Environmental control programs the emergence of distinct product ensembles from unconstrained chemical reaction networks

Andrew J. Surman,[#1] Marc Rodriguez Garcia,[#1] Yousef M. Abul-Haija,[1] Geoffrey J. T. Cooper,[1] Piotr S. Gromski,[1] Rebecca Turk-MacLeod,[1] Margaret Mullin,[2] Cole Mathis,[3] Sara I. Walker,[3] and Leroy Cronin*[1]

[1]*WestCHEM, School of Chemistry, University of Glasgow, Glasgow, UK.*

[2]*School of Life Sciences, MVLS, University of Glasgow, Glasgow, UK.*

[3]*Beyond Center for Fundamental Concepts in Science, Arizona State University, Tempe, AZ, USA.*

Subject terms: Systems Chemistry; Chemomics; Peptides; Origin of Life; Combinatorial.

Harnessing and then controlling combinatorial explosions in uncontrolled condensation reactions of simple building blocks is a key problem to many hypotheses on the origin of life. Much has been achieved in understanding how the building blocks of biopolymers may be formed, and in understanding how macromolecules may produce functional and increasingly life-like systems. How these steps can be joined, and how defined populations of macromolecules can form from mixtures of simple building blocks, instead of an undifferentiated mess, remain open questions. Herein, we show how unconstrained condensation reactions of both amino acids, and prebiotic soup mixtures produced by spark discharge, can be steered by changes in the reaction environment, such as order of reactant addition (mixing history), and addition of salts or minerals. Using techniques akin to untargeted metabolomics to survey product distributions we demonstrate that while these reactions do produce a large range of species, there are distinct, significant, and reproducible differences between the product ensembles. Furthermore, we observe that differences in composition are demonstrated through clearly different structural and functional properties. Using this approach, we demonstrate for the first time that simple variations in environmental parameters can mediate the differentiation of distinct ensembles from both amino acid mixtures and a classic primordial soup model (products of a 'Miller Urey' type spark discharge reaction). This shows that the synthetic complexity produced by such unconstrained reactions is not as intractable as often suggested, when viewed through a chemically agnostic lens. An open approach to complexity can generate compositional, structural, and functional diversity from fixed sets of simple starting materials, suggesting that differentiation of product mixtures can occur in the wider environment without the need for biological machinery.

Modern synthetic chemistry takes a closed approach to complexity, with a focus on making single molecular targets in high yield, purity, and selectivity. Meanwhile, the exploration of complex mixtures or systems is focussed on those formed within, or by, biology,[1,2] since biology imposes boundary conditions on molecular diversity[3] which abiotic chemistry lacks. As researchers interested in how functional ordered chemical systems might be produced from an inorganic world, to ultimately form biological/life-like systems,[4,5] we cannot avoid heterogeneity.[3,6-9] In recent decades, however, most chemists researching life-like systems[10] have moved from exploring high-energy unconstrained 'soup' reactions,[11,12] to examining the intricate mechanisms required for abiotic synthesis of nucleotides,[13] polynucleotides,[14-16] and peptides,[17-20] and on towards the assembly of protocells,[21-23] enzyme-mediated systems,[24] and exploration of autocatalysis.[25,26] This transition arose from the expectation that unconstrained multicomponent reactions would undergo combinatorial explosion.[3] Without some means of control, this would result in analytically intractable, undifferentiated mixtures in which any specific functional molecules would be vanishingly dilute, with no mechanism for the emergence of distinct functional systems or structures.[1,7,8]

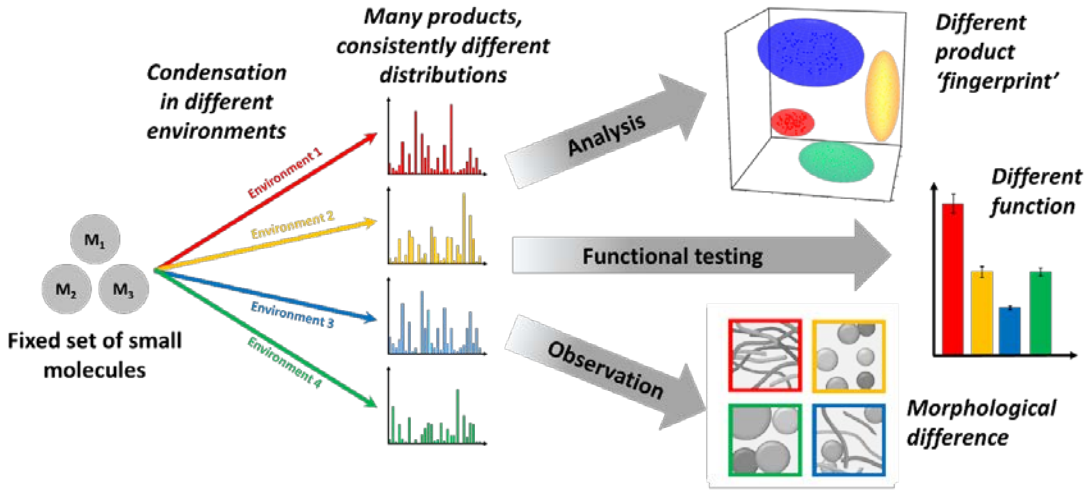

Fig. 1. Concept: uncontrolled condensation reactions make a mess, but they can be steered. *Reactions where multifunctional building blocks yield combinatorial explosions may be steered by different environmental conditions to consistently yield different product distributions. These different product ensembles can be shown to have consistently different structural and functional properties.*

