Systems Biology 551-1174-00L

Modeling Metabolic Networks: Flux Balance Analysis

30 Mar, 2017

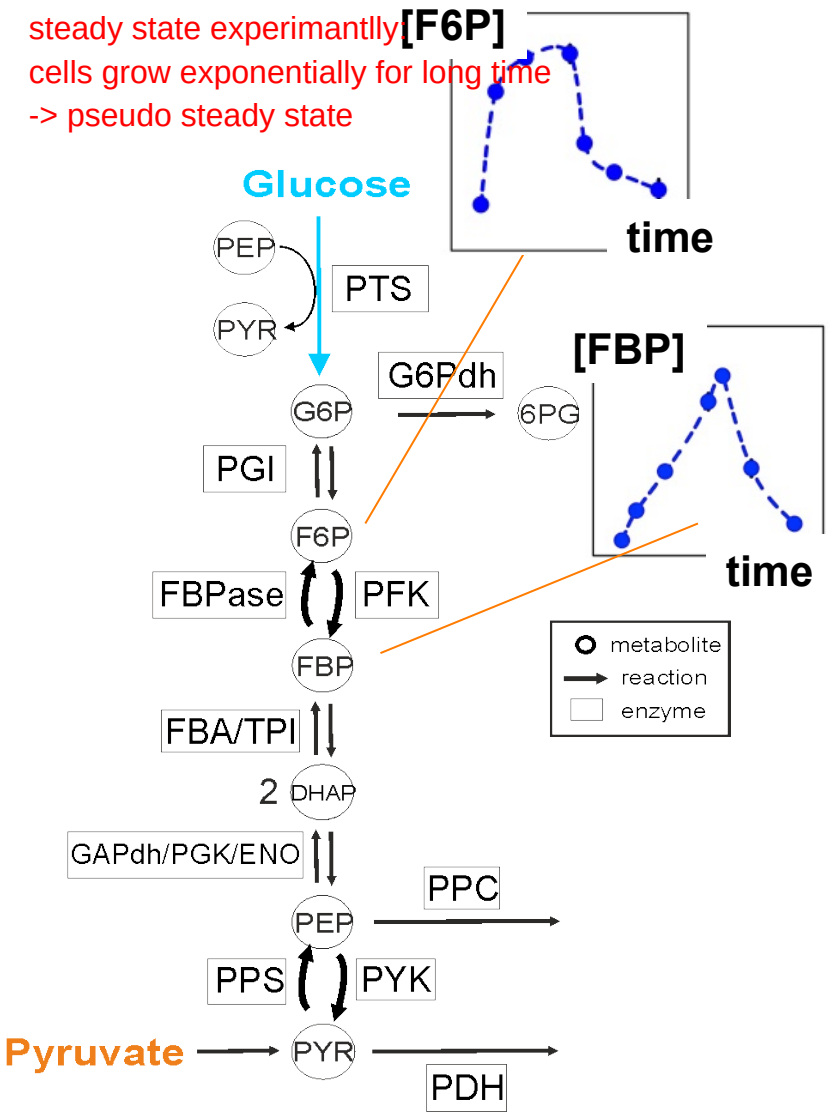
Uwe Sauer, Molecular Systems Biology
Jörg Stelling, BSSE

Content:

- **From textbook biochemistry to genome-scale stoichiometric models (US)**
- The solution space of stoichiometric models (JS)
- Flux balance analysis (JS)

Recap: Relationship between metabolite [conc] and fluxes are mass balances and kinetic equations

steady state experimantlly
cells grow exponentially for long time
-> pseudo steady state



