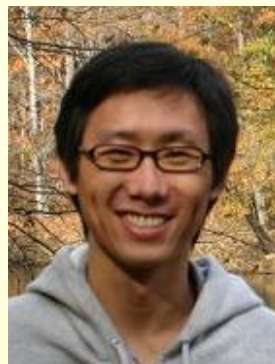# Control and Optimization of Lignin Biosynthesis in Plant Cell Walls
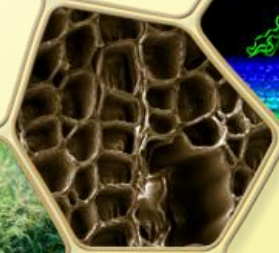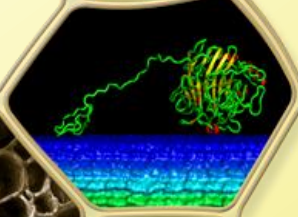
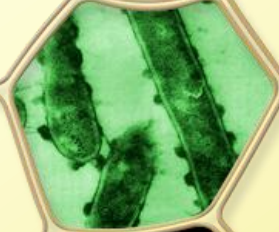**Yun Lee and Eberhard O. Voit**

Wallace H. Coulter Department of Biomedical Engineering
Georgia Institute of Technology
Atlanta, GA 30332
www.bst.bme.gatech.edu

ICM Satellite Meeting
Hyderabad, August 16-18, 2010

$$V_i = \gamma_i \prod_{j=1}^{n+m} X_j^{f_{ij}}$$

BESC
BioEnergy Science Center

OAK RIDGE
National Laboratory

Preamble

**In College I was told that
Biology was too complicated
to use Math.**

**We have learned by now that
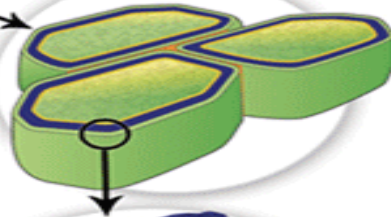Biology is too complicated
not to use Math.**

Overview

- **Context: Sustainable biofuel production**

- **Plant cell wall composition; role of lignin; monolignols**

- **Recalcitrance directly related to lignin content and composition (S/G/H)**

- **Study lignin biosynthesis in:**
  - **Poplar**
  - **Alfalfa**
  - **Switchgrass**

- **Methods**
  - **Stoichiometric analysis (static network analysis)**
  - **Optimization analysis (FBA, MOMA; constrained network analysis)**
  - **Biochemical Systems Theory (BST; fully kinetic, dynamic analysis)**
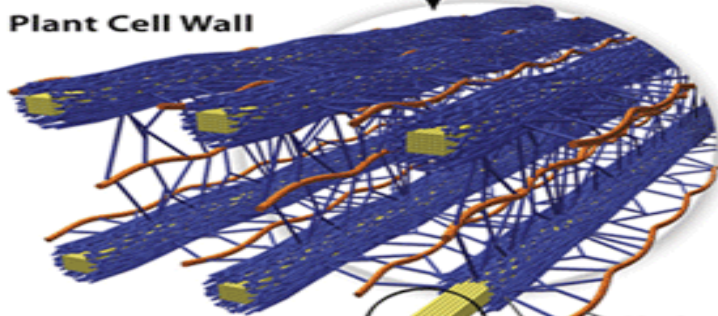
- **Results**

Biofuel Production

Switchgrass   Poplar
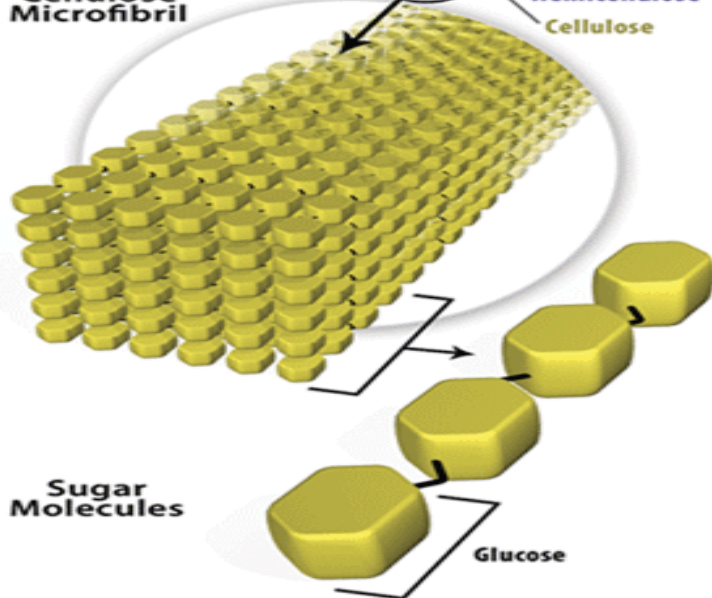
Fermentation
→ **Ethanol**
Bacteria, Yeast

Bioenergy.msu.edu; Virginia Tech news; jcwinnie.biz/wordpress/?p=1934
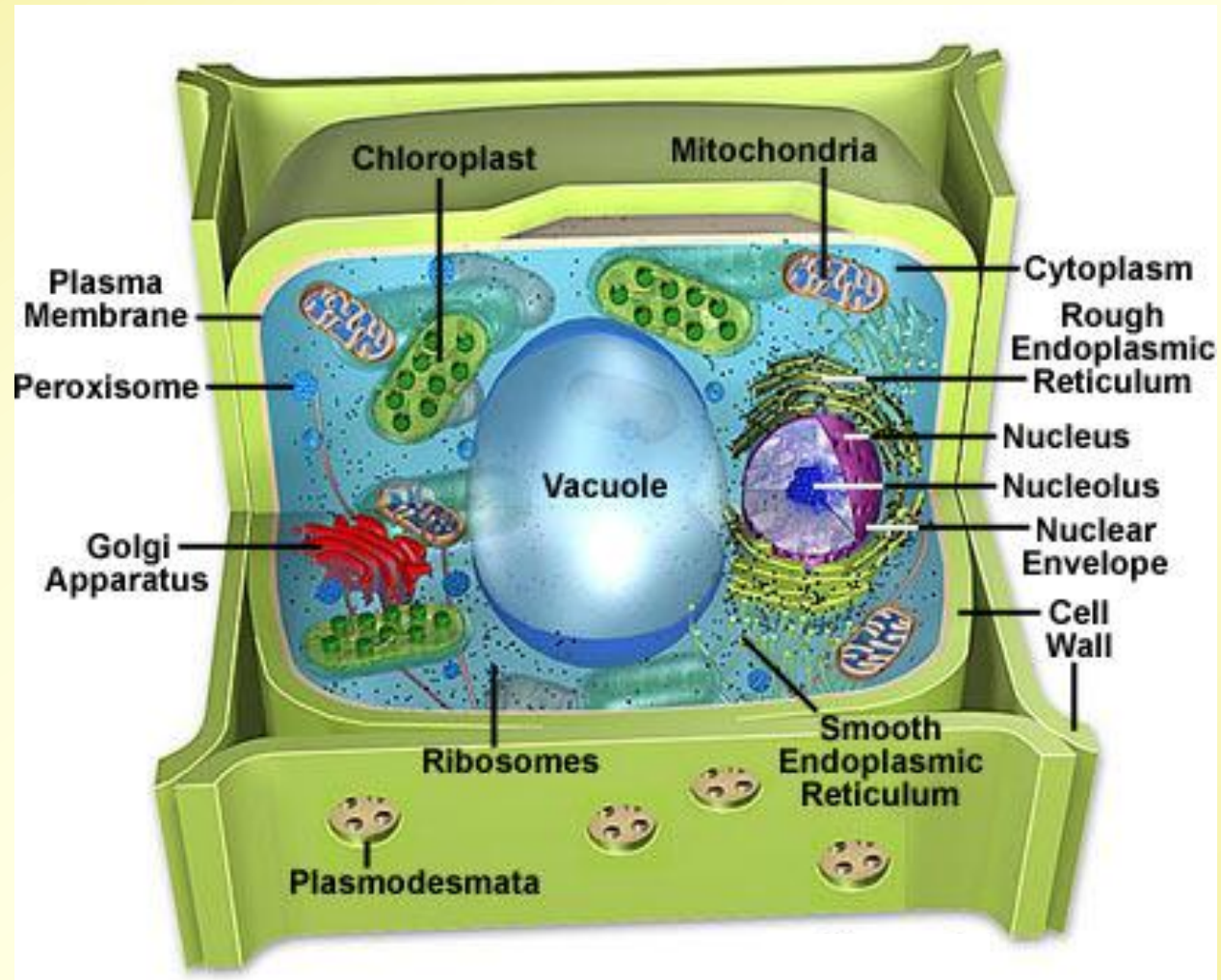
Plant Cell Walls

Plant Cell Walls
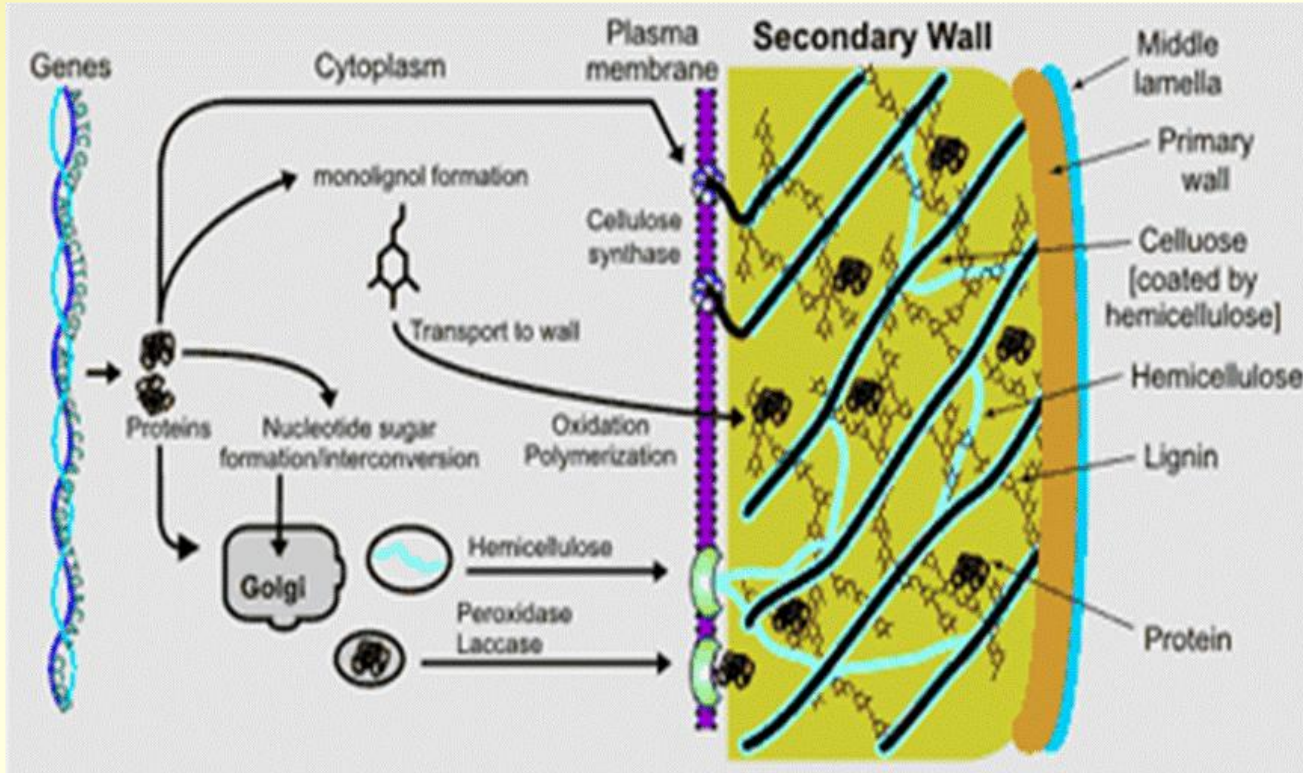
Lignin is a Natural Polymer ("Wood")

Task: remove or <u>reduce</u> lignin to access cellulose

Monolignols
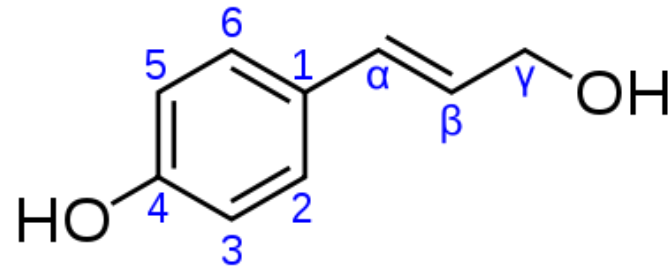
p-Coumaryl alcohol **(H)**
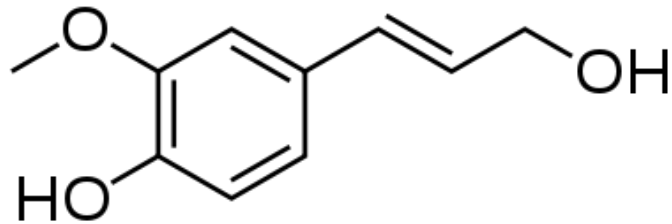
Coniferyl alcohol **(G)**
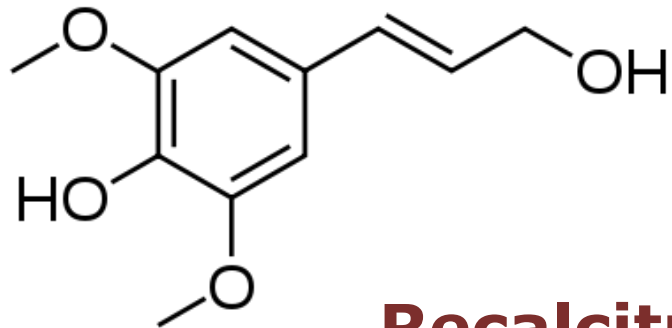
Sinapyl alcohol **(S)**

