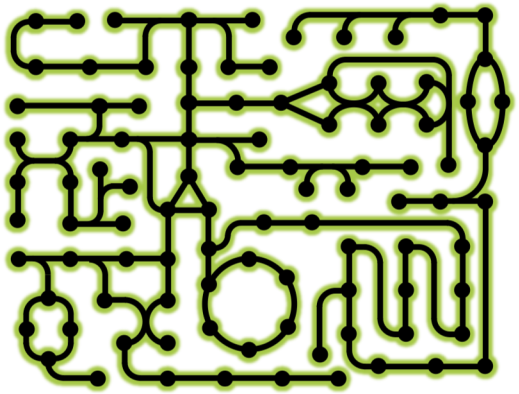


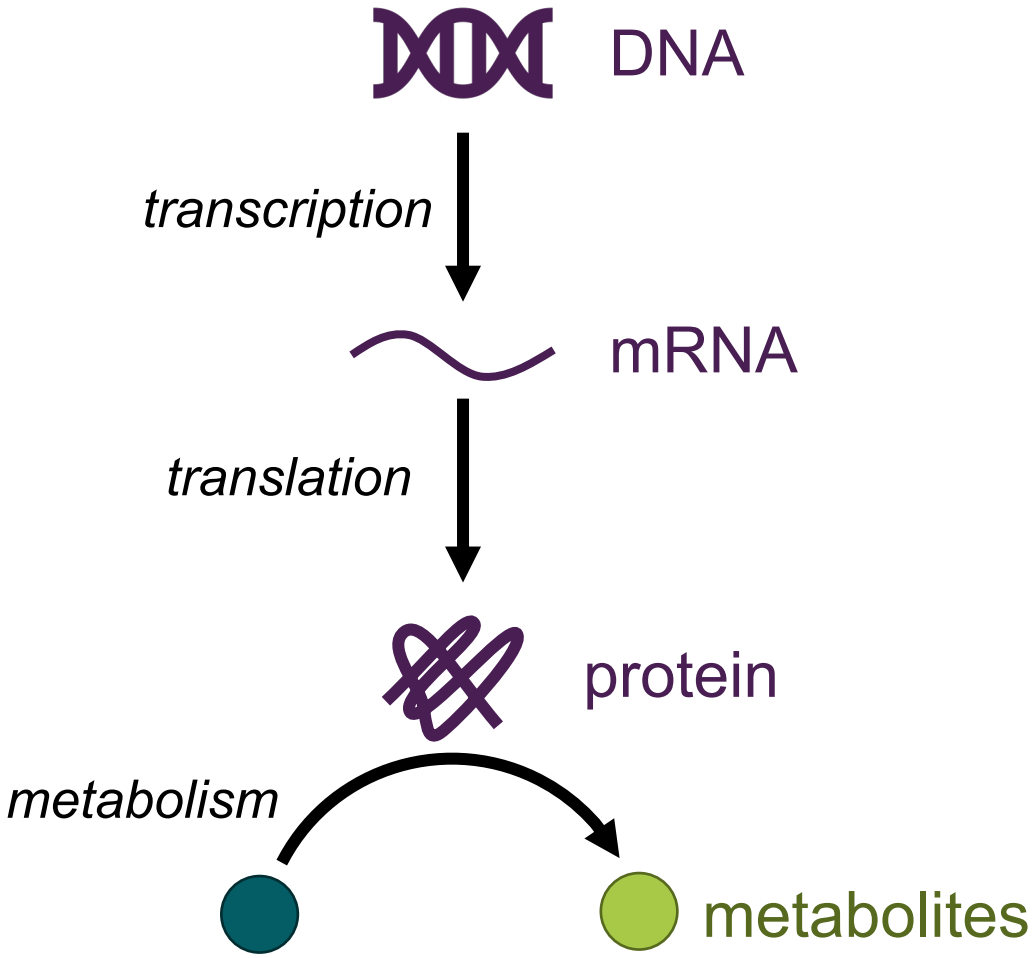
GEMs

Structure & Simulation

NBIS Omics Integration and Systems Biology workshop
Fall 2020, Lund University

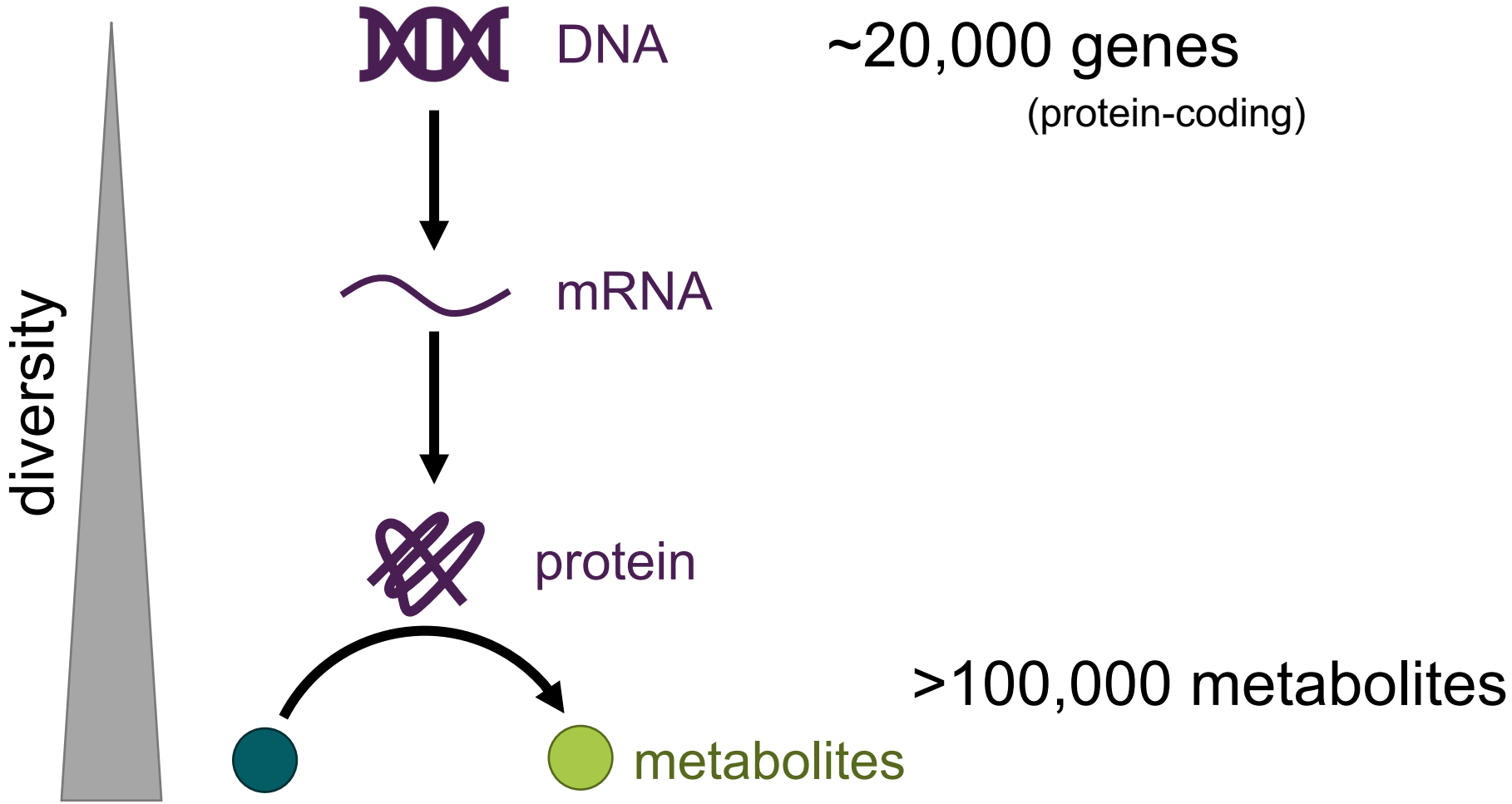
Jonathan Robinson
National Bioinformatics Infrastructure Sweden (NBIS)
Science for Life Laboratory (SciLifeLab)
Chalmers University of Technology
jonathan.robinson@scilifelab.se

