Complex (Biological) Networks

Analyzing Metabolic Networks

Elhanan Borenstein

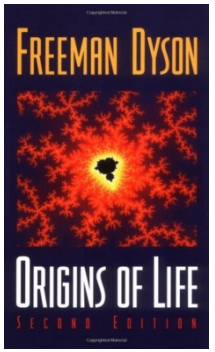
Some slides are based on slides from courses given by Roded Sharan and Tomer Shlomi

Metabolism

"Metabolism is the process involved in the maintenance of life. It is comprised of a vast repertoire of enzymatic reactions and transport processes used to convert thousands of organic compounds into the various molecules necessary to support cellular life"

Schilling et al. 2000

Why study metabolism? (II)



- **It's the essence of life**
  (and maybe its origins)
- Tremendous importance in Medicine
  - Inborn errors of metabolism cause acute symptoms
  - Metabolic diseases (obesity, diabetes) are on the rise (and are major sources of morbidity and mortality)
  - Metabolic enzymes becoming viable drug targets
- Bioengineering applications
  - Design strains for production of biological products
  - Generation of bio-fuels
- **The best understood of all cellular networks**

Metabolites & Biochemical Reactions

- *Metabolite*: an organic substance
  - Sugars (e.g., glucose, galactose, lactose)
  - Carbohydrates (e.g., glycogen, glucan)
  - Amino-acids (e.g., histidine, proline, methionine)
  - Nucleotides (e.g., cytosine, guanine)
  - Lipids
  - Chemical energy carriers (e.g., ATP, NADH)
  - Atoms (e.g., oxygen, hydrogen)
- *Biochemical reaction*: the process in which one or more substrate molecules are converted (usually with the help of an enzyme) to produce molecules

Pathways

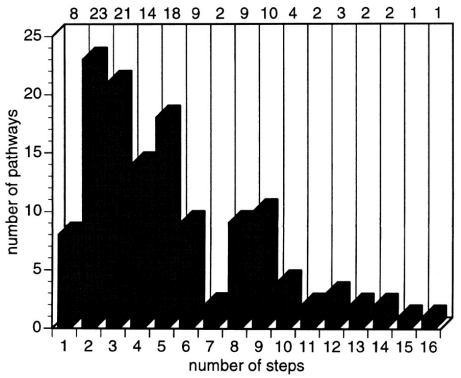
- EcoCyc describes 131 pathways

Table 2. List of All Known *E. coli* Metabolic Pathways as Described by EcoCyc

(Deoxy)ribose phosphate metabolism	Isoleucine biosynthesis
3-Phenylpropionate and 3-(3-hydroxyphenyl)propionate degradation	KDO biosynthesis
4-Aminobutyrate degradation	L-alanine degradation
Aerobic electron transfer	L-arabinose catabolism
Aerobic respiration, electron donors reaction list	L-cysteine catabolism
Alanine biosynthesis	L-lyxose metabolism
Anaerobic electron transfer	L-serine degradation
Anaerobic respiration	Lactose degradation
Anaerobic respiration, electron acceptors reaction list	Leucine biosynthesis
Anaerobic respiration, electron donors reaction list	Lipid A precursor biosynthesis
Arginine biosynthesis	Lysine and diaminopimelate biosynthesis
Asparagine biosynthesis and degradation	Mannitol degradation
Aspartate biosynthesis and degradation	Mannose and GDP-mannose metabolism
Betaine biosynthesis	Mannose catabolism
Biosynthesis of proto- and siroheme	Menaquinone biosynthesis
Biotin biosynthesis	Methionine biosynthesis
Carnitine metabolism	Methyl-donor molecule biosynthesis
Carnitine metabolism, CoA-linked	Methylglyoxal metabolism
Cobalamin biosynthesis	NAD phosphorylation and dephosphorylation
Colanic acid biosynthesis	Nonoxidative branch of the pentose phosphate pathway
Cyanate catabolism	Nucleotide metabolism
Cysteine biosynthesis	O-antigen biosynthesis
D-arabinose catabolism	Oxidative branch of the pentose phosphate pathway
D-galactarate catabolism	Pantothenate and coenzyme A biosynthesis
D-galacturonate catabolism	Peptidoglycan biosynthesis
D-glucarate catabolism	Phenylalanine biosynthesis
D-glucuronate catabolism	Phenylethylamine degradation
Degradation of short-chain fatty acids	Phosphatidic acid synthesis
Deoxypyrimidine nucleotide/side metabolism	Phospholipid biosynthesis
Deoxyribonucleotide metabolism	Polyamine biosynthesis
dTDP-rhamnose biosynthesis	Polyisoprenoid biosynthesis
Enterobacterial common antigen biosynthesis	ppGpp metabolism
Enterobactin synthesis	Proline biosynthesis
Entner-Doudoroff pathway	Proline utilization
Fatty acid biosynthesis, initial steps	Propionate metabolism, methylmalonyl pathway
Fatty acid elongation, saturated	Purine biosynthesis
Fatty acid elongation, unsaturated	Pyridine nucleotide cycling
Fatty acid oxidation pathway	Pyridine nucleotide synthesis
Fermentation	Pyridoxal 5'-phosphate biosynthesis
Folic acid biosynthesis	Pyridoxal 5'-phosphate salvage pathway
FormylTHF biosynthesis	Pyrimidine biosynthesis
Fucose catabolism	Pyrimidine ribonucleotide/ribonucleoside metabolism
Galactitol catabolism	Pyruvate dehydrogenase
Galactonate catabolism	Pyruvate oxidation pathway
Galactose metabolism	Removal of superoxide radicals
Galactose, galactoside and glucose catabolism	Rhamnose catabolism
Gluconeogenesis	Riboflavin, FMN and FAD biosynthesis
Glucosamine catabolism	Ribose catabolism
Glucose 1-phosphate metabolism	Serine biosynthesis
Glutamate biosynthesis	Sorbitol degradation
Glutamate utilization	Sulfate assimilation pathway
Glutamine biosynthesis	TCA cycle, aerobic respiration
Glutamine utilization	Thiamine biosynthesis
Glutathione biosynthesis	Thioredoxin pathway
Glutathione-glutaredoxin redox reactions	Threonine biosynthesis
Glycerol metabolism	Threonine catabolism
Glycine biosynthesis	Trehalose biosynthesis
Glycine cleavage	Trehalose degradation, low osmolarity
Glycogen biosynthesis	Tryptophan biosynthesis
Glycogen catabolism	Tryptophan utilization
Glycolate metabolism	Tyrosine biosynthesis
Glycolysis	Ubiquinone biosynthesis
Glyoxylate cycle	UDP-N-acetylglucosamine biosynthesis
Glyoxylate degradation	Valine biosynthesis
Histidine biosynthesis	Xylose catabolism
Histidine degradation	

The reactions and enzymes within each pathway can be determined using the EcoCyc WWW server that is available at http://ecocyc.DoubleTwist.com/ecocyc/.

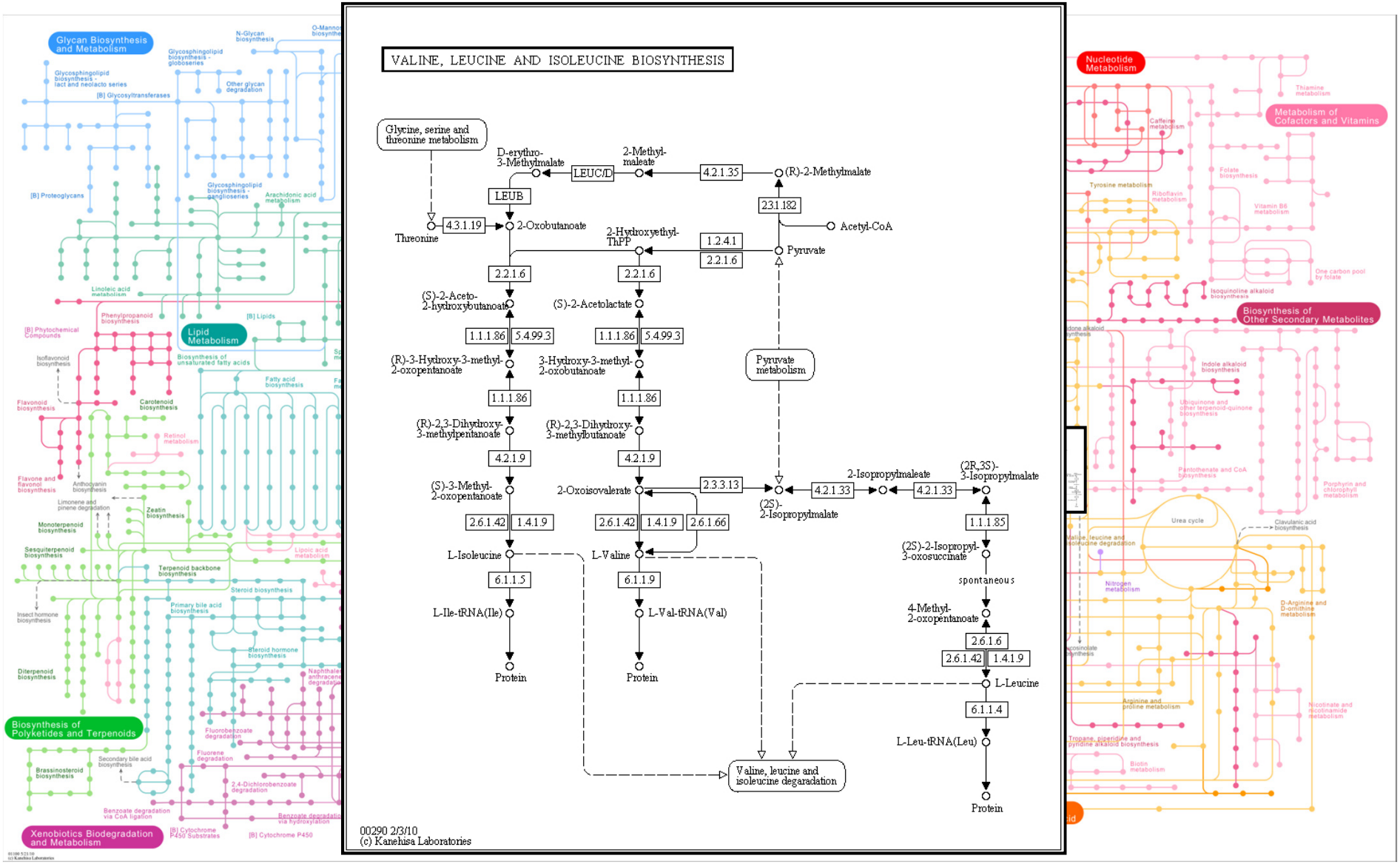
- Pathways vary in length from a single step to 16 steps (ave 5.4)
  - But … no precise biological definition and partitioning of the metabolic network into pathways is somehow arbitrary



Ouzonis, Karp, Genome Res. 10, 568 (2000)

