About OMICS Group

OMICS Group International is an amalgamation of Open Access publications and worldwide international science conferences and events. Established in the year 2007 with the sole aim of making the information on Sciences and technology 'Open Access', OMICS Group publishes 400 online open access scholarly journals in all aspects of Science, Engineering, Management and Technology journals. OMICS Group has been instrumental in taking the knowledge on Science & technology to the doorsteps of ordinary men and women. Research Scholars, Students, Libraries, Educational Institutions, Research centers and the industry are main stakeholders that benefitted greatly from this knowledge dissemination. OMICS Group also organizes 300 International conferences annually across the globe, where knowledge transfer takes place through debates, round table discussions, poster presentations, workshops, symposia and exhibitions.

About OMICS Group Conferences

OMICS Group International is a pioneer and leading science event organizer, which publishes around 400 open access journals and conducts over 300 Medical, Clinical, Engineering, Life Sciences, Pharma scientific conferences all over the globe annually with the support of more than 1000 scientific associations and 30,000 editorial board members and 3.5 million followers to its credit.

OMICS Group has organized 500 conferences, workshops and national symposiums across the major cities including San Francisco, Las Vegas, San Antonio, Omaha, Orlando, Raleigh, Santa Clara, Chicago, Philadelphia, Baltimore, United Kingdom, Valencia, Dubai, Beijing, Hyderabad, Bengaluru and Mumbai.

The 3rd International Conference and Exhibition on Metabolomics & Systems Biology March 24-26, 2014 San Antonio, USA

"Integration of omics into metabolic flux distribution by complementary elementary mode analysis for large-scale metabolic networks"

Md. Bahadur Badsha* and Hiroyuki Kurata

Department of Bioscience and Bioinformatics
Kyushu Institute of Technology (Kyutech),Japan
mbahadur_stat_ru@yahoo.com
m791703b@bio.kyutech.ac.jp

Kyutech
Kyushu Institute of Technology

Contents

1. Background

2. Problems and objectives

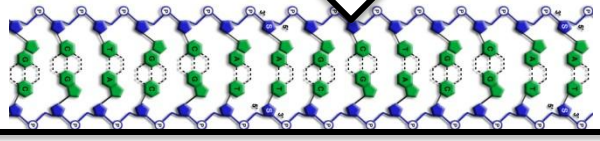
3. Method

4. Results and discussion

5. Concluding remark

Introduction

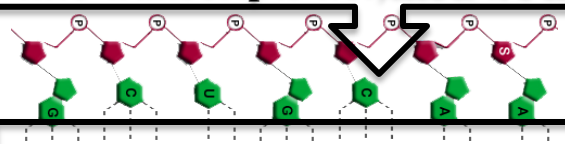
Entirety of all genes under the given studies
i.e., **DNA sequence of an organism**



DNA

Genomics – 40,000 Genes

Entirety of all genes that are converted into transcripts i.e.,**mRNA molecules**



RNA

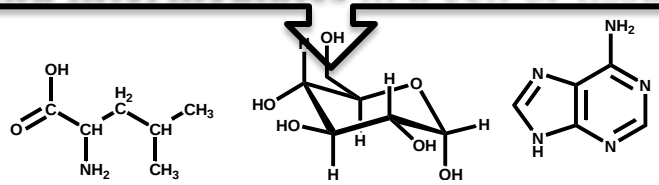
Transcriptomics – 1,00000 Transcripts

Entirety of **all proteins** found in a given cell or tissue

Protein

Proteomics – 10,00,000 Proteins

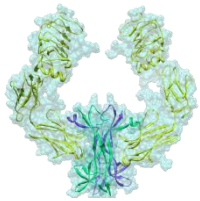
Entirety of **all metabolism products and intermediates** in a cell or tissue



Biochemicals (Metabolites)

Metabolomics – 2,400 Compounds

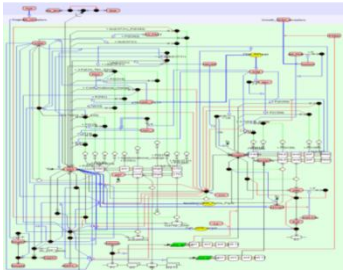
Research Concept



Genome
Omics
Molecular Biology
Medical data

Bioinformatics (Top-down): Parameter values are estimated to reproduce biological behaviors

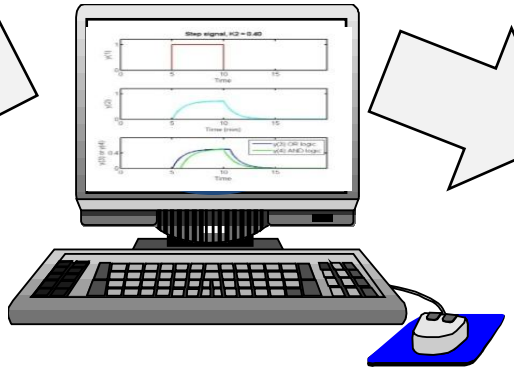
Systems Biology (Bottom-up): Mechanistic model construction based on molecular architecture



$$\frac{dy(1)}{dt} = 0$$

$$\frac{dy(2)}{dt} = km(2) \cdot \frac{y(1)}{K(2) + y(1)} - kd(1) \cdot y(2)$$

$$\frac{dy(3)}{dt} = km(3) \cdot OR(y(1), y(2), K(1), K(2)) - kd(1) \cdot y(3)$$



$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \end{pmatrix}$$

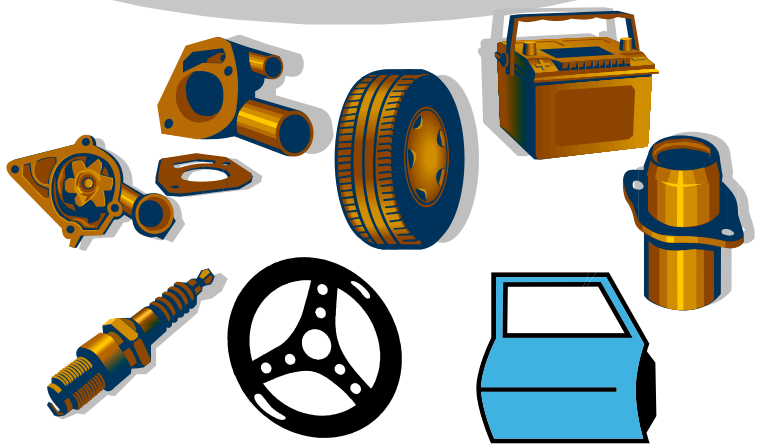
Computer-aided drug design/development; disease diagnosis, prevention, and treatment; and patient care/welfare

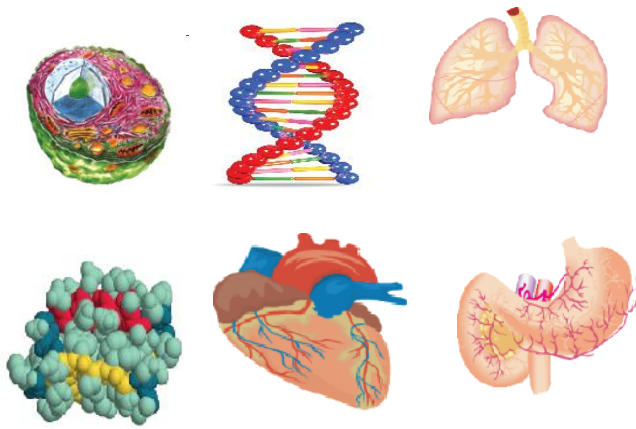
Seeds

Needs from Medical school, Hospitals, and Companies

Bioengineering (Problem-based Research)

Compare two systems





- We want to understand the system, we want to repair when some parts are failing; we want to improve the performance more.
- For example in human body, How genes and proteins that are working in the organism has been accumulated, such that to understand as a system of complex biological reaction is the systems biology.

