Graph Transformations and Algorithmic Cheminformatics for an Atom-Level Modelling of Metabolic Networks

Daniel Merkle

Department of Mathematics and Computer Science, University of Southern Denmark

Metabolism and mathematical models: Two for a tango
November 19, 2021

Pentose Phosphate Pathway

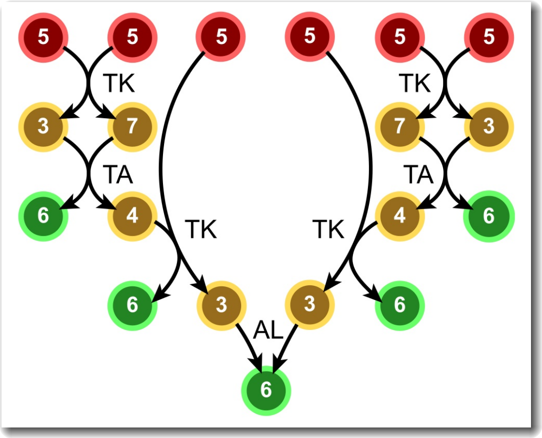
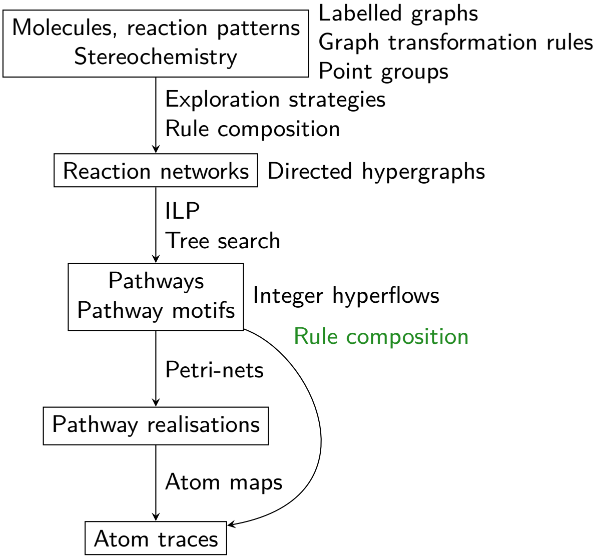


Figure adapted from Noor, E et al (2011) Central Carbon Metabolism as a Minimal Biochemical Walk between Precursors for Biomass and Energy, *J Mol Cell* **39**:809-820 DOI:10.1016/j.molcel.2010.08.031

A Graph Transformation Framework (for (Bio-)Chemistry)



Category theory
Double Pushout
Rule composition
Mono- and Isomorphisms
Canonicalisation
Automorphisms
Quantum Chemistry

Software package: MØD

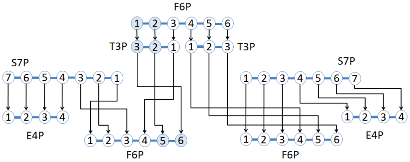
C++, Python, Bash, LaTeX

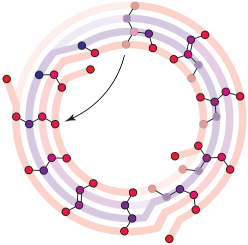
Pentose phosphate pathway
Formose
Glycolysis (EMP and ED)
Non-oxidative glycolysis
Citric acid cycle
Enzyme mechanisms
Prebiotic chemistry (HCN)
Eschenmoser's GLX scenario
DNA templated computing

https://cheminf.imada.sdu.dk

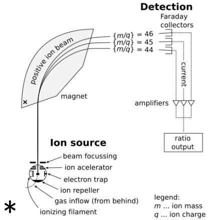
Graph Transformation Applications - Examples

- Isotope labelling experiments

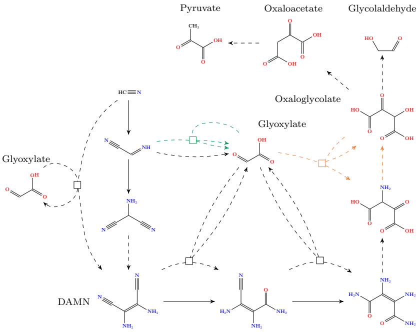




- Mass spectrometry



- Hypothetical (prebiotic) chemistries
- Synthesis Planning
- Multi-enzymatic cascades design
- Enzyme design
- Microbiome analysis and design
- . . . anything with an underlying Chemical Reaction Network

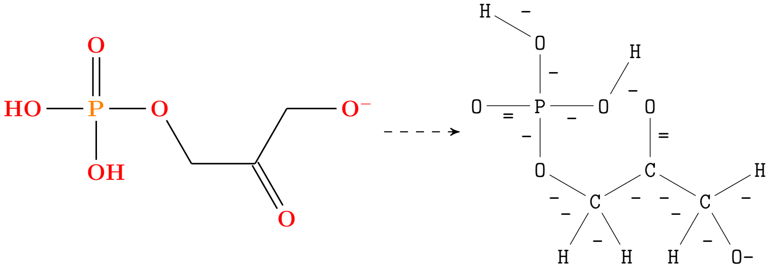


Modelling and Analysis of Chemical Systems

Modelling and Analysis of Chemical Systems

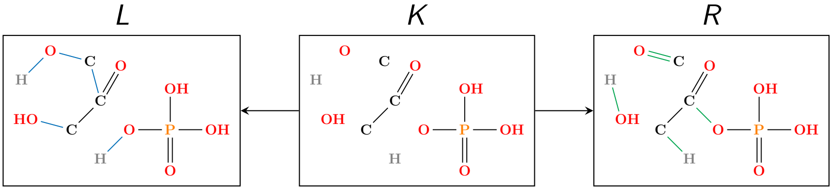
Model molecules as labelled graphs.

- An old idea: [J. J. Sylvester, *Chemistry and Algebra*, Nature 1878]
- Molecule: simple, connected, labelled graph.
- Vertex labels: atom type, charge.
- Edge labels: bond type.



Modelling and Analysis of Chemical Systems

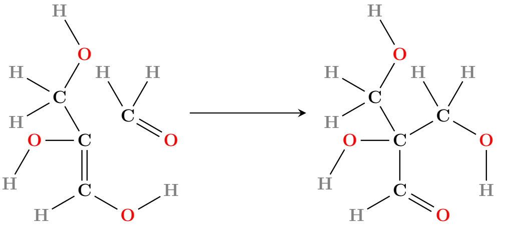
2. Model reaction types and graph transformation rules.

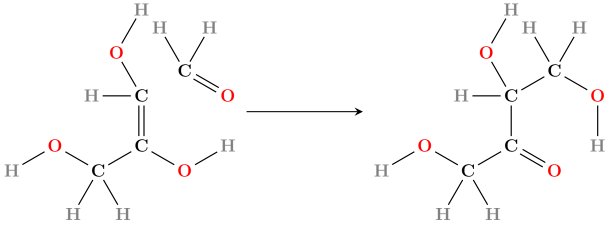


Example: Carbon rearrangement

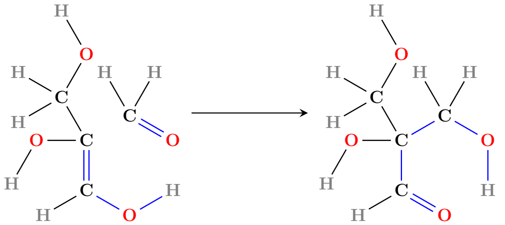
- Aldolase: ketone + aldehyde $\longrightarrow$ ketone
- Aldose-Ketose: aldehyde $\longrightarrow$ ketone
- Ketose-Aldose: ketone $\longrightarrow$ aldehyde
- Phosphohydrolase: H_2O+CnP $\longrightarrow$ Cn+Pi
- Phosphoketolase Pi+ketone $\longrightarrow$ carbonyl + CnP+water
- Transaldolase: Cn+Cm$\longrightarrow$ C(n+3)+C(m-3)
- Transketolase: Cn+Cm$\longrightarrow$ C(n+2)+C(m-2)

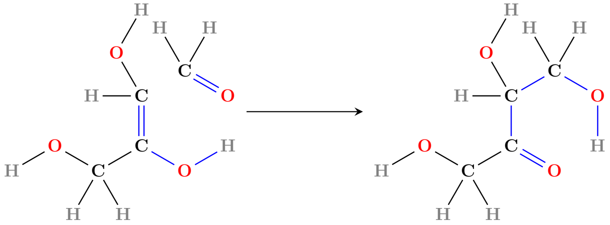
Chemical Reactions (Educts → Products)





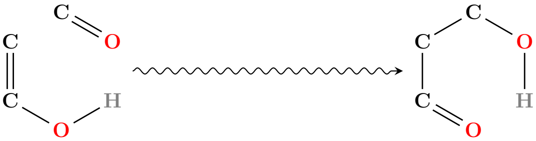
Chemical Reactions (of the Same Type)