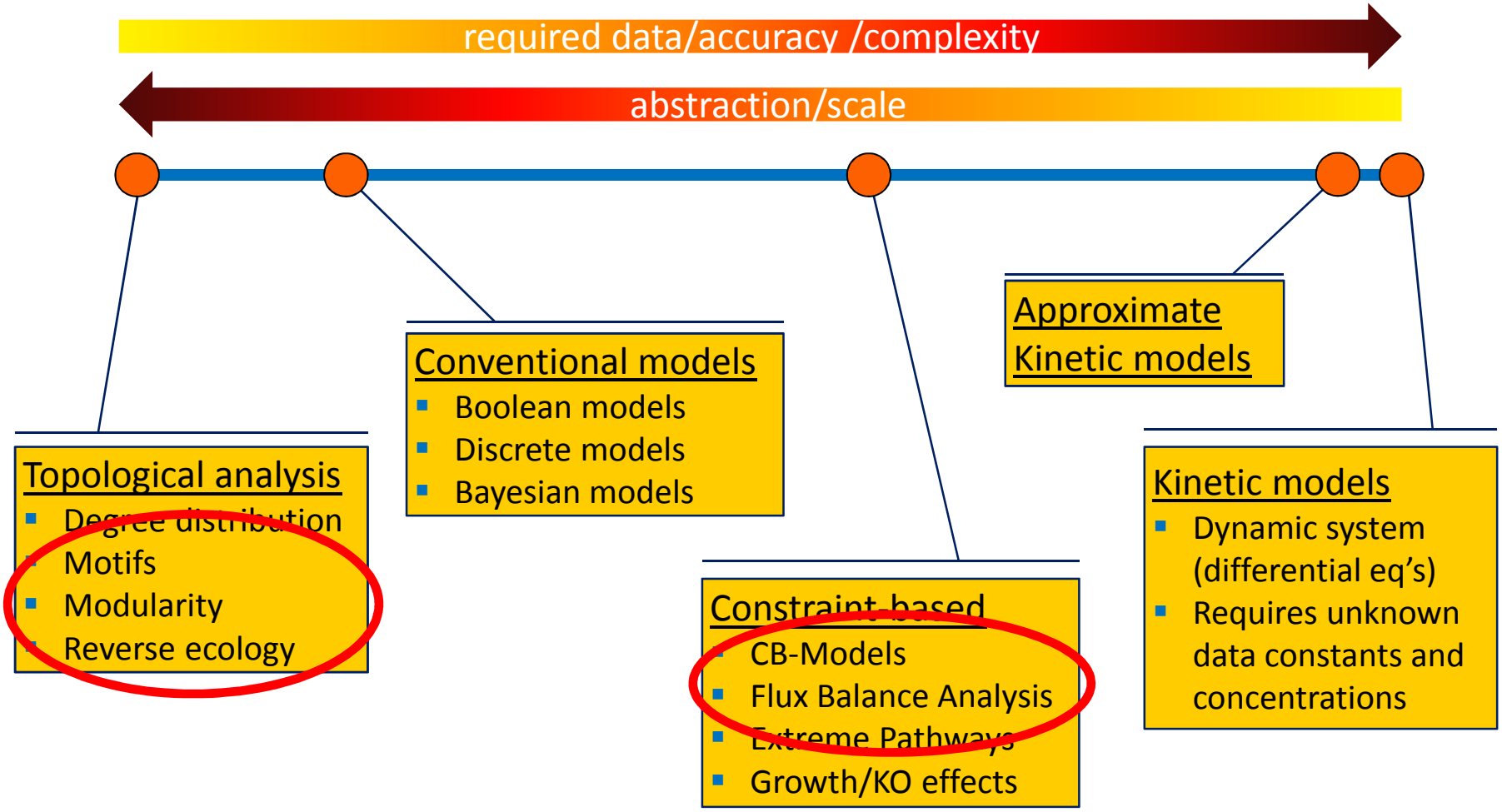
http://www.genome.jp/kegg/pathway/map/map01100.html

From Pathways to a Network



Models
of
Metabolism
(and Metabolic Networks)

Metabolic Network Models

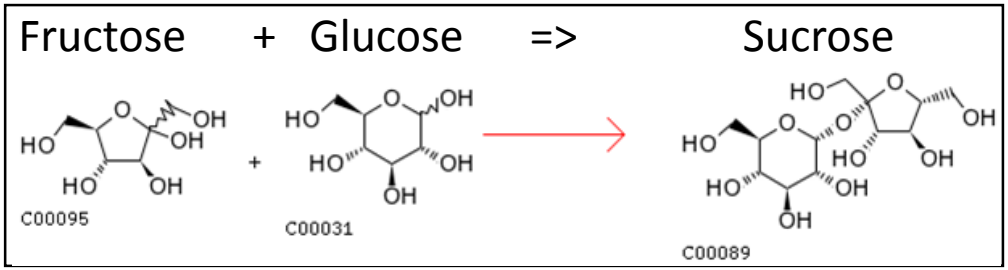


Reverse Ecology

Reconstructing Metabolic Networks

Describing the chemical reactions in the cell and the compounds being consumed and produced

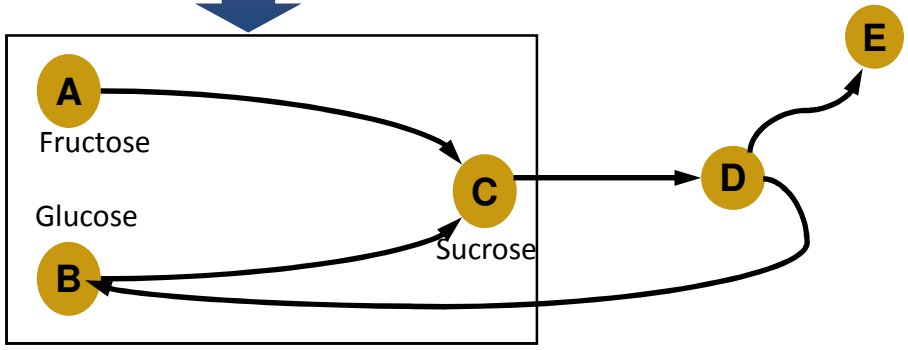
atgaaaaccgtcgttt
ttgcctaccacgatat
gggatgcctcggtatg



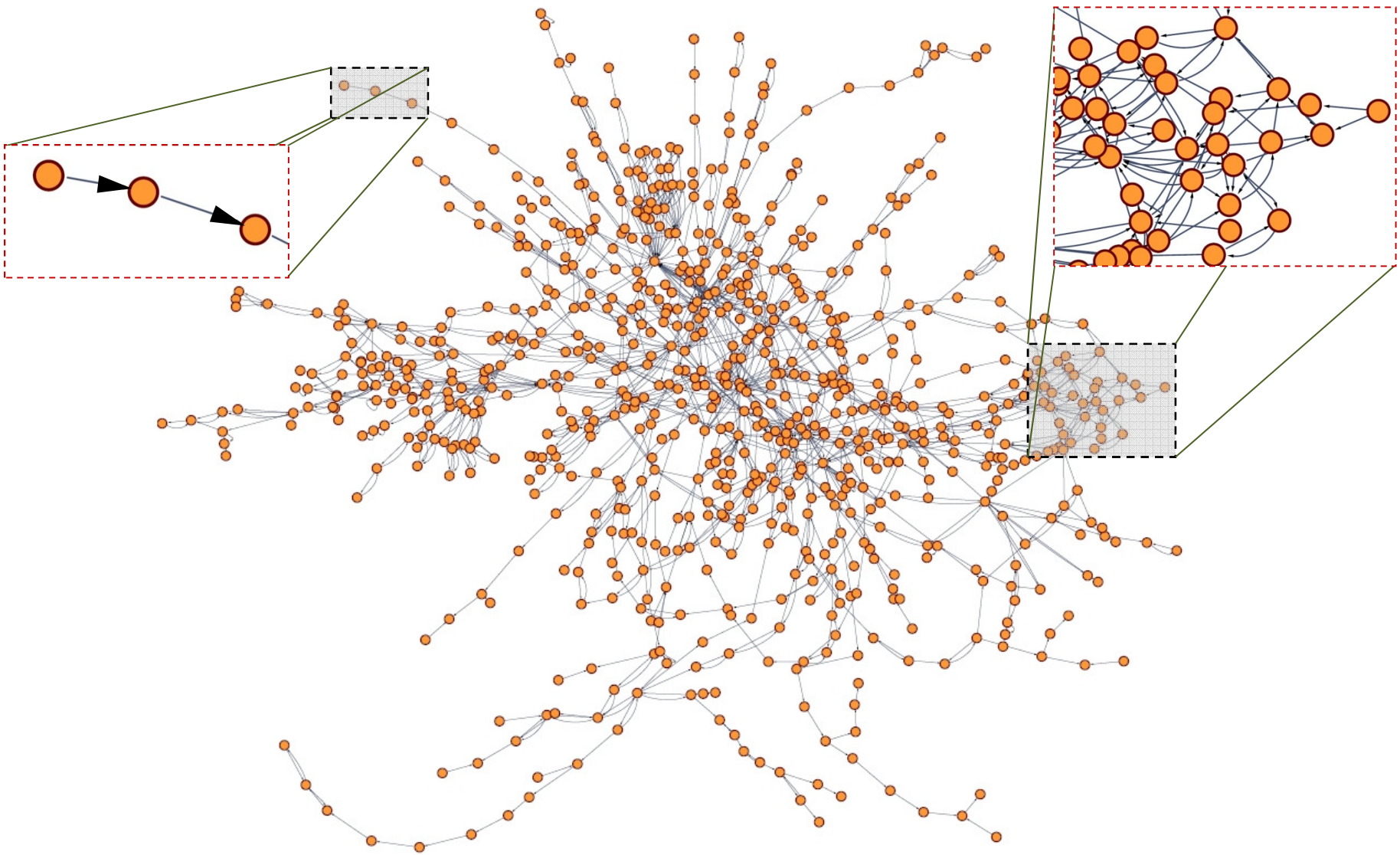
Simple Representation

- ***Nodes=compounds***
- ***Edges=reactions***
- Topology based
- Static

× Incomplete data
× Noise
✓ *Large-scale*
✓ Simple directed graphs

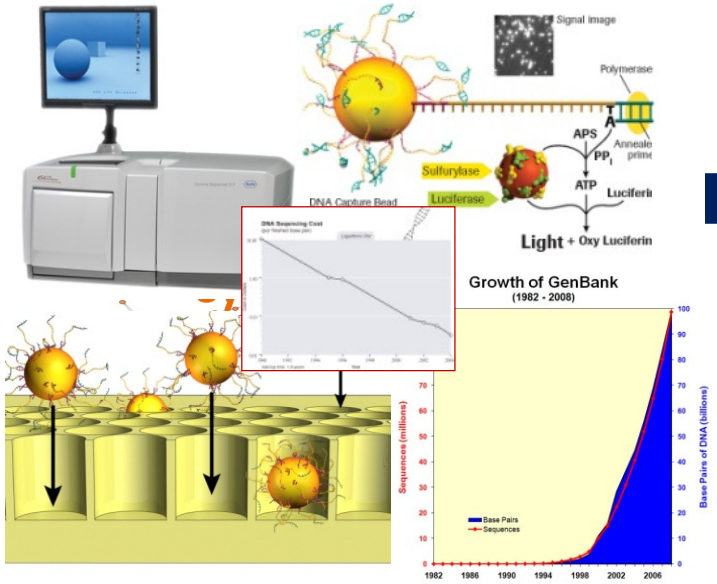


Metabolic Network (*E.Coli*)



Environments from Networks

Can **the structure/topology** of metabolic networks be used to obtain insights into the **ecology** in which species evolved/prevail?