Background

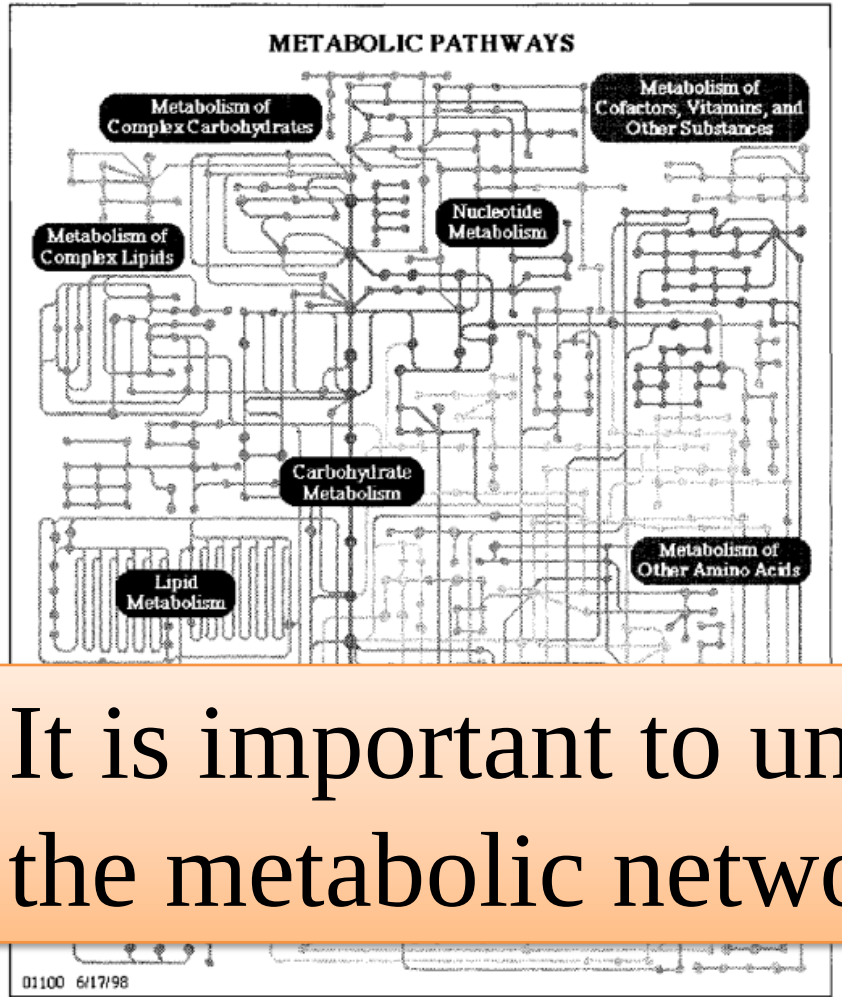


Figure 1:Metabolic pathway MAP

Metabolism is an important **biological processes**

…But

These are complex and highly interconnected

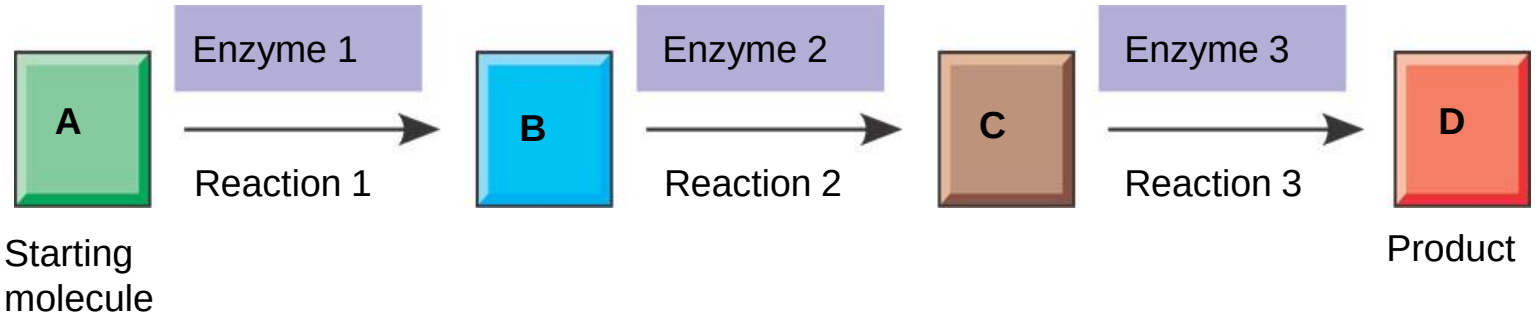
Metabolic disorder is related to a disease

- Diabetes
- high blood pressure
- Cancer …etc

It is important to understand properly of the metabolic networks using EM.

Metabolic Pathways

- Biochemical pathways are the organizational units of metabolism.
- Metabolism is the total of all chemical reactions carried out by an organism.
- A metabolic pathway has many steps that begin with a specific molecule and end with a product, each catalyzed by a specific enzyme.



Elementary mode analysis

Network-based metabolic pathway analysis:

- ❖ Elementary mode (EM) analysis
- ❖ Extreme pathways (ExP) analysis

- ❑ EM is the minimal set of enzymes that can operate at steady state, while the set of extreme pathways is the systemically independent subset of the EMs.

- ❑ The EM coefficients (EMCs) indicate the quantitative contribution of their associated EMs and can be estimated by the maximizing the general objective function.

Problems

- Ordinary EM analysis is that the number of EMs suffers from a combinatorial explosion.
- EMs can be described by many scalar products of each EM, the predicted fluxes must be independent of them.
- Many organisms still do not provide any specific objective biological function.

Objectives

To overcome the existing problems, we proposed a several times faster and efficient EM algorithm named the complementary EMs (cEM) analysis.

