Metabolic flux estimation

- So far in this course we have examined techniques that help us understanding the cell's capabilities:
  - Given genome, what kind of metabolic network (Metabolic reconstruction)
  - Given metabolic network, what kind of behaviour is possible (Flux balance analysis, elementary flux modes)
- Now we turn to a different question: how to analyze quantitatively the activity of metabolic pathways
  - Given some measurements and the stoichiometry, estimate flux vector v

Flux estimation problem

- In flux estimation the goal is to restrict the space of solutions of the steady equations

$$S\mathbf{v} = 0$$

- Ideally, a single rate vector v is left as the solution
- In practise, we will need to resort to constraining the set of solutions in the null space $\mathcal{N}(S)$.

Why don't we just measure the fluxes?

There is currently no practical way to measure internal reaction rates *in vivo, in a living cell* in *quantitative* manner

- Enzyme kinetics studies the reaction rates of individual enzymes *in vitro, in a test tube*, isolated from the rest of metabolism. These rates are in general not the same as in a living cell.
- Microarrays and proteomics can give us estimates of concentration of mRNA or protein. However, these do not directly correlate with the reaction rates. These can be used to obtain qualitative estimates of pathway activity, but exact reaction rates $\mathbf{v}$ cannot be inferred.

Flux estimation strategies

- In flux estimation the goal is to restrict the space of solutions of the steady equations

$$S\mathbf{v} = 0$$

- General strategies:
  - Make biological assumptions ("cell is optimizing biomass growth", "this pathway is not active for reason XYZ")
  - Control some of the exchange fluxes via feeding cell culture certain nutrients at certain rate (e.g. glucose)
  - Measure some of the exchange fluxes (production and consumption of some key metabolites)
  - Isotope tracing experiments

The use of expression data

- Assume a reaction R_j with catalyzing genes $G_1, \ldots, G_k$
- If none of which is expressed, we can infer that the reaction is probably not active
- When solving our fluxes, we can set $v_j = 0$ when estimating the fluxes
- If any of the genes is active, the reaction might be active, but the reaction rate and (even the reaction direction) is hard to infer
- Hardness is due to a non-linear dependency between reaction rate and the enzyme, substrate and product concentrations.

In conclusion, expression data is best used in deducing *inactivity* of pathways

Handling known reaction rates

- Assume we know via measurement or via assumption the rates of reactions $R_{i_1}, \ldots, R_{i_k}$, $v_{i_1} = c_{i_1}, \ldots v_{i_k} = c_{i_k}$, given by the vector equation $\mathbf{v}_{known} = \mathbf{c}_{known}$
- Partition S in to unknown and known part, so that S_{known} (resp. $S_{unknown}$ contains the columns corresponding to reaction rates v_{known} (resp. $v_{unknown}$)
- The steady state equation is now given by

$$\begin{bmatrix} S_{unknown} & S_{known} \end{bmatrix} \cdot \begin{bmatrix} v_{unknown} \\ v_{known} \end{bmatrix} = 0$$

Handling known reaction rates

- Substitute the known rates $\mathbf{v}_{known} = \mathbf{c}_{known}$ into the steady state equation to obtain

$$S_{unknown} \cdot \mathbf{v}_{unknown} = \mathbf{d},$$

  where is a constant vector given by $\mathbf{d} = -S_{known}\mathbf{c}$
- By linear algebra, the complete set of solutions to the simplified steady state equation is given by

$$\mathbf{v}_{unknown} = S^{+}_{unknown}\mathbf{d} + K_{unknown}\mathbf{b},$$

- Above $S^{+} = (S^{T}S)^{-1}S^{T}$ is the pseudo-inverse of matrix S, obtained via command *pinv*() in MATLAB.

Handling known reaction rates

- By linear algebra, the complete set of solutions to the simplified steady state equation is given by

$$\mathbf{v}_{unknown} = S^{+}_{unknown}\mathbf{d} + K_{unknown}\mathbf{b},$$

- $K_{unknown}$ is the kernel matrix of the null space of $S_{unknown}$, and $\mathbf{b}$ is an arbitrary vector.
- Ideally, we would like the kernel to be empty matrix, as then we have fully determined the fluxes $\mathbf{v}_{unknown} = S^{+}_{unknown}\mathbf{d}$

Flux estimation and network topology

- The hardness of flux estimation depends on the metabolic network topology (structure)
- For simple topologies, it suffices to measure rates of the exchange reactions to fully determine fluxes
- Simple topologies include
  - Linear pathway
  - Tree shaped network

Flux estimation and linear pathways

- In a linear pathway, the rate of the exchange reaction $R_j :\Rightarrow M_i$ determines the in-flow towards M_i,
- The steady state requirement

$$\frac{dA}{dt} = s_{ij} v_j + s_{ij'} v_{j'} = 0$$

determines the rate of the sole consumer of M_i, reaction $R_{j'}$

$$v_{j'} = \frac{-s_{ij} b_j}{s_{ij'}}$$

- Following the same procedure, the rates of the linear pathway become fully determined

Flux estimation and tree-shaped topologies

- Utilizing the procedure for determining the reaction rates of a linear pathway generalize to a tree-shaped topology
- Follow linear pathways from the exchange reactions towards interior of the metabolic network. The fluxes will be determined by the above procedure
- The process stops at junctions where two or more linear pathways meet or diverge

Flux estimation and tree-shaped topologies

- We can always find a junction metabolite where we know $k-1$ of the k fluxes of the pathways around the metabolite (Why? Left as exercise).
- Using the $k-1$ known fluxes, we can determine the missing one
- After solving the unknown rate, follow the linear pathway until the next junction is met
- Repeating the linear pathway step and solving fluxes at junctions will eventually determin all the fluxes