In recent work, we[9] and others[27,28] have begun to take a more open approach to complex mixtures; instead of avoiding complexity, we embrace it, and use modern analytical tools to observe otherwise-hidden patterns in complex synthetic systems. Here, we hypothesise that while unconstrained multicomponent reactions do produce a mess, they may be steered to different areas of product space. Here we show that performing a reaction of the same starting materials but under different environmental conditions will consistently produce different product ensembles (Figure 1). These can lead to the emergence of distinct order, structure, and function, 'programmed' by the environment, and

challenge the view that a complexity-first approach, not targeting specific product molecules, will only produce intractable tar.[7]

Initially, we chose activation-free dehydration-driven amino acid (AA) condensation as a model system to explore these ideas. Such reactions can be carried out simply by heating aqueous AA solutions to remove water and drive peptide-forming dehydration reactions,[12,18] but can potentially produce diverse peptide oligomers with a range of structural and functional properties. In preliminary work, to confirm cross reactivity between a range of AAs, we noticed that even small changes, like the presence of different soluble salts, could alter the distribution of products, both in the amount of peptide bonds formed, and the product distribution. This implies that product distribution is neither random,[29] nor completely determined by simple thermodynamic considerations, but rather subject to environmental control.

To test our key hypothesis that environmental programming can produce significantly different product distributions from a fixed set of AAs, we focussed on condensation of glycine (G), alanine (A) and histidine (H). All undergo homo- and cross-oligomerisation with different degrees of reactivity, and their varied incorporation into peptides might be expected to lead to different functional and structural properties. To assess extended structure formation, alanine (A), aspartic acid (D) and valine (V) were specifically chosen for their potential to form oligomers with hydrophobic and hydrophilic blocks, thus increasing the chance of constructing interesting structural motifs. Earlier studies have found the kind of complex mixtures these reactions produce to be analytically intractable, since robust identification and quantification of the many thousands of potential oligomer products is not feasible (e.g. combination of three AAs in oligomers up to ten residues long potentially produces 59049 product permutations). Instead of attempting to identify all products, we have developed a chemomics 'fingerprinting' approach to observe the resulting product differentiation, mirroring the approach of untargeted metabolomics studies in biological systems. Our workflow starts with liquid chromatography coupled with high resolution mass spectrometry (LC-MS), which provides a powerful multi-dimensional means to sensitively resolve large numbers of species.[27,28] Scripted data analysis then allowed automated peak picking to identify features in the LC-MS data (identified by *m/z* and retention time coordinates, and characterised by intensity values for each sample). Finally, dimensionality-reduction approaches to represent data for comparative inspection (principal component analysis (PCA), and principal component differential function analysis (PC-DFA)), allowed us to extract useful observations from the large volumes of data produced, without any attempt to assign molecular structures to features (see SI section 2.2 for full details). This approach follows the example of untargeted metabolomics, rather than proteomics methods, since we also intended to address systems in which products are not restricted to peptides.

We chose three types of environmental condition to vary: i) the presence of soluble salts, ii) the presence of minerals, and iii) the mixing history (the order of precursor addition over multiple reaction cycles). Addition of salts and minerals are both known to interact with AAs in a variety of ways, causing either catalysis, complexation, sequestration, degradation, and/or templating.[30-34] Minerals chosen were alumina, montmorillonite, mica, goethite, quartz, natrolite and silica, while the soluble salts were NaCl, KCl, LiCl, NH_4Cl, $MgCl_2$, $CuCl_2$ and $EuCl_3$. A solution containing equimolar amounts of the three AAs was added to these minerals or soluble salts under successive dehydration-hydration cycles (130 °C for 12 h at pH 2.5). Samples were then dialysed (500-1000 Da cut-off) to remove small species and soluble salts before analysis. Environmental contributions need not be limited to the material additions, or parameters such as temperature; the history of the material and the order of precursor combination/mixing also has a role.[13] To explore this concept, we performed a series of reactions with multiple dehydration/hydration cycles, in which the monomers (G, A and H) were added to the reaction in different orders with dehydration cycles between each addition.

