

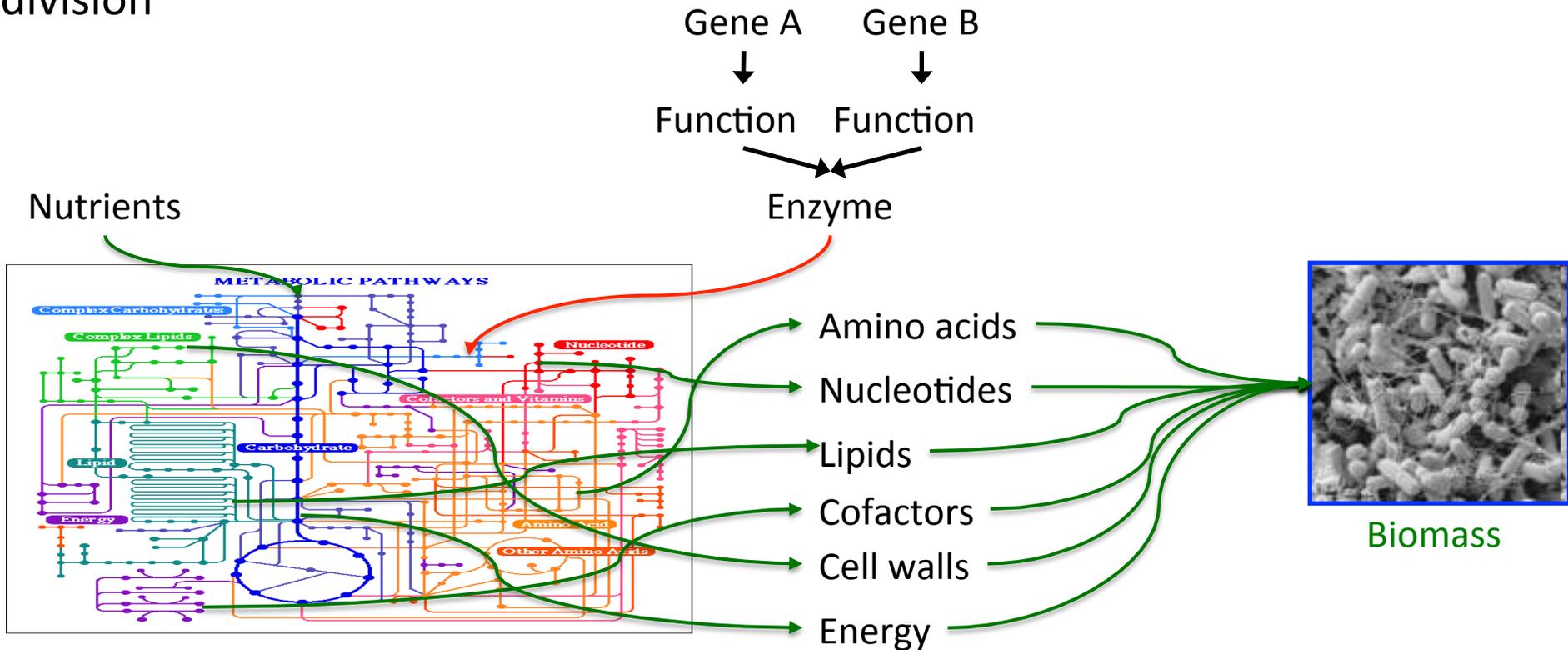
PlantSEED Workshop: Introduction to Metabolic Modeling

Presented by Christopher Henry

Metabolic Modeling to Predict Phenotype from Genotype

What is a metabolic model?

