Article

Navigating the Chemical Space of HCN Polymerization and Hydrolysis: Guiding Graph Grammars by Mass Spectrometry Data

Jakob L. Andersen [1,2]**, Tommy Andersen** [3]**, Christoph Flamm** [4]**, Martin M. Hanczyc** [3]**, Daniel Merkle** [1,]*** and Peter F. Stadler** [2,4,5,6,7,8]

[1] Department of Mathematics and Computer Science, University of Southern Denmark, Odense M DK-5230, Denmark; E-Mail: jlandersen@imada.sdu.dk

[2] Max Planck Institute for Mathematics in the Sciences, Leipzig D-04103, Germany

[3] Institute of Physics, Chemistry and Pharmacy, University of Southern Denmark, Odense M DK-5230, Denmark; E-Mails: tomma06@student.sdu.dk (T.A.); martin@sdu.dk (M.M.H.)

[4] Institute for Theoretical Chemistry, University of Vienna, Wien A-1090, Austria; E-Mails: xtof@tbi.univie.ac.at (C.F.); studla@bioinf.uni-leipzig.de (P.F.S.)

[5] Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Leipzig D-04107, Germany

[6] Fraunhofer Institute for Cell Therapy and Immunology, Leipzig, D-04103, Germany

[7] Center for non-coding RNA in Technology and Health, University of Copenhagen, Frederiksberg C DK-1870, Denmark

[8] Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe NM 87501, USA

* Author to whom correspondence should be addressed; E-Mail: daniel@imada.sdu.dk; Tel.: +45-6550-2322; Fax: +45-6550-2373.

Abstract: Polymers of hydrogen cyanide and their hydrolysis products constitute a plausible, but still poorly understood proposal for early prebiotic chemistry on Earth. HCN polymers are generated by the interplay of more than a dozen distinctive reaction mechanisms and form a highly complex mixture. Here we use a computational model based on graph grammars as a means of exploring the chemical spaces of HCN polymerization and hydrolysis. A fundamental issue is to understand the combinatorial explosion inherent in large, complex chemical systems. We demonstrate that experimental data, here obtained by mass spectrometry, and computationally predicted free energies together can be used

to guide the exploration of the chemical space and makes it feasible to investigate likely pathways and chemical motifs even in potentially open-ended chemical systems.

Keywords: hydrogen cyanide; graph grammars; chemical space; mass spectrometry; chemical motif; polymerization; autocatalysis

1. Introduction

Hydrogen cyanide (HCN) has been recognized as a key molecule in abiogenesis already in the earliest studies of chemical evolution in the 1920s [1–3]. According to this line of reasoning, the organic complexity of Life on Earth arose abiotically through chemical reactions among simple precursor molecules and their products. In a landmark proof of principle experiment Urey and Miller demonstrated in 1956, chemical evolution could have been initialized in a simulated atmosphere of early Earth [1,4,5]. HCN has been highlighted as a prebiotic precursor of particular interest not only in conceptual considerations [6] but also experimentally. HCN has been used to synthesize adenine by Oró in 1961 [7,8], amino acids [9], as well as many other molecules relevant to present-day biology [9–15]. Most recently HCN has also been shown to play a key role in sugar synthesis [16]. HCN and its derivatives and polymers have been detected in the current interstellar medium as chemical components in some extraterrestrial environments, most notably on planetoids, and moons, where it accounts for dark colored regions [17,18], and on Saturn's moon Titan [19]. Tholins, aerosols that are obtained in simulations of the atmosphere of the Saturnian moon Titan, are spectrometrically similar to HCN polymers [20–22].

The HCN monomer is highly reactive and can self-polymerize under certain conditions leading to large dark insoluble precipitates [15,23]. The structure of the resulting complex polymer is not determined, albeit several models for its structure have been proposed [24–29]. The large complex polymers consist mostly of H, C and N according to elemental analysis [27]. The introduction of oxygen atoms occurs largely through the subsequent hydrolysis of the polymer leading to the production of biologically relevant molecules in minor quantities. Early experiments [30] showed that the synthesis of amino acids results from the hydrolytic breakdown of HCN heteropolypeptides. Hydrolysis of HCN polymers gives rise to self-assembling chemical structures that show proto-cell like dynamical behavior [31]. Recently, a high resolution MS analysis of HCN polymers was presented in [32], however without an underlying mathematical model of polymerization.

It is appealing to explore the generative chemistry of HCN monomers when polymerized and then reacted with water as a feasible model of complex chemical evolution. The vast amount of chemical products produced renders a detailed analytical characterization infeasible. Thus computational approaches provide tools to locate interesting regions in this combinatorial complex chemical transformation space and help to characterize them.

We describe here a computational generative chemistry approach, and describe how this combinatorial description of the relevant chemical space can be interfaced closely with empirical data. In Section 2 we summarize mass spectrometry data for the polymerization of HCN and the subsequent hydrolysis of the

resulting black HCN polymer. Our theoretical framework for systematic exploration of chemical spaces is outlined in Section 3. Results that integrate graph grammar approaches and the experimental results are given in Section 4, highlighting some of the complex chemical networks. We conclude in Section 5.

2. Experimental Part—Polymerization, Hydrolysis, and Mass Spectrometry

2.1. Acid-Catalyzed HCN Polymerization

120 mL of milli-Q water was added to a 250 mL round bottom flask. Oxygen and other gases were removed with a flow of argon over night. 10 g of sodium cyanide were dissolved in the milli-Q water under stirring while the gas was changed from argon by flushing the bottle with nitrogen. After the sodium cyanide had dissolved, the system was closed and heated to 60 °C in an oil bath. 10 mL 1 M HCl was slowly added while the system was at 60 °C, for 1 h. The system was then closed and left at 60 °C for 24 h with continuous stirring. Gas was continuously flowed through the system. The gas exiting the reaction flask was flowed into a safety flask containing 200 mL of saturated $NaOH$ solution with 1 M $FeCl_3$ to capture any cyanide gas that might be present. During the 24 h the color of the reaction solution changed from clear to yellow or orange with suspended black particles. After the 24 h the products were harvested and transferred to four 50 mL plastic tubes and centrifuged for 30 min at 3000 g. The supernatant was removed and the black particles dried in a vacuum over night. Note that the supernatant can be transferred back into the system and the synthesis can be repeated.

2.2. Hydrolysis of the HCN Polymer

HCN polymer was synthesized and hydrolyzed under varying but systematic conditions allowing for a wider exploration of the chemical space that will then be modeled.

Long term hydrolysis with different condition: Nine samples were prepared, each containing 0.1 g dried hydrogen cyanide polymer carbon (provided by Robert Minard from Penn State University). Each sample was subjected to different hydrolysis conditions. The first five samples were incubated at pH-values 1, 4, 7, 10, and 12, respectively. The next three samples were incubated at pH 12 in the presence of 1 mM of a metal-salt: aluminium(III) chloride, magnesium(II) chloride, and tin(II) chloride. All samples were prepared in volumes of 7.5 mL and stored in 15 mL closed tubes. The samples were stored at room temperature for 357 d. The long term samples are denoted as L-x-y, where $x \in \{1, 4, 7, 10, 12\}$ denotes the pH-value and $y \in \{Al, Mg, Sn\}$ refers to a metal salt if used.

Short term hydrolysis with different condition: Five additional samples were prepared from 0.2 g dried hydrogen cyanide polymer, as described above. As a control, one of the samples was suspended in milli-Q water without any metal-salts. Three of the samples were suspended in 1 mM metal-salt solutions ($FeCl_2$, $CaCl_2$, KCl). The fifth sample was suspended in a solution containing 1 % of hydrolysis product from sample L-12. All samples are prepared in volumes of 15 mL and stored in 15 mL closed tubes and have a pH-value of 12. The samples were left at room temperature for 114 d. The samples are denoted as S-Fe, S-Ca, S-K depending on the salt, S-control for the control sample, and S-12 for the fifth sample.

All samples were centrifuged, the supernatant was removed and filtered using a 0.2 μm filter. After filtration the samples were analyzed by liquid chromatography and mass spectrometry (LC-MS).

