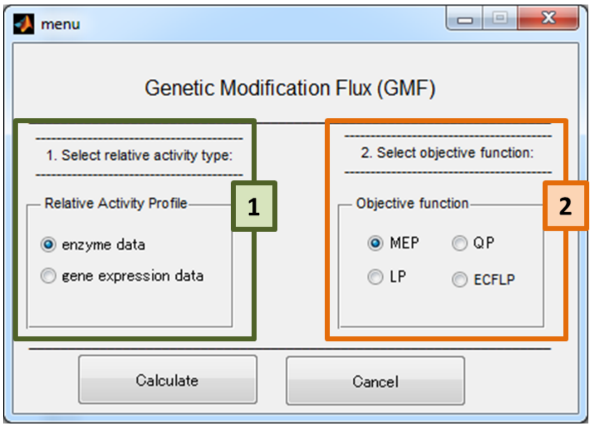
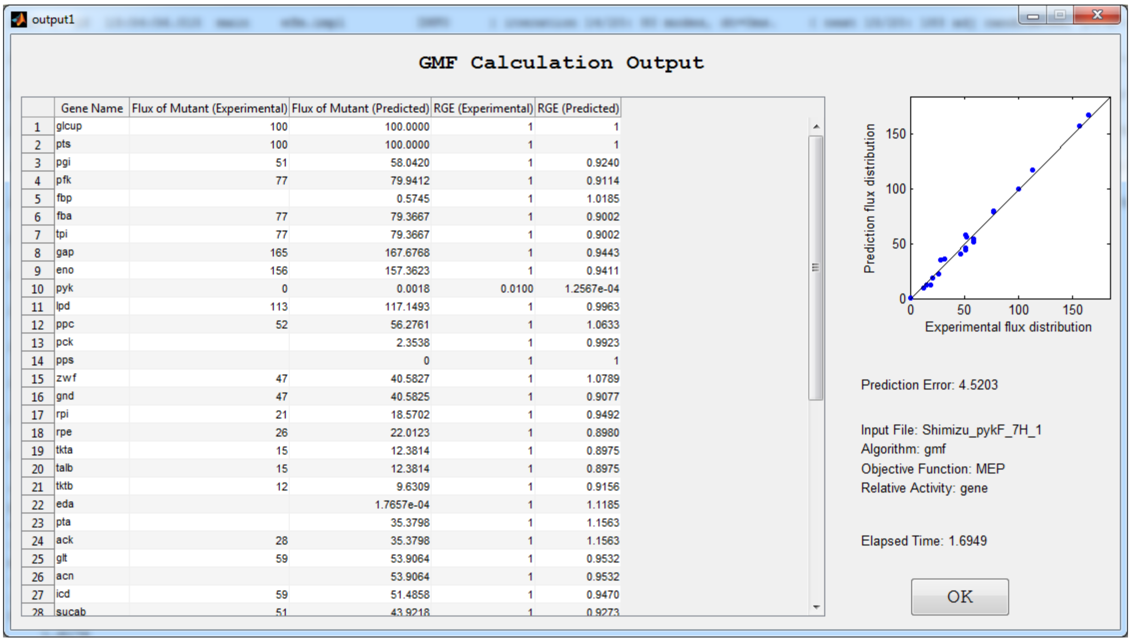

menu
Genetic Modification Flux (GMF)
1. Select relative activity type:
Relative Activity Profile
● enzyme data
○ gene expression data
1
2. Select objective function:
Objective function
● MEP
○ QP
○ LP
○ ECFLP
2
Calculate
Cancel

Fig. B6: The GMF selection menu if GMF algorithm button is selected in Fig. S5

1. Select the relative activity type
2. Select the desired objective function
3. Click on the ‘Calculate’ button



	Gene Name	Flux of Mutant (Experimental)	Flux of Mutant (Predicted)	RGE (Experimental)	RGE (Predicted)
1	glcup	100	100.0000	1	1
2	pts	100	100.0000	1	1
3	pgi	51	58.0420	1	0.9240
4	pfk	77	79.9412	1	0.9114
5	fbp		0.5745	1	1.0185
6	fba	77	79.3667	1	0.9002
7	tpi	77	79.3667	1	0.9002
8	gap	165	167.6768	1	0.9443
9	eno	156	157.3623	1	0.9411
10	pyk	0	0.0018	0.0100	1.2567e-04
11	lpd	113	117.1493	1	0.9963
12	ppc	52	56.2761	1	1.0633
13	pck		2.3538	1	0.9923
14	pps		0	1	1
15	zwf	47	40.5827	1	1.0789
16	gnd	47	40.5825	1	0.9077
17	rpi	21	18.5702	1	0.9492
18	rpe	26	22.0123	1	0.8980
19	tkta	15	12.3814	1	0.8975
20	talb	15	12.3814	1	0.8975
21	tktb	12	9.6309	1	0.9156
22	eda		1.7657e-04	1	1.1185
23	pta		35.3798	1	1.1563
24	ack	28	35.3798	1	1.1563
25	glt	59	53.9064	1	0.9532
26	acn		53.9064	1	0.9532
27	icd	59	51.4858	1	0.9470
28	sucab	51	43.9218	1	0.9273

