Exploring Chemistry Using Satisfiability Modulo Theories

Rolf Fagerberg[1], Christoph Flamm[2], Daniel Merkle[1], Philipp Peters[1]

Dept. of Mathematics and Computer Science University of Southern Denmark

Institute for Theoretical Chemistry, University of Vienna, Austria

February 16, 2012

Submitted to *Computer Aided Verification 2012*

Content

Exploring Chemistry

Example: Retrosynthetic Analysis

Exploring Chemistry

Example: Retrosynthetic Analysis

Given: The (properties of a) target molecule

Find: Synthesis mechanism to target molecule

Todo: Simplify molecule in retrosynthetic direction

Exploring Chemistry

Example: Retrosynthetic Analysis

Given: The (properties of a) target molecule

Find: Synthesis mechanism to target molecule

Todo: Simplify molecule in retrosynthetic direction

F

Exploring Chemistry

Example: Retrosynthetic Analysis

Given: The (properties of a) target molecule

Find: Synthesis mechanism to target molecule

Todo: Simplify molecule in retrosynthetic direction

$$D + E \quad \rightarrow \quad F$$

Exploring Chemistry

Example: Retrosynthetic Analysis

Given: The (properties of a) target molecule

Find: Synthesis mechanism to target molecule

Todo: Simplify molecule in retrosynthetic direction

$$\begin{aligned} B + C &\rightarrow D \\ D + E &\rightarrow F \end{aligned}$$

Exploring Chemistry

Example: Retrosynthetic Analysis

Given: The (properties of a) target molecule

Find: Synthesis mechanism to target molecule

Todo: Simplify molecule in retrosynthetic direction

$$\begin{aligned} A &\rightarrow B \\ B + C &\rightarrow D \\ D + E &\rightarrow F \end{aligned}$$

Exploring Chemistry

Example: Inverse Reaction Mechanism Problem

Exploring Chemistry

Example: Inverse Reaction Mechanism Problem

Given: Reaction mechanism and chemical reactions

Find: Compatible molecules according to this mechanism

Todo: Assign chemical reactions to mechanism

Exploring Chemistry

Example: Inverse Reaction Mechanism Problem

Given: Reaction mechanism and chemical reactions

Find: Compatible molecules according to this mechanism

Todo: Assign chemical reactions to mechanism

r_1: $C_5H_{11}O_8P \rightarrow C_5H_9O_8P$
r_2: $C_3H_7O_6P \rightarrow C_3H_7O_6P$
r_3: $C_3H_7O_6P + C_3H_7O_6P \rightarrow C_6H_{14}O_{12}P_2$
r_4: $C_5H_{11}O_8P + C_5H_9O_8P \rightarrow C_7H_{15}O_{10}P + C_3H_7O_6P$
$\vdots$
r_n: $H_2O + C_6H_{14}O_{12}P_2 \rightarrow C_6H_{13}O_9P + H_3PO_4$

$$A \rightarrow B$$
$$B + C \rightarrow D$$
$$D + E \rightarrow F$$

Exploring Chemistry

Example: Inverse Reaction Mechanism Problem

Given: Reaction mechanism and chemical reactions

Find: Compatible molecules according to this mechanism

Todo: Assign chemical reactions to mechanism

$$\begin{aligned}
r_1&: & C_5H_{11}O_8P &\rightarrow C_5H_9O_8P\\
r_2&: & C_3H_7O_6P &\rightarrow C_3H_7O_6P\\
r_3&: & C_3H_7O_6P + C_3H_7O_6P &\rightarrow C_6H_{14}O_{12}P_2\\
r_4&: & C_5H_{11}O_8P + C_5H_9O_8P &\rightarrow C_7H_{15}O_{10}P + C_3H_7O_6P\\
& & &\vdots\\
r_n&: & H_2O + C_6H_{14}O_{12}P_2 &\rightarrow C_6H_{13}O_9P + H_3PO_4
\end{aligned}$$

