Bioinformatics: Network Analysis

Flux Balance Analysis and Metabolic Control Analysis

COMP 572 (BIOS 572 / BIOE 564) - Fall 2013

Luay Nakhleh, Rice University

Flux Balance Analysis (FBA)

- Flux balance analysis (FBA), an optimality-base method for flux prediction, is one of the most popular modeling approaches for metabolic systems.

- Flux optimization methods do not describe *how* a certain flux distribution is realized (by kinetics or enzyme regulation), but *which* flux distribution is optimal for the cell; e.g., providing the highest rate of biomass production at a limited inflow of external nutrients.

- This allows us to predict flux distributions without the need for a kinetic description.

Flux Balance Analysis (FBA)

- FBA investigates the theoretical capabilities and modes of metabolism by imposing a number of constraints on the metabolic flux distributions:

- The assumption of a steady state: $\mathbf{S}\times\mathbf{v}=\mathbf{0}$.

- Thermodynamics constraints: $a_i \leq v_i \leq b_i$.

- An optimality assumption: the flux distribution has to maximize (or, minimize) an objective function $f(v)$

$$f(v) = \sum_{i=1}^{r} c_i v_i$$

Geometric Interpretation of FBA

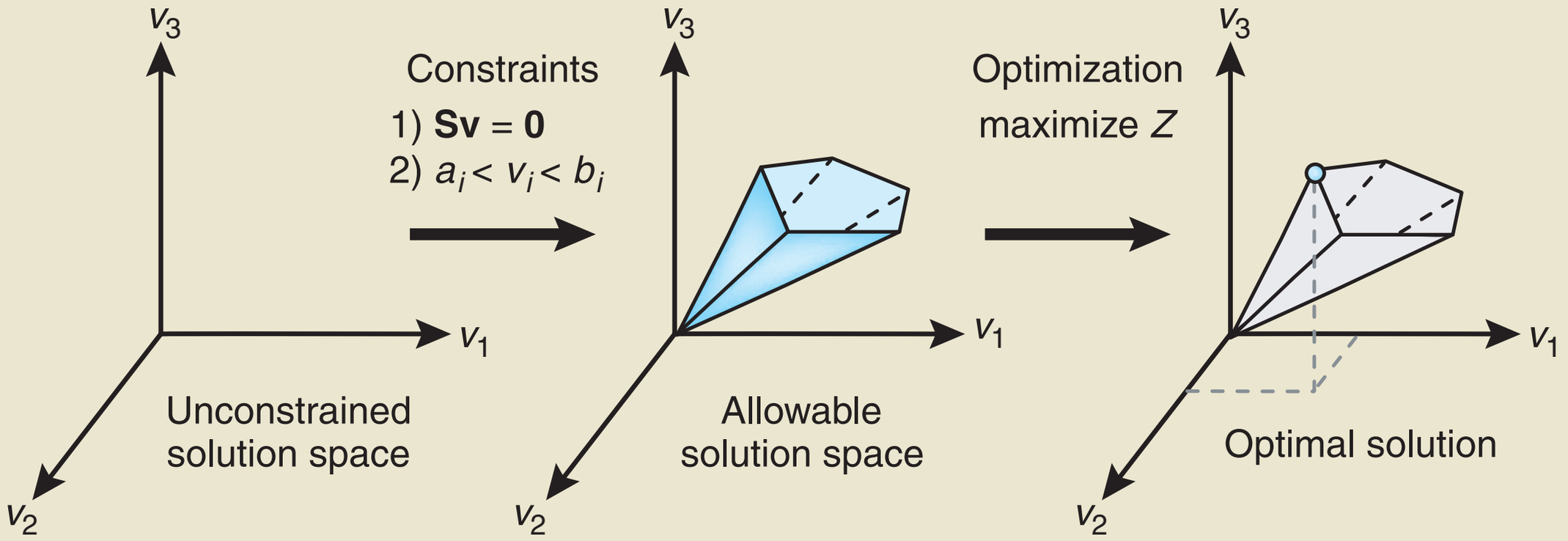
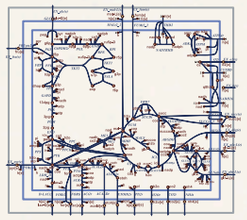


Figure 1 The conceptual basis of constraint-based modeling. With no constraints, the flux distribution of a biological network may lie at any point in a solution space. When mass balance constraints imposed by the stoichiometric matrix **S** (labeled 1) and capacity constraints imposed by the lower and upper bounds (a_i and b_i) (labeled 2) are applied to a network, it defines an allowable solution space. The network may acquire any flux distribution within this space, but points outside this space are denied by the constraints. Through optimization of an objective function, FBA can identify a single optimal flux distribution that lies on the edge of the allowable solution space.

[Source: "What is flux balance analysis?", Nat Biotech.]

