GEMSiRV: A software platform for GEnome-scale Metabolic models Simulation, Reconstruction and Visualization

全基因規模的代謝網路建構、模擬與視覺化平台之介紹與實際操作

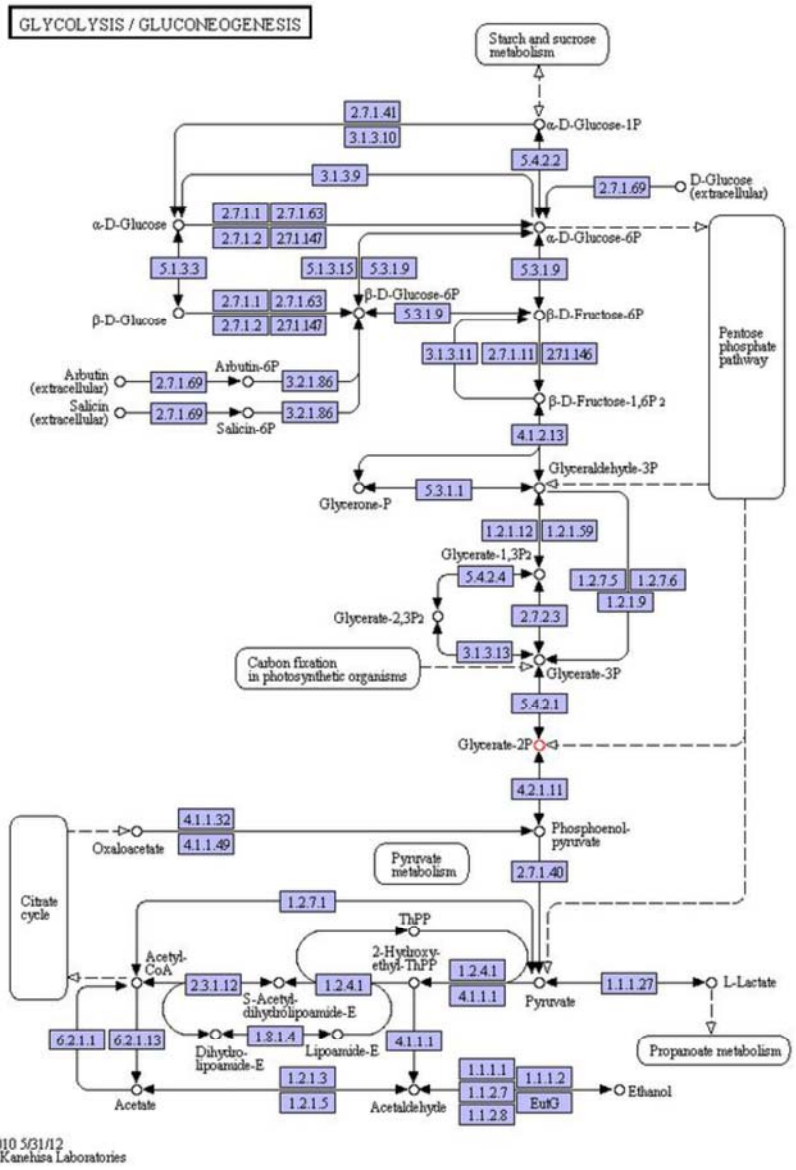
Yu-Chieh Liao, Ph.D.
Assistant Investigator
National Health Research Institutes

http://sb.nhri.org.tw/GEMSiRV

How to simulate a model?

WHAT IS A METABOLIC NETWORK?

KEGG pathway

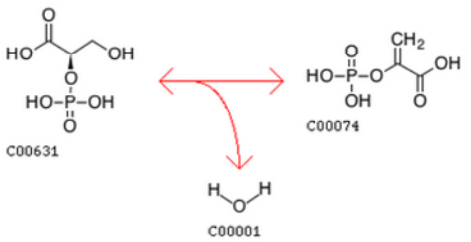


- Enzyme catalyze metabolic reactions
- Reactants and products are involved in a reaction

E.g.

KEGG **REACTION: R00658**

Help

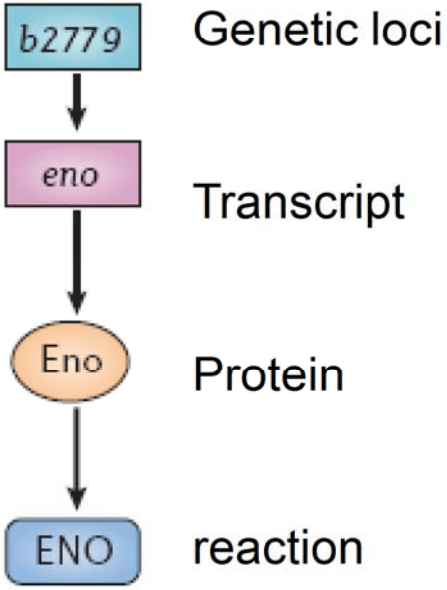
Entry	R00658 Reaction
Name	2-phospho-D-glycerate hydro-lyase (phosphoenolpyruvate-forming)
Definition	2-Phospho-D-glycerate <=> Phosphoenolpyruvate + H2O
Equation	C00631 <=> C00074 + C00001
	 C00631 C00074 C00001
RPair	RP01033 C00074_C00631 main RP06256 C00001_C00631 leave
Enzyme	4.2.1.11
Pathway	rn00010 Glycolysis / Gluconeogenesis rn00680 Methane metabolism rn01100 Metabolic pathways rn01110 Biosynthesis of secondary metabolites rn01120 Microbial metabolism in diverse environments rn01230 Biosynthesis of amino acids
Orthology	K01689 enolase [EC:4.2.1.11]

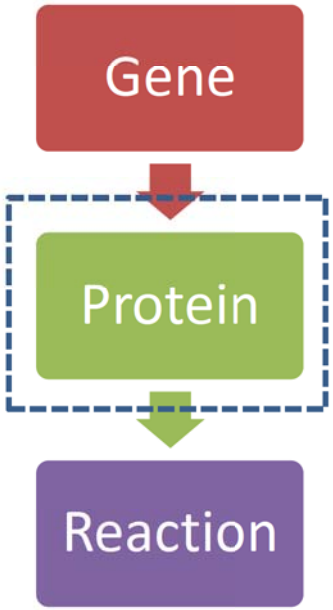
http://www.genome.jp/kegg-bin/show_pathway?ko00010

What is in a metabolic network?

