GEMs

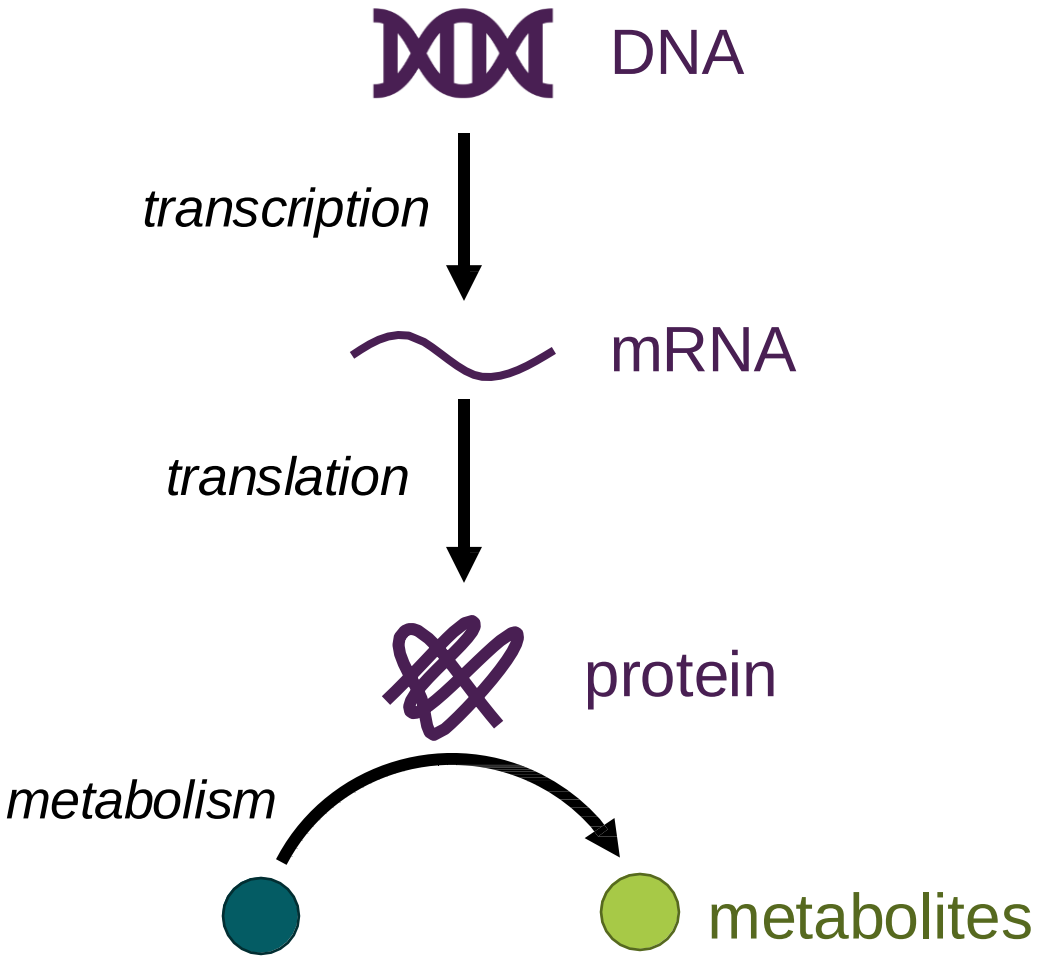
Structure & Developments

NBIS Omics Integration and Systems Biology workshop
Oct 2024, Lund University

Rasool Saghaleyni
National Bioinformatics Infrastructure Sweden (NBIS)
Science for Life Laboratory (SciLifeLab)
Chalmers University of Technology
rasool.saghaleyni@scilifelab.se

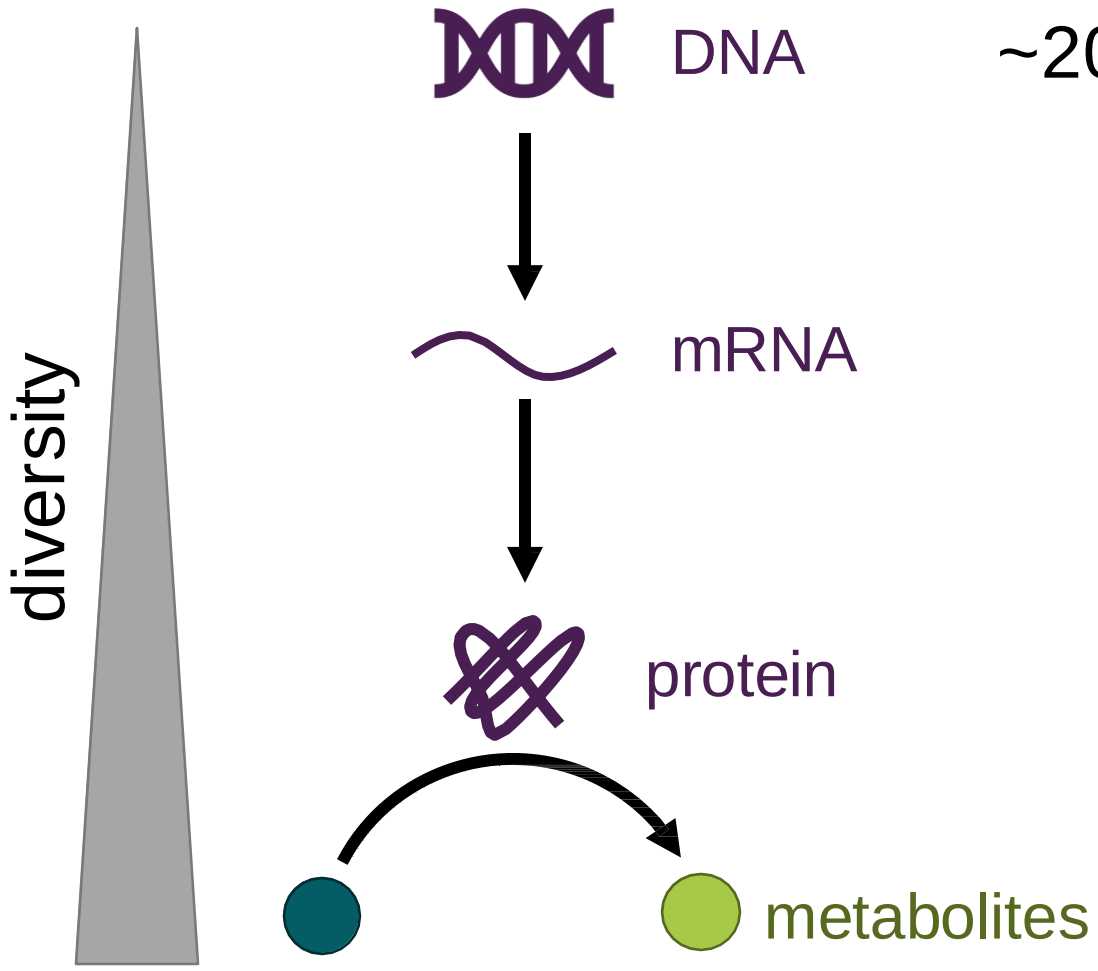
These slides are adapted from previous course materials created by **Jonathan Robinson**. Some updates have been made for this iteration.

Background



Metabolism provides the **energy** and **building blocks** necessary to sustain life.

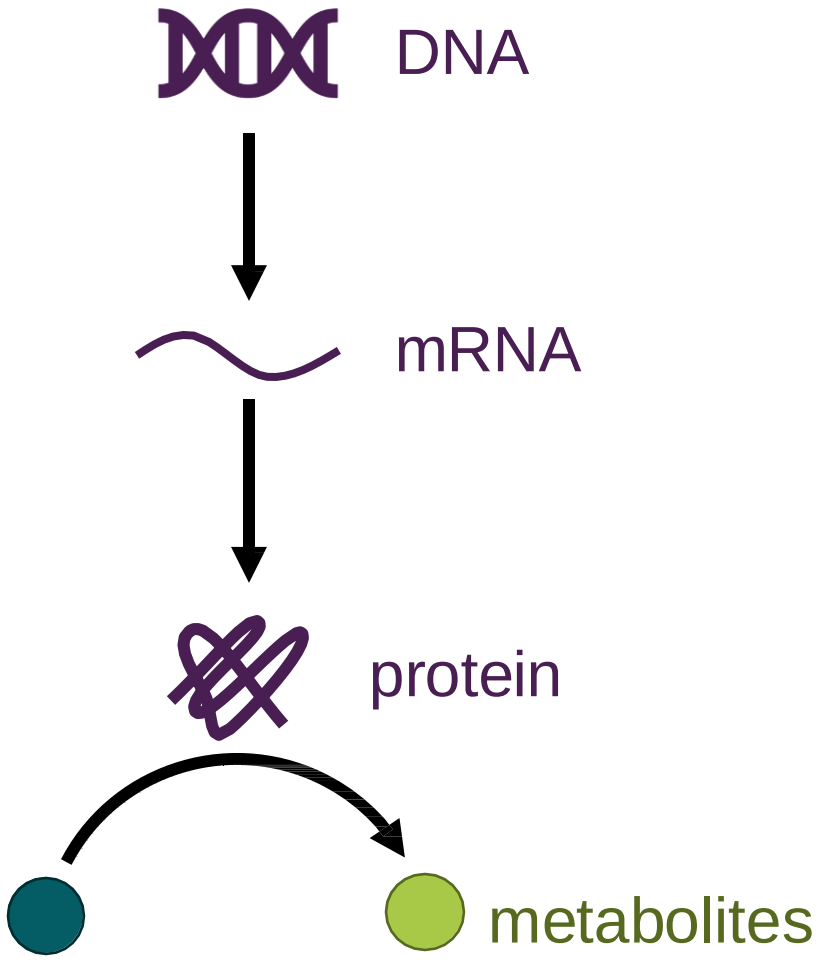
Background



~20,000 genes
(protein-coding)

>100,000 metabolites

Background



stable

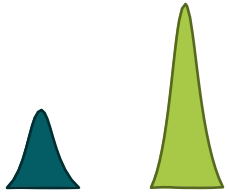
relatively stable

stable

ranges from stable
to very unstable

Background

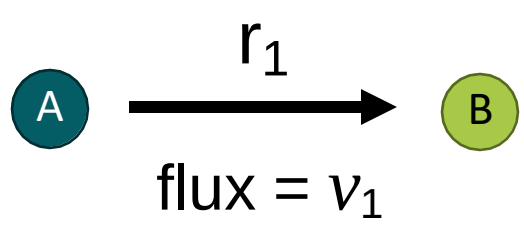
We can generally measure metabolite concentrations



…but what is often important is the flow or **flux** of metabolites through the reactions.

