

# Emergence of Function and Selection from Recursively Programmed Polymerisation Reactions in Mineral Environments

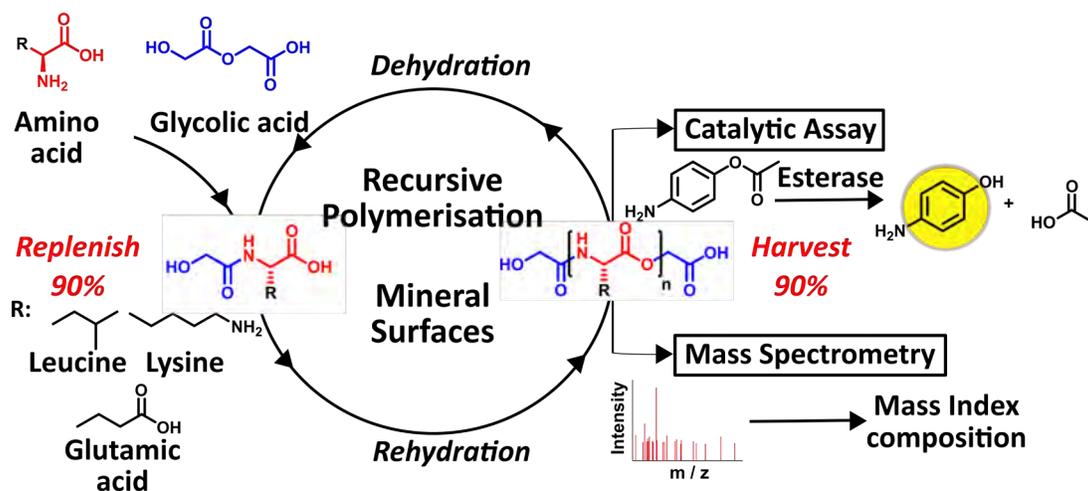
David Doran, Yousef M. Abul-Haija and Leroy Cronin\*

*School of Chemistry, University of Glasgow, Glasgow G12 8QQ (UK)*

\*Corresponding author: [lee.cronin@glasgow.ac.uk](mailto:lee.cronin@glasgow.ac.uk)

**Living systems are characterised by an ability to sustain chemical reaction networks far-from-equilibrium. It is likely that life first arose through a process of continual disruption of equilibrium states in recursive reaction networks, driven by periodic environmental changes allowing the emergence of a memory. Herein, we report the emergence of proto-enzymatic function from recursive polymerisation reactions using amino acids and glycolic acid over four wet-dry cycles. Reactions are kept out of equilibrium by diluting products 9:1 in fresh starting solution at the end of each recursive cycle, and the development of complex high molecular weight species is explored using a new metric, the Mass Index, which allows the complexity of the system to be explored as a function of cycle. This process is carried out on a range of different mineral environments. We explore the hypothesis that disrupting equilibrium *via* recursive cycling imposes a selection pressure and subsequent boundary conditions on products, which may otherwise be prone to uncontrolled combinatorial explosion. After just four reaction cycles, product mixtures from recursive reactions exhibit greater catalytic activity and truncation of product space towards higher molecular weight species compared to non-recursive controls.**

Biological entities can be viewed as self-propagating, autocatalytic networks in a sustained far-from-equilibrium state, in which the stoichiometry of all functionally active components is maintained between generations.<sup>1</sup> This state is only possible due to the recursive nature of biological replication. Recursive chemical pathways, including biological replication, are defined as those in which the functional, bond-forming units are regenerated by the pathway and are thus available for further reaction at the end of each reaction cycle.<sup>2,3</sup> We propose that, for the study of living and potentially life-like artificial systems in the laboratory, this definition should be refined to include a process of regeneration of chemical systems after disruption of the equilibrium state by dilution and transfer to a fresh environment. Ultimately, biological entities are characterised by the ability to sustain far-from-equilibrium states, and regenerate in the presence of what are, relative to the internal components of the system, extremely dilute feedstocks. For the remainder of this communication, the term “chemical recursion” will be used to refer to a process of continuous dilution of products and their replenishment by fresh feedstocks in a new reaction environment.

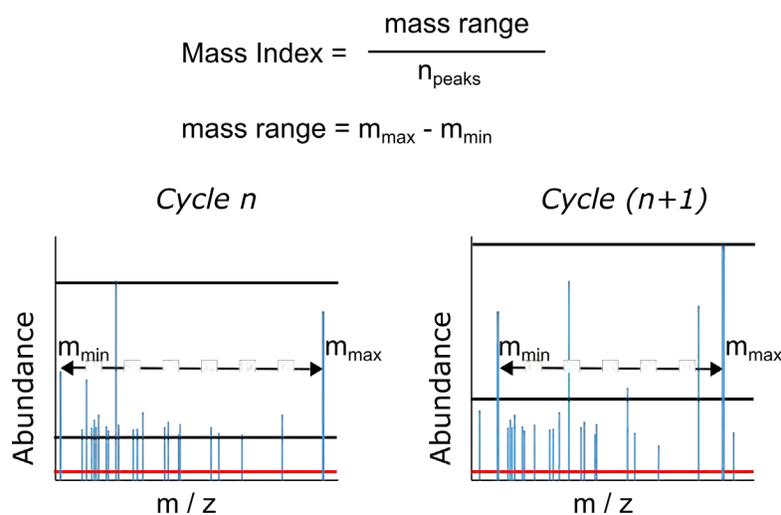


**Figure 1 | Glycolic acid-mediated peptide bond formation and depsipeptide elongation via ester-amide exchange.** This is repeated in an iterative fashion with fresh addition of starting materials on various solid surface matrices. Compositional and functional analyses are carried out at the end of each reaction cycle by mass spectrometry and functional assays, respectively.

Chemical recursion is key not only for maintaining the dynamic, far-from-equilibrium states that characterise biology, but also for providing a means of imparting a “chemical history” on generations of cyclical chemical reactions, in which the outcome of one reaction cycle is partly influenced by those that have preceded it. Without such programmed history, there is no possibility for the emergence of evolving chemical systems and therefore life.<sup>4-6</sup> Despite this, much effort has been devoted to non-recursive syntheses of molecules perceived to be crucial to life’s origins, such as RNA.<sup>7</sup> The aim of such efforts is to recreate the precise pathways *via* which the first biomolecules emerged, often under the assumption that heredity and evolution began only after chance accumulation of these molecules.<sup>8</sup> This may have been confounded by the likely untestable assumption that a template-driven, self-replicating genetic polymer is essential for heredity in the most primitive life or life-like systems. However, we argue here, as we have done previously,<sup>9</sup> that recursion is essential from the earliest onset of a living or artificial living system, and that functionality and evolvability can be induced in chemical networks comprised of much simpler components than have been considered by many in the origins of life field. Indeed, it is likely that life first emerged from a pool of very simple building blocks, but without strict boundary conditions imposed by recursive selection processes, these building blocks would have undergone combinatorial explosion to produce little more than intractable tar.<sup>10</sup>

Experimental frameworks for generating artificial life through recursive selection have been proposed,<sup>1,9</sup> however, the practical utility of these frameworks is yet to be demonstrated. A major obstacle to achieving such a goal is identifying processes that can steer combinatorial explosions towards a narrower pool of functional products.<sup>11-13</sup> Confirming whether this has been achieved may be an equally, if not more, daunting task.<sup>14</sup> In biological systems, a suite of so-called “omics” technologies enable the surveying of system-level changes. However, these tools are built around the constrained chemistry of extant biology and have not been designed for studies on artificial life, although there have been some attempts at developing workflows.<sup>15</sup> Depsipeptides are mixed oligomers of polypeptides and polyesters, produced from co-polymerisation of  $\alpha$ -amino acids and  $\alpha$ -hydroxy acids, with a potential for structural

and functional diversity that is comparable to pure peptides.<sup>16</sup> However, structure formation and functional activity in these reactions are yet to be achieved. From a relatively simple starting mixture of four amino acid and  $\alpha$ -hydroxy acid monomers, assuming equal reactivities and a maximum chain length of 8, a potential 65,536 unique depsipeptides can arise. This is excluding branched or cyclic products, which are common for many side-chain structures.



**Figure 2 | Mass spectrometry data processing workflow.** Raw data is filtered to remove any MS<sup>1</sup> hits not in the depsipeptide product library. An absolute noise threshold (red line) is applied to all remaining peaks. To calculate the Mass Index, a further filtering step is applied, with all species with a total abundance of < 4.55% of the most abundant species discarded. The mass range of remaining peaks is then normalised to the number of peaks within this intensity range (black lines).