Fluxes determine metabolite dynamics

$$\frac{d[FBP]}{dt} = v_{PFK} - v_{FBA} - v_{FBPase} \neq 0$$

Metabolites determine fluxes

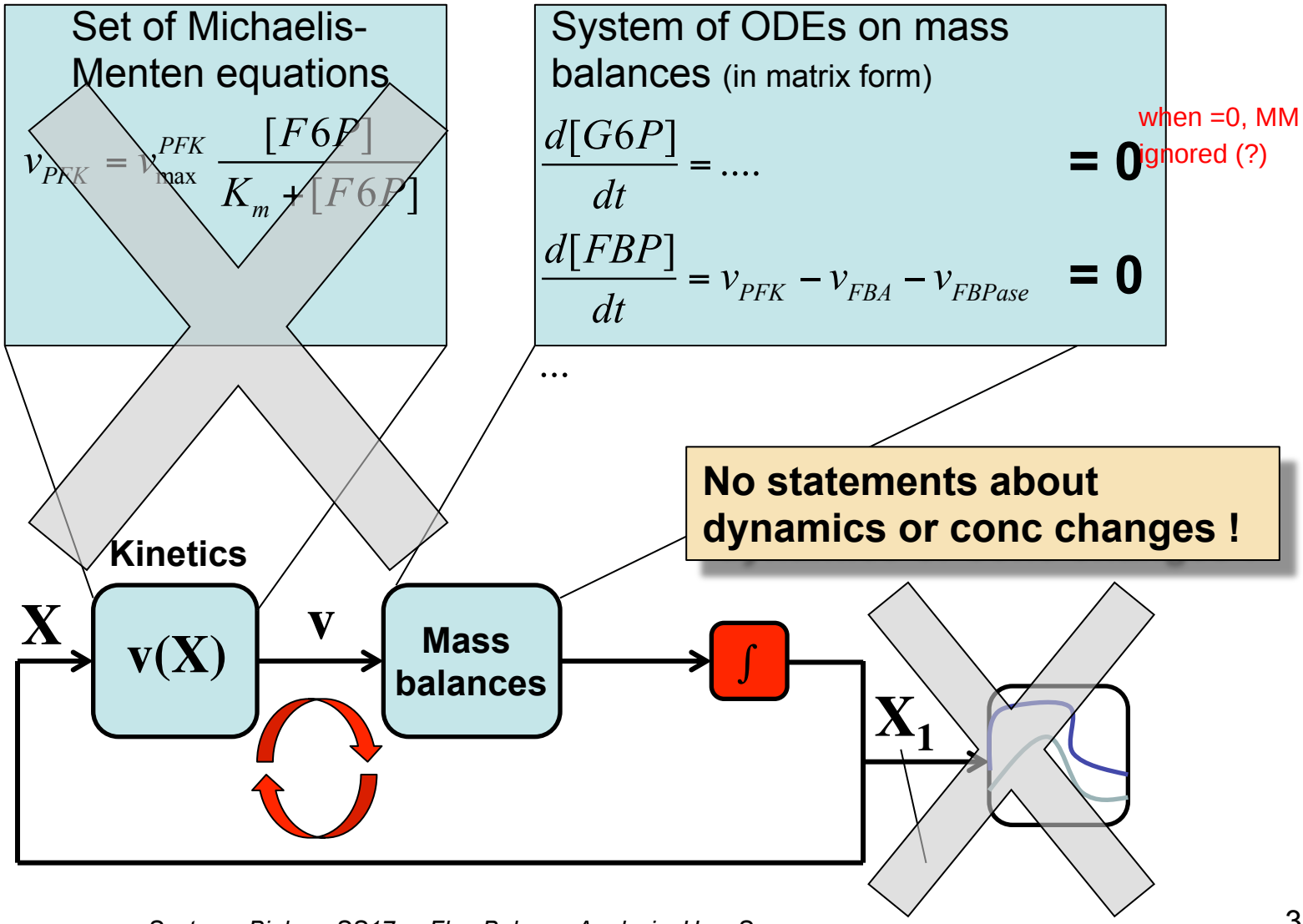
$$v_{PFK} = v_{\max}^{PFK} \frac{[F6P]}{K_m + [F6P]}$$

- concentrations and fluxes are intertwined
- dynamic ODE modelling required for considering metabolite dynamics
- **in steady state we can work with mass balances alone** because all fluxes are balanced and concentrations remain constant.

How to achieve a steady state in practise?

What computational challenges arises when modeling dynamics?

Restricting Scope to Steady State Greatly Simplifies the Analysis (and reduces insights!)



Differential Eqn Modeling vs FBA

- ODE fitting in exercise 5 allowed to test whether the dynamics of metabolite concentrations can be explained by a model (that represents our current knowledge of a biological process, glycolysis in the specific case) – or needs further knowledge such as regulation. There is no other way to understand/check for consistency between dynamic experimental data and existing knowledge.
- ODE models require parameters for each reaction that must be obtained by fitting the model to data. The computational problem of **parameter identification scales poorly with model size and lacking prior knowledge,** because many (often **non-linear**) combinations of parameters lead to the same output. Due to this "curse of dimensionality" for a global, non-linear search in high-dimenstional spaces, mechanistic ODE modeling is typically restricted to smaller systems (eg a single metabolic or signaling pathway)
- If we can ensure that concentrations do not change during the period of analysis (ie system is in steady state), the problem can be simplified in metabolism by considering only balanced material fluxes in the network. Because of this computational simplification (to **linear mass balance equations**) also large networks become tractable. Methods that achieve that are **constraint-based analyses**. We will treat here one of them called **flux balance analysis.**

the more reactions the harder it gets to identify the parameters for the matrix.

5. Modelling Large Metabolic Networks: Flux Balance Analysis (FBA)

will occur in exam 100% (pure pen and paper calculation)

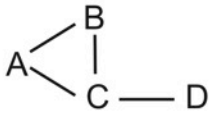
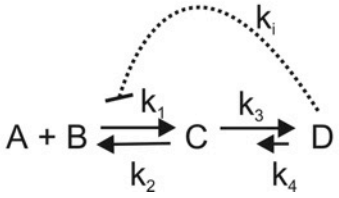
- **Understand biological basis of constructing genome-scale stoichiometric models and their key elements.**
  (eg requirements, optimality assumption ….)
- **Understand principles of steady-state as a key simplification/assumption to make use of large scale models.**
- **Learn to incorporate biological assumptions into steady-state FBA.**

Exercise 6

Familiarize with FBA basics on a simple system with pen-and-pencil and computational modeling

- **apply linear optimization to find a flux distribution.**
- **analyze the effect of constraints and optimization function on flux solution.**

