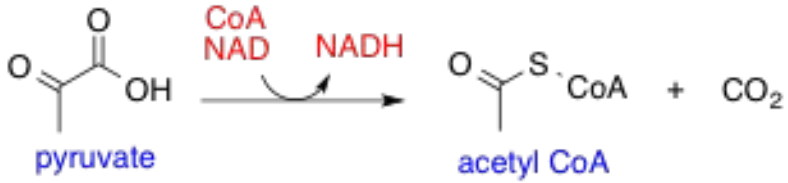
Whole-cell models; Digital evolution

- Flux balance analysis
- Artificial life

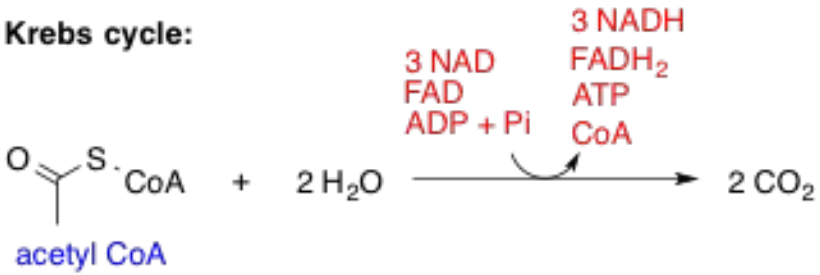
Metabolic networks

Glycolysis:

Oxidative decarboxylation:

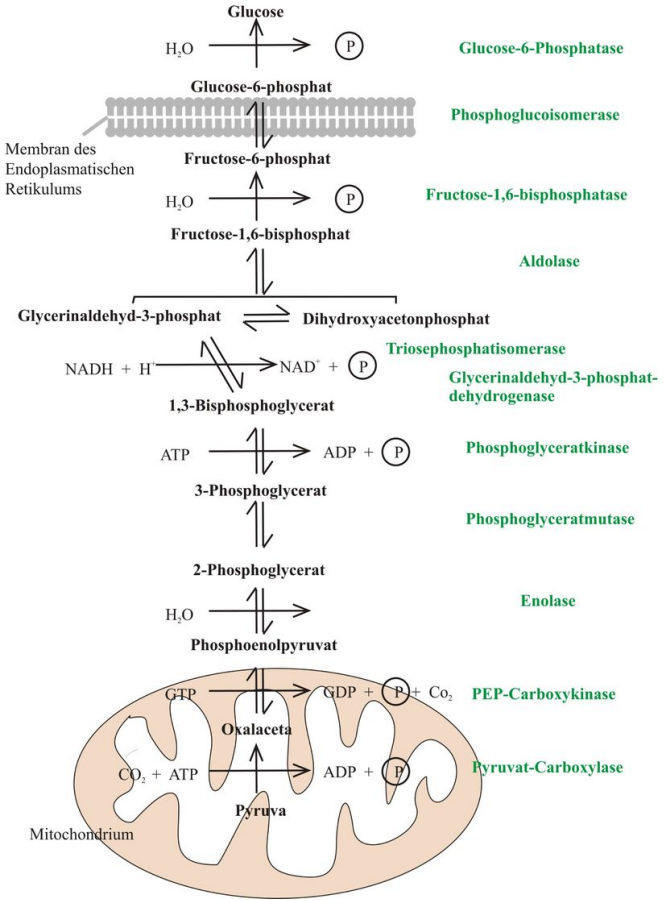


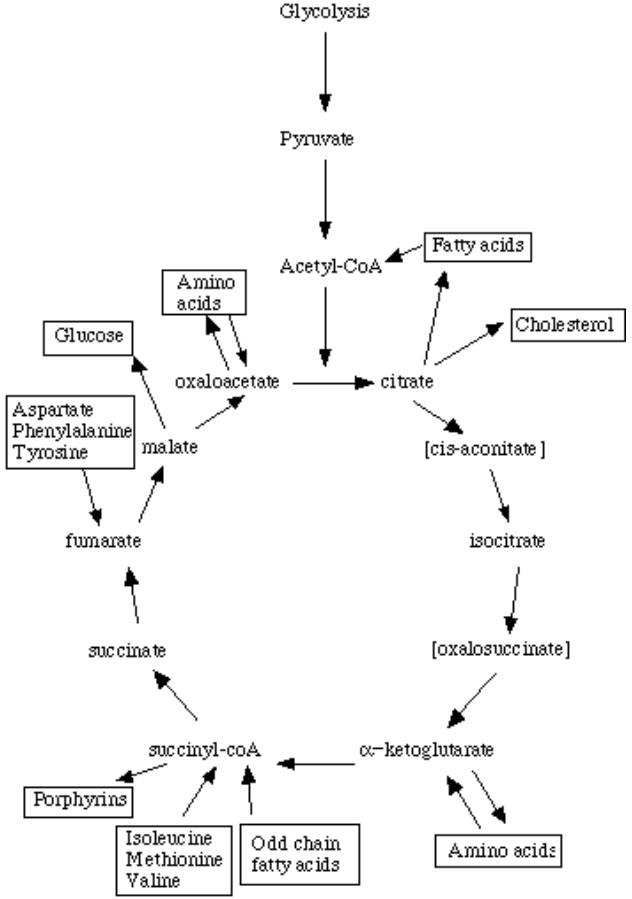
Krebs cycle:



Wikipedia

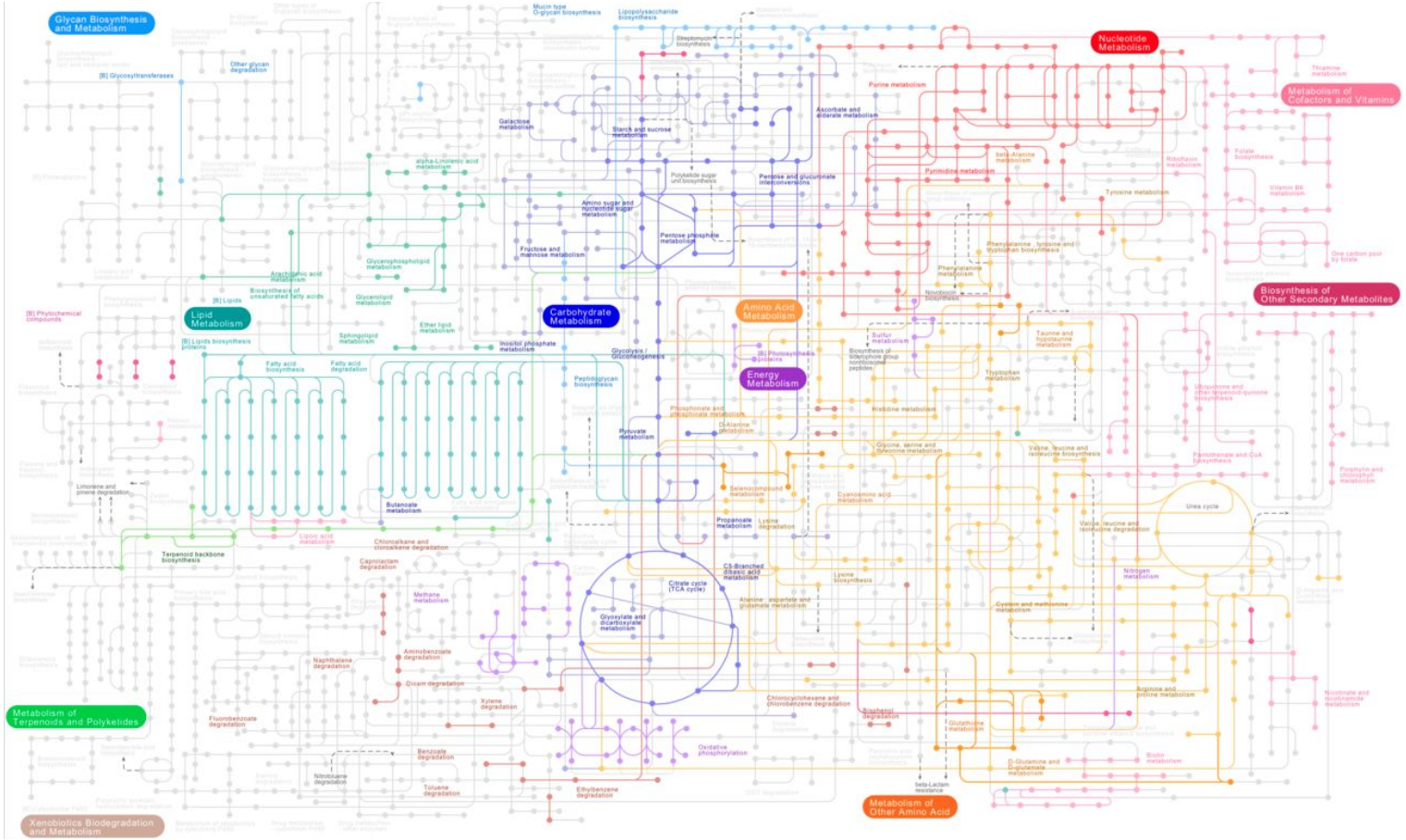
Metabolic networks





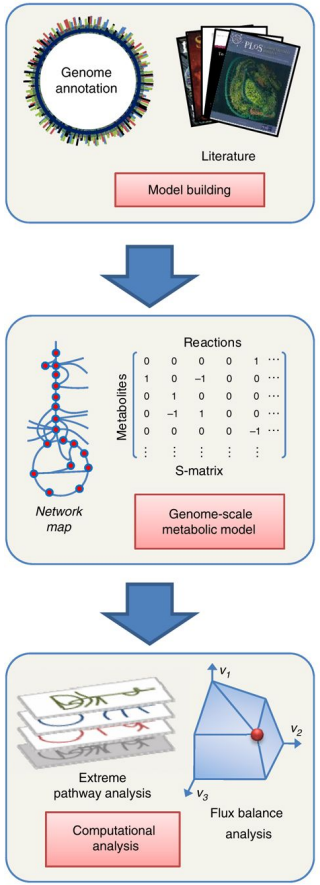
Wikipedia

Metabolic networks



KEGG

Genome-scale metabolic network reconstruction & model



Genome-scale metabolic network **reconstruction**:

- A collection of biochemical transformation derived from the genome annotation and the literature of a particular organism.
- Formed based on an organism-specific knowledge base.
- A network reconstruction is unique to an organism.

Genome-scale metabolic network **model**:

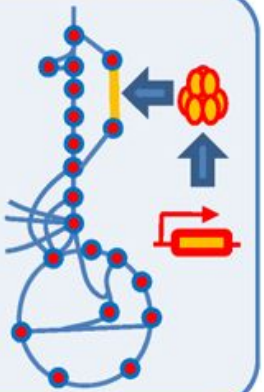
- Derived from a *reconstruction* by converting it into a mathematical form (i.e., an in silico model) and by assessing its phenotypic properties computationally.

Oberhardt (2009) Mol. Sys. biol.
Thiele & Palsson (2010) Nat. Protoc.

Genome-scale metabolic network reconstruction & model

3. Directing hypothesis-driven discovery

A Metabolic GENRE aided in determining pathway usage and discovering a novel citramalate synthase gene in *G. sulfurreducens*. GENREs have also helped study the effects of transposons on downstream genes, and identify transc riptional timing patterns in *S. cerevisiae*.



2. Guidance of metabolic engineering

Metabolic GENREs guided efforts to engineer malate and succinate producing strains of *S. cerevisiae* and *M. succiniciproducens*. GENREs have also helped determine ways to increase the respiration rate of *G. sulfurreducens* and scale-up vaccine production against *N. meningitides*.

Oberhardt (2009) Mol. Sys. biol.
Thiele & Palsson (2010) Nat. Protoc.

Flux balance analysis (FBA)

FBA: metabolic network ⟶ linear programming optimization problem.

