Topology of Chemical Reaction Networks

Christoph Flamm [a,b], and Peter F. Stadler [a,b,c]

[a] *Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Härtelstraße 16-18, D-04107 Leipzig, Germany*

[b] *Department of Theoretical Chemistry University of Vienna, Währingerstraße 17, A-1090 Wien, Austria*

[c] *Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM 87501, USA*

Key words: chemical reaction networks, chemical organization theory, graph rewrite, closure function, stoichiometric matrix

---

A distinguishing feature of chemistry is that the changes of molecules upon interaction are not limited to quantitative physical properties such as free energy, density, or concentrations. Molecular interactions not only produce more of what is already there but novel molecules can be generated. This is the principal difficulty for any theoretical treatment of complex chemical systems. Dynamical systems models start from a fixed set a molecular species and trace the time-dependence of their concentrations, in the same manner as population dynamics describes frequencies of genes or organisms. In this picture, molecules, genes, and biological species are reduced to mere indices of the concentration variables, while their properties are implicitly described by rate constants that must be determined outside these theories. Currently approximately 27 million compounds are described in the chemical databases[1], but the number of conceivable molecules is in principle infinite.

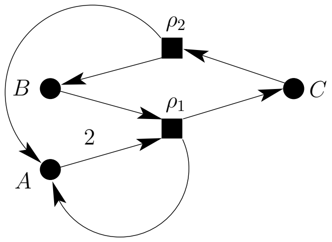
The most interesting question in chemistry is, however, *which* molecules arise in a given situation. To address this point, Walter Fontana and Leo Buss[2,3] introduced the notion of a *constructive dynamical system* in which the interactions, i.e., the chemical reactions, are viewed as algebraic operations on the set of molecular types. Walter Fontana's `AlChemy`[3], for example represents molecules as λ-calculus expressions and reactions are defined by the operations of "application" of one λ-term to its reaction partner. The result is a new λ-term. Since λ-terms are a rather harsh abstraction of 'real' molecules Benkö *et. al.*[4] resort to graphs to represent molecules and graph rewriting to model reactions. A wide variety of different computational paradigms has been used in artificial chemistry models from strings and matrices to Turing machines, see[5] for a review. The approaches which target "real" chemistry are heavily geared towards computer-assisted organic synthesis (CAOS)[6]. These programs however rely on knowledge bases which limit their chemical horizon to the information stored in chemical databases which obviously possess a heavy sampling bias.

The `ToyChem` model represents molecules as vertex and edge labeled graphs, a natural abstraction in the spirit of most basic organic chemistry textbooks. Such a chemical graph now uniquely determines the *orbital graph*, an undirected graph with nodes representing outer atom orbitals, labeled by the atomic element and the hybridization type of the orbital and edges representing overlaps of adjacent orbitals. A simplified Extended Hückel Theory approach based on the orbital graph is used to give the `ToyChem` Model a chemically meaningful energy function that allows the computation of physicochemical properties of molecules and reaction energies. By avoiding the computationally intensive step of embedding the molecules in three-dimensional space and sticking to the graph representation instead all steric information i.e. chirality and Z/E-isomers are deliberately ignored in the `ToyChem` Model. Basic organic reaction mechanisms translate smoothly to graph rewriting rules, which take on the role of named reactions. Explicit collision experiments can be simulated in this universe, so that chemical knowledge in the form of "new" rewriting rules can be extracted[7]. The `ToyChem` universe provides a valuable to study the structure of large chemical networks without the burden and the biased of actually modeling all known described substances and chemical reactions explicitly. It is therefore a computationally tractable proving ground for mathematical and algorithmic approaches.

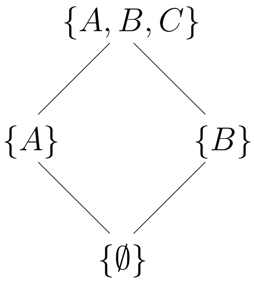
Large chemical reaction networks have been studied extensively in quite different domains, from metabolic networks[8] to the chemistry of planetary athmospheres[9], combinatorial chemistry, see e.g.[10], and thermal cracking[11]. Furthermore, artificial chemistry models at various levels of realism can be used to generate very large reaction networks, with the particular aim to determine generic properties[12].

From a formal point of view, however, these examples can be treated in a unified framework: A chemical reaction network simply consists of a set of chemical species (metabolites in the biochemical setting) X and a system $\mathfrak{R}$ of reactions $\rho_1, \dots, \rho_n$ that inter-converts them. Several "mathematical dialects" have been used to further analyze $(X, \mathfrak{R})$.

