

GEMs

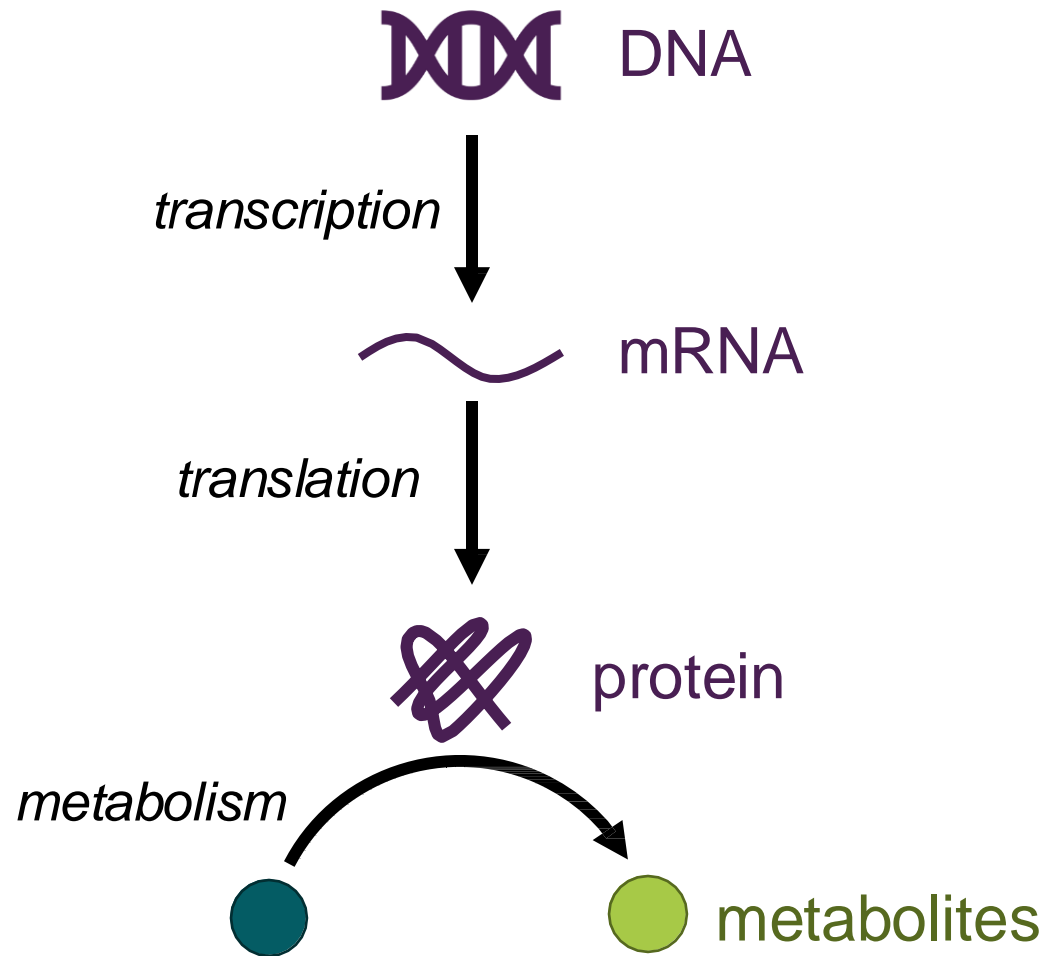
Structure & Developments

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

Rasool Saghaleyni

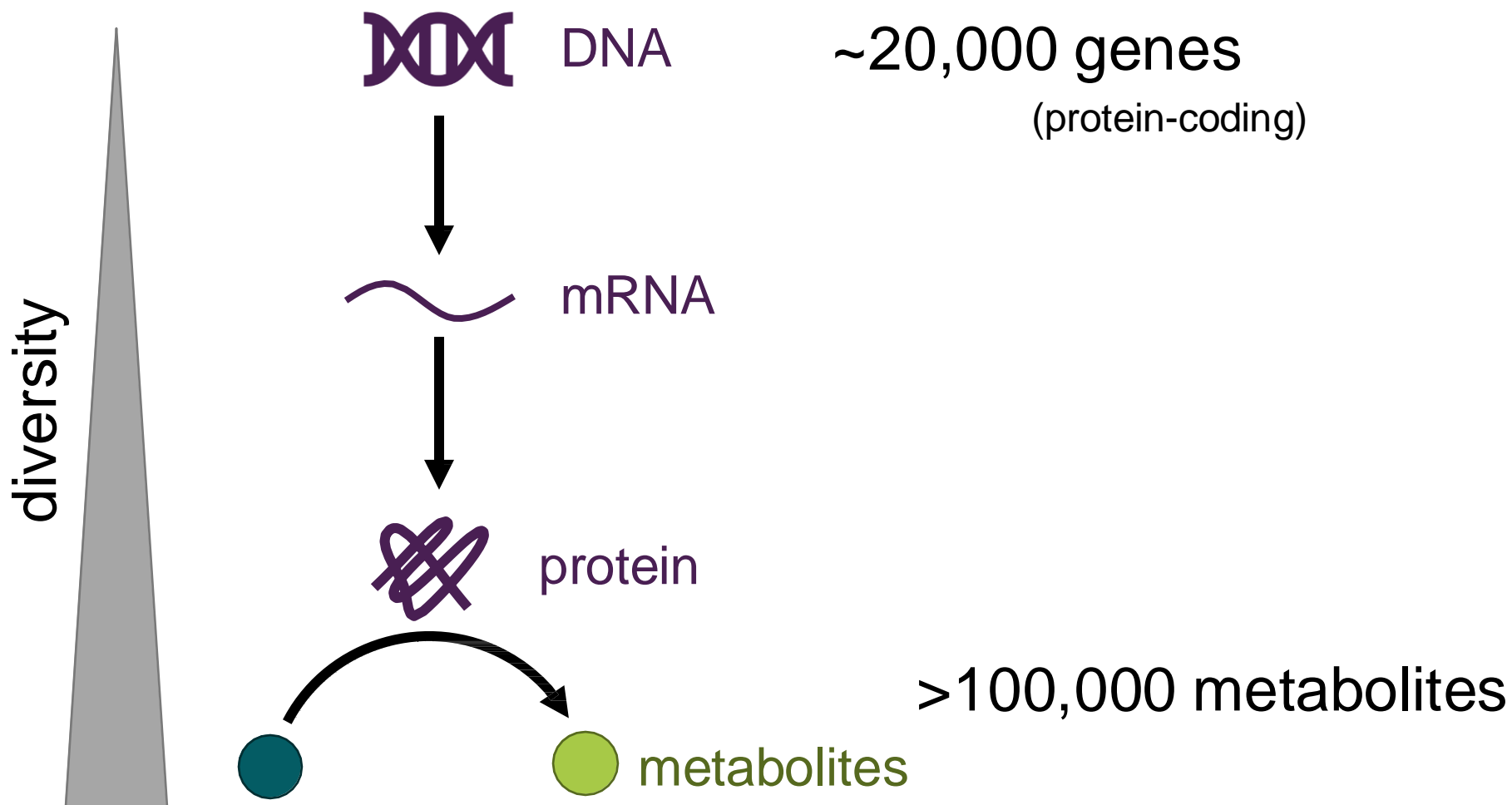
National Bioinformatics Infrastructure Sweden (NBIS)
Science for Life Laboratory (SciLifeLab)
Chalmers University of Technology
rasool.saghaleyni@scilifelab.se

Background

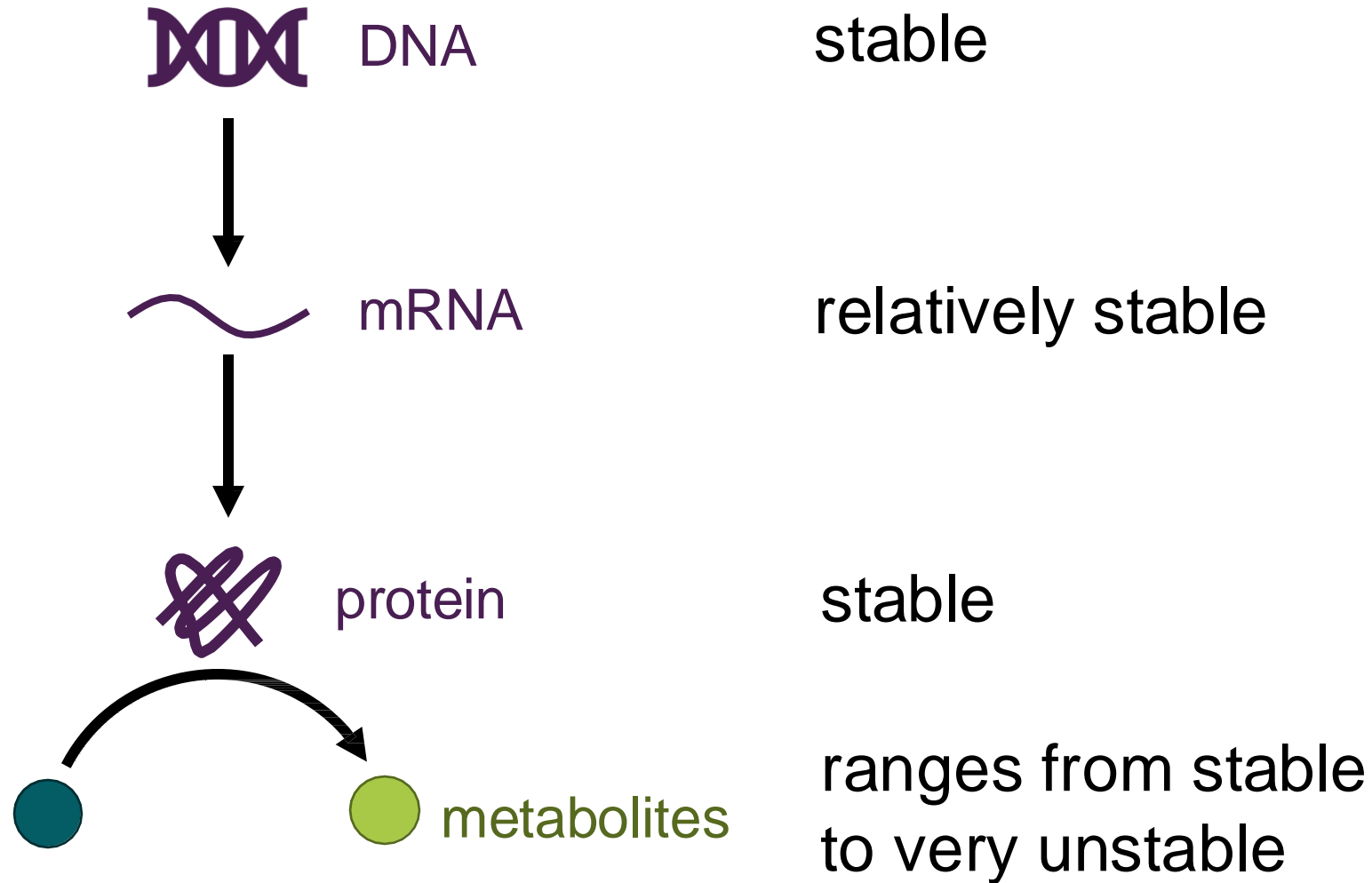


Metabolism provides the **energy** and **building blocks** necessary to sustain life.

Background



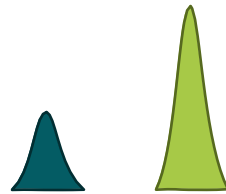
Background



Background



We can generally measure
metabolite concentrations



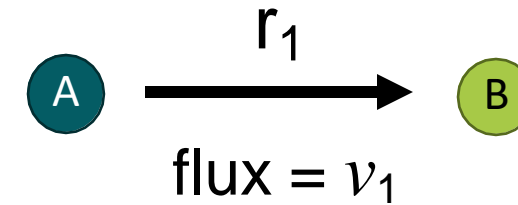
...but what is often important is
the flow or **flux** of metabolites
through the reactions.