2.3. Liquid Chromatography and Mass Spectrometry

The supernatants of the HCN polymer hydrolysis samples were analyzed using a TSQ Vantage Triple Stage Quadrupole Mass Spectrometer with an EASY-nLC 1000 Liquid Chromatography (LC) unit. Solvent A was milli-Q water with 1% formic acid and Solvent B was acetone with 1% formic acid with the following gradient over time: 0, 0, 20, 100, 0% for the time series 0, 2, 20, 25, 30 min. Although differed reaction times and conditions were tried, the same total amount of sample was injected to make the samples comparable. The MS has a resolution of 0.7 atomic units and therefore isotopic peaks with similar mass but different sum formula are not resolved. We used a small percentage of formic acid in the LC because formic acid is volatile and imparts a positive charge on some functional groups. The samples were run in positive scan mode only. We used unfragmented full MS scans for this study. Blank runs (solvent only) were commonly conducted to eliminate the possibility of sample carryover. We have confidence in the peaks observed in the optimum m/z range of 100–400. We nevertheless show all the data as some major peaks are found below $100\,m/z$. The mass spectrometer used eletrospray ionization, *i.e.*, protons are added to charge the molecules. As an approximation we assume a single proton is added, and measured m/z ratios thus correspond approximately to molar mass $+1$. This approximation can of course easily be replaced by a more sophisticated prediction which takes the proton affinity of each molecule into account.

3. Theoretical Part

3.1. Graph Grammars

Graph grammars provide a convenient and efficient method to investigate large, diverse chemical spaces. In this framework, molecules are abstracted to simple edge and vertex labeled graphs while reactions are correspondingly expressed as graph rewrite rules between the educts and products of a chemical reaction [33]. Graph grammars can be thought of as an extension of context-free grammars and their associated term rewriting systems to context-sensitive grammars, where labeled graphs then replace strings as the basic objects [34]. This computational model fully captures the inherent algebraic structure of chemistry, *i.e.*, molecules may react with each other to yield novel molecules.

Multiple approaches to specifying graph transformation rules have been explored, see e.g., [35] for a detailed technical presentation. We use the so-called Double Pushout (DPO) approach, in which a transformation rule p is specified as $p = (L \xleftarrow{l} K \xrightarrow{r} R)$. The graphs L, R, and K are called the left graph, right graph, and context graph, respectively. The maps l and r are graph morphisms. If a rule p is applicable to a graph G with the matching morphism, we can derive a new graph (or molecule) H. More informally, if the left graph L can be found in one or potentially many molecules, then the corresponding edges of the graph (*i.e.*, the chemical bonds of the molecule(s)) can be changed by removing all edges defined by L, and adding all the edges of R. Note, that such an operation can split and merge molecules. An example of a graph grammar rule application is shown in Figure 1.

Figure 1. Illustration of a chemical reaction using the Double Pushout approach. The chemical transformation of complete molecules (*i.e.*, the application of the graph grammar rule as defined in the first row) is represented as the graph derivation $G \Longrightarrow H$ in the second row. This shows an intermediate step in the synthesis of adenine as an HCN pentamer. Above, a possible rule $p = (L \xleftarrow{l} K \xrightarrow{r} R)$ underlying the concrete reaction is shown (*i.e.*, the graph grammar rule defining the chemical reaction). Bonds in L that are removed and bonds in R that are inserted are all colored red. The green vertices form the context K of the rule.
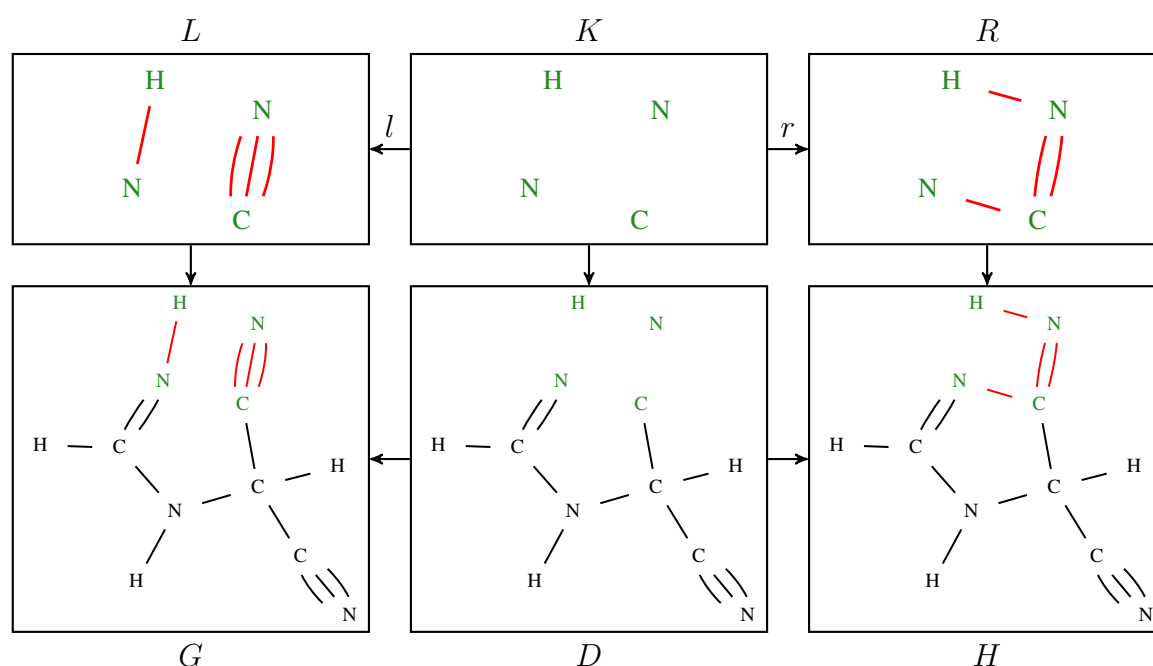
The representation of chemical reactions as graph rewriting rules is closely related to the Dugundji-Ugi theory [36,37], which is based on the concepts of bond and reaction matrices. This approach has proved useful in particular in the analysis of metabolic reactions [38,39]. The matrix-based approach, however, lacks the coherent formal framework that is available for the realm of graph grammars.

3.2. Exploration of the Chemical Space

The Graph Grammar framework utilizes algorithms for well-known computational problems, including the NP-complete subgraph isomorphism problem. Thus, in the general case, no polynomial-time algorithms are known for solving the problem. However, since molecular graphs have limited degree and are relatively small, the framework is efficient enough in practice to explore chemical spaces. Conceptually, we apply all possible graph grammar rules to all possible (combinations of) molecules in an iterative fashion starting from a small number of initial graphs, *i.e.*, a seed set of molecules. In practice, however, even simple chemistries typically generate a plethora of different chemical compounds. This combinatorial explosion makes it necessary to strategically limit the exploration. To this end we have developed a generic framework of exploration strategies that operates

in conjunction with a graph transformation system. A detailed description can be found in [40]. The framework allows us to express pruning strategies for specific reactions and filtering rules for the resulting molecules. Here, we do not aim at a precise and full characterization of specific experimental runs, but we introduce a new approach to constrain the generative exploration steps in order to find pathways and compounds of high interest, while limiting the inherent combinatorial explosion. We make use of filters that link the exploration procedure to the MS data by giving preference to molecules whose mass matches high intensity peaks in the spectrum and discouraging those with low-abundance m/z values. Due to the low-resolution of our available data we cannot infer atomic composition from the MS data. In addition, we use energetic considerations to distinguish between compositional isomers. Thus, for this study we only use the basic data from the MS scans; the coupling of m/z and intensity values, with a basic model of electrospray ionization which assumes a single proton is added to each molecule.

For the calculation of preference of molecules we estimated equilibrium abundances from the Boltzmann factors $p_x = \exp(-G_x/RT)$, where the free energy values G_x of a molecule x were computed with Open Babel [41], employing a Merck Molecular Force Field. Since relative abundances are computed over families of isomers only, we treat families of molecules with different structure but same molecular formula independently. Although this is a rather crude approximation, more refined energy computations are at present still too expensive in terms of computational resources to be used in large-scale network exploration.