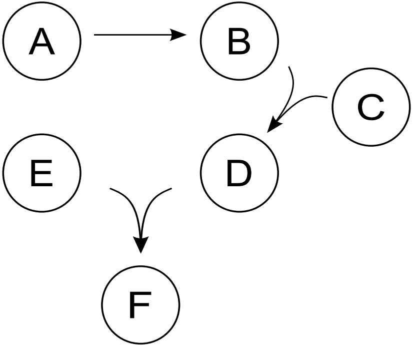
$$\begin{aligned}
A &\rightarrow B\\
B + C &\rightarrow D\\
D + E &\rightarrow F
\end{aligned}$$

$$\begin{aligned}
C_5H_{11}O_8P &\rightarrow C_5H_9O_8P\\
C_5H_9O_8P + CH_4O_4P &\rightarrow C_6H_{14}O_1P_2\\
C_6H_{14}O_{12}P_2 + H_2O &\rightarrow C_6H_{13}O_9P
\end{aligned}$$

Reaction Mechanism

Hypergraph:

$$\begin{aligned} A &\rightarrow B \\ B + C &\rightarrow D \\ D + E &\rightarrow F \end{aligned}$$

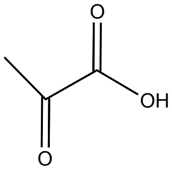


Chemical Rules

A Rule

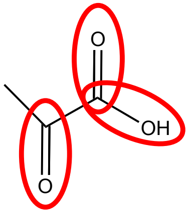
Change of reactive sites (functional groups) of molecules in an elementary reaction

Our Modelling of Molecules

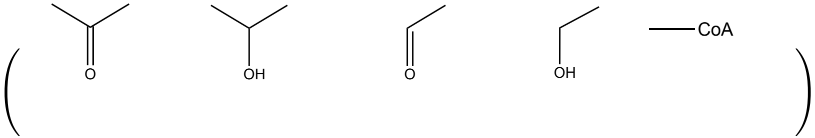


Pyruvate

Our Modelling of Molecules

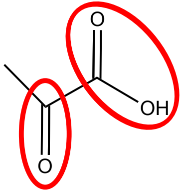


Pyruvate
(1,0,1,1,0)



Vector of functional groups, disregards spatial information

Our Modelling of Molecules

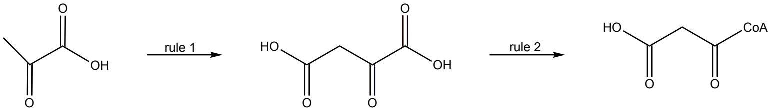


Pyruvate

Vector of functional groups, disregards spatial information

Our Modelling of Chemical Rules

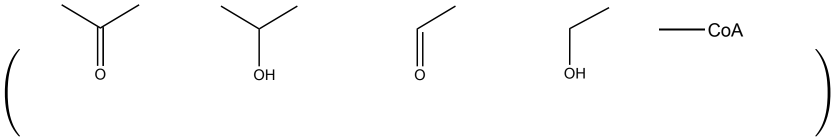
We define a rule as a change and a precondition vector



Pyruvate (1,0,1,1,0)

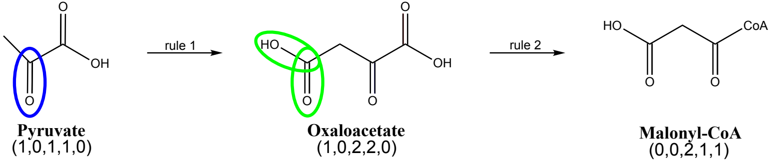
Oxaloacetate (1,0,2,2,0)

Malonyl-CoA (0,0,2,1,1)

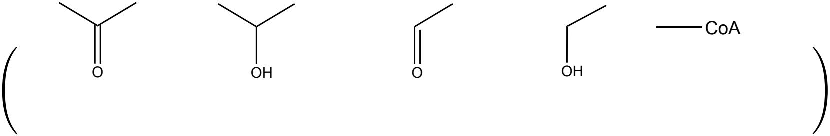


Our Modelling of Chemical Rules

We define a rule as a change and a precondition vector

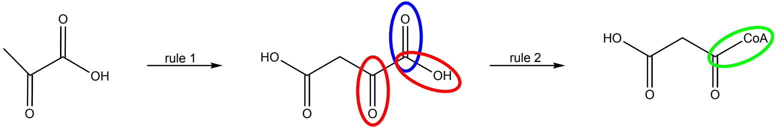


rule 1: precond(pyr) $=$ (1,0,0,0,0) change(pyr) $=$ (0,0,1,1,0)



