Reduction of Graph-Grammar Models by grouping of tautomers

Emil Houtved Andersen, Jakob L. Andersen, Daniel Merkle

Department of Mathematics and Computer Science,
University of Southern Denmark

Bled, February 2019

Overview

In this presentation, we will be talking about:

- Sugar chemistry, in particular, the formose chemistry
- Equivalence Classes
- Auto catalysis
- Gillespie's exact stochastic simulation algorithm
- Bisimulation
- Markov Chains

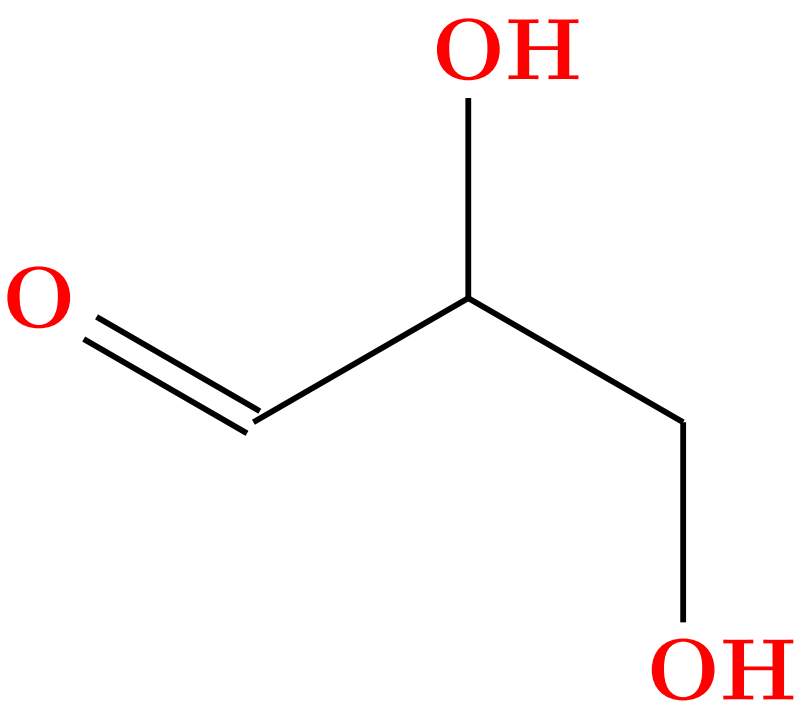
What are we doing?

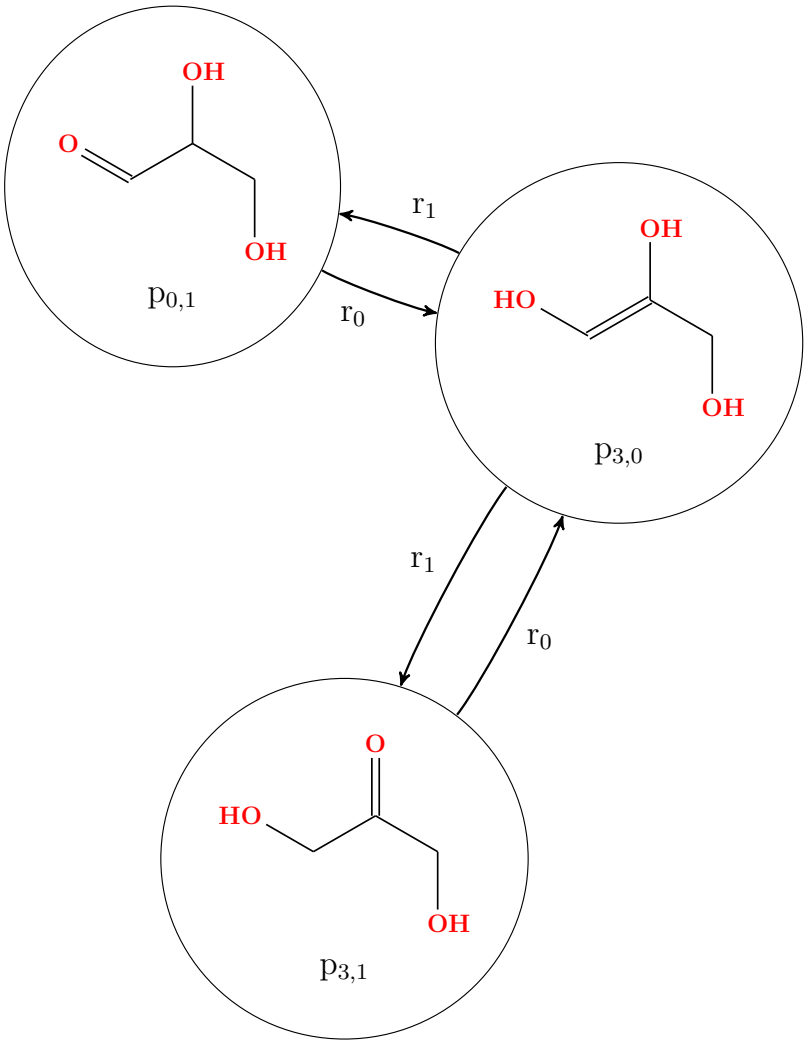
- Derivation graphs can get LARGE
- Many nodes may have extremely similar properties
- Stack these nodes together in a projection

Why do this?

