Carbon Tracing in Chemical Reaction Networks

Jonas Nyrup

University of Southern Denmark

Outline

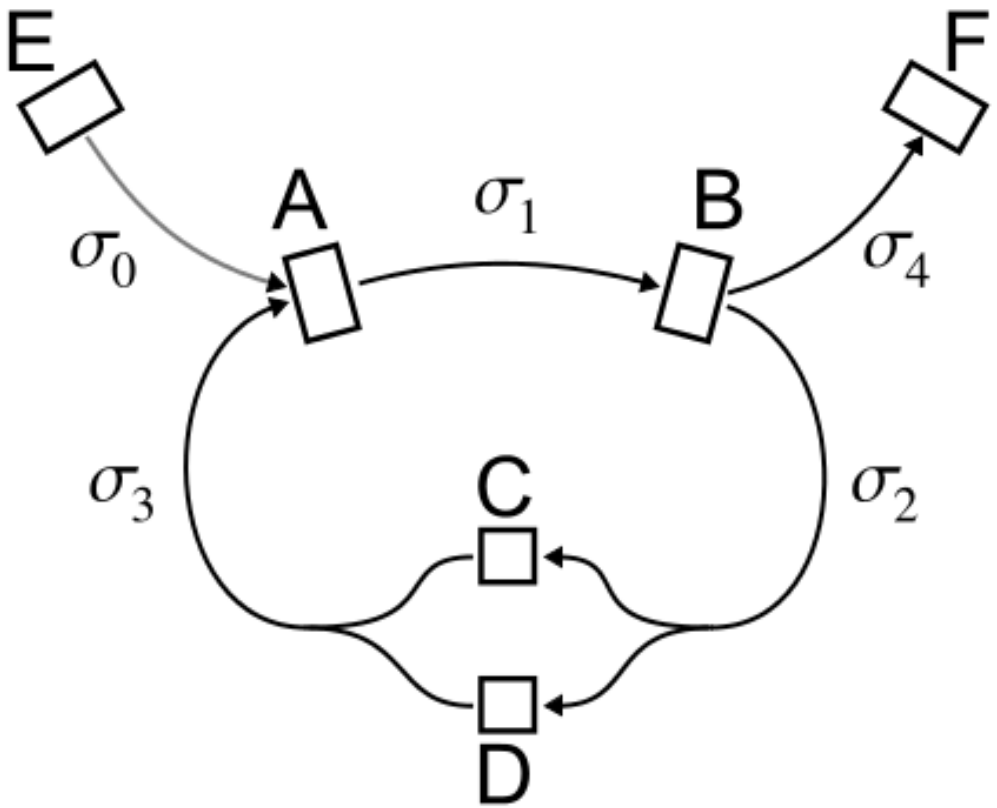
My thesis

- Where can a specific carbon in substrates end up?
- Where could a specific carbon in products come from?
- Complexity questions

Extensions:

- For k specific carbons...
- Expand to probabilistic likelihood model
- Substrates with labeling distribution