Recalcitrance Related to S/G Ratio

wikipedia

Task: Develop Mathematical Pathway Model for Lignin Biosynthesis
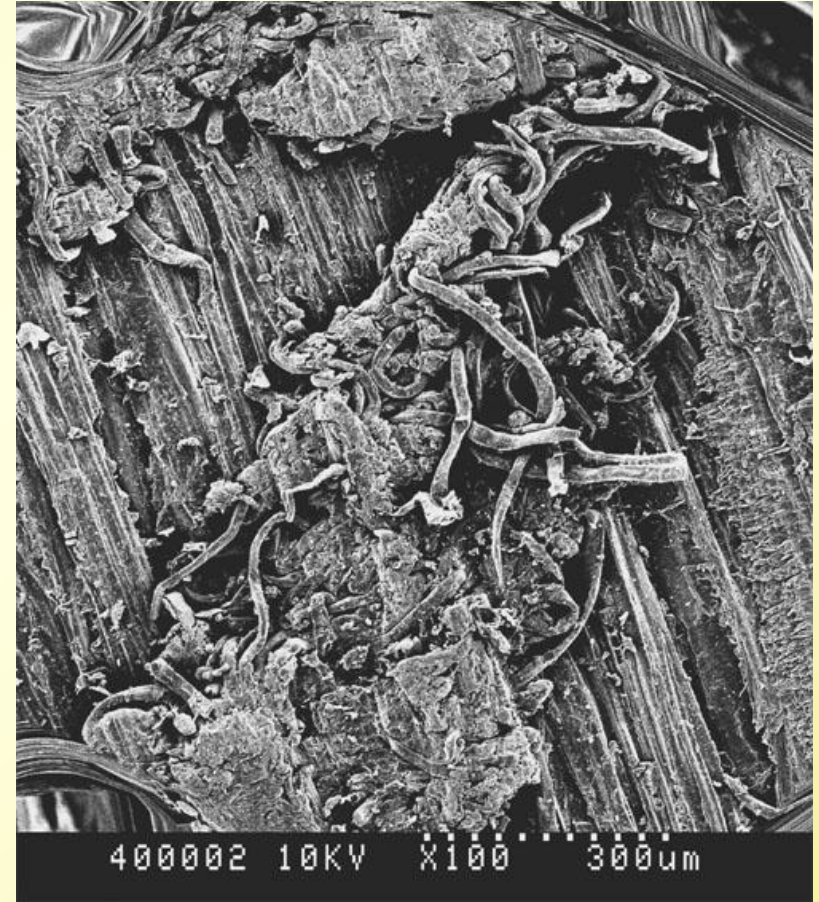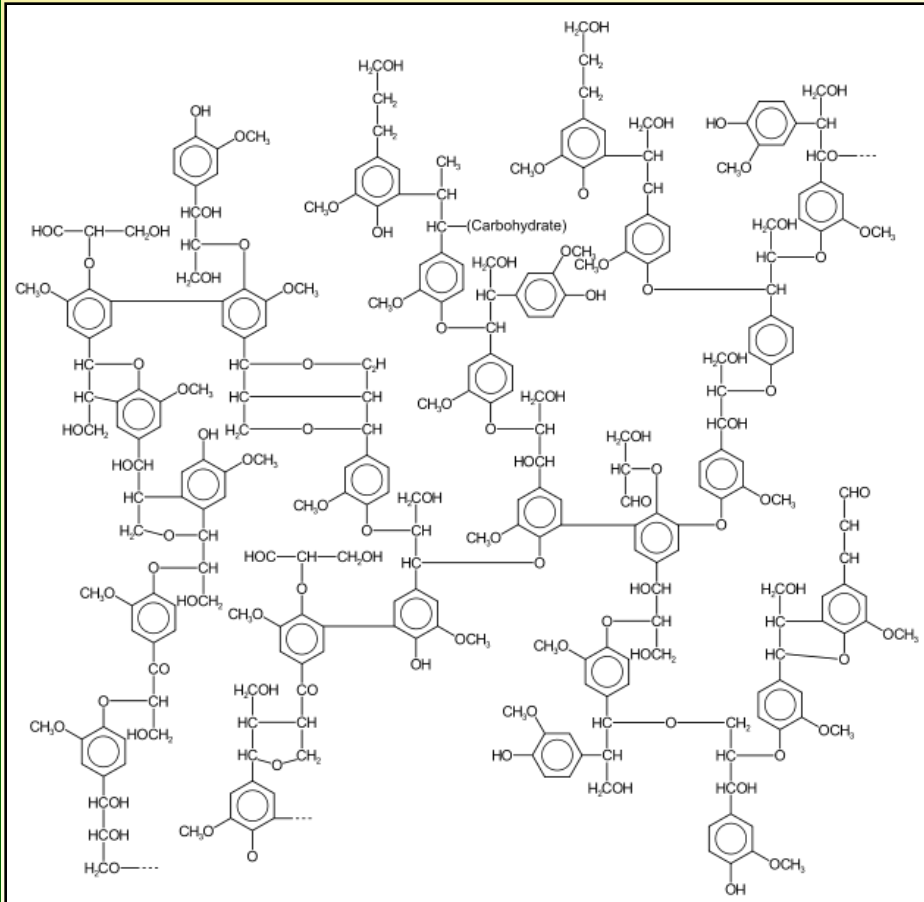
- Model design
    - Choices of models
        - Stoichiometric
        - Dynamic

- Model analysis
    - Insights into pathways
    - Conversion of stoichiometric into dynamic models
    - Optimization

- Results
    - Suggestions from optimization
    - New postulates

General Modeling Procedure

e.o.voit

www.ornl.gov
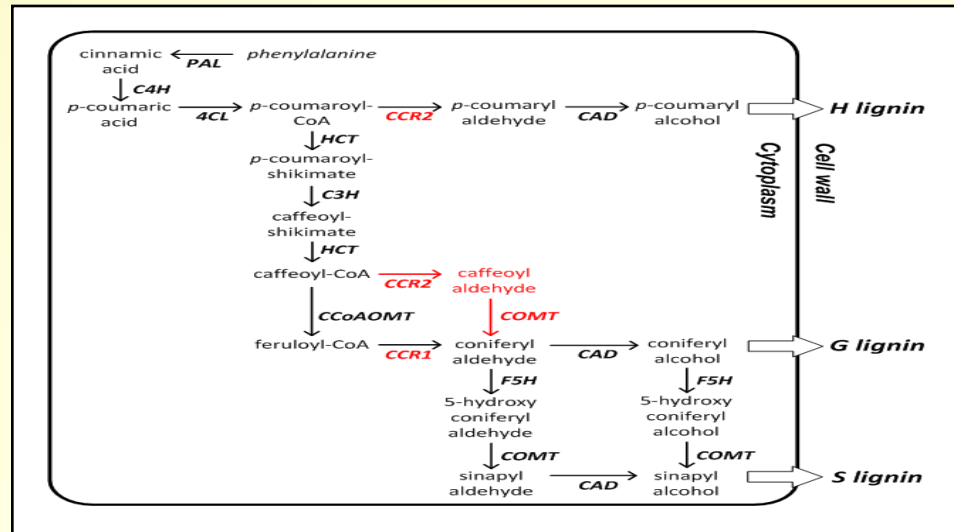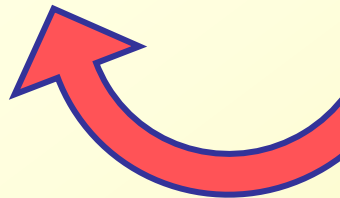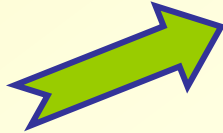
$$\frac{dX_{12}}{dt} = \gamma_{0,12}X_{24}^{f_{0,12,24}}X_{30}^{f_{0,12,30}} + \gamma_{51,12}X_{12}^{f_{51,12,12}}X_{28}^{f_{51,12,28}}X_{51}^{f_{51,12,51}}X_{52}^{f_{51,12,52}}$$

$$+ \gamma_{4,17}X_4^{f_{4,17,4}}X_{22}^{f_{41,17,22}} + \gamma_{6,17}X_6^{f_{6,17,6}}X_{22}^{f_{6,17,22}}$$

```
% model_ode
function dy = PP_ode(t,y)
...
dy(1) = y(1) * (b1 - b2 * y(1) - b3 * y(2));
dy(2) = y(2) * (b4 * y(1) + b5 - b6 * y(2));
end
```

Stoichiometric Analysis
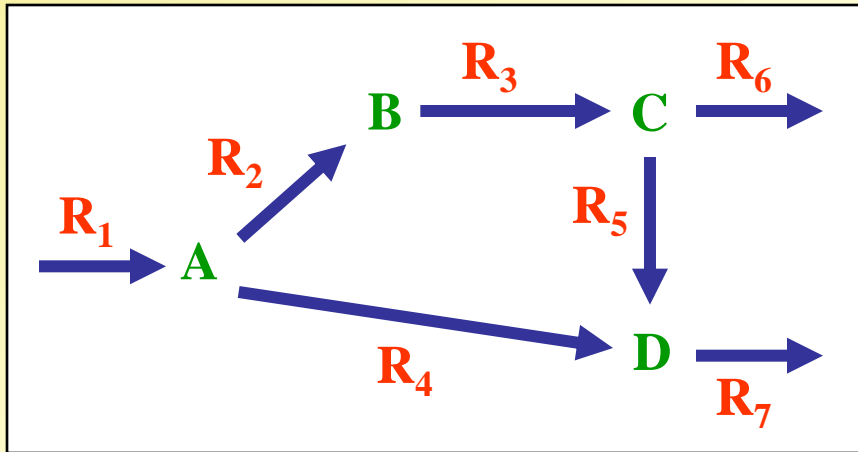
$$N = \begin{pmatrix} 1 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & -1 \end{pmatrix} \begin{matrix} A \\ B \\ C \\ D \end{matrix}$$

- Change in substrate concentrations (S) is a function of fluxes (R) and of the stoichiometric matrix N; at steady state:

$$dS/dt = N \cdot R = 0$$

- No unique solution: optimize some criterion (growth rate)

Constraint-Based Flux-Balance Analysis (FBA)
Reduce solution space with physico-chemical constraints

FBA and MOMA

- Starting Point: $d\mathbf{S}/dt = \mathbf{N} \cdot \mathbf{R} = 0$ (No unique solution)

- FBA: optimize some criterion under additional constraints

- MOMA ("Minimization of Metabolic Adjustment"):

Transgenic strain tries to
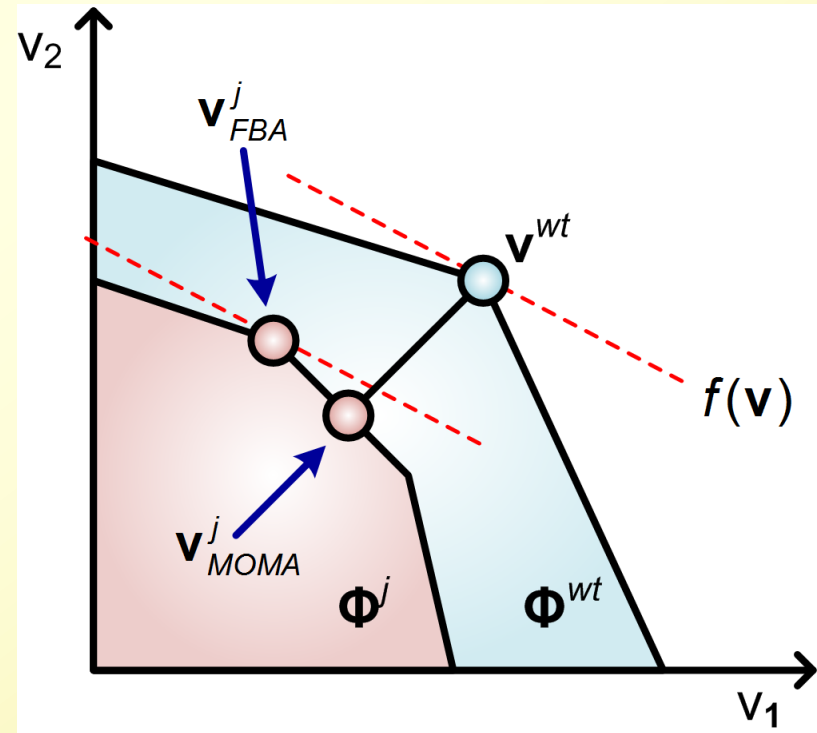emulate wild type
as much as possible;
optimum inferior

Stoichiometric and Flux Balance Analysis

Advantages:

    No kinetic details needed, just topology and fluxes
    Linear system: no real size limitation
    Steady-state solution space given by "kernel"
    Straightforward optimization
    Solution optimizes a desirable criterion (*e.g.*, growth rate)

Limitations:

    Kinetic information cannot be used
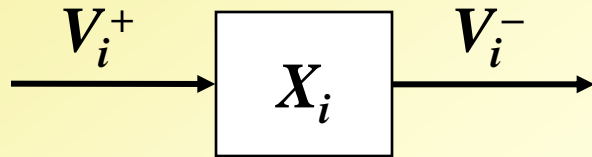   No nonlinearities allowed
    No regulatory signals can be considered, but:
    Optimal strategies of flux alteration affected by signals
    Arbitrariness in objective function for optimization

Choice of Dynamic Model Structure

$$V_i^+ \xrightarrow{\hspace{1cm}} \boxed{X_i} \xrightarrow{V_i^-}$$

$$\dot{X}_i = \frac{dX_i}{dt} = V_i^+ - V_i^-$$

$$V_i^+ = V_i^+(\underbrace{X_1, X_2, \ldots, X_n}_{inside}, \underbrace{X_{n+1}, \ldots, X_{n+m}}_{outside})$$

complicated

inside   *outside*

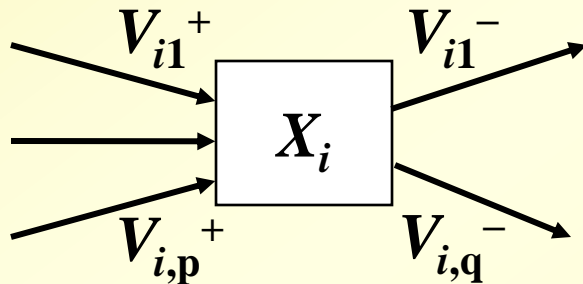**Solution with Potential:**

$$V_{ik}^{+/-} = \gamma_{i,k} \prod_{j=1}^{n} X_j^{f_{k,i,j}}$$

"Biochemical Systems Theory"
(BST)

Alternative Formulations Within BST

S-system Form:

$$\dot{X}_i = \alpha_i X_1^{g_{i1}} X_2^{g_{i2}} \ldots X_{n+m}^{g_{i,n+m}} - \beta_i X_1^{h_{i1}} X_2^{h_{i2}} \ldots X_{n+m}^{h_{i,n+m}}$$

$$\dot{X}_i = \frac{dX_i}{dt} = \sum V_{ij}^+ - \sum V_{ij}^-$$

Alternative Formulations

S-system Form:

$$\dot{X}_i = \alpha_i X_1^{g_{i1}} X_2^{g_{i2}} \ldots X_{n+m}^{g_{i,n+m}} - \beta_i X_1^{h_{i1}} X_2^{h_{i2}} \ldots X_{n+m}^{h_{i,n+m}}$$
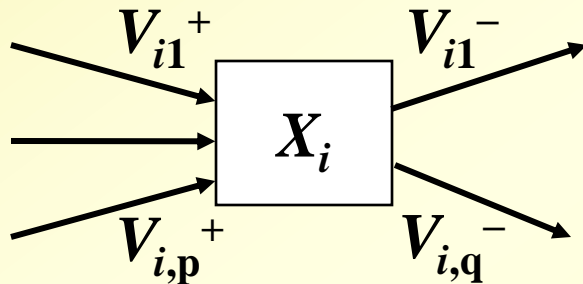

$$\dot{X}_i = \frac{dX_i}{dt} = \sum V_{ij}^+ - \sum V_{ij}^-$$

Generalized Mass Action Form:

$$\dot{X}_i = \sum \pm \gamma_{ik} \prod X_j^{f_{ijk}}$$

Advantages of BST Models

Prescribed model design: Rules for translating diagrams into equations; rules can be automated

Direct interpretability of parameters and other features

One-to-one relationship between parameters and model structure simplifies parameter estimation and model identification
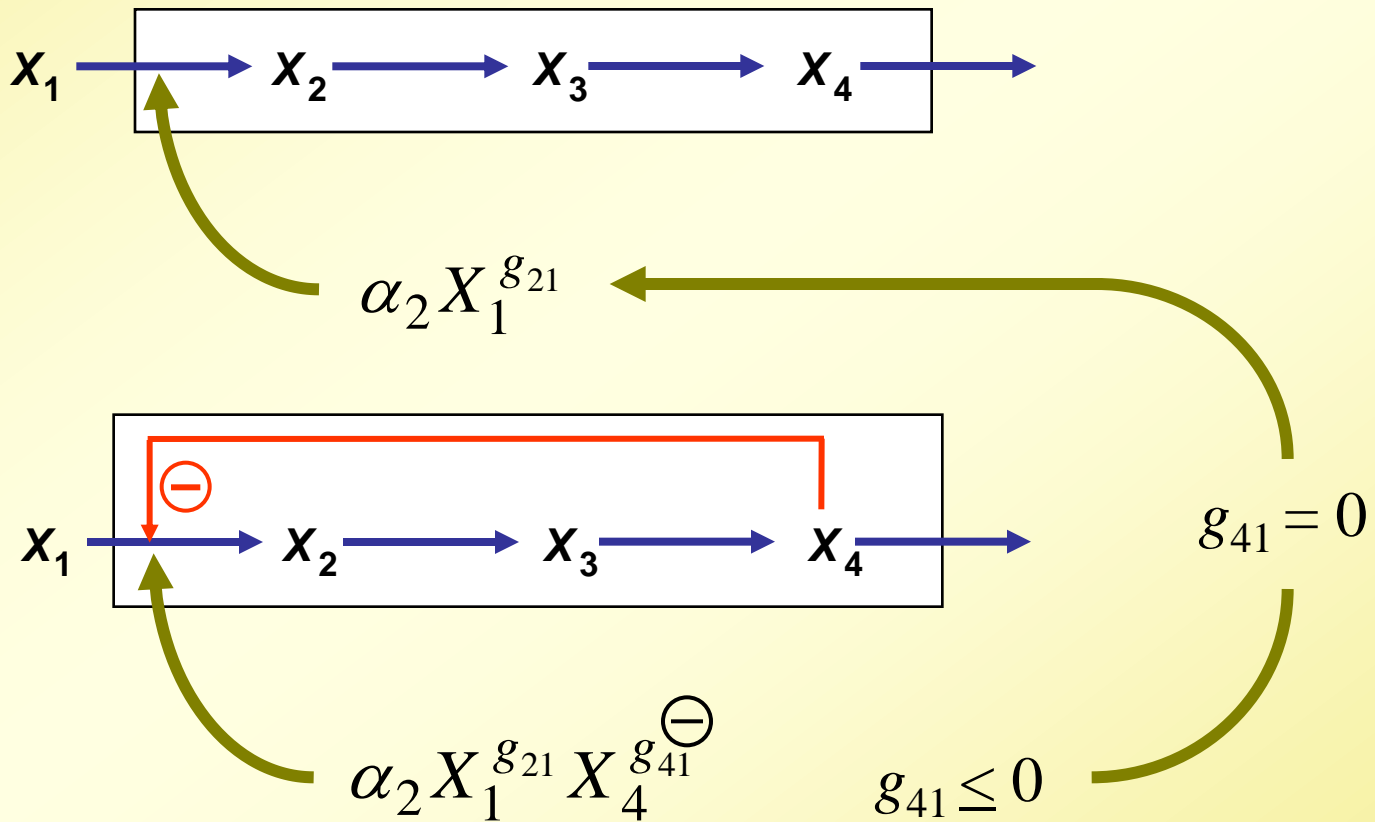
Simplified steady-state computations (for S-systems), including steady-state equations, stability, sensitivities, gains

Simplified optimization under steady-state conditions

Efficient numerical solutions and time-dependent sensitivities

In some sense minimal bias of model choice and minimal model size; easy scalability

Mapping
Structure ⬌ Parameters

S-system Steady-State Equations Linear

$$\dot{X}_i = \alpha_i X_1^{g_{i1}} X_2^{g_{i2}} \ldots X_{n+m}^{g_{i,n+m}} - \beta_i X_1^{h_{i1}} X_2^{h_{i2}} \ldots X_{n+m}^{h_{i,n+m}} = 0$$

Define $Y_i = \log(X_i)$:

$$\log \alpha_i + g_{i1} Y_1 + g_{i2} Y_2 + g_{i,n+m} Y_{n+m}$$
$$= \log \beta_i + h_{i1} Y_1 + h_{i2} Y_2 + h_{i,n+m} Y_{n+m}$$

$$\mathbf{Y}_D = \mathbf{A}_D^{-1} \cdot \mathbf{b} - \mathbf{A}_D^{-1} \cdot \mathbf{A}_I \cdot \mathbf{Y}_I$$

S-system highly nonlinear, but steady-state equations linear.

Pathway Optimization with S-systems

Optimization under steady-state (batch) conditions becomes

Linear Program

even though (nonlinear) kinetics is taken into account:

> *maximize* log(flux)   [or log(variable)]
>
> *subject to:*
>
> Steady-state conditions in log(variables)
>
> Constraints on log(variables)
>
> Constraints on log(fluxes)

Pathway Optimization (continued)

Great Advantage:

Methods of *Operations Research* applicable
- very well understood
- applicable for over 1,000 simultaneous variables
- robust and efficient
- incomparably faster than nonlinear methods

Torres, Alvarez, Voit, …: Applications (*e.g.*, citric acid, ethanol, glycerol, L-carnitine)

Hatzimanikatis, Bailey, Floudas, 1996: Use these features for optimization of pathway structure

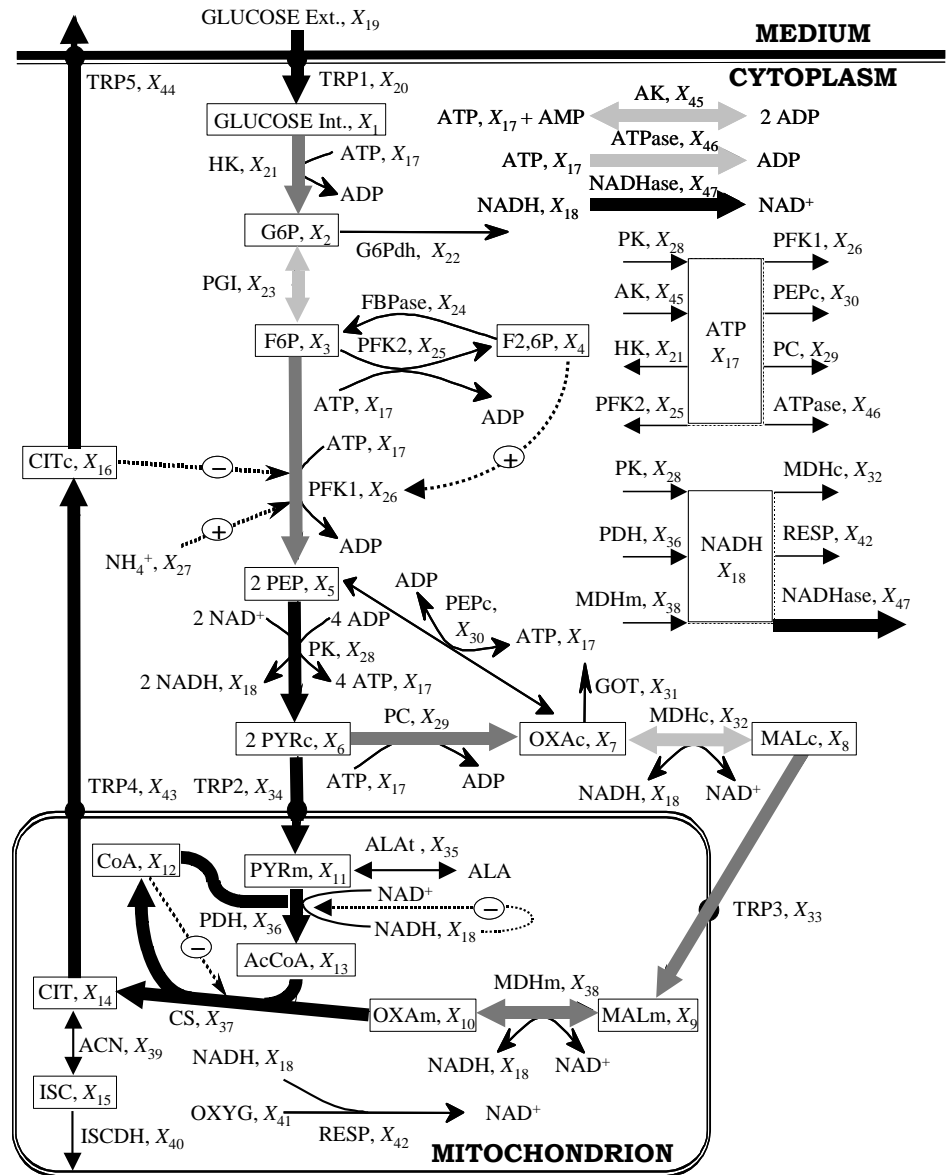*Marin-Sanguino, Torres, Polisetty, Gatzke, Voit, …:* Extension to GMA models via
iterative methods, branch-and-reduce methods, geometric programming

Example

Citric acid yield:

Optimization prescribes enzyme activity levels that lead to maximal citric acid production while satisfying constraints on metabolites and fluxes.