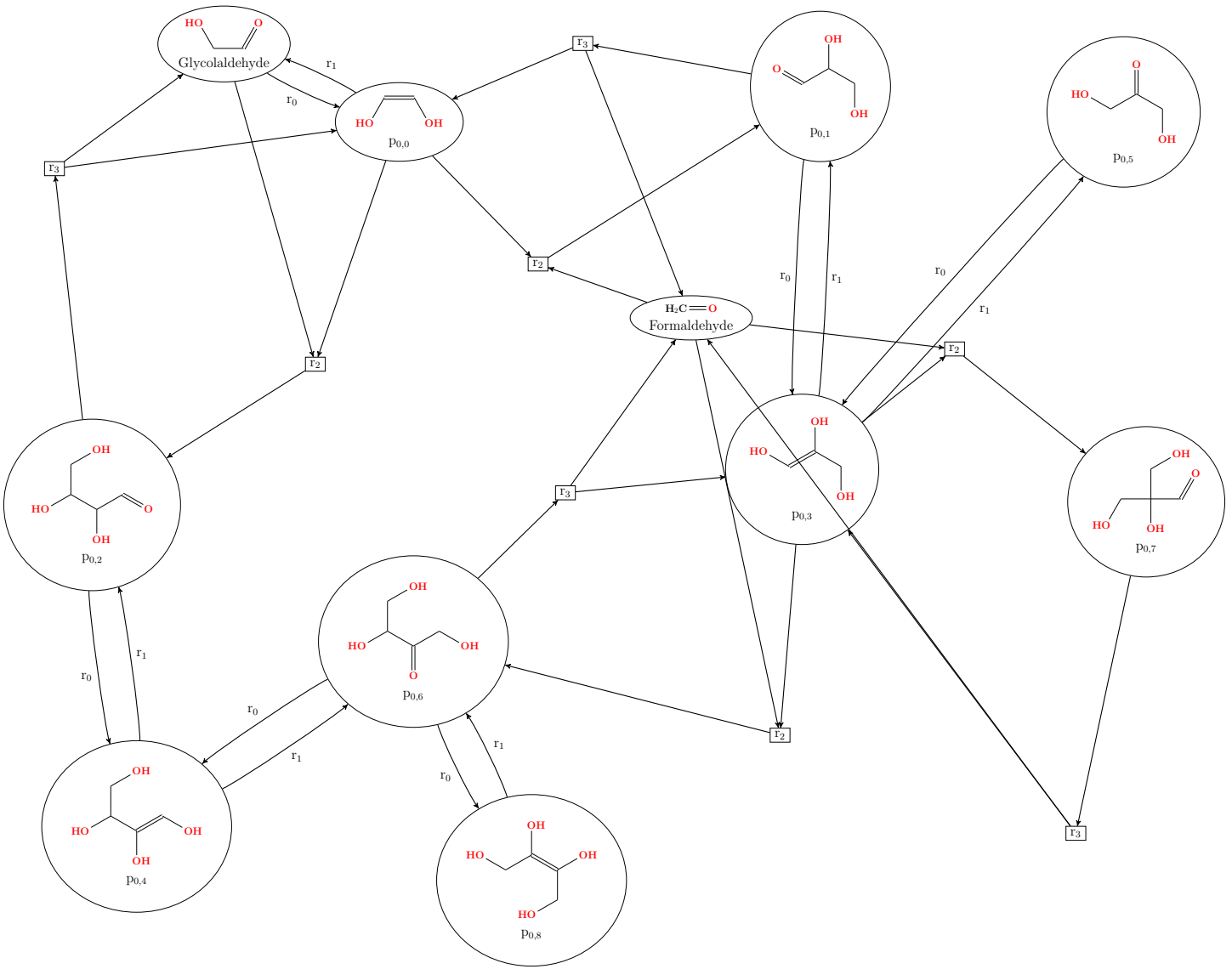
- Projected derivation graphs have much fewer nodes
- Much faster auto catalysis computation, and more correct results
- Stochastic simulations run significantly faster, and potentially produce more correct results