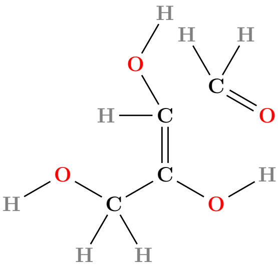




Chemical Reaction Patterns

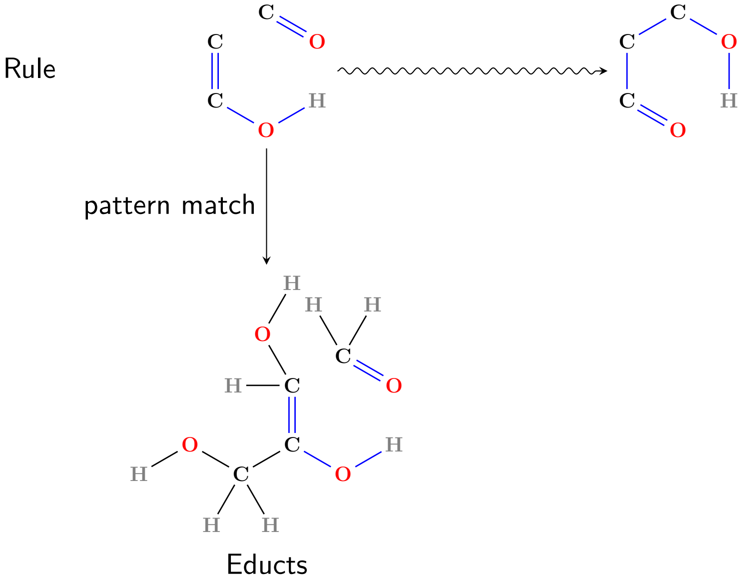
Rule



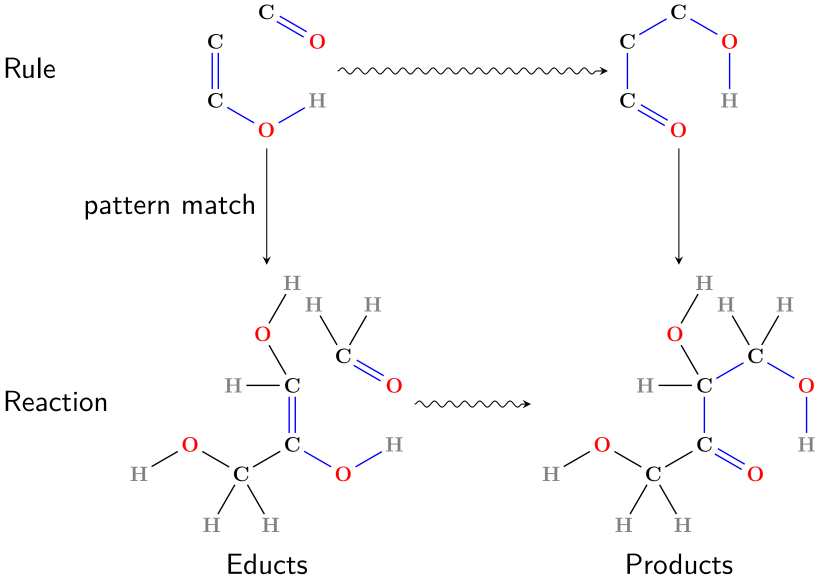


Educts

Chemical Reaction Patterns



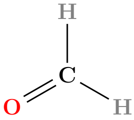
Chemical Reaction Patterns



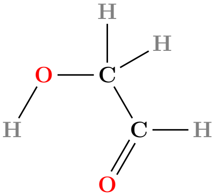
We get a precise atom map!

Grammar Example: The Formose Chemistry

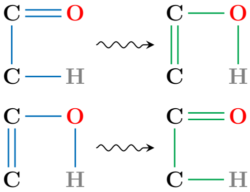
Formaldehyde:



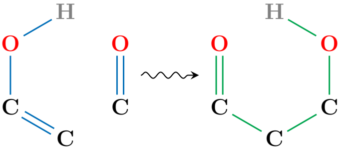
Glycolaldehyde:



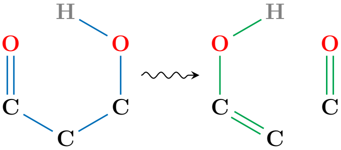
Keto-enol tautomerism:



Aldol addition:



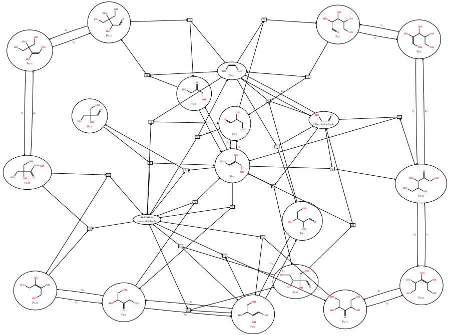
Retro aldol addition:



Modelling and Analysis of Chemical Systems

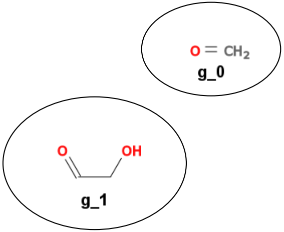
3. Generate a reaction network.

```
dg = dgRuleComp(inputGraphs,
     addSubset(inputGraphs) >> rightPredicate[
          lambda d: all(countCarbon(a) <= 5 for a in d.right)
     ](       repeat(inputRules)    )
)
dg.calc()
```



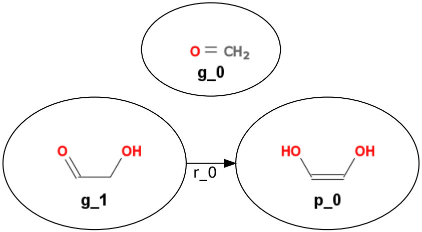
Generic Strategies for Chemical Space Exploration Jakob L. Andersen, Christoph Flamm, Daniel Merkle, and Peter F. Stadler. *International Journal of Computational Biology and Drug Design*, **7(2/3):225-258, 2014.**