General topologies

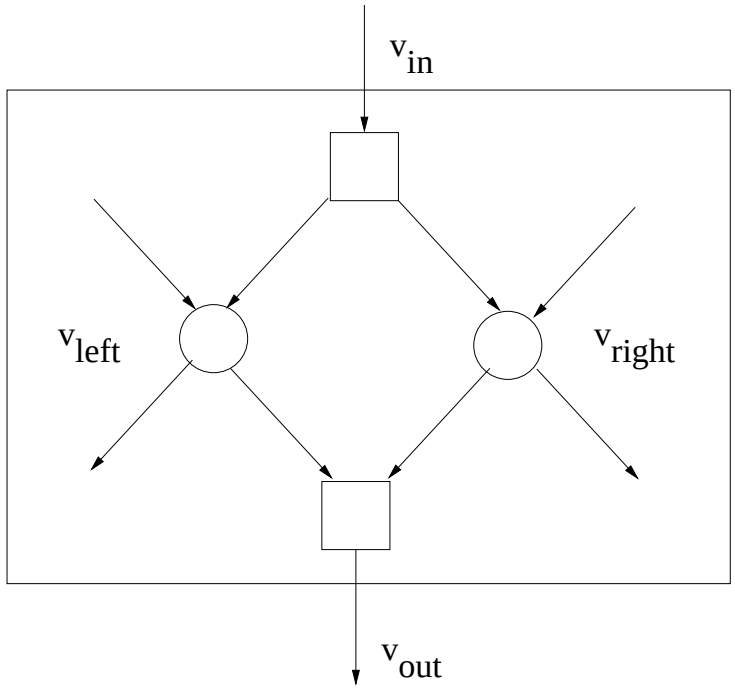
The above described process breaks down in many interesting cases:

- Alternative pathways between two metabolites
- Cycles
- Bi-directional reactions

The problem with alternative routes

- If there are alternative routes to produce some metabolite in the metabolic network, the relative activity of the routes cannot be pinpointed.
- In the example on the right, the fluxes v_{left} and v_{right} cannot be pinpointed just by measuring exchange fluxes, only their sum can be solved.

- In this case the null space of $S_{unknown}$ is non-empty, thus there is a choice of flux vectors that satisfy steady state



Material balance

$$v_{in} = v_{left} + v_{right} = v_{out}$$

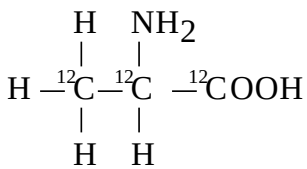
Isotope tracing experiments

- Isotope tracing experiments are the most accurate tool for estimating the fluxes of alternative pathways
- In isotope tracing experiments the cell culture is fed a mixture of natural and ^{13}C labeled substrate (e.g. 90%/10%).
- The fate of the ^{13}C labels is followed by measuring the intermediate metabolites by mass spectrometry or NMR
- From the enrichment of labels in the intermediates the fluxes are inferred

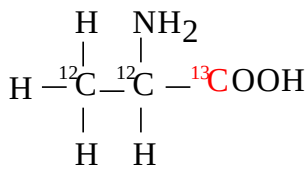
^{13}C-Isotopomers

- In isotope tracing experiments the cell culture is fed a mixture of natural and ^{13}C substrate (e.g. 90%/10%).
- This induces different kinds of ^{13}C labeling patterns, **isotopomers** (isotopic isomers):

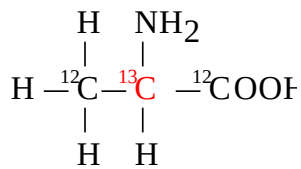
000Ala



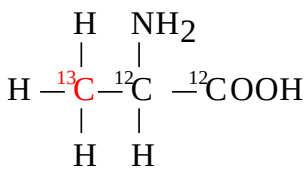
001Ala



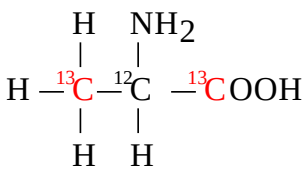
010Ala



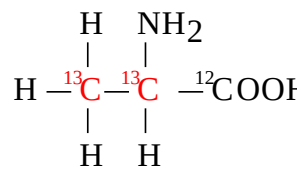
100Ala



101Ala

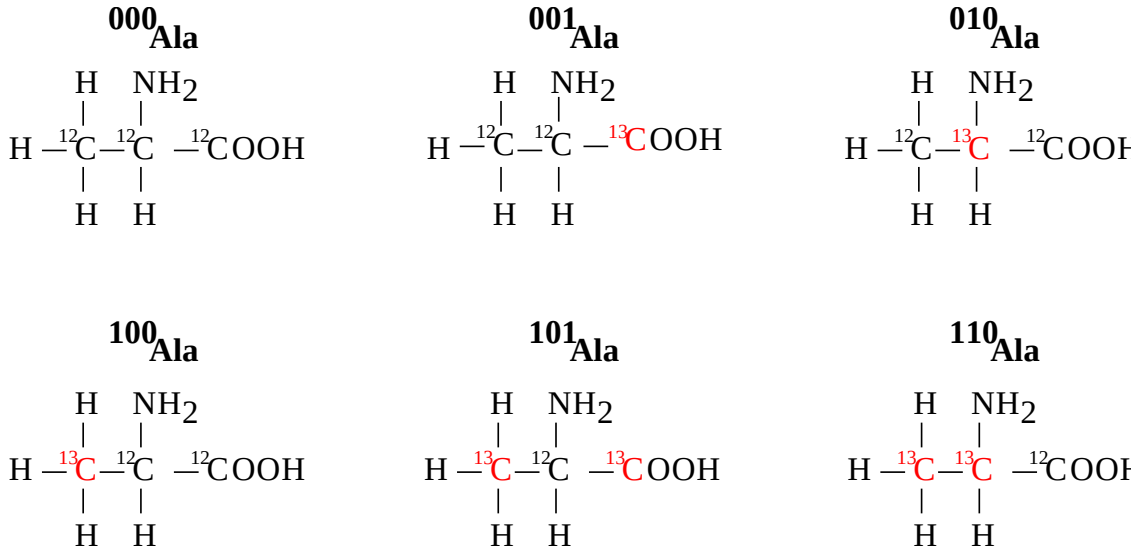


110Ala



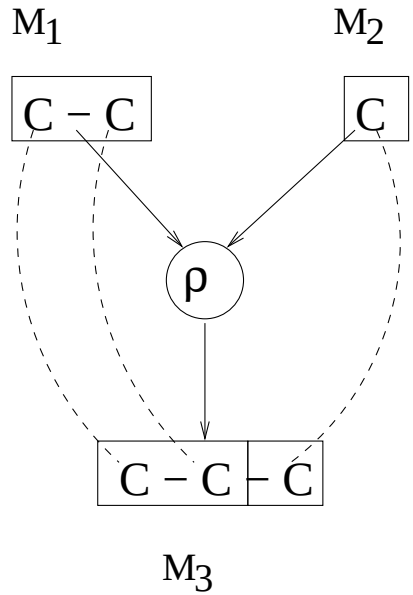
Isotopomers and alternative pathways

- The vector of relative frequencies of the isotopomers $\mathbb{I}_{Ala} = [\mathbb{P}\{^{000}Ala\}, \mathbb{P}\{^{001}Ala\}, \ldots, \mathbb{P}\{^{111}Ala\}] \in [0,1]^{2^3}$, is called an **isotopomer distribution**
- Isotopomer distributions can give information about the fluxes of alternative pathways *if* the pathways manipulate the carbon chains of the metabolites differently