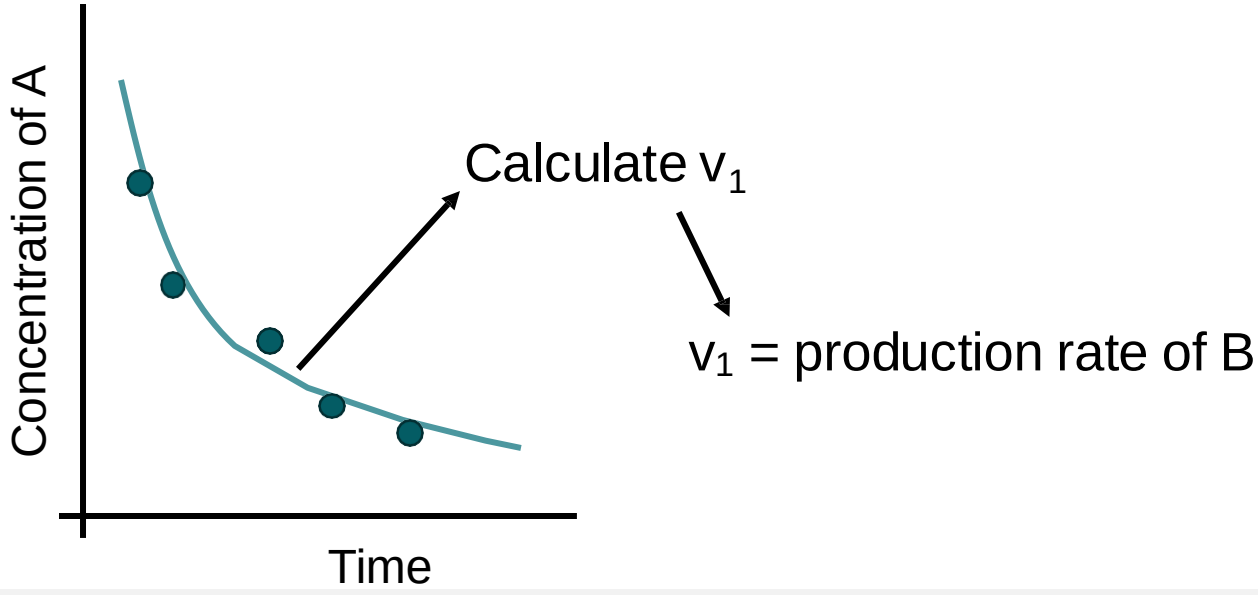
Background

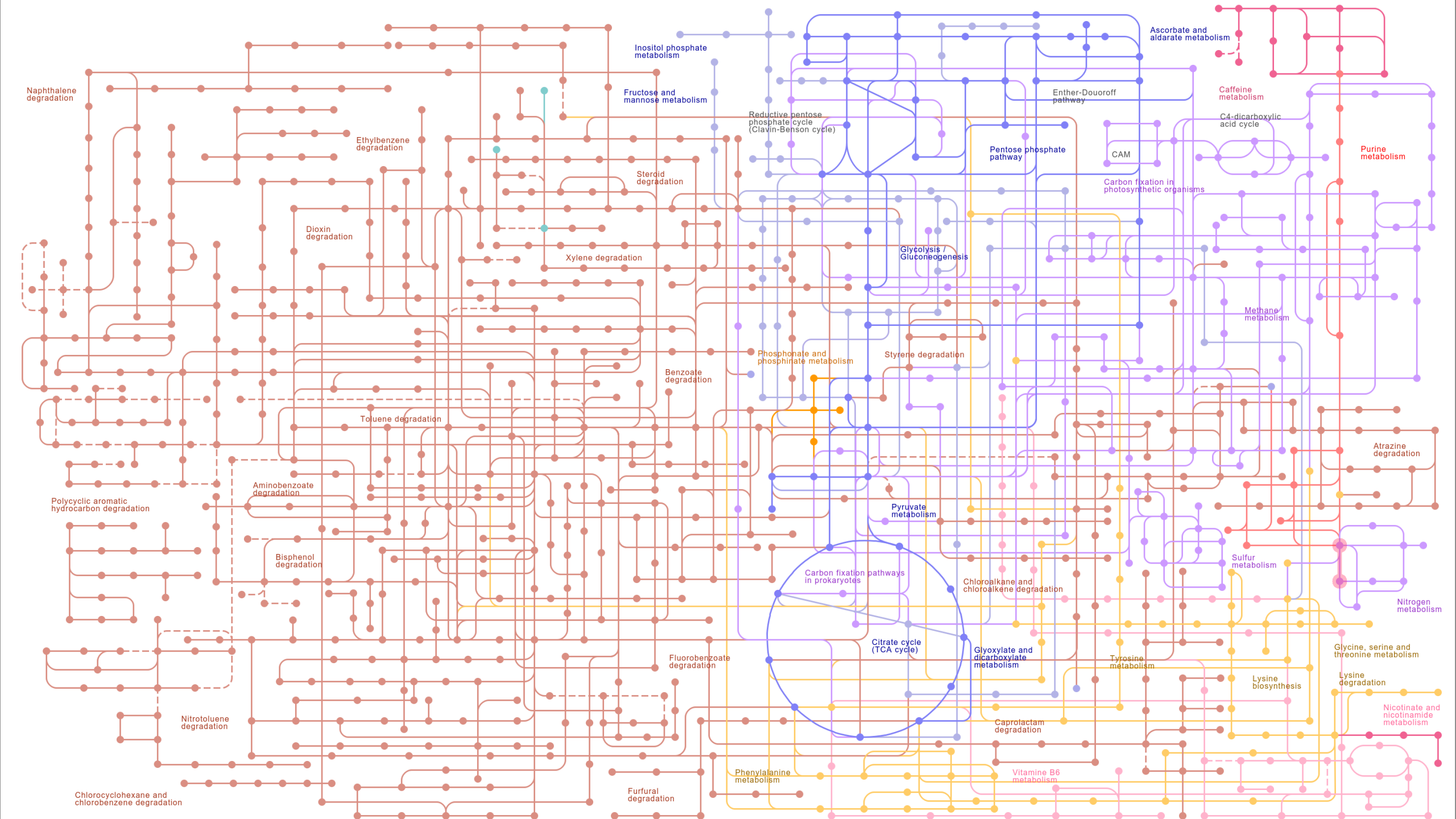
Assume that we want to know the production rate of **B**, but can only measure the concentration of **A**



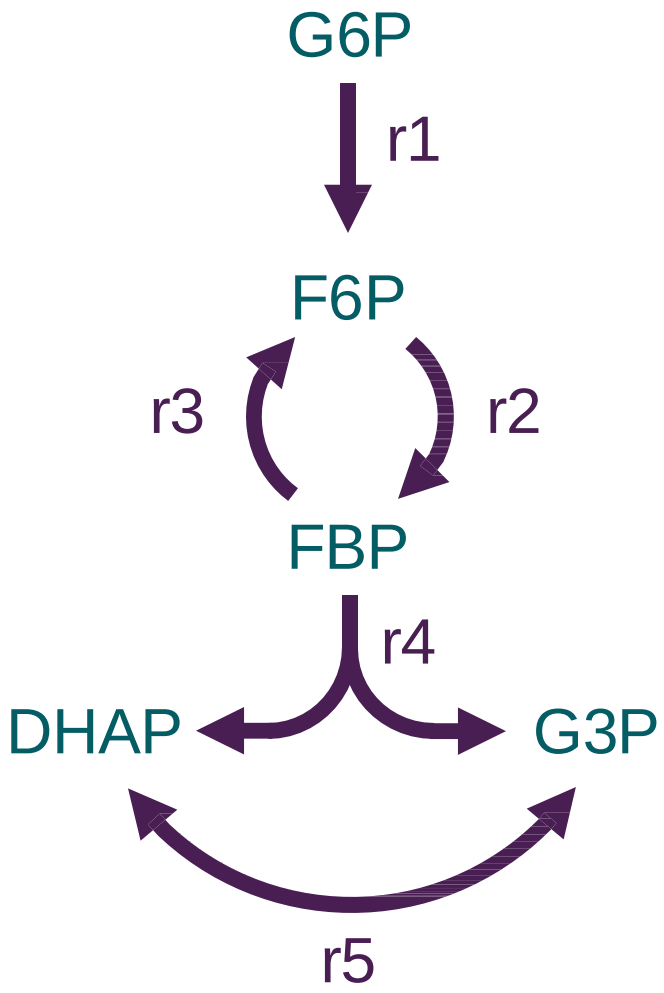
$$\frac{dA}{dt} = -v_1$$

$$\frac{dB}{dt} = v_1$$




Naphthalene degradation
Ethylbenzene degradation
Dioxin degradation
Steroid degradation
Xylene degradation
Benzoate degradation
Toluene degradation
Aminobenzoate degradation
Polycyclic aromatic hydrocarbon degradation
Bisphenol degradation
Nitrotoluene degradation
Chlorocyclohexane and chlorobenzene degradation
Fluorobenzoate degradation
Furfural degradation
Inositol phosphate metabolism
Fructose and mannose metabolism
Reductive pentose phosphate cycle (Clavin-Benson cycle)
Pentose phosphate pathway
Ascorbate and aldarate metabolism
Enther-Douoroff pathway
CAM
Carbon fixation in photosynthetic organisms
Caffeine metabolism
C4-dicarboxylic acid cycle
Purine metabolism
Glycolysis / Gluconeogenesis
Methane metabolism
Phosphonate and phosphinate metabolism
Styrene degradation
Atrazine degradation
Pyruvate metabolism
Carbon fixation pathways in prokaryotes
Chloroalkane and chloroalkene degradation
Sulfur metabolism
Nitrogen metabolism
Citrate cycle (TCA cycle)
Glyoxylate and dicarboxylate metabolism
Tyrosine metabolism
Glycine, serine and threonine metabolism
Lysine biosynthesis
Lysine degradation
Caprolactam degradation
Nicotinate and nicotinamide metabolism
Phenylalanine metabolism
Vitamine B6 metabolism