Background



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

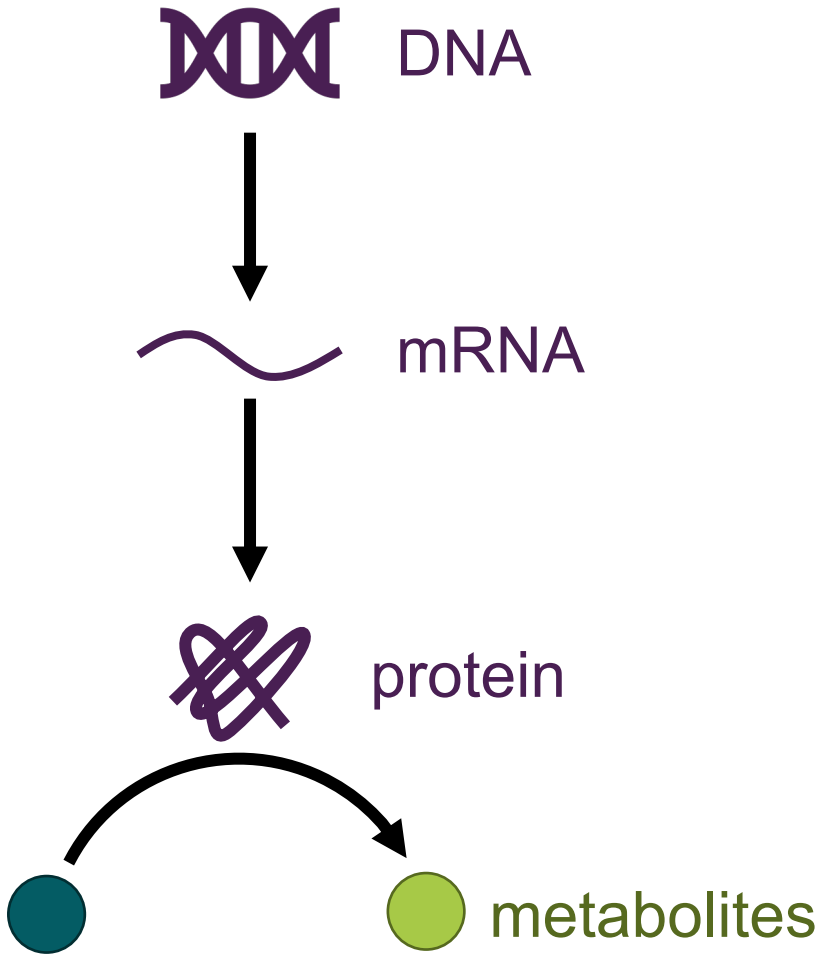
Background



~20,000 genes
(protein-coding)

>100,000 metabolites

Background



stable

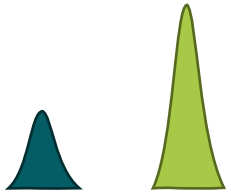
relatively stable

stable

ranges from stable
to very unstable

Background

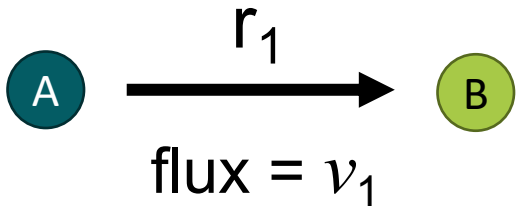
We can generally measure metabolite concentrations



