Enzyme Mechanisms as Synthesis Planning Problem

Bernhard Thiel

Works from others

- Theory of Synthesis Planning
- GGL Graph grammar library by Martin Mann
- MACiE database of enzyme mechanisms
- Reaction mapping as linear program

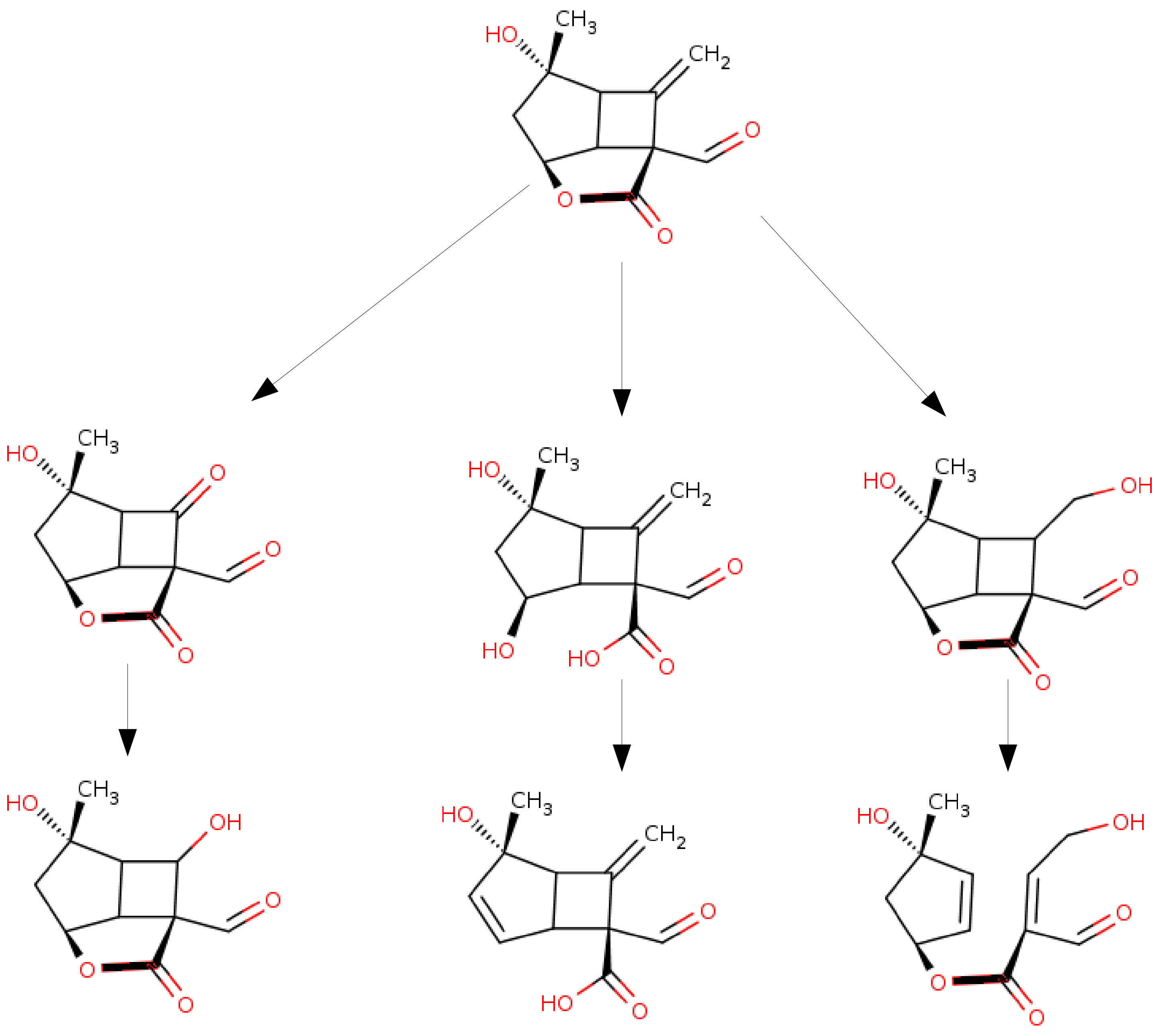
Synthesis planning problem

- Find an optimal sequence of reactions to generate a given target molecule from available starting materials

- LHASA (1970s)
- SYNGEN (1985)
- Route Designer (2009)

1.) PENSAK DAVID A. and COREY E. J. "LHASA—Logic and Heuristics Applied to Synthetic Analysis". In: *Computer-Assisted Organic Synthesis*. Chap. 2, pp. 1–32
2.) James B. Hendrickson, David L. Grier, and A. Glenn Toczko. "A logic-based program for synthesis design". In: *Journal of the American Chemical Society* 107.18 (1985), pp. 5228–5238.
3.) James and Law. "Route Designer: A Retrosynthetic Analysis Tool Utilizing Automated etrosynthetic Rule Generation". In: *J. Chem. Inf. Model.* 49 (2009), pp. 593–602.

Synthesis planning



Graph Grammar Library

- Represent **Molecules** as Boost Adjacency List
- Parse SMILES, write canonical SMILES
- Aromaticity perception

- Apply **Graph Grammar Rules** to Molecules (including SGM)
- Parse Rules from GML

Martin Mann, Heinz Ekker, Christoph Flamm
"The Graph Grammar Library – a generic framework for chemical graph rewrite systems"
In: *Theory and Practice of Model Transformations, Proc of ICMT 2013*