Maximal increase: ~ 12 fold

Notable Results

Citric acid system contains ~ 20 accessible enzymes / genes

Optimize by allowing changes in all enzymes:
Yield increased ~ 12 fold

Q: If only a single enzyme may be changed, which one should it
be? How much could yield be increased?

A: No matter which enzyme is changed, yield does not really
increase!

Q: Change 2, 3, ... enzymes. Yield improvement?

A: 2: none; 3: none, 4-6: almost none; 7 needed for ~3 fold yield!

Interpretation: Standard techniques have found the easy solutions!

Applications of BST

Pathways: purines, glycolysis, citric acid, TCA, red blood cell, trehalose, sphingolipids, lignin, …

Genes: circuitry, regulation, expression patterns, …

Signaling: MAPK, BMP4 (atherosclerosis)

Growth, immunology, pharmaceutical science, forestry, …

Metabolic engineering: optimize yield in microbial pathways

Dynamic labeling analyses possible

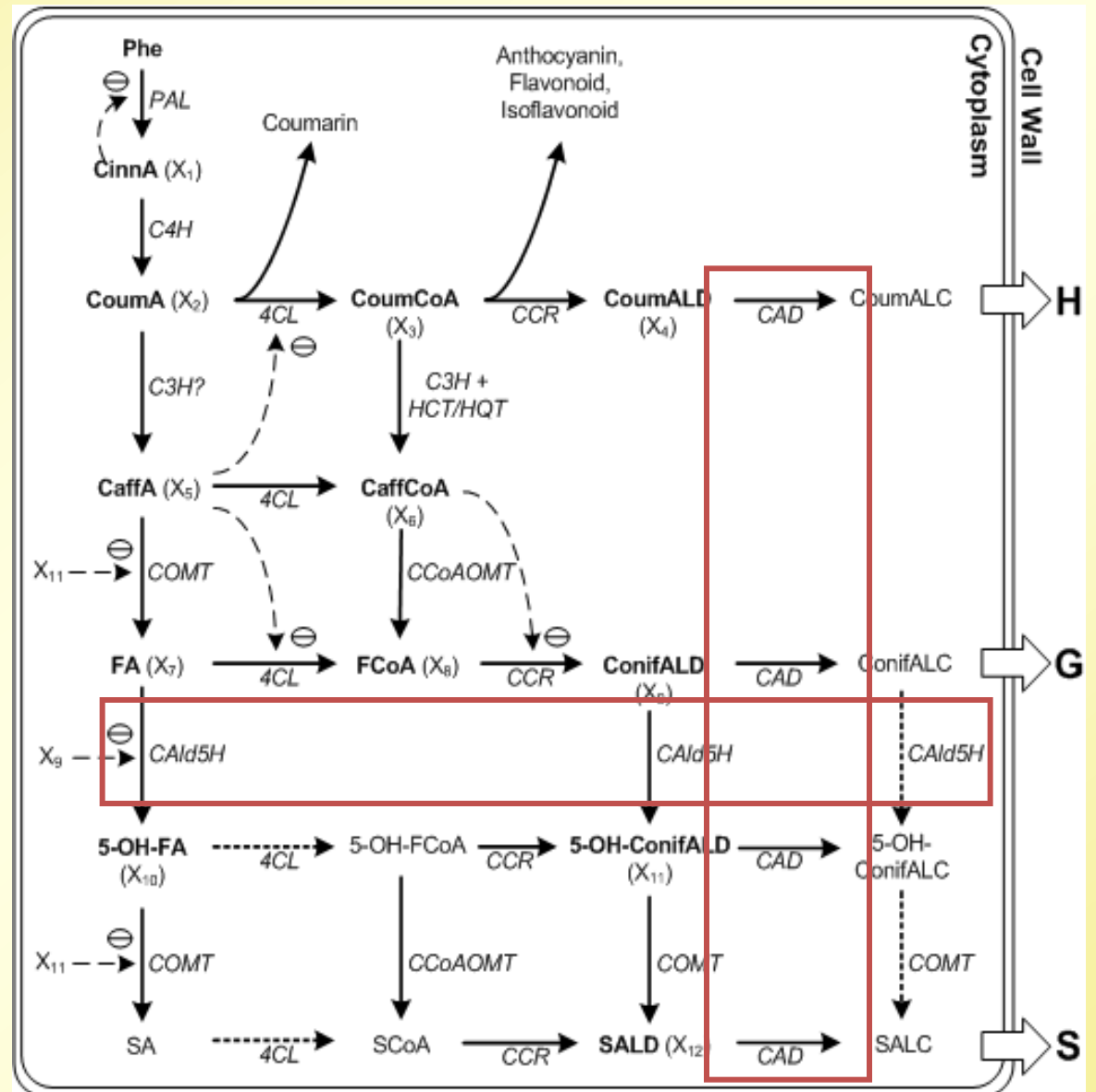Math: recasting, function classification, bifurcation analysis,…

Statistics: S-system representation, S-distribution, trends; applied to seafood safety, marine mammals, health economics

Lignin Biosynthesis in *Populus*

Need for a Model:
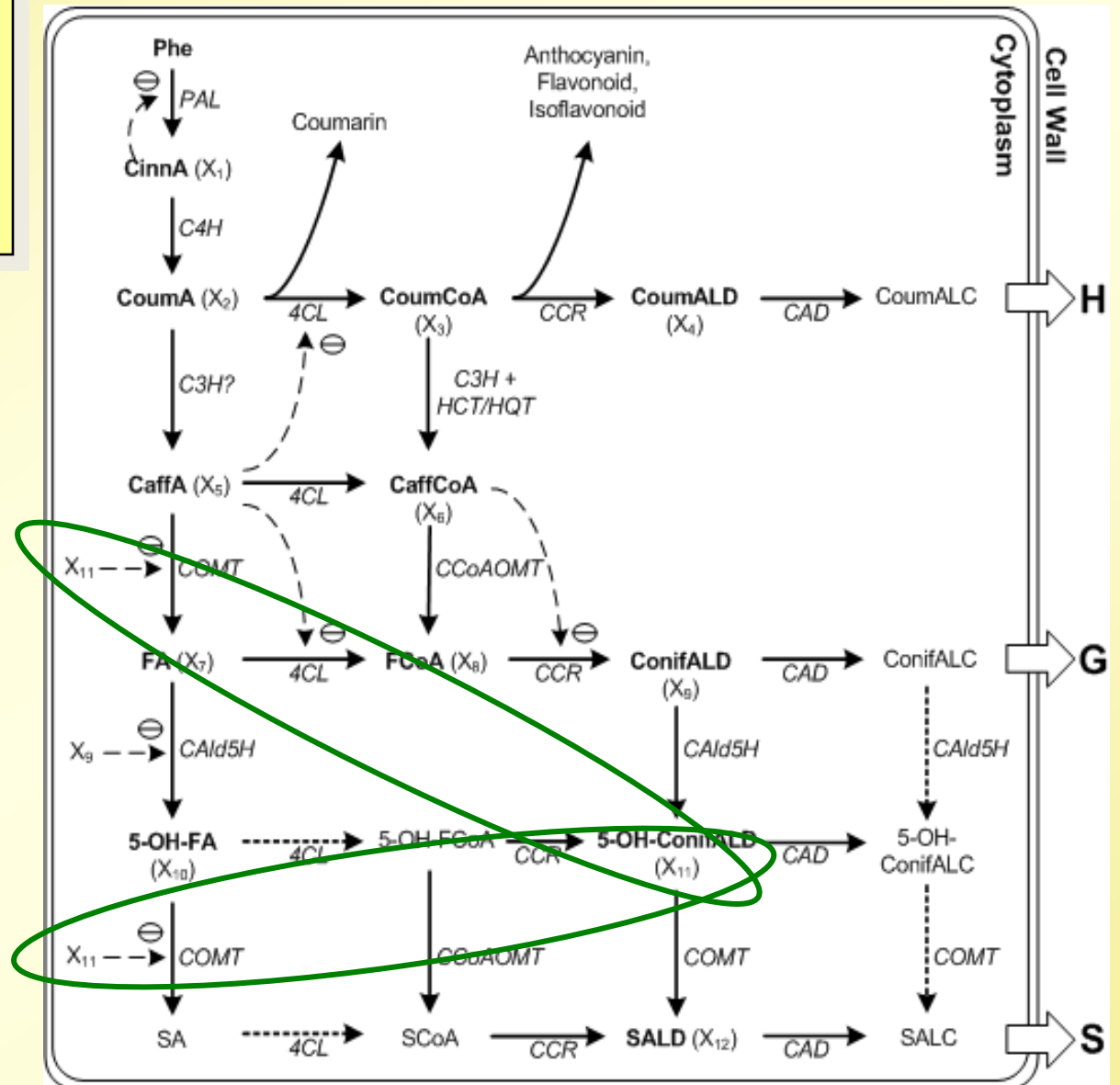
Multiple use of the same enzyme

Lignin Biosynthesis in *Populus*

Need for a Model:

Multiple use of the same enzyme

Regulatory signals (*e.g.*, feedback)

Modeling Lignin Biosynthesis in *Populus*

Five Modeling Steps:

1. Use literature info (alleged pathway structure) to set up stoichiometric / FBA model ⟶ flux distribution

2. Convert FBA model into dynamic BST model, using additional literature information and default assumptions ⟶ dynamic model structure

3. "Train" model with some data (transgenic lines)

4. "Validate" model with other data ⟶ dynamic model

5. Use BST model, for instance, to propose optimized strains

Lignin Biosynthesis in *Populus* (FBA)

Lignin biosynthetic pathway
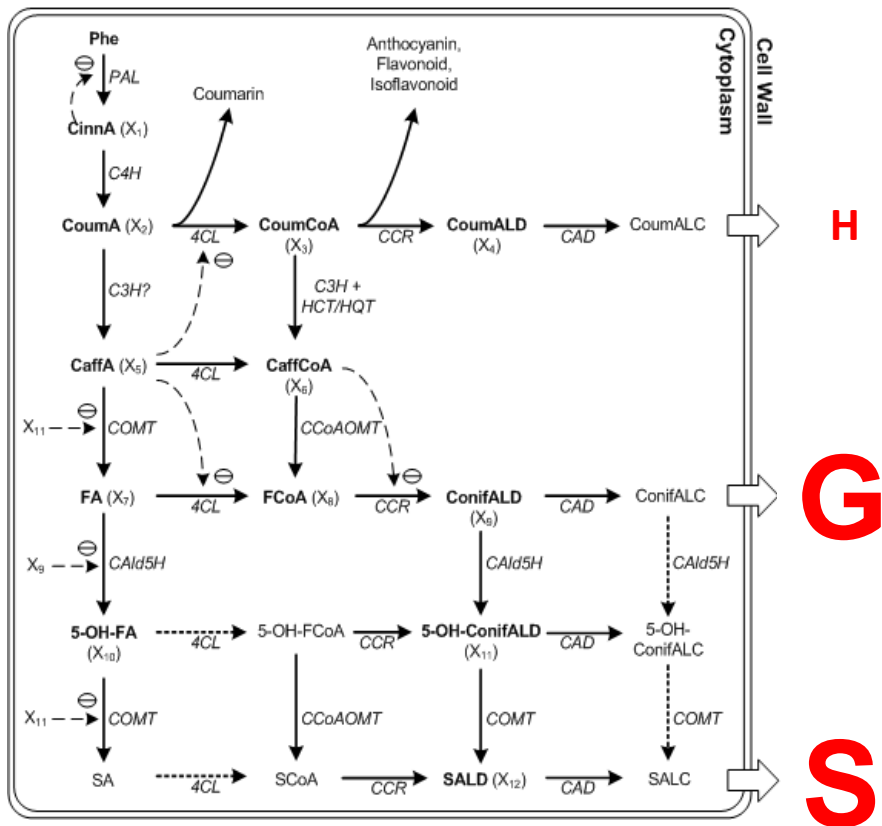
FBA: Computed S/G ratios are optimized to be consistent with experimental S/G ratios