- 1.) A list of all reactions involved in the metabolic pathways
- 2.) A list of rules associating reaction activity to gene activity
- 3.) A biomass reaction listing essential building blocks needed for growth and division



Biochemistry Database for Model Reconstruction

- The PlantSEED biochemistry database is constructed by an integration of biochemistry from numerous source, but the bulk of the data is from AraCyc, PMN, Gramene, KEGG



- 1.) Combine data from all sources in a common format
- 2.) Merge identical compounds from multiple resources based on inchi-key (structure) or exact name match
- 3.) Merge reactions with identical stoichiometry
- 4.) Compute charge for each compound from structure of pH7 using MarvinBeans
- 5.) Compute compound and reaction delta G using group contribution method

Estimating $\Delta_r G'^{\circ}$ with the Group Contribution Method

•Equation for estimating $\Delta_f G'^{\circ}$

$$\Delta_f G'^{\circ} = \sum_{i=0}^{N_{gr}} n_i \Delta_{gr} G'_i$$

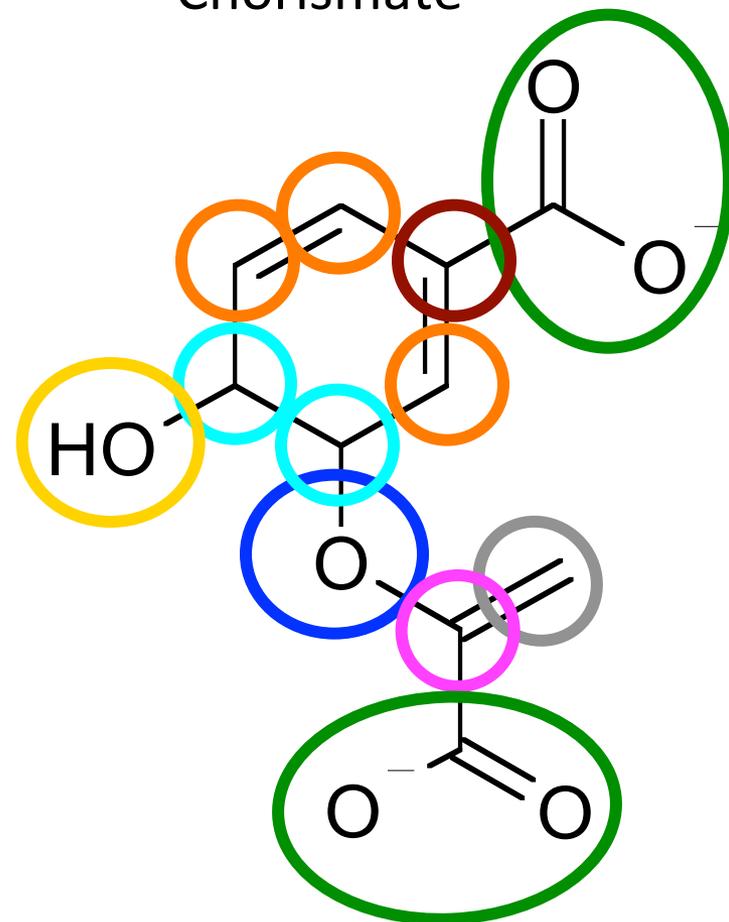
Group	$\Delta_{gr} G'^{\circ}$	n
[-COO-]	-72.0	2
[-OH]	-32.0	1
[-CH=] _{ring}	11.1	3
[-O-]	-22.5	1
[>C=]	5.0	1
[>CH-] _{ring}	-2.2	2
[>C=] _{ring}	8.2	1
[=CH ₂]	18.4	1