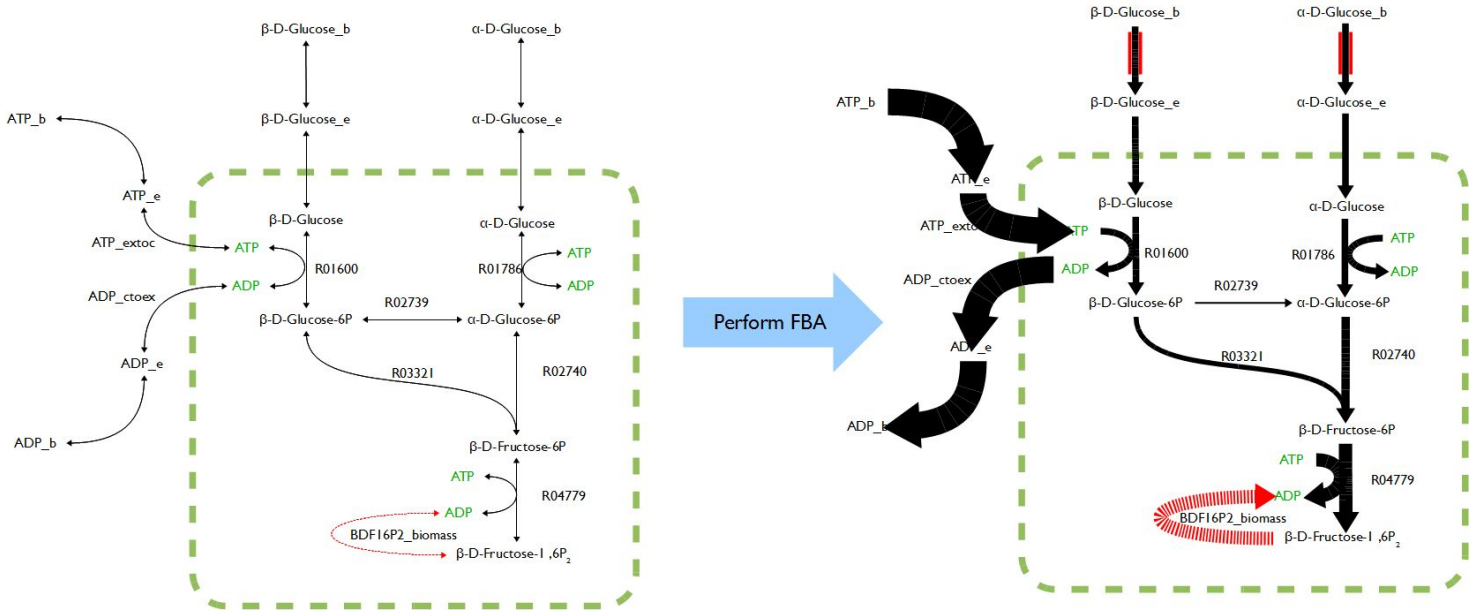
The main constraints in FBA: steady-state mass conservation of metabolites.

- Relies on balancing metabolic fluxes
- Is based on the fundamental law of mass conservation
- Is performed under steady-state conditions (an example of constraint…)
- Requires information only about:
  a. the stoichiometry of metabolic pathways,
  b. metabolic demands, and
  c. a few strain specific parameters
- Does NOT require enzymatic kinetic data

Flux balance analysis

The results of FBA on a metabolic network of the top six reactions of glycolysis.

- The predicted flux through each reaction is proportional to the width of the line.
- **Red springy arrow**: Objective function; **Red bars**: Constraints on **α**-D-glucose and **β**-D-glucose import.

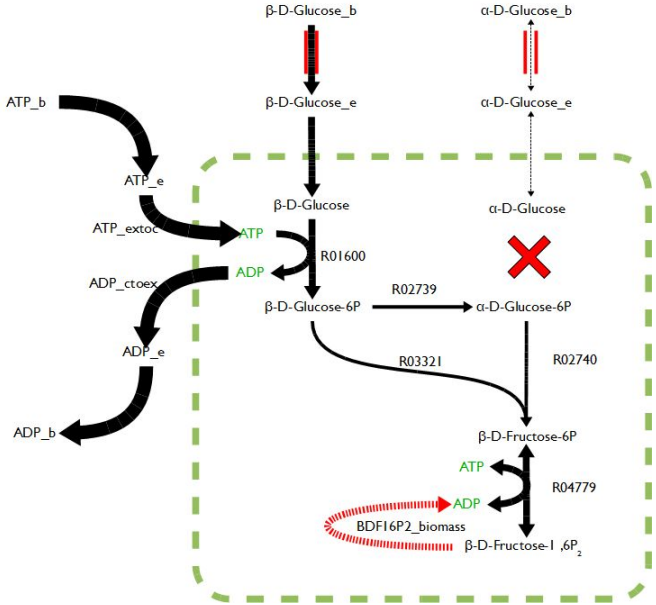


Wikipedia

Flux balance analysis

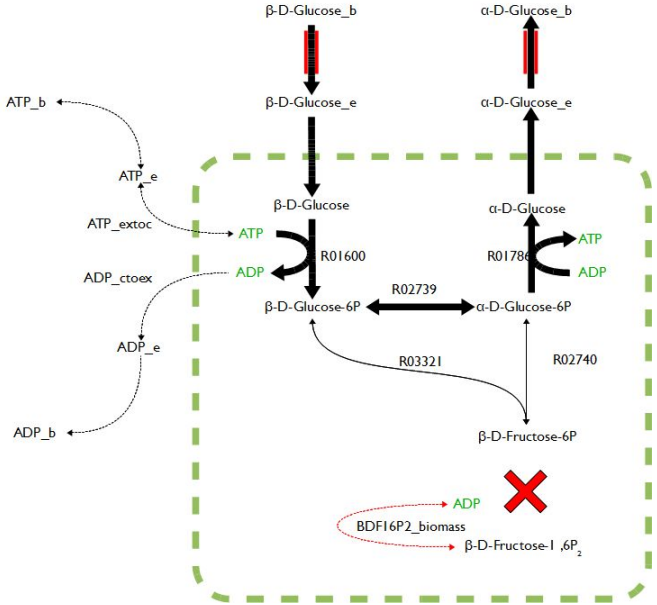
Non-lethal gene deletion in a metabolic network.

- Flux through the objective function is halved but is still present.



Lethal gene deletion in a metabolic network.

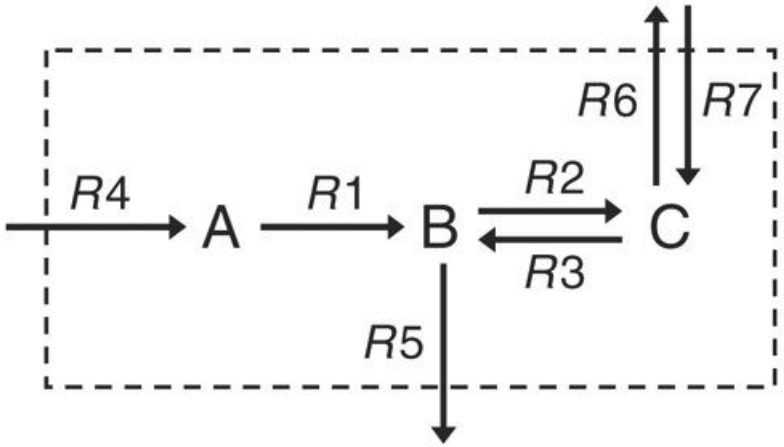
- No flux through the objective function → pathway is no longer functional.