Chemical reaction networks are thus *directed hypergraphs*[13], a generalization of conventional graphs in which an arc (edge) can have more than one beginning and more than one end. However, there exists a one-to-one mapping between a hypergraph and a conventional graph. This so called König repre-



U	cl(U) buffered	cl(U) unbuffered
$\{\emptyset\}$	$\{\emptyset\}$	$\{\emptyset\}$
$\{A\}$	$\{A\}$	$\{\emptyset\}$
$\{B\}$	$\{B\}$	$\{\emptyset\}$
$\{C\}$	$\{A,B,C\}$	$\{A,B\}$
$\{A,B\}$	$\{A,B,C\}$	$\{C\}$
$\{A,C\}$	$\{A,B,C\}$	$\{A,B\}$
$\{B,C\}$	$\{A,B,C\}$	$\{A,B\}$
$\{A,B,C\}$	$\{A,B,C\}$	$\{A,B,C\}$



$$\frac{\mathrm{d}}{\mathrm{dt}}[\vec{X}] = S \cdot \vec{J} = \frac{\mathrm{d}}{\mathrm{dt}} \begin{pmatrix} [A] \\ [B] \\ [C] \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ -1 & 1 \\ 1 & -1 \end{pmatrix} \cdot \begin{pmatrix} k_{\rho_1}[A]^2[B] \\ k_{\rho_2}[C] \end{pmatrix}$$

Fig. 1. Reaction network consisting of the reactions $\rho_1 : 2A + B \rightarrow C + A$ and $\rho_2 : C \rightarrow A + B$. (**left**) König representation of the reaction network; (**middle**) Subsets U of the set of chemical species $\{A, B, C\}$ of the reaction network and the generalized closure functions (cl(U)) for the cases of buffered and unbuffered species concentrations; (**right**) The lattice of organizations of the reaction network. (**bottom**) The stoichiometric matrix $\mathbf{S}$ as linear transformation between the chemical flows $\vec{J}$ and the changes of species concentrations $[\vec{X}]$

sentation of the hypergraph is essentially a bipartite graph where the vertex set is partitioned into two classes, the reactions and the chemical species. If a chemical species participates (either as educt or as product) in a reaction, an edge is introduced between the respective vertices, Fig. 1.

In a recent study, Christian Forst demonstrated that complete metabolic networks of procaryotes can be compared directly, by applying familiar operations from set algebra to their hypergraph representations. This method opens up the route for both the pairwise comparison of metabolic networks and the identification of distinct metabolic features in two set of organisms. By defining a distance based on the symmetric difference between hypergraphs, the phylogenetic tree of the organisms can be reconstructed directly from the reaction networks[14].

The topology of a chemical reaction network is also largely embodied in the stoichiometric matrix $\mathbf{S}$, a special form of an incidence matrix, acting as linear transformation between the space of reaction rates and the time derivatives of concentration space. Compared to the hypergraph representation, $\mathbf{S}$ neglects catalysts, which is problematic in higher level descriptions, but harmless as long as one sticks to elementary reactions. Various methods have been devised to investigate invariant properties of the chemical reaction network based solely on the analysis of the structure of this matrix[15–18]. From the left null space, $\{\vec{z} | \vec{z}^{+}\mathbf{S} = \vec{0}\}$, of $\mathbf{S}$ conservation relations between chemical species can be extracted that do not change over time and form so-called "metabolic pools", while steady-state flux relationships[19] can be derived from the (right) null space $\{\vec{J}|\mathbf{S}\vec{J}=\vec{0}\}$. The (extremal rays of the) cone $\{\vec{J}|\mathbf{S}\vec{J}=\vec{0}\wedge J\geq 0\}$ defines the steady state pathways in the reaction network. Efficient algorithms[20,21] have been proposed for the elementary flux analysis of large-scale reaction networks that circumvent the problem of combinatorial explosion. Although metabolic flux analysis disregards the kinetics of the reactions nevertheless biologically meaningful states of the reaction network can be predicted which coincide nicely with experiments[22,23]. The stoichiometric matrix can be used to derive an ordinary differential equations model of the reaction network describing the dynamics of the concentration vector of the system $\dot{c}=\mathbf{S}J(c)$ by specifying how the flux J_ρ through reaction ρ depends on the concentrations (see Figure 1).

Peter Dittrich recently presented[24] a rigorous mathematical theory of chemical organization, based Fontana's work[3]. The theory captures the purely algebraic aspects of chemistry based on the notions of self-maintaining sets and closures in terms of properties of $\mathbf{S}$ and its flux vectors. The chemical organizations i.e. sets of self-maintaining chemical species, can be organized in a hierarchical lattice. There exists a links between the dynamical features of the reaction network to the static lattice of chemical organization. Every organization in the lattice corresponds exactly to one fixed point of the ordinary differential equation system description the underlying chemical reaction network. The lattice of organizations can therefore be viewed as a compact delineation of the dynamic state space of the system.

The topology of an reaction network can be meaningfully described by a generalized closure function that determines which molecular species can be constructed from a given set of molecular species[25]. Self-maintenance and (algebraic) closure can be expressed in terms of this closure function[26], thereby defining a notions of neighborhood, interior, closure and boundary on chemical reaction networks. These closure spaces are a strict generalization of the "usual" topological spaces for which a rich mathematical theory has accumulated over the last 100 years. For example, it generates natural definitions of connectedness, separation, regularity, compactness, and the like. At the very least the topological "language" is useful to remove ambiguities in the definitions of the most basic notions underlying the analysis of chemical space. Whether it can provide deeper insights into the structure of chemical reaction networks is a topic of active research.

