Graph Transformations in an Enzymatic Context

MASTER THESIS PROJECT

MIKKEL PILEGAARD

DANIEL MERKLE & JAKOB LYKKE ANDERSEN

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE

UNIVERSITY OF SOUTHERN DENMARK

About me

Computer Science master student from the University of Southern Denmark

At the start of thesis

Looking for interesting ideas and difficult problems!

Project Overview

Exploratory work for *hopefully* novel results and applications

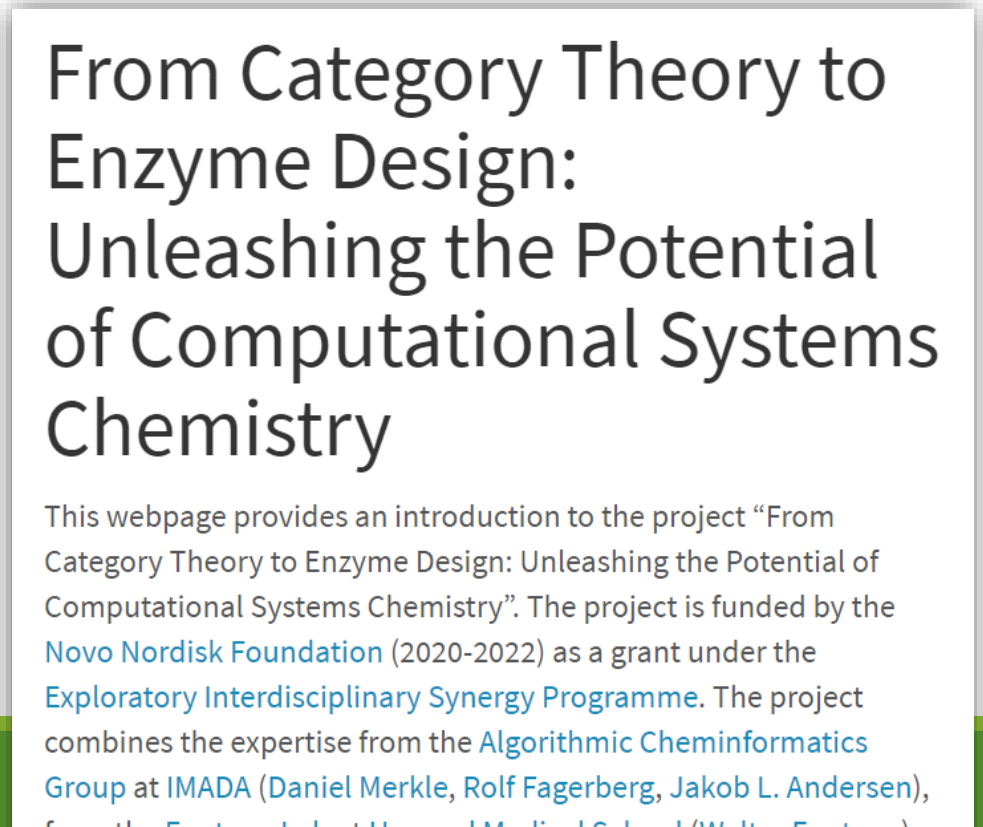
Novo Nordisk Foundation thinks it's exciting too! cheminf.imada.sdu.dk/novo-synergy

A **lot** of work

Presented here is the overarching plan

My role:

- Hopefully help with pre-work
- Method development
- Being kind of naïve on purpose



Applications

Novel enzyme design

Synthetic biology

One-pot designs

MØD + Kappa

Supporting tools for existing discovery techniques

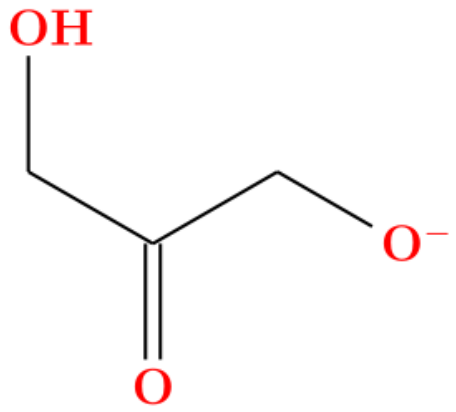
Concrete Steps

1. Rule mining
    1. M-CSA
    2. Infer rules
2. Rule modelling
    1. Generalize to “base knowledge” rule set
    2. Idea: Re-check reaction mechanisms
3. Stochastic simulation
4. Catalytic mechanisms
5. Causality and pathways
6. …

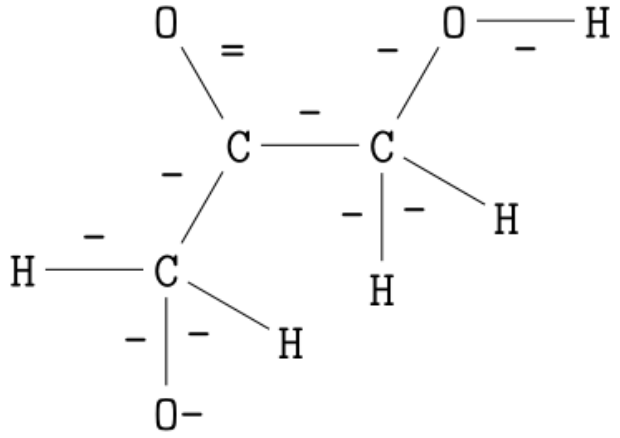
Rules

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

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

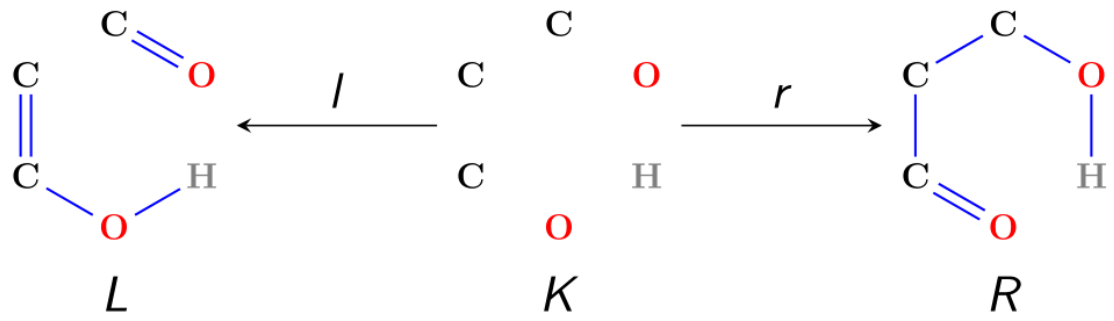
Edge label $\equiv$ bond type

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

Graph Transformation Rules

Vertices and edges are either deleted, preserved, or added.