Metabolite and reaction representation

- We treat metabolites as a set of unique named carbon locations.
- Metabolites M are further divided to fragments, ie. subsets of their carbons $F = M|F$.
- For each reaction a **carbon mapping** describing the transitions of carbon atoms in a reaction event is given.



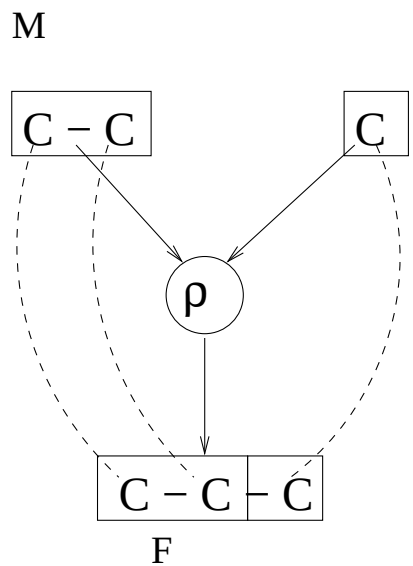
Isotopomers of a product metabolite

We make two assumptions

- Uniform sampling assumption: a reaction draws its reactants independently, uniformly randomly from the reactant pools
- No isotope effects assumption: the reaction does not make any difference between different isotopomer pools