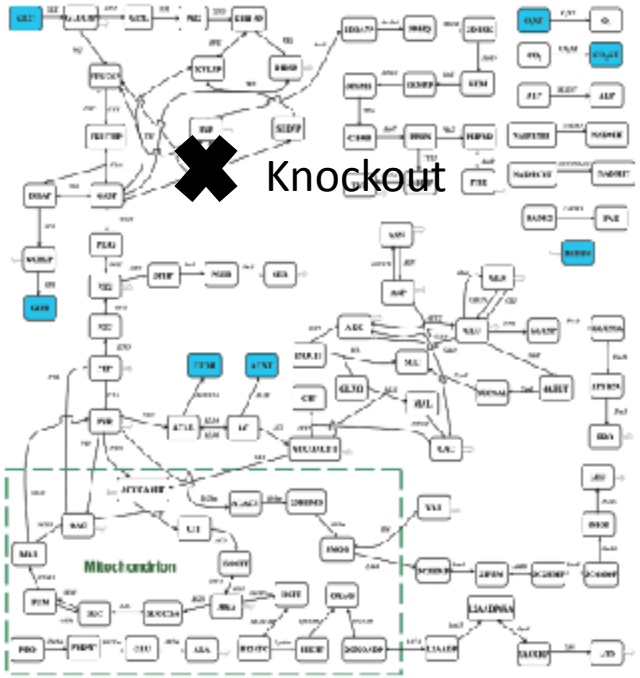
❖EM decomposition [1] which generates the major EMs.

❖Alpha-spectrum [2] can be computed even when the flux is partially unknown.

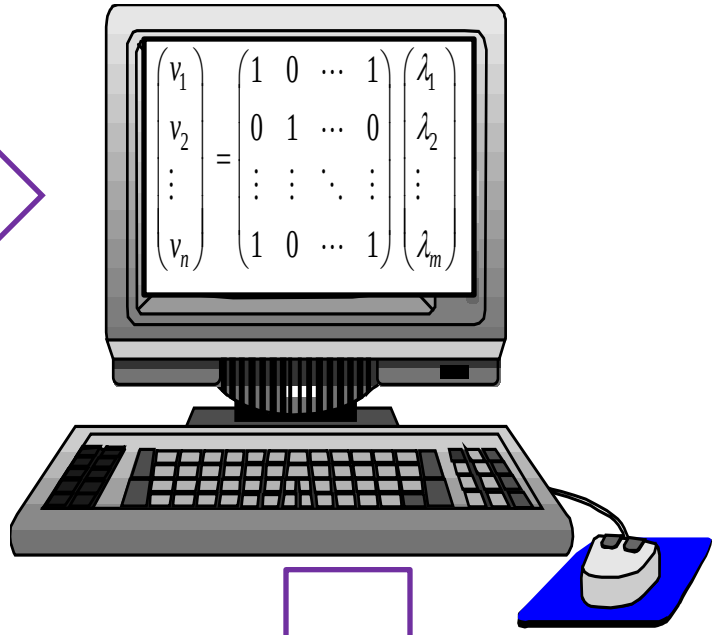
❖MEP [3] to optimize the EMCs

General process

Genetically modified mutant

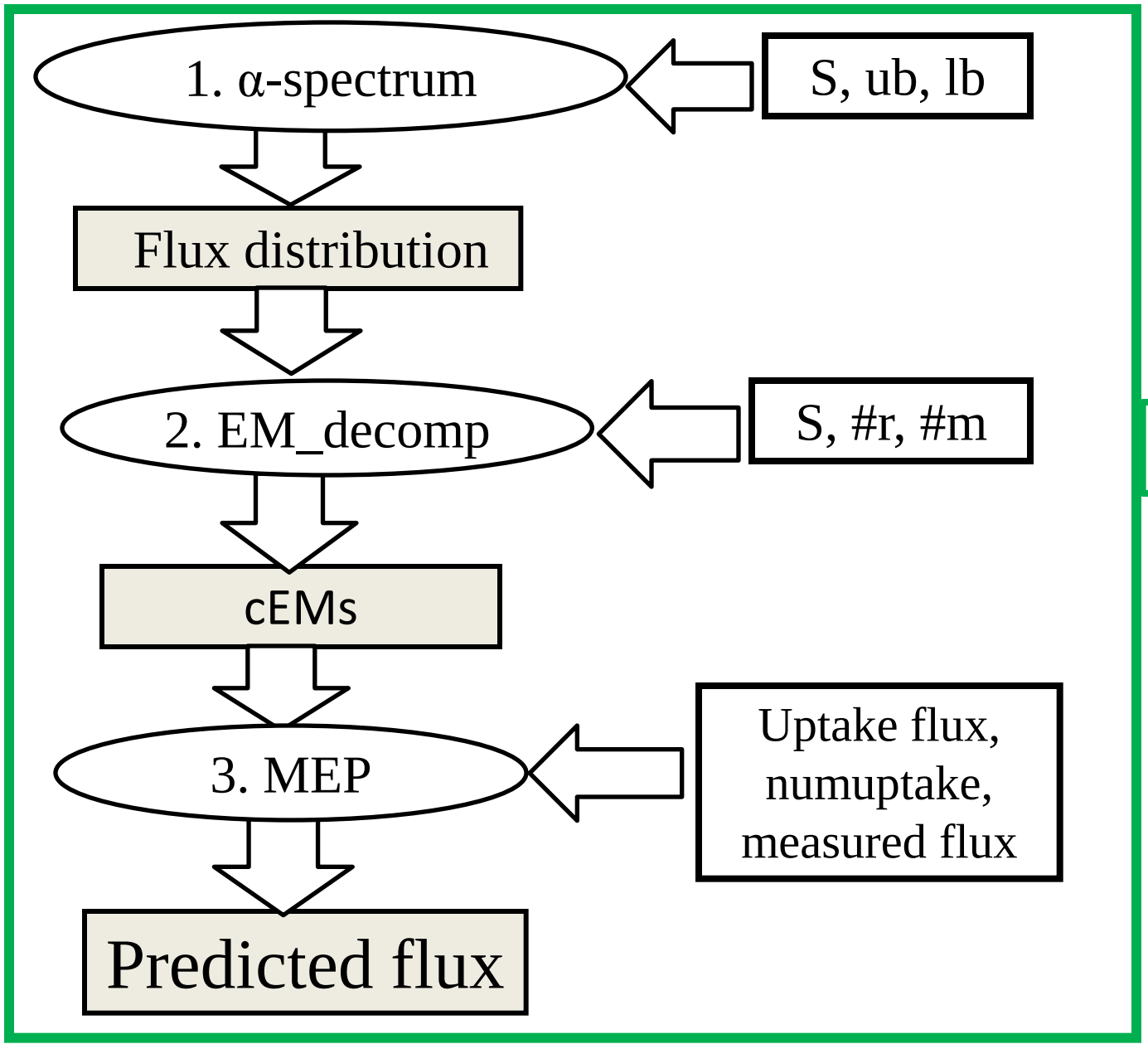


cEM analysis



Prediction of flux distribution

A flow chart of the cEM analysis



The white and grey square boxes are the data and ovals are the algorithms.

