Evolution of Complex Chemical Mixtures Reveals Combinatorial Compression and Population Synchronicity

Kavita Matange[1,2], Vahab Rajaei[1,2], Pau Capera-Aragonès[3], John T Costner[1,2], Adelaide Robertson[1,2], Jennifer Seoyoung Kim[1,2], Anton S Petrov[1,2,4], Jessica C Bowman[1,2,4], Loren Dean Williams*[1,2,4], and Moran Frenkel-Pinter*[3,1]

[1]NASA Center for the Origins of Life, [2]School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332-0400 USA, [3]Institute of Chemistry, The Hebrew University of Jerusalem, Israel 91904, [4]NSF-NASA Center of Chemical Evolution

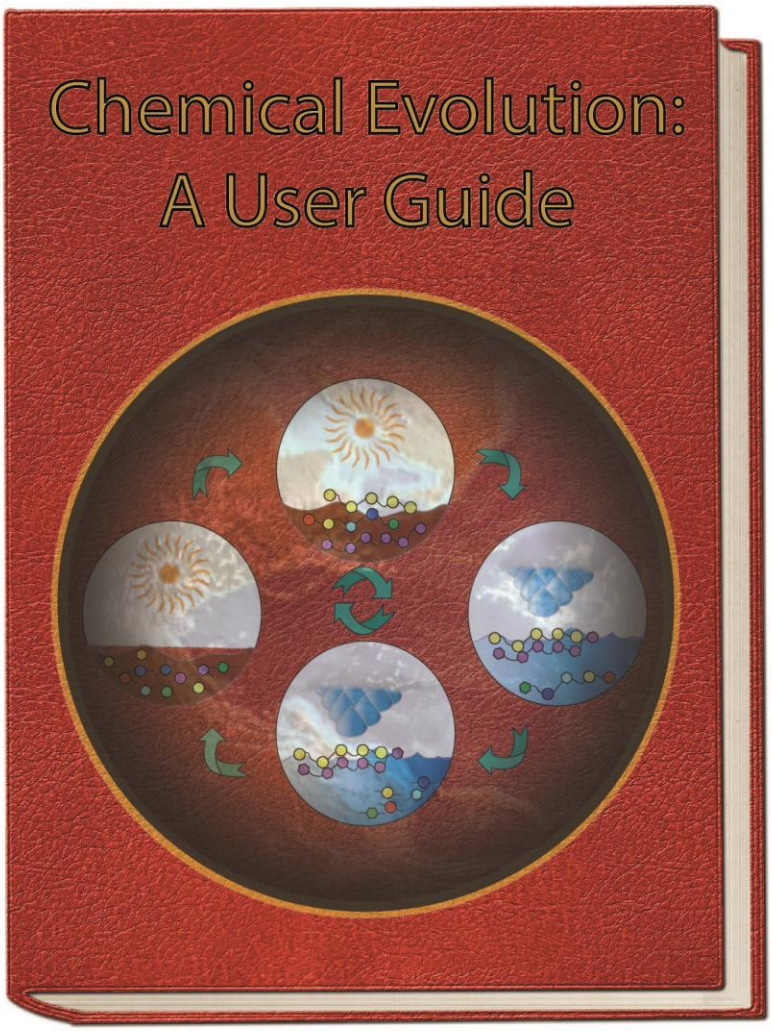
Keywords: Chemical Evolution; Molecular Evolution; Wet-Dry Cycling; Condensation Dehydration; Dynamic Combinatorial Chemistry.

*Corresponding authors:

Prof. Loren Dean Williams
School of Chemistry and Biochemistry
Georgia Institute of Technology
315 Ferst Drive NW
Atlanta, GA 30332-0400
Ph: (+1) (404) 385-6258
Fax: (+1) (404) 894-2295
loren.williams@chemistry.gatech.edu

Dr. Moran Frenkel-Pinter
The Center for Nanoscience and Nanotechnology
Institute of Chemistry
The Hebrew University of Jerusalem
Givat Ram Campus
Jerusalem 91904, Israel
Ph: (+972)-2-6584171
moran.fp@mail.huji.ac.il

TOC Figure



Abstract

Some of the most interesting open questions about the origins of life and molecular sciences center on chemical evolution and the spontaneous generation of complex and functional chemical species. The processes that generated the spectacular biopolymers that underlay biology demonstrate an untapped, by modern science, creative potential. We have established a robust, facile, and generally applicable platform for observing and analyzing chemical evolution using complex mixtures. While previous studies have characterized the formation of proto-polymers via chemical reactions, we systematically studied the process itself. We report empirical outcomes that were not foreseen or predicted. We have constructed an experimental platform to study the evolution of chemical systems that: (i) undergoes continuous recursive change with transitions to new chemical spaces while not converging throughout the course of the experiment, (ii) demonstrates chemical selection, during which combinatorial explosion is avoided, (iii) maintains synchronicity of molecular sub-populations, and (iv) harvests environmental energy that is stored in chemical energy. We have established some general guidelines for conducting chemical evolution. Our results suggest that chemical evolution can be adapted to produce a broad array of molecules with novel structures and functions.

Introduction

Around four billion years ago[1-3], prebiotic chemistry established the molecular keystones of biology, paving a path to life[4-6]. Chemical and geological processes on the ancient Earth[7-9] increased the complexity of organic molecules, ultimately creating RNA, DNA, protein, polysaccharides, membrane-forming amphiphiles, and the roots of biology. These processes harvested environmental energy and invested it in the formation of functional biopolymers. Prebiotic chemistry presents some of the most fascinating, important, and difficult questions in the field of chemical sciences[10-21]. Some of these questions center on the possibility that a form of chemical evolution preceded biological evolution[22,23].

Important advances in understanding chemical evolution have come from systems chemistry[24-34], dynamic combinatorial chemistry (DCC)[35,36], and models of self-organization[37]. Oscillatory networks of organic reactions are sustained by compositional heterogeneity, but not by homogeneity[24]. In DCC, monomers link and shuffle, and exchange between oligomers. DCC has achieved host–guest functionality[38,39], elaborate folds[40], and self-replication[41]. Mechanistic information on chemical evolution has been revealed by mutually catalytic systems[20] and reproducing catalytic micelles[42]. Auto-catalytic production of macrocycles has been used for selecting functional molecules[32,33].

Certain chemical systems, when subject to wet-dry cycling[43-46] (Fig. 1) can undergo striking changes in composition that appear relevant to the origins of life. Wet-dry cycling causes oscillations between linkage of molecules that can condense-dehydrate to form oligomers during the dried phase, and degradation via hydrolyze into smaller fragments during the wet phase.

Glucose oligomerizes during wet-dry cycling to form oligosaccharides[47]. Hydroxy acids oligomerize during wet-dry cycling to form polyesters[43,48-50]. Mixtures of amino acids and hydroxy acids or mercaptoacids oligomerize during wet-dry cycling to form depsipeptides, which contain mixtures of ester and amide bonds[15,21,51-57], and/or thiodepsipsipeptides[58], which contain mixtures of thioester and amide bonds. Depsipeptides formation under wet-dry cycling can be sequence-selective[53]. Mixtures of long chain alcohols and amino acids form amino acid ester amphiphiles[59]. Cycling a mixture of amino and hydroxy acids or a diverse organic set resulted in emergence of function (hydrolase activity)[46,54]. It has been observed that over wet-dry cycles, chemical spaces are progressively explored; first monomers are converted to esters (or thioesters), which are then enriched with amide bonds via ester (or thioester) aminolysis (Fig. 2).