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

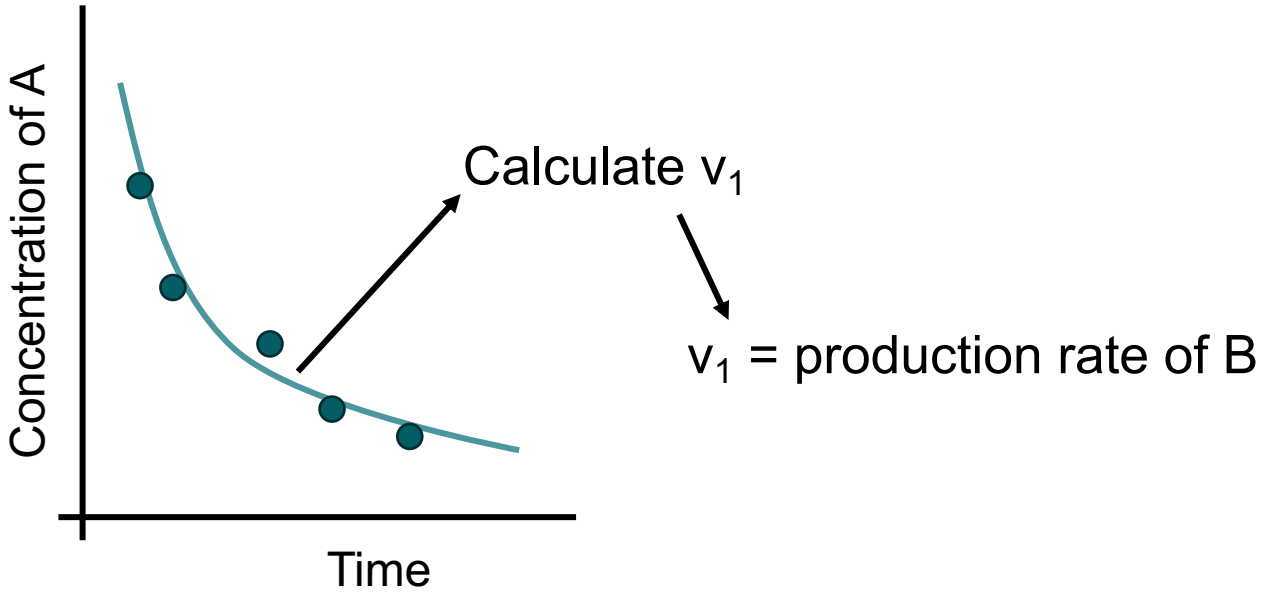
Background

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

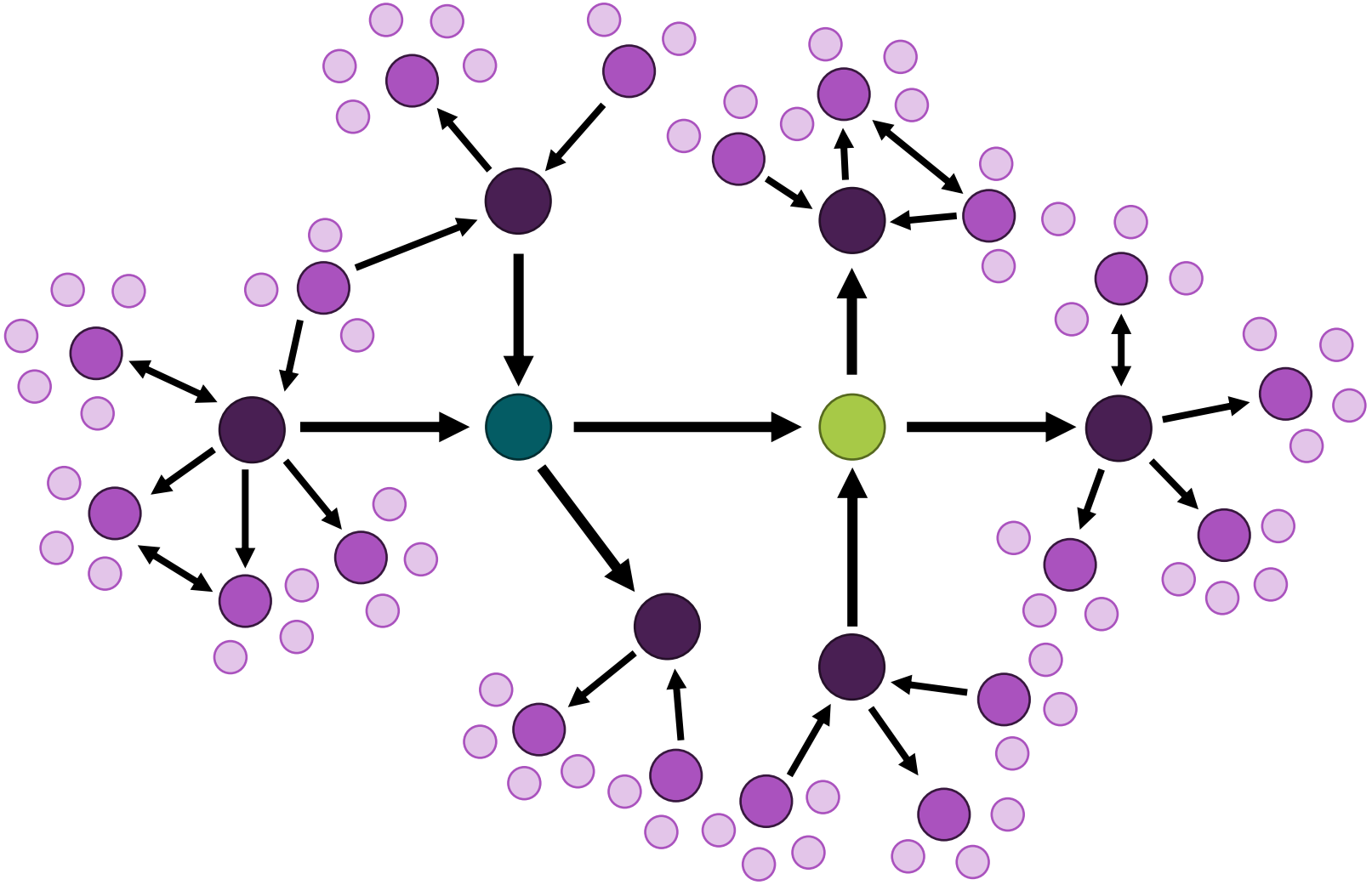


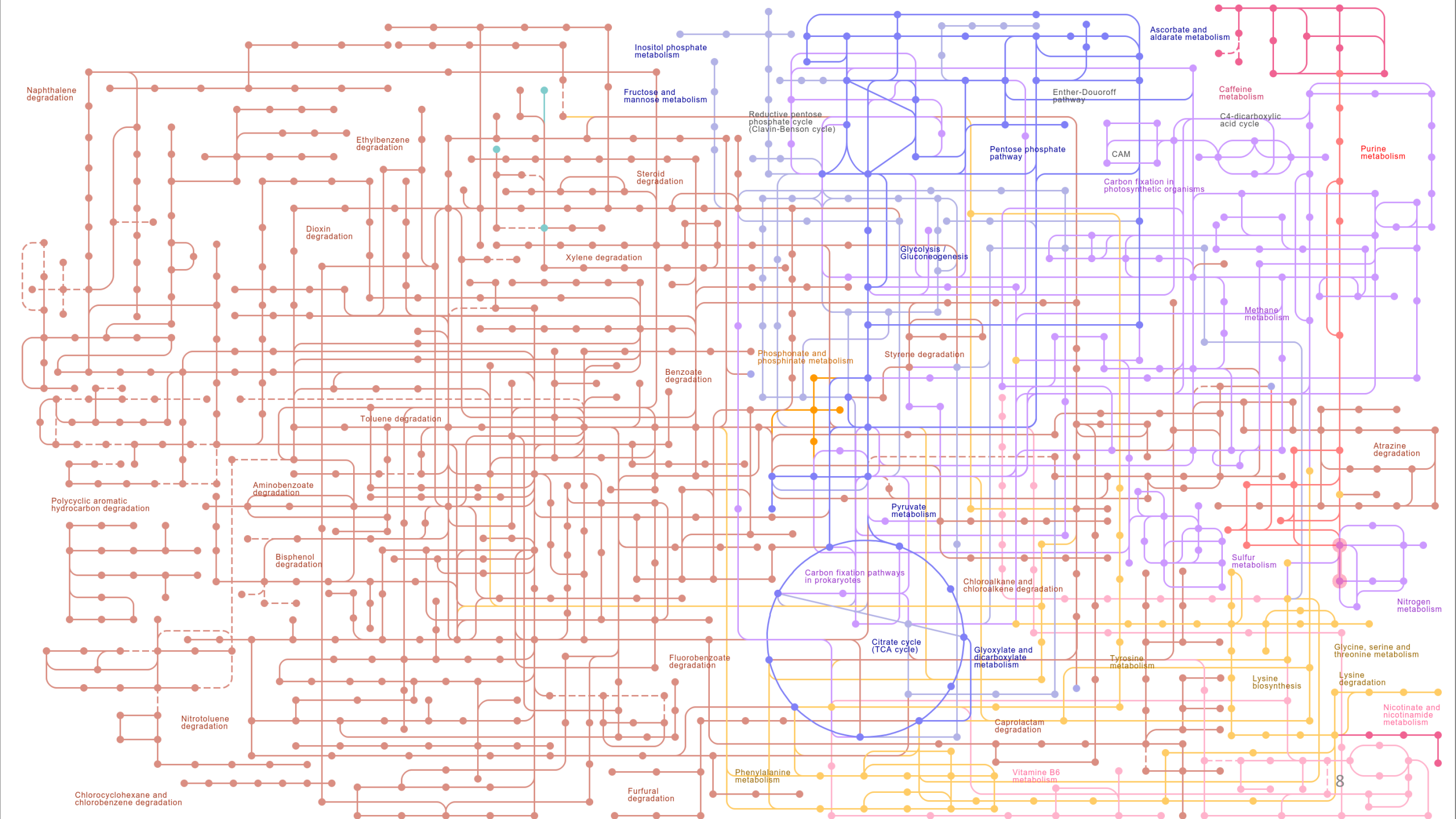
$$\frac{dA}{dt} = -v_1$$

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

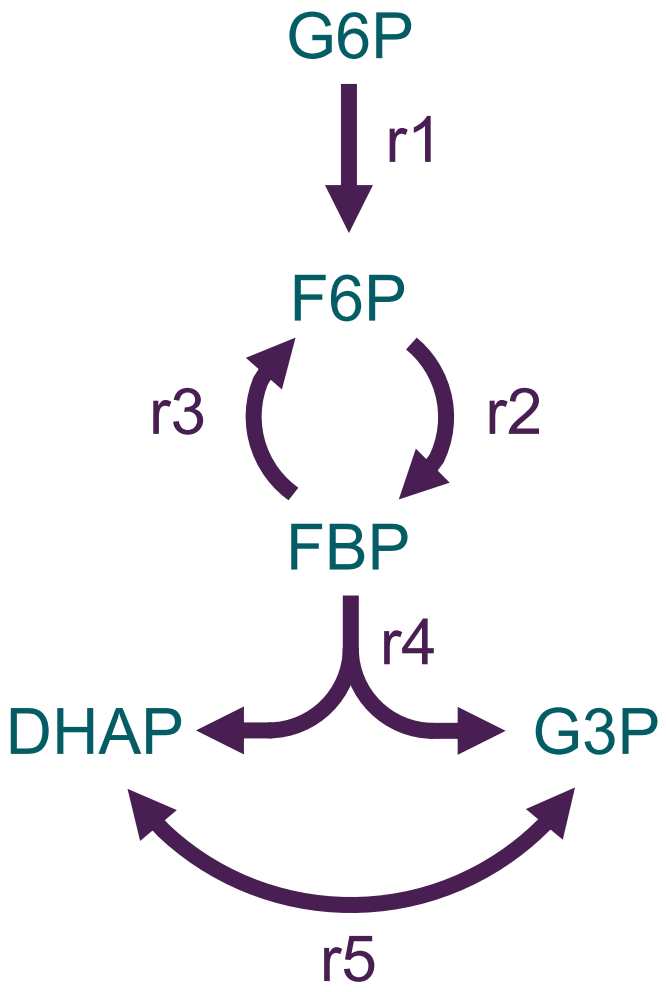