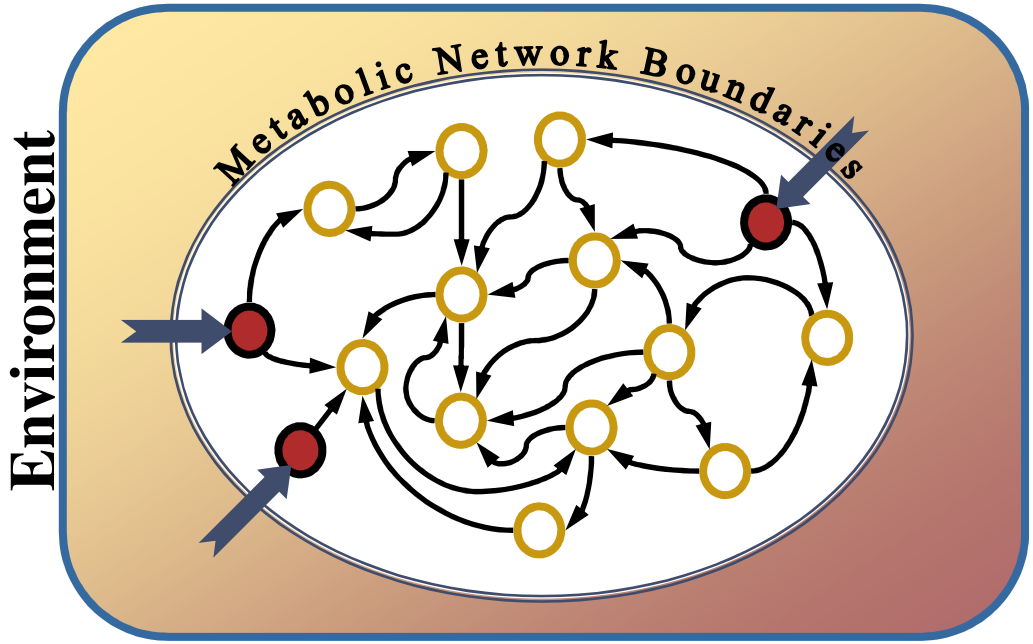
inference

Reverse Ecology of

Metabolic Environments

(Borenstein, et al.PNAS, 2008)

Seed Sets
& Metabolic Environments



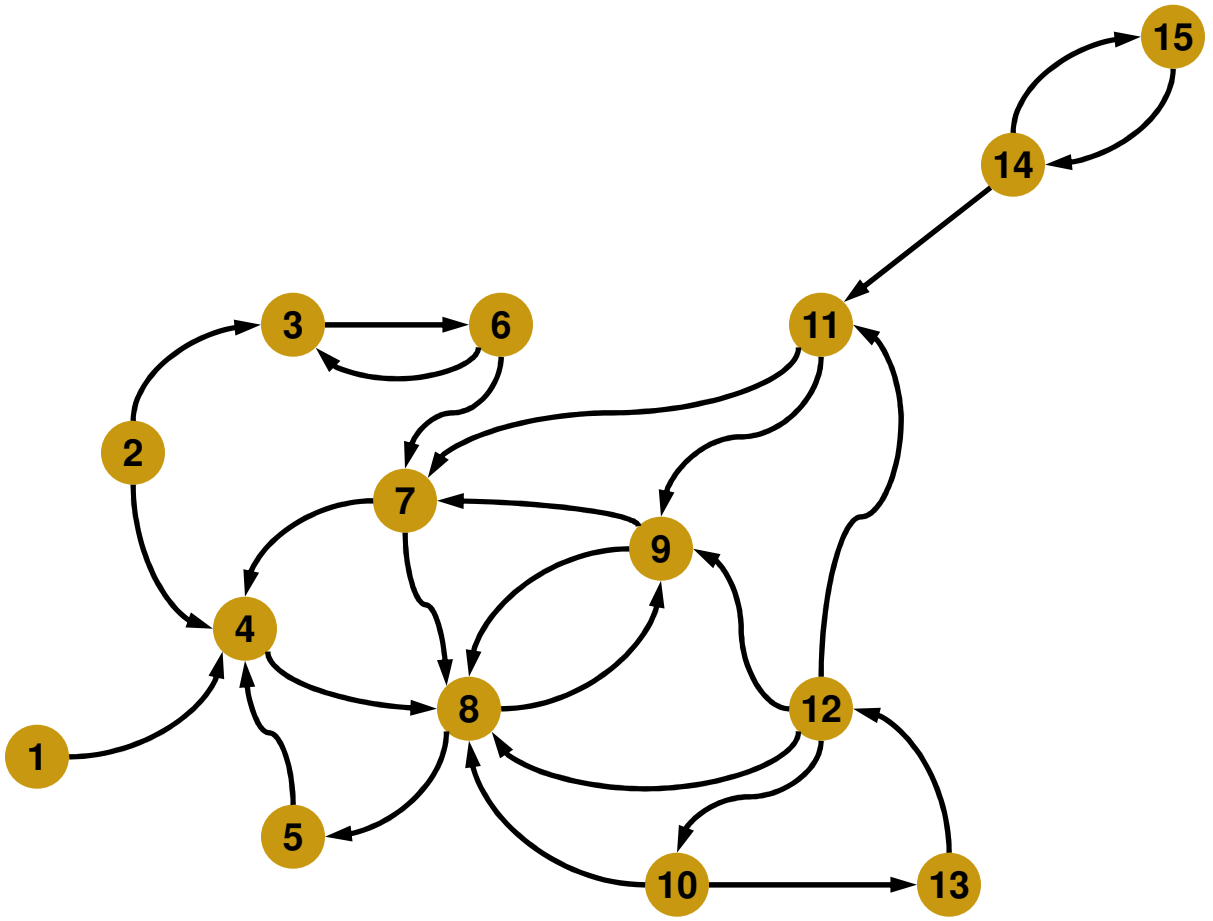
set of exogenously acquired compounds **(seed set)**

proxy for the environment
(operational definition)

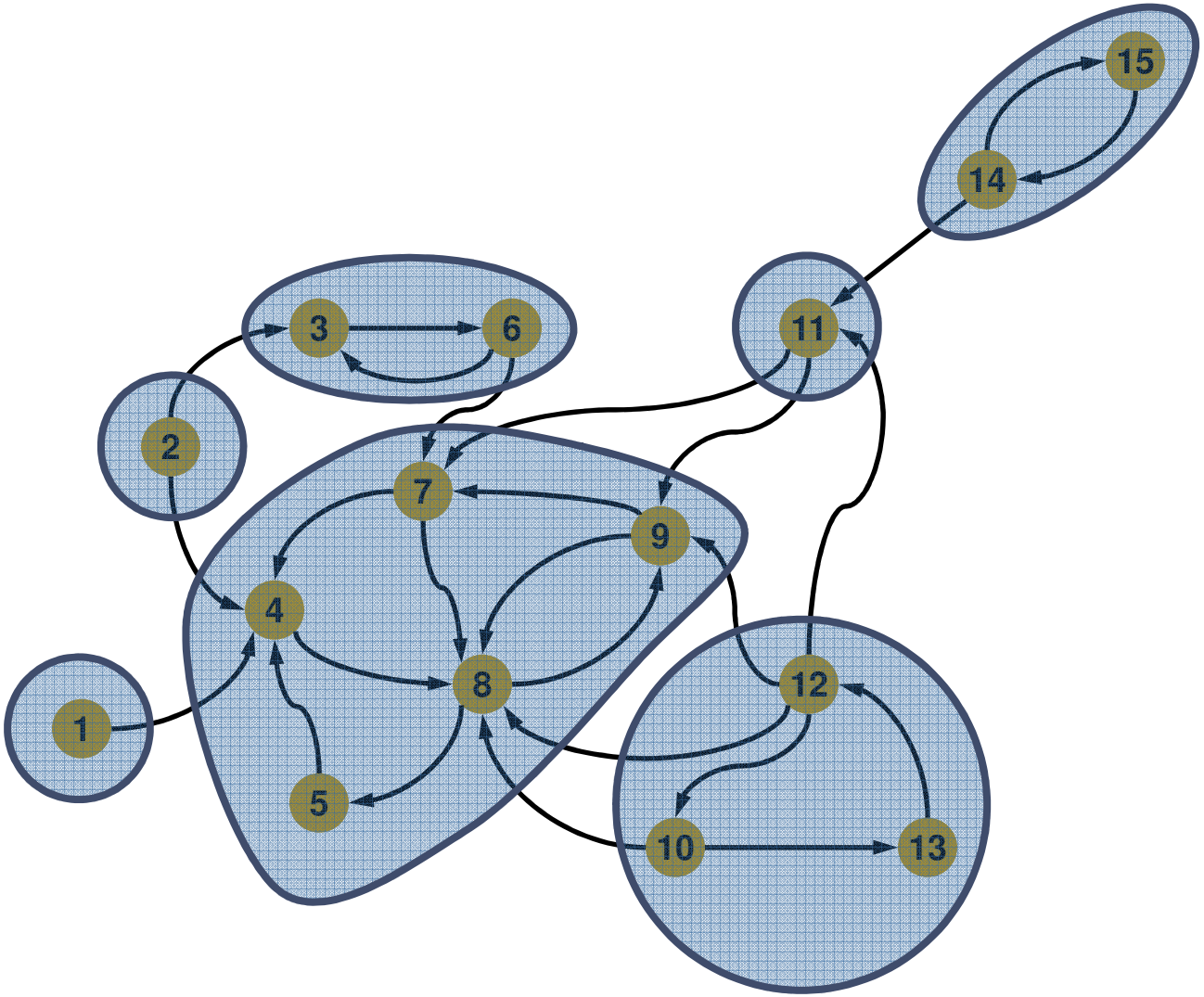
Seed set: *a minimal subset of the compounds that cannot be synthesized from other compounds and whose existence permits the synthesis of all other compounds in the network.*

(Borenstein, et al.PNAS, 2008)

Identifying Seed Compounds: A Simple Synthetic Example



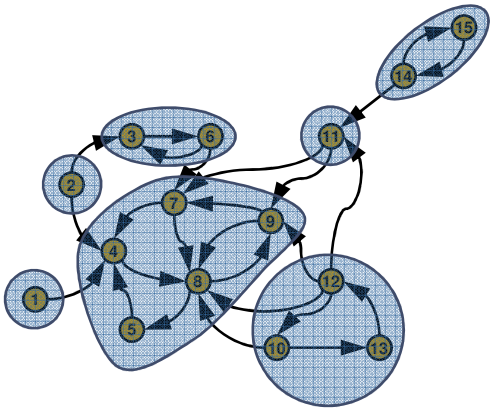
Identifying Seed Compounds: Strongly Connected Components (SCC)