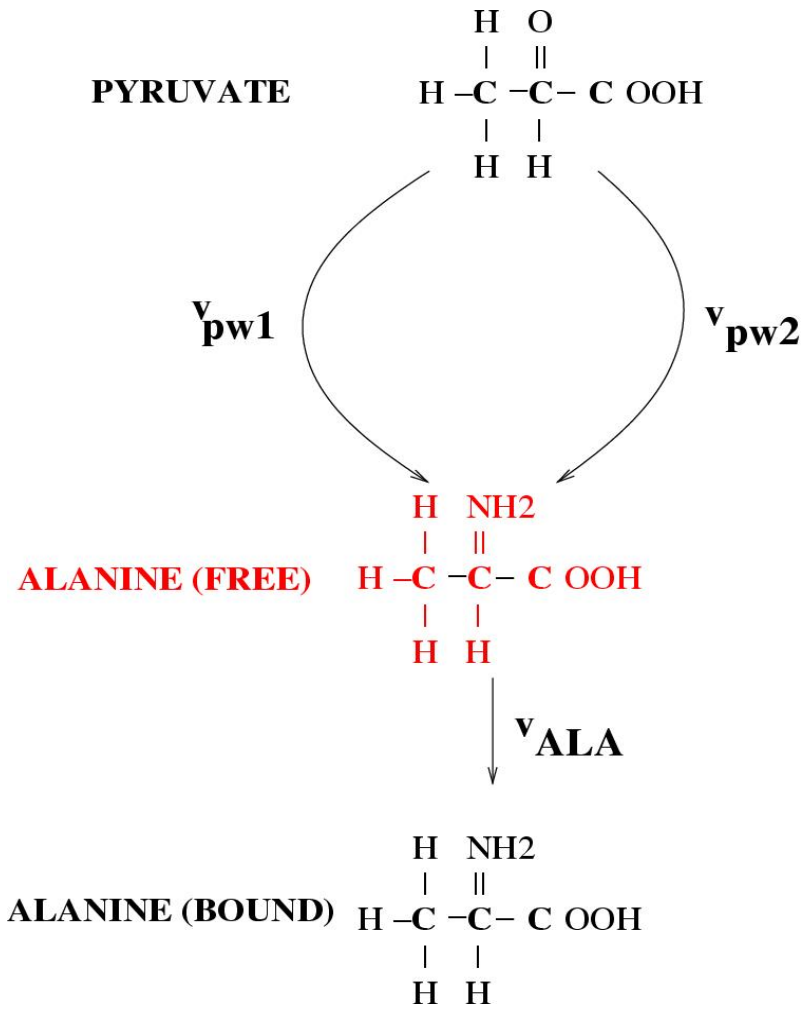
These assumption ensure

$$P(^{xyz}M_3) = P(^{xy}M_1)P(^{z}M_2)$$



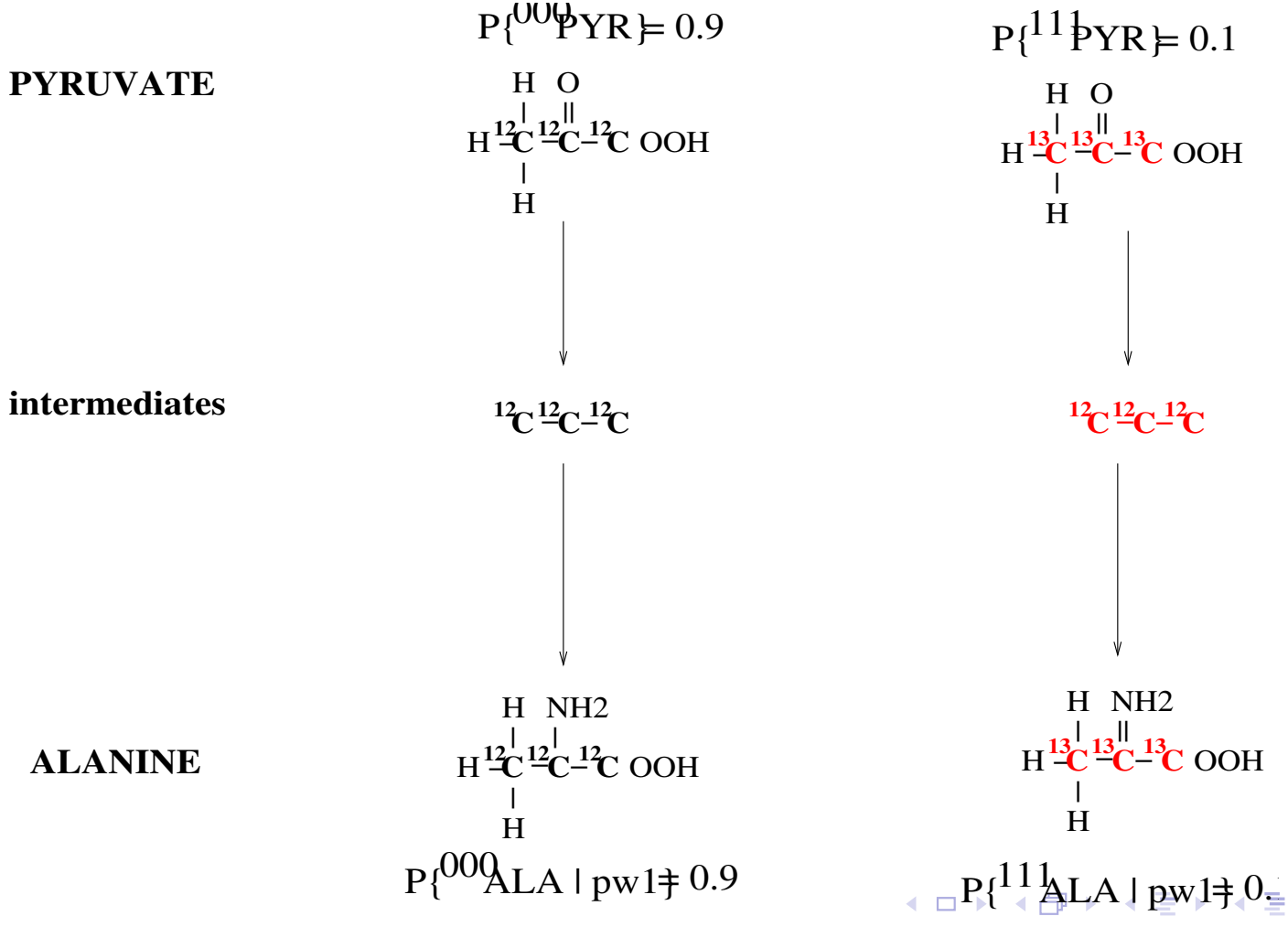
Example

- Left-hand pathway keeps the carbon chain of puryvate intact
- Right-hand pathway cleaves the carbon chain between 2. and 3. carbon
- We have measure the isotopomer distributions of pyruvate and free alanine
  $\mathbb{P}\{^{000}Pyr\} = 0.9, \mathbb{P}\{^{111}Pyr\} = 0.1;$
  $\mathbb{P}\{^{000}Ala\} = 0.855, \mathbb{P}\{^{001}Ala\} = 0.045, \mathbb{P}\{^{110}Ala\} = 0.045, \mathbb{P}\{^{111}Ala\} = 0.055$
- Let us determine $\mathbb{P}\{^{xyz}Ala|pw1\}$ and $\mathbb{P}\{^{xyz}Ala|pw2\}$



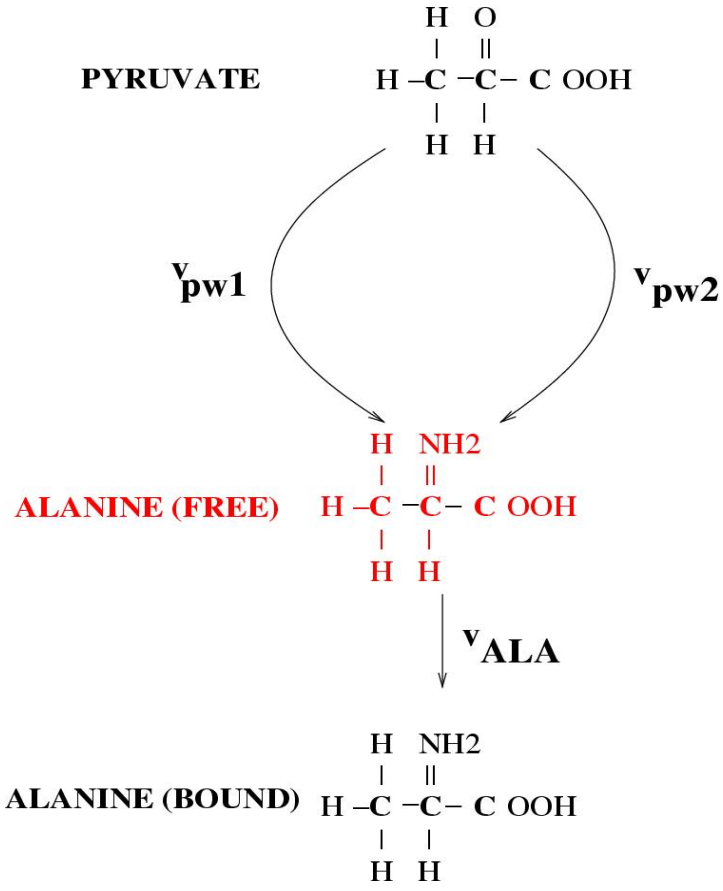
Left-hand pathway

- The left-hand pathway transfer the carbon chains of puryvate intact to alanine
- Labeling patterns and the isotopomer distributions do not change

