ChemPubSoc Europe

DOI: 10.1002/syst.201900014

CHEMSYSTEMSCHEM Communications

Taming the Combinatorial Explosion of the Formose Reaction via Recursion within Mineral Environments

Stephanie Colón-Santos, Geoffrey J. T. Cooper, and Leroy Cronin*[a]

One-pot reactions of simple precursors, such as those found in the formose reaction or formamide condensation, continuously lead to combinatorial explosions in which simple building blocks capable of function exist, but are in insufficient concentration to self-organize, adapt, and thus generate complexity. We set out to explore the effect of recursion on such complex mixtures by 'seeding' the product mixture into a fresh version of the reaction, with the inclusion of different mineral environments, over a number of reaction cycles. Through untargeted UPLC-HRMS analysis of the mixtures we found that the overall number of products detected reduces as the number of cycles increases, as a result of recursively enhanced mineral environment selectivity, thus limiting the combinatorial explosion. This discovery demonstrates how the involvement of mineral surfaces with simple reactions could lead to the emergence of some building blocks found in RNA, ribose and uracil, under much simpler conditions that originally thought.

The mechanism which led to the first genetic-machine, an adaptive chemical system that uses a genetic code to organise metabolic function and propagate that code, is one of the most important outstanding questions in science.[1,2] Modern organisms are genetic machines that take part in open-ended information transfer using biopolymers such as RNA and DNA, which are ubiquitous to all known life forms. *De novo* nucleotide (e.g. monomer of DNA and RNA) synthesis has not been accomplished from a simple or prebiotic route to sugars or purines, even if some progress has been made on the prebiotic synthesis of nucleosides and nucleotides.[3–6] The one-pot synthesis of all the required compounds can be achieved through a diverse set environmental conditions,[7–10] but they always result in a convoluted, and analytically intractable, complex mixture of products.[11–15] Identification of the direct transition of such units into polymers from these mixtures is very challenging analytically, and complete chemical characterization is nearly impossible, as in the case of tholins.[16] As such, the combinatorically large number of products can justify employing a less product-explosive process involving a multi-step synthesis approach. However, the interaction of simple molecules with the environment has been proven to steer the chemical networks into different outcomes or product populations, giving them a higher level of order as a result of environmental constraints (such as inorganic catalysts).[17–20] In particular, the presence of mineral surfaces is known to sometimes truncate the combinatorial explosions generated by one-pot reaction of simple compounds. Two relevant examples are the preferential formation of ribose when borate minerals are added to the formose reaction,[21] a system known for the incredible complexity of its product distribution, and the clear selectivity towards the production of certain nucleobases when formamide condensation is carried out on different mineral surfaces.[22] Notably, these previous results were obtained in batch reactions, leaving the possibility that this effect could be amplified if the reaction mixture was cycled over a given environment.

To investigate this, we set out to explore the effect of reaction cycling by seeding with the products of the previous reaction cycle (recursion) on well-known combinatorial explosions. We carried out the formose reaction in formamide with different mineral environments, see Figure 1, to assess whether the selectivity imparted by the environment can be amplified through recursion, whilst truncating the combinatorial explosion by reducing the overall number of products. We found that the recursive action resulted in a lower number of individual products, with or without a mineral surface, demonstrating that reaction cycling has a significant effect on the product distribution. We also observed a significant increase in the yields of certain species when minerals were present, showing that selection by the environment also plays a role in determining the product mixture, see Figure 2.