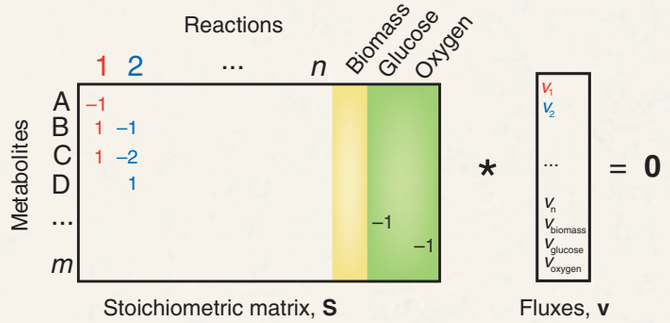
Formulation of an FBA Problem

a Genome-scale metabolic reconstruction



A ↔ B + C — Reaction 1
B + 2C → D — Reaction 2
...
Reaction *n*

b Mathematically represent metabolic reactions and constraints



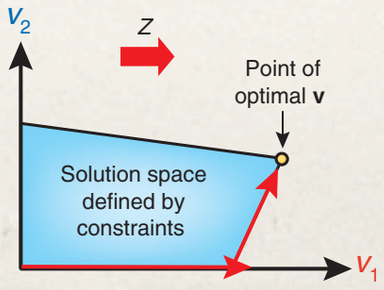
c Mass balance defines a system of linear equations

$$-v_1 + \ldots = 0$$
$$v_1 - v_2 + \ldots = 0$$
$$v_1 - 2v_2 + \ldots = 0$$
$$v_2 + \ldots = 0$$
etc.

d Define objective function ($Z = c_1^* v_1 + c_2^* v_2 \ldots$)

To predict growth, $Z = v_{biomass}$

e Calculate fluxes that maximize Z



What to Optimize?

- Minimize ATP production: the most energy-efficient state
- Minimize nutrient intake: the fittest state under nutrient shortage
- Maximize metabolite production: the biochemical production capabilities of certain desirable metabolites such as lysine, phenylalanine, etc.
- Maximize biomass formation: maximal growth rate
- ...

Producing Biomass

- Growth can be defined in terms of the biosynthetic requirements to make a cell.

- These requirements are based on literature values of experimentally determined biomass composition.

- Thus, biomass generation is defined as a linked set of reaction fluxes draining intermediate metabolites in the appropriate ratios and represented as an objective function Z.

Biomass Formation in *E. coli*

- The requirements for making 1g of *E. coli* biomass from key cofactors and biosynthetic precursors have been documented.

- This means that for *E. coli* to grow, all these components must be provided in the appropriate relative amounts.

- Key biosynthetic precursors are used to make all the constituents of *E. coli* biomass. Their relative requirements to make 1g of *E. coli* biomass are:

$$Z_{precursors} = +0.205V_{g6P}+0.071V_{F6P}+0.898V_{R5P} +0.361V_{E4P}+0.129V_{T3P}+1.496V_{3PG} +0.519V_{PEP}+2.833V_{PYR}+3.748V_{AcCoA} +1.787V_{OAA}+1.079V_{\alpha KG}$$

Biomass Formation in *E. coli*

- In addition to precursors, cofactors are needed to drive the process.
- The cofactors requirement to synthesize the monomers from the precursors (amino acids, fatty acids, nucleic acids) and to polymerize them into macromolecules is

$$Z_{cofactors} = 42.703V_{ATP}\text{-}3.547V_{NADH}\text{+}18.22V_{NADPH}$$

Biomass Formation in *E. coli*

- The mass and cofactor requirements to generate *E. coli* biomass are:

$$Z_{biomass} = Z_{precursors} + Z_{cofactors}$$

Resources for FBA

- The BIGG database: http://bigg.ucsd.edu/
- The COBRA toolbox: http://opencobra.sourceforge.net/
- FASIMU: http://www.bioinformatics.org/fasimu/

Applications of FBA

Modular Epistasis in Yeast Metabolism

- Genes can be classified by categories related to functions of the cell (e.g., translation, energy metabolism, mitosis, etc.) based on textbook knowledge.

- Can we infer functional associations directly from deletion experiments?

- If two gene products can compensate for each other's loss, then deleting both of them will have a much stronger impact on cell fitness than one would expect from their single deletions.

Modular Epistasis in Yeast Metabolism

- On the other hand, if two gene products are essential parts of the same pathway, a single deletion would already shut down the pathway and a double deletion would not have any further effect.

- Accordingly, we may try to infer functional relationships among the gene products by comparing the fitness losses caused by combined gene deletions.

Modular Epistasis in Yeast Metabolism

- *Epistasis* describes how the fitness loss due to a gene mutation depends on the presence of other genes.

- It can be quantified by comparing the fitness of a wild type organism, e.g., the growth rate of a bacteria culture, to the fitness of single and double deletion mutants.

- A single gene deletion (for gene i) will decrease the fitness (e.g., the growth rate) from a value f_{wt} to a value f_i, leading to a growth defect $w_i=f_i/f_{wt}$ (≤ 1).

Modular Epistasis in Yeast Metabolism

- For a double deletion of unrelated genes i and j, we may expect a multiplicative effect $w_{ij}=w_i w_j$ (no epistasis).

- If the double deletion is more severe ($w_{ij}<w_i w_j$), we call the epistasis *aggravating*.

- If the double deletion is less severe ($w_{ij}>w_i w_j$), we call the epistasis *buffering*.

- Both cases of aggravating and buffering epistasis indicate functional associations between the genes in question.

Modular Epistasis in Yeast Metabolism

- Segre *et al.* (Nature Genetics, 37(1):77-83, 2005) recently used FBA to predict growth rates of the yeast *S. cerevisiae* and to calculate the epistatic effects between all metabolic genes.

