Thermodynamic Feasibility of Pathways

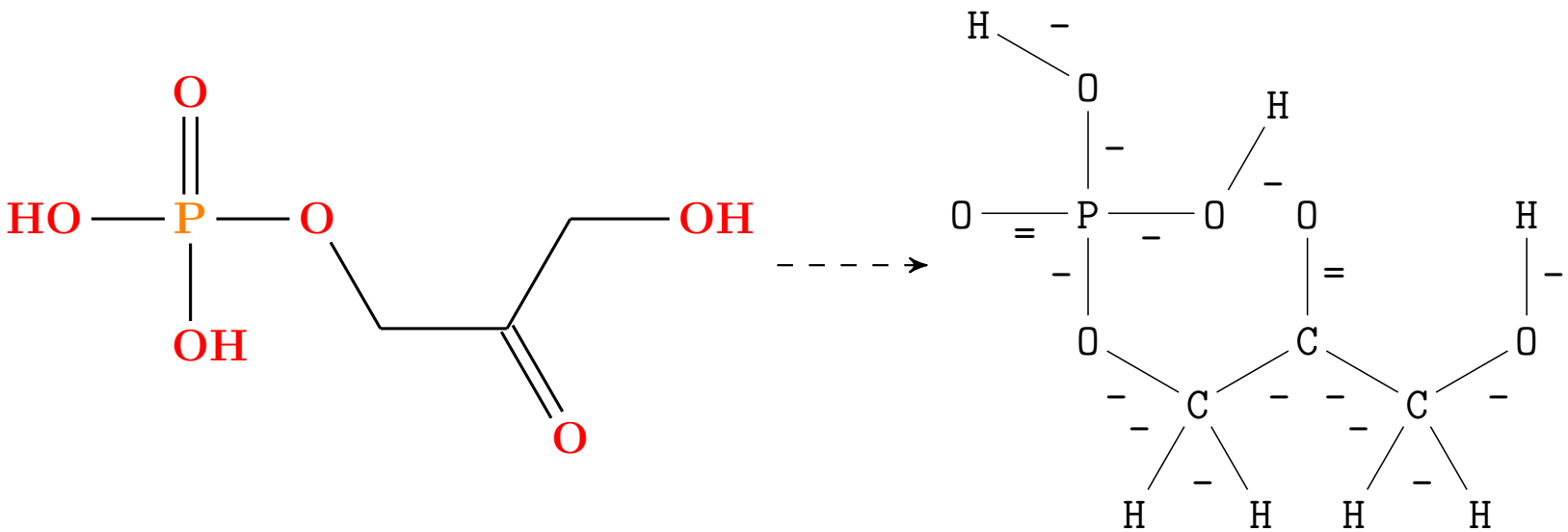
Jakob Lykke Andersen, Daniel Merkle,
Christoph Flamm, Peter F. Stadler

Earth-Life Science Institute, Tokyo Institute of Technology
Department of Mathematics and Computer Science, University of Southern Denmark

Bled, February 2016

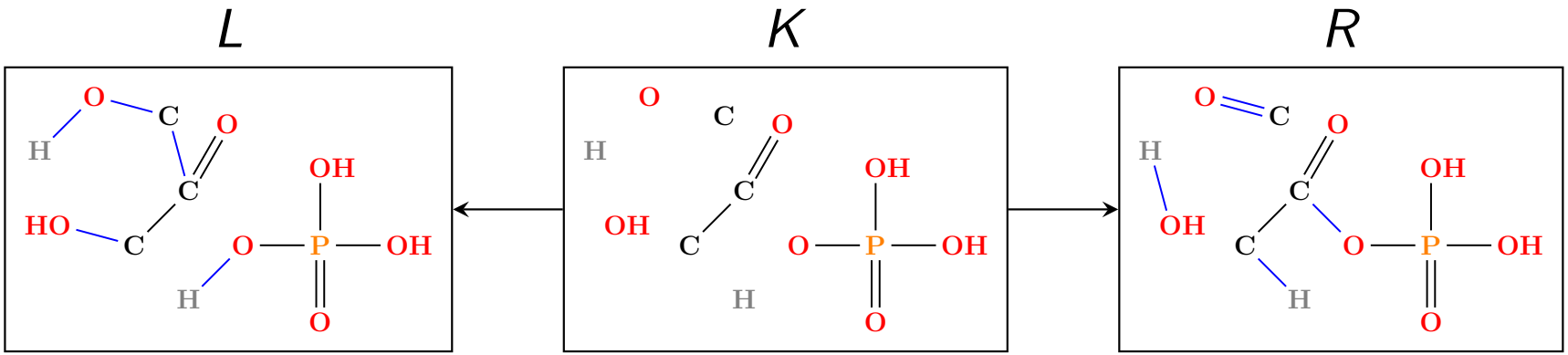
Modelling and Analysis of Chemical Systems