Fig. 1. **Wet-dry cycling.** During dry conditions, molecules condense to form larger, more complex molecules. During wet conditions, complex molecules are partially hydrolyzed back to smaller, simpler molecules. Experimental wet-dry cycling mimics conditions on land surfaces of the earth.

Here we employ a model in which that chemistry is continuous with biology, and use that continuity to provide a working description of chemical evolution that maps principles of biological evolution onto chemical processes. During wet-dry cycling, we say that: (a) a *generation* is a single wet-dry cycle; (b) *heredity* is information passed from one generation to the next; (c) *information* is associated with non-random chemical composition; (d) *selection* is

preferential inheritance of certain molecular compositions; (e) *fitness* is persistence of molecules and specific molecular assemblies over generations; (f) *variation* is spatiotemporal differences in information; and (g) water is the "*energy currency*" that thermodynamically drives and links reactions.

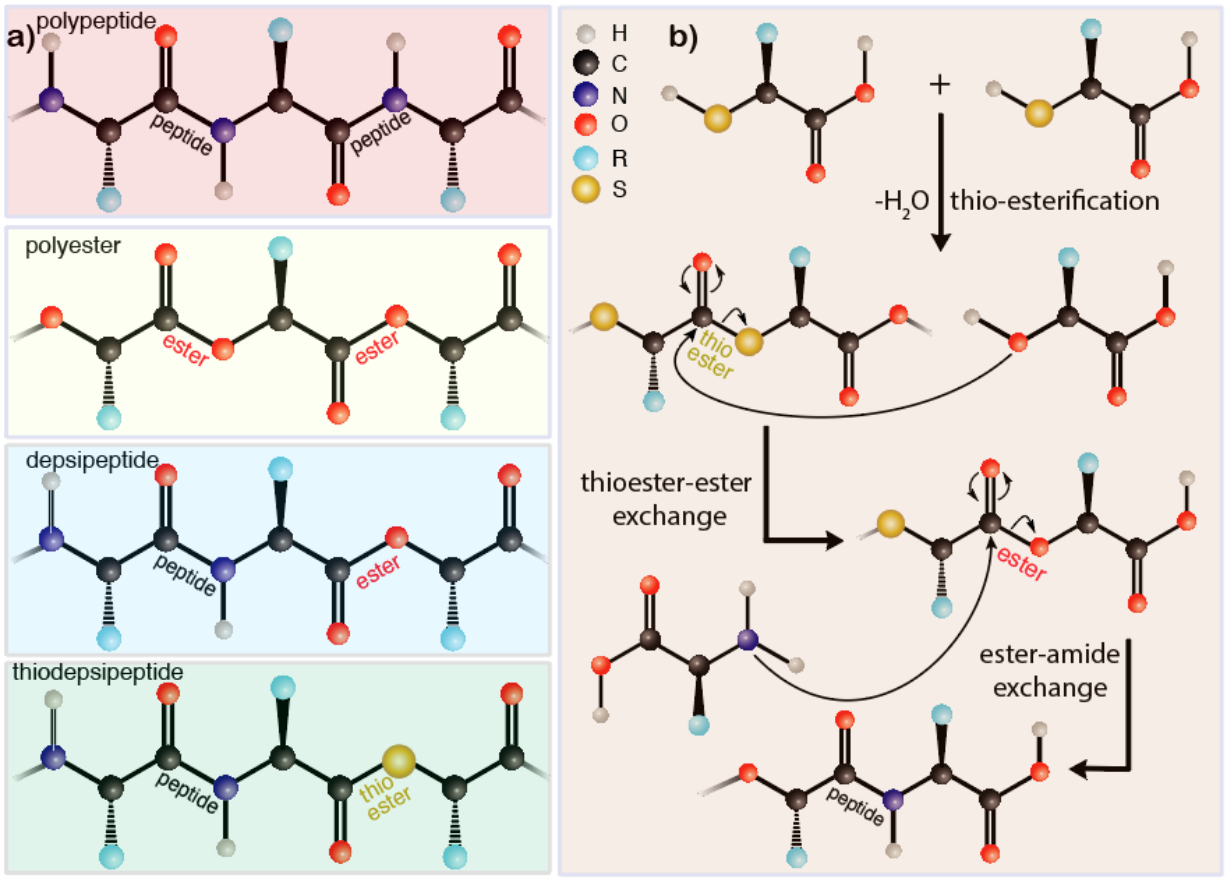


Fig. 2. a) Polypeptide, polyester, depsipeptide, and thiodepsipeptide. b) Progressive exploration of different chemical spaces during chemical evolution of mixtures containing hydroxy acids, mercaptoacids, and amino acids. Thioesters, formed by condensation dehydration reactions between mercaptoacids, are converted to ester by thioester-ester exchange with a hydroxy acid, which are then converted to amide via ester-amide exchange with an amino acid. This image was adapted.[15,58]

In our platform, drying of chemical mixtures generates chemical linkages. Subsequent wetting partially hydrolyzes the linkages. Our experiments: (i) use water as a chemical reactant, a chemical product, and the medium, (ii) cycle water activity to oscillate driving forces, (iii) use

complex mixtures of condensation-active molecules containing carboxylic acids, amines, thiols, and hydroxyl groups, and (iv) use near near-ambient temperatures. When monomers are condensed into oligomers by drying, and are then rehydrated but only partially hydrolyzed, environmental energy has been harvested and converted to chemical energy. Some products of condensation-dehydration, such as peptides, are kinetically trapped in aqueous media and can persist as high energy species during the wet phase. Thus, molecules are selected first on condensation chemistry, then on resistance to hydrolysis. It is possible that such processes, *ad infinitum*, contributed to the rise of biology from chemistry.

A primary goal of our study was to establish a platform for studying and understanding the process of chemical evolution rather than to target specific molecules. Our working definition of chemical evolution is continuous change with exploration of new chemical spaces and avoidance of an equilibrium condition. We did not target extant biopolymers and their functions and assume that replication[60-62] is not necessary for early chemical evolution. We observe novel outcomes from our experiments. Our systems chemically evolve throughout the course of the experiment, do not combinatorially explode, and demonstrate synchronized change in concentrations of sub-populations. Our results suggest that chemical evolution using complex mixtures can be studied and adapted to produce functional molecules.

Results

The experimental platform. Experiments were performed with a variety of initial mixtures containing 2-components, 3-components, 4-components, 5-components, 6-components, 9-components, or 25-components. The mixtures are nested, in that subset mixtures omit reactants from parent reaction mixtures but do not include reactants not found in the parent mixture. Specifically, each 2-, 3-, 4-, 5-, or 6-component mixture is a subset of the 9-component mixture, which is a subset of the 25-component mixture. The primary focus of this report is the 9-component mixture (Fig. 3, MFP Set 3). We consider the 9-component system to be diverse because it contains an alpha-hydroxy carboxylic acid, an amino acid, a mercaptoacid, urea, an aldohexose, a triol, an amphiphilic long chain carboxylic acid, a purine, and a hydrated divalent cation.

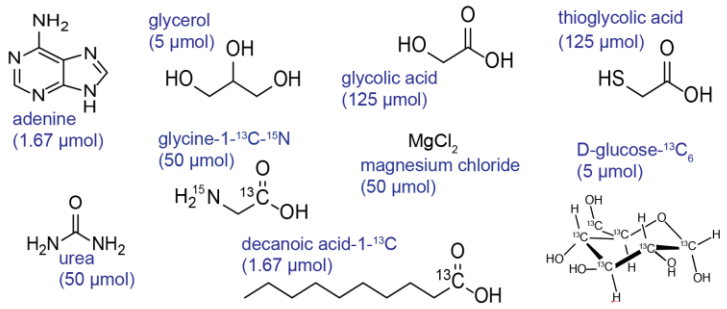


Fig. 3. **The nine-component mixture. Nine chemical components (called MFP Set 3) were used to initiate chemical evolution.** The components include glycolic acid, thioglycolic acid, glycine, glycerol, urea, glucose, decanoic acid, magnesium chloride, and adenine.