In order to investigate and establish the nature of any differences in the product distribution without bias, untargeted analysis of the mixtures was conducted with Hydrophilic Interaction Liquid Chromatography (HILIC). The Ultra-Performance Liquid Chromatography was coupled to tandem mass-spectrometry (UPLC-MS/MS) and carried out in a data dependent fashion, which allowed us to investigate the resulting chemical space without having to target any particular compound. By generating features based on exact mass (m/z) and retention time (RT), we were able to achieve a meaningful representation of the product distribution from mass-spectral data. The features represent unique reaction products and their

[a] S. Colón-Santos, Dr. G. J. T. Cooper, Prof. L. Cronin
School of Chemistry
University of Glasgow
University Avenue, Glasgow, G12 8QQ, UK
E-mail: lee.cronin@glasgow.ac.uk

Supporting information for this article is available on the WWW under https://doi.org/10.1002/syst.201900014

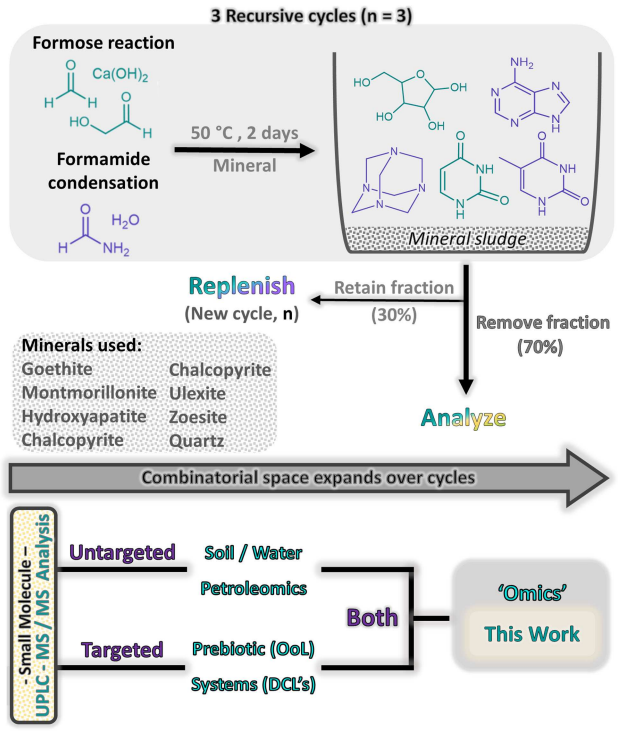

Figure 1. Recursive cycles: A formose reaction (green) in formamide/water (purple) is carried out in the presence of a mineral. After each cycle of 48 h at 50 °C, a fraction of the total volume (70 %, from the top) is removed and the vial is replenished with fresh starting materials to start the next cycle. UPLC-MS/MS analysis: An untargeted analysis, followed by a targeted data processing was conducted in order to explore the resulting product distribution with an 'omics' approach.

number maps to the number of individual species, providing a way to gauge the complexity of the mixture.

To make the large volume of data more accessible, detected features were binned by their m/z values as a means to fingerprint the product distribution (Figure 2). The number of features in each range of molecular weight changes as an effect of recursive action, with a general trend that the number decreases from Cycle 1 to Cycle 3 (Figure 2 a, b). Differences in the distribution also arise as an effect of the environment, as observed in Figure 2 c, between the reaction with no mineral and with the inclusion of a mineral surface. For each environment, the features generate a different pattern which also changes across recursive cycles. The number of detected features decreases from Cycle 2 to Cycle 3 for all reactions, demonstrating that the action of recursive cycles is limiting the combinatorial explosion expected from these reactions. In the case of Chalcopyrite, Quartz and in the absence of any mineral surface (control), the number of features reduces linearly from Cycle 1 to Cycle 3, while all other mineral environments see the number of features peak in Cycle 2 and decrease again in Cycle 3. This suggests that the reactions proceed along different trajectories, towards different product distributions, as a direct result of the mineral environment.

To validate our in-house feature generator and 'omics' based approach to complex mixture analysis, we processed the data using CompoundDiscoverer™ (Thermo Scientific),[23] a conventionally used software for processing untargeted mass-spectral data, which also enabled the extraction of ion chromatograms (EIC's) in a targeted fashion. While this method generated fewer features overall, the trends were consistent throughout the experiments (see Figure 2 and Figure S4).