Background




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

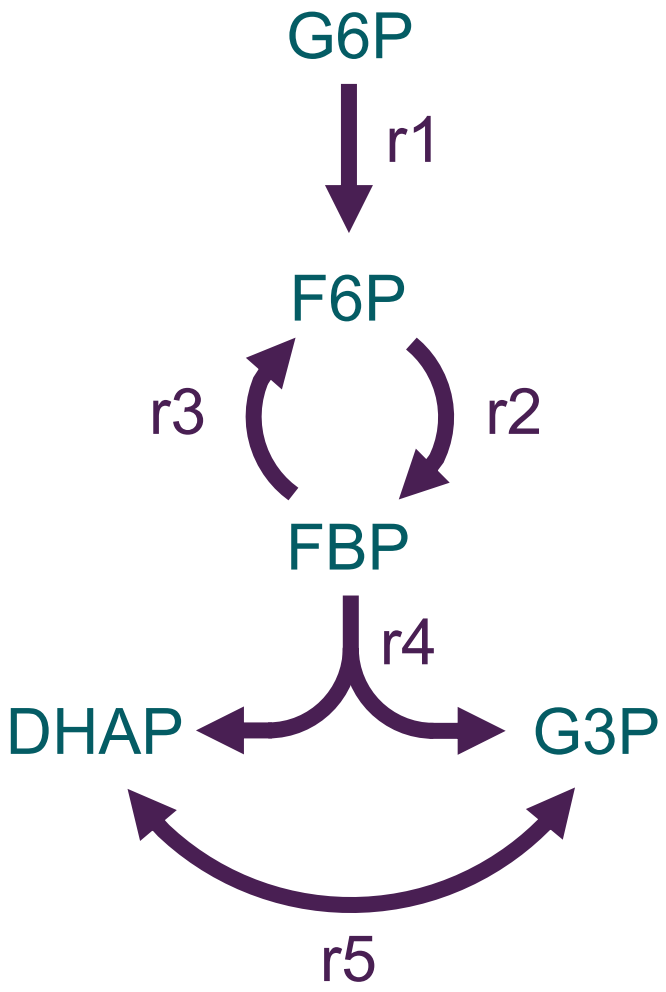
The Stoichiometric Matrix



Reactions

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

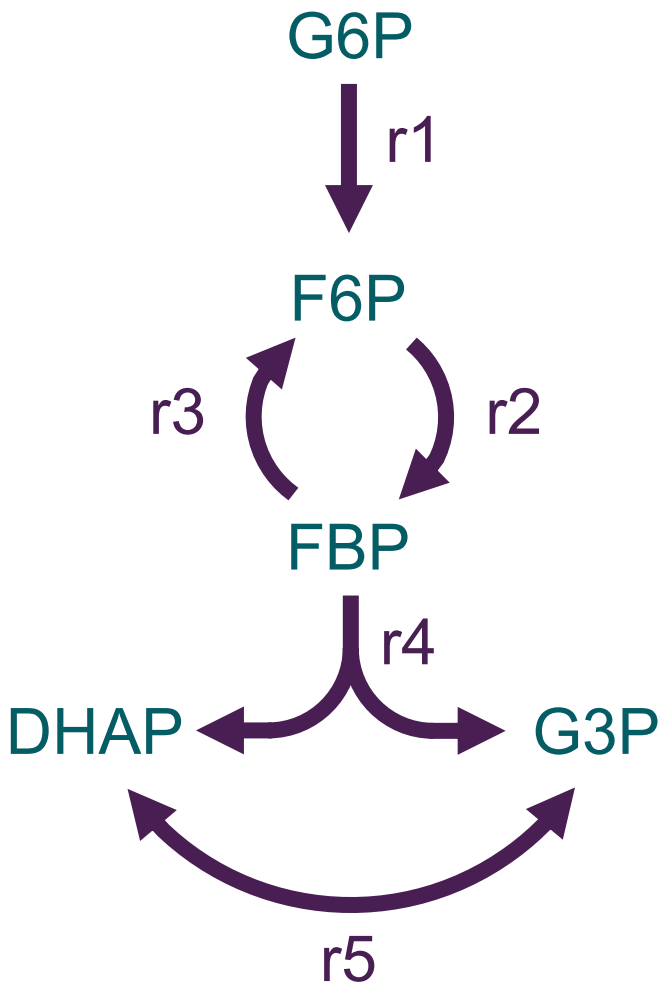
The Stoichiometric Matrix



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 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