Reaction Network for Formose



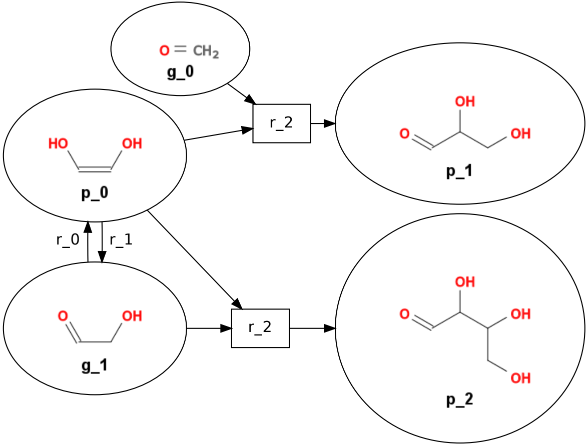
Initial State

Reaction Network for Formose



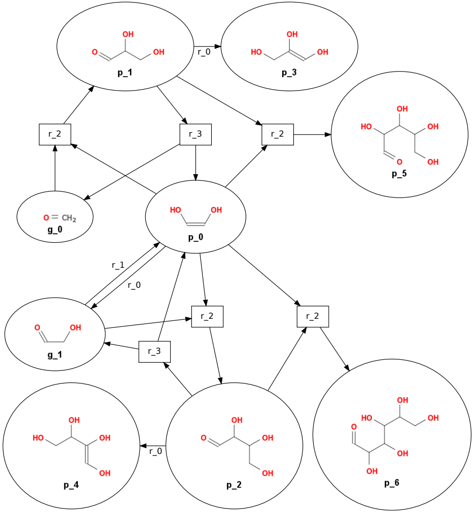
Generation 1

Reaction Network for Formose



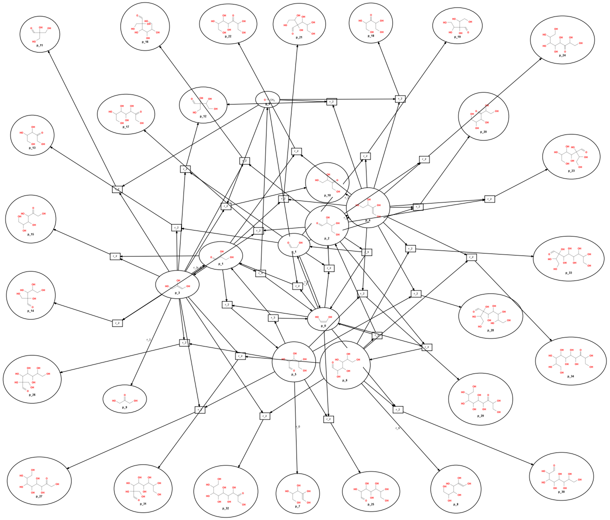
Generation 2

Reaction Network for Formose



Generation 3

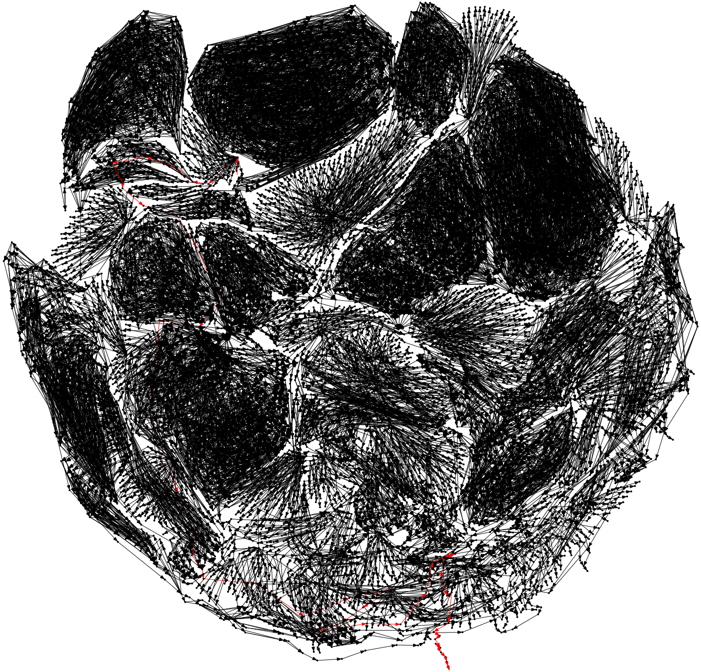
Reaction Network for Formose



Generation 4

The Chemical Space is a Hypergraph

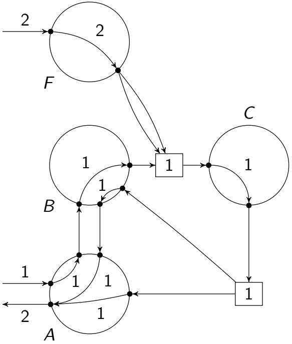
Another Example of a Search Space (Catalan)

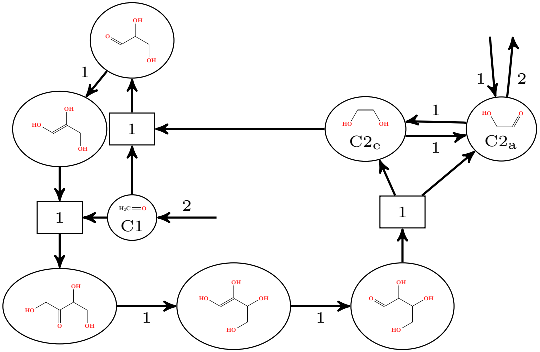


Modelling and Analysis of Chemical Systems

4. Setup pathway model and formulate pathway question.

Example: i.) Given 2 formaldehyde and 1 glycolaldehyde, how can 2 glycolaldehyde be produced? ii.) Or more genral: Is there autocatalysis?





(Demonstration 1)

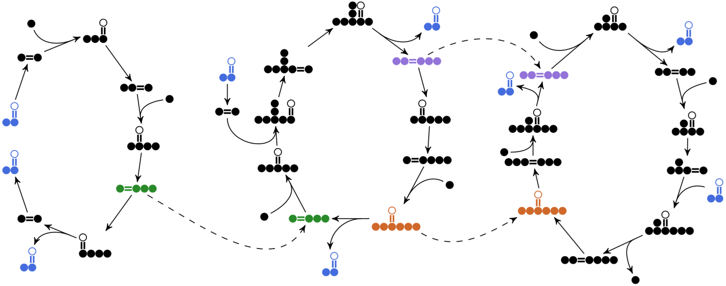
Modelling and Analysis of Chemical Systems

5. Enumerate many alternate pathways.

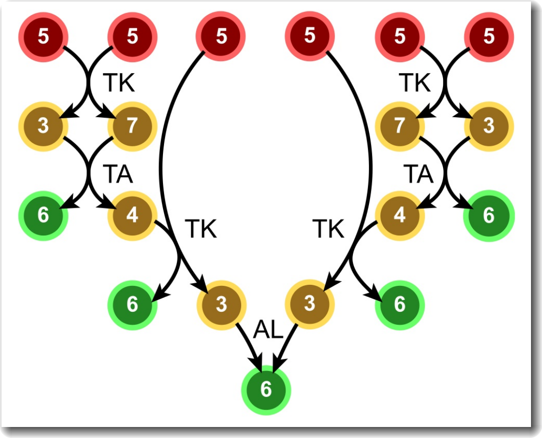
Example (Formose):

Network: all molecules with at most 9 carbon atoms.

Reactions used	Maximum #C 4	5	6	7	8	9	Sum
6	0	0	1	1	1	2	5
7	0	0	0	0	0	2	2
8	1	5	7	17	37	68	135
9	0	0	12	12	37	69	130
10	0	12	50	274	849	—	$\geq$ 1185
11	0	5	41	190	738	—	$\geq$ 974
							$\geq$ 2431

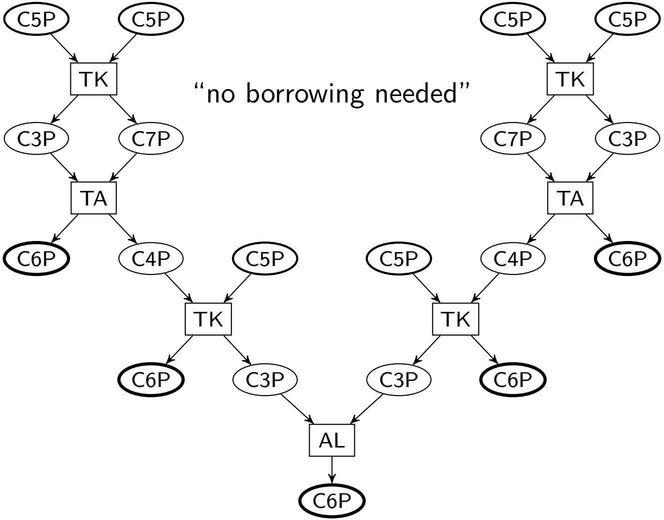


Pentose Phosphate Pathway

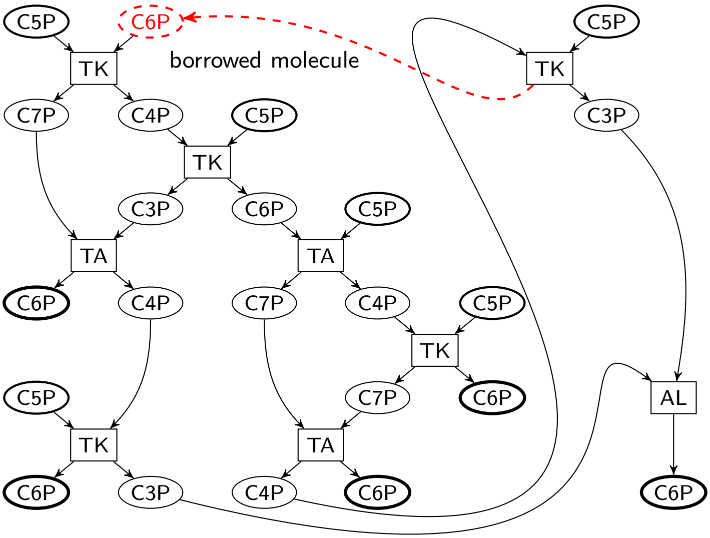


Pentose Phosphate Pathway — Another View

Simplified visualisation (from Integer-Hyperflow to Petri Net analysis)



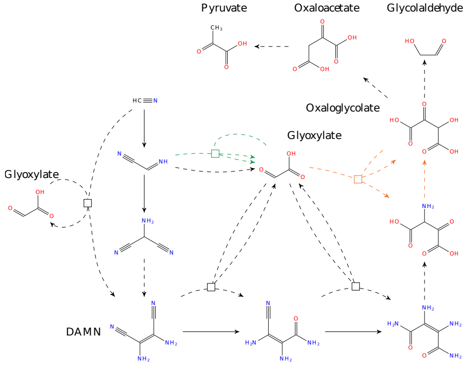
Pentose Phosphate Pathway with Borrowing



(Demonstration 2)

