Multi-Phase Artificial Chemistry

Gil Benkö[1], Christoph Flamm[2], Peter F. Stadler[1,2]
[1]Lehrstuhl für Bioinformatik, Institut für Informatik, Universität Leipzig
Kreuzstrasse 7b, D-04103 Leipzig, Germany
[2]Institut für Theoretische Chemie und Molekulare Strukturbiologie, Universität Wien
Währingerstrasse 17, A-1090 Wien, Austria
gil@bioinf.uni-leipzig.de

Abstract

Artificial chemistries can be used to explore the generic properties of chemical reaction networks. In order to simulate for instance scenarios of prebiotic evolution the model must be close enough to real chemistry to allow at least semi-quantitative comparisons. One example is a previously described Toy Model that represents molecules as graphs, thereby neglecting 3D space, and employs a highly simplified version of the Extended Hückel Theory (EHT) to compute molecular properties. Here we show how the Toy Model can be extended to multiple phases by connecting the EHT calculations with chemical thermodynamics.

1 Introduction

The "Metabolism-First Theory" [15] for the origin of life proposes that the first step of the evolution of life was a "metabolism", i.e., an autocatalytic cycle of small organic molecules. Starting with a first cycle, usually assumed to be (a variant of) the citric acid cycle [15, 26], additional cycles, coupled to or feeding off the first cycle might have formed. The environmental parameters, such as the ocean-atmosphere interface or the composition of the atmosphere on the early earth plays an important role in this scenario. Computer simulations are necessary to explore the plausibility of such models since the time scales of prebiotic evolution are too long to allow for comprehensive experiments.

Artificial chemistries were used repeatedly to explore the self-organization of ensembles of self-replicating species [3, 8]. However, those models are often on a high level of abstraction and far removed from chemistry [7, 23]. The *Toy Model* [4, 5], on the other hand, is an Artificial Chemistry that has been designed deliberately to retain a close resemblance to organic chemistry while at the same time being computationally feasible in large-scale simulation. Furthermore it is independent of chemical databases. Its main features are that: (1) The molecules are determined by their graphs (i.e. the "topology" in the language of theoretical chemistry), (2) Extended Hückel Theory molecular orbital calculations determine all electronic properties in terms of electron densities in each orbital at each individual atom, and (3) binding energies and chemical reactivities are derived from these electron densities.

There are indications that multiple phases and phase boundaries were crucial for the origin of life [22]. As a first step towards simulating prebiotic chemistries in complex environments we describe here how the model can be extended to include the distribution of material between phases. The simplest case consists of two non-mixing solvents. Hence we have to model the solubilities of all possible chemical species.

2 Solvation Simulation

Although a semiempirical calculation might be inadequate to quantitatively predict weak interactions such as solvation processes it is sufficient for our purposes: Recall that we are not attempting to re-create chemical reality in the computer but rather to define a self-consistent artificial chemical universe. In line with the other calculations in the Toy Model, a very simplified model for solvation is used. The starting point is the Kirkwood formula

$$\Delta G_{\text{solv}} = -\frac{\varepsilon - 1}{2\varepsilon + 1}\frac{\mu^2}{a^3}. \tag{1}$$

derived from Reaction Field Theory [6, 11]. Here a is the radius of the molecule and μ is its dipole moment and ε is the dielectric constant of the medium. Since the Toy Model only knows about molecular graphs but not about spatial coordinates we have to re-interpret size and dipole moments in terms of graph parameters.

The Wiener index [27] $W(G) = \frac{1}{2}\sum_{x,y} d(x,y)$ can be used to approximate the volume of the molecule. Here $d(x,y)$ is the graph-theoretical distance, i.e., the length of the shortest path in the graph connecting the vertices x and y. It has been observed that $W(G)$ correlates well with a^3 [18, 19].