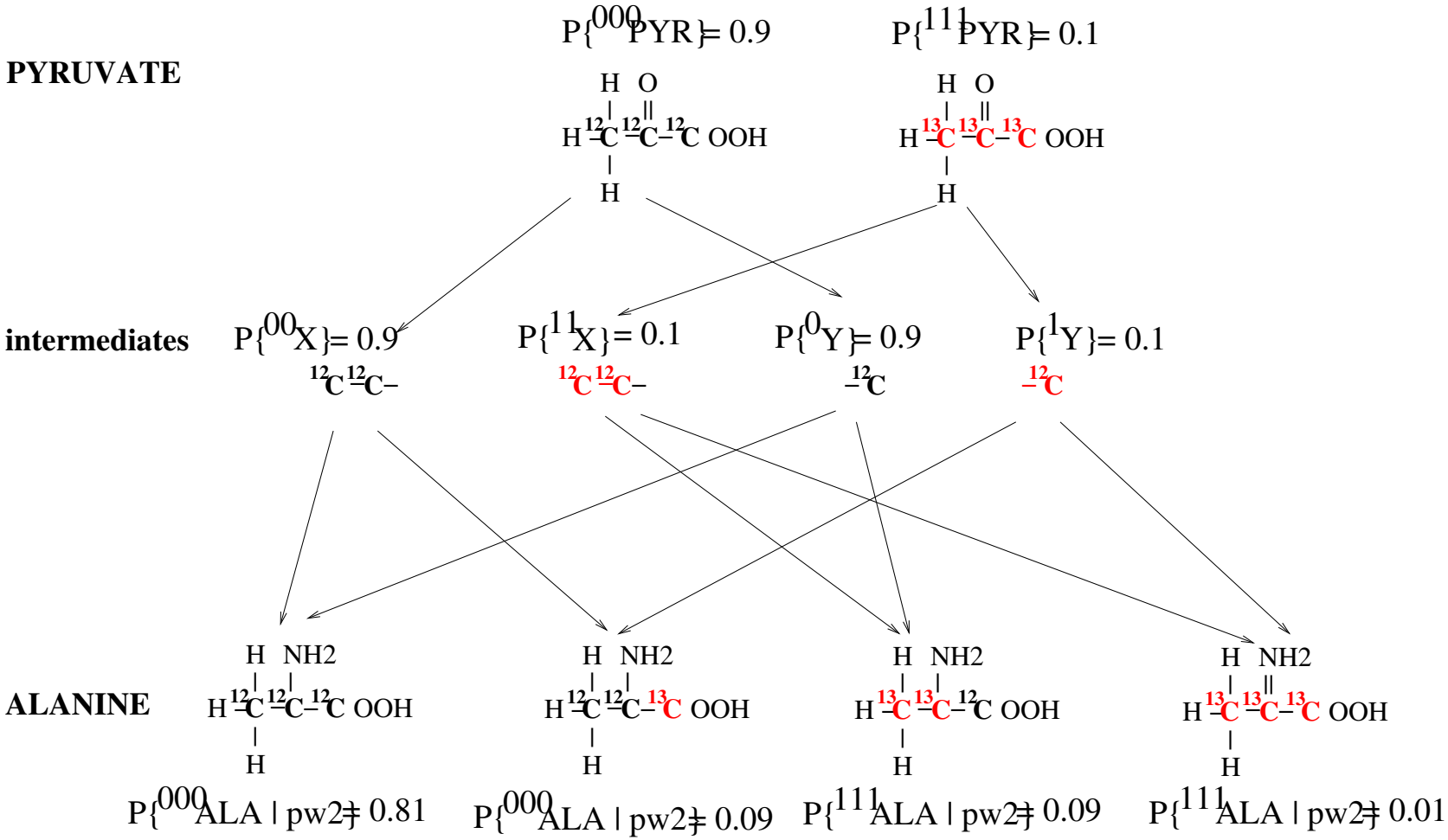


Right-hand pathway

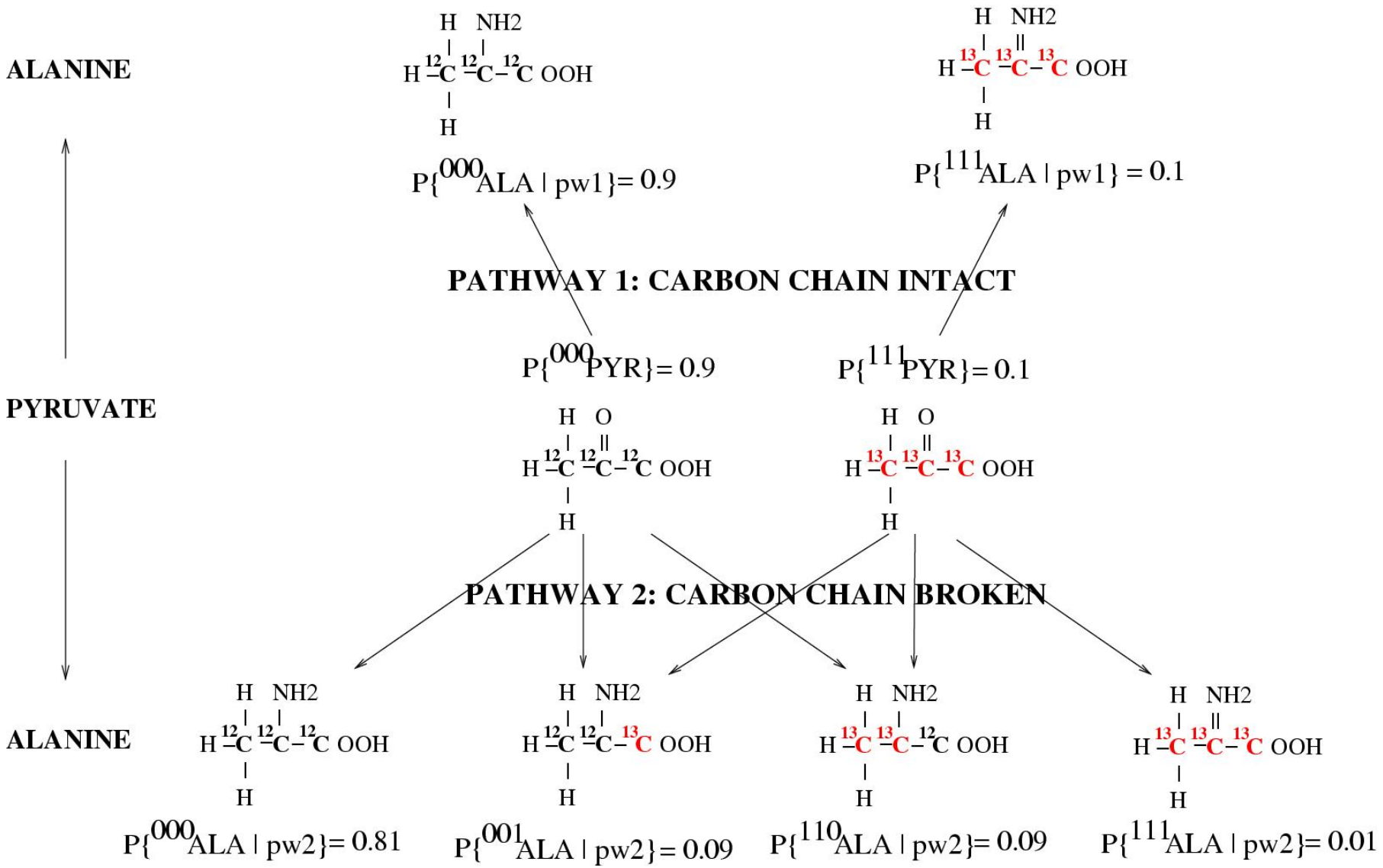
- The right hand pathway clevaes and recombines the carbon chain of puryvate between 2. and 3. carbon
- The fragments X and Y to be recombined are assumed to sampled independently, randomly according to their isotopomer distributions
- The isotopomer frequencies are obtained by multiplying the isotopomer frequencies of the fragments
  $\mathbb{P}\{^{xyz}ALA\} = \mathbb{P}\{^{xy}X\}\mathbb{P}\{^{z}Y\}$