1. Model molecules as labelled graphs.



Modelling and Analysis of Chemical Systems

2. Model reaction types and graph transformation rules.



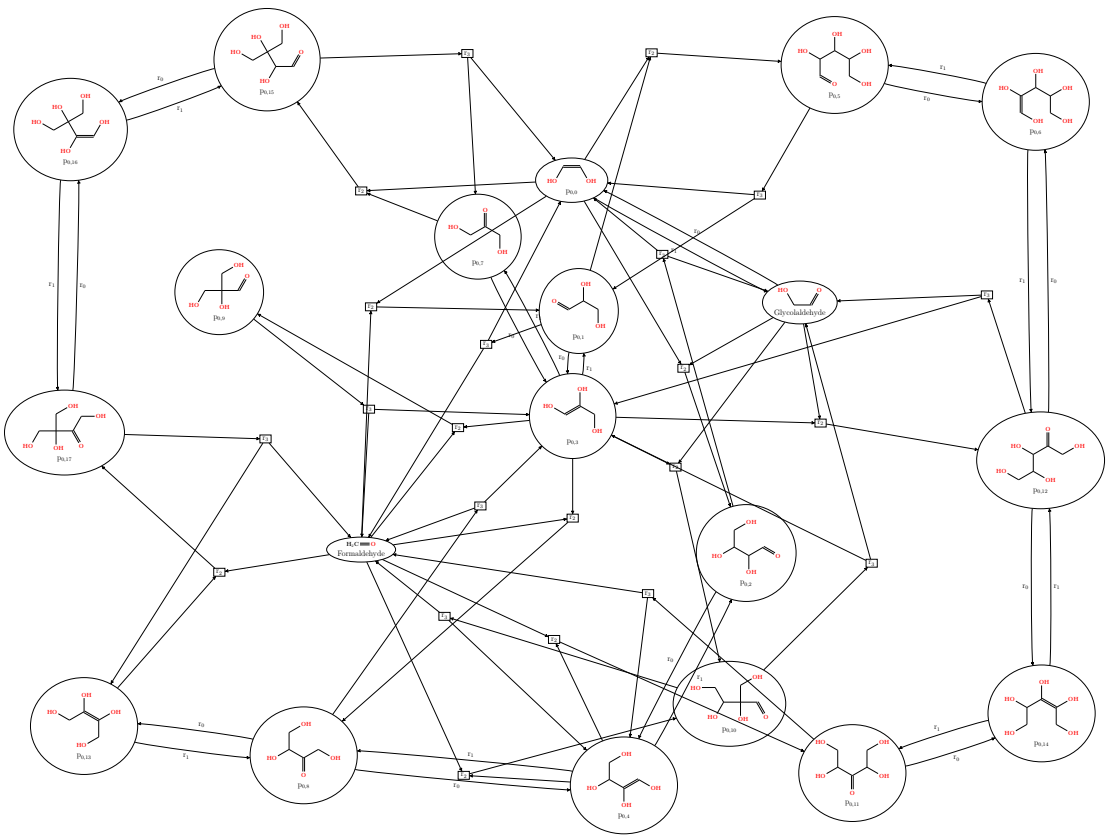
Example: Carbon rearrangement

- Aldolase: $\text{ketone} + \text{aldehyde} \longrightarrow \text{ketone}$
- Aldose-Ketose: $\text{aldehyde} \longrightarrow \text{ketone}$
- Ketose-Aldose: $\text{ketone} \longrightarrow \text{aldehyde}$