Estimated $\Delta_f G'^{\circ}$: -161.6 kcal/mol

•Equation for estimating $\Delta_r G'^{\circ}$

$$\Delta_r G'^{\circ} = \sum_{i=0}^m \nu_i (\Delta_f G'_i)^{\circ}$$

Chorismate



Biochemistry Database for Model Reconstruction

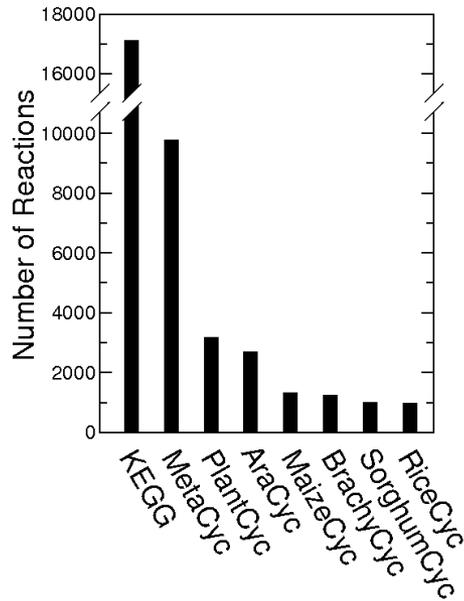
- The PlantSEED biochemistry database is constructed by an integration of biochemistry from numerous source, but the bulk of the data is from AraCyc, PMN, Gramene, KEGG



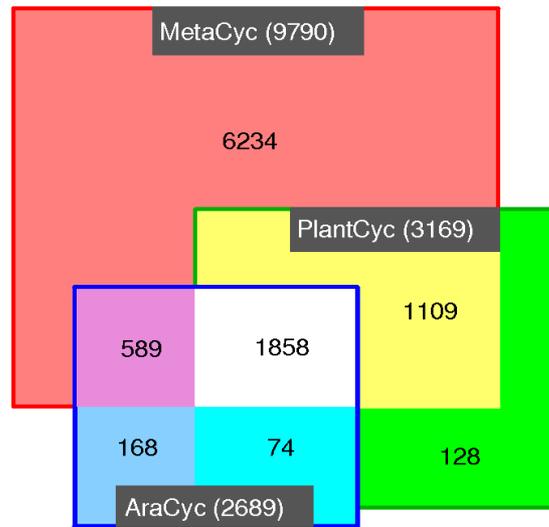
- 1.) Combine data from all sources in a common format
- 2.) Merge identical compounds from multiple resources based on inchi-key (structure) or exact name match
- 3.) Merge reactions with identical stoichiometry
- 4.) Compute charge for each compound from structure of pH7 using MarvinBeans
- 5.) Compute compound and reaction delta G using group contribution method
- 6.) Predict reaction directionality based on delta G
- 7.) Map reactions to exemplar genes and the SEED functional annotation ontology
- 8.) Now moving curation effort to github:
<https://github.com/ModelSEED/ModelSEEDDatabase>

