ChemSystemsChem

Chemistry Europe
European Chemical Societies Publishing

Accepted Article

Title: Evidence of Selection in Mineral Mediated Polymerization Reactions Executed in a Robotic Chemputer System

Authors: Silke Asche, Robert Pow, Hessam Mehr, Geoff Cooper, Abhishek Sharma, and Leroy Cronin

To be cited as: *ChemSystemsChem* **2024**, e202400006

Link to VoR: https://doi.org/10.1002/syst.202400006

WILEY VCH

10.1002/syst.202400006

WILEY VCH

RESEARCH ARTICLE

Evidence of Selection in Mineral Mediated Polymerization Reactions Executed in a Robotic Chemputer System

Silke Asche, Robert W. Pow, Hessam M. Mehr, Geoffrey J. T. Cooper, Abhishek Sharma and Leroy Cronin*

Prof. Lee Cronin
School of Chemistry
The University of Glasgow
Advanced Research Centre, University Avenue, Glasgow G12 8QQ, UK
E-mail: lee.cronin@glasgow.ac.uk

Supporting information for this article is given via a link at the end of the document.

Abstract: It has long been thought that abiogenesis requires a process of selection and evolution at the molecular level, but this process is hard to explore experimentally. One solution could be the use of automation in experiments which could allow for traceability and the ability to explore a larger reaction space. We report a fully programmable and automated platform to explore the reactions of amino acids in the presence of mineral environments. The robotic system is based upon the Chemputer system which has well defined modules, software, and a chemical programming language to orchestrate the chemical processes including analysis. The reaction mixtures were analysed with tandem mass spectrometry and a peptide sequencing algorithm. Each experiment was screened for 1,398,100 possible unique sequences, and more than 550 specifically defined sequences were confirmed experimentally. This work aimed to develop a new understanding of selection in repeated cycles of polymerisation reactions to explore the emergence of well-defined amino acid sequences. We found that the outcome of oligomerisation was significantly influenced by the presence of different minerals, and that a serpentine environment selects glycine and phenylalanine rich fragments that enable the formation of longer oligomers with well-defined sequences as a function of cycle number.

Introduction

The process of abiogenesis that gave rise the emergence of living systems on earth is the subject of much debate but remains difficult to investigate because of the highly contingent nature of biochemistry.[1-4] A central question is which processes are responsible for the emergence of the molecular machinery of biology, and what mechanisms generated this complexity before biological selection and evolution. It has been hypothesised that evolution started at the molecular level. Bottom-up approaches in Origin(s) of Life (OoL) studies aim to understand how small molecules can give rise to highly functional complex reaction networks with features like autocatalysis, goal-directedness, and information storage . While there have been advances in the study of these phenomena, the mechanisms by which these aspects emerged, separately or together, are not understood.[4-6] While selection is a well-defined phenomenon in biology and an essential element of biological evolution,[7] it is not clear how selection emerges on a molecular scale.[8-10] This is because it has been hard to develop hypotheses in OoL experiments that cover a reasonable parameter space,[1] but we hypothesize that oligomerisation reactions could provide a unique model system for the investigation of complex chemical processes.[11,12] Peptides are long chains of amino acids and are one of the key components of living systems and a vital part of biology. The fundamental problem in studying evolutionary effects on a molecular scale is the lack of product space knowledge in complex systems. Peptides are well studied and couple / decouple in a predictable way, making them ideal for this type of study. Furthermore, recent advances have shown that peptides could have been the first molecules to store information during the process which led to the emergence of life.[11-14] This is because information storage on a molecular scale is necessary for process optimisation and developing functionality within the chemical system.

Efforts have been made to study the influence of mineral environments in peptide formation reactions, since minerals were abundant on the early earth[15,16] and may have been the 'beaker' as well as the catalyst for the OoL.[17-19] Clay minerals have been of a particular focus for those studies due to their structural complexity,[20,21] but also iron-containing minerals were shown to assist peptide formation if present in the reaction system.[22,23] While those studies attempted to explore the influence of mineral environments on peptide formation reactions, more work is required to understand the underlying processes that occur in these chemical systems. Further, selection is not deterministic and cannot be predicted by theoretical models,[24] which makes it necessary to use a data-driven approach when studying the emergence of goal-directedness or selective effects on a molecular level.

