Generative Chemistries and MØD

Jakob L. Andersen, Christoph Flamm,
Daniel Merkle, Peter F. Stadler

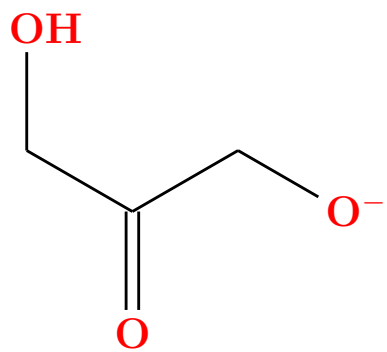
Department of Mathematics and Computer Science
University of Southern Denmark

Bled, February 2019

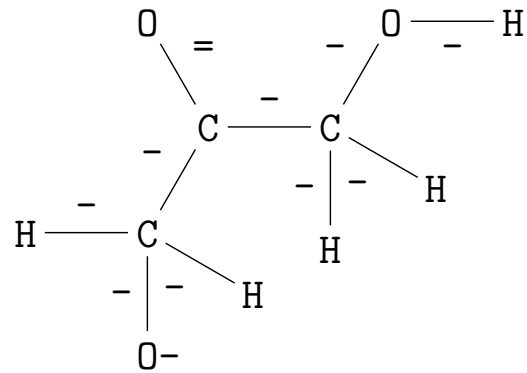
Background

- Modelling and analysis of chemical systems.
- Coherent, flexible models and methods.
- Based on formal methods, instead of chemical "rules".

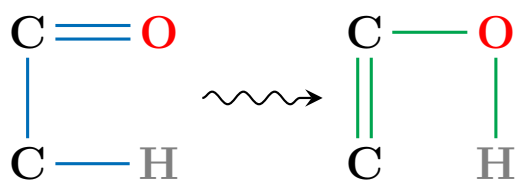
Molecules as graphs



modelled as

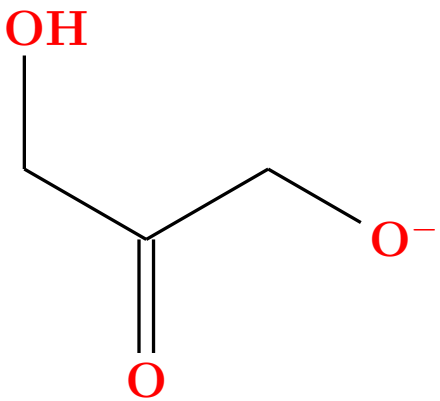


Reactions as rules

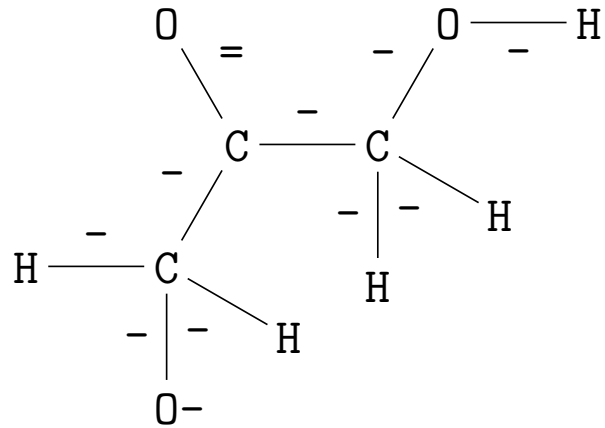


Modelling of Molecules

Def: a molecule is a labelled, connected, simple, undirected graph.



(a) Chemical depiction.



(b) Visualisation of underlying model.