- Phosphohydrolase: $\text{H}_2\text{O}+\text{CnP} \longrightarrow \text{Cn}+\text{Pi}$
- Phosphoketolase $\text{Pi}+\text{ketone} \longrightarrow \text{carbonyl} + \text{CnP}_{+\text{water}}$
- Transaldolase: $\text{Cn}+\text{Cm} \longrightarrow \text{C(n+3)}+\text{C(m-3)}$
- Transketolase: $\text{Cn}+\text{Cm} \longrightarrow \text{C(n+2)}+\text{C(m-2)}$

Modelling and Analysis of Chemical Systems

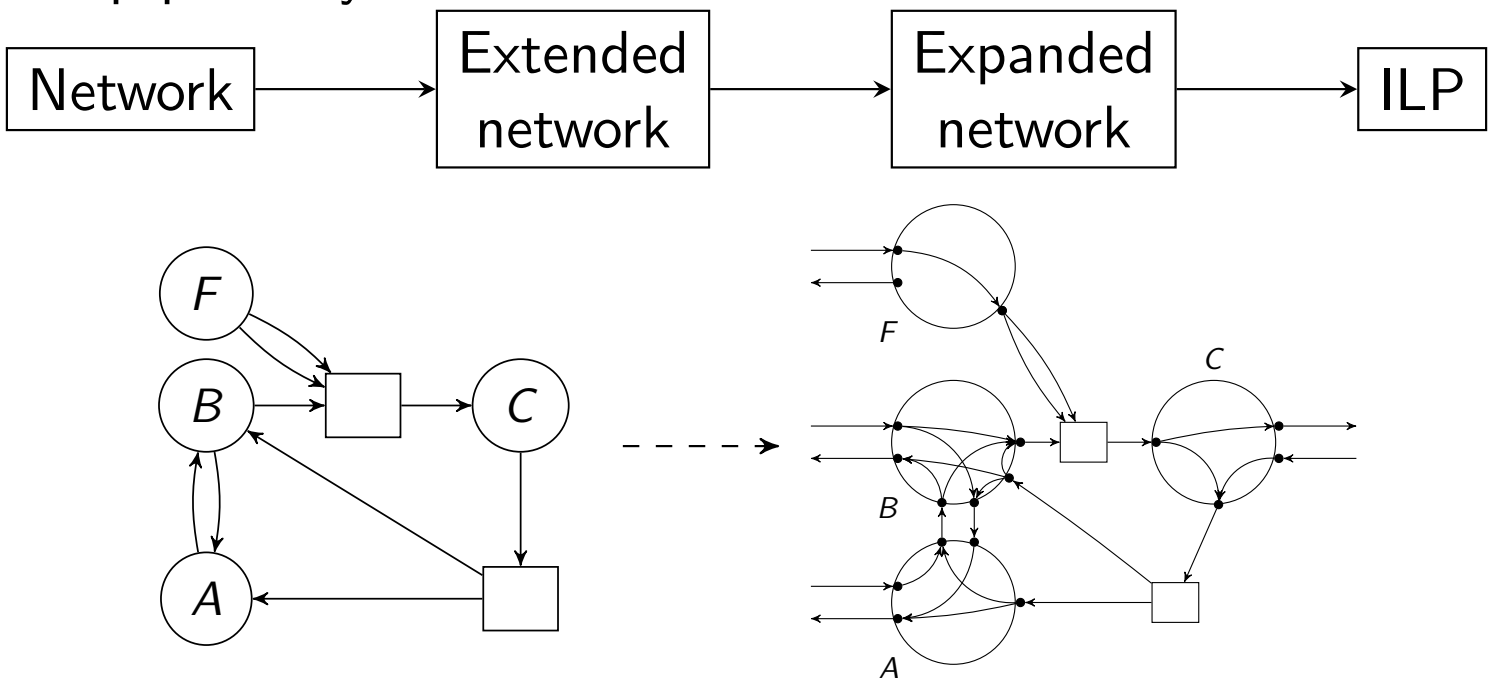
3. Generate a reaction network.

