



Microsoft® Research

Faculty Summit

10
YEAR ANNIVERSARY

Systems Biology and Biotechnology of Microorganisms: Making Systems Biology Work

Sang Yup Lee, Distinguished Professor
Dean, College of Life Science and Bioengineering
Korea Advanced Institute of Science and Technology

KAIST

- Systems biology is the quantitative and qualitative study of interactions among the components of biological systems, and how these interactions give rise to the function and behavior of that system
 - Omics: genome, transcriptome, proteome, metabolome
 - Genome-scale metabolic model and other cellular models
 - Simulation: metabolic flux analysis and dynamic simulation
 - Integrated analysis at a whole system level

Systems Biotechnology: Making Systems Biology Work

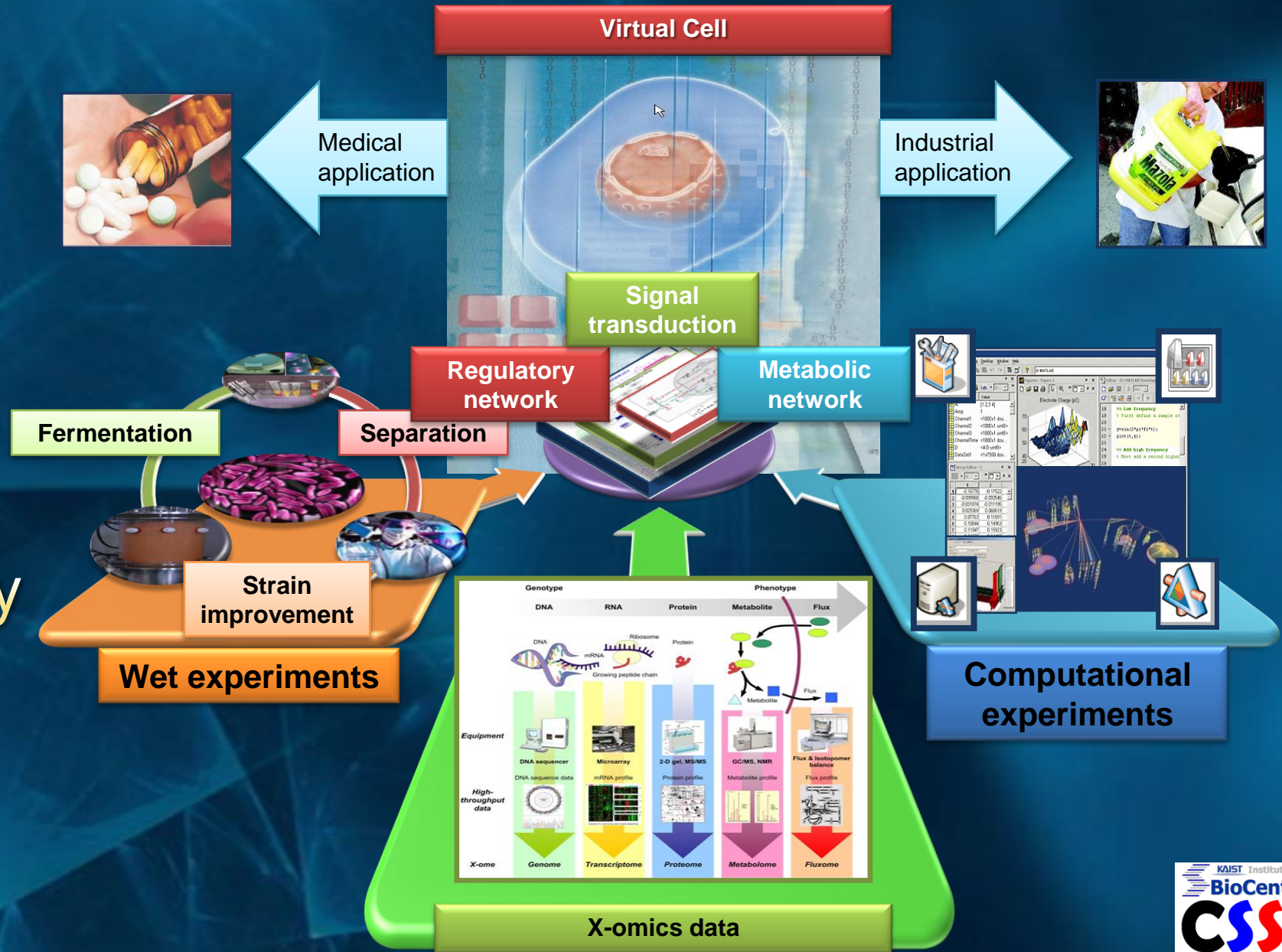
Microsoft
Research

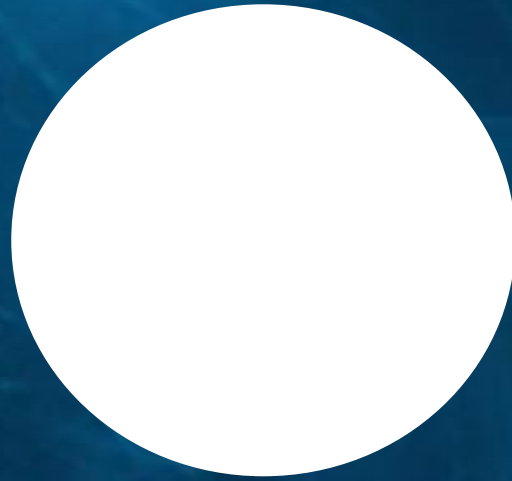


Systems Biology



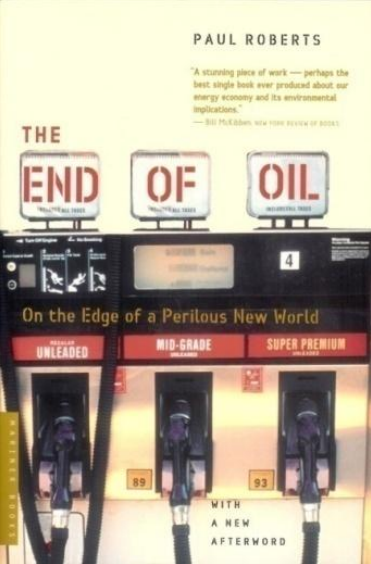
Systems Biotechnology





White Biotechnology = Industrial Biotechnology

enabled through systems and synthetic biotechnology



Global Warming

Kyoto Protocol

Carbon tax

Environmental problems

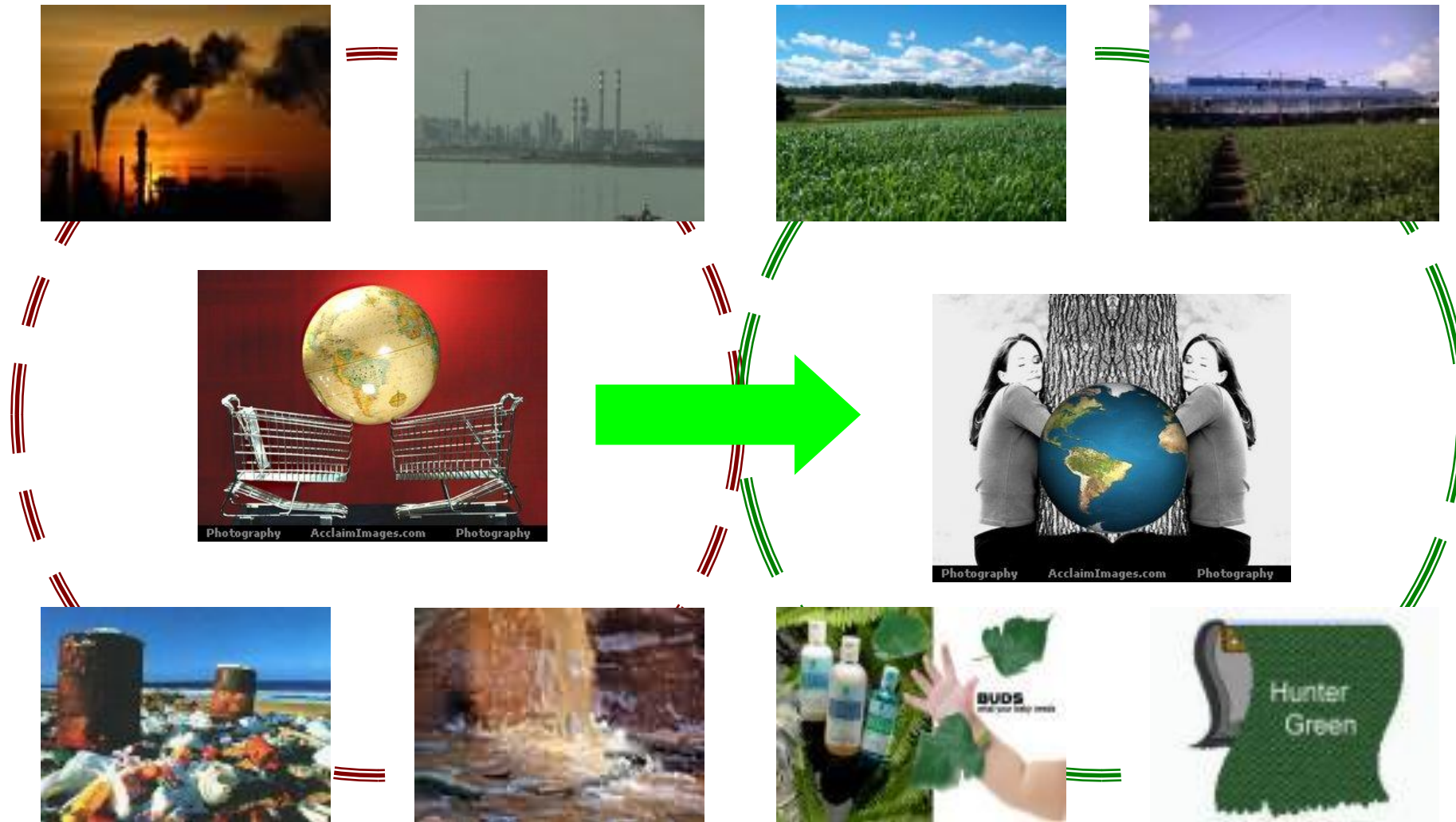
Limited fossil resources



World CO2 output to rise 59 pct by 2030: U.S. Traffic passes downtown Los Angeles on the 10 freeway in this file photo, November 19, 2005. Global emissions of the main gas scientists link to global warming will rise 59 percent from 2004 to 2030, with much of the growth coming from coal burning in developing countries like China, the U.S. government forecast on Monday. REUTERS/Lucy Nicholson

Current system for chemicals and materials

Sustainable system for chemicals and materials



● White Biotechnology

Feedstock: renewable biomass

170 billion tons per year

only 6 billion tons are currently used

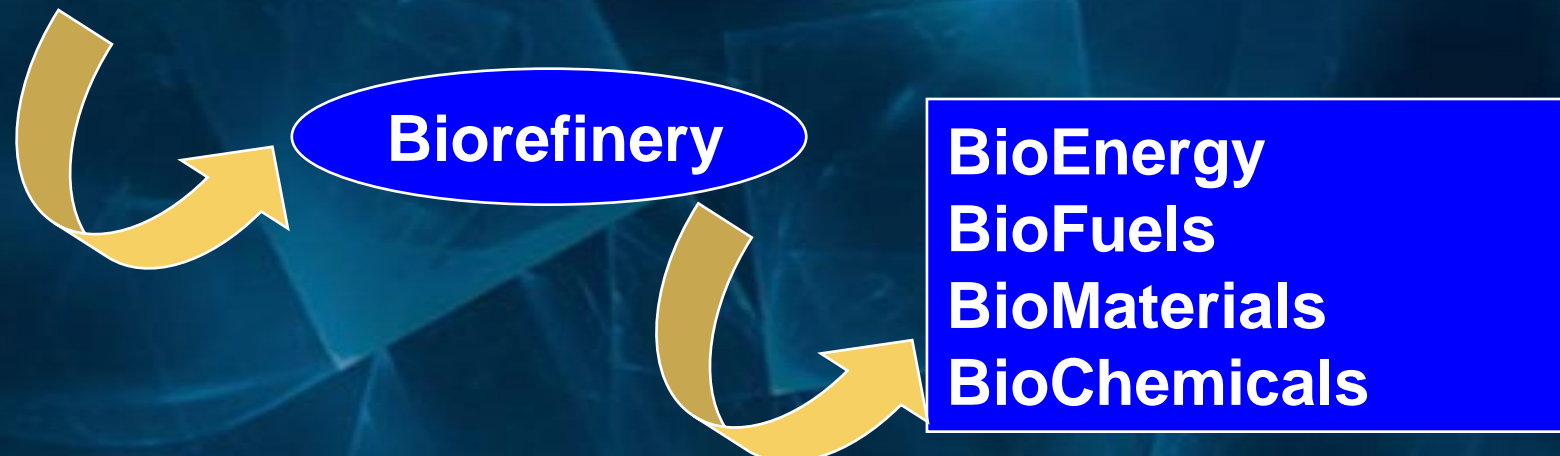
3.5% of this amount is used in non-food applications

170 billion tons of biomass

→ 75% carbohydrates

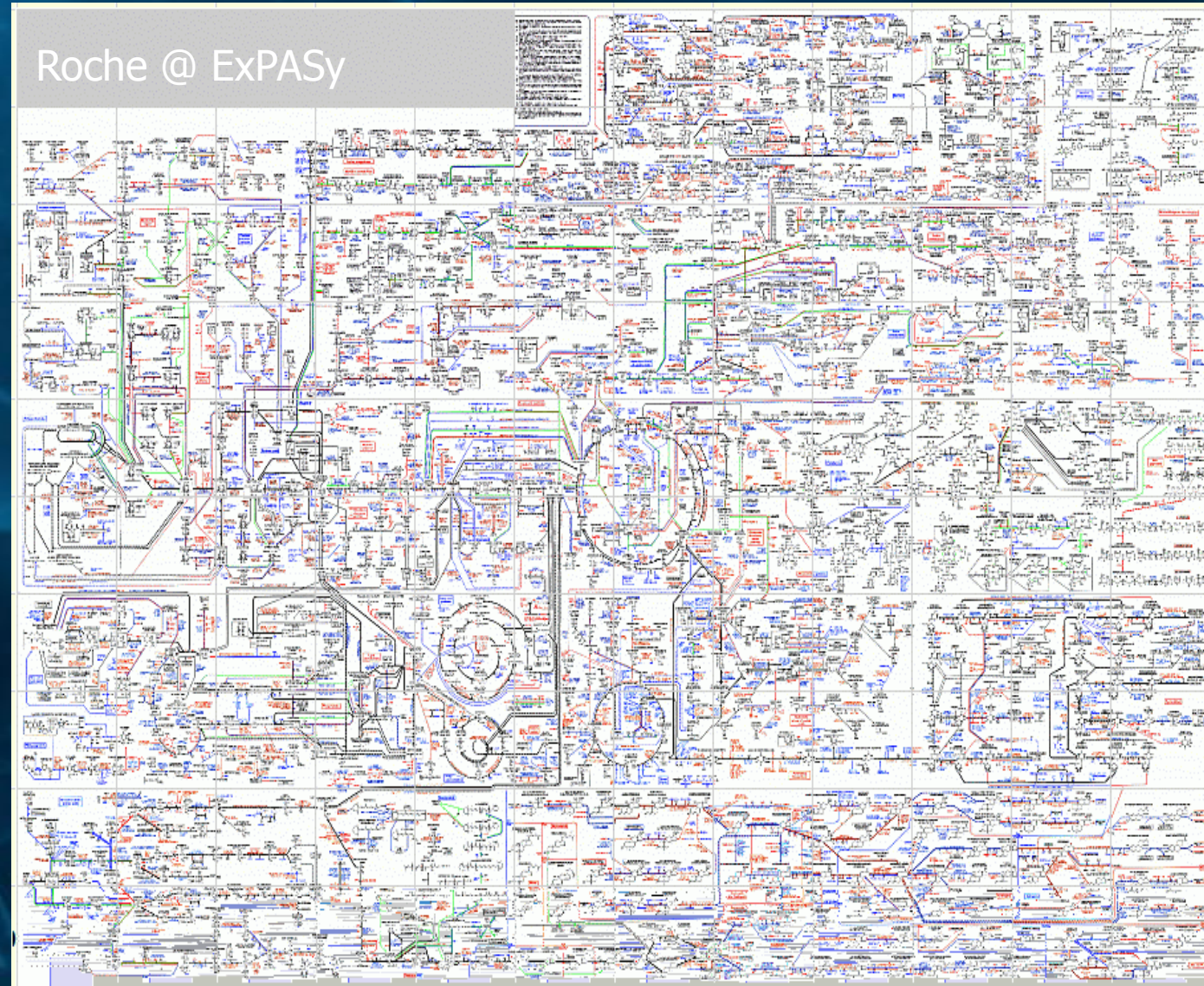
20% lignin

5% others (oils, proteins, etc.)



Chemicals

- Carboxylic acids
- Dicarboxylic acids
- Alcohols
- Diols
- Diamines
- Aromatics
- Amino acids
- and many others
- + Unnatural chemicals



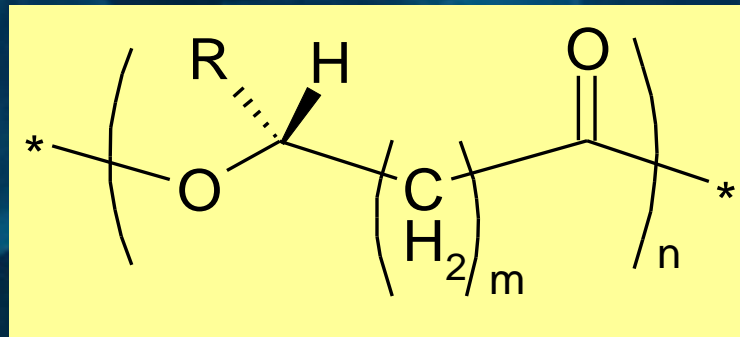
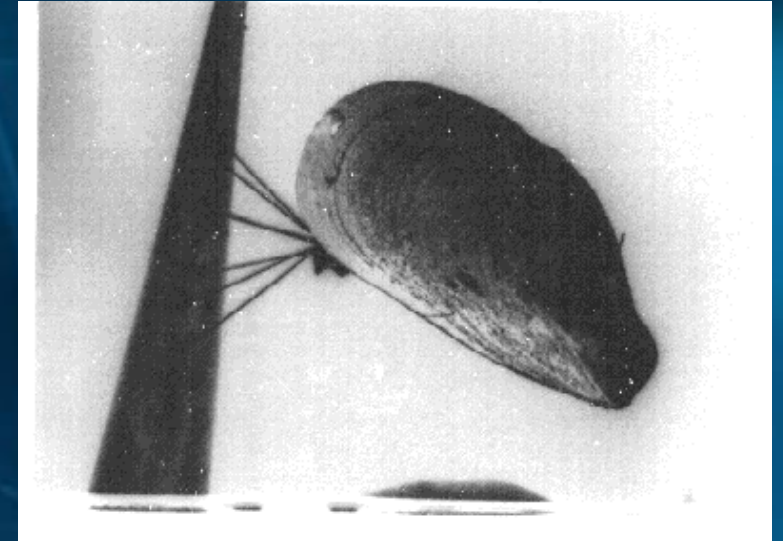
Fuels

- Ethanol
- Diesel
- Butanol/Isobutanol
- Alkanes
- Isoprenoids
- Hydroxycarboxylic esters