- The model predicted relative growth defects of all single and double deletion mutants, from which they computed an epistasis measure for each pair of genes:

$$\hat{\epsilon}_{ij} = \frac{w_{ij} - w_i w_j}{|\hat{w}_{ij} - w_i w_j|}$$

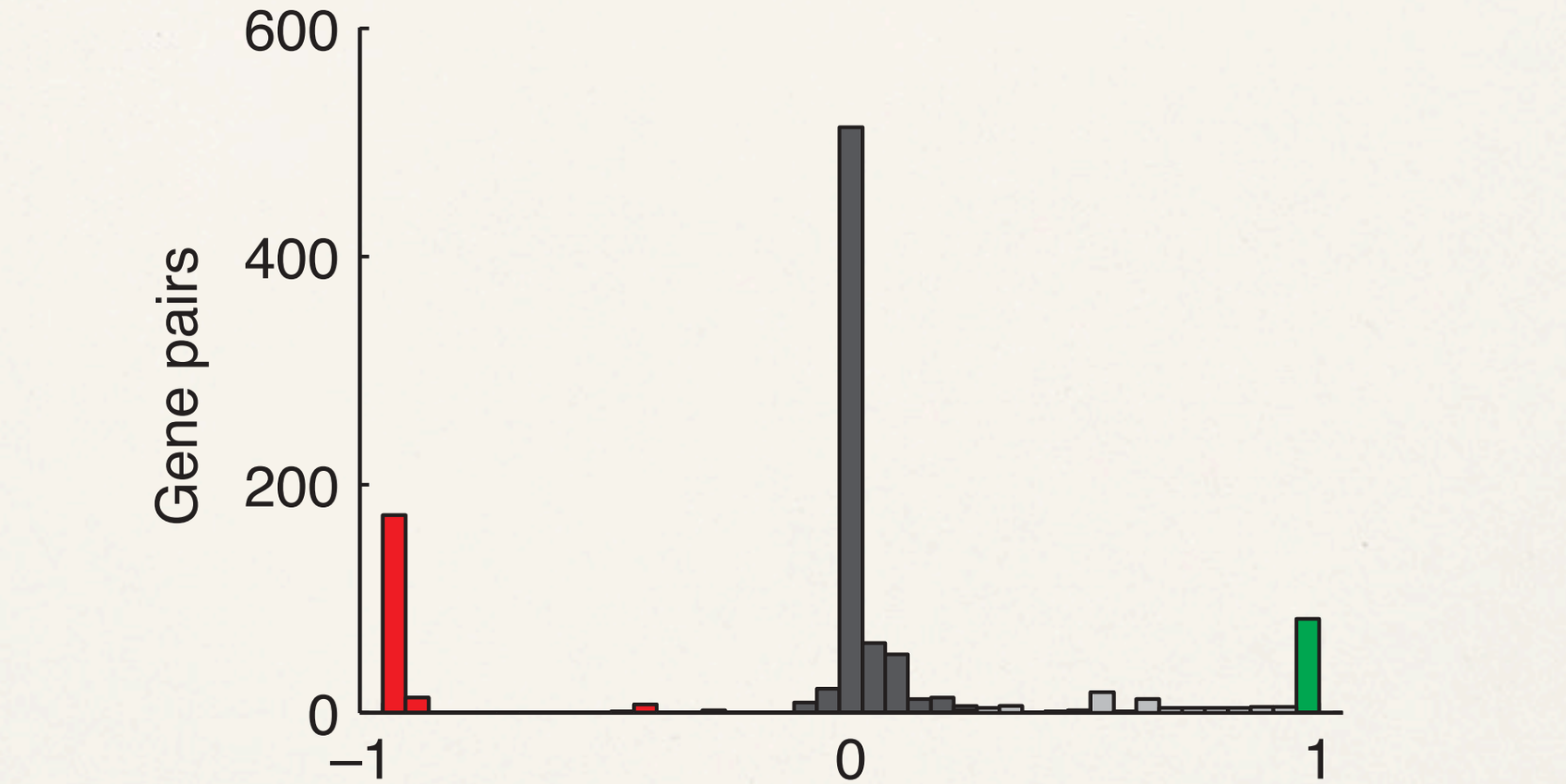
extreme buffering

$$\hat{w}_{ij} = \min\{w_i, w_j\}$$

extreme aggravation

$$\hat{w}_{ij} = 0$$

Modular Epistasis in Yeast Metabolism



Modular Epistasis in Yeast Metabolism

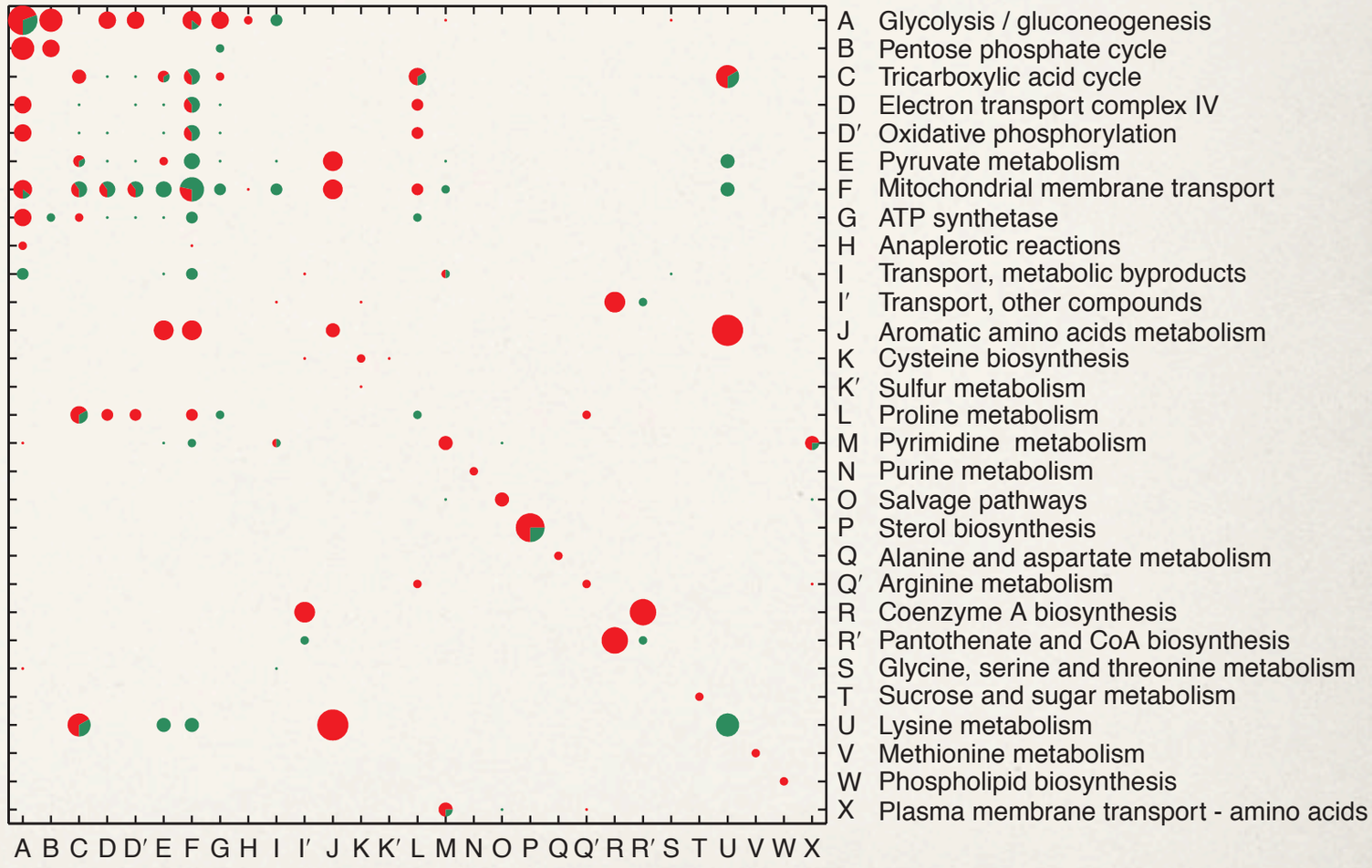


Figure 2 Epistatic interactions between genes classified by functional annotation groups tend to be of a single sign (*i.e.*, monochromatic). (**a**) Representation of the number of buffering and aggravating interactions within and between groups of genes defined by common preassigned annotation from the FBA model. The radii of the pies represent the total number of interactions (ranging logarithmically from 1 in the smallest pies to 35 in the largest). The red and green pie slices reflect the numbers of aggravating and buffering interactions, respectively. Monochromatic interactions, represented by whole green or red pies, are much more common than would be expected by chance. (**b**) Sensitivity

The Interplay Between Metabolism and Gene Regulation

- Recently, Shlomi *et al.* (MSB 3:101) conducted a computational analysis of the interplay between metabolism and transcriptional regulation in *E. coli*.

- To enable such an analysis, the authors proposed a new method, *steady-state regulatory flux balance analysis* (SR-FBA), for predicting gene expression and metabolic fluxes in a large-scale integrated metabolic-regulatory model.