The chemical space of HCN polymerization is modeled by a set of 8 reaction rules that have been extracted from literature [7,13]. An additional 13 rules are added to implement hydrolysis, mostly additions to the $C \equiv N$ and $C = N$ bonds. A full description on the rules is given in the Web Supplement.

3.3. Finding Chemical Motifs

A chemical motif is a collection of reactions which form an interesting pattern in the chemistry. E.g., the reactions connect given input molecules in order to create a specific product, constitute a catalytic cycle, or are collectively autocatalytic. Given a reaction network generated in the space exploration step, chemical motifs are identified using an integer linear programming (ILP) approach. The framework of ILP allows us to model both general pathways such as "the conversion of HCN to adenine" as well as pathways with specific structural properties such as "the presence of autocatalytic cycles" without specifying particular molecules.

The ILP formulations generically model each reaction by a non-negative integer variable that encodes the multiplicity of the reaction. Constraints are introduced to ensure balance of mass. Specific motifs are further encoded as additional linear constraints. The detection of optimal or near-optimal chemical motif is then equivalent to finding an optimal assignment of (integer) variables. We used the software package IBM ILOG CPLEX Optimizer v12 in order to solve the underlying optimization problems. A detailed description of the modeling framework is forthcoming and will be published elsewhere.

4. Results and Discussion

The primary goal of this contribution is to demonstrate the possibility of coupling systematic exploration of complex chemical spaces with empirical data, despite the combinatorial explosions

associated with such an endeavor. Even in extreme cases, such as the HCN chemistry, graph grammars can be employed in combination with suitable exploration strategies. A particularly appealing feature is the flexibility in which the exploration strategies can be written to incorporate data, here from mass spectrometry experiments, in a natural way. This enables the identification of chemical motifs contained in the chemical spaces.

We begin our exposition with a discussion of the Mass Spectrometry (MS) data themselves before proceeding to the simulation results. In order to illustrate the potential of our systematic exploration approach, we first consider alternatives to a well-know pathway for adenine synthesis that are identified by the automatic exploration of the space of HCN polymerization and hydrolysis reactions. As a second example, we describe an autocatalytic loop that can explain experimental data for the hydrolysis of triazine. Again, the network motif was identified automatically.

4.1. Correlations in Mass Spectrometry Experiments

HCN polymer was synthesized and hydrolyzed under varying but systematic conditions allowing for a wider exploration of the chemical space that will then be modeled, see Experimental Part. We processed the 13 MS scans by a simple clustering methods and assigned each sample value an integer-valued m/z value. For each of the 13 scans we then computed values $v_i^{\max}$, $i \in \{10, \ldots, 1500\}$, where $v_i^{\max}$ is the highest intensity that was found at $m/z = i$ among the 1140 different retention times in the specific scan. The correlation matrix of these 13 vectors is shown in the left panel of Figure 2. In general, the individual experiments are highly correlated. The smallest Pearson correlation coefficient, $\rho = 0.21$, was observed between scans L-1 and L-7. Both the short-term and the long-term hydrolysis scans are highly correlated within each of the two groups, $\rho > 0.68$ and $\rho > 0.66$, resp. Tin or magnesium chloride as metal salt leads to the most divergent data sets. Based on the strong correlation of all the scans we decided to merge all scans and use the vector of the highest $v_i^{\max}$-value in *any* of the peaks for further analysis. The highest peaks observed in any of the scans are depicted in Figure 2 (right).

4.2. Exploring the Chemical Space

The common starting points for the exploration of the chemical space are HCN and NH_3. We expanded the chemical space by application of all graph grammar rules 15 times. Due to the combinatorial explosion, the underlying strategy framework restricts the exploration by not allowing recombination of two large molecules (*i.e.*, one of the two molecules in a merging reaction needs to have a molar mass that is smaller than 50). As a further filtering strategy we reject compounds based on their relative equilibrium abundance as estimated from their Boltzmann factors normalized to the interval 0 to 1, such that only high-probability compounds pass the filter. This results in a collection of 356 compounds that, together with H_2O, are taken as starting compounds to explore the chemical space for hydrolysis.
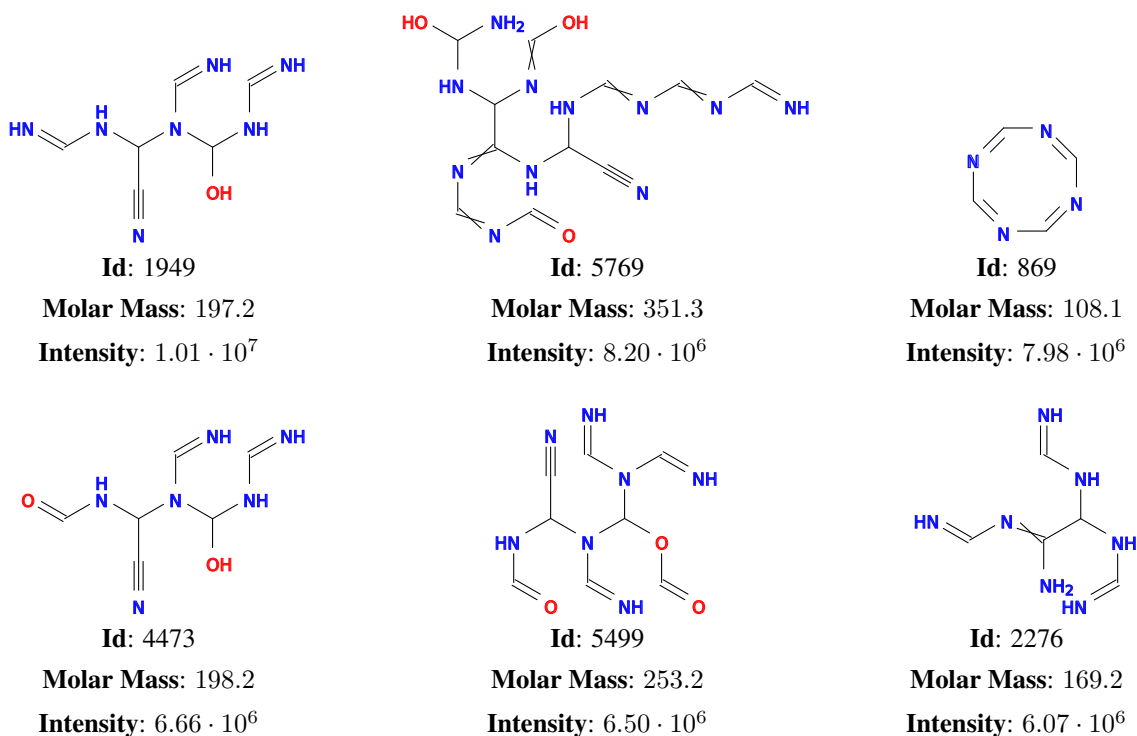
Figure 2. (**Left**): Correlation Matrix for long term and short hydrolysis scans under different reaction conditions; no negative Pearson correlation coefficient ρ was observed; colour of the circles and radii of the circle sizes are scaled linearly with ρ; (**Right**): maximal observed intensity for each m/z-value observed in any of the scans.
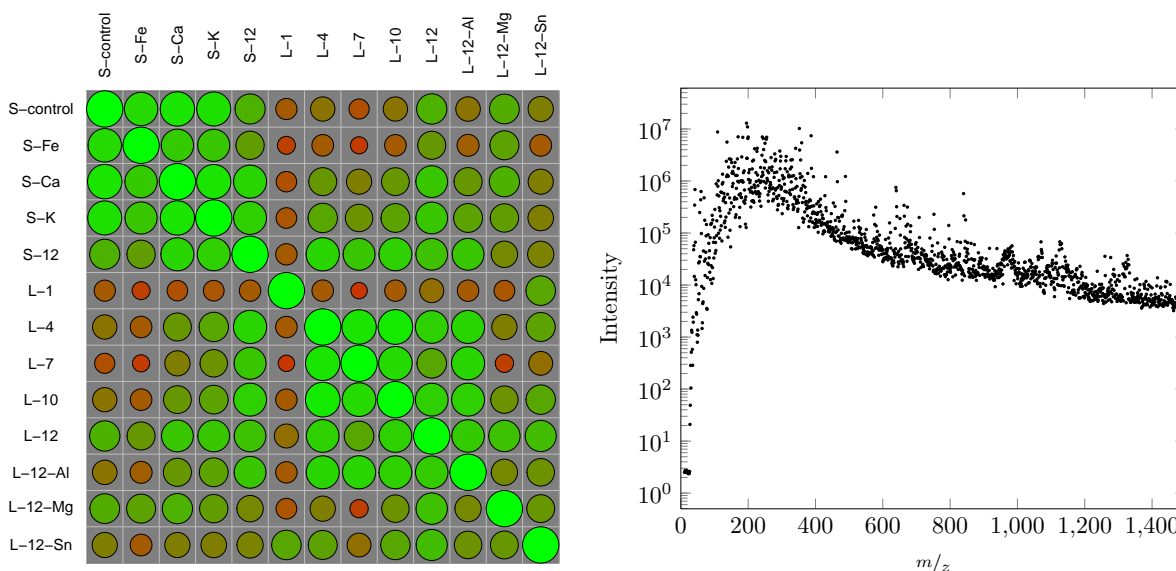
In this step, the expansion strategy is now mainly biased in response to the maximal observed intensities of the MS results. More precisely, only the compounds with the best support from the MS data are not discarded, while isomers are filtered based on their relative abundance, again using the normalized Boltzmann factors. As the intensity values are only available for integer m/z values, we used the molar mass given by Open Babel to calculate an estimated intensity as a weighted average of the intensities of the two nearest integer m/z values, assuming each molecule has a single charge. Due to the use of an electrospray ionization source, the m/z values was shifted by -1 to account for the added proton.