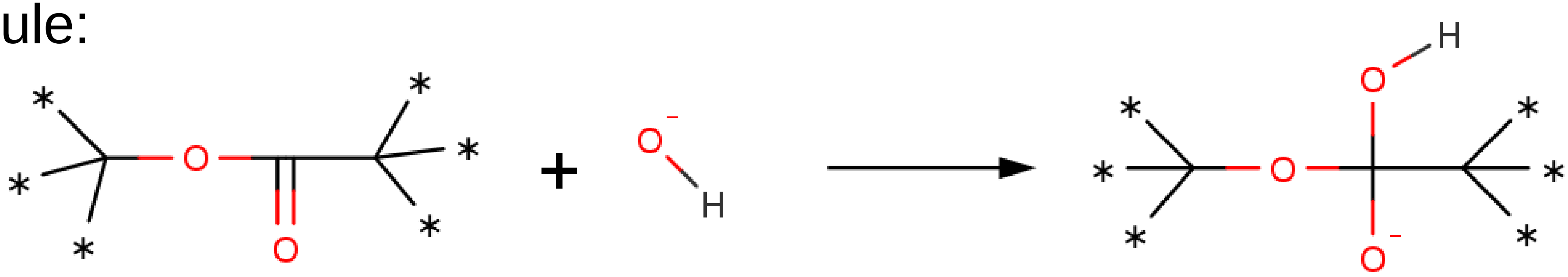
GML Rules

- Left, right
- Context
- Constraints

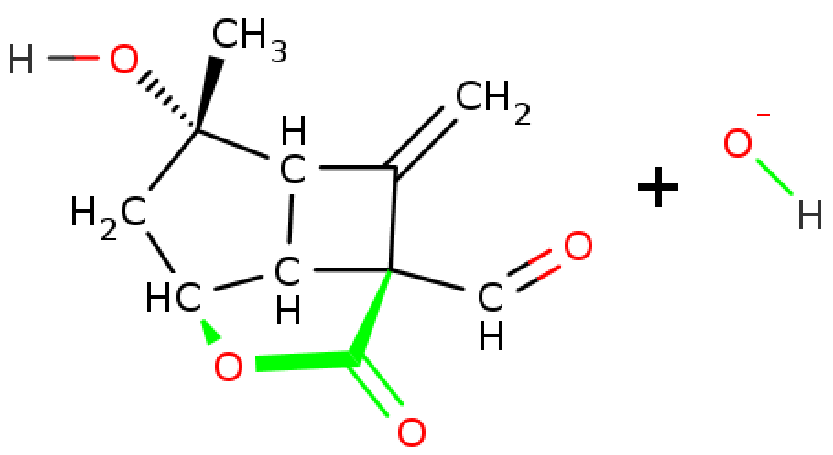
```
rule [
  ruleID "0001.stg01 condensed"
  left [
    node [ id 2 label "O-" ]
    node [ id 3 label "S" ]
    edge [ source 3 target 4 label "-" ]
  ]
  context [
    node [ id 0 label "C" ]
    node [ id 1 label "O" ]
    edge [ source 0 target 1 label "=" ]
    edge [ source 0 target 2 label "-" ]
    node [ id 4 label "H" ]
    node [ id 5 label "C" ]
    edge [ source 0 target 5 label "-" ]
    node [ id 6 label "C" ]
    edge [ source 3 target 6 label "-" ]
  ]
  right [
    node [ id 2 label "O" ]
    node [ id 3 label "S-" ]
    edge [ source 2 target 4 label "-" ]
  ]
  constrainNoEdge [ source 2 target 4 ]
]
```

Rule Application

Rule:

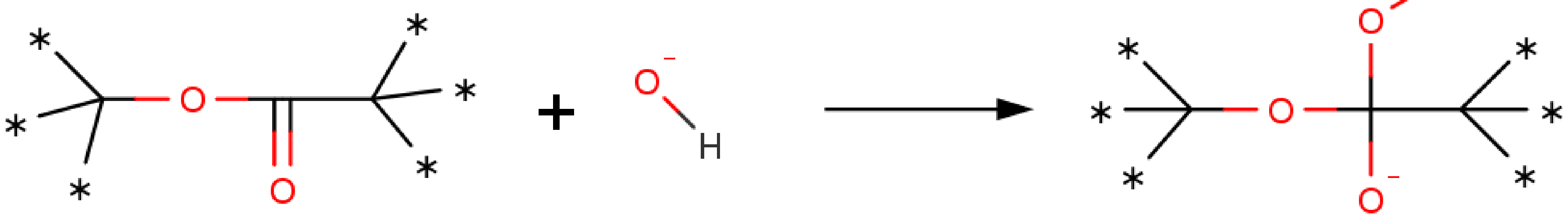


Molecules:

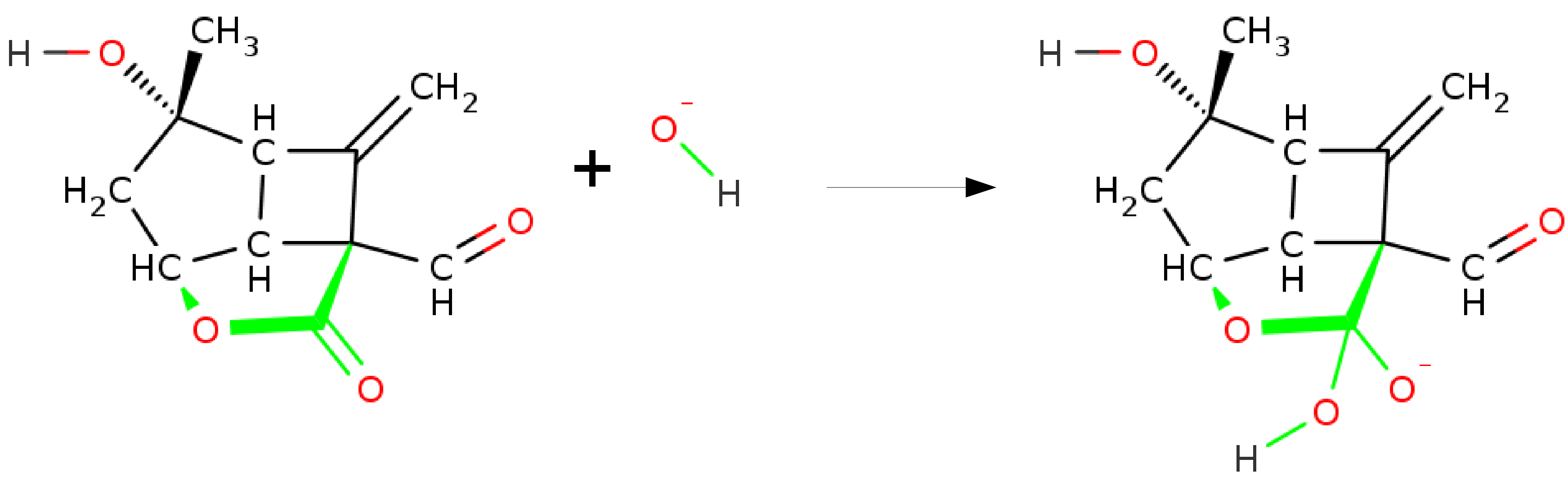


Rule Application

Rule:



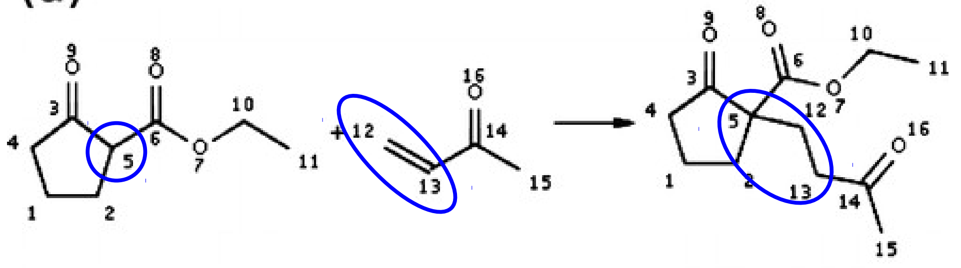
Reaction:



Rule extraction

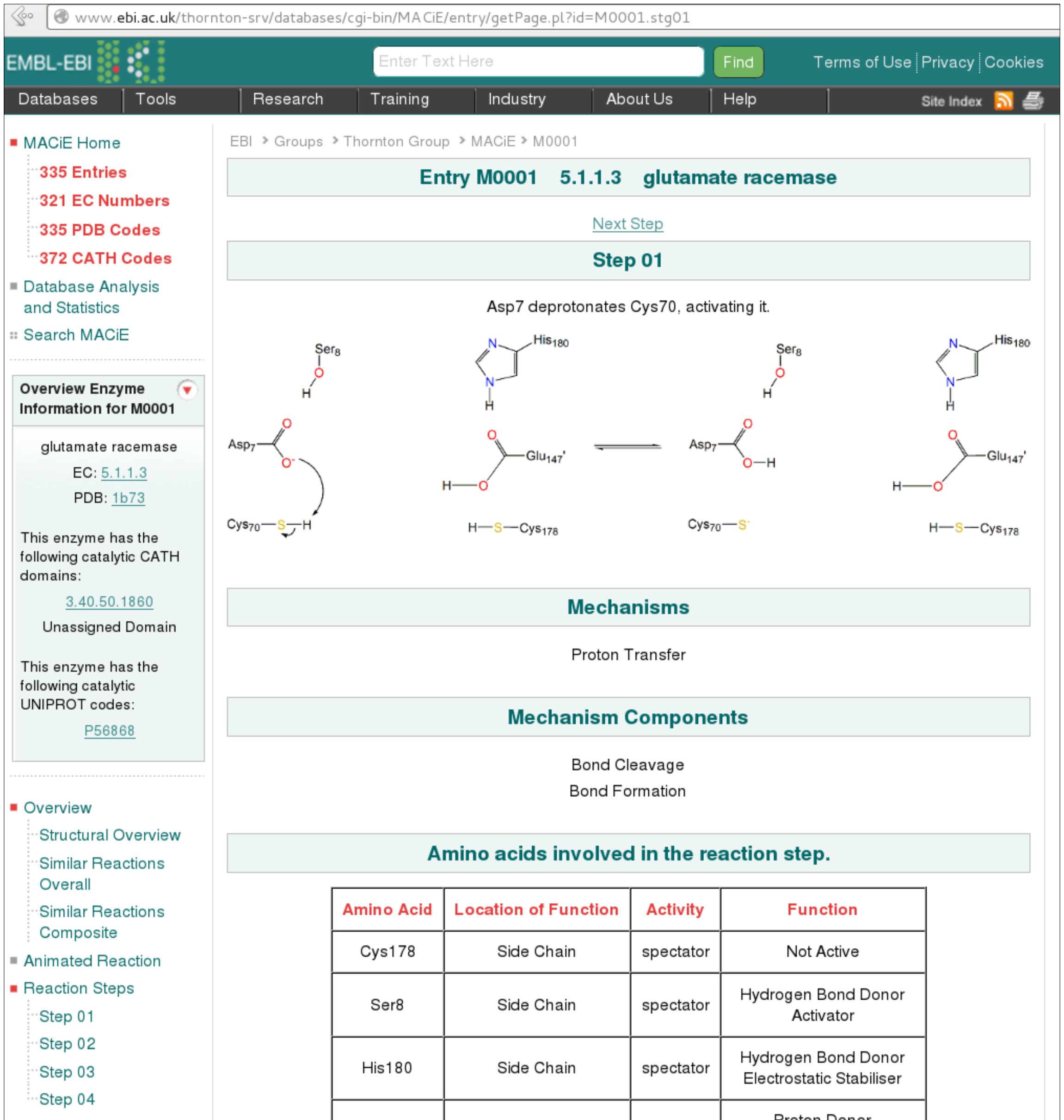
- Get reaction from database
- Find extended reaction core

(a)



3.) James and Law. “Route Designer: A Retrosynthetic Analysis Tool Utilizing Automated etrosynthetic Rule Generation”. In: *J. Chem. Inf. Model.* 49 (2009), pp. 593–602.

MACiE



www.ebi.ac.uk/thornton-srv/databases/cgi-bin/MACiE/entry/getPage.pl?id=M0001.stg01

EMBL-EBI | Enter Text Here | Find | Terms of Use | Privacy | Cookies

Databases | Tools | Research | Training | Industry | About Us | Help | Site Index

- MACiE Home
  - 335 Entries
  - 321 EC Numbers
  - 335 PDB Codes
  - 372 CATH Codes
- Database Analysis and Statistics
- Search MACiE

Overview Enzyme Information for M0001

glutamate racemase

EC: 5.1.1.3

PDB: 1b73

This enzyme has the following catalytic CATH domains:

3.40.50.1860

Unassigned Domain

This enzyme has the following catalytic UNIPROT codes:

P56868

- Overview
  - Structural Overview
  - Similar Reactions Overall
  - Similar Reactions Composite
- Animated Reaction
- Reaction Steps
  - Step 01
  - Step 02
  - Step 03
  - Step 04

EBI > Groups > Thornton Group > MACiE > M0001

Entry M0001 5.1.1.3 glutamate racemase

Next Step

Step 01

Asp7 deprotonates Cys70, activating it.

Mechanisms

Proton Transfer

Mechanism Components

Bond Cleavage

Bond Formation

Amino acids involved in the reaction step.

