Generative Chemistries

DM840 - Week 36/37, 2022

Motivation

- Given a chemical reaction network, which interesting pathways and transformation patterns exists?
- What do we really mean by "transformation pattern" and "pathway"?
- What is interesting? (and what is certainly *not* interesting?)

Examples:

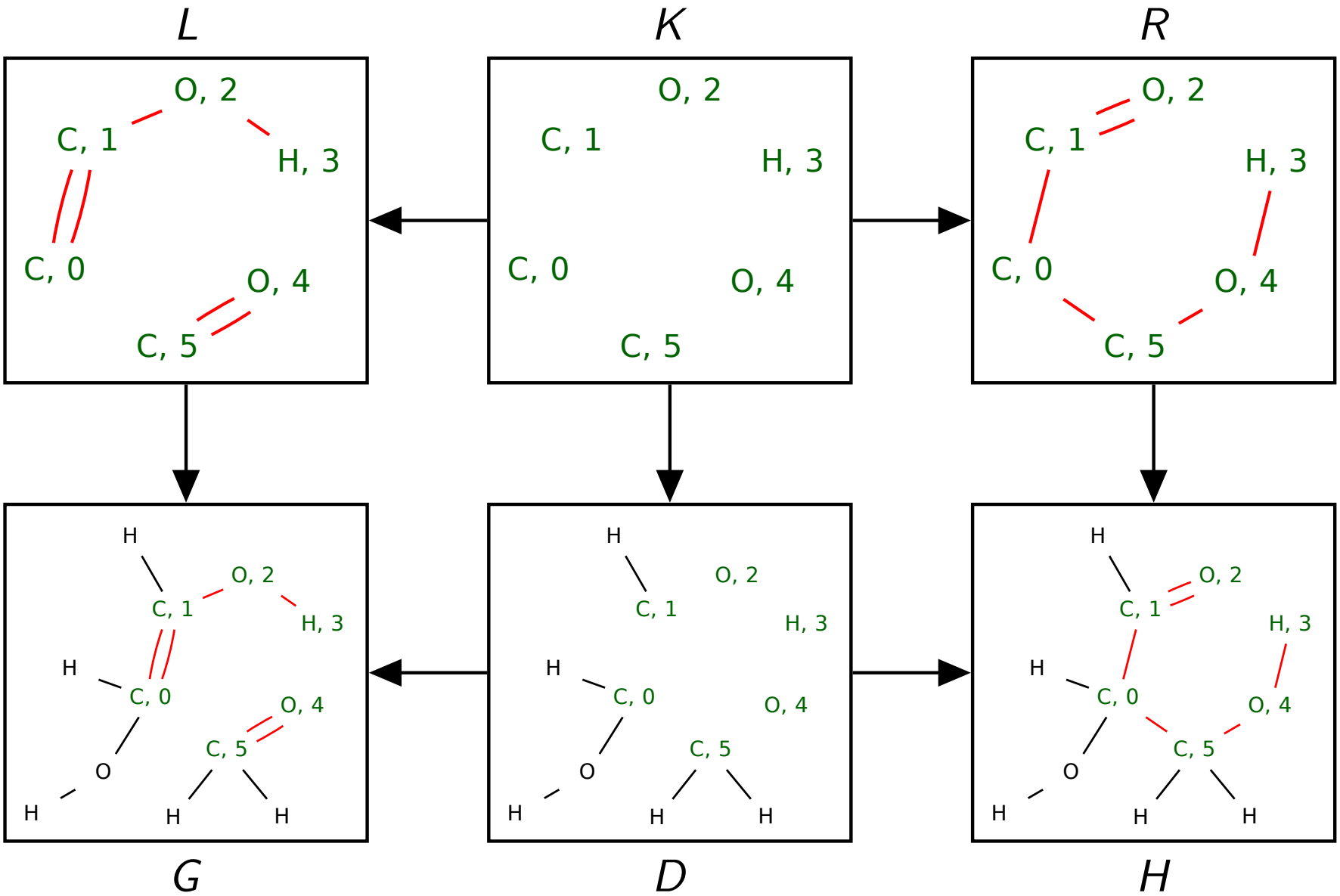
- How can the Pentose Phosphate Pathway be realised?