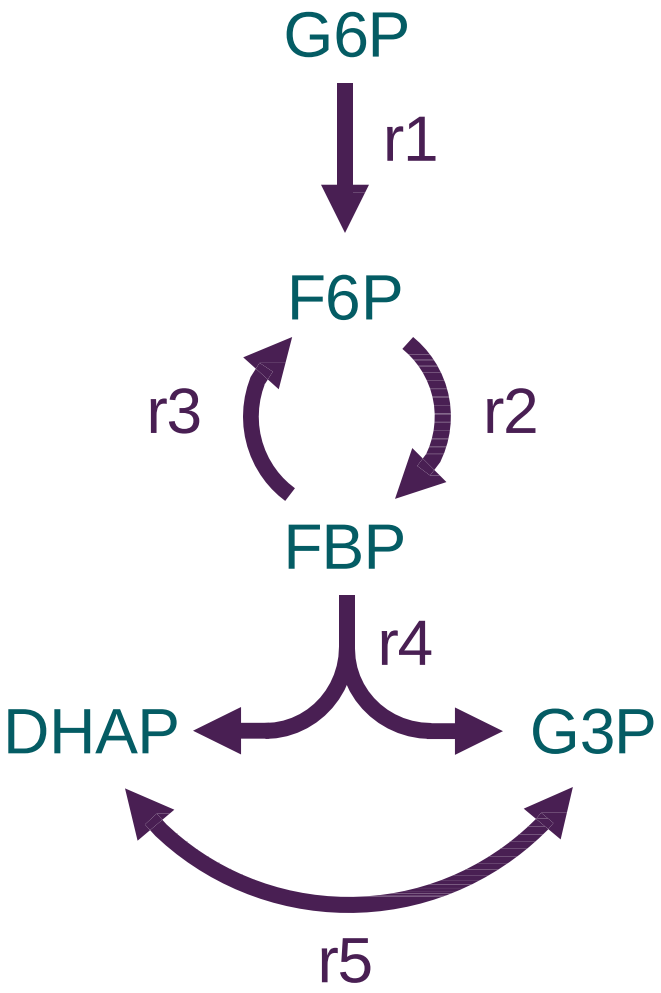
The Stoichiometric Matrix



Reactions

Metabolites	r1
G6P	**−1**
F6P	**1**
FBP	0
DHAP	0
G3P	0

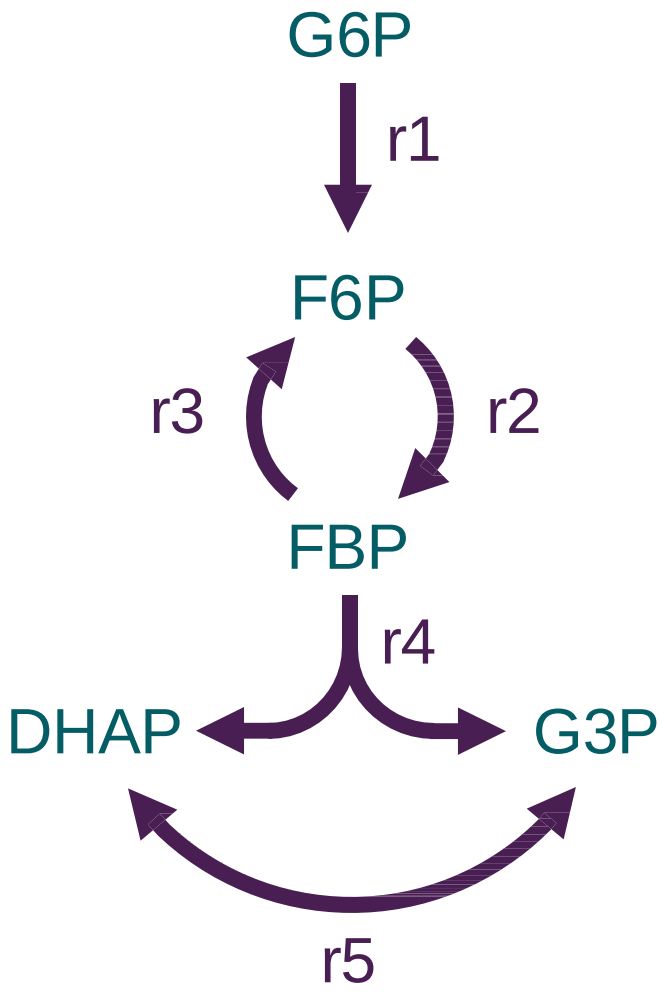
The Stoichiometric Matrix



Reactions

Metabolites	r1	r2
G6P	**−1**	0
F6P	**1**	**−1**
FBP	0	**1**
DHAP	0	0
G3P	0	0

The Stoichiometric Matrix

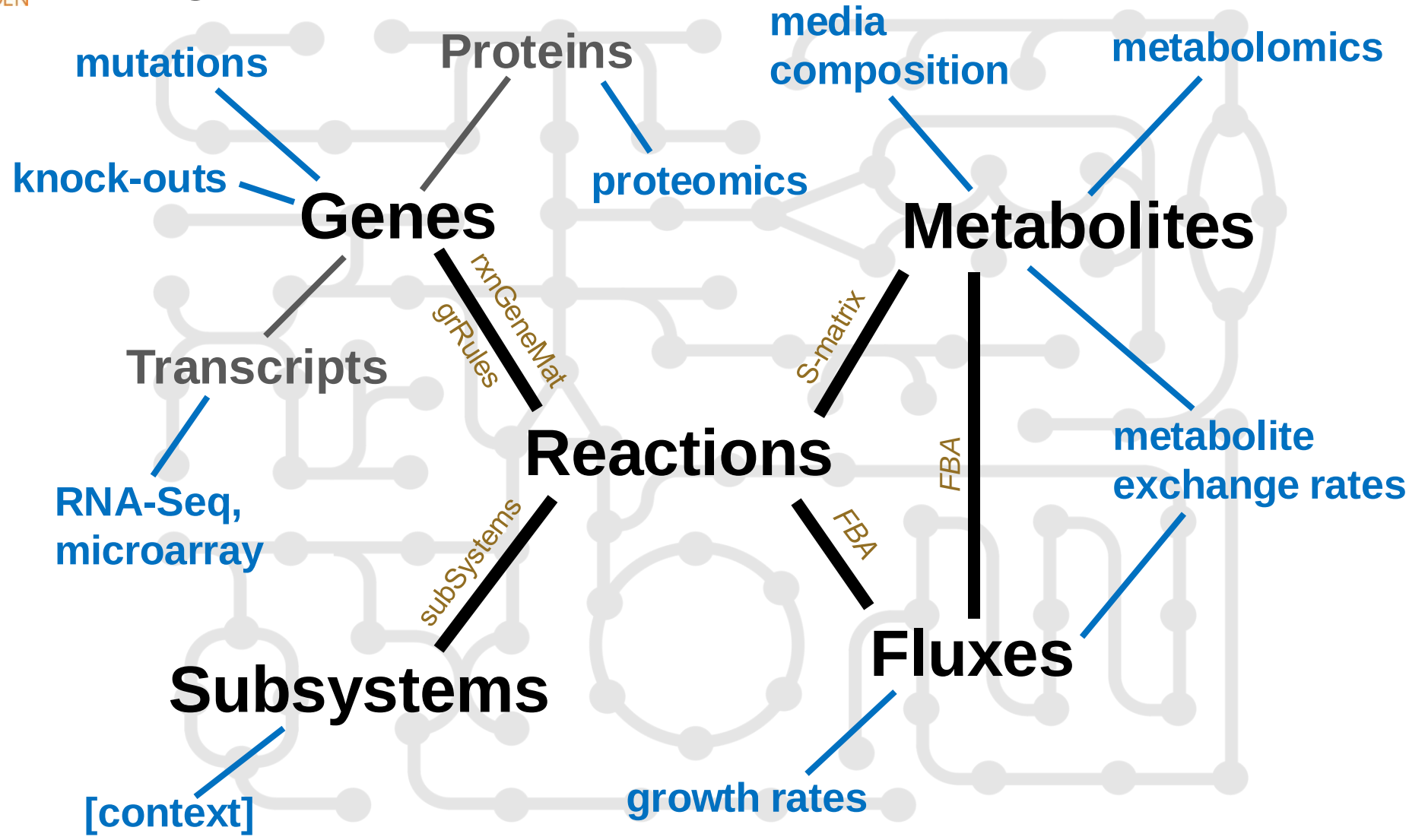


Reactions

Metabolites	r1	r2	r3	r4	r5
G6P	**−1**	0	0	0	0
F6P	**1**	**−1**	**1**	0	0
FBP	0	**1**	**−1**	**−1**	0
DHAP	0	0	0	**1**	**−1**
G3P	0	0	0	**1**	**1**

S

Genome-scale metabolic models (GEMs) for data integration



Genome-scale model (GEM)

Chemical formula
Charge
InChI code
Other external IDs
...

									...
				…	…	…	…	…	**Other IDs**
				…	…	…	…	…	**Name**
KEGG ID	**Compartment**	**Name**	**Symbol**	r1	r2	r3	r4	r5	**Symbol**
C00668	cytosol [c]	glucose 6-phosphate	G6P	**−1**	0	0	0	0	
C00085	cytosol [c]	fructose 6-phosphate	F6P	**1**	**−1**	**1**	0	0	
C00354	cytosol [c]	fructose-1,6-bisphosphate	FBP	0	**1**	**−1**	**−1**	0	
C00111	cytosol [c]	dihydroxyacetone phosphate	DHAP	0	0	0	**1**	**−1**	
C00118	cytosol [c]	glyceraldehyde 3-phosphate	G3P	0	0	0	**1**	**1**	
			...						

Genome-scale model (GEM)

Genes (symbol)	**Proteins** (UniProt ID)
GPI	P06744
n/a	
FBP1, FBP2	P09467, O00757
ALDOA, ALDOB, ALDOC	P04075, P05062, P09972
TPI1	P60174

Transcript IDs

GO Terms

Orthologs

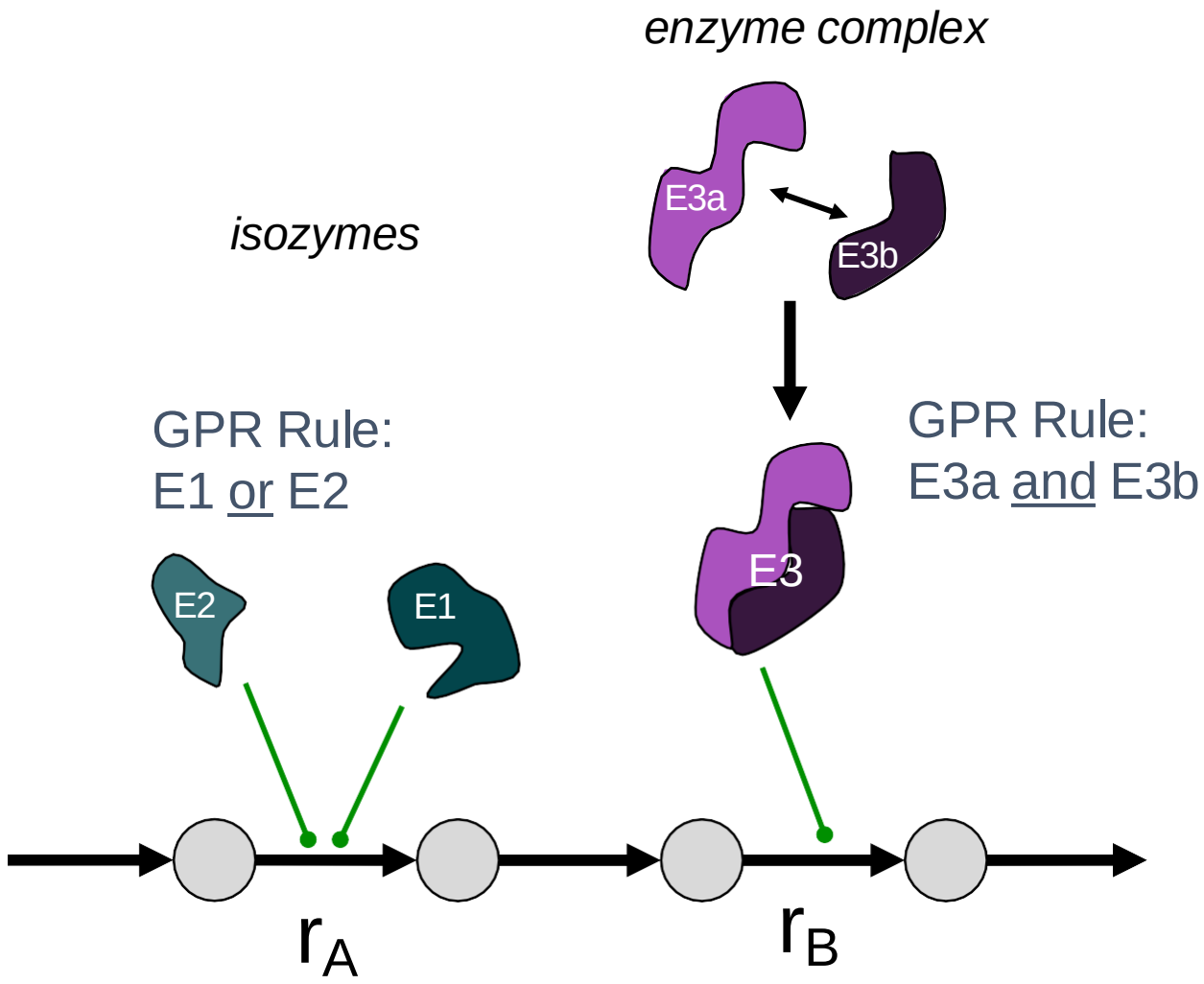
…

Symbol	r1	r2	r3	r4	r5
G6P	**−1**	0	0	0	0
F6P	**1**	**−1**	**1**	0	0
FBP	0	**1**	**−1**	**−1**	0
DHAP	0	0	0	**1**	**−1**
G3P	0	0	0	**1**	**1**

Reactions are linked to genes that encode the enzymes that catalyze the reaction.

These associations are often called “gene-protein rules” (GPR rules)

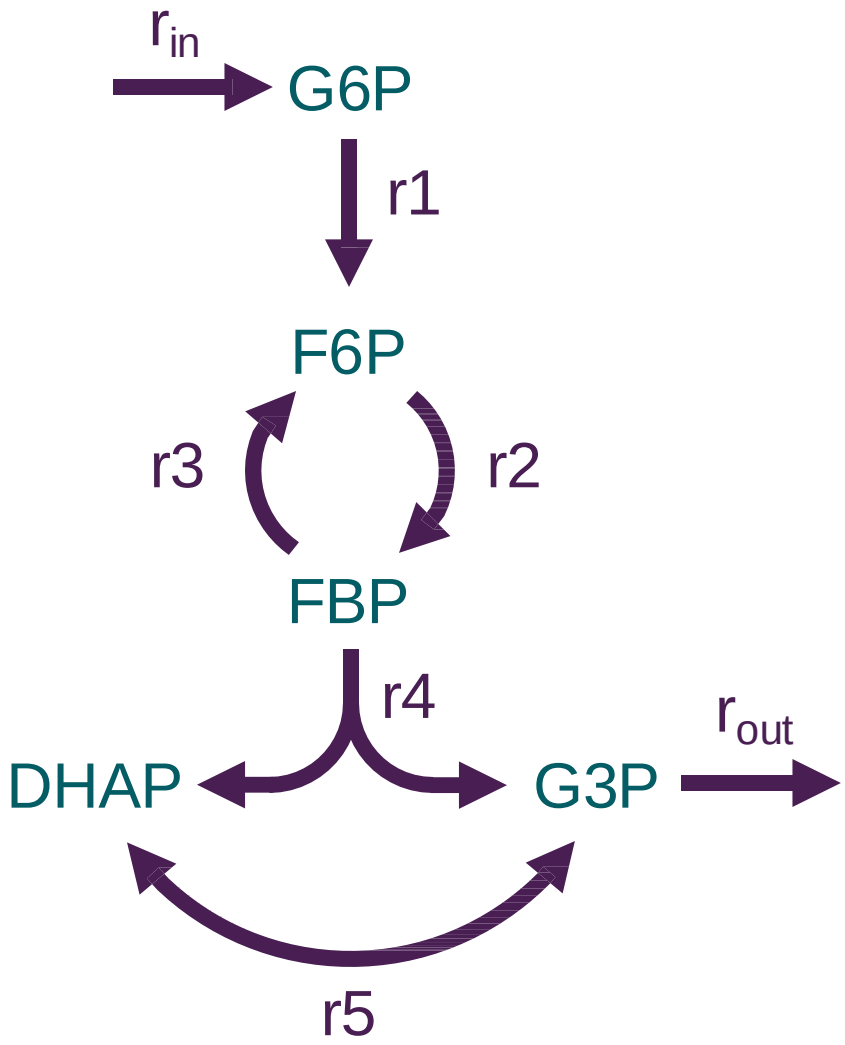
GPR Rules



GPR Rules enable more accurate simulation of gene inactivation/knock-out

Knockout	Effect
E1	none
E2	none
E1 + E2	rA inactive
E3a	rB inactive
E3b	rB inactive
E3a + E3b	rB inactive

Flux Balance Analysis (FBA)