Results of Biochemistry Merging

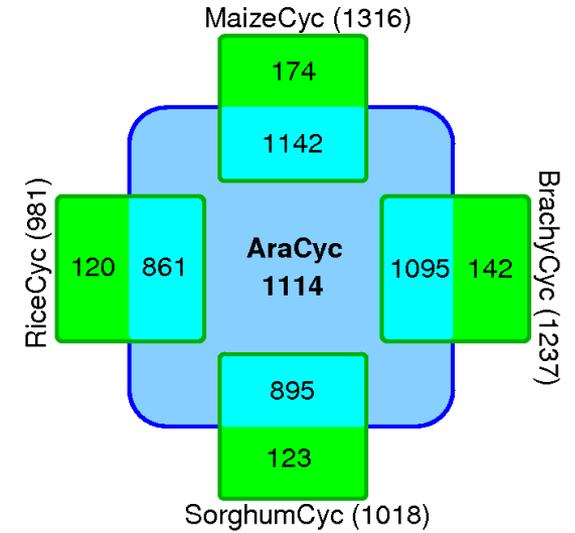
A. Number of Reactions



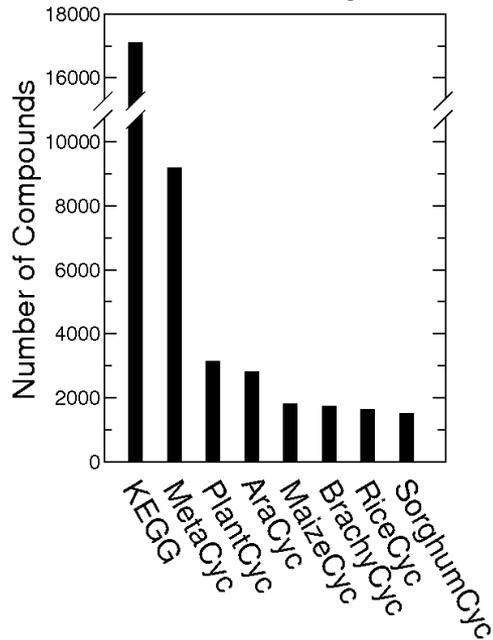
B. Shared Reactions



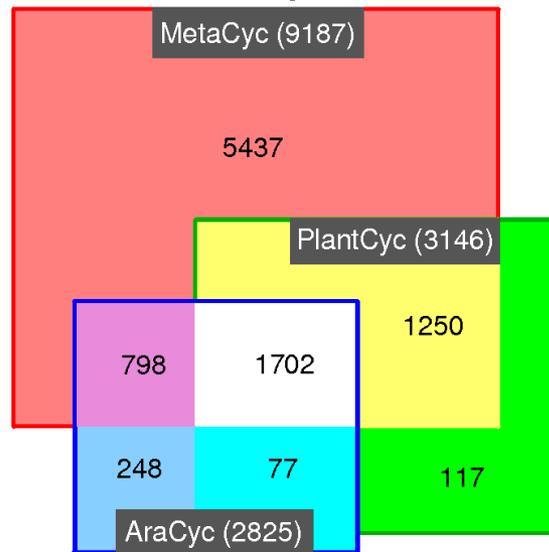
C. Shared Reactions



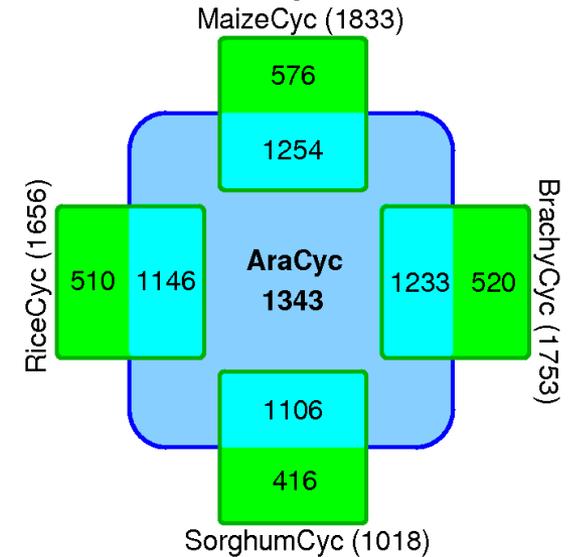
D. Number of Compounds



E. Shared Compounds

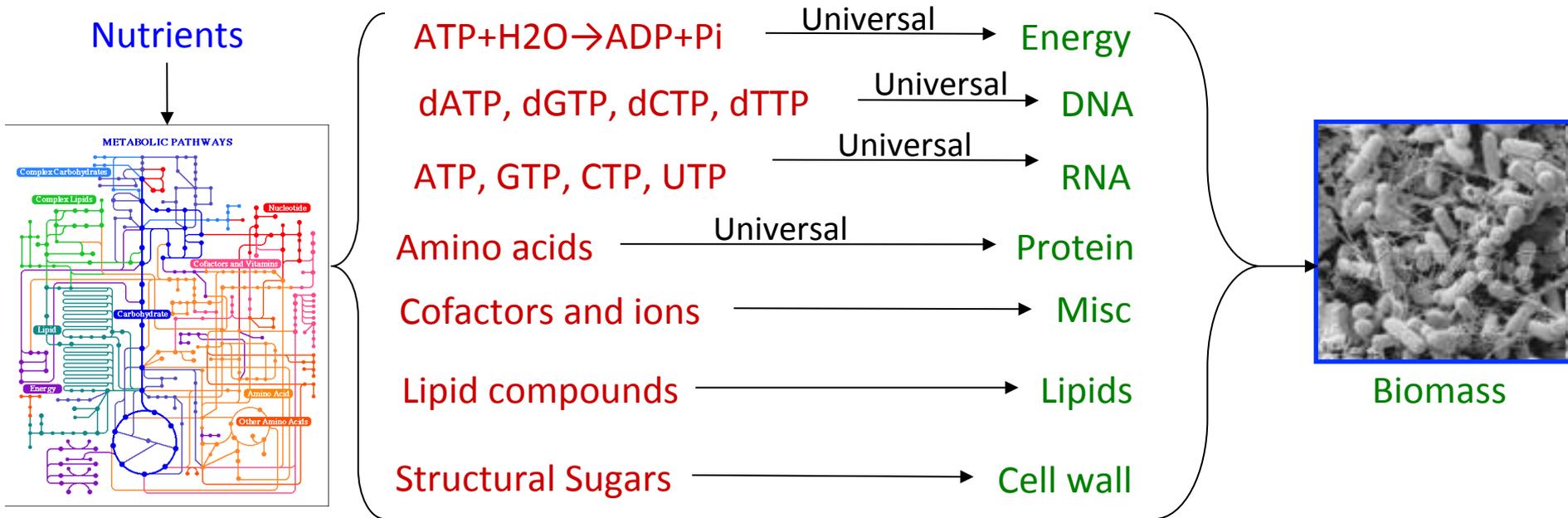