Kosaraju's algorithm for SCC Decomposition

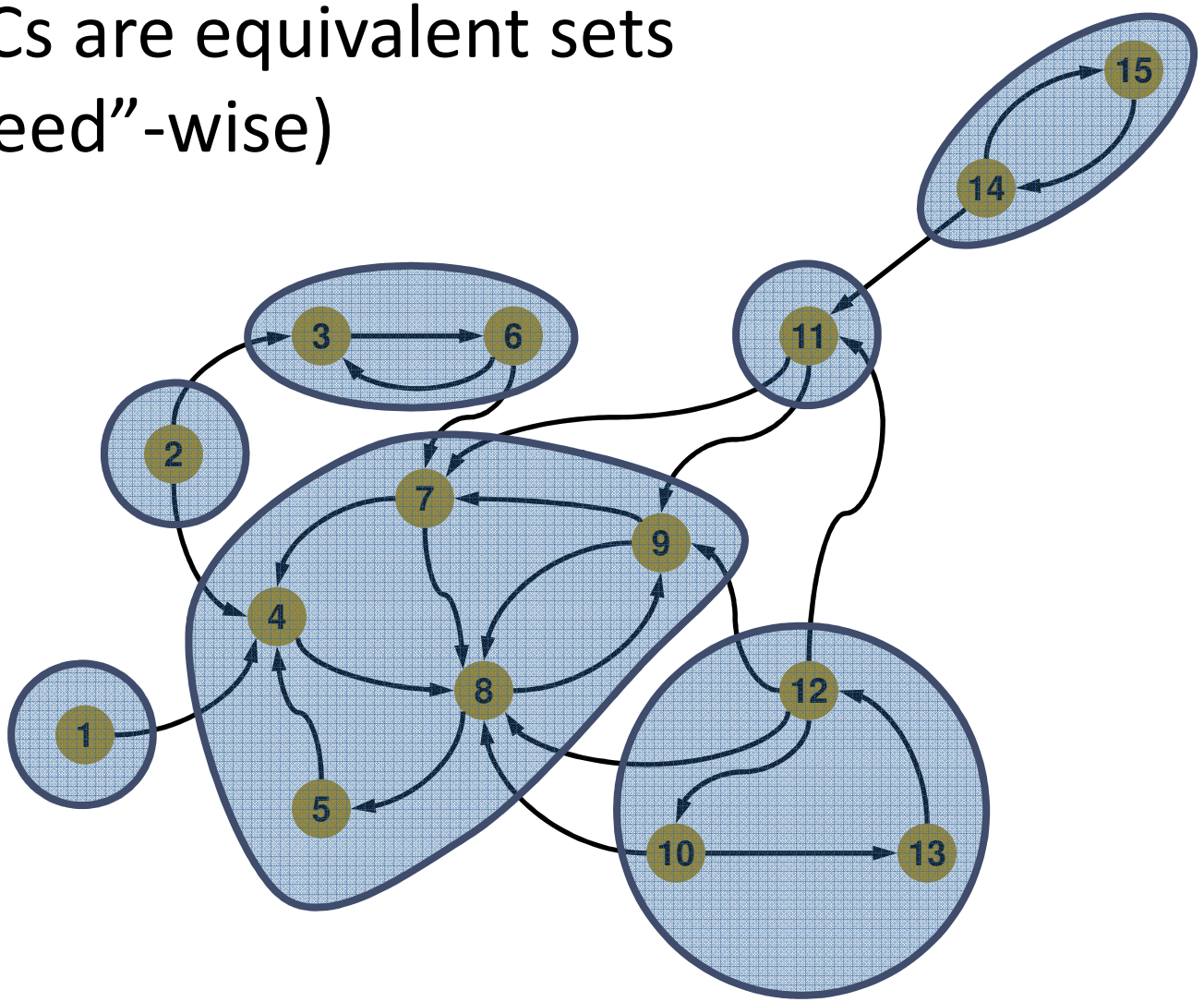
EvoSysBioLab

- Given a graph G:
  1. Run a **Depth-First Search (DFS)** on G to compute finishing times f[v] for each node v
  2. Calculate the **transposed network** G (the network G with the direction of every edge reversed)
  3. Run DFS on G, traversing the nodes in **decreasing** order of f[v]
- Each tree in the DFS forest created by the second DFS run forms a separate SCC



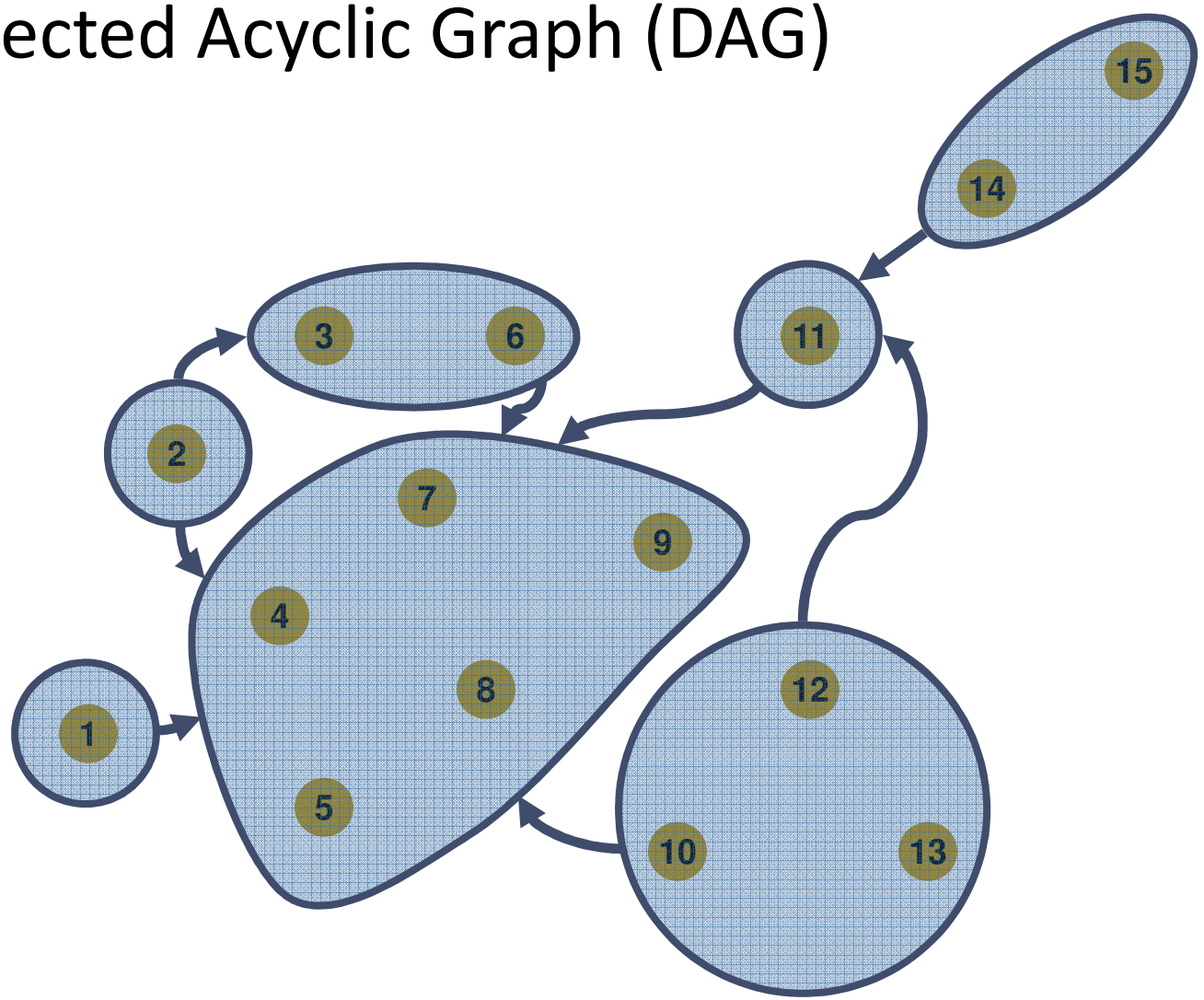
Identifying Seed Compounds: Strongly Connected Components (SCC)

- SCCs are equivalent sets ("seed"-wise)



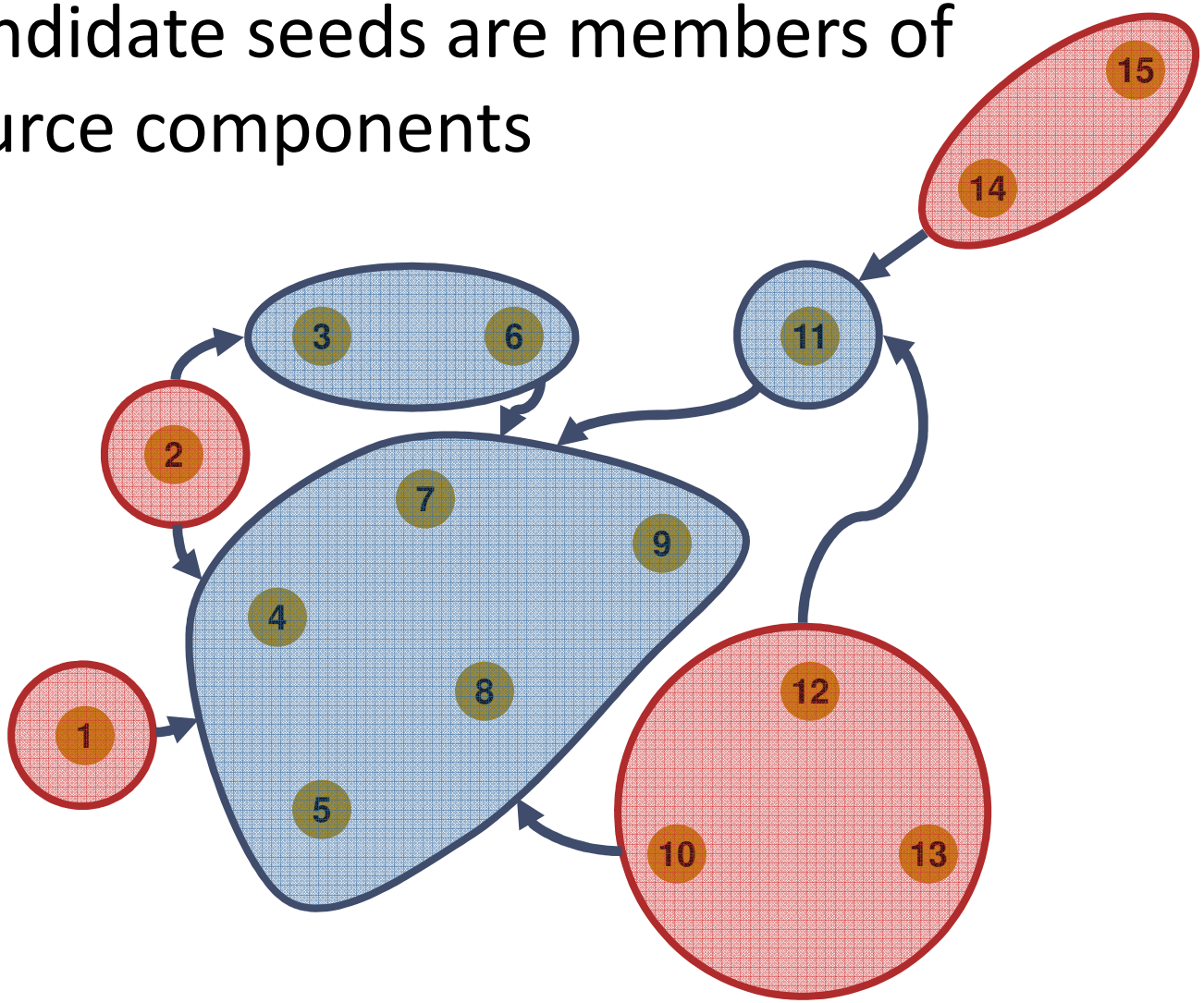
Identifying Seed Compounds: Strongly Connected Components (SCC)

- Directed Acyclic Graph (DAG)