Background

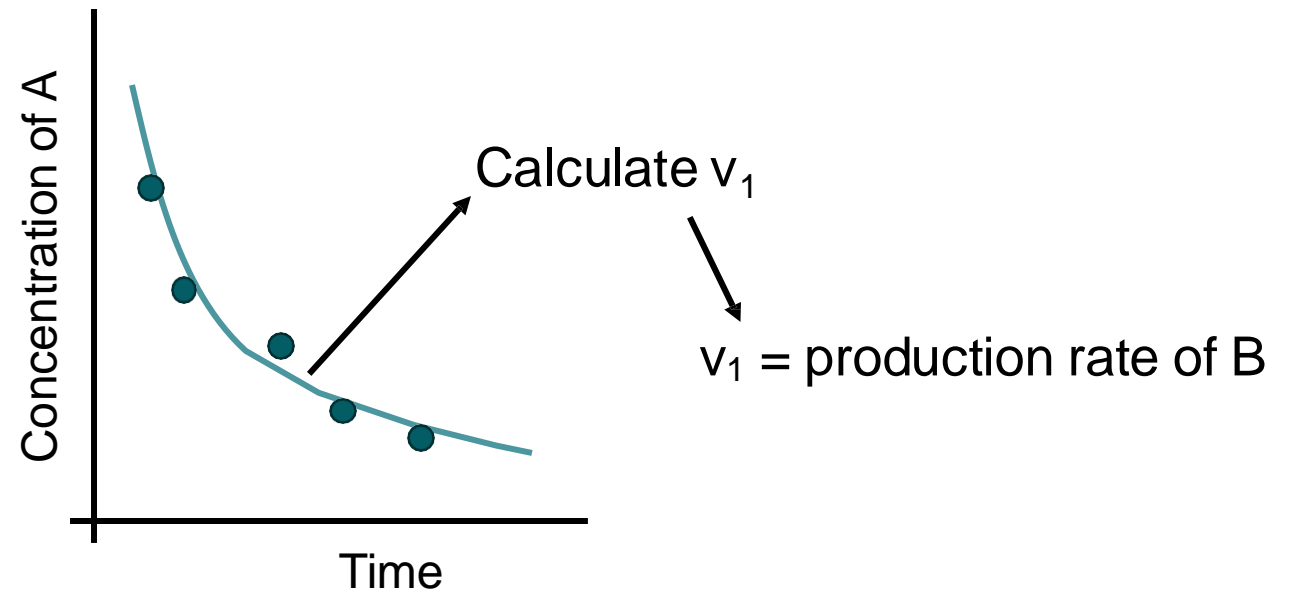


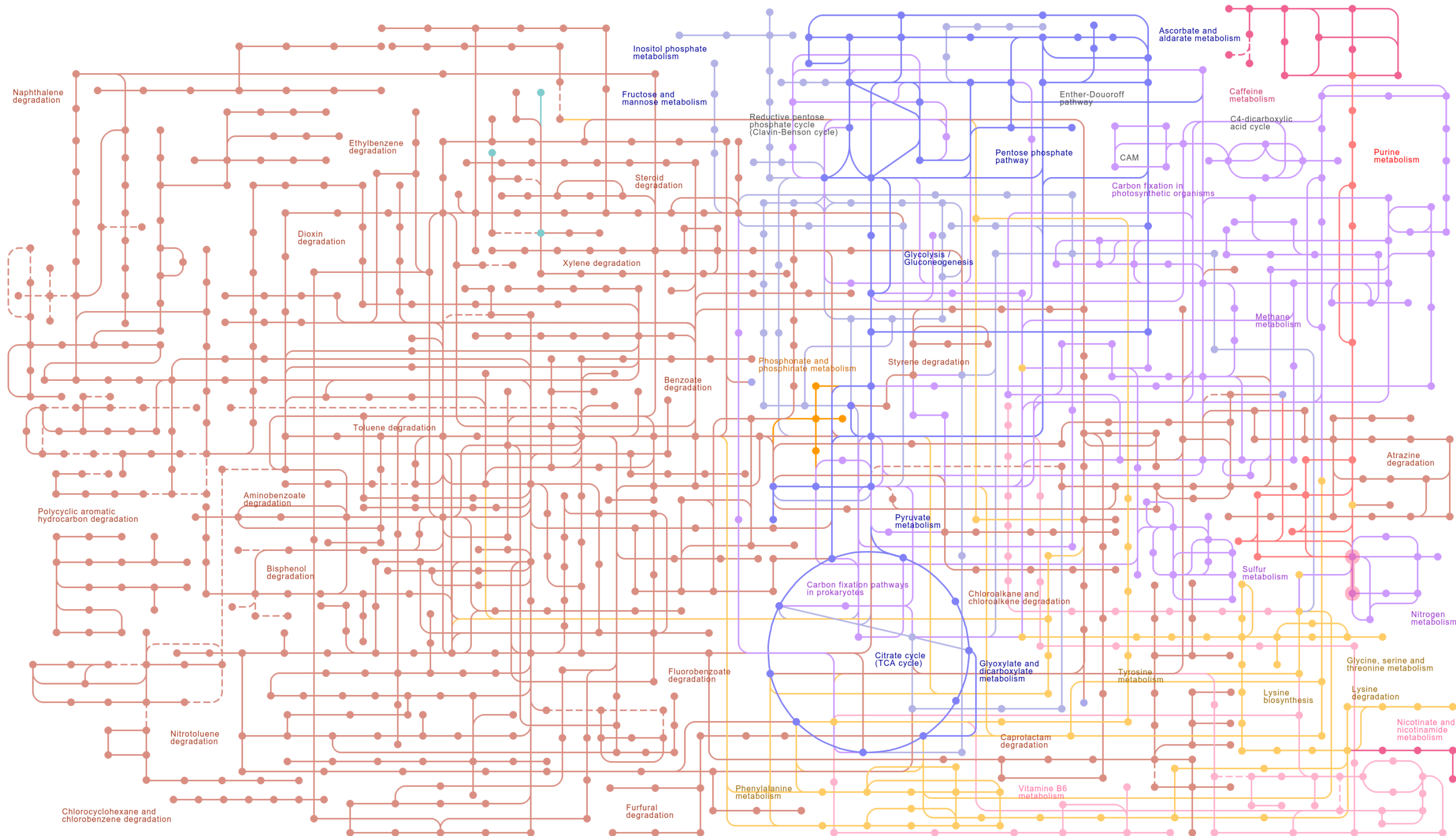
Assume that we want to know the production rate of **B**, but can only measure the concentration of **A**



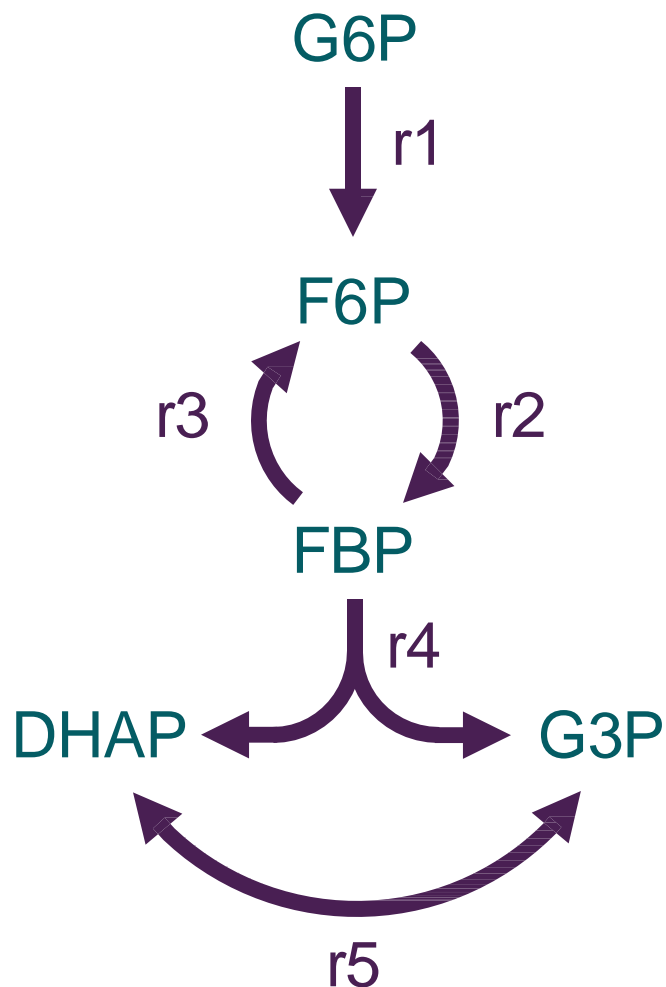
$$\frac{dA}{dt} = -v_1$$

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



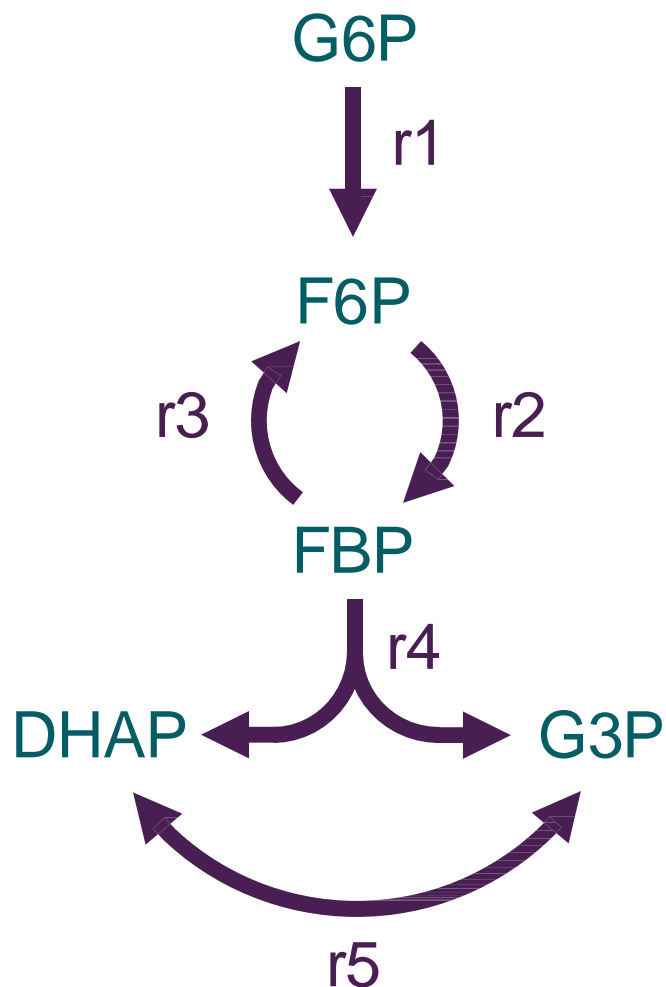


The Stoichiometric Matrix



Metabolites	Reactions	
		r1
	G6P	-1
	F6P	1
	FBP	0
	DHAP	0
	G3P	0

The Stoichiometric Matrix

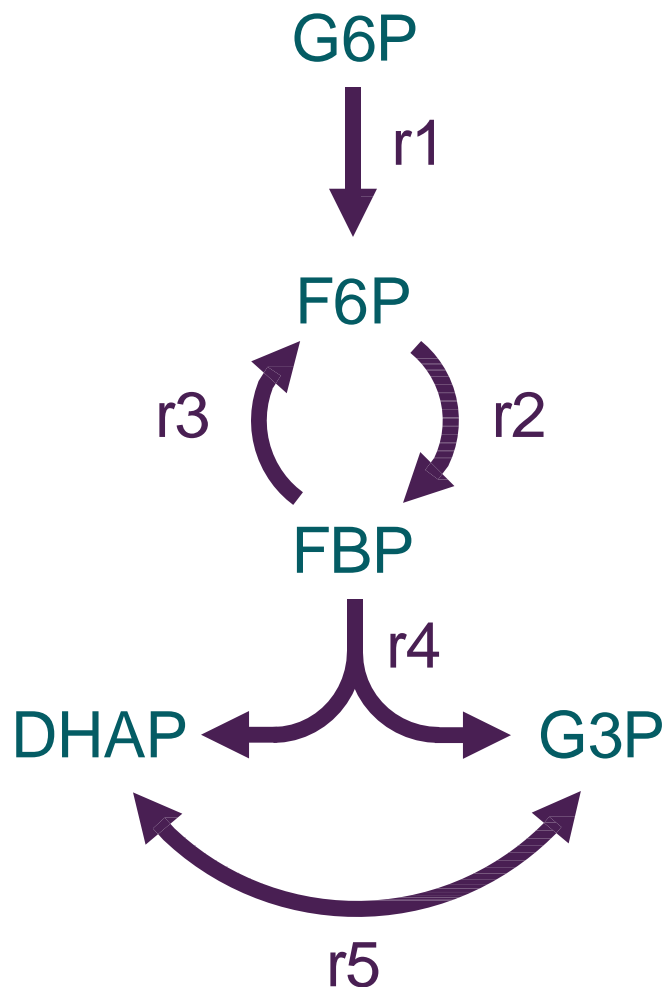


Metabolites

Reactions

	r1	r2
G6P	-1	0
F6P	1	-1
FBP	0	1
DHAP	0	0
G3P	0	0

The Stoichiometric Matrix



Metabolites

Reactions

	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

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)	Transcript IDs	GO Terms	Orthologs
					GPI	P06744			
					<i>n/a</i>				
					FBP1, FBP2	P09467, O00757			
					ALDOA, ALDOB, ALDOC	P04075, P05062, P09972			
					TPI1	P60174			