$$6\,\mathrm{C}_5+? \longrightarrow ? \longrightarrow 5\,\mathrm{C}_6+?$$

- Can we find autocatalysis?

$$\exists A \quad : \quad A+? \longrightarrow 2A+?$$

Graph Transformation - Double Pushout Approach



Graph Transformation - DPO
Connected Components=Molecules

Transformation rule: $p = (L \xleftarrow{l} K \xrightarrow{r} R)$
Derivation with p (and m): $G \overset{p,m}{\Longrightarrow} H$

Assume G has 2 connected components: $G = \{g_1, g_2\}$

$$\{g_1, g_2\} \overset{p,m}{\Longrightarrow} H$$

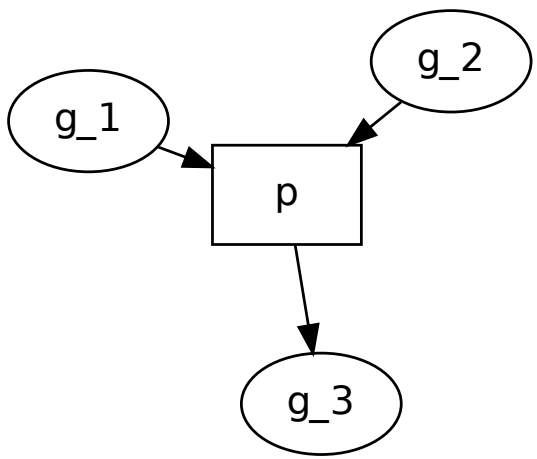
or just

$$\{g_1, g_2\} \overset{p}{\Rightarrow} H$$

Derivation Graphs (Reaction Networks)

Input: a graph grammar (e.g., Formose)
Output: a directed hypergraph of (all) graph derivations

Visualization: $\{g_1, g_2\} \overset{p}{\Rightarrow} g_3$ is represented as



Chemical Space

$\mathcal{G}$: set of input graphs (molecules)
$\mathcal{R}$: set of graph grammar rules (reactions)
$\mathcal{M}(P, k)$: multiset, all elements of P each with multiplicity k
0-th product set: $P_0 = \mathcal{G}$ (starting graphs of the grammar)
Next product sets are defined recursively as

$$P_i = \left\{ p \in Q \,\middle|\, P = \bigcup_{0 \le j < i} P_j \,, t \in \mathcal{R} : \mathcal{M}(P, c_L(t)) \overset{t}{\Rightarrow} Q \right\} \setminus \bigcup_{0 \le j < i} P_j$$

The derivation graph (chemical space) is defined as a directed hypergraph with vertex set $\bigcup_{i \ge 0} P_i$ and the edges being proper derivations.

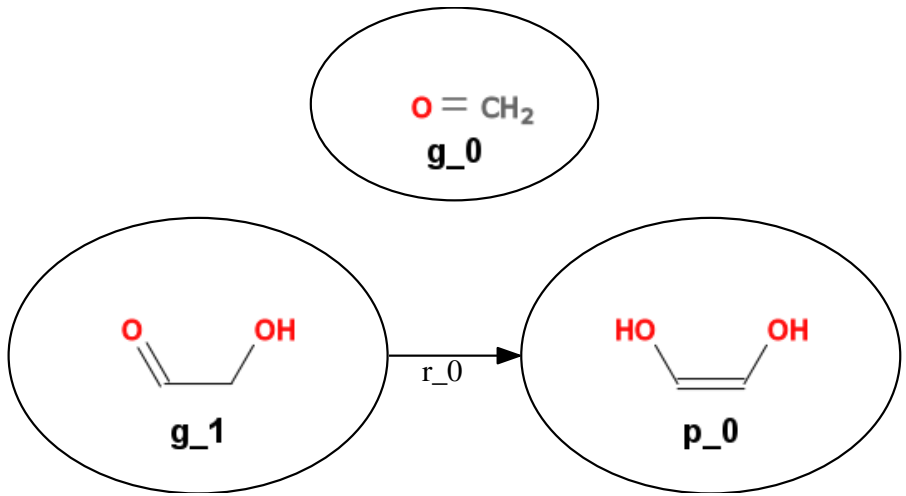
Graph Grammars

Grammar: $\mathcal{H} = (\mathcal{G}, \mathcal{R})$, starting graphs and transformation rules