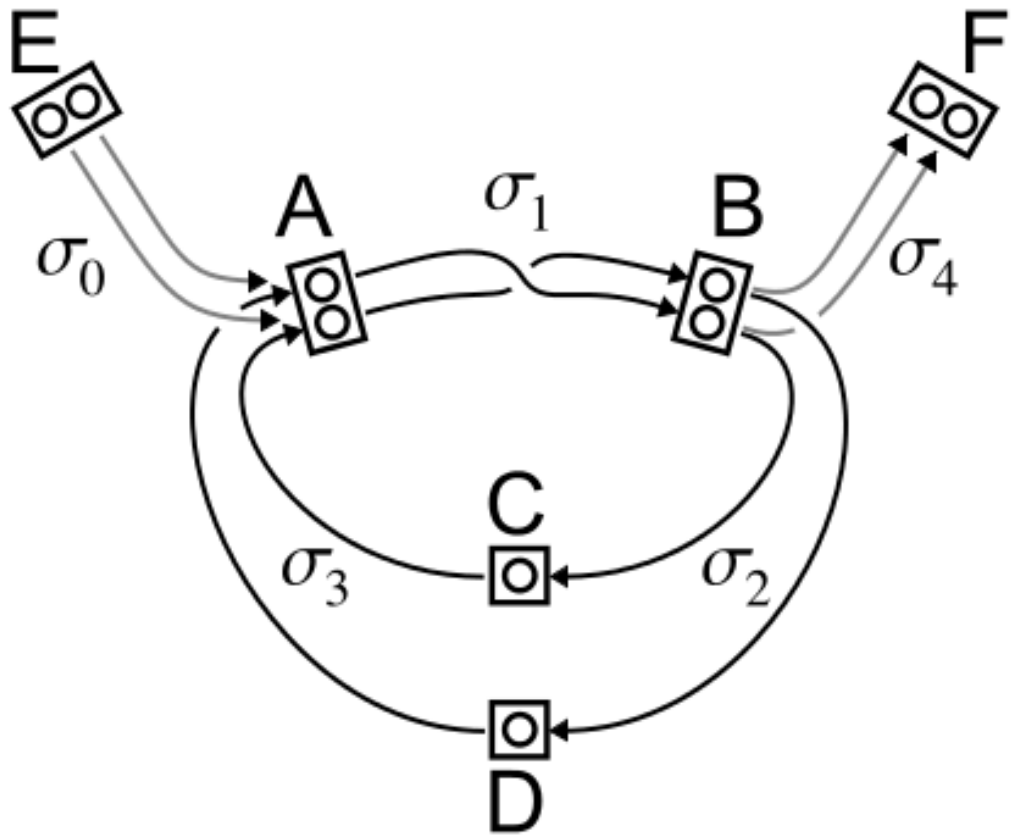
Reaction network



Weitzel, Wiechert, and Nöh, 2007

Atom transition network

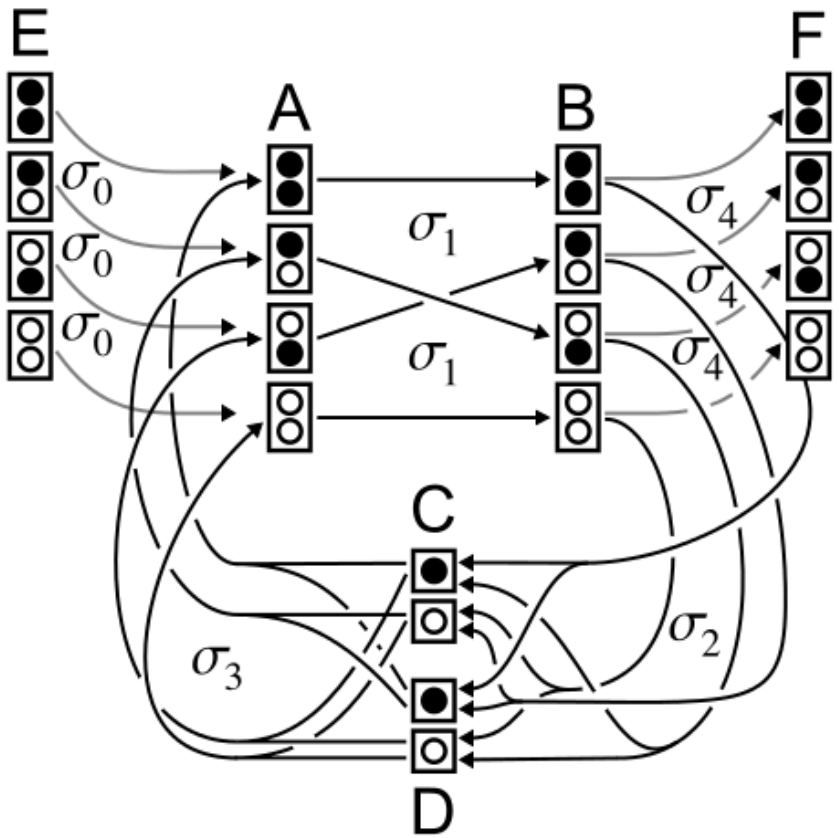
"What is the probability that a carbon is labeled?"



Weitzel, Wiechert, and Nöh, 2007

Isotopomer network

"What is the probability that a subset of carbons are labeled?"



Weitzel, Wiechert, and Nöh, 2007

Why not just use isotopomer model?

- In the atom model the amount of information per molecule is n, where n is the carbon number.

Why not just use isotopomer model?

- In the atom model the amount of information per molecule is n, where n is the carbon number.
- In the isotopomer model this number explodes into 2^n.

Why not just use isotopomer model?

- In the atom model the amount of information per molecule is n, where n is the carbon number.
- In the isotopomer model this number explodes into 2^n.

Limitations of the atom labeling model

From isotopomer distribution to positional label enrichment:

$$A = (\circ\circ, \bullet\circ, \circ\bullet, \bullet\bullet) = (60\,\%, 25\,\%, 15\,\%, 0\,\%)$$

$$\Downarrow$$

$$A_1 = \bullet\circ + \bullet\bullet = 25\,\%$$

$$A_2 = \circ\bullet + \bullet\bullet = 15\,\%$$

Limitations of the atom labeling model

$$A_1 = 25\,\%$$
$$A_2 = 15\,\%$$

Now what is the probability of ●●?

Limitations of the atom labeling model

$$A_1 = 25\,\%$$
$$A_2 = 15\,\%$$

Now what is the probability of ●●?

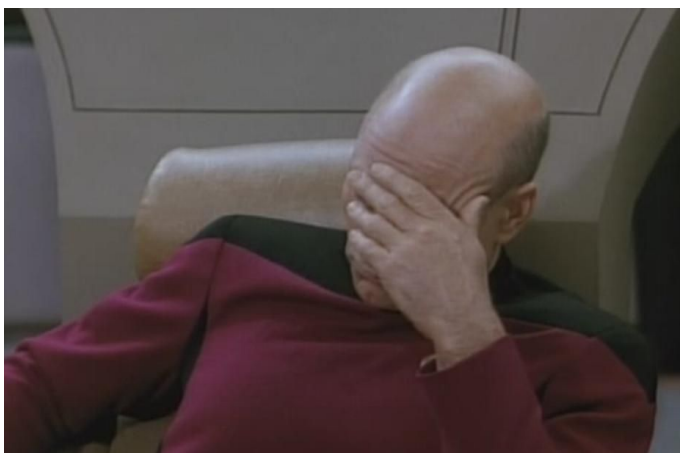
$$A_1 \cdot A_2 = 3.75\,\%$$

Limitations of the atom labeling model

$$A_1 = 25\,\%$$
$$A_2 = 15\,\%$$

Now what is the probability of ❶❸?

~~$A_1 \cdot A_2 = 3.75\,\%$~~



Limitations of the atom labeling model

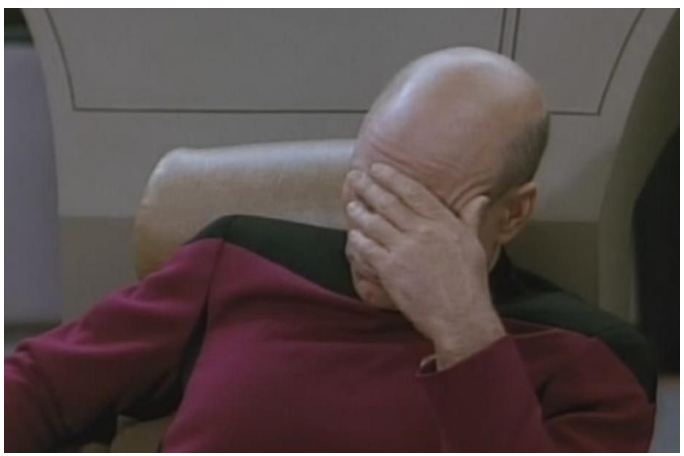
$$A_1 = 25\,\%$$
$$A_2 = 15\,\%$$

Now what is the probability of ●?

~~$A_1 \cdot A_2 = 3.75\,\%$~~

A_1 and A_2 are not independent.