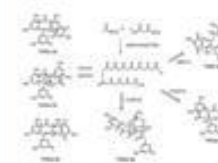
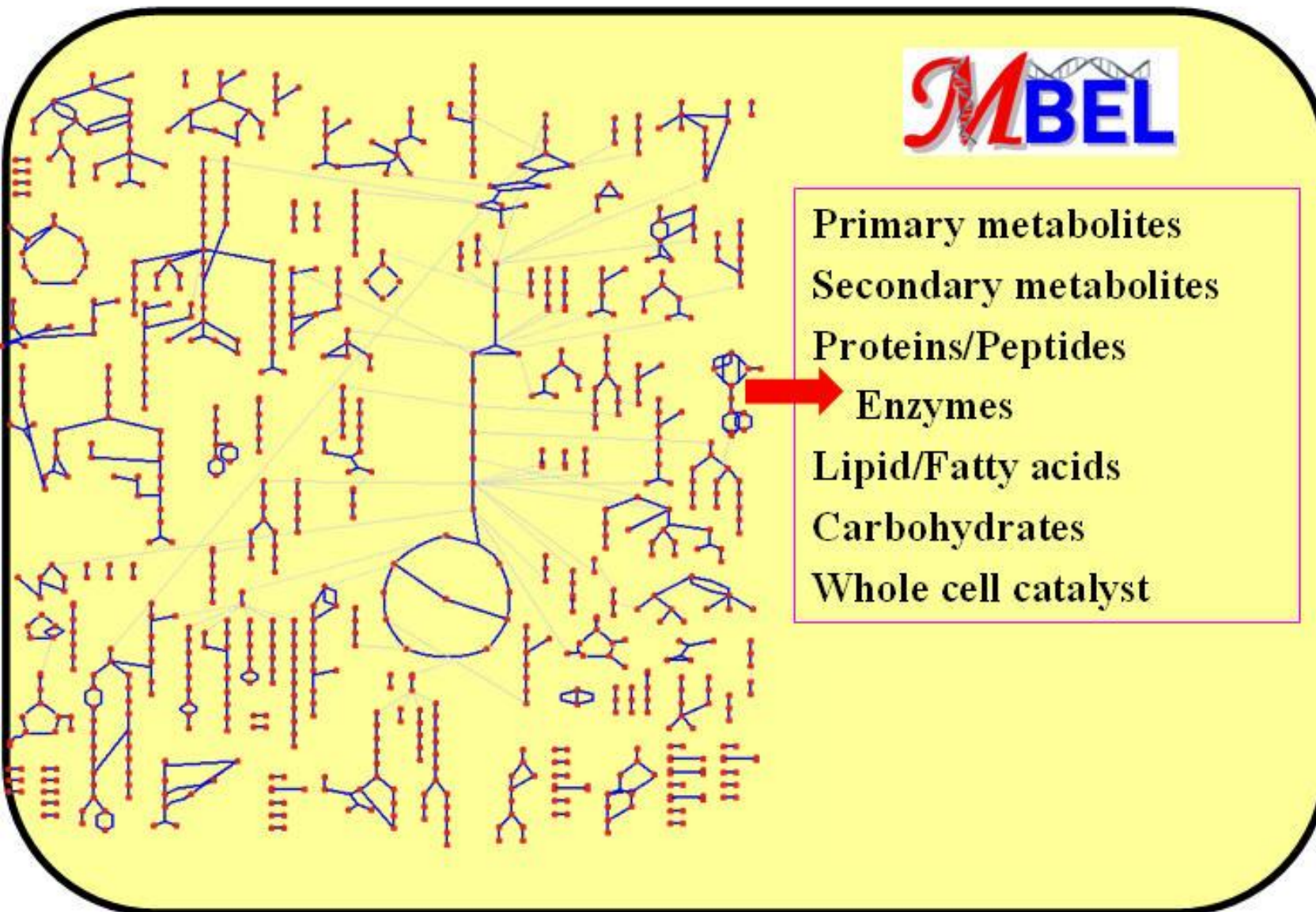
Materials

- Polysaccharides
- Poly (amino acids)
- Polyhydroxyalkanoates
- Unnatural polymers
- *In vitro* polymerization of monomers
- Organic/inorganic hybrids





- Primary metabolites
- Secondary metabolites
- Proteins/Peptides
- Enzymes
- Lipid/Fatty acids
- Carbohydrates
- Whole cell catalyst

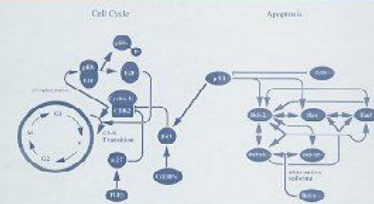


Metabolic Engineering

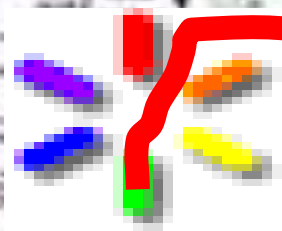
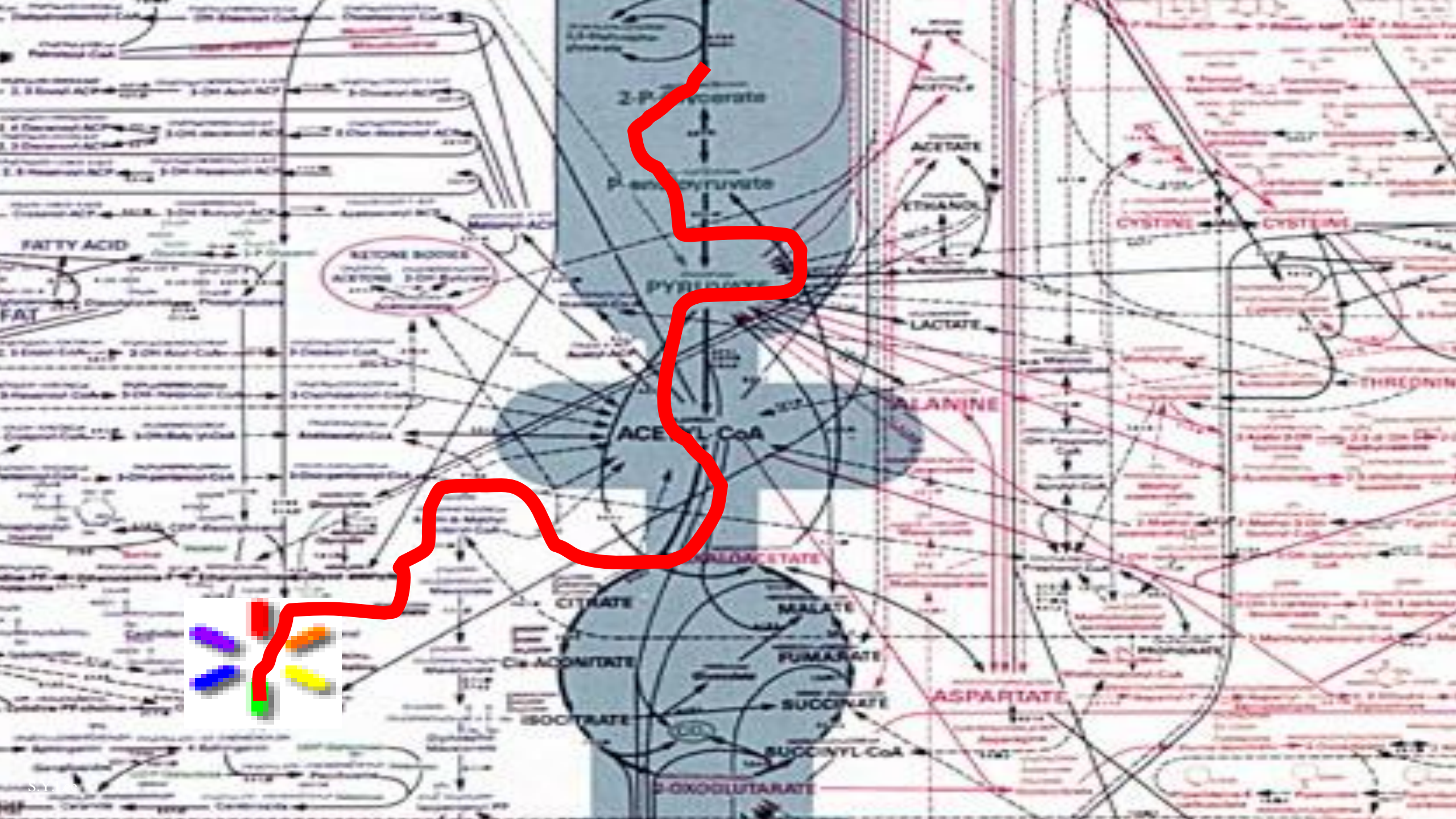
Purposeful modification of metabolic network to achieve... *(after Jay Bailey)*

- Enhanced production of metabolites and other biologicals that are already produced by host organism
- Production of modified or new metabolites and other biologicals that are new to the host organism
- Broadening the substrate utilization range
- Designing improved or new metabolic pathways for degradation of various chemicals, especially xenobiotics
- Modification of cell properties that facilitate bioprocessing (fermentation and product recovery)

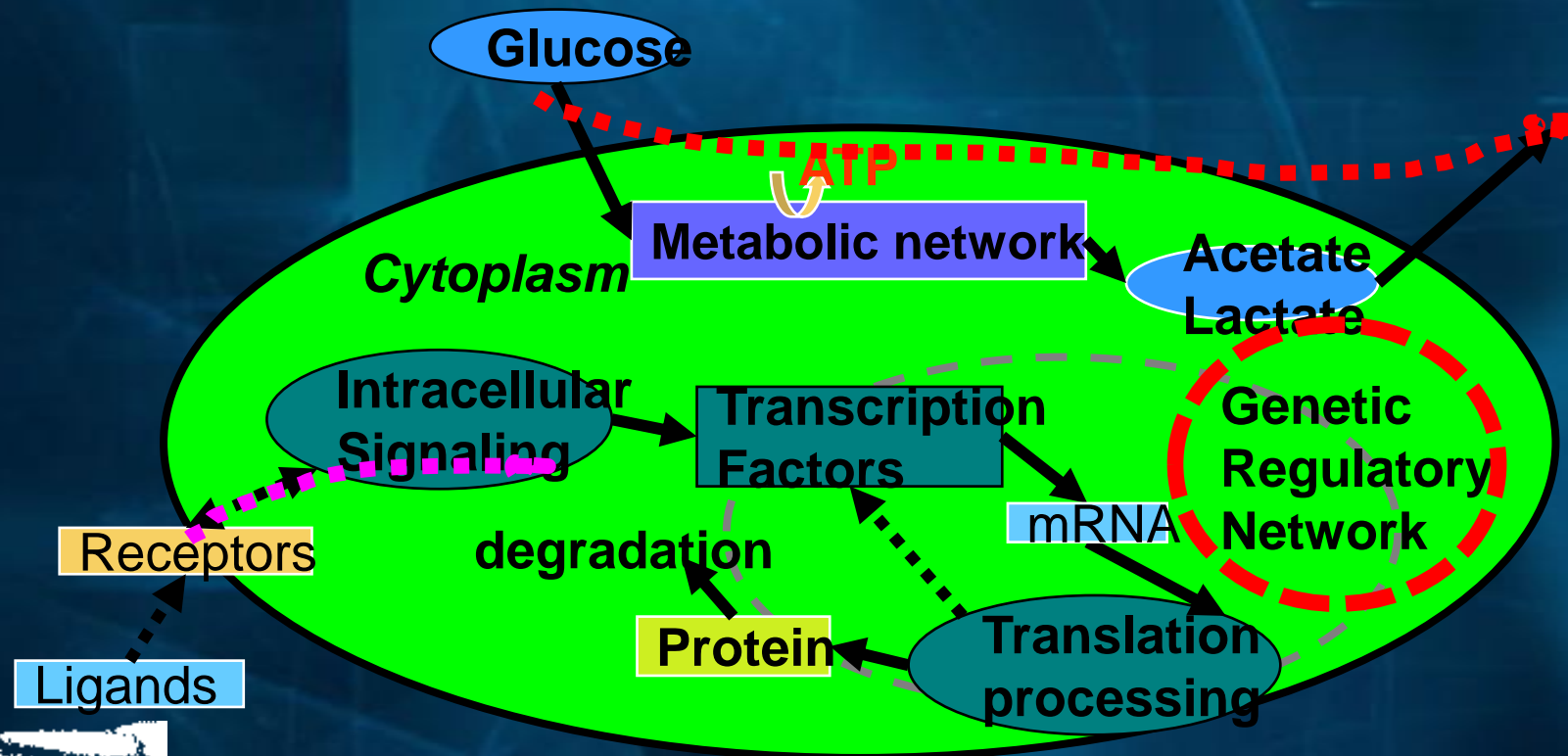
METABOLIC ENGINEERING



edited by
Sang Yup Lee
Eleftherios T. Papoutsakis



Complex Network of Network of Network



Networkⁿ



Genotype



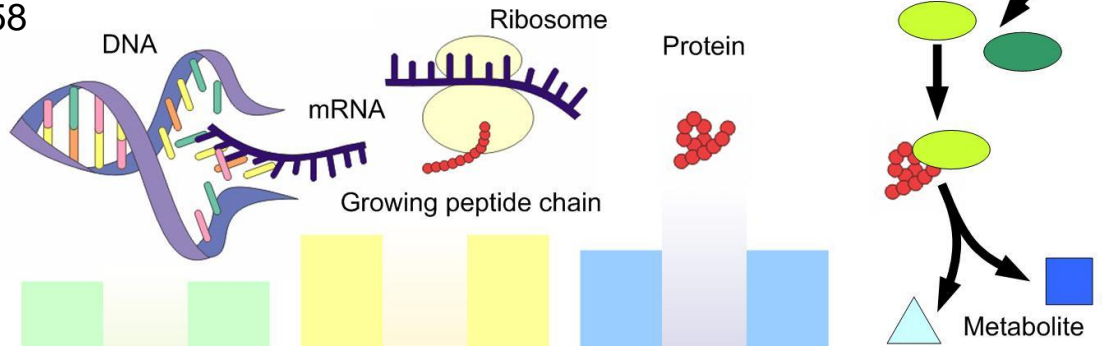
DNA

RNA

Protein

Metabolite

Lee et al.
(2005) Trends
Biotechnol.,
23: 349-358



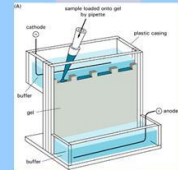
Equipment



DNA sequencer



Microarray



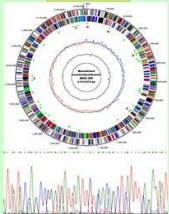
2-D gel, MS/MS



GC/MS, NMR

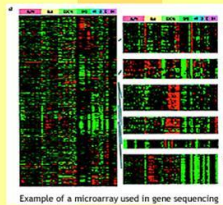
High-throughput data

DNA sequence data



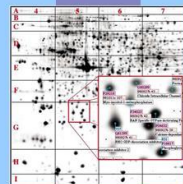
Genome

mRNA profile



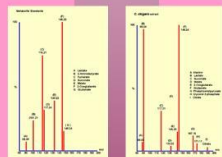
Transcriptome

Protein profile



Proteome

Metabolite profile



Metabolome

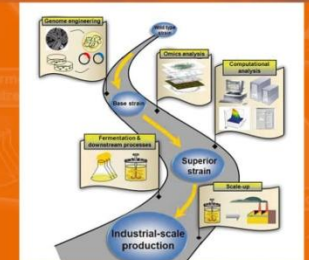
Lee Ed.

Sang Yup Lee
Editor



Systems Biology and Biotechnology of *Escherichia coli*

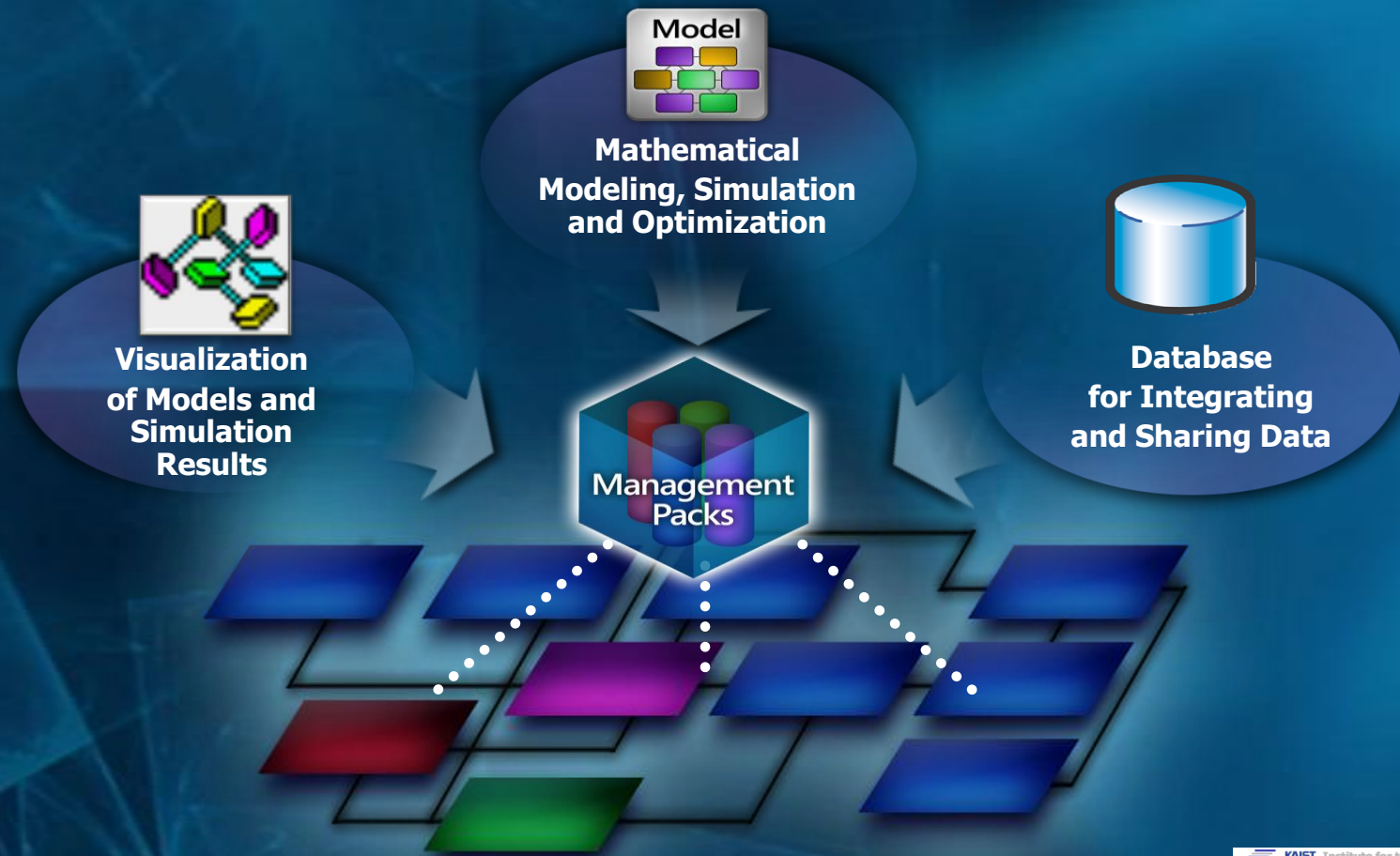
**Systems Biology
and Biotechnology
of *Escherichia coli***



KAIST-Microsoft Collaborative Project for Developing Systems Biology.Net (SB.NET)

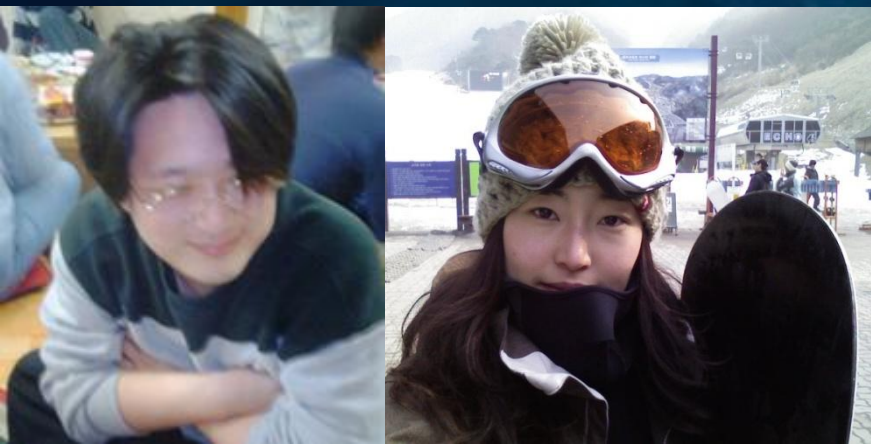
Microsoft
Research

- An integrated computational environment linking all data with theory, modeling, simulation, and experiments to help understand the biological system as a whole

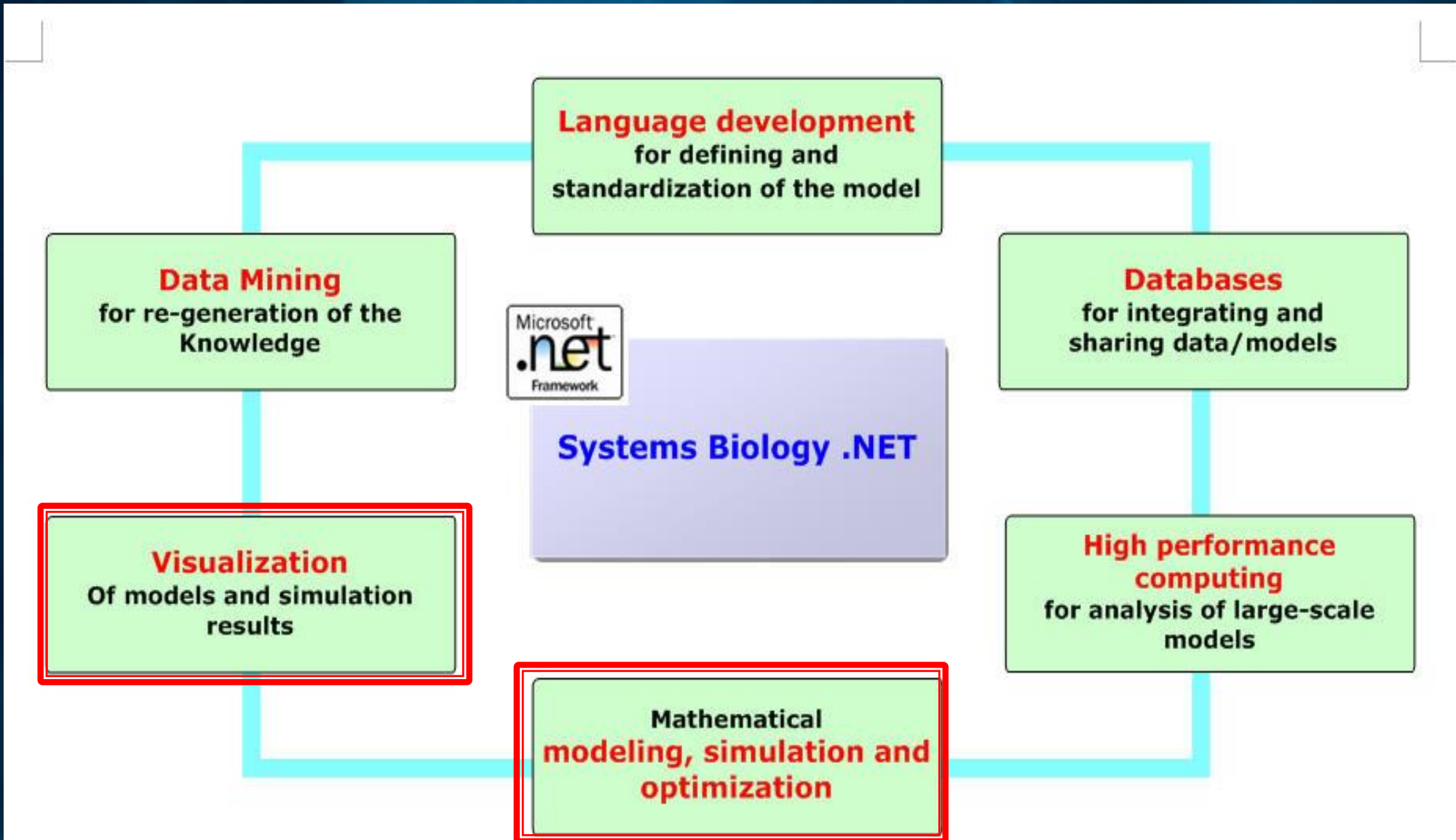


Dr. H. Yoon

Ms. J. Chu



Microsoft
Your potential. Our passion.™



KAIST-Microsoft Project on SB.NET (1st phase)