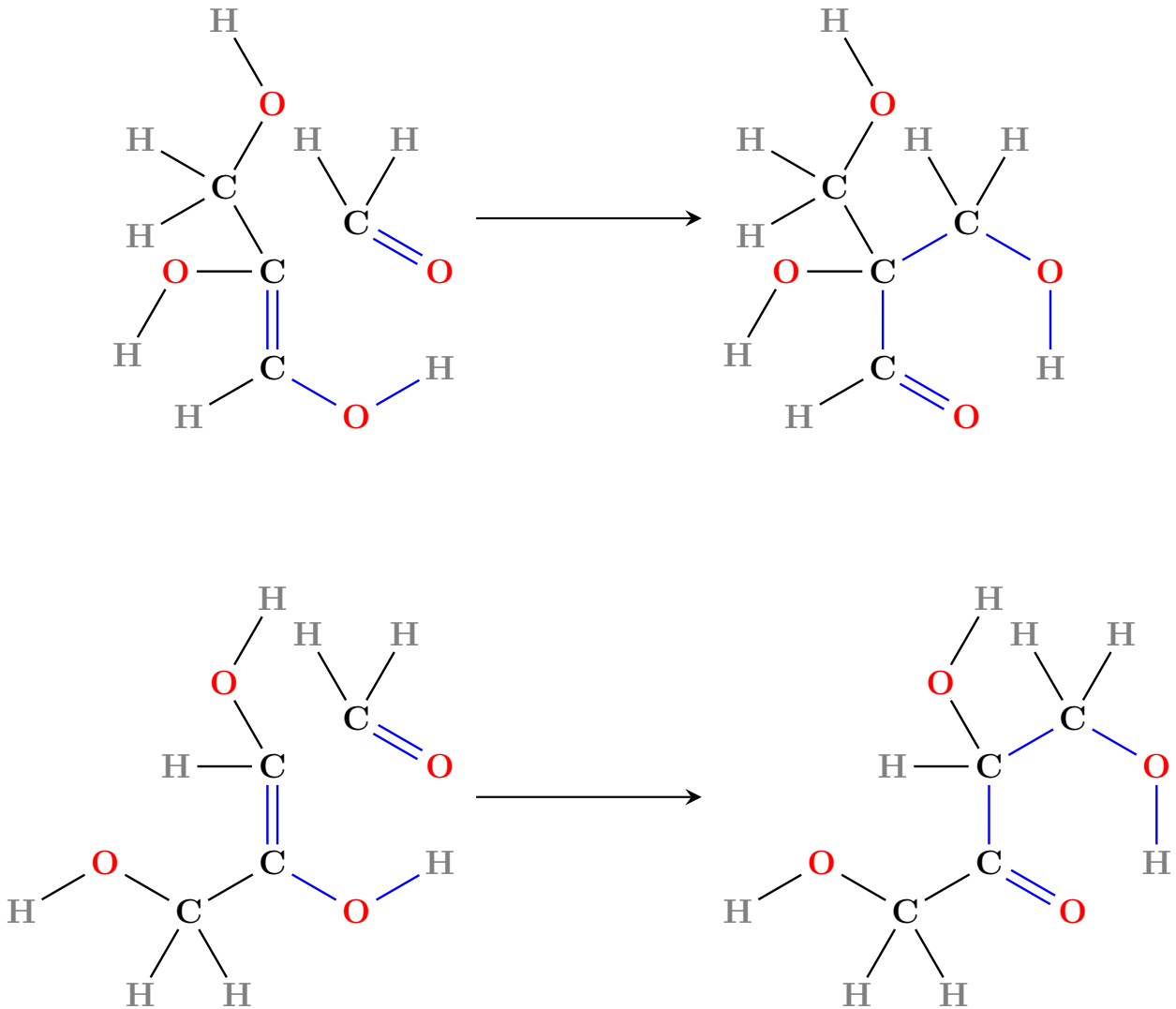
Vertex label $\equiv$ chemical element and charge

$$\{\mathsf{H}, \mathsf{He}, \mathsf{Li}, \mathsf{Be}, \mathsf{B}, \mathsf{C}, \mathsf{N}, \mathsf{O}, \ldots, \mathsf{Uuo}\} \times \mathbb{Z}$$

Edge label $\equiv$ bond type

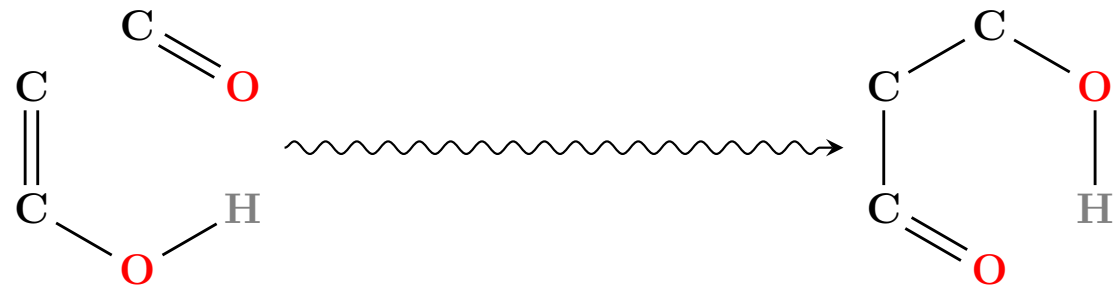
$$\{\textsc{Single}, \textsc{Double}, \textsc{Triple}, \textsc{Aromatic}\}$$

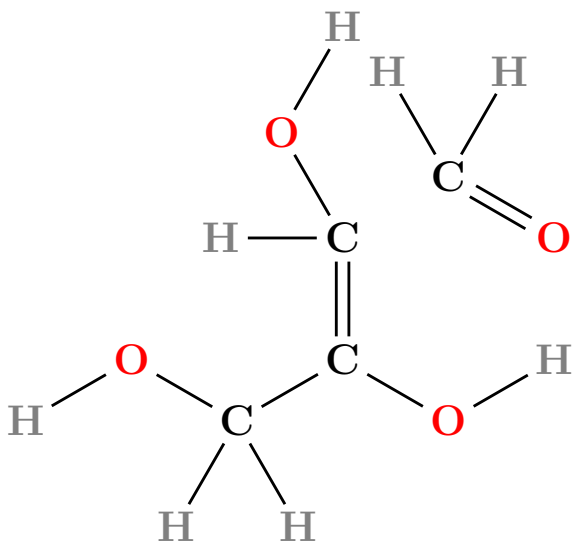
Chemical Reactions (of the Same Type)