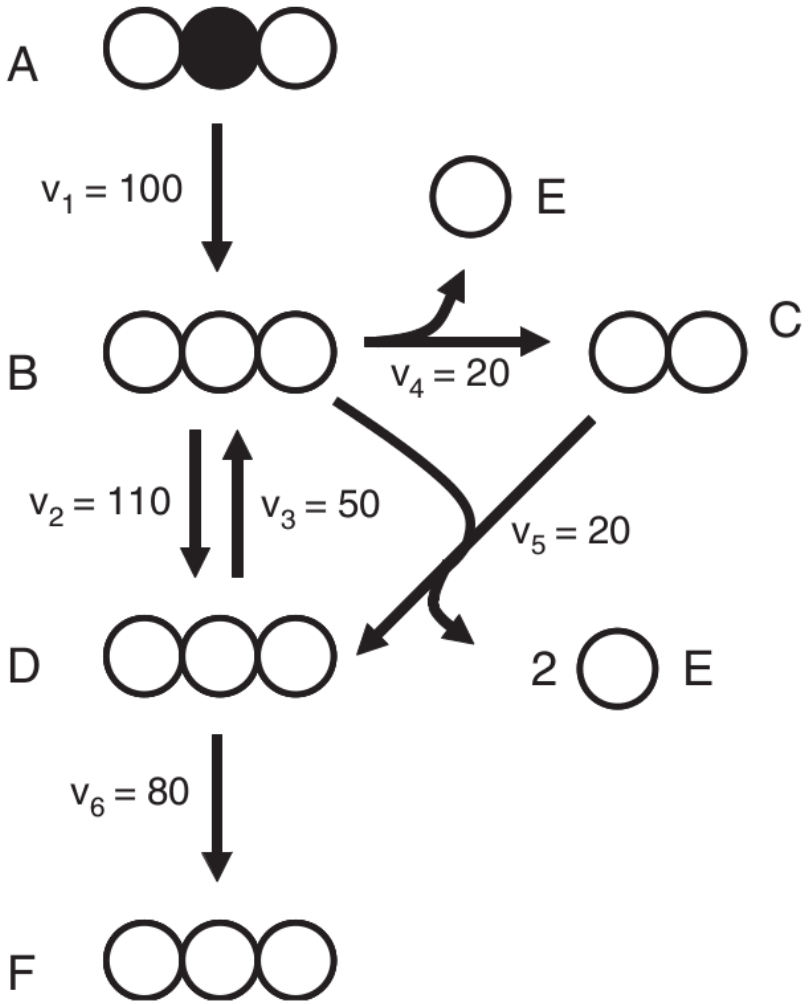
Information is lost!



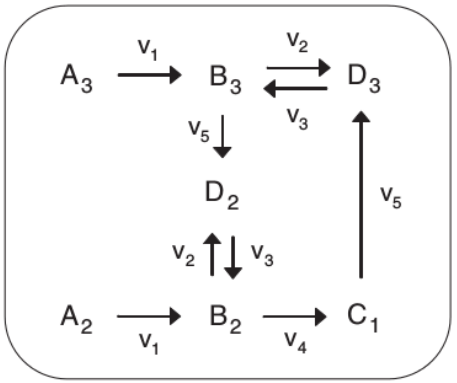
EMU Model

Here, we present a novel framework for the modeling of isotopic labeling systems that significantly reduces the number of system variables without any loss of information... We define an EMU of a compound to be a moiety comprising any distinct subset of the compound's atoms.

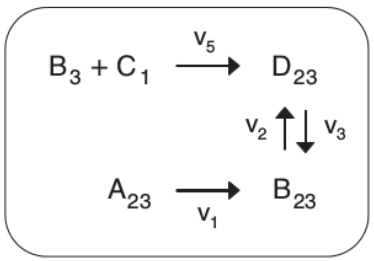
- The isotopomer model usually has $\sim$10x more information than needed.
- Find the minimal amount of information needed to simulate isotopic labeling.