Acknowledgments

This work has been funded by the German DFG Bioinformatics Initiative.

References

[1] Chemical Abstracts Service. CAS Registry; 2006. The CAS Registry provides the largest substance identification system in existence. Available from: http://www.cas.org/cgi-bin/regreport.pl.

[2] Fontana W, Buss LW. "The Arrival of the Fittest":Towards a Theory of Biological Organisation. Bull Math Biol. 1994;56(1):1–64.

[3] Fontana W, Buss LW. What would be conserved if 'the tape were played twice'? Proc Natl Acad Sci USA. 1994;91:757–761.

[4] Benkö G, Flamm C, Stadler PF. A Graph-Based Toy Model of Chemistry. J Chem Inf Comput Sci. 2003;43:1085–1093.

[5] Dittrich P, Ziegler J, Banzhaf W. Artificial Chemistries — A Review. Artificial Life. 2001;7:225–275.

[6] Todd MH. Computer-aided organic synthesis. Chem Soc Rev. 2005;34:247–266.

[7] Benkö G, Flamm C, Stadler PF. Explicit Collision Simulation of Chemical Reactions in a Graph Based Artificial Chemistry. In: Proc. ECAL05, LNAI 3630; 2005. p. 725–733.

[8] Fell DA. Understanding the Control of Metabolism. London: Portland Press; 1997.

[9] Yung YL, DeMore WB. Photochemistry of Planetary Athmospheres. New York: Oxford University Press; 1999.

[10] Höllering R, Gasteiger J, Steinhauer L, Schulz K, Herwig A. The Simulation of Organic Reactions: From the Degradation of Chemicals to Combinatorial Synthesis. J Chem Inf Comput Sci. 2000;40:482–494.

[11] Hillewaert LP, Dierickx JL, Froment GF. Computer generation of reaction schemes and rate equations for thermal cracking. AIChE J. 1988;34(1):17–24.

[12] Benkö G, Flamm C, Stadler PF. Generic Properties of Chemical Networks: Artificial Chemistry Based on Graph Rewriting. In: Banzhaf W, Christaller T, Dittrich P, Kim JT, Ziegler J, editors. Advances in Artificial Life. vol. 2801 of Lecture Notes in Computer Science. Heidelberg, Germany: Springer-Verlag; 2003. p. 10–20. Proceedings of the 7th European Conference of Artifical Life, ECAL 2003, Dortmund, Germany, September 14-17, 2003, Proceedings.

[13] Zeigarnik AV. On Hypercycles and Hypercircuits in Hypergraphs. In: Hansen P, Fowler PW, Zheng M, editors. Discrete Mathematical Chemistry. vol. 51 of DIMACS series in discrete mathematics and theoretical computer science. Providence, RI: American Mathematical Society; 2000. p. 377–383.

[14] Forst CV, Flamm C, Hofacker IL, Stadler PF. Algebraic Comparison of Metabolic Networks, Phylogenetic Inference, and Metabolic Innovation. BMC Bioinformatics. 2006;7:67.

[15] Kacser H, Burns JA. The control of flux. Symp Soc Exp Biol. 1973;32:65–104.

[16] Heinrich R, Rapoport TA. A linear steady-state treatment of enzymatic chains: general properties, control and effector strength. Eur J Biochem. 1974;42:89–95.
[17] Kacser H, Burns JA, Fell DA. The control of flux: 21 years on. Biochem Soc Trans. 1995;23:341–366.
[18] Clark BL. Stoichiometric network analysis. Cell Biophys. 1988;12:237–253.
[19] Reder C. Metabolic control theory: a structural approach. J Theor Biol. 1988;135:175–201.
[20] Schuster S, C H, Woods JH, Fell DA. Reaction routes in biochemocal reaction systems: Algebraic properties, validated calculation procedure and examples from nucleotide metabolism. J Math Biol. 2002;45:153–181.
[21] Urbanczik R, Wagner C. An improved algorithm for stoichiometric network analysis: theory and applications. Bioinformatics. 2005;21(7):1203–1210.
[22] Ibarra R, Edwards J, Palsson B. *Escherichia coli* k-12 undergoes adaptive evolution to achieve *in silico* predicted optimal groth. Nature. 2002;420:186–189.
[23] Stelling J, Klampt S, Bettenbrock K, Schuster S, Gilles E. Metabolic network structure determines key aspects of functionality and regulation. Nature. 2002;420:190–193.
[24] Dittrich P, Speroni di Fenizio P. Chemical organization theory: towards a theory of constructive dynamical systems. WWW. 2005;ArXiv.org q-bio.MN/0501016.
[25] Stadler BMR, Stadler PF. Generalized Topological Spaces in Evolutionary Theory and Combinatorial Chemistry. J Chem Inf Comput Sci. 2002;42:577–585. Proceedings MCC 2001, Dubrovnik.
[26] Benkö G, Centler F, Flamm C, Dittrich P, Stadler BMR, Stadler PF. A Topological Approach to Chemical Organizations; 2006. In preparation.