Interestingly we observe that the exploration method converges for the HCN chemistry in the sense that the constrained exploration reaches a state in which additional exploration steps do not further enlarge the chemical space. There is no theoretical guarantee for convergence in general. In our simulation, we encountered 6,472 compounds through 9,197 reactions during hydrolysis. The final set of molecules after filtering, however, consists of only 94 compounds. Six of these molecules with very strong support from the MS data (intensity at least $6.50 \cdot 10^6$) are shown in Figure 3. Within their isomer class, these compounds are those with the highest support based on their normalized Boltzmann factor values. The complete set of molecules, listed as SMILES strings, their MS support and their equilibrium abundances in the final network is listed in Appendix 5. All the graph grammar rules used can be found in the web supplement.

Figure 3. Six molecules with very strong support from the MS data that were found by chemical space exploration with graph grammars; within their isomer class, these compounds are those with the largest relative equilibrium abundance as predicted using the Open Babel energy calculation; all molecules are listed in Appendix 5.

Id: 1949
Molar Mass: 197.2
Intensity: $1.01 \cdot 10^7$

Id: 5769
Molar Mass: 351.3
Intensity: $8.20 \cdot 10^6$

Id: 869
Molar Mass: 108.1
Intensity: $7.98 \cdot 10^6$

Id: 4473
Molar Mass: 198.2
Intensity: $6.66 \cdot 10^6$

Id: 5499
Molar Mass: 253.2
Intensity: $6.50 \cdot 10^6$

Id: 2276
Molar Mass: 169.2
Intensity: $6.07 \cdot 10^6$

4.3. Adenine Pathway

The formation of adenine as a HCN pentamer is one of the best-studied aspects of HCN polymerization. Plausible pathways were suggested in the literature [7,8,42]. Figure 4 shows a pathway suggested by Oró [7,8], which uses nine reactions steps. To demonstrate the capabilities of our approach in finding multi-step pathways, we ask here whether there are alternative ways to obtain adenine via intermediates that contain oxygen, *i.e.*, that can be formed only after hydrolysis. To this end, we expanded the chemical space after adding H_2O by one exploration step with all rules. The resulting chemical space comprises 262 compounds and 1132 reactions. We then searched for chemical motifs that use HCN, NH_3, and H_2O as input compounds and produce adenine. The corresponding ILP formulation can be used with different optimization functions, e.g., minimizing the overall number of reactions or selecting a pathway with maximum support from the MS data. To search for pathways that have a strong support from the MS scans we used the sum of all the intensities of all compounds of a pathway. Since these numbers cannot be meaningfully compared between pathways with different numbers of steps, we can use it either as a cut-off value instead of an optimization criterion, or we can fix the number of chemical compounds to achieve comparability.

Figure 4. Hypergraph representation of the mechanistic route to adenine as proposed by Oró [7,8]. Under NH_3 catalysing conditions 5 HCN molecules polymerize into 1 molecule of adenine (Note that the consumption of 4 NH_3 molecules is counter balanced by 4 reactions producing NH_3). The pathway uses nine different chemical reactions. The primary route of the pathway is highlighted by thick edges. Each reaction is visualized as a rectangle with in-edges from the educts and out-edges to the products. Parallel edges denote the multiplicity of educts/products in reactions. Additionally, the flux of each reaction is shown, *i.e.*, the reaction $HCN + NH_4 \rightarrow CH_4N_2$ is used twice. See [40] for a formal definition of the encoding and visualization of chemical reaction networks as hypergraphs.
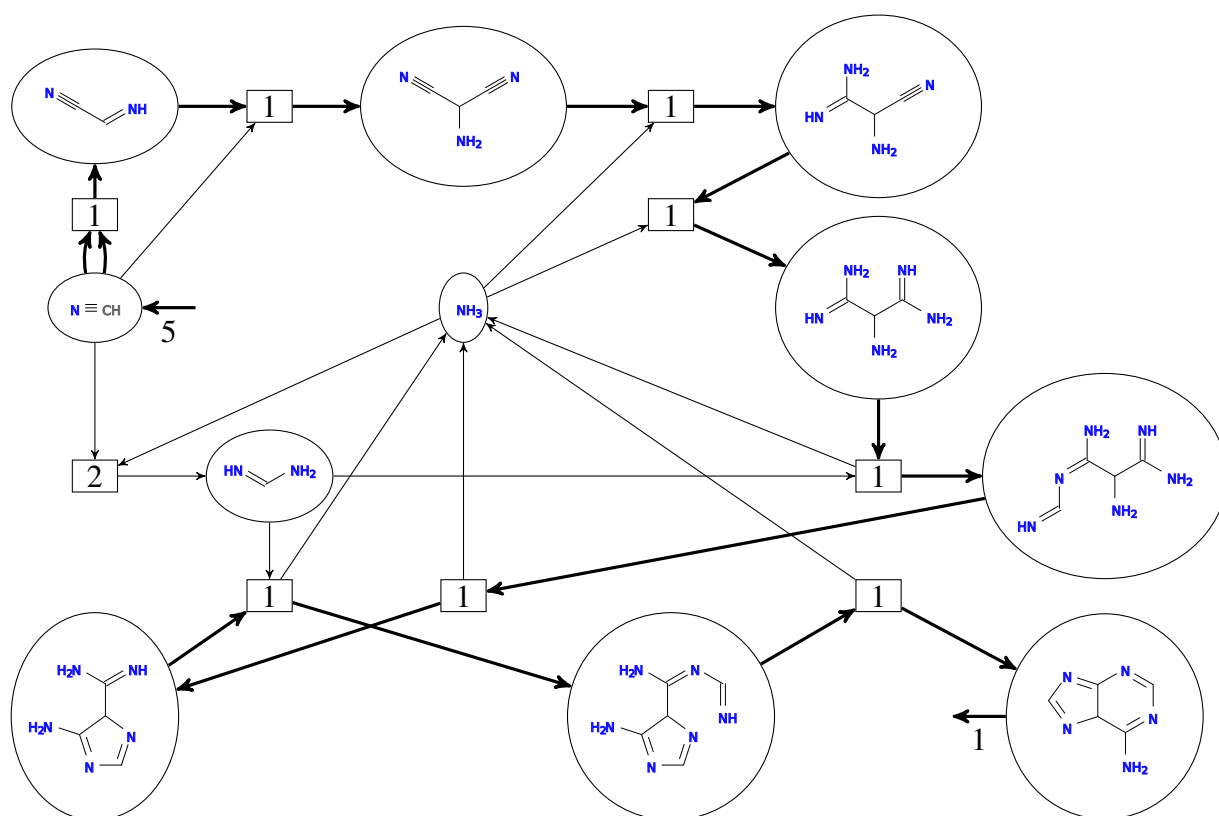