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

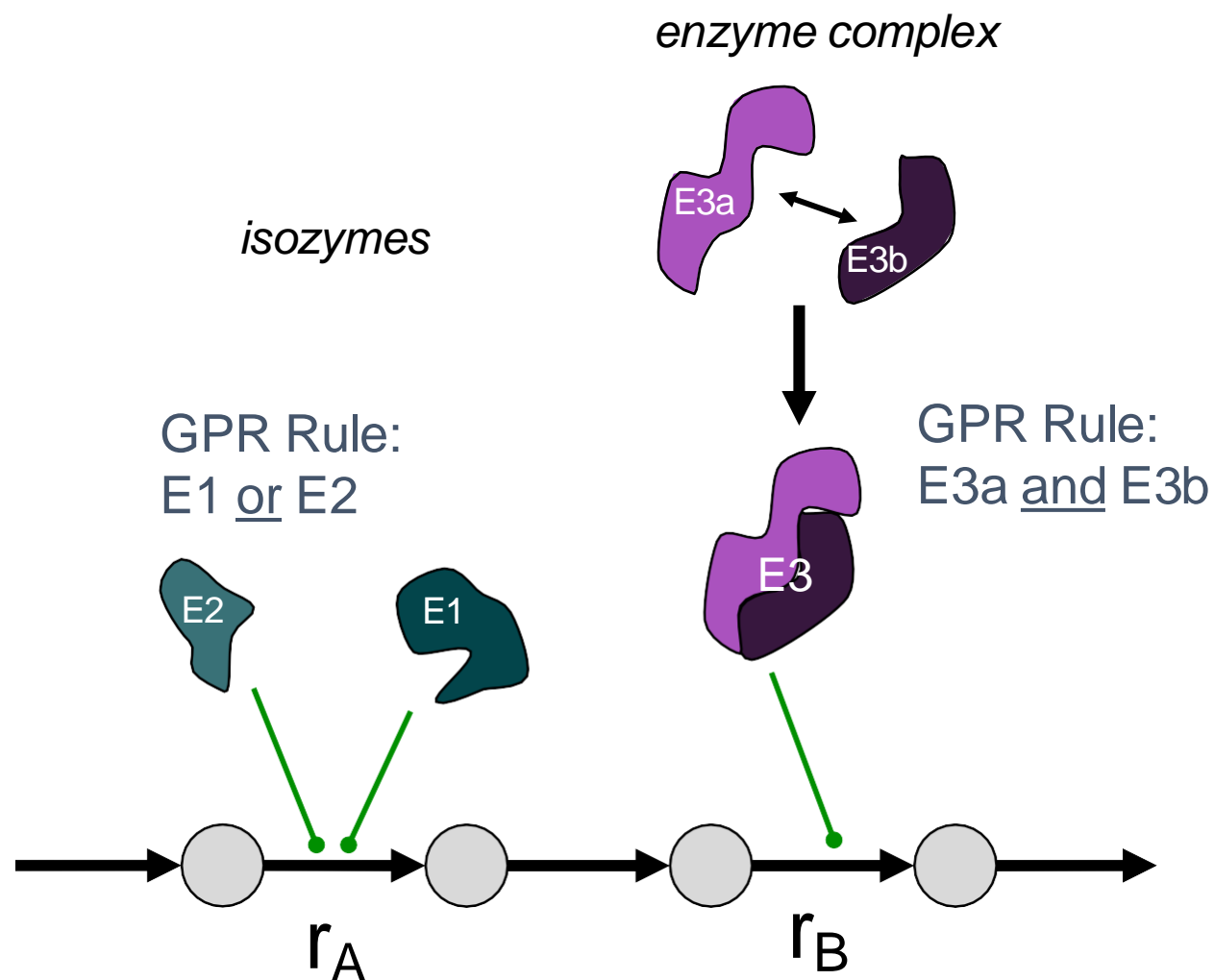
Reactions are linked to genes that encode the enzymes that catalyze the reaction.

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

GPR Rules



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



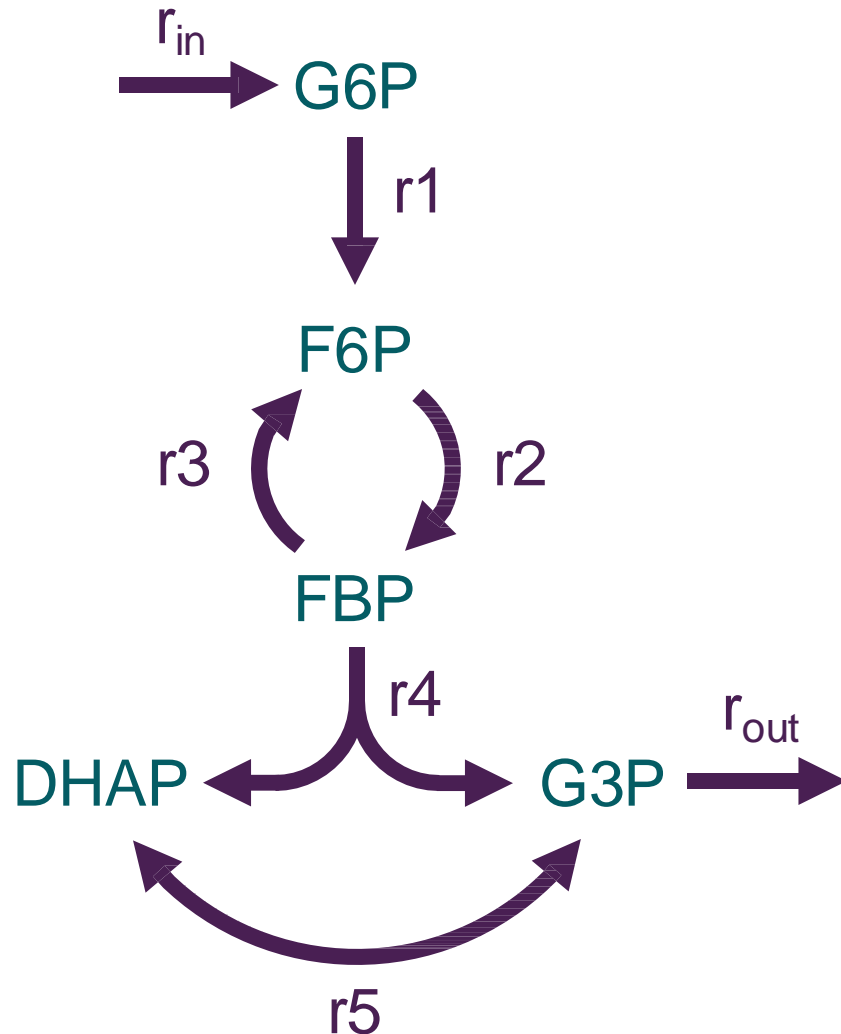
Knockout	Effect
E1	none
E2	none
E1 + E2	rA inactive
E3a	rB inactive
E3b	rB inactive
E3a + E3b	rB inactive

Flux Balance Analysis (FBA)



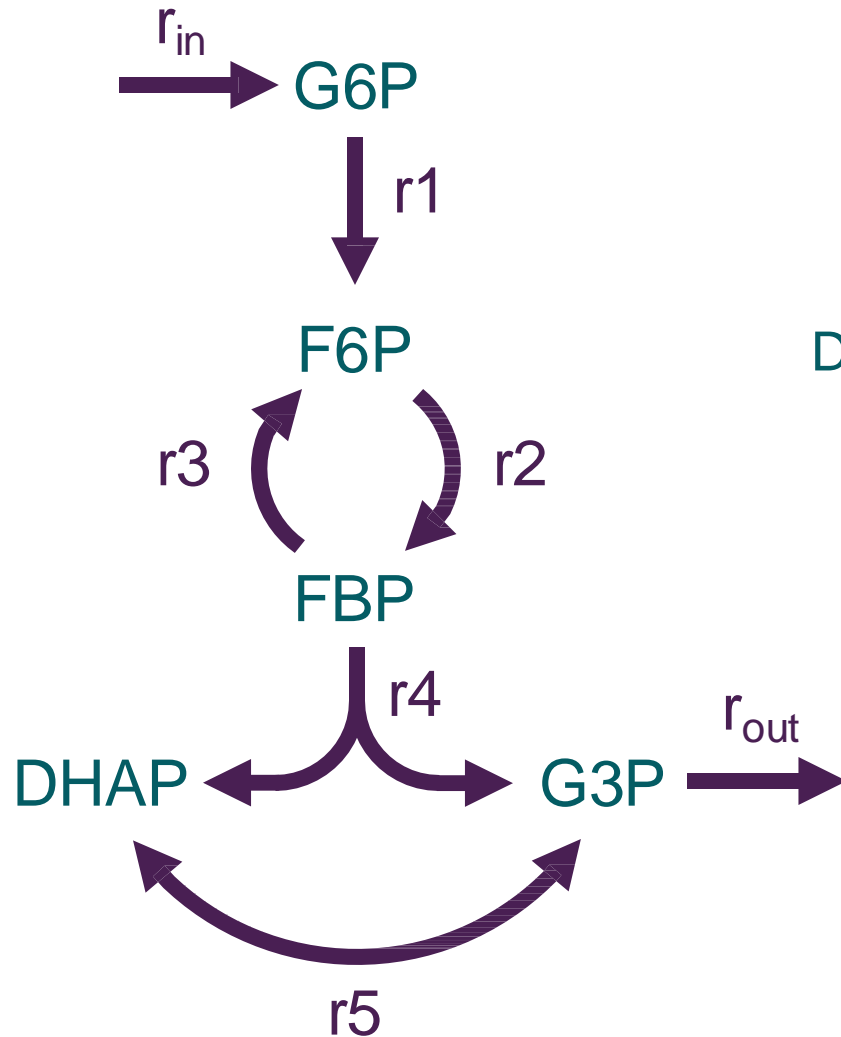
FBA seeks to calculate the reaction **fluxes** (v) of a network

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)



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

$$\frac{d[\text{G6P}]}{dt} = -v_1 + v_{in}$$

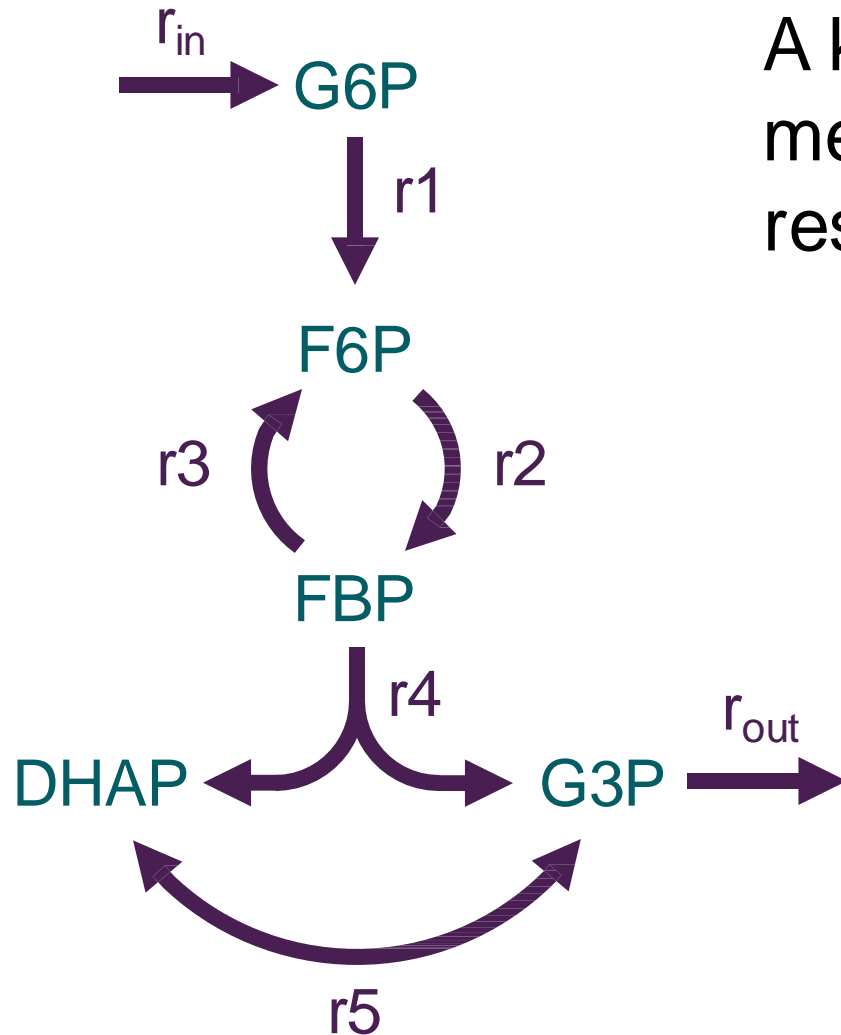
$$\frac{d[\text{G3P}]}{dt} = v_4 + v_5 - v_{out}$$