Abbreviation	Glycolytic reactions	Genes
HEX1	[c]GLC + ATP → G6P + ADP + H	*glk*
PGI	[c]G6P ↔ F6P	*pgi*
PFK	[c]ATP + F6P → ADP + FDP + H	*pfkA, pfkB*
FBA	[c]FDP ↔ DHAP + G3P	*fbaA, fbaB*
TPI	[c]DHAP ↔ G3P	*tpiA*
GAPD	[c]G3P + NAD + PI ↔ 13DPG + H + NADH	*gapA, gapC1, gapC2*
PGK	[c]13DPG + ADP ↔ 3PG + ATP	*pgk*
PGM	[c]3PG ↔ 2PG	*gpmA, gpmB*
ENO	[c]2PG ↔ H_2O + PEP	*eno*
PYK	[c]ADP + H + PEP → ATP + PYR	*pykA, pykF*

Reactant →	Product
2PG →	H2O and **PEP**
-1 →	**1** **1**



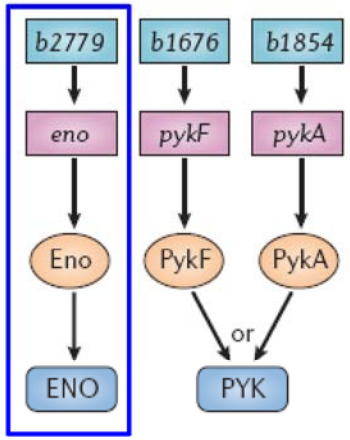


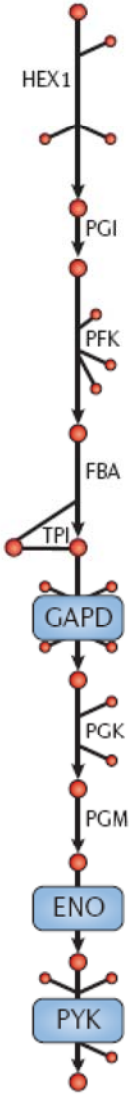
Box 3 | **Assembly and representation**

A list of charge and elementally balanced metabolic reactions can be represented in a stoichiometric matrix (S), where rows and columns correspond to metabolites and reactions and the elements are the stoichiometric coefficients. In genome-scale metabolic networks these stoichiometric matrices contain few non-zero elements, as relatively few metabolites participate in a given reaction. Connections between genes and reactions can be represented as gene–protein–reaction (GPR) associations by using Boolean rules or visualized using graphic images. In the GPR scheme, the first level (teal) corresponds to genetic loci, the second level (pink) to transcripts, the third level (orange) to functional proteins, and the fourth level (blue) to reactions. [c], cytoplasmic reactions.

Abbreviation	Glycolytic reactions	Genes
HEX1	[c]GLC + ATP → G6P + ADP + H	*glk*
PGI	[c]G6P ↔ F6P	*pgi*
PFK	[c]ATP + F6P → ADP + FDP + H	*pfkA, pfkB*
FBA	[c]FDP ↔ DHAP + G3P	*fbaA, fbaB*
TPI	[c]DHAP ↔ G3P	*tpiA*
GAPD	[c]G3P + NAD + PI ↔ 13DPG + H + NADH	*gapA, gapC1, gapC2*
PGK	[c]13DPG + ADP ↔ 3PG + ATP	*pgk*
PGM	[c]3PG ↔ 2PG	*gpmA, gpmB*
ENO	[c]2PG ↔ H_2O + PEP	*eno*
PYK	[c]ADP + H + PEP → ATP + PYR	*pykA, pykF*