As a Double Pushout (DPO) rule $p = (L \xleftarrow{l} K \xrightarrow{r} R)$:

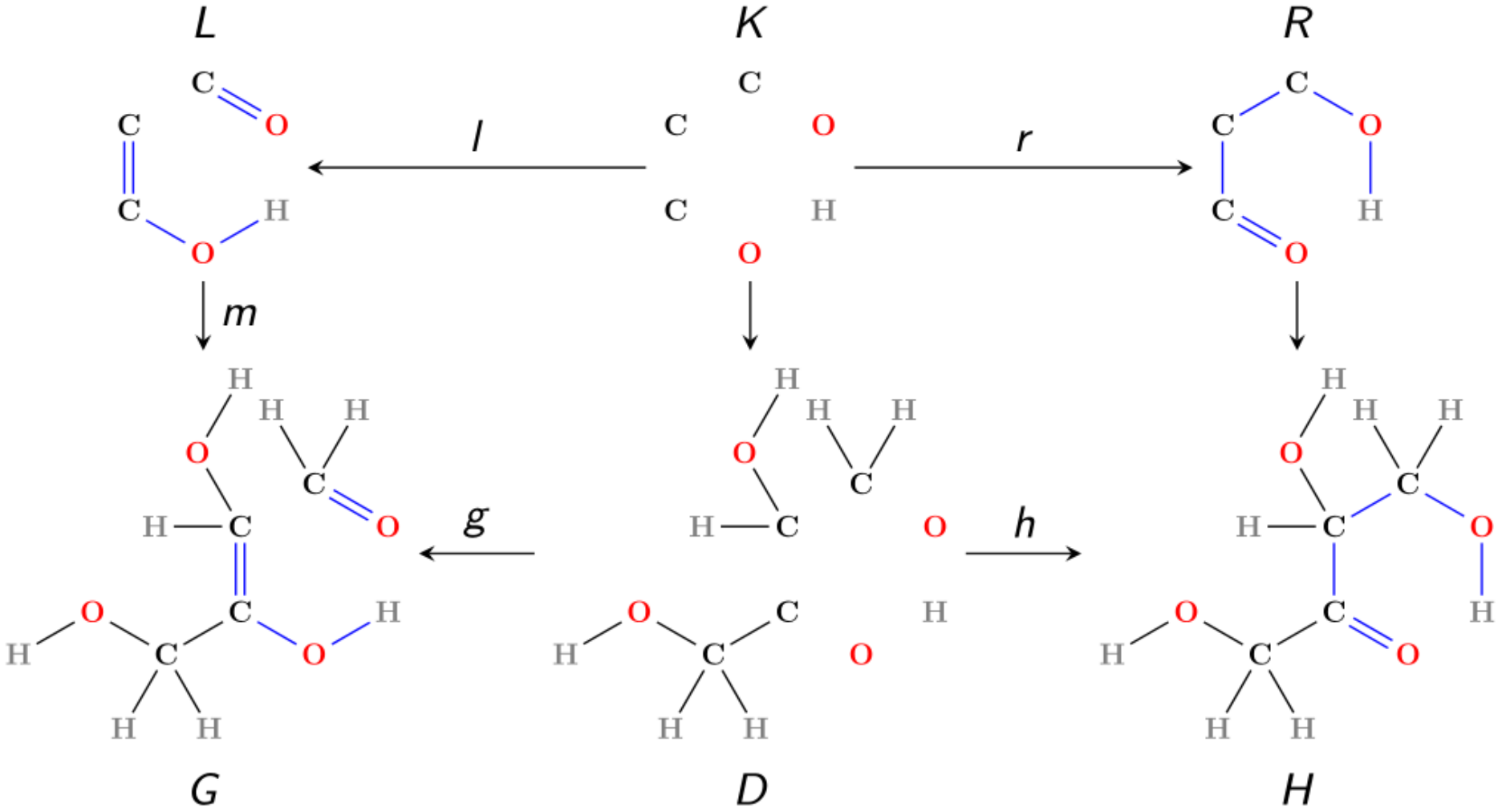


Semantics

- $L \backslash K$ is deleted.
- K is preserved.
- $R \backslash K$ is added.
- Both l and r are monomorphisms.

E.g., see [Ehrig et al., Fundamentals of Algebraic Graph Transformation, 2006]

Rule Application



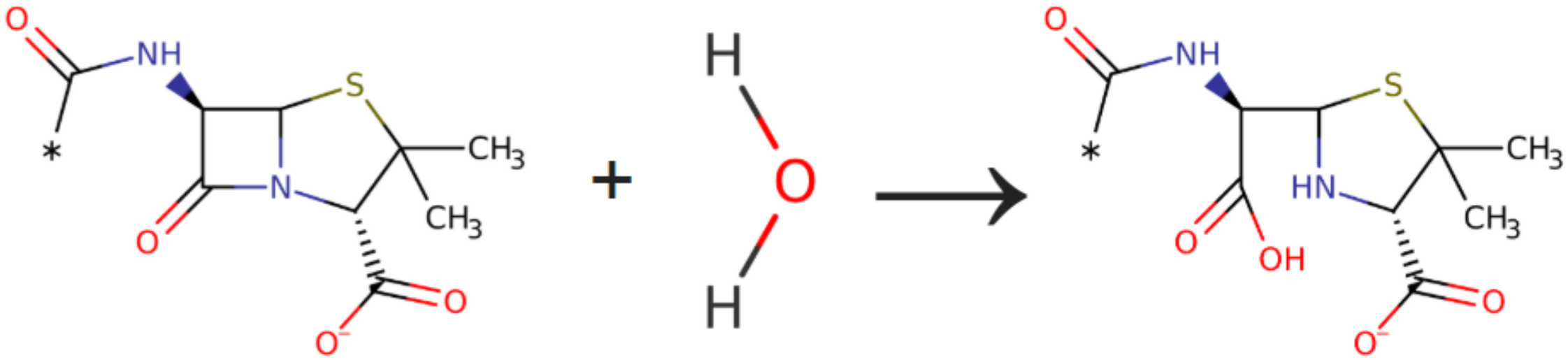
50 Shades of Rule Composition

From Chemical Reactions to Higher Levels of Abstraction

Jakob Lykke Andersen[1], Christoph Flamm[2],
Daniel Merkle[1], and Peter F. Stadler[2–7]