Equivalence Class in practice

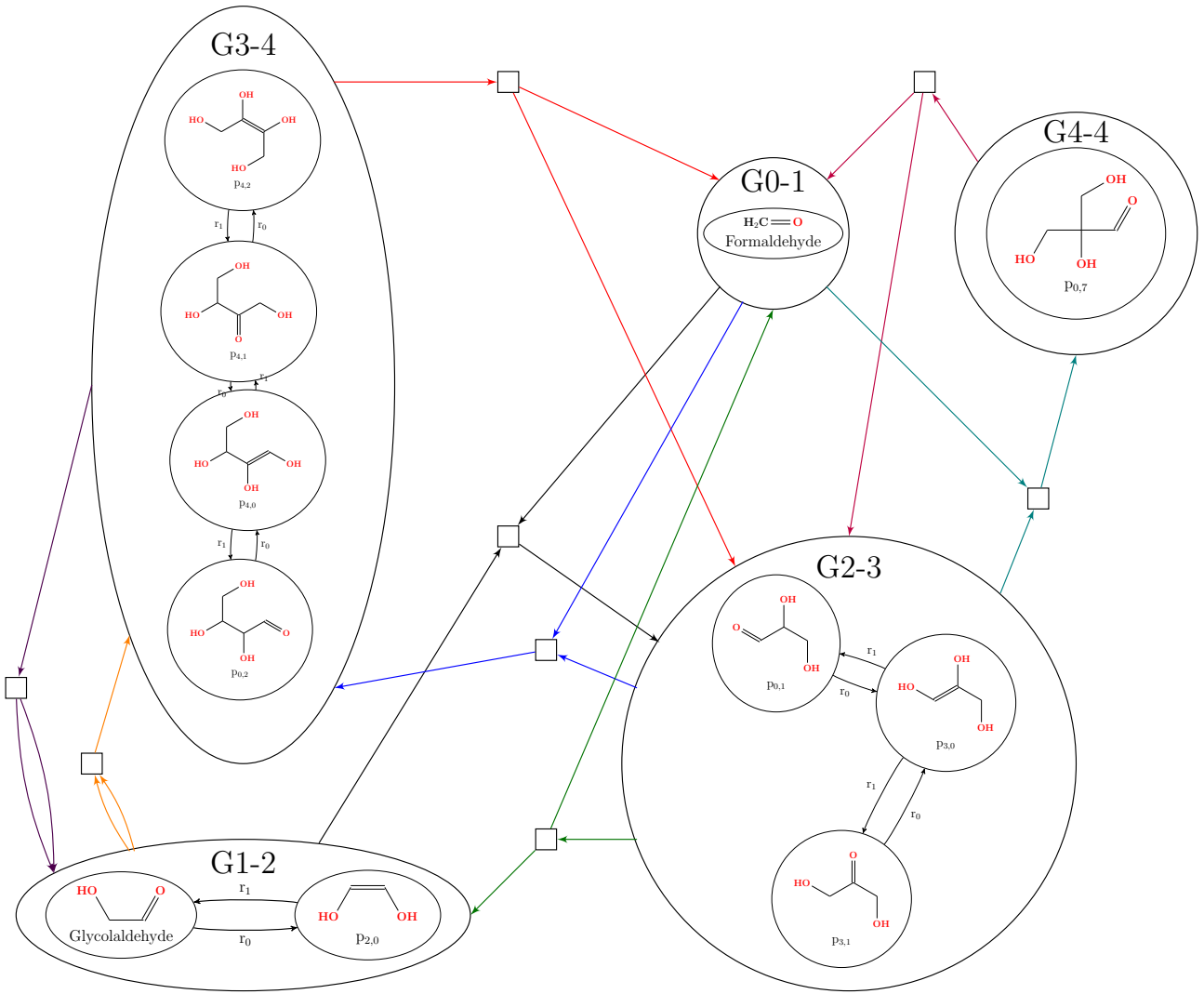




Formose reaction autocatalysis example: Unprojected



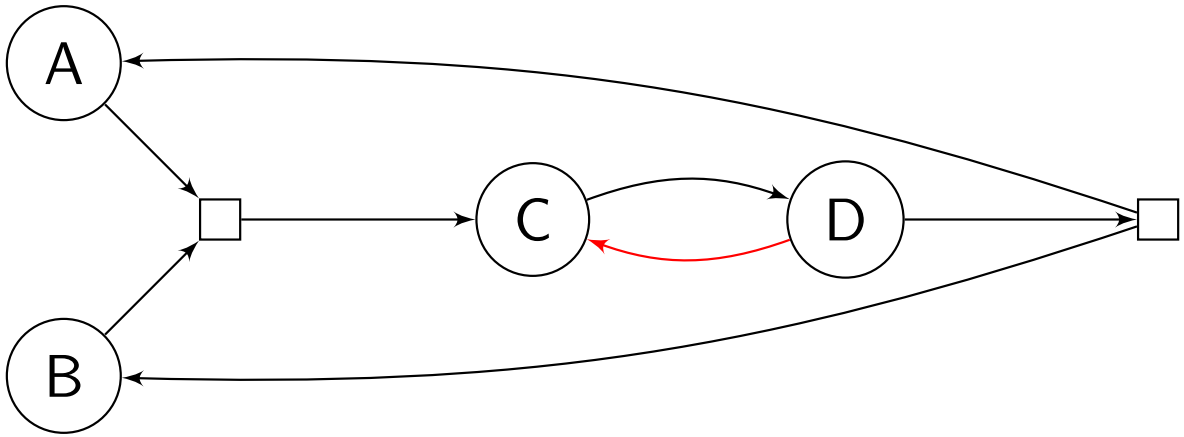
Formose reaction autocatalysis example: Projection

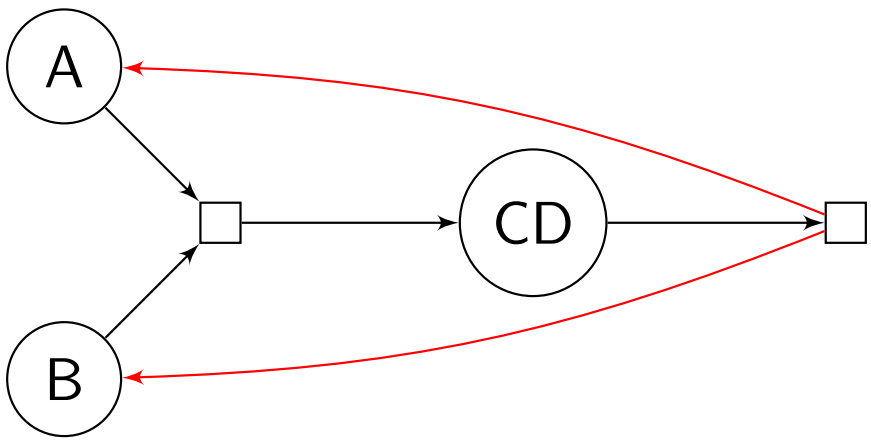


Equivalence Class Theory

- $\{x \in X : xRy\}$
- y is an element of X
- The notation "xRy" means there is an equivalence relation between x and y
- for all $x, y \in X$, we have $xRy \Leftrightarrow x$ and y belong to the same equivalence class

Why is projected auto catalysis more correct?





Why could projected stochastic simulation be more correct?