During the data-dependent acquisition (DDA) of the mass-spectral data, the most intense peaks were fragmented further into MS^2 fragments. This allowed us to identify some of the products using database matching and validation against pure standards to confirm chemical identities. By using the MS^2 data, we were able to do qualitative structural analysis and identify some of the features as Ribose and Uracil, the building blocks of RNA. We found that conventionally analytically targeted products, such as nucleobases, where not only present in our product mixtures but also produced preferentially on mineral surfaces, as observed in the difference between intensity scales for the selected ion in the extracted ion chromatogram (EIC) for Uracil, in the reaction with and without the mineral chalcopyrite, and in the peak areas, shown in Figure 3 c,d,e. Traces of nucleosides (Thymidine and Adenosine) were also detected for most samples in Cycle 3, including the non-mineral control reaction (see Figure S11–S13).

In addition, we detected Hexamethylenetetramine (HMT) across all reactions. HMT was discovered by Aleksandr Butlerov in 1859 and is prepared industrially by combining formaldehyde and ammonia.[24] The significance of HMT in prebiotic chemistry has been discussed previously,[25] particularly in its role of incorporating formaldehyde (from its reaction with ammonia, which is generated *in-situ* by the decomposition of formamide) into a more stable compound, possibly allowing for it to be concentrated in a prebiotic, evaporative environment. The concentration of HMT changed across recursive cycles (Figure 3), with a significant drop being observed after the second cycle for all samples (including the control). We postulate that the HMT is depleted by reaction with the products of Cycle 2, but we currently have no definitive evidence for this, or a mechanism responsible.

In order to show concentrations of specific products, we calculated the relative abundance for the selected ions, which were the features with a matching exact mass to HMT, Uracil and Ribose that were taken to MS^2 by the DDA method, allowing us to validate their identity. This is done in a qualitative manner, as an accurate quantification of this compounds in different complex matrices would require a targeted analysis, which is beyond the scope of this work. We acknowledge this limitation from the untargeted acquisition method and aim to complement it with a targeted workflow in further investigations. However, in a targeted approach we would need to be able to identify the relevant features in a chemical system, but the relevance of certain features over others in complex product distributions is not trivial and it would benefit from a discovery-driven investigation, such as the one used in metabolomics-type workflows and this work. This approach generates a more robust overview of the highly complex product distribution generated in analytically intractable mixtures, as a means to further our understanding of