In order to keep the dipole moment calculation as simple as possible, we want to derive it only from the molecule information the Toy Model already contains, i.e. the topology. The dipole moment of the collection X of points $x \in X$ with charges $q(x)$ such that $\sum_{x\in X} q(x) = 0$ is defined as

$$\vec{\mu} = \sum_{x\in X} q(x)\vec{x} \tag{2}$$

which of course depends on the coordinate system. The square $\mu^2 = \sum_{x,y\in X} q(x)q(y)\langle \vec{x}, \vec{y}\rangle$ on the other hand, is independent of the coordinate system. The magnitude of the dipole moment can be expressed exclusively in terms of charges and distances:

$$\mu^2 = -\frac{1}{2}\sum_{x,y\in X} d^2(x,y)q(x)q(y) \tag{3}$$

Unfortunately, equ.(3) yields non-negative values of μ^2 only if d^2 is of negative type, i.e., if d is a Euclidean distance matrix. Of course, not every metric is euclidean. The following example is due to [9]:

$$\mathbf{D} = \begin{pmatrix} 0 & 2 & 2 & 1.1 \\ 2 & 0 & 2 & 1.1 \\ 2 & 2 & 0 & 1.1 \\ 1.1 & 1.1 & 1.1 & 0 \end{pmatrix} \tag{4}$$

The first three points form an equilateral triangle. The fourth point, which by symmetry must be in or above the center of the triangle, must have a distance of at least $2/\sqrt{3} \approx 1.1547$. The distance matrix from equ.(4) and charges $q = (1,1,1,-3)$ yields $\mu^2 = -1.11$, which does not make physical sense.

The problem of defining a graph-theoretical dipole moment therefore requires the construction of a euclidean distance $\hat{d}$ on G that is as closely related to the graphical distance d as possible. To this end we added a correction for the distance of graph paths of length 2. In order to account for sp^3, sp^2, and sp geometries, their distances where set to 1.633, 1.732, and 2, respectively. In accordance to the contraction of longer paths, i.e. carbon chains, their distance was multiplied with a factor of 0.9. Now the distance matrix $\mathbf{D}$ is diagonalized to Δ, such that $\mathbf{D} = \mathbf{Q}\Delta\mathbf{Q}^T$. By removing the negative eigenvalues, we obtain $\hat{\Delta}$ and a euclidean distance matrix $\hat{\mathbf{D}} = \mathbf{Q}\hat{\Delta}\mathbf{Q}^T$.

Another possibility is generating coordinates from the spectrum of the graph [12], but this approach fails for trees [10] and is thus not suitable for molecules in general.

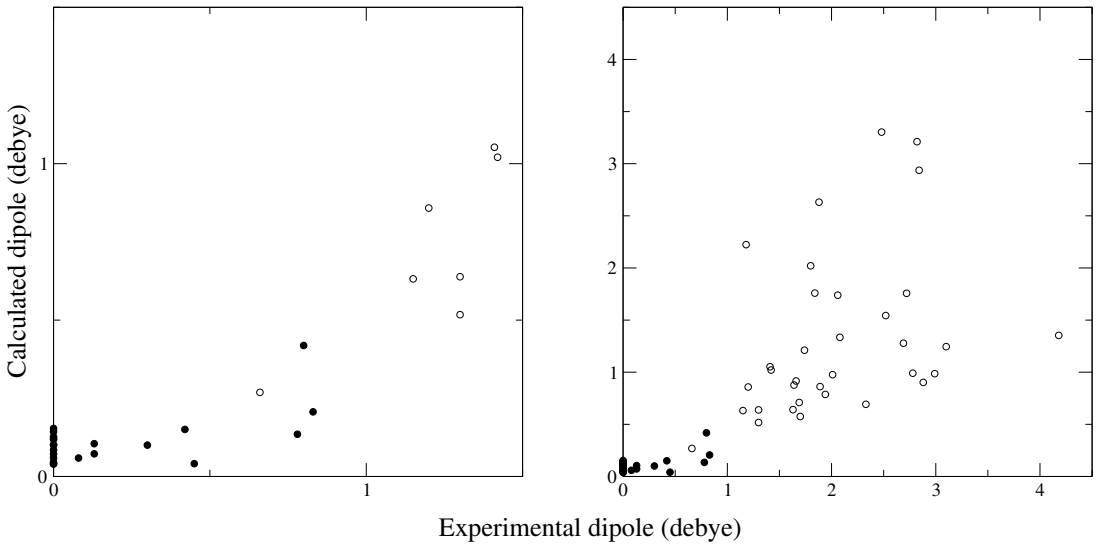

Figure 1: Calculated vs. experimental dipole moments for hydrocarbons (•) and oxygen-containing hydrocarbons (○, left: detail, right: overview).

This method was applied to the calculation of the dipole moments of 25 hydrocarbons and 35 oxygen-containing hydrocarbons from [16]. The results are compared to the experimental values in Fig. 1. Our method often fails to predict apolar molecules, i.e. zero dipole moments. The comparison of our estimates with experimental values shows a reasonable correlation.