…

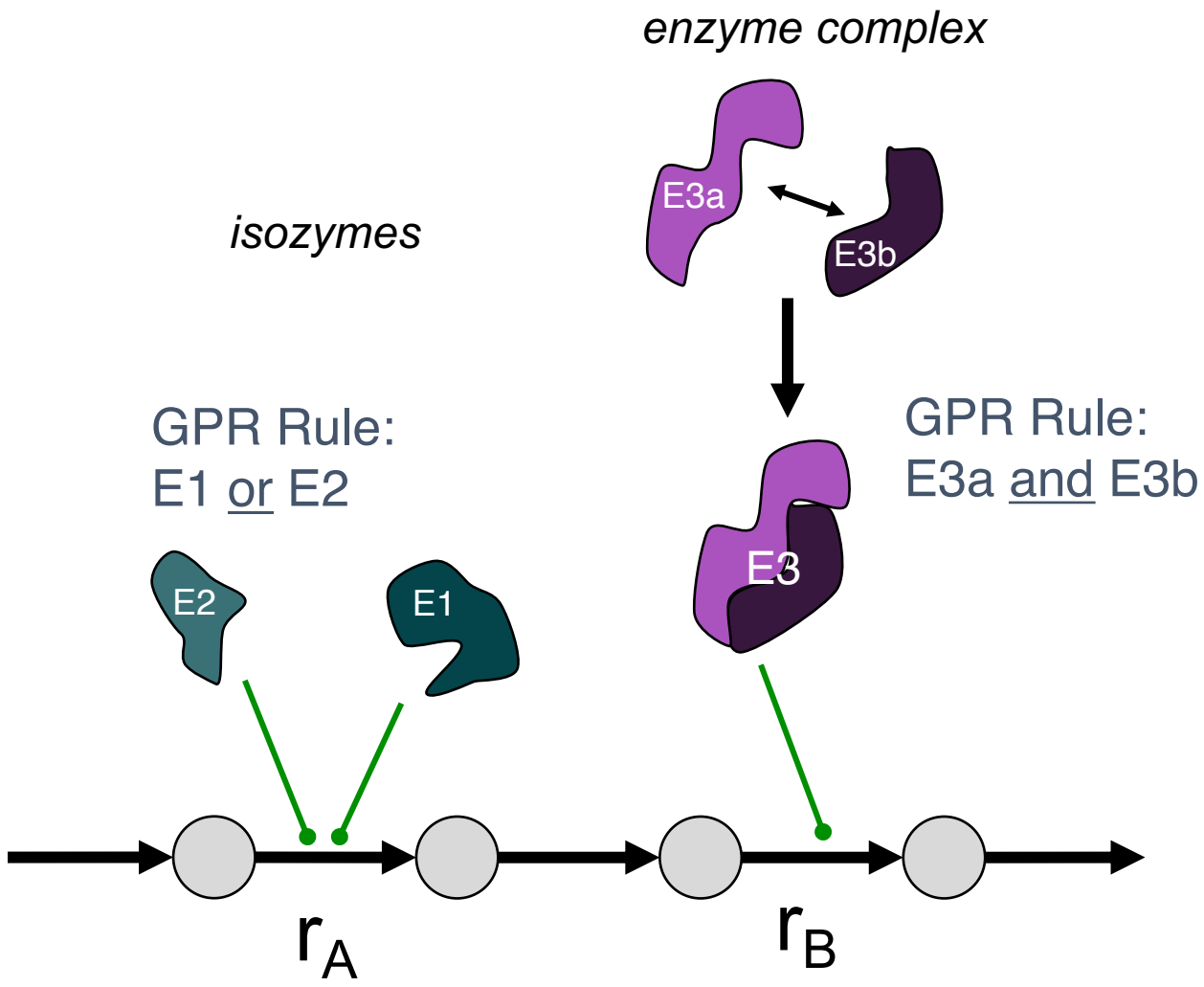
Symbol	r1	r2	r3	r4	r5
G6P	**−1**	0	0	0	0
F6P	**1**	**−1**	**1**	0	0
FBP	0	**1**	**−1**	**−1**	0
DHAP	0	0	0	**1**	**−1**
G3P	0	0	0	**1**	**1**