Figure 5 summarizes five alternatives to Oró's pathways as an illustration of the explorative potential of our method. The straight pathway in the middle of the schematic is one reaction step shorter than the original Oró pathway (which is shown at the bottom) but visits the same check-point intermediates *i.e.*, aminomalononitrile, and 4-Amino-5-imidazolecarboxamide (AICA). The pathways in the upper part use H_2O and accordingly passes through an oxygen-analog of AICA. Detailed descriptions of the five pathways can be found in the web supplement. Note, that we easily can find hundreds of alternative adenine pathways in the chemical space, including solutions that do not use the check-point intermediates. A solution biased towards utilizing hydrolyzed molecules are shown in Figure 6. This pathway uses 12 molecules with a total sum of intensities of 6.62×10^6. The the pathway suggested by Oró (Figure 4) uses 11 molecules with 4.14×10^6 as sum of intensities.

Figure 5. Schematic of the merge of five different pathways to create adenine. Following the lower-most path, utilizing formamidine, yields the pathway proposed by Oró which uses 9 reactions; the R_O group contains oxygen, while the R_N group does not. The horizontal pathway uses eight different chemical reactions. Detailed descriptions of the five pathways can be found in the web supplement.

Figure 6. A chemical pathway to produce adenine by using hydrolyzed molecules as intermediates. The sum of MS data intensities of all 12 molecules in the pathway is $6.62 \cdot 10^6$. Isomerization reactions are visualized as single arrows, without an rectangle. The arrow is annotated with the flux of the reaction. The primary route of the pathway is highlighted by thick edges.

4.4. Autocatalytic Loops

Triazine is a small molecule HCN trimer that is quickly and completely removed from the solution in a simple hydrolysis experiment. After 24 h the absorbance from the ring was no longer detectable by UV-VIS spectrometry (data not shown). This suggests that an autocatalytic network emerges that is fed by triazine or some of its decomposition products under the hydrolysis conditions. Therefore we queried the chemical space of HCN-hydrolysis for small autocatalytic sub networks with our computational methods (for a formal definition of autocatalysis, see [43]). We identified small sub-networks that are autocatalytic in formamide (shown in Figure 7). Interestingly, formamide is the most stable compound with molecular formula CHON. However, its role in prebiotic processes and the origin of life is under heavy debate (see review [44] and accompanying comments).

Figure 7. Putative autocatalytic loop in formamide, identified in the chemical space of HCN-polymer hydrolysis. The autocatalytic loop is fed by cyanide molecules stemming from triazine decomposition. Formamide is initially formed by the reaction shown in grey.
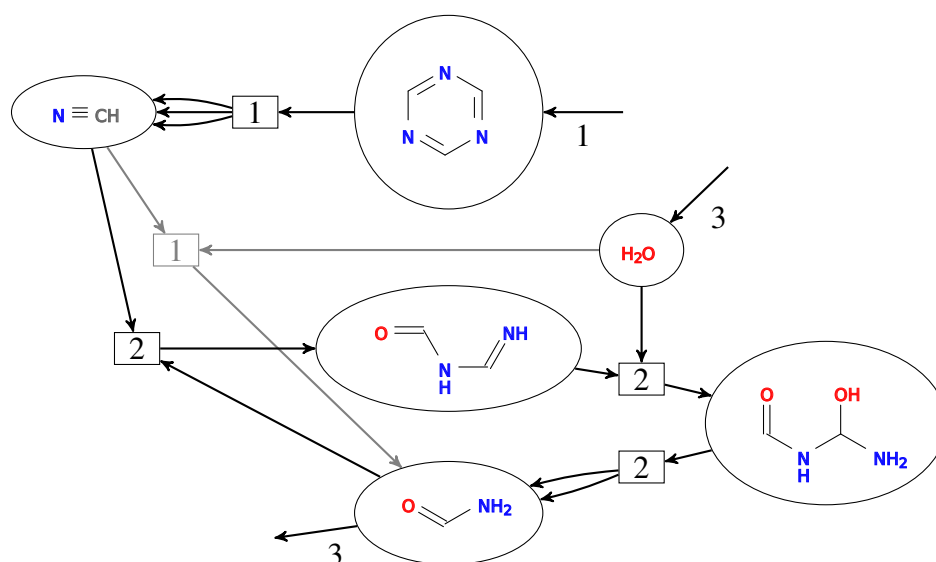
5. Conclusions

We have illustrated the first steps of a novel research program that combines computer science and mathematics with wetlab systems chemistry to dissect complex chemical systems at a hitherto unknown level of detail. Abstracting molecules to vertex and edge labeled graphs and chemical reactions to graph transformation rules enables us to investigate the "language of chemical graphs" that can be generated over a set of molecules and graph rewrite rules, in a rigorous manner. Simple iterative expansion of the language of chemical graphs results quickly in a combinatorial explosion. We overcome this problem by biasing the distribution of generated molecules after each iterative step with experimental data. One such strategy prefers molecules possessing high support in terms of the right mass to charge ratio in the experimentally measured MS data and assuming one charge per molecule. With this interplay between computational and wetlab methodology, we hope to undercover complex chemical system structure such as autocatalytic cycles that would help to structure and bias in a meaningful way a chemical system that

would be considered random. One outcome of this would be a laboratory experiment that starts with simple precursors and in a few steps leads to a limited but desired set of molecules by tuning the system towards certain autocatalytic regions of the chemical landscape.

In traditional synthetic chemistry a chemical pathway can be executed in discrete steps that may be accomplished in days or even months depending on the complexity of the pathway. In contrast, the (bio)chemistry of living systems can produce specific compounds of interest using more than 100 different precursor molecules and catalysts in one pot within minutes. In order to develop traditional chemistry to be able to handle highly complex spaces [13], and to produce practical and desirable products and pathways, a more sophisticated conceptual approach and modelling approach is needed. The system presented here for HCN chemistry will be largely applicable to many types of chemical synthesis, biochemical systems and other complex systems.

In future iterations of this project, essential chemical parameters such as energy landscapes, kinetic components and even catalysts will be built into the system to provide a more realistic virtual chemical experiment. This will develop the system from providing what is possible to what is probable given the starting conditions. This will also help to provide a closer and more interesting coupling between wetlab and *in silico* experiments.

The mass spectrometry data from the available and future wetlab experiments gives many more possibilities for further analysis than what has been done in this study. Examples include analysis of the impact of different wetlab conditions on the network expansion, the use of the time dependent elution data of the mixture (from LC) in the modeling, high-resolution MS data, and fragmentation patterns for specific compounds. Such information can of course be included in our approach and it would obviously lead to tighter constraints for the chemical space exploration. Without conceptually changing our approach, this would allow for a more precise characterization of experimental results of the HCN polymerization and the subsequent hydrolysis. However, this was intentionally not the goal of this initial study and is included in our future research. We instead used varying and systematic conditions for polymerization and hydrolysis and then pooled the data to provide a broad representation of the possible chemical space explored. Nevertheless, our approach shows how analytical chemistry output from wetlab analysis (such as LC-MS) can be used to provide a deeper understanding of this highly complex chemical system. The approach will push synthetic chemistry towards utilizing more complex systems, which are more economical in terms of used resources, and will also advance the understanding of the inherent (self-)organization of chemical networks in living and non-living systems. By the close coupling of analytical chemistry with computer modeling we hope to gain not only an understanding of the origin of complex chemistries but also develop new tools for customizing chemistry. We envision the close coupling of chemical experiments with computer modeling to design fast, efficient, and cheap one pot chemical reactions in the future for the synthesis of desired compounds.