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)

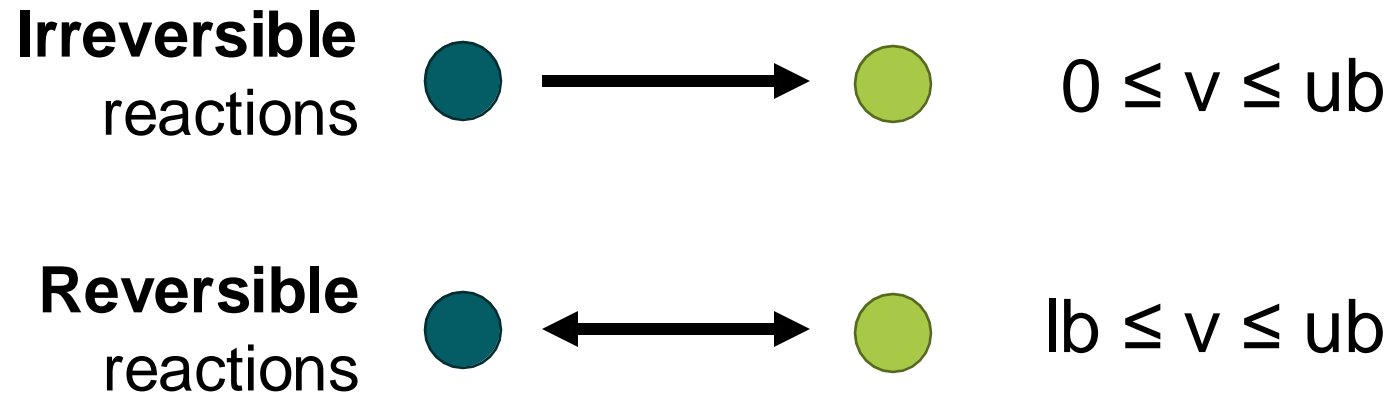
$$\begin{array}{c}
 \text{G6P} \\
 \text{F6P} \\
 \text{FBP} \\
 \text{DHAP} \\
 \text{G3P}
 \end{array}
 \begin{array}{c}
 r_1 \quad r_2 \quad r_3 \quad r_4 \quad r_5 \quad r_{in} \quad r_{out} \\
 \begin{pmatrix}
 -1 & 0 & 0 & 0 & 0 & 1 & 0 \\
 1 & -1 & 1 & 0 & 0 & 0 & 0 \\
 0 & 1 & -1 & -1 & 0 & 0 & 0 \\
 0 & 0 & 0 & 1 & -1 & 0 & 0 \\
 0 & 0 & 0 & 1 & 1 & 0 & -1
 \end{pmatrix}
 \end{array}
 \times
 \begin{array}{c}
 v_1 \\
 v_2 \\
 v_3 \\
 v_4 \\
 v_5 \\
 v_{in} \\
 v_{out}
 \end{array}
 =
 \begin{array}{c}
 d\text{G6P}/dt \\
 d\text{F6P}/dt \\
 d\text{FBP}/dt \\
 d\text{DHAP}/dt \\
 d\text{G3P}/dt
 \end{array}
 =
 \begin{array}{c}
 0 \\
 0 \\
 0 \\
 0 \\
 0
 \end{array}$$

$$S \cdot v = 0$$

Flux Balance Analysis (FBA)



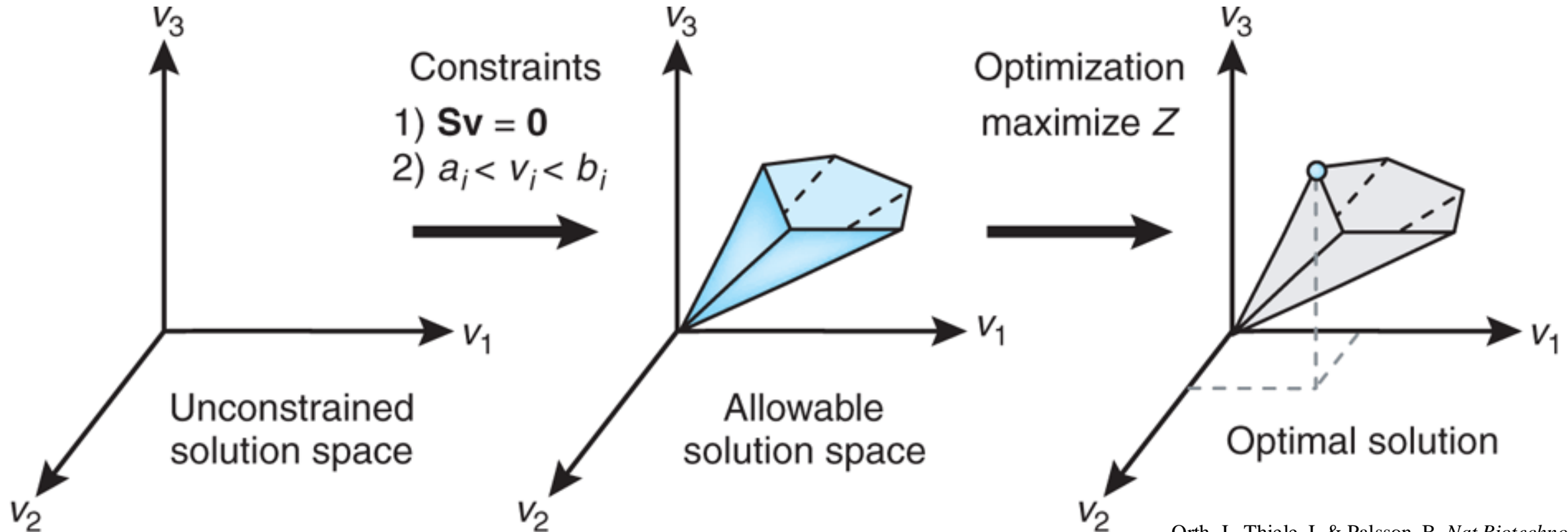
We can further constrain the solution space by limiting reaction fluxes based on their reversibility:



Flux Balance Analysis (FBA)

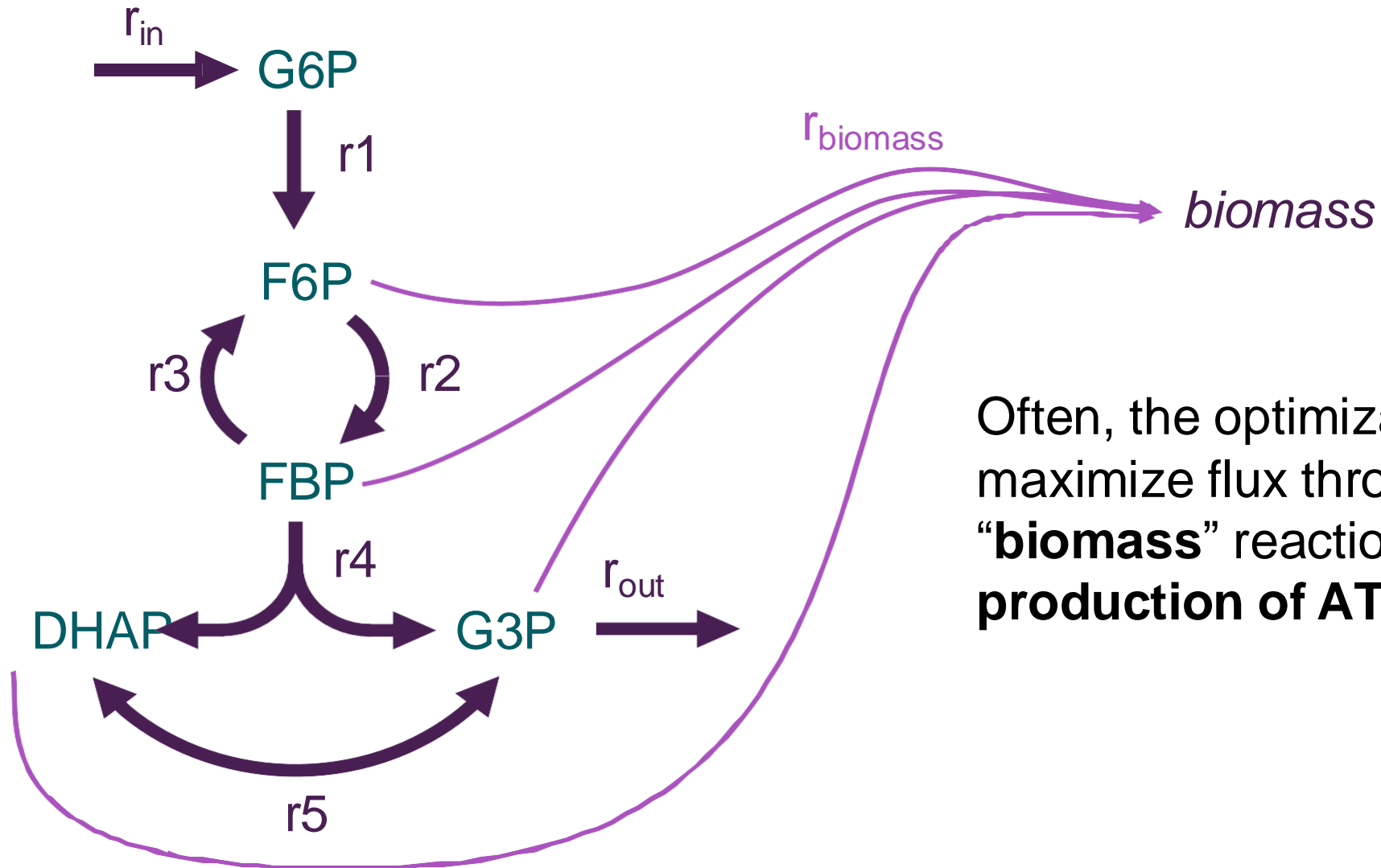


Since the problem is still **under-defined**, FBA uses linear **optimization** to identify a solution that maximizes (or minimizes) some **objective**