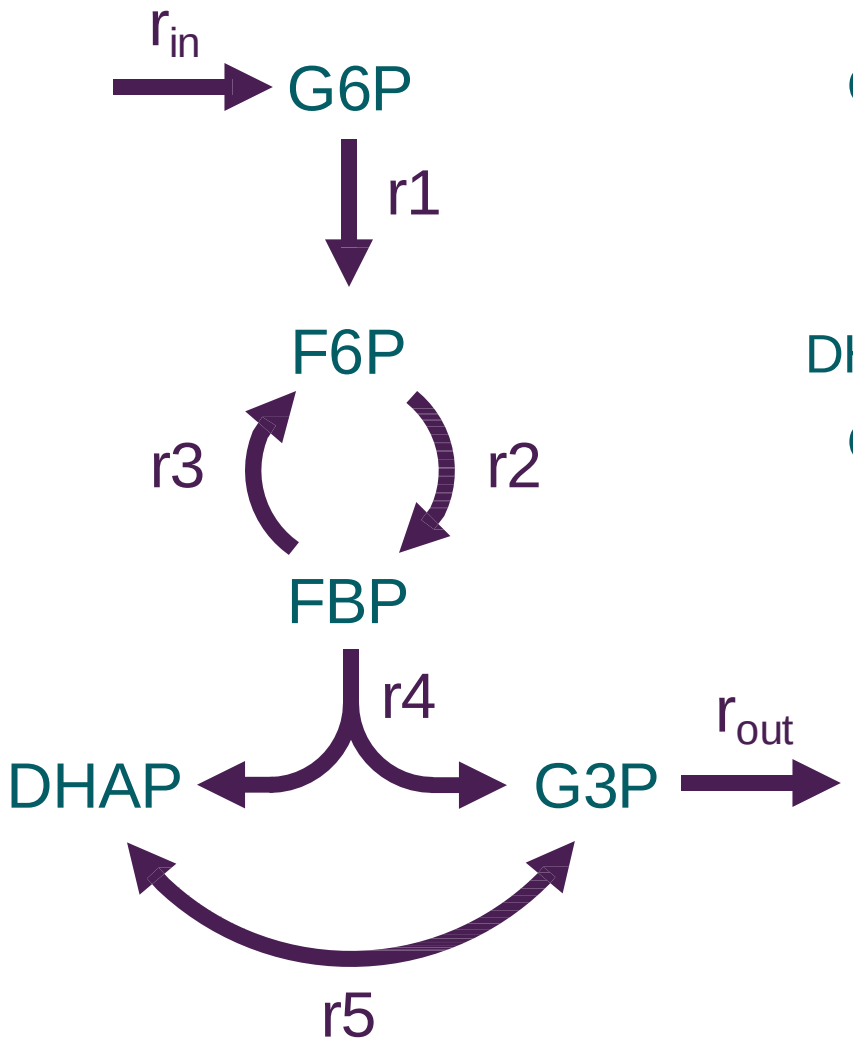


FBA seeks to calculate the reaction **fluxes** (***v***) of a network

The calculation is based on the **conservation of mass**: it cannot be created or destroyed

$$\frac{dX}{dt} = v_{produce} - v_{consume}$$

Flux Balance Analysis (FBA)

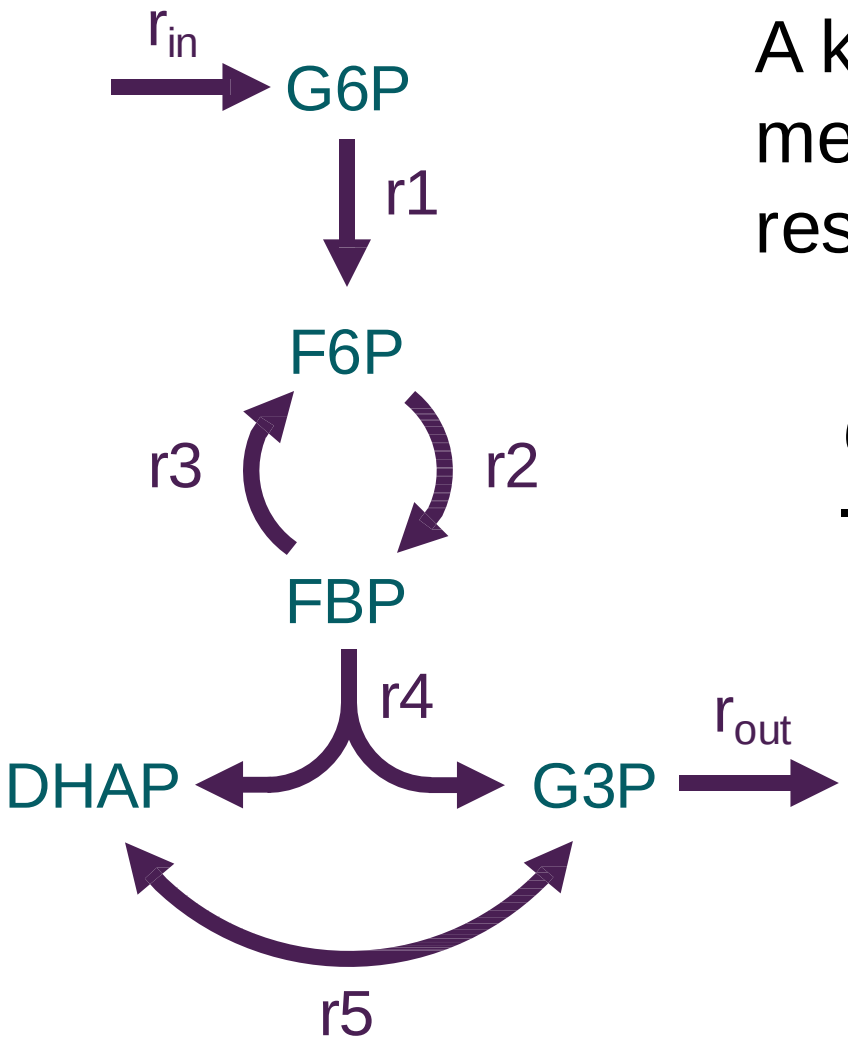


$$
\begin{array}{c} \\ \text{G6P} \\ \text{F6P} \\ \text{FBP} \\ \text{DHAP} \\ \text{G3P} \end{array}
\begin{array}{c}
\begin{array}{ccccccc} \text{r1} & \text{r2} & \text{r3} & \text{r4} & \text{r5} & r_{in} & r_{out} \end{array} \\
\begin{pmatrix}
-1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & -1 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & -1
\end{pmatrix}
\end{array}
\times
\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_{in} \\ v_{out} \end{pmatrix}
=
\begin{pmatrix} \text{dG6P/dt} \\ \text{dF6P/dt} \\ \text{dFBP/dt} \\ \text{dDHAP/dt} \\ \text{dG3P/dt} \end{pmatrix}
$$

$$\frac{d[G6P]}{dt} = -v_{1} + v_{in}$$

$$\frac{d[G3P]}{dt} = v_4 + v_5 - v_{out}$$

Flux Balance Analysis (FBA)



A key assumption to FBA is **steady state**: metabolite concentrations are **constant** with respect to time!

$$\frac{dX}{dt} = v_{produce} - v_{consume} = 0$$

This assumption allows us to **ignore enzyme kinetics**, thus eliminating **many** unknown parameters

Flux Balance Analysis (FBA)

$$
\begin{array}{c}
 \\
\text{G6P} \\
\text{F6P} \\
\text{FBP} \\
\text{DHAP} \\
\text{G3P}
\end{array}
\begin{array}{c}
\begin{array}{ccccccc}
\text{r1} & \text{r2} & \text{r3} & \text{r4} & \text{r5} & r_{in} & r_{out}
\end{array} \\
\begin{pmatrix}
-1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & -1 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & -1 & -1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & -1 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & -1
\end{pmatrix}
\end{array}
\times
\begin{pmatrix}
v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_{in} \\ v_{out}
\end{pmatrix}
=
\begin{pmatrix}
\text{dG6P/dt} \\ \text{dF6P/dt} \\ \text{dFBP/dt} \\ \text{dDHAP/dt} \\ \text{dG3P/dt}
\end{pmatrix}
=
\begin{pmatrix}
0 \\ 0 \\ 0 \\ 0 \\ 0
\end{pmatrix}
$$

$$S \cdot v = 0$$

Flux Balance Analysis (FBA)

We can further constrain the solution space by limiting reaction fluxes based on their reversibility:

Irreversible reactions

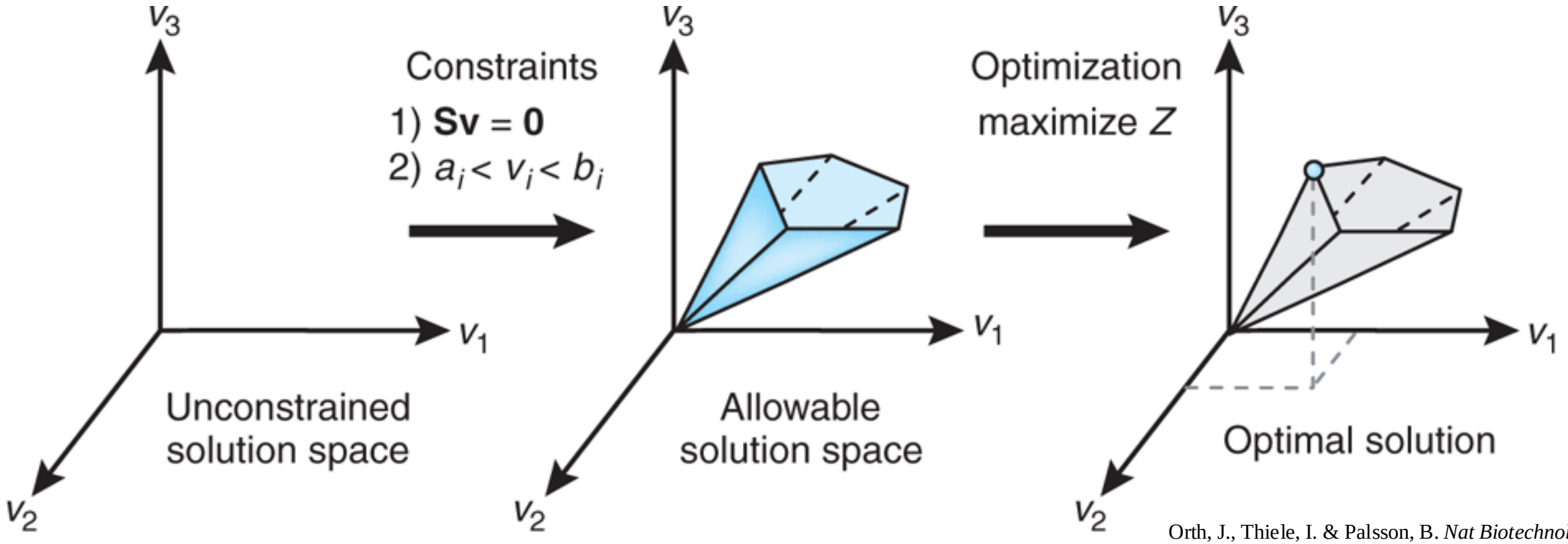
$0 \leq v \leq ub$

Reversible reactions

$lb \leq v \leq ub$

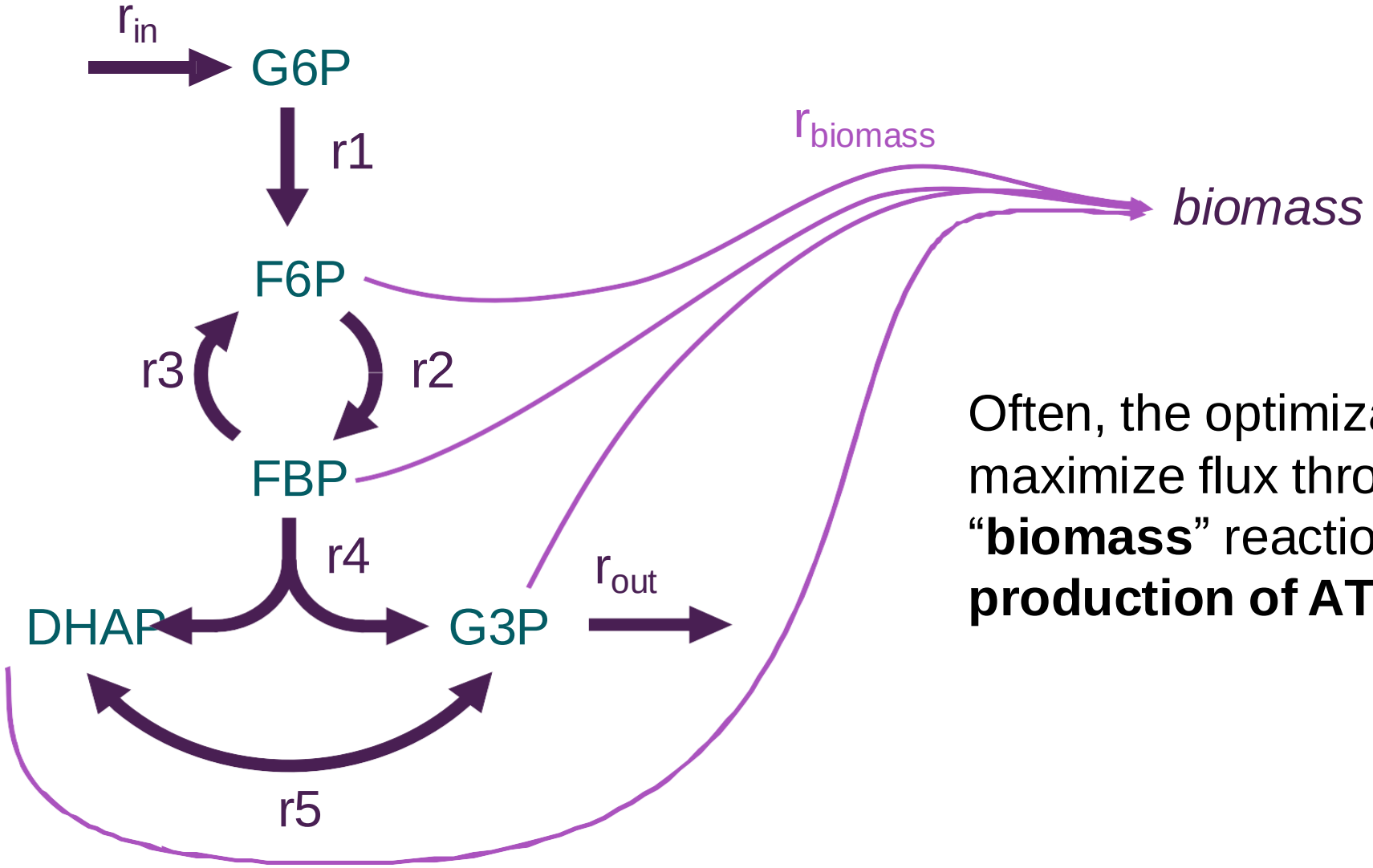
Flux Balance Analysis (FBA)

Since the problem is still **under-defined**, FBA uses linear **optimization** to identify a solution that maximizes (or minimizes) some **objective**



Orth, J., Thiele, I. & Palsson, B. *Nat Biotechnol* (2010).