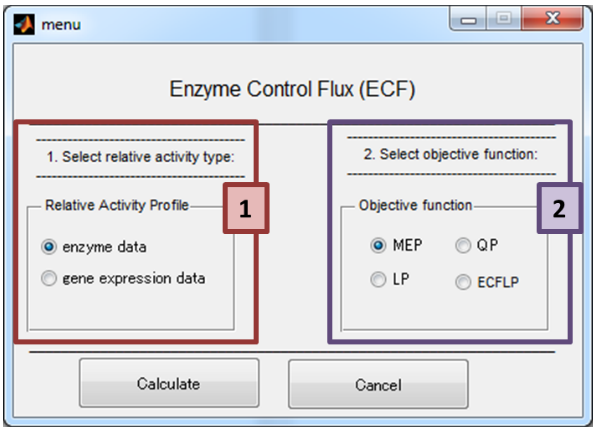
Prediction Error: 4.5203

Input File: Shimizu_pykF_7H_1
Algorithm: gmf
Objective Function: MEP
Relative Activity: gene

Elapsed Time: 1.6949

Fig. B7: The sample of GMF calculation output page

3. Selection of the ECF algorithm



Enzyme Control Flux (ECF)

1. Select relative activity type:

Relative Activity Profile

● enzyme data
○ gene expression data

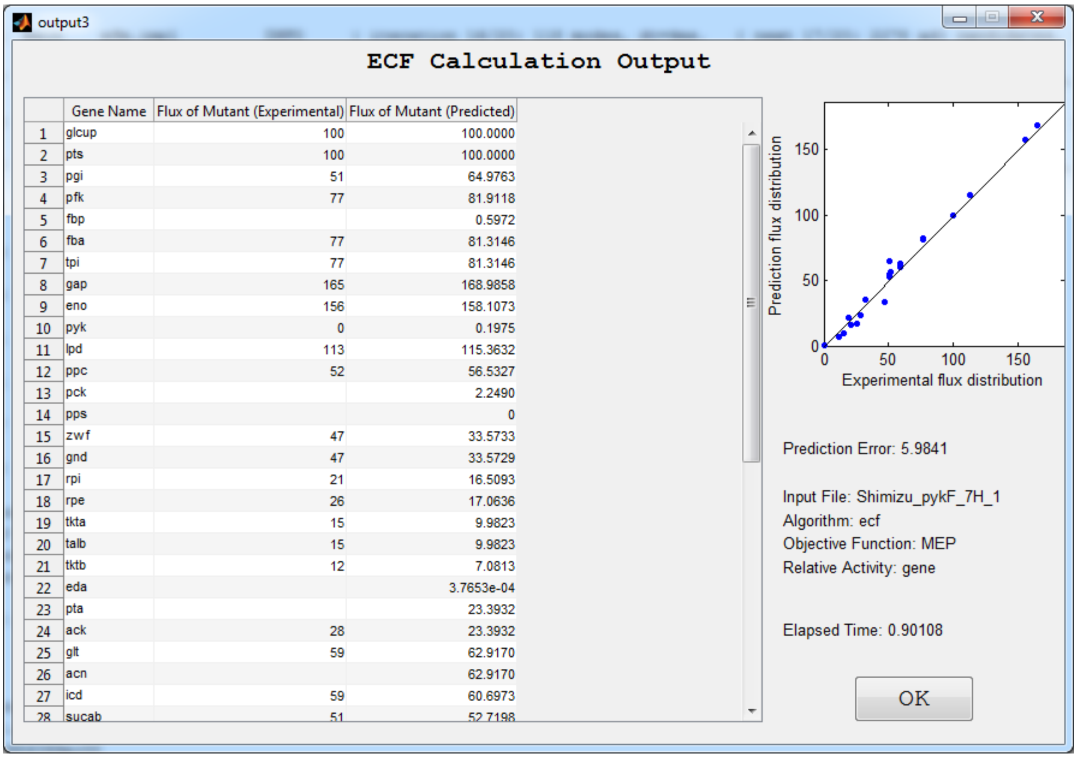
2. Select objective function:

Objective function

● MEP ○ QP
○ LP ○ ECFLP

Fig. B8: The ECF selection menu if ECF algorithm button is selected in Fig. S5

1. Select the relative activity type
2. Select the desired objective function
3. Click on the ‘Calculate’ button



ECF Calculation Output

	Gene Name	Flux of Mutant (Experimental)	Flux of Mutant (Predicted)
1	glcup	100	100.0000
2	pts	100	100.0000
3	pgi	51	64.9763
4	pfk	77	81.9118
5	fbp		0.5972
6	fba	77	81.3146
7	tpi	77	81.3146
8	gap	165	168.9858
9	eno	156	158.1073
10	pyk	0	0.1975
11	lpd	113	115.3632
12	ppc	52	56.5327
13	pck		2.2490
14	pps		0
15	zwf	47	33.5733
16	gnd	47	33.5729
17	rpi	21	16.5093
18	rpe	26	17.0636
19	tkta	15	9.9823
20	talb	15	9.9823
21	tktb	12	7.0813
22	eda		3.7653e-04
23	pta		23.3932
24	ack	28	23.3932
25	glt	59	62.9170
26	acn		62.9170
27	icd	59	60.6973
28	sucab	51	52.7198

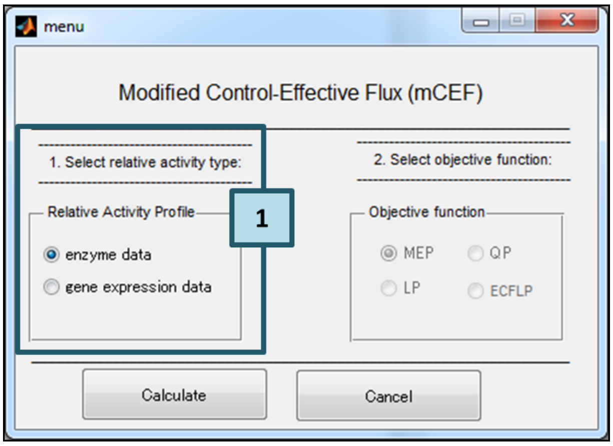
Prediction Error: 5.9841

Input File: Shimizu_pykF_7H_1
Algorithm: ecf
Objective Function: MEP
Relative Activity: gene

Elapsed Time: 0.90108

Fig. B9: The sample of ECF calculation output page

4. Selection of the mCEF algorithm


menu

Modified Control-Effective Flux (mCEF)