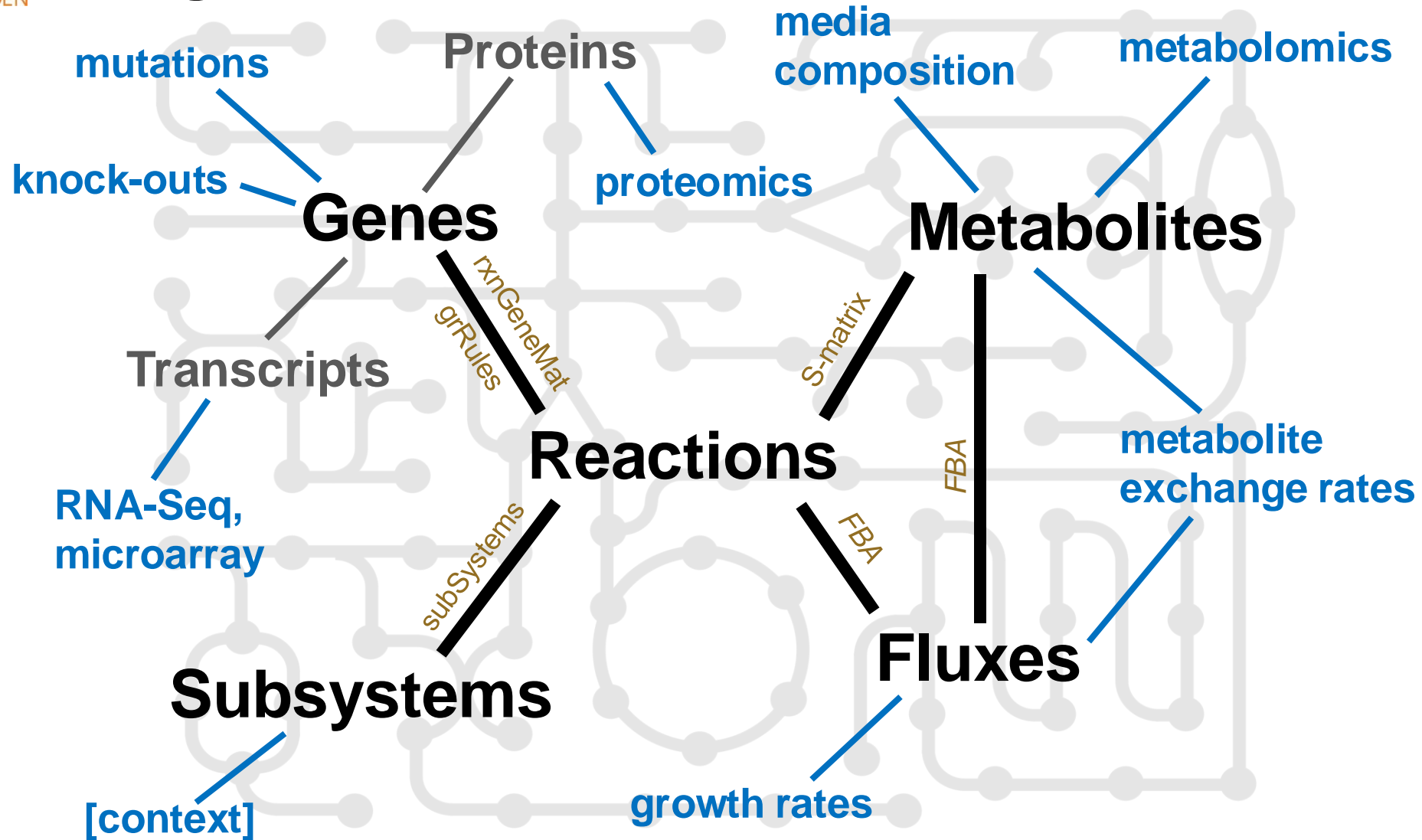
Orth, J., Thiele, I. & Palsson, B. *Nat Biotechnol* (2010).

Flux Balance Analysis (FBA)



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

Genome-scale metabolic models (GEMs) for data integration



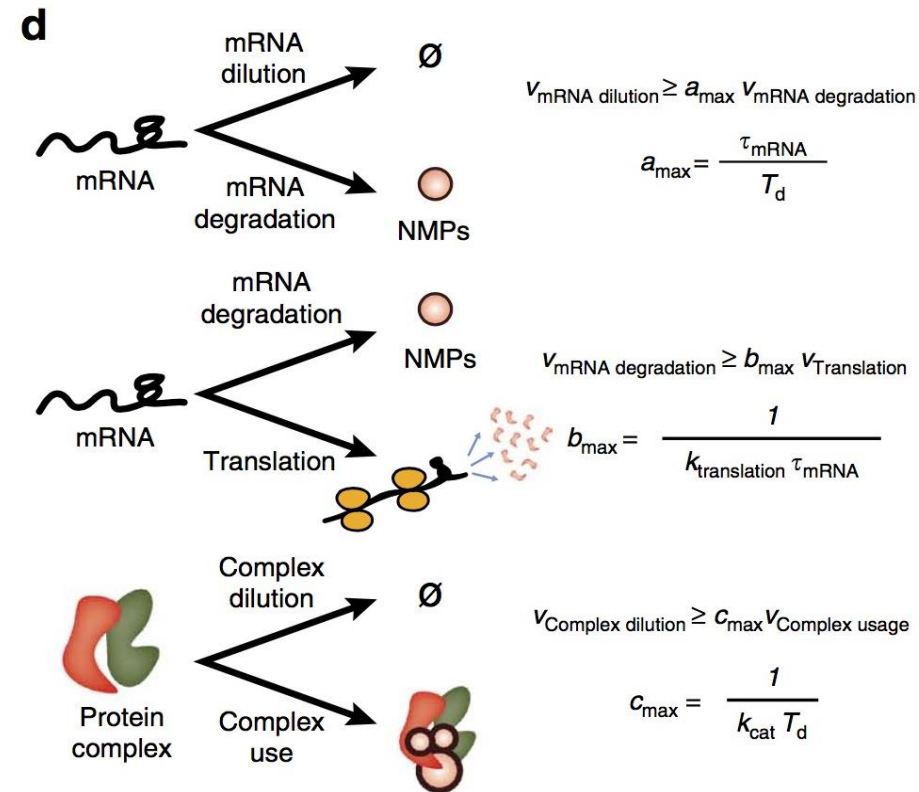
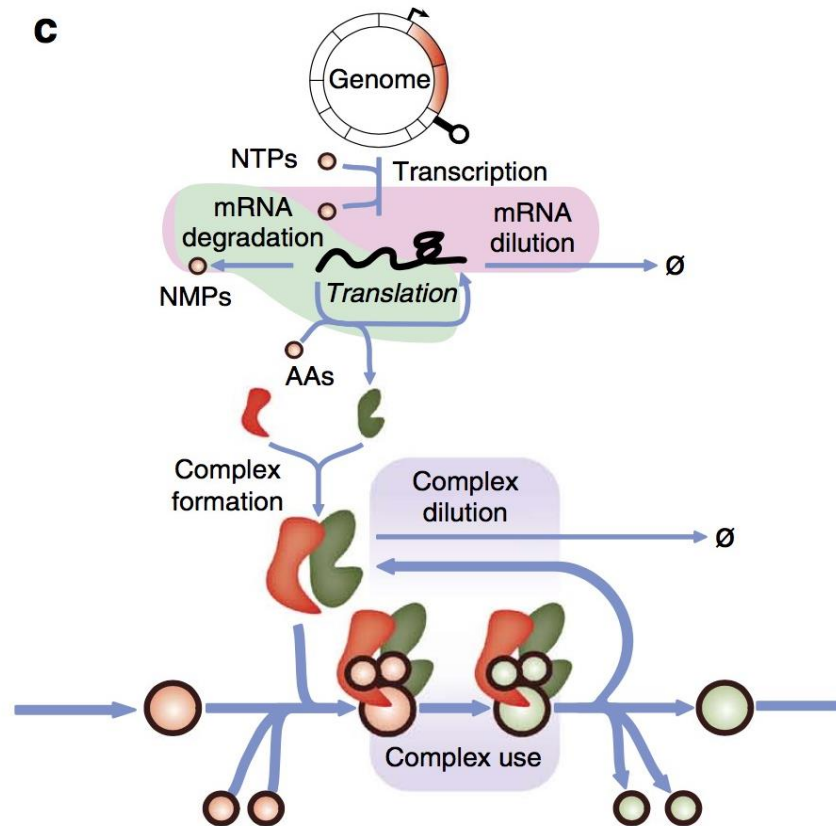


Can GEMs serve as a scaffold for integrating & studying diverse types of (omics) data?

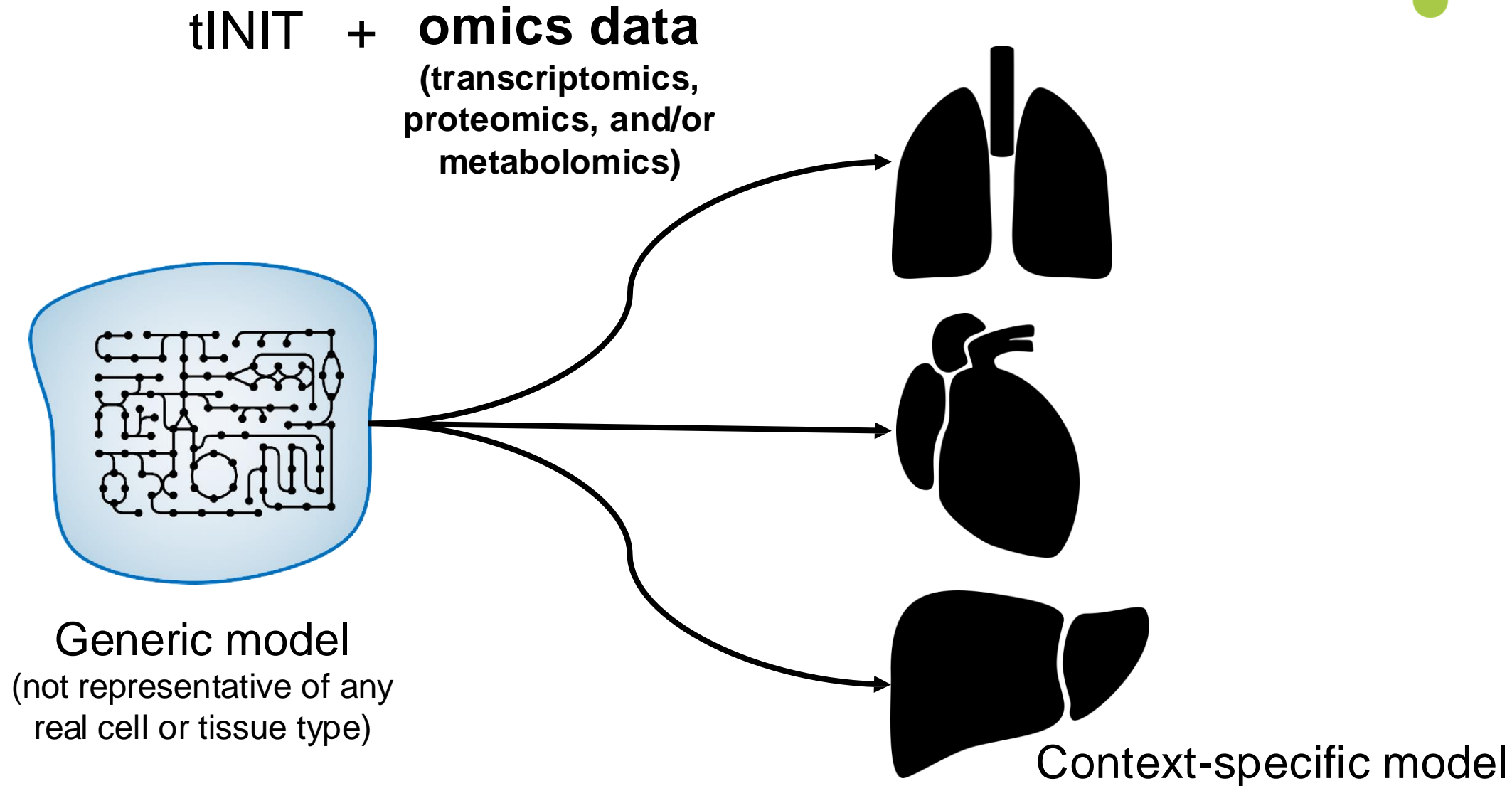
Metabolism and macromolecular expression (ME) model



J Lerman et al, Nat. Commun. 2012



GEM contextualization



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)