Right-hand pathway



Isotopomer distributions & alternative pathways

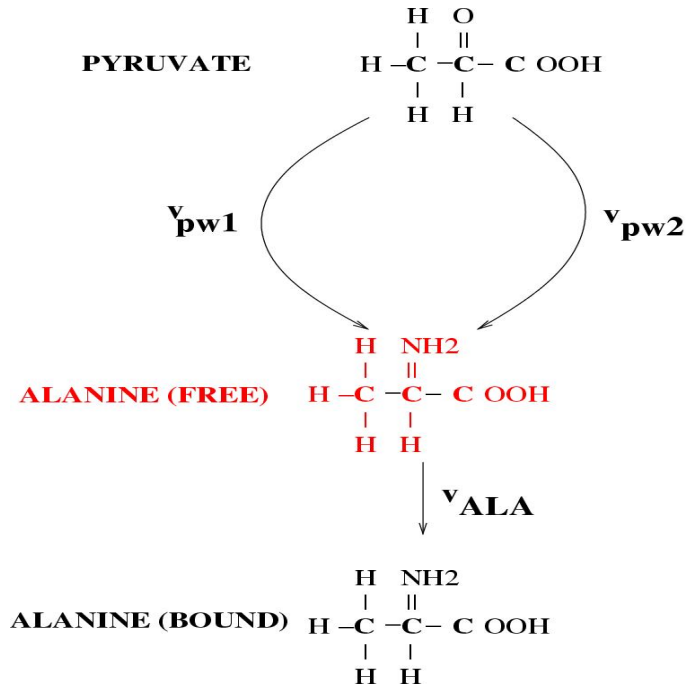


Isotopomeric balance equations

- The steady state condition for free alanine implies:

$$v_{pw1} + v_{pw2} = v_{ALA}$$

- The steady state assumption needs to hold for each isotopomer separately
- We can write balance equations for each isotopomer:



$$P(^{000}ALA|pw1) \cdot v_{pw1} + P(^{000}ALA|pw2) \cdot v_{pw2} = P(^{000}ALA) \cdot v_{ALA}$$
$$P(^{001}ALA|pw1) \cdot v_{pw1} + P(^{001}ALA|pw2) \cdot v_{pw2} = P(^{001}ALA) \cdot v_{ALA}$$
$$P(^{110}ALA|pw1) \cdot v_{pw1} + P(^{110}ALA|pw2) \cdot v_{pw2} = P(^{110}ALA) \cdot v_{ALA}$$
$$P(^{111}ALA|pw1) \cdot v_{pw1} + P(^{111}ALA|pw2) \cdot v_{pw2} = P(^{111}ALA) \cdot v_{ALA}$$

Flux estimation from incomplete isotopomer data

In practice, we are faced with incomplete isotopomer data:

- Not all isotopomer distributions can be measured, due too sensitivity issues of measuring equipment.
- Complete isotopomer distributions can only rerely be measured:
  - MS data groups isotopomers of the same weight:

$$aP(^{010}ALA) + bP(^{100}ALA) = d$$

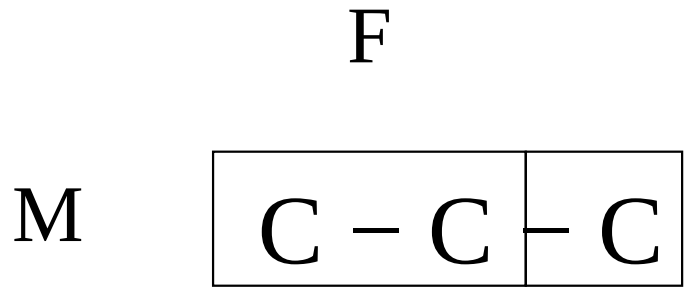
  - NMR measurements require ^{13}C in a specific position e.g. the middle carbon in alanine

$$\frac{P(^{010}ALA)}{\sum_{x1y} P(^{x1y}ALA)} = d.$$

We start by tackling the first difficulty.

Fragment marginals of isotopomer distibutions

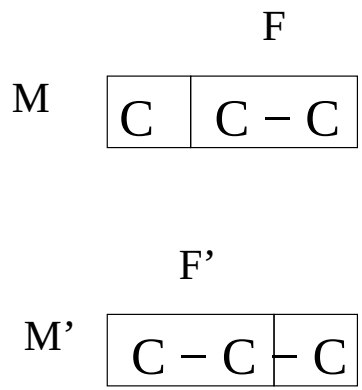
Each fragment $F \subseteq M$ corresponds to a *marginal distribution* to the isotopomer distribution:



$$P(^{00}F) = P(^{000}M_3) + P(^{001}M_3)$$
$$P(^{01}F) = P(^{010}M_3) + P(^{011}M_3)$$
$$P(^{10}F) = P(^{100}M_3) + P(^{101}M_3)$$
$$P(^{11}F) = P(^{110}M_3) + P(^{111}M_3)$$

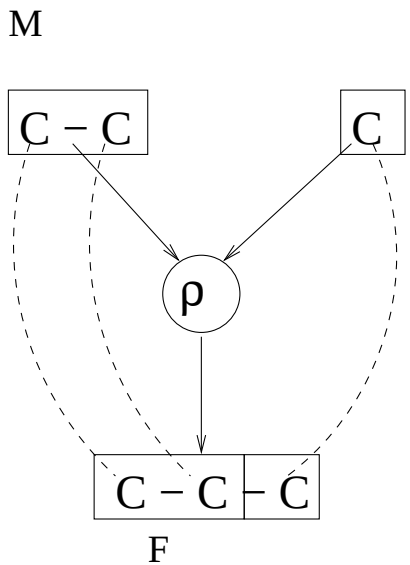
Fragment equivalence

- Two fragments $F \subseteq M$ and $F' \subseteq M'$ are *equivalent* if the fragment marginal distributions of the respective isotopomer distributions of M and M' are equal, irrespectively of the fluxes of the metabolic network
- When does the fragment equivalence hold true?