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



Modelling and Analysis of Chemical Systems

4. Set up pathway model.



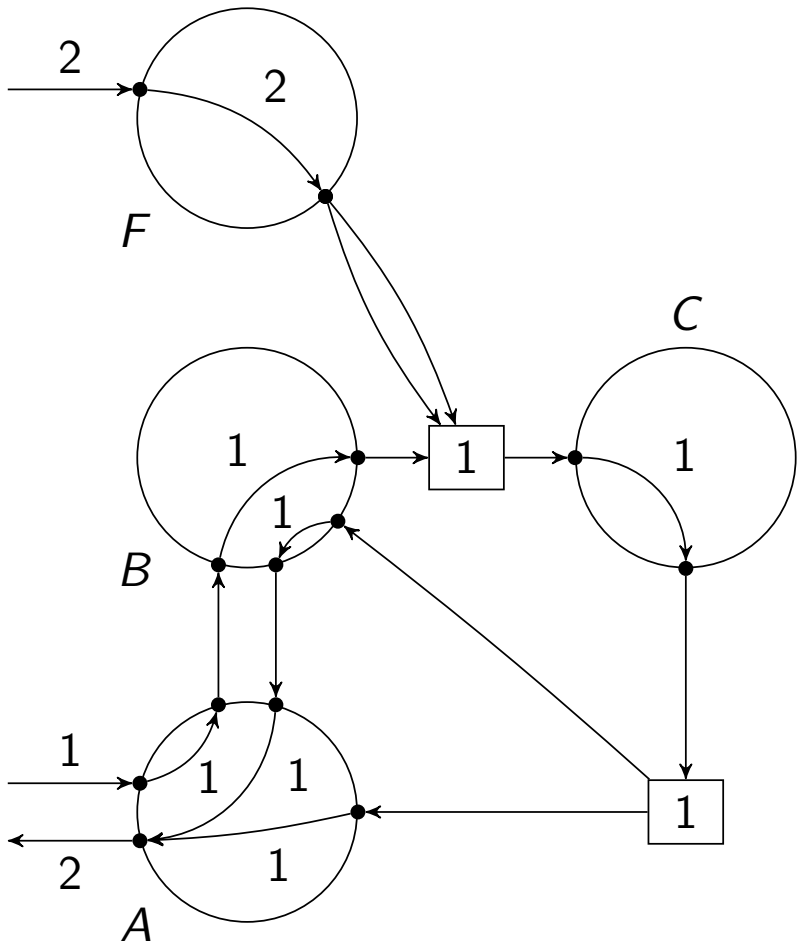
Conservation constraints:

$$\sum_{e \in \delta^+_{\widetilde{E}}(v)} m_v(e^+) f(e) - \sum_{e \in \delta^-_{\widetilde{E}}(v)} m_v(e^-) f(e) = 0 \qquad \forall v \in \widetilde{V}$$

Modelling and Analysis of Chemical Systems

5. Formulate pathway question.

Example: Given 2 formaldehyde and 1 glycolaldehyde, how can 2 glycolaldehyde be produced through autocatalysis.



Modelling and Analysis of Chemical Systems

6. Enumerate many alternate pathways.

Example (Formose):

Network: all molecules with at most 9 carbon atoms (284 molecules, 978 reactions).

	Maximum #C						
Reactions used	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	—	≥ 1185
11	0	5	41	190	738	—	≥ 974
							≥ 2431

Feasibility

How about the laws of physics?
Are all of those pathways feasible in reality?

Feasibility

How about the laws of physics?
Are all of those pathways feasible in reality?

Reality is complicated...