Our Modelling of Chemical Rules

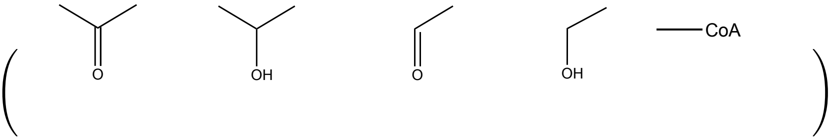
We define a rule as a change and a precondition vector



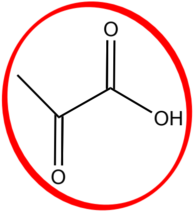
Pyruvate (1,0,1,1,0) **Oxaloacetate** (1,0,2,2,0) **Malonyl-CoA** (0,0,2,1,1)

rule 1: precond(pyr) = (1,0,0,0,0) change(pyr) = (0,0,1,1,0)

rule 2: precond(oxa) = (1,0,1,1,0) change(oxa) = (-1,0,0,-1,1)

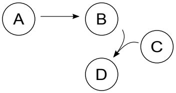


Our Modelling of Molecules

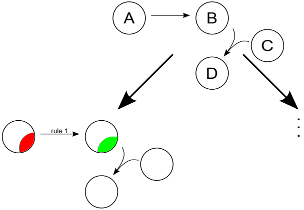


Pyruvate

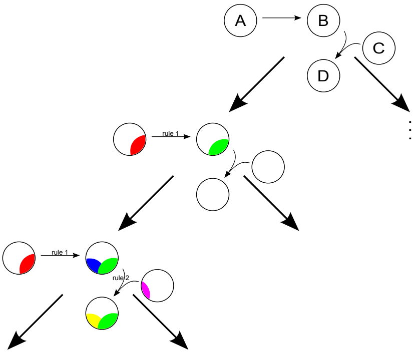
Our Setup



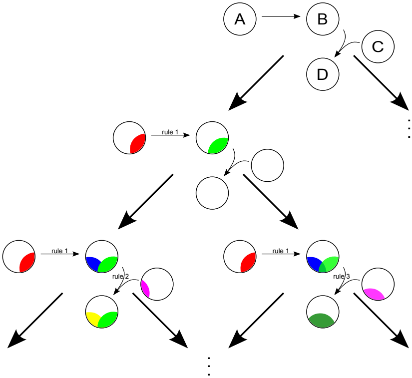
Our Setup



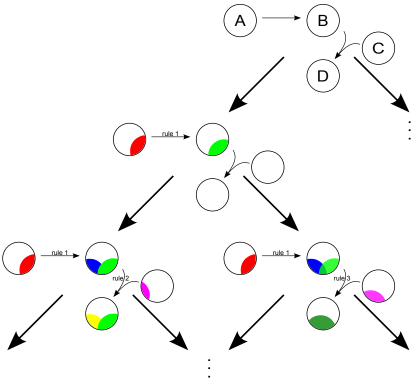
Our Setup



Our Setup



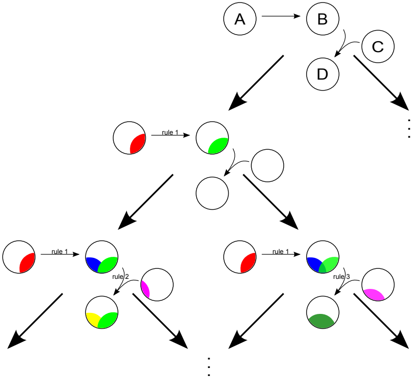
Our Setup



Satisfiability Problem to find compatible assignment of rules

Powerful solvers exist for such problems

Our Setup

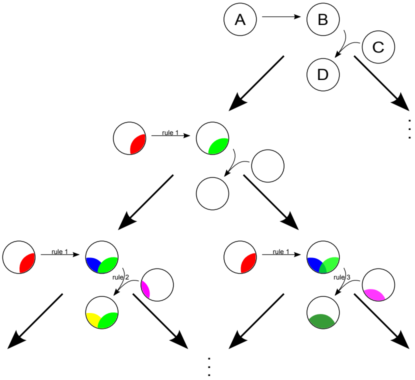


Satisfiability Problem to find compatible assignment of rules

Powerful solvers exist for such problems

$$\text{Definitions} + \text{Constraints} \xrightarrow{\textbf{magic}} \text{Solution}$$