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

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

$$|v_i| \leq 1000 y_i$$

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

$$v_i \geq 0, i \in \text{irreversible rxns}$$

$$b_j \leq 1000 x_j$$

$$b_j + 1000(1 - x_j) \geq \varepsilon$$

$$b_j \geq 0$$

$$x_j = 1, j \in \text{present}$$

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

$$w_{i,j} = 5 \log \left(\frac{\text{Signal}_{i,j}}{\text{Average}_i} \right)$$



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

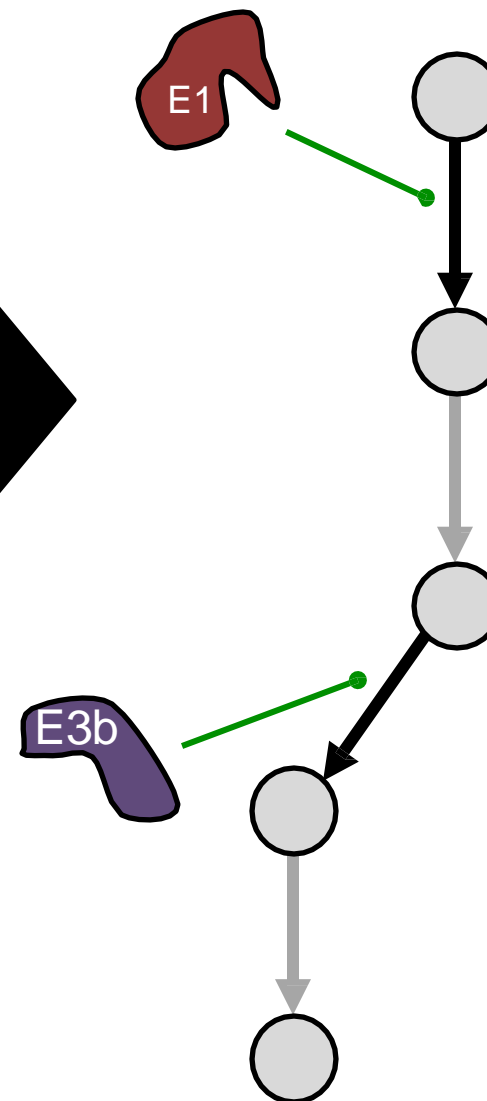
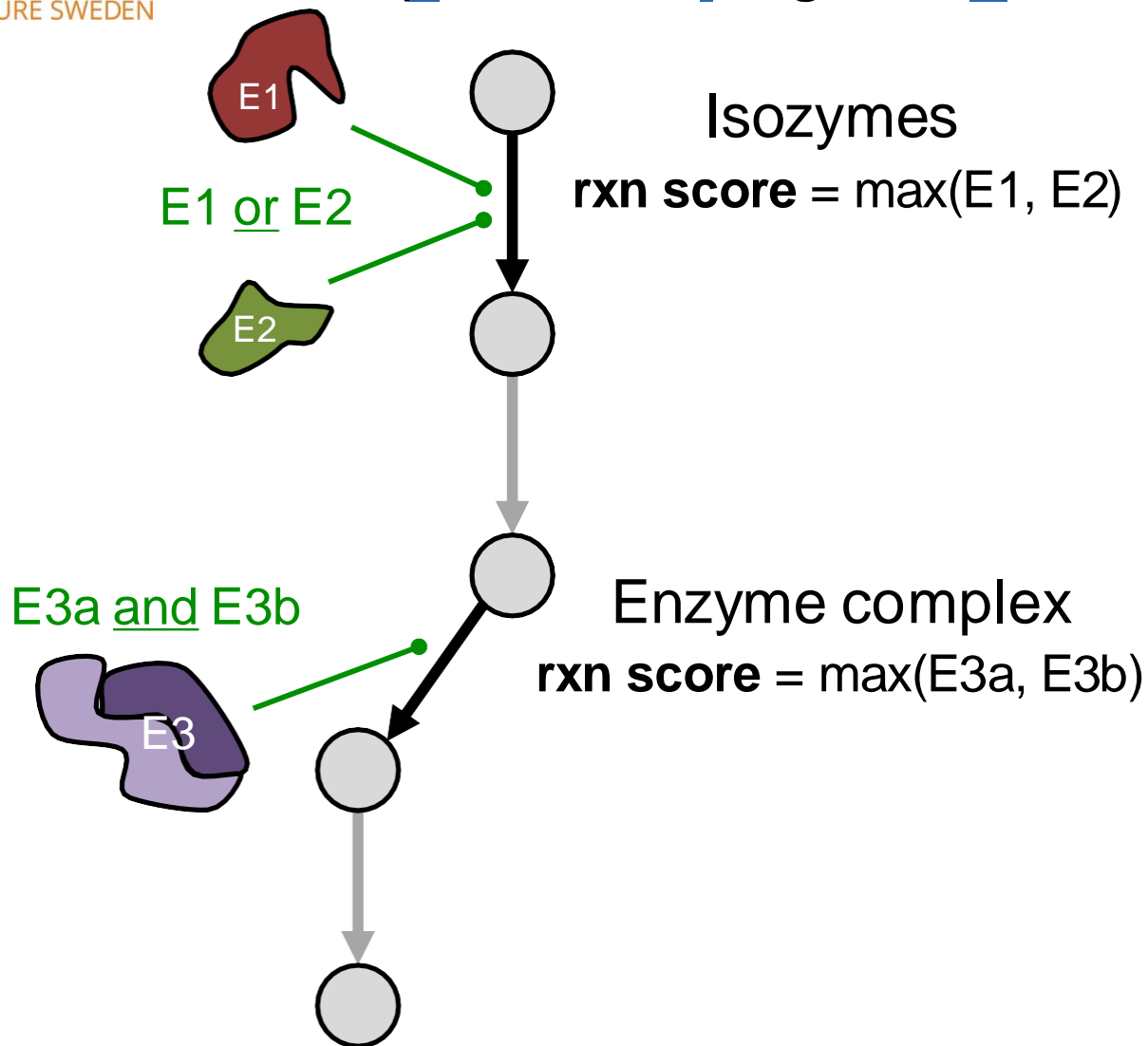
De novo synthesis of key intermediates

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

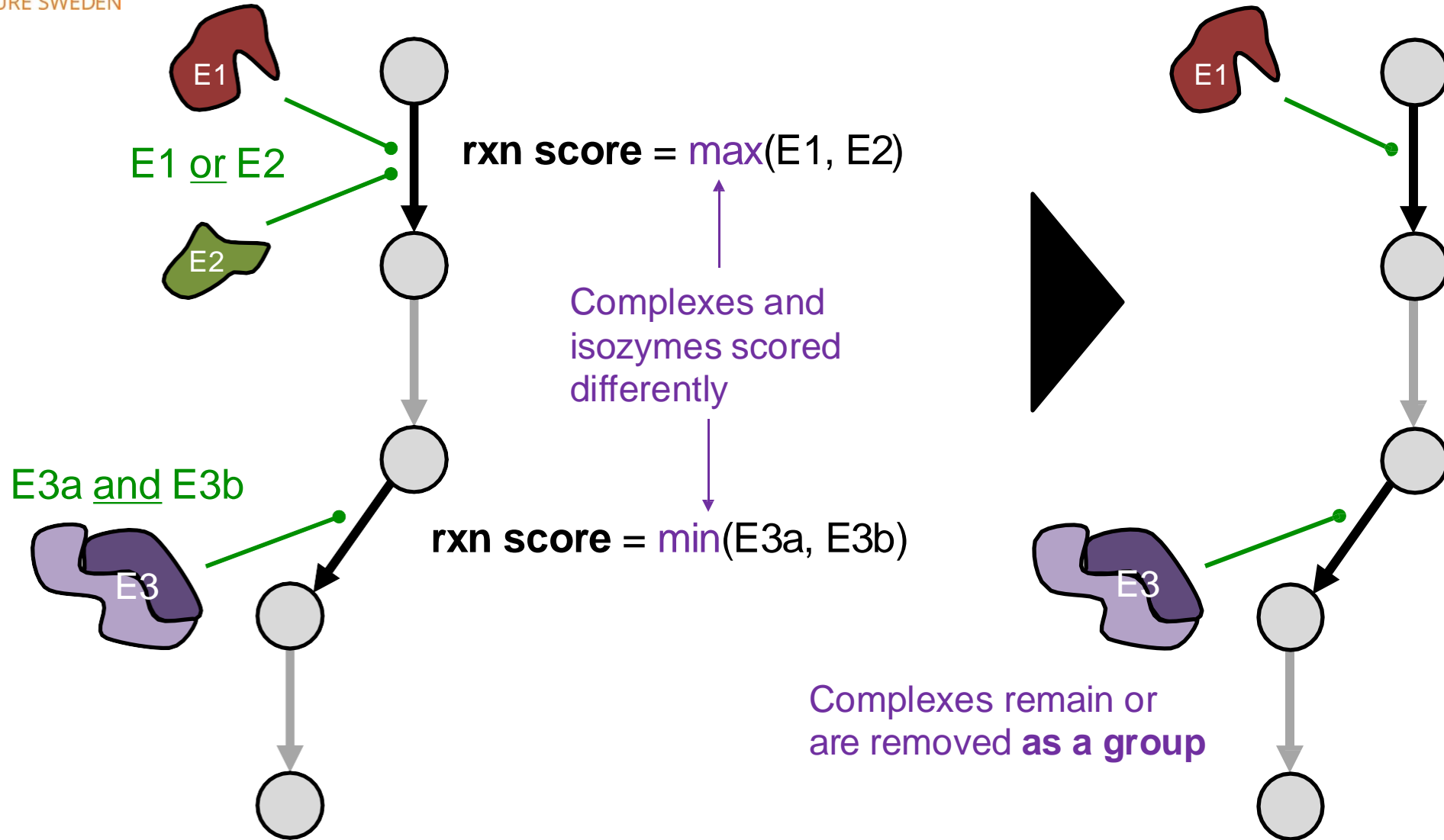


tINIT1 (Task-driven Integrative Network InfERENCE for Tissues)



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:

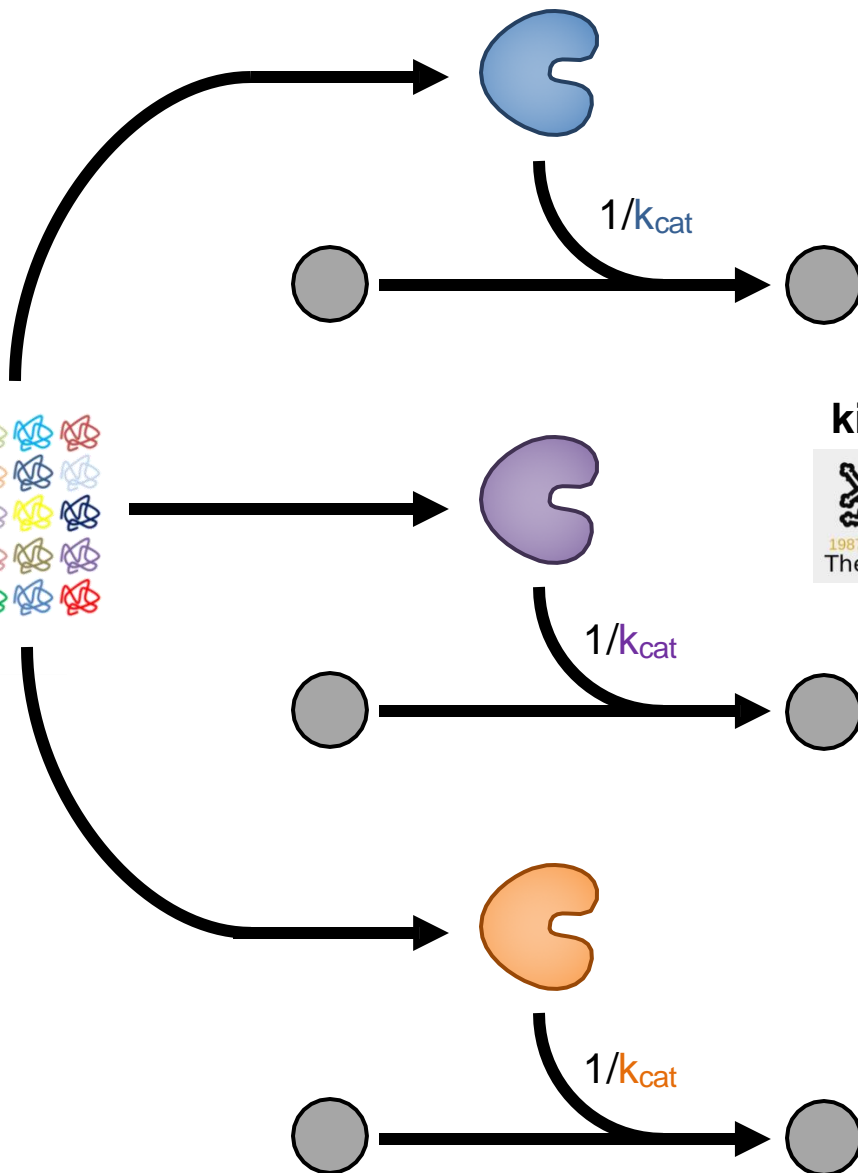
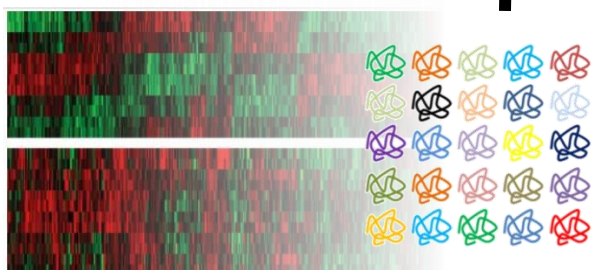
$$\begin{array}{ccccc}
 \text{Flux of reaction} & \longrightarrow & v \leq k_{\text{cat}}[E] & \longleftarrow & \text{Concentration of enzyme} \\
 \text{(from FBA)} & & & & \text{(from absolute proteomics)} \\
 & & \uparrow & & \\
 & & \text{Turnover number} & & \\
 & & \text{(from databases)} & &
 \end{array}$$

However: No simple implementation for connecting proteomics to GEMs...