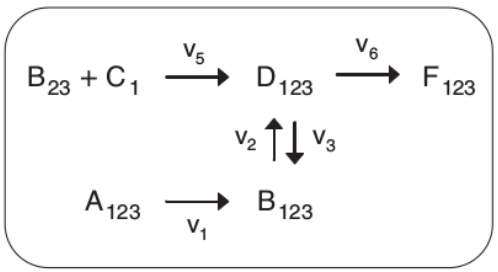
Reaction network for EMU size 1



Reaction network for EMU size 2



Reaction network for EMU size 3



Antoniewicz, Kelleher, and Stephanopoulos, 2007

Stoichiometry matrices for steady state

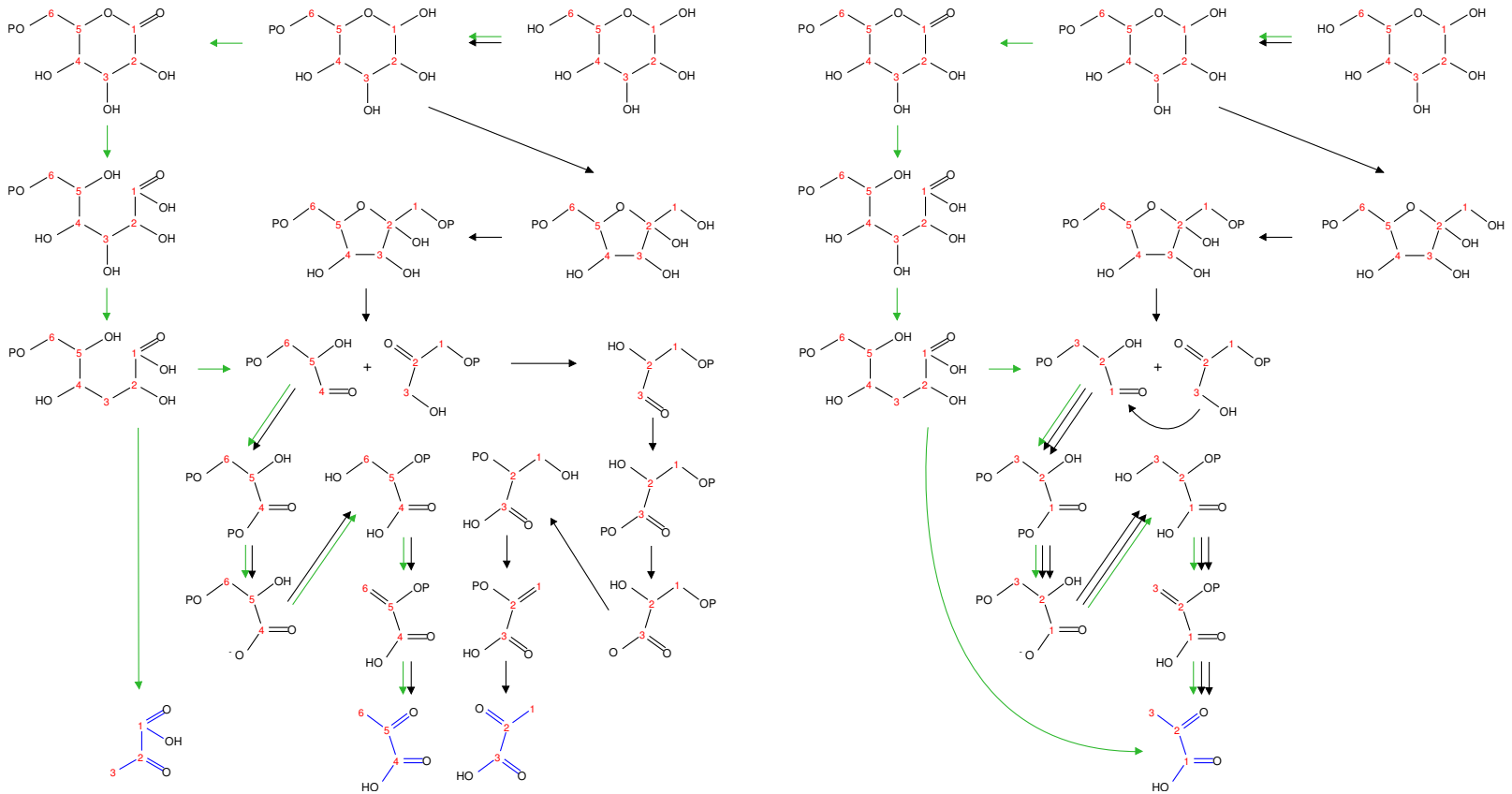
$$\begin{bmatrix} -v_4 & v_4 & 0 & 0 & 0 \\ 0 & -v_1-v_3 & v_3 & 0 & 0 \\ 0 & v_2 & -v_2-v_5 & v_5 & 0 \\ 0 & 0 & 0 & -v_1-v_3 & v_3 \\ v_5 & 0 & 0 & v_2 & -v_2-v_5 \end{bmatrix} \begin{bmatrix} C_1 \\ B_2 \\ D_2 \\ B_3 \\ D_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -v_1 & 0 \\ 0 & 0 \\ 0 & -v_1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} A_2 \\ A_3 \end{bmatrix}$$

$$\begin{bmatrix} -v_5-v_2 & v_2 \\ v_3 & -v_1-v_3 \end{bmatrix} \begin{bmatrix} D_{23} \\ B_{23} \end{bmatrix} = \begin{bmatrix} -v_5 & 0 \\ 0 & -v_1 \end{bmatrix} \begin{bmatrix} B_3 \times C_1 \\ A_{23} \end{bmatrix}$$

$$\begin{bmatrix} -v_6 & v_6 & 0 \\ 0 & -v_5-v_2 & v_2 \\ 0 & v_3 & -v_1-v_3 \end{bmatrix} \begin{bmatrix} F_{123} \\ D_{123} \\ B_{123} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -v_5 & 0 \\ 0 & -v_1 \end{bmatrix} \begin{bmatrix} B_{23} \times C_1 \\ A_{123} \end{bmatrix}$$