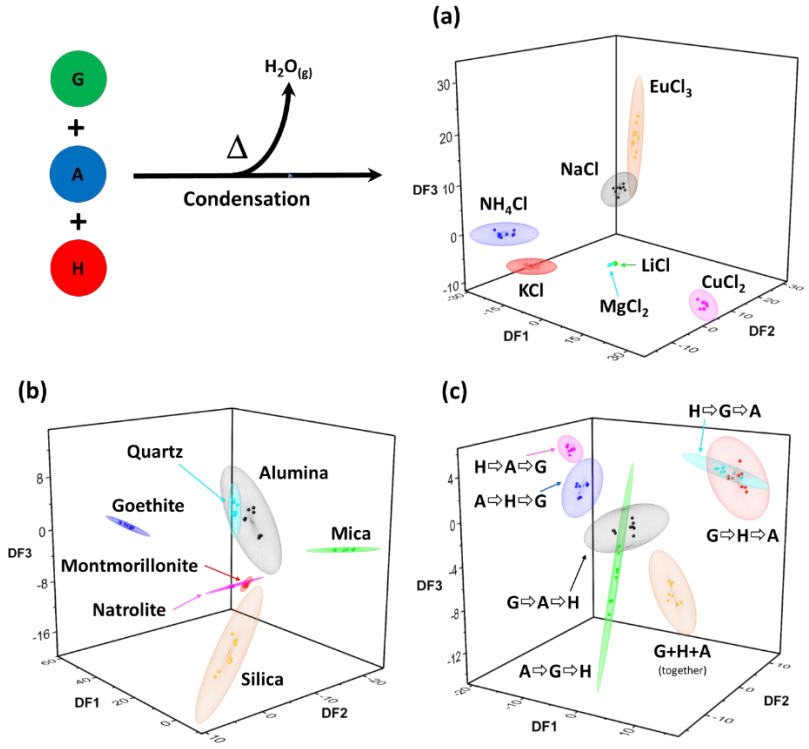

Figure 2. PC-DFA analysis of LC-MS data from condensation of G, A, & H in different environments/conditions: *(a) different soluble salts, (b) different minerals, and (c) different mixing orders. Points represent individual measurements (9 measurements: 3 experimental replicates x 3 analytical replicates), and shaded bubbles represent a two standard deviation space around their mean; see SI section 2.2 for full details and other representations. Plots generated with Origin Pro 2016 (OriginLab).*

Remarkably, we found that all three variations to the environmental conditions led to differentiation of consistently distinct product ensembles in terms of peak distribution and intensities from LC-MS analysis (Figure 2). We used a peak-picking algorithm to define 'features' in LC-MS chromatograms, resulting in hundreds to thousands of features for each data set. Multivariate analysis then allowed us to compare the intensities of these features across the respective environmental parameters, resulting in a 3D interpretation of the uniqueness of chemical compositions between environments. Notably, compositions produced in different environments do not overlap, indicating their uniqueness, and individual measurements from different environments cluster together, indicating reproducibility. Furthermore, inspection of LC-MS data by eye (see SI, Section 2.2, for full data), in the form of plots of feature intensities, and raw extracted ion chromatograms (EICs), confirms that robust systematic differences can be seen directly in the data.

With the exception of Li^+, we observed that addition of monovalent soluble salts produced similar product distributions. Li^+ experiments produced a product distribution similar to that produced in the presence of Mg^{2+}, while presence of Cu^{2+} or Eu^{3+} led to distributions which were clearly distinct from other salts. The reactions incorporating minerals yielded product ensemble distributions which were robustly distinguished in all of the analyses performed (with the exception of quartz and alumina). Broadly, the analyses on the experiments with different mixing orders resolve the ensembles into three pairs (G⇨A⇨H & A⇨G⇨H; G⇨H⇨A & H⇨G⇨A; A⇨H⇨G & H⇨A⇨G), with the reaction in which all amino acids were added together clearly resolved from all others. The reaction pattern is consistent with the trends observed in preliminary binary cross-reactivity tests, where G/A hetero-oligomerisation clearly dominates (see Figure S1). For example, products of G⇨A reactions are likely to resemble A⇨G if G/A hetero-oligomerisation rates are very much larger than either possible homo-oligomerisation. While our approach in this work has been non-deterministic, focussed on observing differences, these observations hint at the potential for deliberate 'programming' using modelling of reaction rates, although simple models accounting for thermodynamic equilibrium alone are not sufficient.

The sequence of peptide oligomers is crucial to their function, and whilst our aim is not to identify individual products, the question of whether product sequence distributions are altered, along with composition and yield, is of interest. While it is possible to match observed masses to consistent oligomer compositions, it is not possible to resolve and quantify all product oligomers in most cases since it requires identifying and separating very similar species, including those of identical mass. In many cases such isomeric species are extremely difficult to resolve using chromatography, especially using a general method, rather than one optimised to resolve specific sequence variants. However, since the shape of the features in the EICs for many masses corresponding to putative oligomers vary dramatically between populations, it is clear that oligomer sequence, as well as composition, is being steered. The five sequence permutations of G_4A (m/z = 318.141) are a rare example where the different

permutations can be resolved through chromatography, and variation of their relative abundances observed (Figure 3). Analysis of synthetic standards of the possible sequence permutations showed that GGGGA and AGGGG could be resolved from GAGGG, GGAGG, and GGGAG, which co-elute. Comparison of the mean intensities of these peaks between samples from different mixing histories showed clear variation in the distribution of the sequence permutations (see SI, Section 2.2.3 for further details).