Alpha spectrum method [2]

Alpha spectrum method is applied to optimize the flux distribution. It is defined by:

$$\text{Max/Min} \quad \mathbf{v}_i = (v_1, v_2, ..., v_n)^t$$

$$\text{where, i} = 1, 2, 3, \ \ldots, \ n \ ; \qquad \begin{cases} -1000 < v_i^{rev} < 1000 \\ 1e-8 < v_i^{irrev} < 1000 \end{cases}$$

$$\text{Subject to: } \mathbf{S} \cdot \mathbf{v} = 0$$

Input: Stoichiometric matrix, ub, lb

Output: V_{max}, V_{min}

EM decomposition method [1]

- EM decomposition method is applied to generate the major EMs responsible for flux distribution.
- Mixed integer linear programming (MILP)
- It is an iterative method.

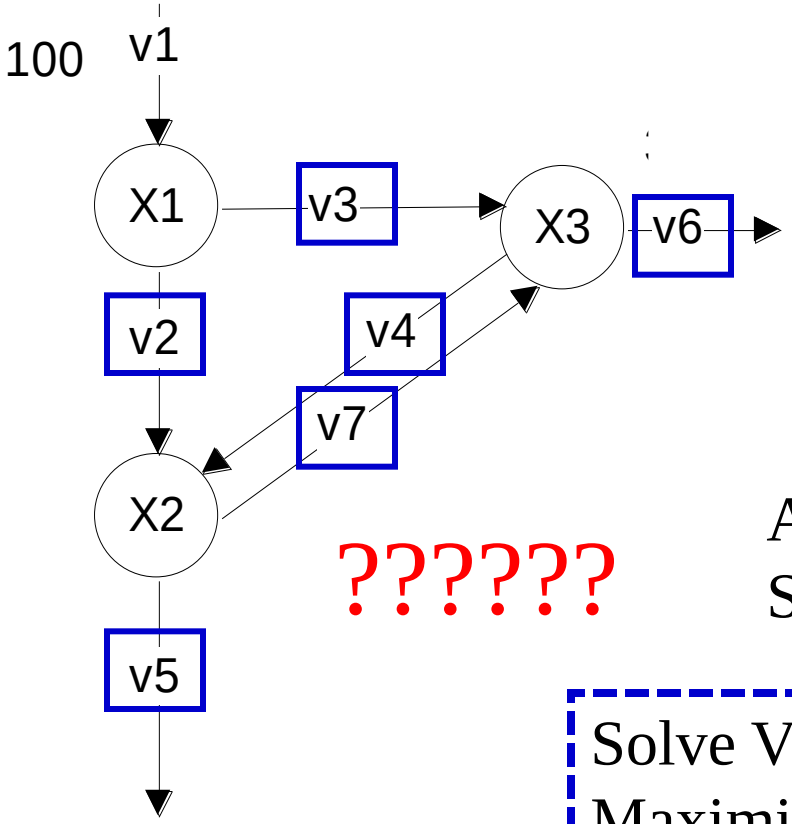
Input: Flux distribution, S, #reaction, #metabolites

Output: EMs

Different objective functions

Methods	Objective Function	Advantage/Disadvantage
LP	$\max \; v_{biomass} = \sum_{i=1}^{ne} p_{biomass,i} \cdot \lambda_i$	Max biomass or ATP production are can vary between different organisms and physiological conditions. Thus the objective function are not good choice.
QP	$\min \sum_{i=1}^{ne} \lambda_i^2$	Suitable when objective function cannot be defined in biological terms. Restricted for small or medium scale and depends on scalar product of EM.
ECFLP	$Max / Min \; \lambda_i \;\; (i = 1,2,\ldots,ne)$	ECFLP is not theoretical but empirical.
MEP	$\max \; -\sum_{i=1}^{ne} \rho_i \ln \rho_i$ s.t. $\sum_{i=1}^{ne} \rho_i = 1; \; \sum_{i=1}^{ne} \rho_i x_{d,i} = v_d$	Straight forward formula and theoretically sound background. No need additional biological hypothesis or objective function and not depends on scalar product of EM

Elementary mode analysis



Example model

Reaction V= (v_1,v_2,v_3,v_4,v_5,v_6,v_7)

Metabolites = X1, X2, X3

$$S = \begin{pmatrix} 1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & -1 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{pmatrix}$$

At steady-state mass balance equation as $S \cdot V = 0$

Solve V:

Maximize or minimize V_i , ($V_i >= 0$, i=2…7)

v_1=100; subject to $S \cdot V = 0$

12 sets of flux distributions will be solved.

$$\mathrm{EM}(\mathbf{P}) = \begin{pmatrix} \mathrm{EM1} & \mathrm{EM2} & \mathrm{EM3} & \mathrm{EM4} & \mathrm{EM5} \\ 0.5774 & 0.5774 & 0.5 & 0.5 & 0 \\ 0.5774 & 0 & 0 & 0.5 & 0 \\ 0 & 0.5774 & 0.5 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & 0.7071 \\ 0.5774 & 0 & 0.5 & 0 & 0 \\ 0 & 0.5774 & 0 & 0.5 & 0 \\ 0 & 0 & 0 & 0.5 & 0.7071 \end{pmatrix}_{7\times5}$$

MEP

EMC (λ)
43.30
43.30
50
50
0

Quantitative contributions $=\lambda_i \cdot P(\text{numuptake}, i)$; where, i=number of EM and numuptake =the row in EM matrix P for uptake / input flux.

The 4 cEM and flux distributions by cEM analysis are as follows:

$$\mathrm{cEM} = \begin{pmatrix} \mathrm{EM1} & \mathrm{EM2} & \mathrm{EM3} & \mathrm{EM4} \\ 0.5774 & 0.5774 & 0.5 & 0.5 \\ 0.5774 & 0 & 0 & 0.5 \\ 0 & 0.5774 & 0.5 & 0 \\ 0 & 0 & 0.5 & 0 \\ 0.5774 & 0 & 0.5 & 0 \\ 0 & 0.5774 & 0 & 0.5 \\ 0 & 0 & 0 & 0.5 \end{pmatrix}_{7\times4}$$