Recap: Requirements for Different Modeling Approaches

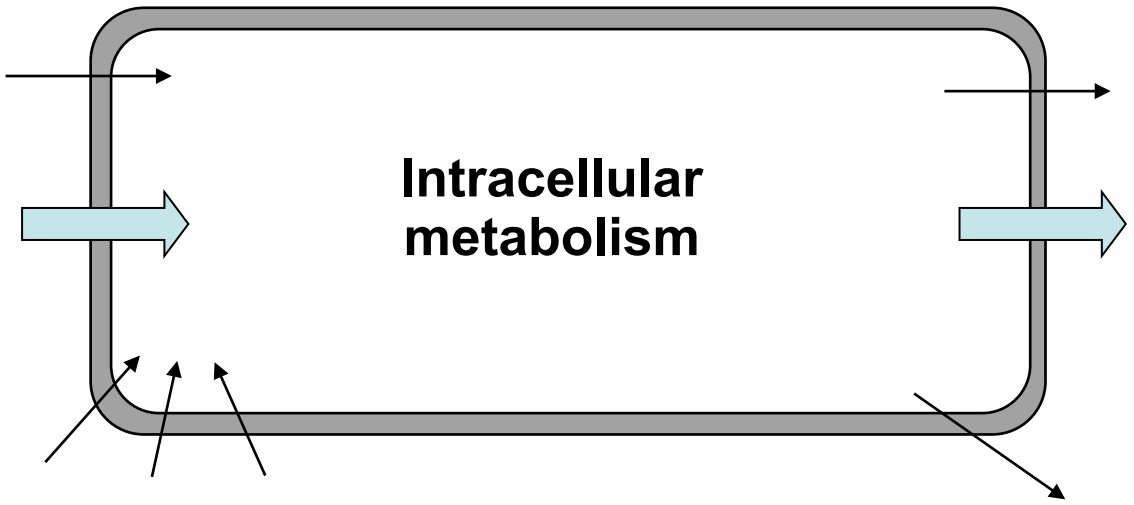
Model class	Level of abstraction	Required information	Example applications
Topological (steady state)	Interaction 	Components and unspecified connections must be known	- Genetic networks - Protein-protein interaction - Metabolite-protein interaction
Steady state			
Stoichiometric (steady state)	Reaction stoichiometry 	Mass and energy balances thermodynamics (directionality)	Metabolic networks - flux balance analysis - elementary flux modes - .....
Mechanistic (dynamic)	Enyzme mechanism and regulation 	Kinetic parameters	Kinetic models (including regulation)

our focus ->

The Problem to be Addressed:
What Goes on Inside the Cell?

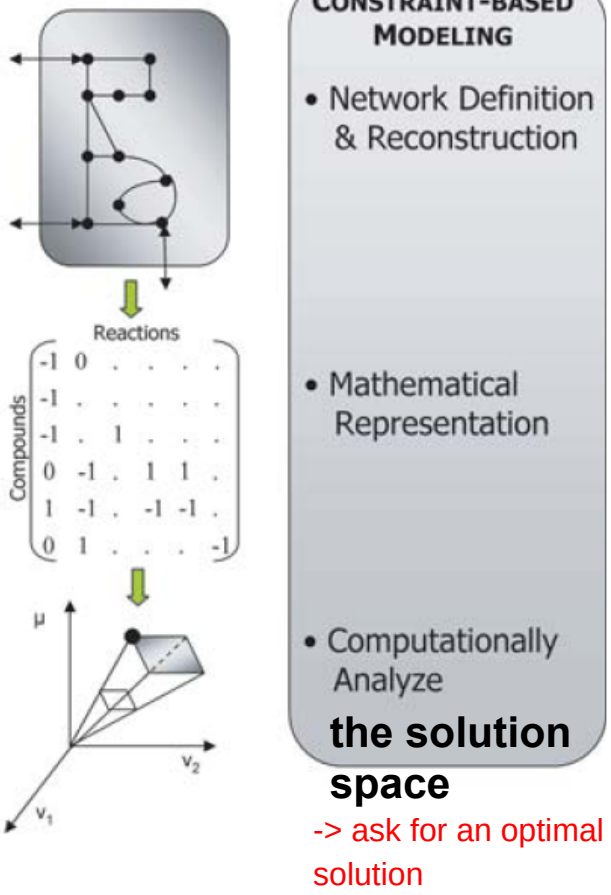
Given some measureable fluxes in/out of a cellular system, what is the underlying intracellular metabolism?
In steady state!

Fluxes IN **Fluxes OUT**



Constraint-Based Stoichiometric Modeling

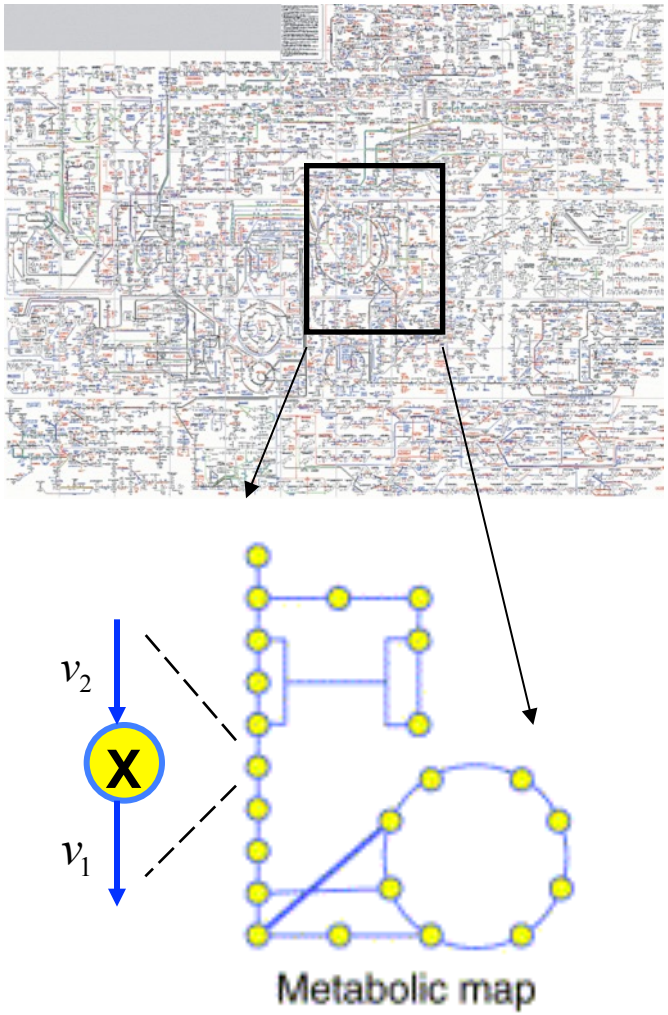
- Modeling and simulation concept primarily suited for metabolism
- Based primarily on mass balances (no parameters!)
- **Many possible solutions exist how fluxes could be balanced!**
- **For steady state only !!**
- What is the idea behind **constraint-based modeling?**
  - Not so much to make specific predictions, but rather to exclude impossible states in the model and analyze „what is left“!