Wikipedia

Flux balance analysis

1. Reaction network formalism



Chemical reactions

Internal	Exchange
*R*1: −1 A → 1 B	*R*4: 1 A
*R*2: −1 B → 1 C	*R*5: −1 B
*R*3: −1 C → 1 B	*R*6: −1 C
	*R*7: 1 C

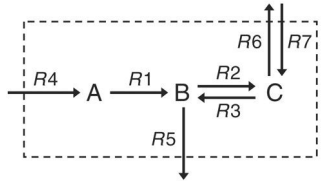
S =

	*R*1	*R*2	*R*3	*R*4	*R*5	*R*6	*R*7
A	−1	0	0	1	0	0	0
B	1	−1	1	0	−1	0	0
C	0	1	−1	0	0	−1	1

Lee (2006) Brief. Bioinfo.

Flux balance analysis

1. Reaction network formalism



S =

	R1	R2	R3	R4	R5	R6	R7
A	–1	0	0	1	0	0	0
B	1	–1	1	0	–1	0	0
C	0	1	–1	0	0	–1	1

$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccccccc} R1 & R2 & R3 & R4 & R5 & R6 & R7 \end{array} \\ \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1 & 1 \end{bmatrix} \end{array} * \begin{bmatrix} v_1 \\ \vdots \\ v_7 \end{bmatrix}$$

?

Lee (2006) Brief. Bioinfo.

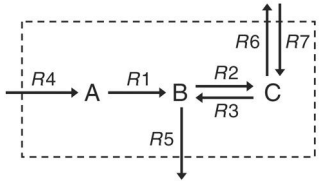
Flux balance analysis

2. FBA formulation

Dynamic mass balance

$$\frac{dC}{dt} = \mathbf{Sv}$$

C : Concentration
t : Time
$\mathbf{S}$: Stoichiometric matrix
$\boldsymbol{v}$: Flux vector



Steady-state assumption $\quad \mathbf{Sv = 0}$

LP formulation

Objective: max $Z = \mathbf{c} \cdot \mathbf{v}$

Constraints:

$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccccccc} R1 & R2 & R3 & R4 & R5 & R6 & R7 \end{array} \\ \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1 & 1 \end{bmatrix} \end{array} \begin{bmatrix} v_1 \\ \vdots \\ v_7 \end{bmatrix} = \mathbf{0}$$

$0 \le v_1, \ldots, v_7 \le 10$

Lee (2006) Brief. Bioinfo.

Flux balance analysis: Objective function

Objective function: Physiologically-meaningful or design-based objective for the interrogation or exploitation of a given system.

Examples:

- Maximizing…
  - biomass or cell growth
  - maximizing ATP production
  - maximizing the rate of synthesis of a particular product
- Minimizing…
  - ATP production
  - nutrient uptake (both to determine conditions of optimal metabolic energy efficiency)

Lee (2006) Brief. Bioinfo.

Flux balance analysis: Constraints

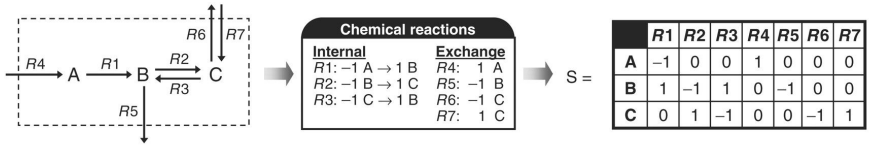
No. of equations (one per reactant) << no. of unknown variables (reaction fluxes).

- An *under-determined* set of linear equations.
- Therefore, optimize fluxes given cellular objective given a bunch of constraints.
- **Principal constraint**: mass balance
- **Additional constraints**:
  - physico-chemical constraints
  - spatial or topological constraints
  - condition dependent environmental constraints
  - regulatory constraints
- All constraints together represent a set of linear equations.

Lee (2006) Brief. Bioinfo.

Flux balance analysis

I. Reaction network formalism



	R1	R2	R3	R4	R5	R6	R7
A	–1	0	0	1	0	0	0
B	1	–1	1	0	–1	0	0
C	0	1	–1	0	0	–1	1

II. FBA formulation

Dynamic mass balance

$$\frac{dC}{dt} = \mathbf{Sv}$$

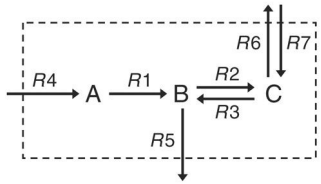
C : Concentration
t : Time
$\mathbf{S}$: Stoichiometric matrix
$\boldsymbol{v}$: Flux vector

Steady-state assumption

$$\mathbf{Sv} = \mathbf{0}$$

LP formulation

Objective: max $Z = \mathbf{c} \cdot \mathbf{v}$

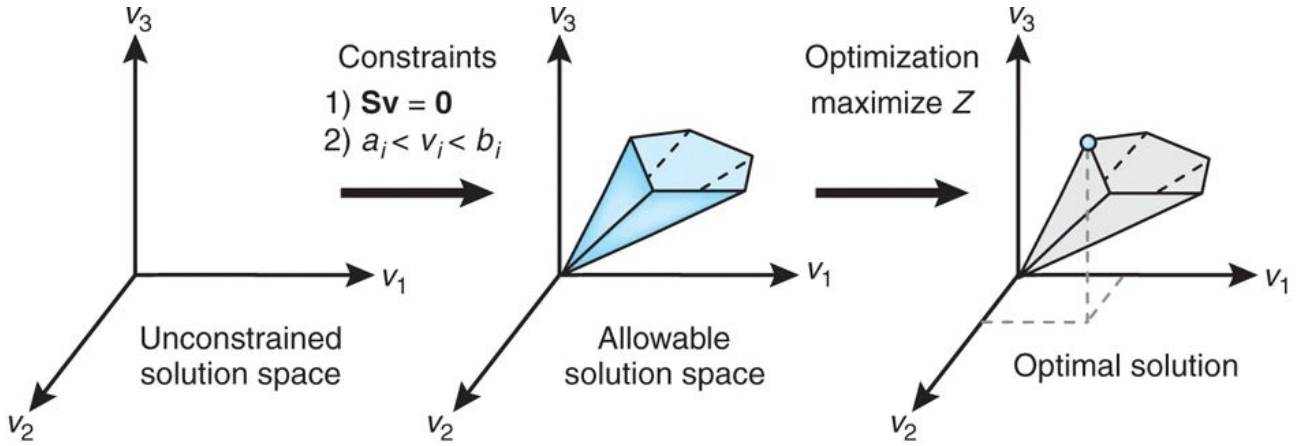


Constraints:

$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccccccc} R1 & R2 & R3 & R4 & R5 & R6 & R7 \end{array} \\ \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1 & 1 \end{bmatrix} \end{array} \begin{bmatrix} v_1 \\ \vdots \\ v_7 \end{bmatrix} = \mathbf{0} \qquad 0 \le v_1, \ldots, v_7 \le 10$$

Lee (2006) Brief. Bioinfo.

