Microsoft® Research

Faculty Summit

10

YEAR ANNIVERSARY

Microsoft Research

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

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

Systems Biology

Microsoft Research

- 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

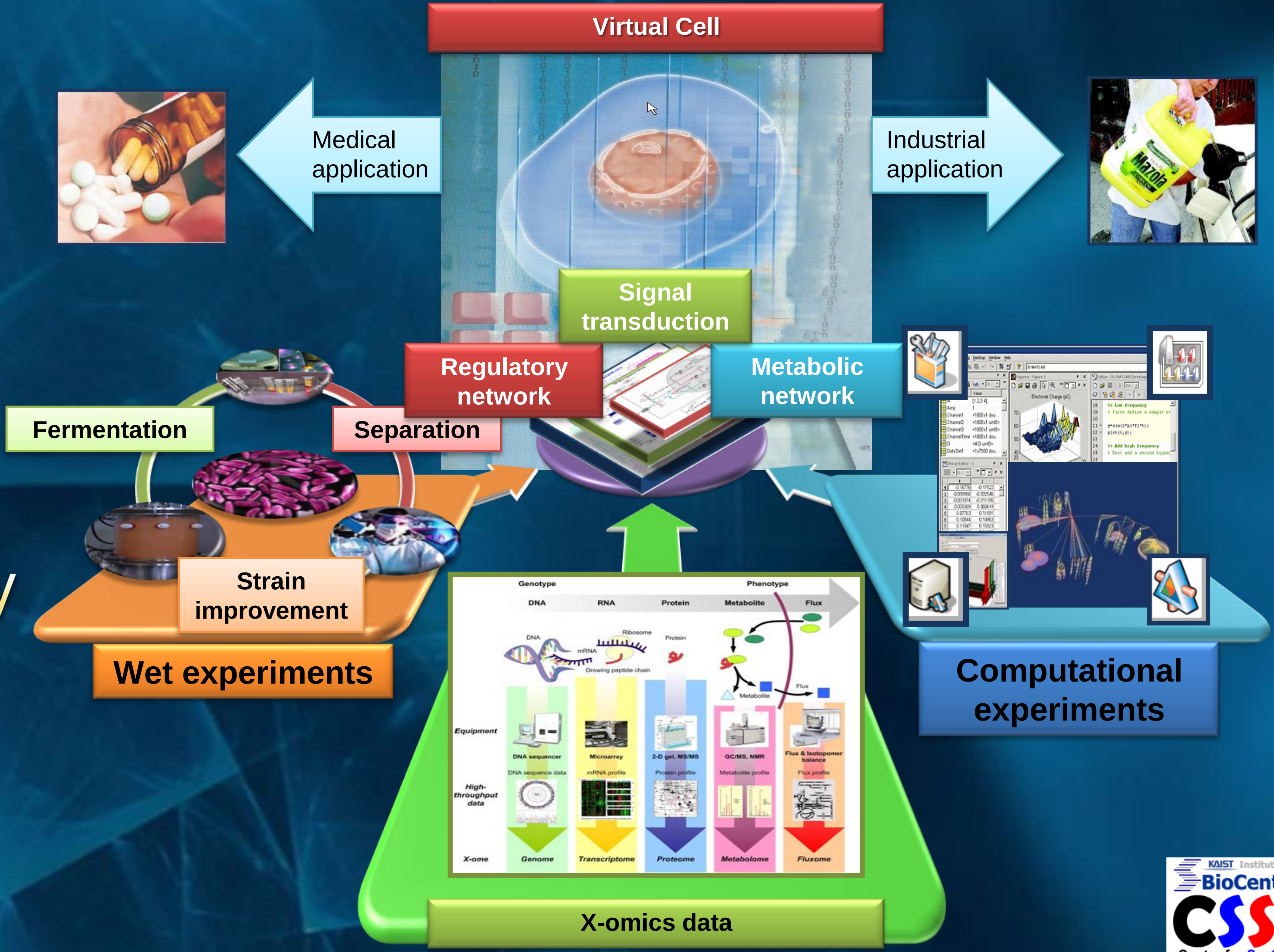
KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

Systems Biotechnology: Making Systems Biology Work

Microsoft Research

Systems Biology

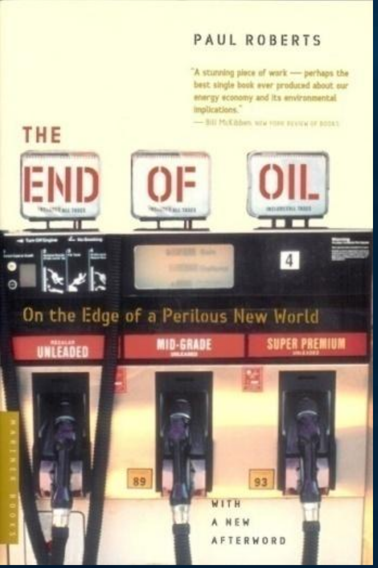
Systems Biotechnology



KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

White Biotechnology = Industrial Biotechnology

enabled through systems and synthetic biotechnology



Global Warming

Kyoto Protocol

Carbon tax

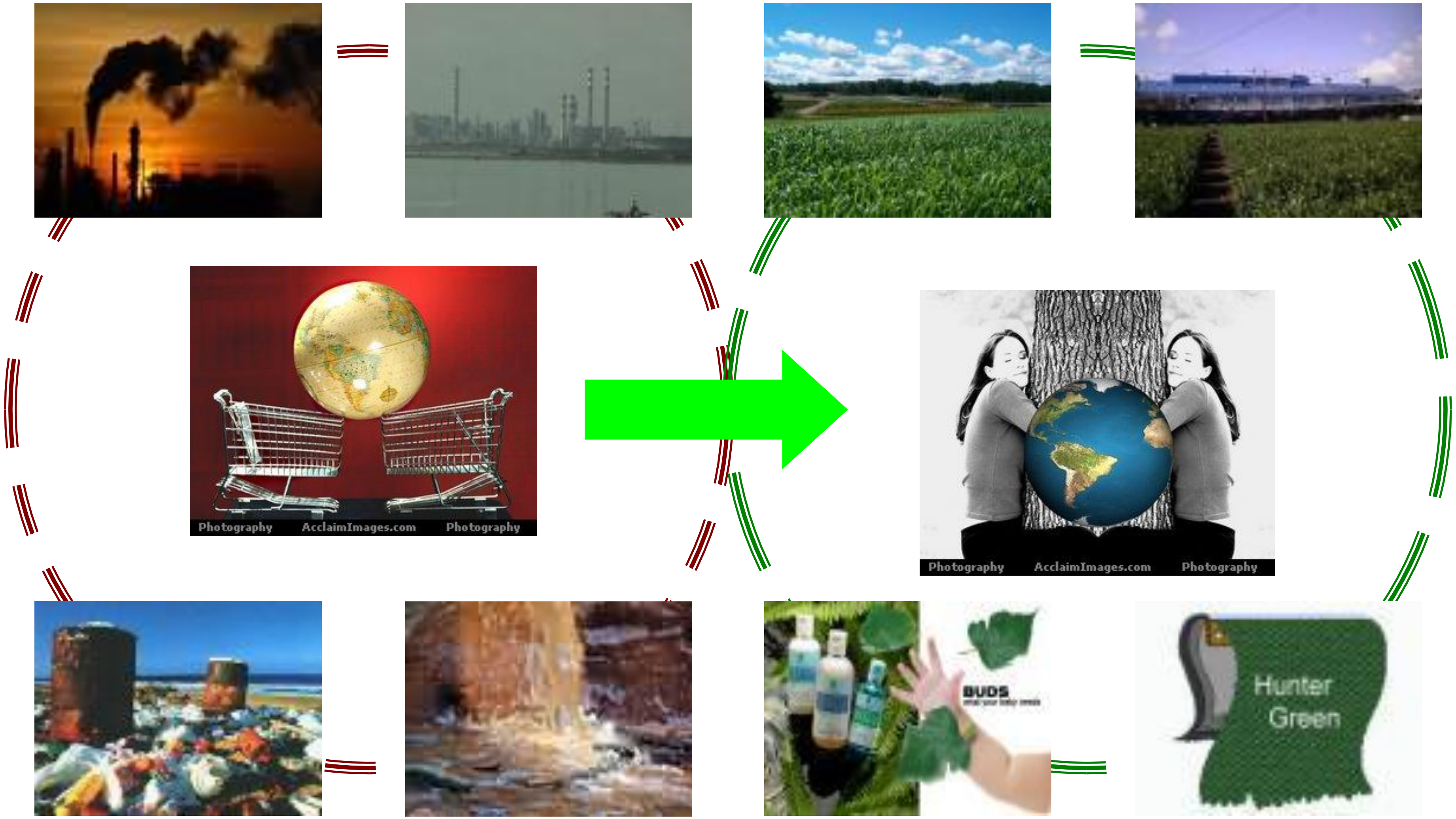
Environmental problems

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



Microsoft Research

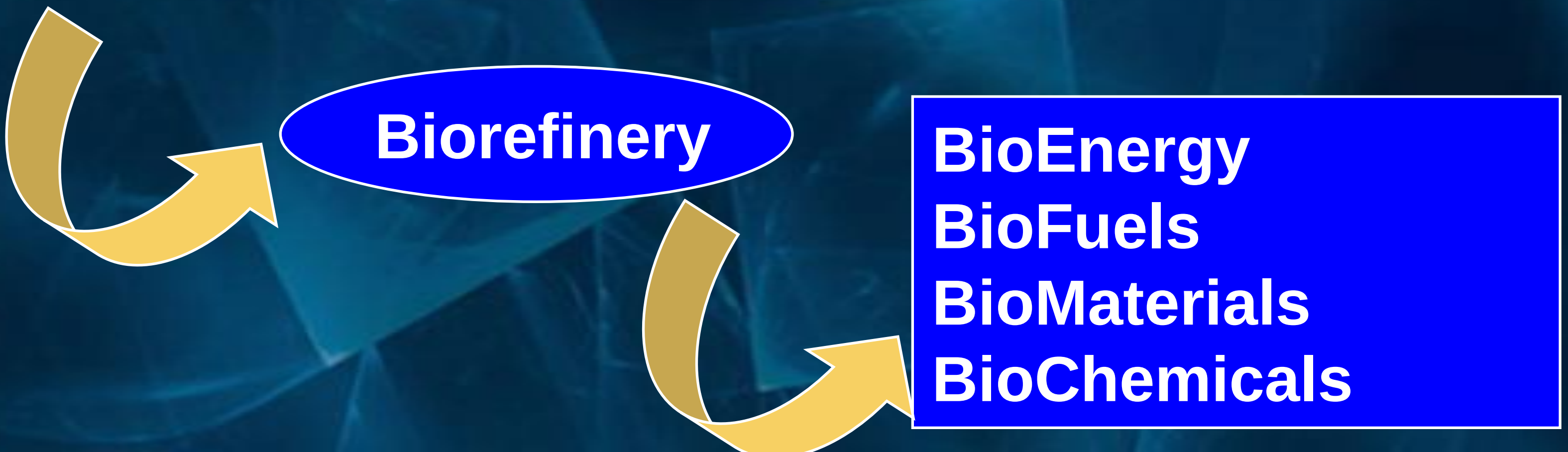
KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

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.)

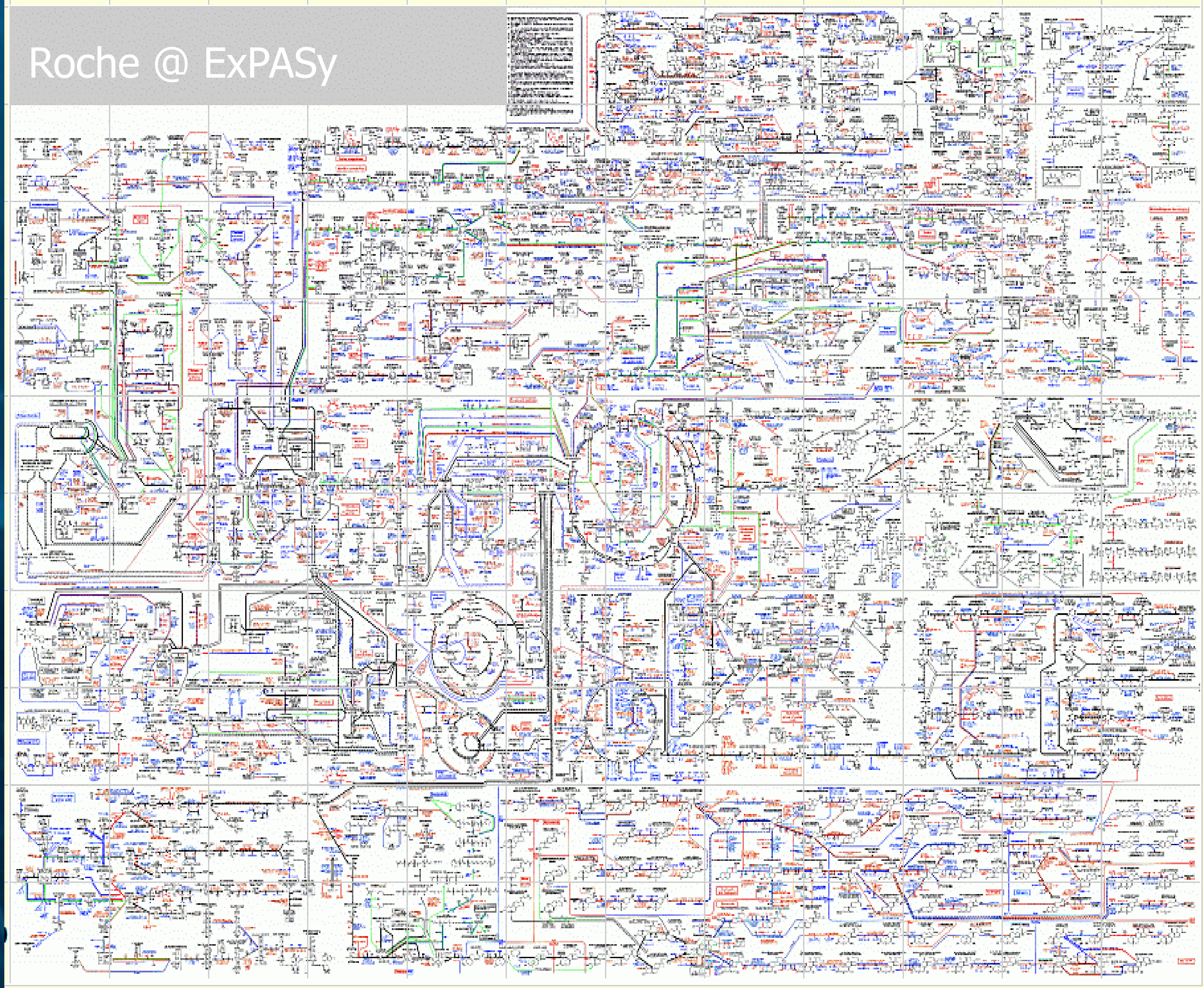


Biorefinery

BioEnergy
BioFuels
BioMaterials
BioChemicals

Chemicals

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



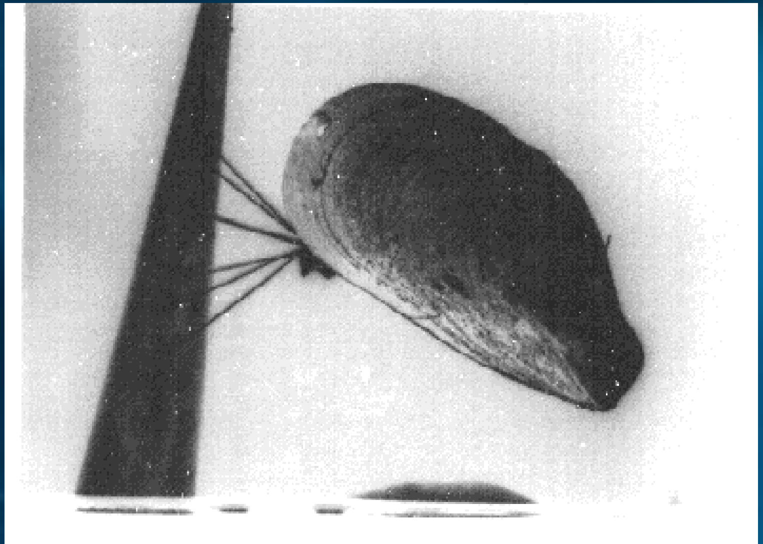
Roche @ ExPASy

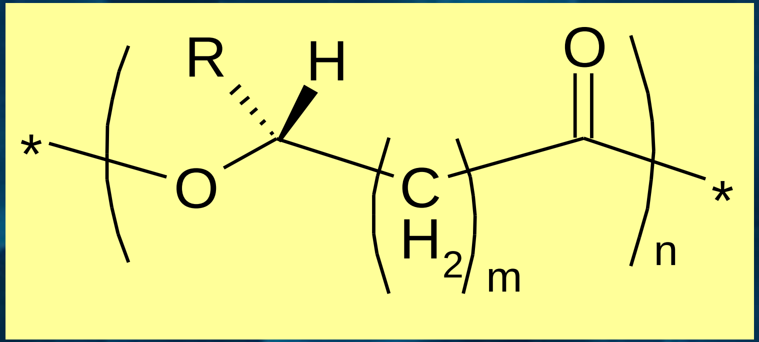
Fuels

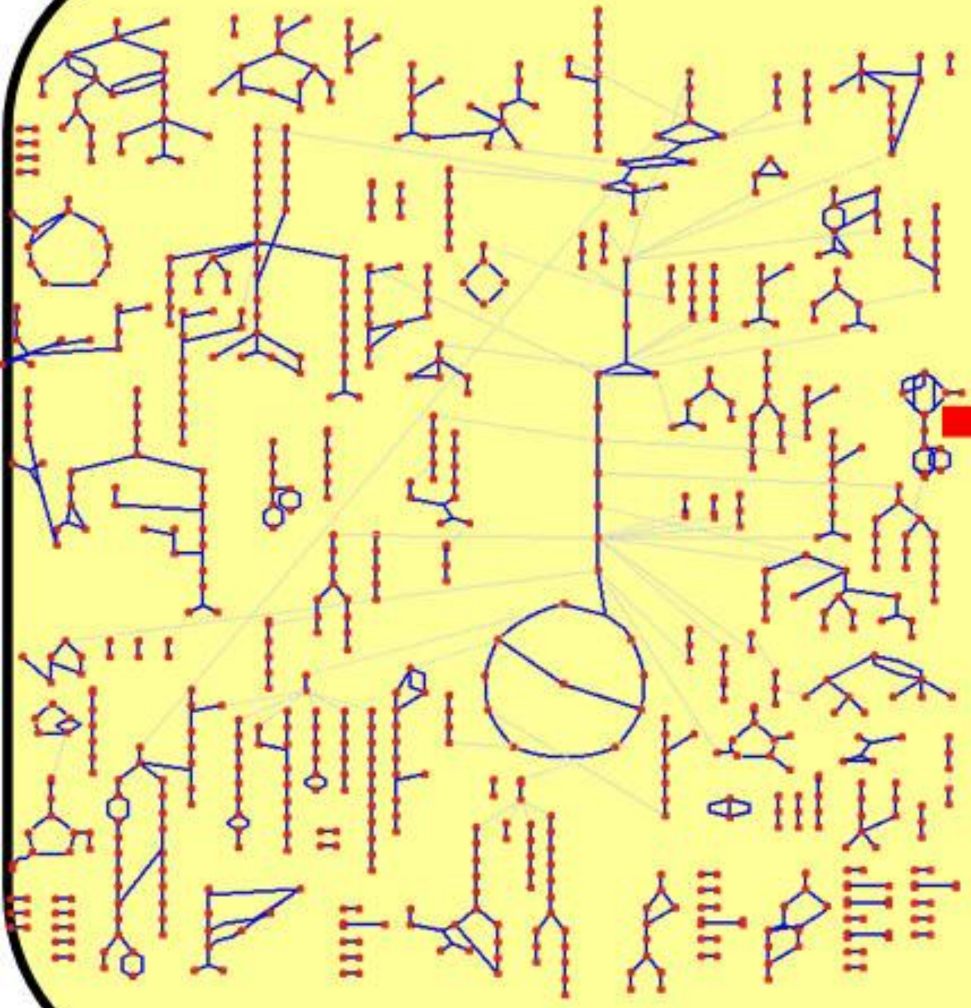
- Ethanol
- Diesel
- Butanol/Isobutanol
- Alkanes
- Isoprenoids
- Hydorxycarboxylic esters

Materials

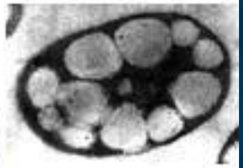
- Polysaccharides
- Poly (amino acids)
- Polyhydroxyalkanoates
- Unnatural polymers
- *In vitro* polymerization of monomers
- Organic/inorganic hybrids

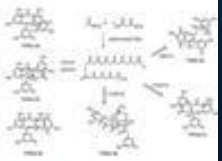






- Primary metabolites
- Secondary metabolites
- Proteins/Peptides
  - Enzymes
- Lipid/Fatty acids
- Carbohydrates
- Whole cell catalyst



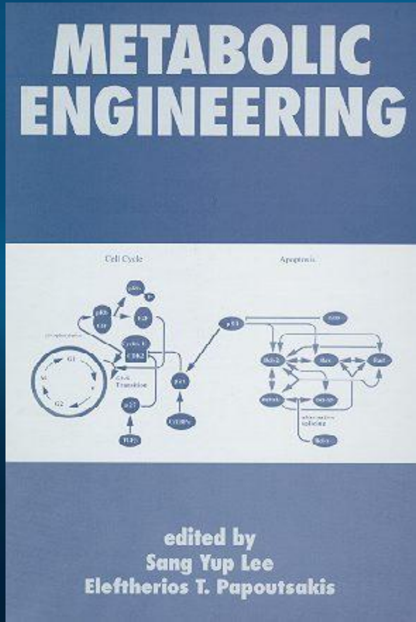


Metabolic Engineering

Microsoft® Research

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

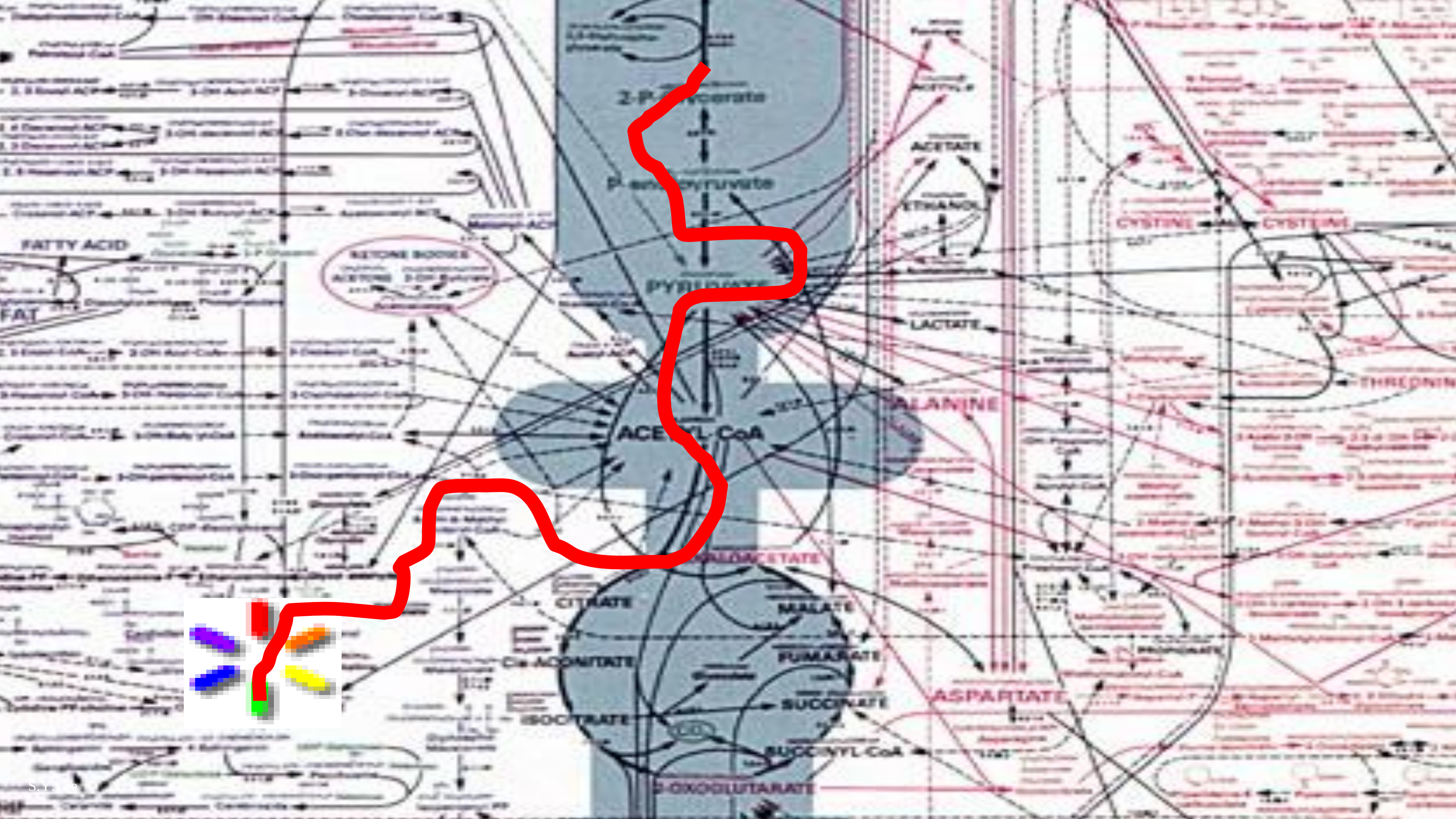
- Enhanced production of metabolites and other biologicals that are already produced by host organism
- Production of modified or new metabolites and other biologicals that are new to the host organism
- Broadening the substrate utilization range
- Designing improved or new metabolic pathways for degradation 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