Training dataset
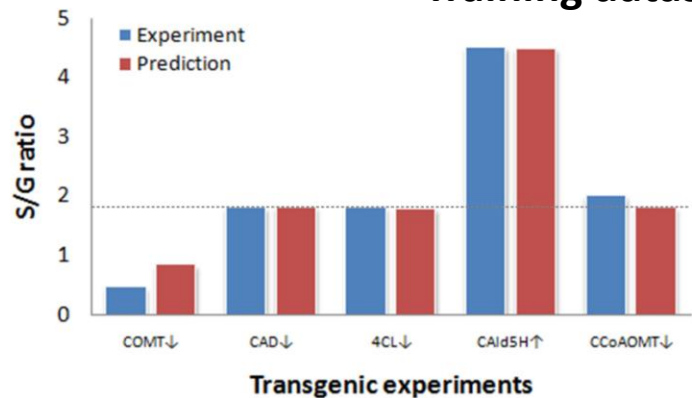
Validation experiments
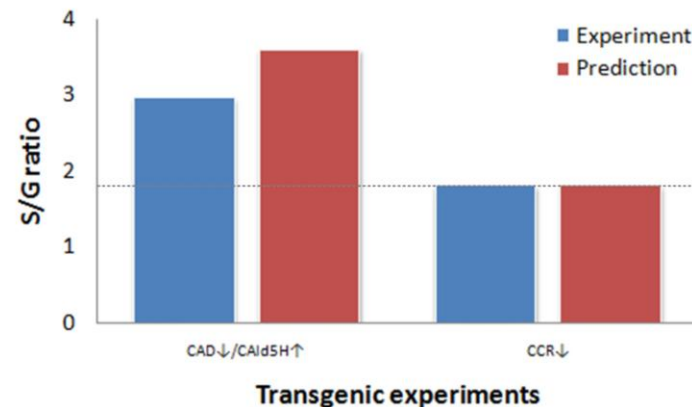
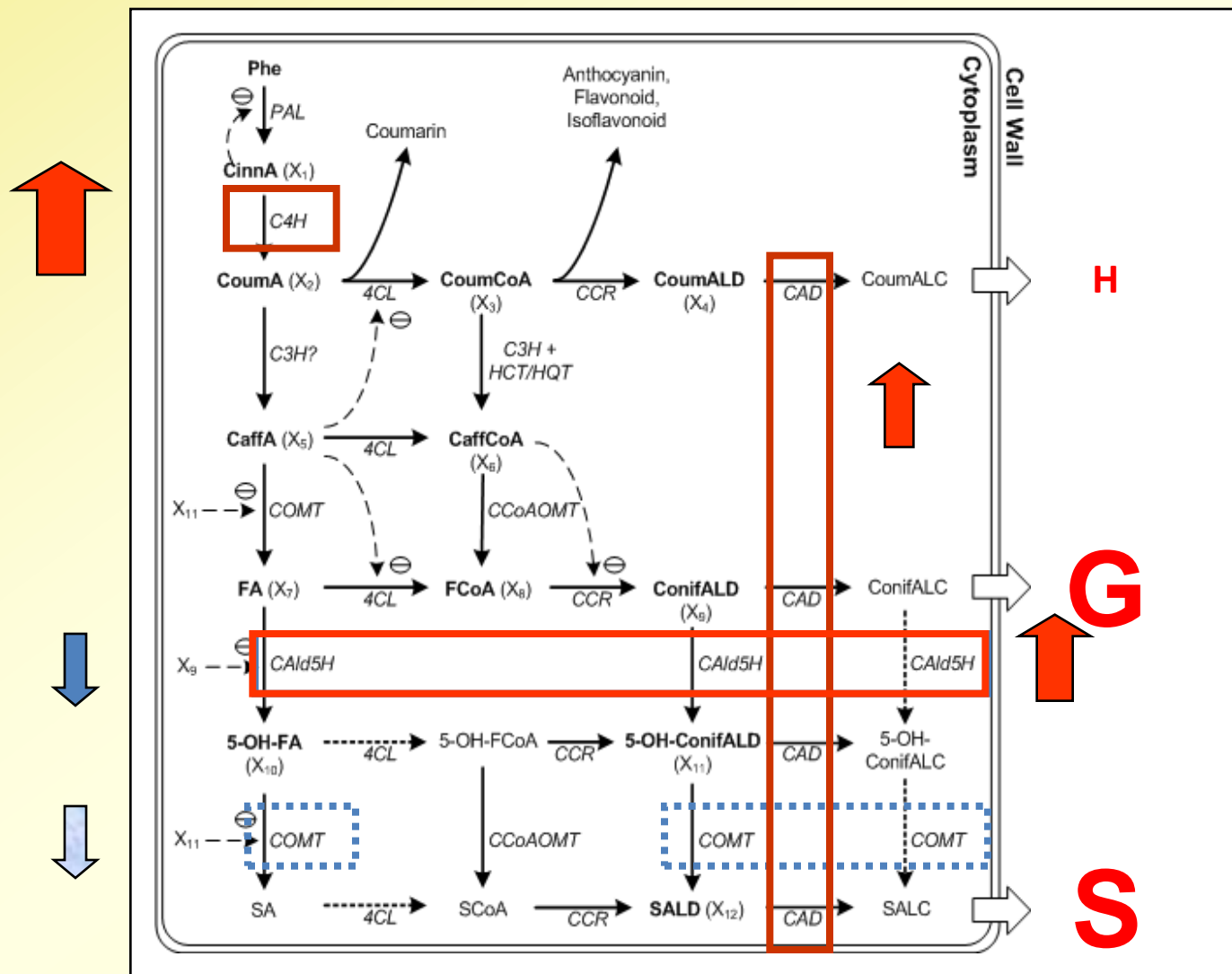# Dynamic BST Modeling of Lignin Biosynthesis in *Populus*

- Convert FBA model into dynamic BST model
- Optimization of the pathway toward higher xylose production by minimization of the S/G ratio

No. of Enzymes	Modified Enzymes	S/G Ratio
Wild-type	N/A	1.8
1	CAld5H (76%)	1.34
2	COMT (96%), CAld5H (71%)	1.29
3	C4H (431%), CAD (167%), CAld5H (134%)	1.11

↓25%

↓38%

- **According to the optimization results, one could achieve ~40% reduction in the S/G ratio by modifying three enzymes and ~25% by modifying just one enzyme (CAld5H)**
- **The set of two enzymes is not a subset of the set of three enzymes!**
- **CAld5H: Scenario 1: decrease; Scenario 2: increase!**

Optimization of Lignin Biosynthesis

Mathematical Modeling of Lignin Biosynthesis in Alfalfa (*Medicago sativa L.*)

- Lignin pathway important:
  - ❑ Model of biofuel production; issues of recalcitrance
  - ❑ Enormously important feedstock; digestibility hindered by lignin
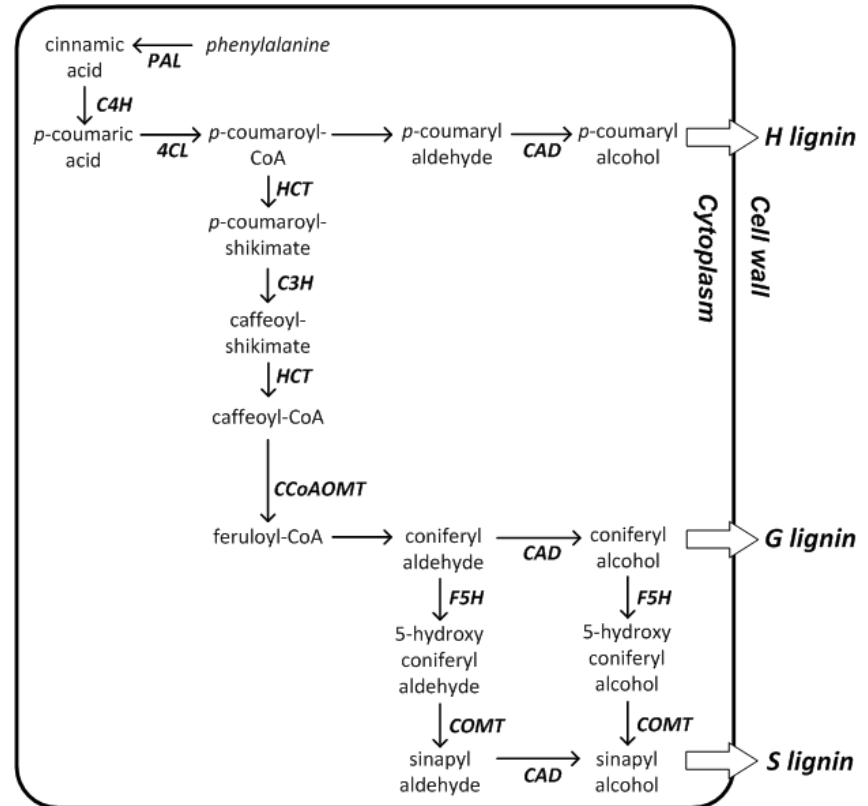
Modeling Development

Steps of Analysis:

1. Use gene/enzyme-modulation data from R. Dixon's group (Noble Center, Oklahoma)

2. Consider differences among internodes

3. Train and validate MOMA model with data

4. Analyze S/G trends during growth

5. Interpret results; formulate postulates

6. Convert FBA model into BST model; propose optimized strains (not yet done)

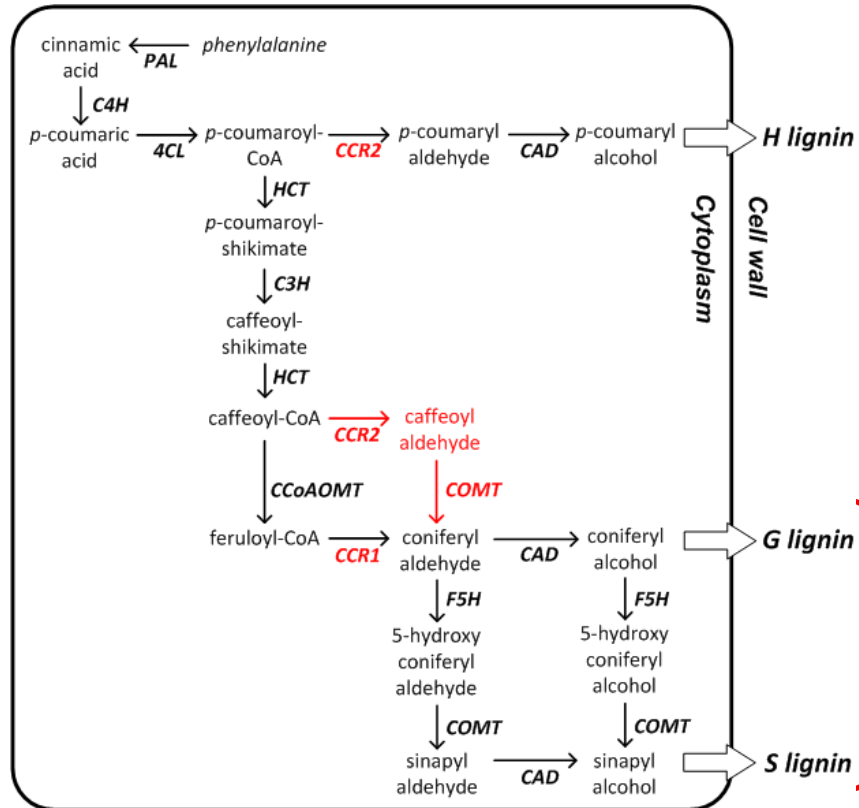"internode"

Data: Enzyme (Gene) Knock-Downs
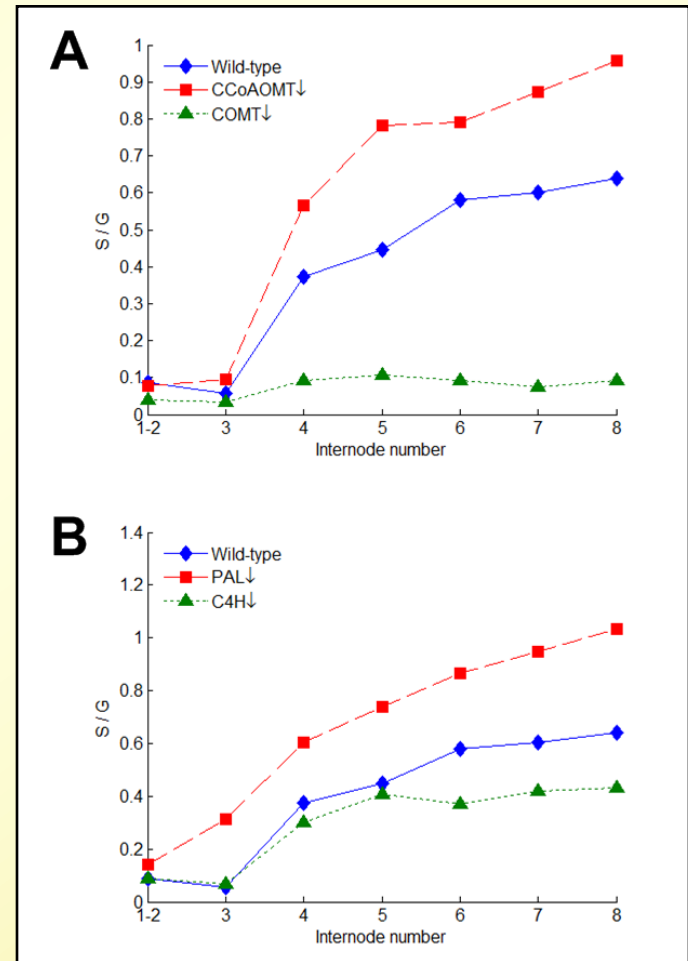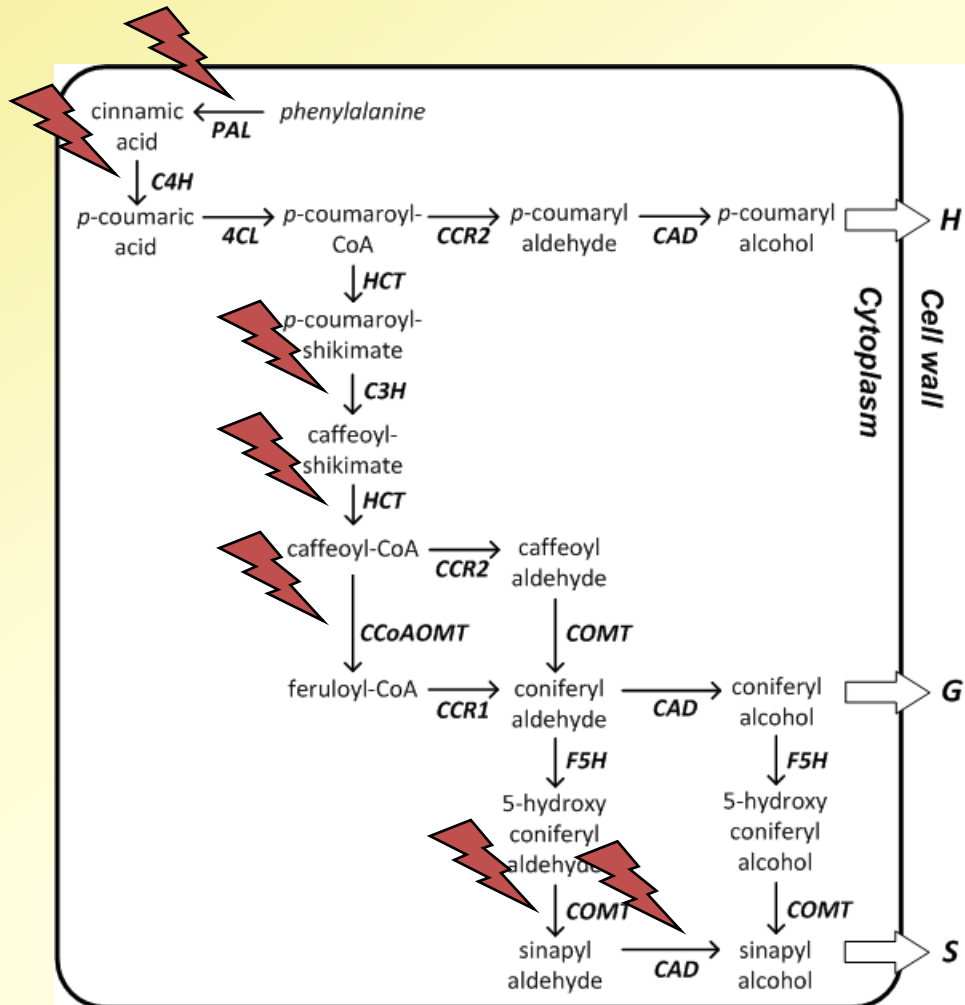
Critical Questions