- Mathematical modeling, simulation and optimization
 - Pseudo-steady state simulation
 - Dynamic simulation
 - Hybrid simulation
 - Final release
- Visualization of models and simulation results
 - Data management
 - Visualization of genome-scale metabolic networks
 - Visualization of regulatory networks
 - Visualization of two combined networks using 3D space
 - User commands
 - Final release

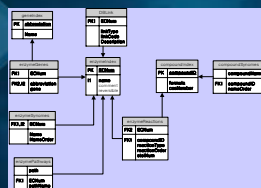
SB.NET Project Overview

Analysis of metabolic networks

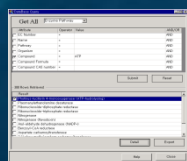
- Construction of reaction models
- Estimation of flux distributions
- Validation of metabolic networks
- Metabolic behavior in response to genetic and/or environmental modifications
- Visualization of metabolic networks

Integration of biochemical data

- Thorough comparison of different DBs
- Standardization
- Integration



Global Schema

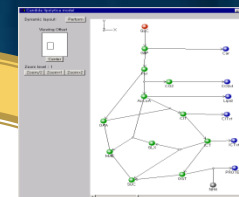


DB Query

MFAML



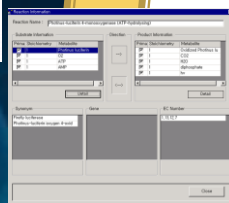
User



Visualization

Time	Glucose	ATP	ADP	...
0	100	0	0	...
10	95	5	5	...
20	90	10	10	...

Dynamic Simulation



Modeling Tool

Reaction	Flux	...
R1	1.0	...
R2	0.5	...
R3	2.0	...

MFA

Other software programs

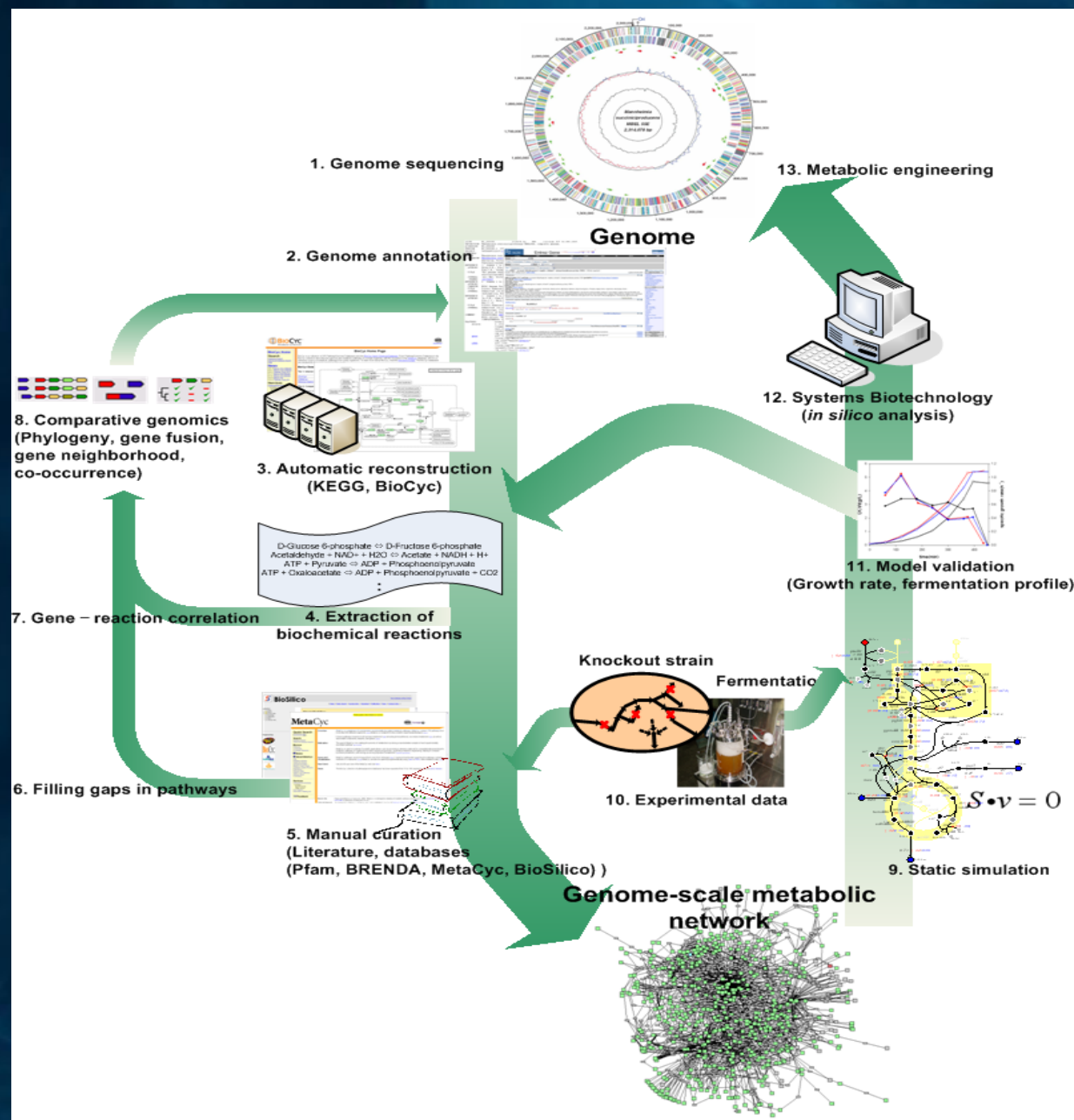


Automatic importing parser

DB Server

Public database

Genome-Scale Metabolic Network Models



Biochemical
Reactions

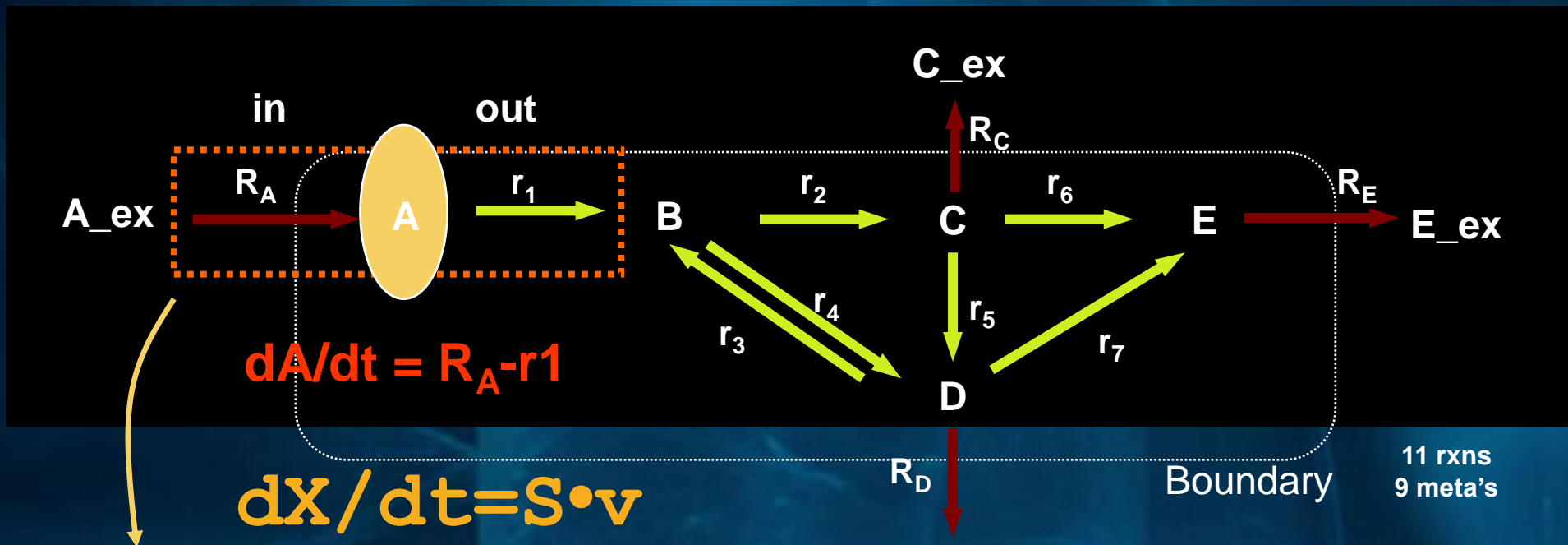
Stoichiometry

Cofactors

Reversibility

Biomass

ATP requirement



Steady state Mass Balance

$$S \cdot v = 0$$

A: $+R_A - r_1 = 0$
 B: $+r_1 - r_4 - r_2 + r_3 = 0$
 C: $+r_2 - r_5 - r_6 - R_C = 0$
 D: $+r_5 + r_4 - r_3 - r_7 - R_D = 0$
 E: $+r_6 + r_7 - R_E = 0$



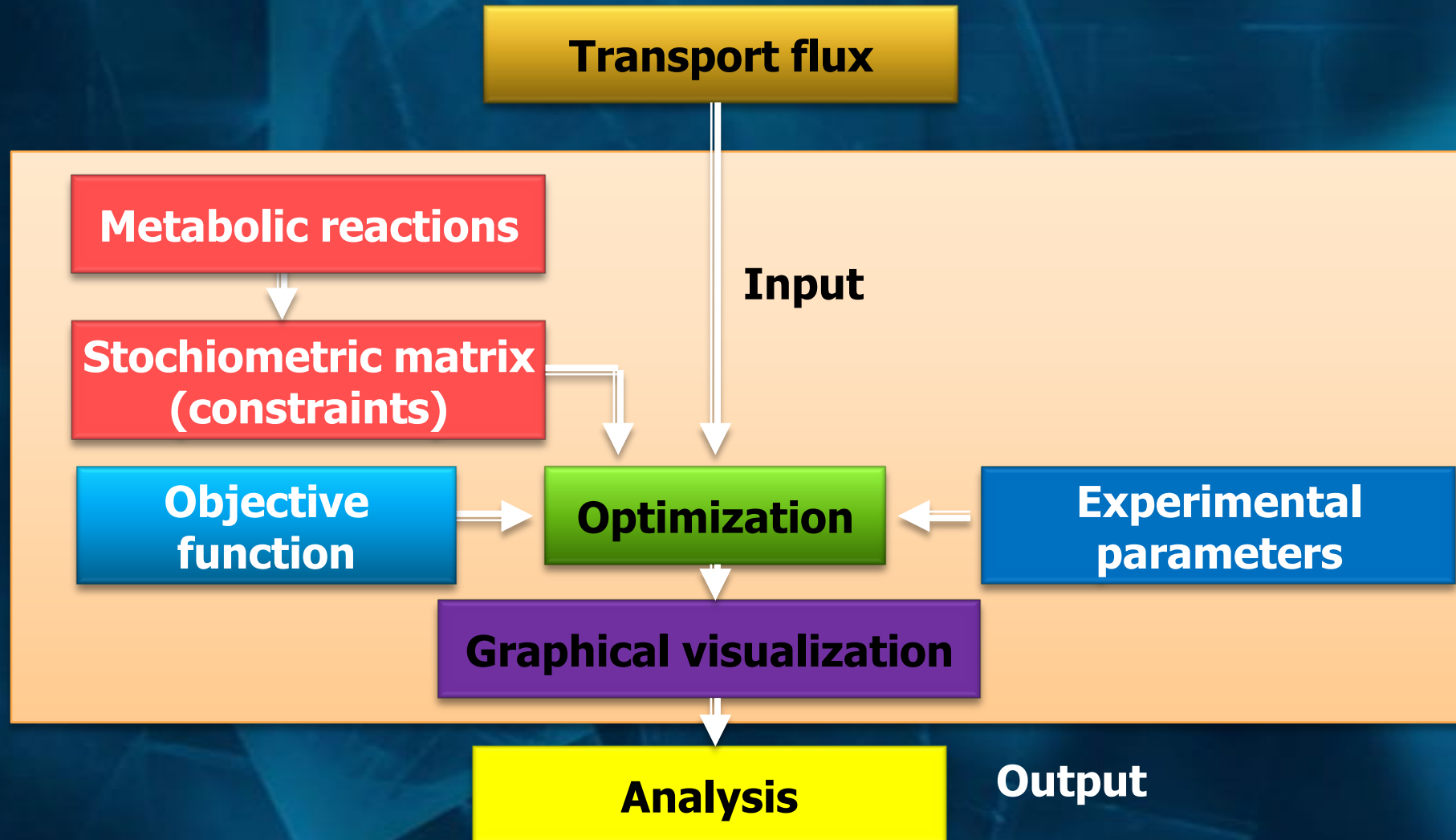
$$\begin{pmatrix}
 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 1 & -1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 1 & 0 & 0 & -1 & -1 & 0 & -1 & 0 & 0 \\
 0 & 0 & 0 & -1 & 1 & 1 & 0 & -1 & 0 & -1 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & -1
 \end{pmatrix}
 \begin{pmatrix}
 R_A \\
 r_1 \\
 r_2 \\
 r_3 \\
 r_4 \\
 r_5 \\
 r_6 \\
 r_7 \\
 R_C \\
 R_D \\
 R_E
 \end{pmatrix}
 =
 \begin{pmatrix}
 0 \\
 0 \\
 0 \\
 0 \\
 0
 \end{pmatrix}$$

S

S = stoichiometric matrix with K metabolites and J reactions $[K \times J]^V$

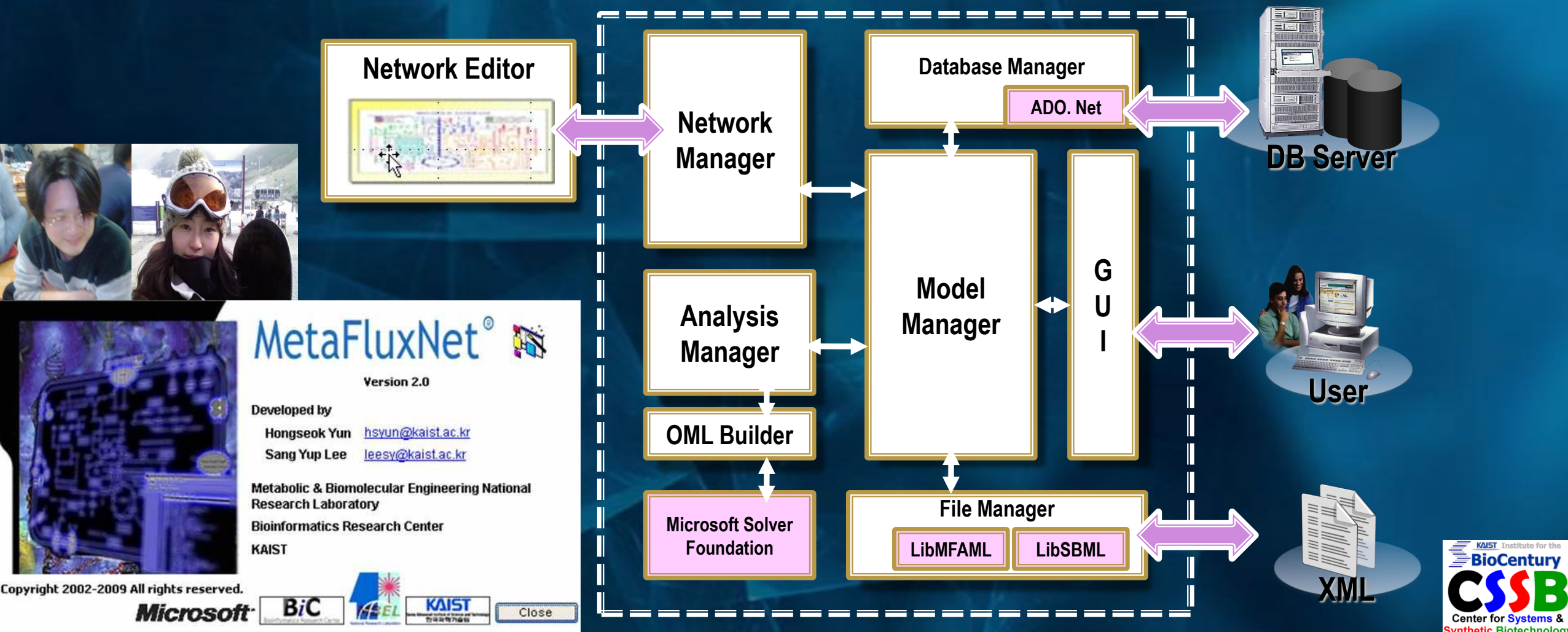
v = flux vector $[J \times 1]$

Schematic Flow Chart of MFA



MetaFluxNet2.0

- Integrated computational environment for modeling and simulation of the genome-scale metabolic model



Amino Acids

Umami and sour tastes	Sweet taste	Bitter taste
Glutamate Aspartic acid	Glycine Alanine Threonine Proline Serine Glutamine	Phenylalanine Tyrosine Arginine Leucine Isoleucine Valine Methionine Histidine



Val, Leu, Ile : effective in hepatic failure
 Glu : antiulcer drug
 Arg : immune-enhancing effect

Val, Leu, Ile : Muscle building, increase of stamina, recovery from fatigue



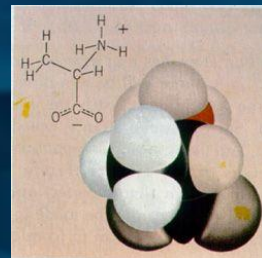
Food and animal feed



Medical care

Sports

Health care



Amino acid

Beauty care



Platform chemicals

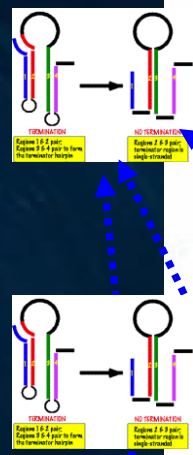
The skin moisturizing effect
 Collagen composition
 Care of damaged cuticles
 Efficient burning of body fat

20 kinds of amino acids that make up proteins

Essential amino acids

Valine Lysine Histidine Leucine
 Methionine Tryptophan Threonine
 Phenylalanine Isoleucine