We carried out multiple recursive cycles of depsipeptide co-polymerisation on various mineral substrates in order to (i) assess the robustness of any effects of recursive cycling, and (ii) provide a means of steering the reaction down different trajectories or ‘chemical histories’. Three amino acids (L-leucine, L-glutamic acid and L-lysine) and one  $\alpha$ -hydroxy acid (glycolic acid) were chosen as a model reaction system. A fixed ratio of monomers was reacted on various mineral surfaces in a recursive manner, with approximately 10% of the product mixture seeded on to fresh mineral and monomer feedstock solutions

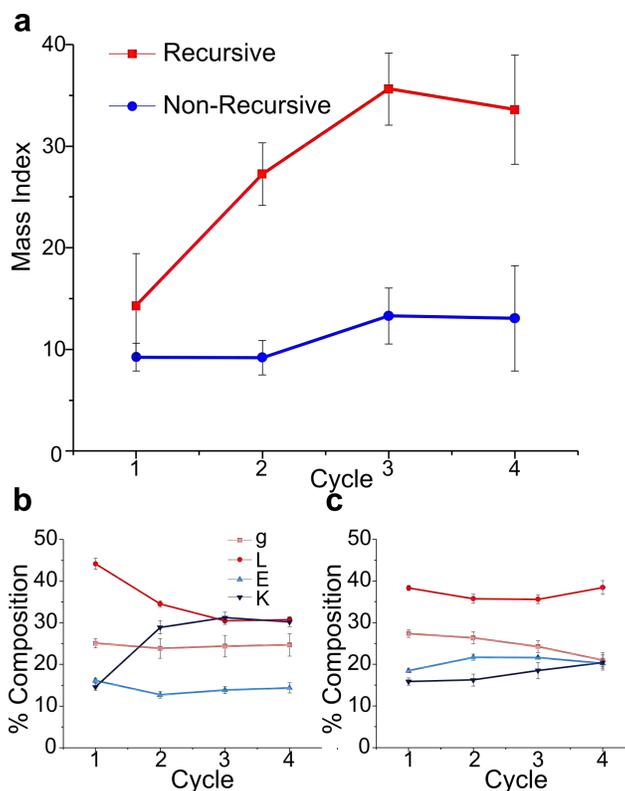
at the end of each reaction cycle. Eight mineral environments were chosen to steer reaction outcomes: crushed glass, montmorillonite, gypsum, quartz, calcite, chalcopyrite, opal, kernite, along with a mineral-free control environment. Products were characterised in a functional assay and via mass spectrometry at the end of each reaction cycle (Figure 1). Due to the vast number of potential products, an extensive mass library of over 19,000 linear, cyclic and branched species, each corresponding to a unique depsipeptide composition of up to 8-unit length, was used to screen the mass spectrometry data. Extracted ion chromatograms (EICs) of each potential product were obtained for all product mixtures. The Mass Index, which normalises the number of observed products to their mass distribution, was used as a simple, objective criterion for assessing differences between product mixtures (Figure 2). The library used to screen products was comprised of 6633 depsipeptide compositions with various common adducts, giving a total of 19,899 masses. This represents a significant portion of feasible product space; however, it is pertinent to note that this screening method does not take into account differences between isobaric sequences. For example, the mass of a trimer containing one L, one E and one K monomer unit could correspond to any one of six unique sequences.

## **Results**

### **Mass Index and Product Composition**

To calculate the Mass Index, raw mass spectrometry data was filtered using a depsipeptide mass library and an absolute noise threshold. After this, a second threshold of relative intensity was applied, retaining only species that exceeded 4.55% of the most intense / abundant species in the mass library. The mass range of species within this intensity range was then divided by the number of remaining species. Thus, an increase in the Mass Index reflects a funnelling of product space from pools of low molecular weight species to a greater relative abundance of higher mass products.

We observed a clear increase in the Mass Index over recursive cycles in the mineral-free environment while the non-recursive controls showed a much smaller increase (Figure 3a). Differences between the various mineral environments were also observed using this metric and showed broadly similar trends (Supplementary Fig. 3), demonstrating the robust effects of chemical recursion in this system. In previous work, we examined the ability of minerals and salts to influence the product distribution of similar uncontrolled condensation reactions,<sup>17</sup> and minerals have been studied extensively for their ability to concentrate amino acids on their surface, as well as catalyse aqueous peptide polymerisation driven by activating agents<sup>18</sup> and wet-dry cycles.<sup>19</sup>



**Figure 3 | Change in desipeptide product distribution over multiple recursive cycles.** Mass Index (a) and relative monomer composition of products (b and c) for recursive and non-recursive samples in a mineral-free reaction environment, respectively) are shown over four reaction cycles. “g” = glycolic acid; “L” = leucine; “E” = glutamic acid; “K” = lysine. Data represent mean of 9 replicates  $\pm$  standard deviation.

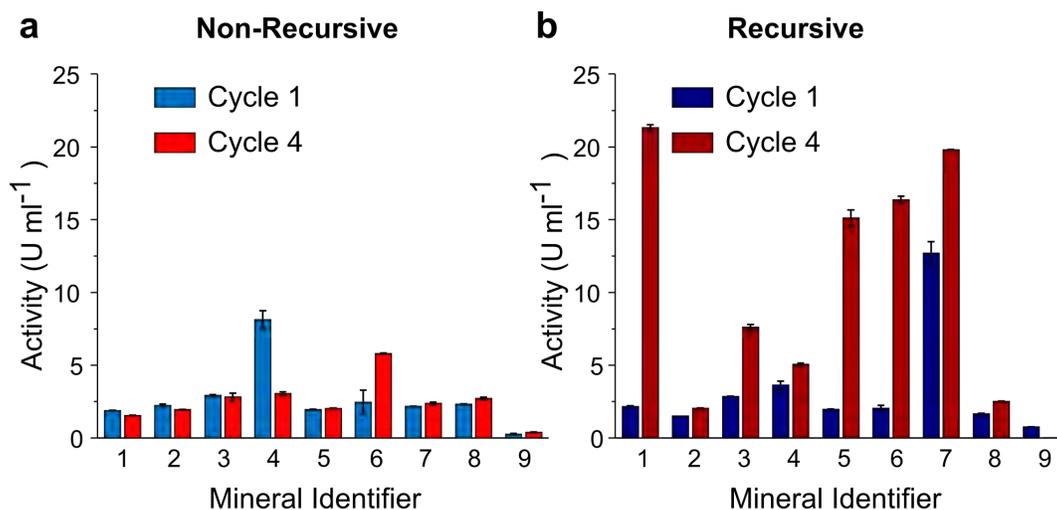
The relative monomer composition of products was also observed from the mass spectrometry data, and a trend towards increasing lysine content over four cycles correlated with a depletion of leucine in products (Figure 3b, Supplementary Figure 4). This trend was not observed in the non-recursive control samples (Figure 3c, Supplementary Figure 4). Interestingly, no evidence of glycolic acid depletion was found in the recursive reactions. This is in contrast to literature reports of an increase in peptide bond character of depsipeptides in similar cyclic, thermally-driven dehydrations of amino acid and  $\alpha$ -hydroxy acid monomers.<sup>20</sup> However, a decrease in glycolic acid content was observed for most of the cyclic, non-recursive control reactions, consistent with the peptide enrichment found in cyclical depsipeptide elongation in literature reports.<sup>3,15</sup>

### **Esterase Activity**

Having observed the effect of recursion on the composition of depsipeptide products, the next step was to test for the emergence of function. A hydrolytic assay was used to determine the esterase activity of products. Ester bond hydrolysis is an important step in the ester-amide exchange reaction that enables depsipeptide elongation,<sup>3</sup> and a vital process in biology. *p*-nitrophenyl acetate (*p*NPA), which breaks down into yellow *p*-nitrophenol (*p*NP) and acetate, was used as an esterase substrate for the hydrolytic assay.<sup>21</sup> The activity of product mixtures in this assay was compared and normalised to a standard curve of  $\alpha$ -chymotrypsin (Figure 4, S6) and activity was measured in esterase units per ml of product solution. Recursive product mixtures of no mineral, montmorillonite, quartz, calcite and chalcopyrite exhibited a sharp increase in activity after four cycles (Figure 4). This increase was reproducible, but not step-wise, occurring sharply between cycles 3 and 4. No comparable increase was observed in the non-recursive control reactions (Figure 4, Supplementary Figure 7).

The increase in esterase activity with chemical recursion suggests some form of selective advantage of functionally active products in a recursive system. Interestingly, this also correlates with the increase in lysine and depletion of leucine content of products (Figure 3b). Lysine is one of the most functionally

active amino acids in modern biochemistry - as well as taking part in acid-base catalysis, lysine is readily modified due to the additional amine sidechain.<sup>22</sup> It is unclear at this stage whether lysine enrichment of depsipeptides plays a role in our observed increase in esterase activity.



**Figure 4 | Effect of solid surface matrices on esterase activity.** Esterase activity (in enzyme units, U, per ml) for (a) non-recursive and (b) recursive products after 1 and 4 reaction cycles on the following mineral environments: 1) no mineral; 2) glass; 3) montmorillonite; 4) gypsum; 5) quartz; 6) calcite; 7) chalcopryrite; 8) opal; 9) kernite. Data represent mean of 2 replicates  $\pm$  standard deviation.

Structural characterisation in solution was carried out to assess any impact of this change in composition on secondary structure formation in products. Circular dichroism profiles of mixtures differed markedly from starting material controls, with a characteristic  $\beta$ -sheet signal at 210nm, indicating potential secondary structure formation.<sup>23</sup> However, no evidence was found of this profile changing after recursive cycling (Supplementary Fig. 8). Similarly, two FTIR absorption bands were identified at 1600 and 1650cm<sup>-1</sup> in products, possibly corresponding to the amide I and amide II structural bands, respectively.<sup>24</sup> Starting material controls exhibited only one band in this region, at approximately 1625 cm<sup>-1</sup>, most likely corresponding to free carboxyl of  $\alpha$ - amino and hydroxy acids. Further investigation would be required to

determine whether absorption bands in products represent secondary structure formation (Supplementary Fig. 9).

## **Discussion**

In this work, the emergence of selection and functionality from unconstrained, recursive depsipeptide polymerisation has been demonstrated for the first time. Further work is ongoing to determine the precise mechanism of this phenomenon. Recursive cycling was used to induce a selection pressure on the system. The ability of mineral surfaces to strongly bind and trap peptide oligomers is well documented, as is the heat lability of many depsipeptides.<sup>3</sup> Thus, only those products not irreversibly trapped on the mineral surface matrices and capable of persisting under continuous heating at 90° C would have been selected for measurement and participation in further cycles. Nevertheless, the successful truncation of product space towards higher mass products through recursive cycling persisted in all but one mineral environment tested for this system, with a majority of environments facilitating increased catalytic activity of products after just four reaction cycles. This demonstrates the potential importance of chemical recursion in truncating combinatorial explosion, which is essential for complex, functional species to arise in sufficient abundance to aid the transition from non-living to living systems. Future work will also demonstrate the applicability of the Mass Index in screening of sequence libraries, and also screening sequence space for sub-populations of products that are responsible for esterase activity and structure formation.