1. Select relative activity type:

Relative Activity Profile

● enzyme data

○ gene expression data

2. Select objective function:

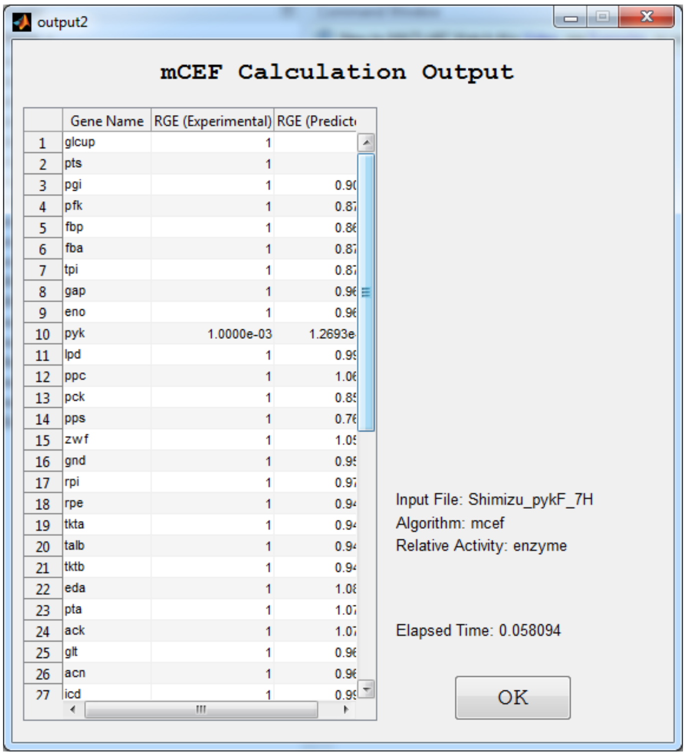
Objective function

● MEP ○ QP

○ LP ○ ECFLP

Fig. B10: The ECF selection menu if ECF algorithm button is selected in Fig. S5

1. Select the relative activity type
2. Click on the ‘Calculate’ button


output2

mCEF Calculation Output

	Gene Name	RGE (Experimental)	RGE (Predict
1	glcup	1	
2	pts	1	
3	pgi	1	0.90
4	pfk	1	0.87
5	fbp	1	0.86
6	fba	1	0.87
7	tpi	1	0.87
8	gap	1	0.96
9	eno	1	0.96
10	pyk	1.0000e-03	1.2693e
11	lpd	1	0.99
12	ppc	1	1.06
13	pck	1	0.85
14	pps	1	0.76
15	zwf	1	1.05
16	gnd	1	0.95
17	rpi	1	0.97
18	rpe	1	0.94
19	tkta	1	0.94
20	talb	1	0.94
21	tktb	1	0.94
22	eda	1	1.08
23	pta	1	1.07
24	ack	1	1.07
25	glt	1	0.96
26	acn	1	0.96
27	icd	1	0.99

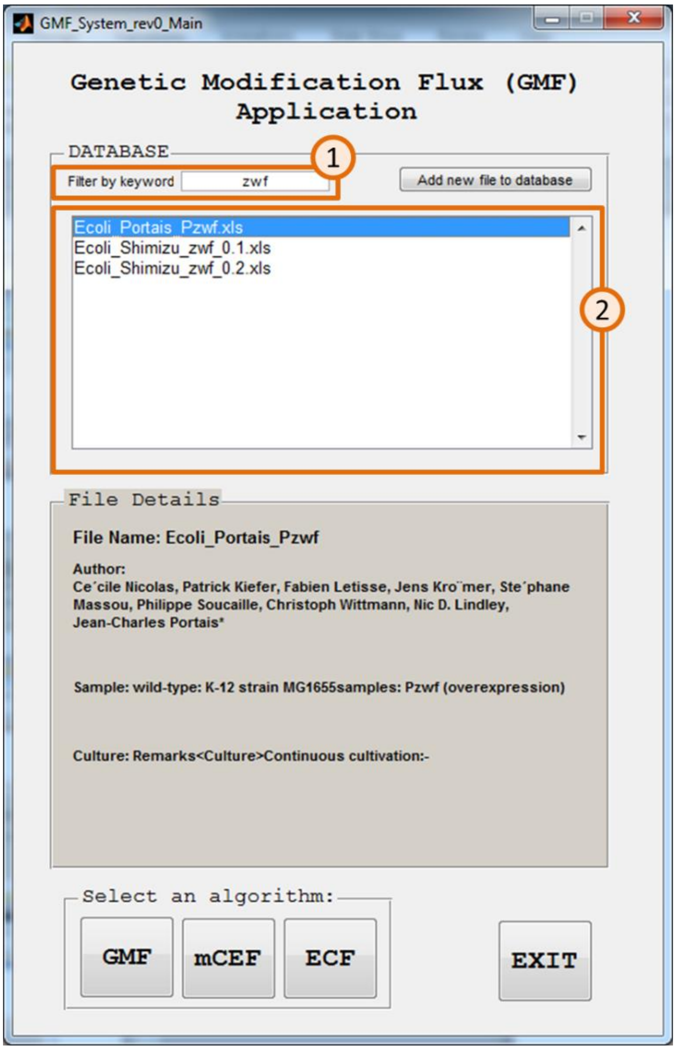
Input File: Shimizu_pykF_7H
Algorithm: mcef
Relative Activity: enzyme

Elapsed Time: 0.058094

Fig. B11:The sample of mCEF calculation output page

5. Additional functions

(a) To search a file by using keyword



Filter by keyword zwf

Ecoli_Portais_Pzwf.xls
Ecoli_Shimizu_zwf_0.1.xls
Ecoli_Shimizu_zwf_0.2.xls

File Details

File Name: Ecoli_Portais_Pzwf

Author:
Ce´cile Nicolas, Patrick Kiefer, Fabien Letisse, Jens Kro¨mer, Ste´phane Massou, Philippe Soucaille, Christoph Wittmann, Nic D. Lindley, Jean-Charles Portais*