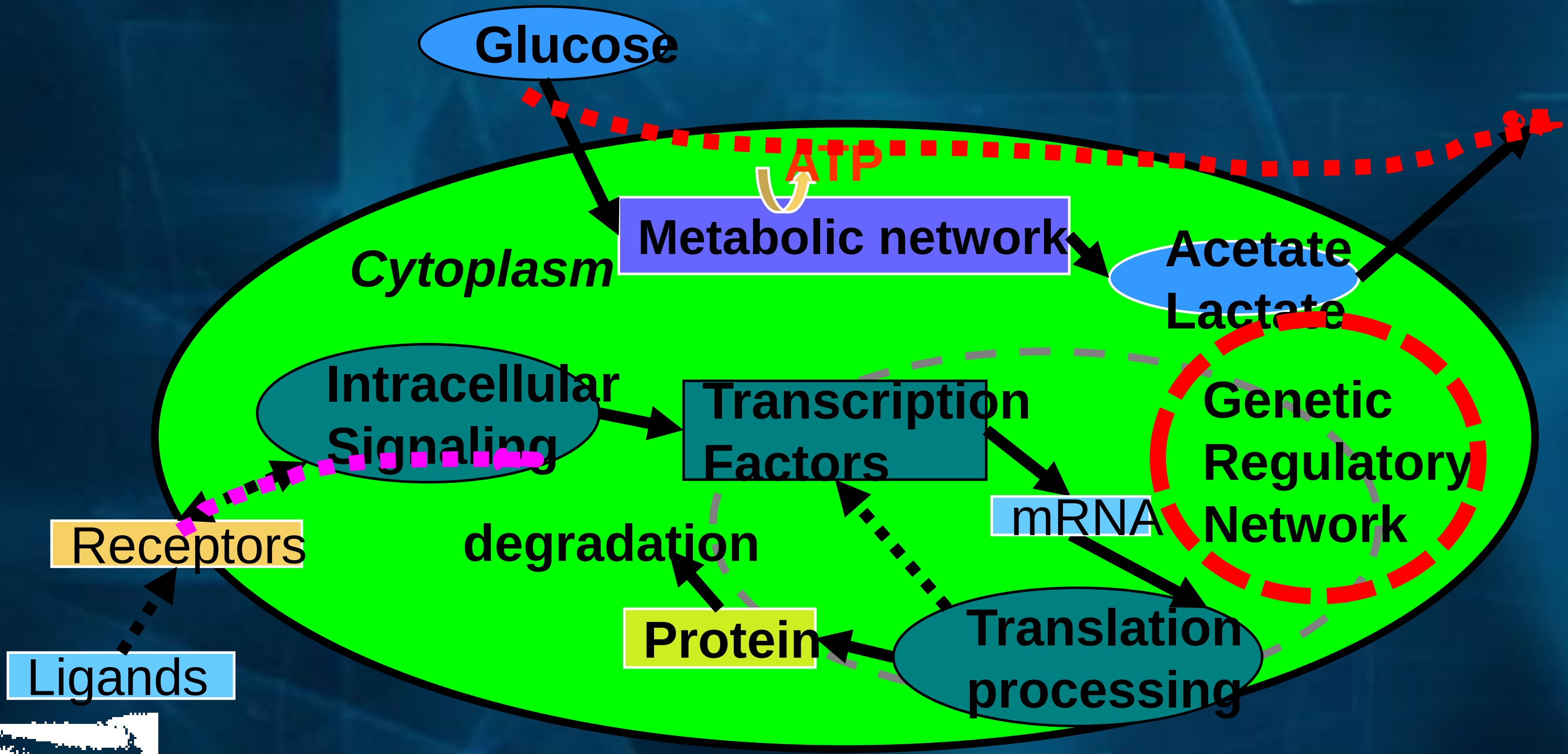
KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology


2-P-Glycerate
P-enolpyruvate
PYRUVATE
ACETYL-CoA
OXALOACETATE
CITRATE
MALATE
cis-ACONITATE
FUMARATE
SUCCINATE
ISOCITRATE
SUCCINYL-CoA
2-OXOGLUTARATE
FATTY ACID
FAT
KETONE BODIES
ACETONE
ACETATE
ETHANOL
LACTATE
ALANINE
ASPARTATE
CYSTINE
CYSTEINE
THREONINE

S.Y.Lee

Complex Network of Network of Network

Microsoft Research



Glucose

ATP

Metabolic network

Cytoplasm

Acetate Lactate

Intracellular Signaling

Transcription Factors

Genetic Regulatory Network

mRNA

Receptors

degradation

Protein

Translation processing

Ligands

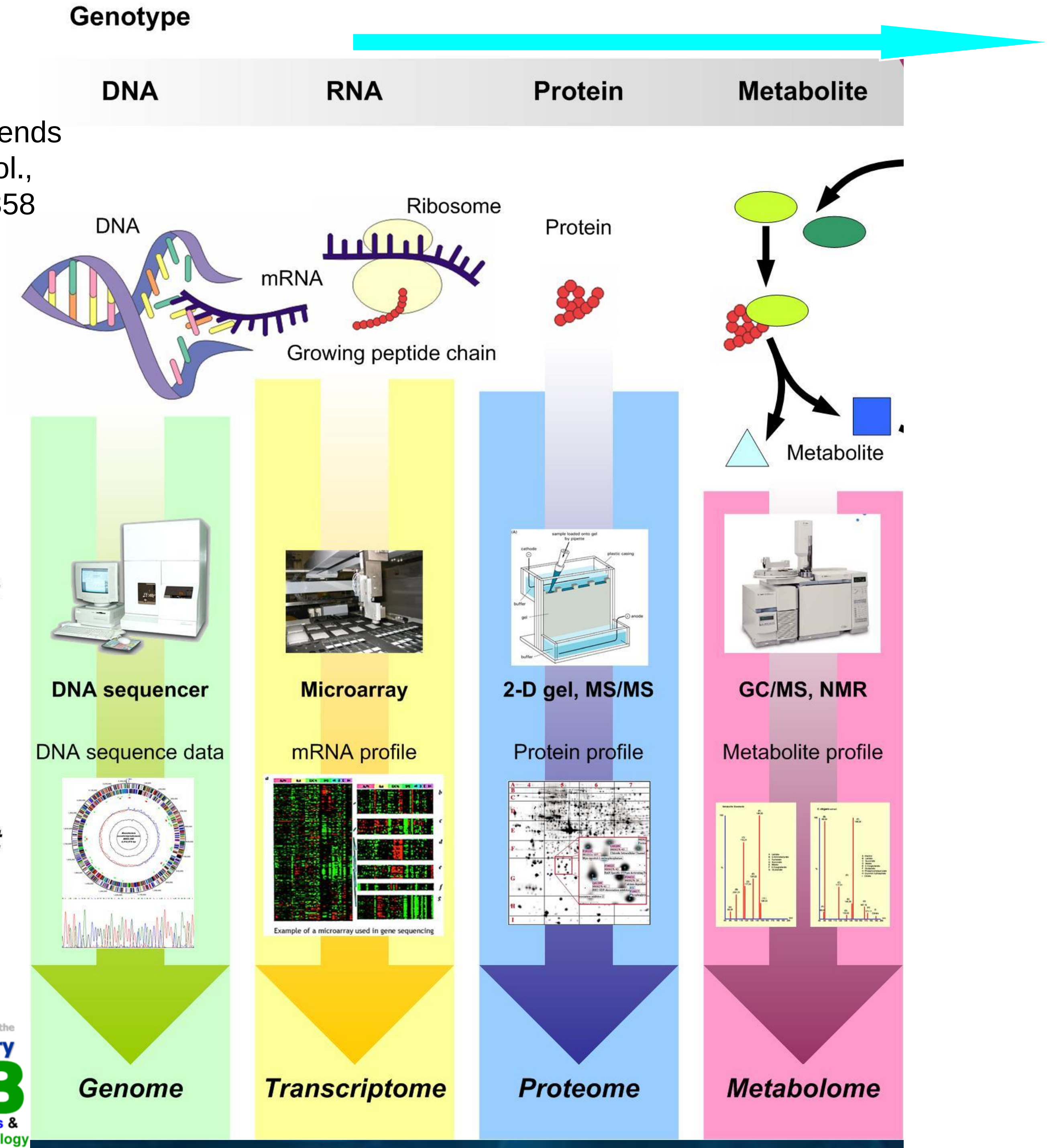
Network^n

KAIST Institute for the BioCentury

CSSB

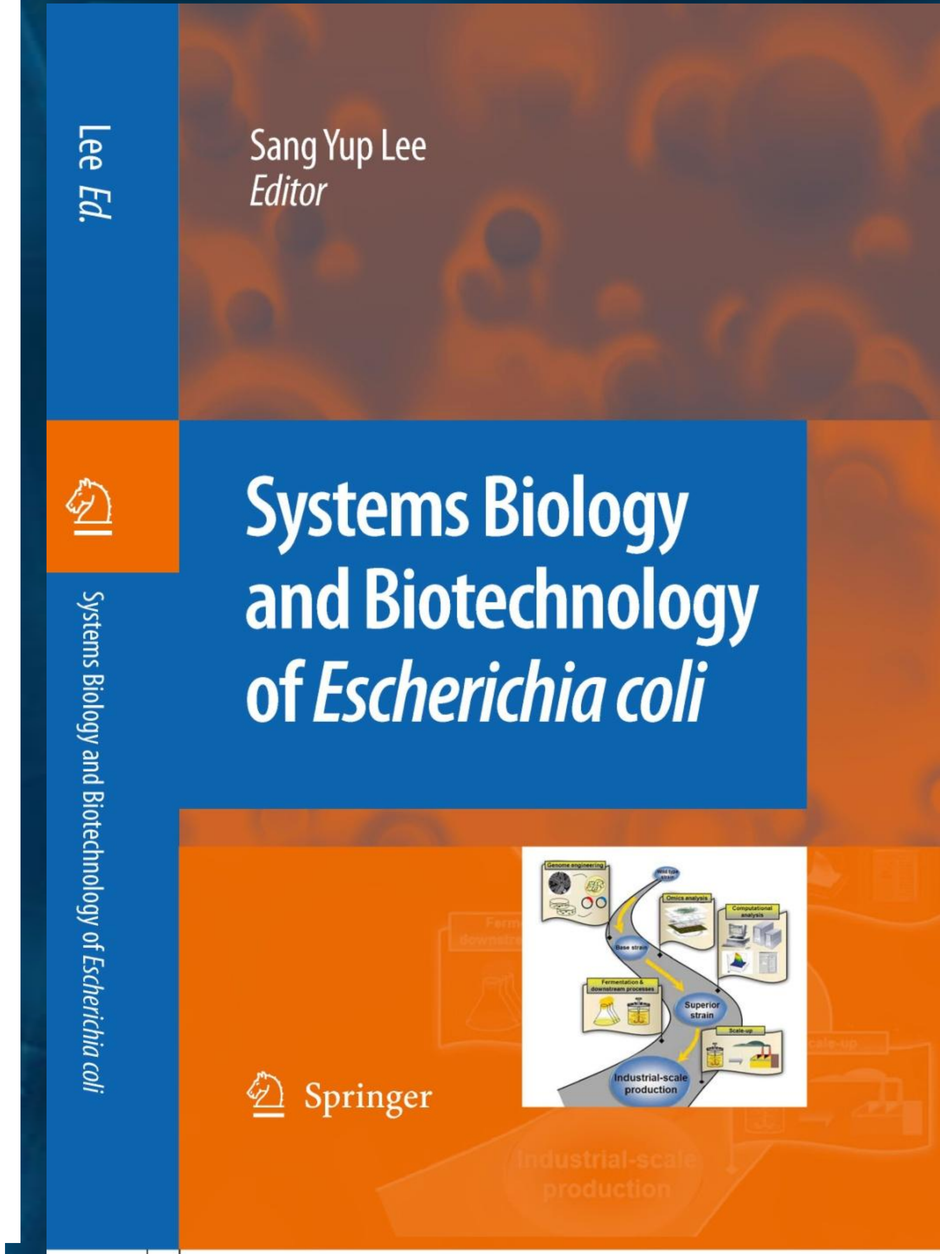
Center for Systems & Synthetic Biotechnology

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



KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

Microsoft Research


Lee Ed.