## **Methods**

### **Depsipeptide Polymerisation**

10 ml solutions of 30 mM L-leucine (Sigma, CAS: 61-90-5), 30mM L-glutamic acid (Sigma, CAS: 56-86-0), 30 mM L-lysine (Sigma, CAS: 56-87-1) and 100 mM glycolic acid (Sigma, CAS: 79-14-1) in HPLC-grade H<sub>2</sub>O were adjusted to pH 2.5 using H<sub>3</sub>PO<sub>4</sub> and heated at 90° C for 15 hours in open cap glass vials on a 75 vial insert heating slab (Supplementary Fig. 10). For recursive reactions, 150 mg of solid matrix was added to the vials at the beginning of the first reaction cycle. After 15 hours of heating, the

reaction solutions had completely evaporated, and products were re-suspended in 10 ml HPLC-grade H<sub>2</sub>O. Re-suspended product was vortexed, and 1 ml of product solution was transferred to 9 ml fresh reagent solution and 135 mg fresh solid matrix in a fresh vial. Solutions were re-adjusted to pH 2.5, and the process was repeated for each subsequent cycle. For non-recursive control reactions, the first cycle was carried out as in the recursive reactions; however, products were not transferred to fresh feedstocks and no further solid matrix material was added to the mixture. Montmorillonite, gypsum, quartz, calcite, chalcopyrite, opal and kernite solid matrix materials were sourced from Richard Tayler minerals. Solids were crushed and passed through a sieve with a 3 µm cut-off prior to addition to the reaction. Crushed glass was generated using the same process from the same glass vials that were used for all reactions.

### **Esterase Assay**

*p*-nitrophenyl acetate (pNPA) (CAS: 830-03-5) and  $\alpha$ -chymotrypsin standards containing 40 enzyme units (U) per mg (CAS: 9004-07-3) were purchased from Sigma. Standard solutions were made up by serial dilution in cold (4° C) 10 mM HCl. Standards were run at the following concentrations (U per ml): 40, 20, 10, 5, 2.5, 1.25. Prior to running on the assay, 1 ml aliquots of product solutions were adjusted to pH 7 *via* addition of aqueous NaOH and centrifuged at 4000 rpm for at least 5 minutes. To each well on a 96 well-plate, 50 µL of product supernatant or chymotrypsin standard and 150 µL of buffered 130 µM *p*NPA solution was added. All standards and samples were run in duplicate. *p*NPA hydrolysis was monitored by measuring absorbance at 405 nm. Measurements were taken every 2 minutes for 2 hours using a Tecan Infinite® 200 Pro plate reader.

### **Circular Dichroism**

Product mixtures were diluted 1:1 in pH 8.0 sodium phosphate buffer (93 mM Na<sub>2</sub>HPO<sub>4</sub>, 7 mM NaH<sub>2</sub>PO<sub>4</sub>) in HPLC-grade H<sub>2</sub>O, vortexed and sonicated at 45° C for approximately 15 minutes, before chilling at 4° C overnight. Immediately prior to CD measurements, buffered product mixtures were diluted 1:20 in fresh pH 8 buffer solution. CD measurements were taken at room temperature using a Jasco J-810 spectropolarimeter in a 0.2 cm path with a data pitch of 0.1 nm, in continuous scanning mode at a scan

speed of 100 nm min<sup>-1</sup>, 2 second response time, accumulation of 1 and bandwidth of 1 nm. Spectra were recorded from 190-400 nm.

### **Fourier Transform Infrared Spectroscopy**

1 ml of product mixture was diluted 1:1 in HPLC-grade H<sub>2</sub>O, frozen at -80° C and lyophilised using a Christ™ Alpha 1-2 LDplus freeze-dryer. Lyophilised product material was then re-suspended in 1 ml deuterated pH 8 sodium phosphate buffer (93 mM Na<sub>2</sub>HPO<sub>4</sub>, 7 mM NaH<sub>2</sub>PO<sub>4</sub>). Deuterated water (CAS: 7789-20-0) was purchased from Sigma. Fourier transform infrared spectroscopy (FTIR) measurements were taken using an IRAffinity-1S FTIR spectrophotometer (Shimadzu). 64 scans were taken for each measurement, with resolution set at 16 cm<sup>-1</sup> for a spectral range of 2000 cm<sup>-1</sup> to 600 cm<sup>-1</sup>.

### **Mass Spectrometry Data Analysis**

Depsipeptide product mass lists were generated for all potential compositions arising from the combination of L, E, K and g monomers up to total oligomer length of 8 monomer units, with a mass range of 200-1000 Da. To account for the presence of cationic contaminants, adducts of each product with NH<sub>4</sub><sup>+</sup>, K<sup>+</sup> and Na<sup>+</sup> were added to the mass lists. Potentially branched and cyclic products were accounted for by removing one H<sub>2</sub>O mass unit (18.0565 Da) from products for every proposed branching point. The final mass list contained 6363 product compositions, plus adducts and branched or cyclic oligomers, giving a total of 19,899 species. 32-bit mzml files containing full MS-1 spectra were generated using Proteowizard MS Convert software. R 3.3 was used to obtain extracted ion chromatograms (EICs) for each species in the mass list. From this data, mass indices and monomer content of products were calculated using Python 3.6. For full mass spectrometry data analysis workflow see Supplementary Information.

### **Mass Spectrometry Data Acquisition**

Product mixtures were diluted 1:100 in HPLC-MS grade H<sub>2</sub>O, filtered through a nylon membrane with 0.2 µm pore size into glass vials (#2-SVW8-CPK, ThermoScientific) and loaded on to an autosampler (#WPS-3000TRS, ThermoScientific) hooked up to a quaternary pump (#LPG-3400RS,

ThermoScientific). Samples were injected in 10  $\mu\text{l}$  aliquots into the Bruker Maxis Impact II in a 1 ml min<sup>-1</sup> flow of HPLC-MS grade H<sub>2</sub>O + 0.1% formic acid (Sigma, CAS: 64-18-6). Measurements were taken in positive ion mode, with the instrument calibrated to a range of 50-1200 Da using sodium formate calibrant solution. Voltage of the capillary tip was set to 4800 V, end plate offset at -500V, funnel 1 RF and funnel 2 RF at 400 Vpp, hexapole RF at 100 Vpp, ion energy at 5.0 eV, collision energy at 5 eV, collision cell RF at 200 Vpp, transfer time at 100.0  $\mu\text{s}$  and pre-pulse storage time at 1.0  $\mu\text{s}$ .

### Data Plotting and Figure Editing

All statistical analyses and data plots were created using Origin Pro 2016 software.

### References

1. Baum, D. A. & Vetsigian, K. An experimental framework for generating evolvable chemical systems in the laboratory. *Orig. Life Evol. Biosph*, **47**, 481-497 (2016).
2. Felnagle, E. A., Chaubey, A., Noey, E. L., Houk, K. N. & Liao, J. C. Engineering synthetic recursive pathways to generate non-natural small molecules. *Nature Chemical Biology* **8**, 518-526 (2012).
3. 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. Chemie - Int. Ed.* **54**, 9871-9875 (2015).
4. Fitz, D., Reiner, H. & Rode, B. M. Chemical evolution toward the origin of life. *Pure Appl. Chem.* **79**, 2101-2117 (2007).
5. Markovitch, O. & Lancet, D. Excess mutual catalysis is required for effective evolvability. *Artif. Life* **18**, 243-66 (2012).
6. Nghe, P. *et al.* Prebiotic network evolution: six key parameters. *Mol. Biosyst.* **11**, 3206-17 (2015).
7. Powner, M. W., Gerland, B. & Sutherland, J. D. Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239-242 (2009).

8. Ruiz-Mirazo, K., Briones, C. & Escosura, A. de la. Prebiotic systems chemistry: new perspectives for the Origins of Life. *Chem. Rev.* **114**, 285–366 (2014).
9. Doran, D., Rodriguez-Garcia, M., Turk-Macleod, R., Cooper, G. J. T. & Cronin, L. A recursive microfluidic platform to explore the emergence of chemical evolution. *Beilstein J. Org. Chem.* **13**, 1702–1709 (2017).
10. Schuster, P. Taming combinatorial explosion. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 7678–80 (2000).
11. Ruiz-Mirazo, K. & Luisi, P. L. Origins of life and evolution of biospheres. *Orig. Life Evol. Biosph.* **40**, 353–497 (2010).
12. Peretó, J. Out of fuzzy chemistry: from prebiotic chemistry to metabolic networks. *Chem. Soc. Rev.* **41**, 5394 (2012).
13. Szymański, J. K., Abul-Hajja, Y. M. & Cronin, L. Exploring strategies to bias sequence in natural and synthetic oligomers and polymers. *Acc. Chem. Res.* **51**, 649–658 (2018).
14. Goodwin, J. A. Y. T., Mehta, A. K. & Lynn, D. G. Digital and analog chemical evolution. **45**, 2189–2199 (2012).
15. Forsythe, J. G. *et al.* Surveying the sequence diversity of model prebiotic peptides by mass spectrometry. *Proc. Natl. Acad. Sci.* **114**, E7652–E7659 (2017).
16. Mamajanov, I. *et al.* Ester formation and hydrolysis during wet-dry cycles: Generation of far-from-equilibrium polymers in a model prebiotic reaction. *Macromolecules* **47**, 1334–1343 (2014).
17. Surman, A. J. *et al.* Environmental control programs the emergence of distinct product ensembles from unconstrained chemical reaction networks. *ChemRxiv. Preprint. DOI: <https://doi.org/10.26434/chemrxiv.6962813.v1>* (2018).
18. Kawamura, K., Takeya, H. & Kushibe, T. Effect of condensation agents and minerals for oligopeptide formation under mild and hydrothermal conditions in related to chemical evolution of proteins. *Adv. Sp. Res.* **44**, 267–275 (2009).