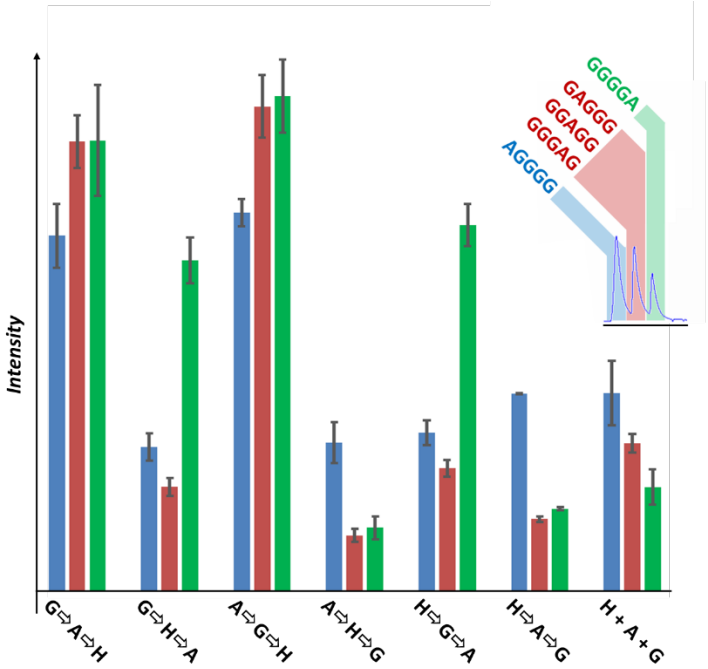

Figure 3. Plots revealing the sequence permutation distribution of G_4A pentamers. *Distribution of mean peak intensity from EICs (m/z = 318.141) of samples with different mixing histories, with error bars representing one standard deviation. Assignment shown in top right inset. [See SI for peak identification and further details. Peak intensities were extracted using Bruker Data analysis; intensity values displayed are means of 3 experimental replicates x 3 analytical repeats].*

Having established that variations to the environment can guide condensation reactions to yield product ensembles which are distinct in composition and sequence, we must then ask if this can also drive functional differences. To assess this, we first observed the effect of the different G/A/H product ensembles on the progression of a simple and well known reaction system, the decomposition of *para*-nitrophenyl acetate (*p*NPA, colourless) to release *para*-nitrophenol (*p*NP, yellow).[35] While absolute rates of the reaction in the presence of our ensembles were much lower than would be expected for catalysis by pure evolved/designed peptide catalysts[35] we found clear and reproducible differences between the effects of many of the ensembles. Interestingly, the parameter most significantly affecting differences in esterase activity appears to be soluble salt content; soluble salts have previously been proposed to direct chemistry in dehydrated environments.[36] Reproducible differences in the rate of *p*NP

release were observed in all of the sets of comparable ensembles, despite *p*NPA being known to interact with a broad range of catalysts (Figure 4). It is important to note that since the same amount of AA starting materials were used in all reactions, all differences in their effect on the reactions are mediated by the environmentally-programmed differences in product ensembles.

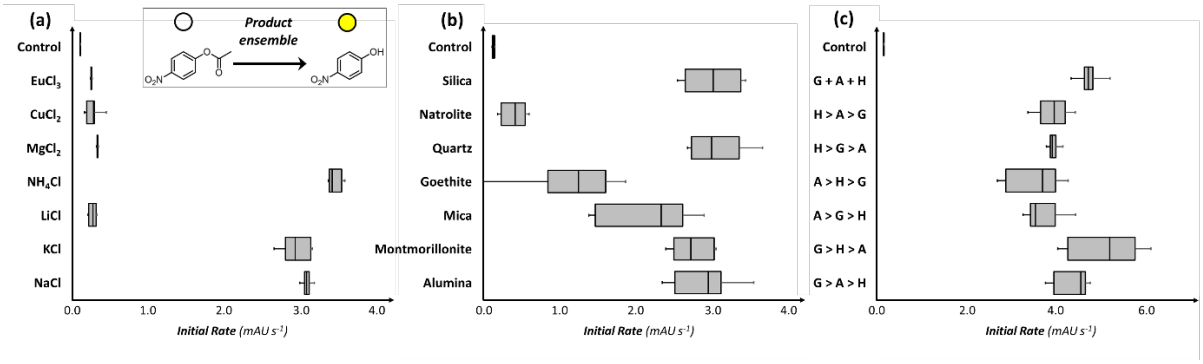

Figure 4. Different product ensembles differentially influence* para*-nitrophenyl acetate conversion. *(Inset) Decomposition of* p*NPA to release* p*NP produces a yellow color (measured as absorbance at 405 nm). Box plots comparing rates of* p*NP release on interaction with ensembles produced in different environments/conditions: an equimolar mixture of G, A & H with (a) different soluble salts and, (b) different minerals, and (c) different mixing orders of G, A & H over multiple cycles. [boxes represent middle quartiles, their middle line represents the mean, whiskers represent outlying quartiles].*

Molecular recognition is another important class of functionality in macromolecules, so a further set of condensation reactions were performed with different mixing histories, this time with alanine (A), aspartic acid (D), and valine (V). The dye Thioflavin T (ThT) is known to be recognised by hydrophobic sites in peptide assemblies/aggregates, and ThT recognition is frequently used to assess the formation of amyloids, where it is bound with some degree of selectivity.[37] Figure 5a shows the fluorescence responses on mixing ThT with ensembles produced, and robust differences are observed between some ensembles in their recognition of the dye. This indicates differential binding between amino acid/peptide assemblies synthesised in different environments, which may be sequence-dependent, based on our previous results with G/A/H oligomerisation. Furthermore, binding of ThT suggests the formation of potentially amyloid-like structures (such assembly is not uncommon in a range of compounds).[40]

We then investigated whether the assembly of nano- or micro-scale structures in our ensembles (as implied from the ThT experiments) could be observed directly. Transmission electron microscopy (TEM) revealed that different product ensembles often tend to assemble different ranges of structures (see Figure 5b and SI). These range from fibre-like structures to larger globular assembles, some of which appear to incorporate internal structures. Furthermore, addition of Ca^{2+} salts to the products leads to gelation in some ensembles, while others remain clear, free flowing, solutions (see Figure 5c and SI).