3 Simulated Networks with two Phases

Now we can explore the effect of having multiple phases of different permittivities on CRNs. Here we simulate a chemistry that is inspired by prebiotic chemistry, but might also exist in the reducing atmosphere of Saturn's moon Titan. During experimental simulations of the Titan atmosphere, compounds such as cyanide and ammonia, and cyanide polymer chemistry were observed [24]. This analogy to early earth might provide insights into prebiotic chemistry.

Fig. 2 shows the two CRNs that differ only by being respectively mono- and biphasic. In the biphasic case, the concentrations of chemical species A in the two phases satisfies

$$\frac{[A]_I}{[A]_{II}} = \kappa_A = e^{-\frac{\Delta G_{\text{solv},I}(A) - \Delta G_{\text{solv},II}(A)}{RT}} \tag{5}$$

where the energies of solvation are computed from equ.(1) using the permittivities ε_I and ε_{II} as external parameters. For simplicity we model a species as present in phase I if $\Delta\Delta G_{\text{solv}}$ is below a threshold value Γ, and is present in phase II otherwise. We use here a non-zero value of $\Gamma = 1.4$kcal/mol to simulate a difference between "hydrophilic" and "hydrophobic" solvents.

As shown in Fig. 2, we obtain a biphasic CRN that is considerably less complex than the monophasic CRN. The phase boundary separates many possible reaction partners. We also observe in the biphasic CRN that some reaction products from educts in phase I go to phase II and vice versa. The CRN 'grows' across the phases. For different thresholds, the CRN might grow for a while solely in the phase I, and then continue completely in the phase II. The reaction products are small at first and thus more soluble in phase I, but the molecules occurring later in the reaction process tend to be bigger and thus to be in phase II. A variation of the threshold thus leads to a different repartition of reaction partners between the phases and finally to entirely different CRNs.