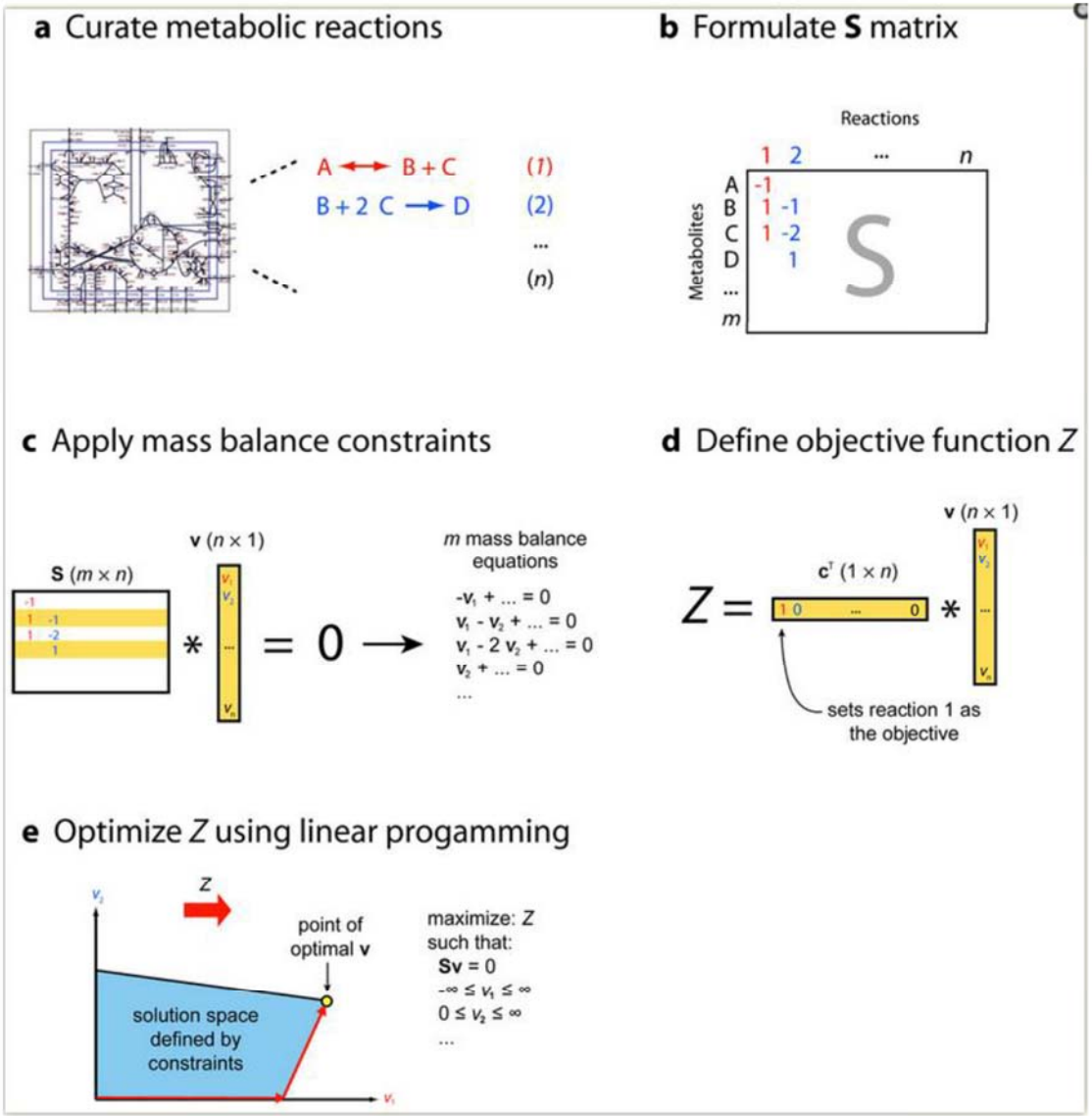
	HEX1	PGI	PFK	FBA	TPI	GAPD	PGK	PGM	ENO	PYK
ATP	−1	0	−1	0	0	0	1	0	0	1
GLC	−1	0	0	0	0	0	0	0	0	0
ADP	1	0	1	0	0	0	−1	0	0	−1
G6P	1	−1	0	0	0	0	0	0	0	0
H	1	0	1	0	0	1	0	0	0	−1
F6P	0	1	−1	0	0	0	0	0	0	0
FDP	0	0	1	−1	0	0	0	0	0	0
DHAP	0	0	0	1	−1	0	0	0	0	0
G3P	0	0	0	1	1	−1	0	0	0	0
NAD	0	0	0	0	0	−1	0	0	0	0
PI	0	0	0	0	0	−1	0	0	0	0
13DPG	0	0	0	0	0	1	−1	0	0	0
NADH	0	0	0	0	0	1	0	0	0	0
3PG	0	0	0	0	0	0	1	−1	0	0
2PG	0	0	0	0	0	0	0	1	−1	0
PEP	0	0	0	0	0	0	0	0	1	−1
H_2O	0	0	0	0	0	0	0	0	1	0
PYR	0	0	0	0	0	0	0	0	0	1





J. L. Reed et al., Nature Reviews Genetics, Vol. 7, pp. 130-141 (2006)

Flux balance analysis



- Linear programming

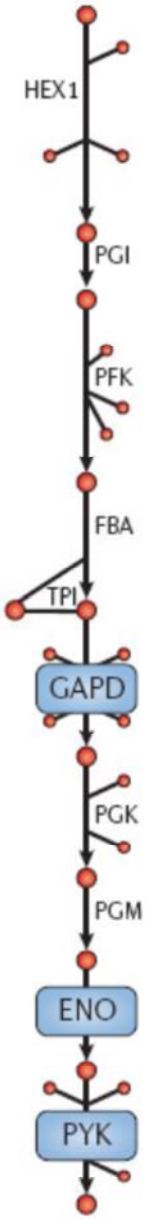
Linear programs are problems that can be expressed in canonical