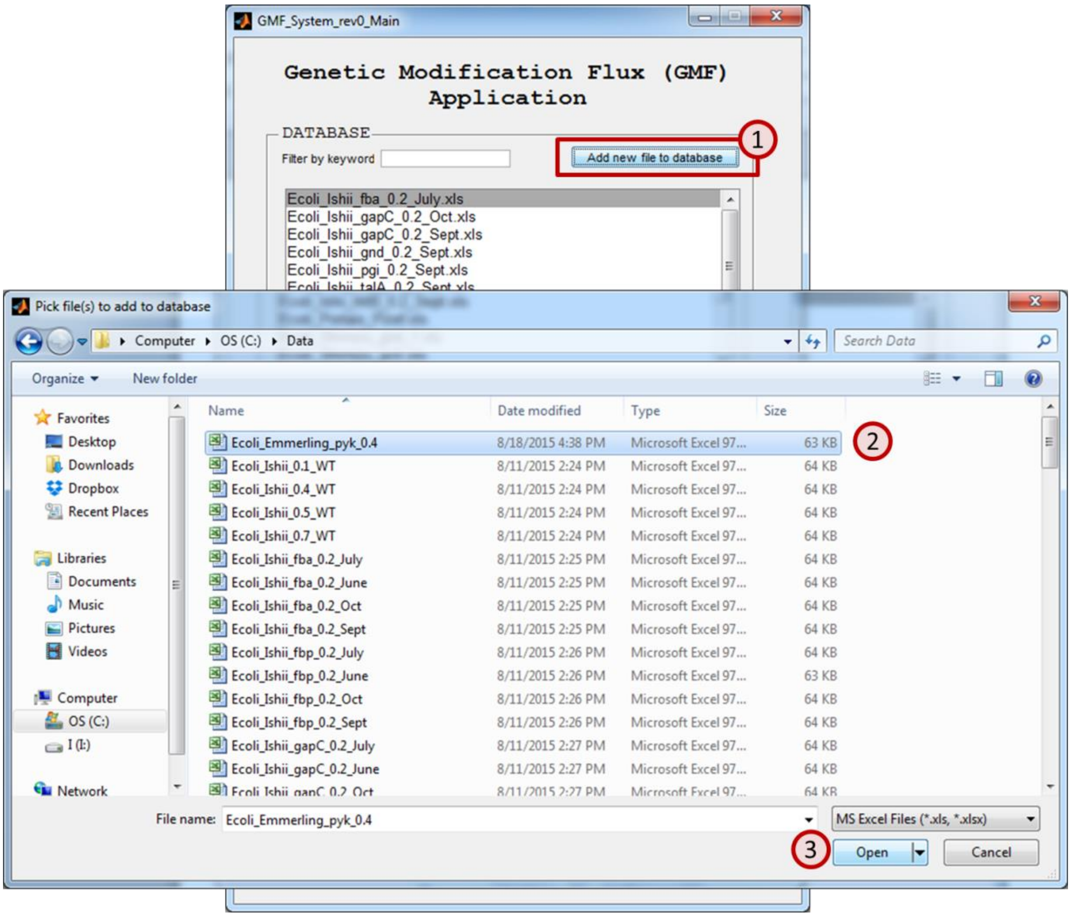
Sample: wild-type: K-12 strain MG1655samples: Pzwf (overexpression)

Culture: Remarks<Culture>Continuous cultivation:-

Fig. B12:The main GUI

1. Put a keyword and press the ‘Enter’ key
2. The files that match to the keyword are listed in 2
3. To perform the calculation, click on the desired file and algorithm

(b) Adding new file to the database



Ecoli_Ishii_fba_0.2_July.xls
Ecoli_Ishii_gapC_0.2_Oct.xls
Ecoli_Ishii_gapC_0.2_Sept.xls
Ecoli_Ishii_gnd_0.2_Sept.xls
Ecoli_Ishii_pgi_0.2_Sept.xls
Ecoli_Ishii_talA_0.2_Sept.xls

Name	Date modified	Type	Size
Ecoli_Emmerling_pyk_0.4	8/18/2015 4:38 PM	Microsoft Excel 97...	63 KB
Ecoli_Ishii_0.1_WT	8/11/2015 2:24 PM	Microsoft Excel 97...	64 KB
Ecoli_Ishii_0.4_WT	8/11/2015 2:24 PM	Microsoft Excel 97...	64 KB
Ecoli_Ishii_0.5_WT	8/11/2015 2:24 PM	Microsoft Excel 97...	64 KB
Ecoli_Ishii_0.7_WT	8/11/2015 2:24 PM	Microsoft Excel 97...	64 KB
Ecoli_Ishii_fba_0.2_July	8/11/2015 2:25 PM	Microsoft Excel 97...	64 KB
Ecoli_Ishii_fba_0.2_June	8/11/2015 2:25 PM	Microsoft Excel 97...	64 KB
Ecoli_Ishii_fba_0.2_Oct	8/11/2015 2:25 PM	Microsoft Excel 97...	64 KB
Ecoli_Ishii_fba_0.2_Sept	8/11/2015 2:25 PM	Microsoft Excel 97...	64 KB
Ecoli_Ishii_fbp_0.2_July	8/11/2015 2:26 PM	Microsoft Excel 97...	64 KB
Ecoli_Ishii_fbp_0.2_June	8/11/2015 2:26 PM	Microsoft Excel 97...	63 KB
Ecoli_Ishii_fbp_0.2_Oct	8/11/2015 2:26 PM	Microsoft Excel 97...	64 KB
Ecoli_Ishii_fbp_0.2_Sept	8/11/2015 2:26 PM	Microsoft Excel 97...	64 KB
Ecoli_Ishii_gapC_0.2_July	8/11/2015 2:27 PM	Microsoft Excel 97...	64 KB
Ecoli_Ishii_gapC_0.2_June	8/11/2015 2:27 PM	Microsoft Excel 97...	64 KB
Ecoli_Ishii_gapC_0.2_Oct	8/11/2015 2:27 PM	Microsoft Excel 97...	64 KB