An Integrated Network

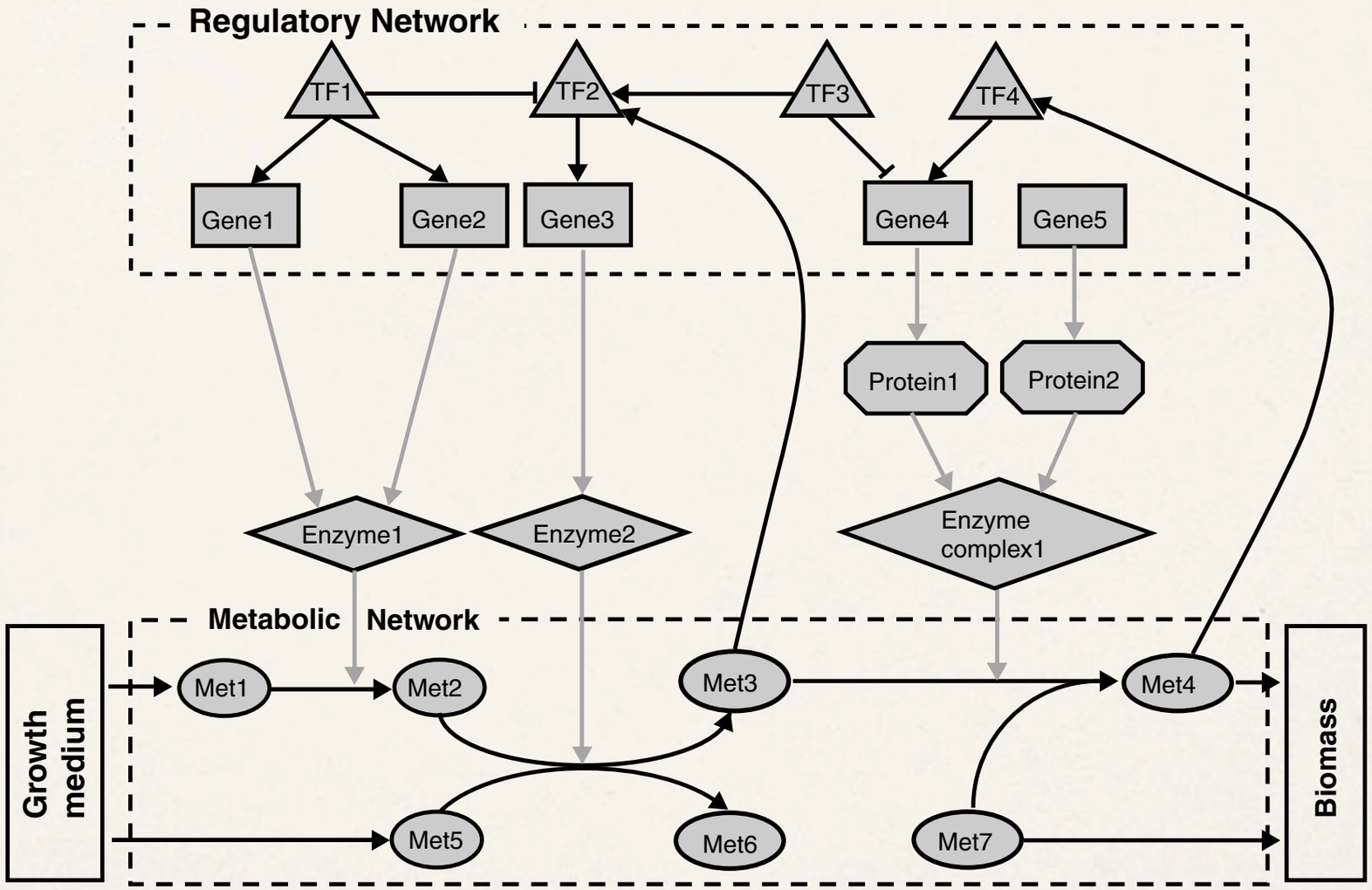


Figure 1 A schematic representation of an integrated metabolic and regulatory network. The regulatory network component consists of a set of interactions between TFs and other TFs and genes. The metabolic network component consists of a set of biochemical reactions between metabolites, with metabolites available from growth medium as input, and a pseudo-metabolite representing biomass production as output. The regulatory component affects the metabolic component through the expression of proteins that catalyze the biochemical reactions (downward pointing arrows). The metabolic component affects the regulatory component via the activation or inhibition of TF expression via the presence of specific metabolites (upwards arrows).

The SR-FBA Method

- In addition to the metabolic constraints, there are
  - *Regulatory constraints*: e.g., `G1=NOT(TF1) AND TF2' (gene G1 is expressed if and only if TF1 is not expressed and TF2 is expressed).
  - *Genes-to-reactions mapping constraints*: e.g., `R1=(P1 AND P2) OR (P3 AND P4)' (reaction R1 is catalyzed by either the enzyme complex P1-P2 or by the enzyme complex P3-P4).
  - *Reaction enzyme state constraints*: The absence of a catalyzing enzyme for a specific reaction should constrain the flux through this reaction to zero.
  - *Reaction predicates constraints*: The reaction predicate b_i represents a rule in the form `FLUX(j)>c', where c$\in$**R**.

The Interplay Between the Two Networks

- The combined functional state of the entire system in a given constant environment, referred to as *metabolic-regulatory steady state* (MRS), is described by a pair of consistent metabolic and regulatory steady states, which satisfy both the metabolic and regulatory constraints.

- The SR-FBA method identifies an MRS for the integrated metabolic-regulatory model.

The Interplay Between the Two Networks

- Each transcription factor (TF) and TF-regulated gene (i.e., genes associated with a regulatory role in the model) can be either in an *expressed* or *non-expressed state*, if it is expressed or non-expressed, respectively, in all alternative MRS solutions attainable within a given growth medium.

- In both cases, the genes are considered to have a *determined expression state*.

- Alternatively, the gene is considered to have an *undetermined expression state* if it is expressed in some of the alternative MRS solutions but non-expressed in others in the same medium.

The Interplay Between the Two Networks

- In parallel, each gene is characterized by its *flux activity state*, which reflects the existence of non-zero flux through one of the metabolic enzymatic reactions that it encodes.

- It can have a *determined* or *undetermined activity state*.

- Obviously, the expression and activity states are inter-dependent as a gene cannot be metabolically active if it is not expressed.

The Interplay Between the Two Networks

- Using the SR-FBA method, the authors quantified the effect of transcriptional regulation on metabolism by measuring the fraction of genes whose flux activity is determined by the integrated model but not by the metabolic component alone.