19. Lahav, N., White, D. & Chang, S. Peptide formation in the prebiotic era: thermal condensation of glycine in fluctuating clay environments. *Science* **201**, 67–69 (1978).
20. Yu, S. S. *et al.* Elongation of model prebiotic proto-peptides by continuous monomer feeding. *Macromolecules* **50**, 9286–9294 (2017).
21. Ascenzi, P., Gioia, M., Fanali, G., Coletta, M. & Fasano, M. Pseudo-enzymatic hydrolysis of 4-nitrophenyl acetate by human serum albumin: PH-dependence of rates of individual steps. *Biochem. Biophys. Res. Commun.* **424**, 451-455 (2012).
22. Hershko, A., Ciechanover, A., Heller, H., Haas, A. L. & Rose, I. A. Proposed role of ATP in protein breakdown: conjugation of protein with multiple chains of the polypeptide of ATP-dependent proteolysis. *Proc. Natl. Acad. Sci.* **77**, 1783–1786 (1980).
23. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc.* **6**, 2876-2890 (2007).
24. Sreerama, N., Venyaminov, S. Y. U. & Woody, R. W. Estimation of the number of  $\alpha$ -helical and  $\beta$ -strand segments in proteins using circular dichroism spectroscopy. *Protein Sci.* **8**, 370-380 (1999).
25. Sojo, V., Herschy, B., Whicher, A., Camprubi, E. & Lane, N. The origin of life in alkaline hydrothermal vents. *Astrobiology* **16**, 181–197 (2016).

**Supplementary Information** is attached as a separate word document alongside main manuscript.

**Author Contributions:** LC conceived the experimental design including the metric, and coordinated the team with YMA-H. DD carried out all reactions, mass spectrometry analysis and structural and functional assays, and wrote Python 3.6 scripts for adduct screening and calculating monomer composition. YMA-H was instrumental in guiding the work, providing expertise in peptide chemistry and, together with DD, developed project ideas. DD, YMA-H and LC wrote the manuscript.

**Author Information** Correspondence and requests for materials should be addressed to Leroy Cronin ([Lee.Cronin@glasgow.ac.uk](mailto:Lee.Cronin@glasgow.ac.uk)).

**Competing Interests** The authors declare no competing interests.

**Acknowledgments** 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 The John Templeton Foundation Grant ID 60625 and Grant ID 61184. The authors would like also to thank Dr. Maria Diana Castro-Spencer and Dr. Jennifer Mathieson for assistance with mass spectrometry data acquisition, Dr. Andrew Surman for writing R scripts used for mass spec data extraction, Dr. Cole Mathis for commenting on the manuscript, Dr. Sebastian Steiner and Dr Jan Szymański for hardware and coding assistance, respectively, and Dr. Piotr S. Gromski for conceiving an initial version of the mass spec index.

## SUPPLEMENTARY INFORMATION

# Emergence of Function and Selection from Recursively Programmed Polymerisation Reactions in Mineral Environments

David Doran, Yousef M. Abul-Haija and Leroy Cronin\*

*WestCHEM, School of Chemistry, University of Glasgow, Glasgow G12 8QQ, UK*

\*Corresponding author: Lee.Cronin@glasgow.ac.uk

### Table of contents

<b>1. Mass Spectrometry Data Analysis.....</b>	<b>2</b>
<b>1.1 Depsipeptide Library Screening and Mass Index.....</b>	<b>2</b>
<b>1.2 Product Composition.....</b>	<b>6</b>
<b>2. Esterase Activity Assay.....</b>	<b>9</b>
<b>3. Structural Characterisation.....</b>	<b>12</b>
<b>3.1 Circular Dichroism.....</b>	<b>12</b>
<b>3.2 Fourier Transform Infrared Spectroscopy.....</b>	<b>13</b>
<b>4. Minerals.....</b>	<b>14</b>
<b>5. Wet-Dry Cycling Reactions.....</b>	<b>15</b>

## 1. Mass Spectrometry Data Analysis

### 1.1 Depsiptide Library Screening and Mass Index

The complete mass list of 6363 depsipeptide products was compiled in Python 3.6. All possible compositions arising from leucine (L), glutamic acid (E), lysine (K) and glycolic acid (g) were calculated from a dictionary of monomer masses read from a .json file. Branched and cyclic products were screened for by removing one water mass (18.01056 Da) for every proposed branching point. Cationic adducts of each branched, cyclic and linear product were accounted for by adding the following masses: 1.007276 (H<sup>+</sup>), 22.989 (Na<sup>+</sup>), 38.963 (K<sup>+</sup>), 18.034 (NH<sub>4</sub><sup>+</sup>). The mass list was compiled as a Python dictionary, with keys corresponding to a string of monomer units plus dehydrations and adducts. Dehydrations were added in the string format “- <n> H<sub>2</sub>O” and adducts were added in the string format “+ <i>”, where n and i correspond to number of dehydrations and cationic adduct, respectively. For example, “LEK – 1 H<sub>2</sub>O + Na” corresponds to a trimer of one L, one E and one K monomer with one extra dehydration (a potential cycling or branching point) plus a sodium adduct in +1 charge state.

```
{
  "E": 129.043,
  "K": 146.10557,
  "L": 131.09467,
  "g": 76.0160
}
```

**Supplementary Figure 1:** Monomer Unit Mass Dictionary. Each potential monomer unit plus masses were stored as a dictionary, with keys corresponding to standard amino acid one letter codes (upper case) or lower-case letter “g” corresponding to the glycolic acid monomer unit. Standard, upper case one letter codes were used for amino acids: “E” = glutamic acid; “K” = lysine; “L” = leucine.

```
n_dhr = 8
monomers = ["L", "E", "K", "g"]
# define paths to input and output JSON file
input_file_path = "C:\\Users\\croningp\\PycharmProjects\\Mass_Spec_Mass_Lists\\Amino
Acid_Mass_Dictionary.json"
output_file_path_full = "C:\\Users\\croningp\\Documents\\MS_Analysis\\Mass_Lists\\DEPSI-
Peptide\\{0}\\{1}\\{2}\\_13012018.json".format(monomers[0], monomers[1], monomers[2])
dhr_output_file_path_full = "C:\\Users\\croningp\\Documents\\MS_Analysis\\Mass
Lists\\DEPSI-Peptide\\{0}\\{1}\\{2}\\{3}\\dhr13012018.json".format(monomers[0], monomers[1],
monomers[2], n_dhr)

# define maximum oligomer length
max_length = 8
water = 18.010565
H = 1.007276

adduct_dict = {"Na": 22.989218, "K": 38.963158, "NH4": 18.033823}
```

**Supplementary Figure 2:** Mass List Compiler. Maximum number of dehydrations per oligomer is set by variable n\_dhr, monomers used for compilation are added to monomers list, and read from monomer mass dictionary (Supplementary Fig. 1). Maximum length of oligomers is set by max\_length variable. Masses of water and H<sup>+</sup> adduct are defined by water

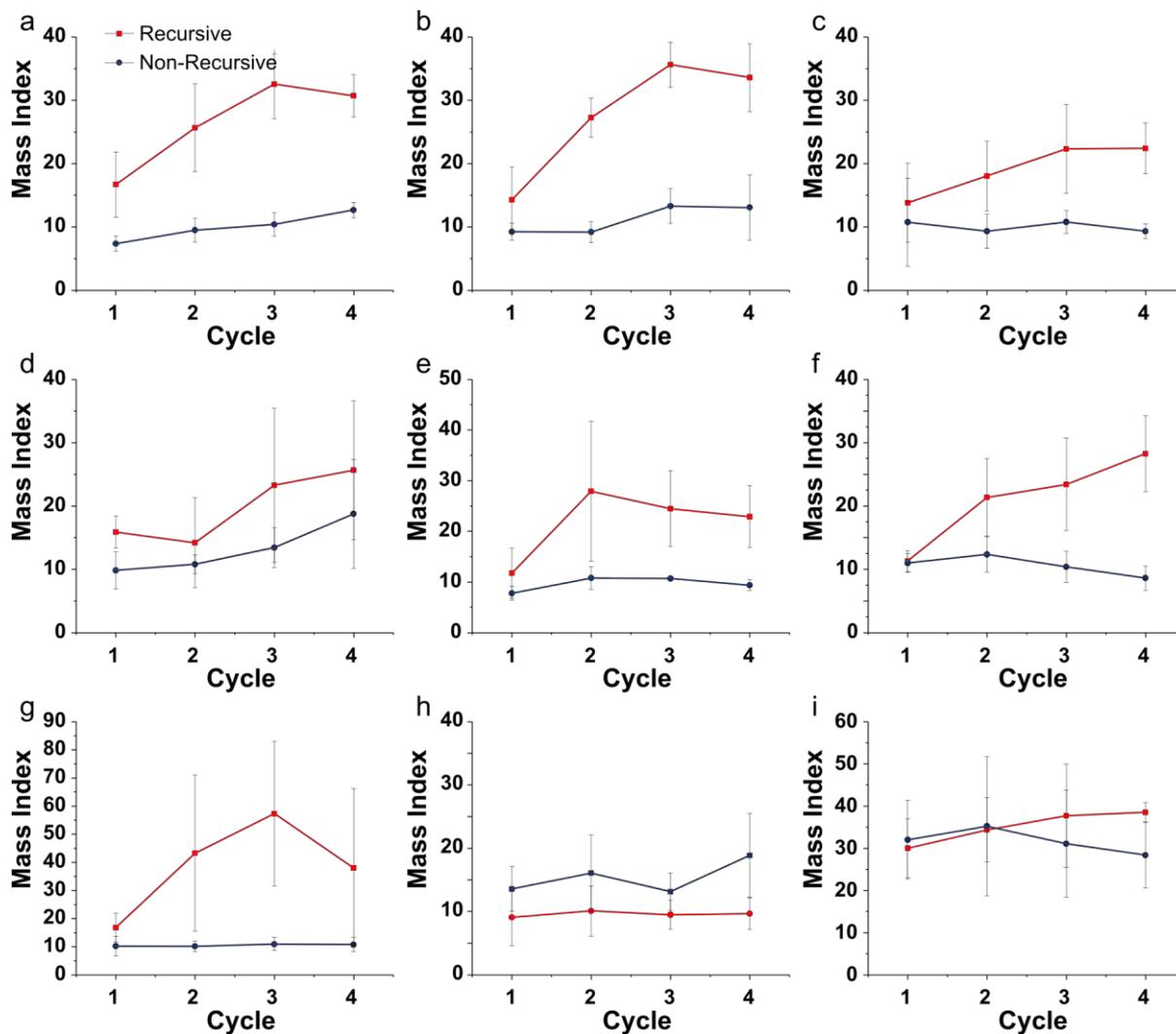
and H variables. Additional adducts are stored as a dictionary `adduct_dict`, with corresponding masses.