For constraint-based modeling, we need to reconstruct a stoichiometric network model, ideally comprising as many metabolic reactions as possible!

Reconstruction (1)

X = metabolite
v = flux



Mass balance for steady-state:

$$\frac{dX}{dt} = -v_1 + v_2 = 0$$

$$\frac{dX_i}{dt} = s_{i1} \bullet v_1 + \ldots + s_{in} \bullet v_n = 0$$

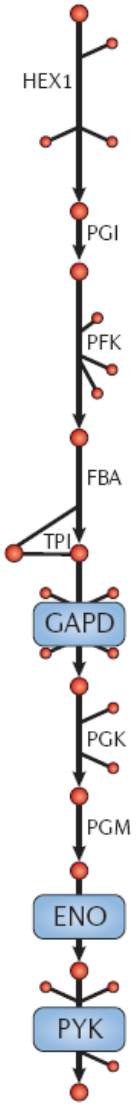
Stoichiometric matrix (S)

n reactions

m metabolites

$$\begin{bmatrix} s_{i1} & \cdots & s_{in} \\ \vdots & \ddots & \vdots \\ s_{m1} & \cdots & s_{mn} \end{bmatrix} \bullet \begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix} = \vec{0}$$

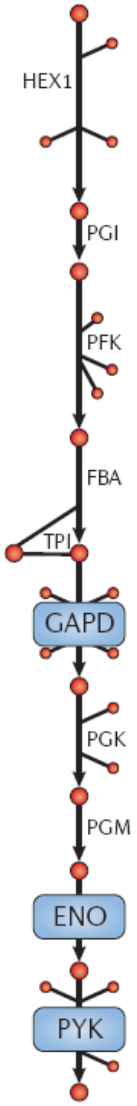
Reconstruction (2)



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
ATP	−1
GLC	−1
ADP	1
G6P	1
H	1
F6P	0
FDP	0
DHAP	0
G3P	0
NAD	0
PI	0
13DPG	0
NADH	0
3PG	0
2PG	0
PEP	0
H_2O	0
PYR	0

Reconstruction (2)

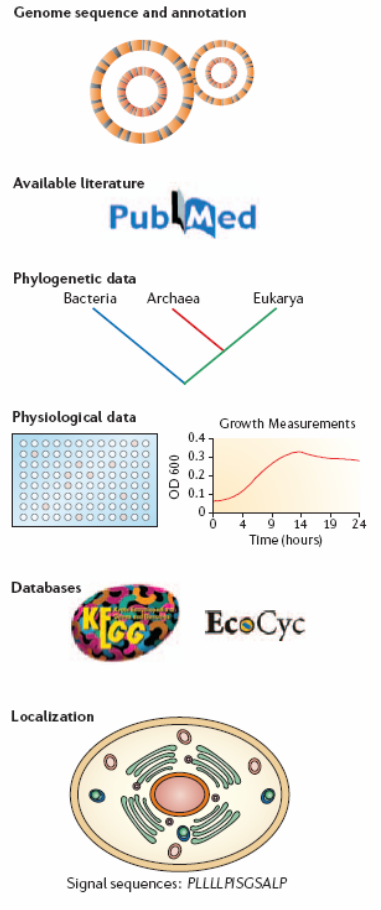


Abbreviation	Glycolytic reactions	Genes
HEX1	[c]GLC + ATP → G6P + ADP + H	*glk*
PGI	[c]G6P ↔ F6P	*pgi*
PFK	[c]ATP + F6P → ADP + FDP + H	*pfkA, pfkB*
FBA	[c]FDP ↔ DHAP + G3P	*fbaA, fbaB*
TPI	[c]DHAP ↔ G3P	*tpiA*
GAPD	[c]G3P + NAD + PI ↔ 13DPG + H + NADH	*gapA, gapC1, gapC2*
PGK	[c]13DPG + ADP ↔ 3PG + 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

Reconstruction (3)

Where do we get the reactions for a given organisms from?



- Textbooks
- Scientific literature
- Databases (KEGG, Brenda, MetaCyc etc)
- Genome sequences