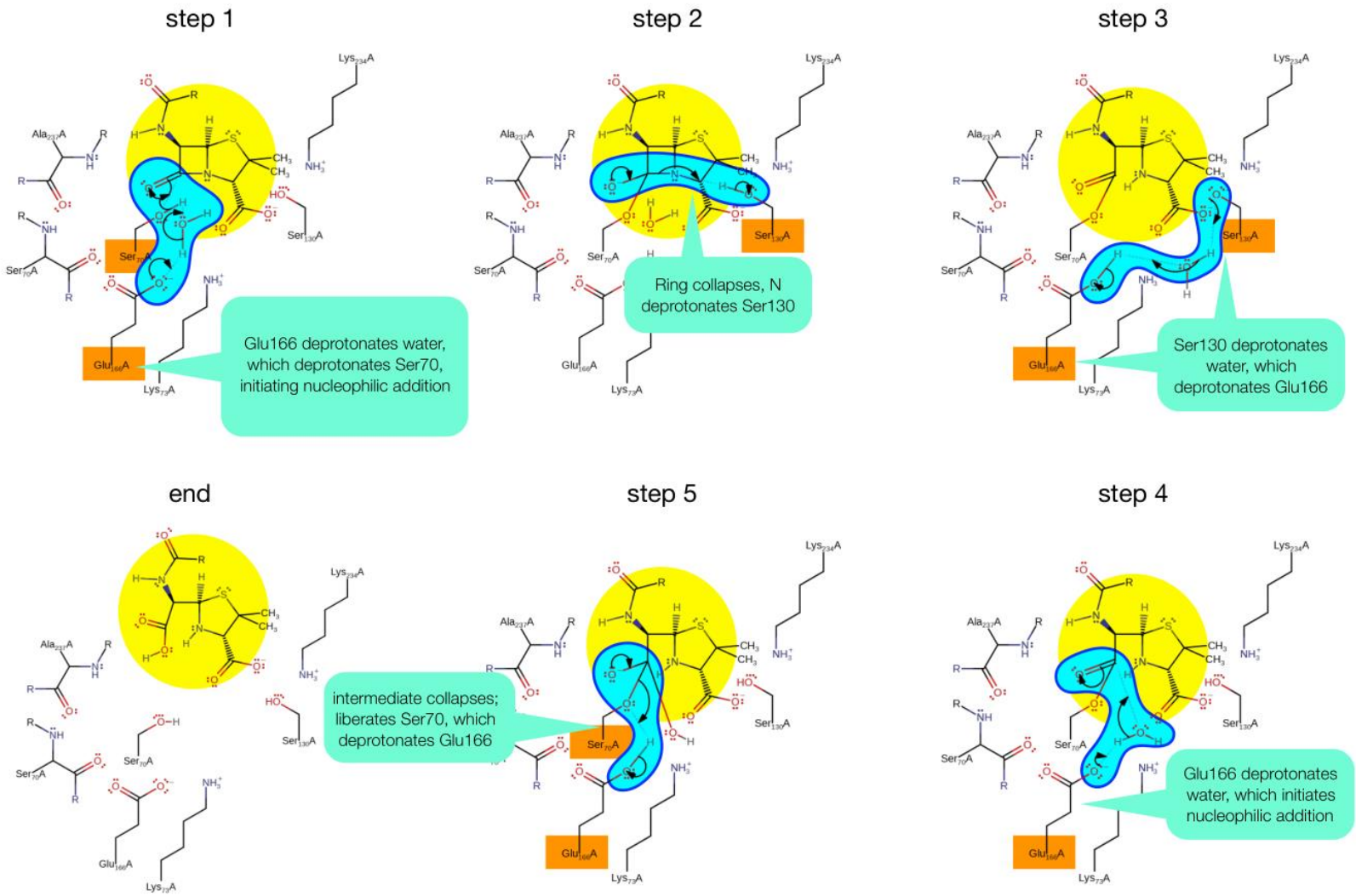
[1] Department of Mathematics and Computer Science
University of Southern Denmark, Denmark
{daniel,jlandersen}@imada.sdu.dk
[2] Institute for Theoretical Chemistry, University of Vienna, Austria
xtof@tbi.univie.ac.at
[3] Bioinformatics Group, Department of Computer Science,
and Interdisciplinary Center for Bioinformatics, Leipzig, Germany
[4] Max Planck Institute for Mathematics in the Sciences, Leipzig, Germany
[5] Fraunhofer Institute for Cell Therapy and Immunology, Leipzig, Germany
[6] Center for Non-coding RNA in Technology and Health
University of Copenhagen, Denmark
[7] Santa Fe Institute, USA
studla@bioinf.uni-leipzig.de

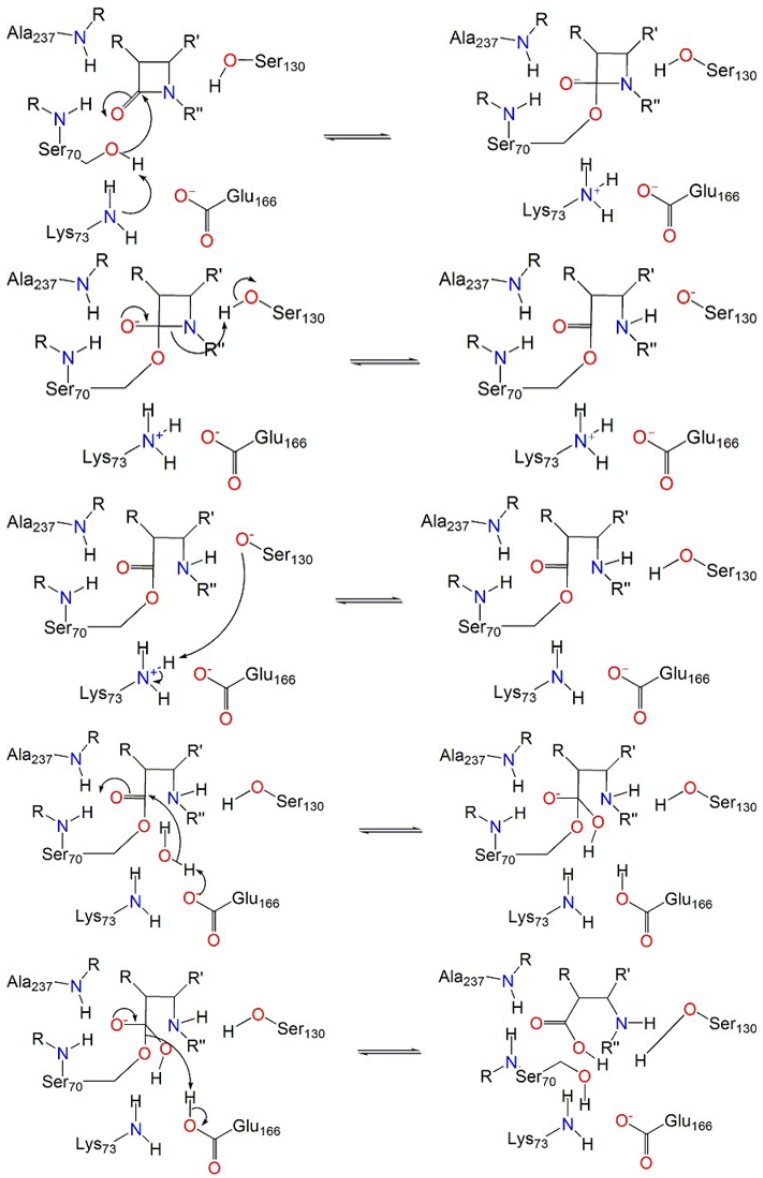
Abstract. Graph rewriting has been applied quite successfully to model