Glycine Serine Glutamine Tyrosine
 Glutamate Alanine Proline
 Cysteine Asparagine Aspartic acid
 Arginine



glucose

L-threonine

pyruvate

pyruvate

2-ketobutyrate

acetoxy acid Synthase I, II, III

ilvA

ilvBN, ilvGM, ilvIH

ilvBN, ilvGM, ilvIH

2-acetolactate

2-aceto-2-hydroxybutyrate

ilvC acetoxy acid isomeroeductase

ilvC

2,3-dihydroxyisovalerate

branched chain amino acid aminotransferase

ilvD

2-ketoisovalerate

L-isoleucine

ilvD dihydroxy acid dehydratase

ilvE

leuA

panB

leuCD

ilvE

panE

leuB

tyrB

panC

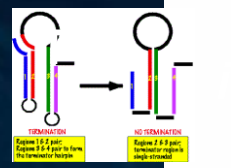
L-leucine

L-valine

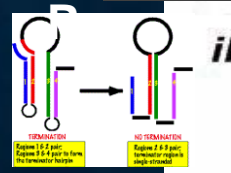
pantothenate



Jin Hwan Park



$\Delta lacl$



41th G → A
50th C → T

2-isopropylmalate synthase

Auxotrophic
2 mM Leu
2 mM Ile
1.5 μM pantoth

L-leucine

glucose



pyruvate



2-acetolactate



2,3-dihydroxyisovalerate



2-ketoisovalerate



L-leucine



L-valine



pantothenate

pyruvate

L-threonine



2-ketobutyrate



2-aceto-2-hydroxybutyrate



L-isoleucine

ilvBN, *ilvGM*, *ilvIH*

ilvBN, *ilvGM*, *ilvIH*

ilvC

ilvC

ilvD

ilvE

ilvD

2-ketoisovalerate

L-isoleucine

leuA

leuCD

leuB

tyrB

panE

panB

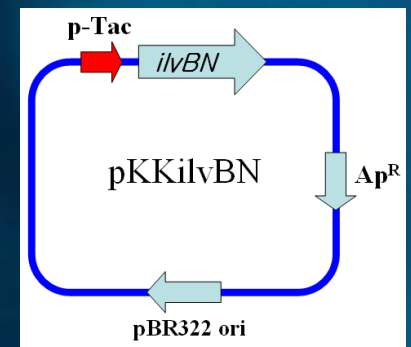
panE

panC

ilvE

3-methyl-2-oxobutanoate hydroxymethyltransferase

ilvA thr dehydratase



Base strain

→ Val {W3110(*attilvG*::*ptac*, *attilvB*::*ptac*, *ilvH*^{A41G, C50T}, $\Delta lacl$, $\Delta ilvA$, $\Delta leuA$, $\Delta panB$)} harboring pKKilvBN

Fermentation and Transcriptome Analysis

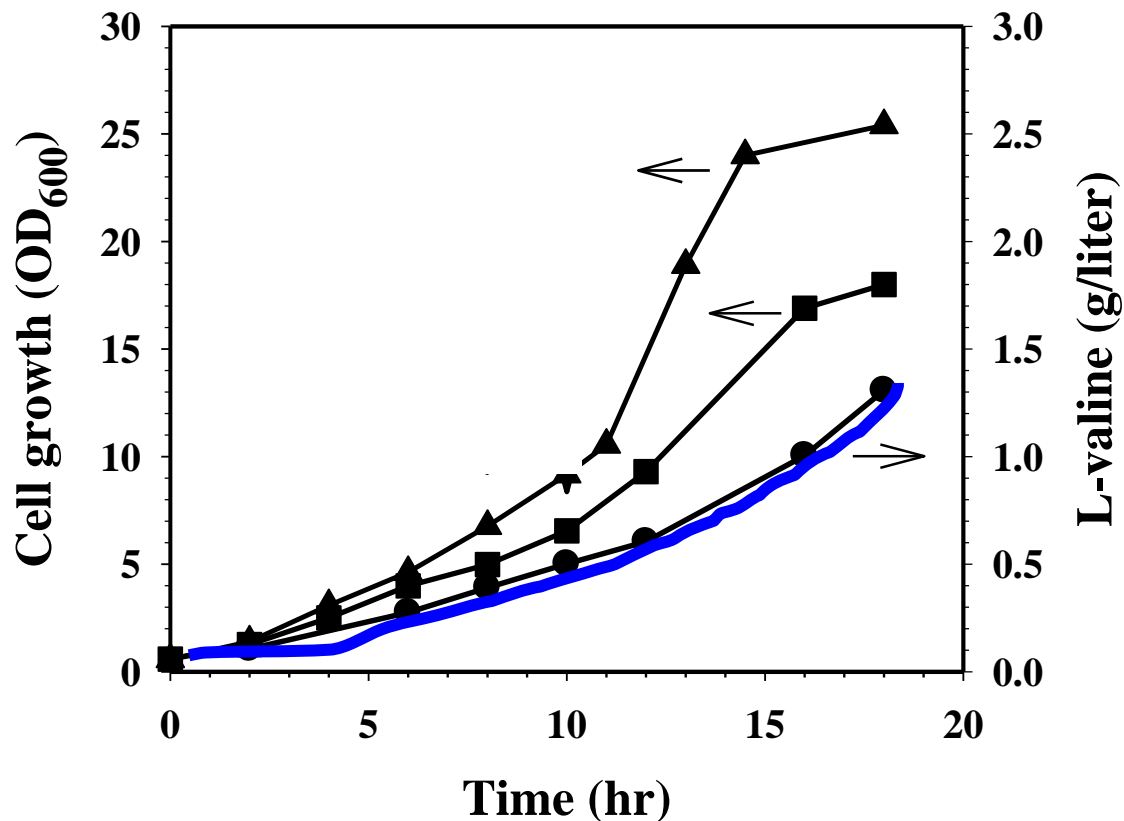
Microsoft
Research

Val : W3110(*attilvG::ptac*, *attilvB::ptac*, *ilvH^{A41G, C50T}*, Δ *lacl*, Δ *ilvA*, Δ *leuA*, Δ *panB*), pKKilvBN

Control : W3110 (Δ *lacl*, pKK223-3)

Medium NM1, Glucose (20 g/L), L-leucine (2mM), L-isoleucine (2mM), D-pantothenate (1.5 μ M)

Condition 31°C, pH 6.0



▶ 1.31 g/L L-valine
by batch fermentation

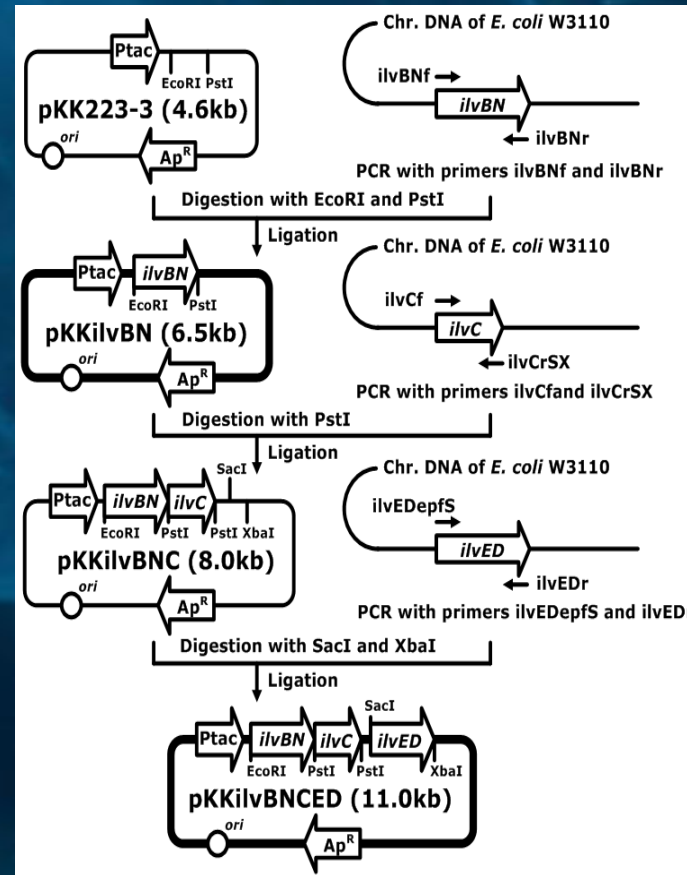


Jin Hwan Park

Overexpression of Genes

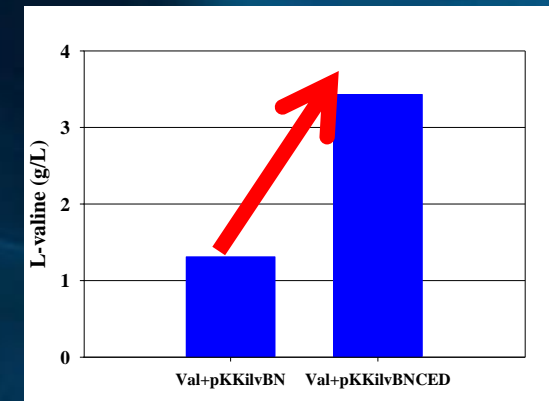
Relative expression level of L-valine biosynthetic pathway genes

gene	enzyme	chip data
<i>ilvB</i>	acetohydroxy acid synthase isoenzyme I	54.46
<i>ilvN</i>	acetohydroxy acid synthase isoenzyme I	32.50
<i>ilvC</i>	acetohydroxy acid isomeroreductase	3.74
<i>ilvD</i>	dihydroxy acid dehydratase	4.28
<i>ilvE</i>	branched chain amino acid aminotransferase	1.32



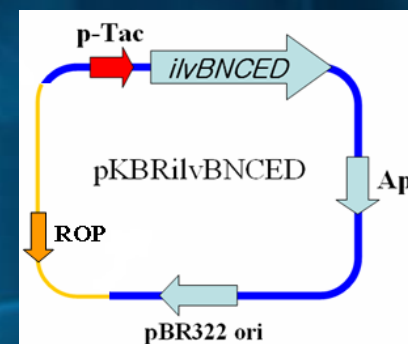
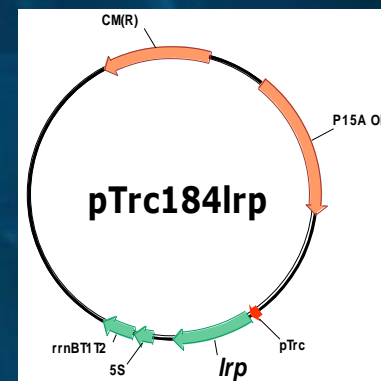
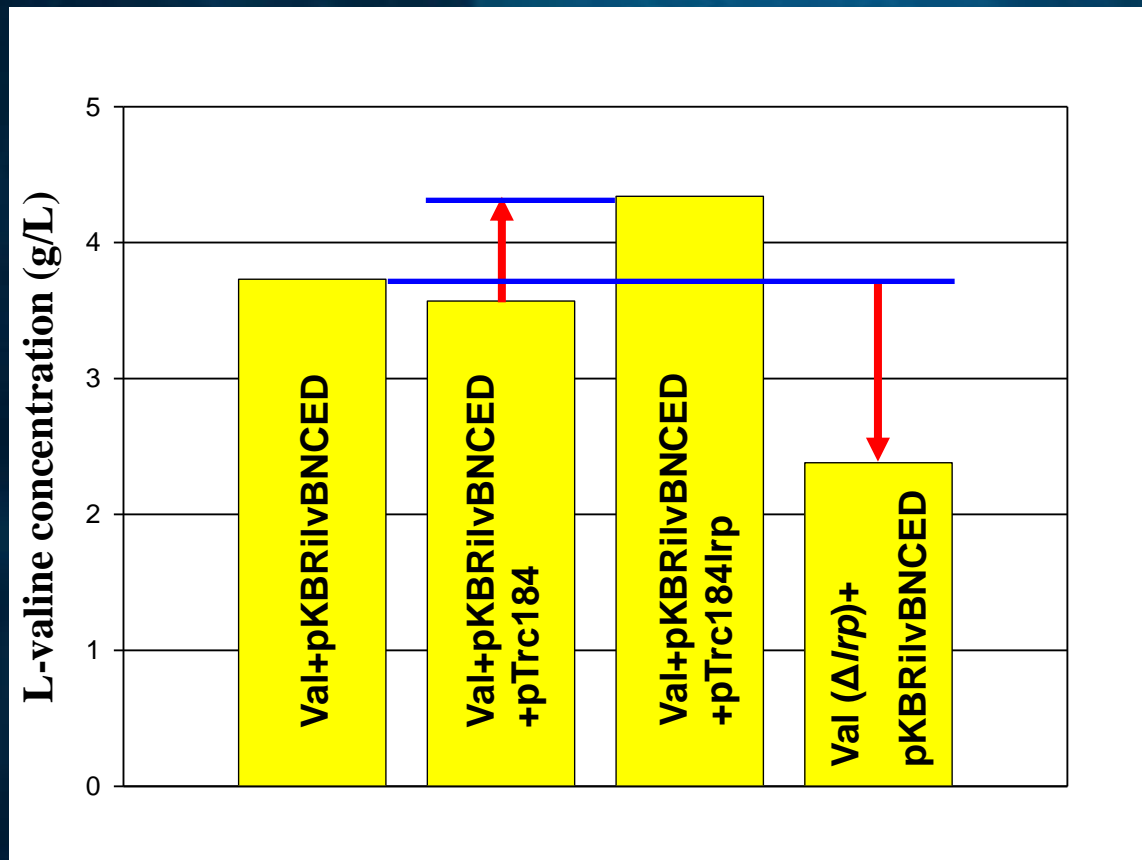
Co-amplification of the *ilvCED* genes in pKKilvBN

Increased L-valine production (1.31 → 3.43 g/L)



Leucine responsive protein: Lrp → Downregulated (ratio: 0.52)

Enhanced production of L-valine by overexpression of *Irp*



21.6% increase (3.57 → 4.34 g/L) in L-valine production with *Irp* overexpression

36.2% decrease (3.73 → 2.38 g/L) in L-valine production with *Irp* deletion

→ Lrp plays an important role in L-valine production

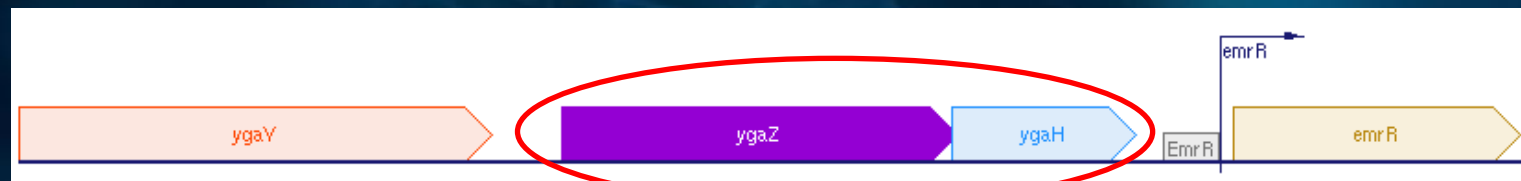
Identification of the *E. coli* gene homologous to *Corynebacterium glutamicum brnF*

Score = 72.4 bits (176), Expect = 2e-13
Identities = 56/204 (27%), Positives = 95/204 (46%), Gaps = 15/204 (7%)

BrnF	34	QGLKTSLAAGLGMYPIGIAFGLLVIOQGYEWAAAPLFSGLIFAGSTEMLVIALVVGAAPL	93
		+G K SL + P+ AFGL + G+ + FS +I+AG+++ ++ A++ + L	
YgaZ	17	EGCKDSLPIVISYIPVAFAFGLNATRLGFSPLSVFFSCIIYAGASQFVITAMLAAGSSL	76
BrnF	94	GAIALTLLVNFVRHVYAFSPLHVV----KNPIARFYVVFALIDEAYAVTAAR----PA	145
		ALT + ++ RHV Y S ++ K+ A + F L DE +A A+	
YgaZ	77	WVAALTVMAMDVRHVLYGPSLRRIIQRLQKSKTALW--AFGLTDEVFAAATAKLVRNMR	134
BrnF	146	GWSAWRLISMQIAFHSYVVF----GGLTGVAIAELIPFEIKGLEFALCSLFVTLTLDSCR	201
		WS +I + + S WVF G +G + + P L F L +LF++ L S +	
YgaZ	135	RWSENMMIGIAFSSWSSWVFGTVIGAFSGSGLLQGYPAVEAALGFMLPALFMSFLLASFQ	194
BrnF	202	TKKQI-PSLLLAGLSFTIALVVIP	224
		K+ + + L G + L IP	
YgaZ	195	RKQSLCVTAALVGALAGVTLFSIP	218

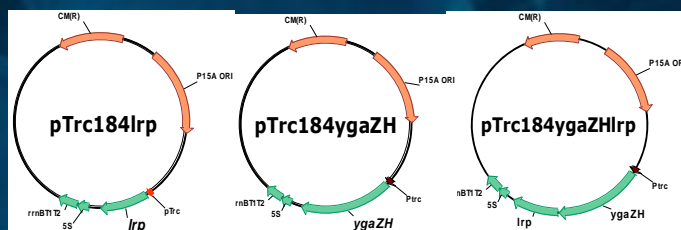
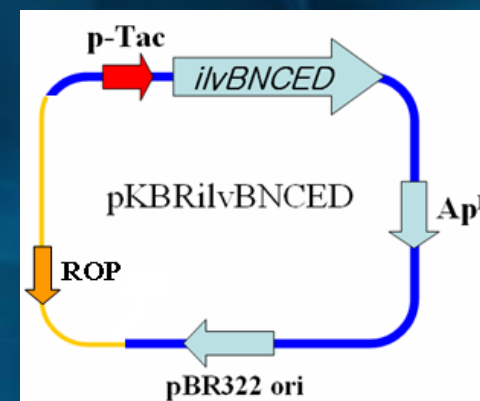
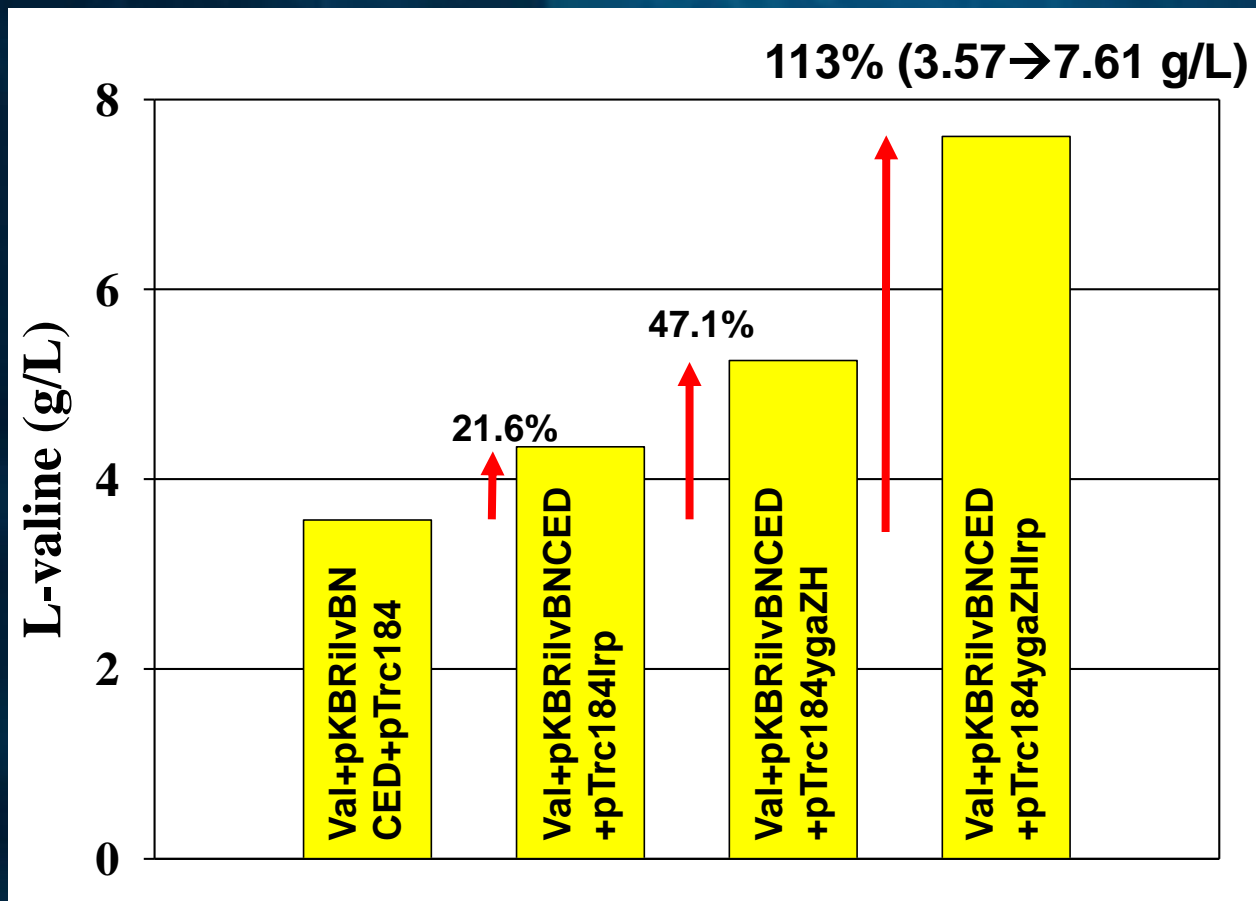
Exporter Engineering

YgaZH: hypothetical protein
 Downregulated to 0.61-0.75 during Val production



L-valine exporter in *E. coli* ?