Chemical Reaction Patterns

Rule

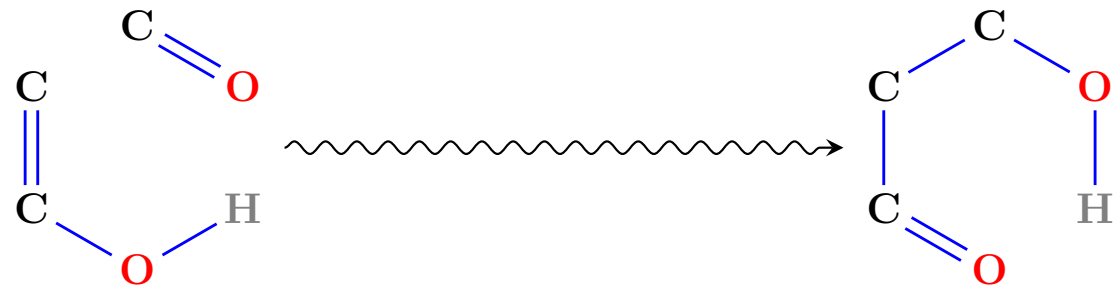




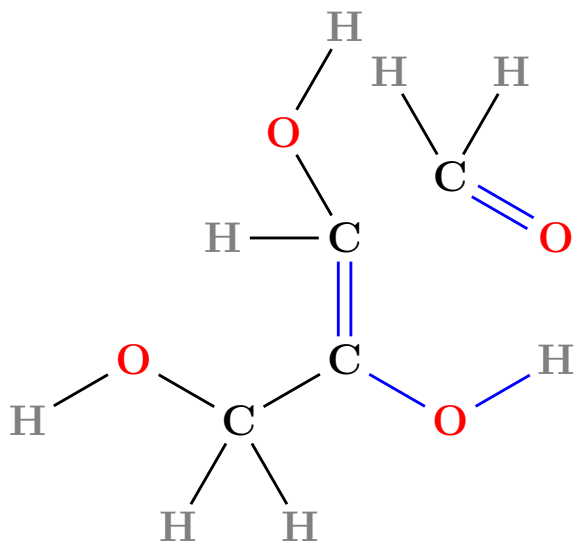
Educts

Chemical Reaction Patterns

Rule

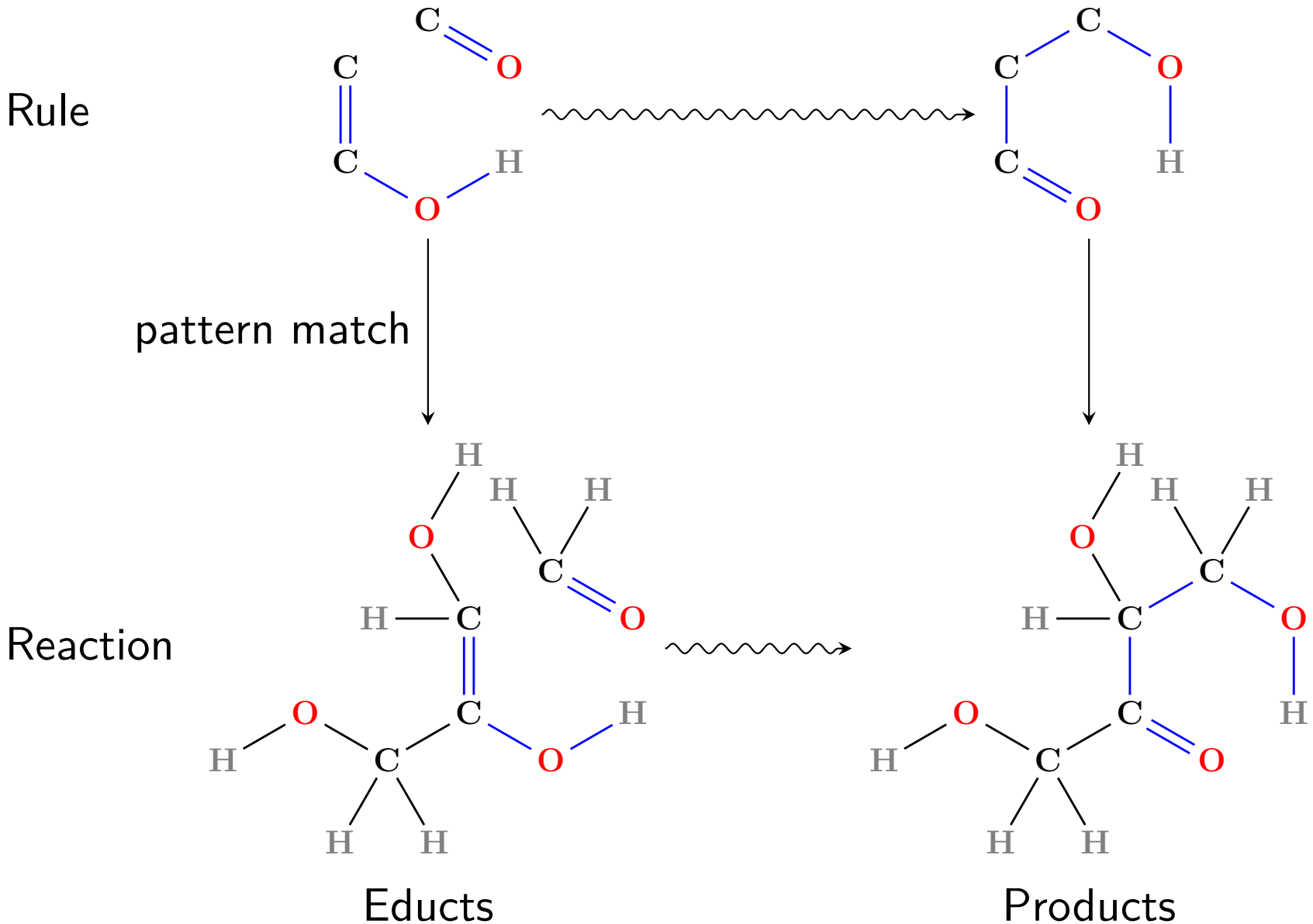


pattern match



Educts

Chemical Reaction Patterns



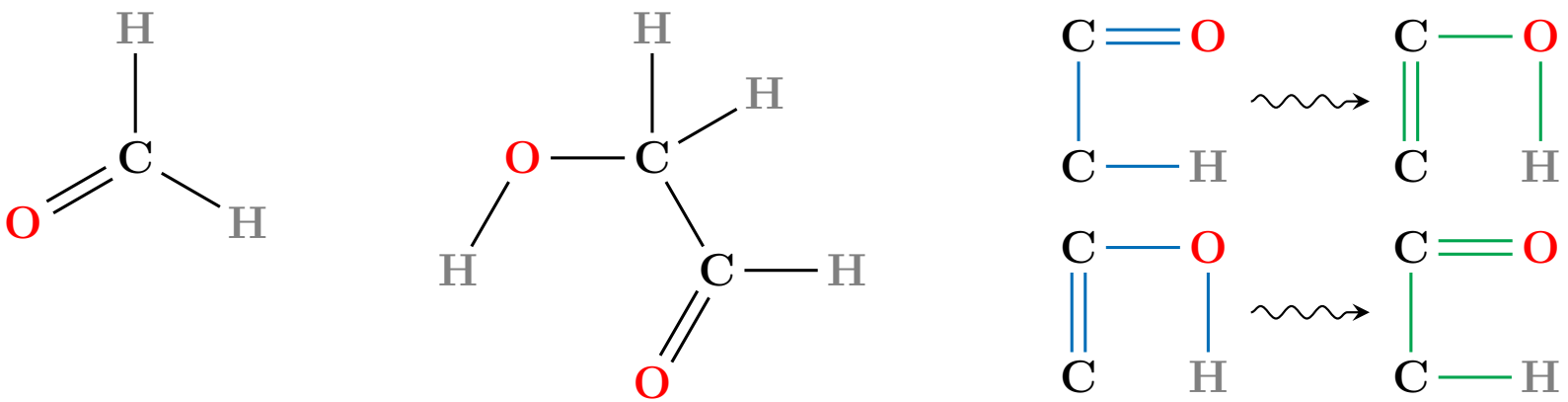