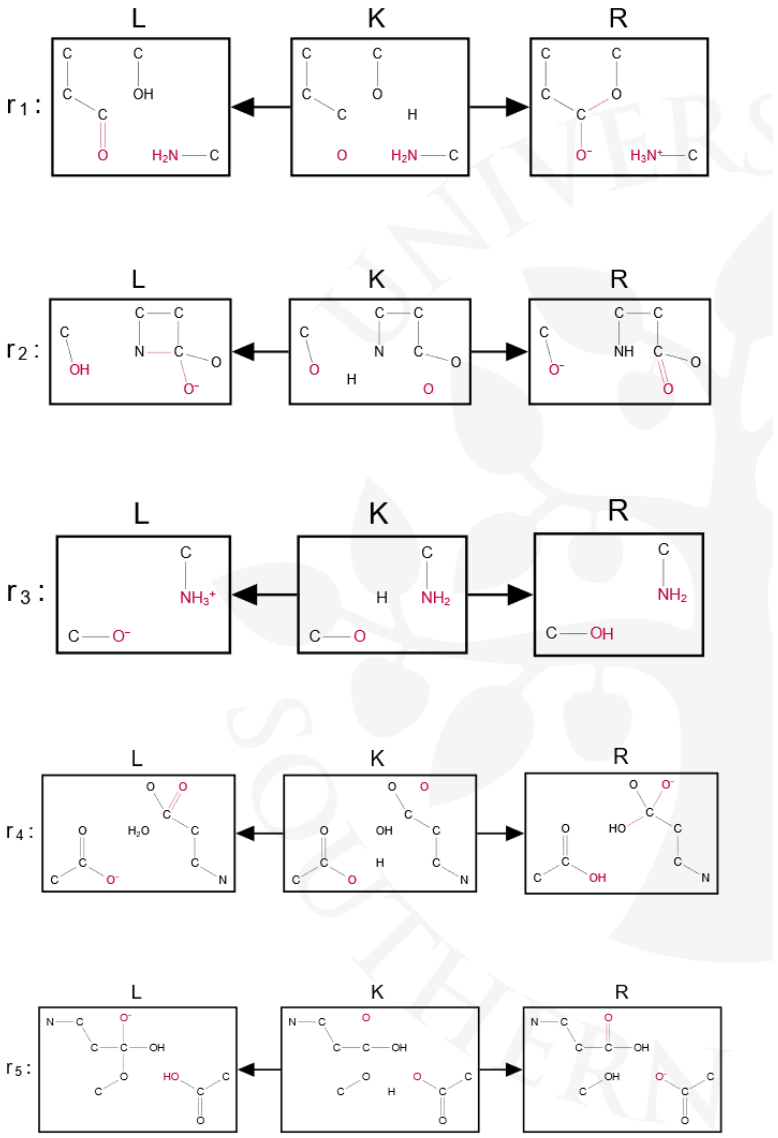


β-Lactamase

The Mechanism of β-lactamase

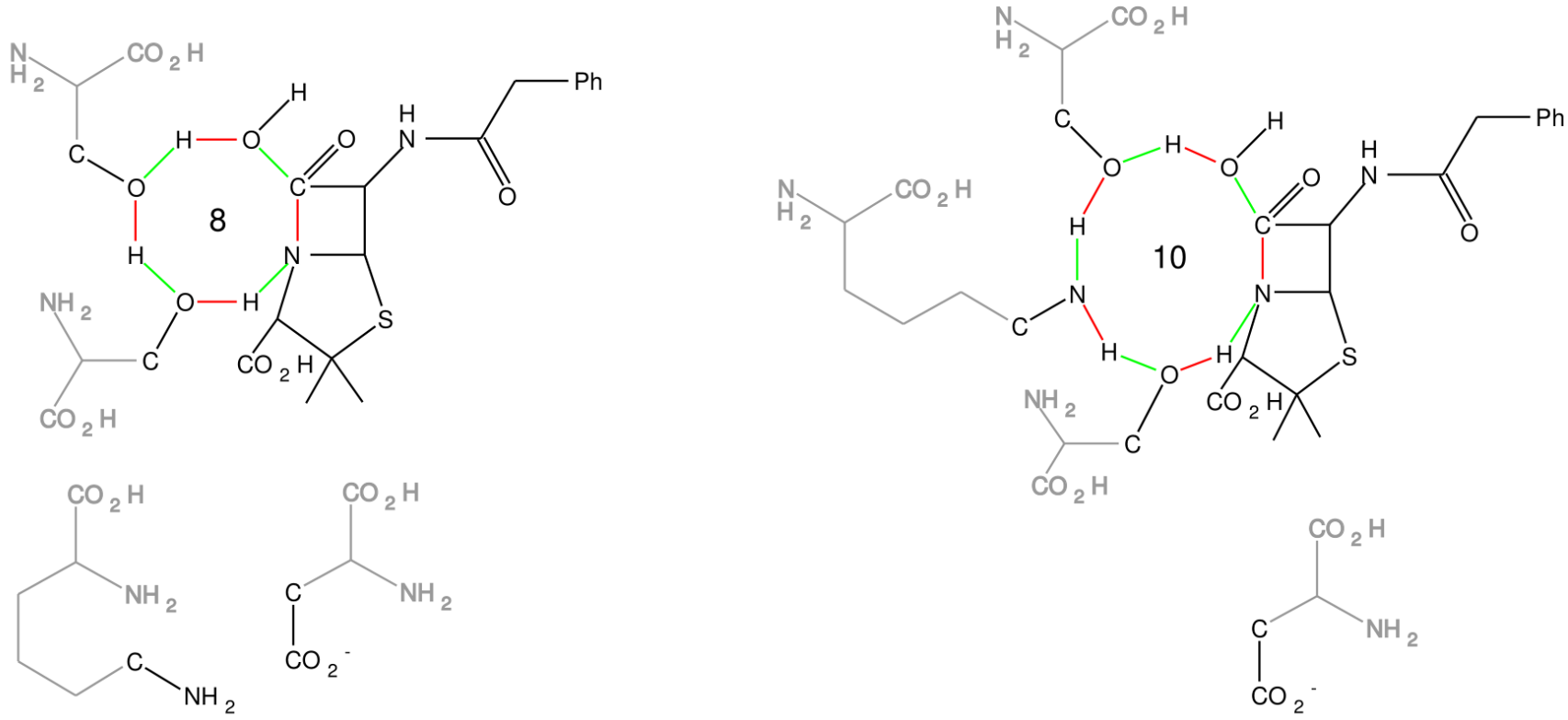



Ala237
Ser130
Ser70
Lys73
Glu166


L
K
R
r1 :
r2 :
r3 :
r4 :
r5 :

Rule Composition - β-Lactamase

$$\iota_G \circ r_1 \circ r_2 \circ r_3 \circ r_4 \circ r_5 \circ \iota_H$$



Stochastic Simulation

Rules implicitly define transitions for stochastic simulations