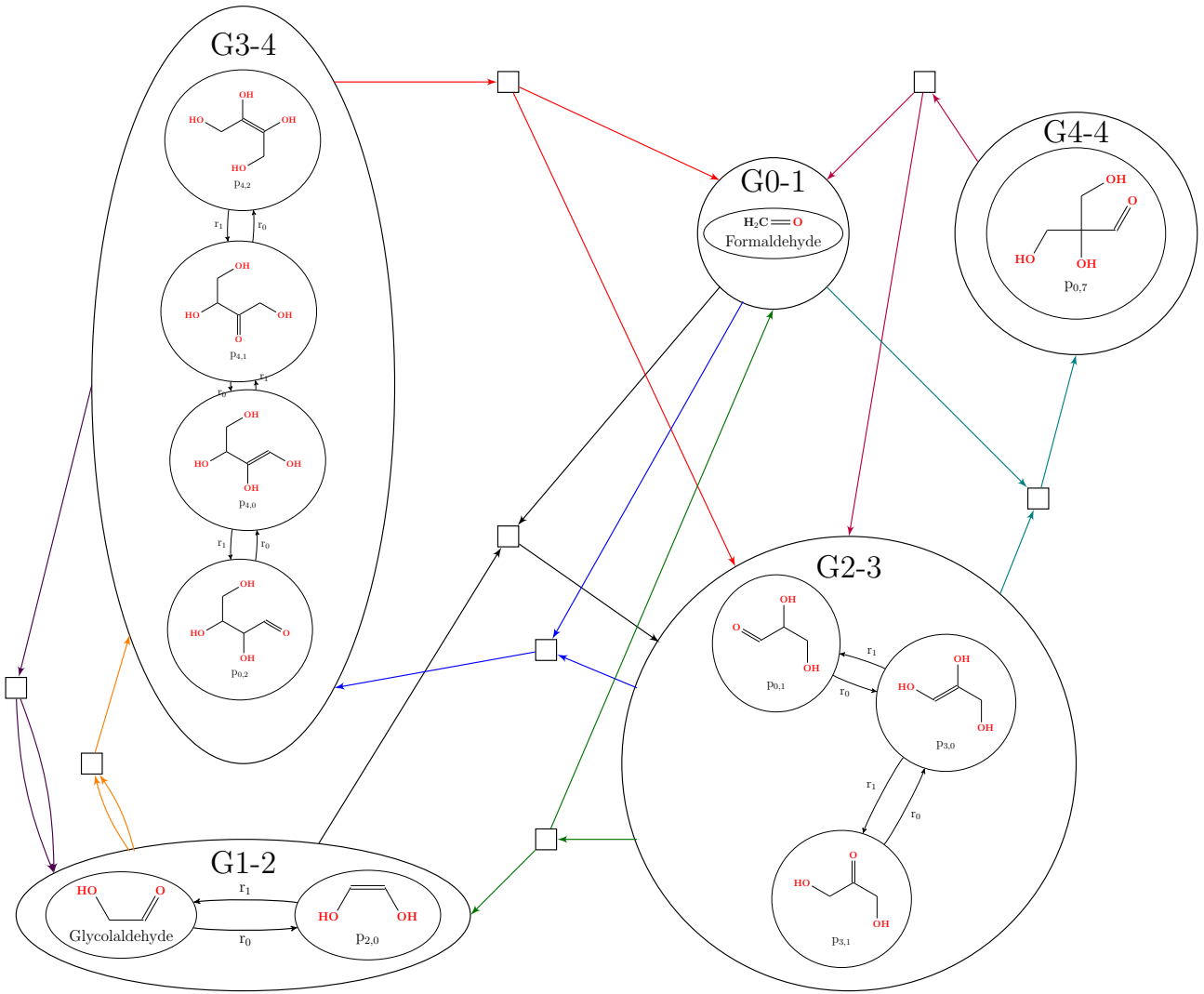
- Keto-Enol reactions are said to be extremely fast, especially comapred to Aldol-addition reactions
- Getting a realistic relative reaction rate, and also a reasonable computation time, is not possible normally
- "Removing" the Keto-Enol reactions from the equation makes it much easier to get something closer to real life

Concurrency Theory: Bisimulations

Bisimilarity definitions:

- *Strong Bisimilarity*: A binary relation $\mathcal{R}$ over the set of states of an Labelled Transition System is a strong bisimulation if and only if whenever $s_1 \mathcal{R} s_2$ and α is an action:
  - if $s_1 \xrightarrow{\alpha} s_1'$, then there is a transition $s_2 \xrightarrow{\alpha} s_2'$ such that $s_1' \mathcal{R} s_2'$, and
  - if $s_2 \xrightarrow{\alpha} s_2'$, then there is a transition $s_1 \xrightarrow{\alpha} s_1'$ such that $s_1' \mathcal{R} s_2'$
- *Weak Bisimilarity*: A binary relation $\mathcal{R}$ over the set of states of a Labelled Transition System is a weak bisimulation if and only if whenever $s_1 \mathcal{R} s_2$ and α is an action (Including τ, the internal action):
  - If $s_1 \xrightarrow{\alpha} s_1'$, then there is a transition $s_2 \stackrel{\alpha}{\Rightarrow} s_2'$ such that $s_1' \mathcal{R} s_2'$ and
  - If $s_2 \xrightarrow{\alpha} s_2'$, then there is a transition $s_1 \stackrel{\alpha}{\Rightarrow} s_1'$ such that $s_1' \mathcal{R} s_2'$.

Concurrency Theory: Bisimulations, example



Results

- A "Unique reaction" in the following table means specifically a 1-2 or 2-1 reaction.
- The autocatalysis results are best described by how many different solutions are found and how fast. This is reflected in how much of the table is actually filled out.
- The stochastic simulation results are best described by how fast a simulation reaches its end, and how far the simulation got towards its end.

Autocatalysis results: Unprojected

Cells: How many different solutions
Columns: How many carbons are contained in the largest molecule in the solution

	4 C	5 C	6 C	7 C	8 C	9 C	Sum
3 Unique Reactions	1	0	0	0	0	0	1
4 Unique Reactions	0	2	8	4	6	12	32
5 Unique Reactions	0	32	119	394	927	—	1472
6 Unique Reactions	1	36	412	1640	4880	—	6969
7 Unique Reactions	0	92	2556	25586	—	—	28234
8 Unique Reactions	0	144	10053	137364	—	—	147561
9 Unique Reactions	0	185	45469	—	—	—	45654
10 Unique Reactions	0	239	170536	—	—	—	170775