Sang Yup Lee
Editor

Systems Biology and Biotechnology of *Escherichia coli*

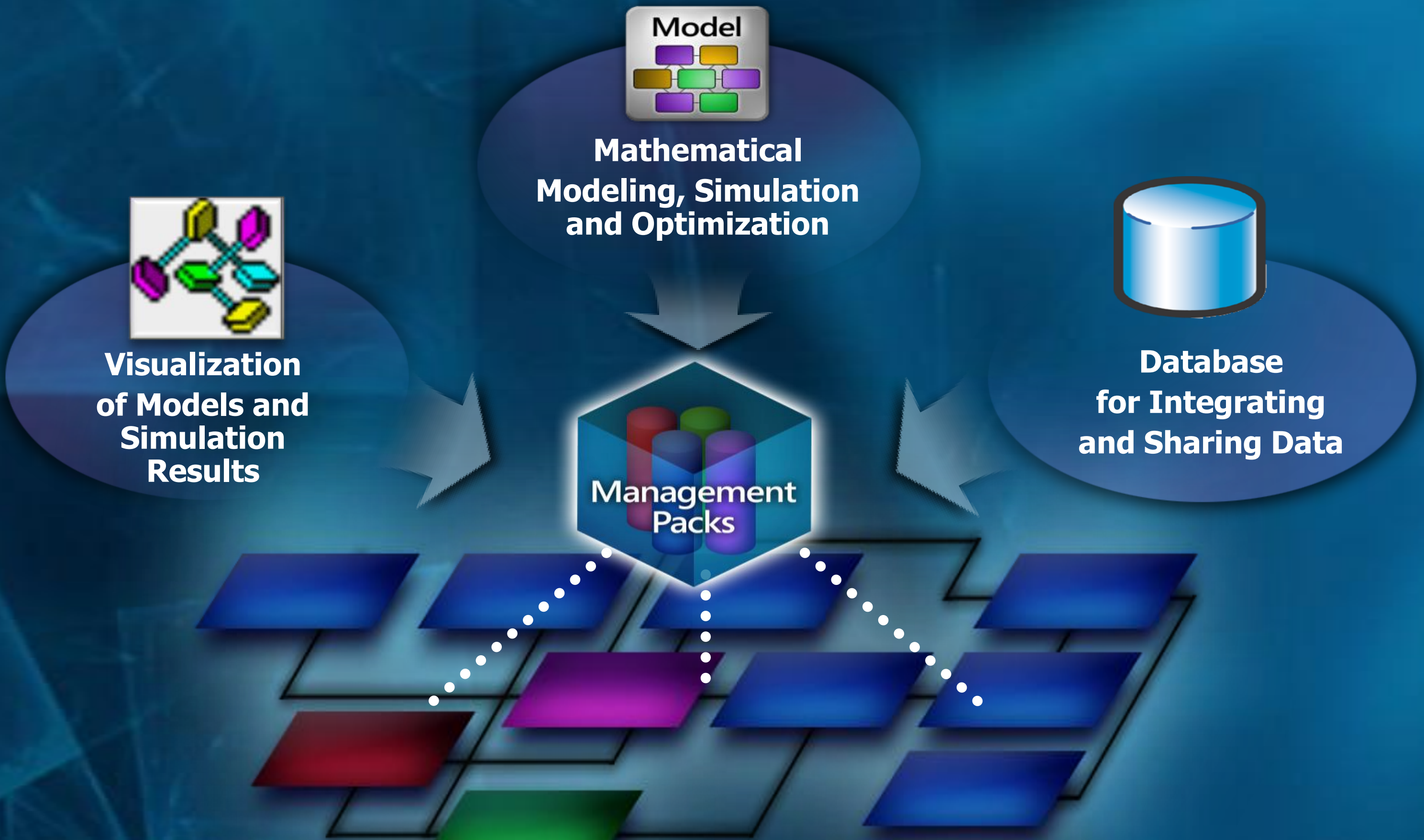
Springer

Systems Biology and Biotechnology of *Escherichia coli*

Microsoft Research

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

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



Model

Mathematical Modeling, Simulation and Optimization

Visualization of Models and Simulation Results

Database for Integrating and Sharing Data

Management Packs

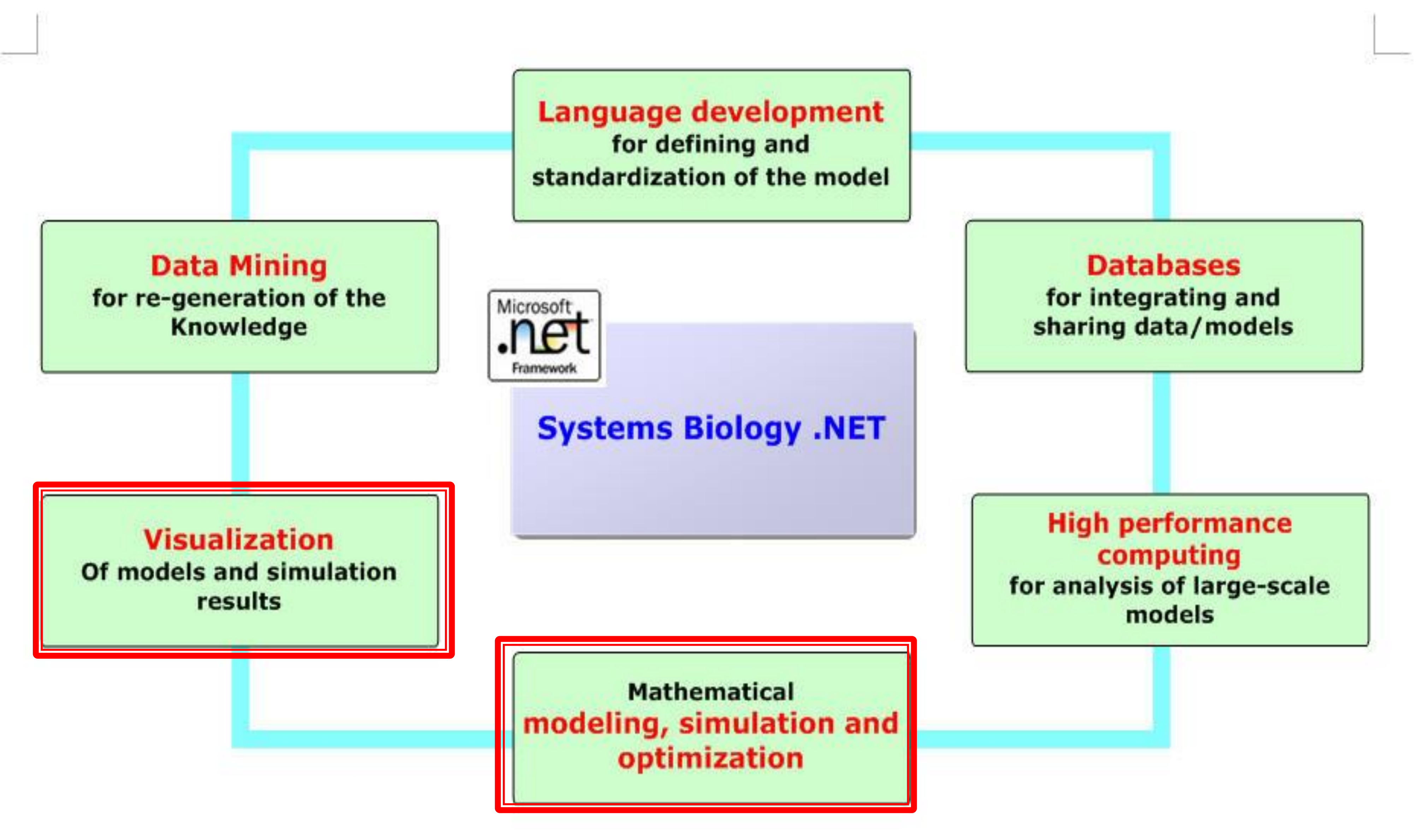
Dr. H. Yoon

Ms. J. Chu

Microsoft
Your potential. Our passion.™

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

Systems Biology.Net



Language development for defining and standardization of the model

Databases for integrating and sharing data/models

High performance computing for analysis of large-scale models

Mathematical **modeling, simulation and optimization**

Visualization Of models and simulation results

Data Mining for re-generation of the Knowledge

Systems Biology .NET

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

Microsoft Research

- 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

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

SB.NET Project Overview

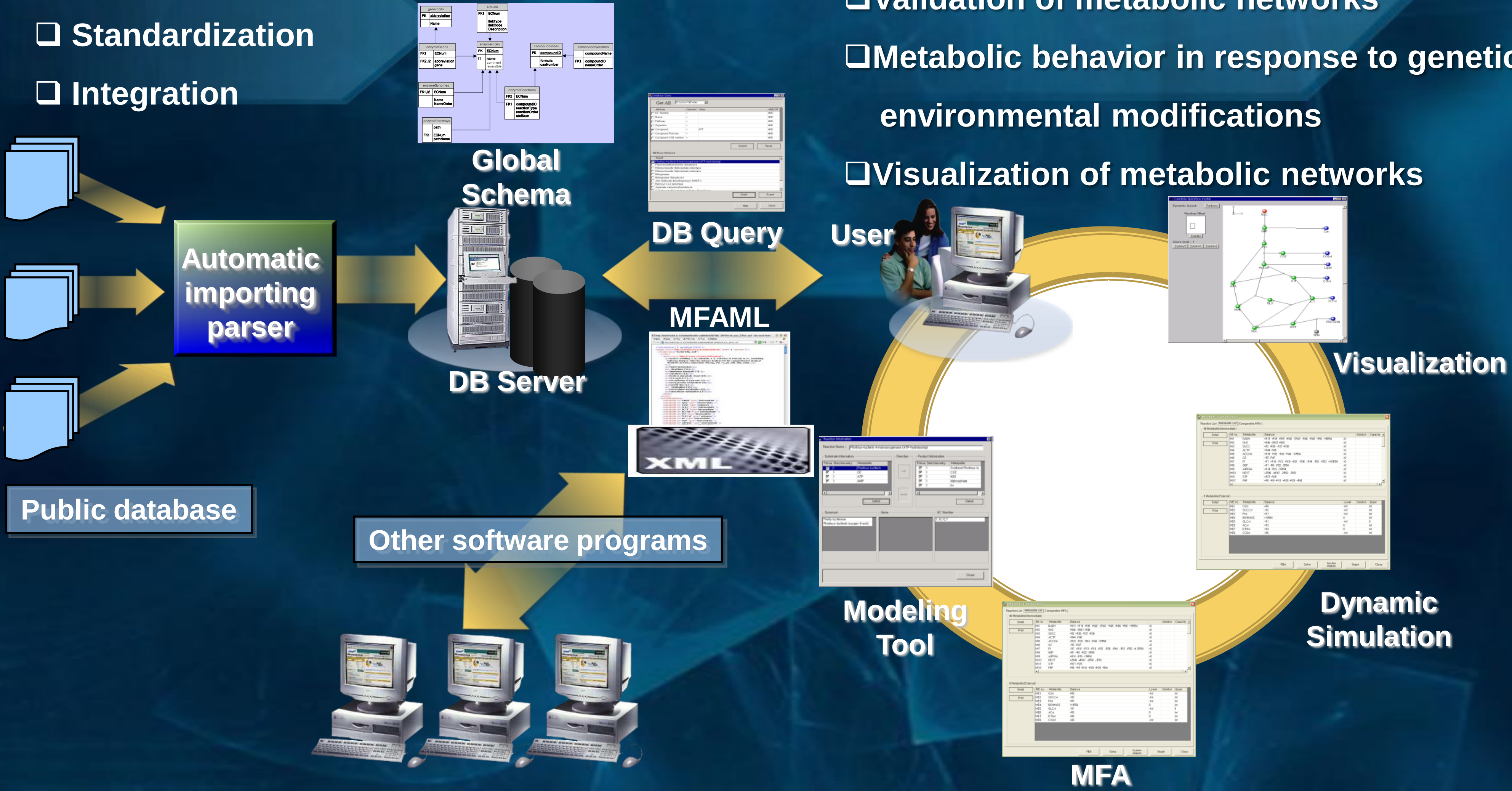
Microsoft® Research

Integration of biochemical data

- ☐ Thorough comparison of different DBs
- ☐ Standardization
- ☐ Integration

Analysis of metabolic networks

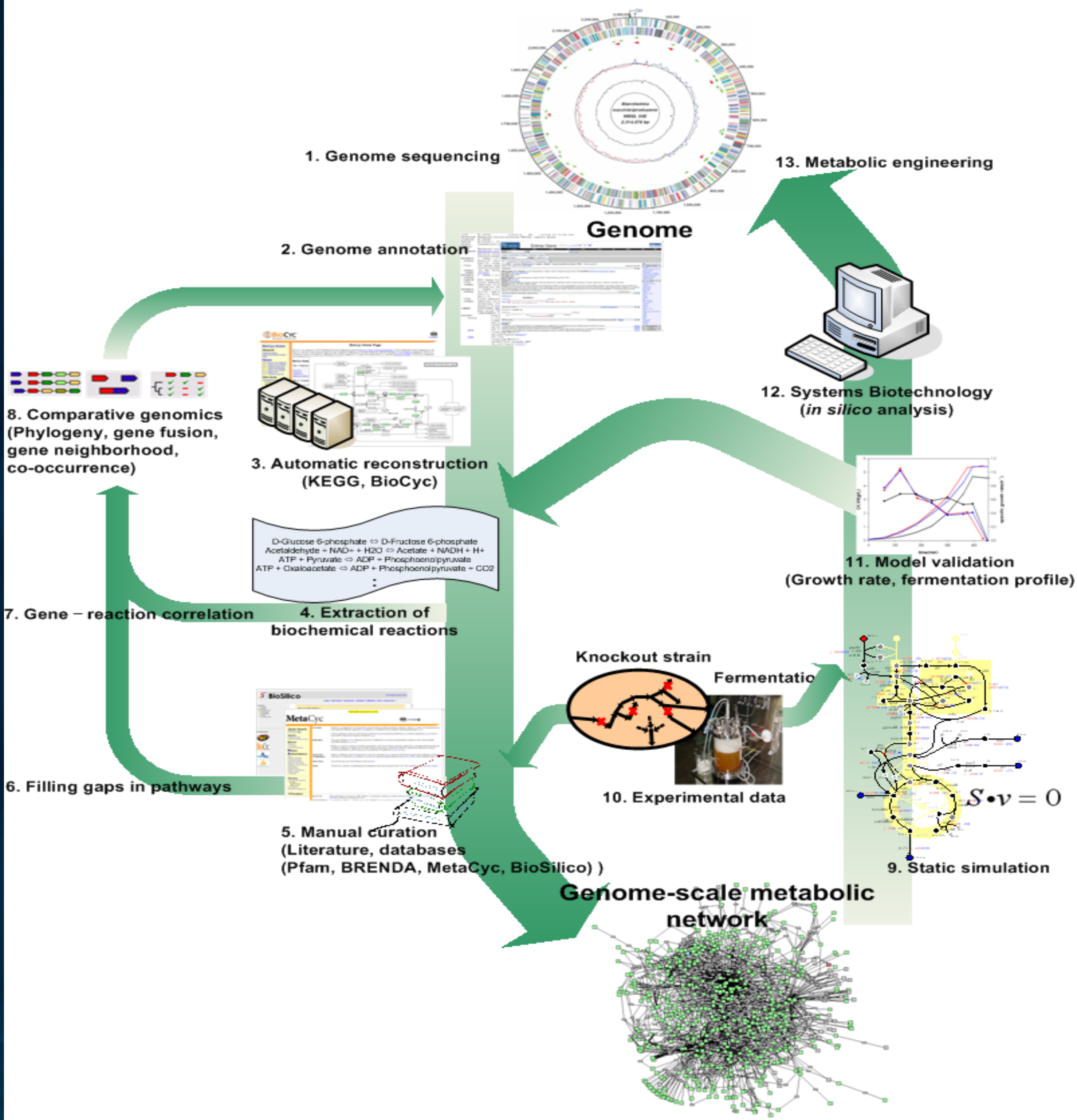
- ☐Construction of reaction models
- ☐Estimation of flux distributions
- ☐Validation of metabolic networks
- ☐Metabolic behavior in response to genetic and/or environmental modifications
- ☐Visualization of metabolic networks



Microsoft®
Your potential. Our passion.™

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

Genome-Scale Metabolic Network Models



Biochemical Reactions

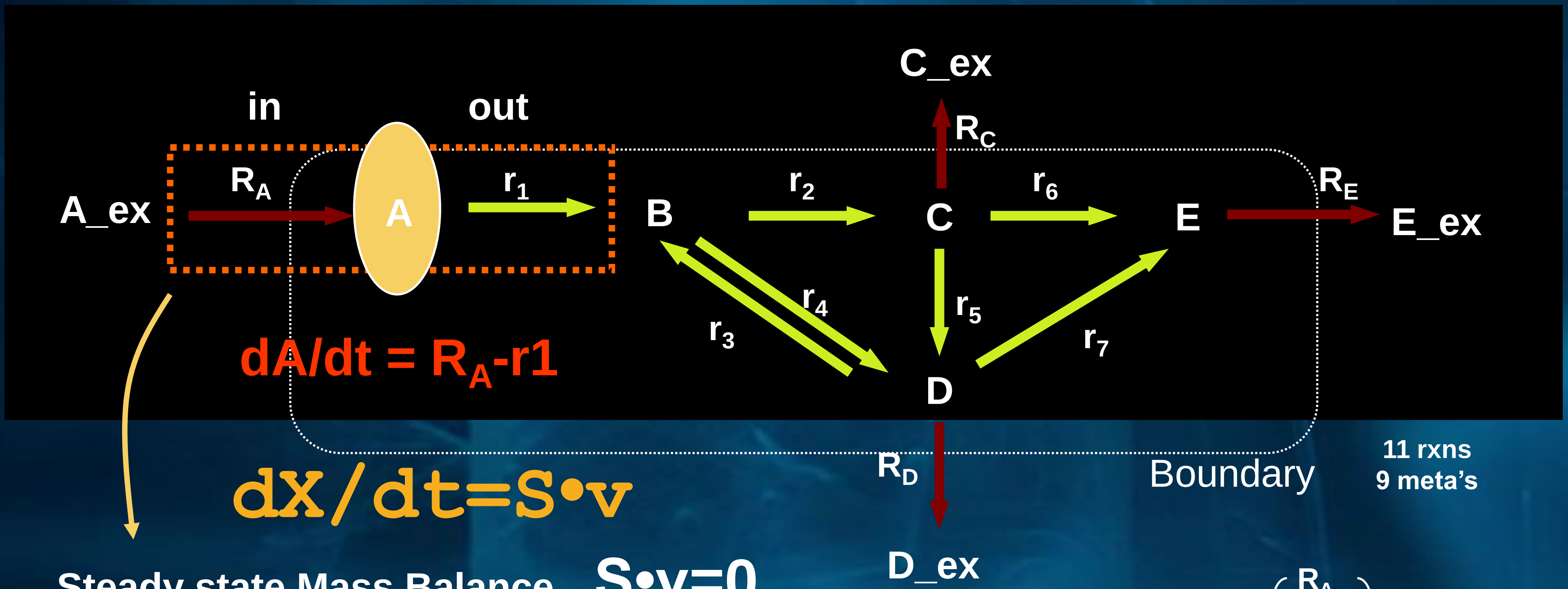
Stoichiometry

Cofactors

Reversibility

Biomass

ATP requirement



$dA/dt = R_A - r1$

$dx/dt = S \bullet v$

Steady state Mass Balance $S \bullet v = 0$

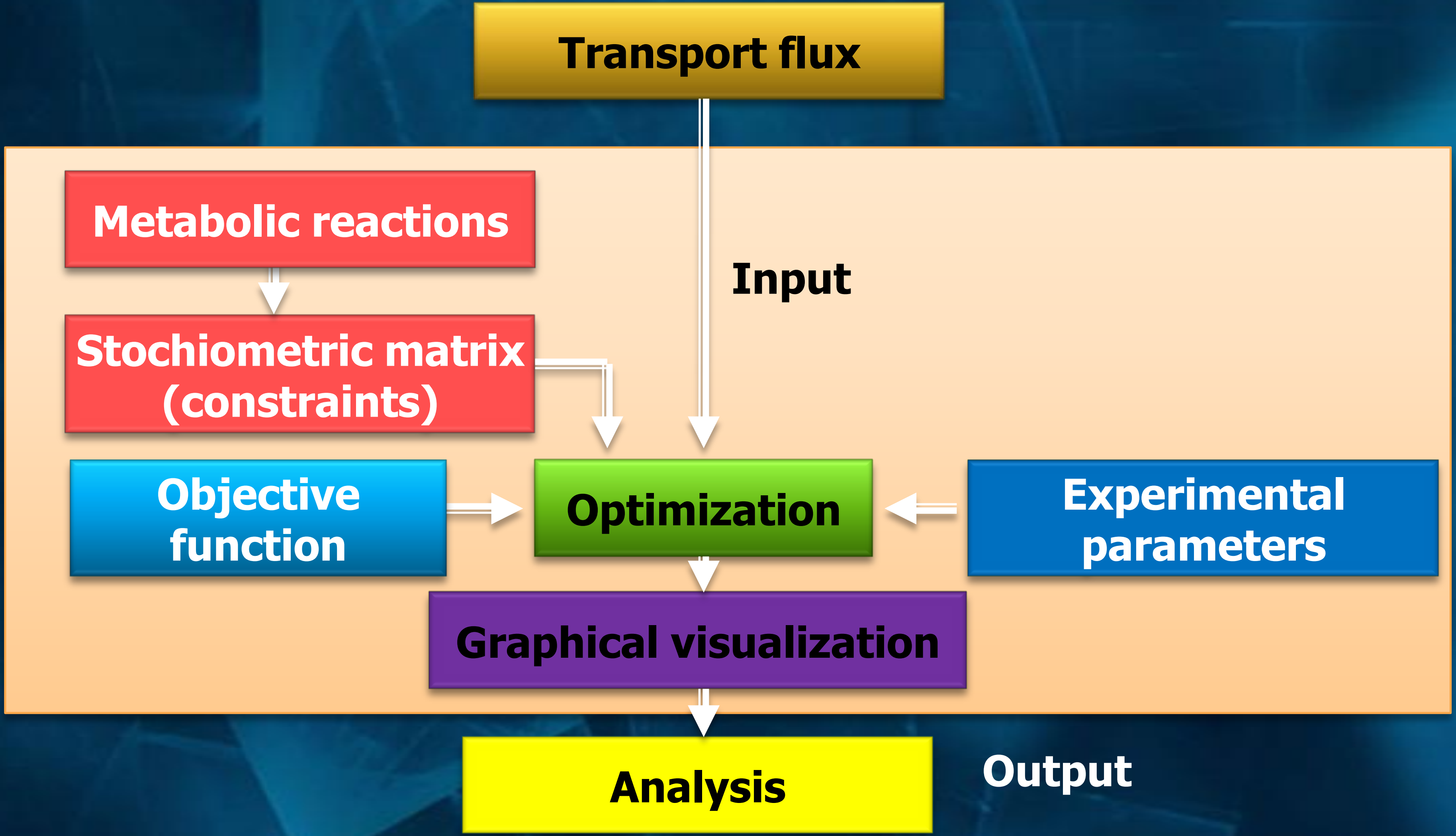
A:+RA-r1=0
B:+r1-r4-r2+r3=0
C:+r2-r5-r6-R_C=0
D:+r5+r4-r3-r7-R_D=0
E:+r6+r7-R_E=0

$$\Rightarrow \underbrace{\begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 & -1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 & 1 & 0 & -1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & -1 \end{pmatrix}}_{S} \underbrace{\begin{pmatrix} R_A \\ r1 \\ r2 \\ r3 \\ r4 \\ r5 \\ r6 \\ r7 \\ R_C \\ R_D \\ R_E \end{pmatrix}}_{v} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

S = stoichiometric matrix with K metabolites and J reactions [K×J]

v = flux vector [J×1]

Schematic Flow Chart of MFA



Transport flux

Input

Metabolic reactions

Stochiometric matrix (constraints)

Objective function

Optimization

Experimental parameters

Graphical visualization

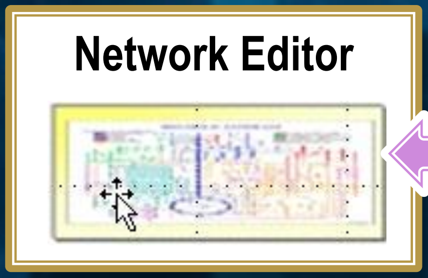
Analysis

Output

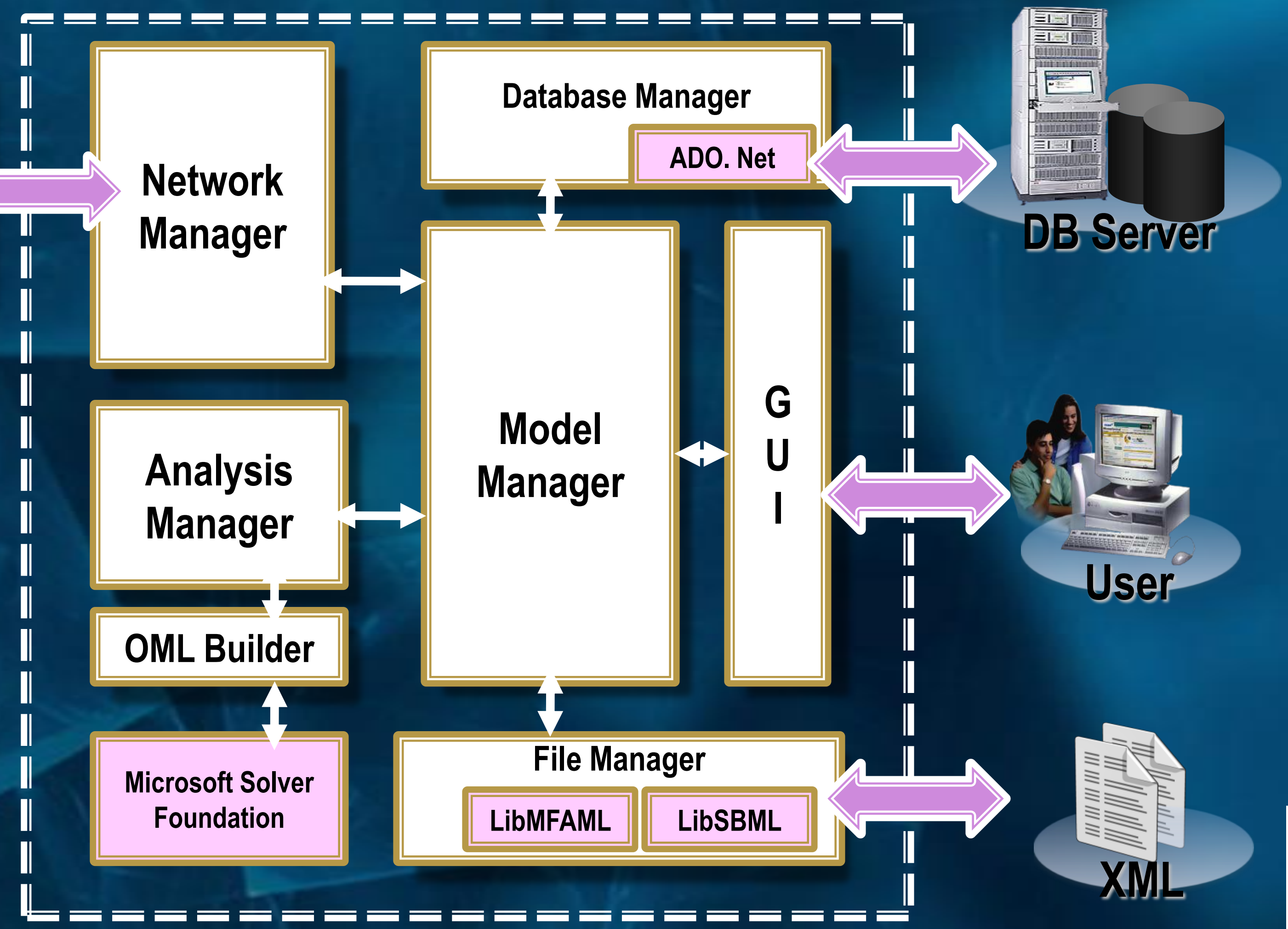
MetaFluxNet2.0

Microsoft Research

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



Network Editor



Network Manager

Database Manager

ADO. Net

DB Server

Model Manager

G U I

User

Analysis Manager

OML Builder

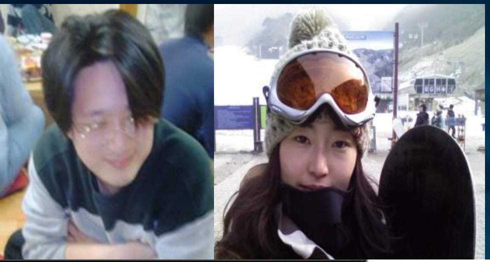
Microsoft Solver Foundation

File Manager

LibMFAML

LibSBML

XML



MetaFluxNet®

Version 2.0

Developed by

Hongseok Yun hsyun@kaist.ac.kr

Sang Yup Lee leesy@kaist.ac.kr

Metabolic & Biomolecular Engineering National Research Laboratory

Bioinformatics Research Center

KAIST

KAIST Institute for the BioCentury

CSSB Center for Systems & Synthetic Biotechnology

Amino Acids

Val, Leu, Ile : effective in hepatic failure

Glu : antiulcer drug

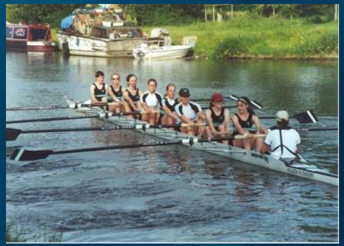
Arg : immune-enhancing effect

Medical care

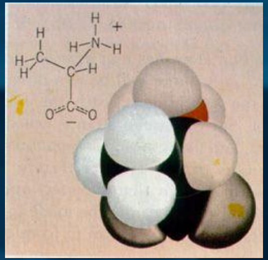
Umami and sour tastes	Sweet taste	Bitter taste
Glutamate Aspartic acid	Glycine Alanine Threonine Proline Serine Glutamine	Phenylalanine Tyrosine Arginine Leucine Isoleucine Valine Methionine Histidine