❑ Is the pathway correct as presently assumed?   **Answer: NO**
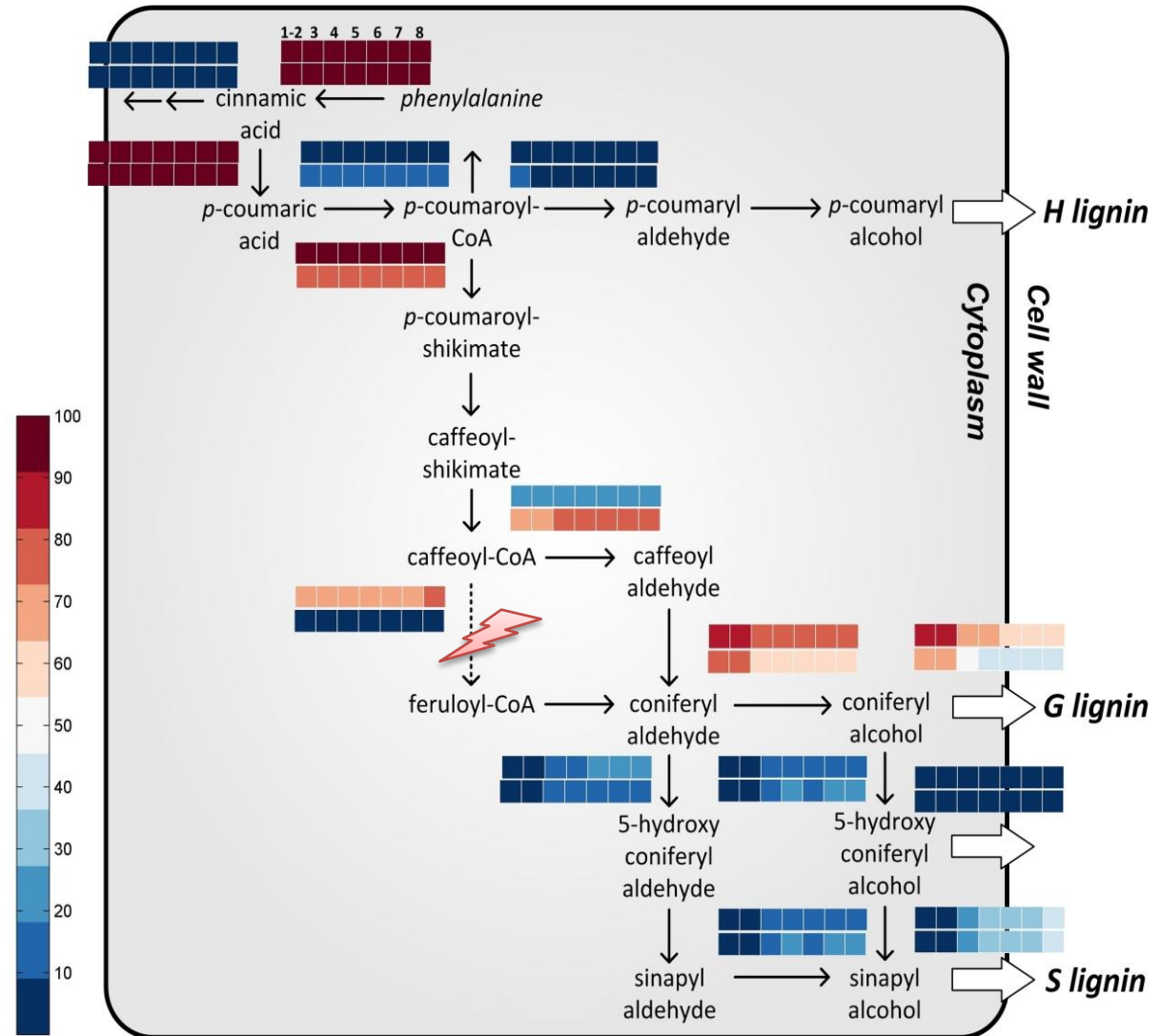❑ Are there independent pathways to G and S lignin?

Revised pathway
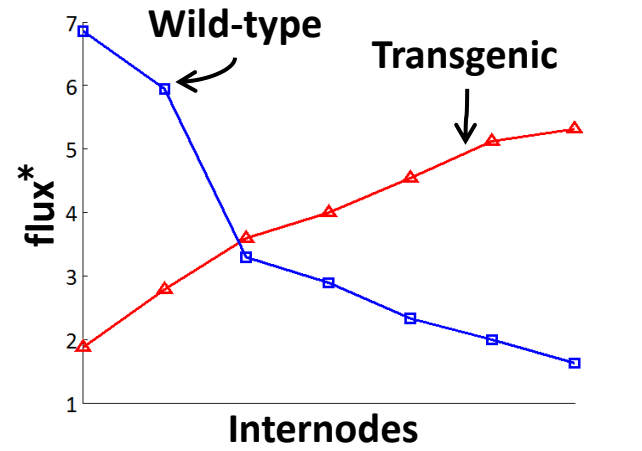

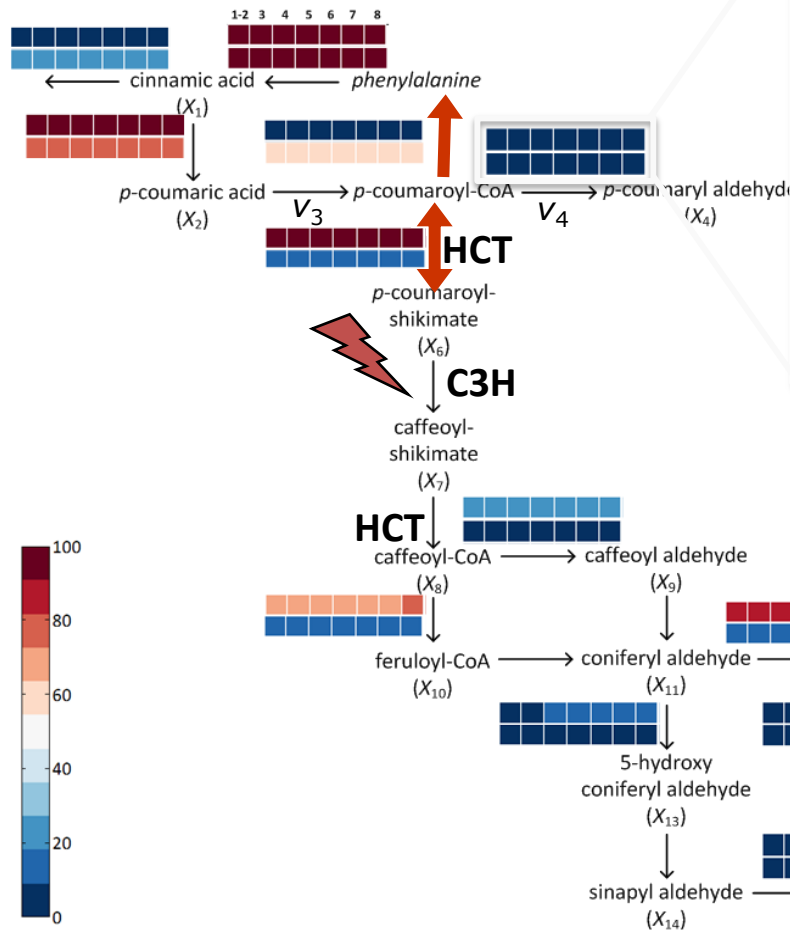S/G ↑ ??

First Result: Dynamic Flux Distributions
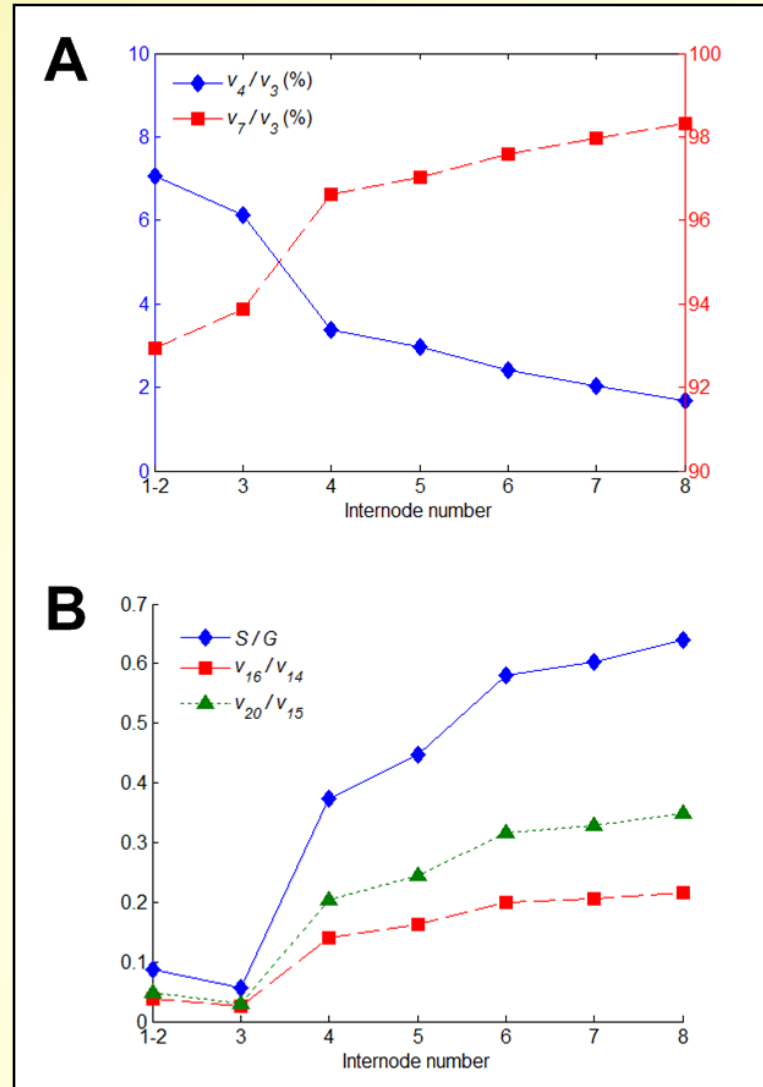
Representation of Model Results