File name: Ecoli_Emmerling_pyk_0.4

MS Excel Files (*.xls, *.xlsx)

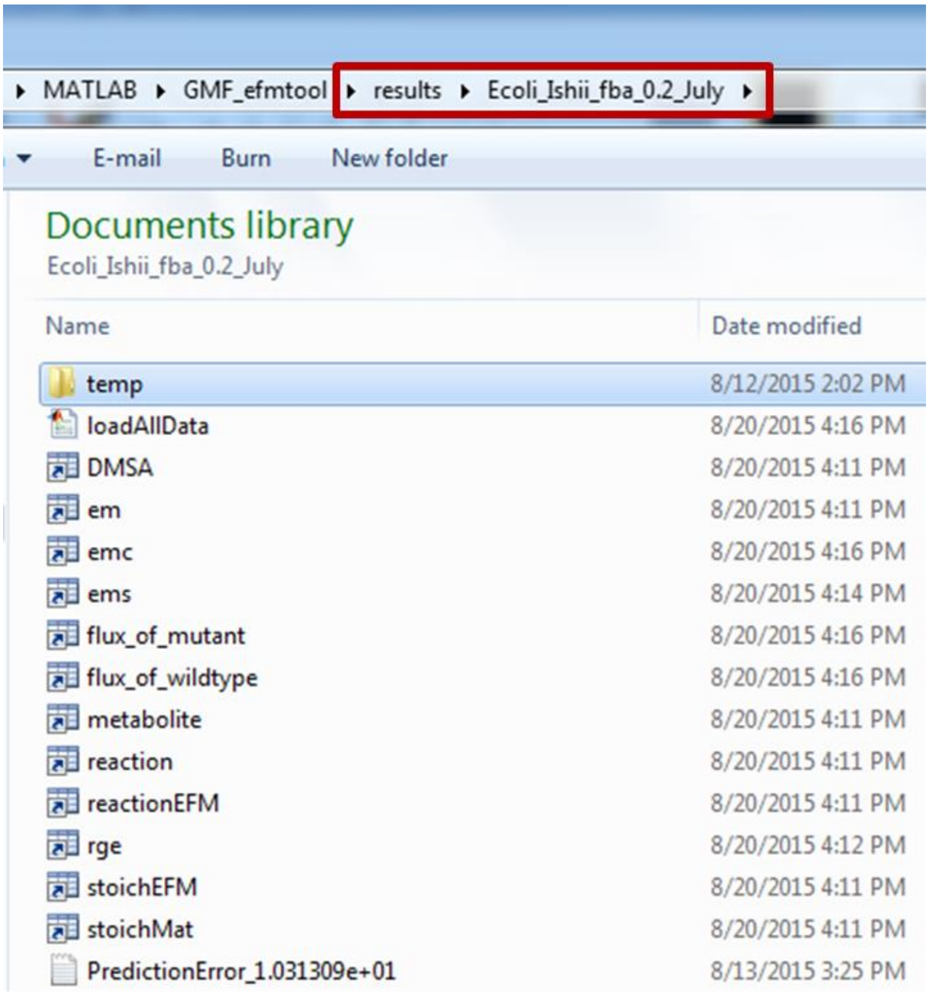
Fig. B13: The sample of file filtering by using keyword in main GUI