Synergistic effect of *ygaZH* and *Irp* on L-valine production



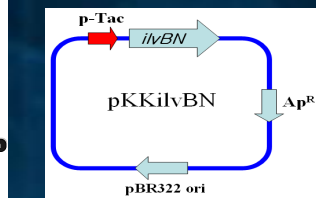
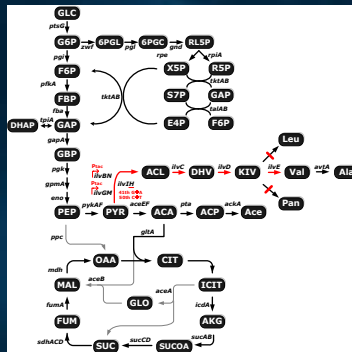
Metabolic engineering of *E. coli* for L-Valine production (flask culture for 48 h @ 31C; 50 g/L glucose)

- Removed feedback inhibition
- Removed transcriptional attenuation control
- Removed major competing metabolic pathways
- Stepwise improvement based on transcriptome analysis

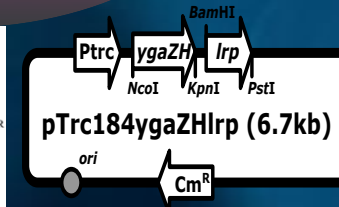
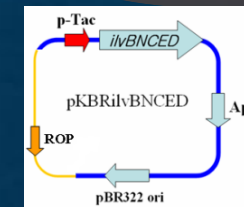
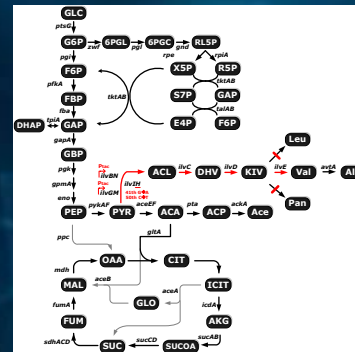
Amplification of
L-valine production
pathways

Global regulator
engineering

Exporter
engineering



Val+pKKilvBN
1.31 g/L



Val+pKBRilvBNCED+pTrc184ygaZHlrp

7.61 g/L

Any more targets to engineer?

Experimental Difficulties

Roche @ ExPASy

Combinatorial gene knockout

$${}^{1000}C_2 = 499,500$$

$${}^{1000}C_3 = 166,167,000$$

$${}^{1000}C_4 = 41,417,124,750$$

...



→ **Genome-scale metabolic flux analysis
to identify new gene-deletion targets**

Developed by

Hongseok Yun hsvun@kaist.ac.kr

Sang Yup Lee leesy@kaist.ac.kr

Metabolic & Biomolecular Engineering National
Research Laboratory

Bioinformatics Research Center

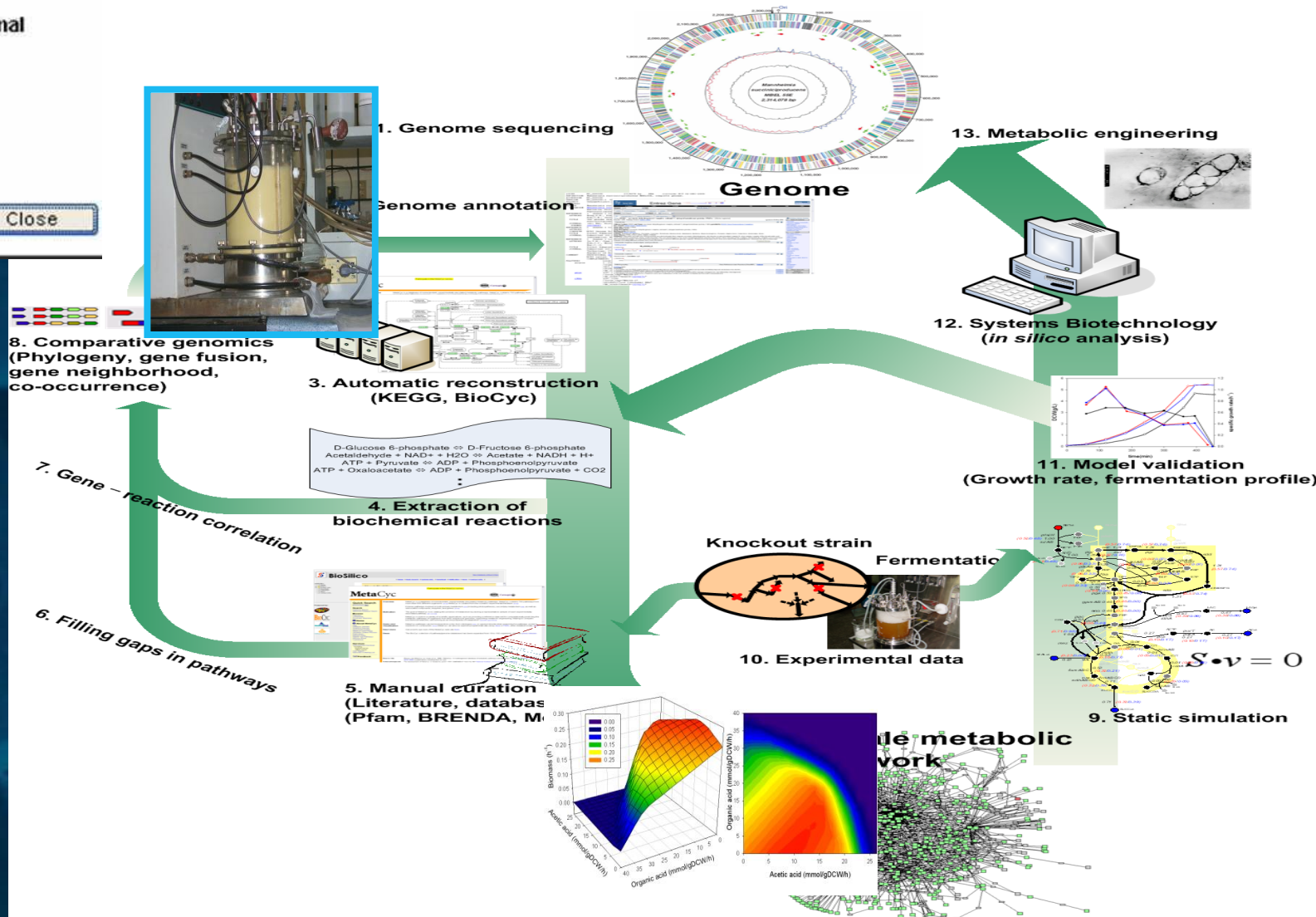
KAIST

Copyright 2002-2009 All rights reserved.

Microsoft



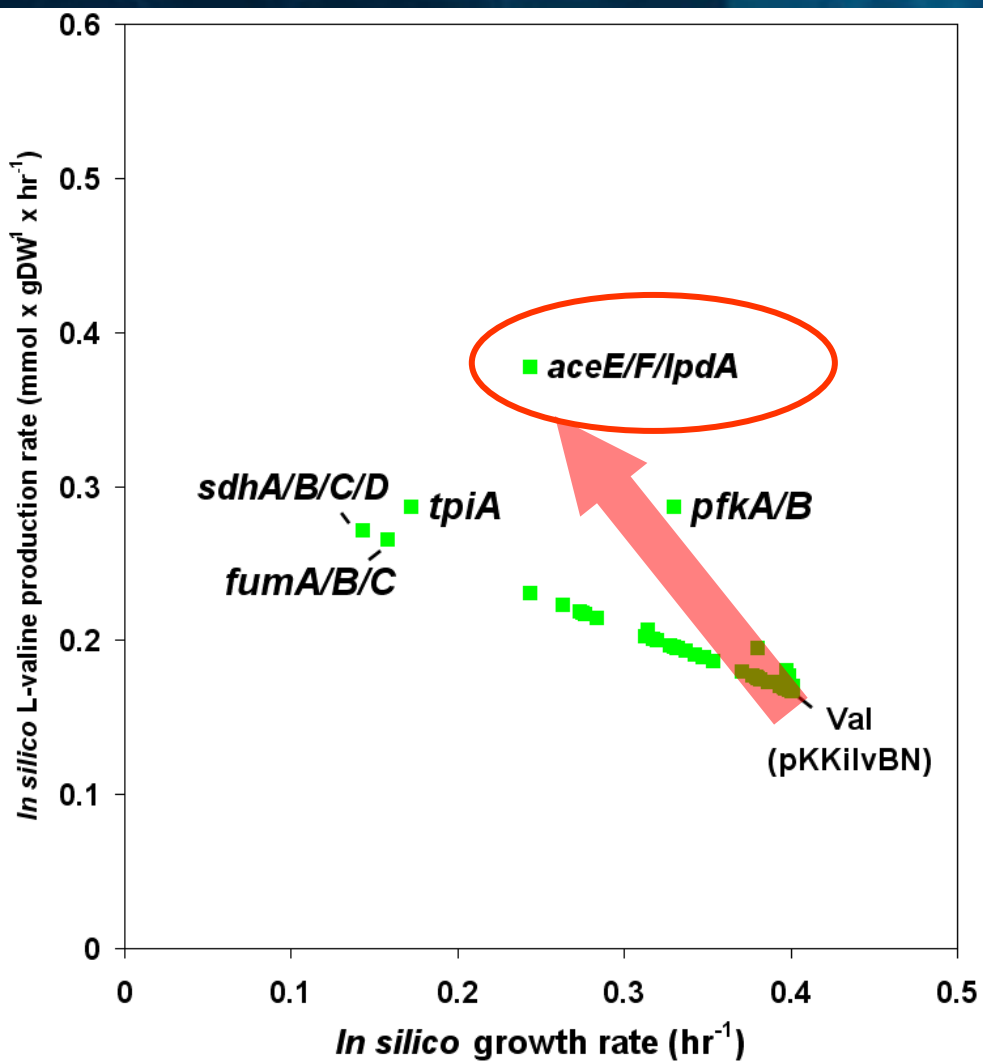
Close



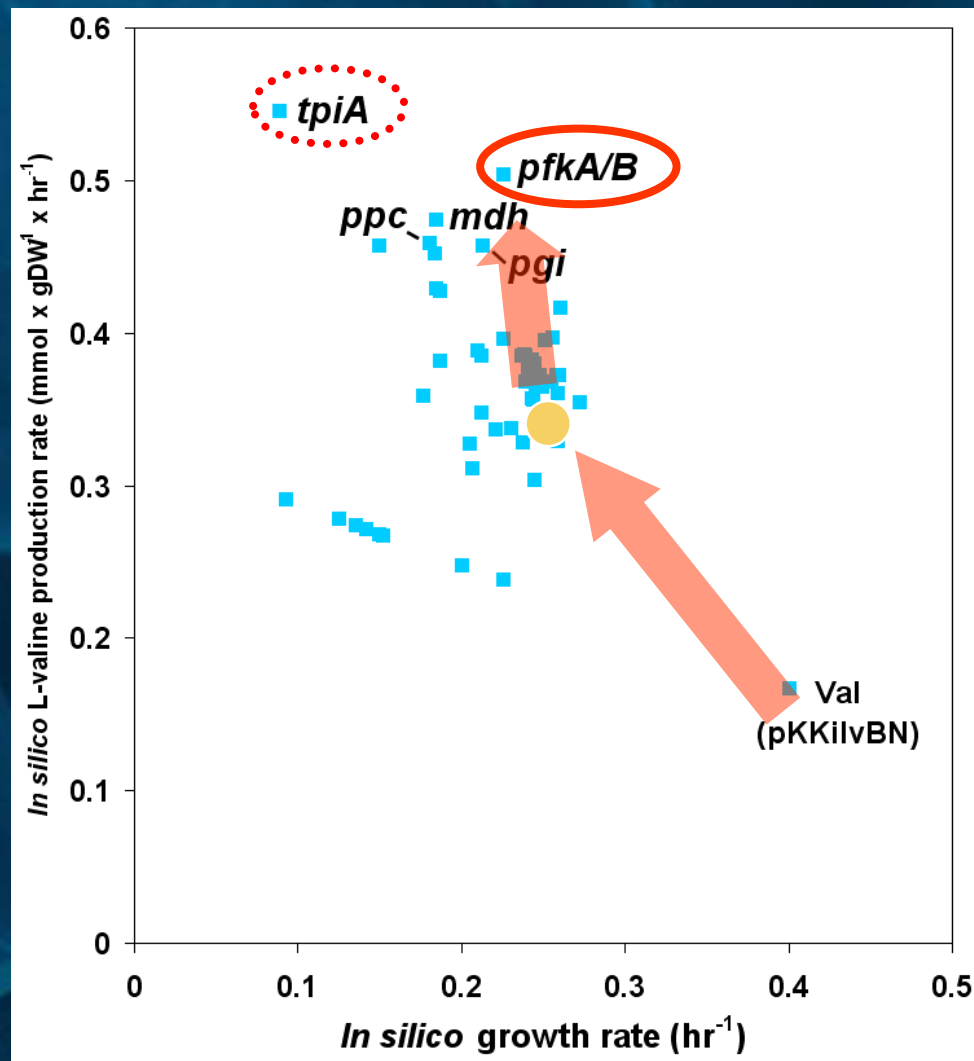
Genome-scale Gene Knockouts



Tae Yong Kim



Single knockout

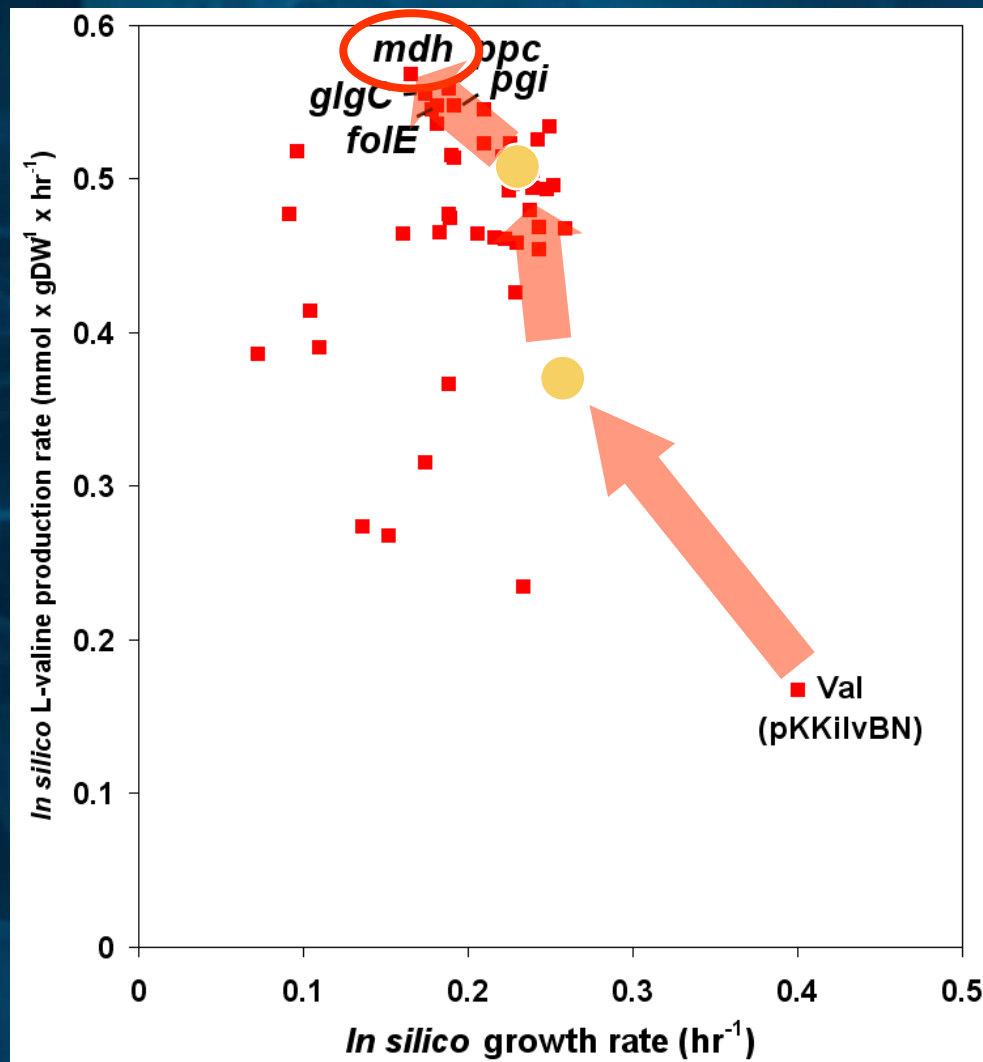


Double knockout

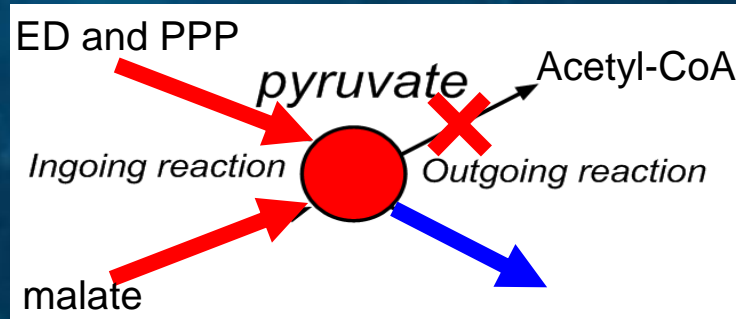
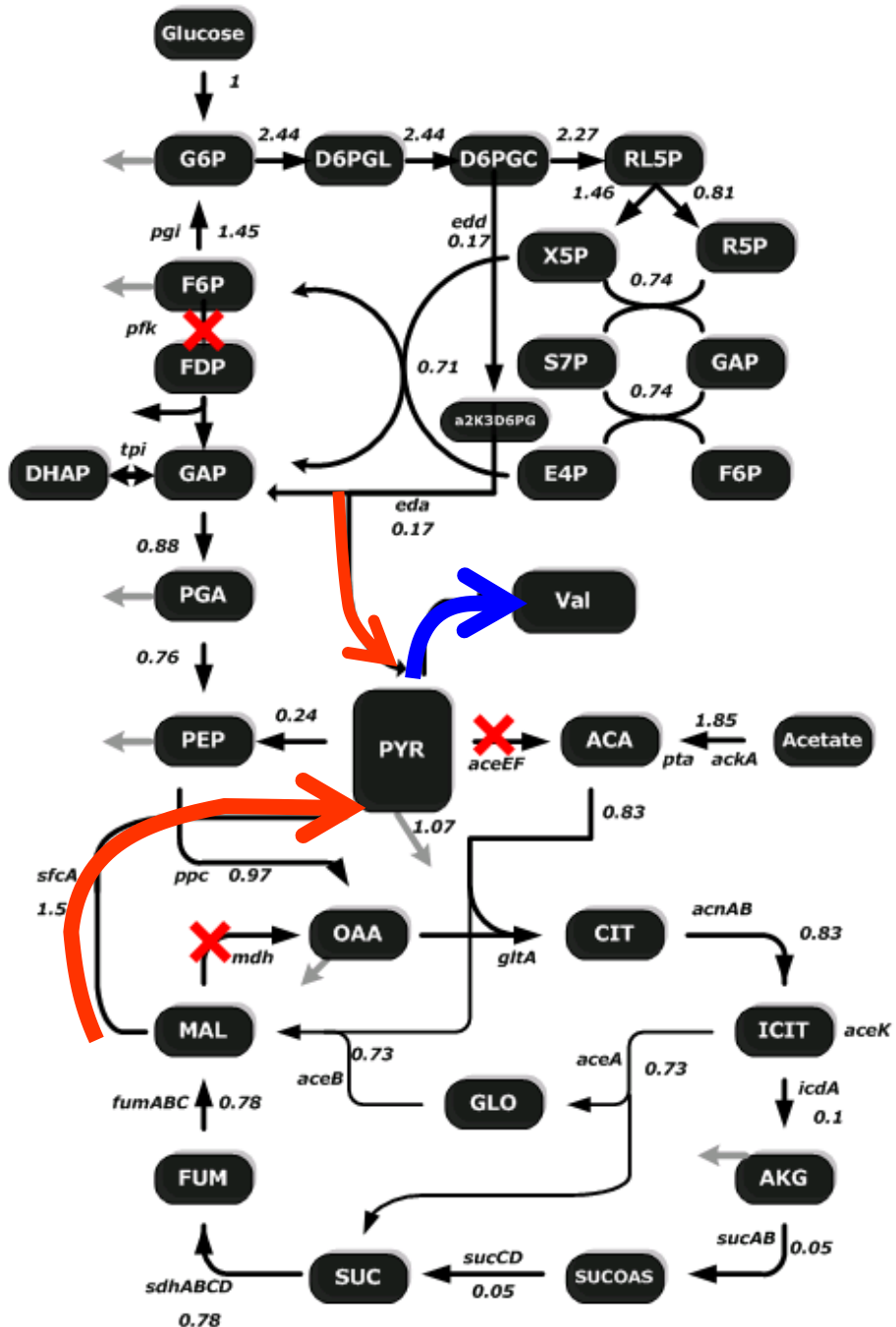
In silico triple gene knockout



Tae Yong Kim



Flux redirection by triple knock-out



**37.75 g L-valine
from 100 g glucose**

<i>aceF/E3pdA</i>	pyruvate dehydrogenase
<i>pfkA/B</i>	phosphofructokinase
<i>mdh</i>	malate dehydrogenase

Selection of genes
for triple knock-out
mutation

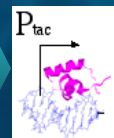
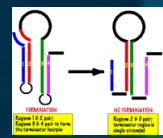
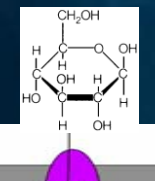