Food and animal feed

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

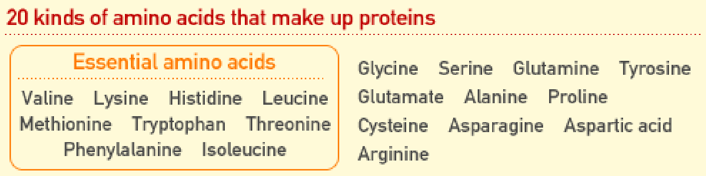


Sports



Amino acid

Health care



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

Platform chemicals

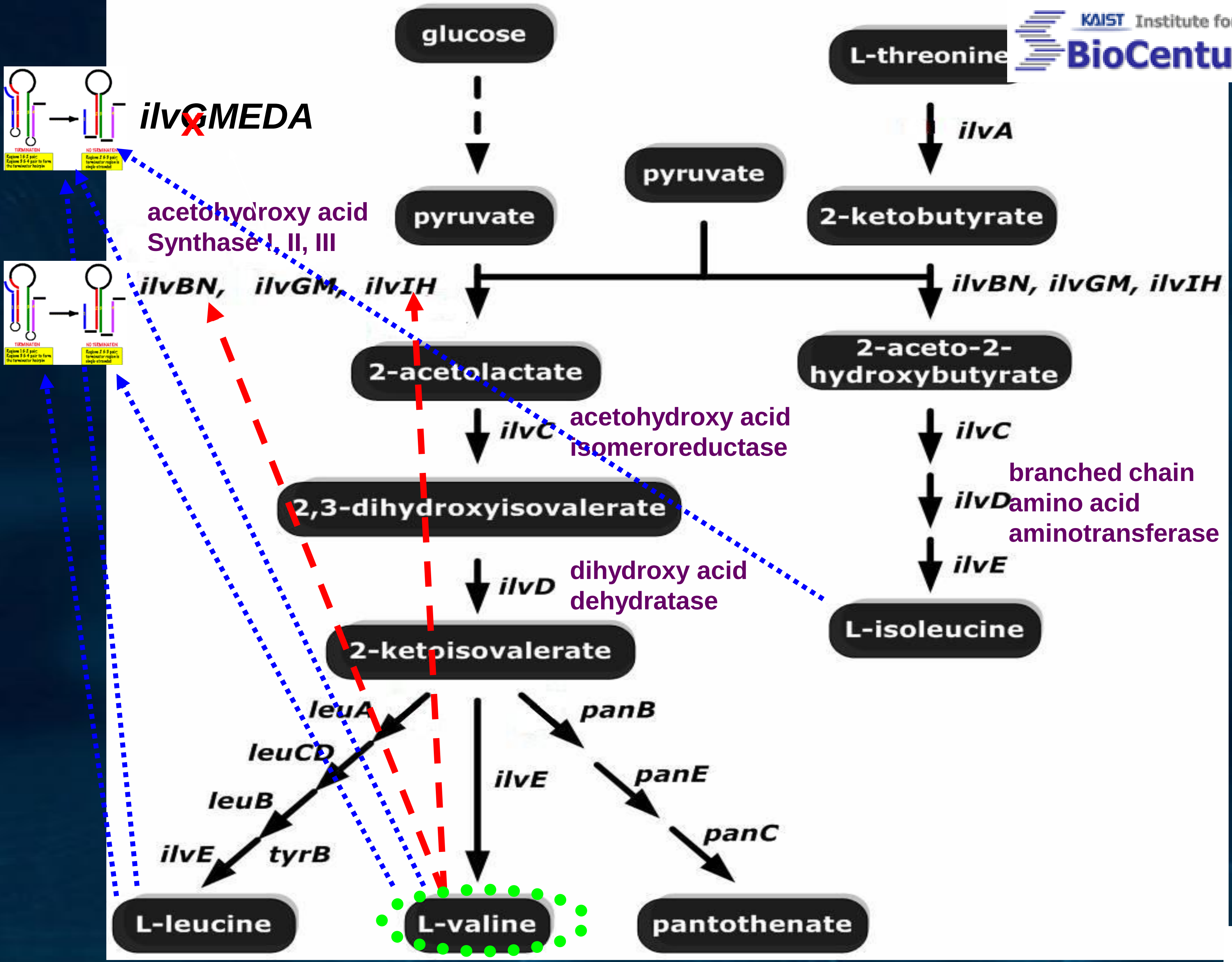
Beauty care

The skin moisturizing effect

Collagen composition

Care of damaged cuticles

Efficient burning of body fat



ilvGMEDA

acetohydroxy acid
Synthase I, II, III

ilvBN, ilvGM, ilvIH

glucose

L-threonine

ilvA

pyruvate

pyruvate

2-ketobutyrate

ilvBN, ilvGM, ilvIH

2-acetolactate

2-aceto-2-hydroxybutyrate

ilvC

acetohydroxy acid
isomeroreductase

ilvC

2,3-dihydroxyisovalerate

ilvD

branched chain
amino acid
aminotransferase

ilvD

dihydroxy acid
dehydratase

ilvE

L-isoleucine

2-ketoisovalerate

leuA

leuCD

leuB

ilvE

tyrB

ilvE

panB

panE

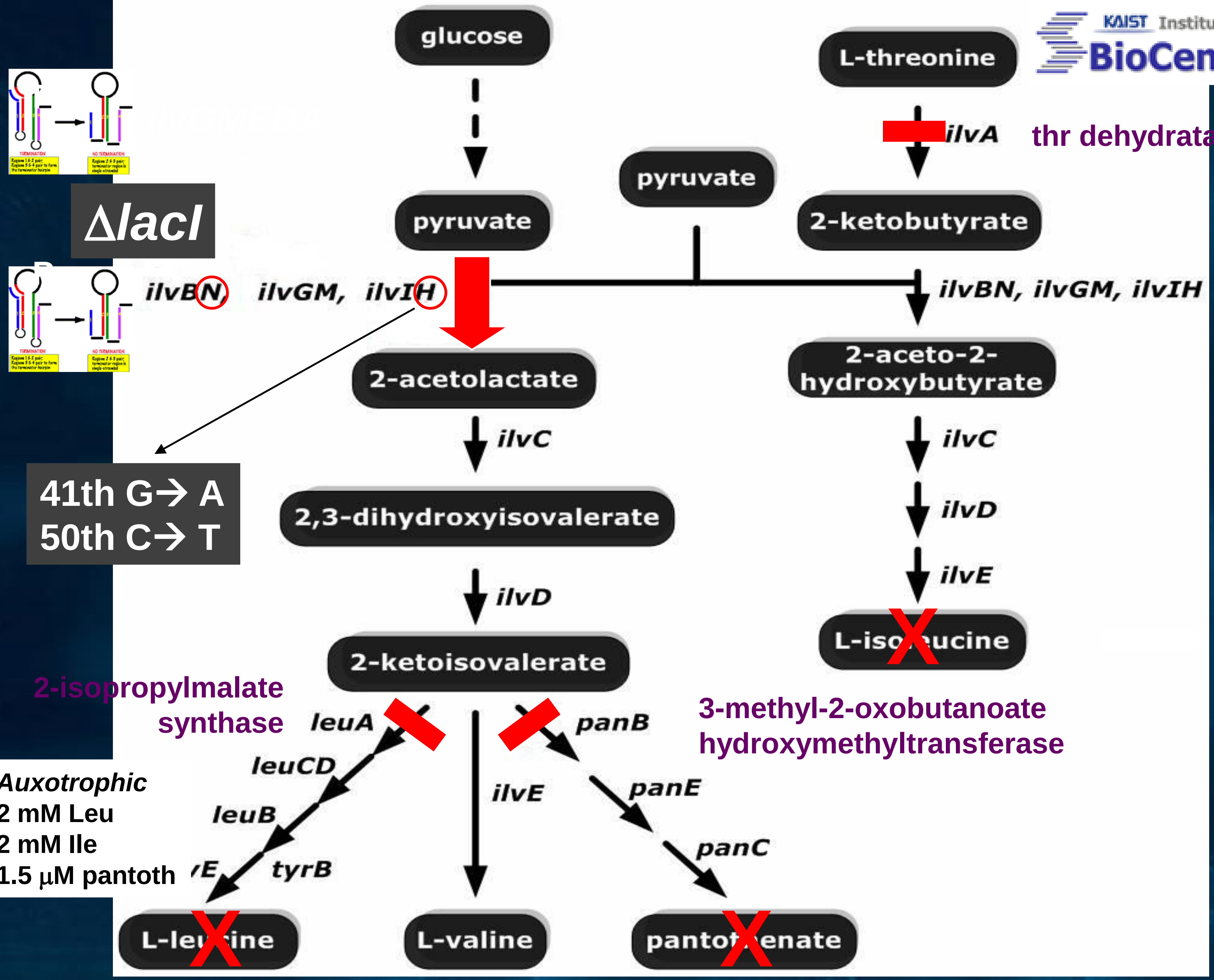
panC

L-leucine

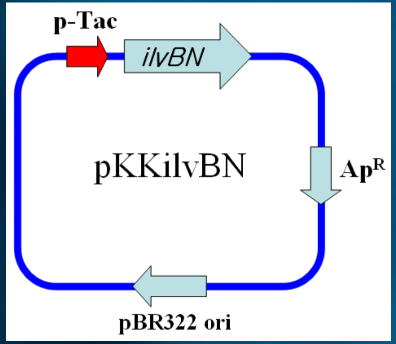
L-valine

pantothenate

Jin Hwan Park



Auxotrophic
2 mM Leu
2 mM Ile
1.5 μM pantoth



Base strain

→Val {W3110(attilvG::p*tac*, attilvB::p*tac*, *ilvH*$^{A41G, C50T}$, Δ*lacI*, Δ*ilvA*, Δ*leuA*, Δ*panB*)} harboring pKKilvBN

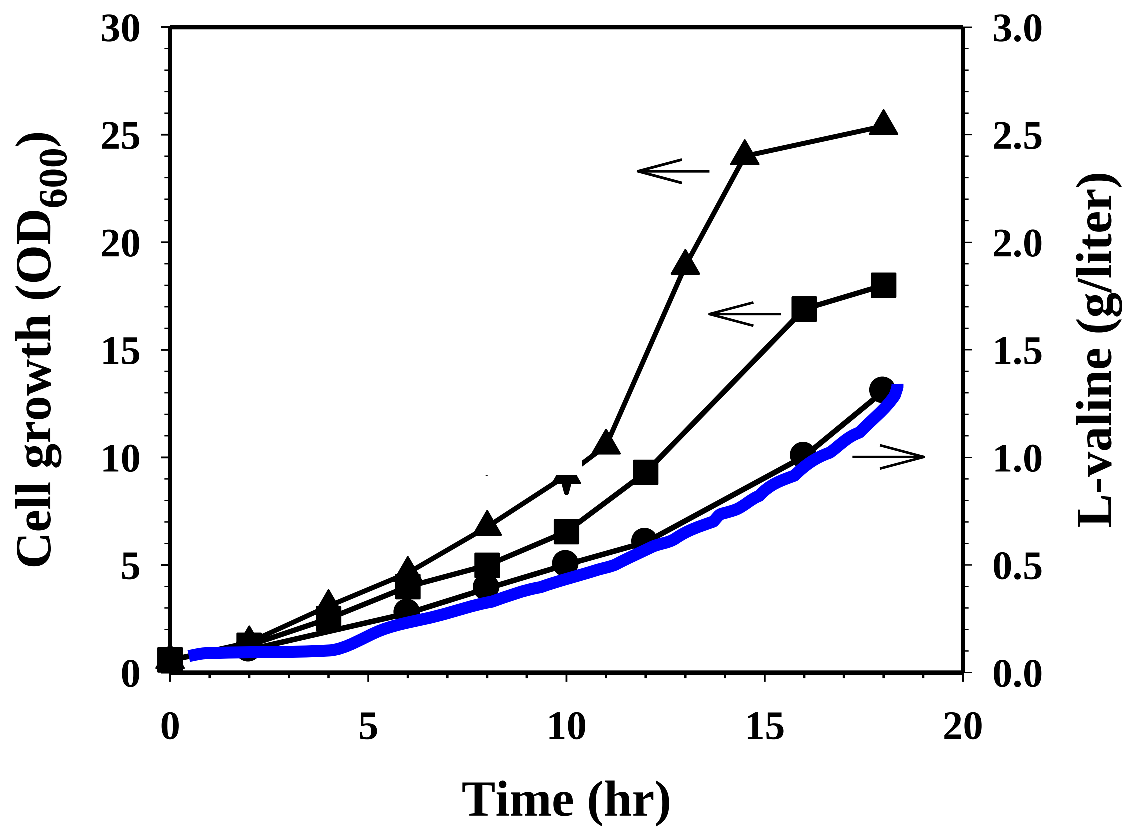
Fermentation and Transcriptome Analysis

Val : W3110(attilvG::p*tac*, attilvB::p*tac*, *ilvH*$^{A41G, C50T}$, Δ*lacI*, Δ*ilvA*, Δ*leuA*, Δ*panB*), pKKilvBN

Control : W3110 (Δ*lacI*, pKK223-3)

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

Condition 31°C, pH 6.0



- **1.31 g/L L-valine by batch fermentation**

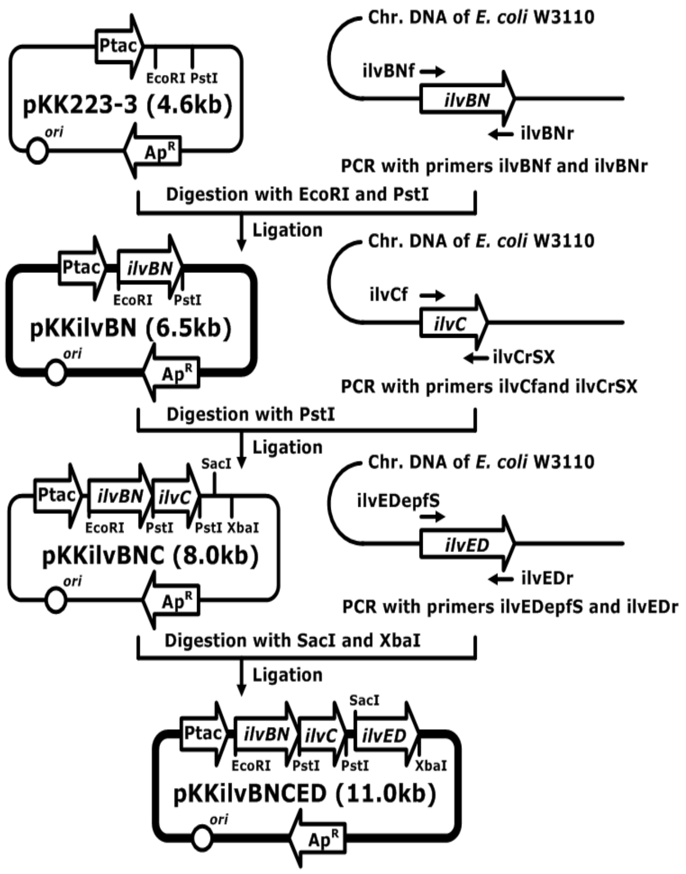
Jin Hwan Park

Overexpression of Genes

Microsoft® Research

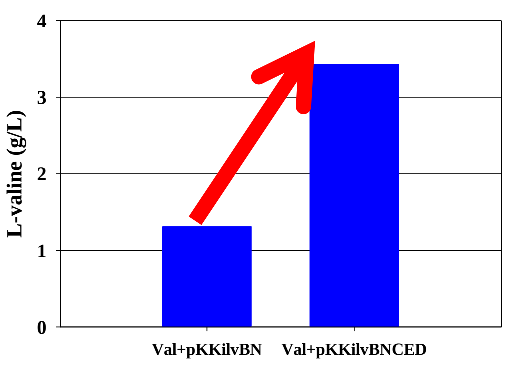
Relative expression level of L-valine biosynthetic pathway genes

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



Co-amplification of the *ilvCED* genes in pKKilvBN

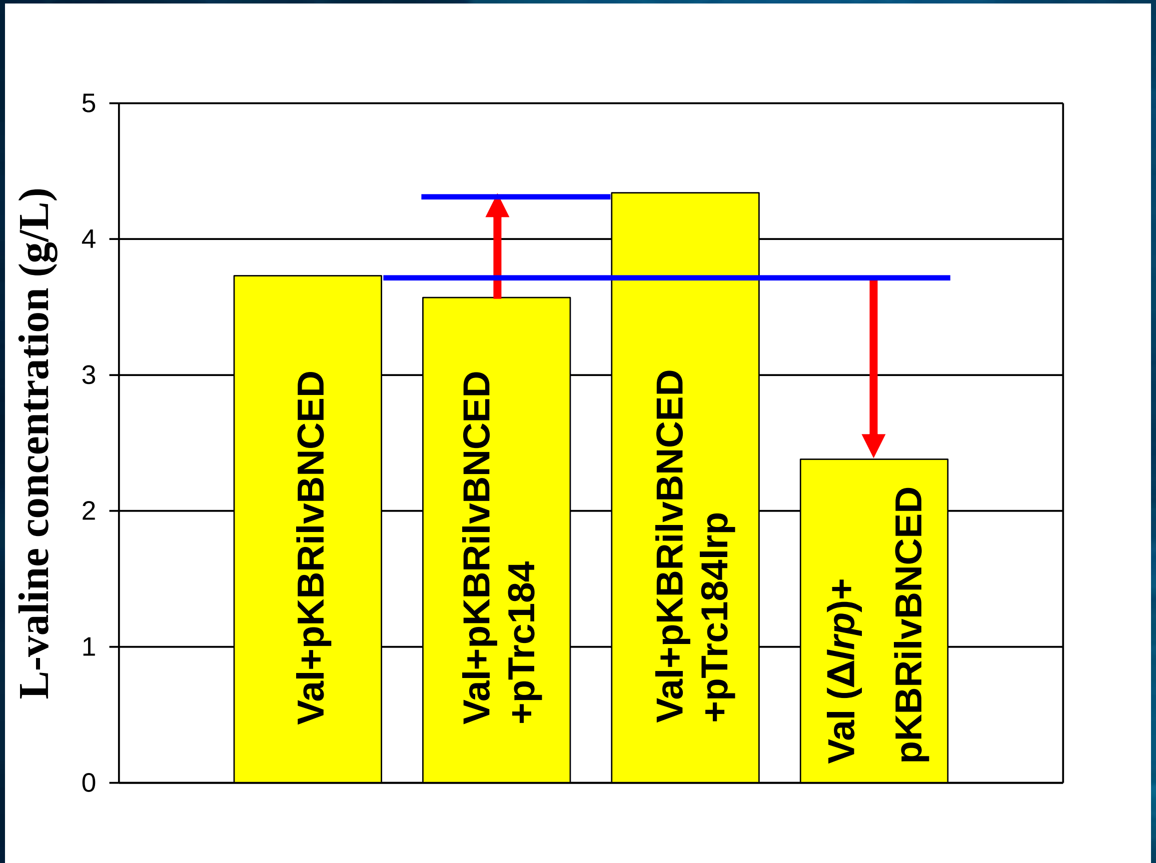
Increased L-valine production (1.31 →3.43 g/L)

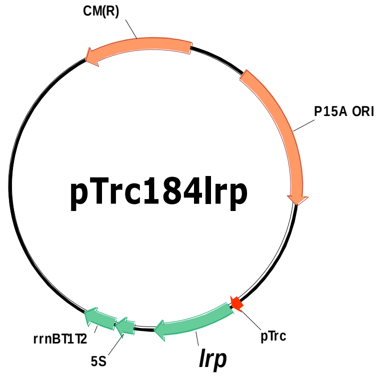


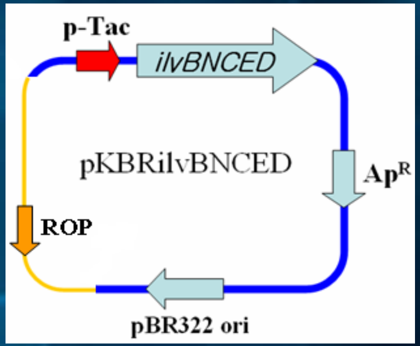
KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

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

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







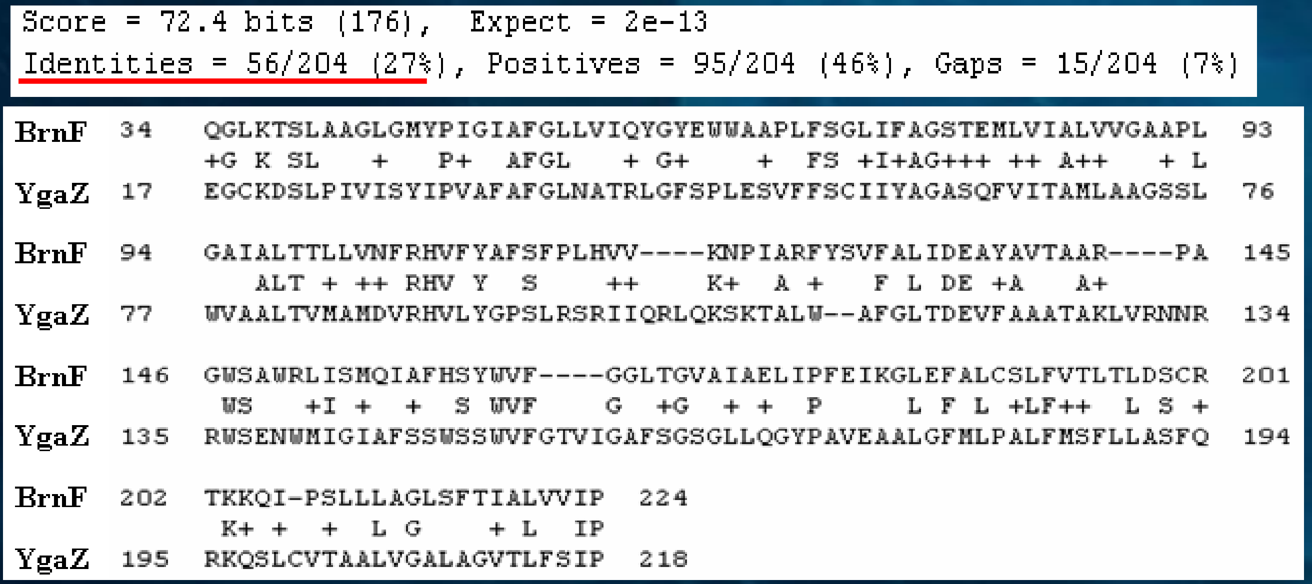
21.6% increase (3.57→4.34 g/L) in L-valine production with *lrp* overexpression

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

→Lrp plays an important role in L-valine production

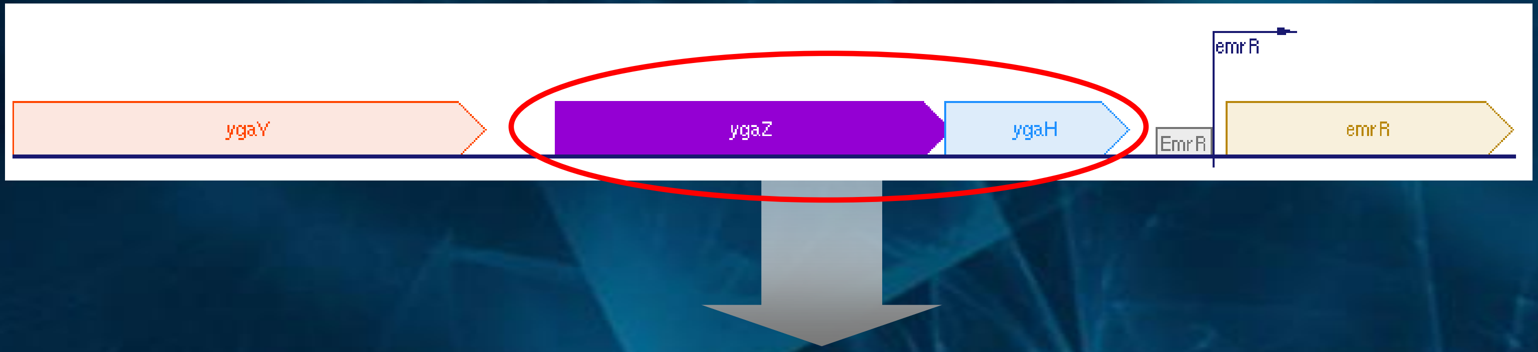
Identification of the *E. coli* gene homologous to *Corynebacterium glutamicum brnF*

Exporter Engineering



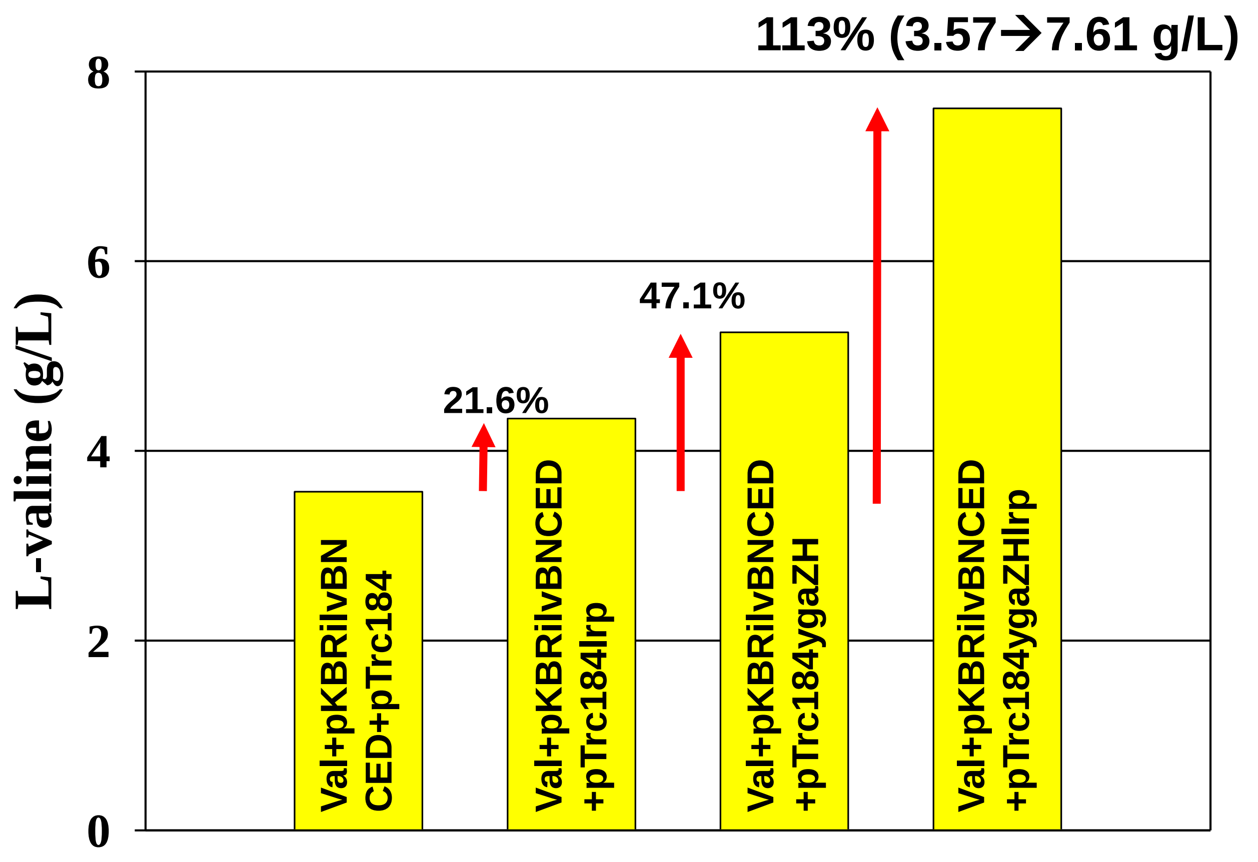
```
Score = 72.4 bits (176),  Expect = 2e-13
Identities = 56/204 (27%), Positives = 95/204 (46%), Gaps = 15/204 (7%)

BrnF  34   QGLKTSLAAGLGMYPIGIAFGLLVIQYGYEWWAAPLFSGLIFAGSTEMLVIALVVGAAPL  93
           +G K SL   +   P+  AFGL   + G+    +  FS +I+AG+++ ++ A++   + L
YgaZ  17   EGCKDSLPIVISYIPVAFAFGLNATRLGFSPLESVFFSCIIYAGASQFVITAMLAAGSSL  76

BrnF  94   GAIALTTLLVNFRHVFYAFSFPLHVV----KNPIARFYSVFALIDEAYAVTAAR----PA  145
              ALT + ++ RHV Y  S    ++    K+  A +   F L DE +A   A+
YgaZ  77   WVAALTVMAMDVRHVLYGPSLRSRIIQRLQKSKTALW--AFGLTDEVFAAATAKLVRNNR  134

BrnF  146  GWSAWRLISMQIAFHSYWVF----GGLTGVAIAELIPFEIKGLEFALCSLFVTLTLDSCR  201
            WS   +I +  +  S WVF    G  +G  + +  P     L F L +LF++  L S +
YgaZ  135  RWSENWMIGIAFSSWSSWVFGTVIGAFSGSGLLQGYPAVEAALGFMLPALFMSFLLASFQ  194

BrnF  202  TKKQI-PSLLLAGLSFTIALVVIP  224
            K+ +  +  L G    + L  IP
YgaZ  195  RKQSLCVTAALVGALAGVTLFSIP  218
```

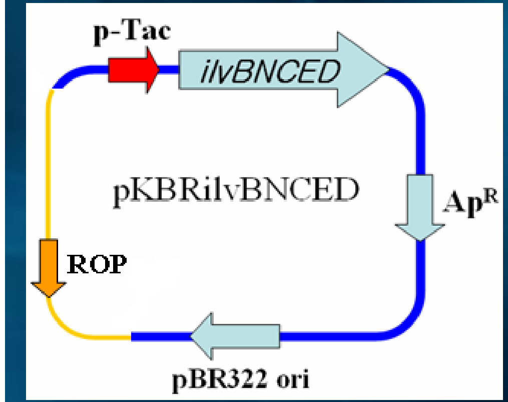
YgaZH: hypothetical protein
Downregulated to 0.61-0.75 during Val production

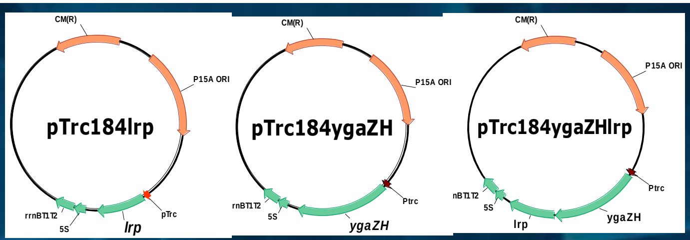


L-valine exporter in *E. coli* ?