Constraint-based modeling



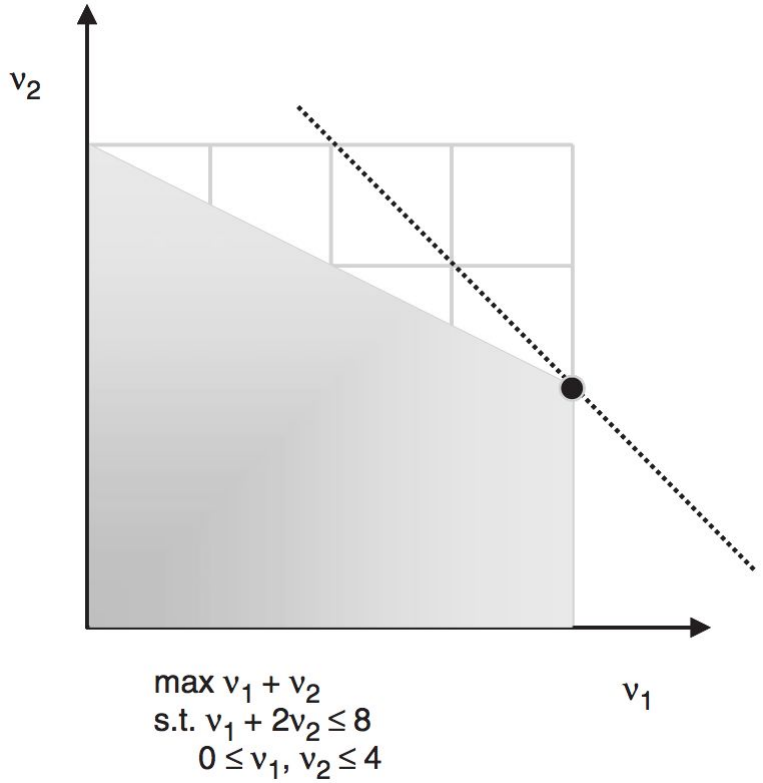
1. No constraints: flux may lie at any point in solution space.
2. Mass balance constraints (imposed by the stoichiometry) and capacity constraints (imposed by the lower and upper bounds: a_i & b_i): defines allowable solution space.
   a. Any flux distribution within this space is allowable; Points outside this space are denied
3. Optimization of an objective function: A single optimal flux distribution that lies on the edge of the allowable solution space.

Orth (2010) Nat. Biotech.

Constraint-based modeling

Linear programming

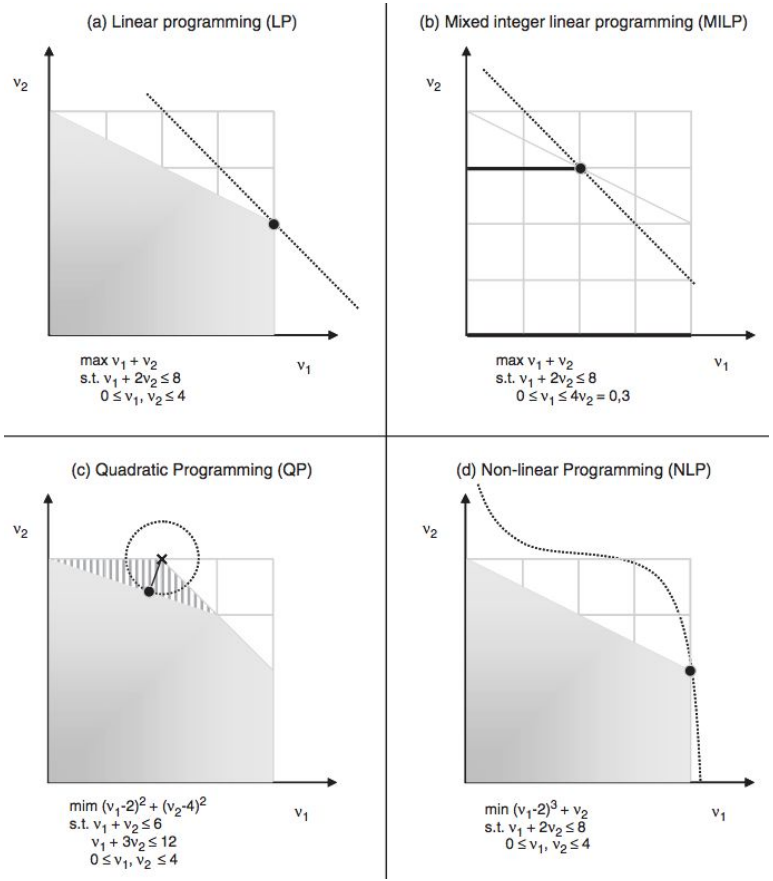
- Feasible solution space:
  - shaded area and solid lines
  - defined by:
    - flux capacities,
    - stoichiometric relationships, and
    - design specification (e.g. gene deletions).
- Objective function: dotted line
- Optimal solution: circular dot



Lee (2006) Brief. Bioinfo.

Constraint-based modeling

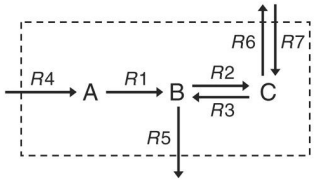
- Mixed Integer LP (MILP):
  - Integer variables are involved in a linear programming problem (e.g. binary variable formulation for gene deletion).
- Quadratic programming (QP):
  - Quadratic objective function subject to linear constraints.
  - This technique is generally used for finding the closest point to a specified point.
- Nonlinear programming (NLP):
  - Nonlinear objectives or constraints.
  - Generally difficult to solve for global optimal solution because of its non-convexity.



Lee (2006) Brief. Bioinfo.