Idea: Add constraints from thermodynamics.
Not an entirely new idea:

Thermodynamics-Based Metabolic Flux Analysis
C. S. Henry, L. J. Broadbelt, V. Hatzimanikatis
Biophysical Journal 2007

Thermodynamics

Each molecule v has

- a standard Gibbs free energy $G^\circ(v)$.
- a concentration $K(v)$

Thermodynamics

Each molecule v has

- a standard Gibbs free energy $G^\circ(v)$.
- a concentration $K(v)$

Change of free energy for a reaction e:

$$\Delta G^\circ(e) = \sum_{v \in_m e^-} G^\circ(v) - \sum_{v \in_m e^+} G^\circ(v)$$

Thermodynamics

Each molecule v has

- a standard Gibbs free energy $G^\circ(v)$.
- a concentration $K(v)$

Change of free energy for a reaction e:

$$\Delta G^\circ(e) = \sum_{v \in_m e^-} G^\circ(v) - \sum_{v \in_m e^+} G^\circ(v)$$

The actual change of energy:

$$\Delta G(e) = \Delta G^\circ(e) + RT \ln K(e)$$

with

$$\ln K(e) = \sum_{v \in_m e^-} \ln K(v) - \sum_{v \in_m e^+} \ln K(v)$$

. . . and energy should go down: $\Delta G(e) < 0$.

Thermodynamics using Mixed Linear Programming

Each molecule v has

- a free energy estimate $G^{\circ}_{est}(v)$, with an error estimate $\sigma(v)$.
- a variable for free energy $x_v^{G^{\circ}} \in [G^{\circ}_{est}(v) - \sigma(v); G^{\circ}_{est}(v) + \sigma(v)]$.
- a variable for concentration $x_v^K \equiv \ln K(v)$.

Thermodynamics using Mixed Linear Programming

Each molecule v has

- a free energy estimate $G^{\circ}_{est}(v)$, with an error estimate $\sigma(v)$.
- a variable for free energy $x_v^{G^{\circ}} \in [G^{\circ}_{est}(v) - \sigma(v); G^{\circ}_{est}(v) + \sigma(v)]$.
- a variable for concentration $x_v^K \equiv \ln K(v)$.

A reaction e has a variable for change in energy:

$$x_e^{\Delta G} = x_e^{\Delta G^{\circ}} + RT \cdot x_e^K$$

with

$$x_e^{\Delta G^{\circ}} = \sum_{v \in_m e^-} x_v^{G^{\circ}} - \sum_{v \in_m e^+} x_v^{G^{\circ}} \qquad x_e^K = \sum_{v \in_m e^-} x_v^K - \sum_{v \in_m e^+} x_v^K$$

Thermodynamics using Mixed Linear Programming

Each molecule v has

- a free energy estimate $G^{\circ}_{est}(v)$, with an error estimate $\sigma(v)$.
- a variable for free energy $x_v^{G^{\circ}} \in [G^{\circ}_{est}(v) - \sigma(v); G^{\circ}_{est}(v) + \sigma(v)]$.
- a variable for concentration $x_v^K \equiv \ln K(v)$.

A reaction e has a variable for change in energy:

$$x_e^{\Delta G} = x_e^{\Delta G^{\circ}} + RT \cdot x_e^K$$

with

$$x_e^{\Delta G^{\circ}} = \sum_{v \in_m e^-} x_v^{G^{\circ}} - \sum_{v \in_m e^+} x_v^{G^{\circ}} \qquad x_e^K = \sum_{v \in_m e^-} x_v^K - \sum_{v \in_m e^+} x_v^K$$

If energy does not go down, then no flow on the reaction:

$$x_e^{\Delta G} \geq 0 \Rightarrow x_e = 0$$

and in linear form (with a "big M"):

$$x_e^{\Delta G} < M \cdot (1 - x_e^+)$$

(with x_e^+ indicating positive flow on e).

Summary

1. Molecule modelling (MØD + a chemist)
2. Reaction type modelling (MØD + more chemists)
3. Network expansion (MØD)
4. Pathway modelling (MØD + CPLEX)
5. Pathway questions (a chemist)
6. Enumeration of thermodynamically feasible pathways (MØD + Nikolai's stuff (and no chemist!))

Thanks

Earth-Life Science Institute
Tokyo Institute of Technology

ELSI Origins Network

Department of Mathematics and Computer Science, University of Southern Denmark