Mass lists in .csv format were read by a script written in R 2.7 and run in R-studio. Extracted ion chromatograms (EICs) for each mass were extracted from MS-1 data stored in 32-bit mzml files, which were generated from raw data using Proteowizard MS Convert. Masses were extracted with an error threshold of  $\pm 0.01$ . Total intensities for each EIC were stored in csv files. The first column of each csv file contained a list of depsipeptide composition strings, each subsequent column contained corresponding intensities for depsipeptide products for each sample. A small example data set is presented in Table S1. A noise threshold of  $2\sigma$  (2.5% of maximum intensity) was applied to this data. Noise filtered data was used for the Mass Index measurement.

**Supplementary Table 1:** Example Data Set for Mass Index Measurement. The top row of each column contains headers, the first of which denotes the column containing m/z values; each subsequent column header denotes the name of a sample measurement file. The first column (left) contains a list of m/z values, each corresponding to a species in the mass library. Subsequent columns contain intensities obtained in the mass spectrometry measurement of each sample corresponding to that m / z value.

m/z	No Mineral_Cycle1_A	No Mineral_Cycle1_B	No Mineral_Cycle1_C
130.049	266141.711	271363.245	249274.3635
132.101	864710.269	863549.7875	854975.3445
135.029	0	0	0
136.076	0	0	0
139.003	0	0	0
147.112	328543.4595	314114.8865	285210.6015
148.060	166482.322	152707.5075	140961.1295
151.087	0	0	0
152.034	62206.188	56934.0725	57648.0905

The Mass Index was used as a metric for assessing the effect of recursion on the reaction carried out in nine mineral environments (Supplementary Fig. 3).



**Supplementary Figure 3.** Mass Index by cycle for recursive and non-recursive depsipeptide product mixtures reacted on various solid surface matrices: A) no mineral; B) crushed glass; C) montmorillonite; D) gypsum; E) quartz; F) calcite; G) chalcopyrite; H) opal; I) kernite. Data represent mean of 9 replicates  $\pm$  2 S.D.

## 1.2 Product Composition

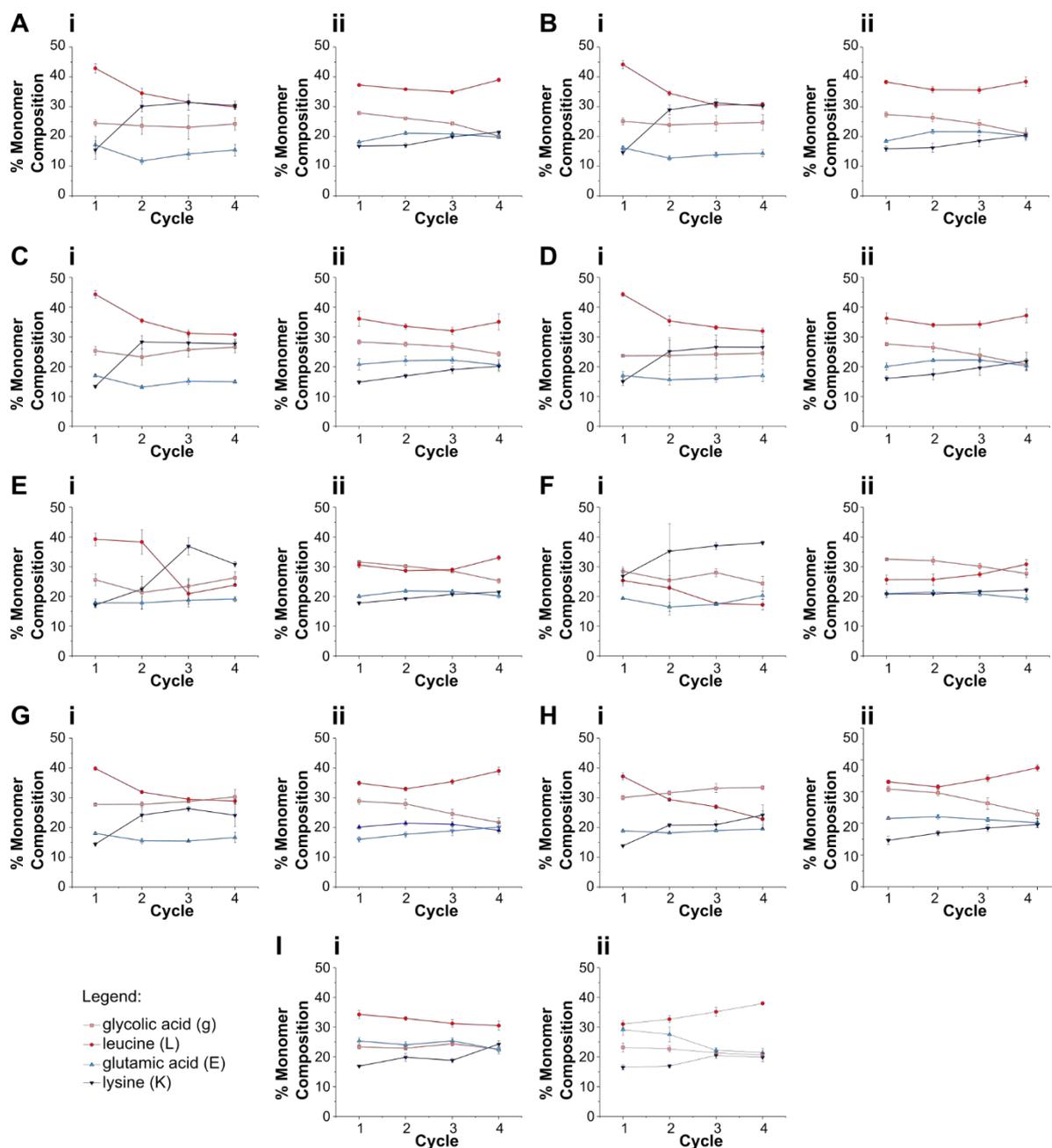
Monomer composition measurements (Figure 3, Supplementary Fig. 4) were obtained from csv files containing depsipeptide composition strings and their corresponding intensities. The following steps were carried to extract monomer composition:

1. Iteration through strings corresponding to depsipeptide compositions, removing extra adducts but keeping dehydrations.
2. For each composition product, addition of the intensities for each adduct.
3. For each composition string, calculation of the relative ratio of each monomer string (“L”, “E”, “K”, “g”) as a decimal fraction, excluding extra string characters for dehydration and whitespace.
4. Multiplication of each of these fractions by the total intensity measured for the composition product.
5. Addition of total monomer intensities over the entire library of 6363 compositions.
6. For each total monomer intensity, division by the total product intensity and multiplication by 100 to obtain % monomer intensity.