The Interplay Between the Two Networks

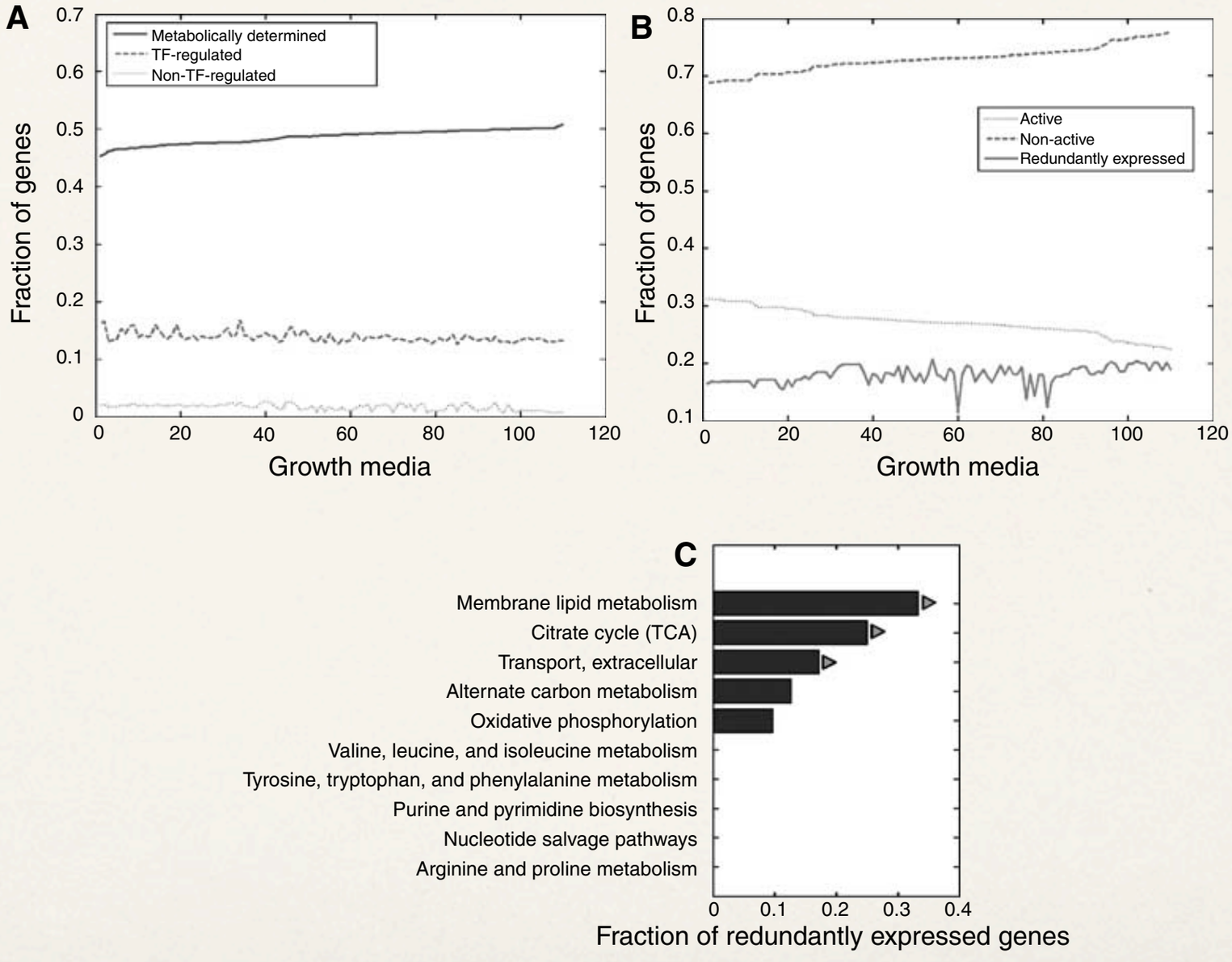


Figure 2 (**A**) The fraction of metabolic-determined genes and the fraction of regulatory-determined genes across different growth media. For the latter, we show the fraction of genes that are TF-regulated and the fraction of non-TF-regulated genes. (**B**) The fraction of genes that are metabolically determined to be active, inactive and redundantly expressed, from the set of metabolically determined genes. (**C**) The distribution of redundantly expressed genes within various functional metabolic categories. Triangles represent a statistically significant enrichment.

Metabolic Control Analysis

Metabolic Control Analysis (MCA)

- MCA characterizes the effects of small perturbations in a metabolic pathway that operates at a steady state.

- MCA was conceived to replace the notion that every pathway has one rate-limiting step, which is a slow reaction that by itself is credited with determining the magnitude of flux through the pathway.

- In MCA, this concept of rate-limiting step was supplanted with the concept of shared control, which posits that every step in a pathway contributes, to some degree, to the control of the steady-state flux.

Metabolic Control Analysis (MCA)

- MCA formalizes this concept with quantities called control coefficients and elasticities, and with mathematical relationships that permit certain insights into the control structure of the pathway.

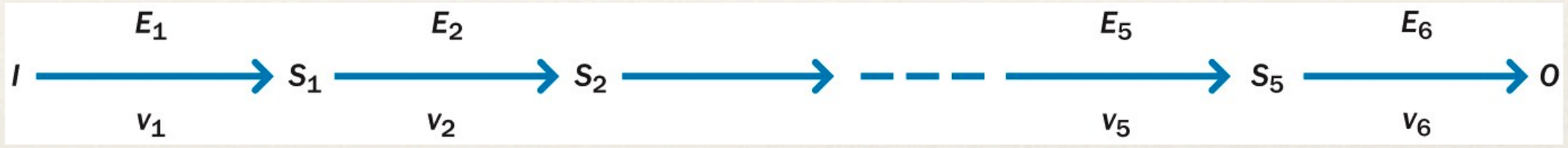
Metabolic Control Analysis (MCA)

- The pathway is assumed to operate at steady state
- All perturbations are required to be infinitesimally small
- Control coefficients quantify the effect of small changes in parameters on features of the system as a whole (recall sensitivity analysis)

Metabolic Control Analysis (MCA)

- The control coefficients measure the relative change in a flux (the flux control coefficient) or substrate concentration (the concentration control coefficient) at steady state that results from a relative or percent change in a key parameter, such as an enzyme activity.

- It is assumed that each v_i is directly proportional to the corresponding E_i, so that the control coefficients may be equivalently expressed in terms of v_i or E_i.



Metabolic Control Analysis (MCA)

- Recall: the steady state corresponds to

$$dS_i/dt = 0 \qquad dI/dt = 0 \qquad dO/dt = 0$$

- In this situation, all reaction rates must have the same magnitude as the overall flux J, namely $v_1 = v_2 = \ldots = v_6 = J$.