Flux distributions
100
50
50
25
50
50
25

Application in large-scale metabolic model

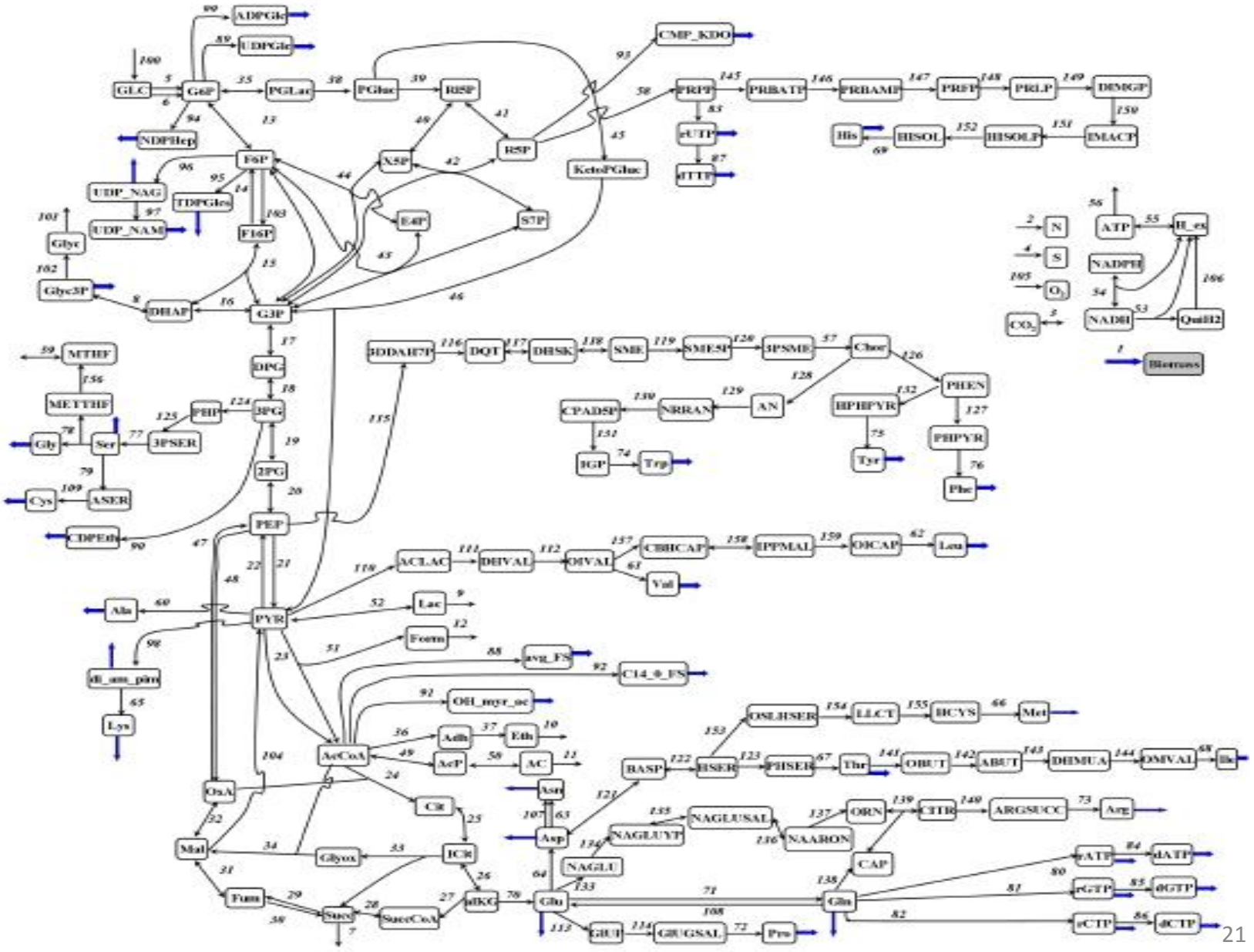
Model-I

The model-I was involved 140 metabolites and 156 reactions for the *E. coli pta-pfkA* gene knockout mutant undergoing adaptive evolution of 30 and 60 days under anaerobic condition.

Model-II

The model-II was involved 140 metabolites and 157 reactions for the *E. coli pta-adhE-pfkA-glk* gene knockout mutant undergoing adaptive evolution of 30 and 60 days under anaerobic condition.

Metabolic network map for *E. coli*



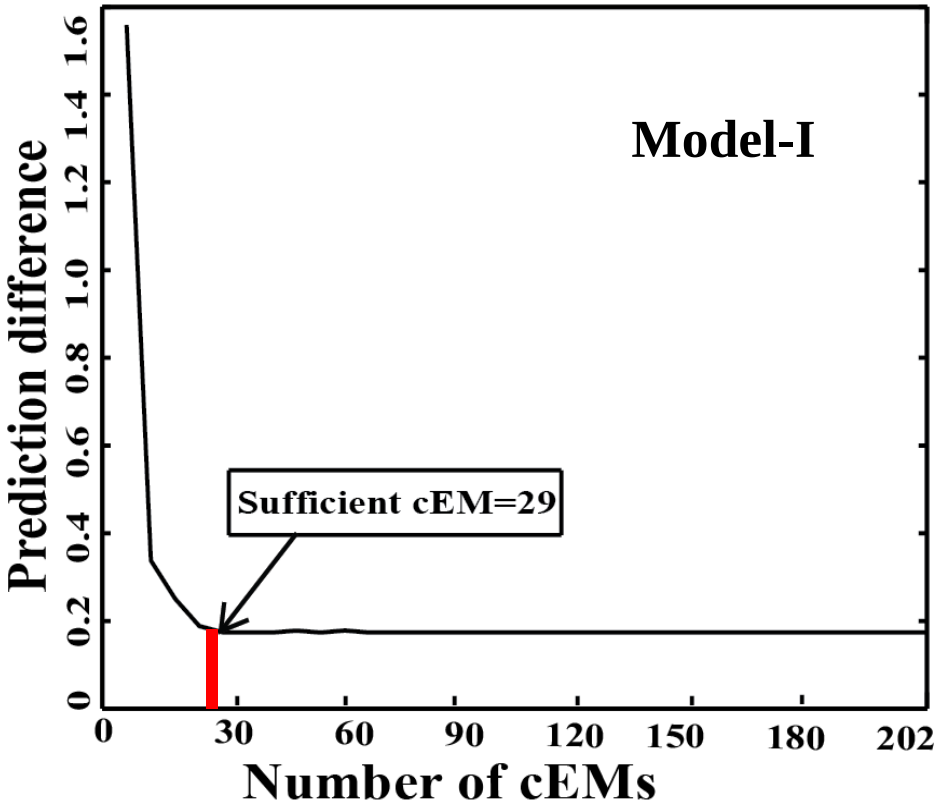
Quantitative contributions for input flux

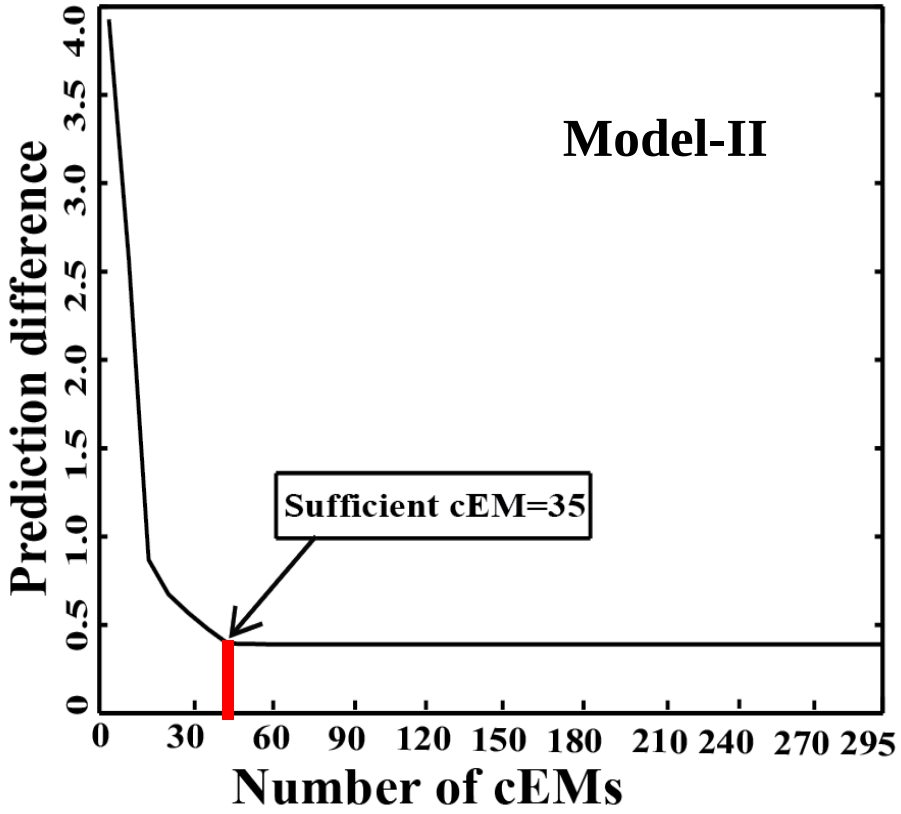
Model	Methods	EM matrix (P)	EMC(λ)
Model-I	Ordinary EM	156 × 122126	122126 × 1
	cEM	156 × 202	202 × 1
Model-II	Ordinary EM	157 × 321416	321416 × 1
	cEM	157 × 295	295 × 1