Modelling Chemistries as Graph Grammars

- Starting graphs (molecules)
- Graph transformation rules (reaction patterns)

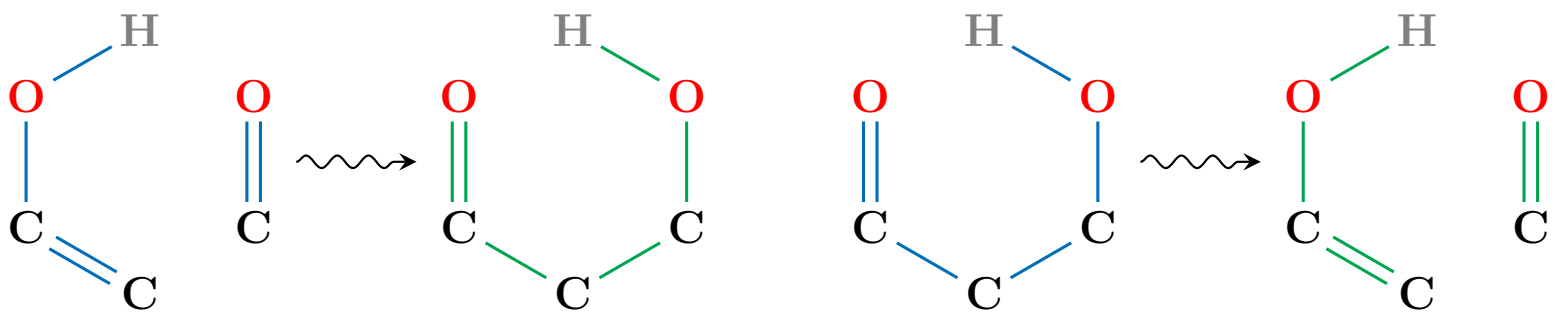
Direct derivation $\equiv$ reaction

Example: The Formose Chemistry

Formaldehyde: Glycolaldehyde: Keto-enol tautomerism:



Aldol addition: Retro aldol addition:



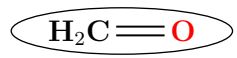
Generation of Derivation Graphs (Reaction Networks)

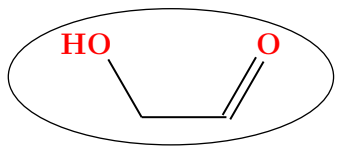
Networks are modelled as directed hypergraphs.

Vertex $\equiv$ molecule (labelled with the molecule graph)

Directed hyperedge, (pair of multisets of vertices)

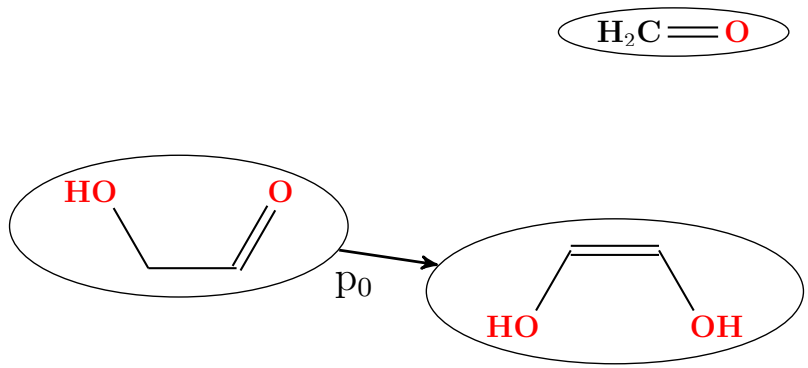
$\equiv$ reaction (labelled with a graph transformation rule)





Generation of Derivation Graphs (Reaction Networks)

Networks are modelled as directed hypergraphs.
Vertex ≡ molecule (labelled with the molecule graph)
Directed hyperedge, (pair of multisets of vertices)
≡ reaction (labelled with a graph transformation rule)



Whenever a new graph/molecule is create, we must check if we have it already: the graph isomorphism problem.

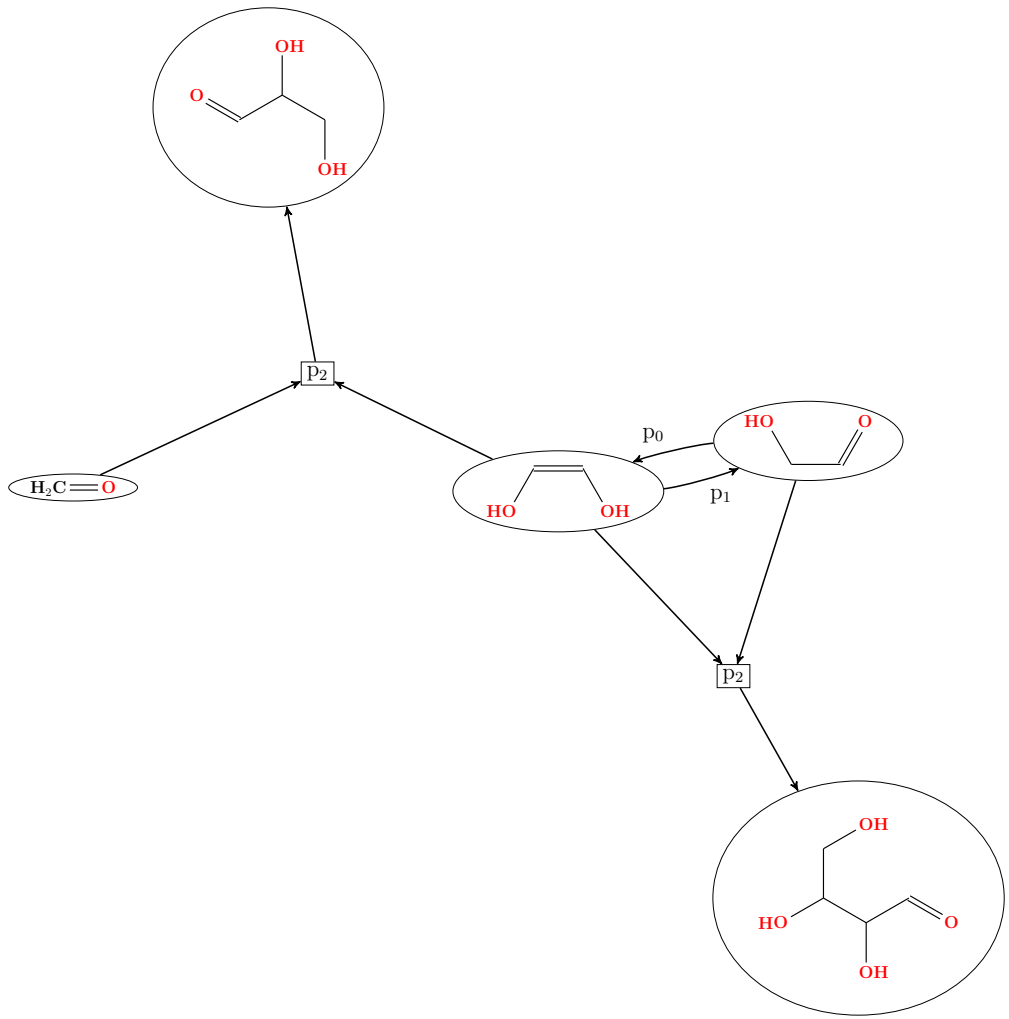
Generation of Derivation Graphs (Reaction Networks)