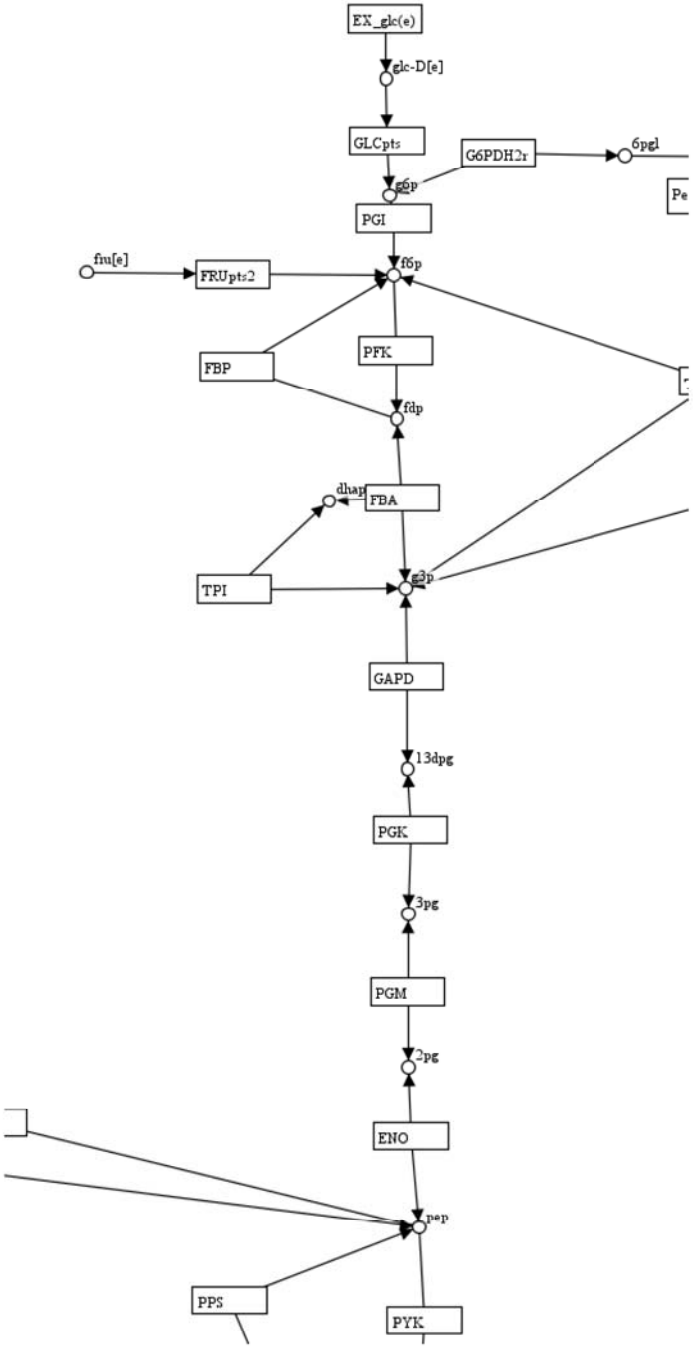
$$\begin{aligned} \text{maximize} \quad & \mathbf{c}^{\mathrm{T}}\mathbf{x} \\ \text{subject to} \quad & A\mathbf{x} \leq \mathbf{b} \\ \text{and} \quad & \mathbf{x} \geq \mathbf{0} \end{aligned}$$

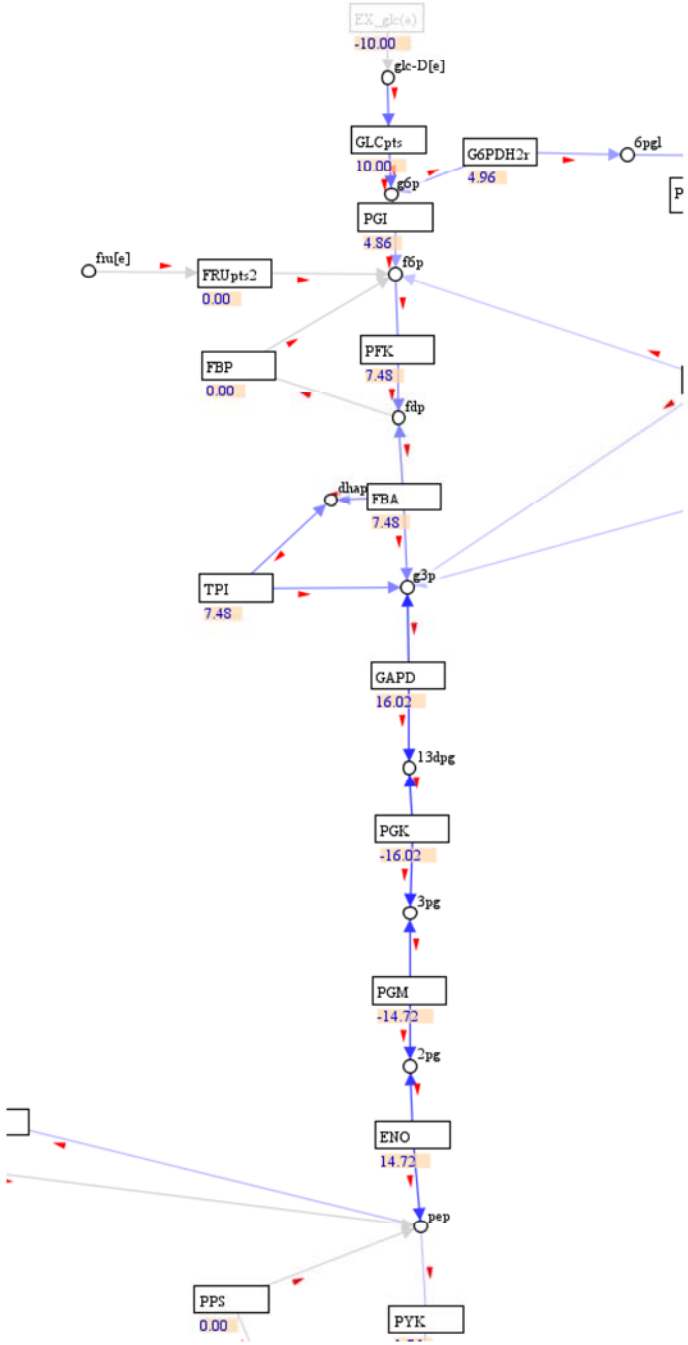
Biomass:

(1.496) 3pg + (3.7478) accoa + (59.81) atp + (0.361) e4p + (0.0709) f6p + (0.129) g3p + (0.205) g6p + (0.2557) gln-L + (4.9414) glu-L + (59.81) h2o + (3.547) nad + (13.0279) nadph + (1.7867) oaa + (0.5191) pep + (2.8328) pyr + (0.8977) r5p
-->
(59.81) adp + (4.1182) akg + (3.7478) coa + (59.81) h + (3.547) nadh + (13.0279) nadp + (59.81) pi

http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3108565/


HEX1
PGI
PFK
FBA
TPI
GAPD
PGK
PGM
ENO
PYK


EX_glc(e)
glc-D[e]
GLCpts
G6PDH2r
6pgl
g6p
PGI
fru[e]
FRUpts2
f6p
FBP
PFK
fdp
dhap
FBA
TPI
g3p
GAPD
13dpg
PGK
3pg
PGM
2pg
ENO
pep
PPS
PYK