Our Setup



Satisfiability Problem to find compatible assignment of rules

Powerful solvers exist for such problems

$$\boxed{\text{Definitions} + \text{Constraints} \xrightarrow{\textbf{SMT}-\textbf{Solver}} \text{Solution}}$$

Satisfiability Modulo Theories (SMT)

Logical Constraints:

$$x, y \in \mathbb{N}$$
$$0 \leq x \leq 4$$
$$x + y = 2$$
$$\forall x, y : f(x) < f(y)$$

Solution:

$$\begin{aligned} x &= ? \\ y &= ? \\ f(x) &= ? \\ f(y) &= ? \end{aligned}$$

SMT formulas provide a rich and powerful modelling language!

SMT

Identity of molecules:

$$\forall v, w \in Mol : (v, w) \in ID$$
$$\Rightarrow \forall x_p \in \textit{functional groups} : v(x_p) = w(x_p)$$

SMT

Identity of molecules:

$$\forall v, w \in Mol : (v, w) \in ID$$
$$\Rightarrow \forall x_p \in \textit{functional groups} : v(x_p) = w(x_p)$$

Equivalence relation:

(declare-fun ID (MOL MOL) Bool)

SMT

Identity of molecules:

$$\forall v, w \in Mol : (v, w) \in ID$$
$$\Rightarrow \forall x_p \in \textit{functional groups} : v(x_p) = w(x_p)$$

Equivalence relation:

(declare-fun ID (MOL MOL) Bool)

(assert (forall (mol MOL) (= (ID mol mol) T))) $\wedge$

SMT

Identity of molecules:

$$\forall v, w \in Mol : (v, w) \in ID$$
$$\Rightarrow \forall x_p \in \textit{functional groups} : v(x_p) = w(x_p)$$

Equivalence relation:

(declare-fun ID (MOL MOL) Bool)

(assert (forall (mol MOL) (= (ID mol mol) T))) ∧
(forall (mol,mol2 MOL) (=(ID mol mol2)(ID mol2 mol))) ∧

SMT

Identity of molecules:

$$\forall v, w \in Mol : (v, w) \in ID$$
$$\Rightarrow \forall x_p \in \textit{functional groups} : v(x_p) = w(x_p)$$

Equivalence relation:

(declare-fun ID (MOL MOL) Bool)

(assert (forall (mol MOL) (= (ID mol mol) T))) ∧
(forall (mol,mol2 MOL) (=(ID mol mol2)(ID mol2 mol))) ∧
(forall (mol,mol2,mol3 MOL) (⇒ (and (=(ID mol mol2) T)
(=(ID mol2 mol3) T)) (=(ID mol mol3) T)))

SMT

Identity of molecules:

$$\forall v, w \in Mol : (v, w) \in ID$$
$$\Rightarrow \forall x_p \in \textit{functional groups} : v(x_p) = w(x_p)$$

Equivalence relation:

(declare-fun ID (MOL MOL) Bool)

(assert (forall (mol MOL) (= (ID mol mol) T))) ∧
(forall (mol,mol2 MOL) (=(ID mol mol2)(ID mol2 mol))) ∧
(forall (mol,mol2,mol3 MOL) (⇒ (and (=(ID mol mol2) T)
(=(ID mol2 mol3) T)) (=(ID mol mol3) T)))

Solver searches for values and assignments to functions

Our Most General Setup

$$
\begin{array}{lrcl}
m_1\times & A+B & \rightarrow & C+D \\
m_2\times & E+F & \rightarrow & G+H \\
m_3\times & I+J & \rightarrow & K+L \\
\hline
\end{array}
$$