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







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

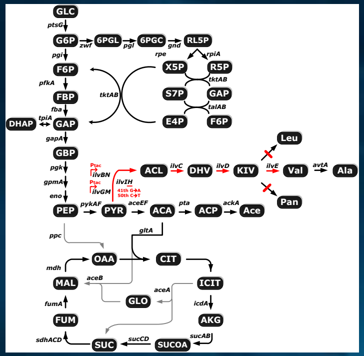
Microsoft Research

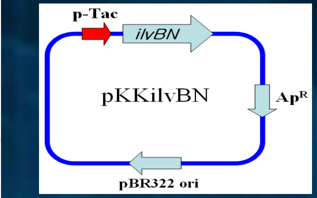
-Removed feedback inhibition
-Removed transcriptional attenuation control
-Removed major competing metabolic pathways
-Stepwise improvement based on transcriptome analysis

Amplification of L-valine production pathways

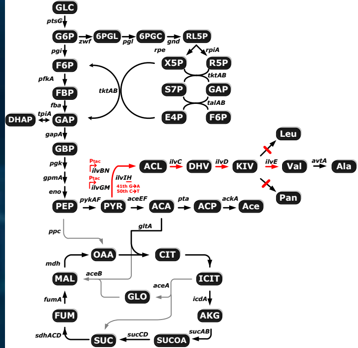
Global regulator engineering

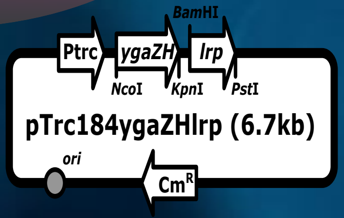
Exporter engineering





Val+pKKilvBN
1.31 g/L



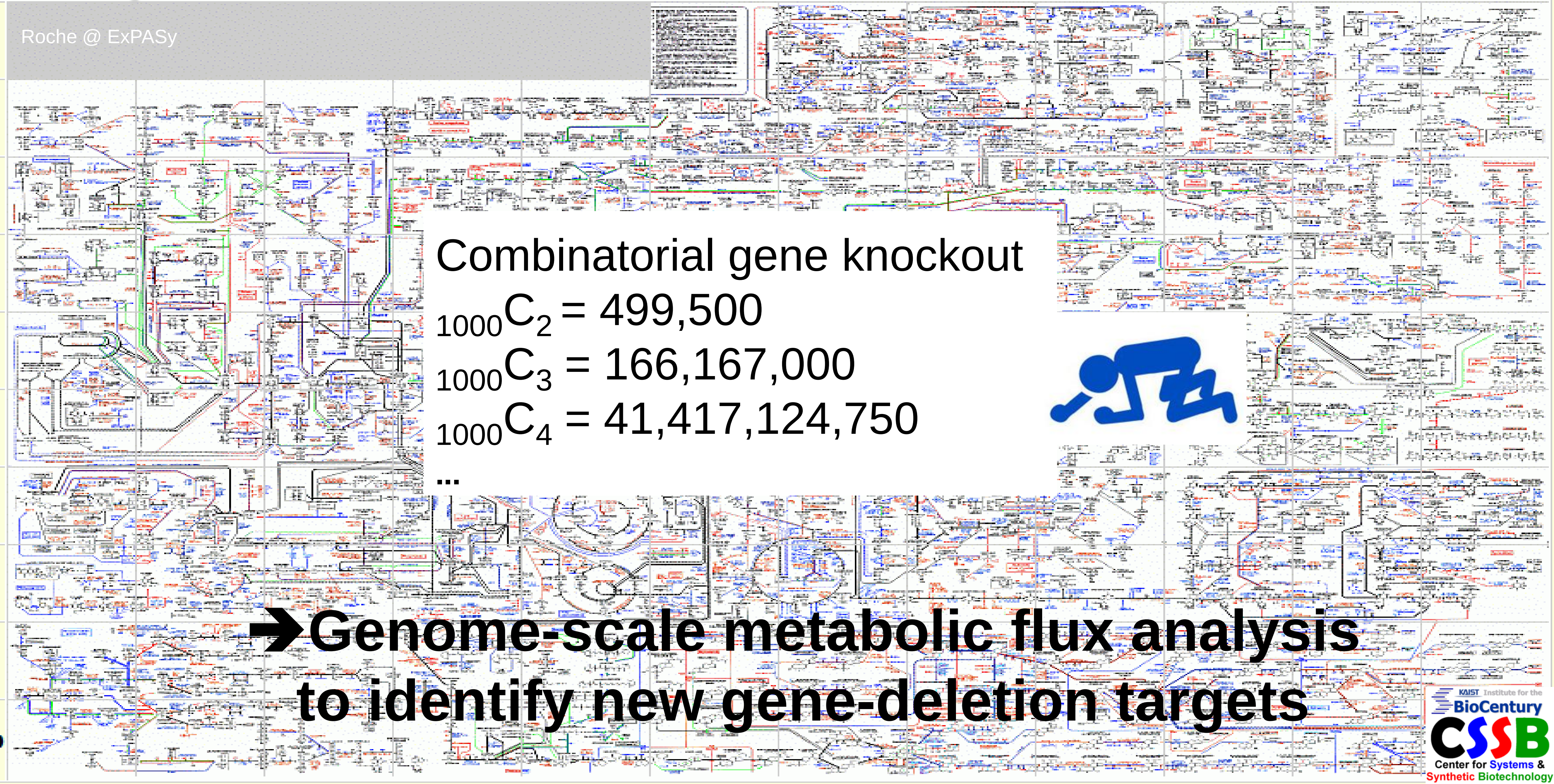


Val+pKBRilvBNCED+pTrc184ygaZHlrp
7.61 g/L

Any more targets to engineer?

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

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

MetaFluxNet®

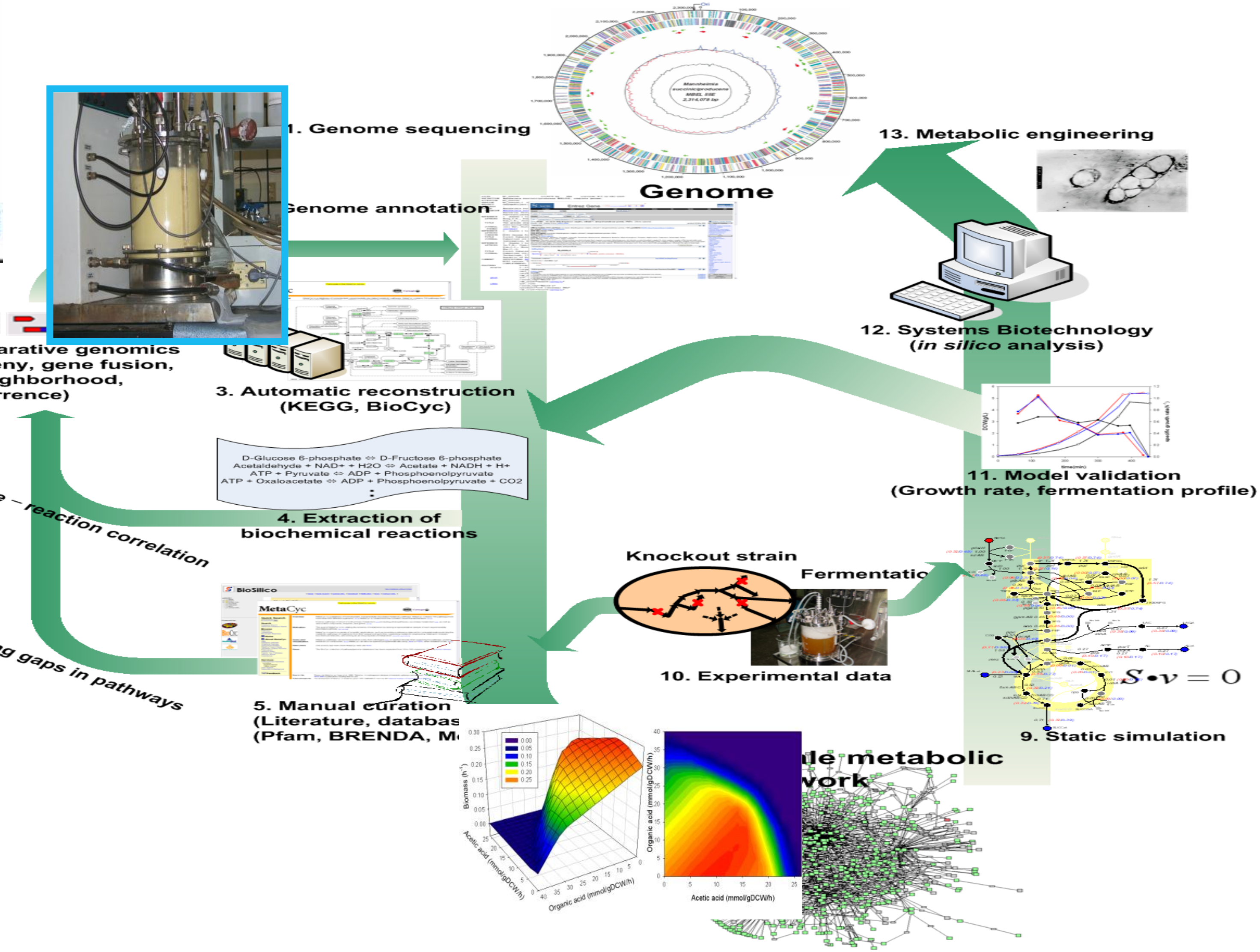
Version 2.0

Developed by

Hongseok Yun hsyun@kaist.ac.kr
Sang Yup Lee leesy@kaist.ac.kr

Metabolic & Biomolecular Engineering National Research Laboratory
Bioinformatics Research Center
KAIST

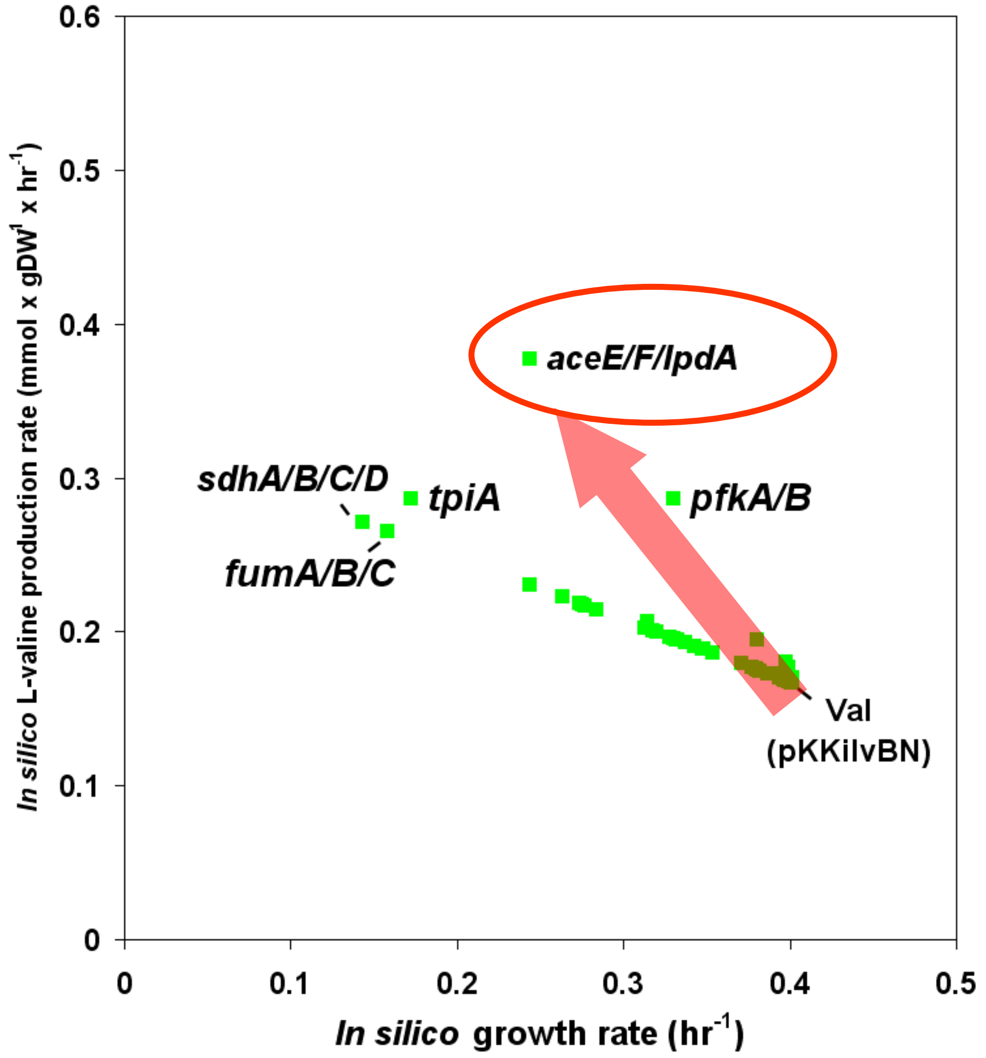
Microsoft® Research



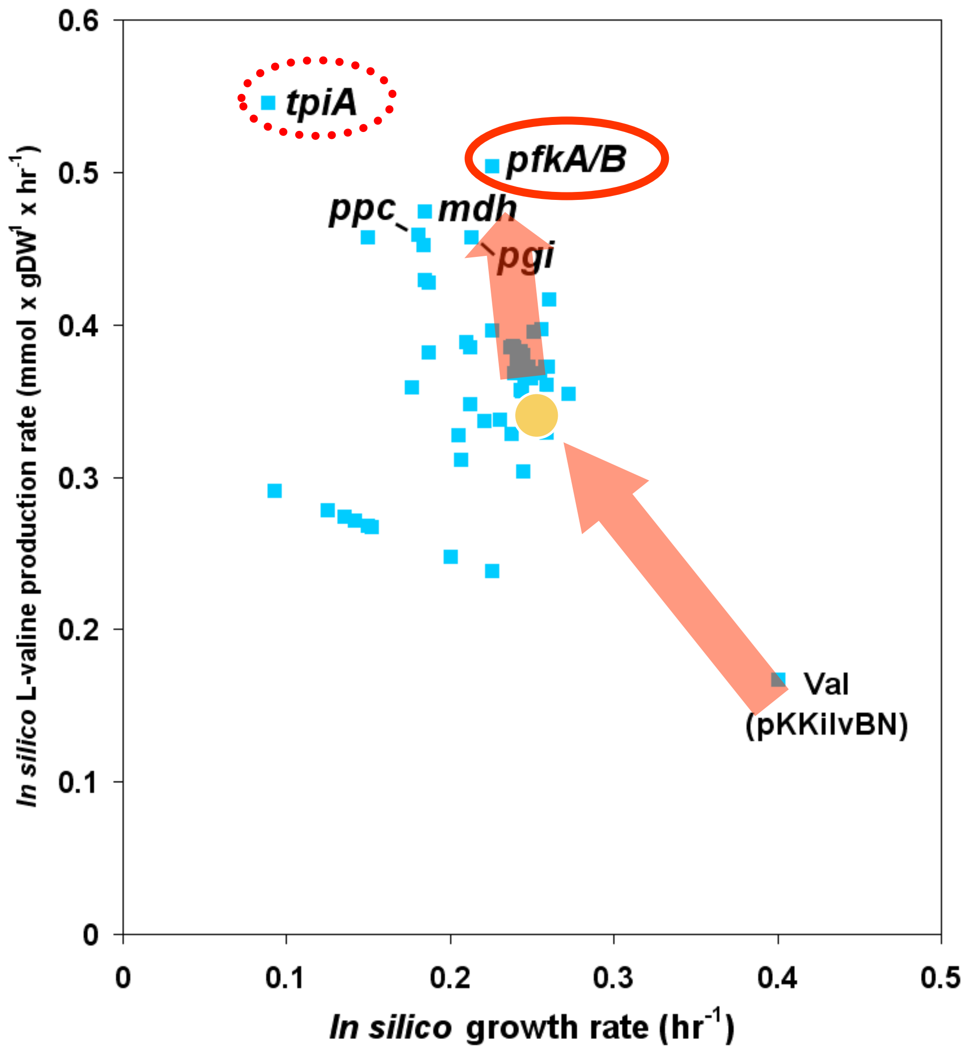
KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

Genome-scale Gene Knockouts

Microsoft Research



Single knockout



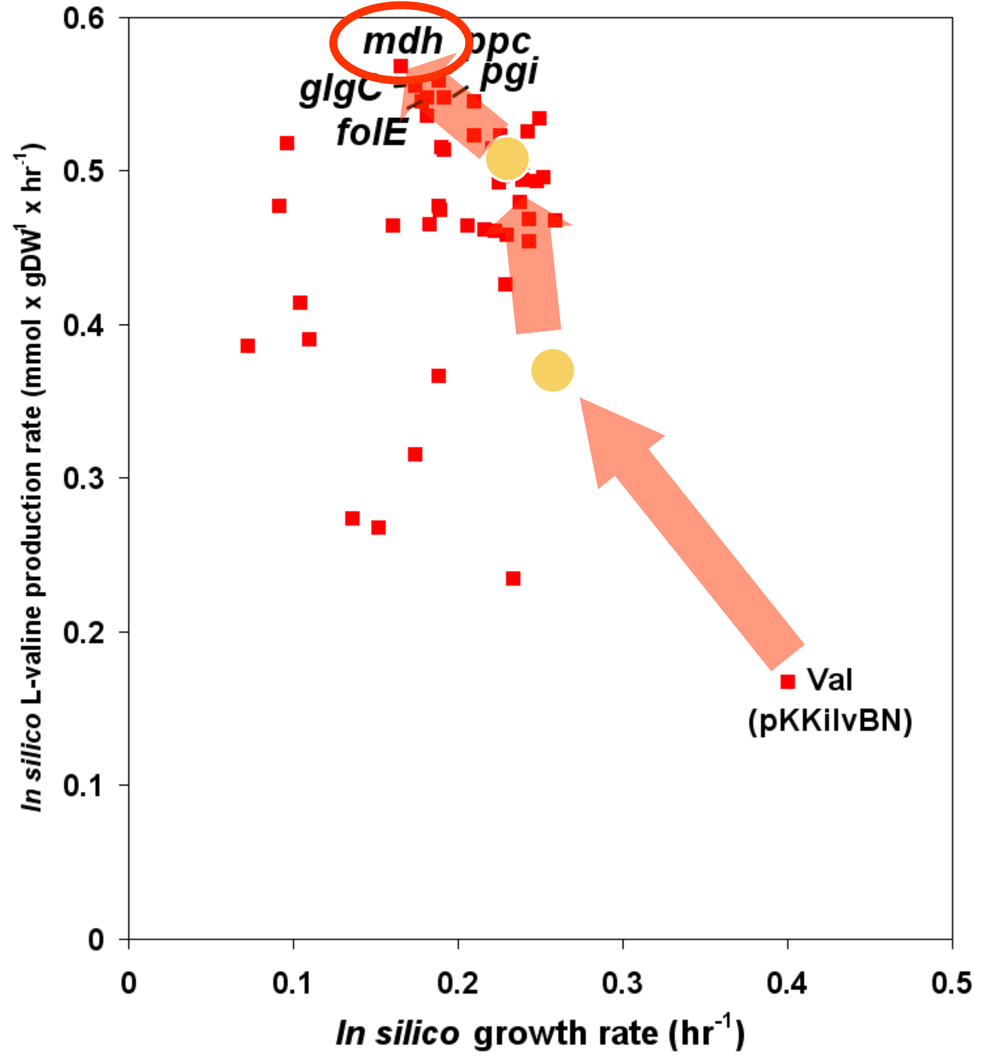
Double knockout

Tae Yong Kim

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

In silico triple gene knockout

Microsoft Research

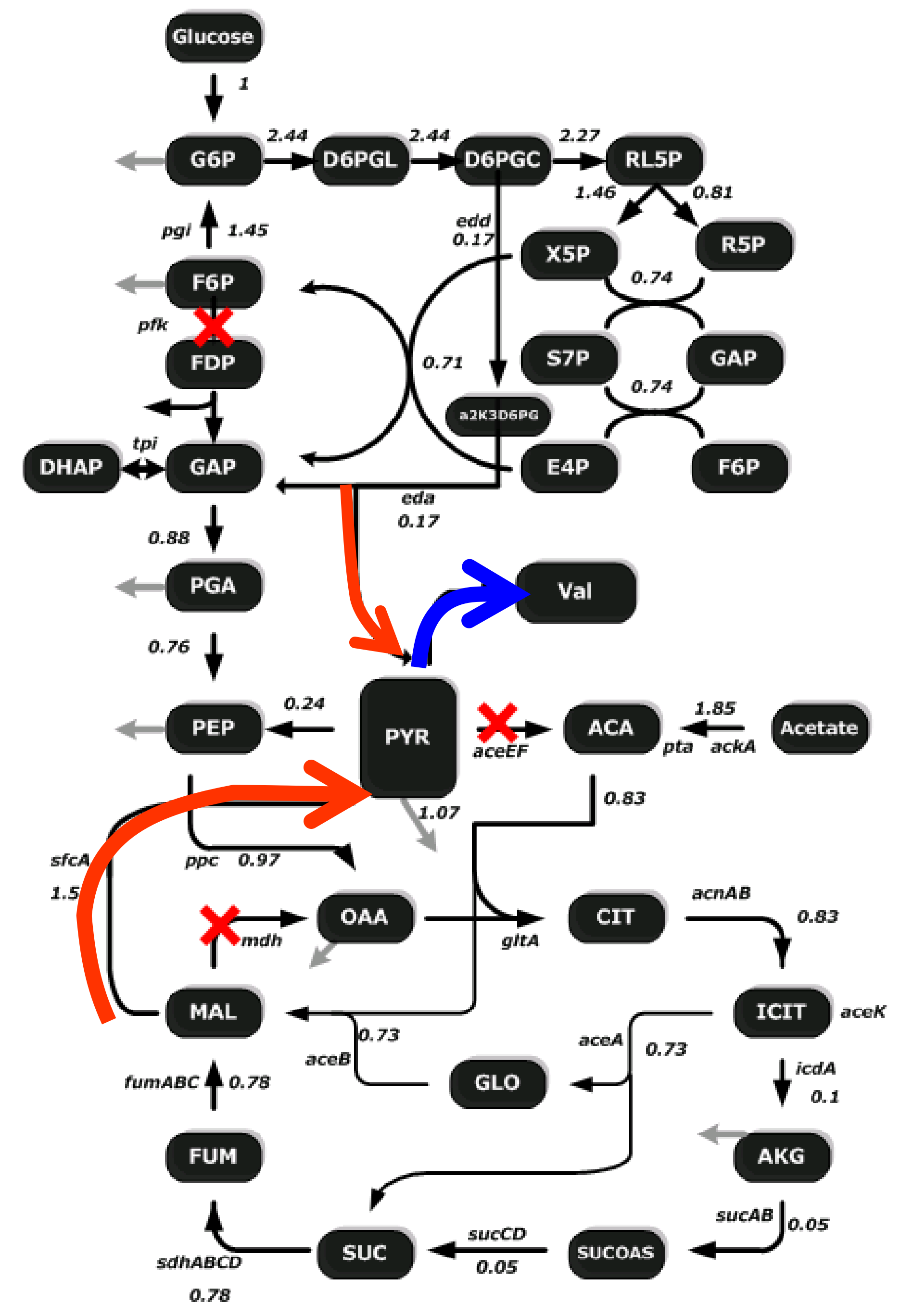


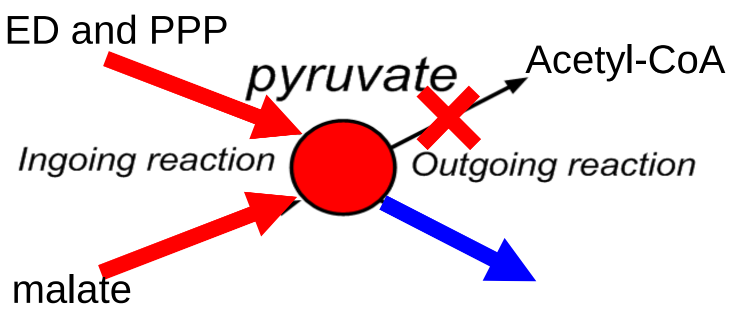
Tae Yong Kim

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

Flux redirection by triple knock-out

Microsoft Research





37.75 g L-valine from 100 g glucose

KAIST Institute for the BioCentury
CSSB Center for Systems & Synthetic Biotechnology