Enzyme-constrained GEMs



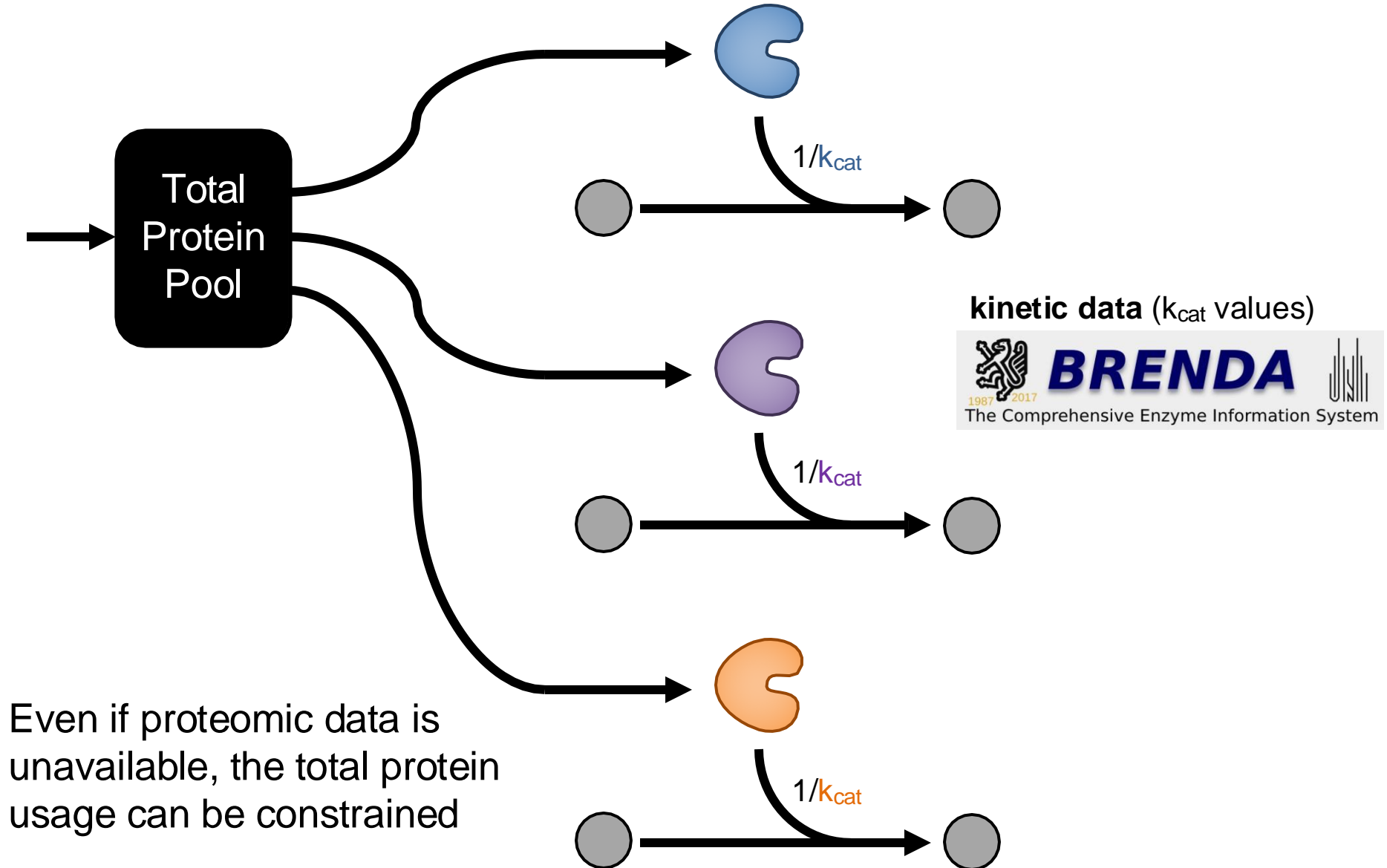
proteomic data



kinetic data (k_{cat} values)



Enzyme-constrained GEMs



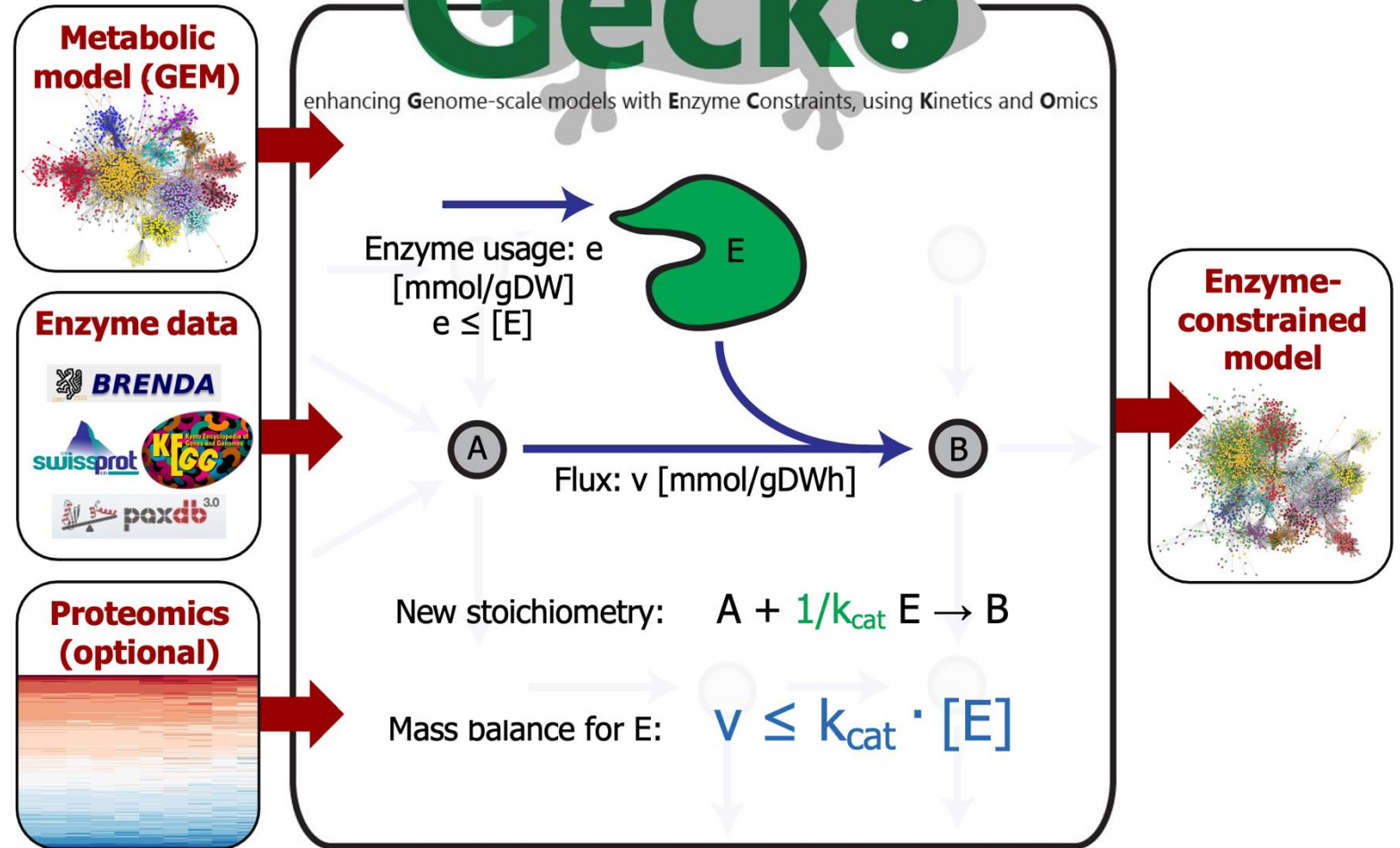
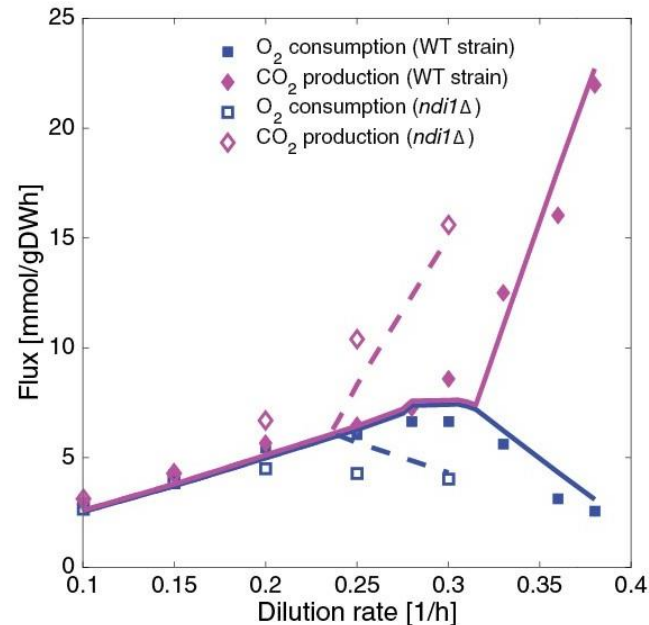
Enzyme-constrained GEMs



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

Applications:

- Improving predictions
- Integrating proteomics data into GEMs

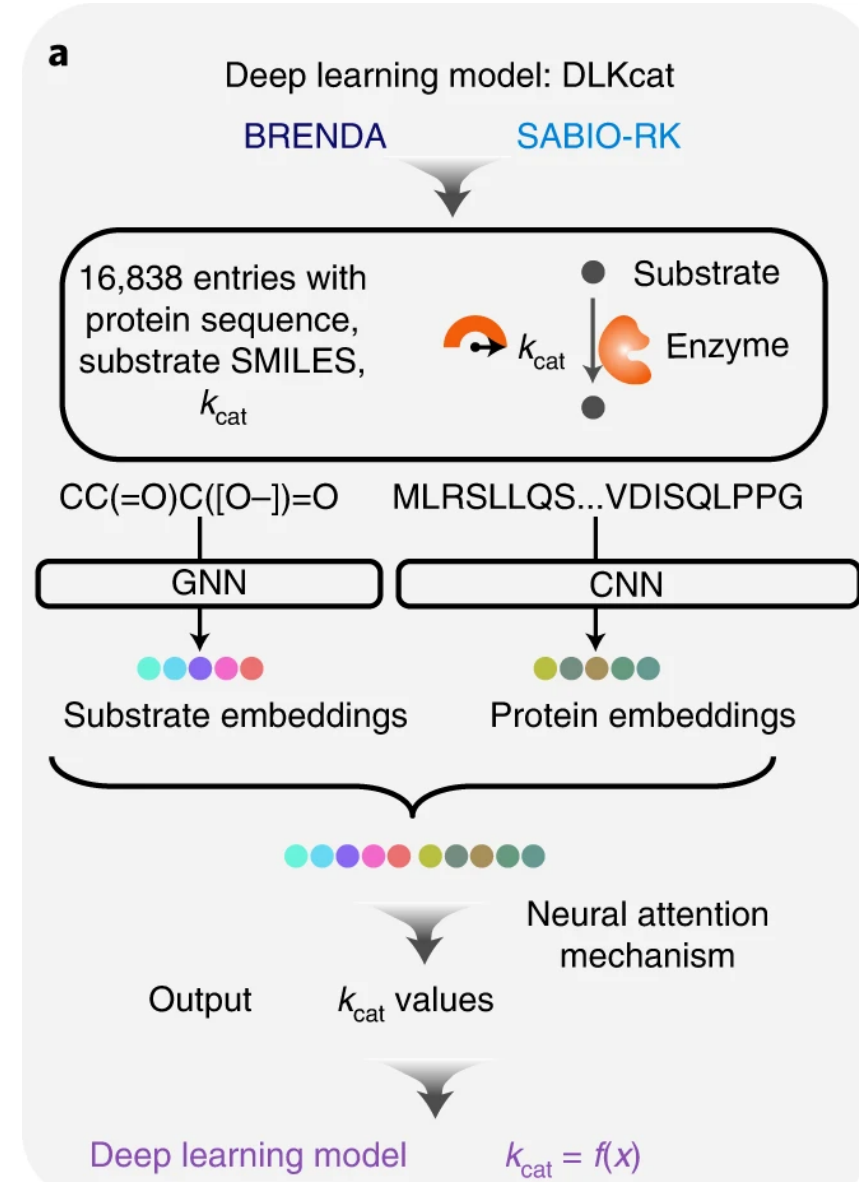
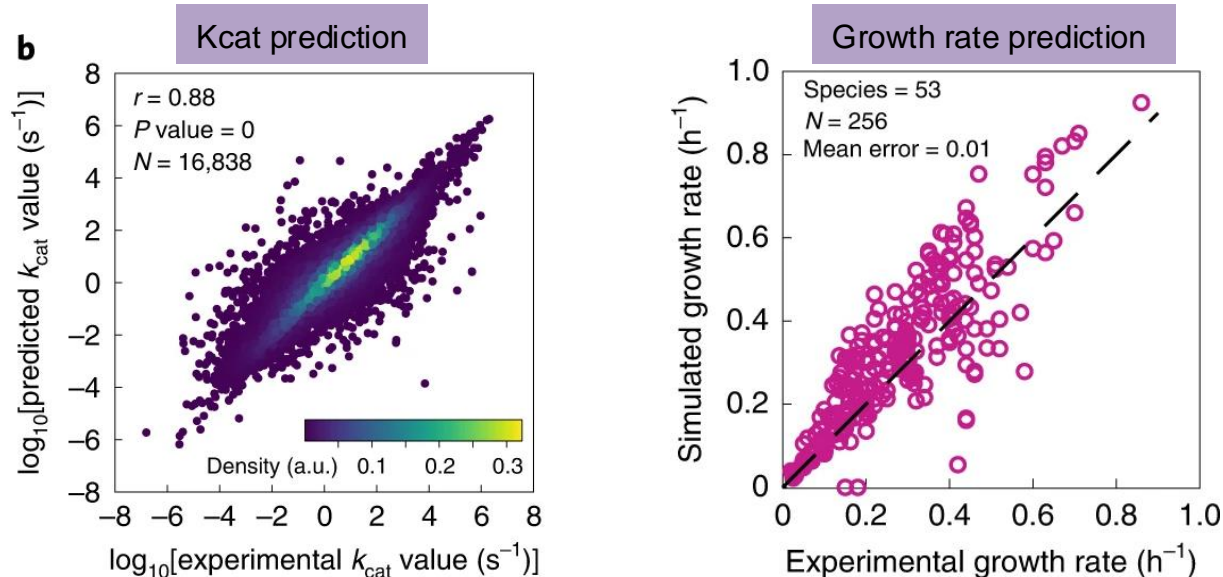