Both single-step dry-down reactions and wet-dry cycling reactions were performed. Reactions were conducted in anoxia at 45° except for experiments investigating temperature-

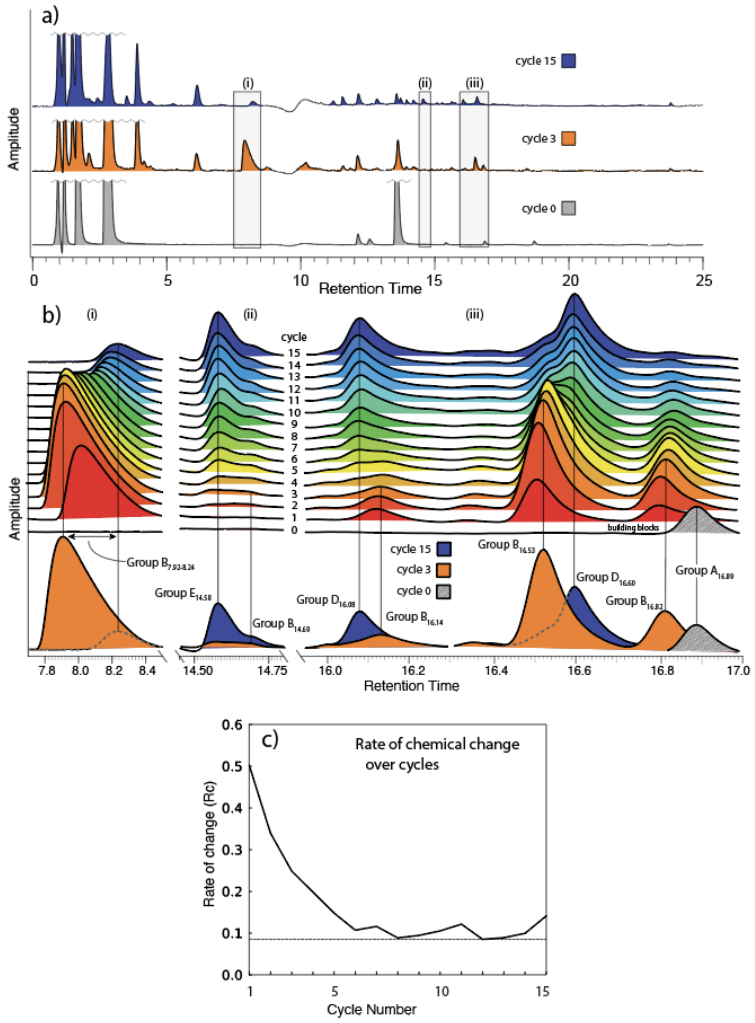


Fig. 4. a) HPLC chromatograms of cycles 0, 3, and 15, initiating with the 9-component mixture. Peaks in the shaded boxes (i), (ii), and (iii) are enlarged in panel b. b) An enlargement of select regions of the chromatogram showing all 15 cycles (top). A superimposition of a common regions of chromatograms of cycles 0, 3, and 15 is also shown. c) Rate of change per cycle (Rc) over 15 wet-dry cycles of the 9-component mixture. The dashed line shows the non-zero asymptote of the rate of evolution. Rc is defined in Eq. (1).

dependence. Wet-dry cycling experiments took 2 days per cycle and were not fed. Aliquots were taken after each cycle for analysis by C18 HPLC-UV-Vis, LC-MS, and NMR (Supplementary Information). The 15-cycle wet-dry cycling of the 9-component mixture was replicated five times (Figs. S1-S2). Chemical characterization is summarized in Table S1. Sections of HPLC chromatograms are shown in Fig. 4. Single-step dry-down reactions, with similar mixtures, were conducted for 3 days. HPLC chromatograms from single-step dry down reactions are shown in Fig. 5, 6, and Fig. S6, S7, S9.

Continuous chemical change.

The realization of continuous change is indicated by inspection of chromatograms and by quantitation of rate of chemical change (Fig. 4c). We define change by the net differences in concentrations between contiguous wet-dry cycles. The *Rate of change per cycle* (*Rc*) is:

(Equation 1) $$Rc = \frac{1}{N}\sum_{i=1}^{N}\frac{|P_i^{c+1} - P_i^c|}{max(P_i^{c+1}, P_i^c)}$$

where *N* is the number of chemical species, and P_i^c is the concentration of species *i* at cycle *c.* The absolute value in Eq 1 accounts for either increasing or decreasing concentration in an equal manner. The *max()* in the denominator is taken such that *R* is defined from '0' to '1', with '0' meaning no change and '1' meaning complete transformation (maximal change detected). Here the concentration of each species is estimated by peak height in HPLC chromatographs. This approximation does not account for differences in extinction coefficients. The experimental rate of chemical evolution over 15 cycles, estimated from equation 1, is shown in Fig. 4c. *Rc* was initially high and tended to a constant non-zero value after around 5 cycles. The system was evolving (changing) at an approximately constant rate from cycle 5 through 15. The chemical composition of the system changed during each of the 15 cycles. Every chromatogram is different from the flanking chromatographs. The system does not appear to converge after 15 cycles. The data show that the difference between cycle 14 and 15 was equivalent to differences between earlier pairs of contiguous cycles.

Combinatorial Compression. Complex or even relatively simple mixtures undergoing chemical transformations tend to combinatorically explode[30,63-65]. The formose reaction, for example, leads to a very large number of carbohydrate species[65,66].

We were surprised to observe that in our experiments, the number of products (above the limit of detection) does not scale with number of reactants. The number of products is not exponential or even additive with number of reactants (Fig. 5). By contrast, a modest increase in the temperature causes the number of products to explode (Fig 6). We use the phrase

184 ‘combinatorial compression’ to describe several interconnected phenomena. A combinatorially-
185 compressed reaction generates few product species from numerous diverse reactants. The
186 number of product species does not increase with the number of reactants. Moreover, addition
187 of reactants causes subtraction of products (Figure 5); initiation of a reaction with addition of
188 new reactants causes the appearance of new products and the disappearance (subtraction) of
189 some products (the latter were formed in reactions involving fewer reactants).

190 Combinatorial compression was observed by comparing single-step dry-down reactions
191 of a 9-component mixture and various 2-, 3-, 4- ,5-, and 6-component subsets of the 9-
192 component mixture (Fig. 5a-c and Fig. S6). For example, a total of 15 distinct primary product
193 peaks were observed in two independent 2-component mixtures (glucose with thioglycolic acid
194 or glycolic acid with glycerol). The 9-component mixture contained all four of these components
195 plus an additional five components but gives only 11 primary product peaks. Ten of the 15
196 product peaks observed in the two 2 component systems are absent (were subtracted) from the
197 9-component reaction.