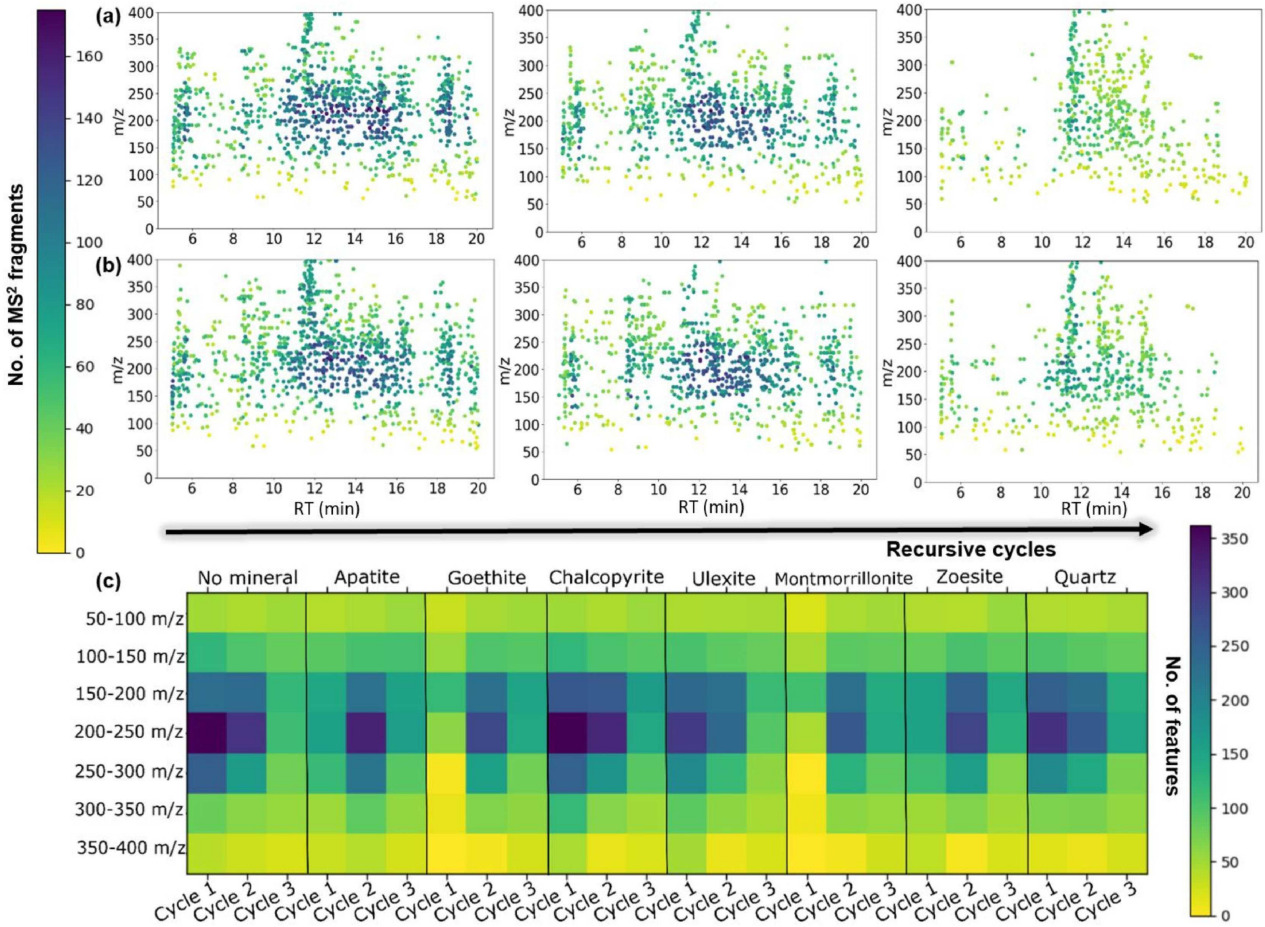

Figure 2. Mass spectral features: Features are based on unique exact mass (m/z) and retention time (RT). Over recursive cycles, differences in the number of features and MS^2 fragments (of each feature) can be observed for both (a) the recursive formose reaction in formamide control (no mineral) and (b) the recursive reaction in the presence of a mineral surface (Chalcopyrite, Cu_2FeS). A heatmap of the features (c) was generated by grouping the features into 50 m/z bins, resulting in a unique pattern for each reaction environment over the three recursive cycles.

complex chemical systems and their intrinsic reproducibility. Due to the high complexity in the product distribution of combinatorial explosions, a satisfactory reproducibility assessment would need a large number of experimental replicates, where a high-throughput experimental design is required. While this is not assessed directly in these work, it has indeed enabled the possibility for such studies, in which a comprehensive overview of the resulting products would be substantial in order to draw any meaningful conclusions.

In summary, we carried out the formose reaction and formamide condensation in a one-pot fashion, under milder conditions than previously reported,[2] while a recursive environment was applied to the resulting mixture in a series of cycles. We found that recursive cycles not only truncated the combinatorial explosion by reducing the number of individual products, but also successfully generated sugars and nucleobases from potentially prebiotic routes, in an integrated fashion. Traces of nucleoside formation were also detected after two recursive cycles, for the first time in this simple-precursor systems (e.g. Formose reaction/ Formamide condensation). Furthermore, we found a molecule with a strong connection to prebiotically-relevant compounds, hexamethylenetetramine (HMT), which might have a non-trivial relationship with the formation of these building blocks. The untargeted analysis of the mixtures allowed for an unprecedented exploration of the chemical space generated in analytically intractable (prebiotic) combinatorial explosions. We believe that recursive experiments bring us one step closer to a plausible 'real-life' scenario and combined with this analytical approach, it provides an improved experimental regime for looking at the evolution of complex mixtures from simple precursors under non-equilibrium conditions.