Amino Acid	Location of Function	Activity	Function
Cys178	Side Chain	spectator	Not Active
Ser8	Side Chain	spectator	Hydrogen Bond Donor Activator
His180	Side Chain	spectator	Hydrogen Bond Donor Electrostatic Stabiliser
Cys70	Side Chain	reactant	Proton Donor

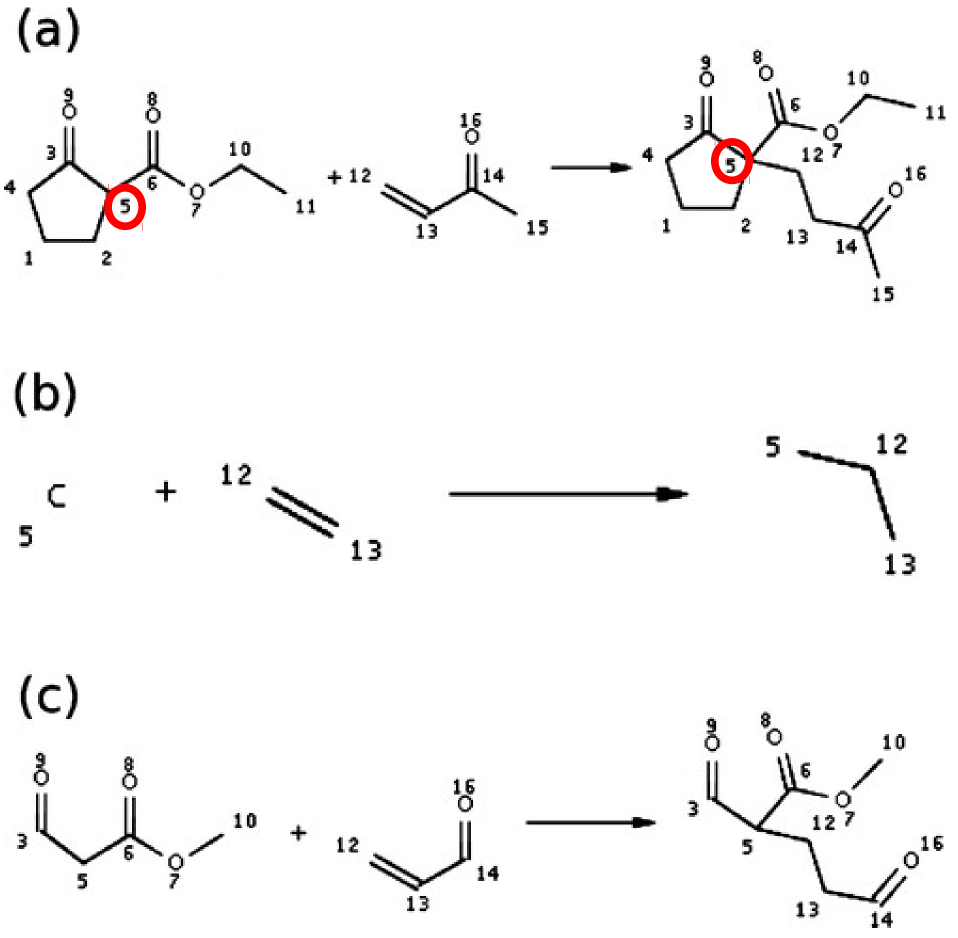
"MACiE: exploring the diversity of biochemical reactions."

G. L. Holliday, C. Andreini, J. D. Fischer, S. A. Rahman, D. E. Almonacid, S. T. Williams and W. R. Pearson.

Nucleic Acids Research, 40, D783-D789, 2012.

Rule extraction

- Get reaction from database
- Find extended reaction core



3.) James and Law. “Route Designer: A Retrosynthetic Analysis Tool Utilizing Automated etrosynthetic Rule Generation”. In: *J. Chem. Inf. Model.* 49 (2009), pp. 593–602.

Integer Linear program

- Minimize a linear equation $\mathbf{c}^T\mathbf{x}$
- Subject to linear constraints $\mathbf{A}x \leq \mathbf{b}$
- x_i within bounds, (integers)

$$\zeta = \min \sum_{(i,j)\in B_R} (1 - \sum_{(k,l)\in B_P} \alpha_{ijkl}) + \sum_{(k,l)\in B_P} (1 - \sum_{(i,j)\in B_R} \alpha_{ijkl}) + 2 \sum_{i\in C_R} \sum_{k\in C_P} \beta_{ik} + 2 \sum_{(i,j)\in D_R} \sum_{(k,l)\in D_P} \gamma_{ijkl} \quad (1)$$

$$\text{s.t.} \sum_{k\in A} y_{ik} = 1 \quad \forall i \in A \quad (2)$$

$$\sum_{i\in A}$$

$$y_{ik} = 0 \quad \forall i,k \in A : T_R^i \neq T_P^k \quad (4)$$

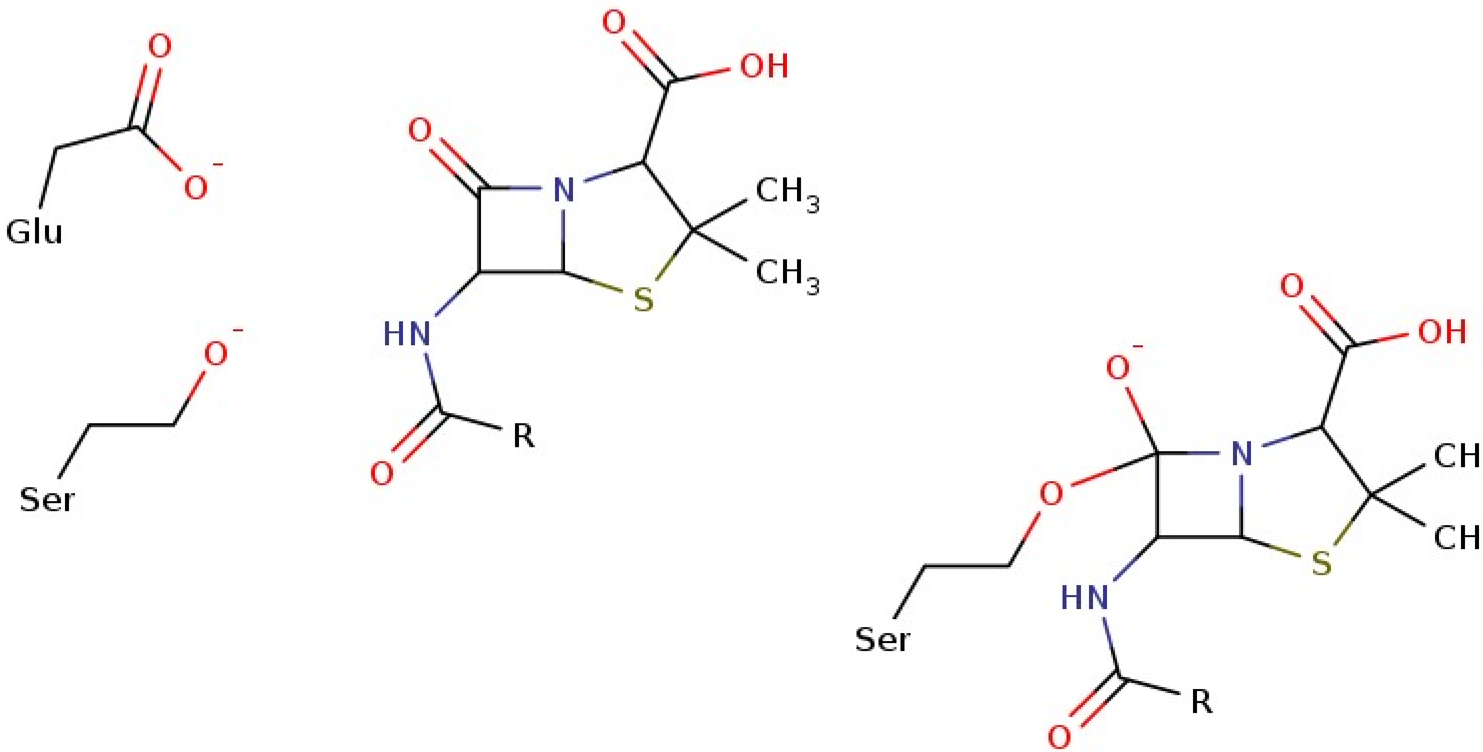
$$\alpha_{ijkl} \leq y_{ik} + y_{il} \quad \forall \in B_R \quad \forall (k,l) \in B_P \quad (5)$$