Flux balance analysis

I. Reaction network formalism



Chemical reactions

Internal	Exchange
*R*1: –1 A → 1 B	*R*4: 1 A
*R*2: –1 B → 1 C	*R*5: –1 B
*R*3: –1 C → 1 B	*R*6: –1 C
	*R*7: 1 C

S =

	*R*1	*R*2	*R*3	*R*4	*R*5	*R*6	*R*7
A	–1	0	0	1	0	0	0
B	1	–1	1	0	–1	0	0
C	0	1	–1	0	0	–1	1

II. FBA formulation

Dynamic mass balance

$$\frac{dC}{dt} = \mathbf{S}\mathbf{v}$$

C : Concentration
t : Time
$\mathbf{S}$: Stoichiometric matrix
$\boldsymbol{v}$: Flux vector

Steady-state assumption

$$\mathbf{S}\mathbf{v} = \mathbf{0}$$

LP formulation

Objective: $\max Z = v_5$

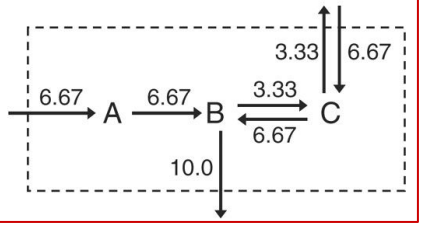
Constraints:

$$\begin{array}{c} \\ A \\ B \\ C \end{array}\begin{array}{c} \begin{array}{ccccccc} R1 & R2 & R3 & R4 & R5 & R6 & R7 \end{array} \\ \begin{bmatrix} -1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1 & 1 \end{bmatrix} \end{array} \begin{bmatrix} v_1 \\ \vdots \\ v_7 \end{bmatrix} = \mathbf{0} \qquad 0 \le v_1,\ldots,v_7 \le 10$$

III. Hypothetical flux distribution at steady-state

$Z = 10$

$\boldsymbol{v} = [6.67\ 3.33\ 6.67\ 6.6710.0\ 3.33\ 6.67]^T$



Lee (2006) Brief. Bioinfo.

Genome-scale metabolic network reconstruction & model

Organism	Strain	Genes	Version	GR	Mets	Rxns	Comp
Bacillus subtilis		4,225	model_v3	844	988	1,020	2 (c,e)
Escherichia coli	K12 MG1655	4,405	iAF1260	1,260	1,039	2,077	3 (c,e,p)
Helicobacter pylori	26695	1,632	iIT341	341	485	476	2 (c,e)
Pseudomonas putida	KT2440	5,350	iNJ746	746	911	950	3 (c,p,e)
Pseudomonas putida	KT2440	5,350	iJP815	815	886	877	2 (c,e)
Pseudomonas aeruginosa	PA01	5,640	iMO1056	1,056	760	883	2 (c,e)
Mycoplasma genitalium	G-37	521	iPS189	189	274	262	2 (c,e)
Lactobacillus plantarum	WCFS1	3,009		721	531	643	2 (c,e)
Streptomyces coelicolor	A3(2)	8,042		700	500	700	2 (c,e)
Leishmania major	Friedlin	8,370	iAC560	560	1,101	1,112	8 (a,f,y,c,e,m,r,n)
Saccharomyces cerevisiae	Sc288	6,183	iMM904	904	713	1,412	8 (c,e,m,x,n,r,v,g)
Homo sapiens		28,783	Recon 1	1,496	2,766	3,311	8 (c,e,m,x,n,r,v,g)

Thiele & Palsson (2010) Nat. Protoc.
http://gcrg.ucsd.edu/InSilicoOrganisms/OtherOrganisms

Constructing a genome-scale metabolic model

Conversion of reconstruction into a model

3. Conversion of reconstruction into computable format

38| Initialize the COBRA toolbox.
39| Load reconstruction into Matlab.
40| Verify S matrix.
41| Set objective function.
42| Set simulation constraints.

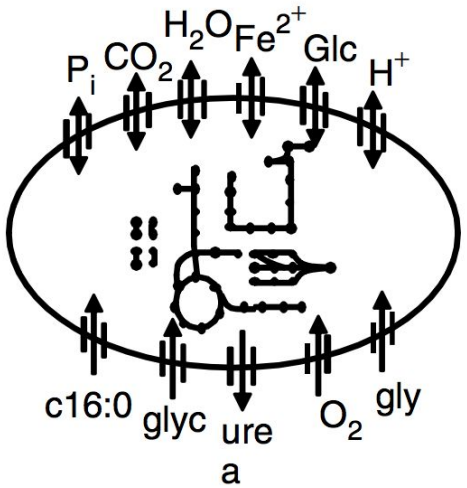
Thiele & Palsson (2010) Nat. Protoc.

Constructing a genome-scale metabolic model

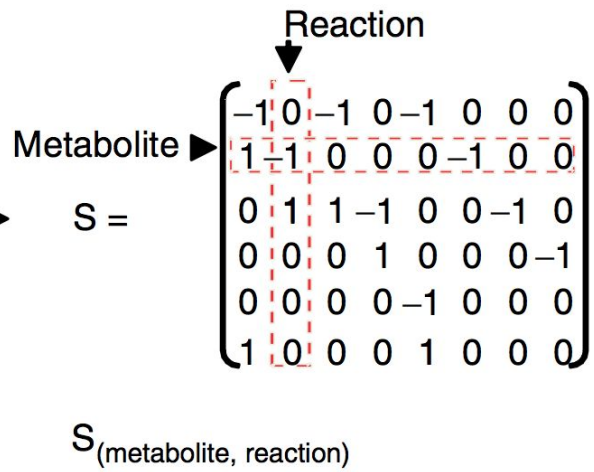
Conversion of reconstruction into a model

Mathematical representation

Metabolic network



Stoichiometric matrix



By definition:

- Substrates have negative coefficients (i.e., they are consumed)
- Products have positive coefficients (i.e., they are produced)

v is a vector of reaction fluxes

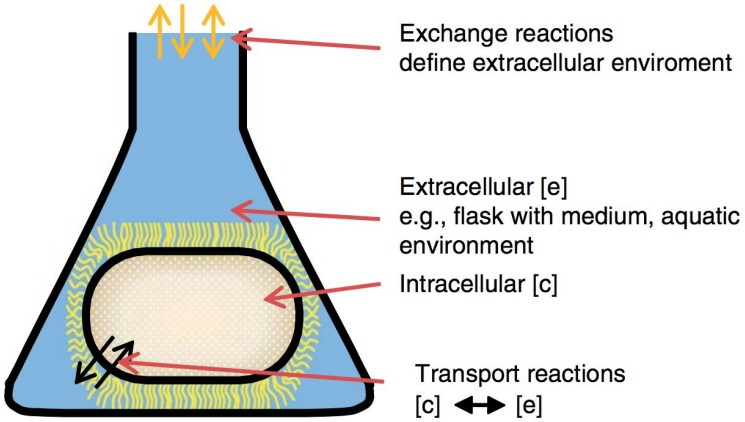
Conservation of mass: All steady states can be described by **Sv** = 0

Thiele & Palsson (2010) Nat. Protoc.