This study presents an investigation of the emergence of selection at the chemical scale in a model system: an oligomerisation reaction using L-phenylalanine, glycine, L-leucine and L-histidine, with CDI as a coupling agent, in the presence of different mineral environments. The experiments were performed on an automated platform derived from the Chemputer concept,[25,26] purpose-built for OoL peptide coupling experiments. In this system the minerals act as 'natural' catalysts and provide a heterogenous system to facilitate reactions, adhesion effects to sequester produced species, and may even template oligomerisation processes directly. These different reaction surfaces may further influence chemical selection through their unique properties. All our reactions were performed through recursive cycling over a prolonged experimental period. In recursive cycles, the majority of the reaction mixture is removed at the end of the cycle and replaced with fresh starting materials, but a small portion of the previous reaction mixture is left to seed further reaction in the

10.1002/syst.202400006

following cycle. This method drives the reaction out of equilibrium in a cyclic manner by exchanging an equivalent of product solution with fresh starting material after every reaction cycle. An analytical algorithm[27] was used to resolve the product space based on the known fragmentation routes of peptides, and a comprehensive analytical workflow allowed the detailed breakdown of the emerging sequences. This, in turn, allowed the investigation of the product space and tracking of the build-up of certain peptide sequences through the experimental cycles. This, in turn, allowed the investigation of the product space and tracking of the build-up of certain peptide sequences through the experimental cycles.

Results and Discussion

In this work, we set out to design chemical experiments that enable the emergence of selection in a polymerisation system. The challenge of observing selection is that it can only be proven by demonstrating that certain products are formed while other products are specifically not formed (Figure 1b). In order for selection to take place, the chemical system must show a certain degree of observable control over the formation of product species. In this regard we suggest that selection enables the production of complex molecules in a very high dimensional combinatorial space, without causing a combinatorial explosion.[24]

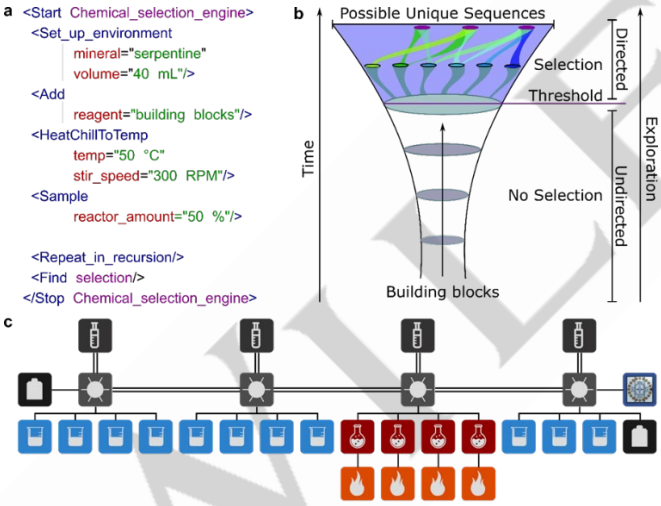


Figure 1. Graphic representation of the conceptual approach. (a) shows pseudo chemical or 'χDL' code for a Chemical Selection Engine (CSE) experiment. (b) represents the experimental approach, searching for selection. The cone represents the possible combinatorial sequence space, while only directed selection can lead to certain sequences forming that undirected reactivity cannot reach. (c) is a representation of the hardware modules in the OoL Chemputer platform. The different modules are depicted as coloured icons, with pumps in dark grey, valves in grey, waste containers in black, input solutions and solvents in blue, reactors in red, stirrer hotplates in orange and the sample wheel in dark blue.

We hypothesise that specific sequences can hold 'information' towards complex molecules necessary for the emergence of functionality. Selectivity can arise in a chemical system over time, and careful observation of the product space can reveal the extent of directedness in its exploration, and thus allow quantification of selection. To further understand information storage and selection, we studied the formation of oligomer sequences influenced by environmental factors in an automated recursive reaction system, the Chemical Selection Engine (CSE). This platform system was derived from the Chemputer, an automated platform system developed for synthesising organic molecules. The Chemputer[25,26,28] is operated by its programming language, χDL (Figure 1a) and hardware is represented through an abstract graph (Figure 1c). We have leveraged the Chemputer system's functionalities and customised it to a platform purposefully built for OoL experiments (Figure 2a).