Predicting K_{cat} for ecGEM parameterization



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



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>.

The screenshot shows the GotEnzymes interface for Compound C00242 (Guanine). It includes a table of cross-references (B), a KEGG Metabolite entry (C), and a table of enzyme reactions (D).

Gene	Organism	Domain	Reaction	EC	kcat[1/s]
144811	hsa	E	R02147	2.4.2.1; 2.4.2.15	12.4392
9615	hsa	E	R01676	3.5.4.3	7.2579
3251	hsa	E	R01229	2.4.2.8; 2.4.2.22	2.3288

$$v = \frac{V_{\max}[S]}{K_m + [S]} \xrightarrow{V_{\max} = K_{cat}[E]_{\text{total}}} v = \frac{K_{cat}[E]_{\text{total}}[S]}{K_m + [S]}$$

Michaelis-Menten equation

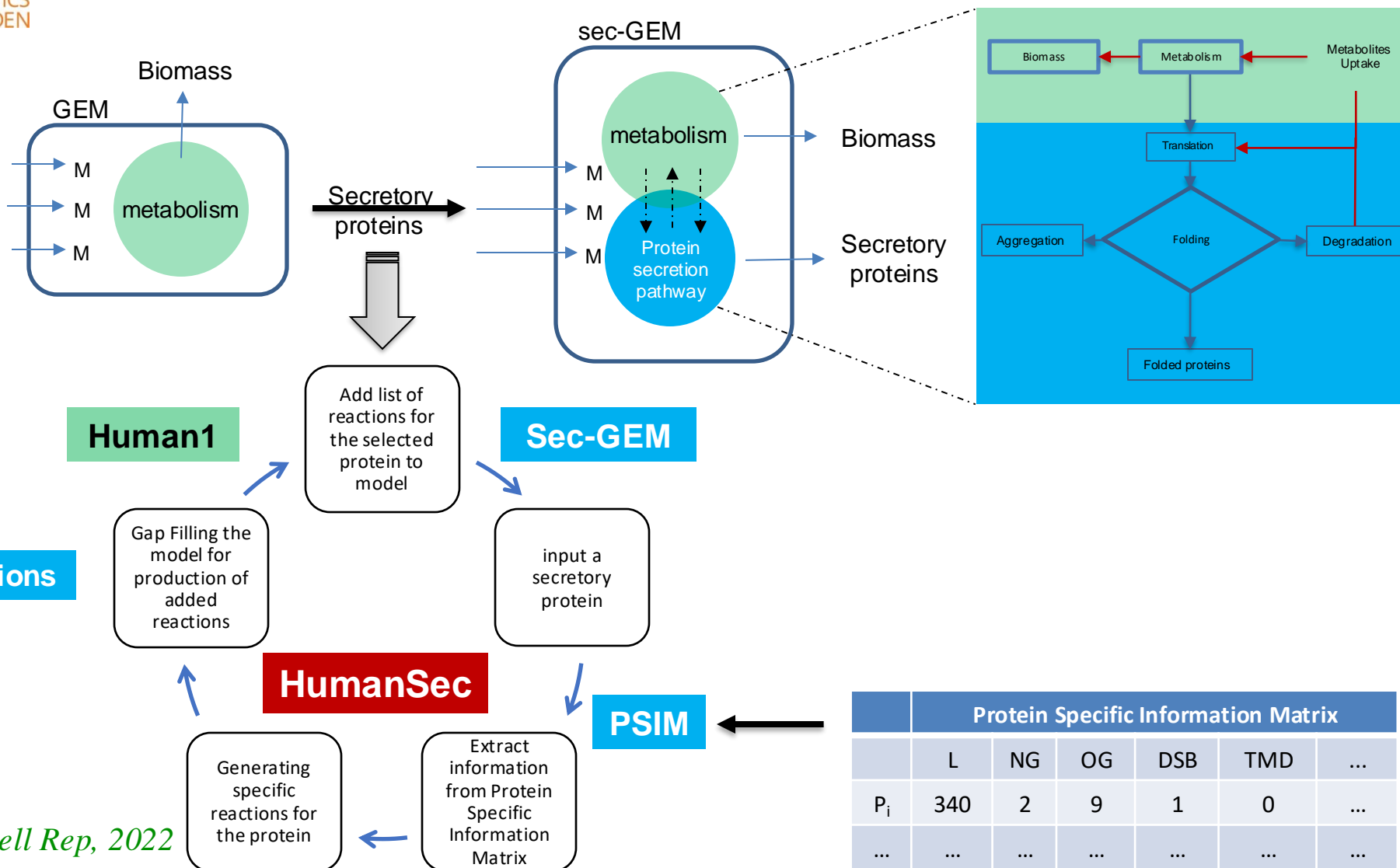
Experimental measurement / **DLKcat**

Proteomics data / whole protein constrain

Experimental measurement / **GotEnzymes** prediction

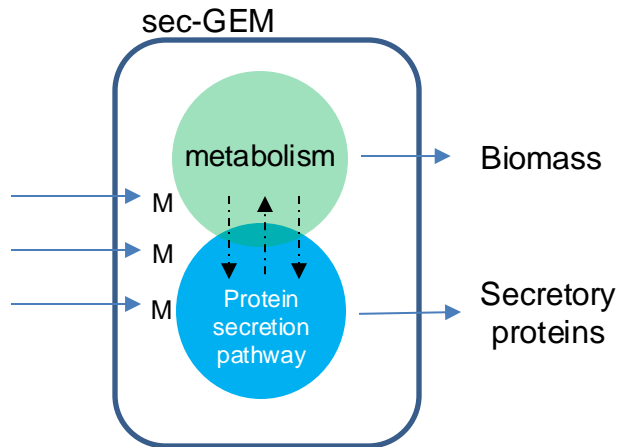
More accurate predictions

Extending the coverage of GEMs: secGEM

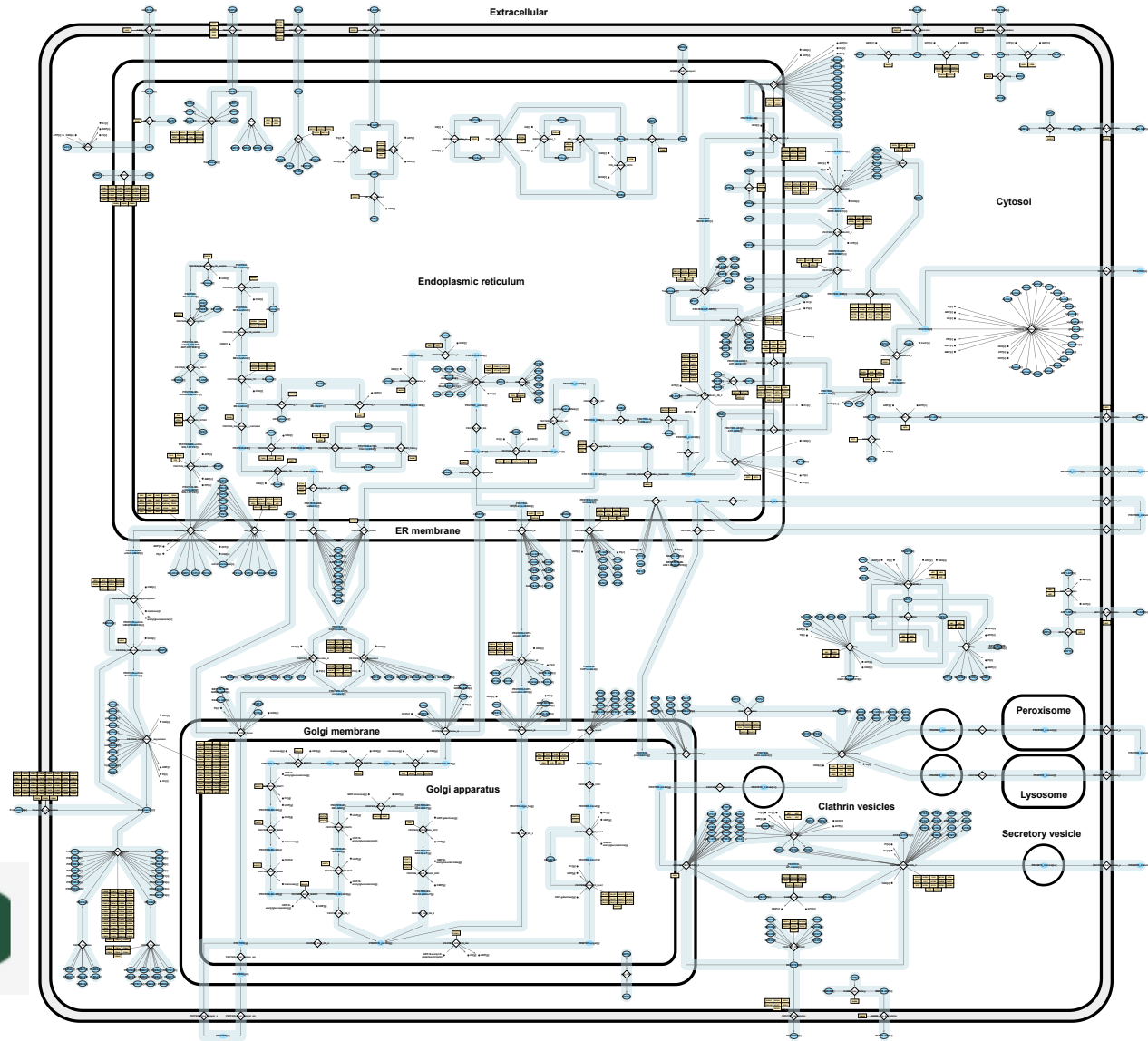


Saghaleyni, et al *Cell Rep*, 2022

Extending the coverage of GEMs: secGEM



metabolic
ATLAS



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 but in the context of metabolism.