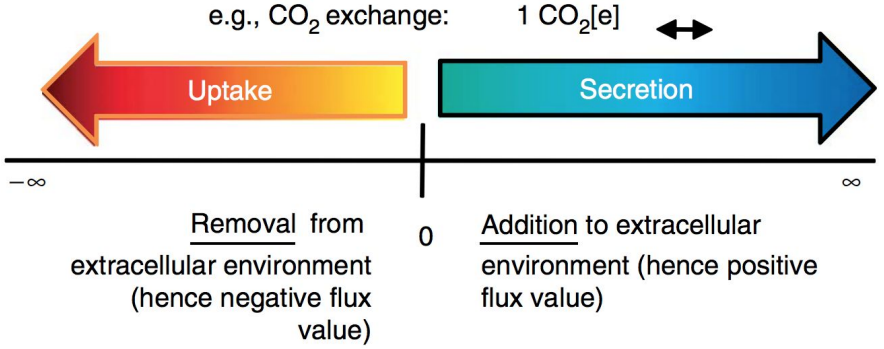
Constructing a genome-scale metabolic model

Conversion of reconstruction into a model

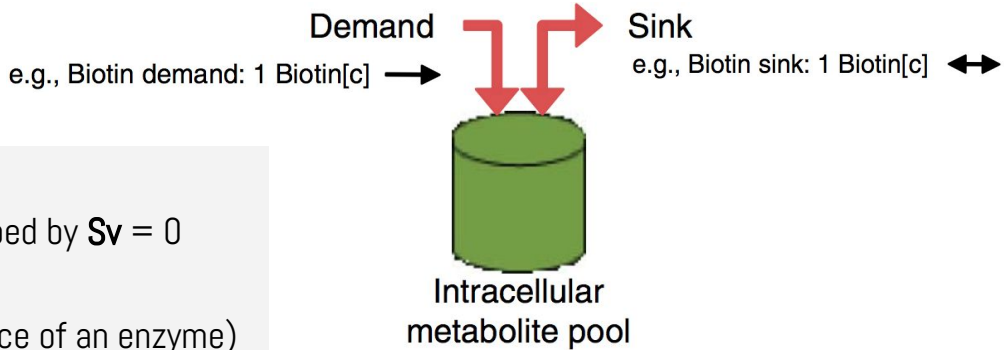
Definition of system boundaries



Exchange reactions



Demand/sink reactions



Constraints:

- Mass conservation: all steady states can be described by $\mathbf{Sv} = 0$
- Thermodynamics (reaction directionality)
- Enzyme capacity or regulation (i.e., presence/absence of an enzyme)

Thiele & Palsson (2010) Nat. Protoc.

Constructing a genome-scale metabolic model

Network evaluation ≃ Debugging

4. Network evaluation

43–44| Test if network is mass-and charge balanced.
45| Identify metabolic dead-ends.
46–48| Perform gap analysis.
49| Add missing exchange reactions to model.
50| Set exchange constraints for a simulation condition.
51–58| Test for stoichiometrically balanced cycles.
59| Re-compute gap list.
60–65| Test if biomass precursors can be produced in standard medium.
66| Test if biomass precursors can be produced in other growth media.
67–75| Test if the model can produce known secretion products.
76–78| Check for blocked reactions.
79–80| Compute single gene deletion phenotypes.
81–82| Test for known incapabilites of the organism.
83| Compare predicted physiological properties with known properties.
84–87| Test if the model can grow fast enough.
88–94| Test if the model grows too fast.

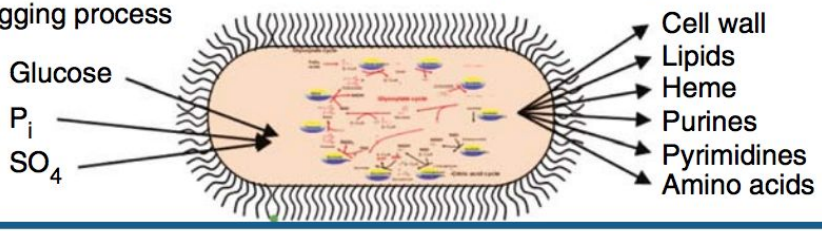
Thiele & Palsson (2010) Nat. Protoc.

Constructing a genome-scale metabolic model

Network evaluation ≃ Debugging

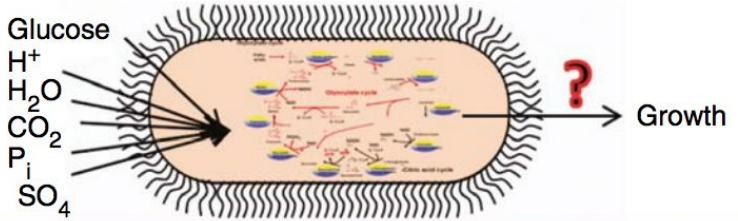
Analysis of biomass precursors synthesis

-Biomass precursors = cellular growth requirements
-Pathways to synthesize precursors must be complete (i.e., functional) for the network to simulate growth
-Testing synthesis of each separate biomass precursor is part of the debugging process



Analysis of growth in minimal medium

-Minimal medium is defined for many organisms and can be found in primary literature
-Contains at least 1 C-, N-, S- and P-source
-Auxotrophs may require the presence of addition metabolites



Test for growth on known carbon sources

-Exchange reactions define medium and environment
-Transport reactions allow network to consume carbon sources
-Biodegradative pathways that are required for carbon utilization

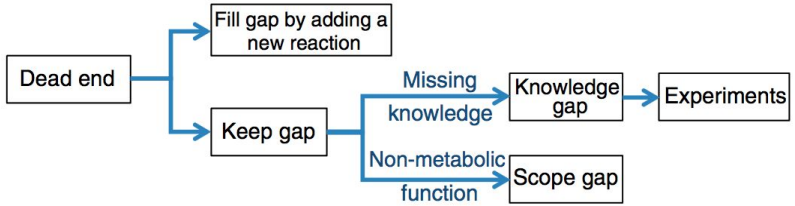
Secretion capability

-Transport and exchange reactions are required in reconstruction to enable secretion
-Secretion may only occur under certain circumstances (e.g., D-lactic acid formation under anoxic conditions)
-Comparison with known secretion pattern of multiple metabolites (e.g., secretion of a certain ratio of CO_2 and acetate)

Thiele & Palsson (2010) Nat. Protoc.