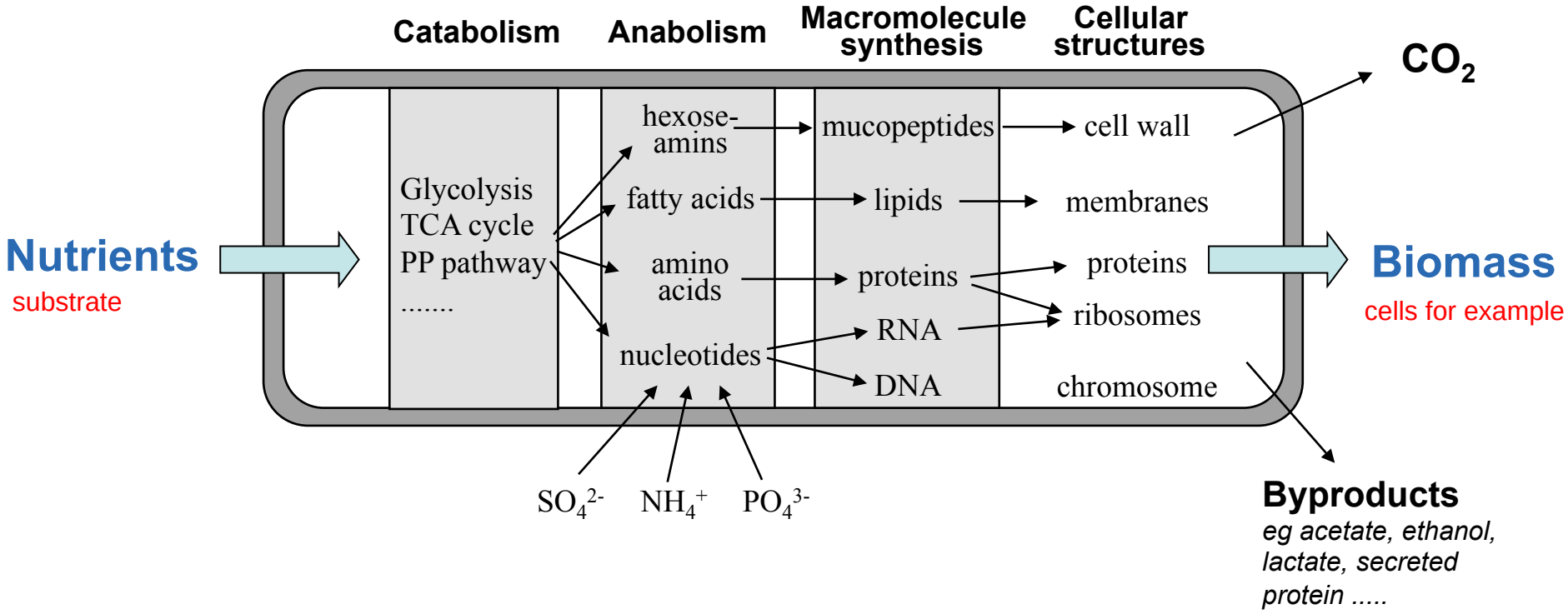
Genome-scale: Reconstruct matrix for the about 1000 reactions of metabolism

What other issues do we need to take care off?

- Reaction stoichiometry and directionality
- Cofactor specificity
- Localization of reactions (compartments)
- Special reactions (eg respiratory chains)

Reconstruction (4) „The Biomass Flux“

To account for new cells being made, we need biomass composition (“stoichiometry“) and energy for assembly/polymerization!



Biosynthetic Building Blocks

All cellular macromolecules are synthesized from 11 carbon precursors and 3 cofactors

Glucose-6-P	P-enol-pyruvate
Fructose-6-P	Pyruvate
Ribose-5-P	Acetyl-CoA
Erythrose-4-P	α-Ketoglutarate
Triose-P	(Succinyl-CoA)
3-Phosphoglycerat	Oxaloacetate

All are intermediates of glycolysis, TCA cycle or PP pathway

But how much is needed of each?

In the following we take a textbook approach. Later we will do it computationally.

Cofactors:
ATP, NADPH, NADH

how many biomass units do we get?

Physiology of the bacterial cell: A molecular approach
Neidhardt, Ingraham, Schaechter, Sinauer Ass. 1990