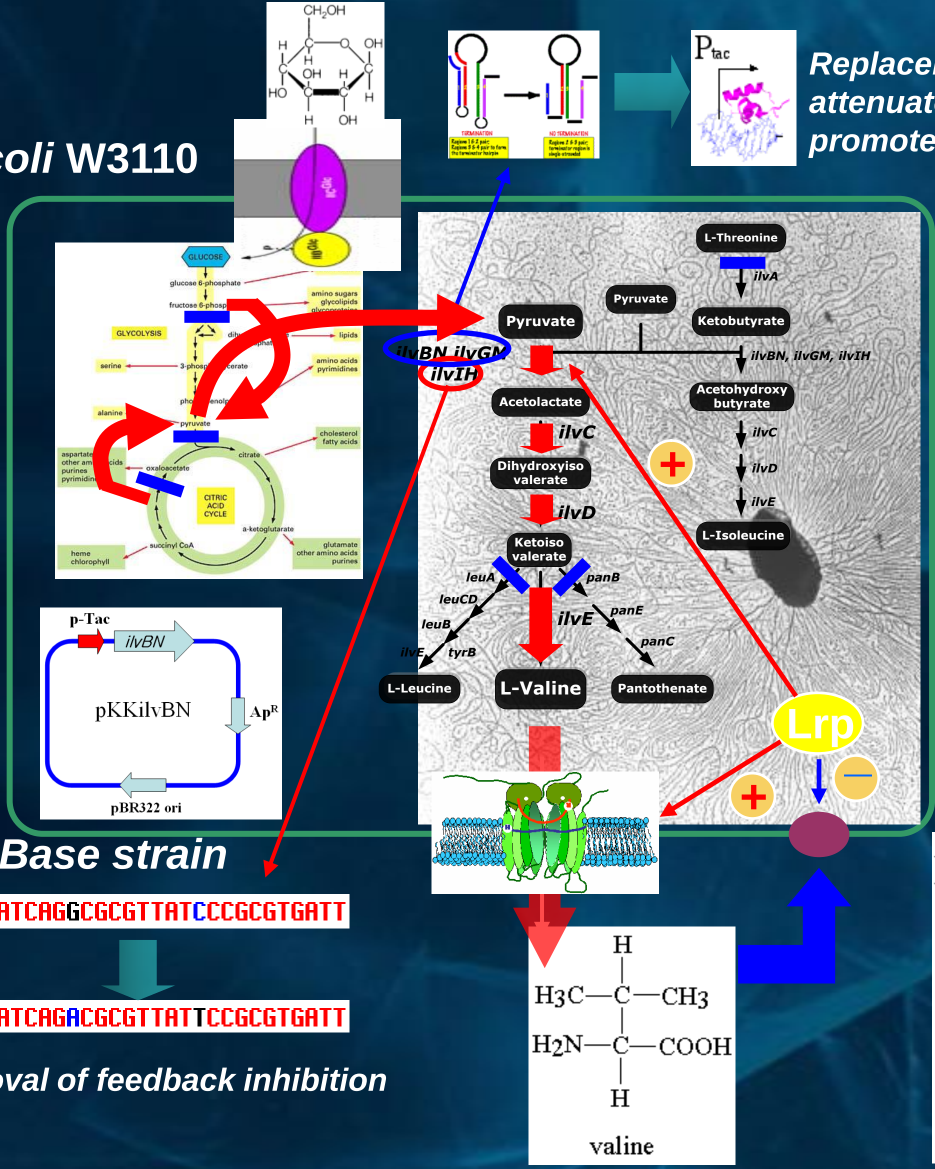
aceF/E/lpdA	**pyruvate dehydrogenase**
pfkA/B	**phosphofructokinase**
mdh	**malate dehydrogenase**

Selection of genes for triple knock-out mutation

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

Microsoft Research

***E. coli* W3110**



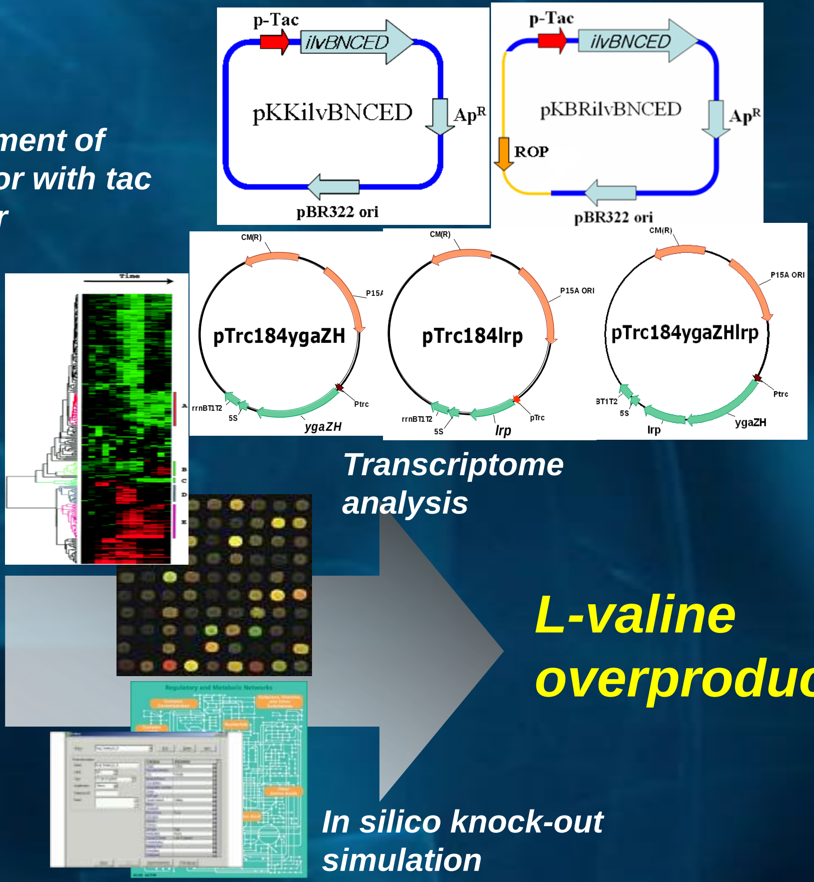
Replacement of attenuator with tac promoter

Base strain

TGAATCAGGCGCGTTATCCCGCGTGATT

TGAATCAGACGCGTTATTCCGCGTGATT

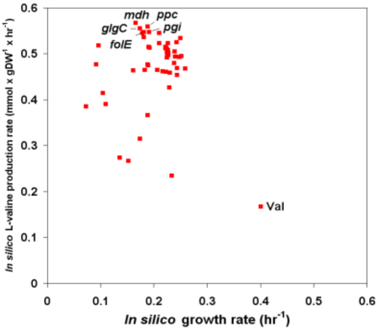
Removal of feedback inhibition

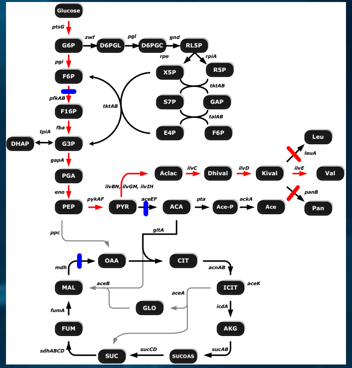


Transcriptome analysis

L-valine overproducer

In silico knock-out simulation



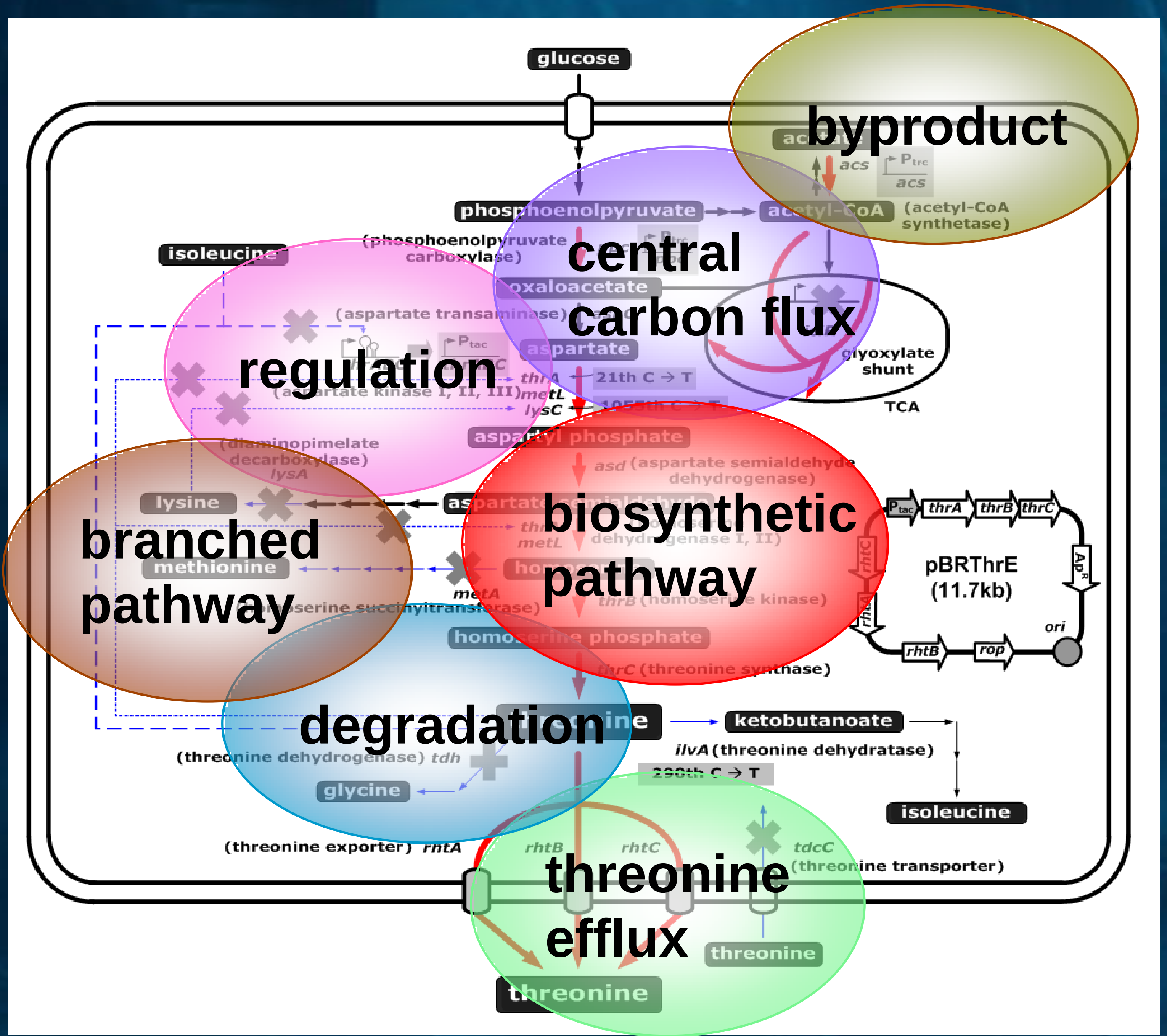


Power of Systems Metabolic Engineering

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

Systems Metabolic Eng for L-Thr

Microsoft Research



byproduct

central carbon flux

regulation

branched pathway

biosynthetic pathway

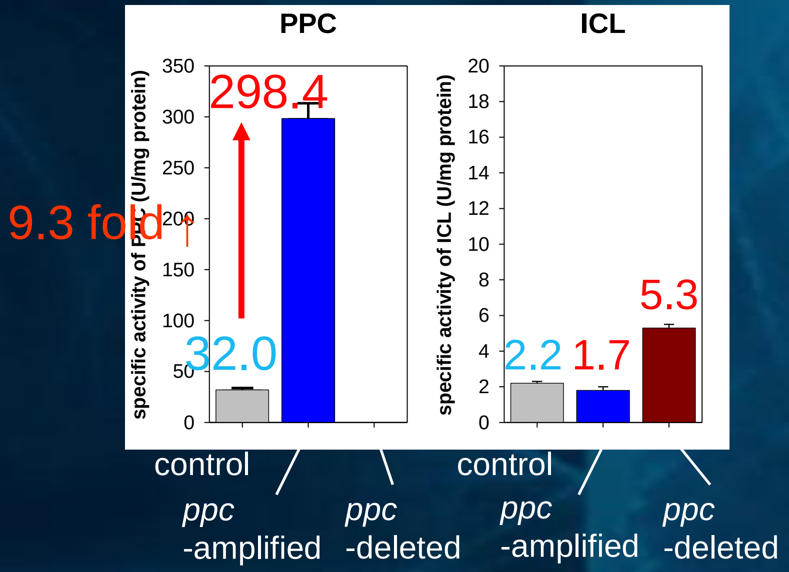
degradation

threonine efflux

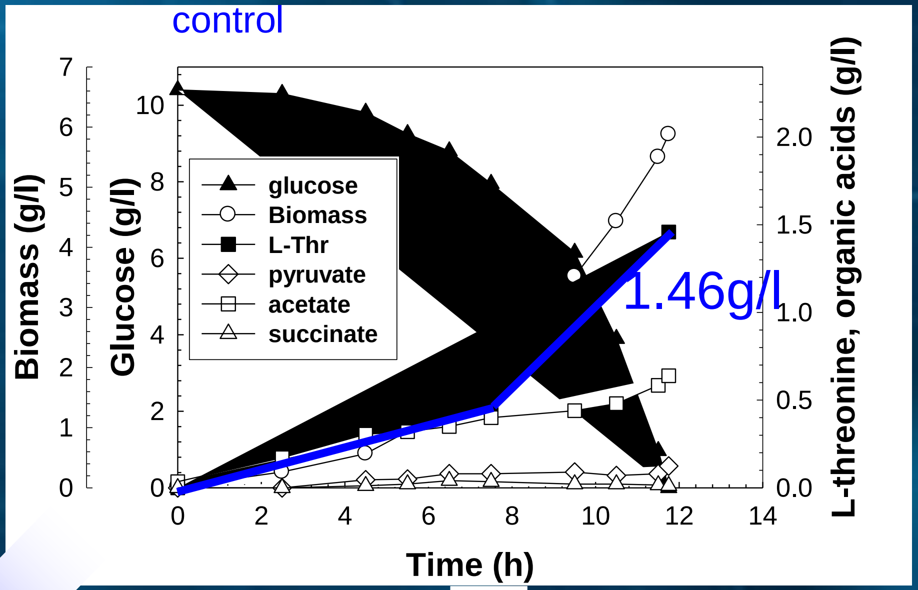
Kwang Ho Lee

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

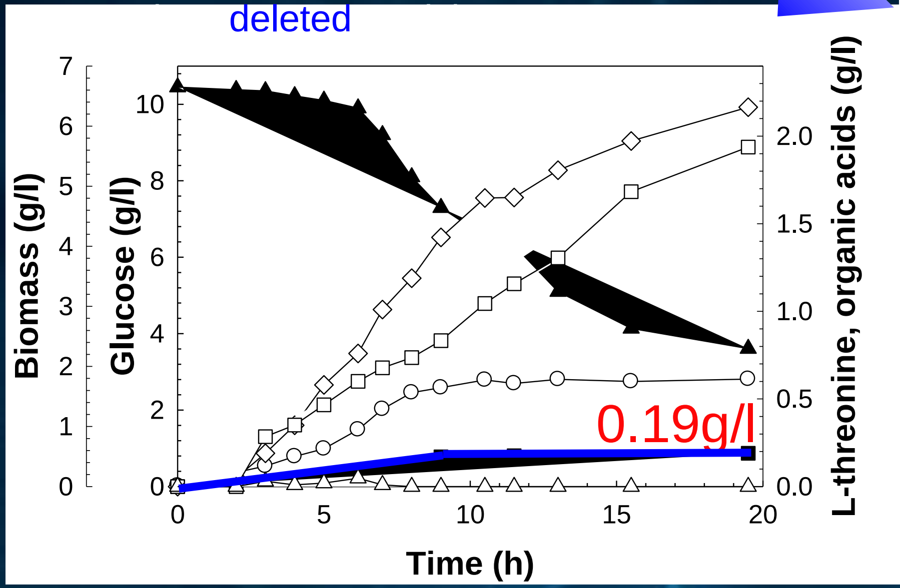
Enzyme assay



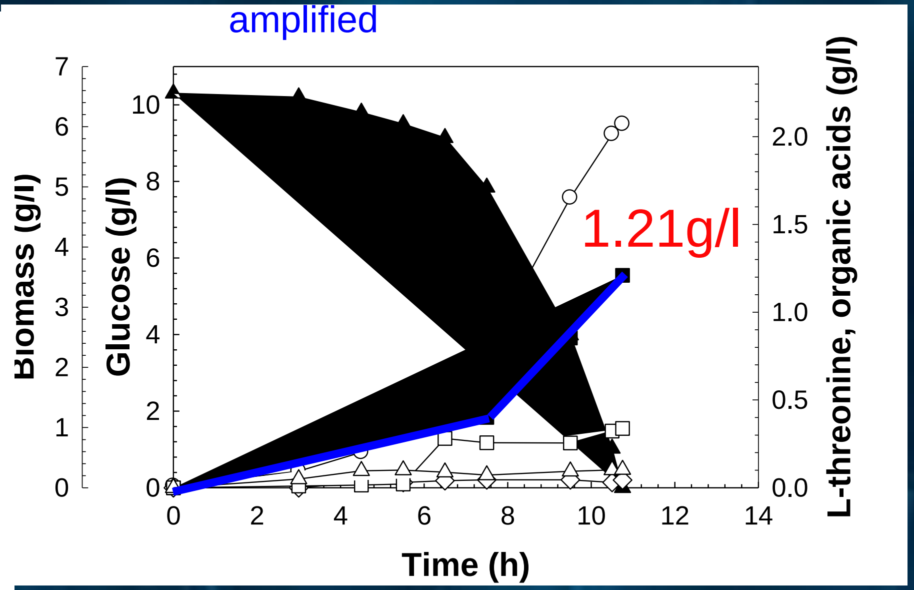
TH07 (pBRThrABC, pACYC177)



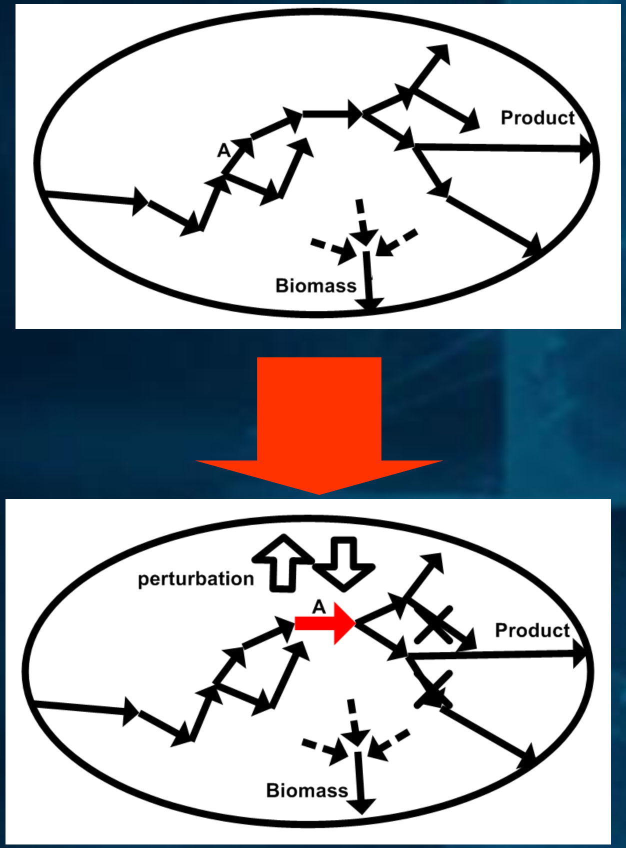
TH11C (pBRThrABC, pACYC177)



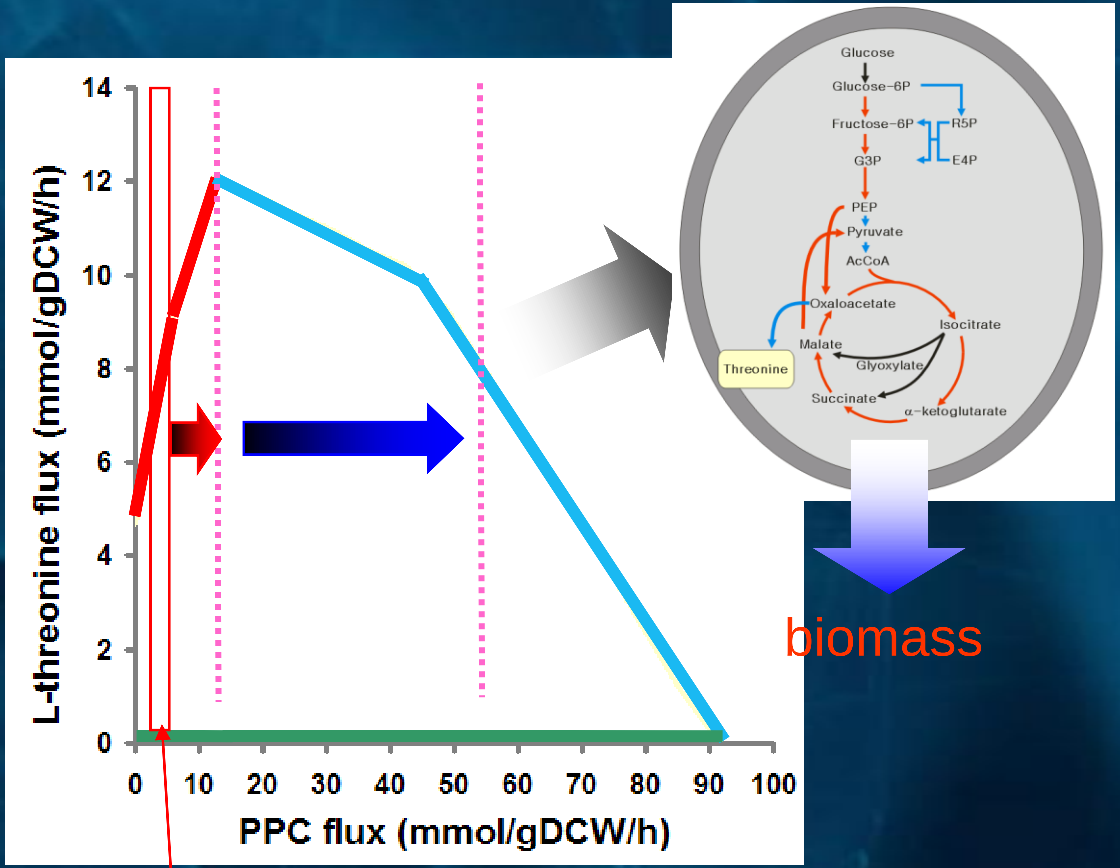
TH07 (pBRThrABC, pACYC-ppc)