EXAMPLE

Limitations of the isotopomer model



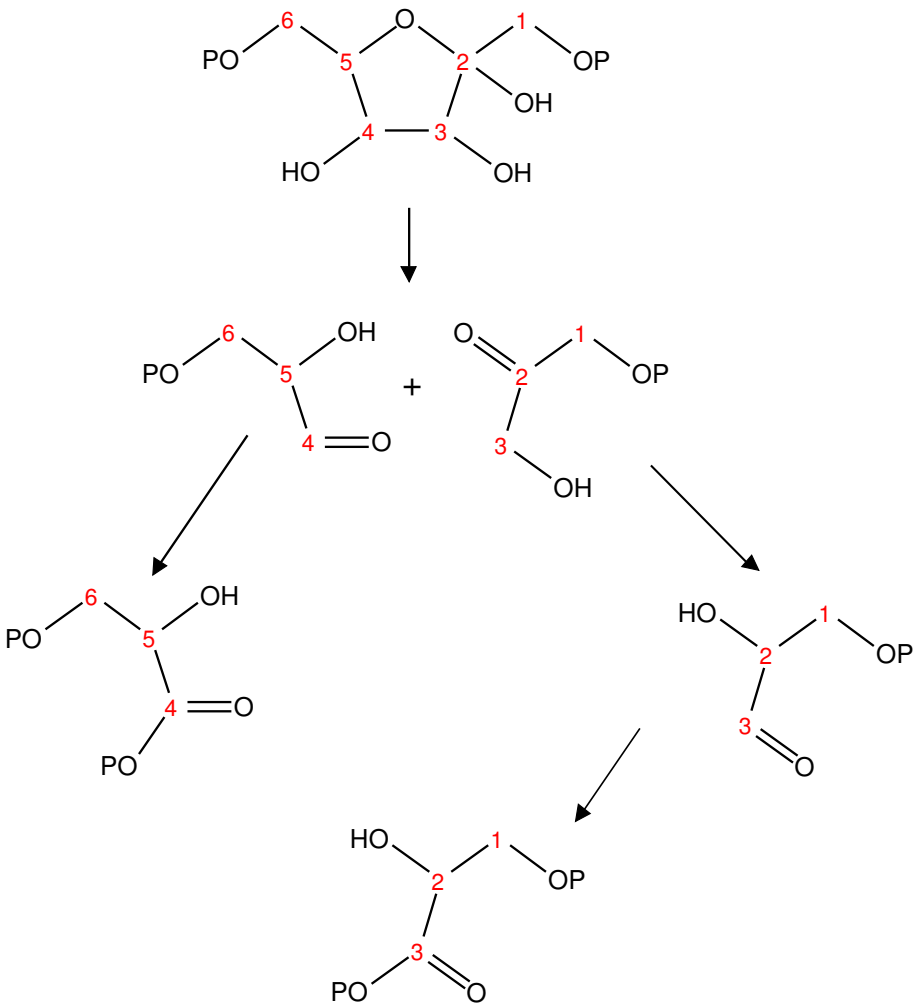
= ED

= EMP

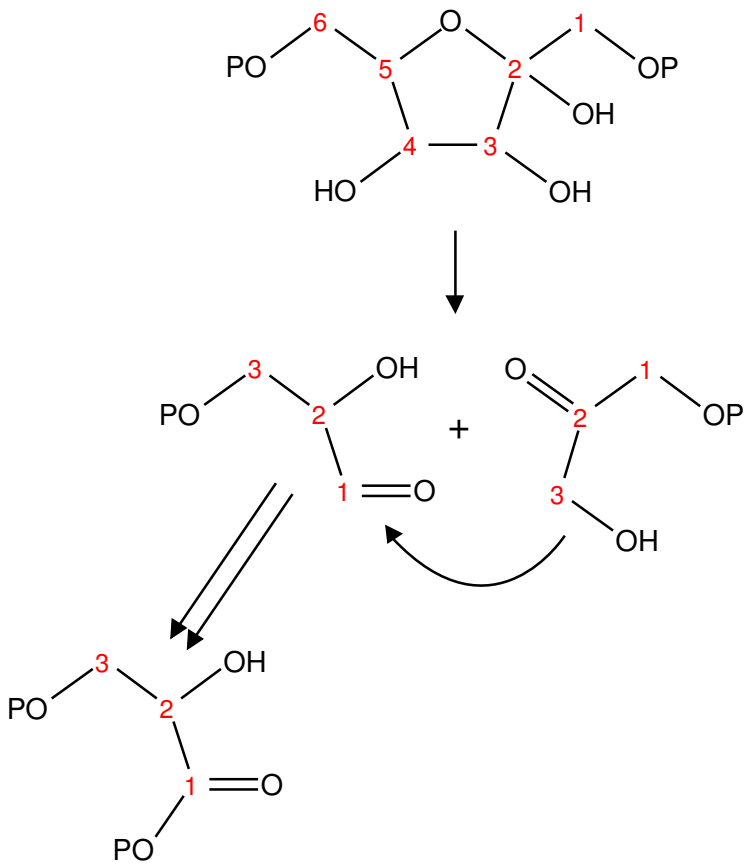
Andersen et al., 2014

Limitations of the isotopomer model

Ideal Model



EMU Model



Limitations of the isotopomer model

The fundamental problem

$$A \rightarrow B, B \qquad \Rightarrow \qquad 2B_1$$
$$ab \rightarrow a, b$$

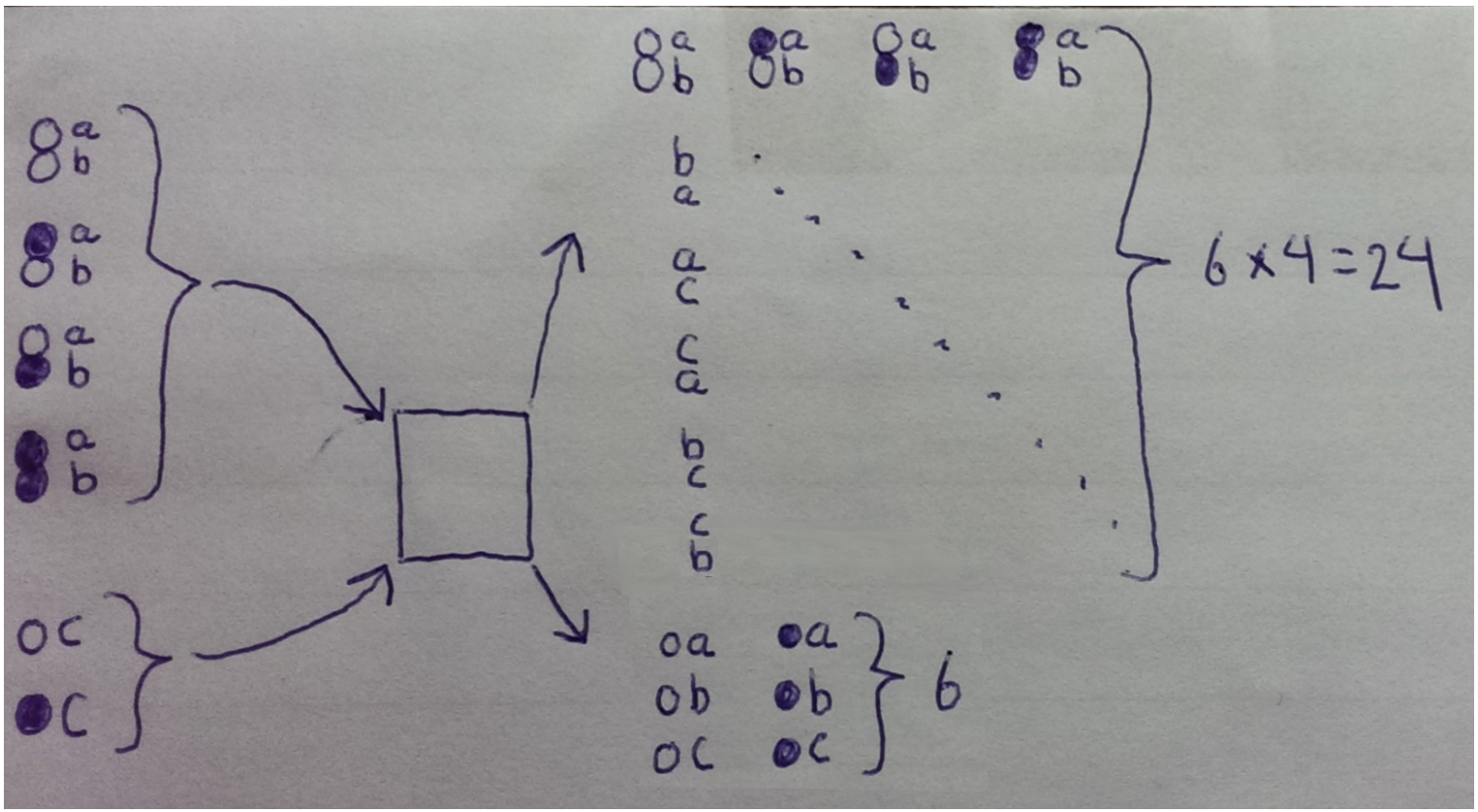
When backtracking from a B_1, does it origin from A_1 or A_2?