198 A 2-component single-step dry-down reaction of glucose with thioglycolic acid gives seven
199 primary product peaks. Four of seven product peaks of these 2-component single-step dry-down
200 were subtracted from the 9-component single-step dry-down. Peaks 3-5 and 7 in Fig. 5a are
201 absent from Fig. 5c. The 2-component single-step dry-down of glycolic acid and glycerol gives

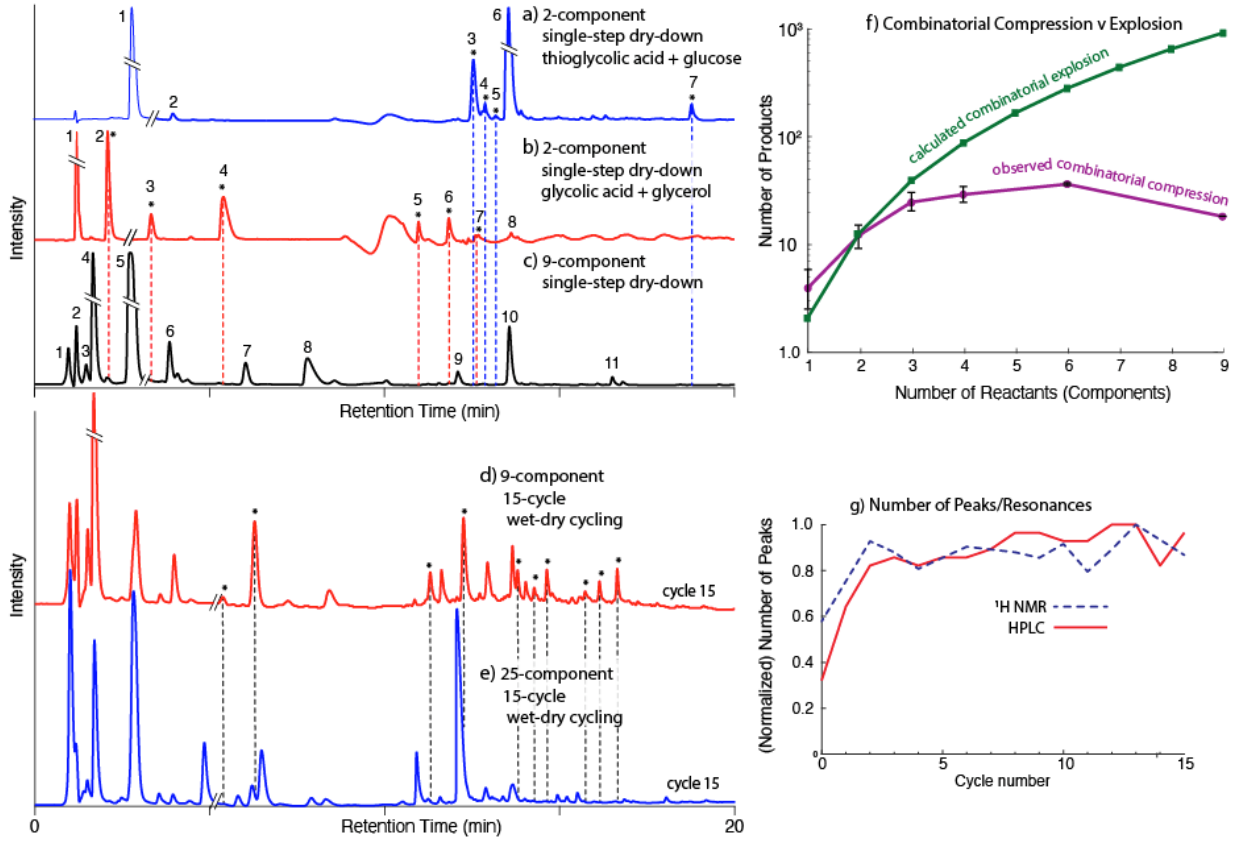


Fig. 5. **Combinatorial compression**. a) A single-step dry-down reaction of a 2-component mixture of thioglycolic acid and glucose reveals 7 primary product peaks. b) A single-step dry-down of a 2-component mixture of glycolic acid and glycerol reveals 8 primary product peaks. c) A single-step dry-down reaction of a 9-component mixture that includes thioglycolic acid, glucose, glycolic acid, and glycerol, plus additional components, reveals only 11 primary product peaks. Product peaks in the 2-component mixtures that are absent from the complex mixture are marked with asterisks. These single-step dry-down reactions were conducted for 72 hours and product mixtures were separated using C18-HPLC. d) A chromatograph of the 9-component system after 15 wet-dry cycles. Product mixtures were separated using C18-HPLC. e) A chromatograph of the 25-component system after 15 wet-dry cycles. Product mixtures were separated using C18-HPLC. Peaks observed in the 9-component system that are absent from the 25-component system are marked by asterisks. f) Observed and theoretical number of products from single-step dry-down reactions at 45°C. Thus far, we have data for 1 component, 2 component, 3 component, 4 component, 6 component, and 9 component experiments. The theoretical number of products was calculated for length < 4 assuming no branching with each permutation being distinct. We have measurements for multiple 2-component systems (error bars are SD). g) Normalized peak count in HPLC and ^{1}H-NMR analyses during wet-dry cycling chemical evolution experiment of the 9-component system. Peak count was normalized to the highest peak count across cycles, obtained at Cycle 13 for both HPLC and NMR data.

202 eight primary product peaks. In the 9-component single-step dry-down, six of eight 2 component

203 peaks were subtracted. Peaks 2-7 in Fig. 5b are absent from Fig. 5c. Product subtraction is

observed when comparing single-step dry-downs of the 9-component set with any 2 component subsets.

Combinatorial compression is also observed during wet-dry cycling. The 9-component MFP Set 3 mixture is a subset of the 25-component MFP Set 4 mixture. After 15 wet-dry cycles many product peaks of the 9-component are absent from the 25-component reaction (Fig. 5d-e). Ten of the 30 product peaks observed in the 9-component dry-down are absent (have been subtracted) from 25-component dry-down, which contained 34 product peaks in total. Only 20 product peaks are shared between the two sets of reactions.