In silico Perturbation Analysis

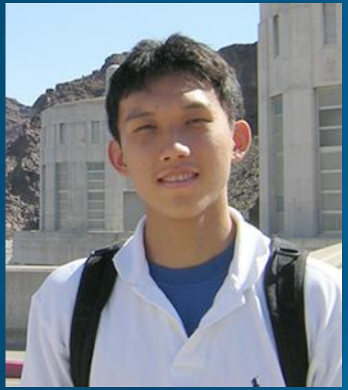


Genome-scale metabolic model
E. coli MBEL979



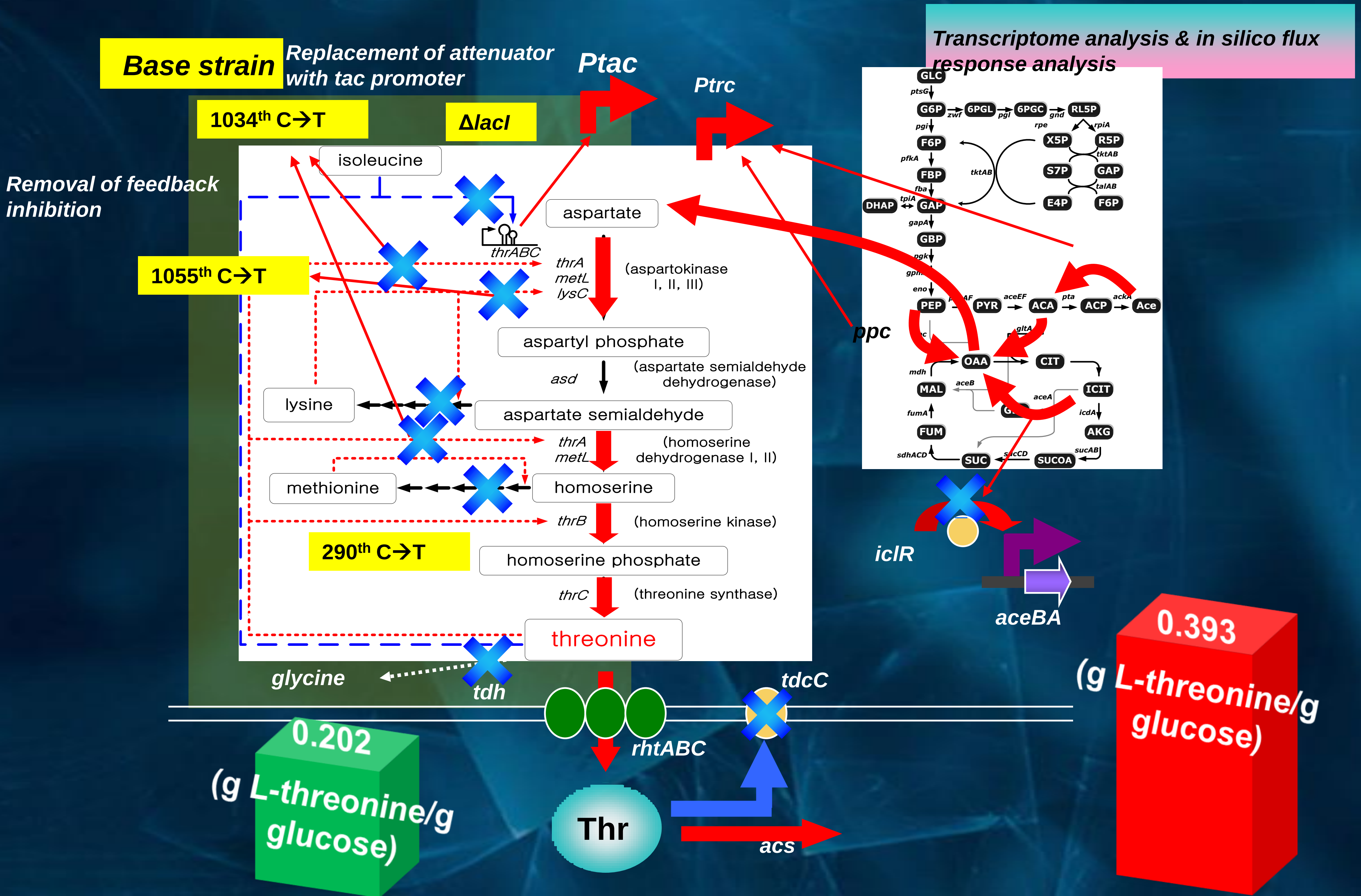
the flux of PPC is presented in the < 4 mmol/gDCW/h in aerobic condition (Fisher et al., Anal. Chem. 2004)

Tae Yong Kim



Hyun Uk Kim

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

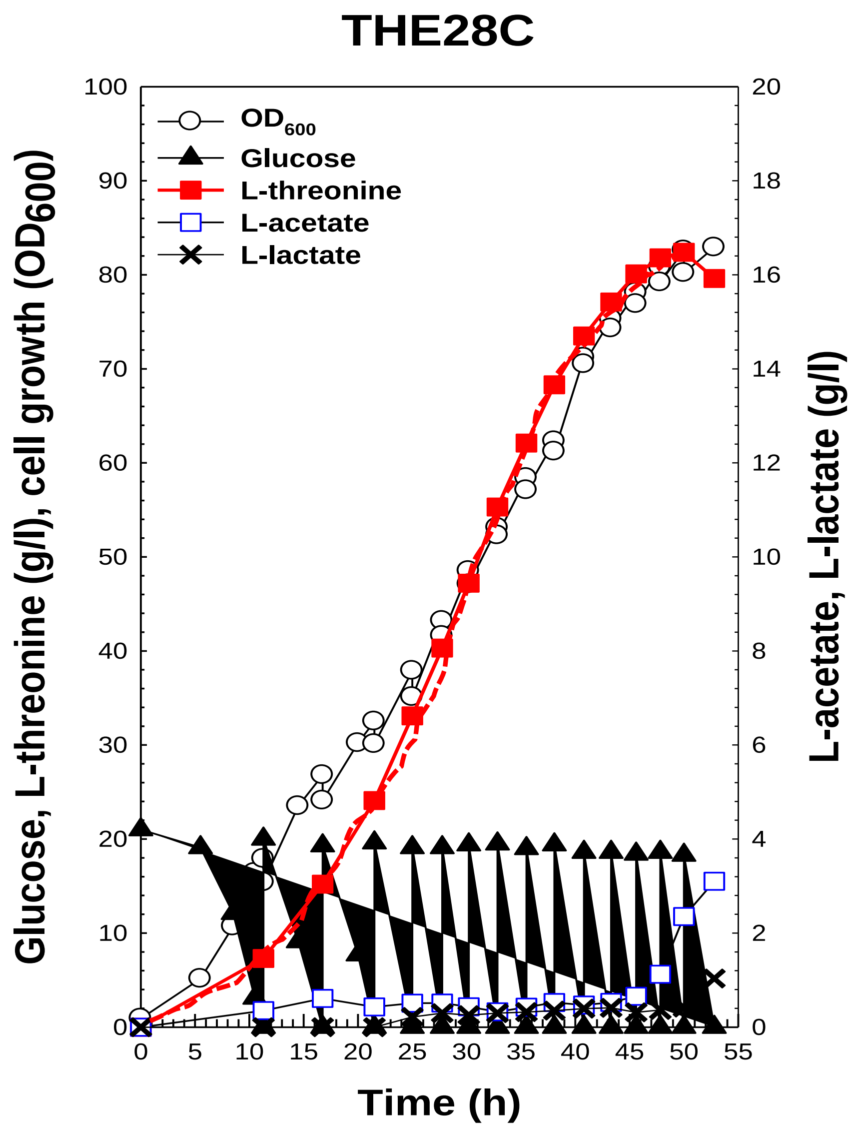


Microsoft Research

KAIST Institute for the BioCentury
CSSB Center for Systems & Synthetic Biotechnology

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

Microsoft Research

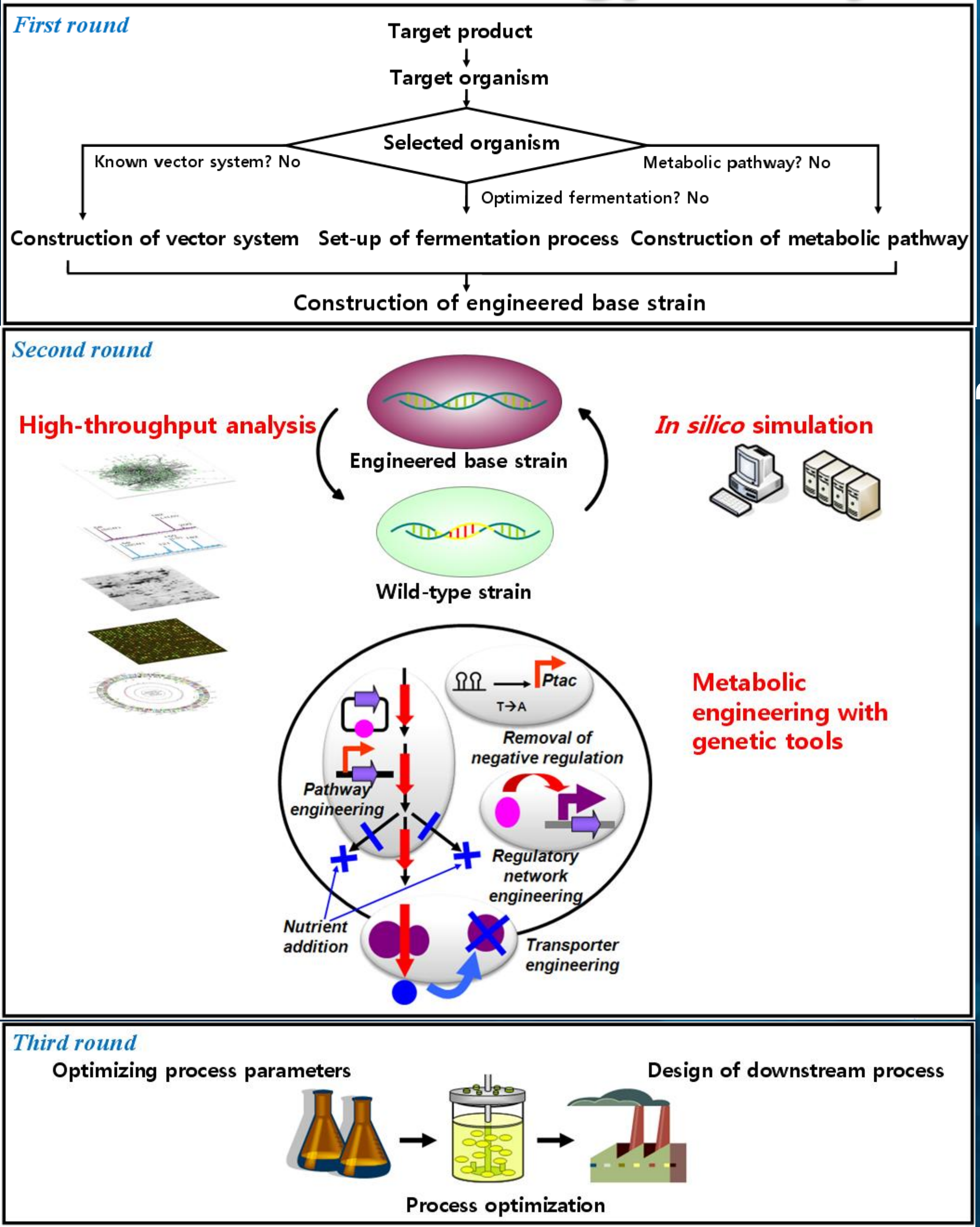


82.4 g/l
L-Thr

Kwang Ho Lee

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

General Strategy for Systems Setabolic Engineering



First round

Target product

Target organism

Selected organism

Known vector system? No

Metabolic pathway? No

Optimized fermentation? No

Construction of vector system

Set-up of fermentation process

Construction of metabolic pathway

Construction of engineered base strain

Second round

High-throughput analysis

Engineered base strain

Wild-type strain

***In silico* simulation**

Metabolic engineering with genetic tools

Third round

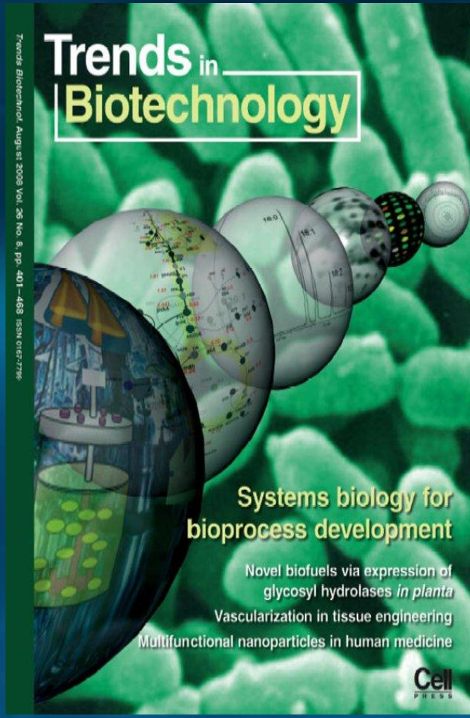
Optimizing process parameters

Design of downstream process

Process optimization

Moving towards genome-scale synthetic biology

“systems metabolic engineering”



Trends in Biotechnology

Systems biology for bioprocess development

Novel biofuels via expression of glycosyl hydrolases *in planta*

Vascularization in tissue engineering

Multifunctional nanoparticles in human medicine

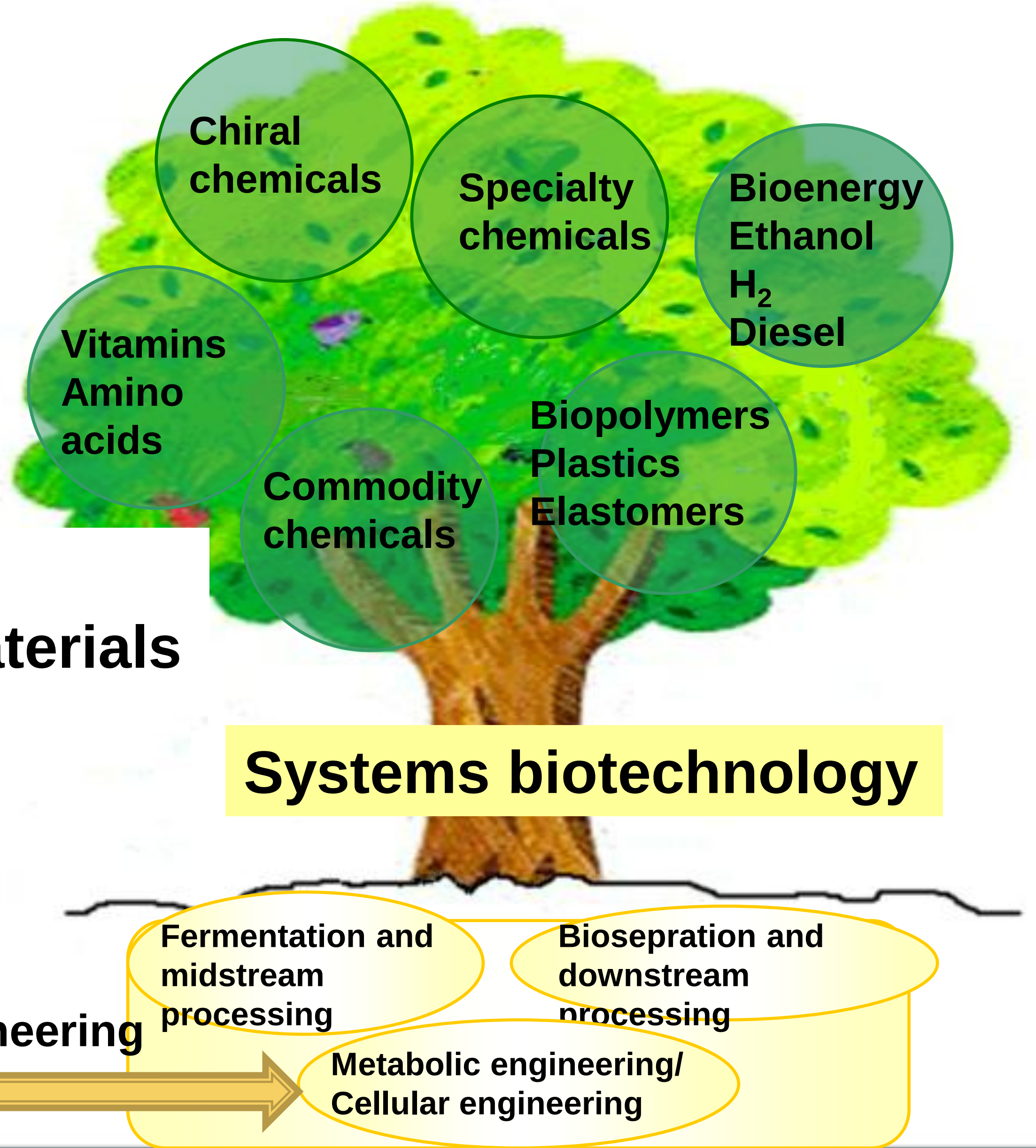
Cell Press

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

White Biotechnology

Green Growth

Sustainable system for chemicals, fuels and materials



Chiral chemicals

Specialty chemicals

Bioenergy
Ethanol
H_2
Diesel

Vitamins
Amino acids

Commodity chemicals

Biopolymers
Plastics
Elastomers

Systems biotechnology

Fermentation and midstream processing

Biosepration and downstream processing

Metabolic engineering/
Cellular engineering

Systems metabolic engineering

Microsoft Research

Red Biotechnology

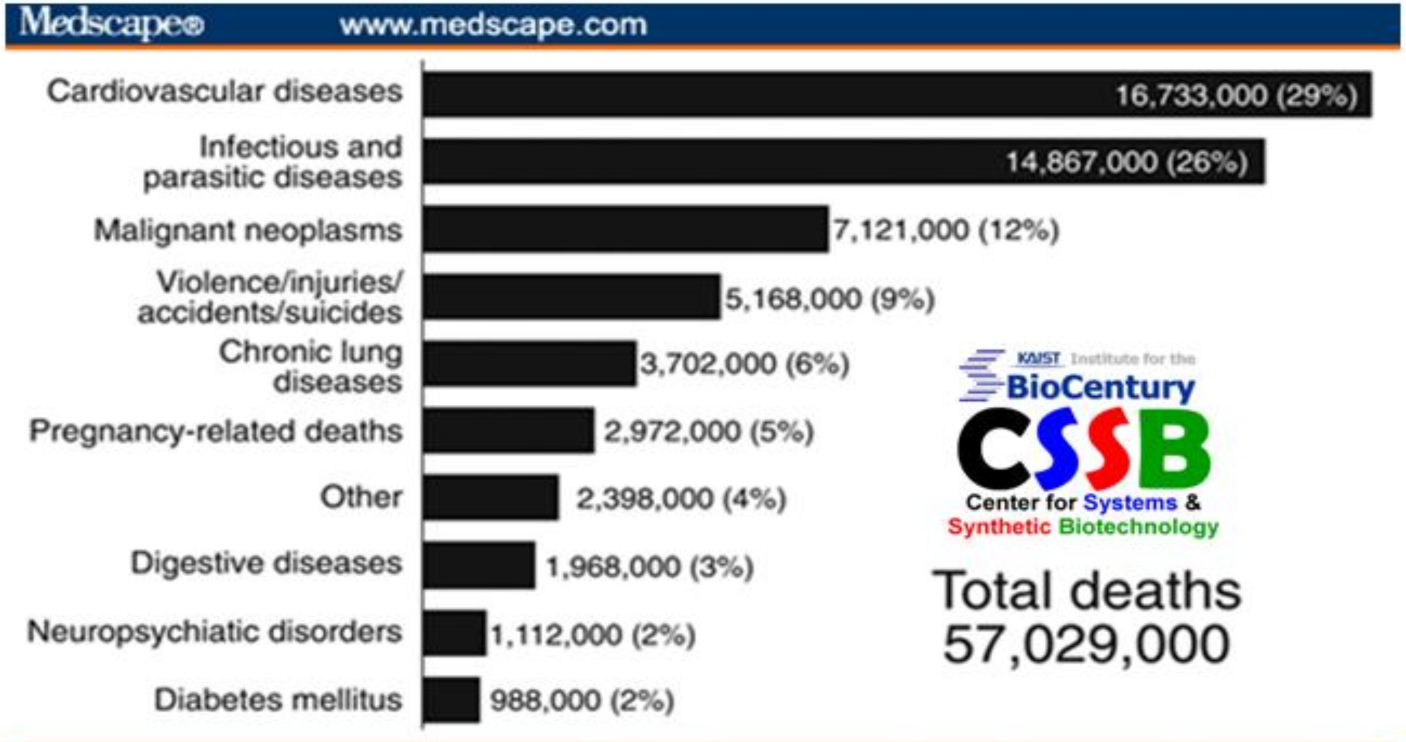
enabled through systems biotechnology

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

Sang Yup Lee

Infectious diseases

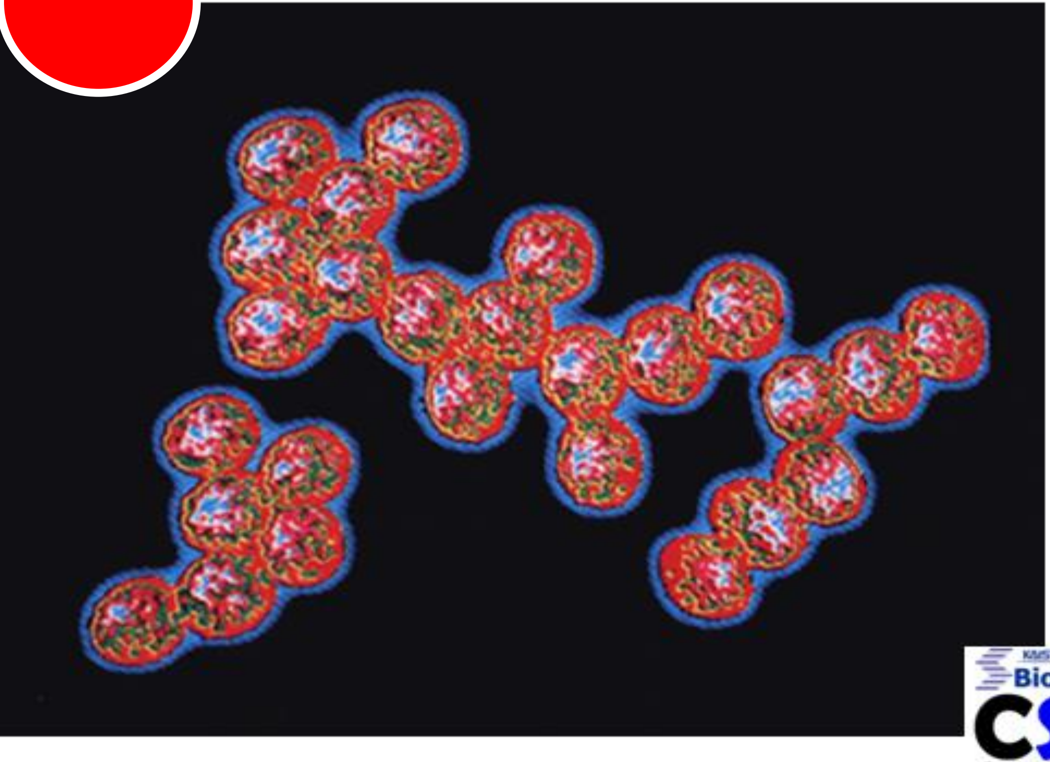
- Incidence of infectious disease: 540 million per year
- 15 million annual deaths worldwide



Total deaths
57,029,000

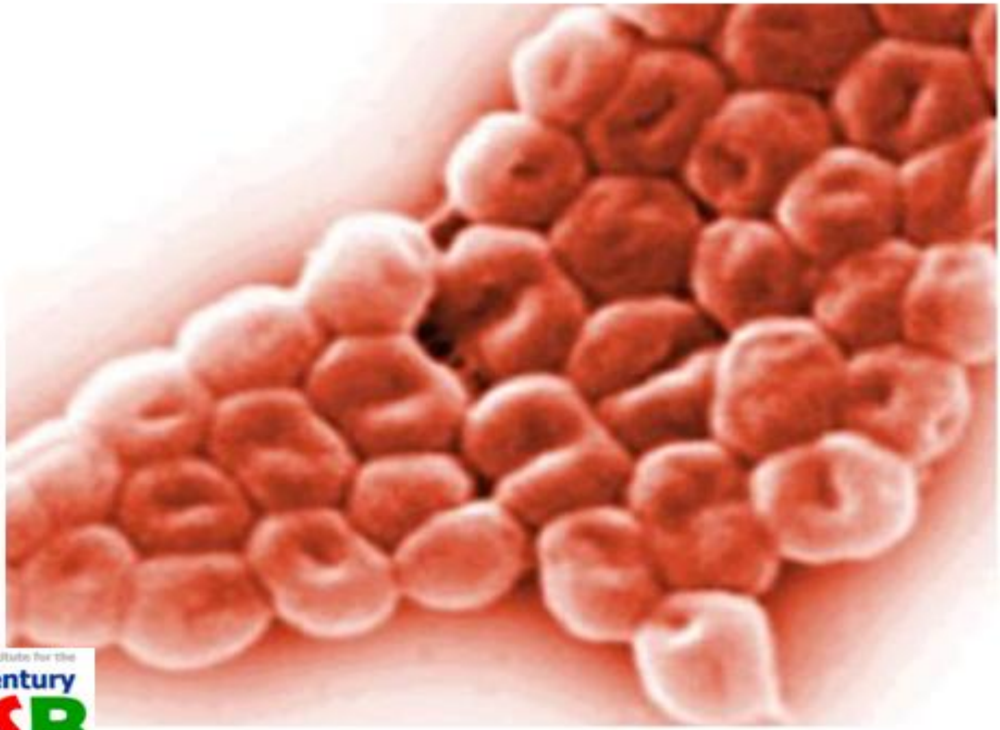
Source: Emerg Infect Dis © 2005 Centers for Disease Control and Prevention (CDC)

Continuous emergence of super bacteria



Staphylococcus sp.