Quantitative contributions $=\lambda_i \cdot P(\text{numuptake}, i)$ ……(*)

where, i=number of EM and numuptake =the row in EM matrix P for uptake / input flux.

Necessary and sufficient cEM

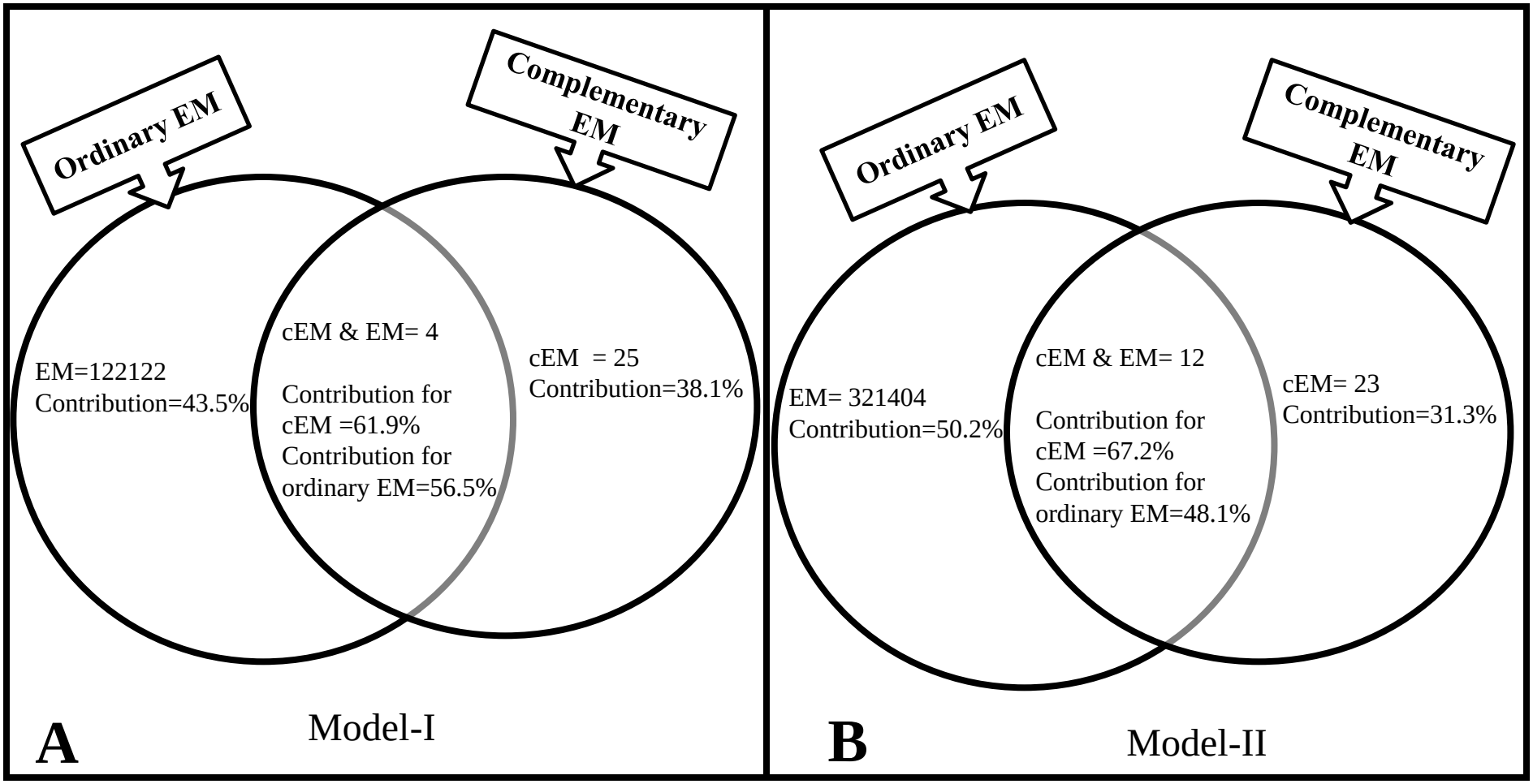




$$\text{Predection difference} = \sqrt{\frac{1}{u}\sum_{t=1}^{u}\left(v_{t,cEM} - v_{t,\text{ordinary EM}}\right)^2}$$

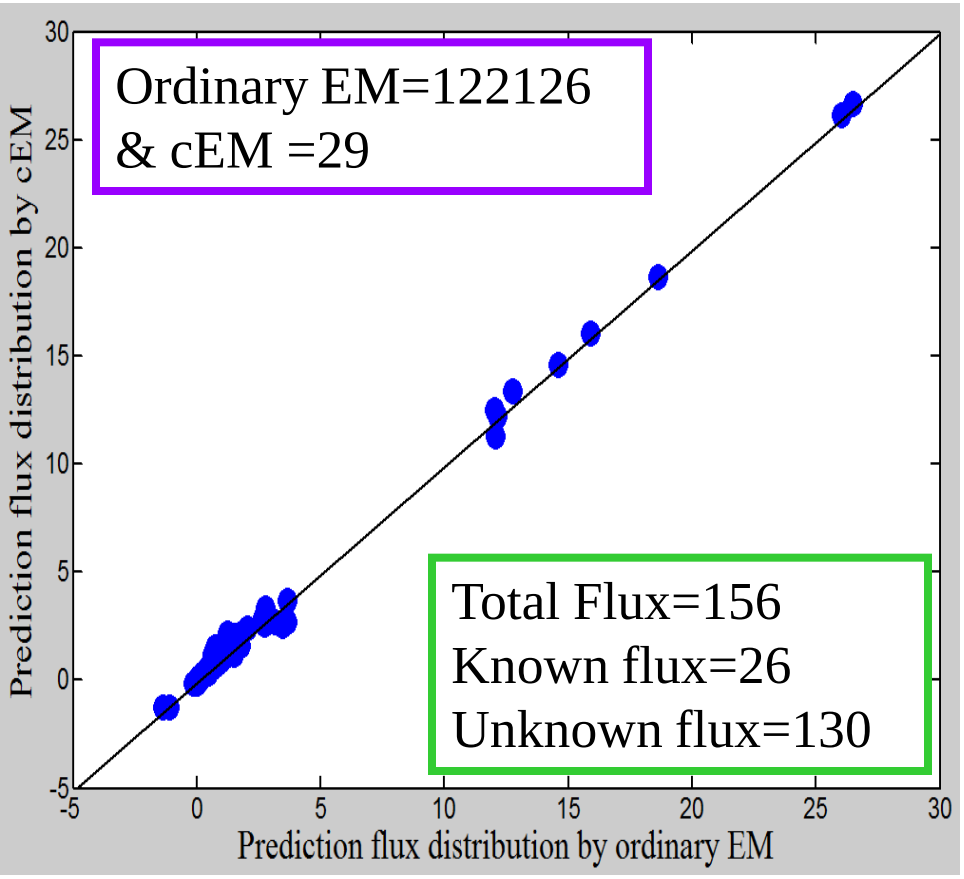
where $v_{t,cEM}$ and $v_{t,\text{ordinary EM}}$ are the predicted fluxes for the *t*th reaction by the cEM and ordinary EM analyses, respectively; u is the number of the unmeasured fluxes.