Experimental

Experimental Methods

A formose reaction (formaldehyde, glycolaldehyde, calcium hydroxide) was carried out in in formamide-water (50:50 v/v) on seven different mineral surfaces (*see SI*, *Page 2*), as well as, in the absence of any mineral surface (*e.g.* control). The reactions were stirred at 1200 rpm and heated at 50 °C, for 48 hours. Then, about

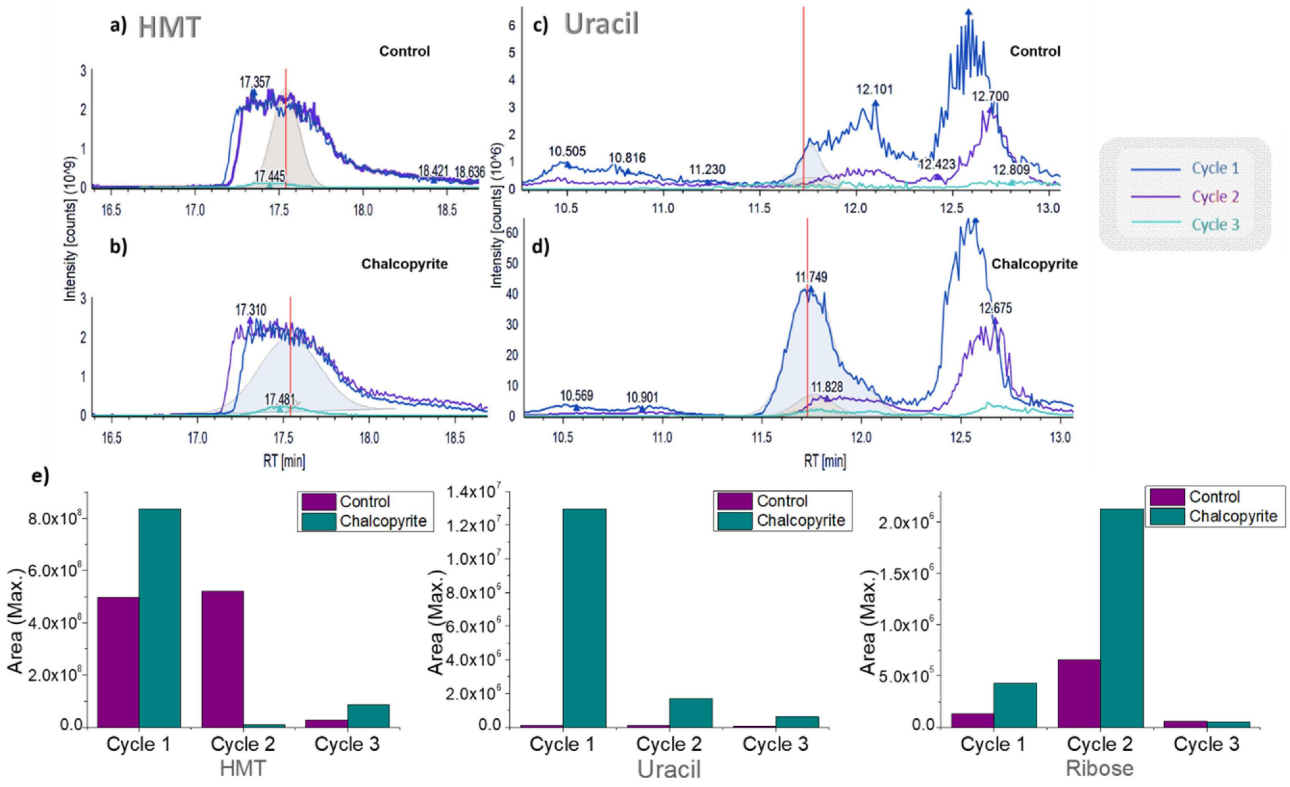

Figure 3. Identification of RNA building blocks and HMT: Extracted Ion Chromatograms (EICs) of HMT (m/z: 141.11, Adduct: [M + H]) for (a) the control reaction and (b) in the presence of a mineral surface (Chalcopyrite). EICs for Uracil (m/z: 113.03, Adduct: [M + H]) (c) in the control reaction and (d) in the presence of a mineral surface. Relative abundance of HMT, Uracil and Ribose (e) calculated by integration of EIC's for the selected ions.