Limitations of the isotopomer model

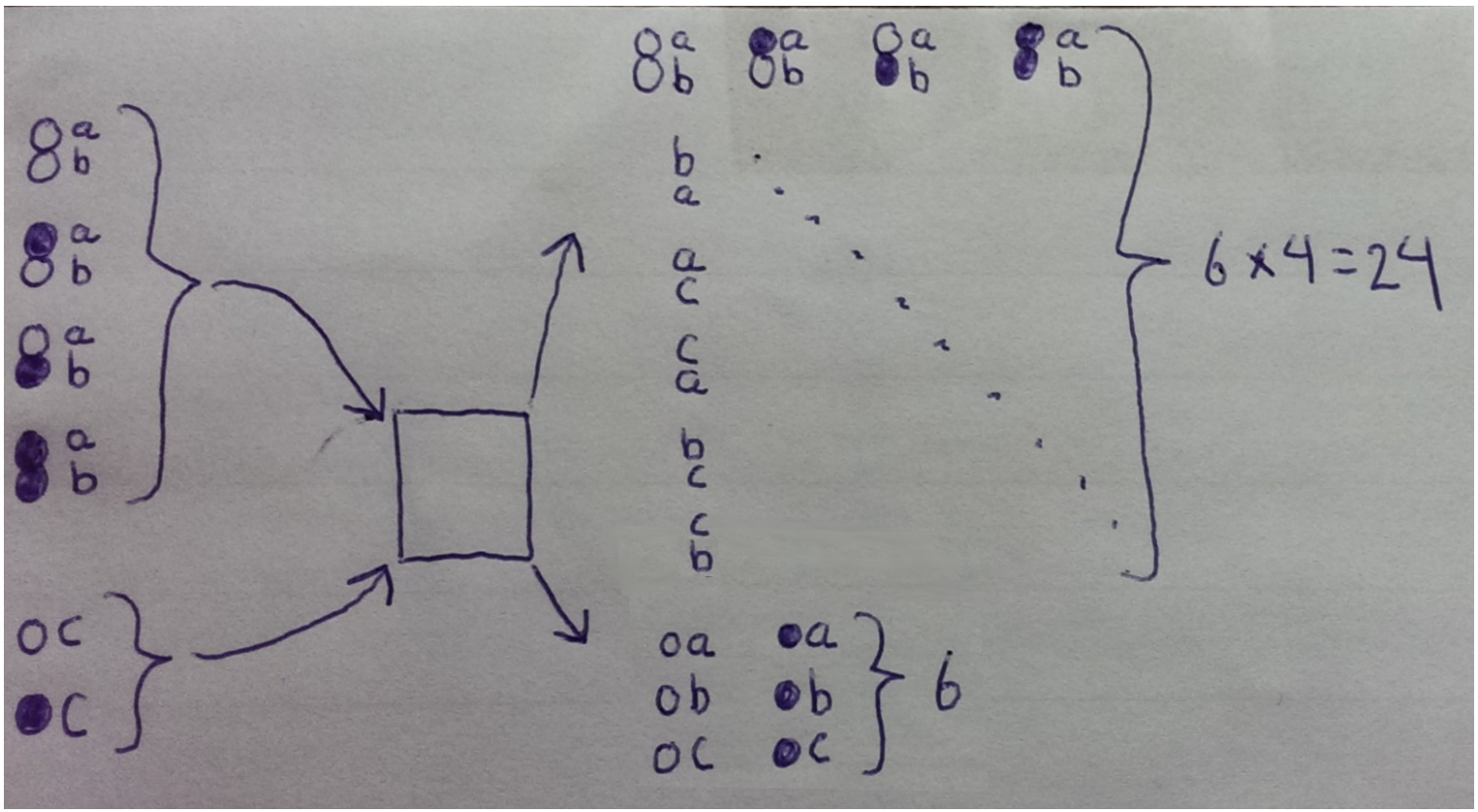
The fundamental problem

$$A \to B, B \qquad \Rightarrow \qquad 2B_1$$
$$ab \to a, b$$

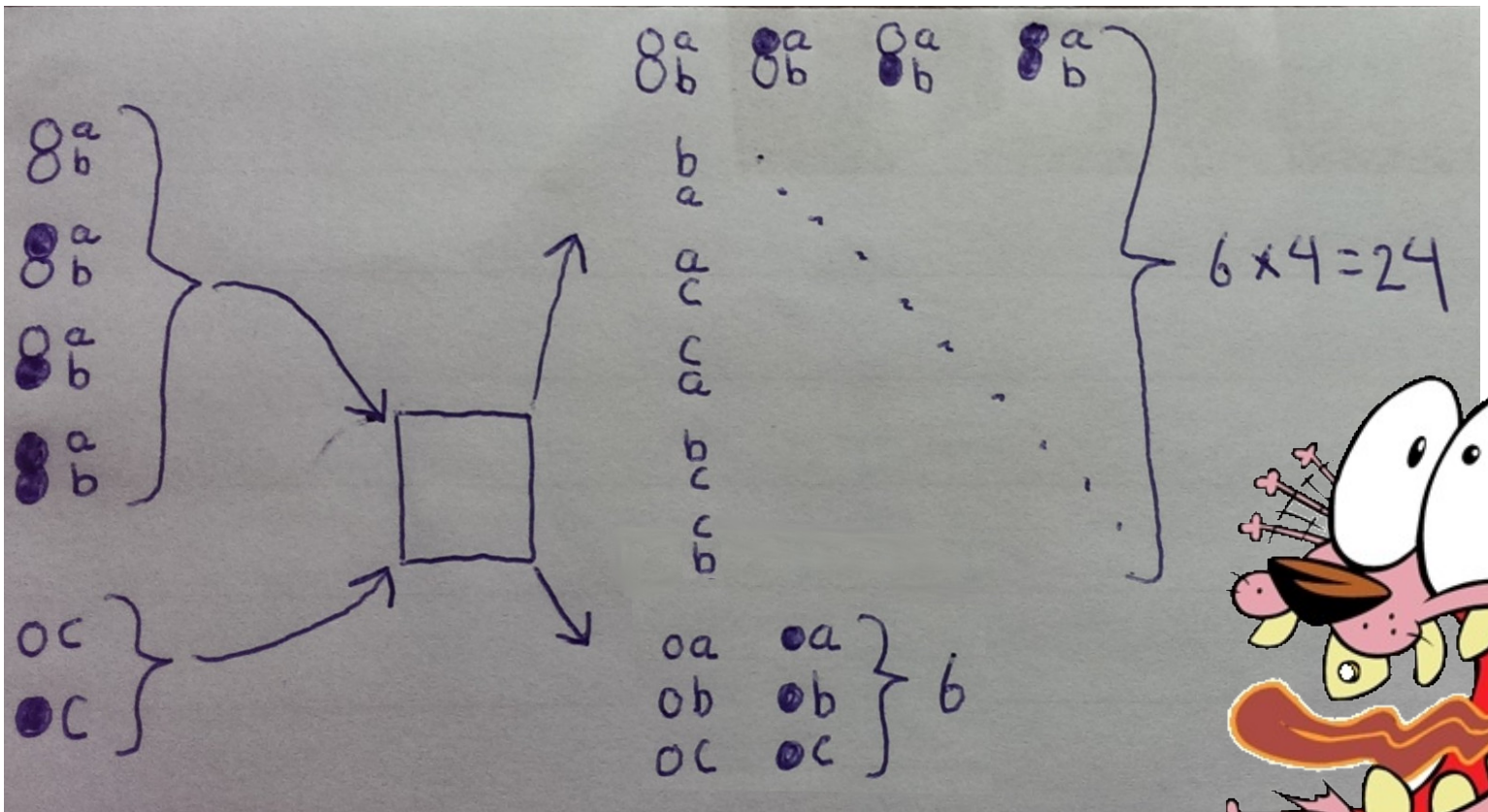
When backtracking from a B_1, does it origin from A_1 or A_2?
Some sort of global labeling is missing, information is lost


