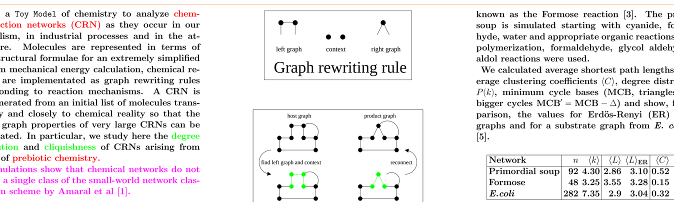
GENERATING PREBIOTIC CHEMISTRIES FOR THE STUDY OF GENERIC GRAPH PROPERTIES

G. Benkö, C.Flamm[1], P.F. Stadler

Bioinformatics Group, Department of Computer Science, University of Leipzig, Germany
[1]Institute for Theoretical Chemistry and Structural Biology, University of Vienna, Austria

Tel: ++49 341 149 5120 Fax: ++49 341 149 5119 Email: {gil, xtof, studla}@bioinf.uni-leipzig.de
WorldWideWeb: http://www.bioinf.uni-leipzig.de/



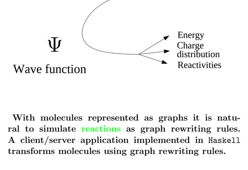
We use a `Toy Model` of chemistry to analyze chemical reaction networks (CRN) as they occur in our metabolism, in industrial processes and in the atmosphere. Molecules are represented in terms of usual structural formulae for an extremely simplified quantum mechanical energy calculation, chemical reactions are implementated as graph rewriting rules corresponding to reaction mechanisms. A CRN is thus generated from an initial list of molecules transparently and closely to chemical reality so that the generic graph properties of very large CRNs can be investigated. In particular, we study here the degree distribution and cliquishness of CRNs arising from models of prebiotic chemistry.
Our simulations show that chemical networks do not fall into a single class of the small-world network classification scheme by Amaral et al [1].

The `Toy Model`

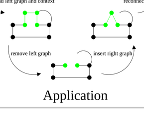
The `Toy Model` [3] is based on a simulation of molecules, reaction mechanisms, and large-scale networks. Molecules undergo reactions, and the set of molecules and reactions can be represented as nodes and edges that form a network.

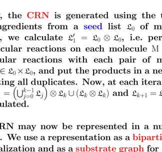
The molecule simulation uses an Extended Hückel Theory-style calculation (EHT). Normally, three-dimensional coordinates are needed for such a calculation. Our simplification of EHT consists in using only the topology of the molecules.

Thus the graph representation of the molecule suffices to calculate a chemically meaningful energy and reactivity.



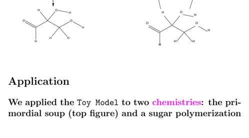
With molecules represented as graphs it is natural to simulate reactions as graph rewriting rules. A client/server application implemented in `Haskell` transforms molecules using graph rewriting rules.





Finally, the CRN is generated using the two previous ingredients from a seed list $\mathfrak{L}_0$ of molecules. Initially, we calculate $\mathfrak{L}'_1 = \mathfrak{L}_0 \otimes \mathfrak{L}_0$, i.e. perform all unimolecular reactions on each molecule $\mathrm{M} \in \mathfrak{L}_0$, all bimolecular reactions with each pair of molecules $(\mathrm{M}_1, \mathrm{M}_2) \in \mathfrak{L}_0 \times \mathfrak{L}_0$, and put the products in a new set $\mathfrak{L}'_1$, eliminating all duplicates. Now, at each iteration, the sets $\mathfrak{L}'_{k+1} = \left(\bigcup_{j=0}^{k-1} \mathfrak{L}_j\right) \otimes \mathfrak{L}_k \cup (\mathfrak{L}_k \otimes \mathfrak{L}_k)$ and $\mathfrak{L}_{k+1} = \mathfrak{L}'_{k+1} \setminus \bigcup \mathfrak{L}_k$ are calculated.

The CRN may now be represented in a number of ways [4]. We use a representation as a bipartite graph for visualization and as a substrate graph for analysis.



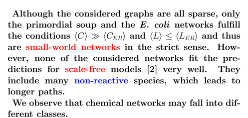
Application

We applied the `Toy Model` to two chemistries: the primordial soup (top figure) and a sugar polymerization known as the Formose reaction [3]. The primodial soup is simulated starting with cyanide, formaldehyde, water and appropriate organic reactions; for the polymerization, formaldehyde, glycol aldehyde and aldol reactions were used.

We calculated average shortest path lengths $\langle L \rangle$, average clustering coefficients $\langle C \rangle$, degree distributions $P(k)$, minimum cycle bases (MCB, triangles Δ and bigger cycles $\mathrm{MCB}' = \mathrm{MCB} - \Delta$) and show, for comparison, the values for Erdös-Renyi (ER) random graphs and for a substrate graph from ***E. coli*** from [5].

Network	n	$\langle k \rangle$	$\langle L \rangle$	$\langle L \rangle_{\mathrm{ER}}$	$\langle C \rangle$	$\langle C \rangle_{\mathrm{ER}}$
Primordial soup	92	4.30	2.86	3.10	0.52	0.04
Formose	48	3.25	3.55	3.28	0.15	0.068
E.coli	282	7.35	2.9	3.04	0.32	0.026

Network	$\lVert\Delta\rVert$	$\lVert\Delta\rVert_{\mathrm{ER}}$	$\lVert\mathrm{MCB}'\rVert$	$\lVert\mathrm{MCB}'\rVert_{\mathrm{ER}}$
Primordial soup	76	10	31	97
Formose	22	2.2	7	26.8
E.coli	282	18	99	363



Although the considered graphs are all sparse, only the primordial soup and the ***E. coli*** networks fulfill the conditions $\langle C \rangle \gg \langle C_{ER} \rangle$ and $\langle L \rangle \leq \langle L_{ER} \rangle$ and thus are small-world networks in the strict sense. However, none of the considered networks fit the predictions for scale-free models [2] very well. They include many non-reactive species, which leads to longer paths.

We observe that chemical networks may fall into different classes.

References

[1] L. A. N. Amaral, A. Scala, M. Barthelemy, and H. E. Stanley. Classes of small-world networks. *Proc. Nat. Acad. Sci.*, 97:11149–11152, 2000.

[2] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[3] G. Benkö, C. Flamm, and P. F. Stadler. A graph-based toy model of chemistry. *J. Chem. Inf. Comput. Sci.*, 2003. in press; presented at *MCC 2002*, Dubrovnik CRO, June 2002; SFI # 02-09-045.

[4] D. Bonchev and O. Mekenyan, editors. *Graph-Theoretical Approaches to Chemical Reactivity*. Kluwer, Dordrecht, NL, 1994.

[5] D. Fell and A. Wagner. The small world inside metabolic networks. *Proc. R. Soc. Lond. Ser. B*, 268:1803–1810, 2001.

Printed by the Universitätsrechenzentrum Leipzig