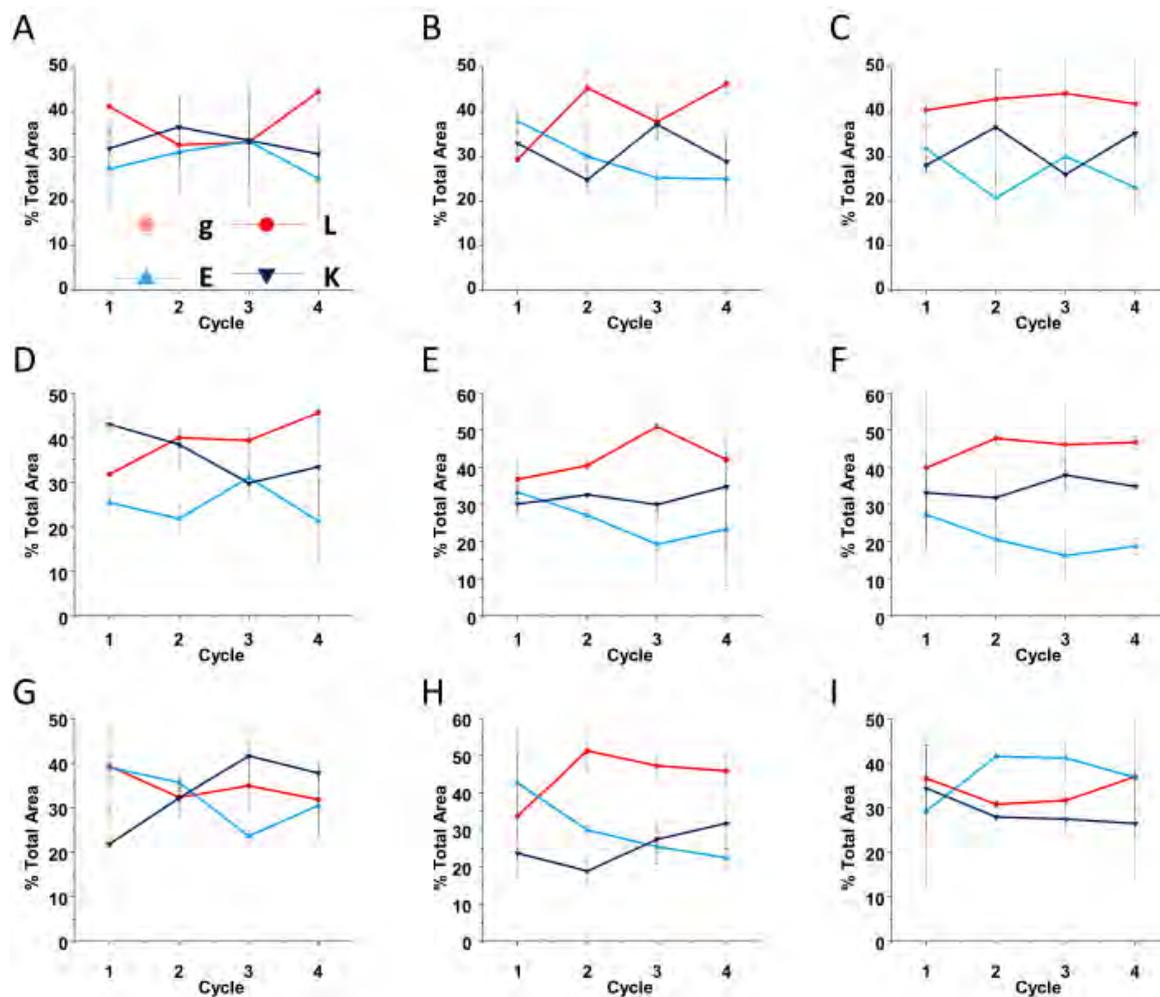
An example of csv data input into the monomer composition script is presented in Table S2. Monomer composition values for mineral environments not shown in the main paper (Figure 3) are given below, in Supplementary Fig. 4.

**Supplementary Table 2:** Combined Adduct Intensity Data for Monomer Composition Extraction. Top table shows depsipeptide strings corresponding to a leucine (L) trimer with one extra dehydration plus adducts (K, NH<sub>4</sub> and Na) with their corresponding intensity values for three samples. Lower table (highlighted in yellow) shows the combined intensity of all adduct species. These combined intensity values were used to calculate percentage intensity contribution of monomer species.

<b>Product Composition</b>	<b>No Mineral_Cycle1_A</b>	<b>No Mineral_Cycle1_B</b>	<b>No Mineral_Cycle1_C</b>
LLL -1 H <sub>2</sub> O	119798.404	94274.3355	85355.094
LLL -1 H <sub>2</sub> O + K	74732.388	80836.3325	73497.701
LLL -1 H <sub>2</sub> O + NH <sub>4</sub>	0	0	0
LLL -1 H <sub>2</sub> O + Na	0	0	0
<b>Product</b>	<b>No Mineral_Cycle1_A</b>	<b>No Mineral_Cycle1_B</b>	<b>No Mineral_Cycle1_B</b>
LLL -1 H <sub>2</sub> O	194530.792	175110.668	158852.795



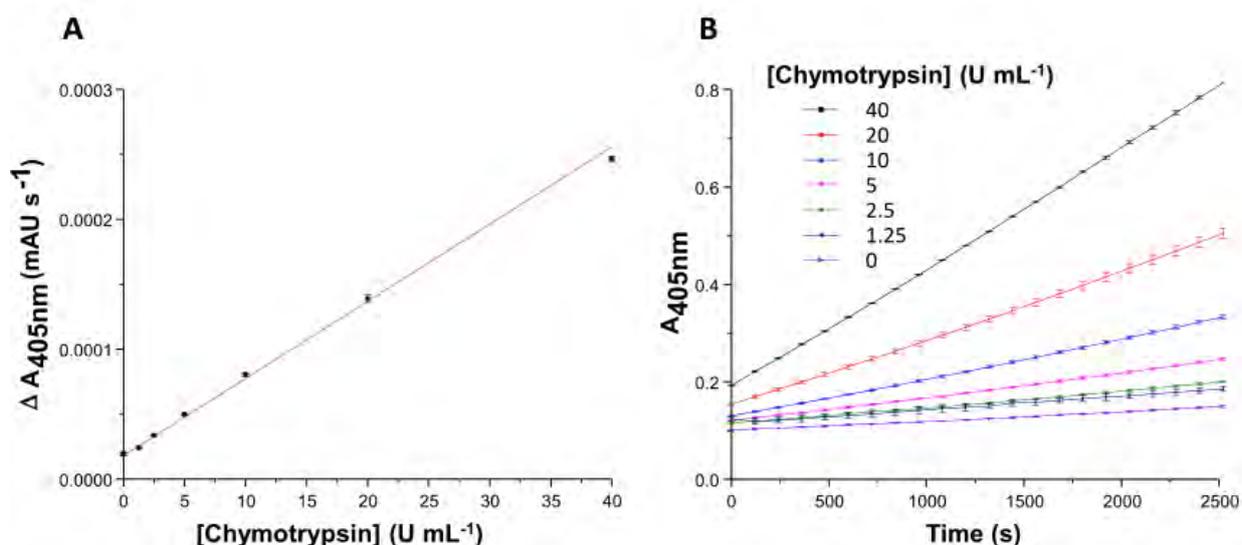
**Supplementary Figure 4: Monomer Composition of Depsipeptide Products over Multiple Recursive Cycles.** Intensity contribution to products as measured by mass spectrometry are shown for depsipeptide reactions carried out on: **A)** no mineral; **B)** crushed glass; **C)** montmorillonite; **D)** gypsum; **E)** quartz; **F)** calcite; **G)** chalcopyrite; **H)** opal; and **I)** kernite for (i) recursive and (ii) non-recursive products. Data points represent the mean of 9 replicates  $\pm$  1 S.D. g = glycolic acid; L = leucine; E = glutamic acid; K = lysine.



**Supplementary Figure 5:** Relative Free Amino Acid Concentration over Multiple Recursive Cycles. Relative concentration of free Leucine (L), glutamic acid (E) and lysine (K) are shown for depsipeptide reactions carried out on: **A)** no mineral, **B)** crushed glass, **C)** montmorillonite, **D)** gypsum, **E)** quartz, **F)** calcite, **G)** chalcopyrite, **H)** opal, **I)** kernite.