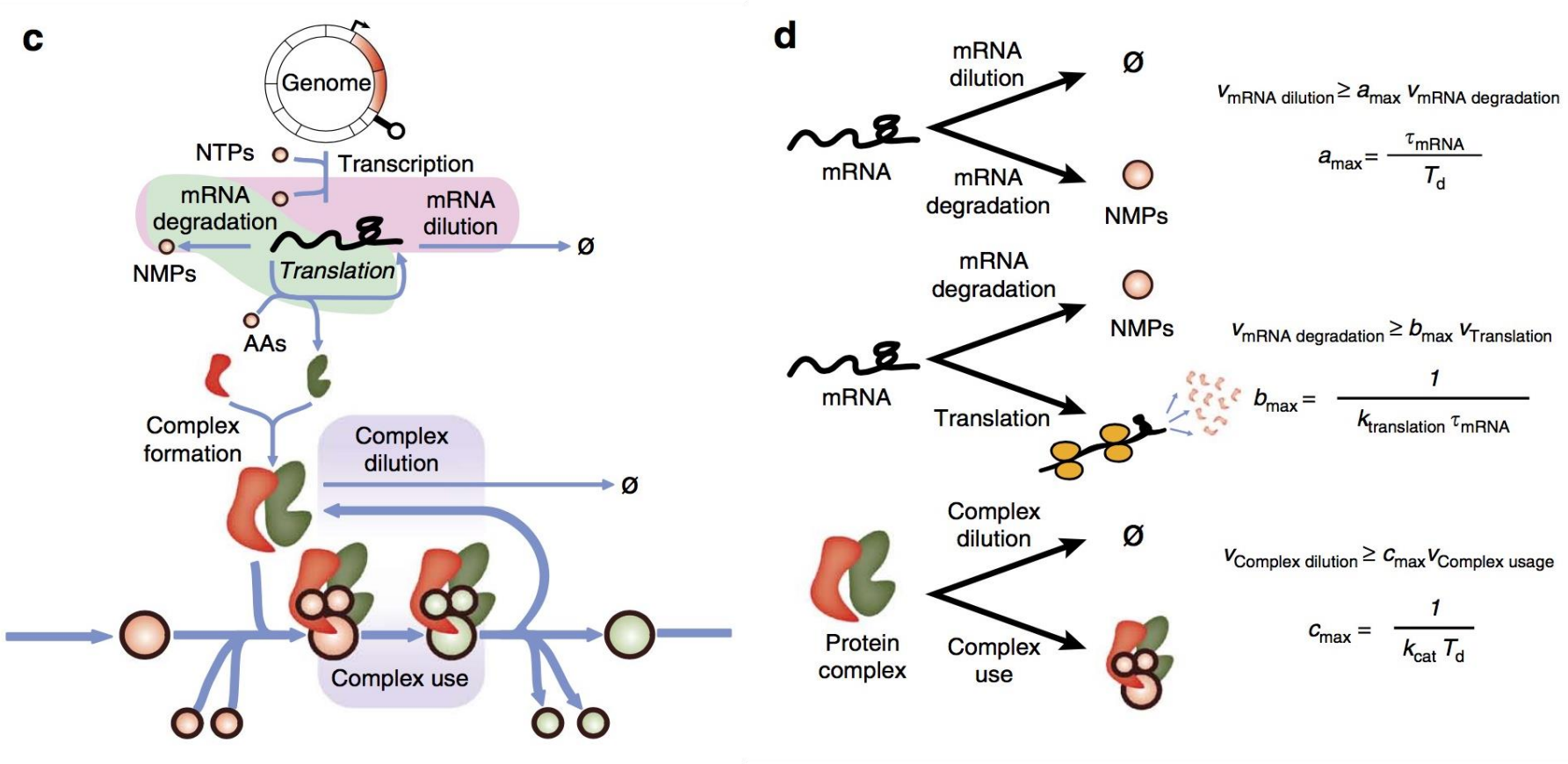
Flux Balance Analysis (FBA)



Often, the optimization objective is to maximize flux through an artificial "**biomass**" reaction, or to maximize **production of ATP**.

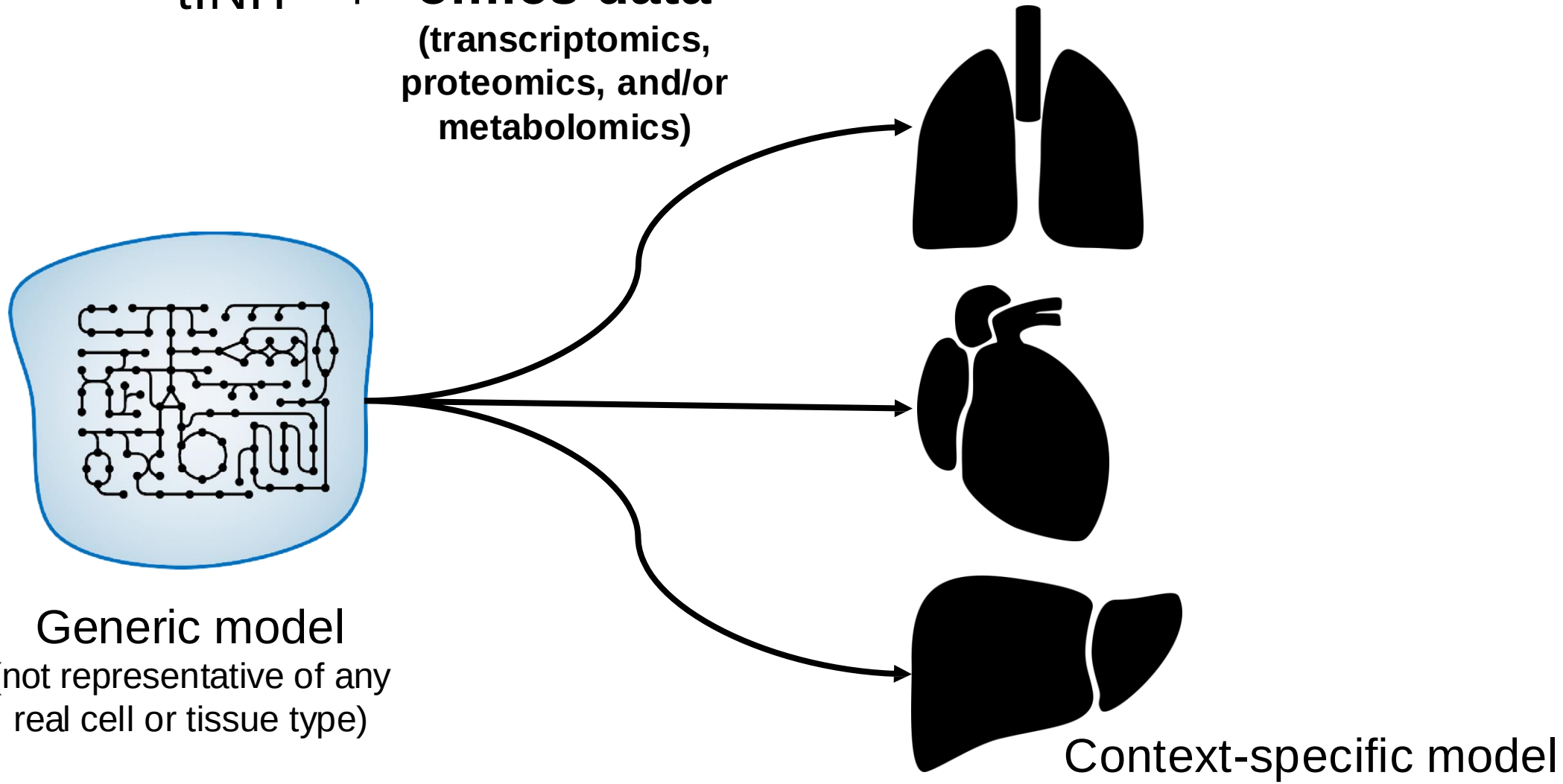
Metabolism and macromolecular expression (ME) model

J Lerman et al, Nat. Commun. 2012



GEM contextualization

tINIT + **omics data**
(transcriptomics, proteomics, and/or metabolomics)



Generic model
(not representative of any real cell or tissue type)

Context-specific model

GEM contextualization

INIT (Integrative Network Inference for Tissues)

R. Ågren, et al. *PLoS Comput Biol* 2012

- Uses proteomic, transcriptomic, and/or metabolomic data
- An optimization is performed to
  - maximize the number of high-confidence (high expression) reactions that are “on”
  - minimize the number of low-confidence (low-expression) reactions that are “on”
- **All reactions in the final model must be able to carry flux**
- **Metabolites are allowed to accumulate** during the optimization
  - An additional term in the algorithm maximizes the number of “present” metabolites that can be produced
  - Distinction of which metabolites should be “present” are based on literature or data (e.g., metabolomics)

$$\max\left(\sum_{i\in R} w_i y_i + \sum_{j\in M} x_j\right)$$

$$S\vec{v} = \vec{b}$$

$$|v_i| \le 1000 y_i$$

$$|v_i| + 1000(1 - y_i) \ge \varepsilon$$

$$v_i \ge 0, i \in irreversible\ rxns$$

$$b_j \le 1000 x_i$$

$$b_j + 1000(1 - x_i) \ge \varepsilon$$

$$b_j \ge 0$$

$$x_j = 1, j \in present$$

$$y_i, x_j \in \{0,1\}$$

$$w_{i,j} = 5\log\left(\frac{Signal_{i,j}}{Average_i}\right)$$

GEM contextualization

tINIT (Task-driven Integrative Network Inference for Tissues)

R. Ågren, et al. *Mol Syst Biol* 2014

- Identical formulation as INIT, with added steps
  - INIT does not necessarily yield simulation-ready models
- User defines a series of metabolic tasks that the model must perform
- Reactions that are required for these tasks are identified
  - A requirement that these reactions are active is included as an additional constraint in the optimization
- A follow-up evaluation of each task is performed
  - If a task fails, a gap-filling algorithm is used to enable task completion

Metabolic Tasks

Rephosphorylation of nucleoside triphosphates
- Aerobic rephosphorylation of ATP from glucose
- Aerobic rephosphorylation of GTP
- Aerobic rephosphorylation of CTP
- Aerobic rephosphorylation of UTP