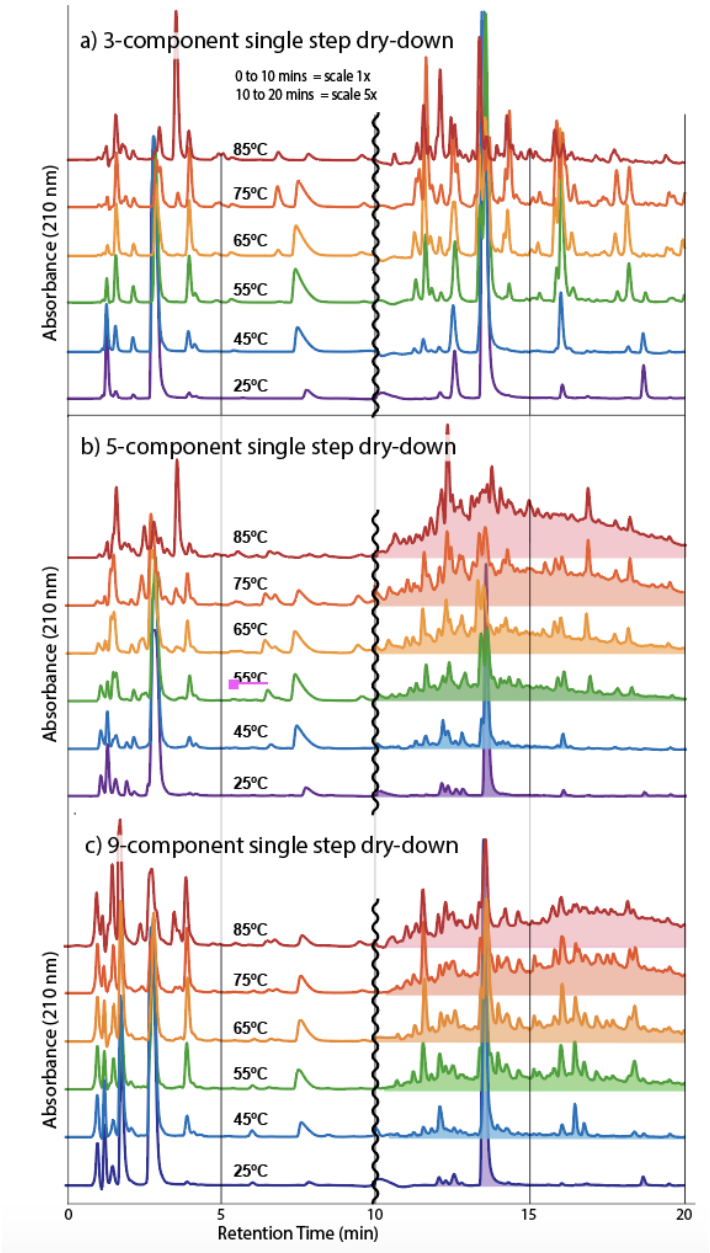


Fig. 6. Combinatorial compression is observed at low temperatures, whereas combinatorial explosion is observed at high temperatures. Single-step dry-down reactions of a 3-componnet system (a), a 5-component system (b), and a 9-component system (c). These nested mixtures were dried for 3 days under anoxia over a range of temperatures. a) Three components: thioglycolic acid, glycolic acid and urea. b) Five components: thioglycolic acid, glycolic acid, urea, glycine, and glucose. c) Nine-components (MFP Set 3). The number of peaks as well as a broad peak seen at high temperature signifies a large number of product species at high temperatures. The broad peak (at high temperature) moves to greater retention time (indicating more hydrophobic species) as the number of components increases.

The balance between combinatorial compression and explosion is temperature-dependent. Chromatograms of single-step dry-down reactions from 3-, 5- and 9-component systems over a range of temperature from 25°C to 85°C are shown in Fig. 6. The results are dramatic. The low temperature (≤45°C) systems are compressed, the number of product peaks remains relatively constant with increasing number of reactants. On the other hand, the high temperature systems (≥55°C) are exploded. At high temperatures, the 5-component and 9-component systems have broad peaks, with a general

increase in absorbance above the baseline, consistent with a large number of product species. The 9-component system has an even broader peak with a maximum at around 18 minutes. This increased breadth of the 9-component peak is consistent with an even greater number of product species than in the 5-component system.

Population Synchronicity. We define population synchronicity as the clustering of population trajectories over wet-cycling. Changes in peak intensities of HPLC chromatograms over 15 wet-dry cycles (Fig. 4) were used to infer changes in molecular populations. The population trajectories clustered[67,68] in well-defined groups A, B, C, D, and E (Fig. 7). Group A peaks correspond to starting components plus their immediate condensation products that form before the first dry-down cycle[57]. These peaks decrease rapidly during the early cycles. Group B peaks increase sharply in population during cycles 1-3 and fall during cycles 5-15. Group C peaks increases during cycles 1-5 and remains nearly constant during cycles 6-15. Group D peaks form in cycle 3, then steadily increases through cycle 15. Group E peaks remain nearly constant over 15 cycles. A dendrogram depicts similarity relationships between the groups (Fig. 7).

Analysis of Complex Chemical Systems. Complex chemical systems impose analytical challenges, as discussed by Mamajanov[30]. However, extensive information regarding the nature of the products and their changes over cycles here was gathered by combining various analytical techniques, including LCMS, HPLC, NMR, and UV absorbance (Table S1). The concordance of relative peak counts by ^{1}H NMR and HPLC (Fig. 5g) supports the utility of our peak counting approach for quantitating numbers of species. Our chromatography technique focused on hydrophobicity-based separation, in which resolution is poor for hydrophilic species. However,

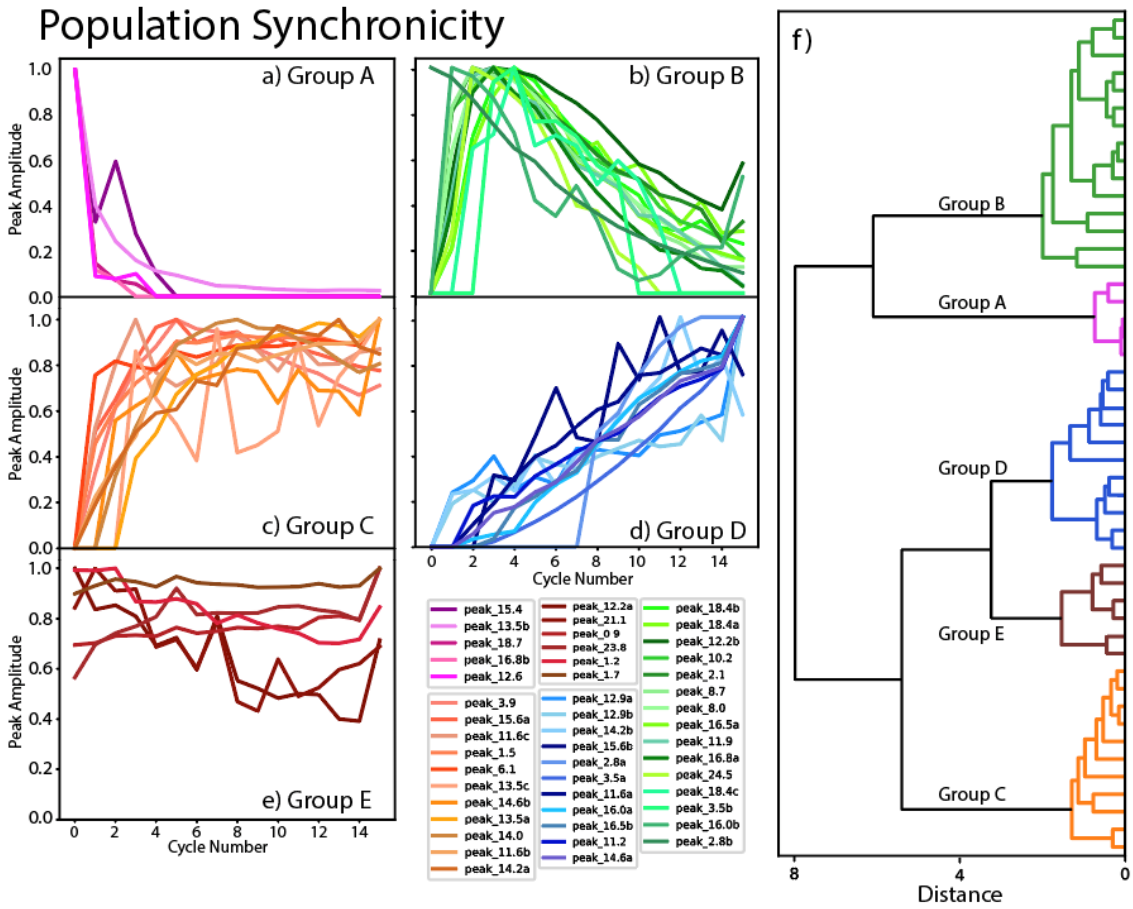


Fig. 7. Synchronous changes in chemical populations during wet-dry cycling. a) Group A molecules (pink) loose population during early cycles. b) Group B molecules (green) reach a maximum population near cycle 3 and fall over the next 12 cycles. c) Group C molecules (orange) increase in population during cycles 1 to 5 and thereafter remain constant. d) Group D molecules (blue) begin increasing in population at cycle 2 or 3 and increase until cycle 15. e) Group E molecules (brown) show minimal change in population over cycles. f) Dendrograms depicting the hierarchical clustering of the changes in peak integrals over cycles. Data were obtained by fitting of modified Gaussians to HPLC chromatograms. Peak assignments are given in the supplementary materials (Table S1).

due to the chemical reactions anticipated (e.g. formation of thioester, ester, and amide bonds), we expect that most of these product peaks will be well resolved in our analytical separation, an expectation that has been met based on peak shape and deconvolution analysis.