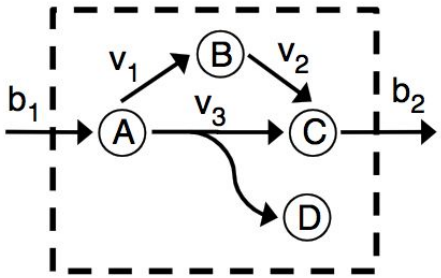
Constructing a genome-scale metabolic model

Network evaluation ≃ Debugging



Identifying gaps

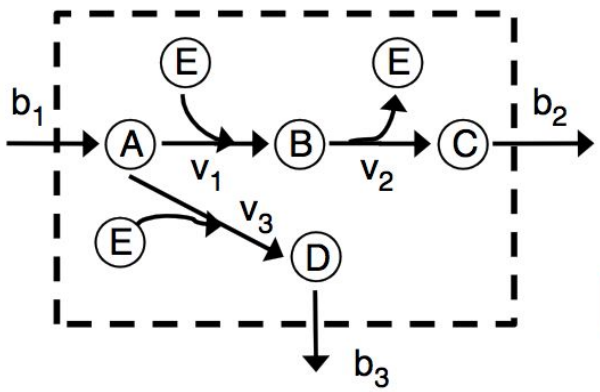
Connectivity based (topology):



	v_1	v_2	v_3	b_1	b_2
A	−1	0	−1	1	0
B	1	−1	0	0	0
C	0	1	1	0	−1
D	0	0	1	0	0

Dead end

Functionality based (computation)



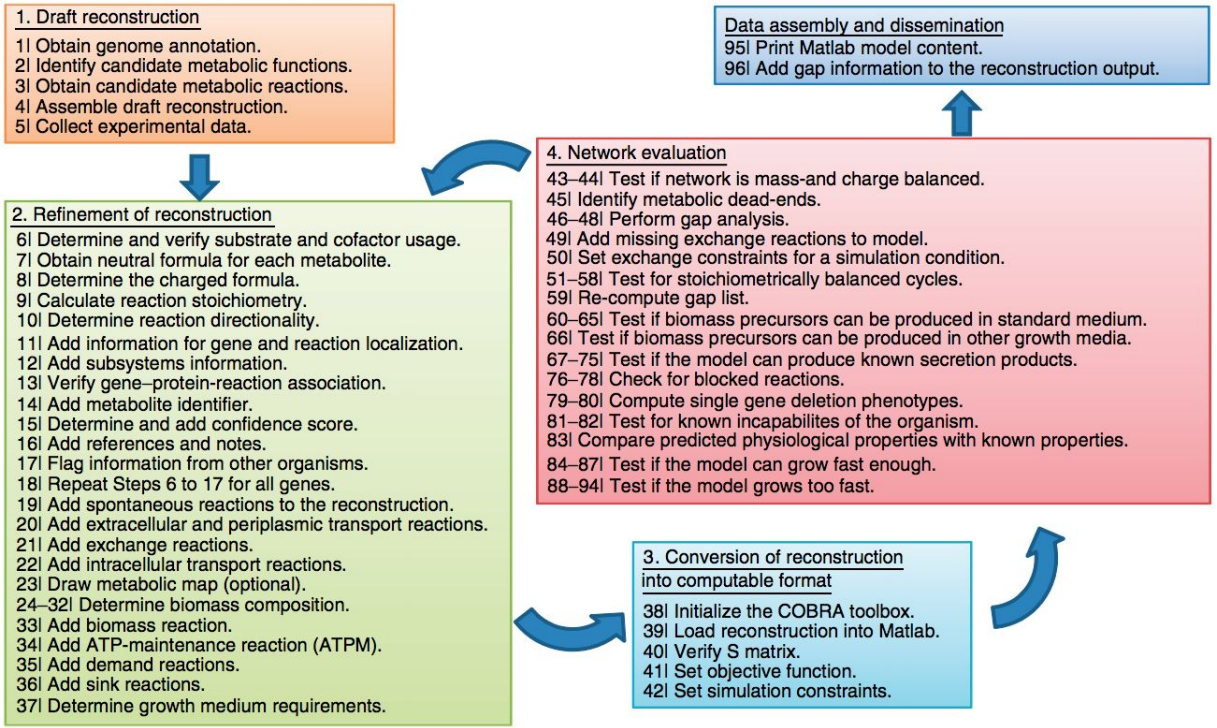
	v_1	v_2	v_3	b_1	b_2	b_3
A	−1	0	−1	1	0	0
B	1	−1	0	0	0	0
C	0	1	0	0	−1	0
D	0	0	1	0	0	−1
E	−1	1	−1	0	0	0

Thiele & Palsson (2010) Nat. Protoc.

Constructing a genome-scale metabolic model

Procedure to iteratively reconstruct metabolic networks.

- Iterate stages 2–4 are continuously…
- …until model predictions are similar to the phenotypic characteristics of the target organism and/or all experimental data for comparison are exhausted.

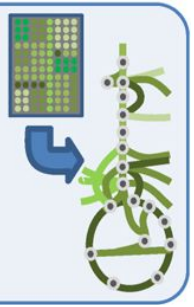


Thiele & Palsson (2010) Nat. Protoc.

Genome-scale metabolic network reconstruction & model

1. Contextualization of HT data

Several studies have overlaid gene microarray data on a metabolic GENRE to determine condition-dependent cell phenotypes. Metabolic GENREs have also been used to interpret metabolomic data, 13C flux data, and to link multiple data types.

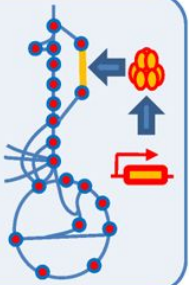


2. Guidance of metabolic engineering

Metabolic GENREs guided efforts to engineer malate and succinate producing strains of *S. cerevisiae* and *M. succiniciproducens*. GENREs have also helped determine ways to increase the respiration rate of *G. sulfurreducens* and scale-up vaccine production against *N. meningitides*.

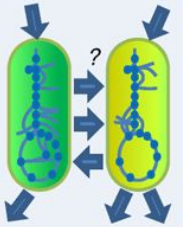
3. Directing hypothesis-driven discovery

A Metabolic GENRE aided in determining pathway usage and discovering a novel citramalate synthase gene in *G. sulfurreducens*. GENREs have also helped study the effects of transposons on downstream genes, and identify transc riptional timing patterns in *S. cerevisiae*.



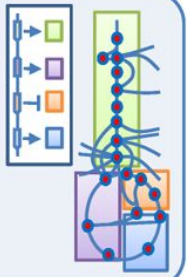
4. Interrogation of multi-species relationships

A dual-species metabolic model was built to study Interactions between the syntrophic bacteria, *D. vulgaris* and *M. maripaludis*. Metabolic models have also been used in comparisons of multiple species, such as an analysis of pathway differences between four halophilic bacteria.



5. Network property discovery

Metabolic GENREs have been used to study metabolite connectivity, and pathway redundancy *in silico*. Pathway-analysis tools have also spawned techniques such as flux coupling analysis, which has helped identify novel drug targets in *M. tuberculosis*.



Oberhardt (2009) Mol. Sys. biol.
Thiele & Palsson (2010) Nat. Protoc.