EX_glc(e)
-10.00
glc-D[e]
GLCpts
10.00
G6PDH2r
4.96
6pgl
g6p
PGI
4.86
fru[e]
FRUpts2
0.00
f6p
FBP
0.00
PFK
7.48
fdp
dhap
FBA
7.48
TPI
7.48
g3p
GAPD
16.02
13dpg
PGK
-16.02
3pg
PGM
-14.72
2pg
ENO
14.72
pep
PPS
0.00
PYK

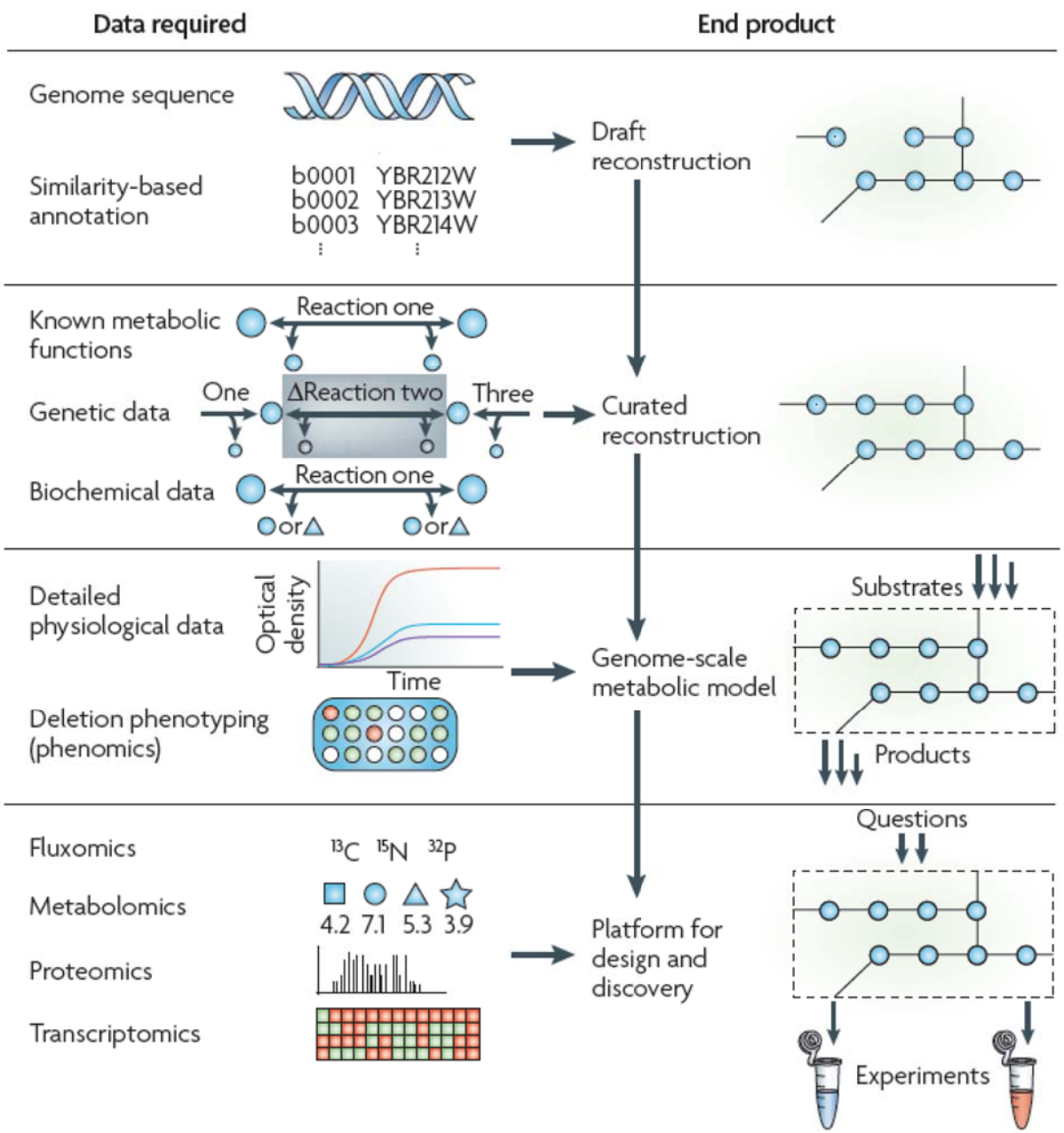


Figure 1 | **Phases and data used to generate a metabolic reconstruction.**

A. M. Feist et al., Nature Reviews Microbiology, Vol. 7, pp. 129-143 (2009)

BIOINFORMATICS ORIGINAL PAPER

Vol. 28 no. 13 2012, pages 1752–1758
doi:10.1093/bioinformatics/bts267

Systems biology

Advance Access publication May 4, 2012

GEMSiRV: a software platform for GEnome-scale metabolic model simulation, reconstruction and visualization

Yu-Chieh Liao[1,*,†], Ming-Hsin Tsai[1,2,†], Feng-Chi Chen[1] and Chao A. Hsiung[1,*]

[1]Division of Biostatistics and Bioinformatics, Institute of Population Health Sciences, National Health Research Institutes, Zhunan 350 and [2]Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan, R.O.C.