Example:
Wild Type *vs.*
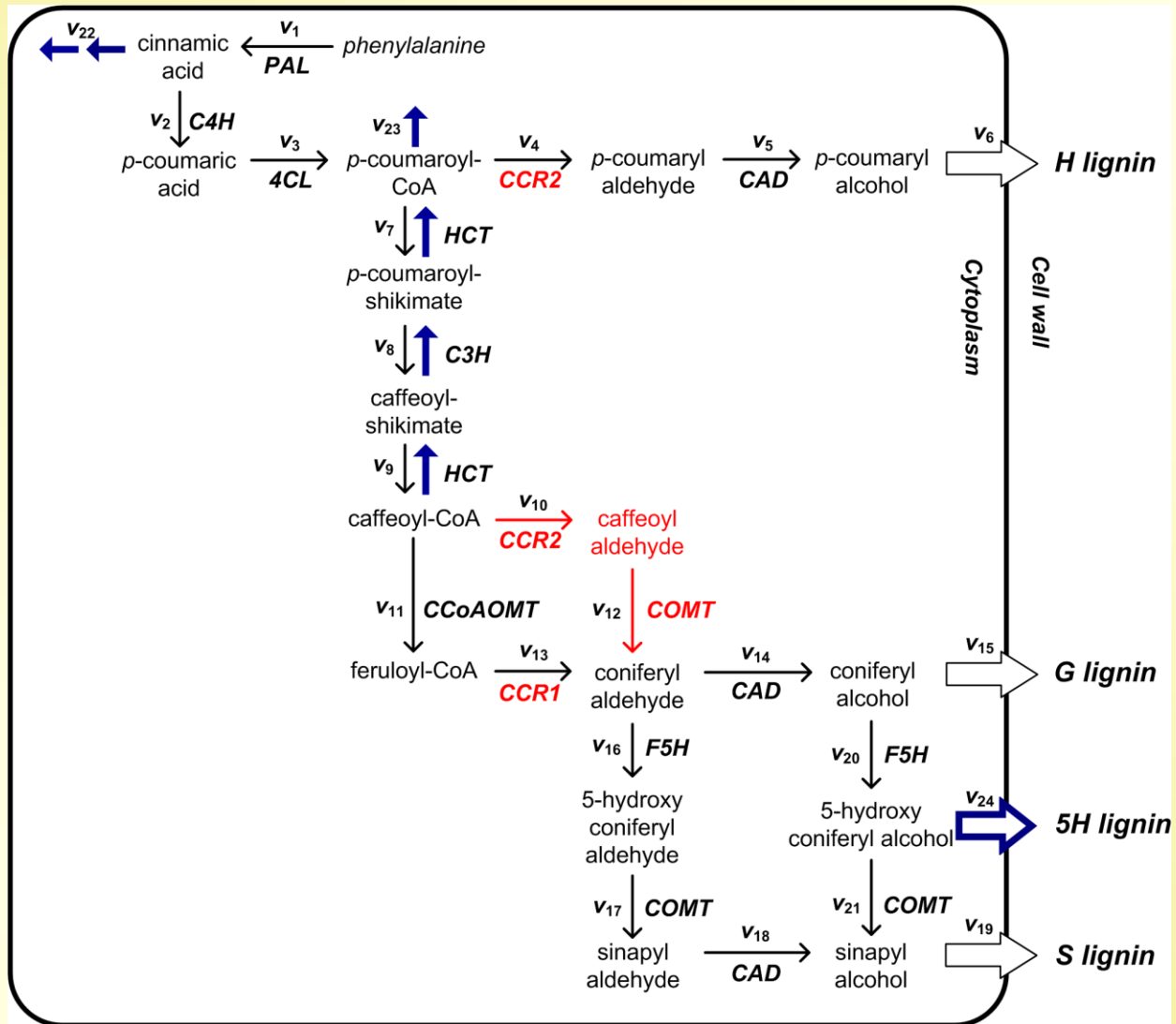**CCoAOMT**
down-regulation

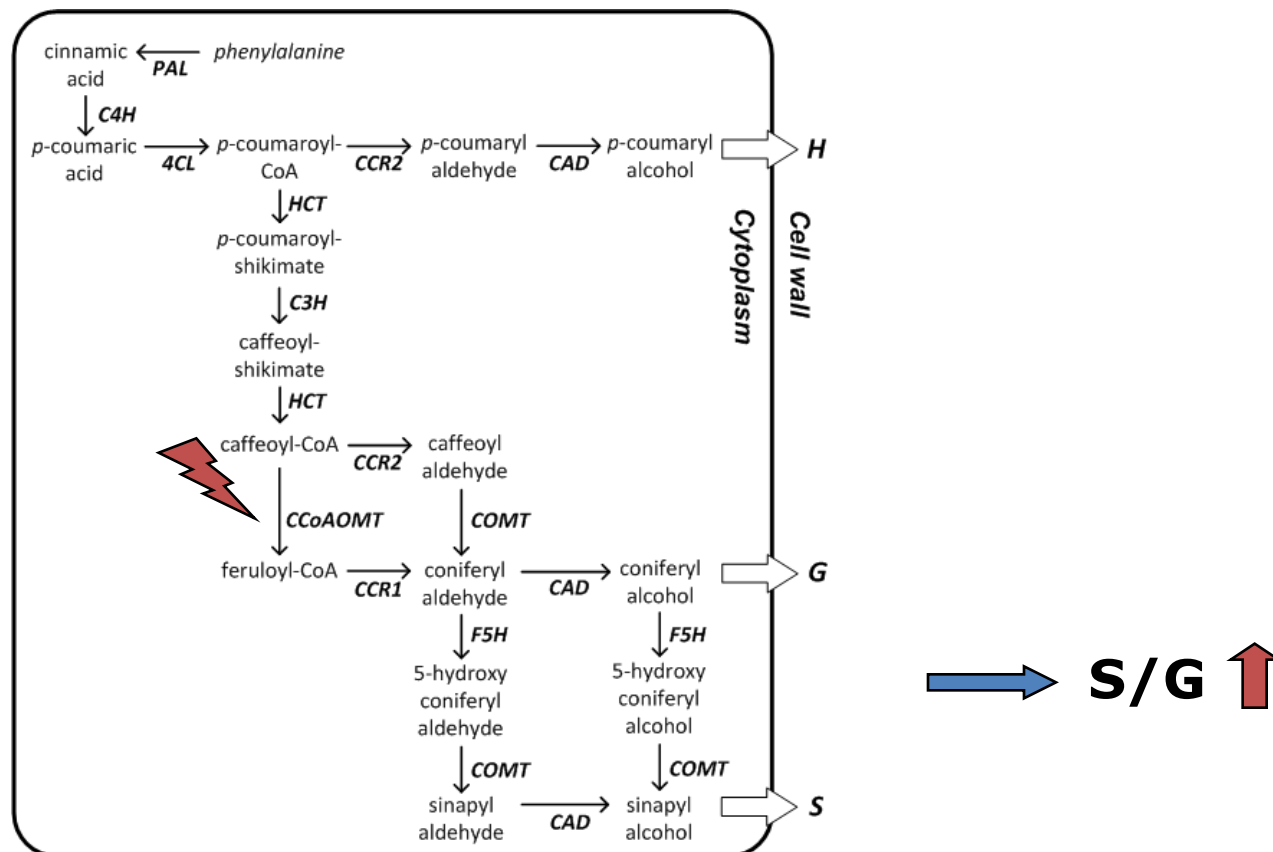# Model-Based Postulates

Example: C3H down-regulation: H

* Postulate 1: HCT must be reversible; excess p-coumaroyl-CoA used elsewhere
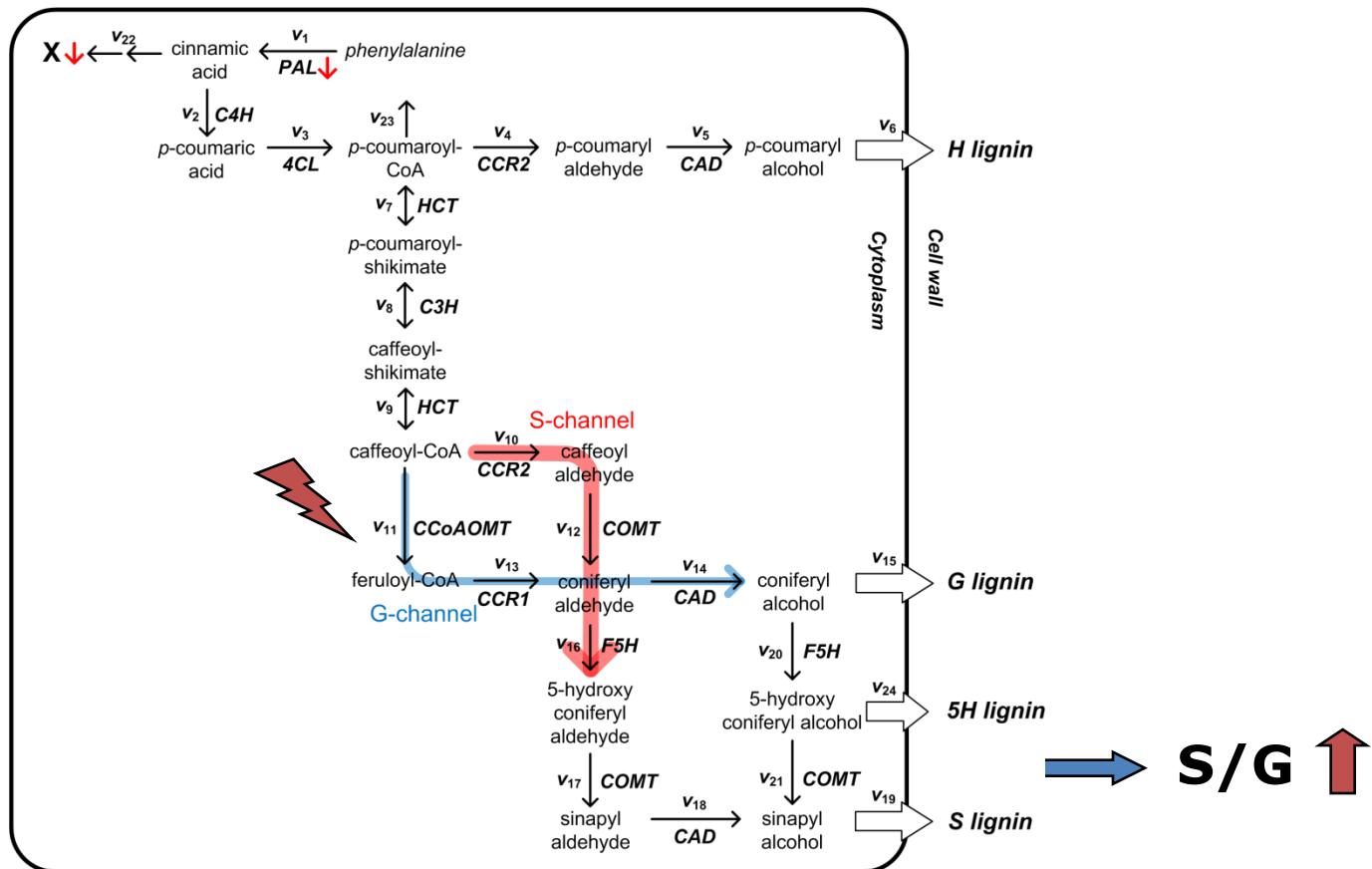
Postulate: Additional Processes

Model–Based Resolution of a Puzzle

Puzzle: Down-regulation of CCoAOMT results in different S/G ratio than wild type, even though alteration occurs before common precursors
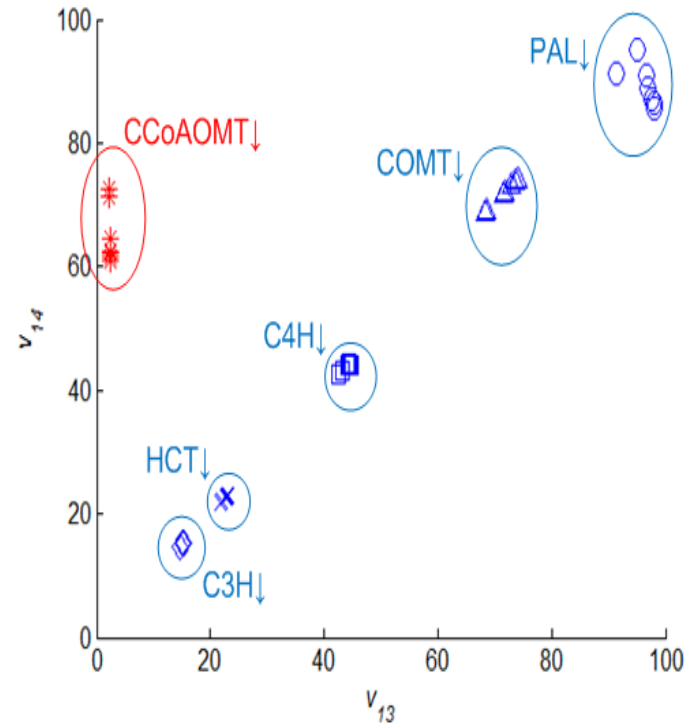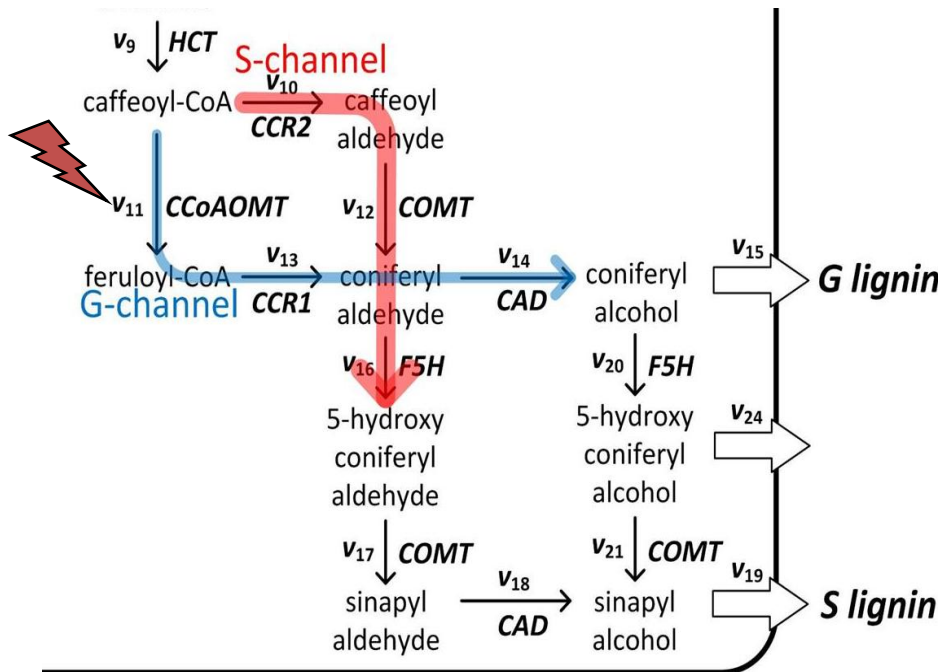
Model-Based Channel Postulate