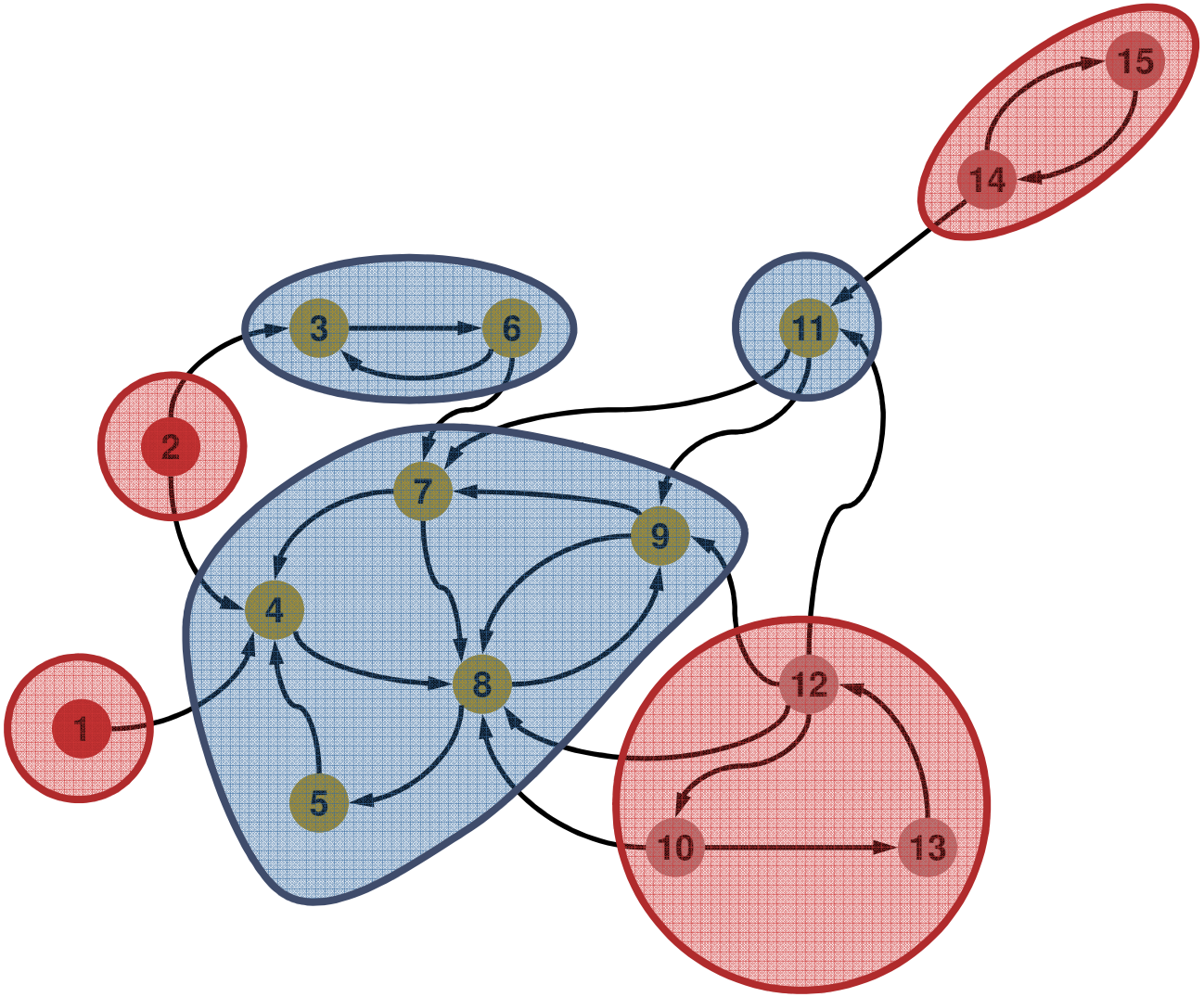


Identifying Seed Compounds: Source Components

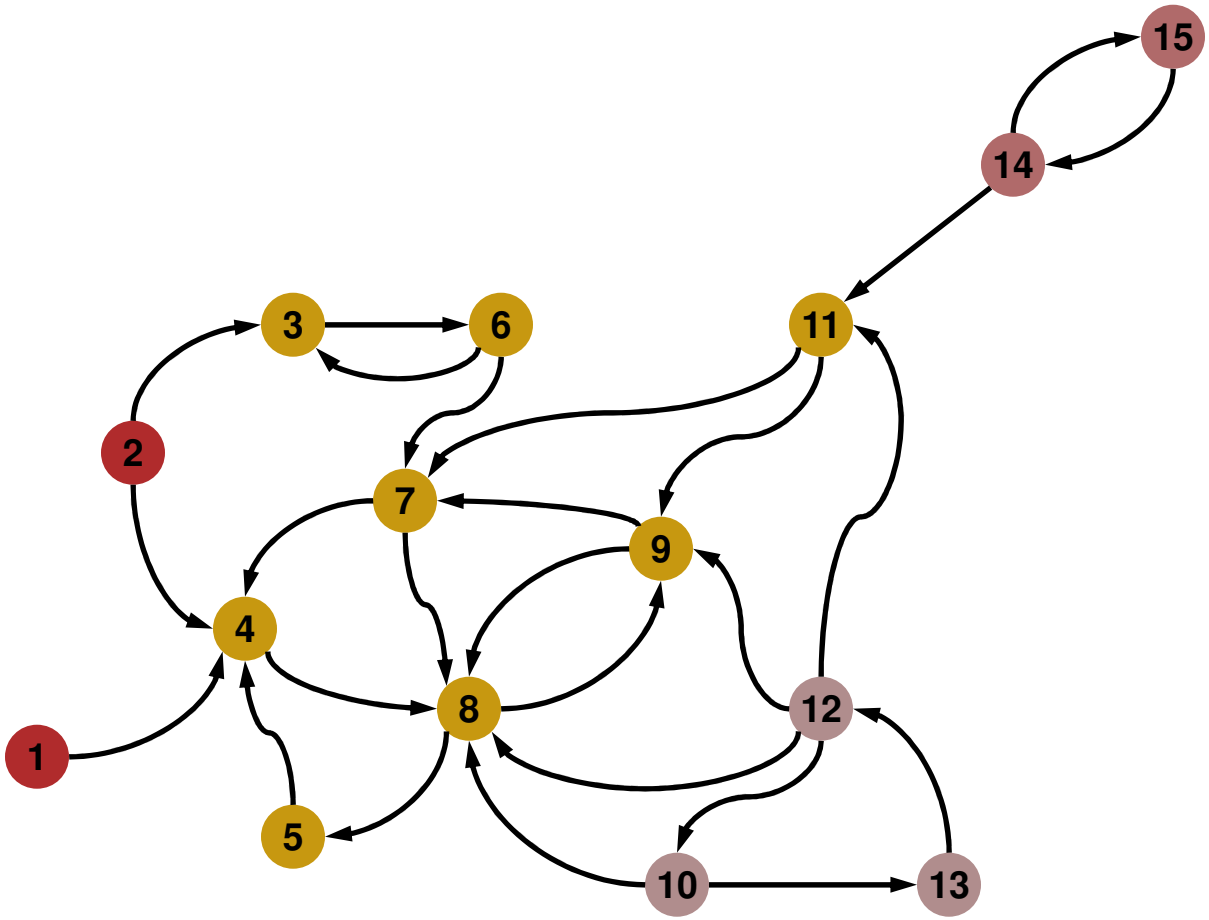
- Candidate seeds are members of source components



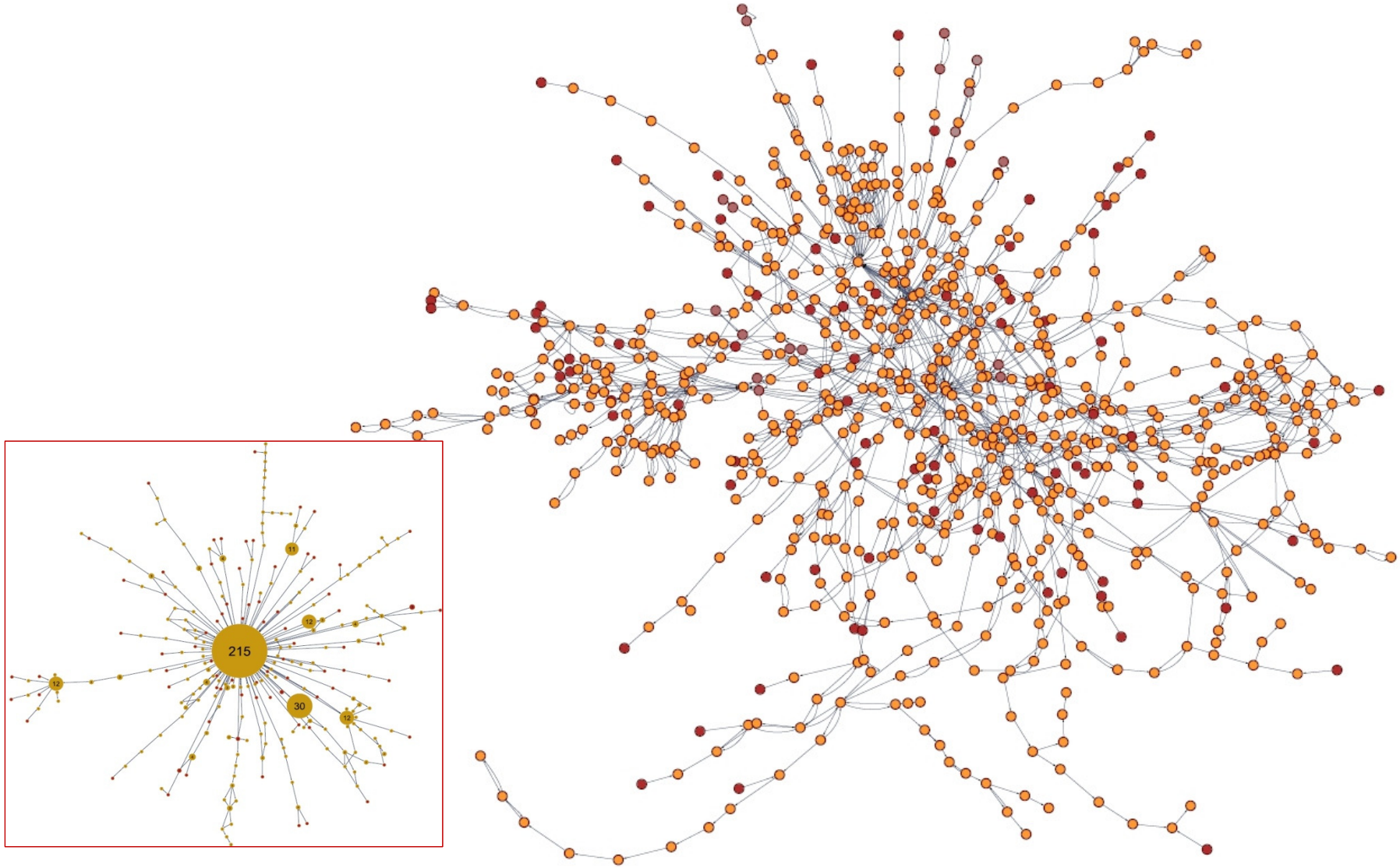
Identifying Seed Compounds: Candidate Seeds