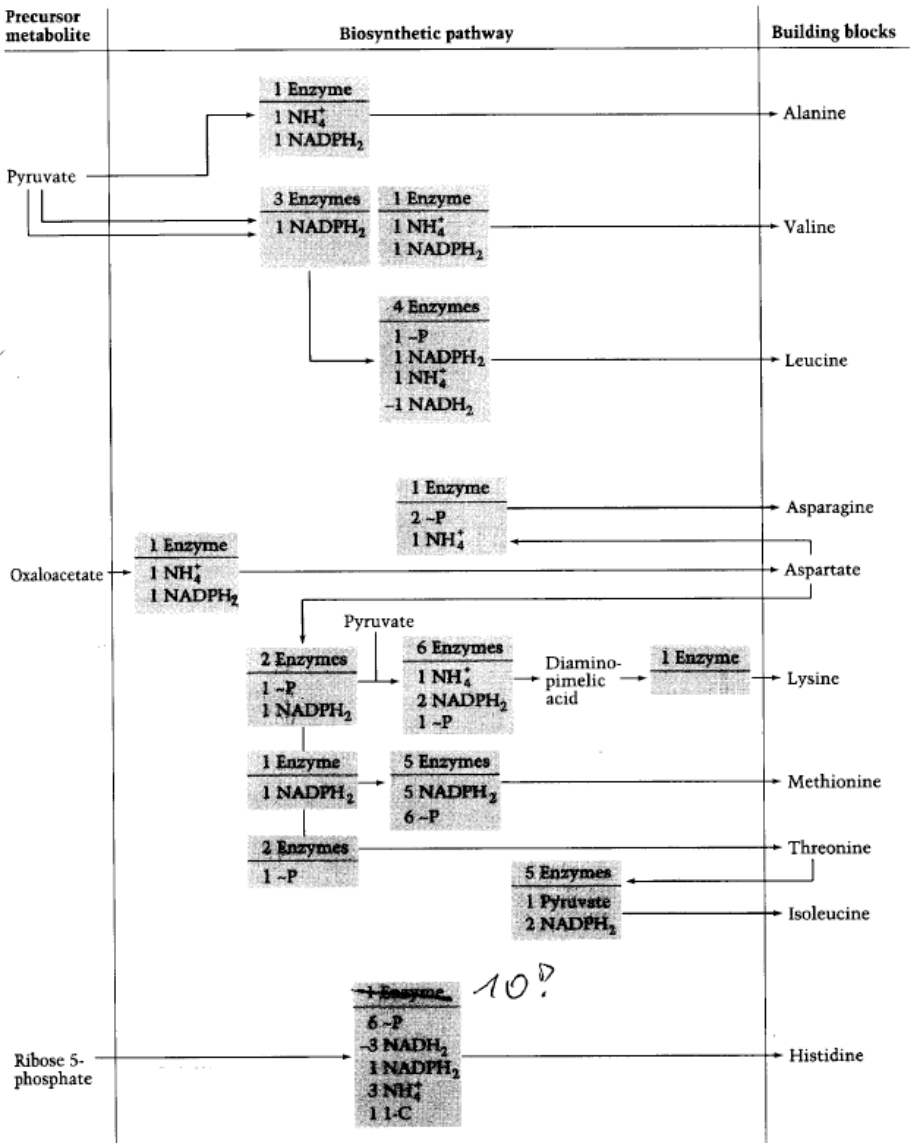
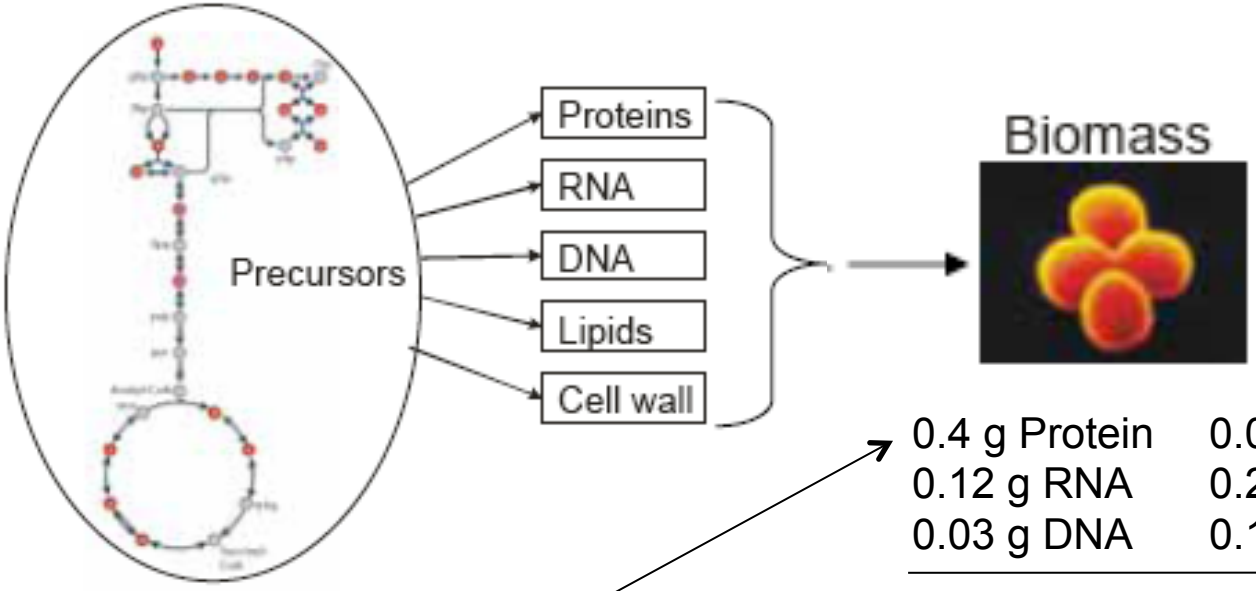


Table 1. Building blocks needed to produce 1 g of *E. coli* protoplasm

Building block	Amount present in *E. coli* B/r (μmol/g dried cells)	Metabolites[a]	Cost of making 1 μmol of each of these building blocks (μmol/μmol) ATP	NADH	NADPH	1-C	NH_4^+	S
Protein amino acids								
Alanine	488	1 pyr	0	0	1	0	1	0
Arginine	281	1 αkg	7	−1	4	0	4	0
Asparagine	229	1 oaa	3	0	1	0	2	0
Aspartate	229	1 oaa	0	0	1	0	1	0
Cysteine	87	1 pga	4	−1	5	0	1	1
Glutamate	250	1 αkg	0	0	1	0	1	0
Glutamine	250	1 αkg	1	0	1	0	2	0
Glycine	582	1 pga	0	−1	1	−1	1	0
Histidine	90	1 penP	6	−3	1	1	3	0
Isoleucine	276	1 oaa, 1 pyr	2	0	5	0	1	0
Leucine	428	2 pyr, 1 acCoA	0	−1	2	0	1	0
Lysine	326	1 oaa, 1 pyr	2	0	4	0	2	0
Methionine	146	1 oaa	7	0	8	1	1	1
Phenylalanine	176	1 eryP, 2 pep	1	0	2	0	1	0
Proline	210	1 αkg	1	0	3	0	1	0
Serine	205	1 pga	0	−1	1	0	1	0
Threonine	241	1 oaa	2	0	3	0	1	0
Tryptophan	54	1 penP, 1 eryP, 1 pep	5	−2	3	0	2	0
Tyrosine	131	1 eryP, 2 pep	1	−1	2	0	1	0
Valine	402	2 pyr	0	0	2	0	1	0
RNA nucleotides								
ATP								0
GTP								0
CTP								0
UTP								0
DNA nucleotides								
dATP	24.7	1 penP, 1 pga	11	−3	2	1	5	0
dGTP	25.4	1 penP, 1 pga	13	−3	1	1	5	0
dCTP	25.4	1 penP, 1 oaa	9	0	2	0	3	0
dTTP	24.7	1 penP, 1 oaa	10.5	0	3	1	2	0
Lipid components								
Glycerol phosphate	129	1 triosP	0	0	1	0	0	0
Serine	129	1 pga	0	−1	1	0	1	0
$C_{16:0}$ fatty acid (43%)		8 acCoA	7	0	14	0	0	0
$C_{16:1}$ fatty acid (33%)		8 acCoA	7	0	13	0	0	0
$C_{18:1}$ fatty acid (24%)		9 acCoA	8	0	15	0	0	0
Average fatty acid	258	8.2 acCoA	7.2	0	14	0	0	0

Stoichiometric table !