Another Example: Eschenmoser's Glyoxylate Scenario

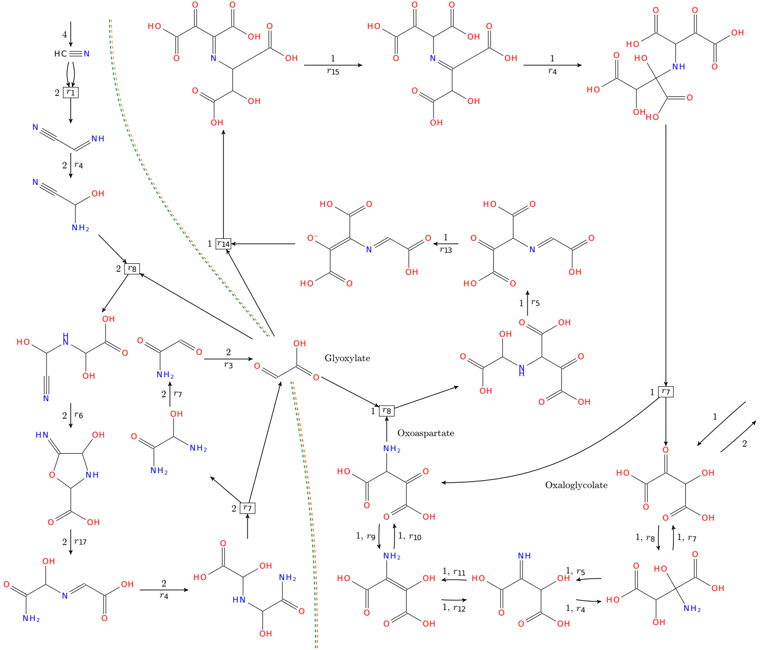
- Prebiotic Chemistry
- Cascade of autocatalytic pathways from HCN to glyoxylate to oxaloglycolate



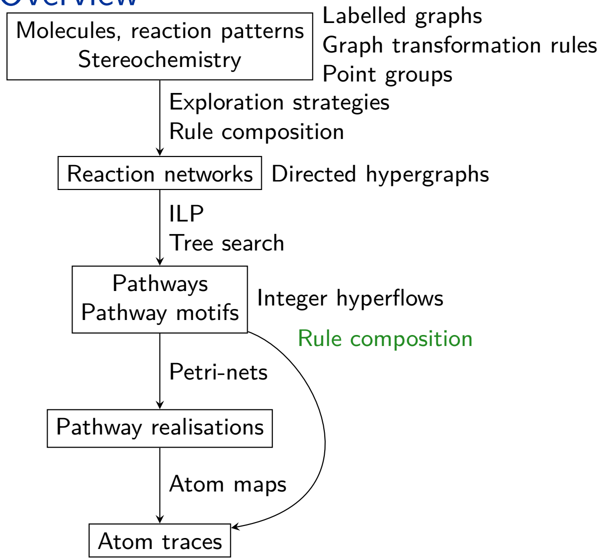
Albert Eschenmoser. On a hypothetical generational relationship between HCN and con- stituents of the reductive citric acid cycle. *Chem. Biodivers.*, **4**:554–573, 2007.

In silico Support for Eschenmoser's Glyoxylate Scenario Jakob L. Andersen, Christoph Flamm, Daniel Merkle, and Peter F. Stadler. *Israel Journal of Chemistry*, **55**(8):919-933, 2015.

Eschenmoser's Glyoxylate Scenario - ILP solution



Overview



https://cheminf.imada.sdu.dk

Category theory
Double Pushout
Rule composition
Mono- and Isomorphisms
Canonicalisation
Automorphisms
Quantum Chemistry

Software package: MØD

C++, Python, Bash, LaTeX

Pentose phosphate pathway
Formose
Glycolysis (EMP and ED)
Non-oxidative glycolysis
Citric acid cycle
Enzyme mechanisms
Prebiotic chemistry (HCN)
Eschenmoser's GLX scenario
DNA templated computing

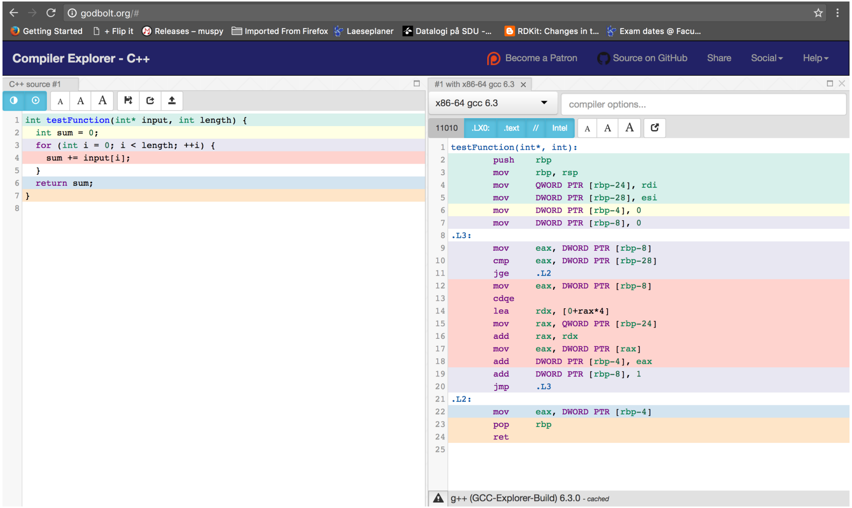
Outline

- Switch of Abstraction Levels: Rule Composition
- Design of Non-Oxidative Glycolysis (Multi-enzymatic Cascade)
- Atom Tracing and Hierarchical Decomposition with Cayley Graphs
- Enzyme Mechanism Design and Analysis
- Future Project: MATOMIC

- Analysis of Enzyme Chemistry (Multi-step, single enzyme)
- Exploring and Evaluating Reaction Mechanisms
- Thermodynamics
- Connection to Quantum Mechanical Methods
- Stochastic Simulations

Rule Composition

Levels of Abstraction in Programming



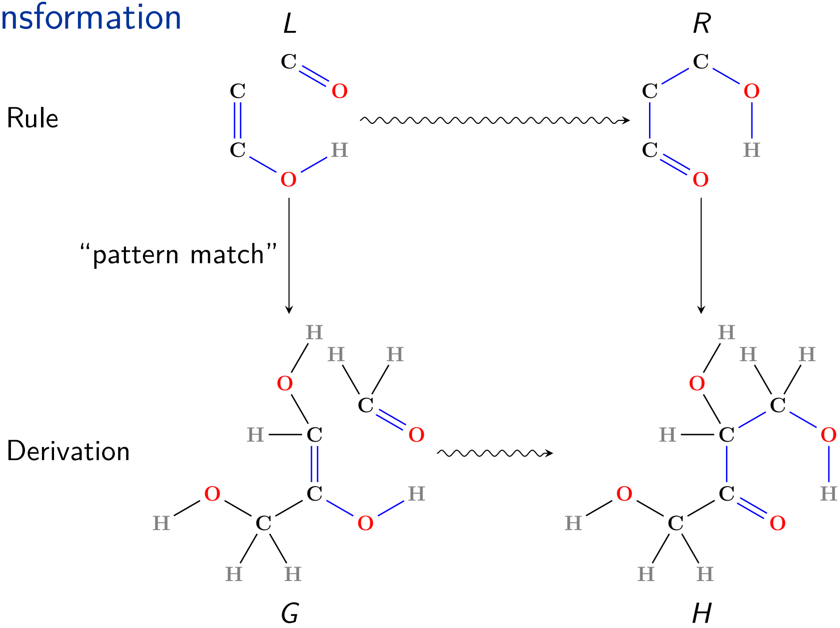
Declarative Description $\leftrightarrow$ DSL $\leftrightarrow$ C++ $\leftrightarrow$ Assembler

Levels of Abstraction in Computer Science

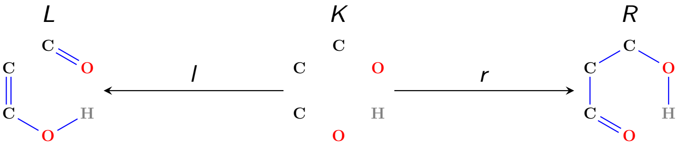
"The psychological profiling [of a Computer Scientist] is mostly the ability to shift levels of abstraction, from low level to high level. To see something in the small and to see something in the large."