Identifying Seed Compounds: Seed Confidence Level



Metabolic Network with Seeds

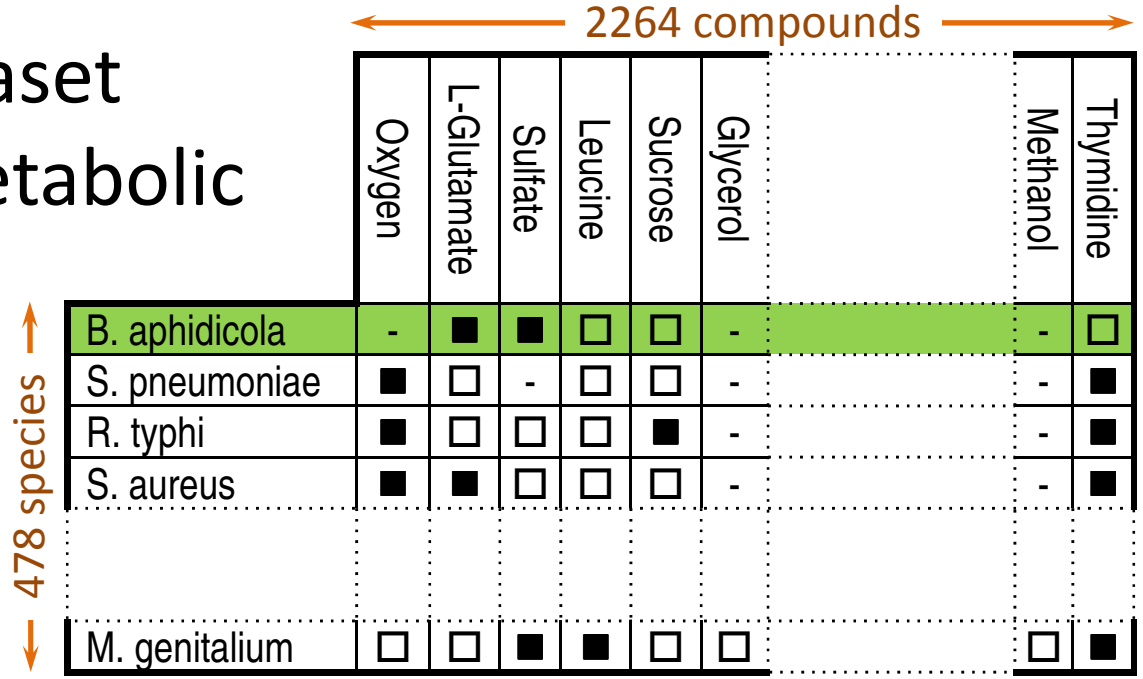


Multi-Species **Large-Scale** Seed Dataset

- **478** species (networks); >**2200** compounds
- Seed compounds for each species

- Large-scale dataset of predicted metabolic environments

accuracy 79%
precision 95%
recall 67%



Applications of Reverse Ecology

- Reconstructing ecology-based phylogeny
- Predicting ancestral environments
- Identifying evolutionary dynamics of networks
- Predicting species interaction
- Analyzing genetic vs. environmental robustness
- Quantifying ecological strategies

Constraint-Based Modeling

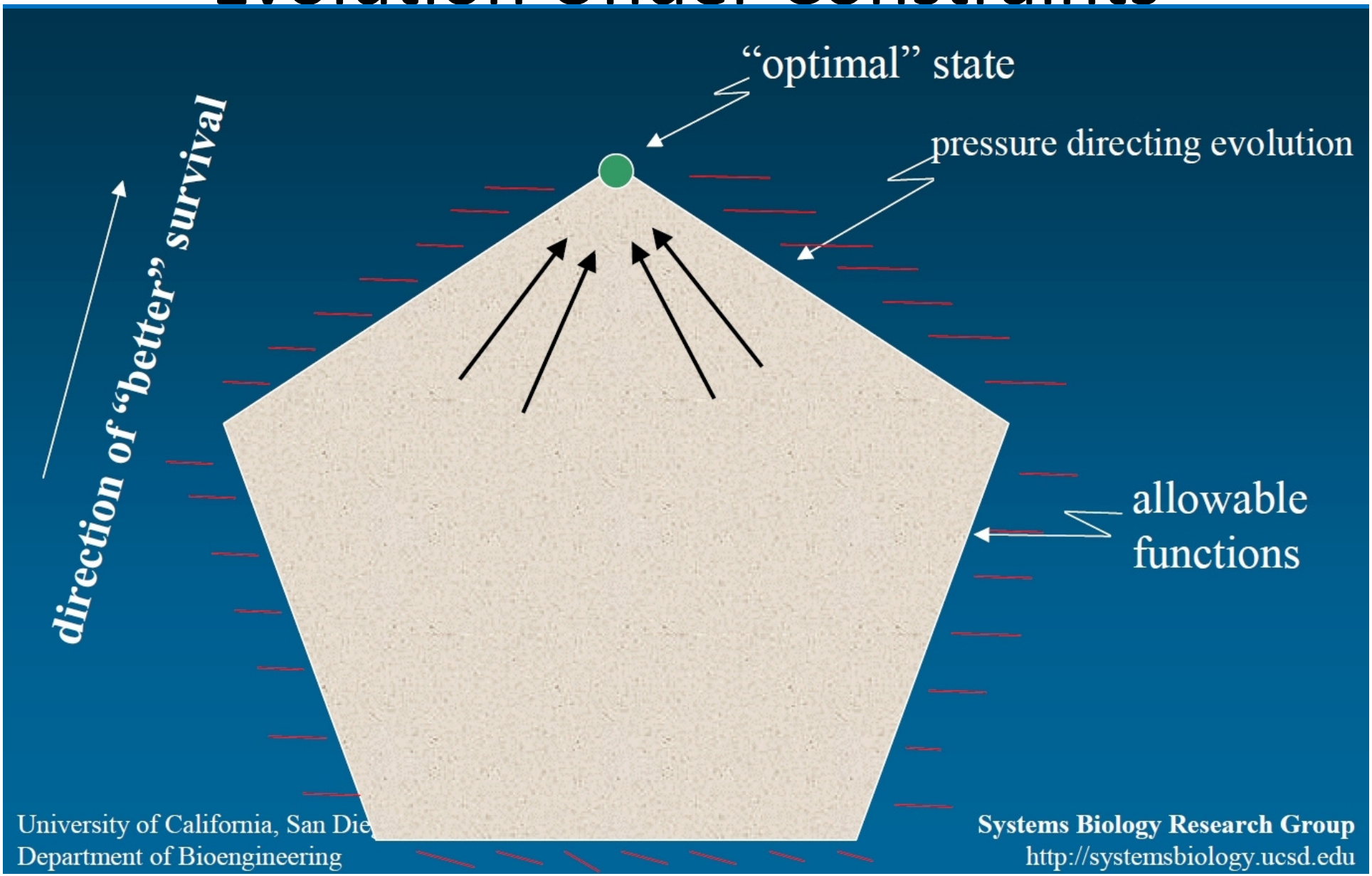
Constraint-Based Modeling

- Living systems obey physical and chemical laws
- These can be used to constrain the space of possible behaviors of the network

How often have I said to you that when you have eliminated the impossible, whatever remains, however improbable, must be the truth?

– Sherlock Holmes (A Study of Scarlet)

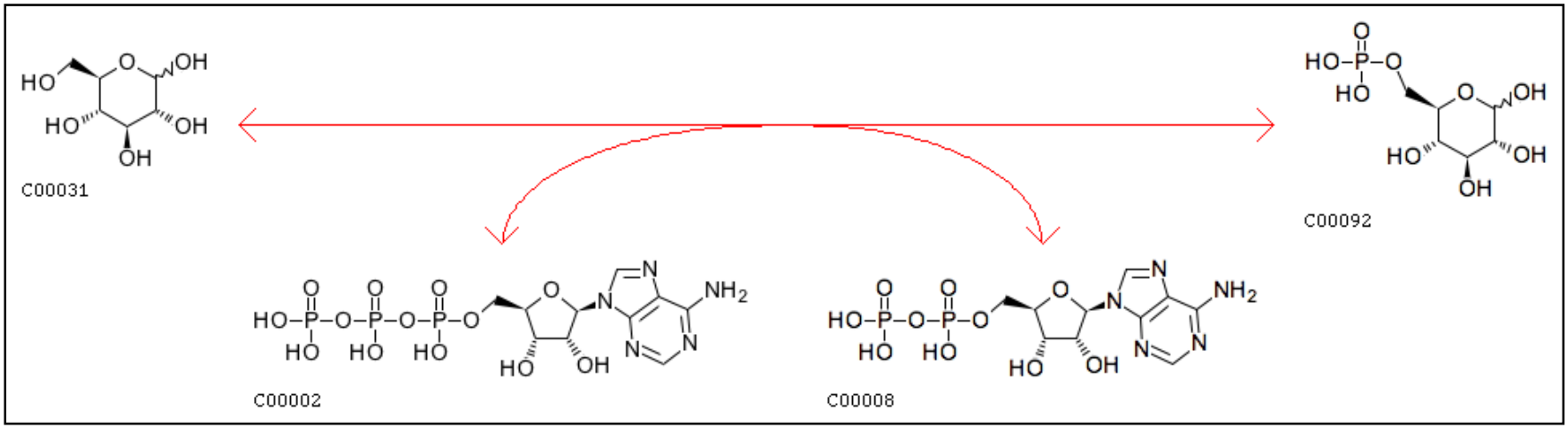
Evolution Under Constraints



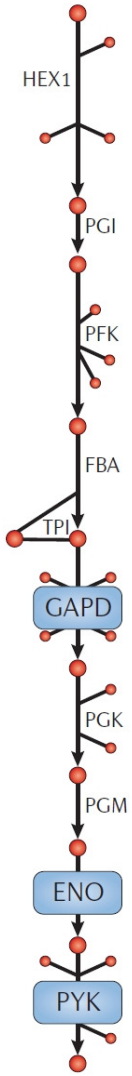
Reaction Stoichiometry

- **Stoichiometry - the quantitative relationships of the reactants and products in reactions**

1 Glucose + 1 ATP ⟷ 1 Glucose-6-Phosphate + 1 ADP



Stoichiometric Matrix

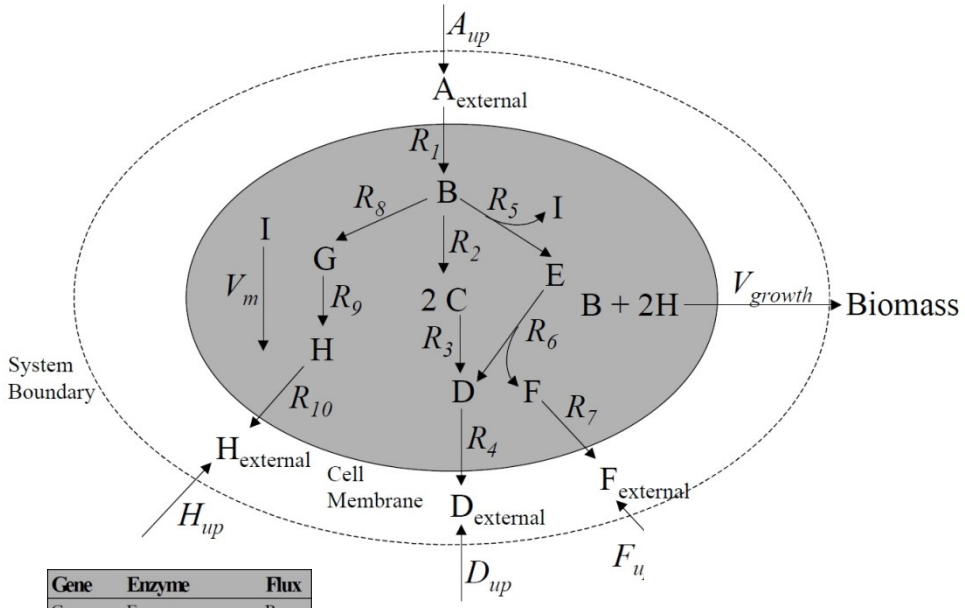


Abbreviation	Glycolytic reactions	Genes
HEX1	[c]GLC + ATP → G6P + ADP + H	*glk*
PGI	[c]G6P ↔ F6P	*pgi*
PFK	[c]ATP + F6P → ADP + FDP + H	*pfkA*, *pfkB*
FBA	[c]FDP ↔ DHAP + G3P	*fbaA*, *fbaB*
TPI	[c]DHAP ↔ G3P	*tpiA*
GAPD	[c]G3P + NAD + PI ↔ 13DPG + H + NADH	*gapA*, *gapC1*, *gapC2*
PGK	[c]13DPG + ADP ↔ 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