Stoichiometric Modeling of Biomass Demands



An OUT-flux in the model

0.4 g Protein
0.12 g RNA
0.03 g DNA
0.07 g Lipid
0.24 g Cell wall
0.14 g Solutes

1 g cells (dry weight)

From the stoichiometric table ....

Protein composition:
0.492 mmol alanine + … + 0.514 mmol valine + 39.94 mmol ATP => 1 g Protein + 39.94 ADP + 39.94 P_i

Definition of biomass flux depends on:

- anabolic biochemistry
- macromolecular synthesis
- composition of macromolecules in a cell
- cellular content of macromolecules
  (does that remain constant?)

Biomass Stoichiometry:
Building Block Requirements for 1 g of Cells

Metabolite	Amount required
Glucose-6-P	205
Fructose-6-P	71
Ribose-5-P	898
Erythrose-4-P	361
Triose-P	129
3-Phosphoglycerat	1496
P-enol-pyruvate	720
Pyruvate	2833
Acetyl-CoA	3748
α-ketoglutarate	1079
Succinyl-CoA	10
Oxaloacetate	1787

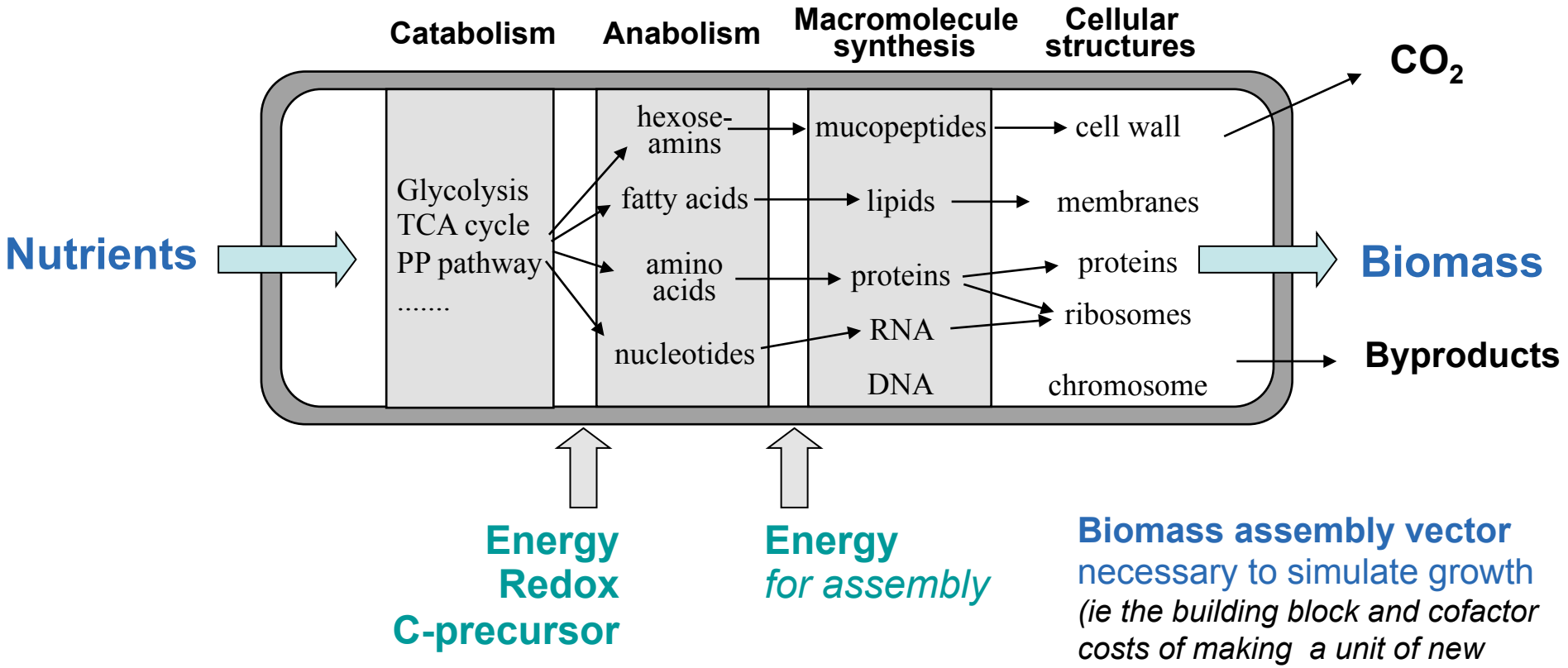
NADH	-3547
NADPH	18225
ATP	18485

μmol / g_{cells}

These are the synthesis costs for building blocks like nucleotides and amino acids .....

What is missing in terms of cost?
What is the currency for these additional costs?