Associate Editor: Trey Ideker

MrBac (http://sb.nhri.org.tw/MrBac)

Home | Draft reconstruction | About | How-to | Contact | Supplementary

*Select a query and a reference species: [Example]

Query Species:	Reference Species:
Acaryochloris marina MBIC11017	Acaryochloris marina MBIC11017
Acholeplasma laidlawii PG-8A	Acholeplasma laidlawii PG-8A
Acidiphilium cryptum JF-5	Acidiphilium cryptum JF-5
Acidithiobacillus ferrooxidans ATCC 23270	Acidithiobacillus ferrooxidans ATCC 23270
Acidithiobacillus ferrooxidans ATCC 53993	Acidithiobacillus ferrooxidans ATCC 53993
Acidobacteria bacterium Ellin345	Acidobacteria bacterium Ellin345
Acidobacterium capsulatum ATCC 51196	Acidobacterium capsulatum ATCC 51196
Acidothermus cellulolyticus 11B	Acidothermus cellulolyticus 11B
Acidovorax citrulli AAC00-1	Acidovorax citrulli AAC00-1
Acidovorax sp. JS42	Acidovorax sp. JS42

*Choose Blast Settings: [Default / Select All]

E value: & Percent Identity: %

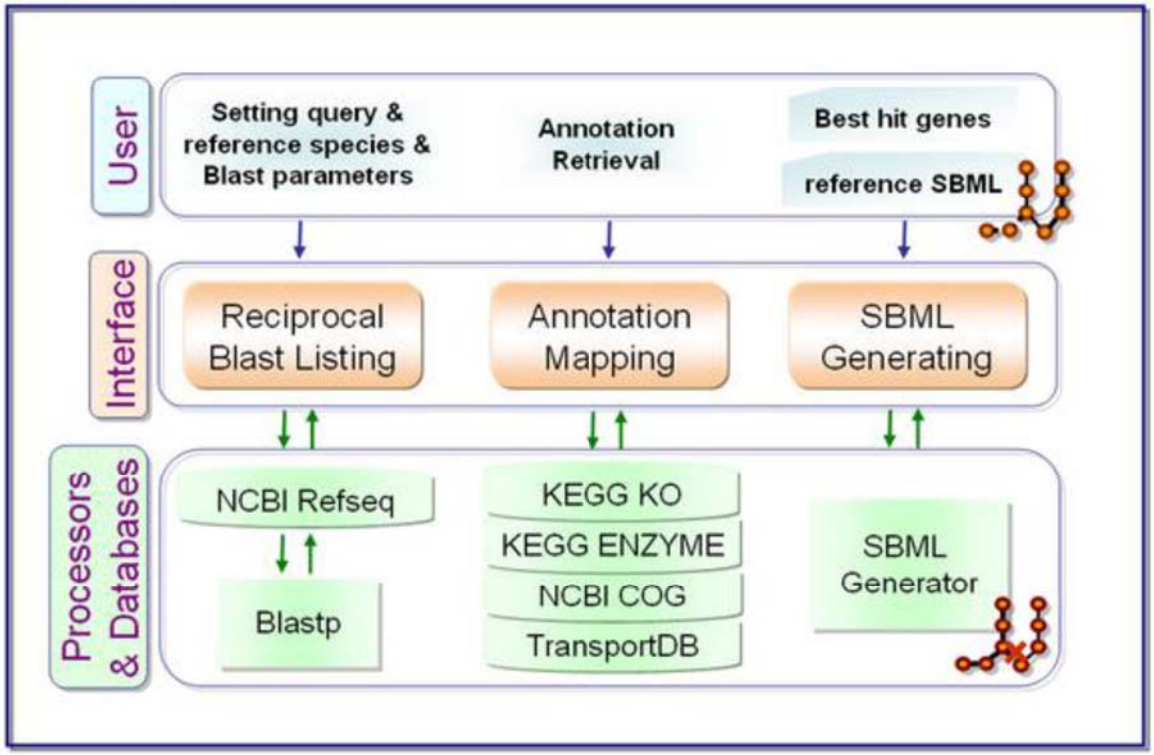
Choose the blast columns to be shown:

☑Ref ☑Query ☑PercIden ☐MLen ☐mismatches ☐gap openings ☐q.start ☐q.end ☐r.start ☐r.end ☑EVal ☐BitScore

* indicates required input.

Submit

Flow Chart


User
Setting query & reference species & Blast parameters
Annotation Retrieval
Best hit genes
reference SBML
Interface
Reciprocal Blast Listing
Annotation Mapping
SBML Generating
Processors & Databases
NCBI Refseq
Blastp
KEGG KO
KEGG ENZYME
NCBI COG
TransportDB
SBML Generator

Metabolic Network Draft Generation

Draft metabolic network generation requires a reference SBML file and the list of reciprocal blast best hits as inputs to proceed.

By taking these two inputs, we are assuming that query ORFs in the best hit list have the same functions as their corresponding reference ORFs. The generator only keeps reactions, in which all associated reference ORFs under "AND" condition and partial reference ORFs under "OR" condition have best hit matches to query ORFs. Only when those two criteria are met, the reference ORF id are replaced by corresponding query id, and the metabolites involved in those reactions are kept in the draft SBML file.

The draft in SBML format can be saved as the final output after query is submitted and can be further curated and edited manually in flux balance analysis softwares such as COBRA.