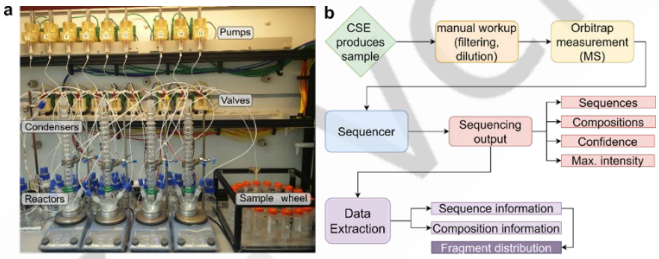


Figure 2. Experimental implementation of CSE. (a) the photograph shows the chemical selection engine setup. The top shelf holds pumps while the second row holds valves. Below are four reactors fitted with air condensers. On the right is the sample wheel with 20 sample spaces. (b) describes the analytical workflow. Once the CSE produces samples, they are manually filtered and diluted before measured via LC-MS on the Thermo Orbitrap Fusion Lumos Tribrid. After analytical measurement, the data is processed to obtain sequence information, compositions, their confidence and the maximum intensity of the parent peaks correlating to the compositions found. From there the data extractor, an additional Python script, calculates further sequence and composition information, such as the fragment distribution in the product space.

The chemical system we chose for this study was an amino acid oligomerisation reaction using a coupling agent, requiring only relatively simple molecular building blocks, and providing a complex, fully resolvable combinatorial library. The building blocks chosen were glycine, L-leucine, L-histidine, and L-phenylalanine. Glycine is a building block that readily couples to other amino acids, and while many amino acid studies have been published with glycine (and alanine) due to their high-water solubility and high affinity to produce oligomers, we also wanted to include more structurally and functionally interesting candidates. L-leucine and L-phenylalanine were chosen because they have hydrophobic side chains, and L-histidine is positively charged, allowing them to impart a greater range of potential functionality to the oligomers produced. Importantly, all these amino acids only yield unbranched linear sequences, as this is a currently a constraint of our analytical sequencing software. The product space was analysed via tandem mass spectrometry (MS/MS), and the resulting sequences and fragment distribution of the product species over time were obtained computationally. This data workflow (Figure 2b) provided knowledge of the product space, including composition and sequence information and the intensity of the parent MS peaks.

We used an in-house developed *de novo* sequencing tool written in Python, 'OLIGOSS',[27] (see supplementary information) to extract sequence data and calculate fragment distributions. The data extractor code generates a library of theoretical oligomer products based on the user input, such as monomer type and maximal oligomer length. This library is compared to the MS data, including the parent species and the fragments generated through MS/MS. This process ensured that each product found was confirmed by several fragments, and since only theoretically calculated products were considered, the possibility of false detections was minimised. The sequencing output was the number of unique sequences found, the number of compositions, a confidence value for the oligomers based on the number of ions confirmed, and an intensity value for the most intense parent peak for each composition.

The workflow using de novo sequencing coupled with tandem MS enables the full analysis of the oligomer product space. In the initial search, 1,398,100 unique sequences were considered. This is the theoretical maximum number of unique products in a system with 4 amino acids and peptide species of up to a decamer in length. Since the longest observed species in the experimental system turned out to be heptapeptides, the chemical space for our model system appeared to include up to 21,844 possible unique sequences. We were also able to infer which conditions led to the production of specific sequences. As such, we were able to find the differences in the produced sequence fragments over cycles based on the mineral environment present in the reactor. This study gives insight into the examined chemical system, identifying every single oligomer in the product space but further applying high-level system distribution analysis, which means the product space is studied both on a systems level and on an individual sequence level.