The employed cEMs and EMs and their quantitative contributions to input flux

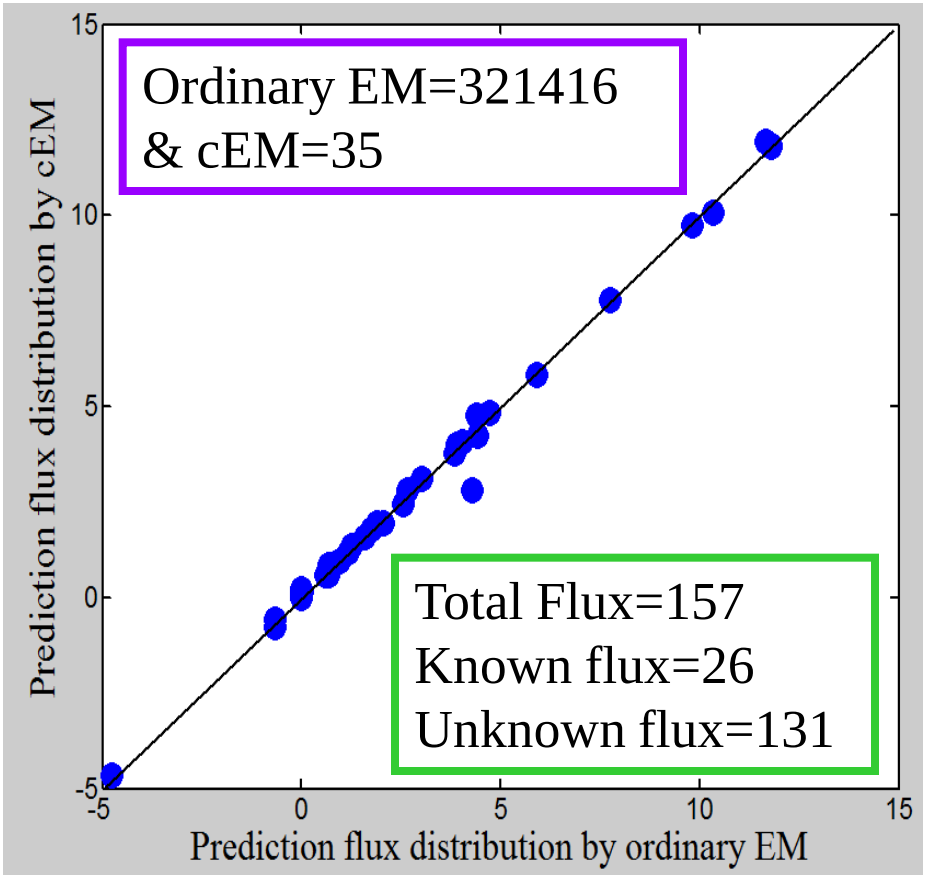


Compare Results (Cont.)

Model-I



Model-II



Genetic Modification of Flux (GMF) [4]

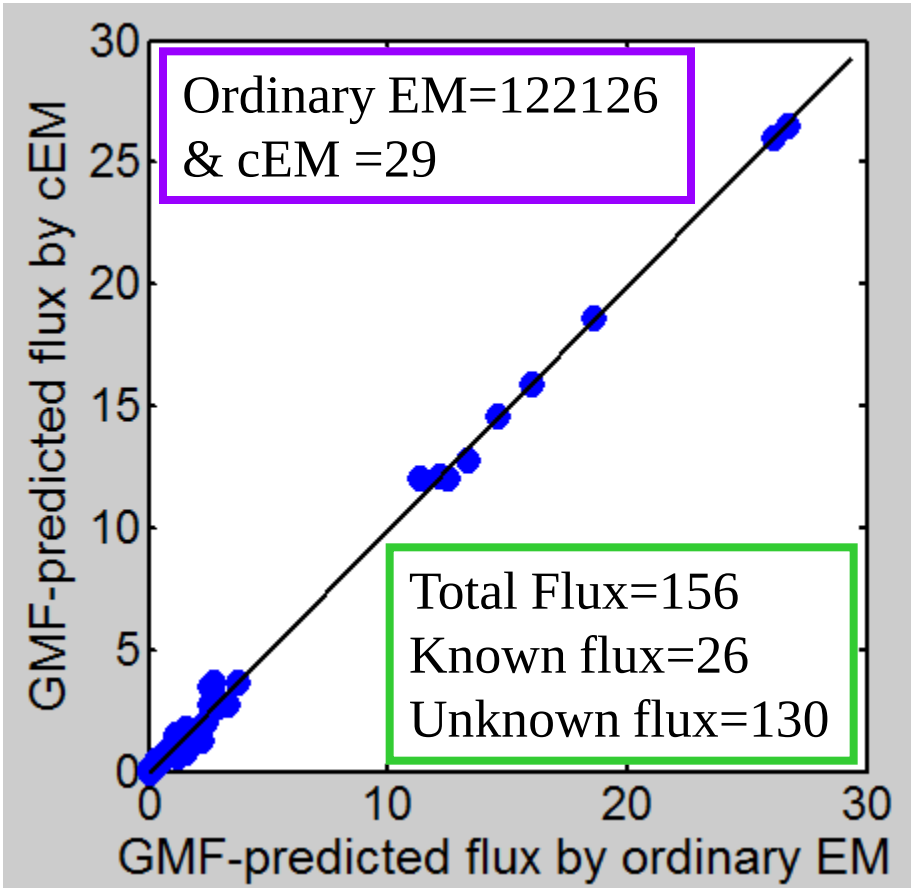
- GMF is an EM-based algorithm that couples with modified control effective flux (mCEF) and enzyme control flux (ECF).

- mCEF was proposed based on CEF to estimate the gene expression patterns in genetically modified mutants in terms of specific biological function.