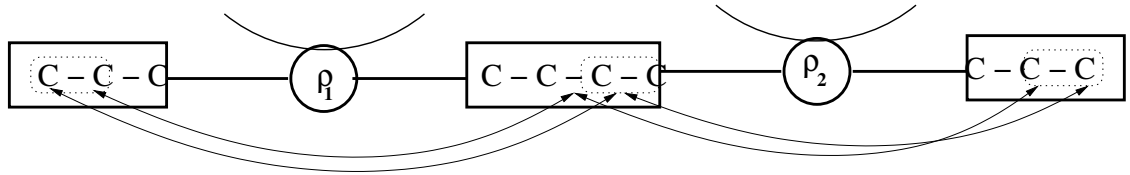
Fragment equivalence for a single reaction

In a single reaction, the reactant and product fragments are equivalent: $P(^{xy}F) = P(^{xy}M_1), x, y = 0, 1$



Fragment equivalance for unbranched pathways

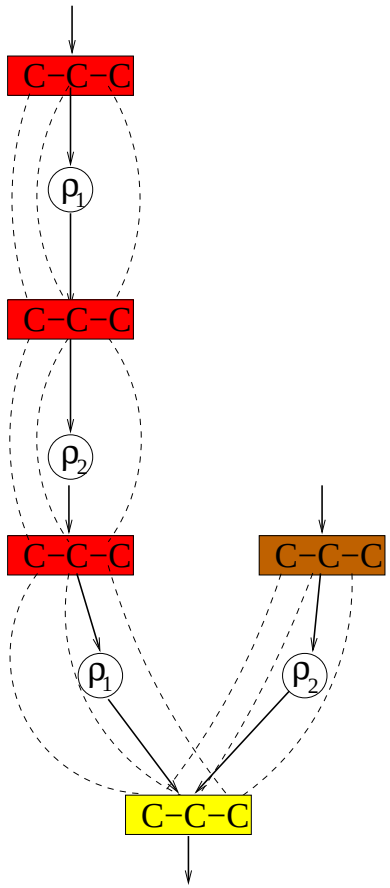
- Transitively, an unbranched pathway defines a one-to-one mapping between carbon atom locations

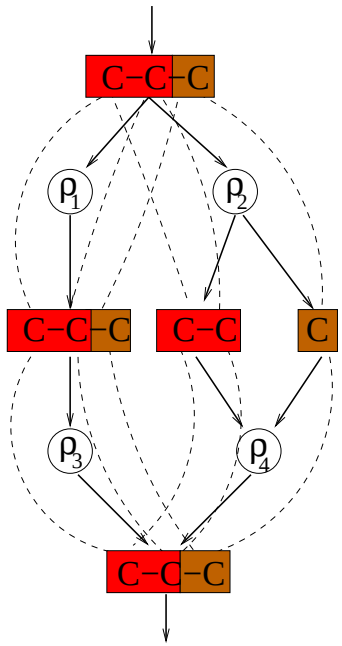


- Given a reactant M and product M' of the pathway, and two fragments $F \subseteq M$ and $F' \subseteq M'$, assume that $F' = \Lambda(F)$ is the image of the fragment F under the atom mapping Λ of the pathway.
- Then F and F' are equivalent

Fragment equivalence in general

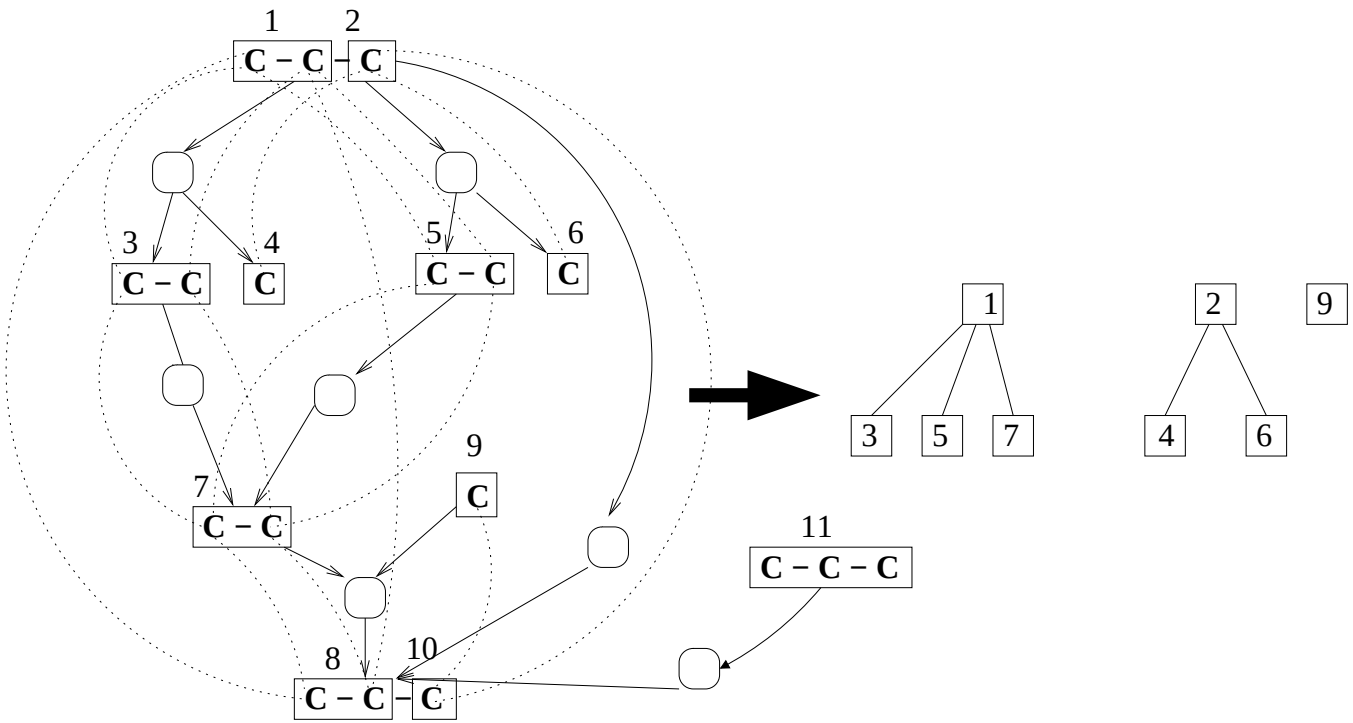
- Assume fragments produced by alternative pathways travel *intact* and *similarly oriented* (i.e. no permutation) starting from the common source fragment
- The isotopomer distribution of that fragment remain equivalent to the source along the alternative pathways