S

Stoichiometric Matrix and Fluxes



Gene	Enzyme	Flux
$Gene_1$	$Enzyme_1$	R_1
$Gene_2$	$Enzyme_2$	R_2
$Gene_3$	$Enzyme_3$	R_3
$Gene_4$	$Enzyme_4$	R_4
$Gene_5$	$Enzyme_5$	R_5
$Gene_6$	$Enzyme_6$	R_6
$Gene_7$	$Enzyme_7$	R_7
$Gene_8$	$Enzyme_8$	R_8
$Gene_9$	$Enzyme_9$	R_9
$Gene_{10}$	$Enzyme_{10}$	R_{10}
$Gene_A$	A Transporter	A_{up}
$Gene_D$	D Transporter	D_{up}
$Gene_F$	F Transporter	F_{up}
$Gene_H$	H Transporter	H_{up}

$$\frac{d\bar{m}}{dt} = S \cdot \bar{v}$$

- **m**: *metabolite concentrations vector (mol/mg)*
- **S**: *stoichiometric matrix*
- **v**: *reaction rates vector*

A Full Model? Not Really

A set of Ordinary Differential Equations (ODE)

$$\frac{d\overline{m}}{dt} = S \cdot \overline{v} = S \cdot f(\overline{m}, k)$$

Reaction rate equation **Kinetic parameters**

Requires knowledge of m, f and k!

Constraint-Based Modeling

- Assumes a quasi steady-state!
  - No changes in metabolite concentrations
  - Metabolite production and consumption rates are equal

$$\frac{d\bar{m}}{dt} = S \cdot \bar{v} = 0$$

	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	V_m	V_{growth}	A_{up}	D_{up}	F_{up}	H_{up}			
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R_1		0
B	1	−1	0	0	−1	0	0	−1	0	0	0	−1	0	0	0	0	R_2		0
C	0	2	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	R_3		0
D	0	0	1	−1	0	1	0	0	0	0	0	0	0	0	0	0	R_4		0
E	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0	0	R_5		0
F	0	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0	R_6		0
G	0	0	0	0	0	0	0	1	−1	0	0	0	0	0	0	0	R_7		0
H	0	0	0	0	0	0	0	0	1	−1	0	−2	0	0	0	0	R_8		0
I	0	0	0	0	1	0	0	0	0	0	−1	0	0	0	0	0	R_9		0
$\mathbf{A}_{external}$	−1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	R_{10}		0
$\mathbf{D}_{external}$	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	V_m		0
$\mathbf{F}_{external}$	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	V_{growth}		0
$\mathbf{H}_{external}$	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	A_{up}		0
																	D_{up}		0
																	F_{up}		0
																	H_{up}		0

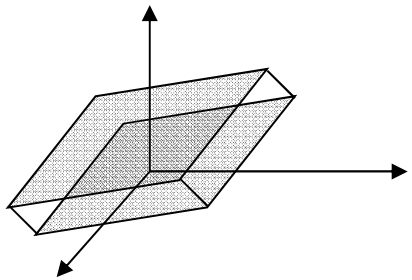
- No need for info on metabolite concentrations, reaction rate functions, or kinetic parameters

Constraint-Based Modeling

- In most cases, *S* is underdetermined:
  → a subspace of R^n (possible flux distributions)

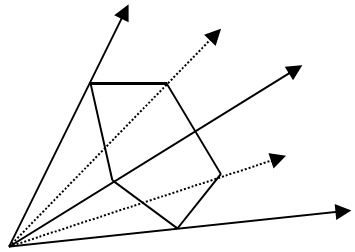
$$\begin{pmatrix}0\\0\\0\end{pmatrix} = \underbrace{\begin{bmatrix}-1 & -1 & 1 & 0 & 1 & 0 & 0\\ 1 & 0 & 0 & 1 & 0 & -1 & 0\\ 0 & 1 & -1 & -1 & 0 & 0 & -1\end{bmatrix}}_{S} \begin{bmatrix}v_1\\v_2\\v_3\\v_4\\b_1\\b_2\\b_3\end{bmatrix} \quad v$$

$S \cdot v = 0$



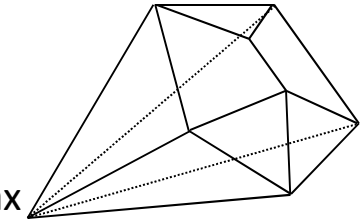
- Thermodynamic constraints:
  → a convex cone

$v_i > 0$



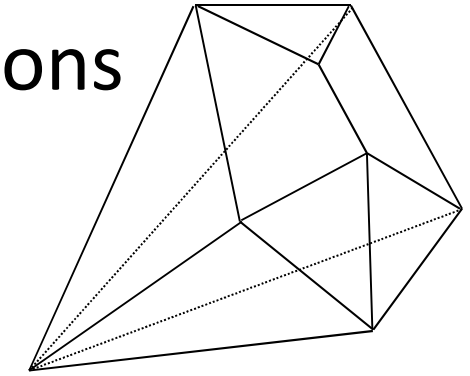
- Capacity constraints:
  → a bounded convex cone

$v_i < v_{max}$



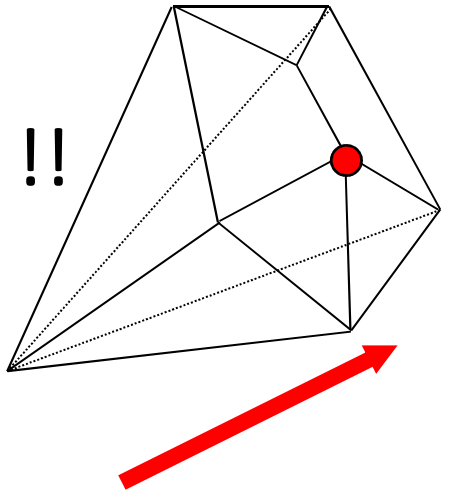
Flux Balance Analysis

- But this still leaves a space of solutions

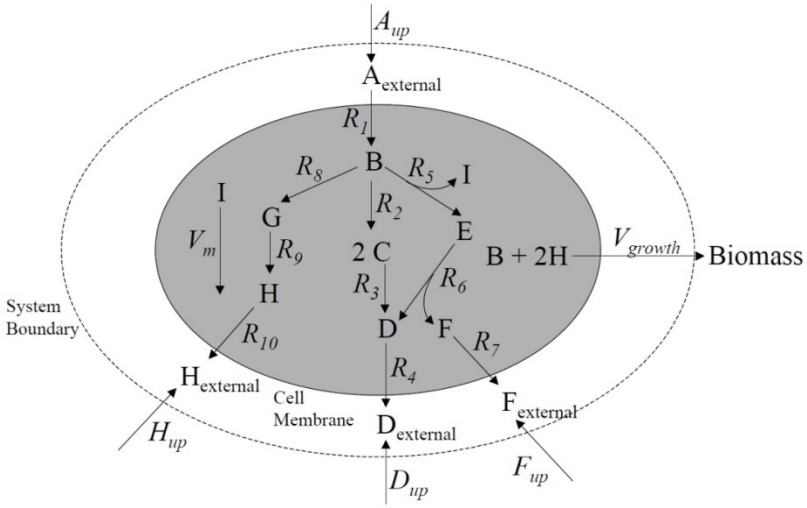


- How can we identify plausible solutions within this space?

- Optimize for maximum growth rate !!



Flux Balance Analysis



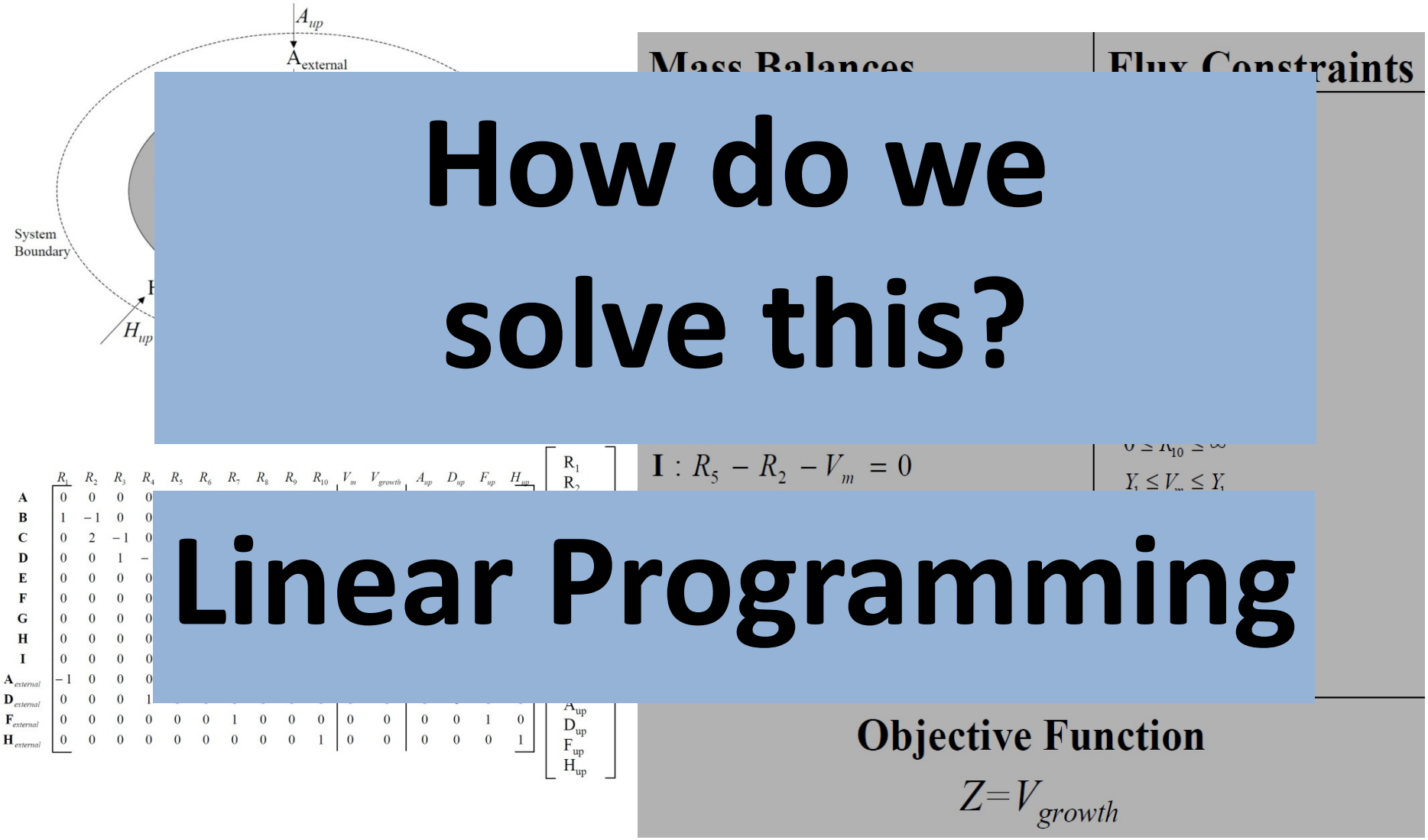
	R_1	R_2	R_3	R_4	R_5	R_6	R_7	R_8	R_9	R_{10}	V_m	V_{growth}	A_{up}	D_{up}	F_{up}	H_{up}	
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	R_1
B	1	−1	0	0	−1	0	0	−1	0	0	0	−1	0	0	0	0	R_2
C	0	2	−1	0	0	0	0	0	0	0	0	0	0	0	0	0	R_3
D	0	0	1	−1	0	1	0	0	0	0	0	0	0	0	0	0	R_4
E	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0	0	R_5
F	0	0	0	0	0	1	−1	0	0	0	0	0	0	0	0	0	R_6
G	0	0	0	0	0	0	0	1	−1	0	0	0	0	0	0	0	R_7
H	0	0	0	0	0	0	0	0	1	−1	0	−2	0	0	0	0	R_8
I	0	0	0	0	1	0	0	0	0	0	−1	0	0	0	0	0	R_9
$\mathbf{A}_{external}$	−1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	R_{10}
$\mathbf{D}_{external}$	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	V_m
$\mathbf{F}_{external}$	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	V_{growth}
$\mathbf{H}_{external}$	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	A_{up}
																	D_{up}
																	F_{up}
																	H_{up}