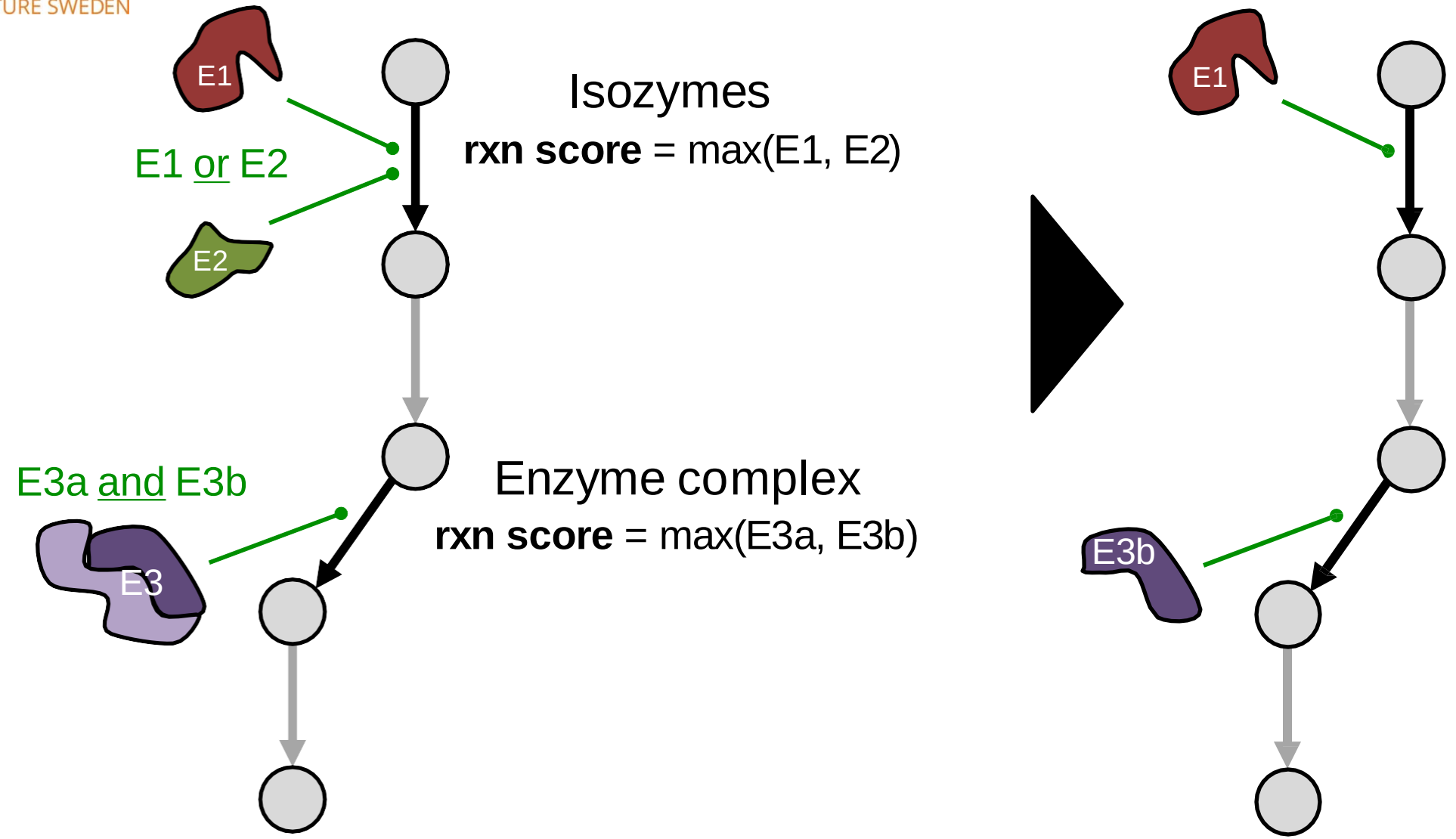
De novo synthesis of nucleotides
- ATP de novo synthesis
- CTP de novo synthesis
- GTP de novo synthesis
- UTP de novo synthesis
- dATP de novo synthesis
- dCTP de novo synthesis
- dGTP de novo synthesis
- dTTP de novo synthesis

Uptake of essential amino acids
- Histidine uptake
- Isoleucine uptake
- Leucine uptake
- Lysine uptake
- Methionine uptake
- Phenylalanine uptake
- Threonine uptake
- Tryptophan uptake
- Valine uptake

De novo synthesis of key intermediates
- Glycerate 3-phosphate de novo synthesis
- Mitochondrial acetyl-CoA de novo synthesis
- Mitochondrial AKG de novo synthesis
- Erythrose 4-phosphate de novo synthesis
- Fructose 6-phosphate de novo synthesis

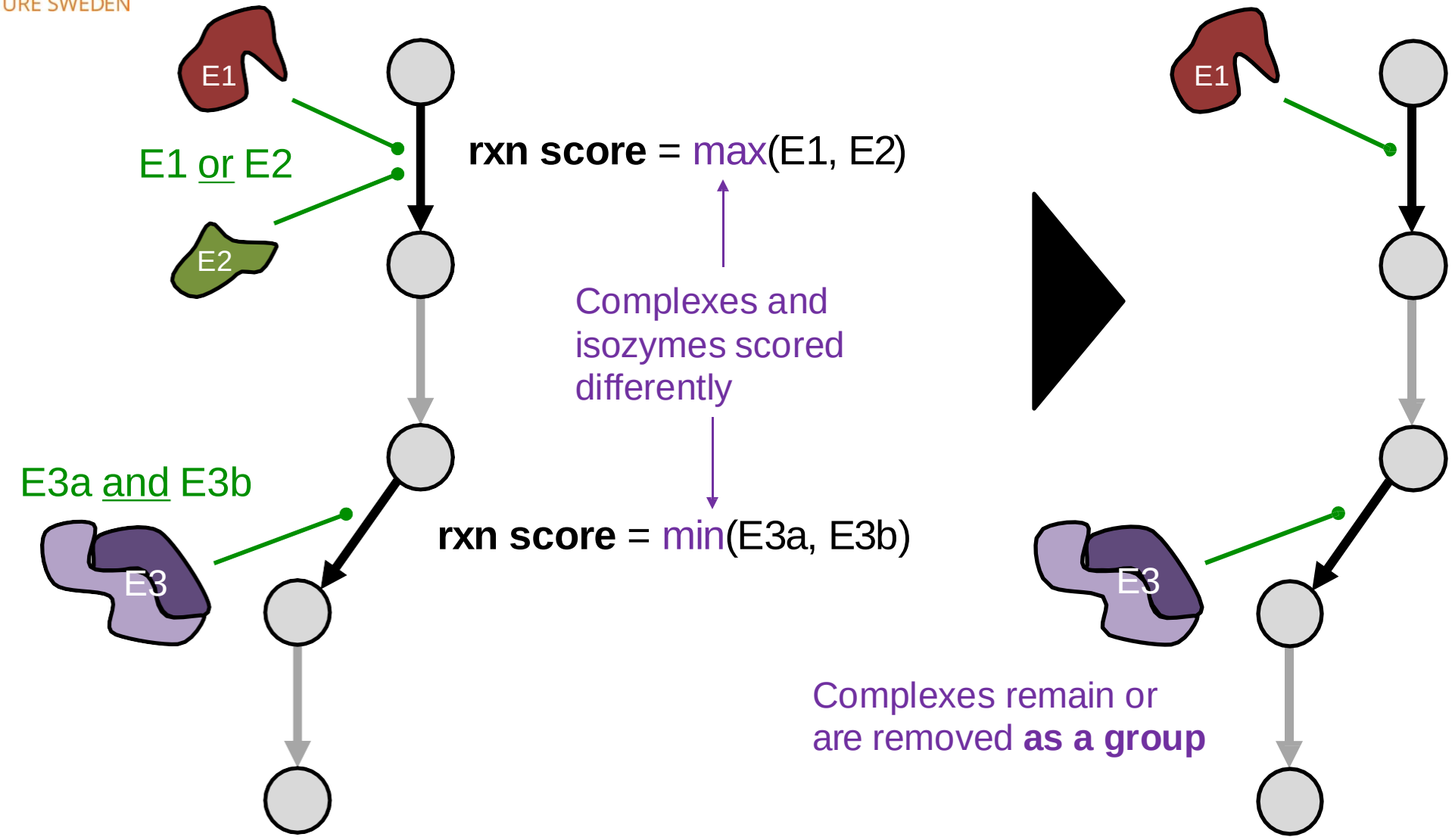
GEM contextualization

tINIT1 (Task-driven Integrative Network Inference for Tissues)



GEM contextualization

tINIT2 (Task-driven Integrative Network Inference for Tissues)



Enzyme-constrained GEMs

- Should any reaction have bounds up to $+\infty$?
- Should these 2 pathways have reactions with the same bounds?

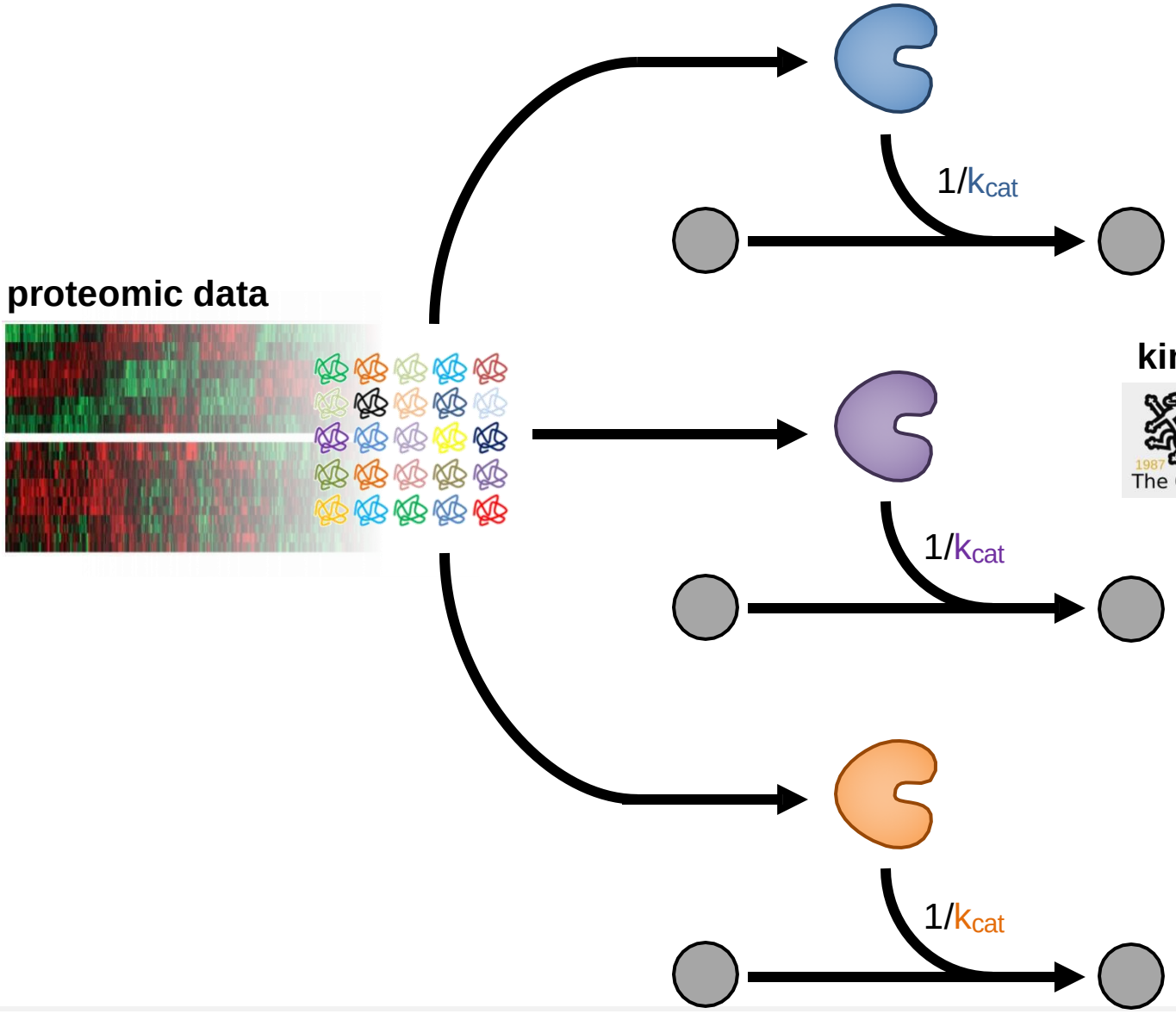
Relationship between enzyme and reaction:

Flux of reaction (from FBA) → $v \leq k_{cat}[E]$ ← Concentration of enzyme (from absolute proteomics)