$$\alpha_{ijkl} \leq y_{jk} + y_{jl} \quad \forall (i,j) \in B_R \quad \forall (k,l) \in B_P \quad (6)$$

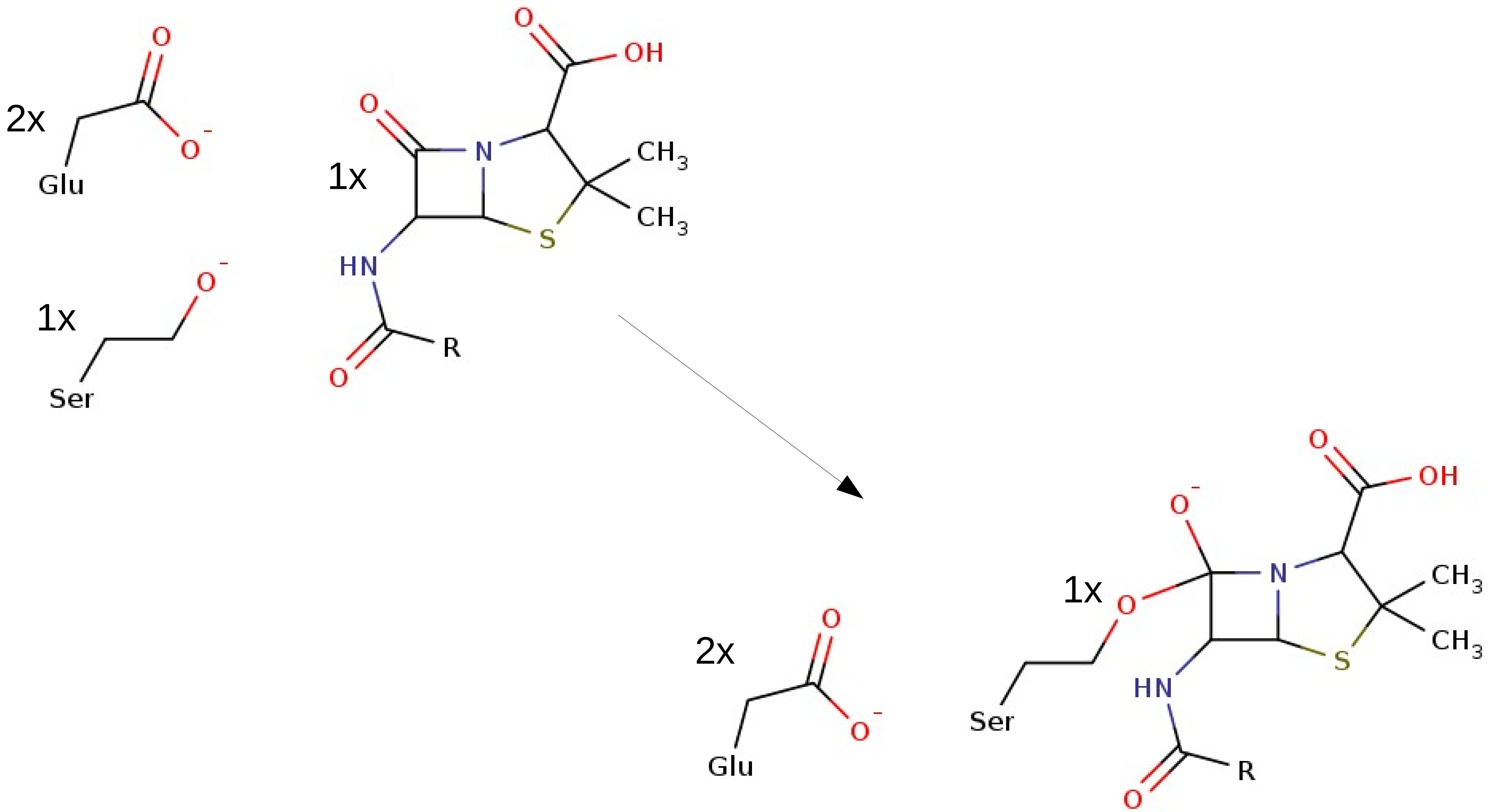
Eric L. First, Chrysanthos E. Gounaris, and Christodoulos A. Floudas.“Stereochemically Consistent Reaction Mapping and Identification of Multiple Reaction Mechanisms through Integer Linear Optimization”. In:*Journal of Chemical Information and Modelin*g 52.1 (2012), pp. 84–92