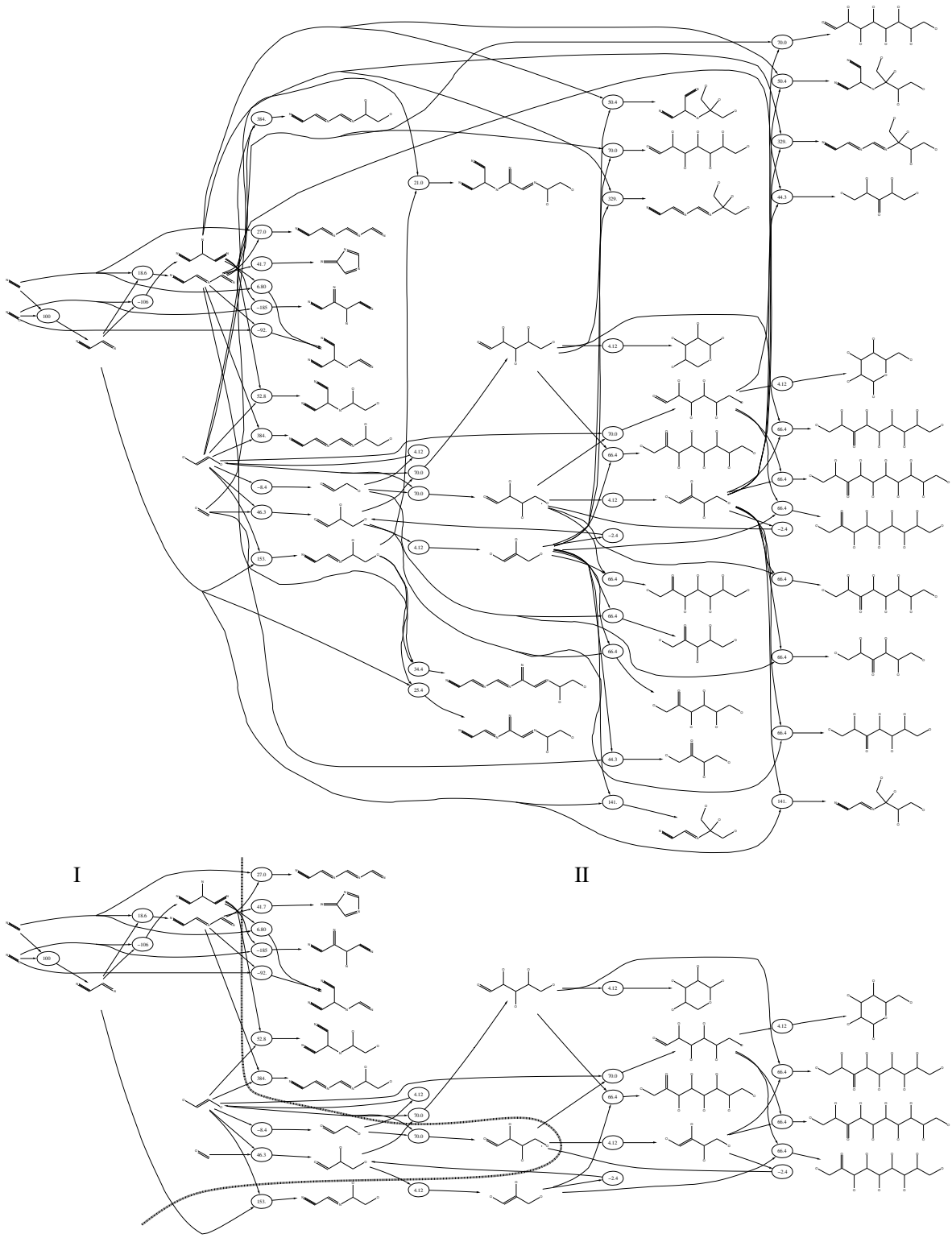

Figure 2: Simulation of the same CRN with one (*top*) and two (*bottom*) phases. Relative permittivities are $\varepsilon_1 = 1.5$ (lower phase, I) and $\varepsilon_2 = 2.5$ (upper phase, II), $\Gamma = 1.4$kcal/mol. In both simulations with start with cyanide, formaldehyde and glycol, and let them react by aldol addition, nucleophile addition, and tautomerization. The thick line in the biphasic case separates the two phases.

4 Discussion

Phase equilibria and the segregation of molecules into different phases probably were of utmost importance in prebiotic evolution. The presence of different phases allows certain species to concentrate to a level necessary for reaction, of which the products can subsequently switch phases to encounter reaction partners which would have before rendered impossible its synthesis. Concentration gradients, in particular pH gradients, between different phases might also have provided the driving force for chemical reactions. Hydrogen cyanide, for example, has been proposed to have polymerized first in an ice phase before its polymerization products continued to react further under warmer conditions [14]. This would avoid the problem posed by its tendency to react with formaldehyde [2]. Allamandola *et al.* [1] and Trinks *et al.* [25] have also proposed models including different ice mantels in comets or ice phases in the sea.

The most accepted model for the origin of life has been proposed by [13, 17] with the primordial 'soup'. Hot deep sea vents as the birth place of life (the primordial 'pizza') were discussed as in interesting alternative [26]. However, all these models need a prebiotic chemistry with complicated syntheses. As described in [20] and [21], they cannot occur in single, unpartitioned environment. A sequence of different environments would be important for the origin of life, just as it is in a traditional organic synthesis. During such a synthesis a reaction mixture is subjected to certain conditions, then some products are extracted, purified, and/or crystallized, new reagents might be added, and the next step with new conditions begins. We can imagine an analogous situation in prebiotic chemistry, where the different conditions and steps are mimicked by different environments, i.e. phases like hot vents, the atmosphere, and ice, and intermittent evaporation, phase change, crystallization or filtration. This might mitigate the problems of complicated synthesis in prebiotic chemistry.

Acknowledgments

This work was performed under the auspices of the COST D27 program "Prebiotic Chemistry and Early Evolution". It was supported in part by the DFG Bioinformatics Initiative BIZ-6/1-2 and the Graduiertenkolleg "Wissensrepräsentation" at the University of Leipzig.

References

[1] L. Allamandola and D. Hudgins. From interstellar polycyclic aromatic hydrocarbons and ice to astrobiology. In V. Pirronello and J. Krelowski, editors, *Solid State Astrochemistry*. NATO ASI, Kluwer, 2003.

[2] T. Arrhenius, G. Arrhenius, and W. Paplawsky. Archean geochemistry of formaldehyde and cyanide and the oligomerization of cyanohydrin. *Orig. Life Evol. Biosph.*, 24:1–17, 1994.