TEM imaging of gelled product ensembles reveals the assembly of fibrous structures, while that of those in clear solutions reveals discrete assemblies (see Figure 5d and SI). It is interesting to note that not all difference/similarity is clearly correlated, their relationship reflecting the complexity of the ensembles. For example, the "V⇨A⇨D" and "V⇨D⇨A" ensembles behave similarly in the ThT interaction measurements (Figure 5a), and yet on addition of Ca^{2+} they behave strikingly differently (one produces a persistent gel-like material, while the other does not; see Figure 5c and 5d). This observation highlights one advantage of an open exploratory approach, over one optimising for only a specific molecular target or parameter.

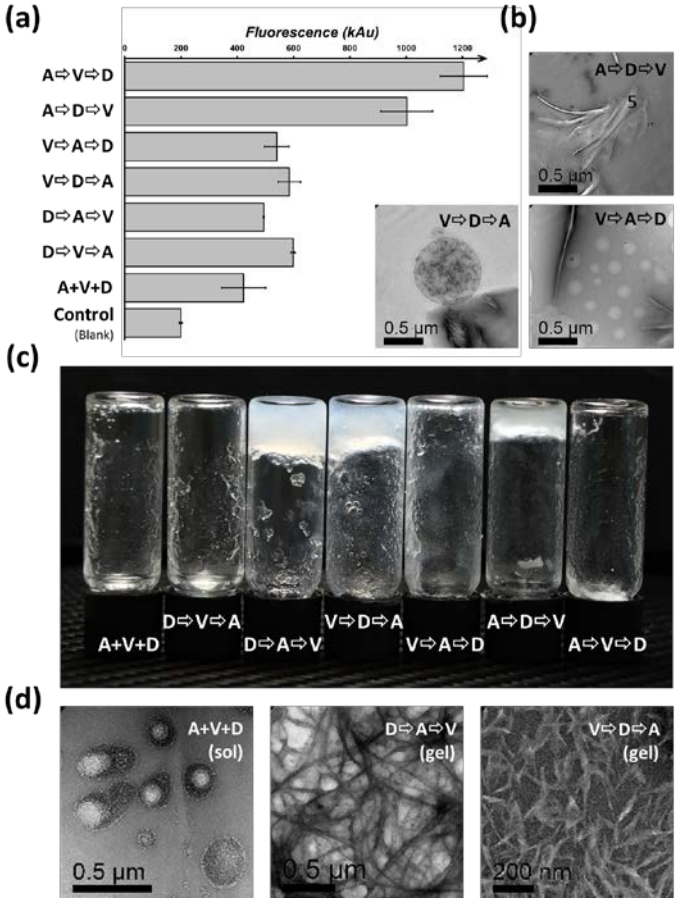

Figure 5. Recognition, assembly, and gelation properties differ between product ensembles. *(a) Thioflavin T (ThT) assay reveals considerable differences in binding properties of product ensembles from A, V & D condensation with different mixing orders, and (b) TEM images of the same products reveal assembly of qualitatively different structures. (c) On addition of Ca^{2+} salts, some ensembles form self-supporting gels, which remain in place when vials are inverted (others leave clear solutions, which flow to the bottom of vials on inversion, see SI), and (d) TEM inspection of these samples reveals the assembly of fibres in the gelled samples ("gel"), and discrete globular structures in the clear solutions ("sol").*