6 × 4 = 24
6

Could be modelled with $2^n P(t,n) = \frac{2^n t!}{(t-n)!}$ states per molecule, where t is the sum of carbon numbers of all input substrates.



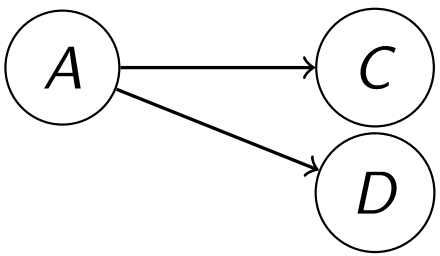
Could be modelled with $2^n P(t,n) = \frac{2^n t!}{(t-n)!}$ states per molecule, where t is the sum of carbon numbers of all input substrates.



Considerations

False positives

Be aware of false positives (if they can exist?)

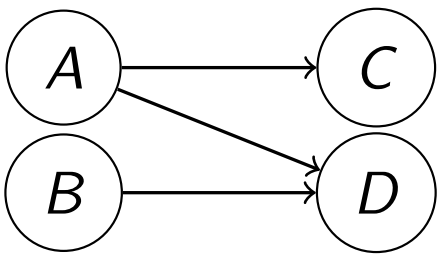


A can go to both C or D.

Considerations

False positives

Be aware of false positives (if they can exist?)



A can go to both C or D.
B can only go to D, so A must go to C.

Considerations

Identity reactions

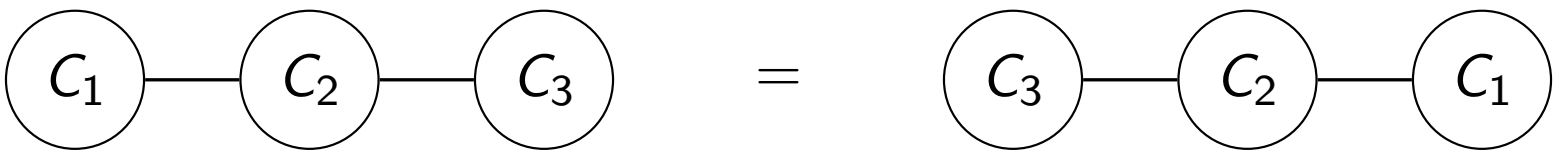
Identity reactions can be removed:

$$abcd \rightarrow abcd \rightarrow abdc \quad \Rightarrow \quad abcd \rightarrow abdc$$

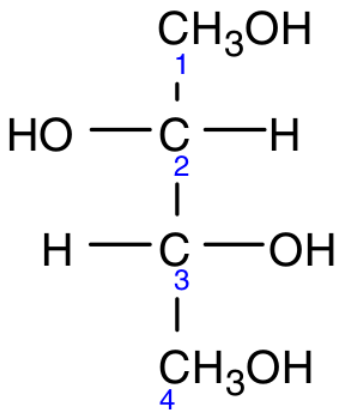
Considerations

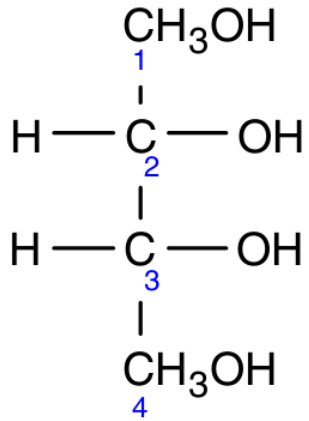
Automorphisms

Indistinguishable carbons, due to automorphisms:



(Bio)chemical equivalence



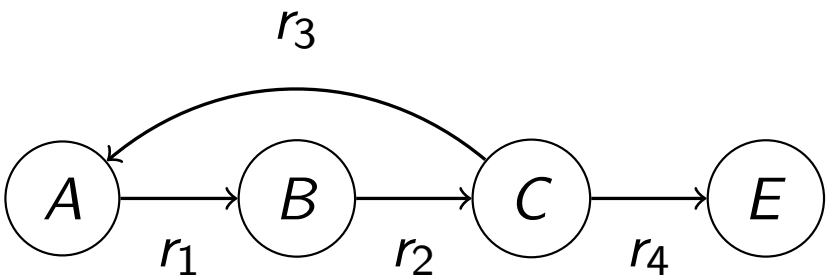


Name :	(2R,3R)-butane-1,2,3,4-tetraol	(2S,3R)-butane-1,2,3,4-tetraol (i.e. erythritol)
Symmetry :	Rotation axis	Center of inversion
C1 vs. C4 :	Chemically and biochemically equivalent	Chemically equivalent, but biochemically distinct

Antoniewicz, Kelleher, and Stephanopoulos, 2007

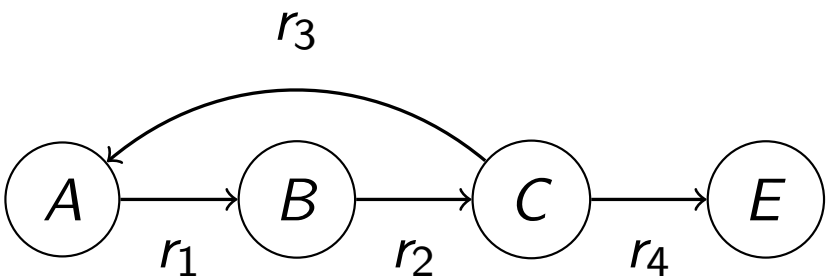
Complexity

A digraph with infinite many non-simple paths.