Upload a reference SBML file (sample, iAF1260)
SBML File: 選擇檔案 未選擇檔案

Upload a one to one gene pair file (sample, SLT2vsECO)
Gene Pair File: 選擇檔案 未選擇檔案

Type in a model name
New Model Name:

Submit Query

```
<?xml version="1.0" encoding="UTF-8"?>
<sbml level="2" version="1" >
    <model id="Ec_iAF1260" name="Ec_iAF1260" >
        <listOfUnitDefinitions >
            <unitDefinition id="mmol_per_gDW_per_hr" >
                <listOfUnits >
                    <unit kind="mole" scale="-3" />
                    <unit exponent="-1" kind="gram" />
                    <unit exponent="-1" kind="second" multiplier=".00027777" />
                </listOfUnits>
            </unitDefinition>
        </listOfUnitDefinitions>
        <listOfCompartments >
            <compartment id="Extra_organism" />
            <compartment id="Periplasm" outside="Extra_organism" />
            <compartment id="Cytosol" outside="Periplasm" />
        </listOfCompartments>
        <listOfSpecies >
            <species boundaryCondition="false" charge="-2" compartment="Cytosol" id="M_10fthf_c" name="M_10_Formyltetrahydrofolate_C2
            <species boundaryCondition="false" charge="0" compartment="Cytosol" id="M_12ppd_R_c" name="M__R__Propane_1_2_diol_C3H8O2"
            <species boundaryCondition="false" charge="0" compartment="Cytosol" id="M_12ppd_S_c" name="M__S__Propane_1_2_diol_C3H8O2"
            <species boundaryCondition="false" charge="-4" compartment="Cytosol" id="M_13dpg_c" name="M_3_Phospho_D_glyceroyl_phospha
            <species boundaryCondition="false" charge="0" compartment="Cytosol" id="M_14glucan_c" name="M_1_4_alpha_D_glucan_C36H62O3
            <species boundaryCondition="false" charge="0" compartment="Periplasm" id="M_14glucan_p" name="M_1_4_alpha_D_glucan_C36H62
            <species boundaryCondition="false" charge="2" compartment="Cytosol" id="M_15dap_c" name="M_1_5_Diaminopentane_C5H16N2" />
            <species boundaryCondition="false" charge="2" compartment="Periplasm" id="M_15dap_p" name="M_1_5_Diaminopentane_C5H16N2"
            <species boundaryCondition="false" charge="0" compartment="Periplasm" id="M_1agpe120_p" name="M_1_Acyl_sn_glycero_3_phospl
            <species boundaryCondition="false" charge="0" compartment="Periplasm" id="M_1agpe140_p" name="M_1_Acyl_sn_glycero_3_phospl
            <species boundaryCondition="false" charge="0" compartment="Periplasm" id="M_1agpe141_p" name="M_1_Acyl_sn_glycero_3_phospl
            <species boundaryCondition="false" charge="0" compartment="Periplasm" id="M_1agpe160_p" name="M_1_Acyl_sn_glycero_3_phospl
            <species boundaryCondition="false" charge="0" compartment="Periplasm" id="M_1agpe161_p" name="M_1_Acyl_sn_glycero_3_phospl
            <species boundaryCondition="false" charge="0" compartment="Periplasm" id="M_1agpe180_p" name="M_1_Acyl_sn_glycero_3_phospl
            <species boundaryCondition="false" charge="0" compartment="Periplasm" id="M_1agpe181_p" name="M_1_Acyl_sn_glycero_3_phospl
            <species boundaryCondition="false" charge="-1" compartment="Periplasm" id="M_1agpg120_p" name="M_1_Acyl_sn_glycero_3_phosp
            <species boundaryCondition="false" charge="-1" compartment="Periplasm" id="M_1agpg140_p" name="M_1_Acyl_sn_glycero_3_phosp
```

Starting materials for manual curation

3	STM1097	--	1184453	1184863	hypothetical protein	136	K07127	--	Unclassifi	--	--	--	--	R	POORLY	General fu	--	--	--	b1970	hiuH
4	STM1098	hpaC	1185708	1185196	4-hydroxyphenylacetate catabolism	170	K00484	4-hydroxy	Metabolis	1.14.13.3	4-hydroxy	Oxidoredu	4-hydroxy	R	POORLY	General fu	--	--	--	--	--
5	STM1099	hpaB	1187288	1185726	4-hydroxyphenylacetate catabolism	520	K00483	4-hydroxy	Metabolis	1.14.13.3	4-hydroxy	Oxidoredu	4-hydroxy	Q	METABC	Secondary	--	--	--	--	--
6	STM1100	hpaR	1187946	1187506	4-hydroxyphenylacetate catabolism	146	--	--	--	--	--	--	--	K	INFORM	Transcript	--	--	--	--	--
7	STM1101	hpaG	1188221	1189510	4-hydroxyphenylacetate catabolism	429	K05921	5-oxopent	Metabolis	4.1.1.68;;	5-oxopent	Lyases; C	5-oxopent	Q	METABC	Secondary	--	--	--	--	--
8	STM1102	hpaE	1189507	1190973	4-hydroxyphenylacetate catabolism	488	K00151	5-carboxy	Metabolis	1.2.1.60	5-carboxy	Oxidoredu	5-carboxy	C	METABC	Energy pr	--	--	--	--	--
9	STM1103	hpaD	1190975	1191826	4-hydroxyphenylacetate catabolism	283	K00455	3,4-dihyd	Metabolis	1.13.11.1	3,4-dihyd	Oxidoredu	3,4-dihyd	S	POORLY	Function	--	--	--	--	--
10	STM1104	hpaF	1191836	1192216	4-hydroxyphenylacetate catabolism	126	K01826	5-carboxy	Metabolis	5.3.3.10	5-carboxy	Isomerase	5-carboxy	E	METABC	Amino aci	--	--	--	--	--
11	STM1105	hpaH	1192359	1193162	4-hydroxyphenylacetate catabolism	267	K02509	2-oxo-hep	Metabolis	4.2.1.-	--	Lyases; C	--	Q	METABC	Secondary	--	--	--	--	--
12	STM1106	hpaI	1193173	1193964	4-hydroxyphenylacetate catabolism	263	K02510	2,4-dihyd	Metabolis	4.1.2.-	--	Lyases; C	--	G	METABC	Carbohydr	--	--	--	--	--
13	STM1107	hpaX	1194036	1195412	4-hydroxyphenylacetate catabolism	458	K02511	MFS trans	Protein Fa	--	--	--	--	G;; E;; P;;	METABC	Carbohydr	Secondary	The Majo	4-hydroxy	--	--
14	STM1108	hpaA	1195422	1196318	4-hydroxyphenylacetate catabolism	298	K02508	AraC fam	Protein Fa	--	--	--	--	K	INFORM	Transcript	--	--	--	--	--
15	STM1109	--	1196332	1197270	putative periplasmic protein	312	--	--	--	--	--	--	--	R	POORLY	General fu	--	--	--	--	--