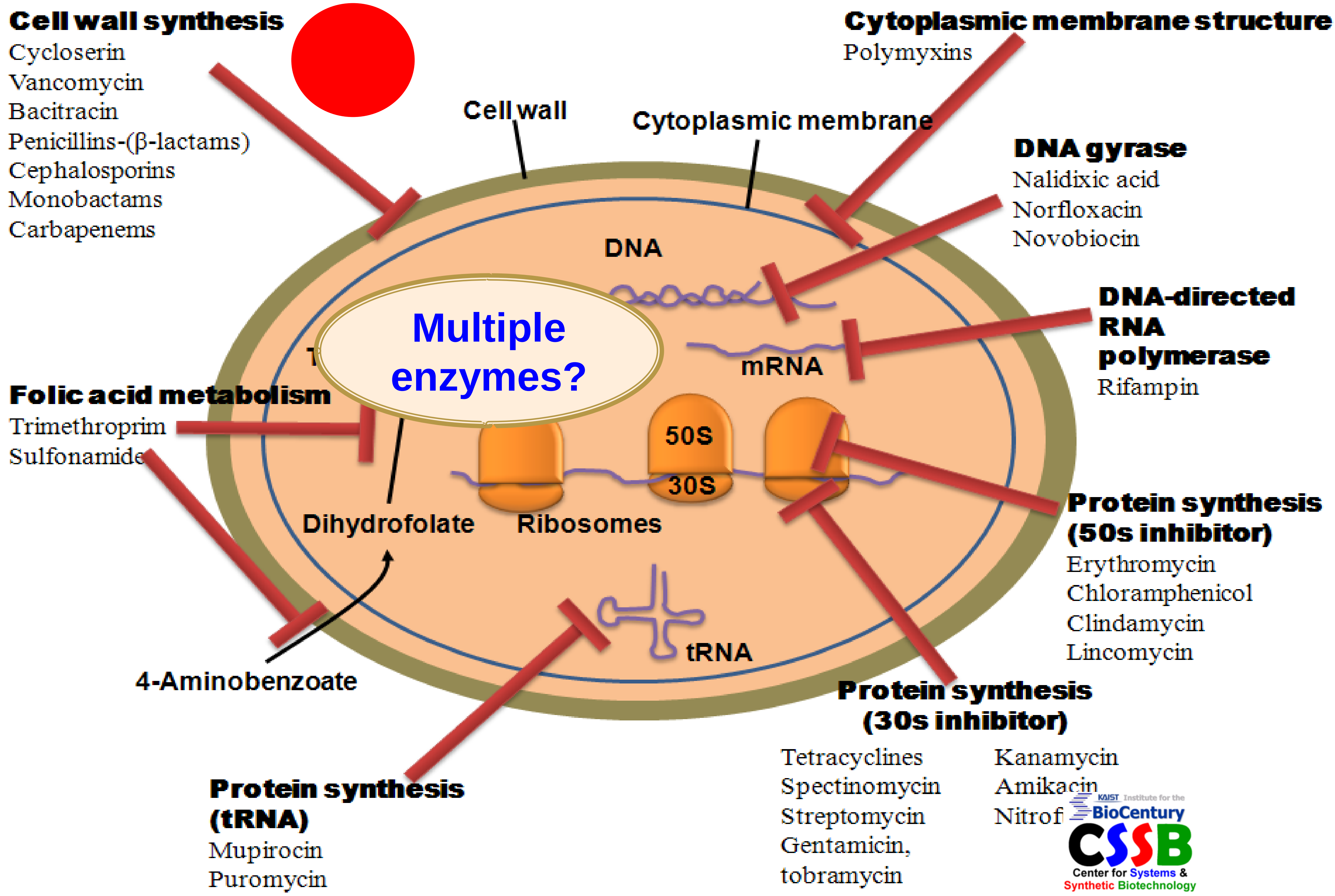
MRSA, VRSA



Acinetobacter baumanii

Increasing nosocomial pathogen

Increased after the Iraqi war



Multiple enzymes?

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

Cytoplasmic membrane structure
Polymyxins

DNA gyrase
Nalidixic acid
Norfloxacin
Novobiocin

DNA-directed RNA polymerase
Rifampin

Folic acid metabolism
Trimethroprim
Sulfonamide

Protein synthesis (50s inhibitor)
Erythromycin
Chloramphenicol
Clindamycin
Lincomycin

Protein synthesis (30s inhibitor)
Tetracyclines
Spectinomycin
Streptomycin
Gentamicin,
tobramycin
Kanamycin
Amikacin
Nitrof

Protein synthesis (tRNA)
Mupirocin
Puromycin

Microsoft Research

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

Breaking the robustness as a way of killing pathogens

Organisms can survive even with many of disrupted genes

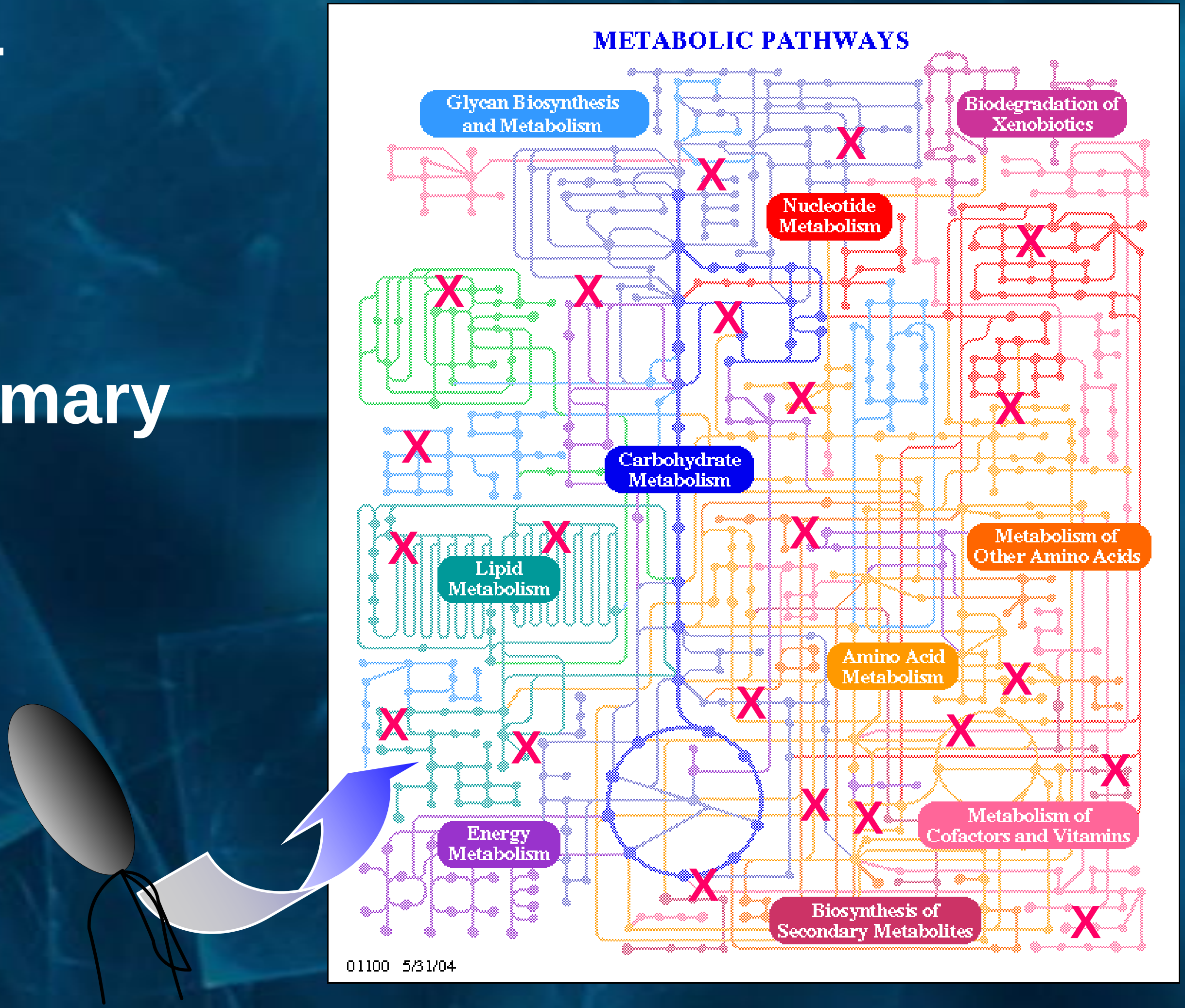
Use of Flux Balance Analysis

<u>Single gene deletion tests</u>

reaction flux **(*j*) = 0**

if biomass = 0

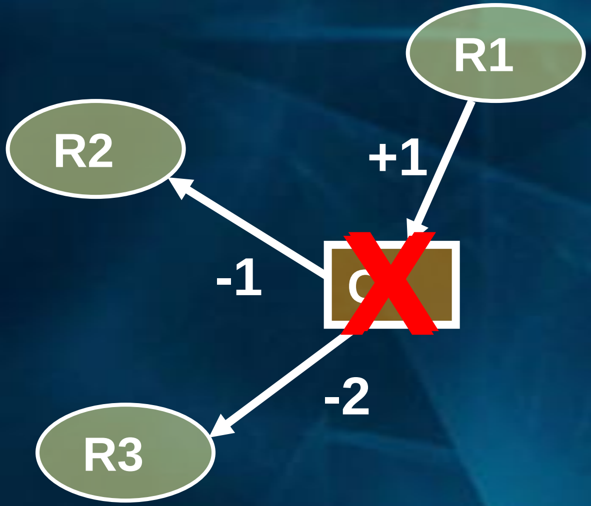
then, *reaction flux* (*j*) = primary drug target candidate

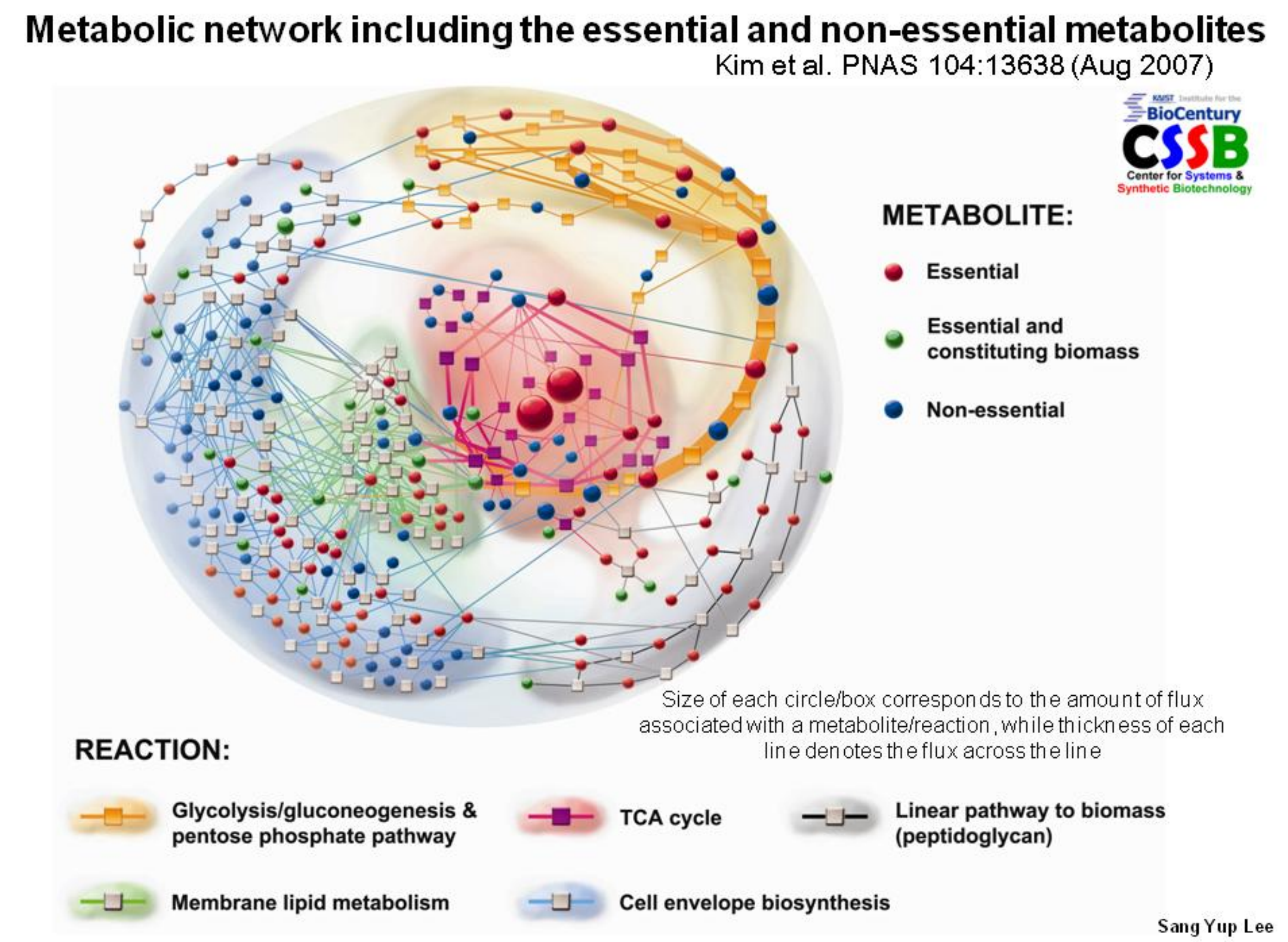


Metabolite Essentiality

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

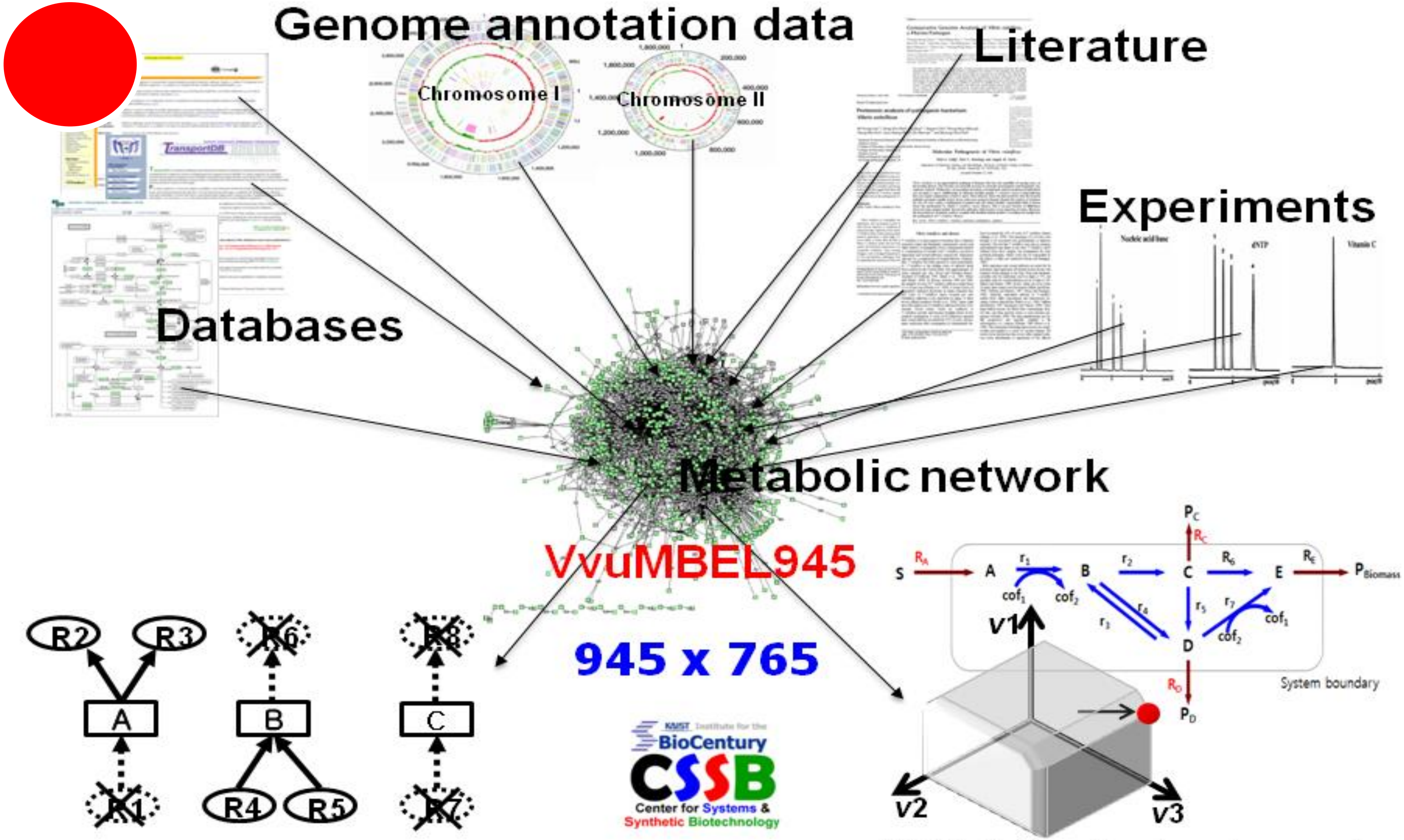
From gene knockout to metabolite knockout ?




Metabolic network including the essential and non-essential metabolites
Kim et al. PNAS 104:13638 (Aug 2007)

Size of each circle/box corresponds to the amount of flux associated with a metabolite/reaction, while thickness of each line denotes the flux across the line

Sang Yup Lee



Genome annotation data

Literature

Experiments

Databases

Metabolic network

VvuMBEL945

945 x 765

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

Network topology-based approach

Chokepoint analysis

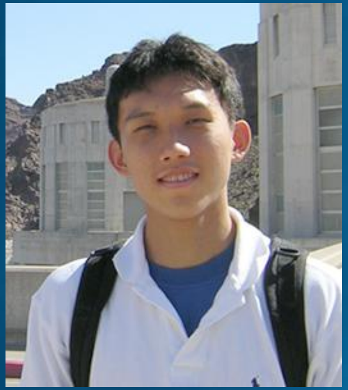
Metabolic function-based approach

Constraints-based flux analysis for single and multiple drug targeting

Sang Yup Lee

Microsoft Research

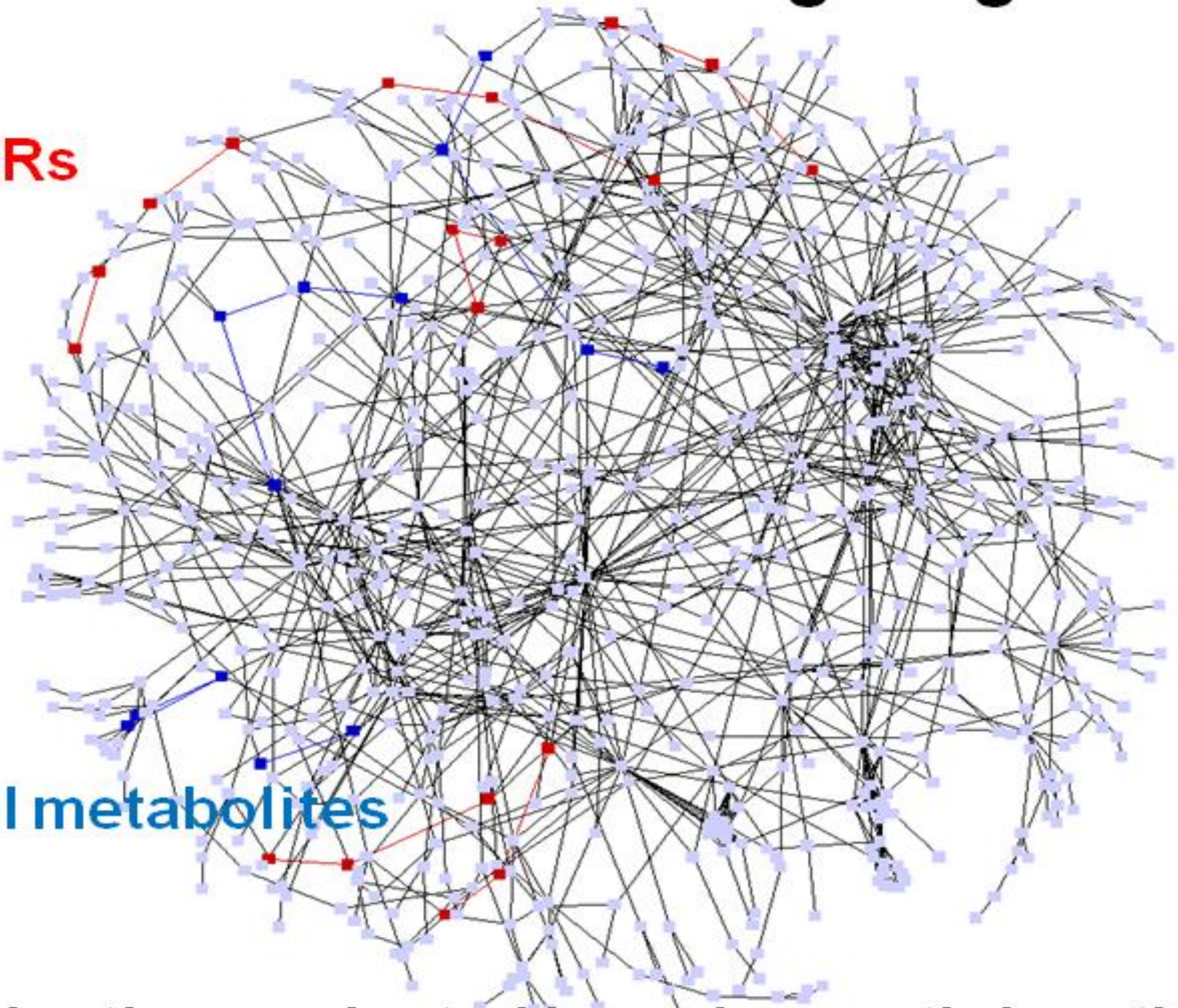
Tae Yong Kim



Hyun Uk Kim

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

Combinatorial drug targets



KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

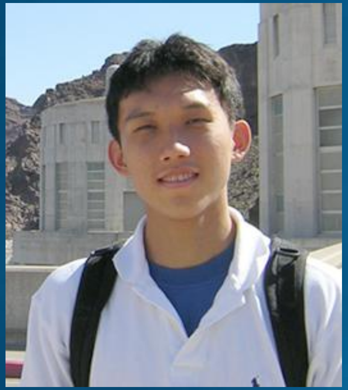
20 combinations selected based on path length

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

Sang Yup Lee

Microsoft® Research

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

Microsoft® Research

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

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

Microsoft Research

KAIST Institute for the BioCentury
CSSB
Center for Systems & Synthetic Biotechnology

http://mbel.kaist.ac.kr

Thank you !

Microsoft® Research

Microsoft®

Your potential. Our passion.™