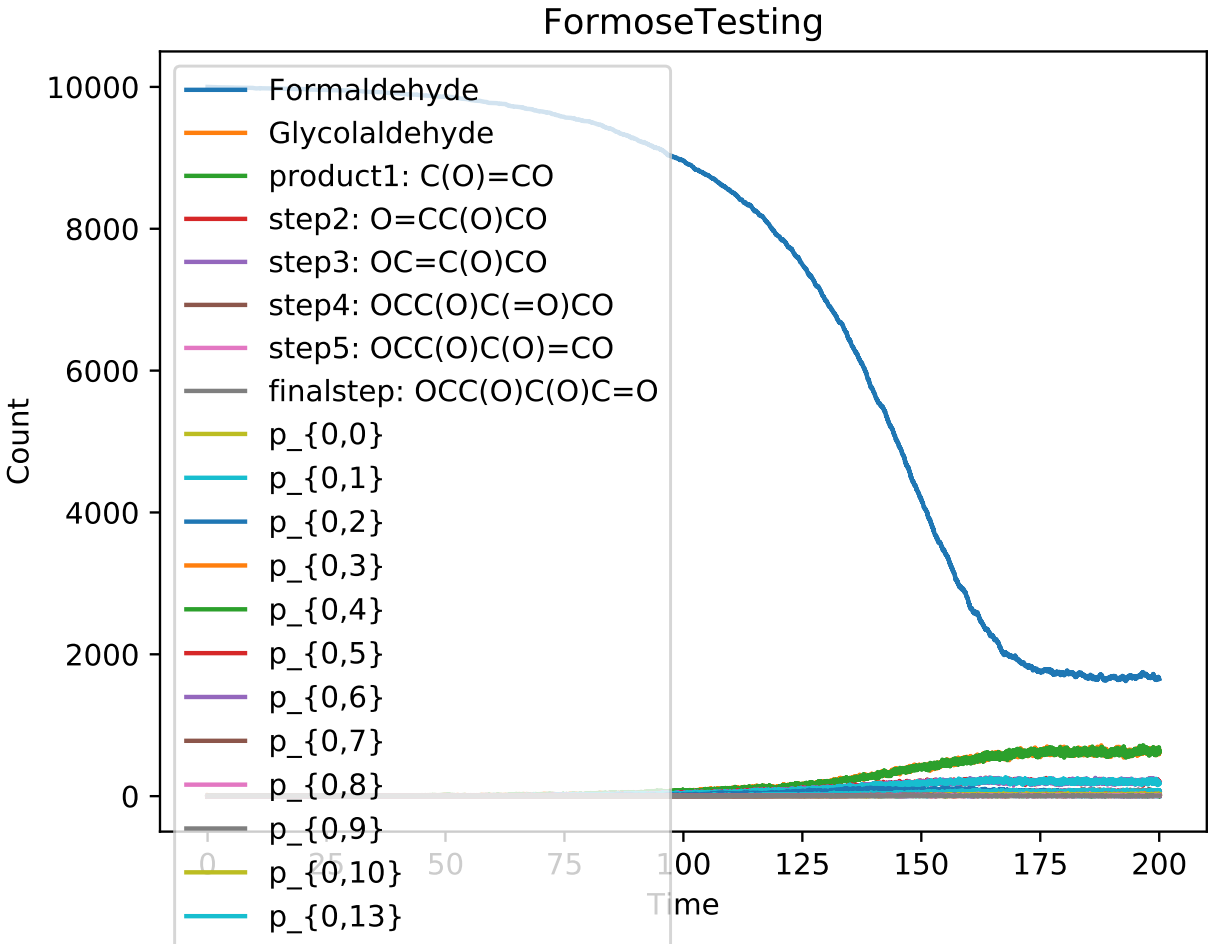
Autocatalysis results: Projection

	4 C	5 C	6 C	7 C	8 C	9 C	Sum
3 Unique Reactions	1	0	0	0	0	0	1
4 Unique Reactions	0	0	2	0	0	0	2
5 Unique Reactions	0	18	29	106	196	357	706
6 Unique Reactions	1	21	162	561	1278	—	2023
7 Unique Reactions	0	57	891	7271	27768	—	35987
8 Unique Reactions	0	102	4012	45817	—	—	49931
9 Unique Reactions	0	125	17529	377398	—	—	395052
10 Unique Reactions	0	166	67407	—	—	—	67573

Stochastic simulation results: Unprojected

Runtime: 81 minutes and 51.395 seconds

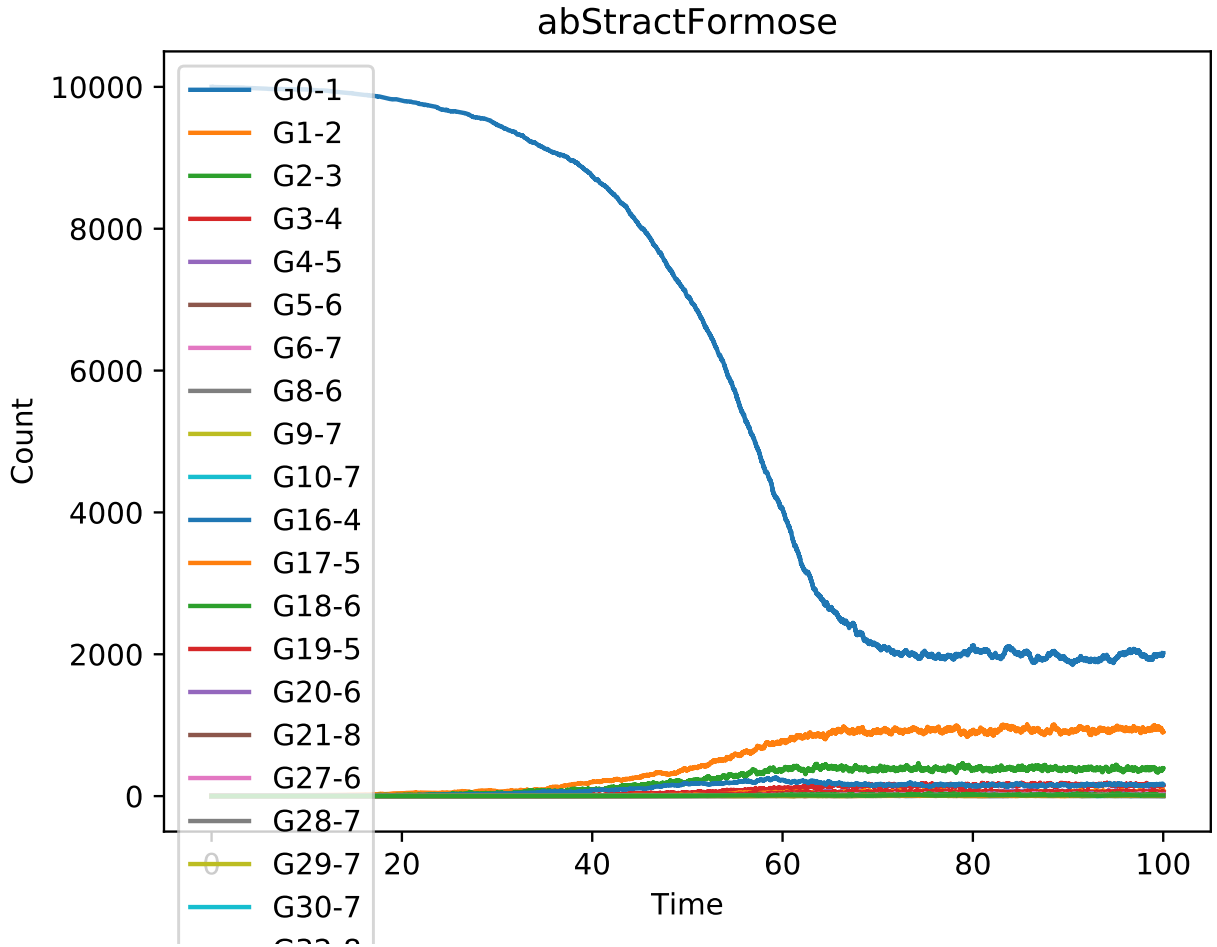
Details: 100 times higher reaction rate on keto-enol than on aldol-addition