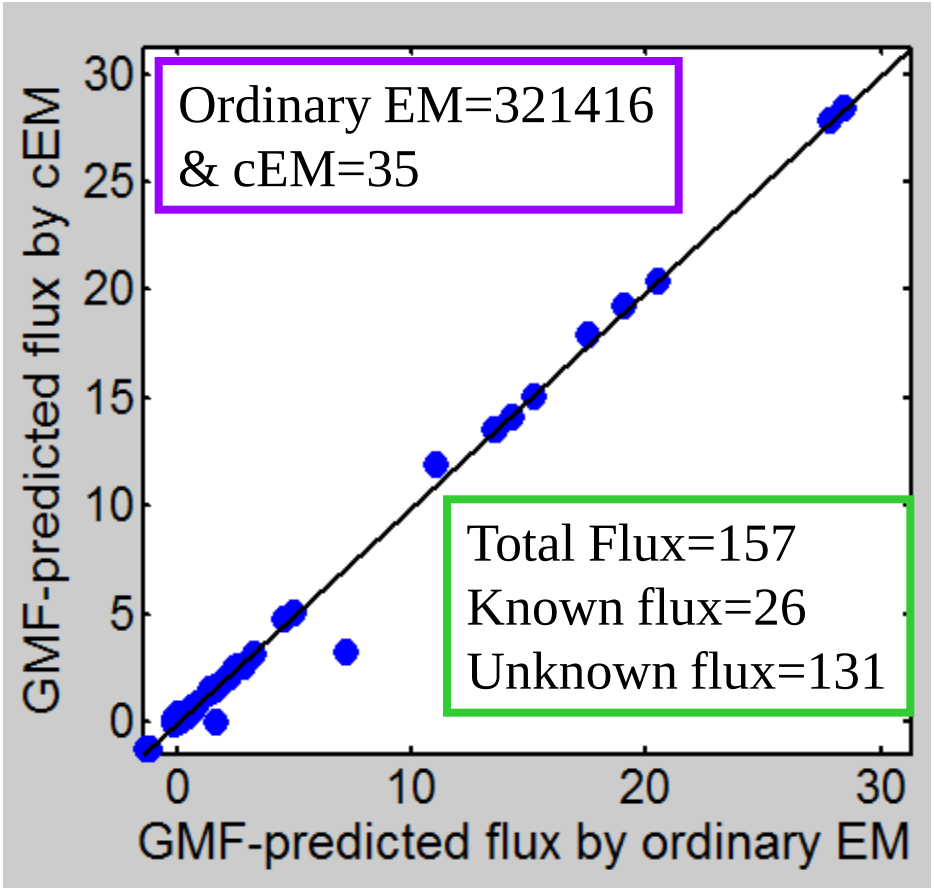
- ECF predicts how change in enzyme profile affects the flux distribution.

Compare Results (Cont.)

Model-I



Model-II



Pearson's correlation coefficient (r) and coefficients of determination (R^2) between the experimental flux and predicted flux.

Model	Condition	Method	Pearson's Correlation (r)	Coefficients of determination (R^2)
Model-I	Wild type	Ordinary EM	0.9980	0.9960
		cEM	0.9982	0.9964
	Mutant type	Ordinary EM	0.9973	0.9947
		cEM	0.9975	0.9950
Model-II	Wild type	Ordinary EM	0.9639	0.9291
		cEM	0.9634	0.9281
	Mutant type	Ordinary EM	0.9989	0.9978
		cEM	0.9989	0.9978

r range between 0.9634 and 0.9989, R^2 ranging from 0.9281 to 0.9978. These statistical analyses demonstrate the r and R^2 remarkably high, and provide statistically significant correlation between the experimental flux and predicted flux by cEM and ordinary EM analyses.

Pearson's correlation coefficient (r) and coefficients of determination (R^2) between the experimental flux and GMF-predicted flux.

Model	Adaptive evolution	Method	Pearson's Correlation (r)	Coefficients of determination (R^2)
Model-I	30 days	Ordinary EM	0.9696	0.9401
		cEM	0.9680	0.9370
	60 days	Ordinary EM	0.9756	0.9519
		cEM	0.9743	0.9492
Model-II	30 days	Ordinary EM	0.8723	0.7610
		cEM	0.9873	0.9747
	60 days	Ordinary EM	0.9908	0.9817
		cEM	0.9841	0.9684

r range between 0.8723 and 0.9908, R^2 ranging from 0.7610 to 0.9817. These statistical analyses demonstrate the r and R^2 remarkably high, and provide statistically significant correlation between the experimental flux and GMF-predicted flux by proposed and ordinary method.

Calculation speed and accuracy

Model	Method		# EM	Total running time(s)	Prediction error
Model-I	Ordinary EM	CNA	122126	600+770.871=1370.871[a]	0.0233
		efmtool	122126	100+780.871=980.871[a]	0.0233
	cEM		202	10+34+1.561=45.561[b]	0.0268
Model-II	Ordinary EM	CNA	321416	6000+1050.56=7050.56[a]	0.0813
		efmtool	321416	200+1000.56=1200.56[a]	0.0233
	cEM		295	12+38+1.805=51.805[b]	0.0475

$$\text{Predection error} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left(v_{i,\text{prediction}} - v_{i,\text{exp}}\right)^2}$$

[a] The ordinary EM analysis consists of two steps (EM extraction and flux prediction by MEP) to predict the flux distributions. [b] The cEM analysis consists of three steps (α-spectrum, cEM extraction, and flux prediction by MEP).

Concluding Remarks

- We found 29 and 35 cEMs for model-I and model-II are enough to estimate the flux distributions, at which the prediction difference almost converges.

- The predicted (MEP & GMF) flux distribution by cEMs was very consistent with that by the ordinary EM analyses.

- The cEM method, where neither requires the initial generation of a full set of EMs nor any objective biological function, which is often computationally demanding, memory improvements and reduced the cost.

References

[1]. Lp K, Colijn C, Lun DS: Analysis of complex metabolic behavior through pathway decomposition. *BMC Syst Biol* 2011, **5**:91.

[1]. Wiback SJ, Mahadevan R, Palsson BO: Reconstructing metabolic flux vectors from extreme pathways: defining the alphaspectrum. *J Theor Biol* 2003, **224**:313-324.

[3]. Zhao QY, Kurata H: Use of maximum entropy principle with Lagrange multipliers extends the feasibility of elementary mode analysis. *J Biosci Bioeng* 2010, **110 (2)**: 254–261.

[4] Zhao QY, Kurata H: Genetic Modification of Flux for flux prediction of mutants, Bioinformatics **25** (2009) 1702–1708.

Acknowledgement

❖ This work is supported by Grant-in-Aid for Scientific Research (B) (25280107) from Japan Society for the Promotion of Science.

Thanks' for your kind attention!!!!!

Let Us Meet Again

We welcome you all to our future conferences of OMICS Group International

Please Visit:

www.omicsgroup.com

www.conferenceseries.com

www.pharmaceuticalconferences.com

www.metabolomicsconference.com