Our Most General Setup

$$
\begin{array}{llcl}
m_1\times & A+B & \rightarrow & C+D \\
m_2\times & C+D & \rightarrow & G \\
m_3\times & A & \rightarrow & K+B \\
\hline
\end{array}
$$

$$id(C,E), id(D,F), id(A,I), id(B,L), H=\emptyset, J=\emptyset$$

Solver assigns equivalence classes

Our Most General Setup

$$
\begin{array}{lrcl}
2\times & A + B & \rightarrow & C + D \\
2\times & C + D & \rightarrow & G \\
1\times & A & \rightarrow & K + B \\
\hline
\end{array}
$$

$$id(C,E), id(D,F), id(A,I), id(B,L), H = \emptyset, J = \emptyset$$

$$m_1 = 2, m_2 = 2, m_3 = 1$$

Solver assigns equivalence classes
Solver assigns multiplicities of reactions

Our Most General Setup

$$\begin{array}{lrcl} 2\times & A+B & \rightarrow & C+D \\ 2\times & C+D & \rightarrow & G \\ 1\times & A & \rightarrow & K+B \\ \hline \sum: & \mathbf{3}\cdot\mathbf{A}+\mathbf{B} & \rightarrow & \mathbf{2}\cdot\mathbf{G}+\mathbf{K} \end{array}$$

$id(C,E), id(D,F), id(A,I), id(B,L), H=\emptyset, J=\emptyset$

$m_1 = 2, m_2 = 2, m_3 = 1$

Solver assigns equivalence classes
Solver assigns multiplicities of reactions
Solver assigns overall reaction

Our Most General Setup

$2\times$	$A + B$	$\rightarrow$	$C + D$
$2\times$	$C + D$	$\rightarrow$	G
$1\times$	A	$\rightarrow$	$K + B$
$\sum$:	$\mathbf{3 \cdot A + B}$	$\rightarrow$	$\mathbf{2 \cdot G + K}$

$$id(C, E), id(D, F), id(A, I), id(B, L), H = \emptyset, J = \emptyset$$

$$m_1 = 2, m_2 = 2, m_3 = 1$$

Solver assigns equivalence classes
Solver assigns multiplicities of reactions
Solver assigns overall reaction
Solver assigns molecule vectors

Our Most General Setup

$$\begin{array}{lrcl} 2\times & A+B & \rightarrow & C+D \\ 2\times & C+D & \rightarrow & G \\ 1\times & A & \rightarrow & K+B \\ \hline \sum: & \mathbf{3\cdot A+B} & \rightarrow & \mathbf{2\cdot G+K} \end{array}$$

$id(C,E), id(D,F), id(A,I), id(B,L), H=\emptyset, J=\emptyset$

$m_1=2, m_2=2, m_3=1$

Solver assigns equivalence classes
Solver assigns multiplicities of reactions
Solver assigns overall reaction
Solver assigns molecule vectors

Also allow user to specify constraints

An SMT Solution

Provides:

Rule Mapping $A + B \xrightarrow{rule1} C + D$

Equivalence classes: identical molecules $id(C, E)$

Multiplicities $m_1 = 2$

Molecule vectors $Pyruvate : (1, 0, 1, 1, 0)$

Results

Pentose Phosphate Pathway

m_1	$A + B$	$\rightarrow$	$C + D$
m_2	$E + F$	$\rightarrow$	$G + H$
m_3	$I + J$	$\rightarrow$	$K + L$
m_4	$M + N$	$\rightarrow$	$O + P$
m_5	$Q + R$	$\rightarrow$	$S + T$
m_6	$U + V$	$\rightarrow$	$W + X$
m_7	$Y + Z$	$\rightarrow$	Ä + Ö
m_8	Ü + Æ	$\rightarrow$	Å + Ø

Starting with general mechanism

Pentose Phosphate Pathway

$$
\begin{array}{llcl}
m_1 & A + B & \rightarrow & C + D \\
m_2 & E + F & \rightarrow & G + H \\
m_3 & I + J & \rightarrow & K + L \\
m_4 & M + N & \rightarrow & O + P \\
m_5 & Q + R & \rightarrow & S + T \\
m_6 & U + V & \rightarrow & W + X \\
m_7 & Y + Z & \rightarrow & \ddot{A} + \ddot{O} \\
m_8 & \ddot{U} + \text{Æ} & \rightarrow & \mathring{A} + \text{Ø} \\
\hline
\sum: & \mathbf{6} \cdot \mathbf{A} & \rightarrow & \mathbf{5} \cdot \textbf{Ø}
\end{array}
$$

Starting with general mechanism
Define overall reaction
Using set of 10 rules
Molecule as 1 functional group