Recursive cycling[3] was used to move the system out of equilibrium in cycles – 20 cycles of 4 hours were performed for each mineral, with 75% of the total volume being removed / refreshed on each cycle. The experiment was performed in a batch reactor charged with different environments, but after each full cycle was completed, a proportion of the product mixture (50%) was removed from the reactor and replenished with an equal volume of fresh starting material. Through this recursive experimental approach, the reaction was repeatedly moved out of equilibrium by combining fresh reactants with the complex mixture left from the previous cycle(s). Additionally, we retained the mineral environment between each experiment and only exchanged product solutions by adding new solutions containing the amino acid building blocks and the coupling agent. This workflow also naturally provided routine time points for sample withdrawal, after the completion of every cycle, without interrupting the experiment. This allowed us to pay attention to the history of a given prolonged experiment in which the reaction mixture can undergo processes that lead to the increased complexity of the product space, assisted by remnants of preceding chemical processes. We used carbonyldiimidazole (CDI) as a coupling agent to oligomerise the monomers. While CDI is not prebiotically plausible, it is a more kinetically favoured route than peptide coupling through wet-dry cycling and allows better control over the reaction process.[29] Different mineral surfaces were added to the amino acid mixtures in an aqueous environment, and the mixture was replenished recursively with fresh starting materials at the beginning of every new cycle. The minerals chosen for this experiment were olivine, axinite, almandine, staurolite, albite, vesuvianite, serpentine and chalcopyrite. These were curated to provide a diverse range of different environments in terms of crystal structure, elemental composition, and mineral class. All minerals were crushed and sieved to a specific grain size, to ensure the surface areas of each were approximately equal, and washed before use to minimise impurities - details in Mineral Preparation section below.

By designing the experiments in this way, we aimed to test whether mineral surfaces could promote selection in a peptide coupling system by catalysis or templating effects, and how distinct mineral surfaces might influence polymerisation reactions differently. Minerals acting as 'natural' catalysts provide both active sites for catalysis with a possible tendency to template oligomerisation processes, as well as enrichment of certain species in the reactor due to adhesion effects, which further increases reactivity. While peptide condensation experiments on mineral surfaces were reported before,[30-33] no experiments were carried out in recursion over a prolonged period explicitly investigating selective effects in the product space. The extracted data provides information about the product compositions, which are the monomer building blocks in no specific order and the exact sequences inferred from their MS/MS fragments. As an initial screening experiment (Figure 3), the eight different mineral species were first manually evaluated in 24-hour cycles for four cycles.

The reactions we carried out aimed to investigate whether different recursive mineral species alter the reaction product space over successive cycles. The results of those initial experiments strongly indicated that the presence of minerals had a considerable effect on the chemical system. The data in Figure 3 shows a decline in sequences in the curve at cycle two in the presence of almandine, staurolite, axinite, and serpentine. We hypothesise that mineral adsorption effects cause this phenomenon. While only the liquid phase was considered in the measurements, it is evident that the solid surface influenced the solutions' products. The data shows by the decrease in the first cycles that there was a period where the mineral surface was freely available for monomers and products to adsorb until the surface was saturated. We measured a product increase from cycle 3 in the data for the mentioned minerals, which might have been the point at which fewer product species adsorbed to the mineral surface due to its saturation and were, therefore, detectable in the product solution. While it is likely that some products remained adsorbed to the mineral surface, this study focuses only on the impact of the minerals being present on the oligomerisation reactions based on the product species that are detected in solution.