Turnover number (from databases) ↑

However: No simple implementation for connecting proteomics to GEMs…

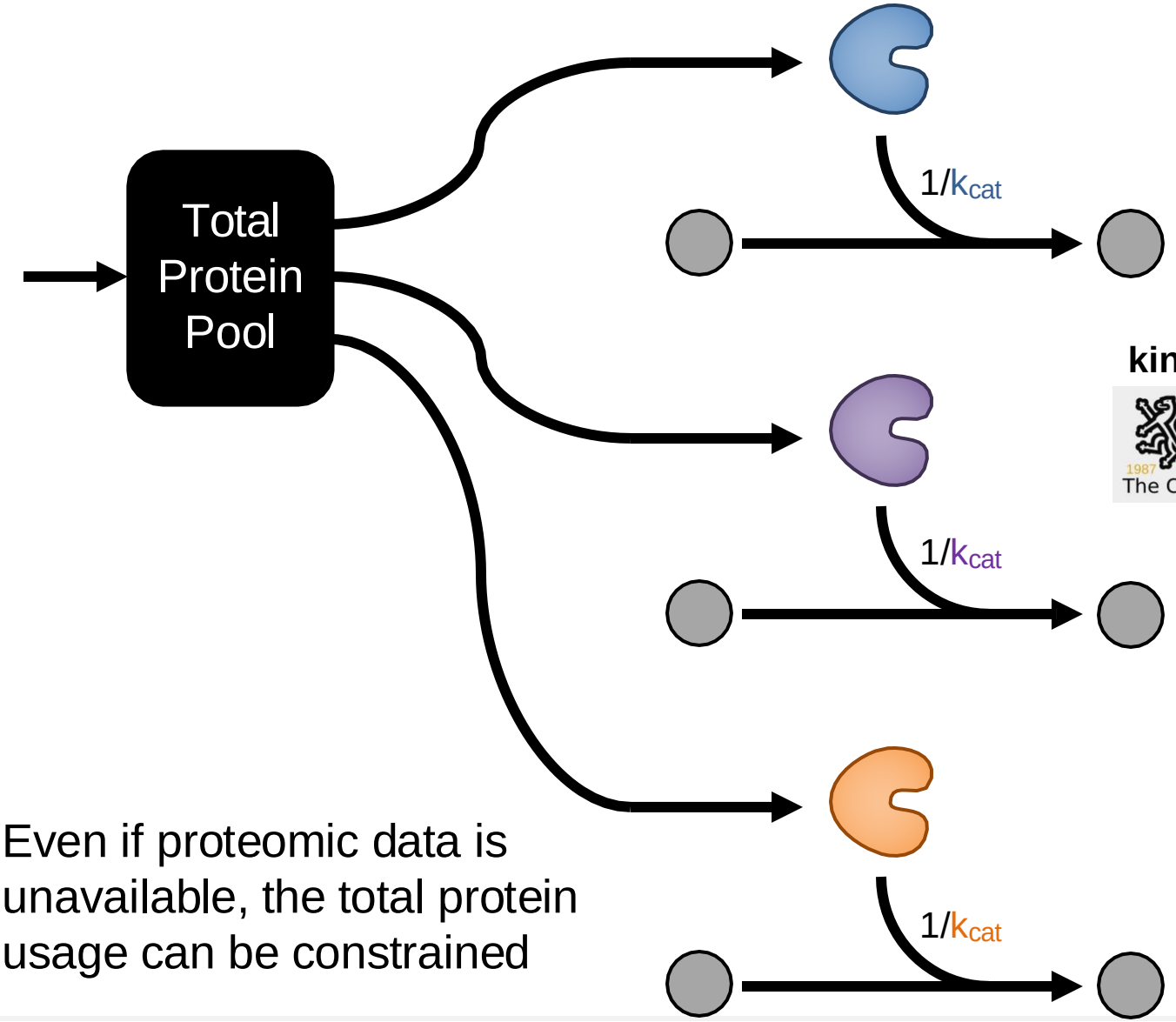
Enzyme-constrained GEMs



kinetic data (k_{cat} values)

BRENDA
1987 2017
The Comprehensive Enzyme Information System

Enzyme-constrained GEMs



kinetic data (k_{cat} values)

BRENDA 1987 2017
The Comprehensive Enzyme Information System

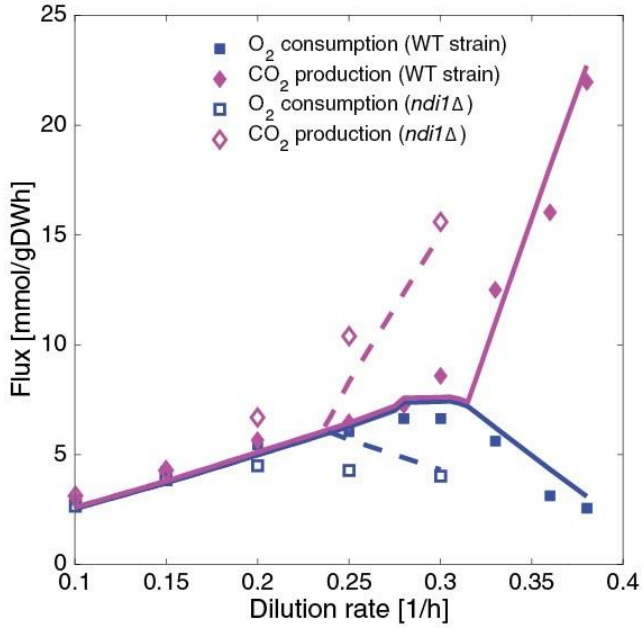
Even if proteomic data is unavailable, the total protein usage can be constrained

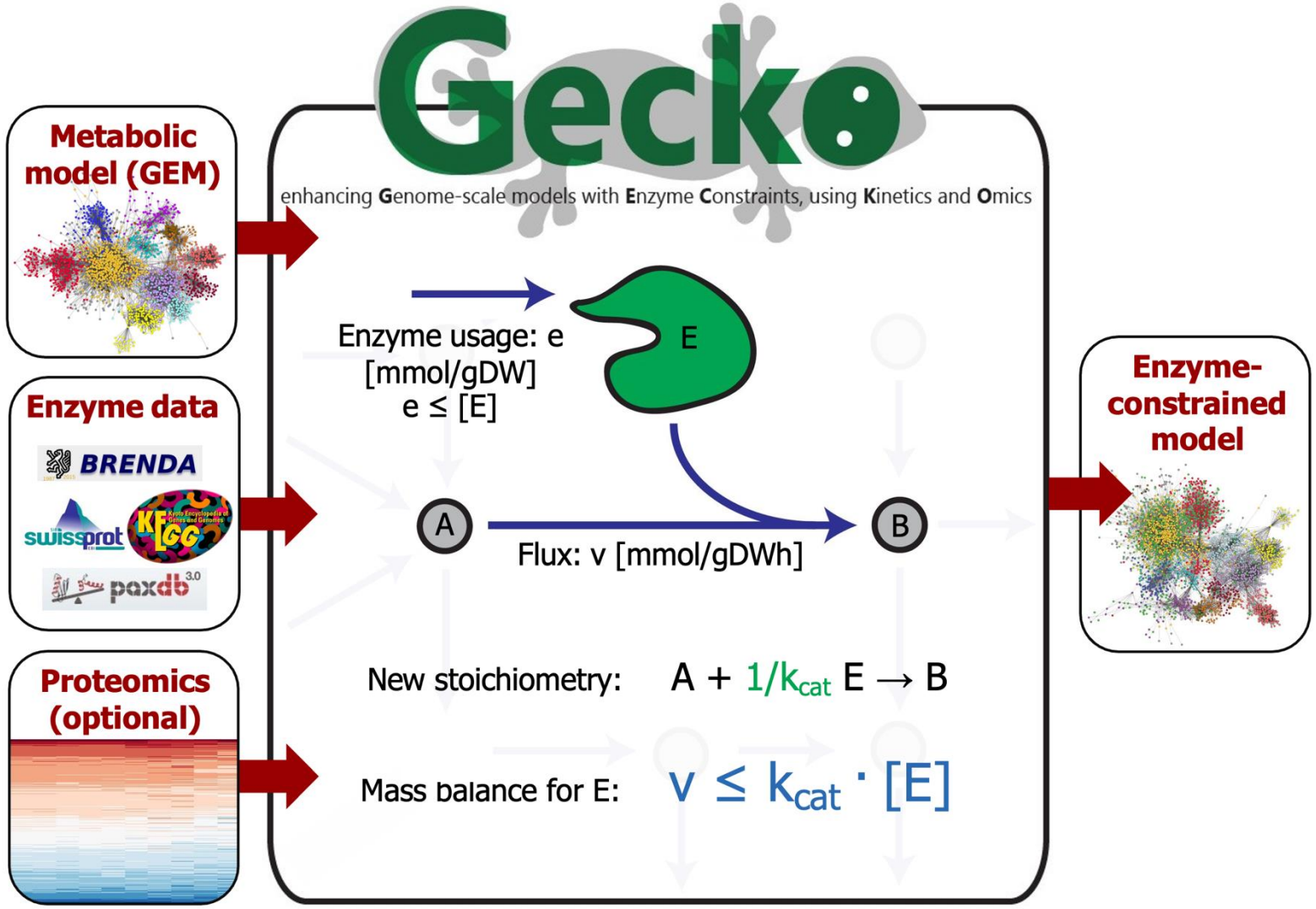
Enzyme-constrained GEMs

B. Sánchez, et al *Mol Syst Biol* 2017

Applications:

- Improving predictions
- Integrating proteomics data into GEMs

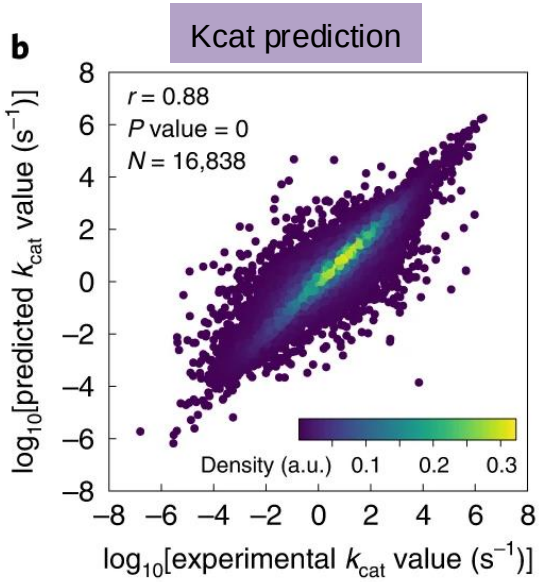




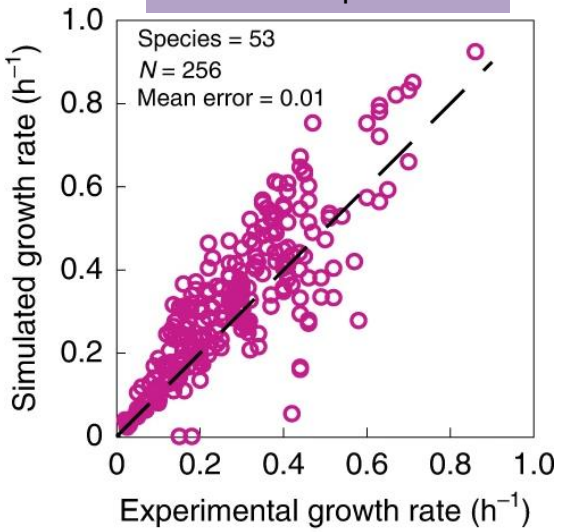
Predicting K_{cat} for ecGEM parameterization

Li F, et al *Nat Cat 2022*

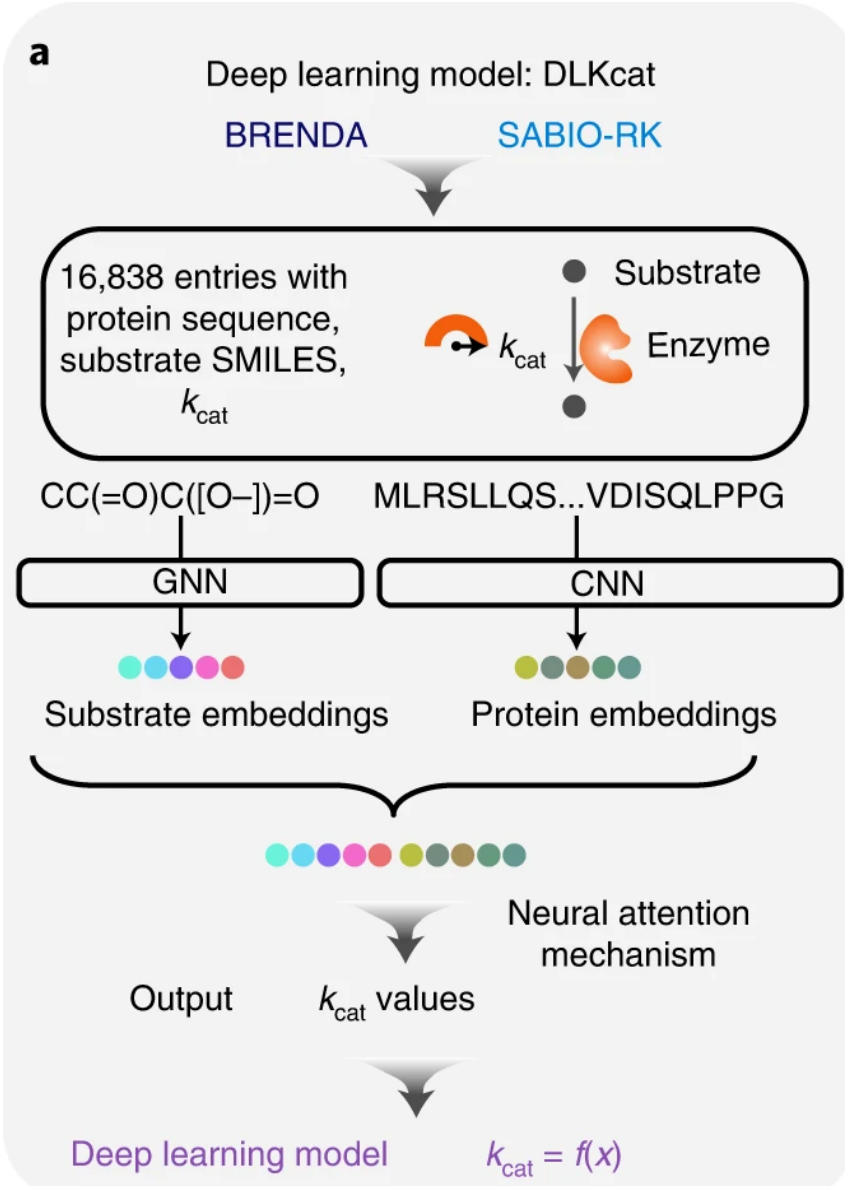
- Experimentally measured k_{cat} data are sparse and noisy
- Deep learning approach (DLKcat) for high-throughput k_{cat} prediction for metabolic enzymes
- They designed a Bayesian pipeline to parameterize enzyme-constrained genome-scale metabolic models from predicted k_{cat} values