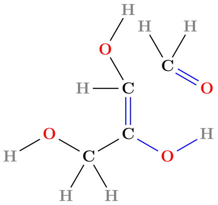
Donald Knuth

Graph Transformation



Rule Application

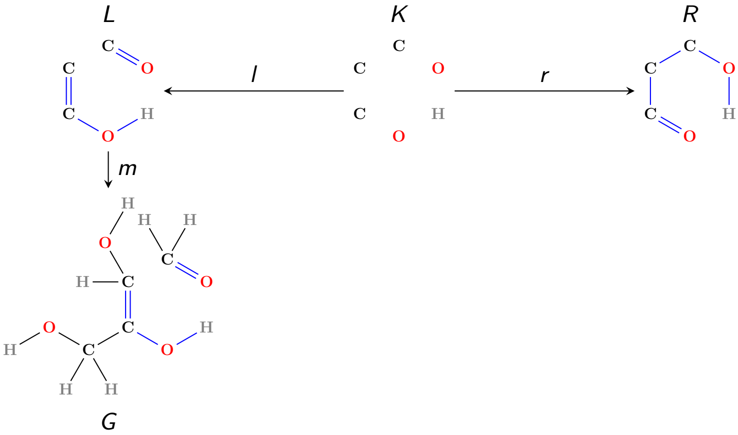




G

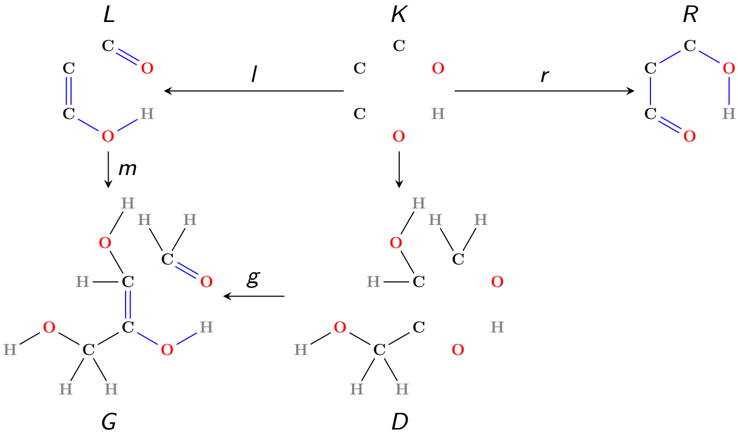
Given a rule $p = (L \stackrel{l}{\leftarrow} K \stackrel{r}{\rightarrow} R)$ and a graph G,

Rule Application



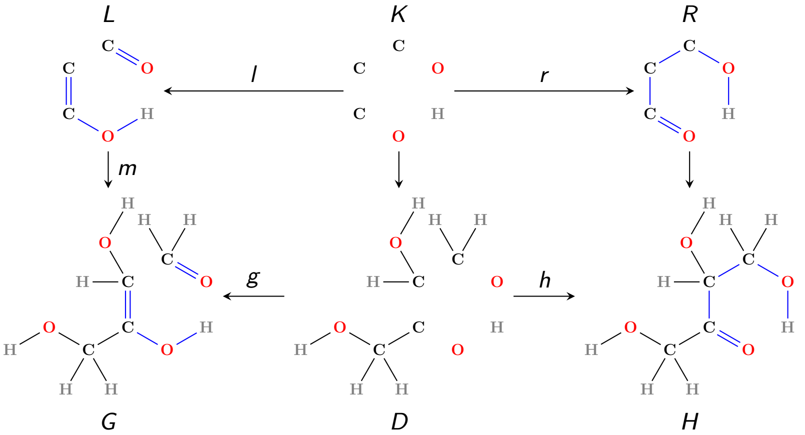
find a monomorphism $m\colon L \to G$,

Rule Application



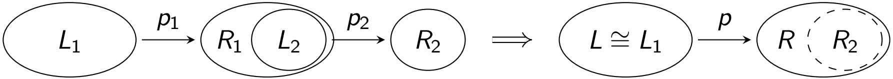
construct D as the pushout complement of $K \to L \to G$,

Rule Application

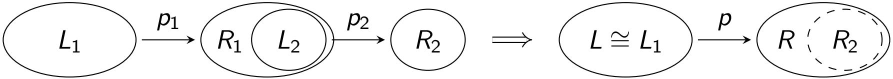


and construct H as the pushout object of $D \leftarrow K \rightarrow R$.

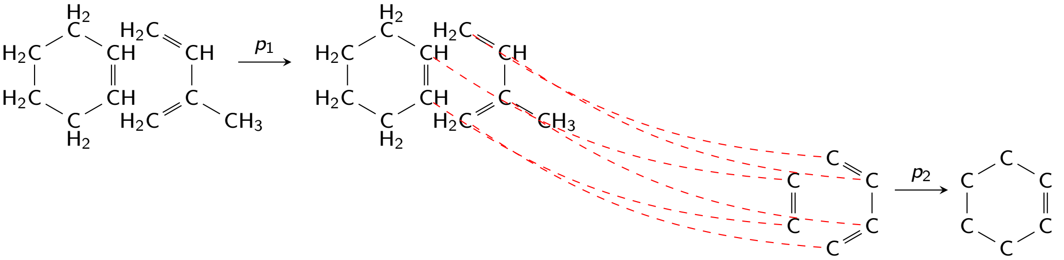
Full Rule Composition

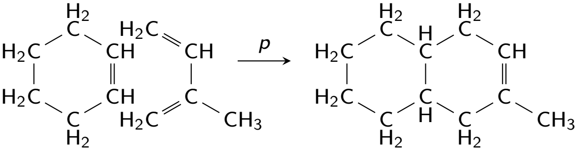


Full Rule Composition

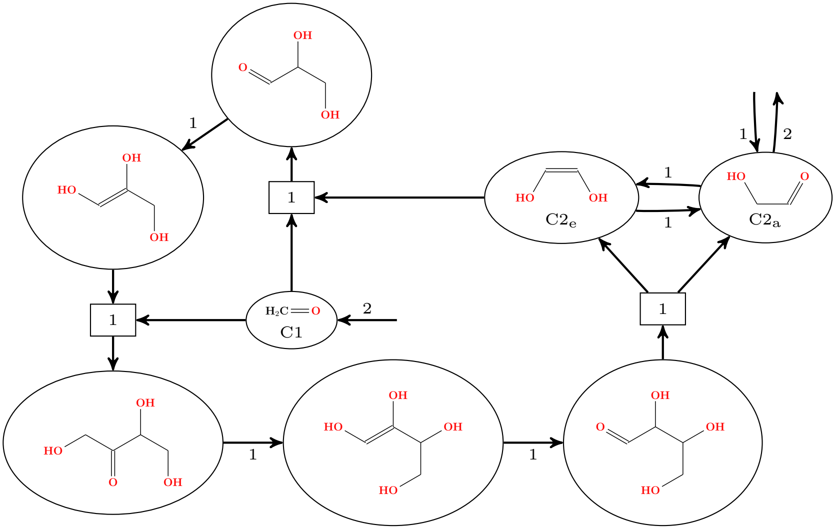


$p_1 = (G \leftarrow G \rightarrow G)$ $G = \{\!\{\text{Cyclohexene, Isoprene}\}\!\}$ $p_2 =$ Diels-Alder reaction

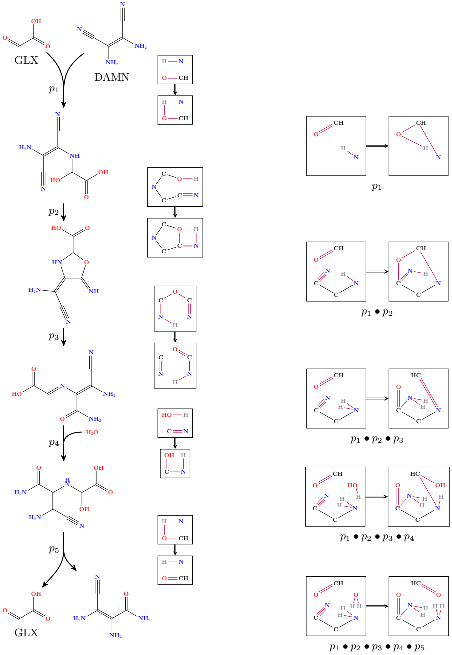




Demonstration 3 : Rule Composition in Formose



Levels of Abstraction in Computational Chemistry



Automatic Abstraction:

Aldehydes acting as catalysts for the hydrolysis of an HCN-tertramer.

[Eschenmoser, 2007]

Atom-Tracing Glycolysis Pathways

Glucose $\longrightarrow$ 2 Pyruvate, two implementations:

- Embden–Meyerhof–Parnas (EMP) pathway.
- Entner–Doudoroff (ED) pathway.

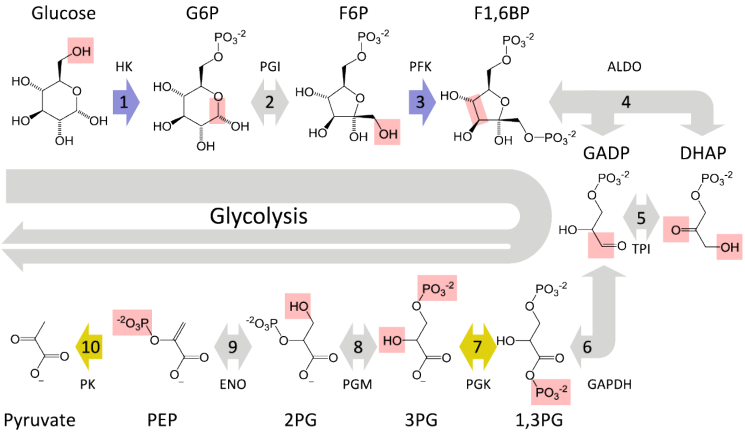
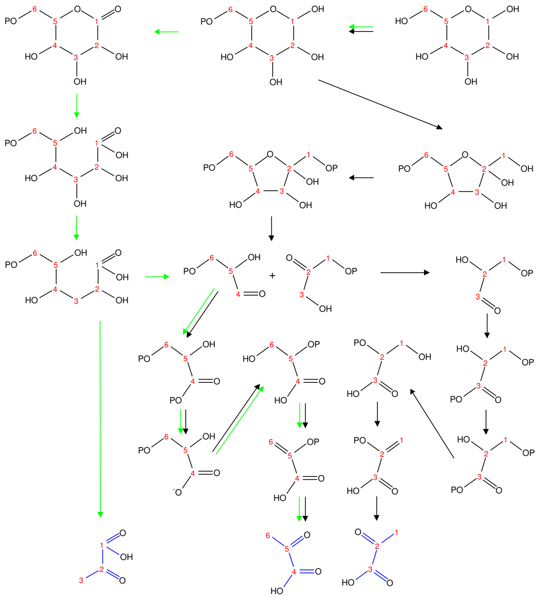


Figure: Embden–Meyerhof–Parnas (EMP) pathway [Wikipedia].