Networks are modelled as directed hypergraphs.

Vertex ≡ molecule (labelled with the molecule graph)

Directed hyperedge, (pair of multisets of vertices)

≡ reaction (labelled with a graph transformation rule)



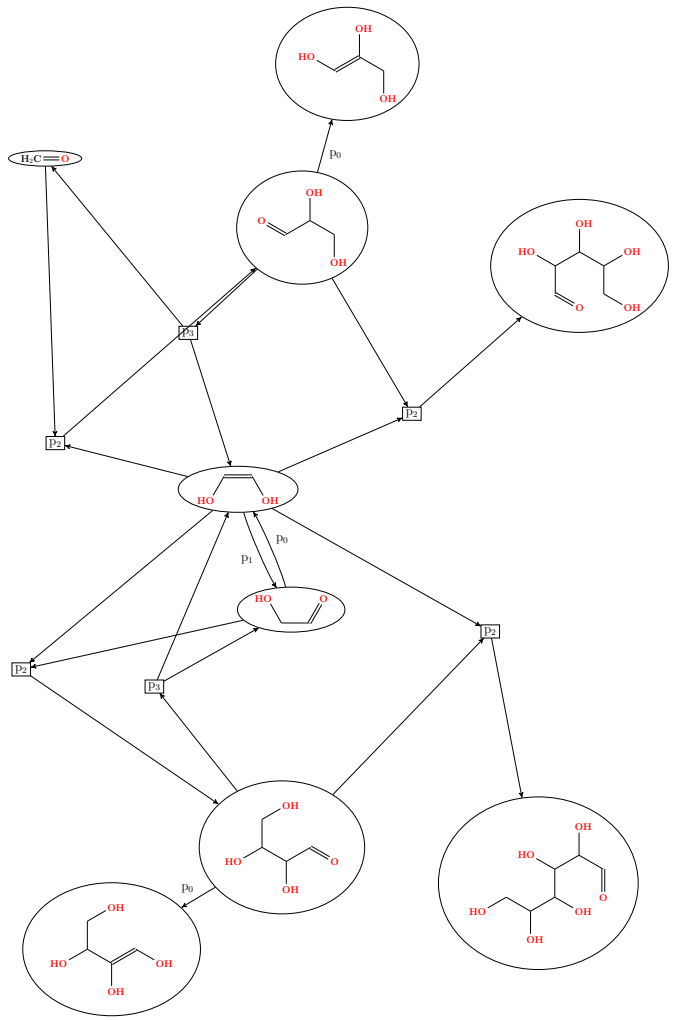
Generation of Derivation Graphs (Reaction Networks)

Networks are modelled as directed hypergraphs.

Vertex ≡ molecule (labelled with the molecule graph)

Directed hyperedge, (pair of multisets of vertices)

≡ reaction (labelled with a graph transformation rule)