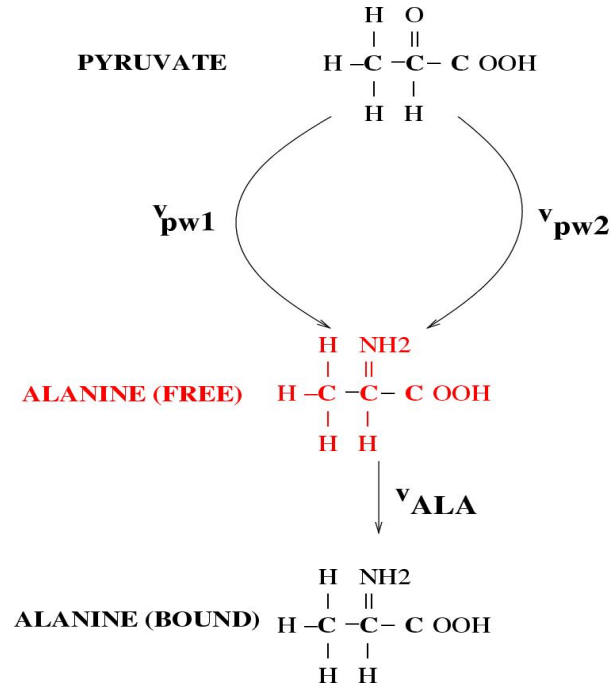
Equivalence classes

- The equivalance relation for fragments induces equivalence classes of fragments to the metabolic networks
- The isotopomer distribution is the theoretically the same for the whole equivalence class



Balance equations for fragments

- Assume we have deduced fragment marginals of ALA_{12} for both pathways
- Balance equations for the fragment ALA_{12}:



$$P(^{00}ALA_{12}|pw1) \cdot v_{pw1} + P(^{00}ALA_{12}|pw2) \cdot v_{pw2} = P(^{00}ALA_{12}) \cdot v_{ALA_{12}}$$

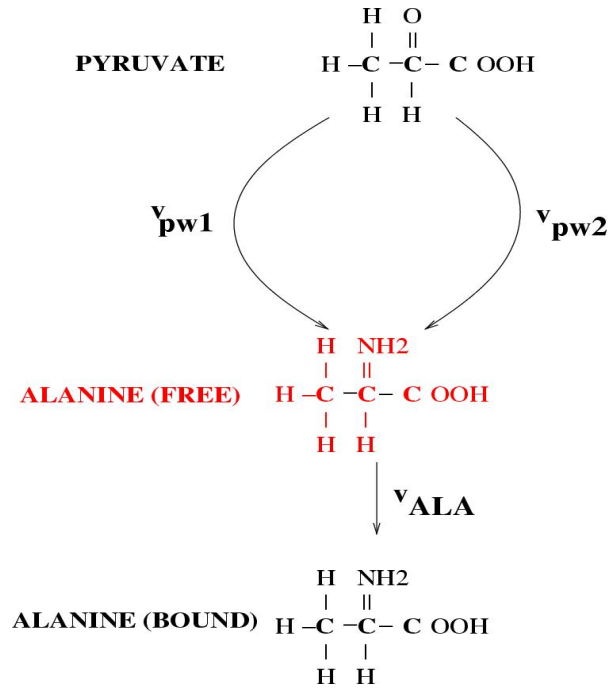
$$P(^{01}ALA_{12}|pw1) \cdot v_{pw1} + P(^{01}ALA_{12}|pw2) \cdot v_{pw2} = P(^{01}ALA_{12}) \cdot v_{ALA_{12}}$$

$$P(^{10}ALA_{12}|pw1) \cdot v_{pw1} + P(^{10}ALA_{12}|pw2) \cdot v_{pw2} = P(^{10}ALA_{12}) \cdot v_{ALA_{12}}$$

$$P(^{11}ALA_{12}|pw1) \cdot v_{pw1} + P(^{11}ALA_{12}|pw2) \cdot v_{pw2} = P(^{11}ALA_{12}) \cdot v_{ALA_{12}}$$

Balance equations for fragments

- In order the balance equations to be useful, it is required that ALA_{12} is not equivalent with the fragments $ALA_{12}|pw_1$ or $ALA_{12}|pw_2$ produced by the two pathways
- What will happen if this assumption is not satisfied? Left as exercise.



$$P(^{00}ALA_{12}|pw1) \cdot v_{pw1} + P(^{00}ALA_{12}|pw2) \cdot v_{pw2} = P(^{00}ALA_{12}) \cdot v_{ALA_{12}}$$

$$P(^{01}ALA_{12}|pw1) \cdot v_{pw1} + P(^{01}ALA_{12}|pw2) \cdot v_{pw2} = P(^{01}ALA_{12}) \cdot v_{ALA_{12}}$$

$$P(^{10}ALA_{12}|pw1) \cdot v_{pw1} + P(^{10}ALA_{12}|pw2) \cdot v_{pw2} = P(^{10}ALA_{12}) \cdot v_{ALA_{12}}$$

$$P(^{11}ALA_{12}|pw1) \cdot v_{pw1} + P(^{11}ALA_{12}|pw2) \cdot v_{pw2} = P(^{11}ALA_{12}) \cdot v_{ALA_{12}}$$