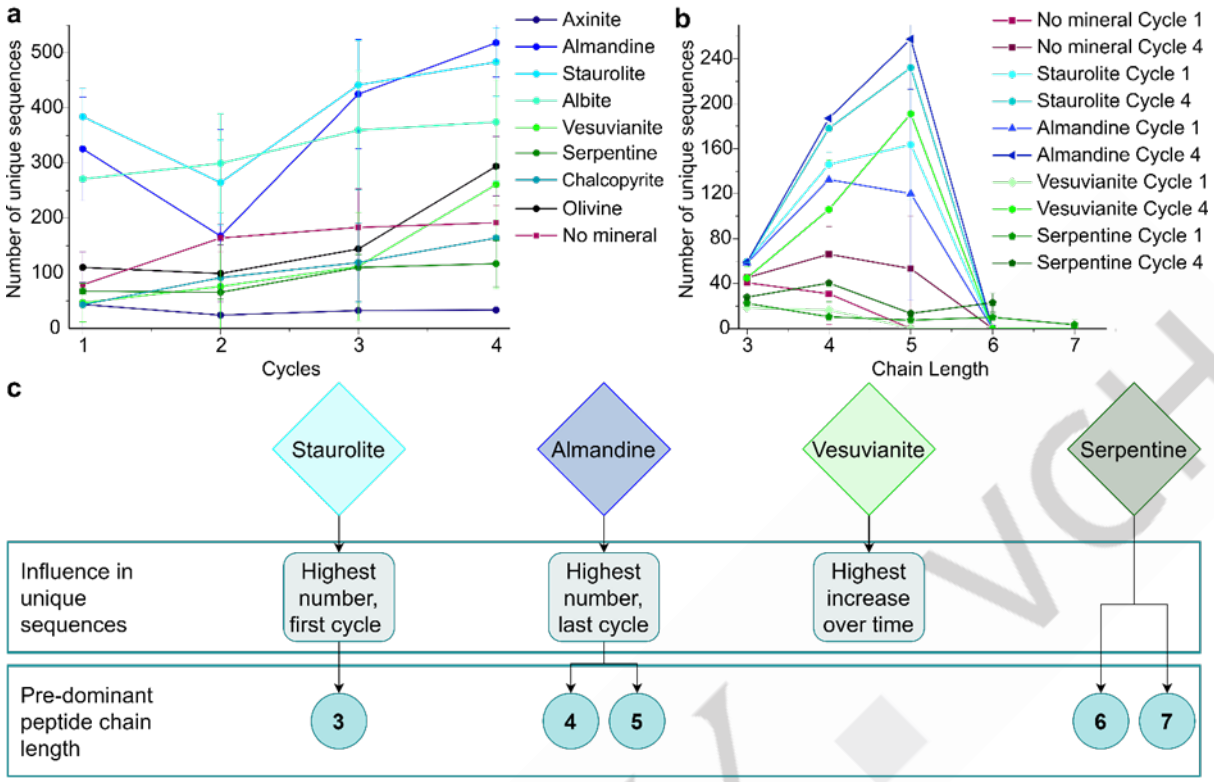


Figure 3. Initial screening results showing distinct differences in the chemical system caused by the mineral environment. (a) shows the number of unique sequences observed per cycle between the different mineral environments. (b) compares the first and last cycle of each experiment by the chain length of the produced oligomers and shows the differences of product species length by environment. The red lines in (a) and (b) correlate to standard experiments without minerals present. (c) Summarises the findings of (a) and (b) by representing the 4' most interesting' minerals. These mineral environments were selected as they all showed different unique behaviour, their product space trends, and the pre-dominant observed peptide chain length are presented.

By looking into trends emerging from this first screening experiment, it became apparent that certain minerals showed distinct behaviours. Those were summarised in mineral profiles in Figure 3c. We observed that staurolite showed the highest number of unique sequences in the first cycle (Figure 3a), with 384 different sequences measured after thresholding (only sequences with over 50% confirmed fragments) and the highest number of compositions, with 49 different produced compositions observed. The highest number of unique sequences and compositions at the end of the last cycle of the experiment was observed in the samples containing almandine, with 518 observed sequences and 56 compositions. Vesuvianite appears to have the highest increase in unique sequences and compositions throughout the experiment, with an increase of around 214 sequences and 26 compositions over 96 hours. The trends in the length of the oligomer products were also considered, as shown in Figure 3b. The staurolite experiment produced many species very quickly and the highest number of tripeptides. Almandine produced the highest number of unique products throughout the experiment and the highest amount of tetra- and pentapeptides. Serpentine did not produce a comparably high number of different species but made the longest oligomers in the screening experiment with hexa- and heptapeptides in detectable abundance. While this showed very clearly that the mineral environments heavily influenced the product space, comparing the unique sequences produced did not provide enough information to understand selective effects in the system.

In addition to studying the emergence of different compositions and sequences, we aimed to examine the causation between fragments incorporated into the products and how this led to longer sequences. We wanted to inquire if there were specific patterns within the fragment distribution space extracted from the data over the recursive cycles and across distinct mineral surfaces. The fragment distribution provides information about which monomer, dimer and trimer species were observed and how many product species the fragment occurred in each cycle. An in-silico library based on the amino acid monomer inputs, including all theoretically possible linear fragments, was built to calculate the fragment distribution. Once the library was established, the theoretically calculated sequence and composition fragments were compared with the experimentally observed product sequences.