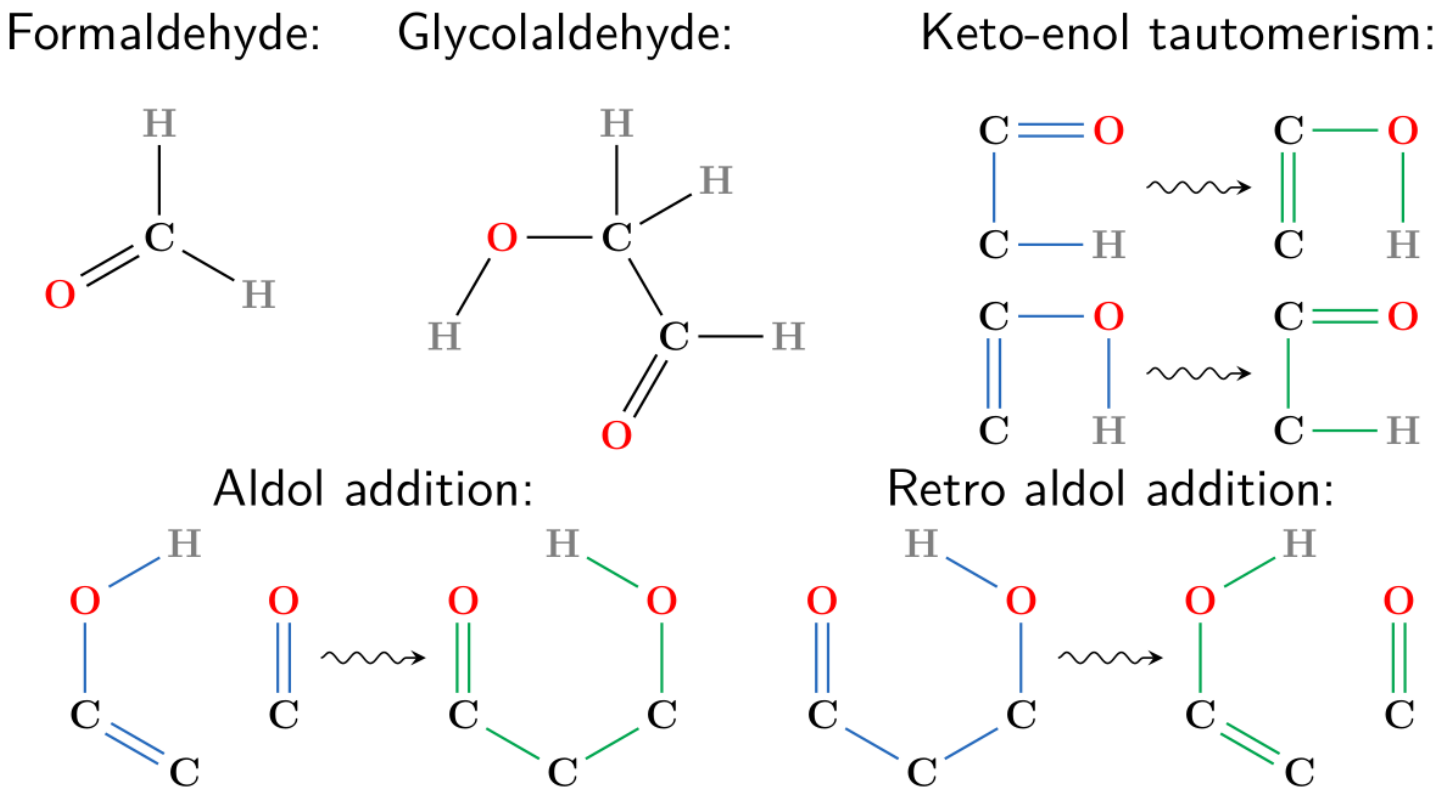
Network-free Gillespie-like simulations

- Reaction rates are needed

Can potentially be used to verify one-pot systems

Stoch-sims are in development for MØD

Reaction Network



Example: Formose

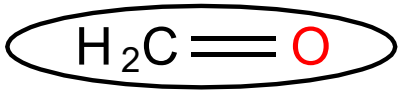
Starting graphs:

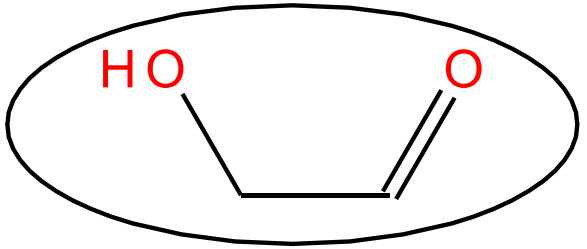
- Formaldehyde
- Glycolaldehyde

Transformation rules:

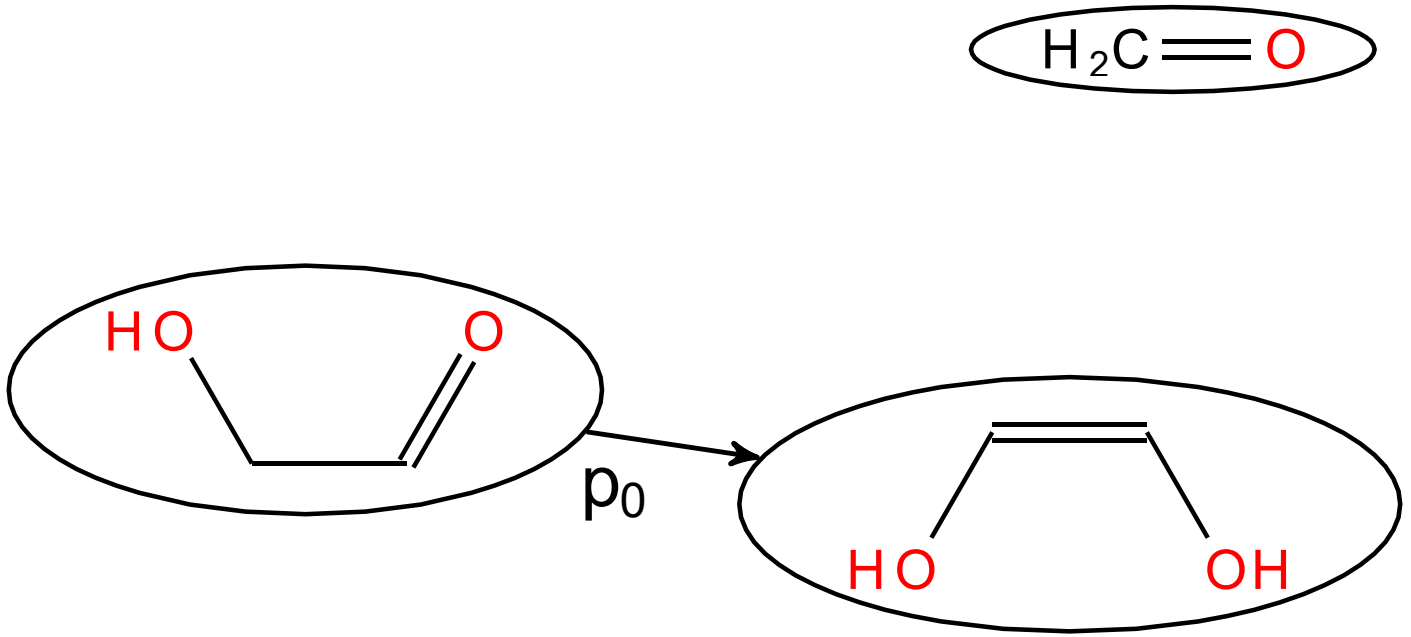
- $\mathbf{p_0}$: keto-enol-tautomerism, one direction
- $\mathbf{p_1}$: keto-enol-tautomerism, the other direction
- $\mathbf{p_2}$: aldol addition, one direction
- $\mathbf{p_3}$: aldol addition, the other direction