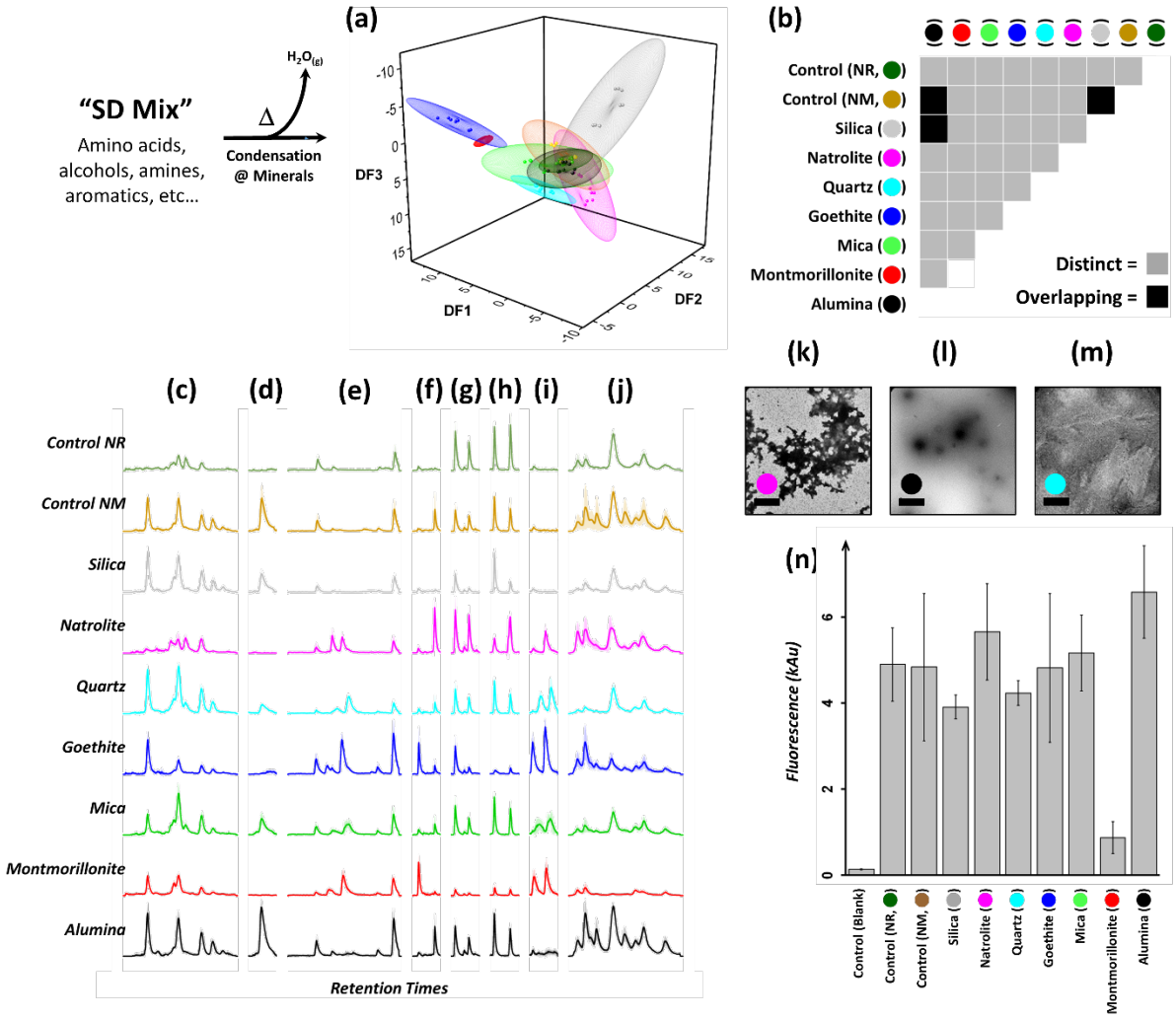

Figure 6. Minerals can direct emergence of distinct ensembles from a 'Primordial Soup' model. *Complex mixtures (SD Mix) were dehydrated in the presence of different minerals. LC-MS data shows almost all ensembles are compositionally different: (a) LC-MS features as PC-DFA analysis; plot generated with Origin Pro 2016 (OriginLab) (b) matrix demonstrating ensemble overlap in PC-DFA ["Control NM" = no mineral present; "Control NR" = no condensation reaction; "Control (Blank)" = assay buffer blank]; Example EICs (c to j, ordered by ascending m/z) selected to demonstrate a range of diversity in the data. Some show almost-complete selectivity between three isobaric species. [c, m/z = 101.0715; d, m/z = 102.0918; e, m/z = 166.0245; f, m/z = 174.0582; g, m/z = 244.1907; h, m/z = 278.0520; i, m/z = 255.0590; j, m/z = 321.0014; see SI for further examples]; TEM of ensembles produced in the presence of Natrolite (k), Alumina (l), and Quartz (m) reveals marked morphological differences [scale bar = 2 μm; see SI for full collection of TEM images]; and (n) some ensembles interact differently with ThT.*

Taken together, these results show that the reactivity, assembly, and molecular recognition properties of condensation products from fixed sets of amino acids can be consistently controlled, or programmed, by their reaction environment, with functional consequences. However, these reactions have only

incorporated a small set of pure amino acid building blocks, producing a large but restricted range of products. We then wondered whether the same phenomenon could be observed applying our approach to a far more complex mixture of starting materials. For this we chose a classic "Primordial Soup": the products of the Miller-Urey type spark discharge experiment ("SD Mix", hereafter).[11,39] Starting from a simple mixture of gases (H_2, CH_4, NH_3) and refluxing water, with the gases passing through a high energy spark discharge between two tungsten electrodes,[9] this experiment produces a notoriously complex product mixture. Since SD Mix is known to contain a very wide variety of species, including carboxylic acids, alcohols, aromatic species, and amino acids,[11,39] the possibility of forming condensation products is clear.

To explore whether condensation reactions of such a complex mixture could be directed by the environment, we performed a series of experiments in which we subjected a standardised SD Mix to similar condensation conditions in the presence of the same series of minerals with which we previously treated AA reactions (see SI for full details). LC-MS fingerprinting of the product ensembles[9] (following dialysis to remove salts and small molecules) revealed that reaction in different mineral environments yielded a range of product populations which are almost all reproducibly distinct (Figure 6a and 6b). Closer analysis revealed a wide range of product species (ca. 1800 features identified, see Figure S23 for full plot), many of which are not observed in SD Mix controls where no condensation reaction had been carried out. Selected EICs of specific masses (a representative selection) from these mineral-programmed product ensembles (see Figure 6c to 6j, and SI for further examples) reveal a range of selectivities, with some ensembles manifesting almost complete selectivity for the presence/absence of possible isomeric products. Inspection of EICs and raw data also confirm that while the ensembles are diverse, none is completely lacking in products – their compositions are simply different. Furthermore, ThT recognition tests (Figure 6n) revealed that some product ensembles possess different recognition and assembly properties. For example, ensembles formed in the presence of Montmorillonite, Silica, and Alumina all reproducibly recognise ThT very differently (and all distinct to a control lacking condensation products). TEM morphological examination also shows differences (see Figure 6k to m for representative samples, and SI for additional data). Where some populations have similar results in the ThT recognition assay, they do not necessarily appear to produce morphologically similar assemblies when observed by TEM (e.g. those with Goethite and Natrolite are clearly distinct). Demonstrating the differentiation of distinct ensembles from what have been seen as classic 'intractable'[8] reaction systems adds strength to the increasing argument that "it is possible that the difficulty that chemical heterogeneity presents to early life has been exaggerated"[16] and that this open approach to systems complexity may be more instructive in these contexts than traditional metrics of success in organic chemistry (yield, purity).