1. Click on 'Add new file to database' button
2. Click on a new file from the computer drive
3. Click on Open button

6. Retrieving the Calculated Results

   (a) MATLAB

      1. Refer to `\results\<Excel_file_name>` folder



Documents library

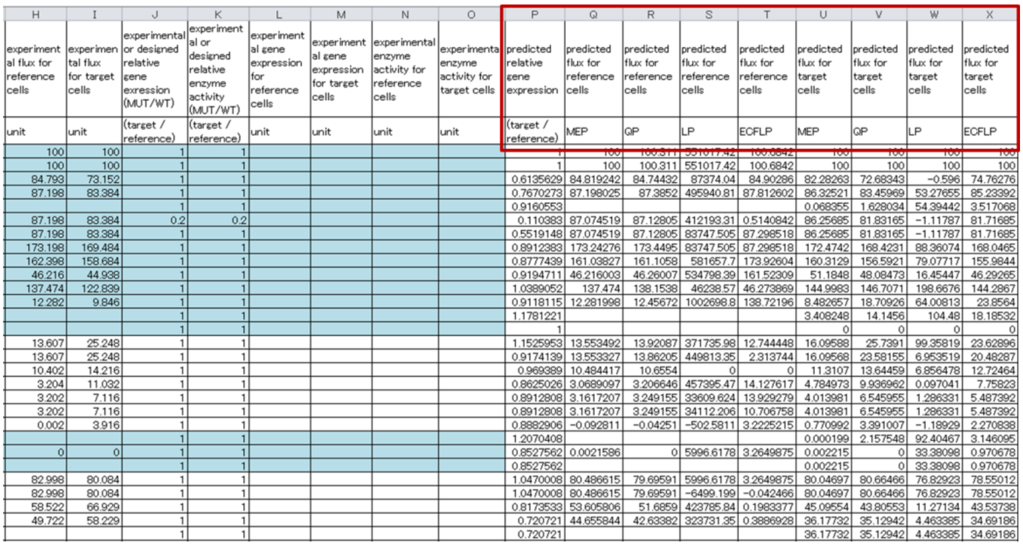
Ecoli_Ishii_fba_0.2_July

Name	Date modified
temp	8/12/2015 2:02 PM
loadAllData	8/20/2015 4:16 PM
DMSA	8/20/2015 4:11 PM
em	8/20/2015 4:11 PM
emc	8/20/2015 4:16 PM
ems	8/20/2015 4:14 PM
flux_of_mutant	8/20/2015 4:16 PM
flux_of_wildtype	8/20/2015 4:16 PM
metabolite	8/20/2015 4:11 PM
reaction	8/20/2015 4:11 PM
reactionEFM	8/20/2015 4:11 PM
rge	8/20/2015 4:12 PM
stoichEFM	8/20/2015 4:11 PM
stoichMat	8/20/2015 4:11 PM
PredictionError_1.031309e+01	8/13/2015 3:25 PM