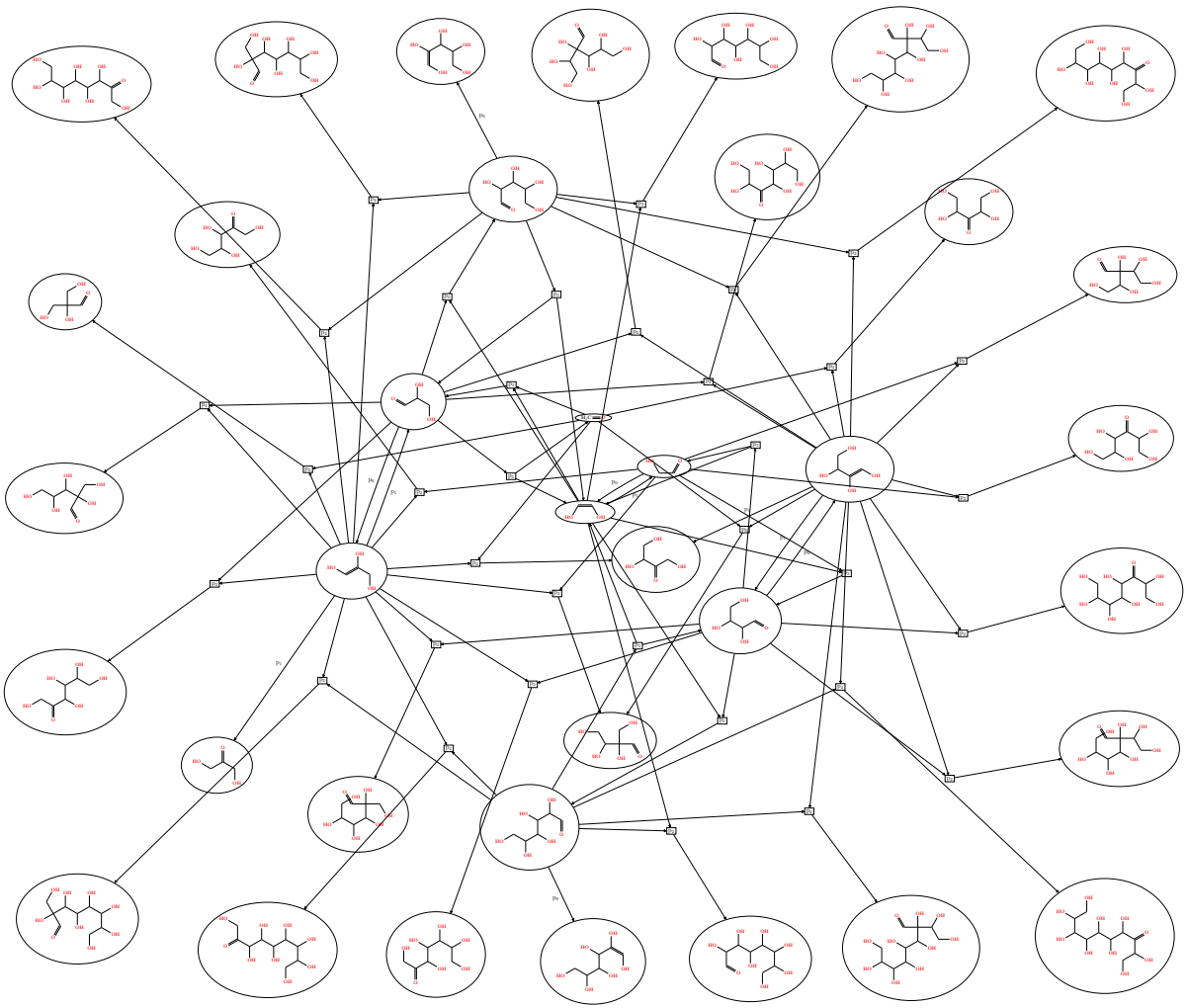
Generation of Derivation Graphs (Reaction Networks)

Networks are modelled as directed hypergraphs.

Vertex $\equiv$ molecule (labelled with the molecule graph)

Directed hyperedge, (pair of multisets of vertices)

$\equiv$ reaction (labelled with a graph transformation rule)



Chemical Rules

- What is a "good" reaction rule?
- It should represent the underlying mechanism.
- But how do we know the mechanism?
- Well, maybe there is a paper/book somewhere ...
- But what if not?
- Maybe we can decompose the reaction into elementary steps.
- What is an elementary step?
- What is a good decomposition?

[Andersen et al., MathChemComp, 2018]

Conversion Routes in Networks, i.e., Pathways

Given a network, how can $\langle input \rangle$ be converted to $\langle output \rangle$?

- 6 ribulose 5-phosphate $\longrightarrow^*$ 5 fructose 6-phosphate?
- sugar $\longrightarrow^*$ $CO_2 + \langle$energy molecules$\rangle$?
- $HCN + H_2O \longrightarrow^* \langle$biomolecules$\rangle$?