Pentose Phosphate Pathway

$$
\begin{array}{lrcl}
2\times & A & \rightarrow & C \\
2\times & A + C & \rightarrow & D + E \\
2\times & D + E & \rightarrow & G + H \\
2\times & A + G & \rightarrow & E + H \\
1\times & E & \rightarrow & F \\
1\times & E + F & \rightarrow & I \\
1\times & B + I & \rightarrow & J + K \\
1\times & J & \rightarrow & H \\
\hline
\sum: & \mathbf{6 \cdot A + B} & \rightarrow & \mathbf{5 \cdot H + K}
\end{array}
$$

Starting with general mechanism
Define overall reaction
Using set of 10 rules
Molecule as 1 functional group

Pentose Phosphate Pathway

SMT-Solution

Using rules with 6 functional groups

$$
\begin{array}{lccc}
2\times & (1,2,0,1,0,1) & \rightarrow & (0,3,1,0,0,1) \\
2\times & (1,2,0,1,0,1) + (0,3,1,0,0,1) & \rightarrow & (1,4,0,1,0,1) + (0,1,1,0,0,1) \\
2\times & (1,4,0,1,0,1) + (0,1,1,0,0,1) & \rightarrow & (0,2,1,0,0,1) + (1,3,0,1,0,1) \\
2\times & (1,2,0,1,0,1) + (0,2,1,0,0,1) & \rightarrow & (0,1,1,0,0,1) + (1,3,0,1,0,1) \\
1\times & (0,1,1,0,0,1) & \rightarrow & (1,0,0,1,0,1) \\
1\times & (0,1,1,0,0,1) + (1,0,0,1,0,1) & \rightarrow & (0,3,0,0,1,2) \\
1\times & B + (0,3,0,0,1,2) & \rightarrow & (0,3,0,1,1,1) + K \\
1\times & (0,3,0,1,1,1) & \rightarrow & (1,3,0,1,0,1)
\end{array}
$$

(1,2,0,1,0,1) can refer to Ribulose-5-P

(1,3,0,1,0,1) can refer to Fructose-6-P

Postprocessing

SMT-Solution provides:

Synthesis mechanism with underspecified molecules

$$
\begin{array}{lccccc}
2\times & (1,2,0,1,0,1) & & \rightarrow & (0,3,1,0,0,1) & \\
2\times & (1,2,0,1,0,1) & + (0,3,1,0,0,1) & \rightarrow & (1,4,0,1,0,1) & + (0,1,1,0,0,1) \\
2\times & (1,4,0,1,0,1) & + (0,1,1,0,0,1) & \rightarrow & (0,2,1,0,0,1) & + (1,3,0,1,0,1) \\
 & \vdots & & & \vdots &
\end{array}
$$

Postprocessing

SMT-Solution provides:

Synthesis mechanism with underspecified molecules

$$
\begin{array}{lccccccc}
2\times & & (1,2,0,1,0,1) & & \rightarrow & & (0,3,1,0,0,1) & \\
2\times & (1,2,0,1,0,1) & + & (0,3,1,0,0,1) & \rightarrow & (1,4,0,1,0,1) & + & (0,1,1,0,0,1) \\
2\times & (1,4,0,1,0,1) & + & (0,1,1,0,0,1) & \rightarrow & (0,2,1,0,0,1) & + & (1,3,0,1,0,1) \\
 & & \vdots & & & & \vdots &
\end{array}
$$

E.g. KEGG can be used to find all fitting real chemical rules

Postprocessing

SMT-Solution provides:

Synthesis mechanism with underspecified molecules

$$
\begin{array}{lccccc}
2\times & & (1,2,0,1,0,1) & \rightarrow & (0,3,1,0,0,1) & \\
2\times & (1,2,0,1,0,1) + (0,3,1,0,0,1) & & \rightarrow & (1,4,0,1,0,1) + (0,1,1,0,0,1) & \\
2\times & (1,4,0,1,0,1) + (0,1,1,0,0,1) & & \rightarrow & (0,2,1,0,0,1) + (1,3,0,1,0,1) & \\
 & & \vdots & & \vdots &
\end{array}
$$

E.g. KEGG can be used to find all fitting real chemical rules

One reaction mechanism - different synthesis mechanisms!