Some components of the 9-component system were much more reactive than others; various products were observed that contained linkages to glycolic acid, glycerol, and/or thioglycolic acid. UV absorbance spectra of various product peaks match spectra of known esters

252 and thioesters. On the other hand, it appeared that decanoic acid and adenine were not reactive
253 under our conditions.

254 Discussion

255 In this study we have established a platform for studying and understanding a process of
256 chemical evolution. Using complex mixtures and low temperatures (45 °C), we have begun to
257 characterize chemical evolution during wet-dry cycling reactions. The results suggest that
258 chemical systems can; (i) undergo continuous change, while avoiding equilibrium, (ii) exhibit
259 combinatorial compression and population synchronicity, (iii) transition to new chemical spaces,
260 from monomers, to thioesters and esters to amides, and (iv) harvest environmental energy in
261 kinetically trapped molecules. Thus far, chemical evolution in our experiments did not require
262 feeding; an initial ensemble of components underwent unceasing chemical change during
263 environmental cycling throughout the course of the experiment.

264 We hypothesize that species that are formed readily but are resistant to hydrolysis by
265 kinetic trapping are selected during wet-dry cycling; their populations selectively increase over
266 cycles. Thioesters and esters are made readily during the dry phase (low activation energies) and
267 hydrolyzed readily during the wet phase. Amides are formed via ester or thioester intermediates
268 during the dry phase, but are hydrolyzed very slowly during the wet phase due to high activation
269 energies. Wet-dry cycling can selectively grow populations of amides from esters or
270 thioesters[51,56,58,69]. Although it has not been demonstrated here, we hypothesize that more
271 prolonged wet-dry cycling will select for molecules that are recalcitrant, that gain persistence by

assembling[70]. It is known that assembled amphiphiles hydrolyze more slowly than isolated amphiphiles[71]. Assembled peptides hydrolyze more slowly than free peptides[72,73]. Folded RNA hydrolyzes more slowly than single-stranded RNA[74].

Our intent was to enable chemical evolution without steering the system to specific molecular targets. To study chemical evolution, we utilized the following elements:

Wet-Dry Cycling. Liquid water is both the medium of biology and the chemical nexus of biochemistry[75]. Water is by far the most frequent and abundant chemical reagent in biology. Between a third and a half of known biochemical reactions involve chemical consumption or production of water, and all universal biopolymers and most metabolites are produced by condensation-dehydration reactions. The centrality of water as both the medium and primary reactive species in biology is consistent with a model in which water governed chemical processes during the origins of life. No other substance known to science can play active roles as both a solvent with unique physical properties and as a hyperactive chemical reagent[75]. During wet-dry cycling, water is a shared reactant, a shared product, and the reaction medium. Water is an energy currency that can cause near-equilibrium reactions to oscillate in direction and to ratchet in energy and complexity.

Complex Libraries of Condensable Components. To explore chemical evolution, we used chemical components whose reactivities, in both thermodynamic and kinetic senses, are expected to be dependent on water activity. Components that can covalently link by reactions that absorb water and delink by reactions that release water are candidates for chemical

evolution here. Thus, component libraries used as inputs in our experiments contain combinations of hydroxyl groups, carboxylic acids, thiols, and amines.

Combinatorial Compression. Combinatorial compression appears to facilitate chemical evolution. The mild temperatures of our reactions account, in part, for differences in our results (compression) from many previous approaches (explosion). For complex mixtures at low temperature, only a subset of many possible products are selected. For thermodynamically-controlled reactions at 45°C, a small difference of around 1.5 kcal/mol in reaction free energies ($\Delta\Delta G_{rxn}$) gives 10-fold selectivity in terms of equilibrium concentration change[76]. Similarly, for kinetically-controlled reactions, a difference of around 1.5 kcal/mol in the activation free energies ($\Delta\Delta G^{‡}$) gives 10-fold selectivity in terms of reaction rates (Fig. 5). Selection of certain product species over others is based on these small differences in reaction free energies or activation energies. We have observed that the compression/explosion balance depends on temperature. As temperature rises, the systems tip towards explosion (Fig. 5c-e). As temperature decreases, the systems tip towards compression.

Population Synchronicity. Chemical evolution causes synchronicity of population trajectories; populations of groups of molecules rise and fall together during wet-dry cycling (Fig. 7). The observed dynamics of chemical evolution here appear similar in many respects to allele frequency trajectories in the long term evolution of *E. coli* by Lenski and Desai[77].

Ratcheting and Energy Harvesting. Wet-dry cycling causes oscillations between the linkage of components to form oligomers during the dried phase, and the partial depolymerization of oligomers into smaller fragments during the wet phase (Figs. 1-2). Chemical

evolution in our system harvests energy made accessible by environment cycling of water activity. In our system, environmental cycling causes chemical ratcheting, in which a system advances toward an equilibrium state, but without reaching equilibrium, is redirected by the environmental cycle toward a different equilibrium state, then is redirected again, and again. Constant oscillations in the system can be sufficient to keep the system always frustrated and prevent it from equilibrate. The system moves into progressively deeper kinetic traps. It is possible that such processes ad infinitum led to the rise of biology from chemistry.

Selection and Fitness. The prologue to biology was a chemical process that generated complex organic molecules that persist under environmental conditions. We propose that wet-dry cycling can facilitate production of progressively sophisticated methods to enable persistence. During wet-dry cycling, molecules that readily hydrolyze have lower fitness, and do not persist. Chemical transformation, as in ester-amide exchange or thioester-amide exchange, is a mechanism of persistence. All biopolymers are kinetically trapped, and thermodynamically unstable in water. Assembly, a second mechanism of persistence, is also utilized by all biopolymers[11,56,71-74,78-87]. Replication, a third mechanism, enables molecules to persist far beyond their chemical lifetimes. Replication has enabled the ribosome to persist for nearly 4 billion years. As predicted for products of chemical evolution as outlined here, the ribosome is a hydrolytically resistant assembly of kinetically trapped amides and phosphodiesters.

Future Prospects. Our focus here is early-stage chemical evolution, rather than the production of highly evolved biopolymers such as RNA or protein. As noted by François Jacob, "the really creative part in biochemistry must have occurred very early"[4]. We do not currently

understand the limits of chemical evolution. We do not yet know if chemical evolution can lead to proto-biological structures and functions. What happens after 150 cycles or 15,000 cycles? What happens if we feed the system? In the next series of experiments, currently in progress, these issues will be explored.

Funding

This research was supported by the National Science Foundation [1724274], the Azrieli Foundation Early Career Faculty Grant (to MFP), the Israel Science Foundation grant (1611/22, to MFP), and FEBS Foundation Excellence Award (to MFP).

Acknowledgements

We thank Drs. Nicholas Hud, M.G. Finn, Gary Schuster, Michael Travisano, and Eric Smith for helpful comments. All data generated or analyzed during this study are included in this published article and the supplementary information.

References

1 Bell, E. A., Boehnke, P., Harrison, T. M. & Mao, W. L. Potentially biogenic carbon preserved in a 4.1 billion-year-old zircon. *Proc. Natl. Acad. Sci. USA* **112**, 14518-14521, (2015).

2 Knoll, A. H. & Nowak, M. A. The timetable of evolution. *Science advances* **3**, e1603076, (2017).

3 Benton, M. J. & Harper, D. A. *Introduction to paleobiology and the fossil record*. (John Wiley & Sons, 2020).

4 Jacob, F. Evolution and tinkering. *Science* **196**, 1161-1166, (1977).

5 Woese, C. R. On the evolution of cells. *Proc. Natl. Acad. Sci. USA* **99**, 8742-8747, (2002).

6 Orgel, L. E. The origin of life on the earth. *Sci. Am.* **271**, 76-83, (1994).

7 Trainer, M. G. *et al.* Organic haze on Titan and the early Earth. *Proc. Natl. Acad. Sci. USA* **103**, 18035-18042, (2006).