Example: Formose

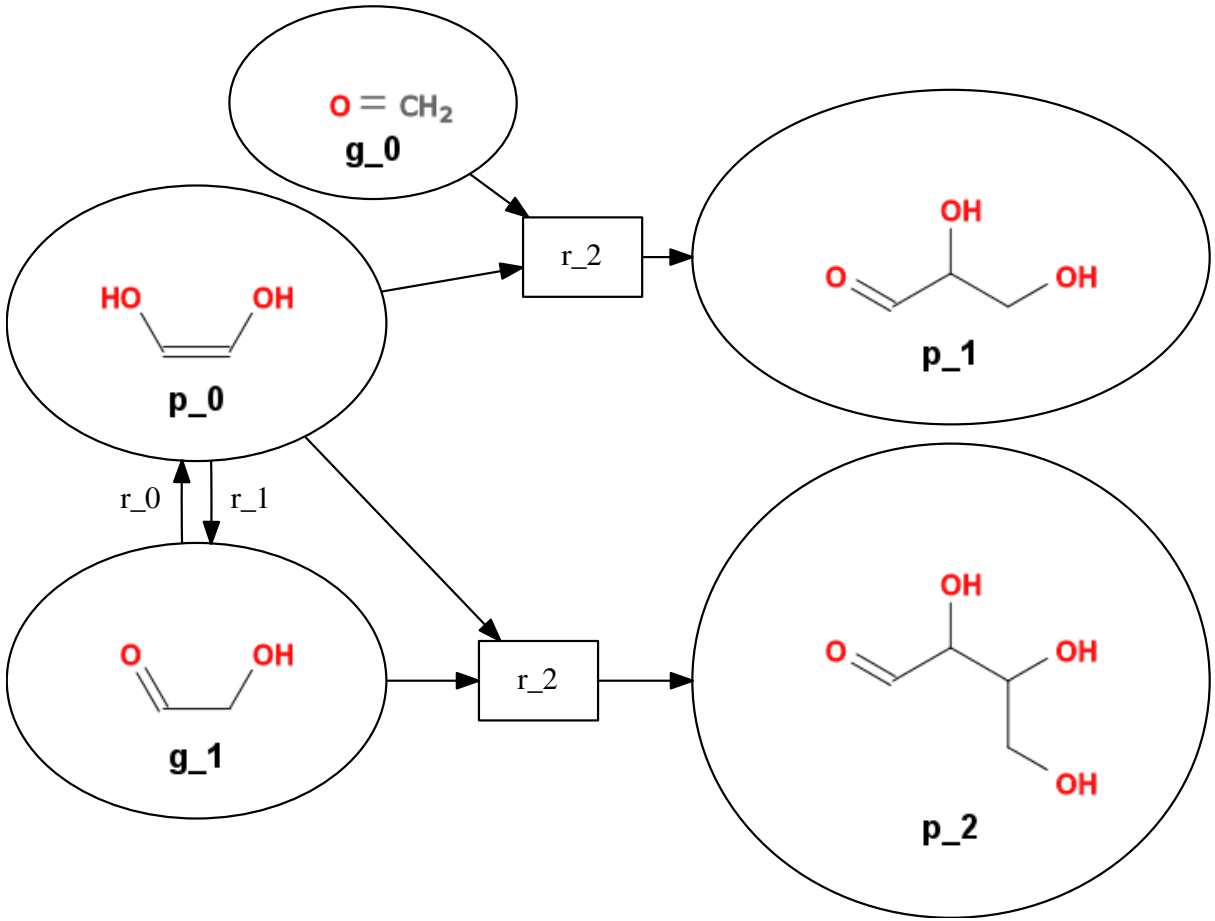
- Starting graphs:
  - g_0 formaldehyde
  - g_1 glycolaldehyde
- Transformation rules:
  - r_0 keto-enol-tautomerism, one direction
  - r_1 keto-enol-tautomerism, the other direction
  - r_2 aldol addition, one direction
  - r_3 aldol addition, the other direction