[3] R. J. Bagley and J. D. Farmer. Spontaneous emergence of a metabolism. In C. G. Langton, C. Taylor, J. D. Farmer, and S. Rasmussen, editors, *Artificial Life II*, Santa Fe Institute Studies in the Sciences of Complexity, pages 93–141, Redwood City, CA, 1992. Addison-Wesley.

[4] G. Benkö, C. Flamm, and P. F. Stadler. Generic properties of chemical networks: Artificial chemistry based on graph rewriting. In W. Banzhaf, T. Christaller, P. Dittrich, J. T. Kim, and J. Ziegler, editors, *Advances in Artificial Life - Proceedings of the 7th European Conference on Artificial Life (ECAL)*, pages 10–20. Springer, 2003.

[5] G. Benkö, C. Flamm, and P. F. Stadler. A graph-based toy model of chemistry. *J. Chem. Inf. Comput. Sci.*, 43:1085–1093, 2003.

[6] C. J. Cramer and D. G. Truhlar. Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem. Rev.*, 99:2161–2200, 1999.

[7] P. Dittrich, J. Ziegler, and W. Banzhaf. Artificial chemistries — a review. *Artificial Life*, 7:225–275, 2001.

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

[9] J. C. Gower and P. Legendre. Metric and euclidean properties of dissimilarity coefficients. *J. Classification*, 5:5–48, 1986.

[10] D. Harel and Y. Koren. Graph drawing by high-dimensional embedding. In M. Goodrich and S. Kobourov, editors, *Graph Drawing: 10th International Symposium, GD 2002*, volume 2528, pages 207 – 219. Springer, 2002.

[11] J. G. Kirkwood. Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *J. Chem. Phys.*, 2:351–361, 1934.

[12] D. Manolopoulos and P. Fowler. Molecular graphs, point groups and fullerenes. *J. Chem. Phys.*, 96:7603–7614, 1992.

[13] S. L. Miller. A production of amino acids under possible primitive earth conditions. *Science*, 117:528–529, 1953.

[14] S. Miyakawa, H. J. Cleaves, and S. L. Miller. The cold origin of life: A. Implications based on the hydrolytic stabilities of hydrogen cyanide and formamide. *Orig. Life Evol. Biosph.*, 32:195–208, 2002.

[15] H. J. Morowitz, J. D. Kostelnik, J. Yang, and G. D. Cody. The origin of intermediary metabolism. *PNAS*, 97:7704–7708, 2000.

[16] Computational chemistry comparison and benchmark database, release 6a, May 2002. `http://srdata.nist.gov/cccbdb/`.

[17] A. I. Oparin. *Proiskhozhdenie Zhizny [The Origin of Life]*. Izd. Moskoviskiy Rabochiy, 1924.

[18] J. R. Platt. Prediction of isomeric differences in paraffin properties. *J. Phys. Chem.*, 56:328–336, 1952.

[19] N. Raos. Mean molecular radius and the Wiener number: A quest for meaning. *Croat. Chem. Acta*, 76:81–85, 2003.

[20] D. Segrè and D. Lancet. Composing life. *EMBO Reports*, 1:217–222, 2000.

[21] R. Shapiro. The improbability of prebiotic nucleic acid synthesis. *Orig. Life Evol. Biosph.*, 14:565–570, 1984.

[22] S. J. Sowerby, C. A. Cohn, W. M. Heckl, and N. G. Holm. Differential adsorption of nucleic acid bases: Relevance to the origin of life. *PNAS*, 98:820–822, 2001.

[23] H. Suzuki, N. Ono, and K. Yuta. Several necessary conditions for the evolution of complex forms of life in an artificial environment. *Artificial Life*, 9:537–558, 2003.

[24] B. N. Tran, J. C. Joseph, J. P. Ferris, P. D. Persans, and J. J. Chera. Simulation of titan haze formation using a photochemical flow reactor. The optical constants of the polymer. *Icarus*, 165:379–390, 2003.

[25] H. Trinks, W. Schröder, and C. K. Biebricher. *Eis und die Entstehung des Lebens [Ice and the origin of life]*. Shaker, 2003.

[26] G. Wächtershäuser. Before enzymes and templates: theory of surface metabilism. *Microbiol. Rev.*, 52:452–484., 1988.

[27] H. Wiener. Structure determination of paraffine boiling points. *J. Am. Chem. Soc.*, 69:17–20, 1947.