S

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

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

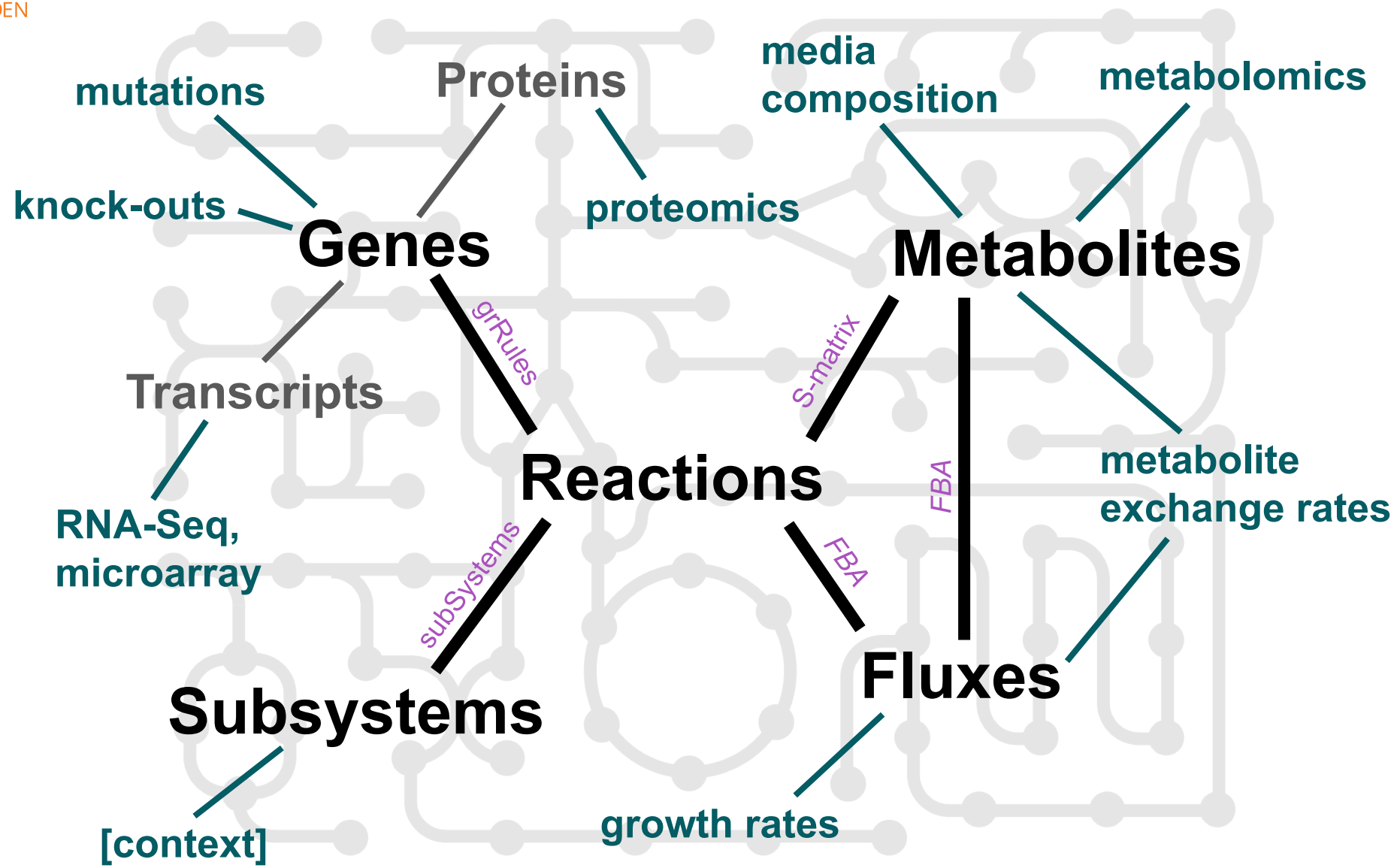
GPR Rules



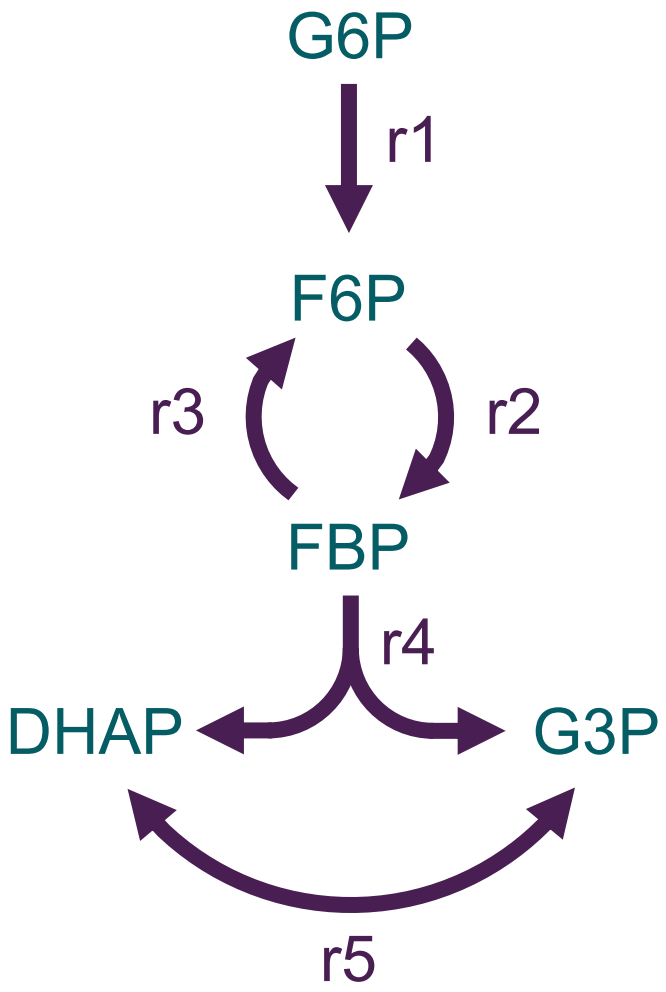
GPR Rules enable more accurate simulation of gene inactivation/knock-out

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

GEMs as an integrative tool



Flux Balance Analysis (FBA)

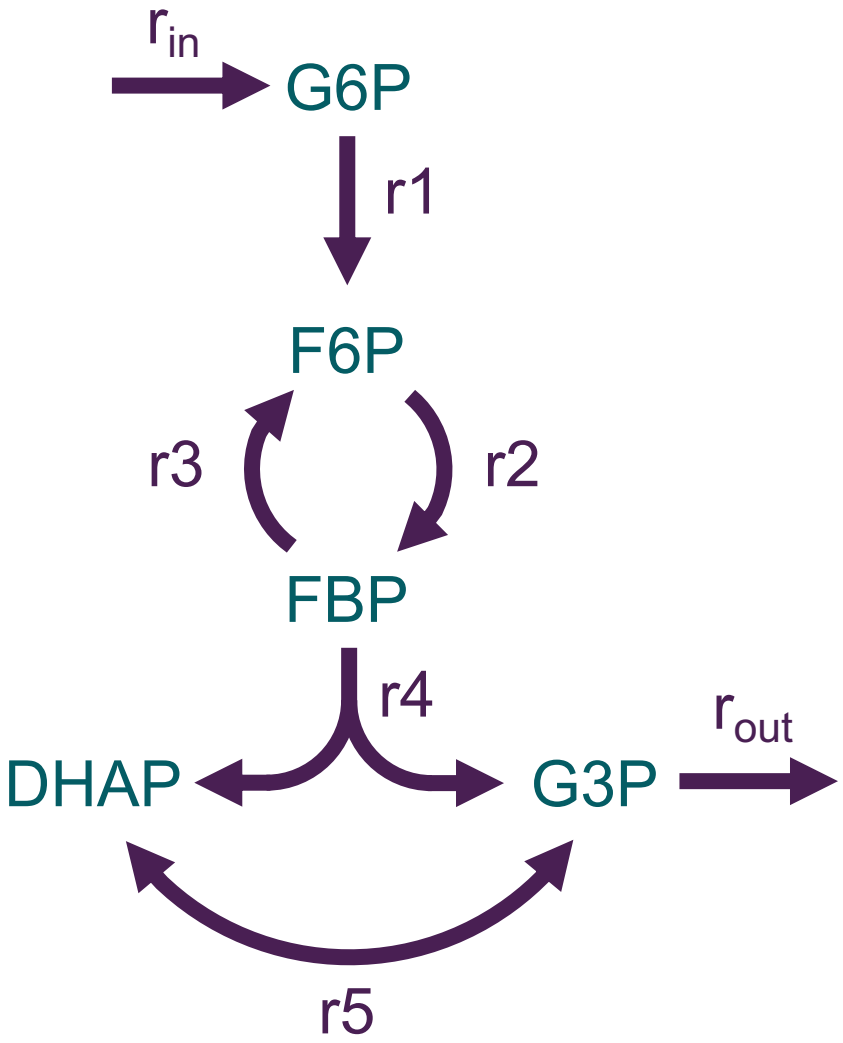


Reactions

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**

Flux Balance Analysis (FBA)