Systems metabolic engineering of *E. coli* for L-valine production

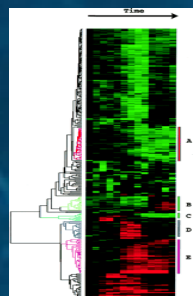
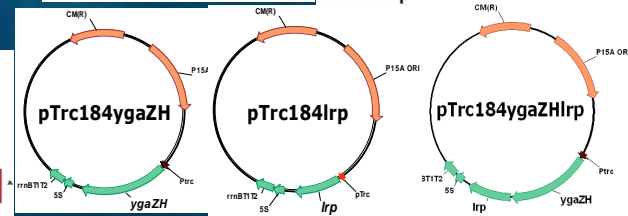
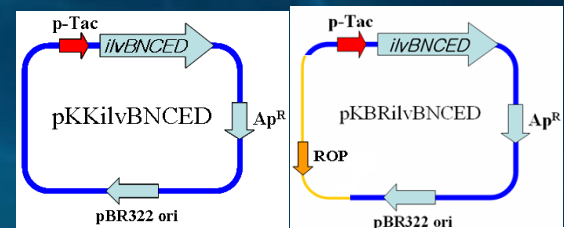
Microsoft
Research

Power of
Systems
Metabolic
Engineering

E. coli W3110



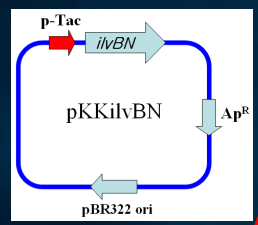
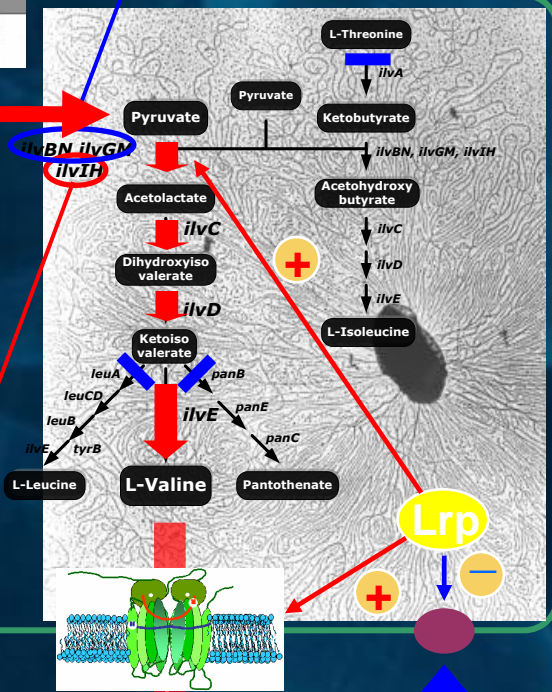
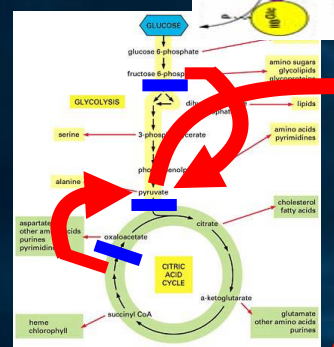
Replacement of attenuator with *tac* promoter



Transcriptome analysis

L-valine overproducer

In silico knock-out simulation

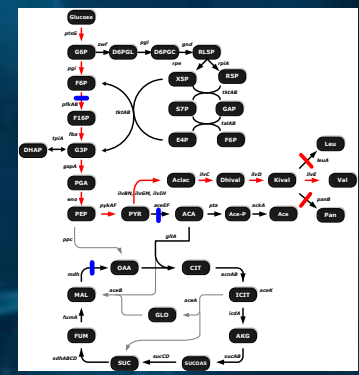
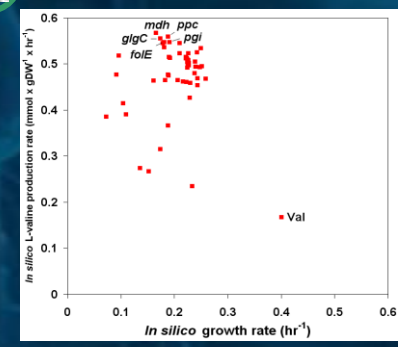
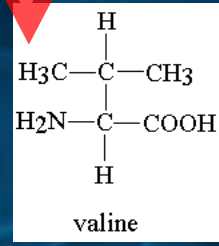


Base strain

TGAATCAGGCGCGTTATCCGCGTGATT

TGAATCAGACGCGTTATCCGCGTGATT

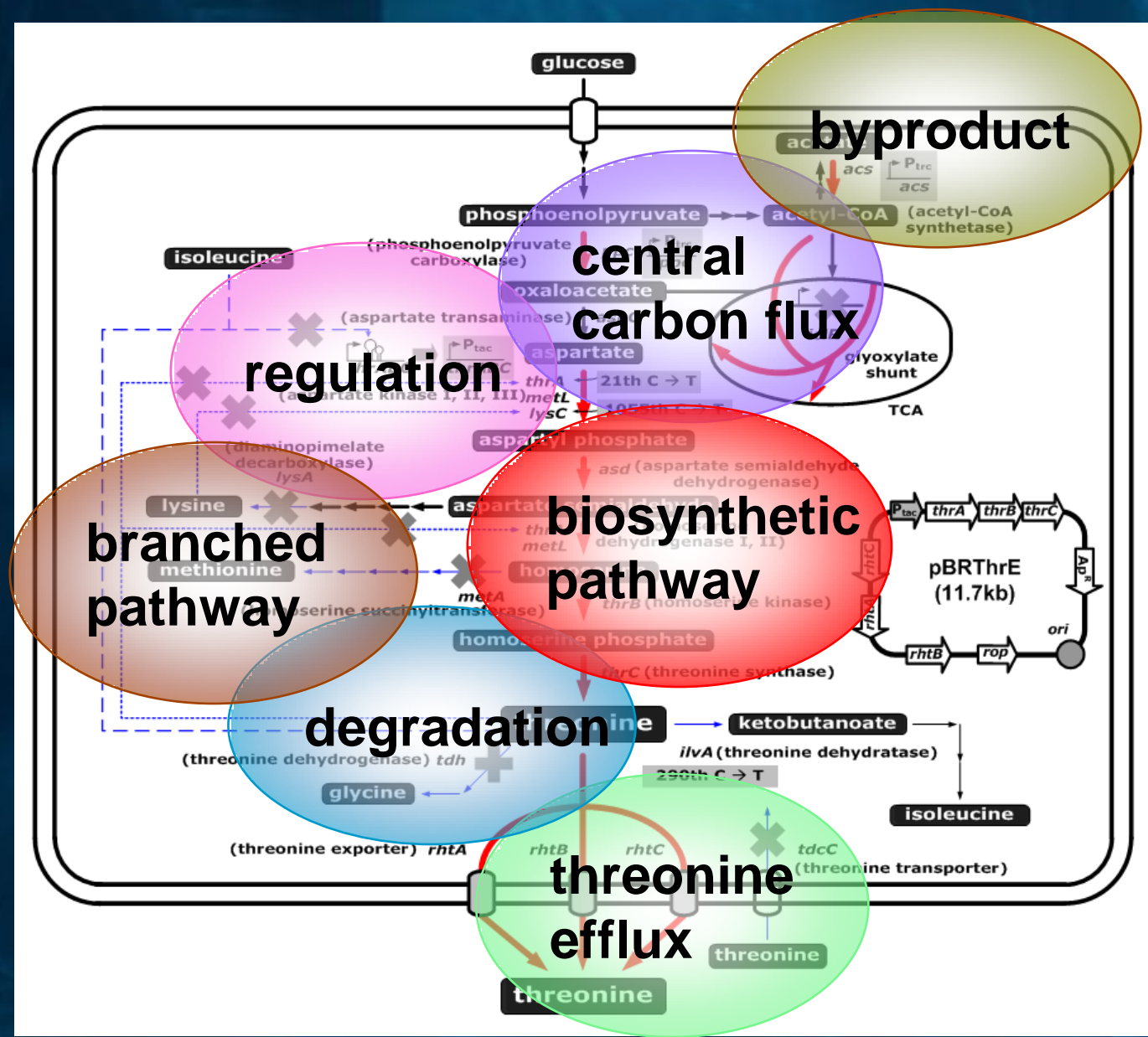
Removal of feedback inhibition



Systems Metabolic Eng for L-Thr

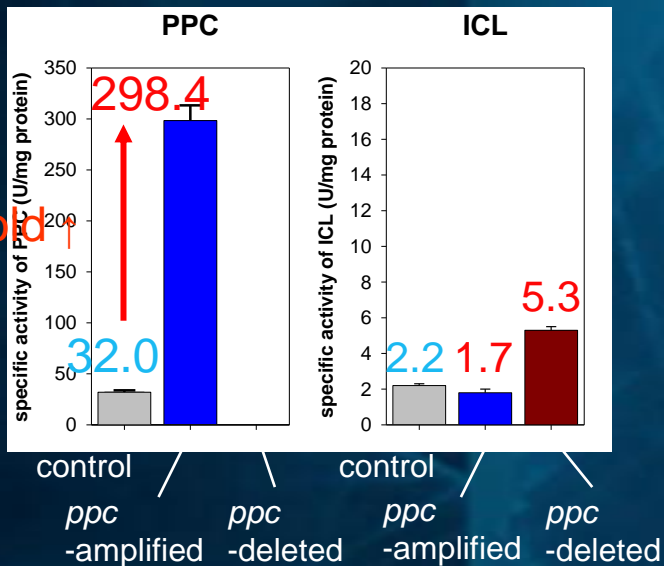


Kwang Ho Lee



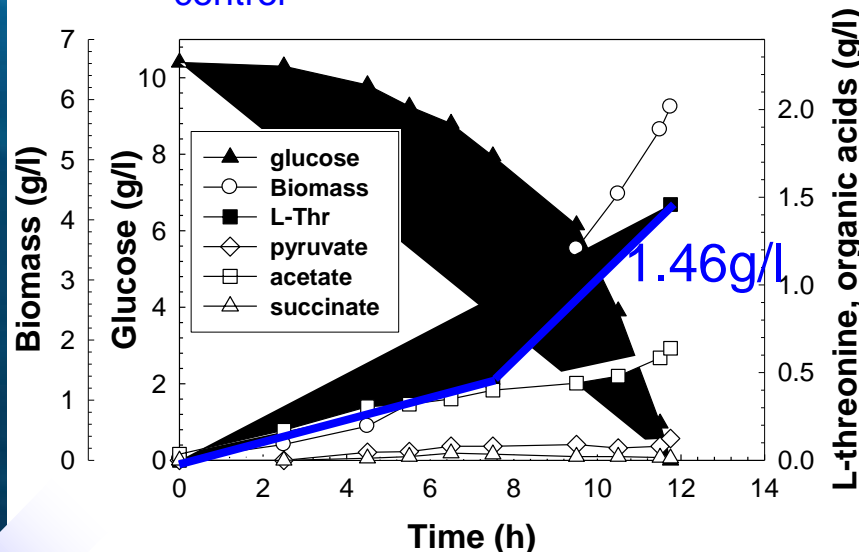
Enzyme assay

9.3 fold ↑



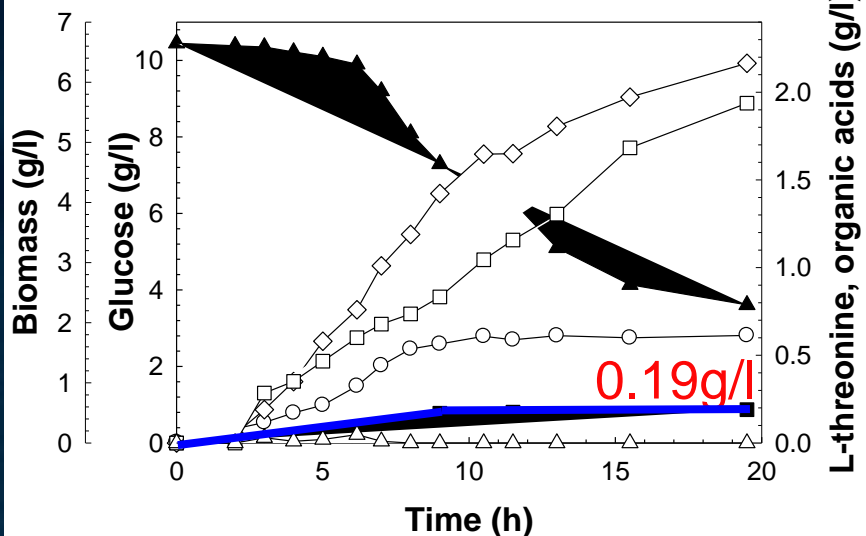
TH07 (pBRThrABC, pACYC177)

control



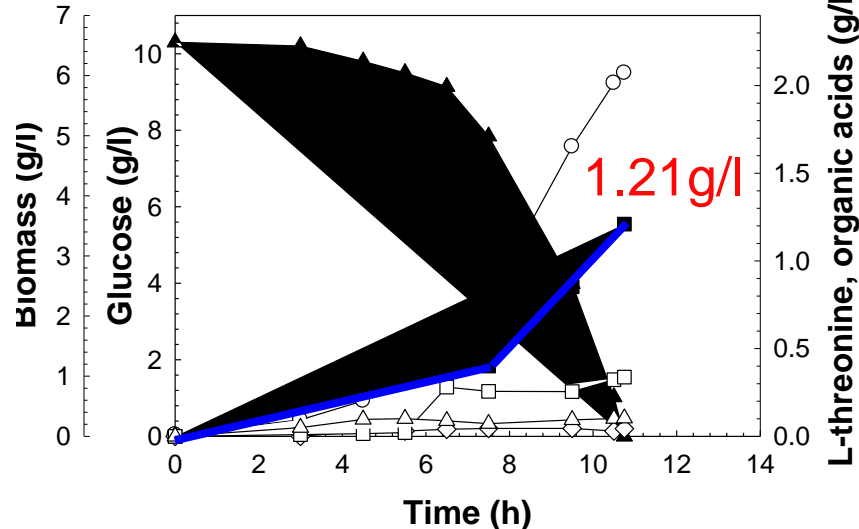
TH11C (pBRThrABC, pACYC177)

deleted

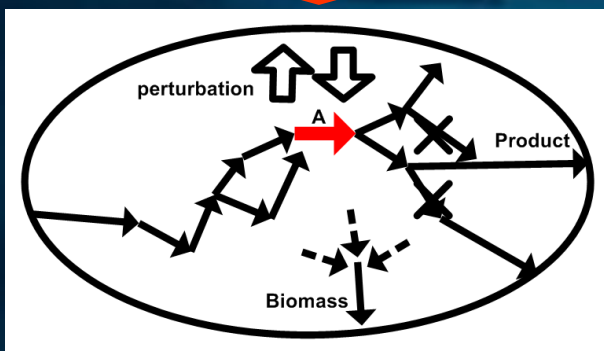
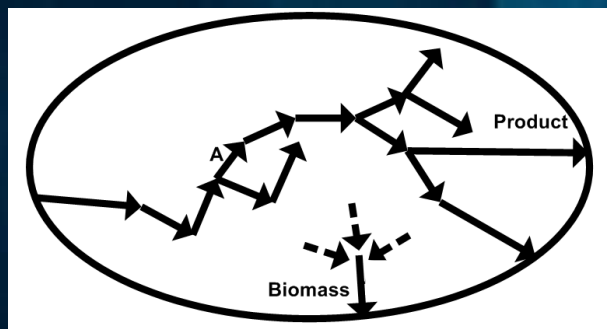


TH07 (pBRThrABC, pACYC-ppc)

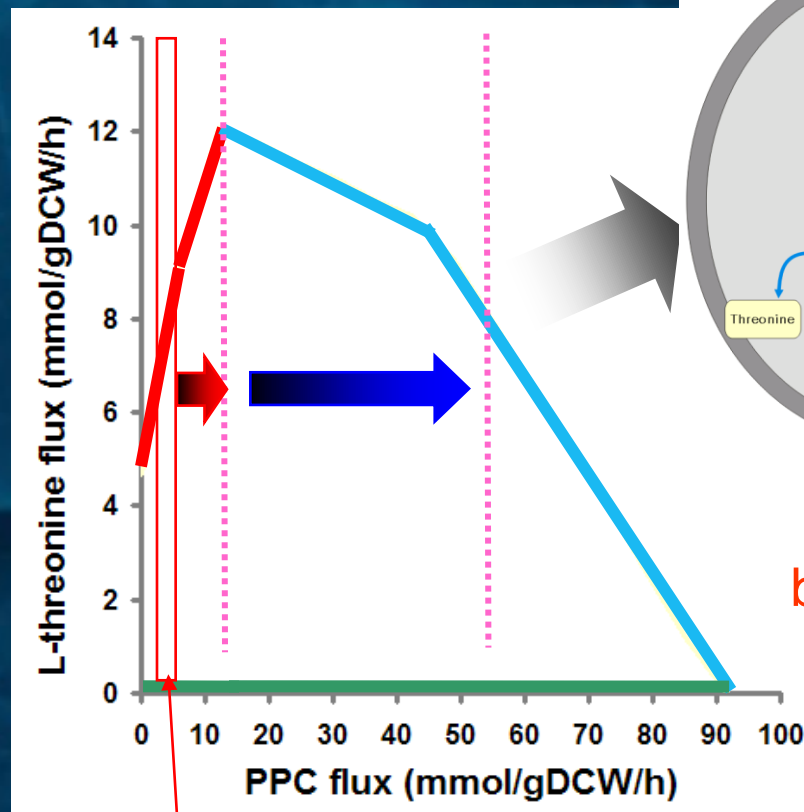
amplified



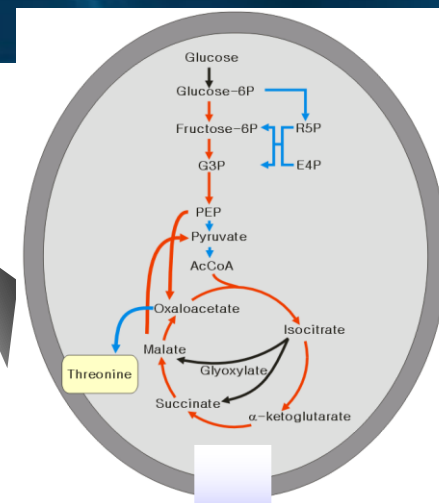
In silico Perturbation Analysis



Genome-scale metabolic model
E. coli MBEL979



the flux of PPC is presented in the < 4 mmol/gDCW/h
in aerobic condition (Fisher et al., Anal. Chem. 2004)



biomass



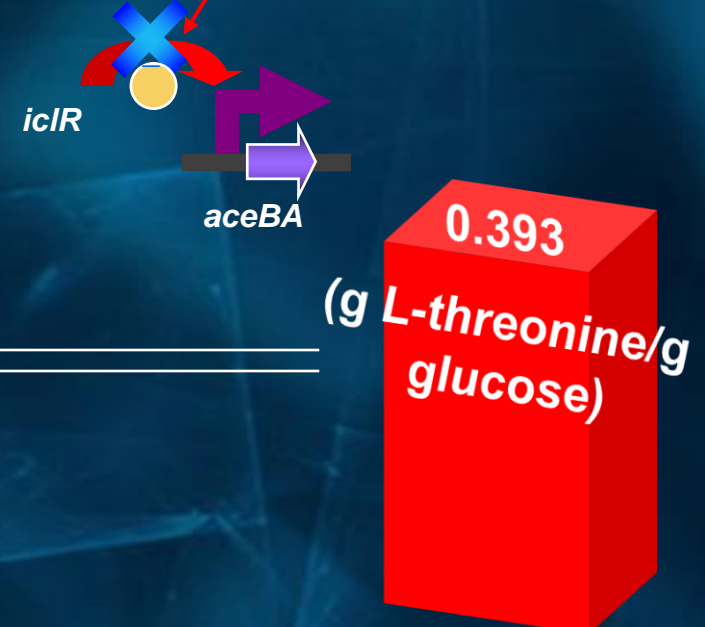
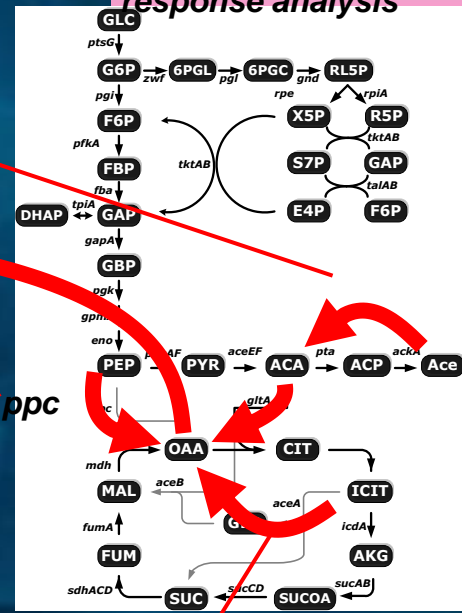
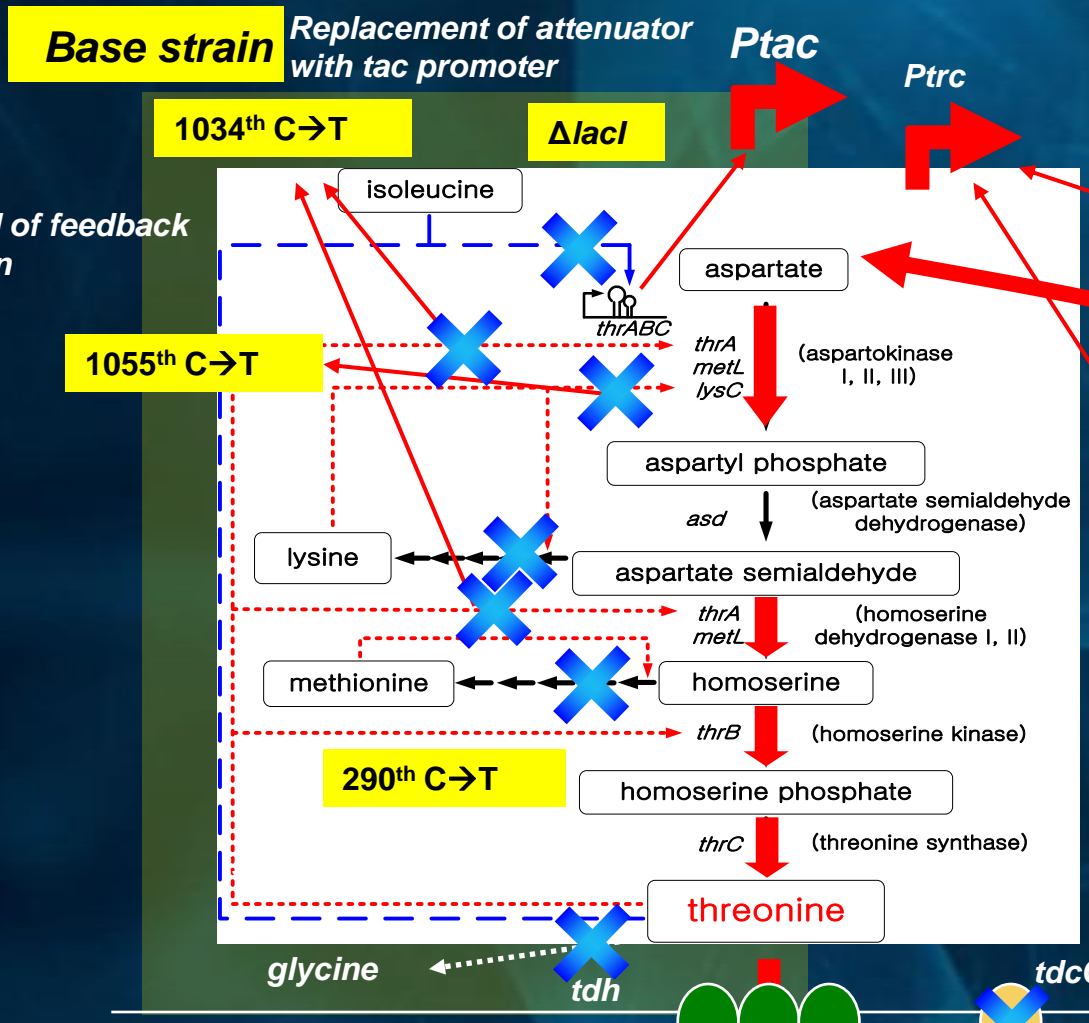
Tae Yong Kim