Atom-Tracing of Glycolysis Pathways



Andersen, Flamm, Merkle, and Stadler. 50 Shades of rule composition: From chemical reactions to higher levels of abstraction. *Formal Methods in Macro-Biology*, volume 8738 of LNCS., pages 117–135, 2014

Atom-Tracing of Glycolysis Pathways

1. Model each reaction as a rule, with correct atom-map: 13 rules.
2. Let $G(EMP)$, $G(ED)$, $H(EMP)$, and $H(ED)$ be the combined educt and product graphs.
3. Express the pathways as rule composition expressions:

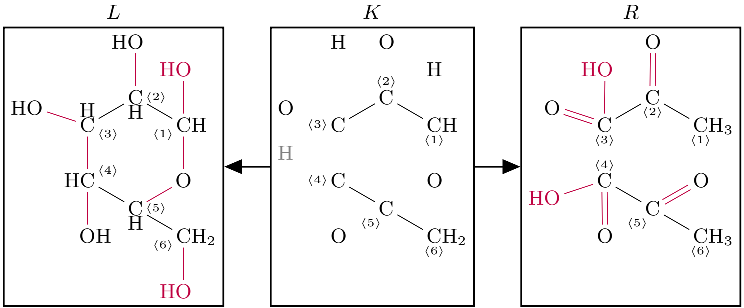
EMP:

$$\imath_{G(EMP)} \circ \overbrace{p_4 \circ p_1 \circ p_4 \circ p_2 \circ p_{13} \circ p_3}^{\text{Glucose} \to \text{2 G3P}}$$
$$\circ \underbrace{(p_6 \circ_\emptyset p_6) \circ (p_5 \circ_\emptyset p_5) \circ (p_7 \circ_\emptyset p_7) \circ (p_8 \circ_\emptyset p_8) \circ (p_5 \circ_\emptyset p_5) \circ (p_9 \circ_\emptyset p_9)}_{\text{2 G3P} \to \text{2 Pyruvate}} \circ \imath_{H(EMP)}$$

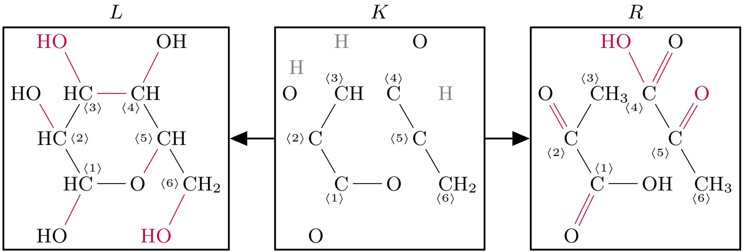
ED:

$$\imath_{G(ED)} \circ \underbrace{r_4 \circ r_{10} \circ r_{11} \circ r_{12} \circ r_{13}}_{\text{Glucose} \to \text{G3P + Pyruvate}} \circ \underbrace{r_6 \circ r_5 \circ r_7 \circ r_8 \circ r_5 \circ r_9}_{\text{G3P + Pyruvate} \to \text{2 Pyruvate}} \circ \imath_{H(ED)}$$

Overall EMP Pathway:

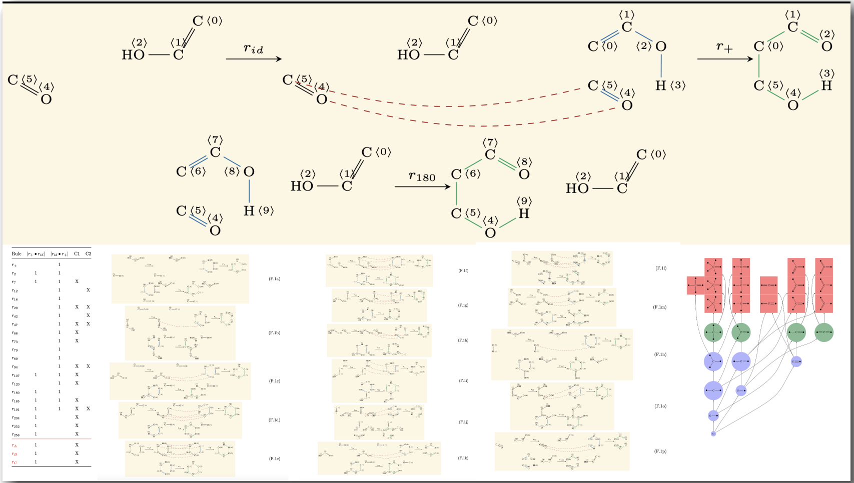


Overall ED Pathway:



More Applications of Rule Composition

Universal theory of continuous-time Markov chains for stochastic rewriting systems.



Allows for "static analysis" of biochemical (rule-based) models.

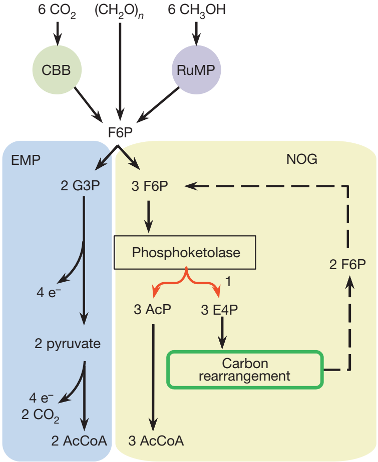
Rewriting Theory for the Life Sciences: A Unifying Theory of CTMC Semantics (2021)
Theoretical Computer Science Behr, Krivine, Andersen, Merkle

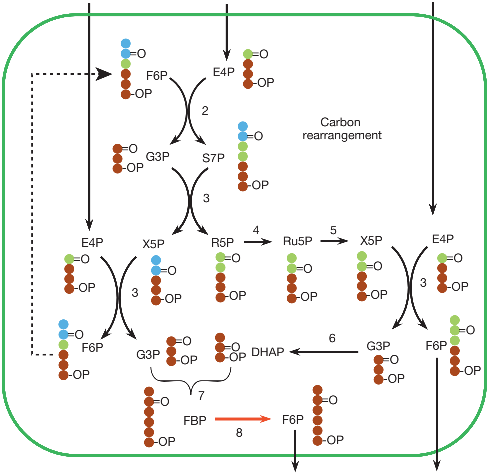
Multi-Enzymatic Cascade Design

Synthetic non-oxidative glycolysis enables complete carbon conservation

Igor W. Bogorad[1,2], Tzu-Shyang Lin[1] & James C. Liao[1,3]

[Nature, 2013]





Non-Oxidative Glycolysis (all 100% carbon yield)

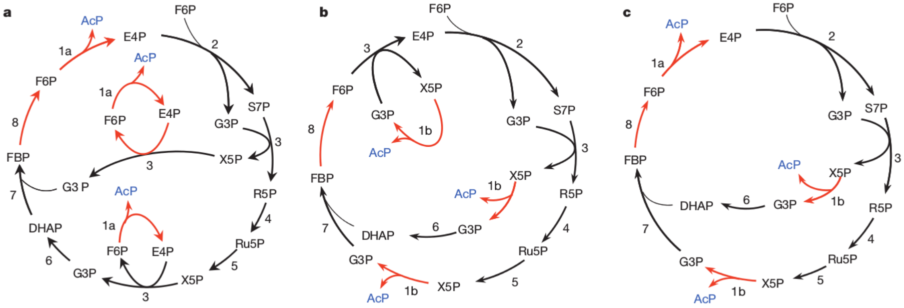


Figure 2 | Three FBP-dependent NOG networks. a–c, NOG using Fpk only (**a**), NOG using Xpk only (**b**) and NOG using F/Xpk (**c**). These configurations differ from those shown in Fig. 1 because the Xpk-linked Tkt has been integrated with carbon rearrangement. The red arrows in **a**–**c** indicate irreversible reactions that drive the cycle. Enzyme numbers are defined in Fig. 1 legend, except: 1a, Fpk; 1b, Xpk.

These networks are **engineered**.
Works *in vitro* and *in vivo*.