Mass Balances	Flux Constraints
$\mathbf{B}: R_1 - R_2 - R_5 - R_8 - V_{growth} = 0$	$0 \le R_1 \le \infty$
	$0 \le R_2 \le \infty$
$\mathbf{C}: 2R_2 - R_3 = 0$	$0 \le R_3 \le \infty$
$\mathbf{D}: R_3 + R_6 - R_4 = 0$	$0 \le R_4 \le \infty$
$\mathbf{E}: R_5 - R_6 = 0$	$0 \le R_5 \le \infty$
	$0 \le R_6 \le \infty$
$\mathbf{F}: R_6 - R_7 = 0$	$0 \le R_7 \le \infty$
$\mathbf{G}: R_8 - R_9 = 0$	$0 \le R_8 \le \infty$
$\mathbf{H}: R_9 - R_{10} - 2V_{growth} = 0$	$0 \le R_9 \le \infty$
	$0 \le R_{10} \le \infty$
$\mathbf{I}: R_5 - R_2 - V_m = 0$	$Y_1 \le V_m \le Y_1$
$\mathbf{A}_{external}: A_{up} - R_1 = 0$	$0 \le V_{growth} \le \infty$
$\mathbf{D}_{external}: D_{up} + R_4 = 0$	$Y_2 \le A_{up} \le Y_2$
$\mathbf{F}_{external}: F_{up} + R_7 = 0$	$-\infty \le D_{up} \le 0$
$\mathbf{H}_{external}: H_{up} + R_{10} = 0$	$-\infty \le F_{up} \le 0$
	$-\infty \le H_{up} \le 0$

Objective Function

$$Z = V_{growth}$$

Flux Balance Analysis

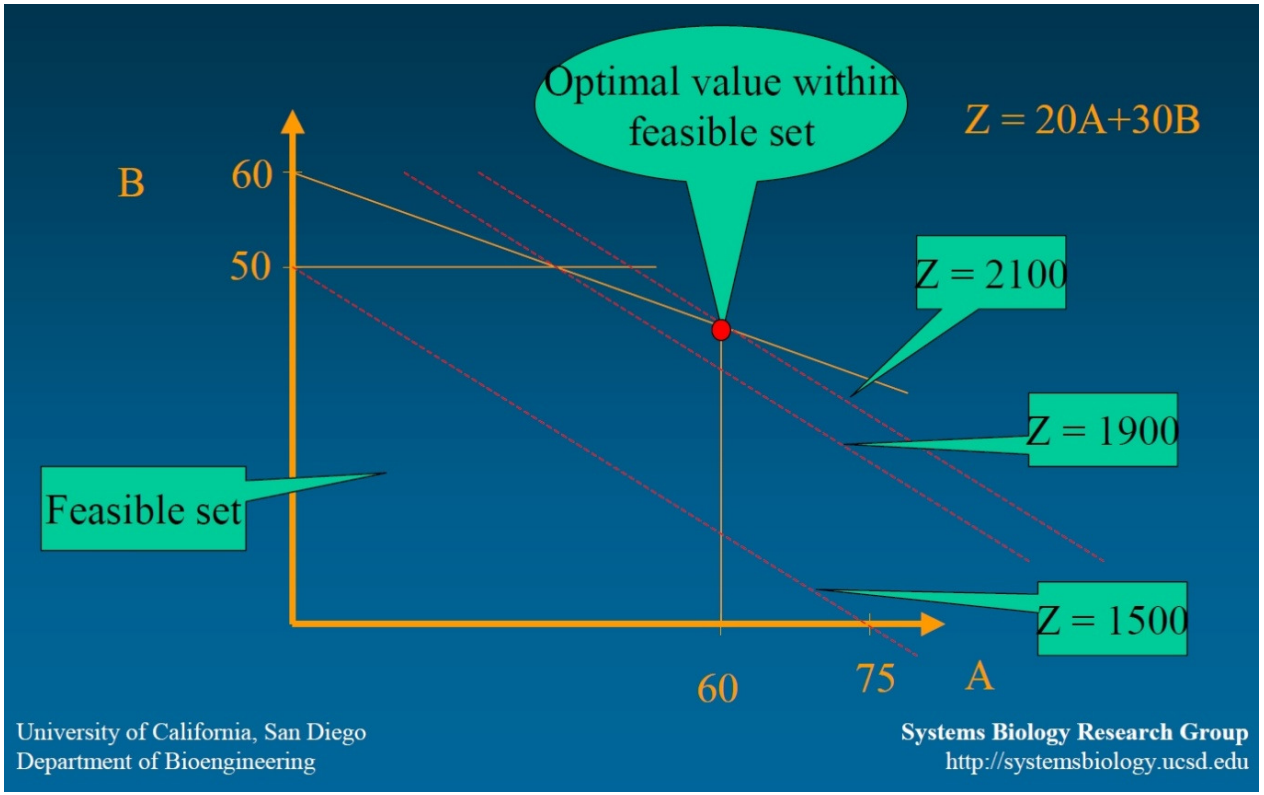


How do we solve this?

Linear Programming

Linear Programming (LP)

- Assume the following constraints:
  - 0<A<60
  - 0<B<50
  - A+2B<120
- Optimize:
  - Z=20A+30B



Application of CBM & FBA

- Predict metabolic fluxes on various media
- Predict growth rate
- Predict gene knockout lethality
- Characterize solution space
- Many more …

Available CBM Metabolic Models

Organism	Strain	Genes	Status	Version	Genes	Metabolites	Reactions
BACTERIA							
Acinetobacter baumannii	AYE	3,760	F	AbyMBEL891	650	778	891
Acinetobacter baylyi	ADP1	3,287	F	iAbaylyiV4	774	701	875
Bacillus subtilis		4,114	F	model_v3	844	988	1020
Bacillus subtilis		4,114	F		534	456	563
Bacillus subtilis	168	4,114	F	iBsu1103	1103	1138	1437
Buchnera aphidicola	APS	574	F	iGT196	196	240	263
Clostridium acetobutylicum	ATCC 824	3,848	F		474	422	552
Clostridium acetobutylicum	ATCC 824	3,848	F		432	479	502
Clostridium thermocellum	ATCC 27405	3,307	F	iSR432	432	525	577
Corynebacterium glutamicum	ATCC 13032	3,002	F			411	446
Corynebacterium glutamicum	ATCC 13032	3,002	F		227	423	502
Escherichia coli	K12 MG1655	4,405	F	iJE660	660	438	627
Escherichia coli	K12 MG1655	4,405	F	iJR904	904	625	931
Escherichia coli	K12 MG1655	4,405	F	iAF1260	1260	1039	2077
Geobacter metallireducens		3,532	F		747	769	697
Geobacter sulfurreducens		3,530	F		588	541	523
Haemophilus influenzae	Rd	1,775	F	iJE296	296	343	488
Haemophilus influenzae	Rd	1,775	F	iCS400	400	451	461
Helicobacter pylori	26695	1,632	F	iCS291	291	340	388
Helicobacter pylori	26695	1,632	F	iIT341	341	485	476
Lactobacillus plantarum	WCFS1	3,009	F		721	531	643
Lactococcus lactis	ssp. lactis *IL1403*	2,310	F		358	422	621
Mannheimia succiniciproducens	MBEL55E	2,384	F		335	332	373
Mannheimia succiniciproducens	MBEL55E	2,384	F		425	519	686
Mycobacterium tuberculosis	H37Rv	4,402	F	iNJ661	661	828	939
Mycobacterium tuberculosis	H37Rv	4,402	F	GSMN-TB	726	739	849
Mycoplasma genitalium	G-37	521	F	iPS189	189	274	262
Neisseria meningitidis	serogroup B	2,226	F		555	471	496
Porphyromonas gingivalis	W83	2,015	F	iVM679		564	679
Pseudomonas aeruginosa	PA01	5,640	F	iMO1056	1056	760	883
Pseudomonas putida	KT2440	5,350	F	iNJ746	746	911	950
Pseudomonas putida	KT2440	5,350	F	iJP815	815	886	877
Rhizobium etli	CFN42	3,168	F	iOR363	363	371	387
Rhodoferax ferrireducens		4,770	F		744	790	762
Salmonella typhimurium	LT2	4,489	F	iRR1083	1083	774	1087
Salmonella typhimurium	LT2	4,489	F	iMA945	945	1036	1964
Staphylococcus aureus	N315	2,588	F	iSB619	619	571	641
Staphylococcus aureus	N315	2,588	F	iMH551	551	604	712
Staphylococcus aureus	N315	2,588	F		546	1431	1493
Streptococcus thermophilus	LMG18311	1,889	F		429		522
Streptomyces coelicolor	A3(2)	7,825	F		700	500	700
Streptomyces coelicolor	A3(2)	7,825	F		789	759	1015
Synechocystis sp. PCC6803	PCC 6803	3,221	F		633	704	831
Thermotoga maritima	MSB8	1,917	F		478	503	562
Yersinia pestis	91001	4,037	F	iAN818m	818	825	1020

Organism	Strain	Genes	Status	Version	Genes	Metabolites	Reactions
ARCHAEA							
Methanosarcina barkeri	Fusaro	5,072	F	iAF692	692	558	619
Halobacterium salinarum	R-1	2,867	F		490	557	711
EUKARYOTES							
Arabidopsis thaliana		27,379	F	AraGEM	1419	1748	1567
Aspergillus nidulans		9,451	F		666	732	794
Homo sapiens		28,783	F	Recon 1	1,496	2,766	3,311
Leishmania major	Friedlin	8,370	F	iAC560	560	1,101	1,112
Mus musculus		28,287	F		473	872	1,220
Mus musculus		28,287	F		724	1287	1494
Saccharomyces cerevisiae	Sc288	6,183	F	iFF708	708	584	1,175
Saccharomyces cerevisiae	Sc288	6,183	F	iND750	750	646	1,149
Saccharomyces cerevisiae	Sc288	6,183	F	iLL672	672	636	1,038
Saccharomyces cerevisiae	Sc288	6,183	F	iIN800	800	1013	1446
Saccharomyces cerevisiae	Sc288	6,183	F	iMM904	904	713	1,412

Bernhard Palsson
UCSD