Acknowledgements

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Miller, S.L. A production of amino acids under possible primitive earth conditions. *Science* **1953**, *117*, 528–529.
2. Cleaves, H.J. Prebiotic chemistry: What we know, what we don't. *Evol. Educ. Outreach* **2012**, *5*, 342–360.
3. Summers, D.P. The prebiotic chemistry of nitrogen and the origin of life. In *Genesis—In The Beginning: Precursors of Life, Chemical Models and Early Biological Evolution*; Seckbach, J., Ed.; Springer: Heidelberg, Germany, 2012; Volume 22, pp. 201–216.
4. Miller, S.L. Production of some organic compounds under possible primitive earth conditions. *J. Am. Chem. Soc.* **1955**, *77*, 2351–2361.
5. Miller, S.L. The mechanism of synthesis of amino acids by electric discharges. *Biochim. Biophys. Acta* **1957**, *23*, 480–489.
6. Oparin, A.I. *The Origin of Life*; Dover: Mineola, NY, USA, 1953.
7. Oró, J.; Kimball, A. Synthesis of purines under possible primitive earth conditions. I. Adenine from hydrogen cyanide. *Arch. Biochem. Biophys.* **1961**, *94*, 217–227.
8. Oró, J.; Kimball, A. Synthesis of purines under possible primitive earth conditions: II. Purine intermediates from hydrogen cyanide. *Arch. Biochem. Biophys.* **1962**, *96*, 293–313.
9. Ferris, J.P.; Wos, J.D.; Nooner, D.W.; Oró, J. Chemical evolution. XXI. The amino acids released on hydrolysis of HCN oligomers. *J. Mol. Evol.* **1974**, *3*, 225–231.
10. Ferris, J.P.; Joshi, P.C.; Edelson, E.H.; Lawless, J.G.J. HCN: A plausible source of purines, pyrimidines and amino acids on the primitive earth. *J. Mol. Evol.* **1978**, *11*, 293–311.
11. Voet, A.; Schwartz, A. Prebiotic adenine synthesis from HCN-Evidence for a newly discovered major pathway. *Bioorg. Chem.* **1983**, *12*, 8–17.
12. Miyakawa, S.; Cleaves, H.J.; Miller, S.L. The cold origin of life: B. Implications based on pyrimidines and purines produced from frozen ammonium cyanide solutions. *Orig. Life Evol. Biosph.* **2002**, *32*, 209–218.
13. Saladino, R.; Crestini, C.; Costanzo, G.; DiMauro, E. Advance in the prebiotic synthesis of nucleic acids bases: Implications for the origin of life. *Curr. Org. Chem.* **2004**, *8*, 1425–1443.
14. Borquez, E.; Cleaves, H.J.; Lazcano, A.; Miller, S.L. An investigation of prebiotic purine synthesis from the hydrolysis of HCN polymers. *Orig. Life Evol. Biosph.* **2005**, *35*, 79–90.
15. Matthews, C.N.; Minard, R.D. Hydrogen cyanide polymers, comets and the origin of life. *Faraday Discuss.* **2006**, *133*, 393–401, 427–452.
16. Ritson, D.; Sutherland, J.D. Prebiotic synthesis of simple sugars by photoredox systems chemistry. *Nat. Chem.* **2012**, *4*, 895–899.

17. Matthews, C.N. Hydrogen cyanide polymers from the impact of comet P/Shoemaker-Levy 9 on Jupiter. *Adv. Space Res.* **1997**, *19*, 1087–1091.

18. Khare, B.N.; Bakes, E.L.; Cruikshank, D.; McKay, C.P. Solid organic matter in the atmosphere and on the surface of outer Solar System bodies. *Adv. Space Res.* **2001**, *27*, 299–307.

19. Hébrard, E.; Dobrijevic, M.; Loison, J.C.; Bergeat, A.; Hickson, K.M. Neutral production of hydrogen isocyanide (HNC) and hydrogen cyanide (HCN) in Titan's upper atmosphere. *Astron. Astrophys.* **2012**, *541*, A21.

20. Cable, M.L.; Hörst, S.M.; Hodyss, R.; Beauchamp, P.M.; Smith, M.A.; Willis, P.A. Titan tholins: Simulating Titan organic chemistry in the Cassini-Huygens era. *Chem. Rev.* **2012**, *112*, 1882–1909.

21. Israël, G.; Szopa, C.; Raulin, F.; Cabane, M.; Niemann, H.B.; Atreya, S.K.; Bauer, S.J.; Brun, J.F.; Chassefière, E.; Coll, P.; *et al.* Complex organic matter in Titan's atmospheric aerosols from *in situ* pyrolysis and analysis. *Nature* **2005**, *438*, 796–799.

22. Quirico, E.; Montagnac, G.; Lees, V.; Mcmillan, P.F.; Szopa, C.; Cernogora, G.; Rouzaud, J.N.; Simon, P.; Bernard, J.M.; Coll, P.; *et al.* New experimental constraints on the composition and structure of tholins. *Icarus* **2008**, *198*, 218–231.

23. Mamajanov, I.; Herzfeld, J. HCN polymers characterized by SSNMR: Solid state reaction of crystalline tetramer (diaminomaleonitrile). *J. Chem. Phys.* **2009**, *130*, 134504.

24. Matthews, C.N.; Moser, R.E. Prebiological protein synthesis. *Proc. Natl. Acad. Sci. USA* **1966**, *56*, 1087–1094.

25. Evans, R.A.; Lorencak, P.; Ha, T.K.; Wentrup, C. HCN dimers: Iminoacetonitrile and N-cyanomethanimine. *J. Am. Chem. Soc.* **1991**, *113*, 7261–7276.

26. Minard, R.D.; Hatcher, P.G.; Gourley, R.C.; Matthews, C.N. Structural investigations of hydrogen cyanide polymers: New insights using TMAH thermochemolysis/GC-MS. *Orig. Life Evol. Biosph.* **1998**, *28*, 461–473.

27. Eastman, M.P.; Helfrich, F.S.E.; Umantsev, A.; Porter, T.L.; Weber, R. Exploring the structure of a hydrogen cyanide polymers by electron spin resonance and scanning force microscopy. *Scanning* **2003**, *25*, 19–24.

28. Ruiz-Bermejo, M.; de la Fuente José, L.; Rogero, C.; Menor-Salván, C.; Osuna-Estebana, S.; Martín-Gago, J.A. New Insights into the Characterization of 'Insoluble Black HCN Polymers'. *Chem. Biodiv.* **2012**, *9*, 25–40.

29. He, C.; Lin, G.; Upton, K.T.; Imanaka, H.; Smith, M.A. Structural investigation of HCN polymer isotopomers by solution-state multidimensional NMR. *J. Phys. Chem. A* **2012**, *116*, 4751–4759.

30. Matthews, C.; Nelson, J.; Varma, P.; Minard, R. Deuterolysis of amino acid precursors: Evidence for hydrogen cyanide polymers as protein ancestors. *Science* **1977**, *198*, 622–625.

31. Hanczyc, M.M. Metabolism and motility in prebiotic structures. *Phil. Trans. R. Soc. B* **2011**, *366*, 2885–2893.

32. Bonnet, J.Y.; Thissen, R.; Frisari, M.; Vuitton, V.; Quirico, E.; Orthous-Daunay, F.R.; Dutuit, O.; Roy, L.L.; Fray, N.; Cottin, H.; *et al.* Compositional and structural investigation of HCN polymer through high resolution mass spectrometry. *Int. J. Mass Spectrom.* **2013**, doi:10.1016/j.ijms.2013.06.015.

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

34. Rozenberg, G.; Ehrig, H. *Handbook of Graph Grammars and Computing by Graph Transformation*; World Scientific: Singapore, 1997; Volume 1.

35. Ehrig, H.; Ehrig, K.; Prange, U.; Taenthzer, G. *Fundamentals of Algebraic Graph Transformation*; Springer-Verlag: Berlin, Germany, 2006.

36. Dugundji, J.; Ugi, I.K. An algebraic model of constitutional chemistry as a basis for chemical computer programs. *Top. Curr. Chem.* **1973**, *39*, 19–64.