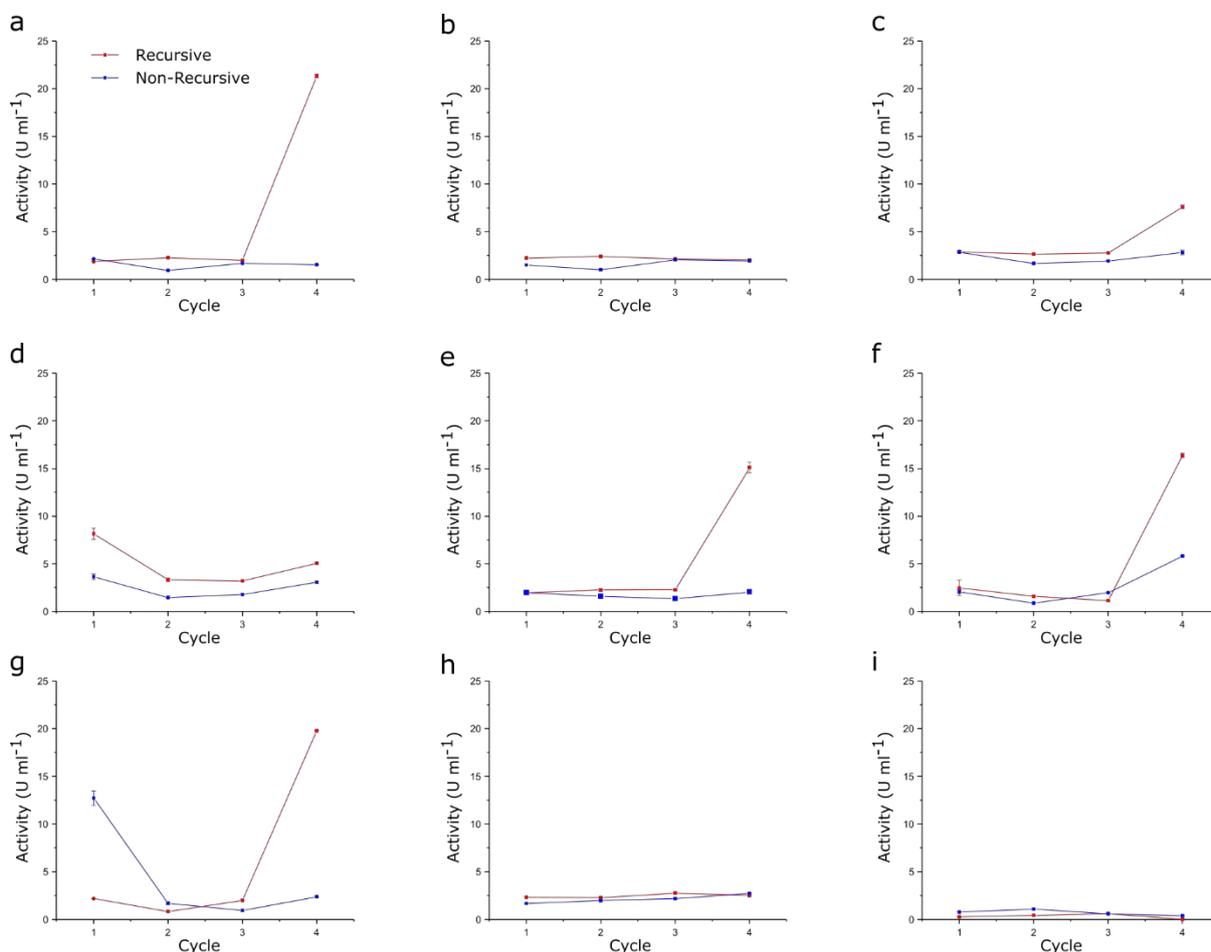
## 2. Esterase Activity Assay

Esterase activity of product mixtures was normalised to an  $\alpha$ -chymotrypsin standard curve (Supplementary Fig. 6a). Rate of *p*NPA hydrolysis was measured by rate of increase in absorbance at 405 nm (Supplementary Fig. 6b). To calculate the esterase activity of products, rate of increase in absorbance at 405 nm was substituted into the equation of the standard curve fit in Supplementary Fig. 6A.



**Supplementary Figure 6:** Chymotrypsin Standards for *p*NPA Hydrolytic Assay. A) Rate of *p*NPA hydrolysis (measured by increase in absorbance at 405 nm) plotted for  $\alpha$ -chymotrypsin standards at 40, 20, 10, 5, 2.5, 1.25 and 0 enzyme units (U) per ml. Equation of curve fit:  $y = mx + c$ , where  $y = \Delta A_{405\text{ nm}}$ ,  $m = 5.9793 \times 10^{-6}$ ,  $x = [\text{chymotrypsin}]$ ,  $c = 1.82694 \times 10^{-5}$ . B) Raw data from *p*NPA hydrolytic assay for  $\alpha$ -chymotrypsin standards showing absorbance at 405 nm over time.

The esterase activities of all recursive and non-recursive product mixtures were ran for cycles 1-4. Recursive product mixtures showed a trend towards increased activity from cycles 1 to 4 (Supplementary Fig. 8). Esterase activity of product mixtures exceeded that of the controls, and also unreacting starting material (main text Figure 4), demonstrating that *p*NPA hydrolysis was not due to hydrolysis by H<sub>3</sub>PO<sub>4</sub> or monomers.

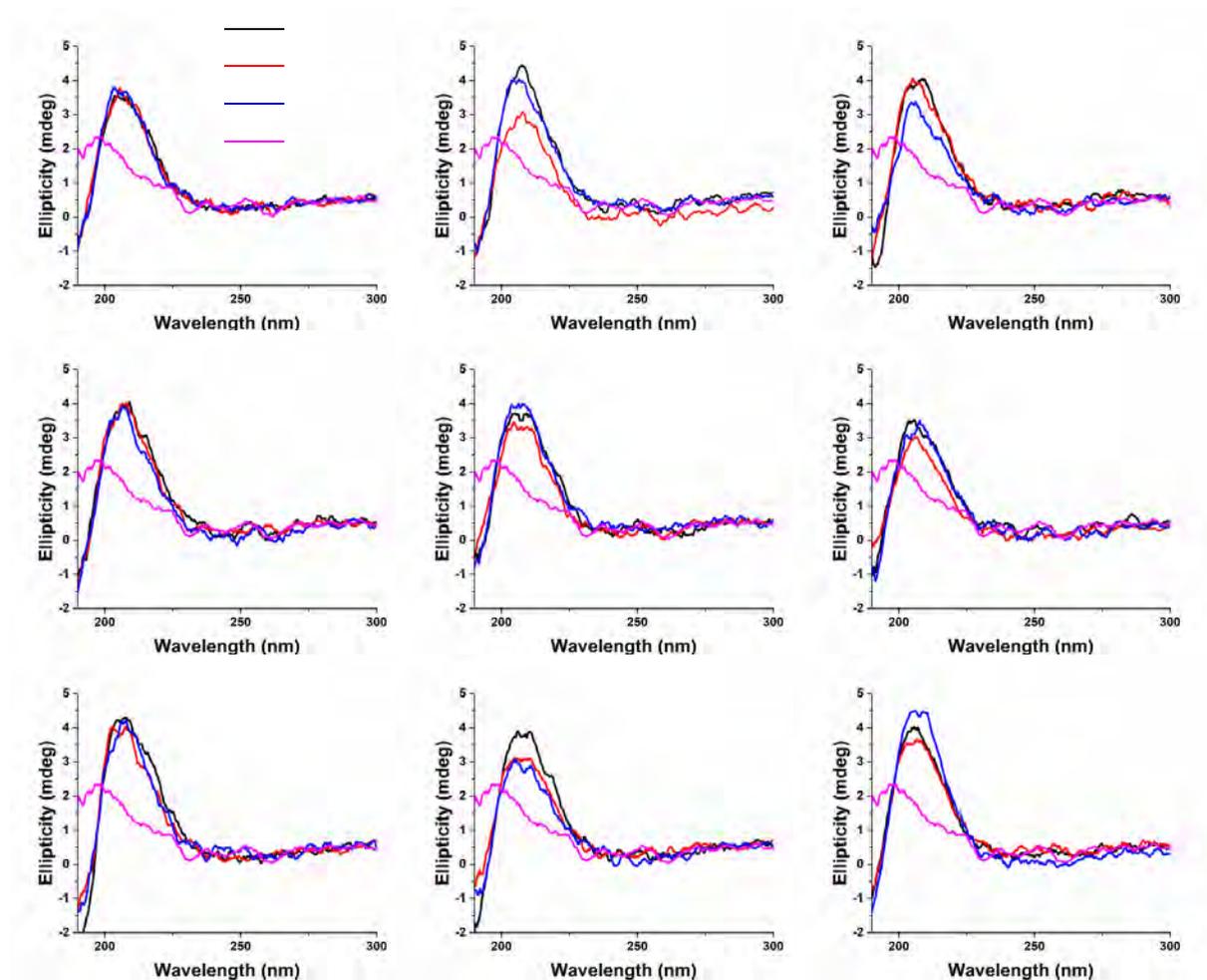


**Supplementary Figure 7:** Esterase Activity of Depsipeptide Products Measured by *p*NPA Hydrolytic Assay. Activity was measured in enzyme units (U) per ml for product mixtures after 1-4 cycles carried out on: A) no mineral; B) crushed glass; C) montmorillonite; D) gypsum; E) quartz; F) calcite; G) chalcopryrite; H) opal; I) kernite for both recursive (red) and non-recursive (blue) product mixtures. Data show mean of 2 replicates + / - 2 S.D.

### 3. Structural Characterisation

#### 3.1 Circular Dichroism

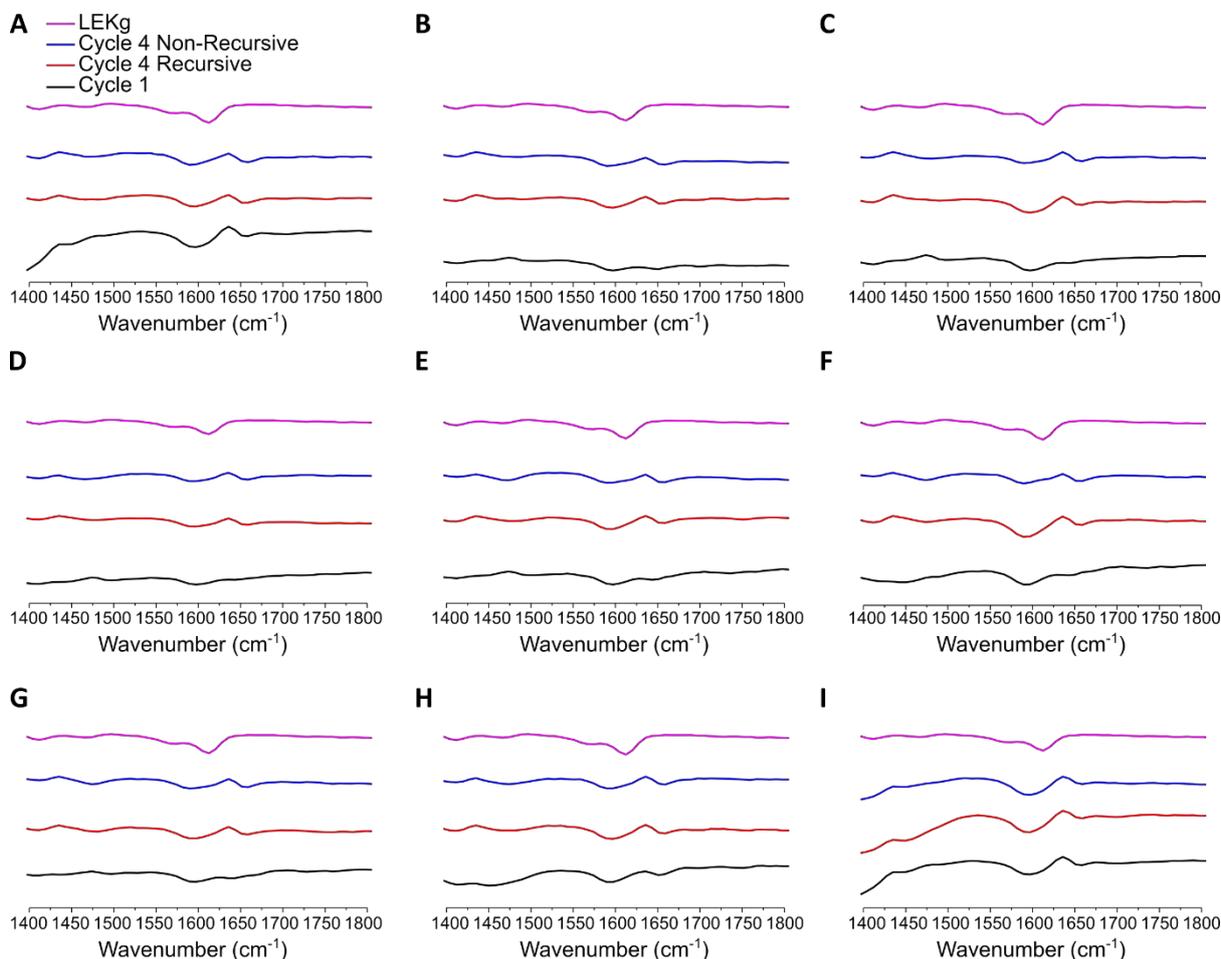
Circular dichroism (CD) was used to screen for secondary structures in recursive product mixtures. CD spectra of product mixtures were taken after cycles 1, 3 and 5, and compared to unreacted starting material. Data in Supplementary Fig. 9 show some form of secondary structure - possibly  $\beta$ -sheets, which are common for lysine-rich sequences and also hydrophobic peptides.