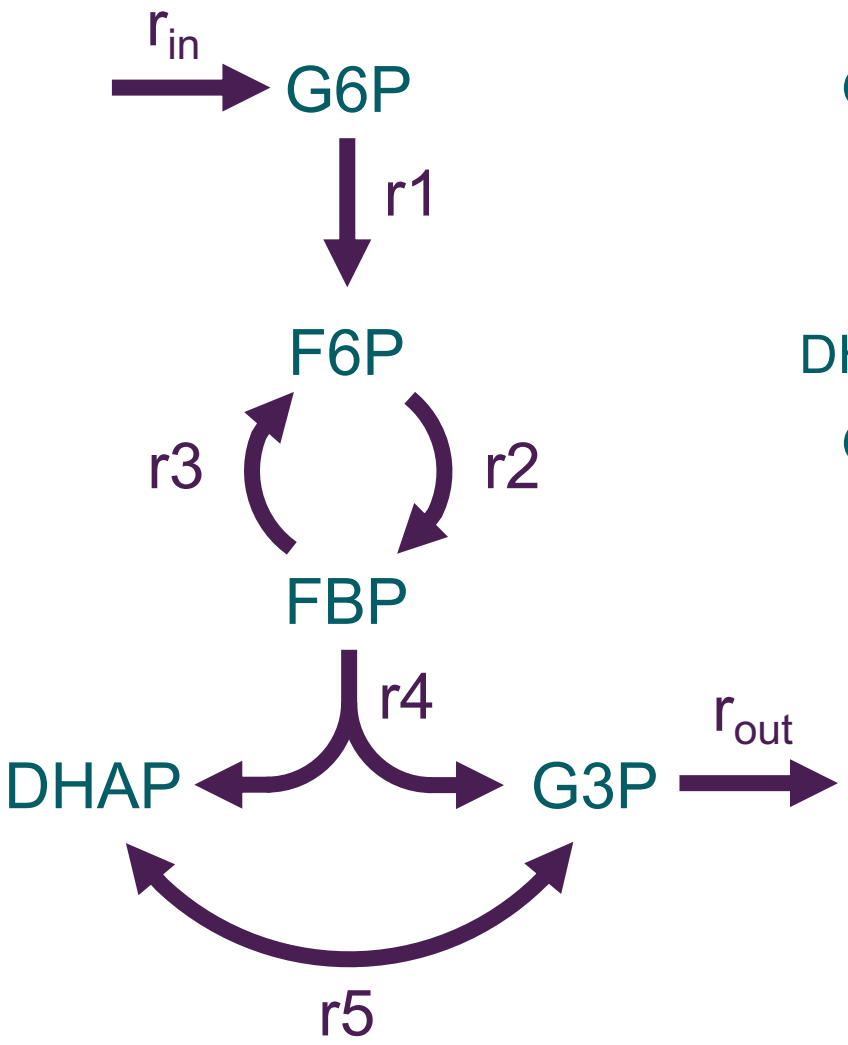


FBA seeks to calculate the reaction **fluxes** (***v***) of a network

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)

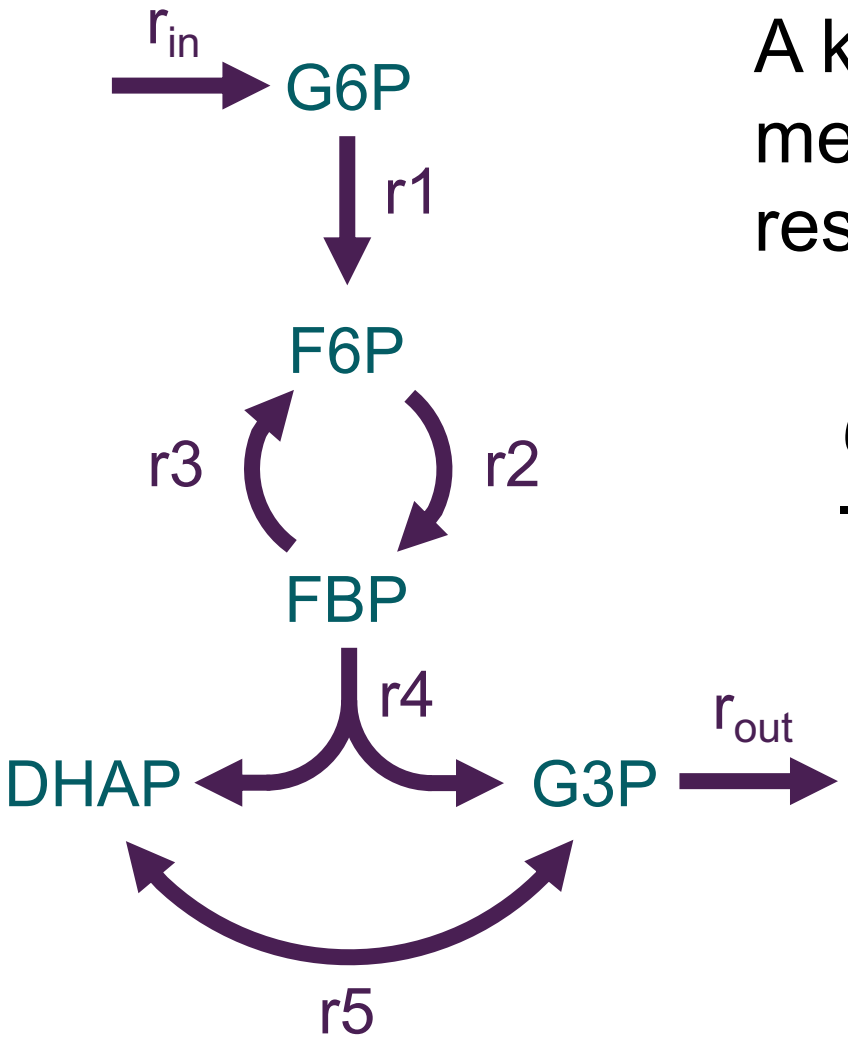


$$\begin{array}{c} \\ G6P \\ F6P \\ FBP \\ DHAP \\ G3P \end{array} \begin{array}{c} \begin{array}{ccccccc} r1 & r2 & r3 & r4 & r5 & r_{in} & r_{out} \end{array} \\ \begin{pmatrix} -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \\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