Reconstruction (4) Biomass Vector



Biomass assembly vector
necessary to simulate growth
*(ie the building block and cofactor costs of making a unit of new cells, including stoichiometric composition **and** energy for polymerization!*

To quantify the biomass (OUT) flux, we need:
Biomass composition ("stoichiometry")
& energy for polymerization!

Reconstruction (5) Adding constraints

Defining reaction directions

- **Irreversibility/Reversibility (internal reactions)**

$0 \leq v_i \leq \beta_i$ for v_i irreversible

$\alpha_i \leq v_i \leq \beta_i$ for v_i reversible

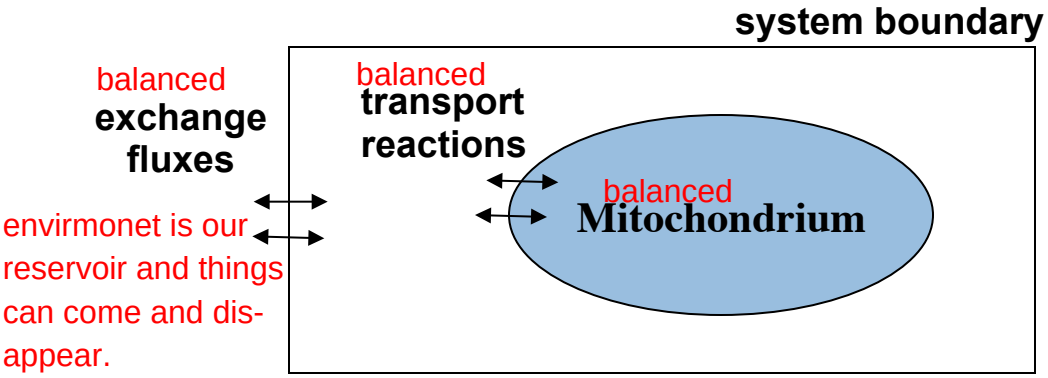
we can set reversibility under the assumed conditions of the cell and by using biochemical logic

Ex. when there is ample ATP, reaction surely wont go the other way around too (reactions assumed to be irreversible)

On what principle is that definition based on?

Matrix Consists of 3 Sets of Processes

- Intracellular reactions R
- Intracellular transport reactions T
- Exchange fluxes E
  - used to define growth conditions (eg glucose uptake rate, lactate secretion rate, CO_2 evolution rate, O_2 consumption rate)
  - biomass production rate μ is part of this set of fluxes



envirmonet is our reservoir and things can come and dis-appear.

Ex. glucose disappears, but there are more cells at the end.
in itself all is balanced, but the conc. of glucose in medium changes, and there are more cells and conc. of CO2 changes too

Constraints on exchange fluxes define growth conditions

$0 \le e_i \le \beta_i$ **for products**

$\alpha_i \le e_i \le 0$ **for substrates**

Which fluxes are balanced – which are not ?

ie which metabolite [conc] are in steady state?

Genome-scale metabolic reconstructions

Organism	Genes	SKI	N_g	N_m	N_r	Status	Refs
Bacteria							
Bacillus subtilis	4,225	4.8	614	637	754	C, E	95
Escherichia coli	4,405	55.1	904 720 961	625 438 NA	931 627 1,107	C, E C, E C	39 92 53
Francisella tularensis	1,804	ND	350‡	NA	429	C	68
Geobacter sulfurreducens	3,530	ND	588	541	523	C, E	105
Haemophilus influenzae	1,775	8.9	296 400	343 451	488 461	C, E C, E	96 97
Helicobacter pylori	1,632	13	341 291 301‡‖	485 340 442‖	476 388 533‖	C, E C, E C	61 98 63
Lactococcus lactis	2,310	ND	358	422	621	C,E	99
Mannheimia succiniciproducens	2,463	ND	335	352	373	C, E	100
Pseudomonas aeruginosa	5,640	5.7	546 718	467 623	542 800	C, E C	¶ 67
Staphylococcus aureus	2,702	16	619	571	641	C, E	4
Streptomyces coelicolor	8,042	0.13	700	500	700	C, E	36
Archaea							
Methanococcus jannaschii	1,821	0.3	436‡	510	609	C	64
Methanosarcina barkeri	5,072	ND	692	558	619	C, E	106
Eukarya							
Arabidopsis thaliana	28,848	ND	1,418	NA	894	C	66
Homo sapiens	28,783	48.5	2,709‡	661	1,093	C	65
Mus musculus	28,287	15.6	1,156§	872	1,220	C, E	94
Plasmodium falciparum	5,342	ND	737‡	525	697	C	3
Saccharomyces cerevisiae	6,183	10.6	750 708	646 584	1,149 1,175	C, E C, E	45 93

*Several non-curated automated reconstructions are also available from KEGG[52] and BioCyc[69]. ‡Only enzyme numbers were reported. §Genes as reported[101]. ‖Latest numbers from HpCyc[63]. ¶J. Edwards, personal communication. C, a curated network; E, a network that is evaluated using computational modelling methods that are based on a stoichiometric matrix; NA, not available; ND, not determined; N_g, number of genes; N_m, number of metabolites; N_r, number of reactions; SKI, reported species knowledge index[12].

At present > 1000 models

- We have learned …
  - what genome-scale metabolic networks models are (they only consist of reaction stoichiometry!)
  - how they are reconstructed from known biochemistry
- What do we have now? A model consisting of many mass balance equations that is greatly **underdetermined** – i.e. **many possible solutions exist** !!!

Exercise 6: FBA

Goal

- Learning the basic ingredients of Flux Balance Analysis
- Simulate and interpret flux distributions in a toy metabolic model