Enzyme Mechanism as Synthesis Planning Problem

Reaction in Solution



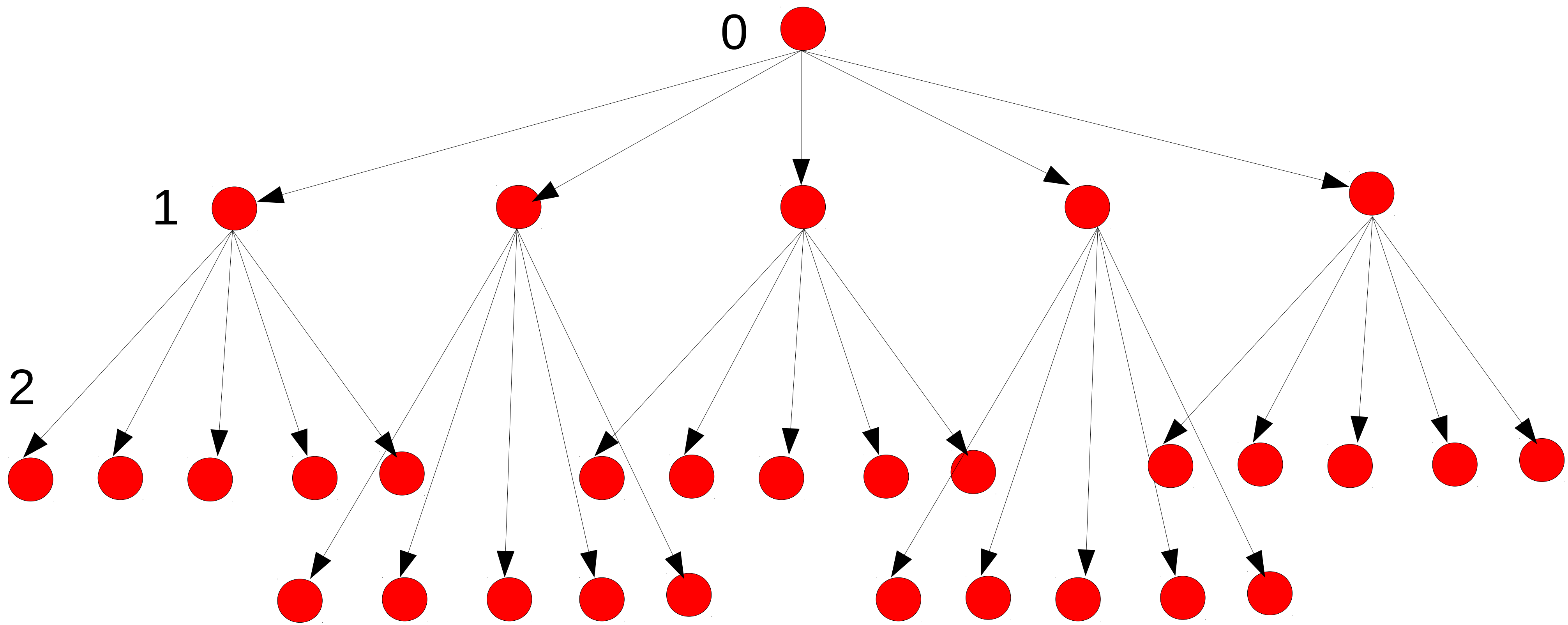
Reaction in Enzyme



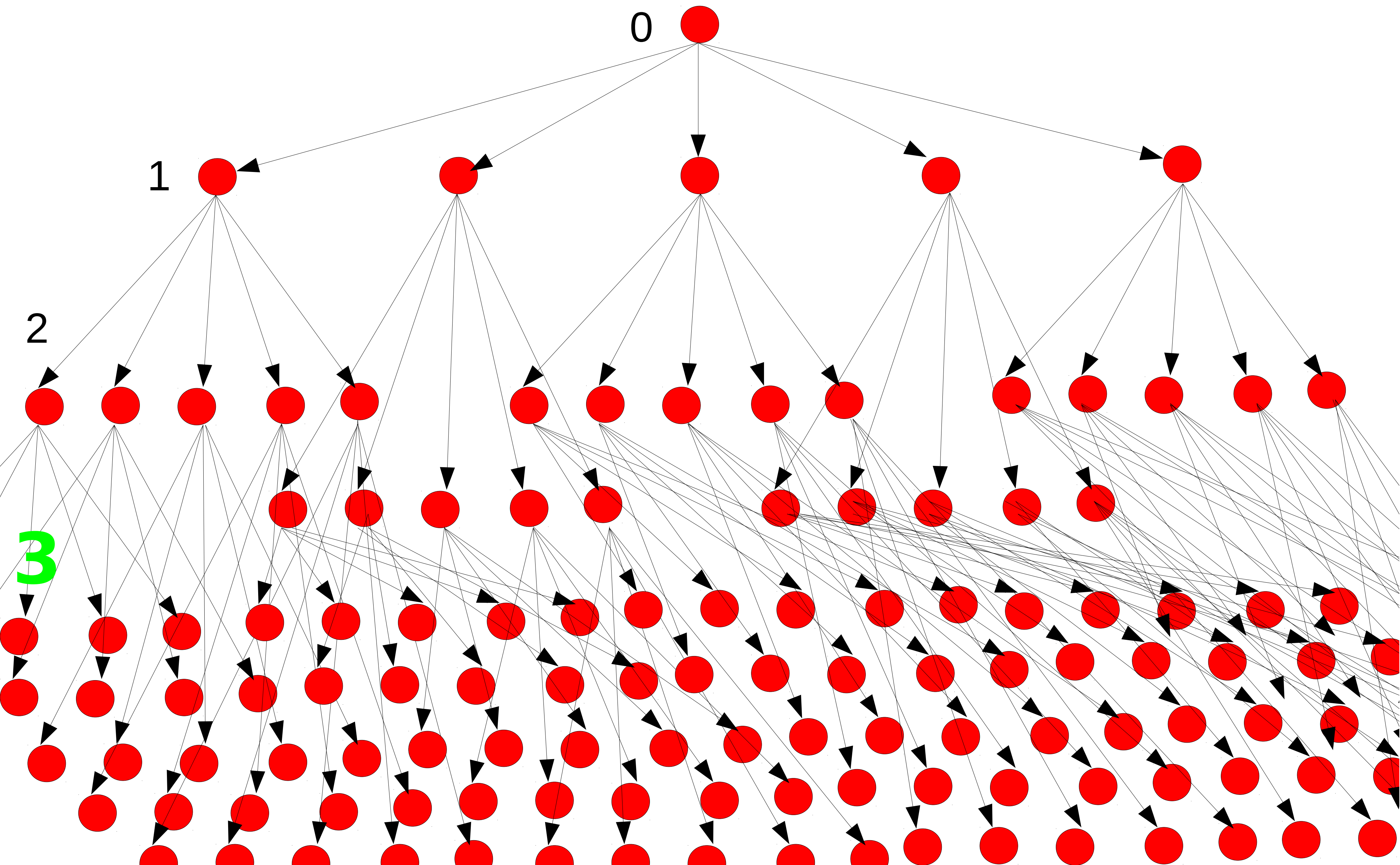
Bidirectional Search

- Only search half the total depth from each side.
- Substrate(s) transferred to product(s)
- Cofactors might be transformed
- Amino acid side chains (from the enzyme) can react in intermediate steps but have to be restored in the end.