8 Reed, N. W., Wing, B. A., Tolbert, M. A. & Browne, E. C. Trace H2S Promotes Organic Aerosol Production and Organosulfur Compound Formation in Archean Analog Haze Photochemistry Experiments. *Geophys. Res. Lett.* **49**, e2021GL097032, (2022).

9 Zahnle, K. J., Lupu, R., Catling, D. C. & Wogan, N. Creation and evolution of impact-generated reduced atmospheres of early Earth. *The Planetary Science Journal* **1**, 11, (2020).

10 Hud, N. V., Cafferty, B. J., Krishnamurthy, R. & Williams, L. D. The origin of RNA and "my grandfather's axe". *Chem. Biol.* **20**, 466-474, (2013).

11 Brack, A. Selective emergence and survival of early polypeptides in water. *Orig. Life Evol. Biosph.* **17**, 367-379, (1987).

12 Schneider, C. *et al.* Noncanonical RNA nucleosides as molecular fossils of an early Earth—Generation by prebiotic methylations and carbamoylations. *Angew. Chem. Int. Ed.* **57**, 5943-5946, (2018).

13 Deamer, D. The role of lipid membranes in life's origin. *Life* **7**, 5, (2017).

14 Kim, S. C. *et al.* A model for the emergence of RNA from a prebiotically plausible mixture of ribonucleotides, arabinonucleotides, and 2'-deoxynucleotides. *J. Am. Chem. Soc.* **142**, 2317-2326, (2020).

15 Frenkel-Pinter, M. *et al.* Selective incorporation of proteinaceous over nonproteinaceous cationic amino acids in model prebiotic oligomerization reactions. *Proc. Natl. Acad. Sci. USA* **116**, 16338-16346, (2019).

16 Runnels, C. M. *et al.* Folding, assembly, and persistence: the essential nature and origins of biopolymers. *J Mol Evol* **86**, 598-610, (2018).

17 Lemmon, R. M. Chemical evolution. *Chem. Rev.* **70**, 95-109, (1970).

18 Cleaves II, H. J. The origin of the biologically coded amino acids. *J. Theor. Biol.* **263**, 490-498, (2010).

19 Eschenmoser, A. Etiology of potentially primordial biomolecular structures: from vitamin B12 to the nucleic acids and an inquiry into the chemistry of life's origin: a retrospective. *Angew. Chem. Int. Ed.* **50**, 12412-12472, (2011).

20 Vincent, L. *et al.* Chemical ecosystem selection on mineral surfaces reveals long-term dynamics consistent with the spontaneous emergence of mutual catalysis. *Life* **9**, 80, (2019).

21 Rodriguez-Garcia, M. *et al.* Formation of oligopeptides in high yield under simple programmable conditions. *Nat. Commun.* **6**, 1-7, (2015).

22 Gould, S. J. *The structure of evolutionary theory*. (Harvard University Press, 2002).

23 Gould, S. J. & Vrba, E. S. Exaptation—a missing term in the science of form. *Paleobiology* **8**, 4-15, (1982).

24 Cafferty, B. J. *et al.* Robustness, entrainment, and hybridization in dissipative molecular networks, and the origin of life. *J. Am. Chem. Soc.* **141**, 8289-8295, (2019).

25 Segré, D., Ben-Eli, D. & Lancet, D. Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies. *Proc. Natl. Acad. Sci. USA* **97**, 4112-4117, (2000).

400 26 Ruiz-Mirazo, K., Briones, C. & de la Escosura, A. Prebiotic systems chemistry: new
401 perspectives for the origins of life. *Chem. Rev.* **114**, 285-366, (2014).

402 27 Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. Common origins of
403 RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **7**, 301-
404 307, (2015).

405 28 Kroiss, D., Ashkenasy, G., Braunschweig, A. B., Tuttle, T. & Ulijn, R. V. Catalyst: can systems
406 chemistry unravel the mysteries of the chemical origins of life? *Chem* **5**, 1917-1920,
407 (2019).

408 29 Ashkenasy, G., Hermans, T. M., Otto, S. & Taylor, A. F. Systems chemistry. *Chemical*
409 *Society Reviews* **46**, 2543-2554, (2017).

410 30 Guttenberg, N., Virgo, N., Chandru, K., Scharf, C. & Mamajanov, I. Bulk measurements of
411 messy chemistries are needed for a theory of the origins of life. *Philos. Trans. R. Soc.*
412 *London, Ser. A* **375**, 20160347, (2017).

413 31 Islam, S. & Powner, M. W. Prebiotic systems chemistry: Complexity overcoming clutter.
414 *Chem* **2**, 470-501, (2017).

415 32 Wołos, A. *et al.* Synthetic connectivity, emergence, and self-regeneration in the network
416 of prebiotic chemistry. *Science* **369**, (2020).

417 33 Miao, X. *et al.* Kinetic selection in the out-of-equilibrium autocatalytic reaction networks
418 that produce macrocyclic peptides. *Angew. Chem.* **133**, 20529-20538, (2021).

419 34 Semenov, S. N. *et al.* Autocatalytic, bistable, oscillatory networks of biologically relevant
420 organic reactions. *Nature* **537**, 656-660, (2016).

421 35 Cao, Y. *et al.* Self-Synthesizing Nanorods from Dynamic Combinatorial Libraries against
422 Drug Resistant Cancer. *Angew. Chem.* **133**, 3099-3107, (2021).

423 36 Komáromy, D., Nowak, P. & Otto, S. Dynamic combinatorial libraries. *Dynamic covalent*
424 *chemistry: principles, reactions, and applications*, 31-119, (2017).

425 37 Chvykov, P. *et al.* Low rattling: A predictive principle for self-organization in active
426 collectives. *Science* **371**, 90-95, (2021).

427 38 Otto, S., Furlan, R. L. & Sanders, J. K. Selection and amplification of hosts from dynamic
428 combinatorial libraries of macrocyclic disulfides. *Science* **297**, 590-593, (2002).

429 39 Lam, R. T. *et al.* Amplification of acetylcholine-binding catenanes from dynamic
430 combinatorial libraries. *Science* **308**, 667-669, (2005).

431 40 Liu, B. *et al.* Complex molecules that fold like proteins can emerge spontaneously. *J. Am.*
432 *Chem. Soc.* **141**, 1685-1689, (2018).

433 41 Carnall, J. M. *et al.* Mechanosensitive self-replication driven by self-organization. *Science*
434 **327**, 1502-1506, (2010).

435 42 Kahana, A. & Lancet, D. Self-reproducing catalytic micelles as nanoscopic protocell
436 precursors. *Nature Reviews Chemistry* **5**, 870-878, (2021).

437 43 Mamajanov, I. *et al.* Ester formation and hydrolysis during wet–dry cycles: generation of
438 far-from-equilibrium polymers in a model prebiotic reaction. *Macromolecules* **47**, 1334-
439 1343, (2014).

440 44 Ross, D. & Deamer, D. Dry/wet cycling and the thermodynamics and kinetics of prebiotic
441 polymer synthesis. *Life* **6**, 28, (2016).

442 45 Damer, B. & Deamer, D. The hot spring hypothesis for an origin of life. *Astrobiology* **20**,
443 429-452, (2020).

444 46 Foster, K., Hillman, B., Rajaei, V., Seng, K. & Maurer, S. Evolution of Realistic Organic
445 Mixtures for the Origins of Life through Wet–Dry Cycling. *Sci* **4**, 22, (2022).