37. Ugi, I.K.; Bauer, J.; Blomberger, C.; Brandt, J.; Dietz, A.; Fontain, E.; Gruber, B.; von Scholley-Pfab, A.; Senff, A.; Stein, N. Models, concepts, theories, and formal languages in chemistry and their use as a basis for computer assistance in chemistry. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 3–16.

38. Hatzimanikatis, V.; Li, C.; Ionita, J.A.; Henry, C.S.; Jankowski, M.D.; Broadbelt, L.J. Exploring the diversity of complex metabolic networks. *Bioinformatics* **2005**, *21*, 1603–1609.

39. Leber, M.; Egelhofer, V.; Schomburg, I.; Schomburg, D. Automatic assignment of reaction operators to enzymatic reactions. *Bioinformatics* **2009**, *25*, 3135–3142.

40. Andersen, J.L.; Flamm, C.; Merkle, D.; Stadler, P.F. Generic strategies for chemical space exploration. *Int. J. Comput. Biol. Drug Des.* **2013**, in press; see also arXiv:1302.4006.

41. O'Boyle, N.M.; Banck, M.; James, C.A.; Morley, C.; Vandermeersch, T.; Hutchison, G.R. Open Babel: An open chemical toolbox. *J. Cheminformatics* **2011**, *3*, 33.

42. Roy, D.; Najafian, K.; von Ragué Schleyer, P. Chemical evolution: The mechanism of the formation of adenine under prebiotic conditions. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 17272–17277.

43. Andersen, J.L.; Flamm, C.; Merkle, D.; Stadler, P.F. Maximizing output and recognizing autocatalysis in chemical reaction networks is NP-complete. *J. Syst. Chem.* **2012**, *3*, 1–9.

44. Saladino, R.; Crestini, C.; Pino, S.; Costanzo, G.; Di Mauro, E. Formamide and the origin of life. *Phys. Life Rev.* **2012**, *9*, 84–104.

Appendix

Molecules with High Support from MS Scan Data

The following table lists all the molecules which passed the filters in each step of the chemical space expansion after polymerization and hydrolysis; given is the molar mass, the support based on the intensity of the MS data, and the normalized Boltzmann factor within the class of isomers. The molecules are grouped by isomer class and ordered by decreasing intensity. Each isomer class is internally ordered by decreasing Boltzmann factor. The listed molecules correspond to many of the high-intensity peaks in the data set (comp. Figure 2 right). Further calculations with less strict filters may provide molecule suggestions for other peaks.

Id	SMILES	Molar Mass	Intensity	Boltzmann factor (normalized)
1949	OC(NC=N)N(C=N)C(NC=N)C#N	197.2	$1.01 \cdot 10^7$	0.56
3173	NC(O)N(C=NC=N)C(NC=N)C#N	197.2	$1.01 \cdot 10^7$	0.44
428	OC(NC=NC=N)NC(NC=N)C#N	197.2	$1.01 \cdot 10^7$	< 0.01
5769	NC(O)NC(N=CO)C(NC(NC=NC=NC=N)C#N)=NC=NC=O	351.3	$8.20 \cdot 10^6$	0.55
5310	OC=NC(NC(O)NC=O)C(NC(NC=NC=NC=N)C#N)=NC=N	351.3	$8.20 \cdot 10^6$	0.42
4363	OC=NC(O)NC(NC=NC=NC=NC=N)C(NC(NC=O)C#N)=N	351.3	$8.20 \cdot 10^6$	0.03
4090	OC=NC(O)NC(NC=NC=NC=NC=N)C(NC(N=CO)C#N)=N	351.3	$8.20 \cdot 10^6$	< 0.01
3472	OC=NC(O)N(C=N)C(NC=NC=NC=N)C(NC(N=CO)C#N)=N	351.3	$8.20 \cdot 10^6$	0.00
869	C1N=CN=CN=CN=1	108.1	$7.98 \cdot 10^6$	1.00
4473	OC(NC=N)N(C=N)C(NC=O)C#N	198.2	$6.66 \cdot 10^6$	0.39
3243	NC(O)N(C=N)C(NC=NC=O)C#N	198.2	$6.66 \cdot 10^6$	0.36
3260	NC(O)N(C=NC=O)C(NC=N)C#N	198.2	$6.66 \cdot 10^6$	0.21
2382	OC(NC=NC=N)NC=NC=NC=O	198.2	$6.66 \cdot 10^6$	0.02
1743	N=CN=CNC(NC=NC=N)OC=O	198.2	$6.66 \cdot 10^6$	0.02
5499	N=CN(C=N)C(OC=O)N(C=N)C(NC=O)C#N	253.2	$6.50 \cdot 10^6$	0.49
4659	OC(NC=NC=O)N(C=N)C(NC=NC=O)C#N	253.2	$6.50 \cdot 10^6$	0.37
3701	OC(NC=NC=N)N(C=NC=O)C(NC=O)C#N	253.2	$6.50 \cdot 10^6$	0.13
3392	OC=NC(C#N)N(C=NC=O)C(O)NC=NC=N	253.2	$6.50 \cdot 10^6$	0.01
2276	NC(C(NC=N)NC=N)=NC=N	169.2	$6.07 \cdot 10^6$	0.31
878	N=CNC(NC=N)C(NC=N)=N	169.2	$6.07 \cdot 10^6$	0.28
1358	NC(NC=NC=N)C(NC=N)=N	169.2	$6.07 \cdot 10^6$	0.22
5338	NC=NC(N)=C(N)NC=NC=N	169.2	$6.07 \cdot 10^6$	0.18
4135	NC=NC(NC=NC=N)C(N)=N	169.2	$6.07 \cdot 10^6$	0.01
4995	NC(O)NC(O)N(C=NC=O)C(NC=O)C#N	244.2	$6.05 \cdot 10^6$	0.88
3769	OC(NC=NC=O)NC(NC=NC=N)OC=O	244.2	$6.05 \cdot 10^6$	0.12
4535	NC(O)N(C=NC=O)C(NC=O)C#N	199.2	$6.02 \cdot 10^6$	0.43
4541	N=CN=CNC(NC=NC=O)OC=O	199.2	$6.02 \cdot 10^6$	0.43
3829	N=CNC(NC=NC=NC=O)OC=O	199.2	$6.02 \cdot 10^6$	0.14
3468	OC=NC(O)N(C=N)C(N=CO)C#N	199.2	$6.02 \cdot 10^6$	< 0.01
5259	OC(NC=NC=O)N(C=NC=O)C(NC=O)C#N	254.2	$5.78 \cdot 10^6$	0.63
5494	N=CN(C(NC=NC=O)OC=O)C(NC=O)C#N	254.2	$5.78 \cdot 10^6$	0.36
4359	NC(OC=O)N(C=NC=O)C(NC=NC=O)C#N	254.2	$5.78 \cdot 10^6$	0.01
4114	O=COC(NC=NC=NC=O)NC=NC=NC=O	254.2	$5.78 \cdot 10^6$	< 0.01
5785	NC=NC(NC(N)=C(N)N)=C1NC(O)NC=N1	241.3	$5.63 \cdot 10^6$	0.47
5383	NC(N)=C(N)NC(NC=N)=C1NC(O)NC=N1	241.3	$5.63 \cdot 10^6$	0.45
4733	NC(N)=C(N)NC(C1N=CNC(O)N1C=N)=N	241.3	$5.63 \cdot 10^6$	0.07
2136	NC(O)NC(N(C=N)C=NC=N)C(NC=N)=N	241.3	$5.63 \cdot 10^6$	< 0.01
3927	NC(NC(C1N=CNC(O)N1C=N)=N)C(N)=N	241.3	$5.63 \cdot 10^6$	< 0.01
457	NC(O)NC(C(NC=N)NC=NC=NC=N)=N	241.3	$5.63 \cdot 10^6$	< 0.01
6342	NC(N)=C(N)NC(O)NC=NC(C(NC=O)NC(O)NC=NC=O)=NC=N	386.4	$5.33 \cdot 10^6$	0.65
6222	NC(N)=C(N)NC(O)NC=NC(C(NC=O)N(C=N)C(O)NC=NC=O)=N	386.4	$5.33 \cdot 10^6$	0.25
6193	NC(N)=C(N)NC(O)NC=NC(C(N=CO)N(C=N)C(O)NC=NC=O)=N	386.4	$5.33 \cdot 10^6$	0.10
5802	NC(NC(O)NC=NC(C(N=CO)N(C=N)C(O)NC=NC=O)=N)C(N)=N	386.4	$5.33 \cdot 10^6$	< 0.01
4905	NC(C(NC(O)NC=N)NC(C(N=CO)N(C=N)C(NC=N)OC=O)=N)=N	386.4	$5.33 \cdot 10^6$	< 0.01
3281	NC=NC(N(C=N)C(NC=NC=N)C#N)C(NC(C#N)N(C=N)C=N)=NC=O	386.4	$5.30 \cdot 10^6$	0.84
2407	N=CNC(N(C=N)C(NC=NC=N)C#N)C(NC(C#N)N(C=N)C=N)=NC=O	386.4	$5.30 \cdot 10^6$	0.16
5550	NC=NC(N)=C(N)NC(NC=N)C#N	196.2	$4.49 \cdot 10^6$	0.58
1670	N=CNC(NC=N)C(NC=N)=NC=N	196.2	$4.49 \cdot 10^6$	0.32
1323	NC(C(NC=N)N(C=N)C=NC=N)=N	196.2	$4.49 \cdot 10^6$	0.11
2330	NC(C(NC=N)NC(C(NC=N)NC=N)=N)=N	240.3	$4.18 \cdot 10^6$	1.00
5169	NC(NC(O)NC=NC(C(N=CO)N(C=N)C(O)NC=N)=N)C(N)=N	358.4	$3.92 \cdot 10^6$	0.98
4452	NC(C(NC(O)NC=N)NC(C(N=CO)N(C=N)C(O)NC=N)=N)=N	358.4	$3.92 \cdot 10^6$	0.02
744	N=CNC(N(C=N)C(NC=NC=N)C#N)C(NC(C#N)N(C=N)C=N)=N	358.4	$3.90 \cdot 10^6$	1.00
4829	OC(N(C=N)C=N)N(C=N)C(NC=O)C#N	225.2	$3.83 \cdot 10^6$	0.48
4828	OC(NC=N)N(C=N)C(C#N)N(C=N)C=O	225.2	$3.83 \cdot 10^6$	0.37
2826	OC(NC=NC=N)N(C=N)C(NC=O)C#N	225.2	$3.83 \cdot 10^6$	0.09
3007	OC(NC=N)N(C=N)C(NC=NC=O)C#N	225.2	$3.83 \cdot 10^6$	0.07