F. Shared Compounds



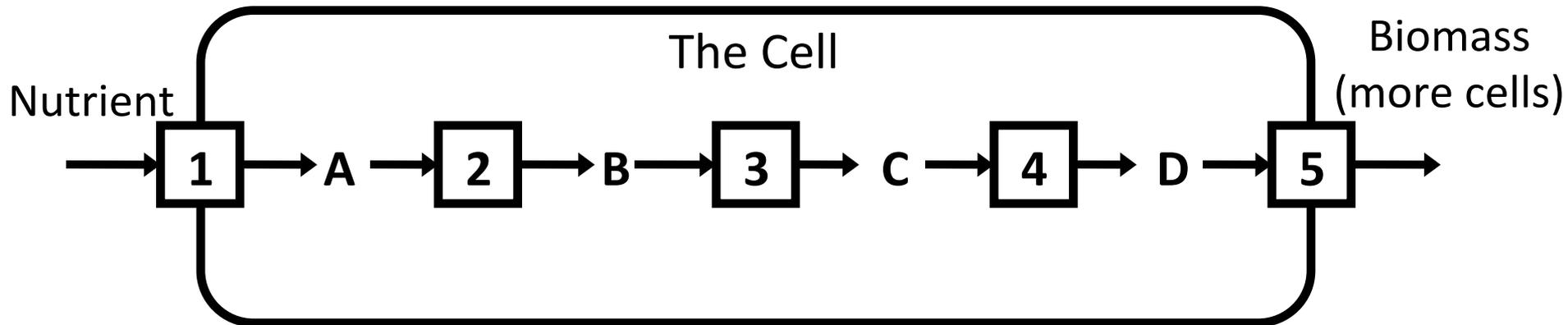
Building Biomass Composition Reaction for Plants

We curated a biomass composition reaction for our plant models



Constructed based on curation of literature on biomass composition of leaf cells in plants

Application of Linear Optimization to Biology: Flux Balance Analysis



Assuming Steady State:

No internal metabolite is
allowed to accumulate

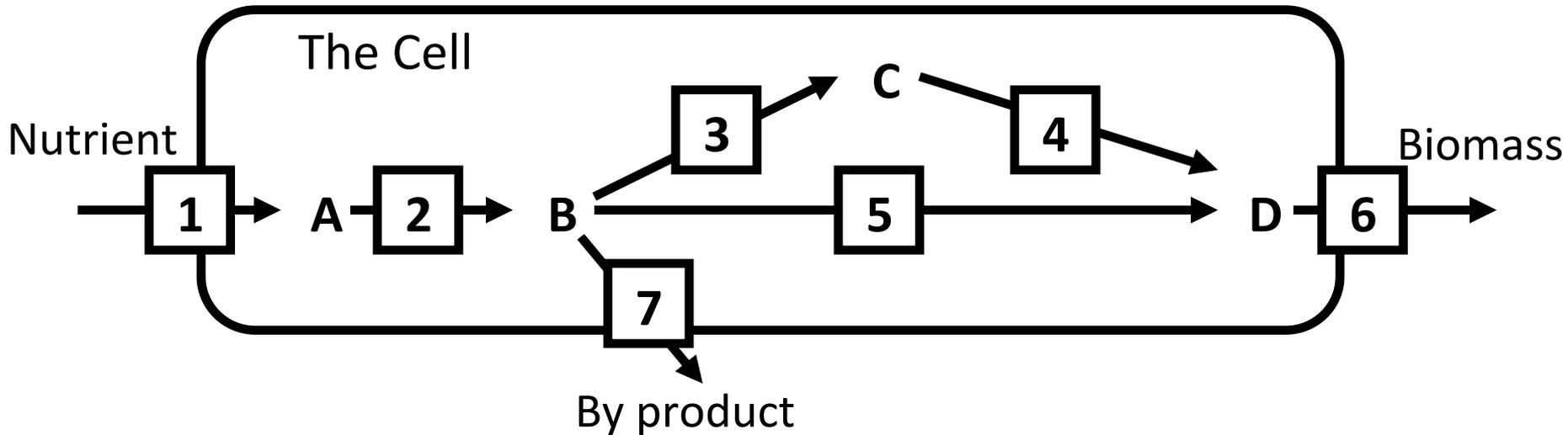
Thus, reaction rates are constrained
by mass balances

At Steady State:

$$v_1 = v_2 = v_3 = v_4 = v_5$$

If we measure one reaction rate,
we can calculate every other
rate

Flux Balance Analysis



Assuming Steady State:

No internal metabolite is allowed to accumulate

Thus, reaction rates are constrained by mass balances

For example:

At Steady State:

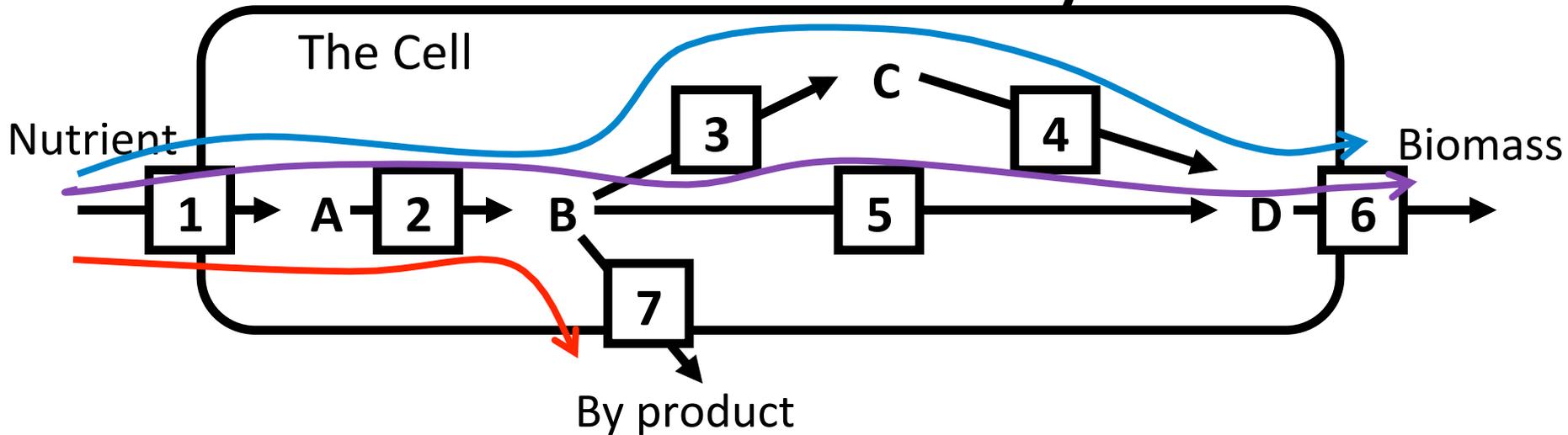
$$v_1 = v_2$$

$$v_2 = v_3 + v_5 + v_7$$

$$v_3 = v_4$$

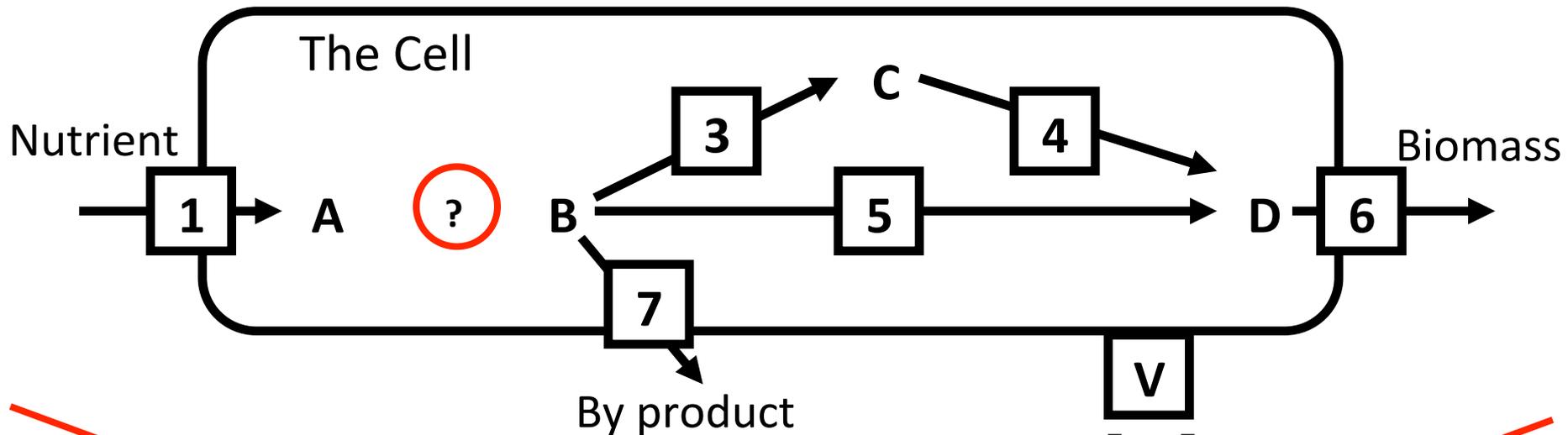
$$v_4 + v_5 = v_6$$

Flux Balance Analysis



- Now we have an underdetermined system
- There are three completely independent and equally valid solutions, with infinite combinations
- How can we tell which one is correct?
 - More measurements
 - Predict the correct solution using an objective function (e.g. growth)

Gapfilling Annotations to Identify Missing Functions



By product

	1	2	3	4	5	6	7
A	1	-1	0	0	0	0	0
B	0	1	-1	0	-1	0	-1
C	0	0	1	-1	0	0	0
D	0	0	0	1	1	-1	0

$$\begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Gapfilling Annotations to Identify Missing Functions

Objective:

$$\text{Minimize } \sum_{i=1}^r (p_{lump} + p_{thermo} + p_{SEED} + p_{Path}) z_{\text{not in model},i}$$

Penalty for lumped reactions

Penalty for thermodynamically infeasible reactions

Penalty for reactions not associated with SEED roles

Bonus for reactions completing pathways

Subject to:

Mass balance constraints:

Compounds in model

Compounds not in model

$$\begin{bmatrix} N_{core} & N_{db} \\ \mathbf{0} & N_{db} \end{bmatrix} \cdot \begin{bmatrix} v_{core} \\ v_{db} \end{bmatrix} = \mathbf{0}$$

Reactions in model
Reactions not in model

Use variable constraints:

$$0 \leq v_{i,forward} \leq v_{Max} z_{i,forward}$$

$$0 \leq v_{i,reverse} \leq v_{Max} z_{i,reverse}$$

Forcing at least one inactive reaction to be active:

$$v_{biomass} > 0$$