Stochastic simulation results: Projected

Runtime: 6 minutes and 26.545 seconds

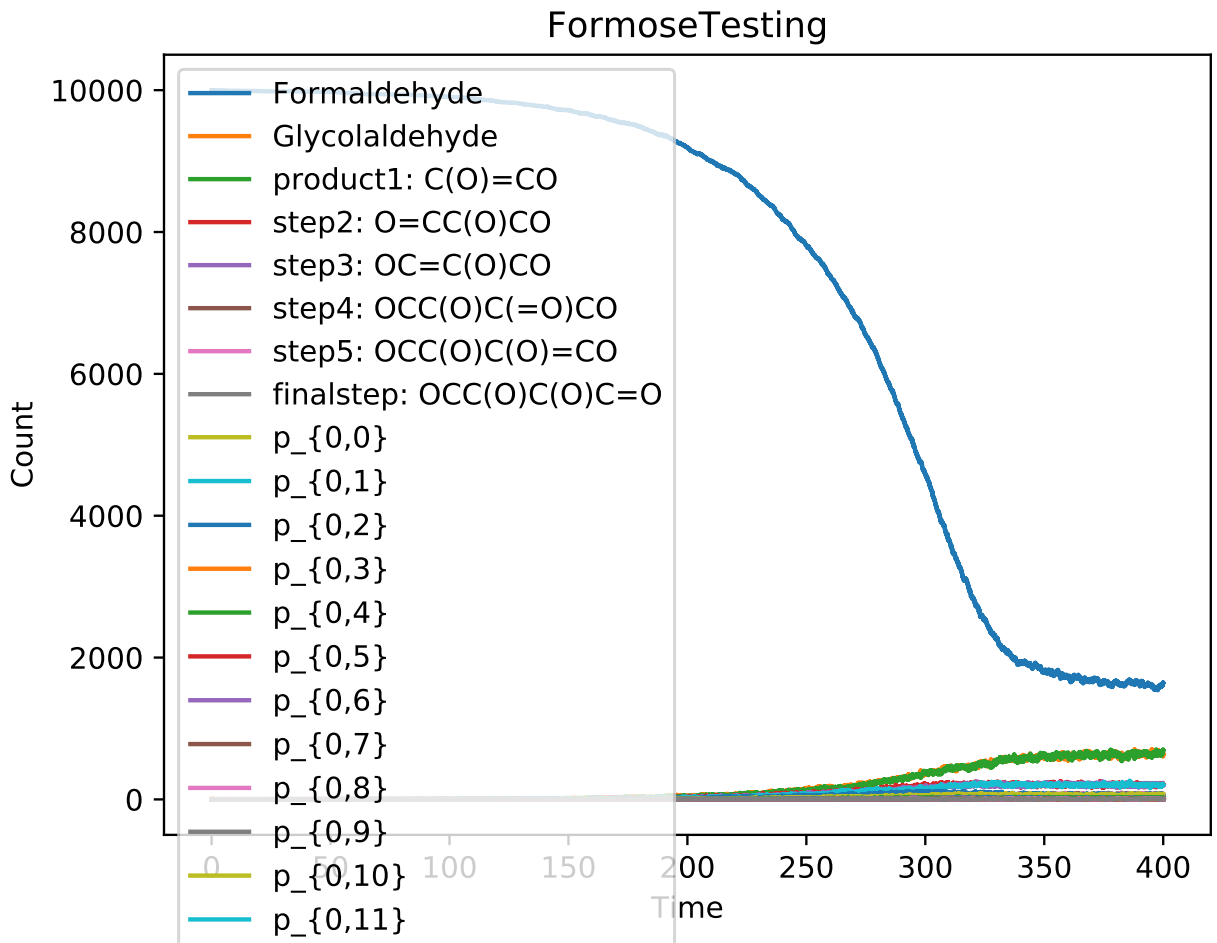
Details: Same reaction rate on aldol-addition as for unprojected