446 47 Li, Z. *et al.* The oligomerization of glucose under plausible prebiotic conditions. *Orig. Life*
447 *Evol. Biosph.* **49**, 225-240, (2019).

448 48 Jia, T. Z. *et al.* Incorporation of basic α-hydroxy acid residues into primitive polyester
449 microdroplets for RNA segregation. *Biomacromolecules* **22**, 1484-1493, (2021).

450 49 Jia, T. Z. & Chandru, K. Recent progress in primitive polyester synthesis and membraneless
451 microdroplet assembly. *Biophysics and Physicobiology* **20**, e200012, (2023).

452 50 Jia, T. Z. *et al.* Membraneless polyester microdroplets as primordial compartments at the
453 origins of life. *Proc. Natl. Acad. Sci. USA* **116**, 15830-15835, (2019).

51 Forsythe, J. G. *et al.* Ester-mediated amide bond formation driven by wet-dry cycles: a possible path to polypeptides on the prebiotic earth. *Angew. Chem. Int. Ed.* **54**, 9871-9875, (2015).

52 Yu, S. *et al.* Elongation of model prebiotic proto-peptides by continuous monomer feeding. *Macromolecules* **50**, 9286–9294, (2017).

53 Forsythe, J. G. *et al.* Surveying the sequence diversity of model prebiotic peptides by mass spectrometry. *Proc. Natl. Acad. Sci. USA* **114**, E7652-E7659, (2017).

54 Doran, D., Abul-Haija, Y. M. & Cronin, L. Emergence of function and selection from recursively programmed polymerisation reactions in mineral environments. *Angew. Chem. Int. Ed.* **58**, 11253-11256, (2019).

55 Bouza, M. *et al.* Compositional Characterization of Complex Protopeptide Libraries via Triboelectric Nanogenerator Orbitrap Mass Spectrometry. *Rapid Commun. Mass Spectrom.*, (2019).

56 Frenkel-Pinter, M. *et al.* Mutually stabilizing interactions between proto-peptides and RNA. *Nat. Commun.* **11**, 1-14, (2020).

57 Frenkel-Pinter, M., Sargon, A. B., Glass, J. B., Hud, N. V. & Williams, L. D. Transition metals enhance prebiotic depsipeptide oligomerization reactions involving histidine. *RSC Advances* **11**, 3534-3538, (2021).

58 Frenkel-Pinter, M. *et al.* Thioesters provide a plausible prebiotic path to proto-peptides. *Nat. Commun.* **13**, 1-8, (2022).

59 Lago, I., Black, L., Wilfinger, M. & Maurer, S. E. Synthesis and Characterization of Amino Acid Decyl Esters as Early Membranes for the Origins of Life. *Membranes* **12**, 858, (2022).

60 Gilbert, W. Origin of life: The RNA world. *Nature* **319**, 618, (1986).

61 Benner, S. A., Ellington, A. D. & Tauer, A. Modern metabolism as a palimpsest of the RNA world. *Proc. Natl. Acad. Sci. USA* **86**, 7054-7058, (1989).

62 Joyce, G. F. & Szostak, J. W. Protocells and RNA self-replication. *Cold Spring Harb. Perspect. Biol.* **10**, a034801, (2018).

63 Asche, S., Cooper, G. J. T., Keenan, G., Mathis, C. & Cronin, L. A robotic prebiotic chemist probes long term reactions of complexifying mixtures. *Nat. Commun.* **12**, 3547, (2021).

64 Marshall, S. M. *et al.* Identifying molecules as biosignatures with assembly theory and mass spectrometry. *Nat. Commun.* **12**, 3033, (2021).

65 Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit Rev Biochem Mol Biol* **39**, 99-123, (2004).

66 Colón-Santos, S., Cooper, G. J. & Cronin, L. Taming combinatorial explosion of the formose reaction via recursion within mineral environments. *ChemSystemsChem* **1**, (2019).

67 van Duppen, P., Daines, E., Robinson, W. E. & Huck, W. T. Dynamic Environmental Conditions Affect the Composition of a Model Prebiotic Reaction Network. *J. Am. Chem. Soc.* **145**, 7559-7568, (2023).

68 Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **17**, 261-272, (2020).

69 Yu, S. S. *et al.* Kinetics of prebiotic depsipeptide formation from the ester-amide exchange reaction. *Physical Chemistry Chemical Physics* **18**, 28441-28450, (2016).

70 Edri, R., Fisher, S., Menor-Salvan, C., Williams, L. D. & Frenkel-Pinter, M. Assembly-Driven Protection from Hydrolysis as Key Selective Force during Chemical Evolution. *FEBS Lett.*, (2023).

71 Kensil, C. R. & Dennis, E. A. Alkaline hydrolysis of phospholipids in model membranes and the dependence on their state of aggregation. *Biochemistry* **20**, 6079-6085, (1981).

72 Abkevich, V. I., Gutin, A. M. & Shakhnovich, E. I. How the first biopolymers could have evolved. *Proc. Natl. Acad. Sci. USA* **93**, 839-844, (1996).

73 Prusiner, S. B. Prions. *Proc. Natl. Acad. Sci. USA* **95**, 13363-13383, (1998).

74 Nahvi, A. *et al.* Genetic control by a metabolite binding mRNA. *Chem. Biol.* **9**, 1043-1049, (2002).

75 Frenkel-Pinter, M., Rajaei, V., Glass, J. B., Hud, N. V. & Williams, L. D. Water and Life: The Medium is the Message. *J. Mol. Evol.*, 1-10, (2021).

76 Lowry, T. & Richardson, K. (Harper & Row, New York, 1981).

77 Good, B. H., McDonald, M. J., Barrick, J. E., Lenski, R. E. & Desai, M. M. The dynamics of molecular evolution over 60,000 generations. *Nature* **551**, 45-50, (2017).

78 Guth-Metzler, R. *et al.* Goldilocks and RNA: where Mg^{2+} concentration is just right. *Nucleic Acids Res.* **51**, 3529-3539, (2023).

79 Habibi, Y., Lucia, L. A. & Rojas, O. J. Cellulose nanocrystals: chemistry, self-assembly, and applications. *Chem. Rev.* **110**, 3479-3500, (2010).

80 Hattori, S. *et al.* Preservation of collagen in the soft tissues of frozen mammoths. *PLoS One* **16**, e0258699, (2021).

81 Hud, N. V. & Anet, F. A. Intercalation-mediated synthesis and replication: a new approach to the origin of life. *J Theor Biol* **205**, 543-562, (2000).

82 Jain, S. S. & Tullius, T. D. Footprinting protein–DNA complexes using the hydroxyl radical. *Nat. Protoc.* **3**, 1092-1100, (2008).

83 McKinley, M. P., Bolton, D. C. & Prusiner, S. B. A protease-resistant protein is a structural component of the scrapie prion. *Cell* **35**, 57-62, (1983).

84 Peters, J. W. & Williams, L. D. The origin of life: look up and look down. *Astrobiology* **12**, 1087-1092, (2012).

85 Reynaud, C. *et al.* In-place molecular preservation of cellulose in 5,000-year-old archaeological textiles. *Proc. Natl. Acad. Sci. USA* **117**, 19670-19676, (2020).

86 Shoulders, M. D. & Raines, R. T. Collagen structure and stability. *Annu. Rev. Biochem* **78**, 929-958, (2009).

87 Ulijn, R. V., Baragaña, B., Halling, P. J. & Flitsch, S. L. Protease-catalyzed peptide synthesis on solid support. *J. Am. Chem. Soc.* **124**, 10988-10989, (2002).