Reaction Network of Formose



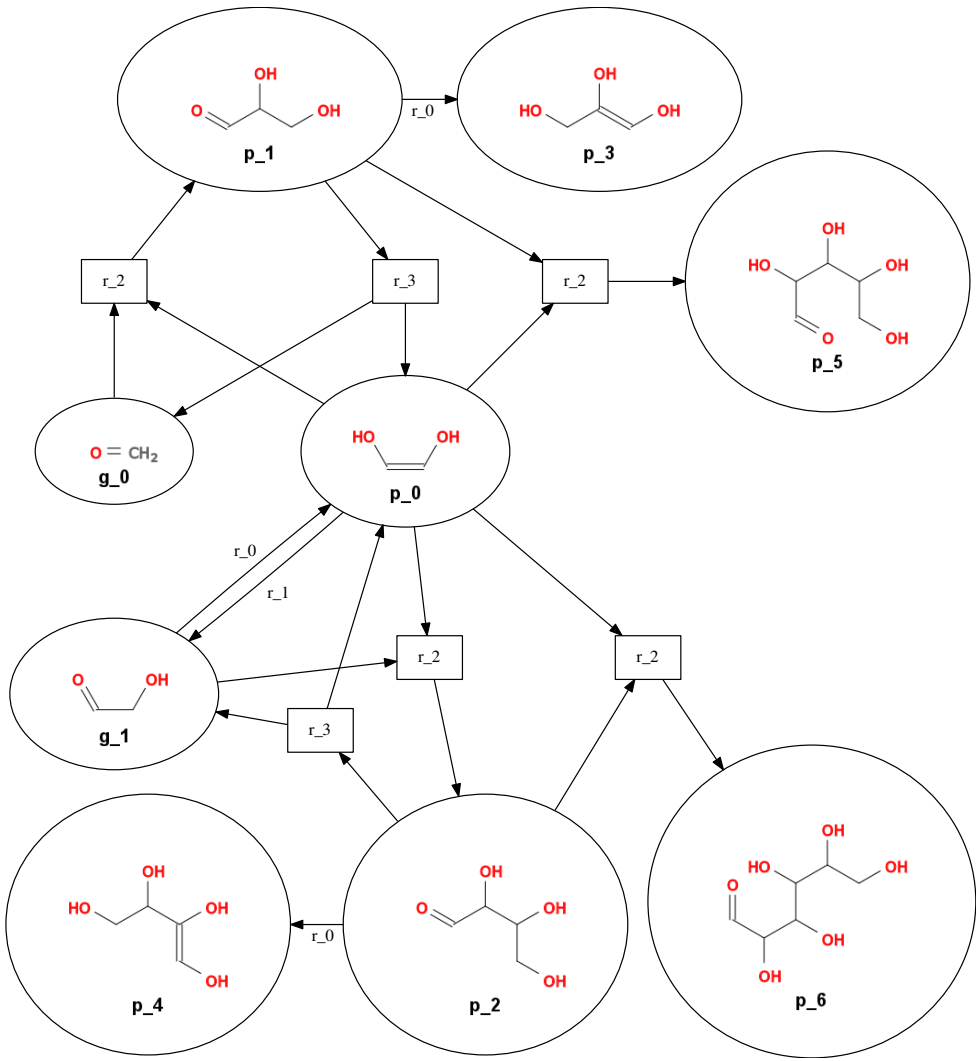
Generation 1

Reaction Network of Formose



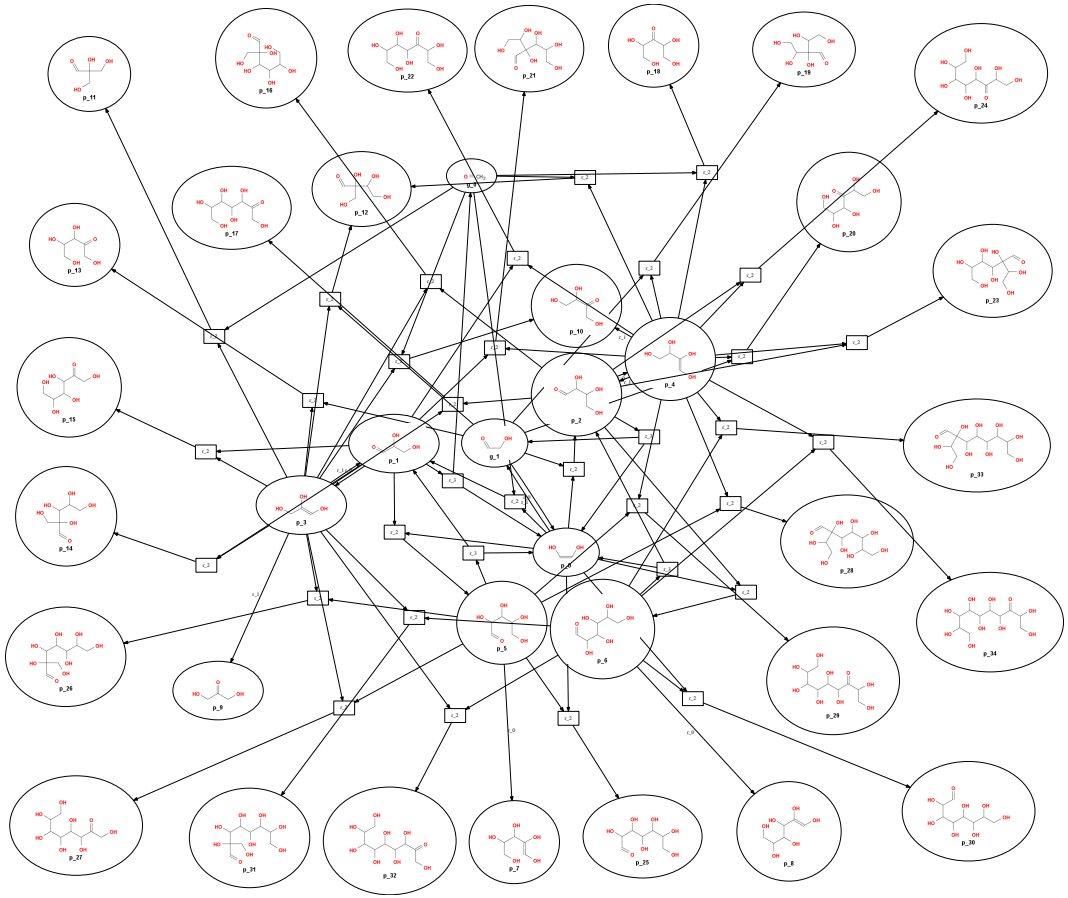
Generation 2

Reaction Network of Formose



Generation 3

Reaction Network of Formose



Generation 4

Pathways in Reaction Networks

Idea: use network flows as model for chemical pathways,
and find interesting flows
Problem: derivation graphs are hypergraphs
Solution: use integer linear programming (ILP)

Examples of interesting questions::

General pathway:

$$6 \text{ ribulose-5-phosphate} \longrightarrow 5 \text{ fructose-6-phosphate}$$

Is it possible? and how?

Autocatalysis:

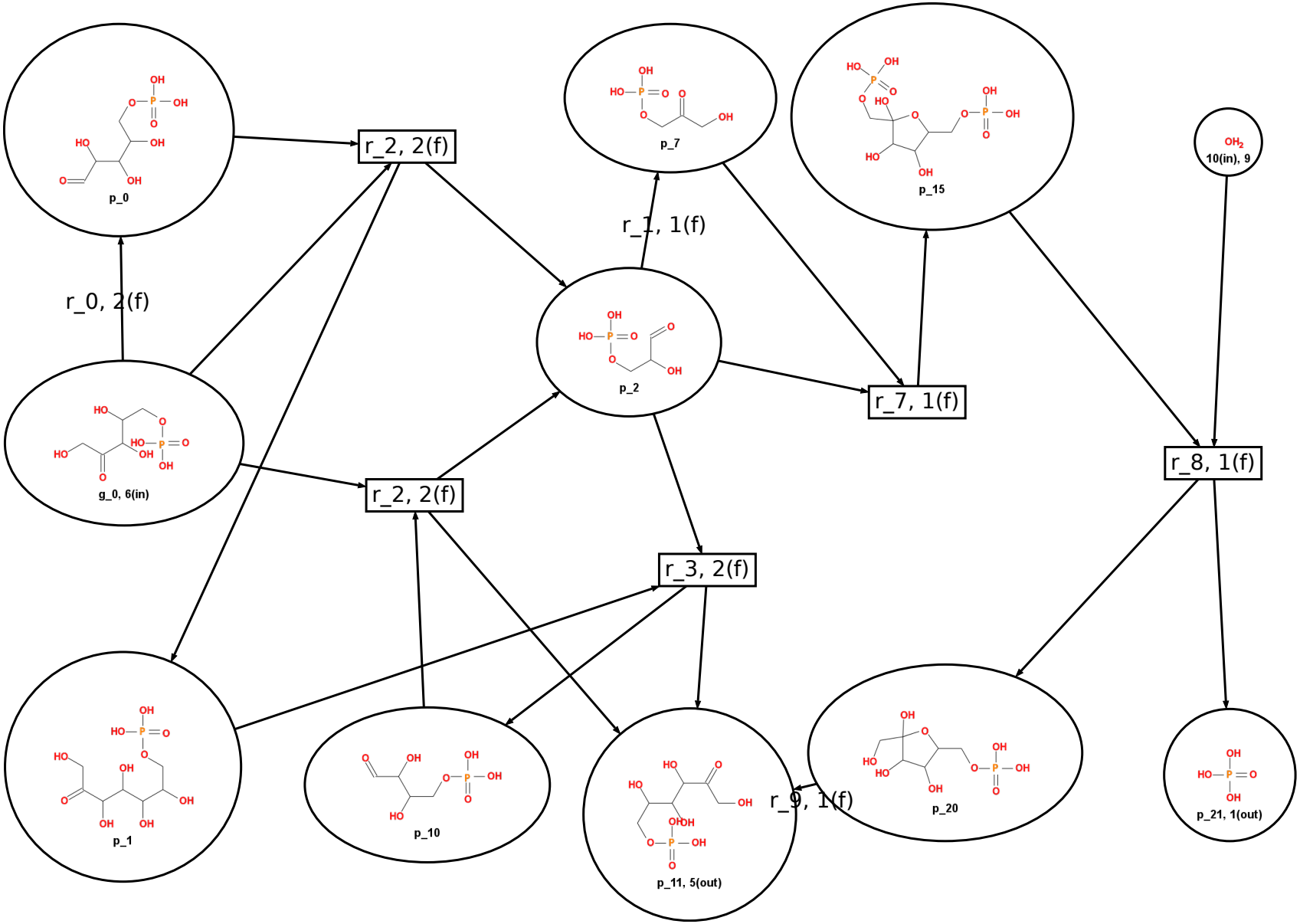
$$2 \text{ formaldehyde} + 1 \text{ glycolaldehyde} \longrightarrow 2 \text{ glycolaldehyde}$$

Is it possible? and how?