Stochastic simulation results: Unprojected 2

Runtime: 25 minutes and 18.185 seconds

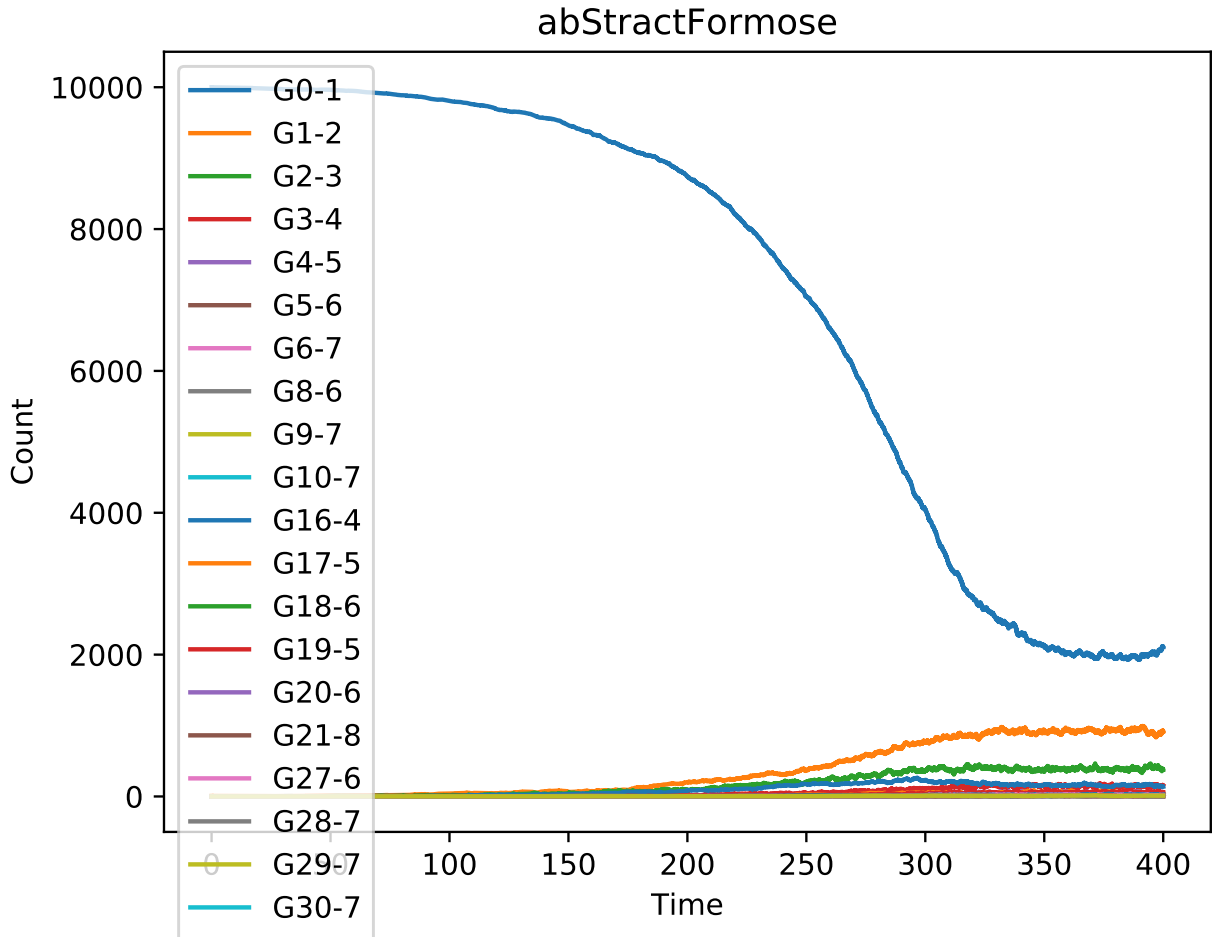
Details: Keto-enol reaction rate same as aldol-addition



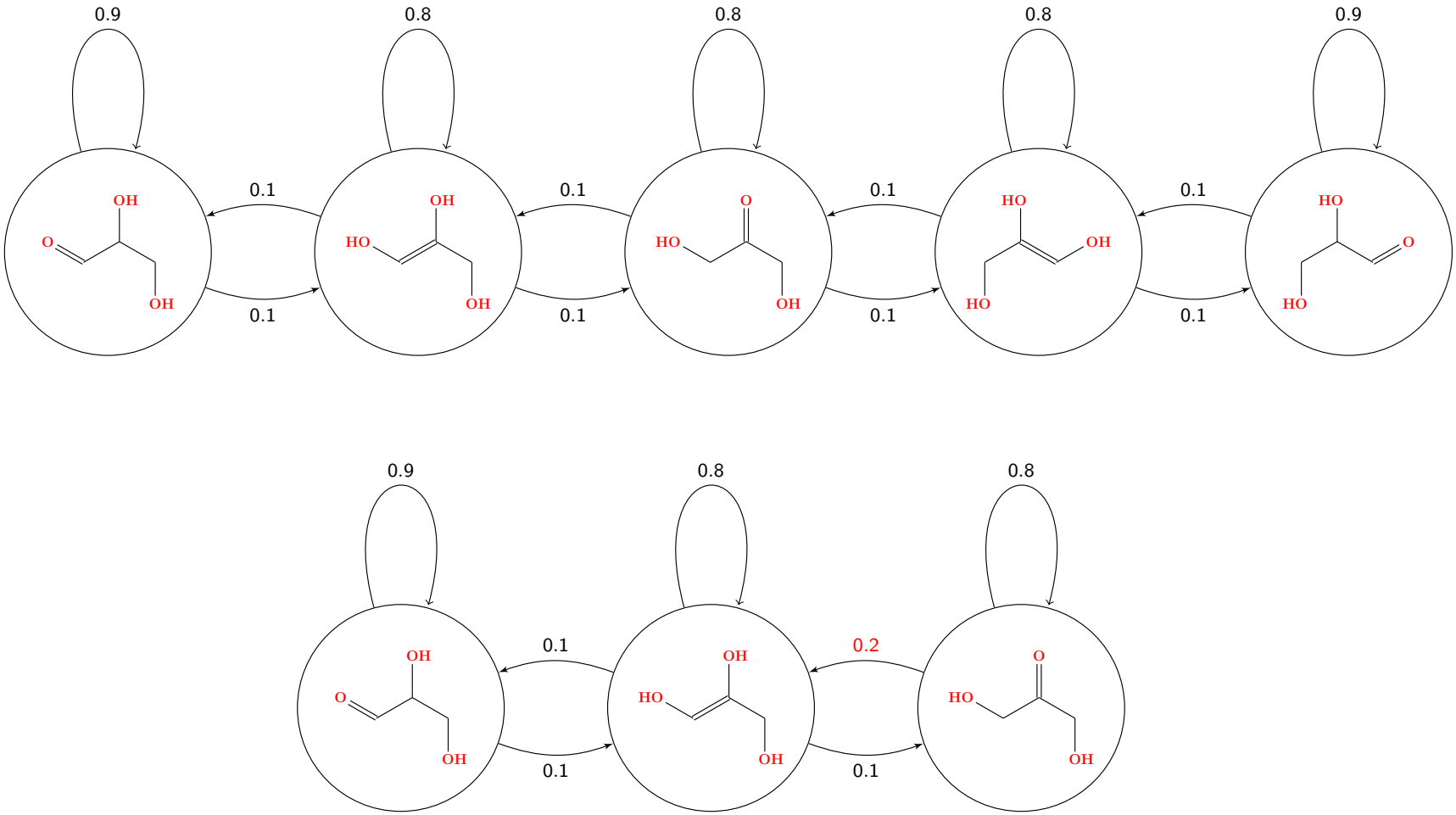
Stochastic simulation results: Projected 2