Generation 0

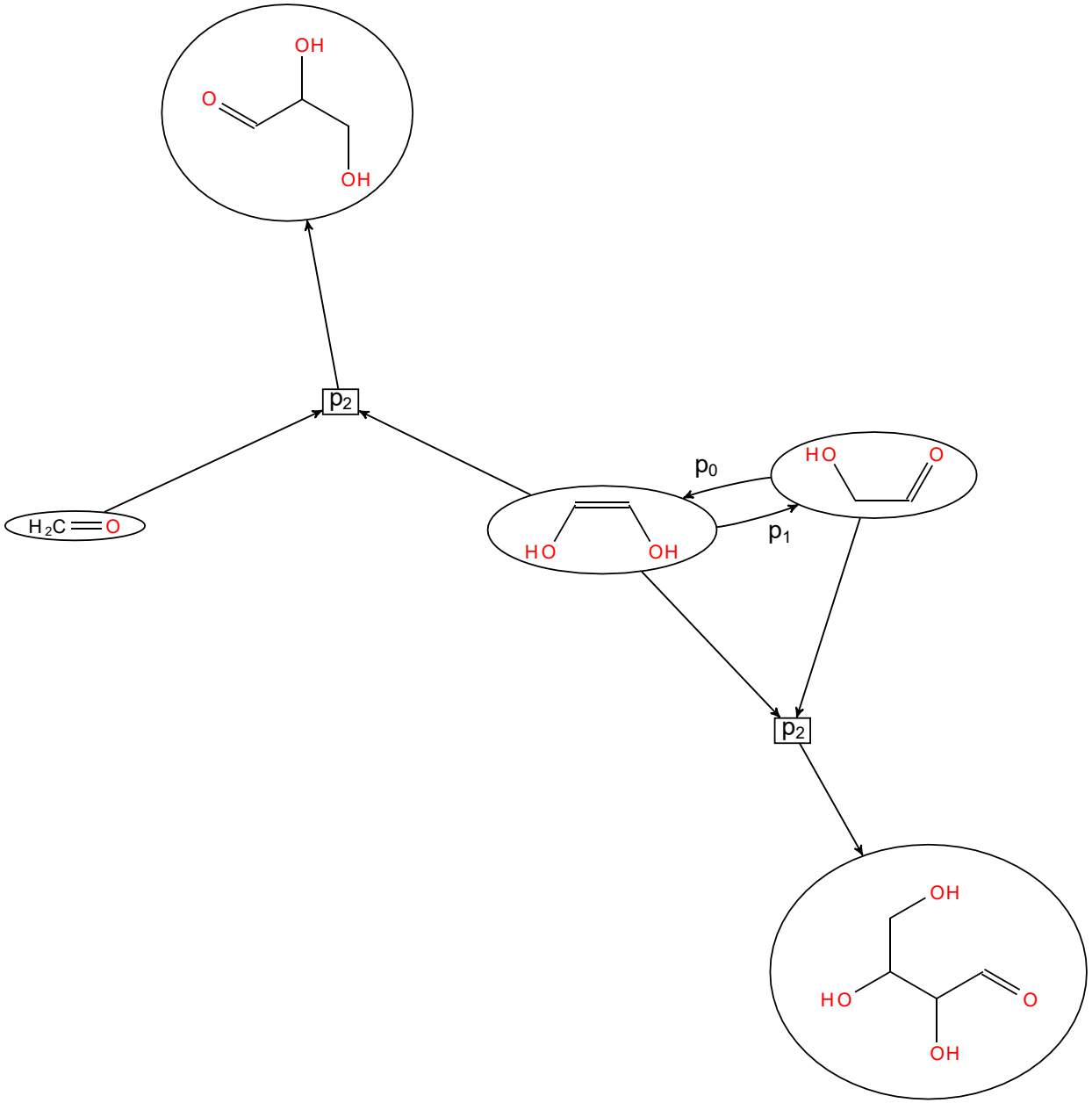




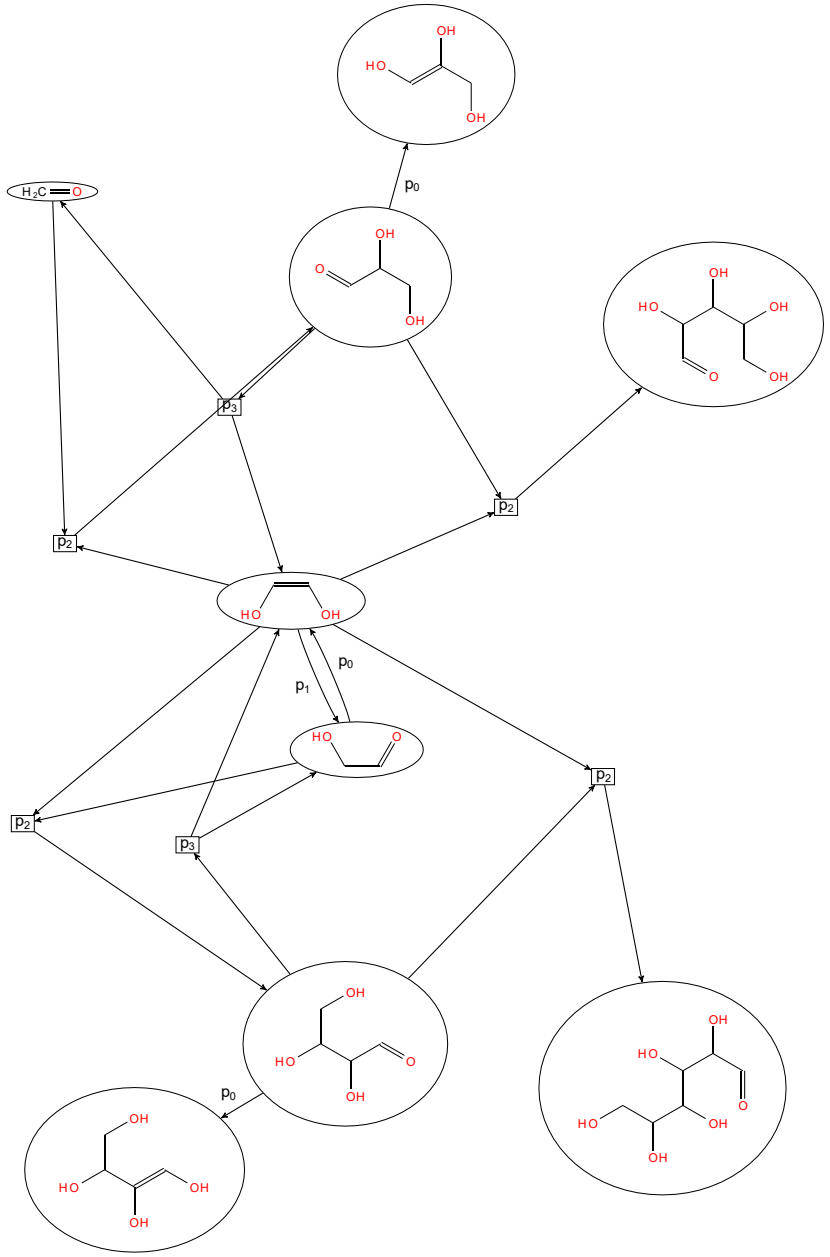
Generation 1



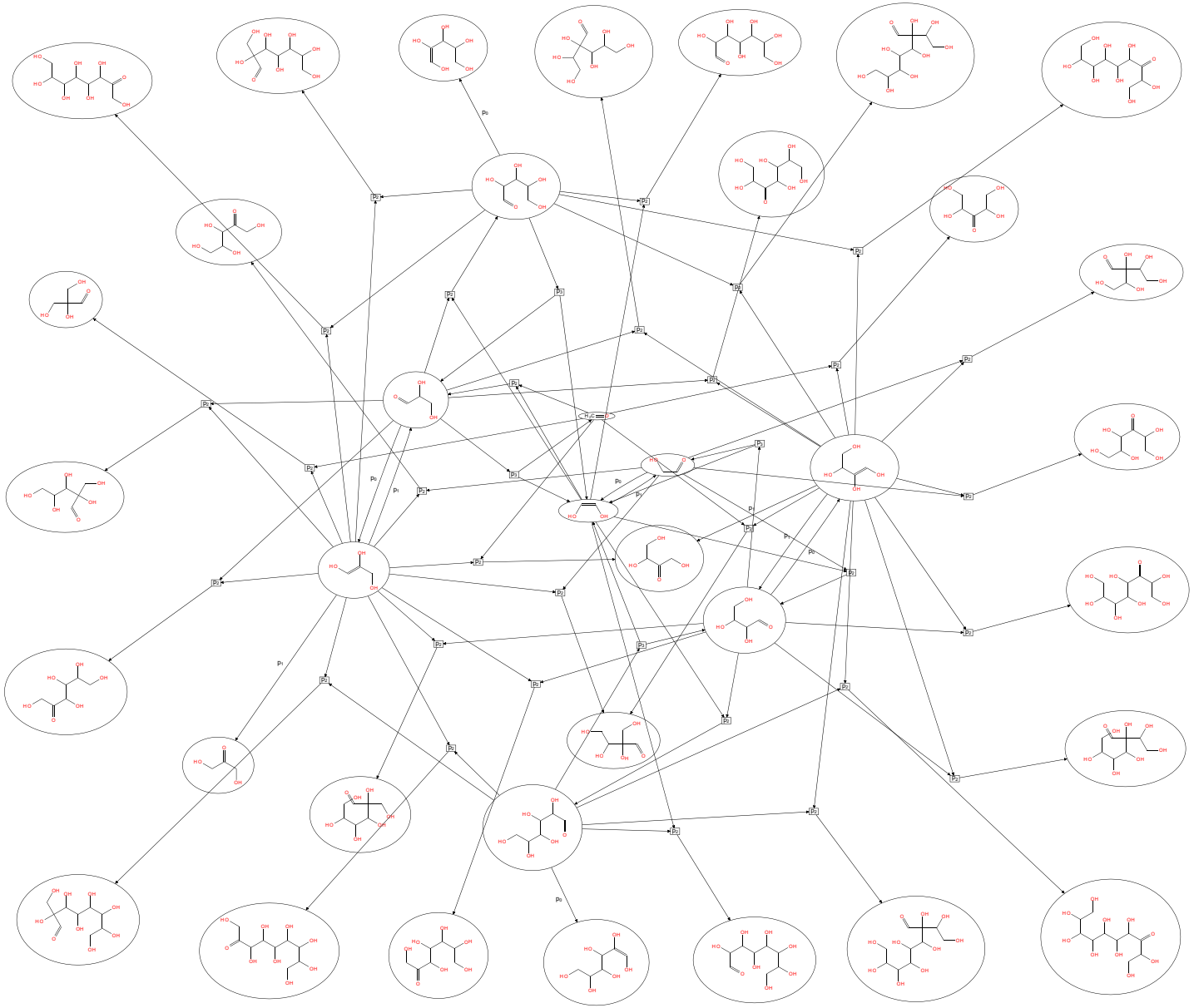
Generation 2

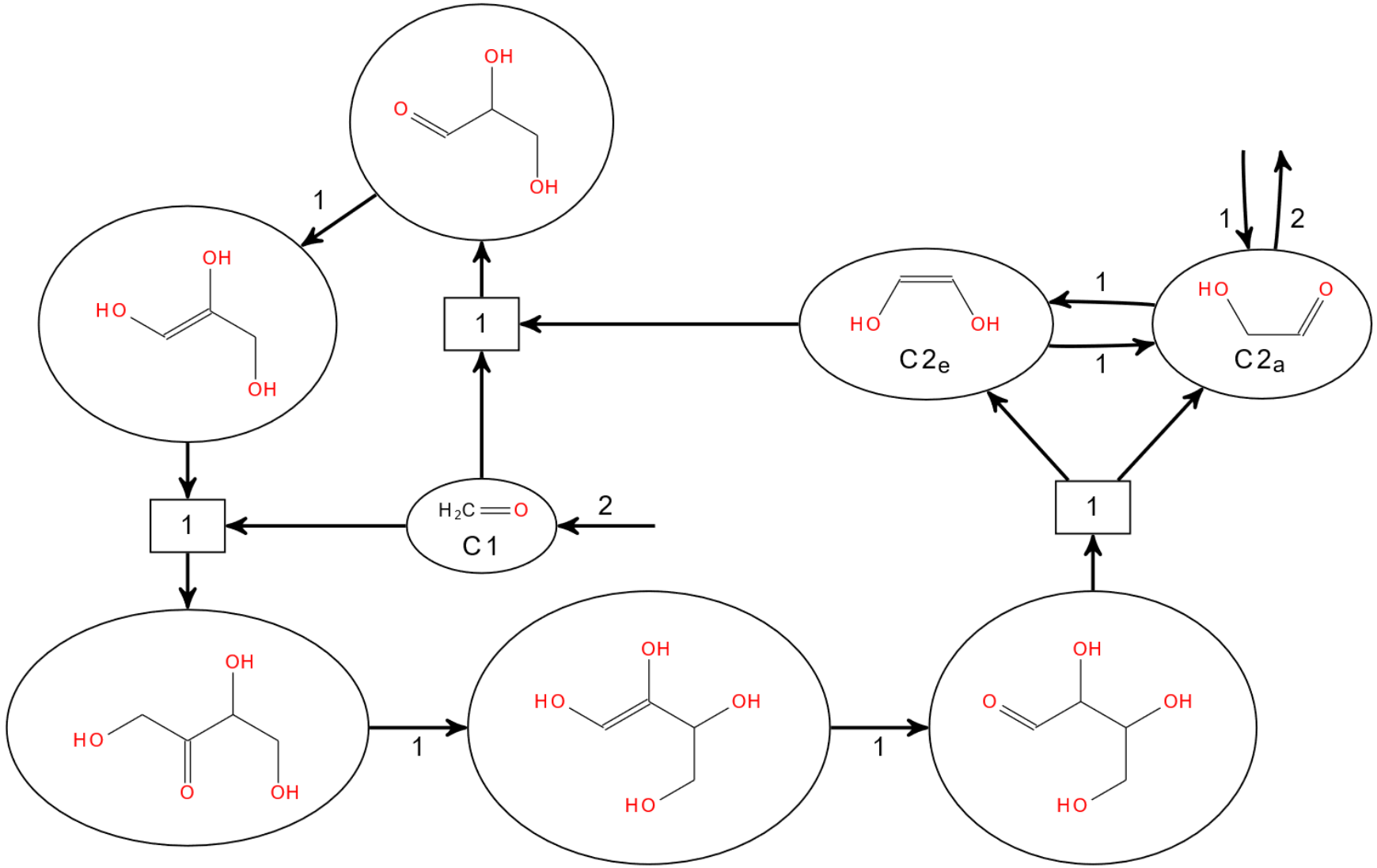


Generation 3



Generation 4




OH
O
OH
1
OH
HO
OH
1
1
2
1
HO
OH
C2e
1
HO
O
C2a
1
H2C O
C1
2
1
1
OH
HO
OH
O
OH
1
OH
HO
OH
HO
1
OH
O
OH
HO

Catalytic Mechanisms

From networks we find catalytic reactions

- What is a rigorous definition of catalysis? Is it useful?

Essential for enzyme-like action

From base rule set, compute super-set of likely amino-acid catalyzed reactions

- Can hopefully suggest amino acids in active site based on educt and product
- Will represent the non-3D embedded active site in beginning

Summary

What is already there:

- Modelling formalisms for chemical reactions
- Tools for computing reaction networks
- Drive to push the methods forward

Next steps:

- Building of “base” rule set
- Explore the nuances of enzyme mechanisms

Come talk to me about what is hard in enzymatic design!

Thanks!