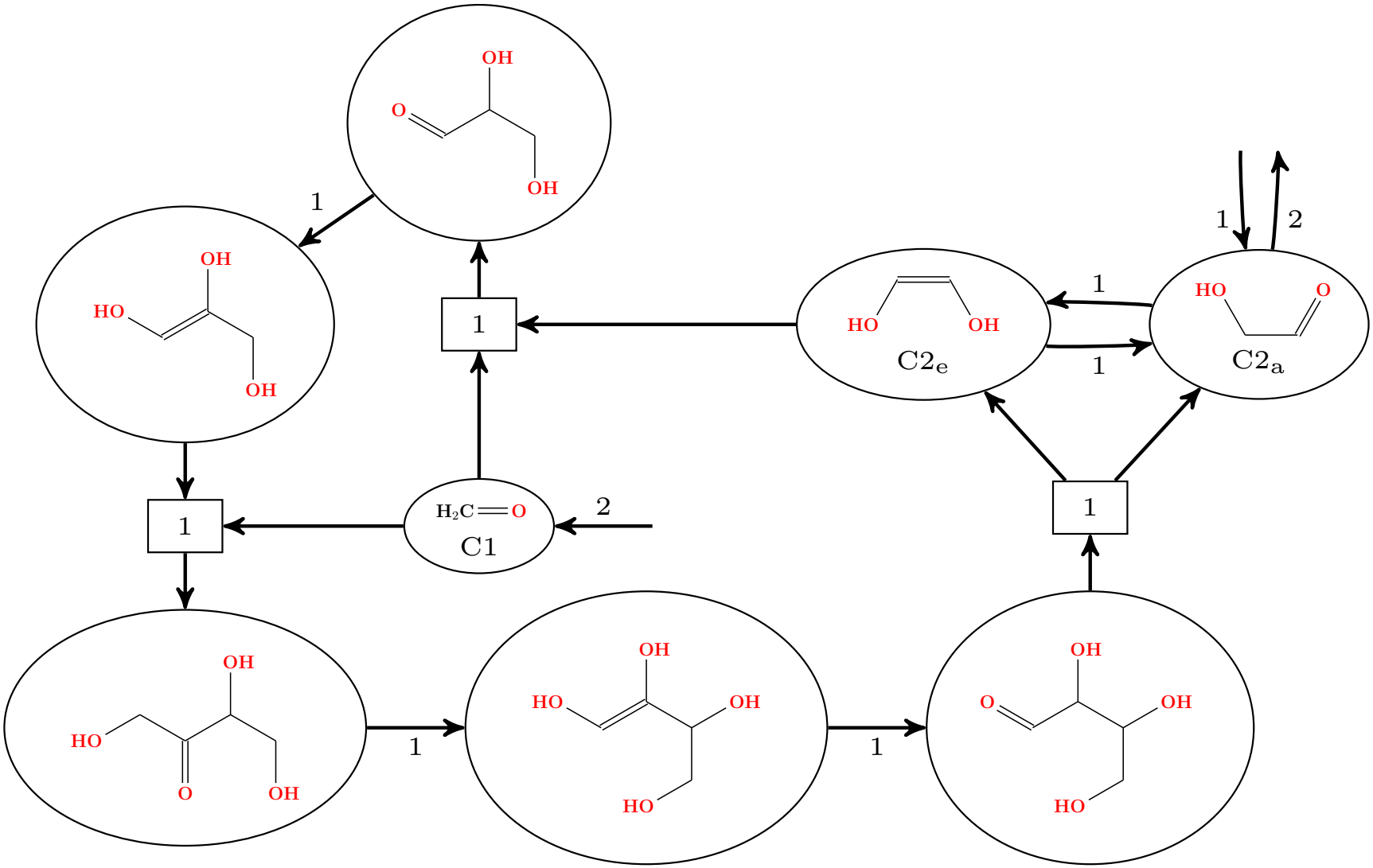
Specialised routes (pathway motifs):

- Catalysis
- Autocatalysis

Can we find a set of necessary constraints?

- Pathway model implemented with Integer Linear Programming.
- Enumeration of (sub)optimal solutions.

The Formose Reaction — $\mathrm{C2_a} + 2\,\mathrm{C1} \longrightarrow 2\,\mathrm{C2_a}$



Network Generation

Problem

- Sometimes the network is implicitly defined.
- An explicit version may be *very* large.
- It may even be infinite, so how much to generate?

Rule-Based ("Network-Free") Simulation

- Use the stochastic simulation algorithm by Gillespie.
- Only enabled reactions are required in each step.
- Generate reactions (and corresponding molecules) as needed.
- Compute reaction rates dynamically.
  (Probably needs specialization for each individual system)

Rule-Based Simulation

- What is the structure of a "molecule"?
- What rule formalism to use for generating reactions?

Examples

- Molecule: just a name.
  Rules: lazy reaction enumeration.
- Kinfold, for kinetic folding of RNA:
  Molecule: RNA sequence with base pairs.
  Rules: hard-coded base-paring rules.
- Kappa (or BioNetGen), designed for cell signalling networks:
  Molecule: site-graph.
  Rules: single-pushout (SPO) formalism.

How about "ordinary" atomic-level (bio)chemistry?
E.g., simulation of stable isotope experiments.

[Flamm et al., RNA, 2000], [Danos et al., FSTTCS, 2012],

[Harris et al., Bioinformatics, 2016], [Suderman et al., B. Math. Biol]

Rule-Based Simulation Using MØD

Simulation as a network computation.

```
strat = (   addSubset(initialMolecules)
            >> repeat(
              inputRules
              >> simulationCallback
              >> clear
              >> addUniverse(lambda: M_U)
              >> addSubset(lambda: M_S)
)           )
```

- If the cached network becomes too large: simply discard it and start a new.
- Network generation and simulation is separated.
- The simulation can be customized on the user side.

[Andersen et al., VEMDP, 2018]