Id	SMILES	Molar Mass	Intensity	Boltzmann factor (normalized)
2375	OC(NC=NC=N)NC(NC=NC=O)C#N	225.2	$3.83 \cdot 10^6$	< 0.01
2478	OC=NC(C#N)N(C=N)C(O)NC=NC=N	225.2	$3.83 \cdot 10^6$	< 0.01
4362	OC(NC=O)NC(NC=N)C(NC(NC=NC=NC=N)C#N)=NC=N	350.3	$3.80 \cdot 10^6$	0.98
4061	OC=NC(O)NC(NC=N)C(NC(NC=NC=NC=N)C#N)=NC=N	350.3	$3.80 \cdot 10^6$	0.02
3343	OC=NC(O)NC(NC=NC=NC=N)C(NC(NC=N)C#N)=NC=N	350.3	$3.80 \cdot 10^6$	< 0.01
2489	OC=NC(O)N(C=N)C(NC=NC=NC=N)C(NC(NC=N)C#N)=N	350.3	$3.80 \cdot 10^6$	< 0.01
4475	OC(NC=N)N(C=NC=O)C(NC=O)C#N	226.2	$3.71 \cdot 10^6$	0.73
4122	NC(O)N(C=NC=O)C(NC=NC=O)C#N	226.2	$3.71 \cdot 10^6$	0.26
2820	N=CN=CNC(NC=NC=NC=O)OC=O	226.2	$3.71 \cdot 10^6$	0.01
4921	OC(NC=NC=O)NC(NC=O)OC=O	218.2	$3.60 \cdot 10^6$	0.96
4709	OC=NC(NC(O)NC=NC=O)OC=O	218.2	$3.60 \cdot 10^6$	0.04
4619	NC(C(NC=NC=N)NC(C(NC(O)NC=NC=N)N=CO)=N)=N	340.3	$3.17 \cdot 10^6$	0.96
4173	NC(C(NC=NC=N)NC(C(N=CO)N(C=N)C(O)NC=N)=N)=N	340.3	$3.17 \cdot 10^6$	0.04
1914	OC1NC=NC(N1C=N)C(NC(NC=NC=N)C#N)=N	278.3	$3.08 \cdot 10^6$	1.00
3066	NC=NC=NC(N(C=N)C(NC=NC=N)C#N)C(NC(C#N)N(C=N)C=N)=N	385.4	$3.01 \cdot 10^6$	0.86
581	N=CN=CNC(C#N)N(C=N)C(NC=NC=N)C(NC(C#N)N(C=N)C=N)=N	385.4	$3.01 \cdot 10^6$	0.14
548	OC(NC=NC=N)NC(N(C=N)C=NC=N)C(NC=N)=N	295.3	$2.60 \cdot 10^6$	0.66
2754	NC(C(NC=NC=N)NC(C1N=CNC(O)N1C=N)=N)=N	295.3	$2.60 \cdot 10^6$	0.34
2040	OC(NC=NC=N)N1C=NC1C#N	180.2	$2.59 \cdot 10^6$	1.00
2945	OC(NC=N)NC(O)NC(NC=NC=O)C#N	243.2	$2.48 \cdot 10^6$	0.53
2966	OC(O)NC(C#N)N(C=N)C(O)NC=NC=N	243.2	$2.48 \cdot 10^6$	0.47
482	OC(NC=N)NC=NC=N	143.1	$2.35 \cdot 10^6$	1.00
1706	OC(NC=NC=NC=N)NC(NC=NC=N)C(NC=O)=N	296.3	$2.32 \cdot 10^6$	1.00
2243	N=CNC(NC=N)C(NC(NC=N)C(NC=N)=N)=N	267.3	$2.20 \cdot 10^6$	1.00
2379	OC(NC=NC=N)N(C=N)C(NC=NC=O)C#N	252.2	$2.19 \cdot 10^6$	0.59
2381	OC(NC=NC=N)N(C=NC=O)C(NC=N)C#N	252.2	$2.19 \cdot 10^6$	0.41
488	OC(NC=NC=NC=N)NC(NC=NC=N)C#N	251.2	$2.00 \cdot 10^6$	0.64
487	OC(NC=N)NC(NC=NC=NC=NC=N)C#N	251.2	$2.00 \cdot 10^6$	0.36
1345	NC(C(NC=N)NC(C(NC=N)N(C=N)C=NC=N)=N)=N	294.3	$1.94 \cdot 10^6$	1.00
722	N#CC1NC=NC=NC=N1	135.1	$1.77 \cdot 10^6$	1.00
469	OC(NC=NC=N)NC=NC=N	170.2	$1.61 \cdot 10^6$	1.00
464	OC(NC=NC=N)N(C=N)C(NC=N)C#N	224.2	$1.27 \cdot 10^6$	1.00
1053	N=CN=CNC(NC(C1N=CN=CN1C=N)=N)C#N	260.3	$1.23 \cdot 10^6$	1.00
890	N=CN1C=NC=NC=NC1C#N	162.2	$1.12 \cdot 10^6$	1.00
529	OC(NC=NC=N)N(C=N)C(NC=NC=NC=N)C(NC(NC=N)C#N)=N	376.4	$9.94 \cdot 10^5$	1.00
803	N=CNC(NC(C(NC=N)N(C=N)C=NC=N)=N)C(NC=N)=N	321.3	$9.93 \cdot 10^5$	1.00
565	OC(NC=NC=N)NC(NC=NC=N)C(NC=N)=N	268.3	$9.59 \cdot 10^5$	1.00