Hyun Uk Kim

Systems metabolic engineering of *E. coli* for the production of L-threonine

Transcriptome analysis & in silico flux response analysis

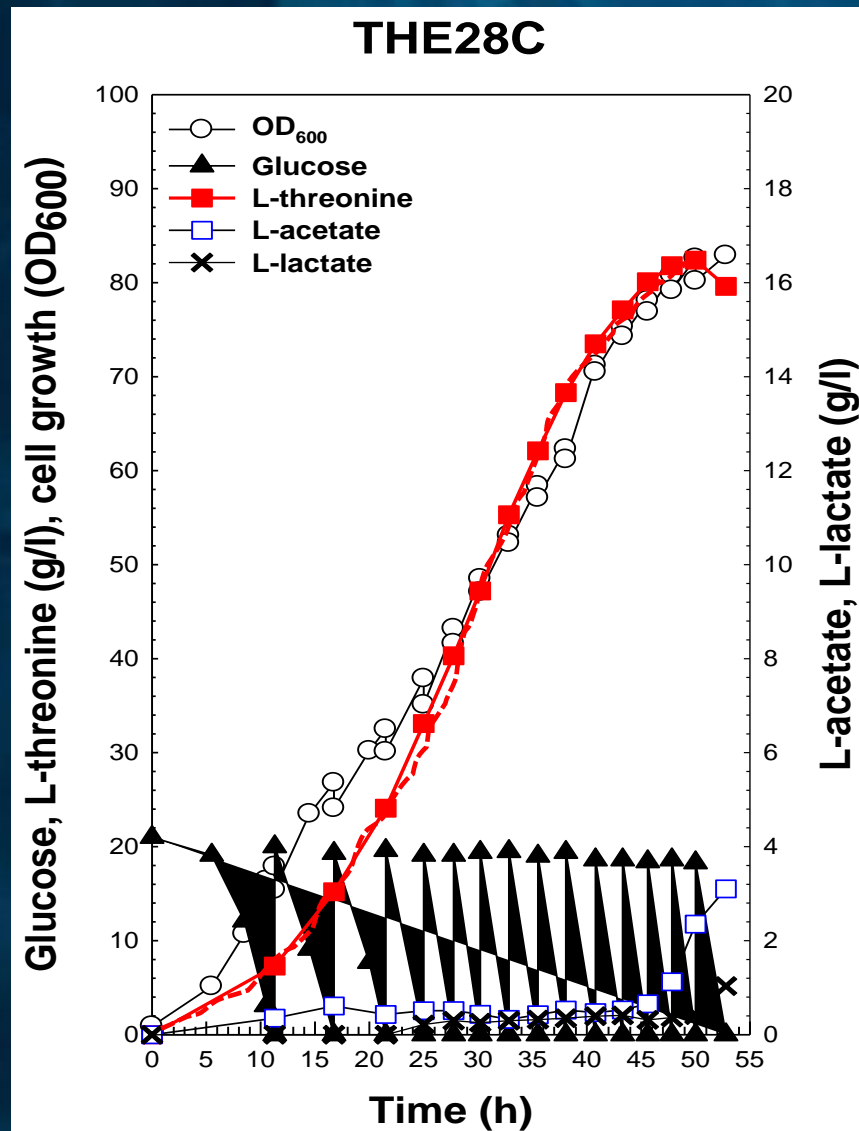


0.202 (g L-threonine/g glucose)

Fed-batch Fermentation of 100% Rationally Engineered *E. coli*



Kwang Ho Lee

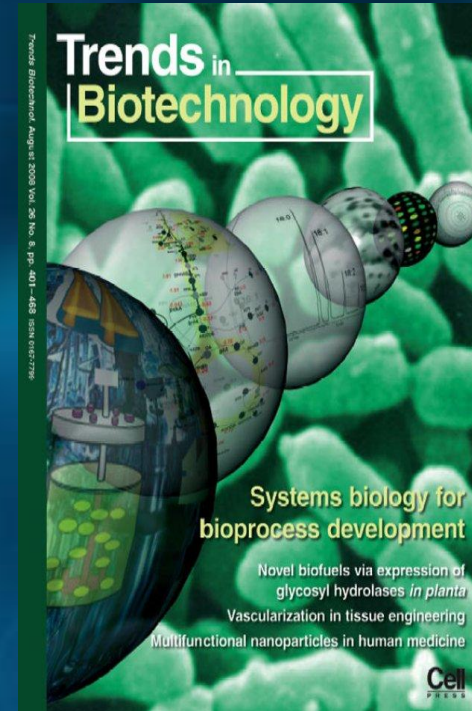


82.4 g/l
L-Thr

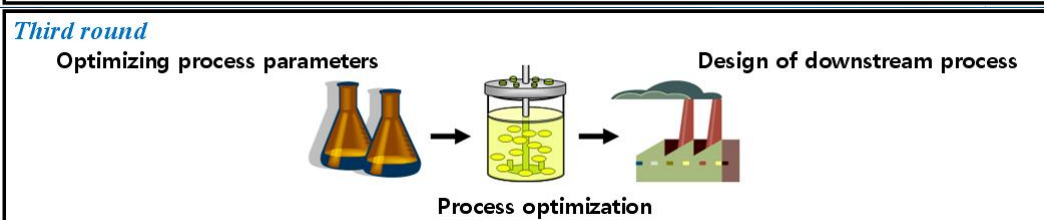
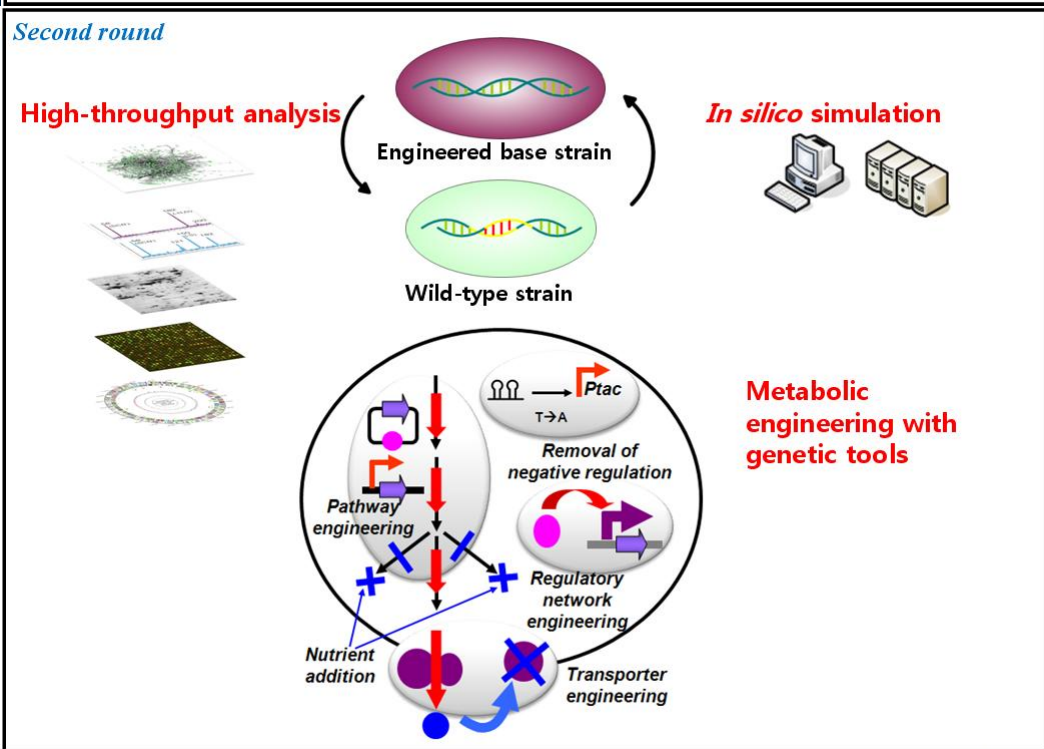
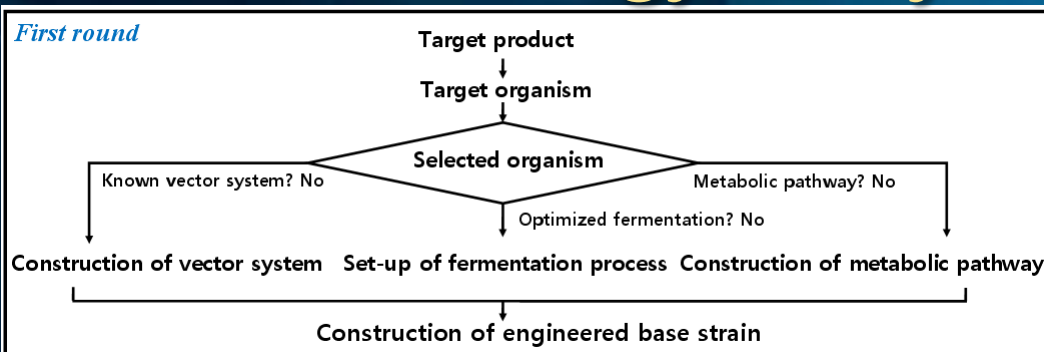
General Strategy for Systems Metabolic Engineering

Moving towards
genome-scale synthetic biology

“systems metabolic engineering”

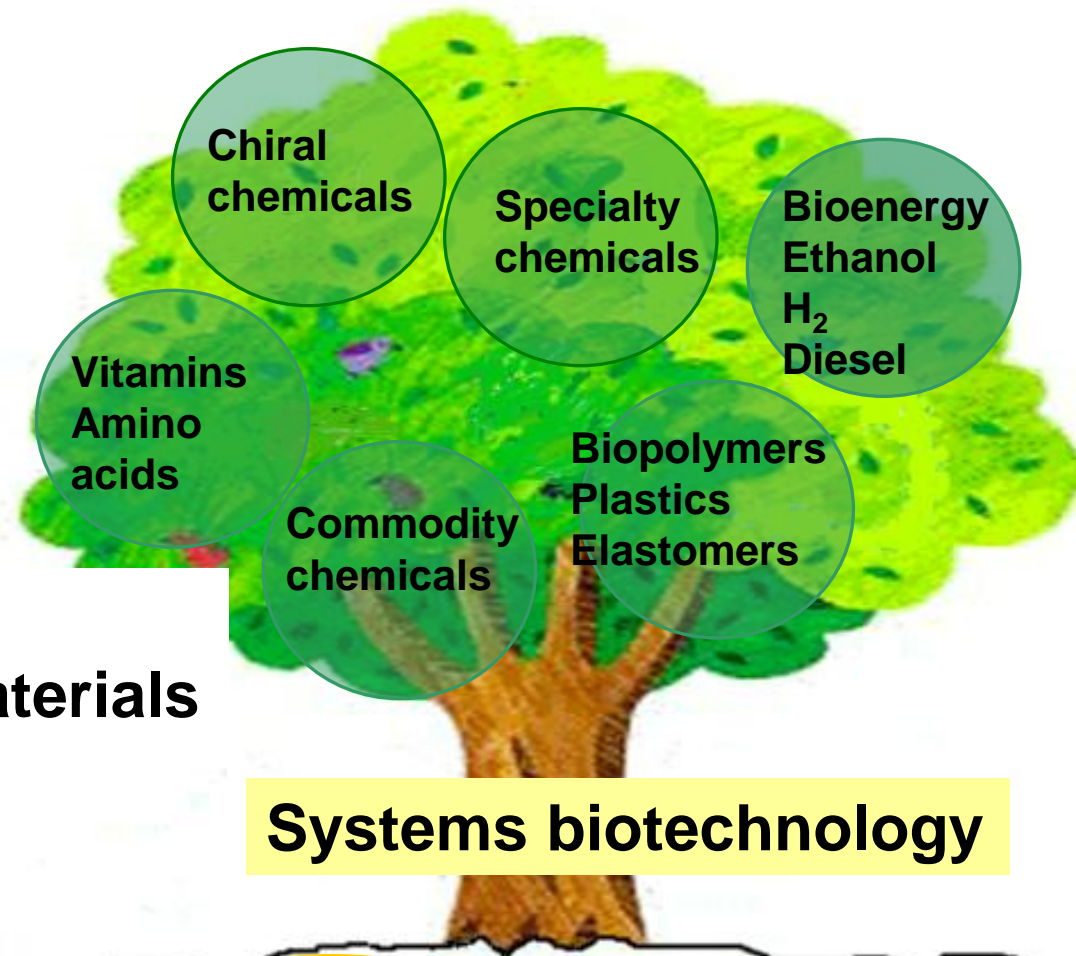
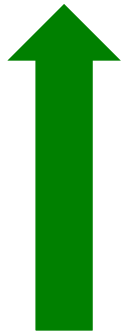


Park et al. 2008. Trends Biotechnol 26: 404-412



White Biotechnology

Green Growth



Chiral
chemicals

Specialty
chemicals

Bioenergy
Ethanol
H₂
Diesel

Vitamins
Amino
acids

Commodity
chemicals

Biopolymers
Plastics
Elastomers

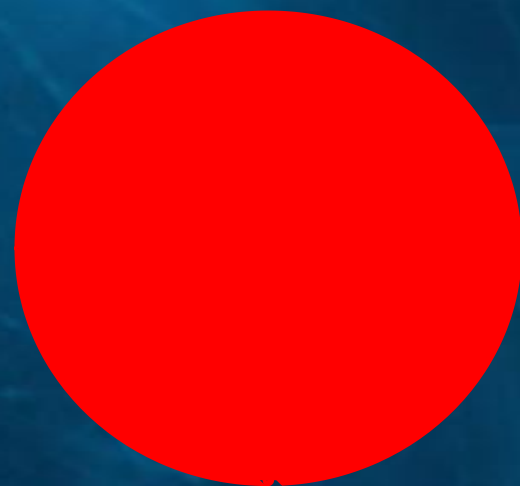
Systems biotechnology

Systems metabolic engineering

Fermentation and
midstream
processing

Bioseparation and
downstream
processing

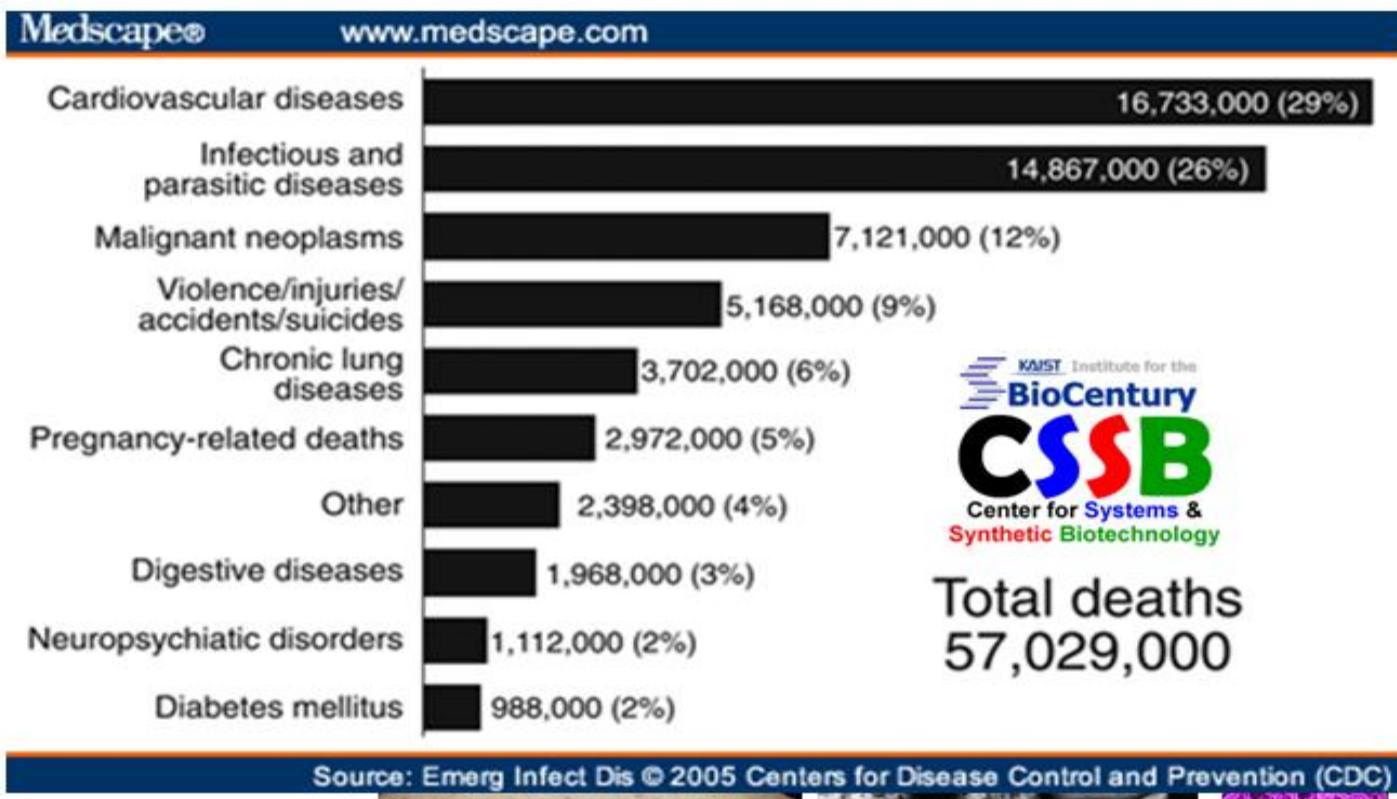
Metabolic engineering/
Cellular engineering



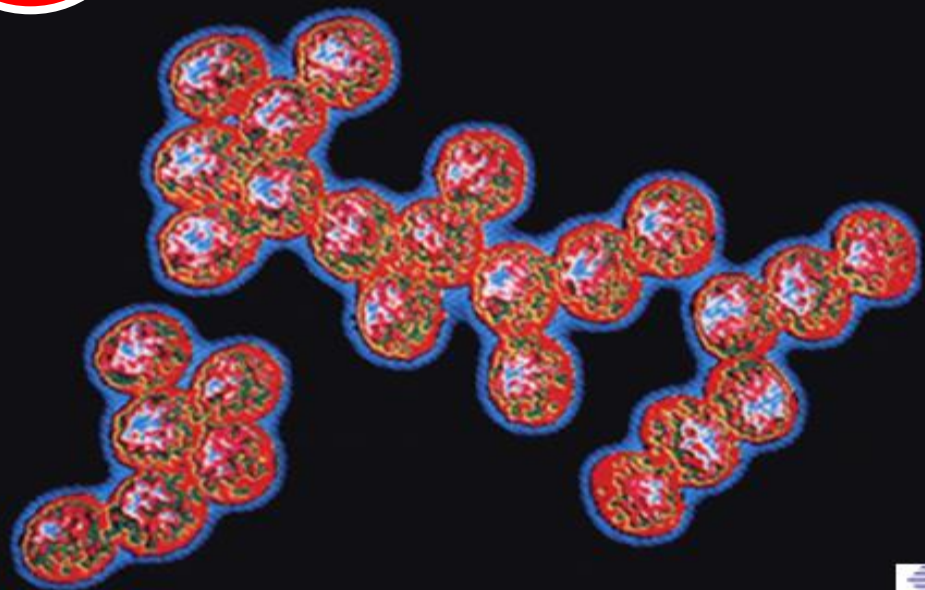
Red Biotechnology

enabled through systems biotechnology

- Incidence of infectious disease: 540 million per year
- 15 million annual deaths worldwide