The experimental data showed differences based on the mineral environment present. Differences in the species were observed when the fragments of the single mineral samples were compared. Monomers and dipeptides were not compared in detail since the production of those can still be solely based on reactivity. However, tripeptide fragments offer a large enough combinatorial space to start detecting divergence based on their mineral

environments beyond base reactivity. Therefore, the tripeptide fragments were compared in detail to search for some early selective behaviour. Serpentine and staurolite showed the most apparent differences in their fragment distribution patterns. When compared directly (Figure 4a), staurolite showed a high abundance of various peptides, while serpentine exhibited a much lower overall abundance of any products. The most abundant fragments in the staurolite trimer space appeared to be HGG (shown in Supplementary Figure 6), followed by GHG and GGH. The most abundant fragments in the serpentine environmental samples appeared to be GGG, followed by GGF and FGG, but overall, the serpentine fragments occurred in drastically lower abundance than the staurolite fragments. It is interesting to observe that most of the fragments incorporated into the longer sequences appear to have mostly glycine and phenylalanine monomers. While the assumption can be made that glycine is easily incorporated into the sequences through its very low steric hindrance, the favoured incorporation of phenylalanine as one of the most complex amino acid used in this study was an unexpected result.

While there appears to be selectivity in which tripeptide fragment species are produced in higher abundance over others in specific mineral environments, at this point, it was unclear whether this also influenced the assembly of specific sequences and the production of longer oligomer chains. Therefore, the occurrence of the already observed tripeptides was compared with pentapeptide products. Pentapeptide fragment strings that occurred in the longer oligomer sequences observed were considered for this comparison, as shown in Figure 4. Although the staurolite environment produced all fragments shown in Figure 4a in much higher abundance than other mineral environments, the data suggests that the production occurred in an unspecific manner. In contrast, the serpentine environment, which produced tripeptide fragments more selectively (marked in bold Figure 4), seemed to also generate higher abundance pentapeptide fragments that contain those specific tripeptide fragments, indicating a preferential incorporation process into longer oligomer chains.

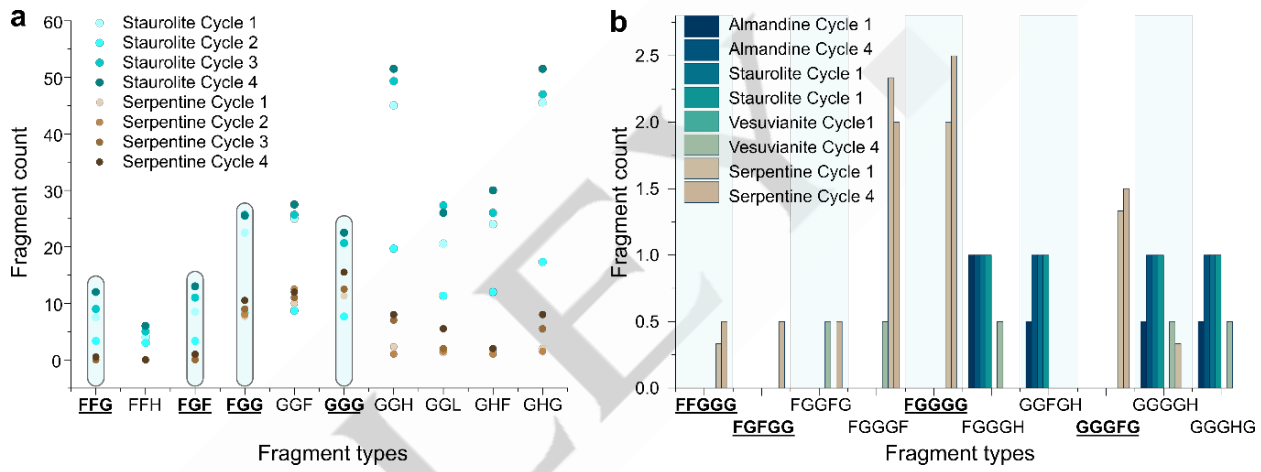


Figure 4. Fragment distribution of selected sequences. (a) Shows selected trimer fragments of staurolite and serpentine in comparison. (b) Presents pentamer fragments of almandine, staurolite, vesuvianite and serpentine. The fragments marked in bold are species also found in longer oligomer chains. Specific fragments were selected for readability; the full fragment distribution is shown in Supplementary Figure 6.

This effect becomes even more evident in the pentamer fragment space, where serpentine produced more fragments that are required to build longer sequences (Figure 4). At the same time, other mineral environments led to the formation of other pentamer sequences that were not incorporated in longer sequences and that were not formed in the serpentine environment. This observation suggested that the serpentine environment selected specific favourable pentamer sequences over others, through the possible incorporation of preferred trimers. Together, this shows a selective synthetic process driven by the mineral environment, in which specificity in the sequence space of short oligomers ultimately led to further selection in the production of longer oligomers.