Kcat prediction



Growth rate prediction

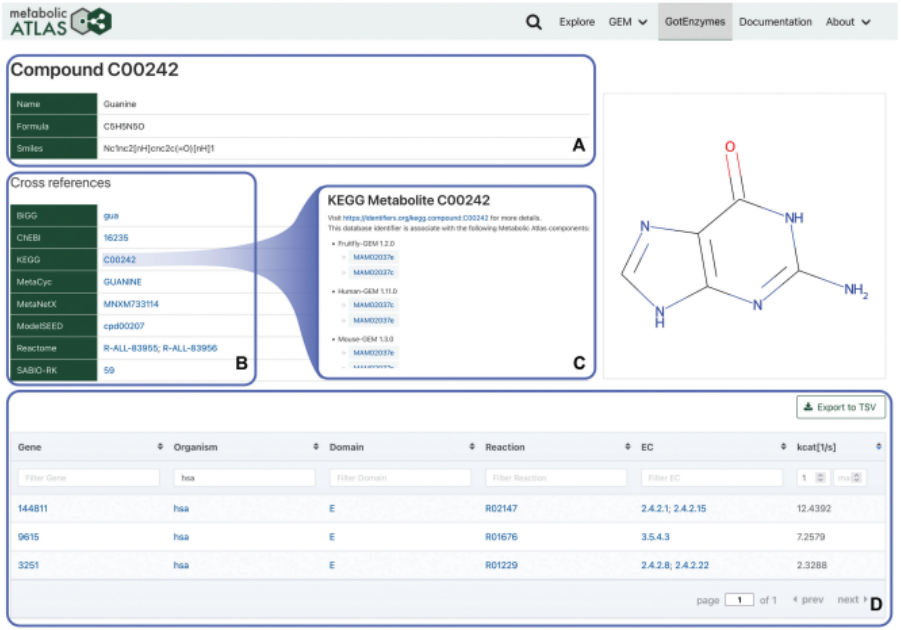


Predicting kcat for ecGEM parameterization

Li F, et al *Nucleic Acids Res, 2023*

- Enzyme performance can be quantitatively described by parameters such as enzyme turnover number kcat and Michaelis constant K_M.
- The ratio kcat/Km is a measure of enzyme efficiency, combining both the affinity for the substrate and the rate of catalysis. It is often used as a benchmark for comparing the performance of different enzymes.

GotEnzymes provides a comprehensive database with enzyme parameter predictions available at https://metabolicatlas.org/gotenzymes.



$$v = \frac{V_{max}[S]}{K_m+[S]} \xrightarrow{\text{Vmax=Kcat[E]total}} v = \frac{Kcat[E]total[S]}{K_m+[S]}$$

Michaelis-Menten equation

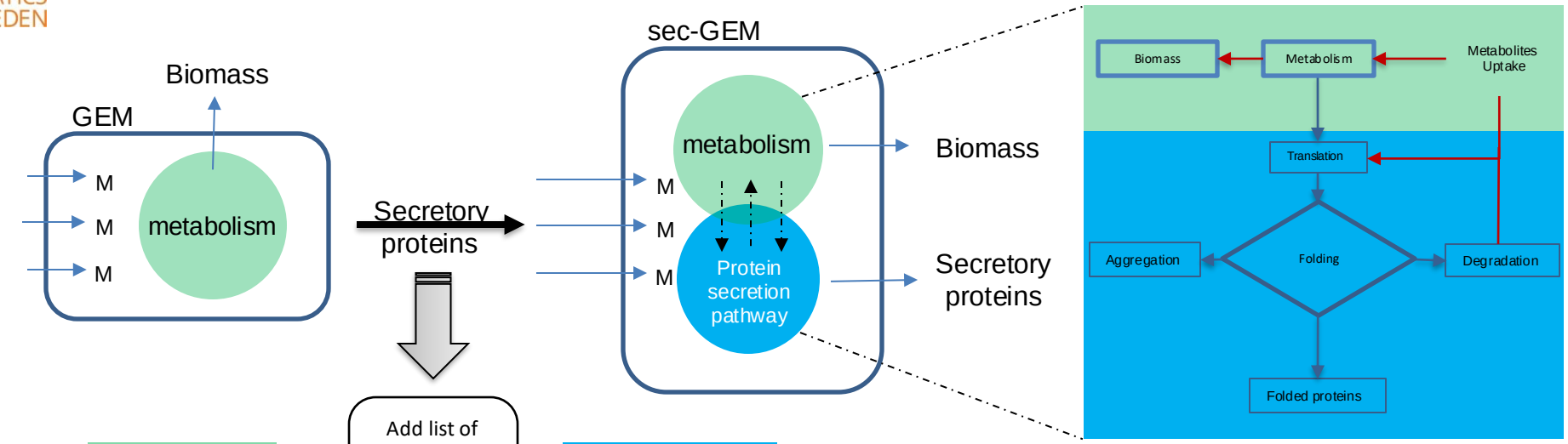
Experimental measurement / DLKcat prediction

Proteomics data / whole protein constrain

Experimental measurment / GotEnzymes prediction

More accurate predictions

Extending the coverage of GEMs: secGEM



Human1

Sec-GEM

Add list of reactions for the selected protein to model

Gap Filling the model for production of added reactions

input a secretory protein

Template reactions

HumanSec

PSIM

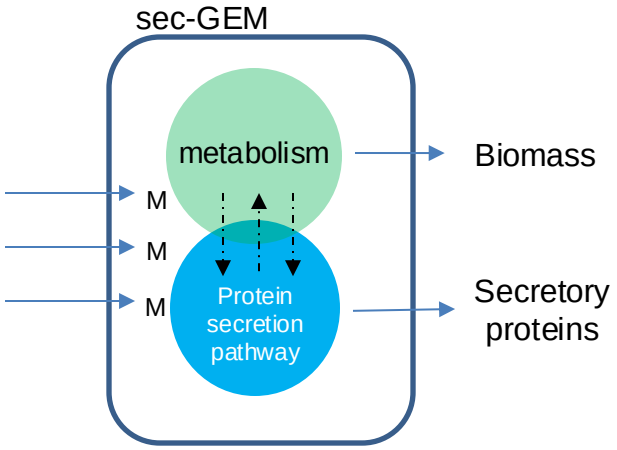
Generating specific reactions for the protein

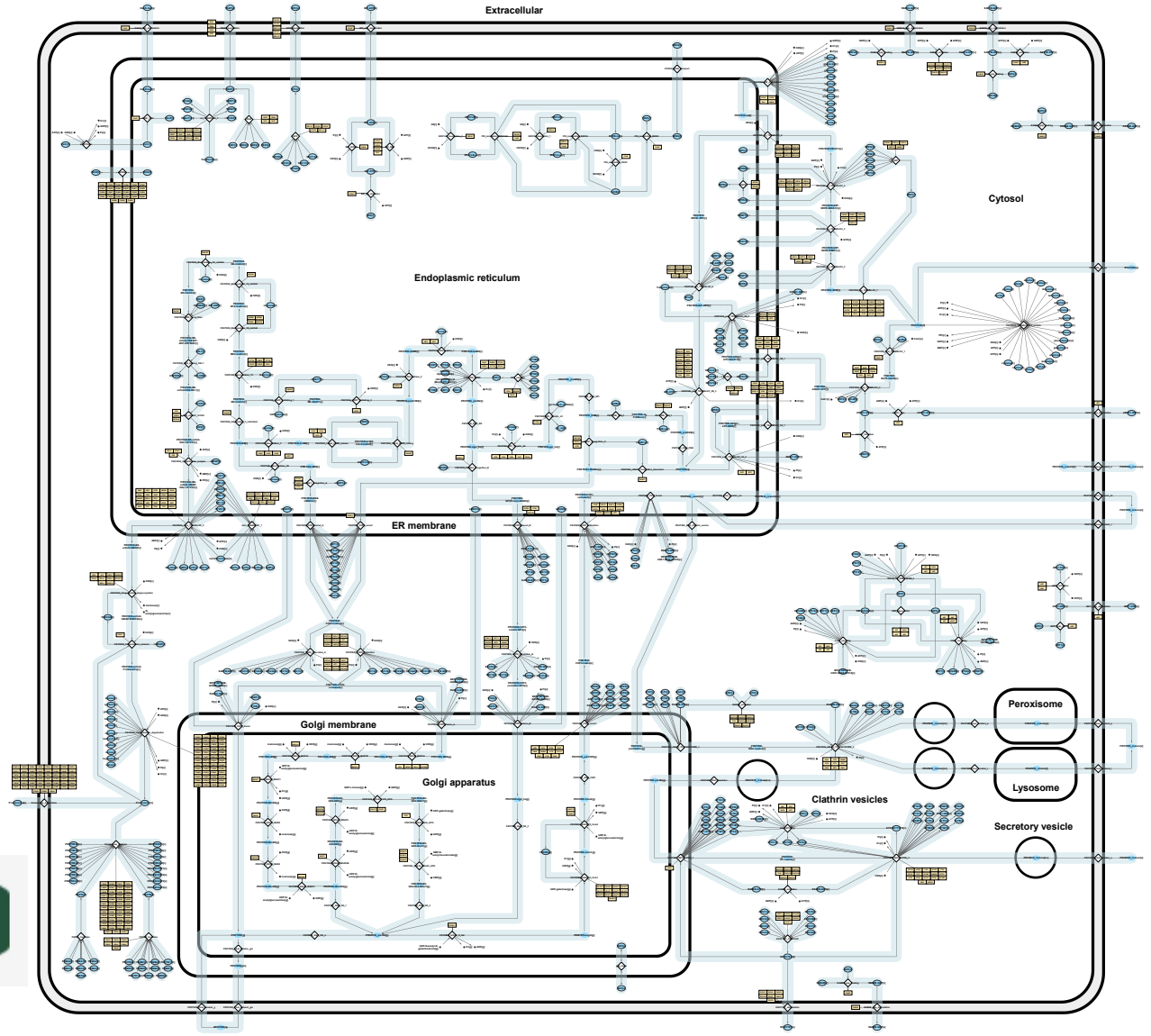
Extract information from Protein Specific Information Matrix

	Protein Specific Information Matrix					
	L	NG	OG	DSB	TMD	...
P_i	340	2	9	1	0	...
...	...	...	...	...	...	...

Saghaleyni, et al *Cell Rep, 2022*

Extending the coverage of GEMs: secGEM





Integrating Single cell transcriptomics into GEMs

Single-cell omics analysis with genome-scale metabolic modeling,

J Gustafsson. Et al
Current Opinion in Biotechnology, 2024

Generation and analysis of context-specific genome-scale metabolic models derived from single-cell RNA-Seq data

J Gustafsson. Et al
PNAS, 2023

Johan Gustafsson

Postdoctoral Fellow, Broad institute, USA

Talk Title: *Generation of context-specific genome-scale metabolic models using single-cell RNA-Seq data*

Time: October 17, 13:00 – 14:15 CET online on zoom

Link to Talk: BIG talk event, Link, pass:spd996

Description of the Talk:
The metabolic networks in cells vary across tissues and cell types, and to accurately model the metabolism of cells, the full generic metabolic network defined in the genome needs to be reduced to a context-specific network representing the network expressed specifically in the cells of interest. Single-cell RNA-Seq promises to provide the information needed for such a reduction, but noise in the form of data sparsity is a challenge. Here, we present methods to handle data sparsity and estimate the uncertainty of modeling results.

About the Speaker:
Johan is an expert in modeling cancer metabolism and analyzing single-cell RNA/DNA sequencing data, aiming to uncover vulnerabilities in cancer. With a background in both computer science and biochemistry, Johan has completed a PhD in metabolic modeling at Chalmers University of Technology and now works as a postdoc in the Getz lab at the Broad Institute, focusing on CLL/Richter's syndrome and hypoxia in solid tumors.

Take home Messages

- Developing GEMs is an iterative process.
- GEMs can serve as a scaffold for integrating & studying diverse types of (omics) data.
- GEMs are simulation based and (FBA) and depending on the objective functions can provide deeper insights into metabolism.
- GEMs enables the analysis of omics data in the context of metabolism.