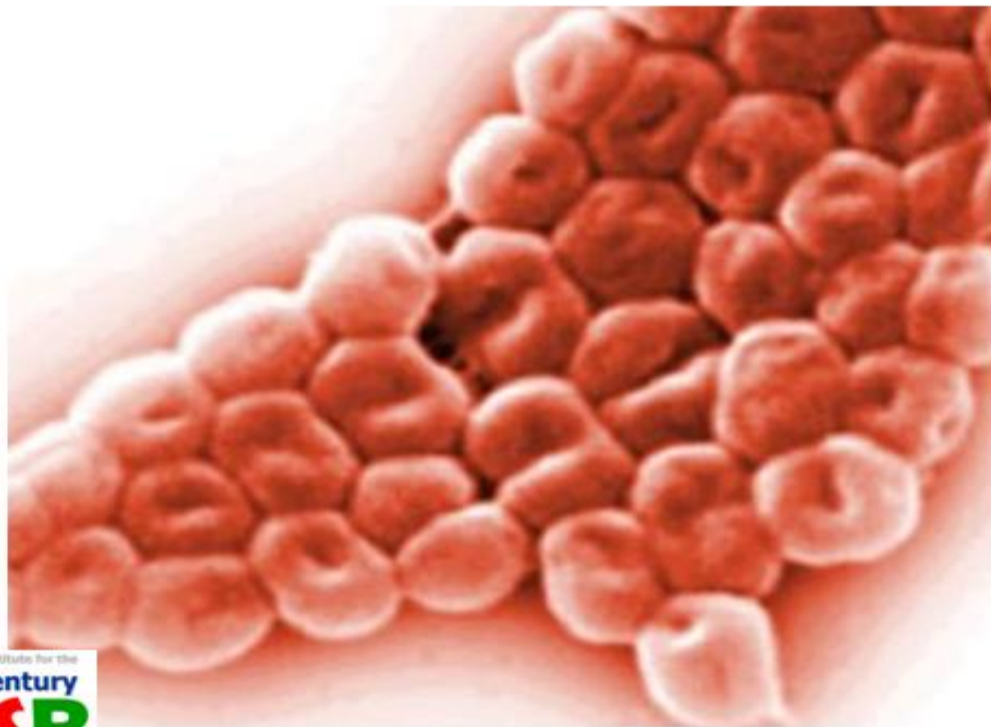


Continuous emergence of super bacteria



Staphylococcus sp.

MRSA, VRSA



Acinetobacter baumannii

Increasing nosocomial pathogen

Increased after the Iraqi war

Cell wall synthesis

- Cycloserin
- Vancomycin
- Bacitracin
- Penicillins-(β -lactams)
- Cephalosporins
- Monobactams
- Carbapenems

Cytoplasmic membrane structure

- Polymyxins

Cell wall

Cytoplasmic membrane

DNA gyrase

- Nalidixic acid
- Norfloxacin
- Novobiocin

DNA-directed RNA polymerase

- Rifampin

Protein synthesis (50s inhibitor)

- Erythromycin
- Chloramphenicol
- Clindamycin
- Lincomycin

Protein synthesis (30s inhibitor)

- Tetracyclines
- Spectinomycin
- Streptomycin
- Gentamicin, tobramycin
- Kanamycin
- Amikacin
- Nitrofurantoin

Protein synthesis (tRNA)

- Mupirocin
- Puromycin

Folic acid metabolism

- Trimethoprim
- Sulfonamide

Dihydrofolate

Ribosomes

DNA

mRNA

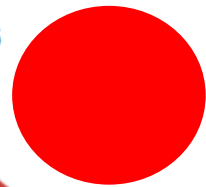
50S

30S

tRNA

Multiple enzymes?

4-Aminobenzoate



Breaking the robustness
as a way of killing pathogens



Organisms can survive even with many of disrupted genes

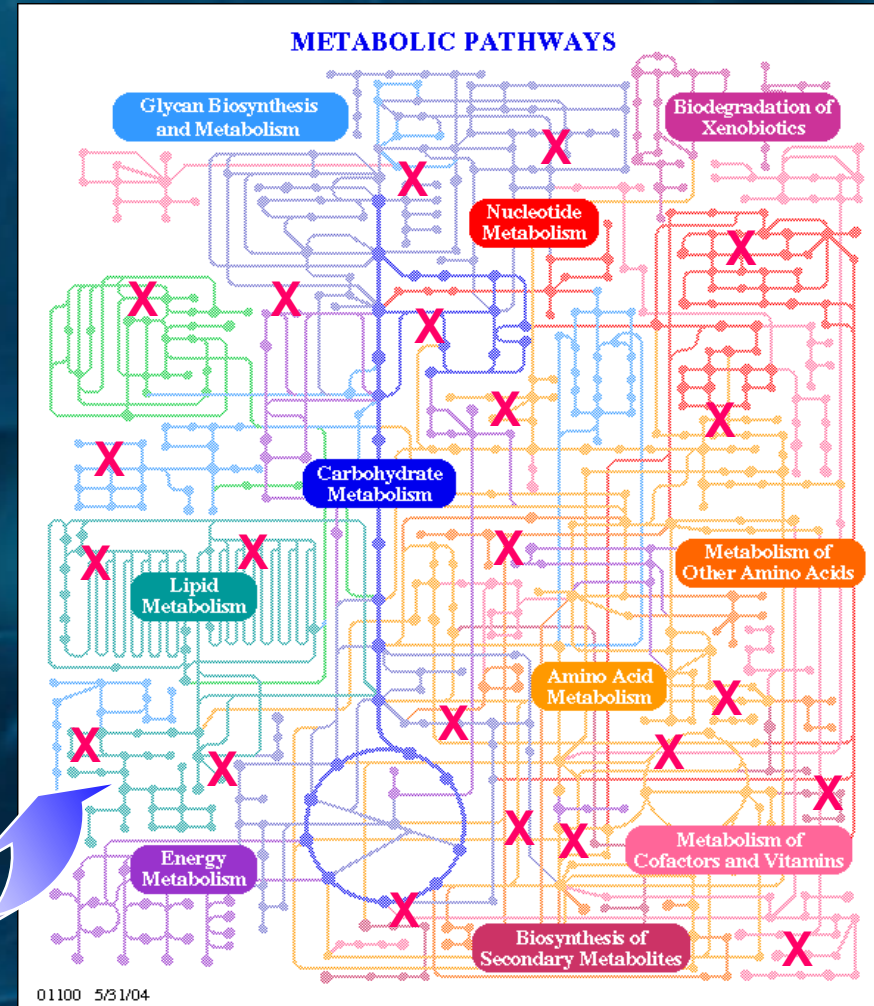
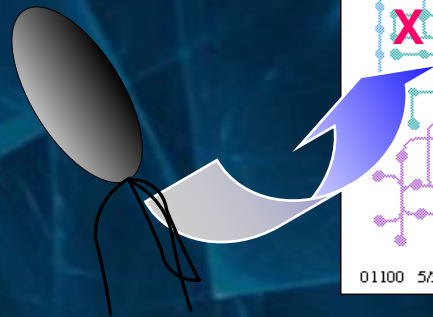
Use of Flux Balance Analysis

Single gene deletion tests

reaction flux (j) = 0

if biomass = 0

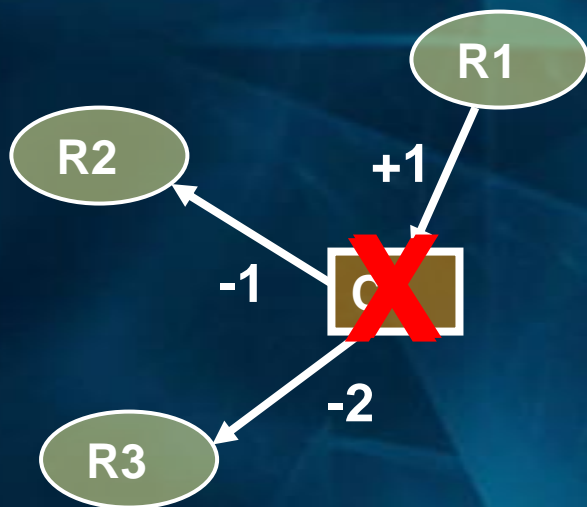
then, *reaction flux* (j) = primary
drug target candidate



Metabolite Essentiality

Kim et al. PNAS 104:13638 (Aug 2007)
In collaboration with Prof. H. Jeong, KAIST

From gene knockout to metabolite knockout ?

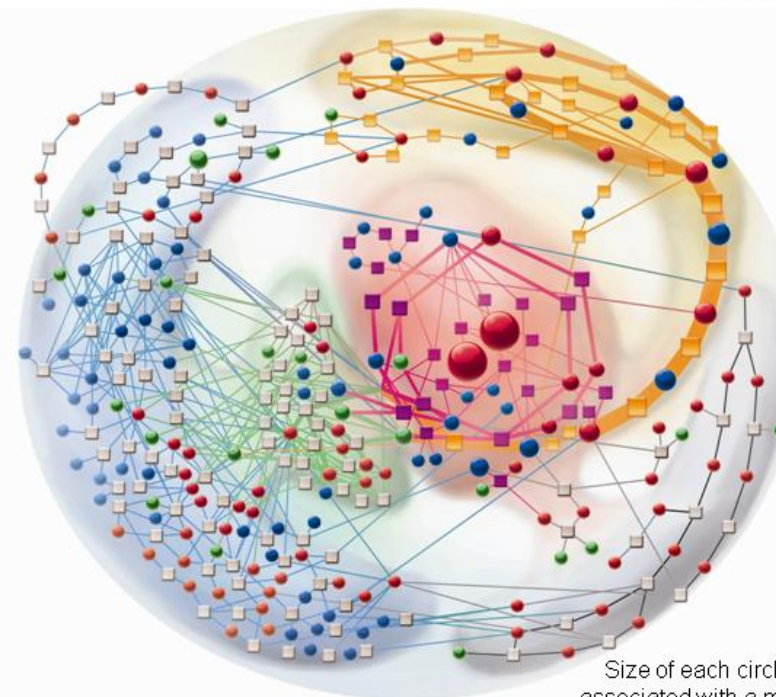


Metabolic network including the essential and non-essential metabolites

Kim et al. PNAS 104:13638 (Aug 2007)

METABOLITE:

- Essential
- Essential and constituting biomass
- Non-essential



REACTION:

- Glycolysis/gluconeogenesis & pentose phosphate pathway
- TCA cycle
- Linear pathway to biomass (peptidoglycan)
- Membrane lipid metabolism
- Cell envelope biosynthesis

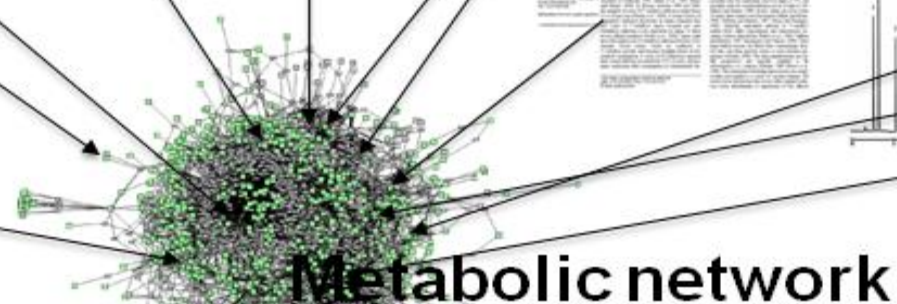
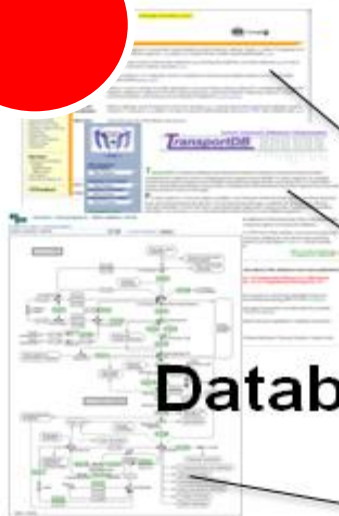
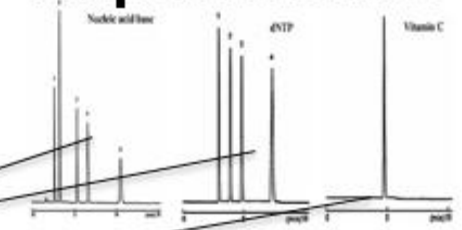
Sang Yup Lee

Genome annotation data

Literature

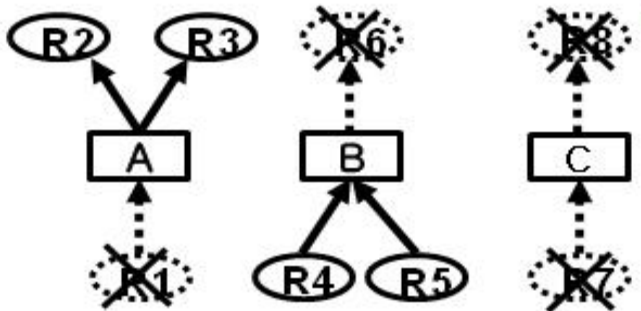
Experiments

Databases

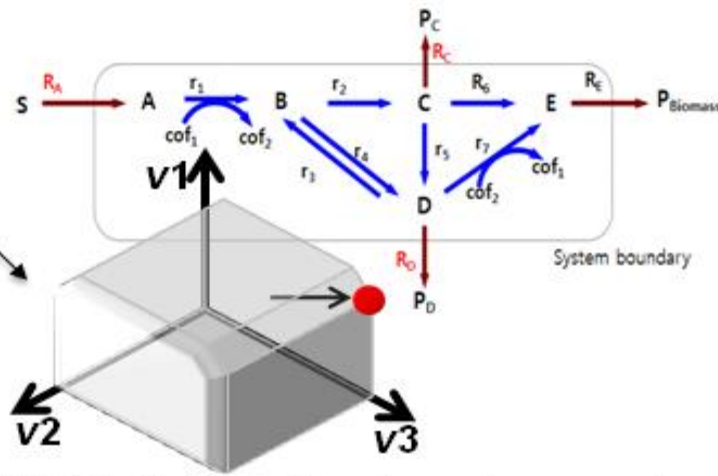


VuMBEL945

945 x 765



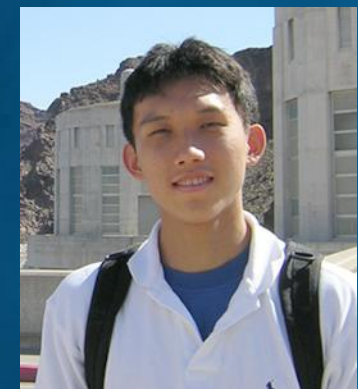
Network topology-based approach
Chokepoint analysis



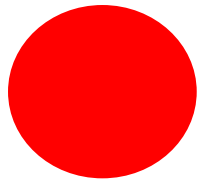
Metabolic function-based approach
Constraints-based flux analysis for single and multiple drug targeting



Tae Yong Kim

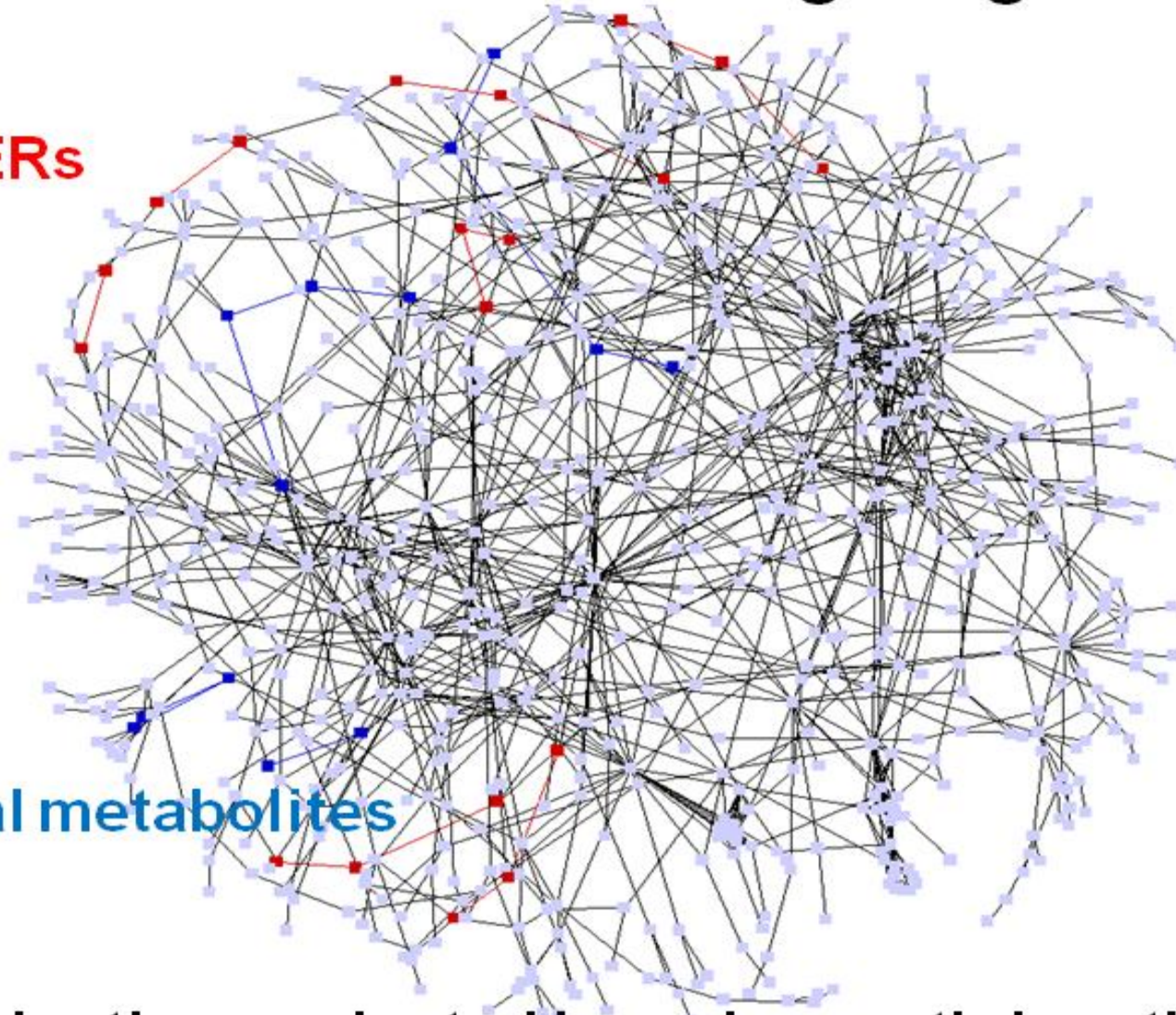


Hyun Uk Kim



Combinatorial drug targets

UERS



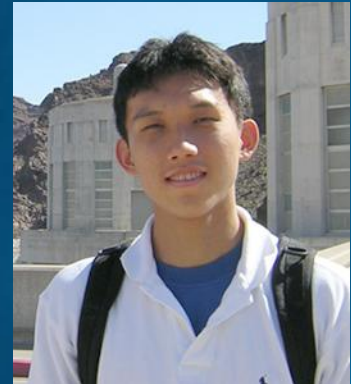
Essential metabolites

20 combinations selected based on path length

Combined analysis of universally essential reactions, choke points, and metabolite essentiality provides potential gene targets !



Tae Yong Kim



Hyun Uk Kim



Conclusions

- Systems biology allows systems-level understanding of cellular and metabolic characteristics
- Systems biotechnology allows development of bioprocesses that are much more efficient than those developed by traditional strategies
- Systems biological analysis of cellular network allows identification of new targets for treatment
- KAIST-Microsoft Collaborative Research Project on SB.NET provides a computational platform for systems-level studies

Acknowledgements

SB.NET team

Dr. Hongseok Yoon

Dr. Tae Yong Kim

Dr. Jin Hwan Park

Mr. Hyun Uk Kim

Mr. Jong Myung Park

Mr. Seung Bum Sohn

Ms. Jung Suk Chu

Mr. Jung Ho Park

Microsoft[®]

Your potential. Our passion.[™]

Ministry of Education, Science & Technology

Korean Systems Biology Research

Genome-based Integrated Bioprocess Development

BK21 Program

WCU Program



Microsoft[®]

Your potential. Our passion.[™]