Metabolic Control Analysis (MCA)

The flux control coefficient:

$$C^J_{v_i} = \frac{\partial J}{J} \Big/ \frac{\partial v_i}{v_i} = \frac{v_i}{J}\frac{\partial J}{\partial v_i} = \frac{\partial \ln J}{\partial \ln v_i}$$

change in flux

change in enzyme activity

The concentration control coefficient:

$$C^{S_k}_{v_i} = \frac{v_i}{S_k}\frac{\partial S_k}{\partial v_i} = \frac{\partial \ln S_k}{\partial \ln v_i}$$

Metabolic Control Analysis (MCA)

- The control exerted by a given enzyme either on a flux or on a metabolite concentration can be distinctly different, and it is indeed possible that a concentration is strongly affected but the pathway flux is not.

- Therefore, the distinction between the two types of control coefficients is important.

Metabolic Control Analysis (MCA)

- Further, the distinction is also pertinent for practical considerations, for instance in biotechnology, where gene and enzyme manipulations are often targeted either toward an increased flux or toward an increased concentration of a desirable compound.

Metabolic Control Analysis (MCA)

- An elasticity coefficient measures how a reaction rate v_i changes in response to a perturbation in a metabolite S_k or some other parameter.

- With respect to S_k, it is defined as

$$\varepsilon_{S_k}^{v_i} = \frac{S_k}{v_i}\frac{\partial v_i}{\partial S_k} = \frac{\partial \ln v_i}{\partial \ln S_k}$$

Metabolic Control Analysis (MCA)

- The elasticity with respect to the Michaelis constant K_k of an enzyme for the substrate S_k is the complement of the metabolite elasticity:

$$\varepsilon_{S_k}^{v_i} = -\varepsilon_{K_k}^{v_i}$$

Metabolic Control Analysis (MCA)

- Because only one metabolite (or parameter) and one reaction are involved in this definition, but not the entire pathway, each elasticity is a local property, which can in principle be measured *in vitro*.

Metabolic Control Analysis (MCA)

- The main insights provided by MCA are gained from relationships among the control and elasticity coefficients.

- We may be interested in the overall change in the flux *J*, which is mathematically determined by the sum of responses to possible changes in all six enzymes:

Metabolic Control Analysis (MCA)

- It has been shown that all effects, in the form of flux control coefficients, sum to 1, and that all concentration control coefficients with respect to a given substrate *S* sum to 0:

$$\sum_{i=1}^{n+1} C_{v_i}^{J} = 1 \qquad \sum_{i=1}^{n+1} C_{v_i}^{S} = 0$$

Metabolic Control Analysis (MCA)

$$\sum_{i=1}^{n+1} C_{v_i}^{J} = 1$$

- Implications:

- The control of metabolic flux is shared by all reactions in the system; this global aspect identifies the control coefficients as systemic properties.

- If a single reaction is altered and its contributions to the control of flux changes, the effect is compensated by changes in flux control by the remaining reactions.

Metabolic Control Analysis (MCA)

$$\sum_{i=1}^{n+1} C_{v_i}^{J} = 1$$

- Use:
  - One measures in the laboratory the effects of changes in various reactions of a pathway, and if the sum of flux control coefficients is below 1, then one knows that one or more contributions to the control structure are missing.

Metabolic Control Analysis (MCA)

- A second type of insight comes from connectivity relationships, which establish constraints between control coefficients and elasticities.

- These relationships have been used to characterize the close connection between the kinetic features of individual reactions and the overall responses of a pathway to perturbations.

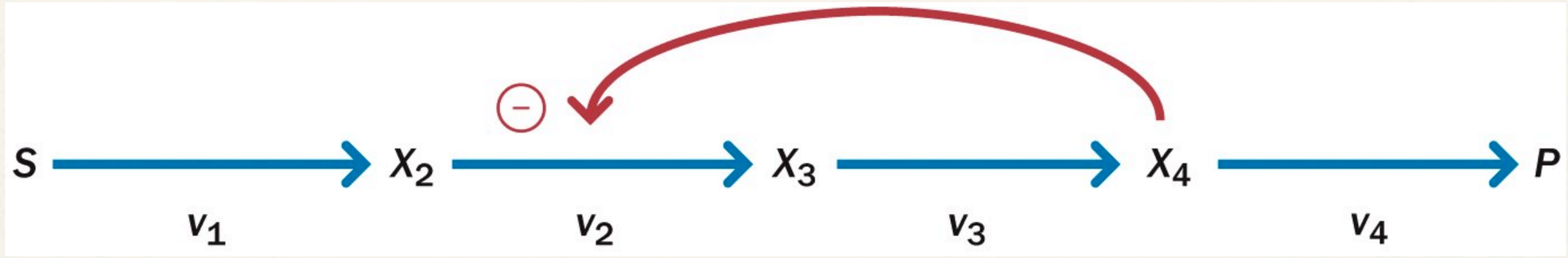
Metabolic Control Analysis (MCA)

- The most important of these connectivity relationships is

$$\sum_{i=1}^{n+1} C_{v_i}^{J} \varepsilon_{S_k}^{v_i} = 0$$

- Let us consider an example of using this relationship.