Runtime: 3 minutes and 42.391 seconds

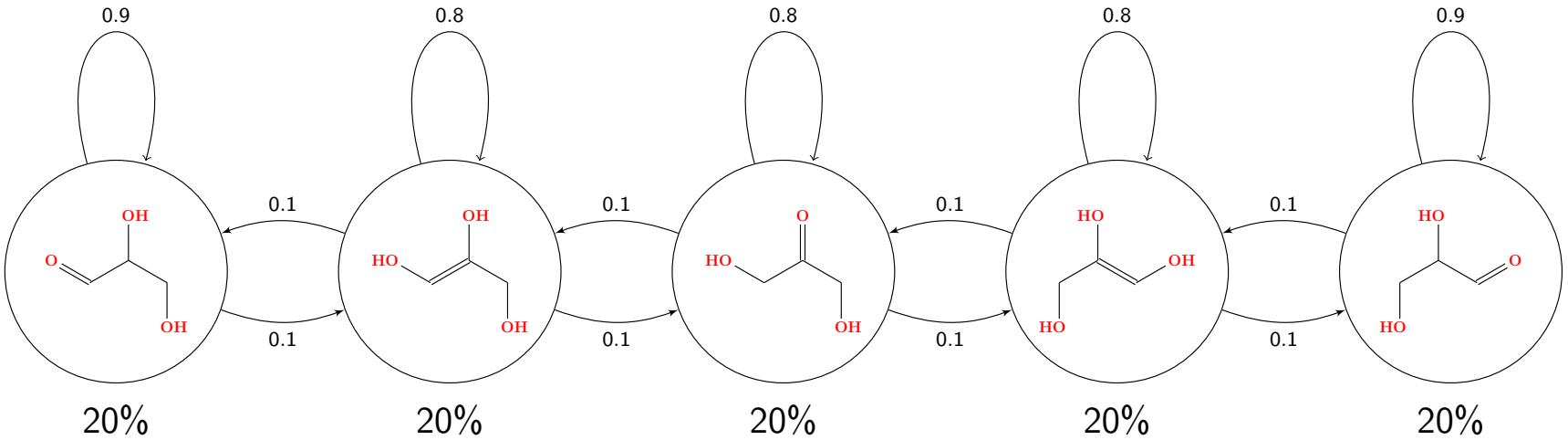
Details: reaction rate 1/5th of unprojected.

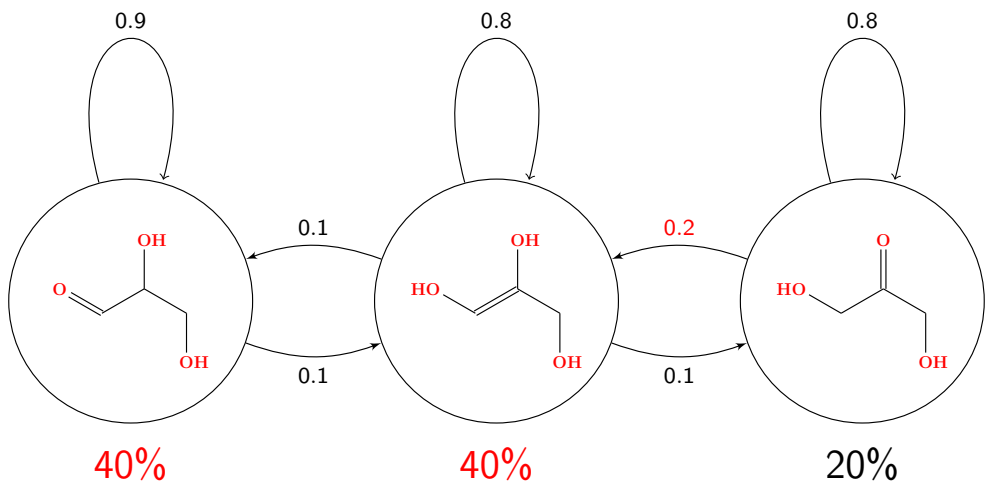


Markov Chains: Example 1: Formose reaction



Markov Chains: Example 1: Formose reaction





Summary

- Equivalence classes can be used to reduce derivation graphs based on (for instance) tautomerisms
- Such projections improve both the computation time and quality of auto catalysis and
- also improves the computation time of Stochastic simulations, and may sometimes also improve the quality of the stochastic simulation itself
- Markov Chains can be used to show that the projection does not appear to cause any negative effect on the concentration differences in such a stochastic simulation
- The comparison between the unprojected and projected solutions can be proven to be weakly bisimilar