Generation of Reaction Network

Expansion strategy: apply rules unless we make too large molecules

Python (using PyMØD):

```
stratBFS = (
  addSubset(water, Pi, AcP, G3P, E4P, moreSugarPhosphates)
  >> rightPredicate[
    lambda d: all(a.vLabelCount("C") <= 8 for a in d.right)
  ](
    repeat(nogRules)
  )
)
dg = dgRuleComp(inputMolecules, stratBFS)
dg.calc()
```

Enumerating NOG Pathways

I/O Constraints:

- 1 F6P as input (the only carbon source)
- AcP as output (the only carbon sink)
- $\mathrm{P_i}$ and water as input and output (food/waste molecules)

Python:

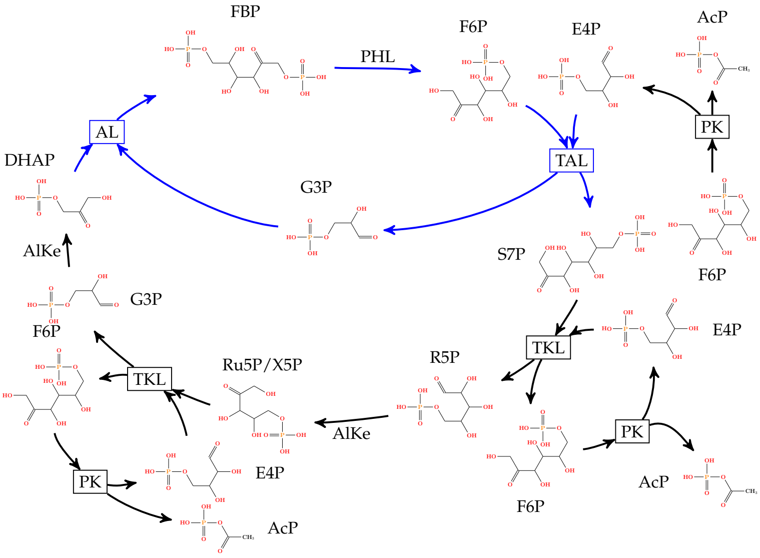
```
flow = dgFlow(dg)
flow.objectiveFunction = edge # implicitly minimised
for a in {Pi, water}:
    flow.addSource(a)
    flow.addSink(a)
flow.addSource(F6P)
flow.addConstraint(inFlow(F6P) == 1)
flow.addSink(AcP)
# enumerate solutions (optimal and optimal+1)
flow.setSolverEnumerateBy(absGap=1)
flow.calc()
```

Central Carbon Metabolism – Alternate Pathways?

	Only FBP				Other Bisphosphates							
	8 Unique React.				7 Unique React.				8 Unique React.			
PK Type	Reactions				Reactions				Reactions			
X, F, S, O	8	9	10	11	7	8	9	10	8	9	10	11
0, 0, 0, 3	–	–	–	–	–	–	–	–	–	–	4	16
0, 0, 1, 2	–	–	–	–	–	–	–	–	–	3	2	–
0, 0, 2, 1	–	–	–	–	–	–	–	–	–	4	–	–
0, 0, 3, 0	–	–	1	2	–	–	1	2	–	–	9	20
0, 1, 0, 2	–	–	–	–	–	–	–	–	–	4	4	–
0, 1, 1, 1	–	–	–	–	–	–	–	–	3	–	–	–
0, 1, 2, 0	–	1	–	–	–	1	–	–	–	8	2	–
0, 2, 0, 1	–	–	–	–	–	–	–	–	–	6	–	–
0, 2, 1, 0	–	1	–	–	–	1	–	–	–	9	–	–
0, 3, 0, 0	–	–	2	4_a	–	–	2	4	–	–	14	24
1, 0, 0, 2	–	–	–	–	–	–	–	–	–	2	4	–
1, 0, 1, 1	–	–	–	–	–	–	–	–	1	–	–	–
1, 0, 2, 0	–	1	–	–	–	1	–	–	–	6	2	–
1, 1, 0, 1	–	–	–	–	–	–	–	–	2	–	–	–
1, 1, 1, 0	1	–	–	–	1	–	–	–	3	–	–	–
1, 2, 0, 0	–	2	–	–	–	2	–	–	–	10	–	–
2, 0, 0, 1	–	–	–	–	–	–	–	–	–	4	–	–
2, 0, 1, 0	–	1	–	–	–	1	–	–	–	7	–	–
2, 1, 0, 0	–	2_c	–	–	–	2	–	–	–	10	–	–
3, 0, 0, 0	–	–	2_b	4	–	–	2	4	–	–	12	20

(263 solutions)

Example Pathway: 3 FPK, Only FBP, shortest

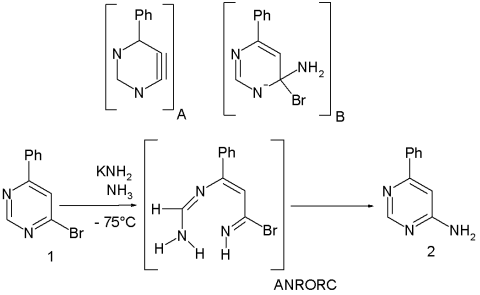


Chemical Transformation Motifs — Modelling Pathways as Integer Hyperflows, *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 16(2), 510-523, 2019

Atom Tracing and Hierarchical Decomposition with Cayley Graphs

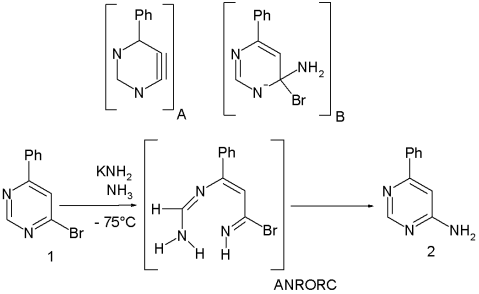
Atom Tracing and Hierarchical Decomposition with Cayley Graphs

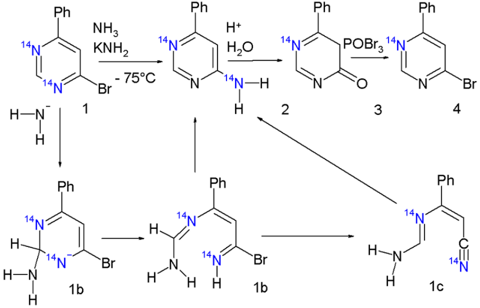
Addition of the Nucleophile, Ring Opening, and Ring Closure (ANRORC mechanism)



Atom Tracing and Hierarchical Decomposition with Cayley Graphs

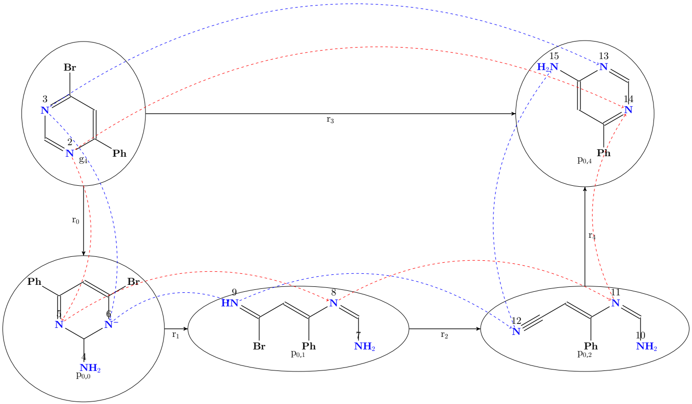
Addition of the Nucleophile, Ring Opening, and Ring Closure (ANRORC mechanism)





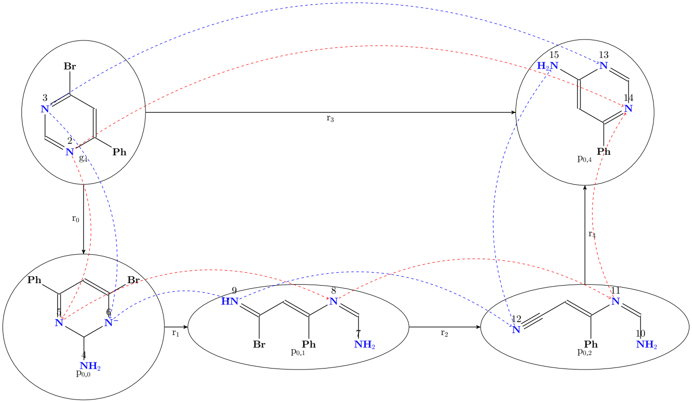
Atom Tracing: Example ANRORC mechanism

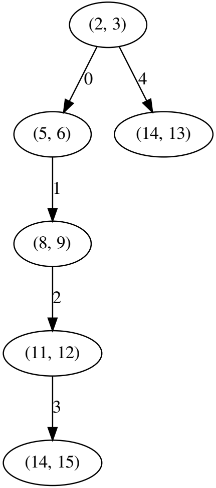
Mechanism alternatives and the (right) Cayley Graph