**Supplementary Figure 8:** Circular Dichroism of Recursive L, E, K and g Depsipeptide Products. CD measurements were taken of products from cycles 1, 3 and 5 and unreacted starting material (S.M.) mixture at equivalent concentration from A) no mineral; B) crushed glass; C) montmorillonite; D) gypsum; E) calcite; F) chalcopryrite; G) opal; H) kernite.  $\beta$ -sheet profile is found in all product mixtures, but not unreacted starting material.

### 3.2 Fourier Transform Infrared Spectroscopy

FTIR measurements (Supplementary Fig. 10) were taken to confirm the presence of secondary structures measured by CD.



**Supplementary Figure 9:** Fourier Transform Infra-Red (FTIR) Spectroscopy of Depsiptide Mixtures. FTIR absorption spectra of product mixtures are shown after 1 cycle (black) and 4 recursive (red) and non-recursive (magenta) cycles, with starting material control (magenta) of L-Leucine (L), L-glutamic acid E, L-Lysine (K) and glycolic acid (g). A = no mineral; B = crushed glass; C = montmorillonite; D = gypsum; E = quartz; F = calcite; G = chalcopyrite; H = opal; I = kernite. Data represent mean of 2 measurements + / - 2 S.D.

#### 4. Minerals

Minerals were sourced from Richard Tayler minerals. Composition, structure and water solubility are given below in Table S3.

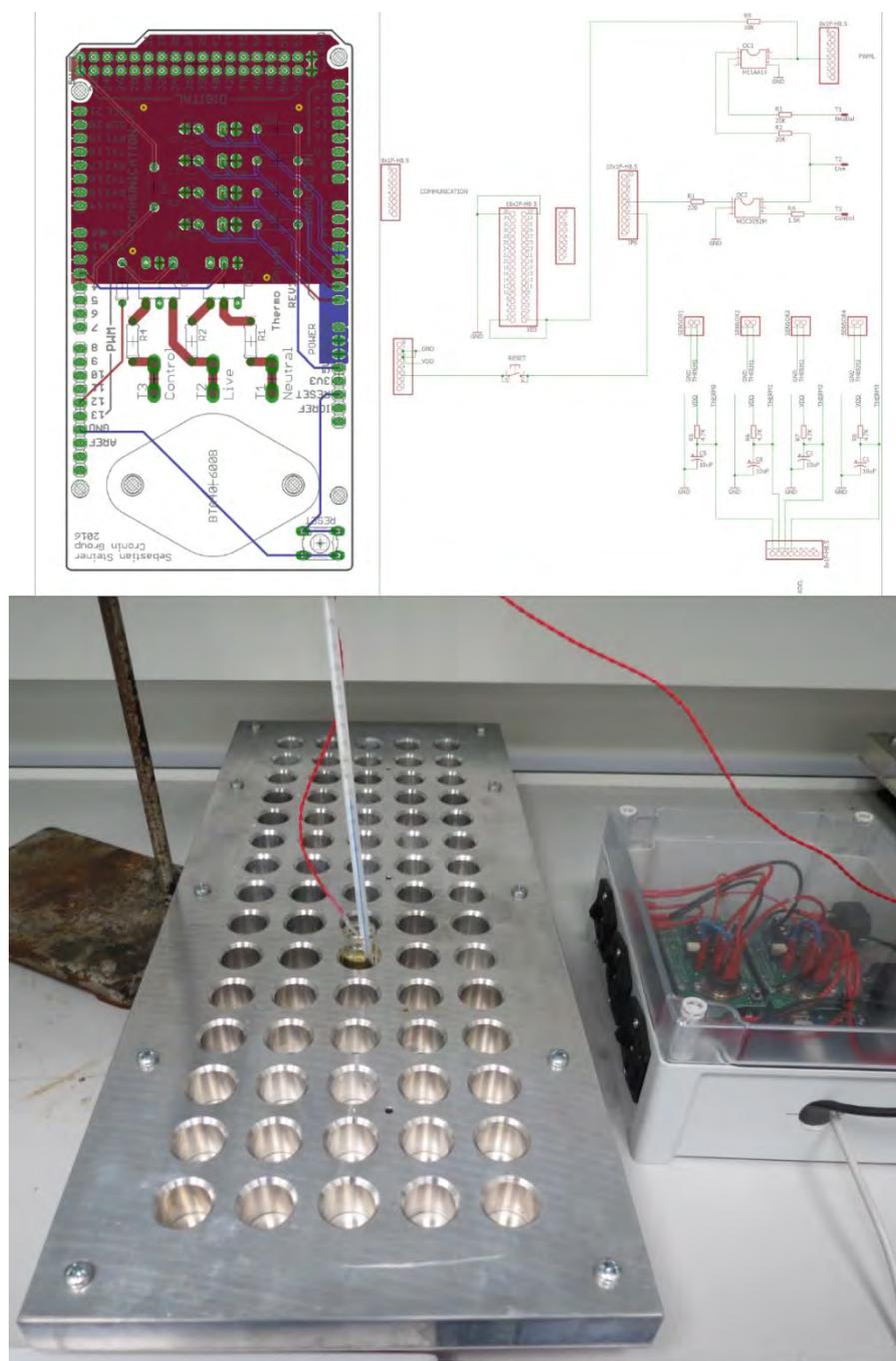
**Supplementary Table 3:** Chemical and Physical Properties of Minerals used for Solid Surface Matrices.

Mineral	Composition	Structure	Water Soluble?
Montmorillonite	$(\text{Na,Ca})_{0.33}(\text{Al,Mg})_2(\text{Si}_4\text{O}_{10})(\text{OH})_2 \cdot n\text{H}_2\text{O}$	Crystalline - monoclinic prismatic	No
Gypsum	$\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$	Crystalline - monoclinic prismatic	2-2.5 g / L
Calcite	$\text{CaCO}_3$	Crystalline - hexagonal scalenohedral	Highly
Chalcopyrite	$\text{CuFeS}_2$	Crystalline - tetragonal scalenohedral	No
Quartz	$\text{SiO}_2$	Crystalline - trigonal or hexagonal	No
Opal	$\text{SiO}_2 \cdot n\text{H}_2\text{O}$	Amorphous	No
Crushed glass	$\text{SiO}_2$	Amorphous	No
Kernite	$\text{Na}_2\text{B}_4\text{O}_6(\text{OH})_2 \cdot 3\text{H}_2\text{O}$	Crystalline - monoclinic prismatic	Yes

#### 5. Wet-Dry Cycling Reactions

Wet-dry cycling reactions were carried out in 15 ml glass vials placed on to 75-well heating slabs (Supplementary Fig. 11). The temperature of the heating slabs was set to and maintained at 90 °C via a custom PID controller designed and built in-house, the ThermoShield. Fine power control of resistive loads such as the silicone heating mat employed in this project is achieved via phase angle control utilizing a TRIAC. An EPCOS B57560G104F NTC Thermistor submerged in a vial filled with mineral oil (CAS: 8042-47-5, Sigma) was used to monitor the current temperature. A custom PCB featuring circuitry for reading the NTC, sensing the AC zero crossing, and firing the TRIAC was designed to fit on top of an Arduino Mega 2560 microcontroller board. A simple firmware utilizing the Arduino PID library (<http://playground.arduino.cc/Code/PIDLibrary>) provides communication and control. Since no complicated temperature programs were required, the ThermoShield was controlled through the terminal program Termite ([https://www.compuphase.com/software\\_termite.htm](https://www.compuphase.com/software_termite.htm)). Temperature was monitored both through Termite and through a thermometer placed in the same vial.

Gerber files, Bill of Materials, Arduino sketch and documentation can be found on github:  
<http://datalore.chem.gla.ac.uk/Origins/ThermoShield.git>



**Supplementary Figure 10:** Temperature Control System for Wet-Dry Cycling Reactions. Thermistor was placed in a vial of mineral oil in the centre of the heating slab. Temperature was continuously monitored throughout the 15 hour heating cycles, through Termite and manually using a standard analogue thermometer inserted into the vial of mineral oil.