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

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

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

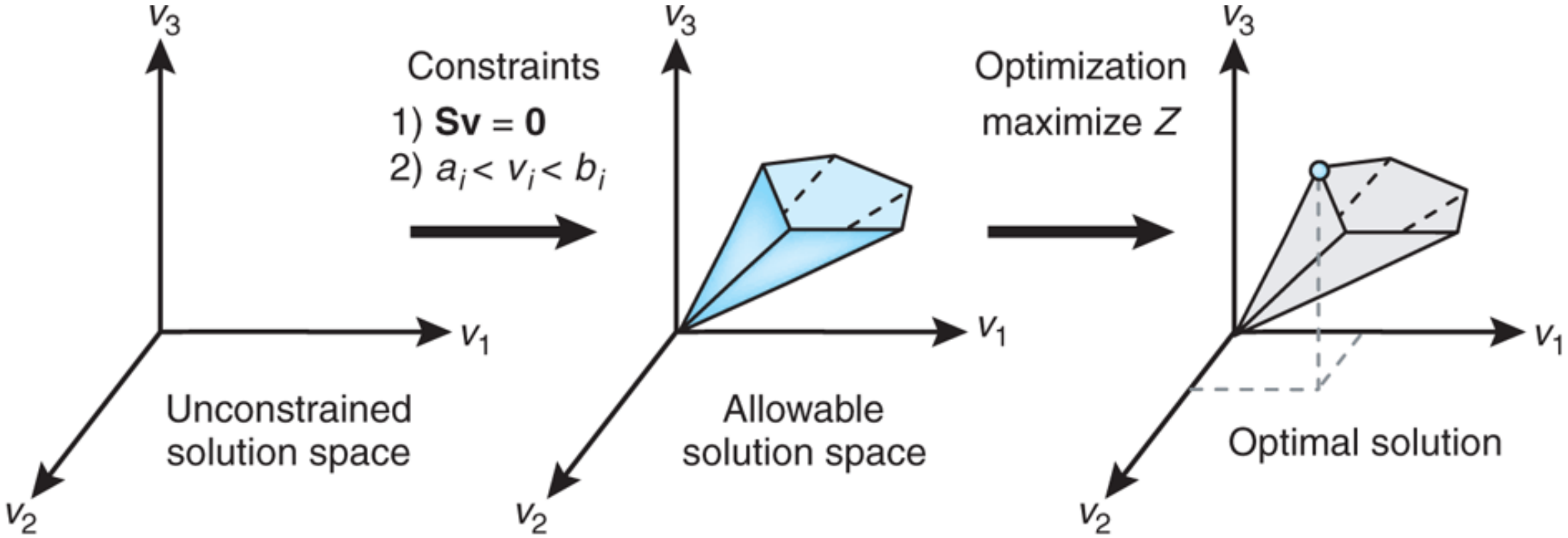
$0 \leq v \leq ub$

Reversible reactions

$lb \leq v \leq ub$

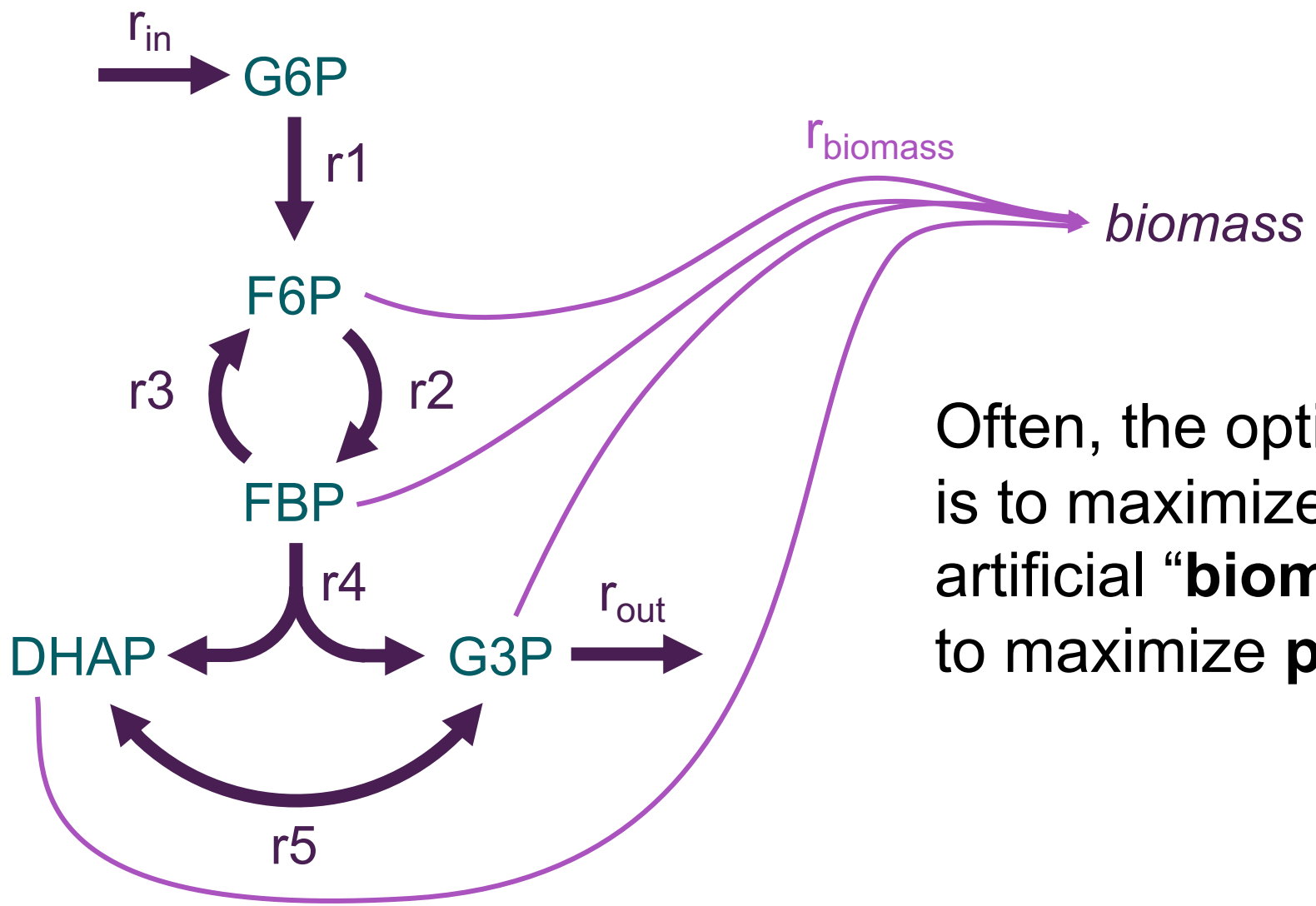
Flux Balance Analysis (FBA)

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**.

Exercise: COBRApy

COBRApy (COnstraint-Based Reconstruction and Analysis) is a package that facilitates the use of GEMs in python

The exercise will walk through the basics of GEM structure, functionality, flux balance analysis, and gene knockouts.