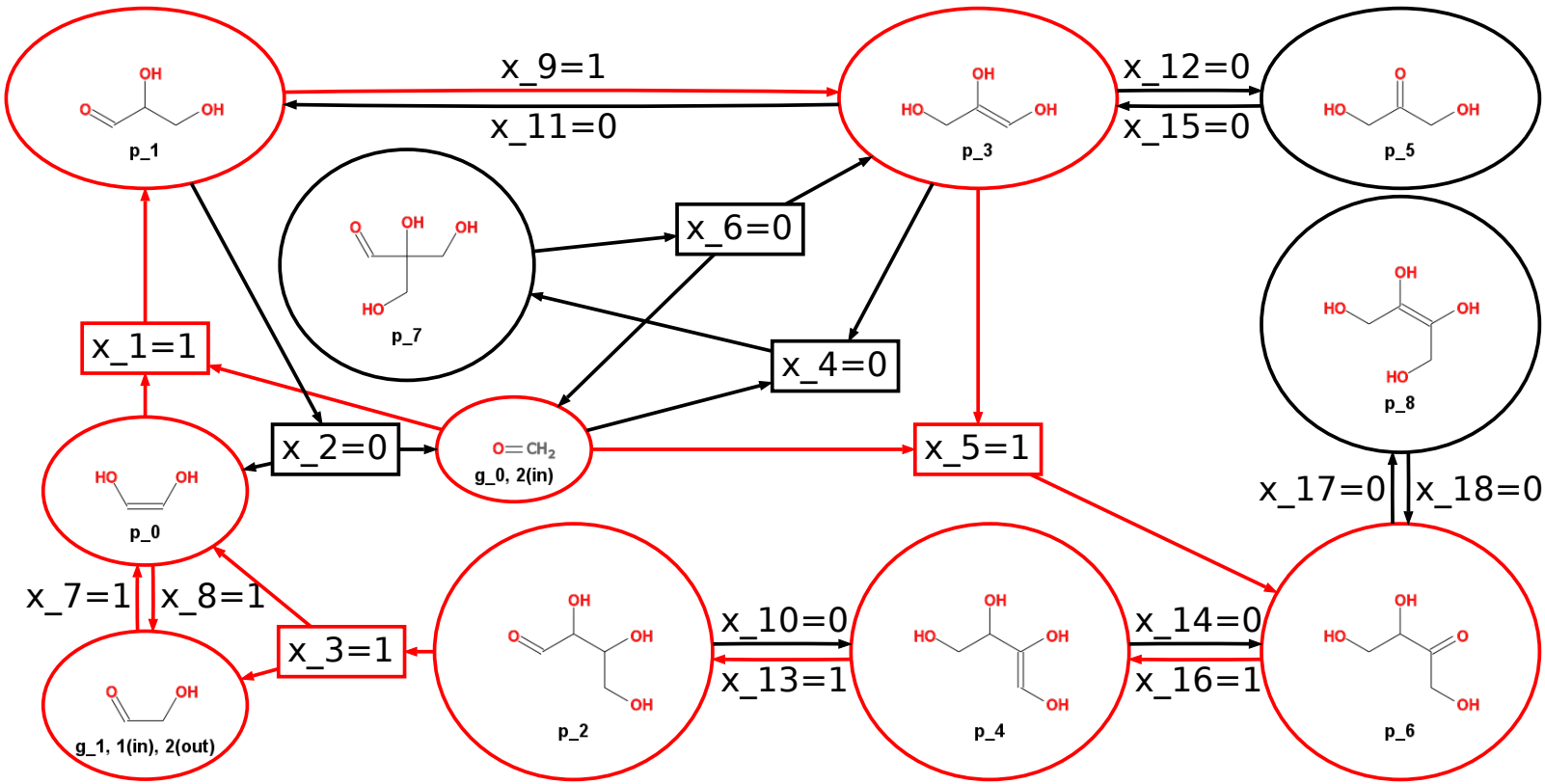
The Pentose Phosphate Pathway

$\rightarrow$ presentation of MØD (intro + query + PPP)

“A” solution to PPP



Simplistic version of the ILP formulation (Formose)



Basis Model

Input: a derivation graph, (V, E)

Augment with input and output edges:

$$E_I = \{e_{in}^v = (\emptyset, \{v\}) | v \in V\} \qquad \bar{E} = E \cup E_I \cup E_O$$
$$E_O = \{e_{out}^v = (\{v\}, \emptyset) | v \in V\}$$

Variables:

$$x_e \in \mathbb{N}_0 \quad , \quad \forall e \in \bar{E}$$

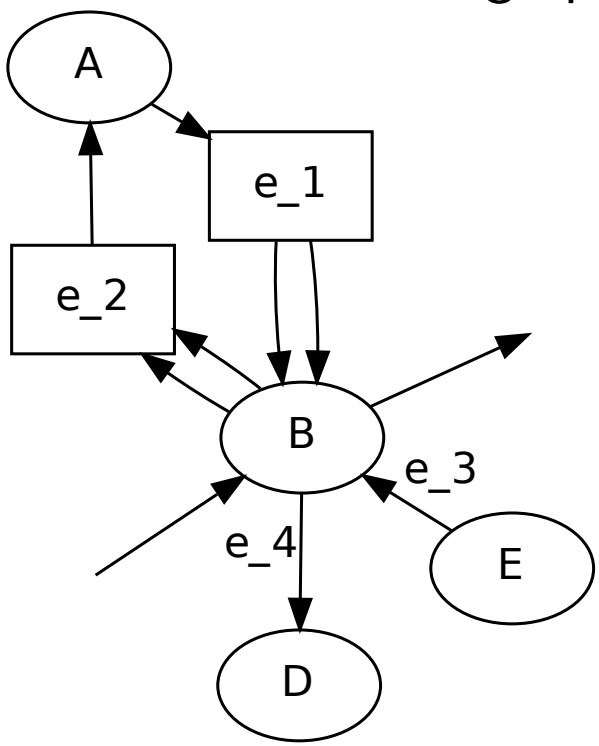
Constraints:

$$\sum_{e \in in(v)} x_e = \sum_{e \in out(v)} x_e \quad , \quad \forall v \in V$$

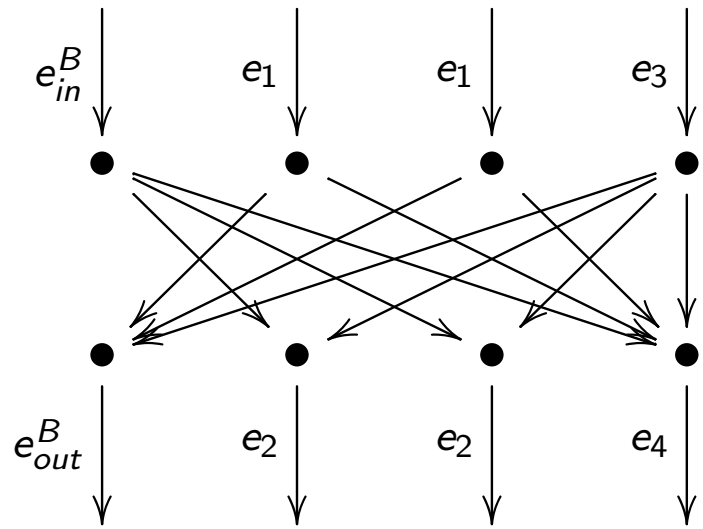
Change of the Basis Model

Replace each vertex with a bipartite graph

Part of a derivation graph



B



Autocatalysis in MØD

Autocatalysis: when a molecule catalyse its own production
Informal Example:
$A + B \rightarrow 2B$
B can not be produced by A or any other input molecules
We say, B is autocatalytic.

Input:

- a derivation graph, (V, E)
- input molecules, $S \subseteq V$
- output molecules, $T \subseteq V$

Does a flow exist such that $\exists v \in V : 0 < x_{in}^{v} < x_{out}^{v}$?

Autocatalysis - ILP formulation

Variables:

$$Z_v \in \{0,1\} \quad , \quad \forall v \in V$$
$$Z_v^{in} \in \{0,1\} \quad , \quad \forall v \in V$$

Desired constraints:

$$Z_v = 1 \Leftrightarrow 0 < x_{in}^v < x_{out}^v$$
$$Z_v^{in} = 1 \Leftrightarrow 0 < x_{in}^v$$

$$\sum_{v \in V} Z_v \geq 1$$

Autocatalysis - ILP formulation

Constraints:

$$Z_v^{in} \leq x_{in}^v \tag{1}$$

$$M \cdot Z_v^{in} \geq x_{in}^v \tag{2}$$

$$Z_v \leq x_{in}^v \tag{3}$$

$$x_{in}^v < x_{out}^v + M \cdot (1 - Z_v) \tag{4}$$

$$M \cdot Z_v \geq x_{out}^v - x_{in}^v - M \cdot (1 - Z_v^{in}) \tag{5}$$

x_{in}^v		x_{out}^v	${}^1Z_v^{in}$	${}^2Z_v^{in}$	${}^3Z_v$	${}^4Z_v$	${}^5Z_v$
0		0	0	–	0	0	–
0		42	0	–	0	–	–
42		0	–	1	–	0	–
	$=$		–	1	–	0	–
	$<$		–	1	–	–	1
	$>$		–	1	–	0	–

MØD – Detection of Autocatalysis

$\rightarrow$ presentation of MØD (autocatalysis)

Autocatalysis in Formose