70% (~3.5 mL) of the reaction volume (supernatant) was removed for analysis.

Recursive Cycles

The remaining fraction (~1.5 mL) was used to seed the next reaction. Topping up with the same concentration of starting materials (3.5 mL), but conserving the total reaction volume (5 mL); we repeated the process.

Sample Preparation

The removed fraction was allowed to cool to room temperature. Then, a 100μL aliquot was taken for each analysis; to which an ion-exchange resin was employed to

remove excess cations in solutions (*e.g.* Ca^{2+}) and the supernatant transferred to glass vial, followed by a 1 in a 100 dilution with MS grade water. Finally, the solution was filtrated with a syringe filter (0.22 μm cut-off)].

Ultra-Performance Liquid Chromatography and Tandem Mass Spectrometry

Chromatographic separation was achieved using a Thermo Vanquish UPLC with a ZIC-HILIC column, eluted in a linear gradient mixture of solvents A (water w/20 mM Ammonium Acetate, pH = 5) and B (100% acetonitrile w/0.1% v/v formic acid) over 25 min, coupled to a Thermo Fusion Orbitrap for mass-spectral analysis. Spectra were collected for 30 minutes in positive mode over a scan range of 50–500 m/z. Ion transfer tube was set to 275 °C, RF lens 60%, and acquisition was performed in a Data-dependent (DDA) manner. The Fragmentation data was collected at top speed (3 second window) with an intensity threshold of 5.0E4 and dynamic exclusion, after one time for 15 seconds, using the ion trap isolation at HCD collision energy of 35 eV and resolution 15000.

Interpretation of Raw Data

All raw files were converted to mzML and centroided using Proteowizard's[26] convert function (with a vendor-specific algorithm). The converted files (mzML) were processed in Python using Pymzml. In each file, (m/z, intensity, rt) features were extracted using pymzml feature detection algorithm, with default parameter values used for both the centroiding and mass trace detection. Performance of the feature detection and extraction algorithm was evaluated by comparing them with those generated in an analogous processing software, CompoundDiscoverer™, which was developed particularly for data acquired in Thermo-Orbitrap instruments and used to automatically detect features across samples; which were comparable with those obtained with Pymzml.

Data Analysis

After aligning the peaks detected across all samples and removing those present in the blanks, duplicate features were removed by eliminating values that had the same exact mass (to the third decimal value) and were within an acceptable retention time window (+/− 30 s) of each other. Filtering of the features was achieved by a 2-step procedure, with in-house scripts developed in

python: (1) All detected features were filtered for those that had MS/MS spectra appended and (2) which were not present in any sample blanks. The DDA fashion in which the data was acquired, allows for this filtering to be possible without losing any of the most abundant compounds and allows for plausible chemical identification of the features. The relative abundance was carried out in a qualitative manner, as only the features (e.g. Retention Time – *m/z*)

Acknowledgements

We gratefully acknowledge financial support from the EPSRC (Grant Nos EP/P00153X/1, EP/J015156/1, EP/K021966/1, EP/K038885/1, EP/L015668/1, EP/L023652/1), BBSRC (Grant No. BB/M011267/1), ERC (project 670467 SMART-POM), and the John Templeton Foundation Grant ID 60625 and Grant ID 61184. We thank Dario Caramelli, Dr. Davide Angelone and Graham Keenan for their help in writing the python scripts.(School of Chemistry, University of Glasgow)

Conflict of Interest

The authors declare no conflict of interest.