Atom Tracing: Example ANRORC mechanism

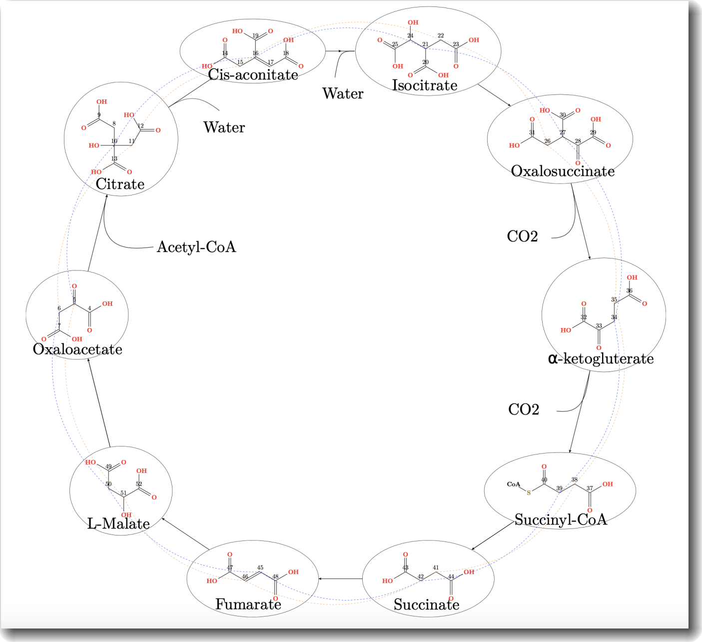
Mechanism alternatives and the (right) Cayley Graph





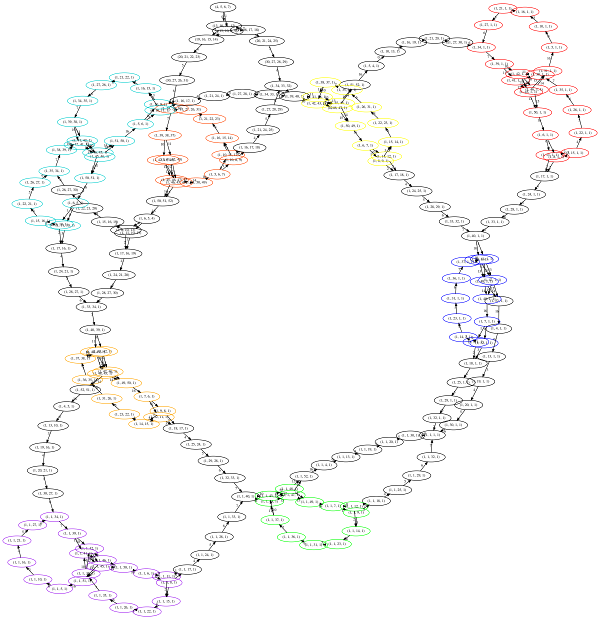
- Automatic enumeration of distinguishing subgraphs (e.g. $^{14}NH_2$)
- NMR spectroscopy / mass spectrometry

Atom Tracing: Example Citric Acid Cycle



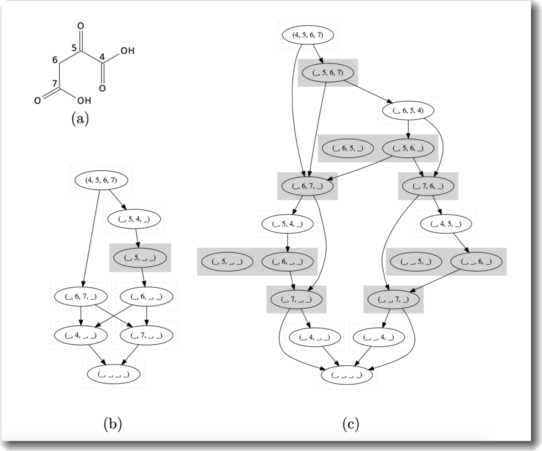
Atom Tracing: Example Citric Acid Cycle

Cayley graph (of the 4 C atoms of OAA)



Atom Tracing: Example TCA Cycle

"Components" of the Cayley graph

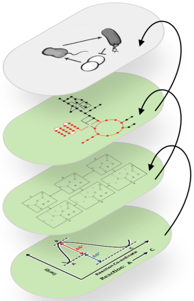


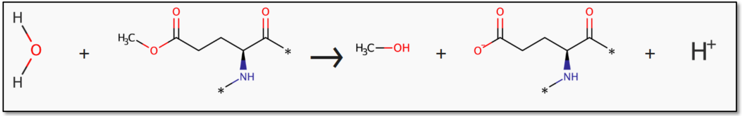
- Automated analysis of hierarchical structures
- Natural formal approach for isotope labelling design (ongoing: M/S and NMR spec. prediction)

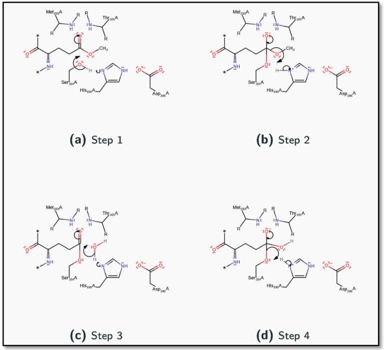
Cayley Graphs of Semigroups Applied to Atom Tracking in Chemistry (2021), *Journal of Computational Biology*, Nøjgaard, Fontana, Hellmuth, Merkle

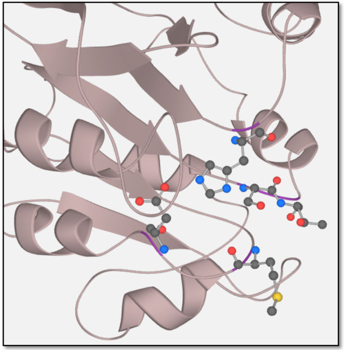
Enzyme Mechanisms Analysis and Design

Enzymatic Mechanisms - Analysis and Design



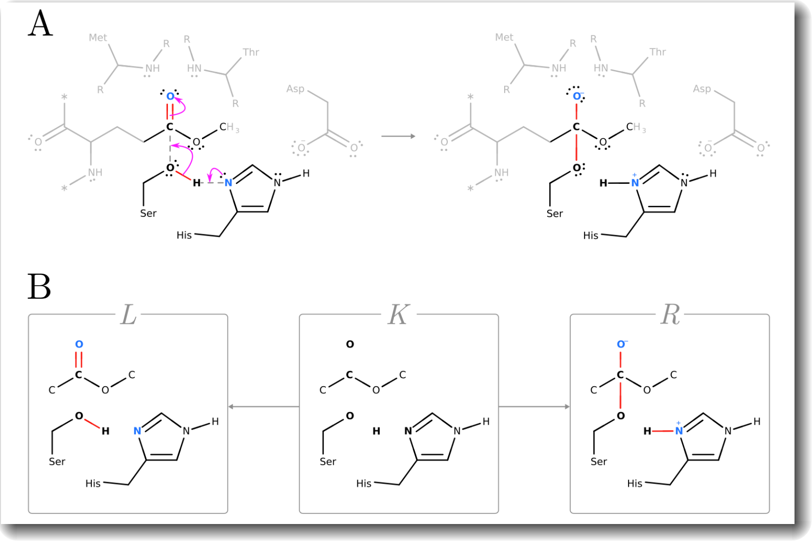



(a) Step 1
(b) Step 2
(c) Step 3
(d) Step 4



- Illustrated: protein-glutamate methylesterase (EC3.1.1)
- Novo Nordisk Foundation Exploratory Synergy Grant
  Harvard Medical School (Fontana), University of Vienna (Flamm), 2020-2022

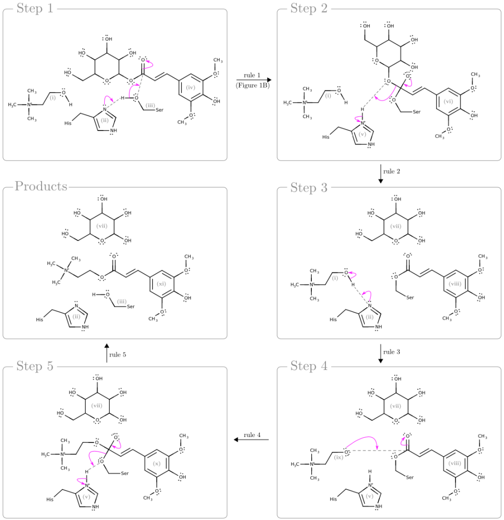
First step of protein-glutamate methylesterase



A Proposed New Mechanism

Proposed mechanism for the conversion of choline (i) and sinapoyl-glucose (iv) into glucose (vii) and sinapoyl-choline (xi) (RHEA:12024 entry)

- **Catalytic Mechanism**: a sequence of steps that is cyclical in the participating amino acids and whose traversal converts substrate(s) into product(s)
- Input to theozymes and compuzymes design methods.



Graph transformation for enzymatic mechanisms, Andersen et al., Bioinformatics, ECCB, 2021