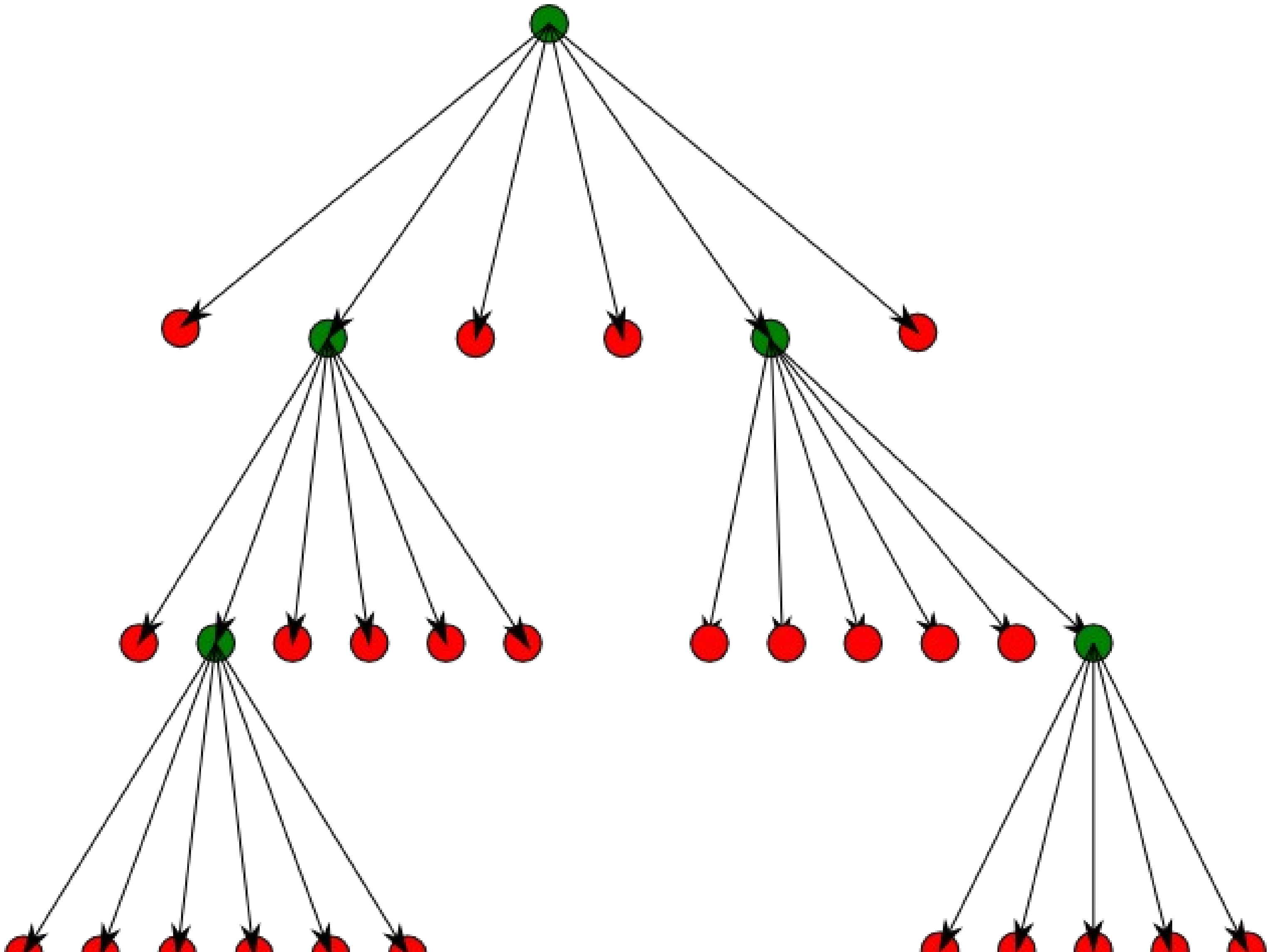
Combinatorial explosion



Combinatorial explosion



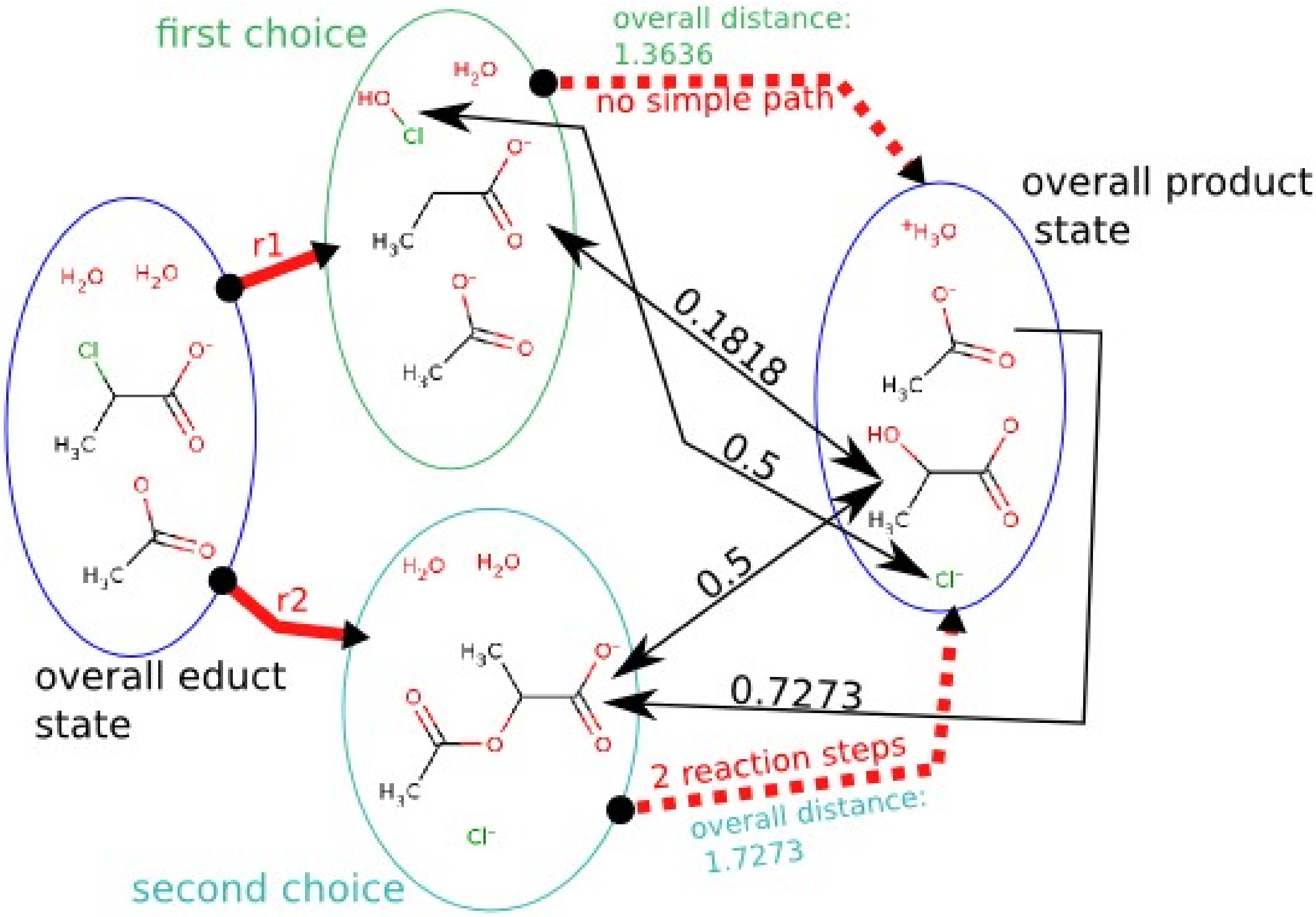
Heuristic



Heuristic

- Catalytic residues can react and be restored several times
- **Substrates** should not be restored, once they have reacted.