Complexity

A digraph with infinite many non-simple paths.

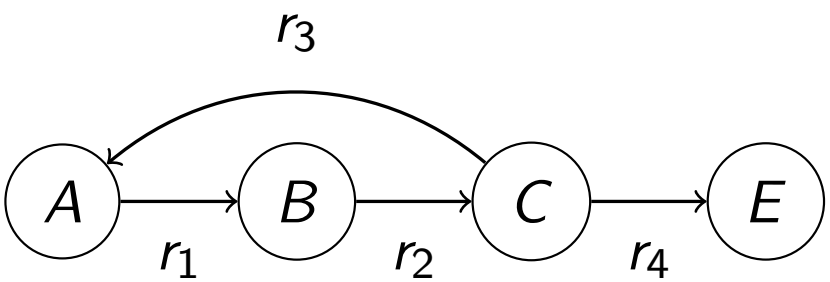


$$r_1 = (3, 4)$$
$$r_2 = (1, 5)$$
$$r_3 = (1, 2)$$
$$r_4 = ()$$

If we model reactions with permutations there are a finite number of non-simple paths before repeating itself.

Complexity

A digraph with infinite many non-simple paths.



$$r_1 = (3, 4)$$
$$r_2 = (1, 5)$$
$$r_3 = (1, 2)$$
$$r_4 = ()$$

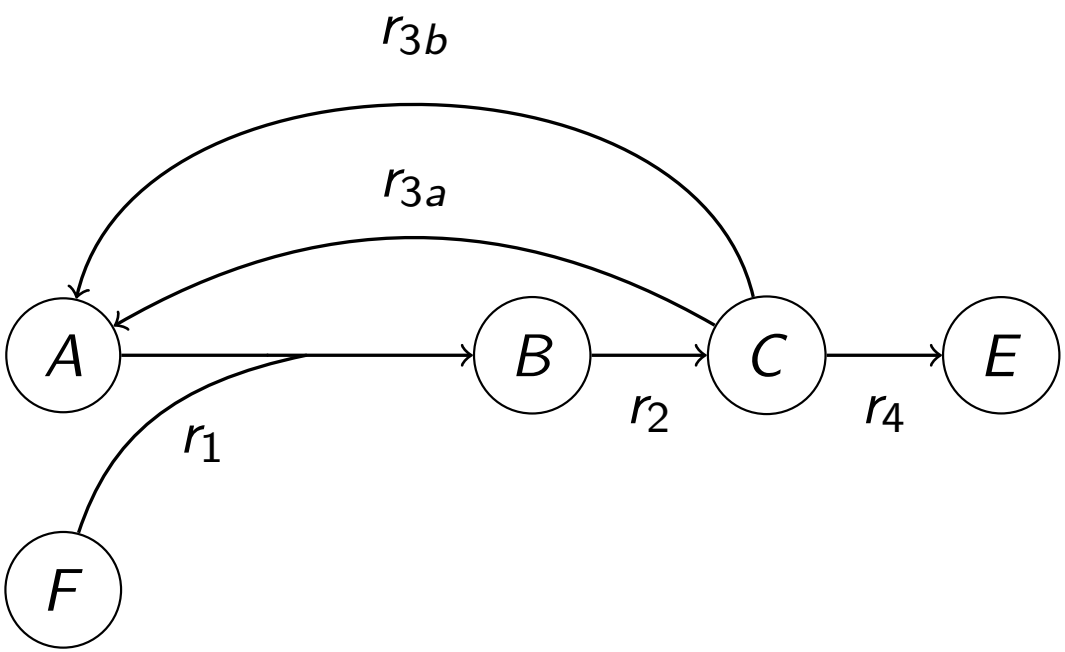
If we model reactions with permutations there are a finite number of non-simple paths before repeating itself.

$$r_{312} = r_3 r_1 r_2 = (1, 2, 5)(3, 4)$$
$$|r_{312}| = \mathrm{lcm}(3, 2) = 6$$

There are 6 ways to take the loop with distinct outcome.

Complexity

Getting more complicated



Flow of computation

- Generative chemistry
  - Input: substrates, targets, reactions
  - Output: reaction network
  - E.g. derive all glycolysis pathways.
- Compute integer flow solution.
- Chemistry happens sequentially, use Petri nets to capture time aspect
- Do reachability analysis with Petri nets
  - If possible: A concurrency graph is returned
- Trace carbon in each concurrency graph

Conclusion

- Current models are lacking information about global labeling.
- By adding this, we can trace atoms through a reaction network.
- Improve simulation of real chemistry

References I

Michael Weitzel, Wolfgang Wiechert, and Katharina Nöh. "The topology of metabolic isotope labeling networks". In: BMC Bioinformatics 8.1 (2007), p. 315. DOI: 10.1186/1471-2105-8-315. URL: http://dx.doi.org/10.1186/1471-2105-8-315.

Maciek R. Antoniewicz, Joanne K. Kelleher, and Gregory Stephanopoulos. "Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions". In: Metabolic Engineering 9.1 (Jan. 2007), pp. 68–86. DOI: 10.1016/j.ymben.2006.09.001. URL: http://dx.doi.org/10.1016/j.ymben.2006.09.001.

References II

Jakob Lykke Andersen et al. “50 Shades of Rule Composition”. In: Formal Methods in Macro-Biology. Springer Science + Business Media, 2014, pp. 117–135. DOI: 10.1007/978-3-319-10398-3_9. URL: http://dx.doi.org/10.1007/978-3-319-10398-3_9.