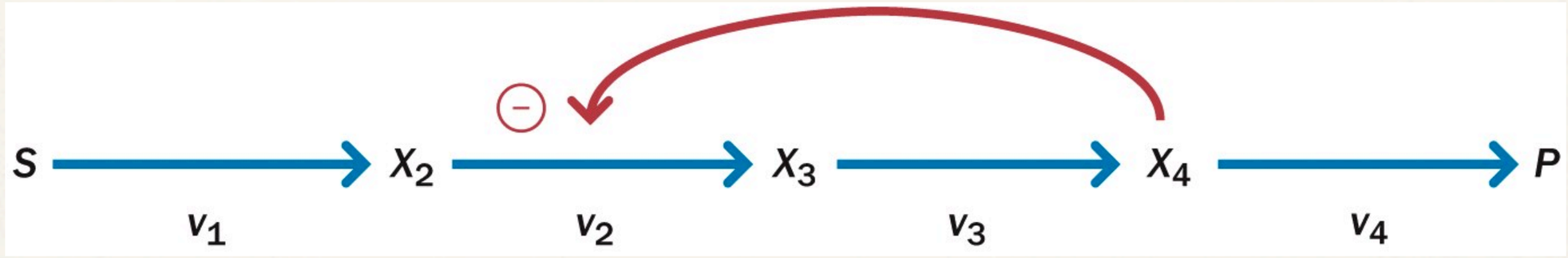
Metabolic Control Analysis (MCA)



$$\varepsilon_{X_2}^{v_1} = -0.9 \quad \varepsilon_{X_2}^{v_2} = 0.5 \quad \varepsilon_{X_3}^{v_2} = -0.2 \quad \varepsilon_{X_3}^{v_3} = 0.7 \quad \varepsilon_{X_4}^{v_2} = -1 \quad \varepsilon_{X_4}^{v_4} = 0.9$$

all other elasticities are 0

Metabolic Control Analysis (MCA)



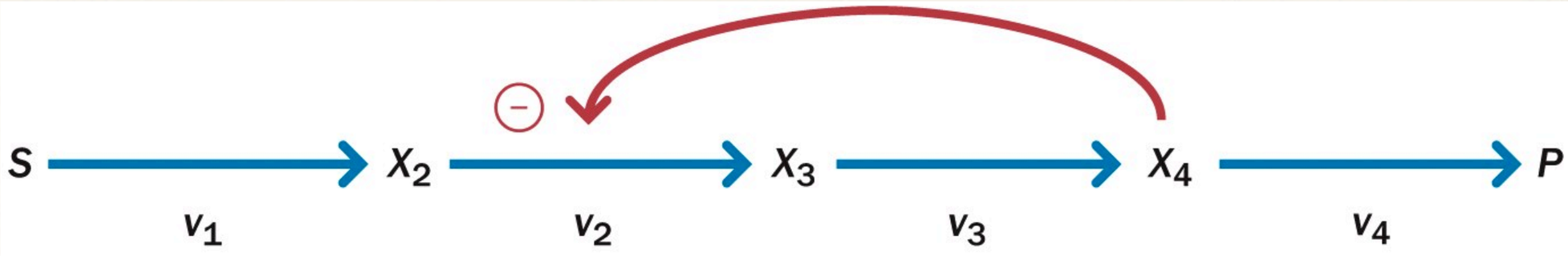
$$\varepsilon_{X_2}^{v_1} = -0.9 \quad \varepsilon_{X_2}^{v_2} = 0.5 \quad \varepsilon_{X_3}^{v_2} = -0.2 \quad \varepsilon_{X_3}^{v_3} = 0.7 \quad \varepsilon_{X_4}^{v_2} = -1 \quad \varepsilon_{X_4}^{v_4} = 0.9$$

all other elasticities are 0

using the connectivity theorem

$$C_{v_1}^J = 0.19 \quad C_{v_2}^J = 0.34 \quad C_{v_3}^J = 0.10 \quad C_{v_4}^J = 0.38$$

Metabolic Control Analysis (MCA)



- Suppose reaction v_2 is a bottleneck and that a goal of the analysis is to propose strategies for increasing the flux through the pathway.

- Let's consider two strategies:

1. modify the enzyme E_2 so that it is less affected by the inhibition exerted by X_4

2. increase the activity of E_4 to reduce the inhibition of v_2

Metabolic Control Analysis (MCA)

- Strategy 1: Suppose we could alter the effect of X_4 by changing the binding constant K^2_4 of the enzyme E_2 by $p\%$

- K^2_4 has an effect on v_2, which is quantified by the elasticity (of the Michaelis constant formula).

- Further, the effect of changes in v_2 on the pathway flux J is given by the flux control coefficient.

Metabolic Control Analysis (MCA)

- Putting it all together:

$$C^{J}_{v_2}\varepsilon^{v_2}_{K_4^2} = -\frac{\partial \ln J}{\partial \ln v_2}\frac{\partial \ln v_2}{\partial \ln K_4^2} = -\frac{\partial \ln J}{\partial \ln K_4^2}$$

- Rearranging:

$$\partial \ln J = -C^{J}_{v_2}\varepsilon^{v_2}_{K_4^2}\partial \ln K_4^2 \approx -C^{J}_{v_2}\varepsilon^{v_2}_{K_4^2}p\%$$

- Substituting numerical values:

a relative change in flux J of -0.34 per percent change in the binding constant of enzyme E2.

⇒to increase J, the constant must be decreased!

Metabolic Control Analysis (MCA)

- Strategy 2:

$$\partial \ln J = -C^J_{v_4} \partial \ln v_4 \approx C^J_{v_4} q\% = 0.38 q\%$$

- The effect is about 10% stronger than in the previous strategy.

Acknowledgments

- “A First Course in Systems Biology,” by E.O. Voit.