• **Postulate 2: Existence of specific "S-channel" and "G-channel"**
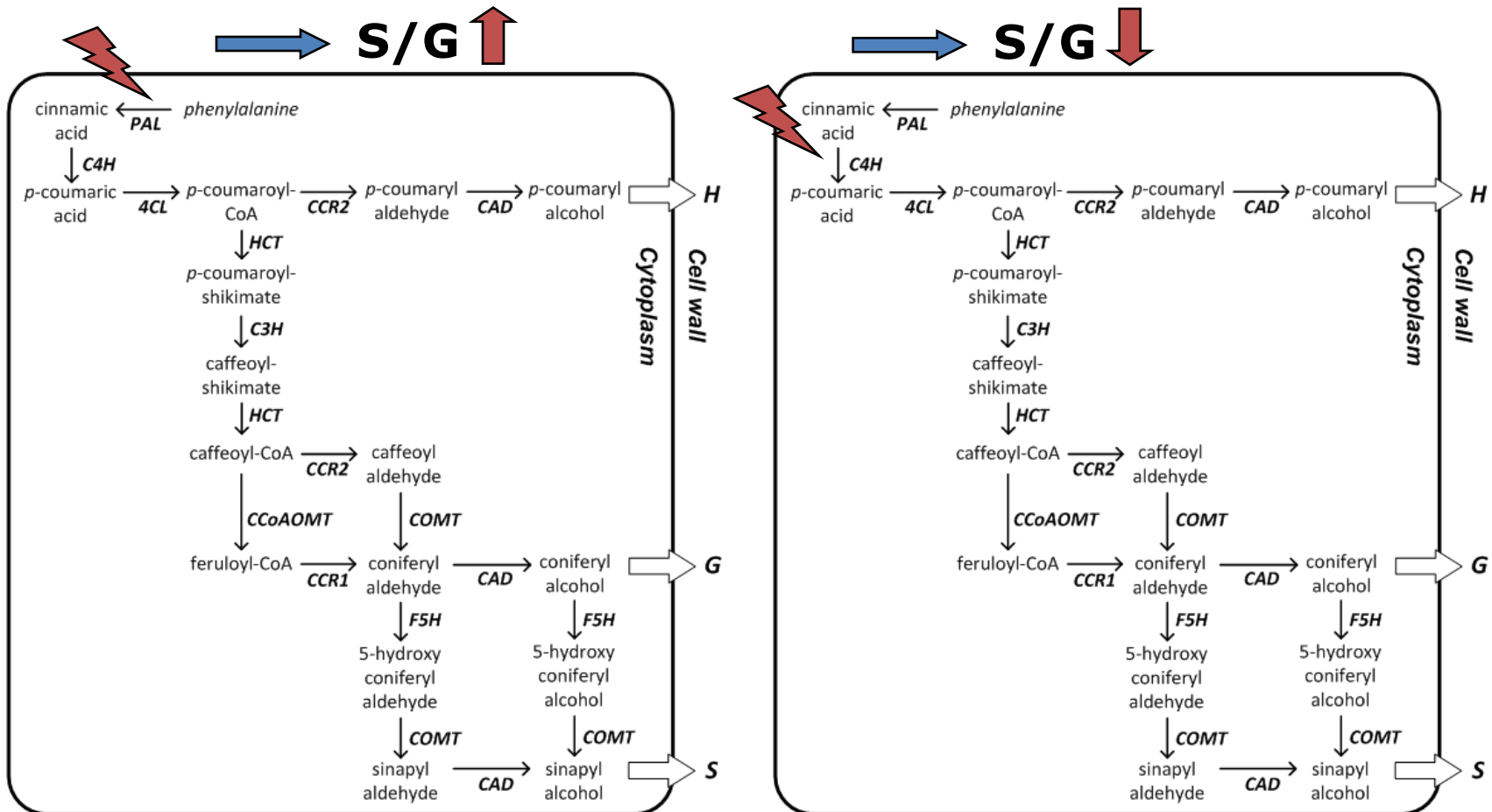
Support for Channel Postulate

- "G-channel" fluxes very strongly correlated, except for ΔCCOAMT
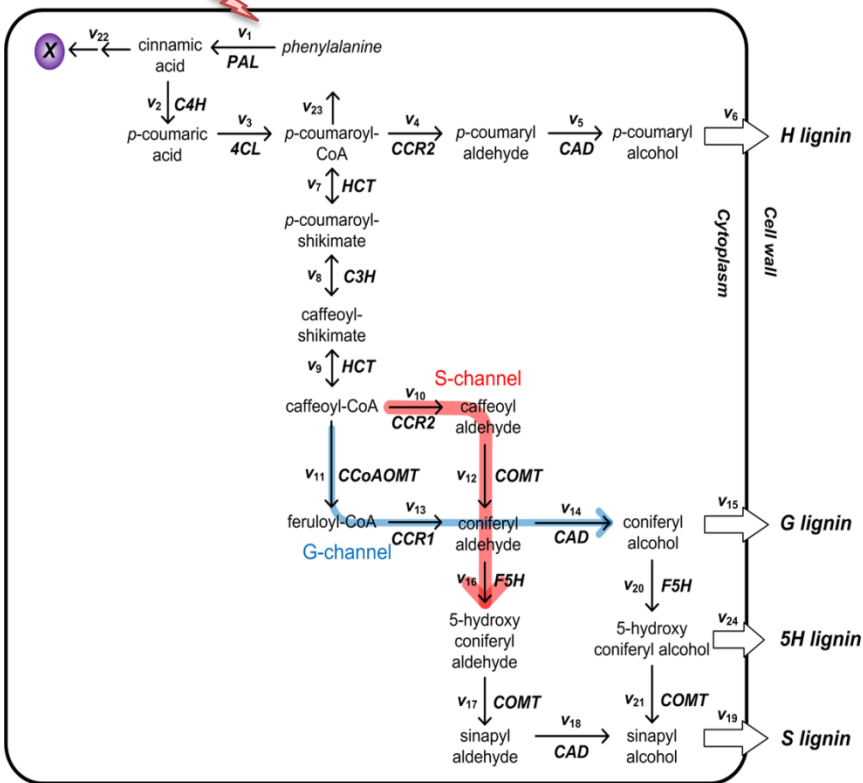

S/G

Model-Based Resolution of a Puzzle

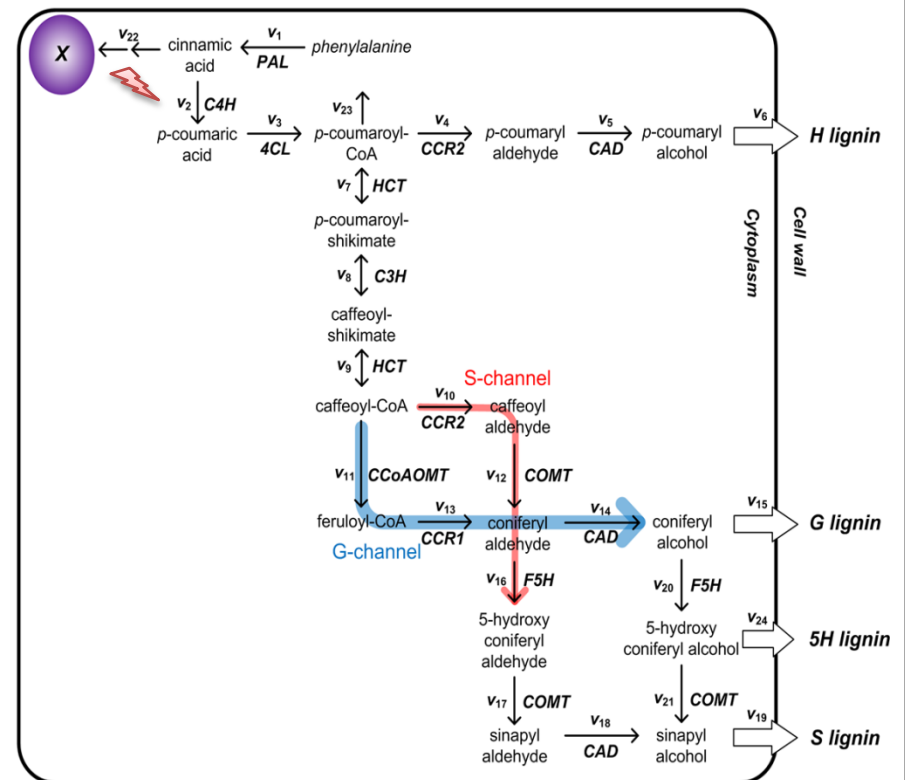**Puzzle: Down-regulation of PAL *vs.* C4H results in different S/G ratios**

Model-Based Signal Hypothesis

- **Down-regulation of PAL or C4H results in different S/G ratios**
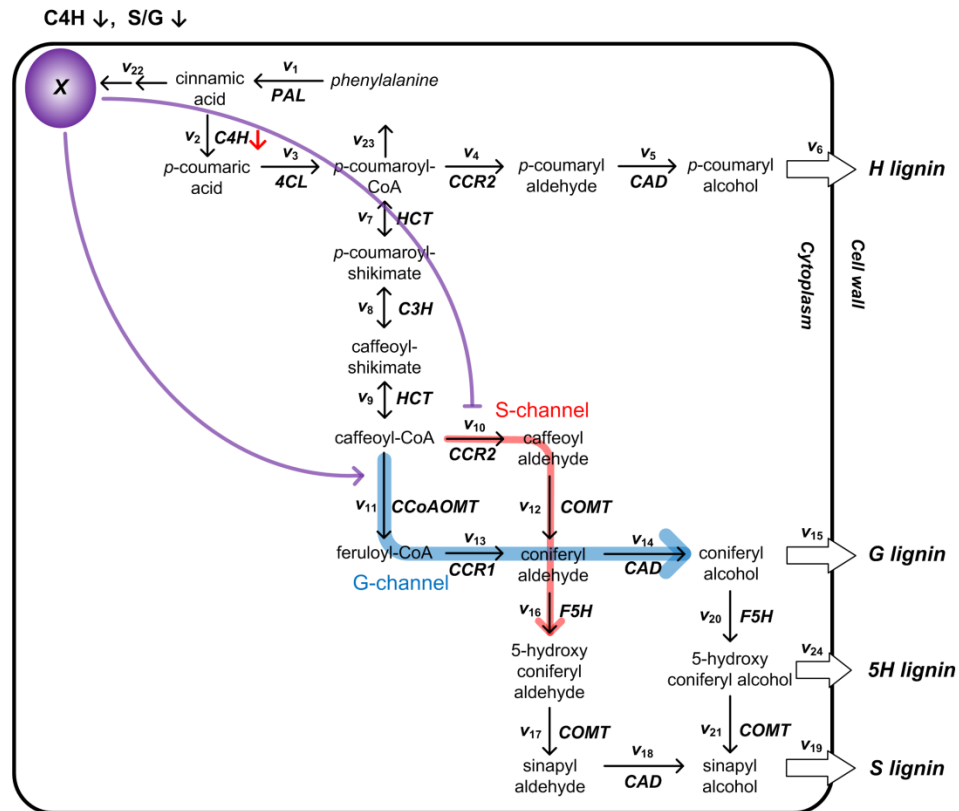

- **Hypothesis: "S-channel" and "G-channel" are affected**

Model-Based Signal Postulate

- **Down-regulation of PAL or C4H results in different S/G ratios**

- **Postulate:**

  Some cinnamic acid derivative X controls "S-channel" and "G-channel"

Next Steps

Experimental verification / refutation of postulates

Conversion of FBA/MOMA model into dynamic BST model

Optimization of BST model toward reduced S/G ratio

Experimental verification / refutation of knock-down
        combinations suggested by optimization

Execute similar analysis for switchgrass and other
        bioenergy crops

Summary

o   Pathways of lignin biosynthesis are not fully understood

o   Intuitive predictions are problematic because of multiple uses of the same enzymes and because of regulation

o   Modeling can add genuine value to experimental data

o   Kinetic pathway information in the literature is scarce

o   Gene modulation data are of tremendous benefit

o   Principles and methods shown here also apply to modeling of lignin degradation

o   We are very grateful for DOE-BESC support

www.bst.bme.gatech.edu