Exploring the mechanism by which complexity and function emerge, and are differentiated, in chemical systems is important for establishing potential origins of evolution,[40,41] pointing to how a variety of

ordered systems might emerge from the "clutter wrought by prebiotic chemistry".[42] Indeed, our demonstration of salts and minerals guiding the differentiation of distinct functional ensembles from simple building blocks are the first experimental demonstration of Cairns-Smith's ideas that inorganic materials can program complex organic chemistry to yield differently-fit populations,[30] beyond simply selecting for particular molecular targets.[13,43] This should be seen as a complement to research identifying particular sets of target molecules[13,43] which may have been involved in a historical origin of life.[44] Our approach to make these observations, using tools from omics sciences with no requirement for target products,[9] represents a promising alternative approach to understanding the emergence of complex functional systems where outcomes are tuned by the environment.[45] Unlike more familiar approaches,[27] it is expandable to address increasingly complex systems, wherein selectivity may be driven by competition and complexity,[46] and where approaches based on expectations of particular products are limited (e.g. those relying on databases of known species, or *de novo* assignment of peptides).

Supplementary Information is linked to the online version of the paper on www.nature.com/nchem.

Acknowledgements: We would like to thank Mr Jonas Kube, Mr. José María Planchat i Barbarà & Dr. Mohammed Hezwani for help with the protocols & reactions, Mr. Kevin Donkers, Mr. James McIver, Dr. Maria Diana Castro & Dr. Anna Andreou for technical support, Mr. Igor Menghini (NValue Analytics), Dr Jonathan Grizou for guidance on data analysis, Dr. Karl E. Burgess (Glasgow Polyomics) & Dr. Emma Carrick for guidance on LC-MS & data analysis. 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), the EC (projects 610730 EVOPROG, 611640 EVOBLISS, 318671 MICREAGENTS), ERC (project 670467 SMART-POM).

Author Contributions: LC conceived the idea, designed the project and coordinated the efforts of the research team, with AS. AS and MRG did most of the peptide research together, contributing equally: MRG focussing more on developing the synthetic approach, and AS focussing more on developing the analytical approaches. GC performed the spark discharge experiments, and worked with AS on downstream processing. YMA and GC helped perform some experiments and analysis. PSG, CM and SIW helped with exploratory data analysis and modelling. YMA introduced ThT assays and gelation. MM performed TEM imaging and sample preparation, with YMA and AS. AS, GC, RT-M and LC co-wrote the paper with input from all the authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Leroy Cronin (Lee.Cronin@glasgow.ac.uk).

References

1. Chuang, H. Y., Hofree, M. & Ideker, T. A Decade of Systems Biology. *Annu. Rev. Cell Dev. Biol.*, **26**, 721-744 (2010).
2. de Lorenzo, V. & Galperin, M. Microbial systems biology: bottom up and top down. *FEMS Microbiol. Rev.*, **33**, 1-2 (2009).
3. Schuster, P. Taming combinatorial explosion. *Proc. Natl. Acad. Sci.*, **97**, 7678-7680 (2000).
4. Cronin, L. Reaction: A New Genesis for Origins Research? , **2**, 601-603 (2017).
5. Ashkenasy, G., Hermans, T. M., Otto, S. & Taylor, A. F. Systems chemistry. *Chem. Soc. Rev.*, **46**, 2543-2554 (2017).
6. Benner, S. A., Kim, H. J. & Carrigan, M. A. Asphalt, Water, and the Prebiotic Synthesis of Ribose, Ribonucleosides, and RNA. *Acc. Chem. Res*, **45**, 2025-2034 (2012).
7. Benner, S. A. Paradoxes in the Origin of Life. *Orig. Life Evol. Biosp.*, **44**, 339-343 (2014).
8. Schwartz, A. W. Intractable mixtures and the origin of life. *Chem. Biodiversity*, **4**, 656-664 (2007).
9. Cooper, G. J. T. *et al.* Miller–Urey Spark-Discharge Experiments in the Deuterium World. *Angew .Chem. Int. Ed.*, **56(28)**, 8079-8082 (2017).
10. Ruiz-Mirazo, K., Briones, C. & de la Escosura, A. Prebiotic Systems Chemistry: New Perspectives for the Origins of Life. *Chem. Rev.*, **114**, 285-366 (2014).
11. Miller, S. L. A Production of Amino Acids under Possible Primitive Earth Conditions. *Science*, **117**, 528-529 (1953).
12. Fox, S. W. & Harada, K. Thermal Copolymerization of Amino Acids to a Product Resembling Protein. *Science*, **128**, 1214-1214 (1958).
13. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature*, **459**, 239-242 (2009).
14. Ferris, J. P., Hill, A. R., Liu, R. H. & Orgel, L. E. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature*, **381**, 59-61 (1996).
15. Rajamani, S. *et al.* Lipid-assisted synthesis of RNA-like polymers from mononucleotides. *Orig. Life Evol. Biosp.*, **38**, 57-74 (2008).
16. Pressman, A., Blanco, C., Chen, I. A. The RNA World as a Model System to Study the Origin of Life. *Current Biology*, **25**, R953-R963 (2015).
17. Beaufils, D., Jepaul, S., Liu, Z. W., Boiteau, L. & Pascal, R. The Activation of Free Dipeptides Promoted by Strong Activating Agents in Water Does not Yield Diketopiperazines. *Orig. Life Evol. Biosp.*, **46**, 19-30 (2016).
18. Rodriguez-Garcia, M. *et al.* Formation of oligopeptides in high yield under simple programmable conditions. *Nat. Commun.*, **6**, 8385 (2015).
19. 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).
20. Schwendinger, M. G. & Rode, B. M. Possible Role of Copper and Sodium-Chloride in Prebiotic Evolution of Peptides. *Anal. Sci.*, **5**, 411-414 (1989).
21. Walde, P., Wick, R., Fresta, M., Mangone, A. & Luisi, P. L. Autopoietic Self-Reproduction of Fatty-Acid Vesicles. *J. Am. Chem. Soc.*, **116**, 11649-11654 (1994).
22. Adamala, K. & Szostak, J. W. Nonenzymatic Template-Directed RNA Synthesis Inside Model Protocells. *Science*, **342**, 1098-1100 (2013).
23. DeClue, M. S. *et al.* Nucleobase Mediated, Photocatalytic Vesicle Formation from an Ester Precursor. *J. Am. Chem. Soc.*, **131**, 931-933 (2009).
24. Pappas, C.G. *et al.* Dynamic Peptide Libraries for the Discovery of Supramolecular Nanomaterials. *Nat. Nanotechnol.*, **11**, 960 (2016).
25. Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K. & Ghadiri, M. R. A self-replicating peptide. *Nature*, **382**, 525-528 (1996).
26. Semenov, S. N. *et al.* Autocatalytic, bistable, oscillatory networks of biologically relevant organic reactions. *Nature*, **537**, 656-660 (2016).