- **Products** should not be destroyed/ modified again once they are created.

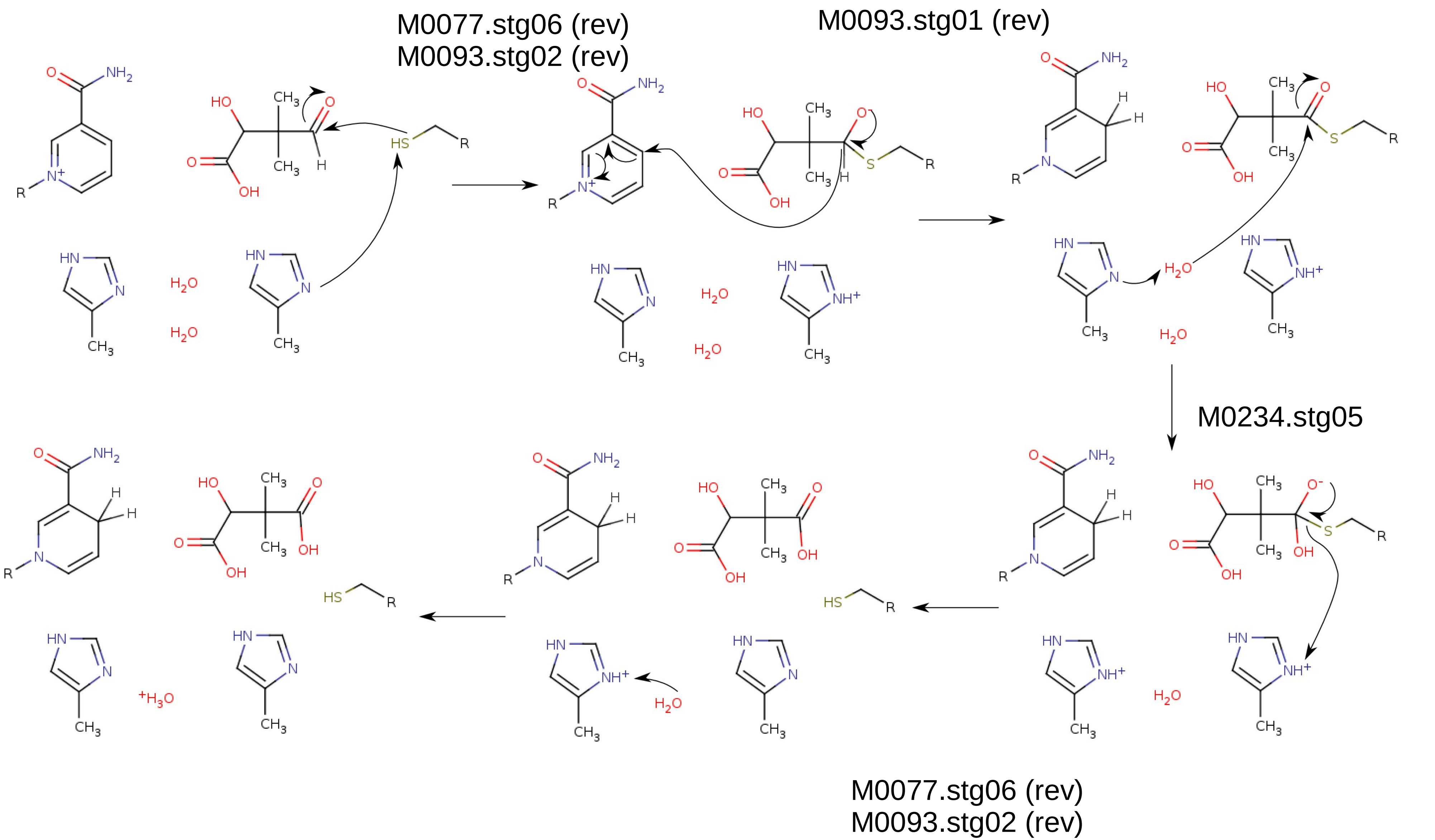
A distance between States



Application & Results

- Applied to some orphan enzymes

Results


M0077.stg06 (rev)
M0093.stg02 (rev)
M0093.stg01 (rev)
M0234.stg05
M0077.stg06 (rev)
M0093.stg02 (rev)