Fig. B14: The sample of calculated result in ‘results’ folder

(b) MS Excel file

1. Refer to column P – Q in corresponding input file



H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
experimental flux for reference cells	experimental flux for target cells	experimental or designed relative gene exression (MUT/WT)	experimental or designed relative enzyme activity (MUT/WT)	experimental gene expression for reference cells	experimental gene expression for target cells	experimental enzyme activity for reference cells	experimental enzyme activity for target cells	predicted relative gene expression	predicted flux for reference cells	predicted flux for reference cells	predicted flux for reference cells	predicted flux for reference cells	predicted flux for target cells	predicted flux for target cells	predicted flux for target cells	predicted flux for target cells
unit	unit	(target / reference)	(target / reference)	unit	unit	unit	unit	(target / reference)	MEP	QP	LP	ECFLP	MEP	QP	LP	ECFLP
100	100	1	1					1	100	100.311	551017.42	100.6842	100	100	100	100
100	100	1	1					1	100	100.311	551017.42	100.6842	100	100	100	100
84.793	73.152	1	1					0.6135629	84.819242	84.74432	87374.04	84.90286	82.28263	72.68343	-0.596	74.76276
87.198	83.384	1	1					0.7670273	87.198025	87.3852	495940.81	87.812602	86.32521	83.45969	53.27655	85.23392
		1	1					0.9160553					0.068355	1.628034	54.39442	3.517068
87.198	83.384	0.2	0.2					0.110383	87.074519	87.12805	412193.31	0.5140842	86.25685	81.83165	-1.11787	81.71685
87.198	83.384	1	1					0.5519148	87.074519	87.12805	83747.505	87.298518	86.25685	81.83165	-1.11787	81.71685
173.198	169.484	1	1					0.8912383	173.24276	173.4495	83747.505	87.298518	172.4742	168.4231	88.36074	168.0465
162.398	158.684	1	1					0.8777439	161.03827	161.1058	581657.7	173.92604	160.3129	156.5921	79.07717	155.9844
46.216	44.938	1	1					0.9194711	46.216003	46.26007	534798.39	161.52309	51.1848	48.08473	16.45447	46.29265
137.474	122.839	1	1					1.0389052	137.474	138.1538	46238.57	46.273869	144.9983	146.7071	198.6676	144.2867
12.282	9.846	1	1					0.9118115	12.281998	12.45672	1002698.8	138.72196	8.482657	18.70926	64.00813	23.8564
		1	1					1.1781221					3.408248	14.1456	104.48	18.18532
		1	1					1					0	0	0	0
13.607	25.248	1	1					1.1525953	13.553492	13.92087	371735.98	12.744448	16.09588	25.7391	99.35819	23.62896
13.607	25.248	1	1					0.9174139	13.553327	13.86205	449813.35	2.313744	16.09568	23.58155	6.953519	20.48287
10.402	14.216	1	1					0.969389	10.484417	10.6554	0	0	11.3107	13.64459	6.856478	12.72464
3.204	11.032	1	1					0.8625026	3.0689097	3.206646	457395.47	14.127617	4.784973	9.936962	0.097041	7.75823
3.202	7.116	1	1					0.8912808	3.1617207	3.249155	33609.624	13.929279	4.013981	6.545955	1.286331	5.487392
3.202	7.116	1	1					0.8912808	3.1617207	3.249155	34112.206	10.706758	4.013981	6.545955	1.286331	5.487392
0.002	3.916	1	1					0.8882906	-0.092811	-0.04251	-502.5811	3.2225215	0.770992	3.391007	-1.18929	2.270838
		1	1					1.2070408					0.000199	2.157548	92.40467	3.146095
0	0	1	1					0.8527562	0.0021586	0	5996.6178	3.2649875	0.002215	0	33.38098	0.970678
		1	1					0.8527562					0.002215	0	33.38098	0.970678
82.998	80.084	1	1					1.0470008	80.486615	79.69591	5996.6178	3.2649875	80.04697	80.66466	76.82923	78.55012
82.998	80.084	1	1					1.0470008	80.486615	79.69591	-6499.199	-0.042466	80.04697	80.66466	76.82923	78.55012
58.522	66.929	1	1					0.8173533	53.605806	51.6859	423785.84	0.1983377	45.09554	43.80553	11.27134	43.53738
49.722	58.229	1	1					0.720721	44.655844	42.63382	323731.35	0.3886928	36.17732	35.12942	4.463385	34.69186
		1	1					0.720721					36.17732	35.12942	4.463385	34.69186

Fig. B15: The sample of calculated result in MS Excel file

List of Abbreviations

API	: Application Programming Interface
B.subtilis	: *Bacillus Subtilis*
C.glutamicum	: *Corynebacterium glutamicum*
CHO	: *Chinese Hamster Ovary*
CNA	: CellNetAnalyzer
DNA	: Deoxyribonucleic acid
E.coli	: *Escherichia coli*
ECF	: Enzyme control flux
ECFLP	: Enzyme control flux linear programming
EM	: Elementary Mode
EMC	: Elementary mode coefficient
FBA	: Flux Balance Analysis
GEM	: genome-scale model
GMF	: Genetic modification of flux
KEGG	: Kyoto Encyclopedia of Genes and Genomes
LP	: Linear programming
MEP	: Maximum entropy principle
MFA	: Metabolic Flux Analysis
MILP	: Mixed-integer linear programming
MOMA	: Minimization of Metabolic Adjustment
MeanLP	: Linear Programming based on alpha spectrum
QP	: Quadratic programming
ROOM	: Regulatory On/Off Minimization
S.cerevisiae	: *Saccharomyces cerevisiae*
efmtool	: Elementary flux mode tool
mCEF	: Modified control effective flux

List of Tables

List of Figures

Acknowledgement

My endless appreciations goes to the respective supervisor; Professor Hiroyuki Kurata for the guidance, patience, encouragement, and immerse knowledge during this study.

Special thanks to the thesis committee; Professor Yamazaki Toshimasa, Professor Sueda Shinji, and Professor Ohashi Takeshi for the insightful comments and encouragement, critical-thinking enquiries to widen the study perspectives.

This warm appreciation also dedicates to Associate Professor Kazuhiro Maeda and Yu Matsuoka for the guidance and stimulating discussions, that really helpful in completing the study. I owe my gratitude to all Kurata Laboratory members for the help, support and team working spirit that we shared in the past memorable years.

Special appreciation is extended to Malaysia Government for providing the sponsorship and support throughout the study.