3-Hydroxypropanoate (3HP) Synthesis Pathway

Goal: Synthesize 3HP from Pyruvate

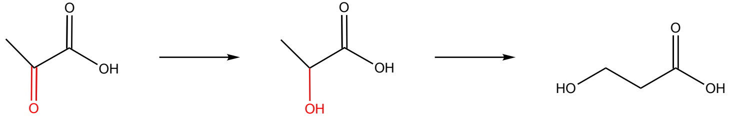
We followed approach of Henry et al., 2010
We modelled a pathway of n cascading 1-to-1 reactions

$$\begin{array}{ccc} A & \rightarrow & B \\ B & \rightarrow & C \\ & \vdots & \end{array}$$

Using 16 rules and 9 functional groups

3HP Synthesis Pathway

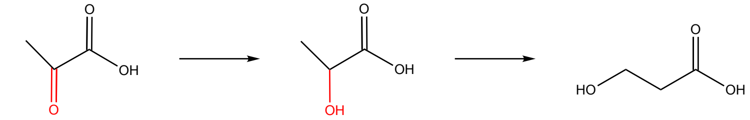
Our solution fits to Henry's solutions



Pyruvate (1,0,1,1,0,0,0,0,0) **Lactate** (0,1,1,1,0,0,0,0,0) **3HP** (0,0,1,2,0,0,0,0,0)

3HP Synthesis Pathway

Our solution fits to Henry's solutions



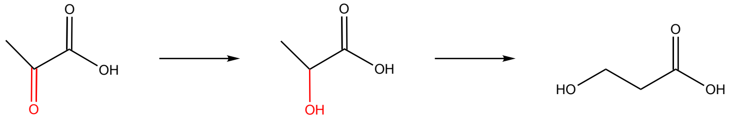
Pyruvate
(1,0,1,1,0,0,0,0,0)

Lactate
(0,1,1,1,0,0,0,0,0)

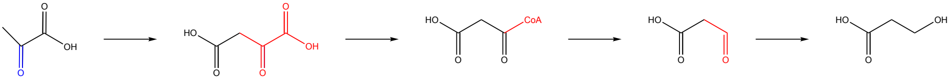
3HP
(0,0,1,2,0,0,0,0,0)

3HP Synthesis Pathway

Our solution fits to Henry's solutions

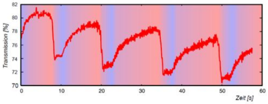


Pyruvate (1,0,1,1,0,0,0,0,0) **Lactate** (0,1,1,1,0,0,0,0,0) **3HP** (0,0,1,2,0,0,0,0,0)



Pyruvate (1,0,1,1,0,0,0,0,0) **Oxaloacetate** (1,0,2,2,0,0,0,0,0) **Malonyl-CoA** (0,0,2,1,0,0,0,0,1) **3-oxopropanoate** (0,0,2,1,0,0,0,0,0) **3HP** (0,0,1,2,0,0,0,0,0)

Conclusion



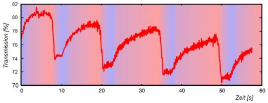
$\leftarrow$

$$A \rightarrow B$$
$$B + C \rightarrow D + E$$
$$D \rightarrow F$$

$\rightarrow$

$$\begin{aligned}
\frac{d[A]}{dt} &= -k_1[A][X] \\
\frac{d[X]}{dt} &= k_1[A][X] - k_2[X][Y] \\
\frac{d[Y]}{dt} &= k_2[X]YX] - k_3[Y] \\
\frac{d[B]}{dt} &= k_3[Y]
\end{aligned}$$

Conclusion



←

$$A \rightarrow B$$
$$B + C \rightarrow D + E$$
$$D \rightarrow F$$

→

$$\frac{d[A]}{dt} = -k_1[A][X]$$
$$\frac{d[X]}{dt} = k_1[A][X] - k_2[X][Y]$$
$$\frac{d[Y]}{dt} = k_2[X]YX] - k_3[Y]$$
$$\frac{d[B]}{dt} = k_3[Y]$$

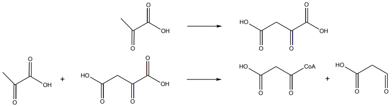
↑↓

SMT

↓

Postprocessing

↙ ↓ ↘



⋮

Thank you for attention!