27. Keller, M. A., Driscoll, P. C., Messner, C., Ralser, M. Primordial Krebs-cycle-like non-enzymatic reactions detected by mass spectrometry and nuclear magnetic resonance. W*ellcome Open Res.*, **2**, 52 (2017).
28. Keller, M. A., Turchyn, A. V. & Ralser, M. Non-enzymatic glycolysis and pentose phosphate pathway-like reactions in a plausible Archean ocean. *Mol. Syst. Biol.*, **10** (2014).
29. Nakashima, T., Jungck, J. R., Fox, S. W., Lederer, E. & Das, B. C. A Test for Randomness in Peptides Isolated from a Thermal Polyamino Acid. *Int. J. Quantum Chem.*, **12**, 65-72 (1977).
30. Cairns-Smith, A. G. *Genetic Takeover And The Mineral Origins Of Life*. (Cambridge University Press, 1982).
31. Cleaves, H. J. *et al.* Mineral-organic interfacial processes: potential roles in the origins of life. *Chem. Soc. Rev.*, **41**, 5502-5525 (2012).
32. Kitadai N. Dissolved divalent metal and pH effects on amino acid polymerization: A thermodynamic evaluation. *Orig. Life Evol. Biosph.*, **47**, 13-37 (2016).
33. Fuchida, S., Masuda, H., Shinoda, K. Peptide formation mechanism on montmorillonite under thermal conditions. *Orig. Life Evol. Biosph.* **1**, 13-28 (2014).
34. Kitadai, N., Oonishi, H., Umemoto, K., Usui, T., Fukushi, K. & Nakashima, S. Glycine Polymerization on Oxide Minerals, *Orig. Life Evol. Biosph.*, **47(2)**, 123-143 (2016).
35. Rufo, C. M. *et al.* Short peptides self-assemble to produce catalytic amyloids. *Nature Chem.*, **6**, 303-309 (2014).
36. Lahav, N. and Chang, S. The possible role of soluble salts in chemical evolution. *J. Mol. Evol.* **19**, 36-46 (1982).
37. Biancalana, M. & Koide, S. Molecular mechanism of Thioflavin-T binding to amyloid fibrils. *Biochim. Biophys. Acta, Proteins Proteomics*, **1804**, 1405-1412 (2010).
38. Shaham-Niv, S., Adler-Abramovich, L., Schnaider, L. & Gazit, E. Extension of the generic amyloid hypothesis to nonproteinaceous metabolite assemblies, *Science Adv.*, **1(7)**, e1500137 (2015).
39. Johnson, A. P. et al. The Miller volcanic spark discharge experiment. *Science*, **322**, 404 (2008).
40. Cairns-Smith, A. G. *Seven Clues To The Origin Of Life : A Scientific Detective Story*. (Cambridge University Press, 1985).
41. Baum, D. A. Selection and the Origin of Cells. *Bioscience*, **65**, 678-684 (2015).
42. Joyce, G. F. The antiquity of RNA-based evolution. *Nature*, **418**, 214–221 (2002).
43. Ricardo, A., Carrigan, M. A., Olcott, A. N. & Benner, S. A. Borate minerals stabilize ribose. *Science*, **303**, 196-196 (2004).
44. Orgel, L. E. Some consequences of the RNA world hypothesis. *Orig. Life Evol. Biosp.*, **33**, 211-218 (2003).
45. Horowitz, J. M. & England, J. L. Spontaneous fine-tuning to environment in many-species chemical reaction networks, *Proc. Natl. Acad. Sci.*, **114(29)**, 7565-7570 (2017).
46. Kovaříček, P. *et al.* Competition-driven selection in covalent dynamic networks and implementation in organic reactional selectivity. *Chem Sci*, **7**, 3215-3226 (2016).