- We can identify, for instance, the metabolic genes that are present in SLT2 but absent in ECO.
- Examples include the cob (STM2016-STM2035), hpa (STM1098-STM1108) and pdu (STM2037-STM2058).
- These genes have been reported to confer SLT2 the abilities to synthesize cobalamin and to utilize 4-hydroxylphen-ylacetate and propanediol.

http://sb.nhri.org.tw/GEMSiRV/

GEMSiRV

登入 註冊

文章 註解 歷史 連結 列印

HOME

Guide

Download
Installation
Manual
Users' Comments

Repository

Metabolic Models
Reference Databases
Metabolic Maps

Toolbox

TextReplacer
GBKParser
MrBac

搜尋 移至

GEMSiRV

A software platform for GEnome-scale Metabolic models Simulation, Reconstruction and Visualization.

This work has been published in Bioinformatics (Pubmed). If you have any question, please do not hesitate to contact the author by emailing to sysbio@nhri.org.tw.

Start GEMSiRV. *An example model (E. coli textbook model) and its map can be directly imported by click on the Help to Import an Example.*

Genome-scale metabolic network models have become an indispensable part of the increasingly important field of systems biology.

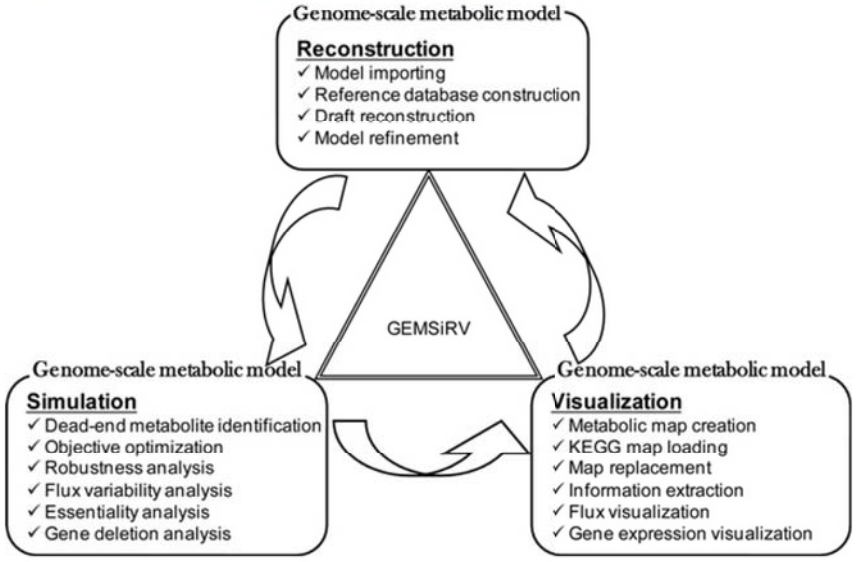
Metabolic systems biology studies usually include three major components – network model construction, objective- and experiment-guided model editing and visualization, and simulation studies based mainly on flux balance analyses.

Here we present a software platform, GEMSiRV, to provide functionalities of easy metabolic network drafting and editing, amenable network visualization for experimental data integration, and flux balance analysis tools for simulation studies.

GEMSiRV is an open-source software for building users' metabolic systems biology project, it provides interactive features in model management, simulation, visualization and integration of omics data.

Furthermore, all of the GEMSiRV-generated metabolic models and analysis results, including projects in progress, can be easily exchanged in the research community. GEMSiRV is a powerful integrative resource that may facilitate the development of systems biology studies.

A schematic overview of the GEMSiRV outlines the key features implemented in the GEMSiRV.



Acknowledgements

- Dr. Chao A. Hsiung
- Dr. Feng-Chi Chen
- Mr. Ming-Hsin Tsai
- Ms. Julie Chih-Yu Chen
- Ms. Yueh-Hsia Tang
- **National Health Research Institutes**
- **National Science Council**

Practices

Metabolic models

- Gene-protein-reaction association
  - Excel model
- Gene/protein information is not necessary
  - Excel or xml model

http://sb.nhri.org.tw/GEMSiRV/en/Manual

- **Dead-end metabolite identification** (Step-by-step)

Select a metabolic model and a map (if you have), GEMSiRV can identify dead-end metabolites and tag them with crosses in the map.

- **Objective optimization** (Step-by-step)

Select a metabolic model and a map (if you have) for objective optimization, the flux result can be visulized in the map.

- **Flux variability analysis** (Step-by-step)

Select a metabolic model and a map (if you have) for flux variability analysis, the min and max fluxes of reaction can be plotted in the map and the blocked reaction are tagged with crosses.

- **Robustness analysis** (Step-by-step)

Select the reactions of interest in a model to see how sensitive the objective is to the particular reactions.

- **Essentiality analysis** (Step-by-step)

Select a metabolic model for essentiality analysis, the computational essential genes or reactions can be identified.

- **Gene deletion analysis** (Step-by-step)

Select a metabolic model for gene deletion analysis, the gene-deletion model can be saved and imported for further evaluation.