Keywords: combinatorial chemistry · complex mixtures · prebiotic chemistry · recursive chemistry · systems chemistry

[1] L. Cronin, S. I. Walker, *Science* **2016**, *352*, 1174–1175.
[2] N. Guttenberg, N. Virgo, K. Chandru, C. Scharf, I. Mamajanov, *Philos. Trans. R. Soc. London* **2017**, *375*, 20160347.
[3] R. Saladino, G. Botta, S. Pino, G. Costanzo, E. Di Mauro, *Biochimie* **2012**, *94*, 1451–1456.
[4] M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **2009**, *459*, 239–242.
[5] S. C. Kim, D. K. O'Flaherty, L. Zhou, V. S. Lelyveld, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 13318–13323.
[6] I. Suárez-marina, Y. M. Abul-haija, R. Turk-macleod, P. S. Gromski, G. J. T. Cooper, A. O. Olivé, S. Colón-santos, L. Cronin, *Commun. Chem.* **2019**, *2*, 28.
[7] J. D. Sutherland, *Angew. Chem. Int. Ed.* **2015**, *54*, 104–121.
[8] D. Ross, D. Deamer, S. Lower, *Life* **2019**, *6*, 28.
[9] S. A. Benner, H. J. Kim, M. a. Carrigan, *Acc. Chem. Res.* **2012**, *45*, 2025–2034.
[10] N. V. Hud, *Synlett* **2017**, *28*, 36–55.
[11] K. Ruiz-Mirazo, J. Peretó, A. Moreno, *Origins Life Evol. Biospheres* **2004**, *34*, 323–346.
[12] E. Wollrab, S. Scherer, F. Aubriet, V. Carré, T. Carlomagno, L. Codutti, A. Ott, *Origins Life Evol. Biospheres* **2016**, *46*, 149–169.
[13] S. Scherer, E. Wollrab, L. Codutti, T. Carlomagno, S. G. da Costa, A. Volkmer, A. Bronja, O. J. Schmitz, A. Ott, *Origins Life Evol. Biospheres* **2017**, *47*, 381–403.
[14] M. Ferus, A. Knížek, S. Civiš, *Proc. Mont. Acad. Sci.* **2015**, *112*, 7109–7110.
[15] W. Martin, M. J. Russell, *Philos. Trans. R. Soc. London* **2003**, *358*, 59–83; discussion 83–5.
[16] M. Ruiz-Bermejo, C. Menor-Salván, E. Mateo-Martí, S. Osuna-Esteban, J. Á. Martín-Gago, S. Veintemillas-Verdaguer, *Icarus* **2008**, *198*, 232–241.
[17] D. A. Baum, K. Vetsigian, *Origins Life Evol. Biospheres* **2017**, *47*, 481–497.
[18] L. Boiteau, R. Pascal, *Origins Life Evol. Biospheres* **2011**, *41*, 23–33.
[19] R. M. Hazen, D. A. Sverjensky, *Cold Spring Harb. Perspect. Biol.* **2010**
[20] A. J. Surman,, M. R. Garcia, Y. M. Abul-Haija, G. Cooper, P. S. Gromski, R. Turk-MacLeod, M. Mullin, C. Mathis, S. Walker, L. Cronin, *PNAS* **2019**, *116*, 5387–5392.
[21] A. Ricardo, *Science* **2004**, *303*, 196–196.
[22] R. Saladino, J. M. G. Ruiz, E. Di Mauro, *Chem. Eur. J.* **2019**, *25*, 3181–3189.
[23] E. Hung, *ThermoFisher Scientific*, **2017**.
[24] D. A. Butlerow, *Ber. Dtsch. Chem. Ges.* **1860**, *8*, 398–416.
[25] H. J. Cleaves, *Precambrian Res.* **2008**, *164*, 111–118.
[26] M. C. Chambers, B. Maclean, R. Burke, D. Amodei, D. L. Ruderman, S. Neumann, L. Gatto, B. Fischer, B. Pratt, J. Egertson, *Nat. Biotechnol.* **2012**, *30*, 918–920.

Manuscript received: April 11, 2019
Accepted manuscript online: April 30, 2019
Version of record online: June 11, 2019