No fully quantitative observations were made concerning the abundance of the oligomer products but estimates from MS data were made. The focus of this study was the exploration of selective effects, represented by the unique sequences formed in different environments over recursive cycles, and how their build-up process worked. While MS is generally not a quantitative method for sample evaluation, due to variations in ionisation, the assumption can be made that similar chemical systems in the same solvent and product matrix will ionise relatively similarly.[34] This assumption is critical, as the amount of different product sequences and compositions emerging in the system is enormous, and separating those products for truly quantitative measurements would be far from trivial. Since the chemistry and the solvent were identical in all samples, and the extracted data was heavily filtered to only consider theoretical calculations and

possible products, quantitative estimates of abundance could be made among similar products for a rough comparison. When investigating selection, the abundance of selected products, or copy number, is important for determining how selectively emerged. The first plot, in Figure 5a, shows the average maximum intensity of tetra- and pentapeptides on a normalised, logarithmic scale. While the data is still slightly skewed, a trendline shows an increase in the intensity of compositions throughout the experiment. A decreasing trend can be observed compared to the same plot in the vesuvianite environment Figure 5b. In Figure 5c the plot shows the maximum intensity of the observed tetra- and pentapeptides compositions in the serpentine environment.

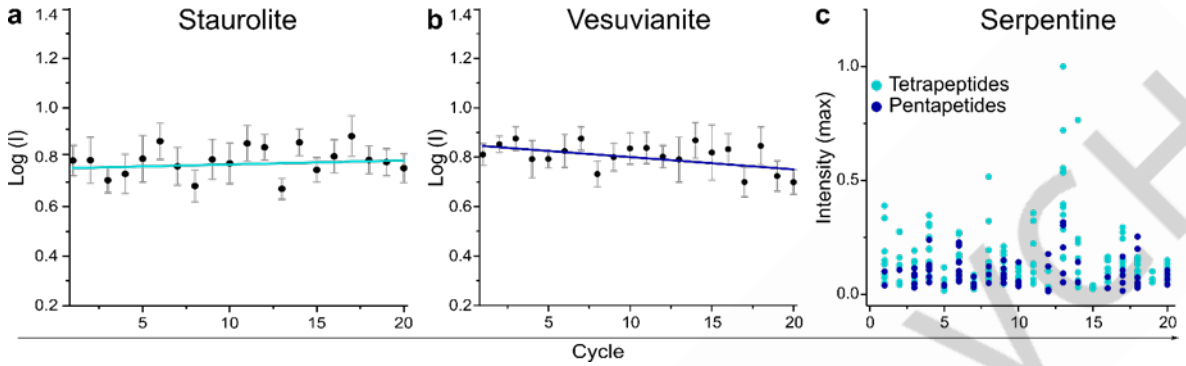


Figure 5. Intensity trends in tetra- and pentapeptide composition products of single mineral CSE experiments. (a) and (b) show the average maximum intensity (I) of observed compositions in a normalised logarithmic scale. (a) Represents the staurolite sample, which shows a slight incline of intensity throughout the experiment, while (b) vesuvianite shows a slight decline. (c) Shows the normalised maximum intensities observed over time. The serpentine environment is presented differently, compared to staurolite and vesuvianite, to highlight the noticeably different behaviour in cycle 13, which was not sufficiently presented in a logarithmic plot.

The serpentine plot in Figure 5c shows changes in the intensity of product species throughout the experiment. A very distinct spike appeared in cycle 13 (both duplicate experiments), in which the intensity of tetrapeptides rose by more than 50% of all observed maximum intensity values. Smaller spikes were also observed in cycles 1, 4 and 8. Those intensity differences were observed repeatedly during experiments in the serpentine environment. The data shows that the intensity of the compositions was affected by the recursive cycles; therefore, we could suggest that the copy number, which could be approximated based on the parent peak intensities, of the observed compositions, also changed.

Conclusion

Indications of mineral driven selection behaviour were found in an amino acid system in the presence of mineral surfaces, bringing us one step further in explaining how selectivity can emerge on a molecular scale. We found that serpentine promoted the production of favourable fragments for the formation of longer oligomer species. This study has shown that the presence of mineral environments not only leads to a change in the product space, but also in the distribution of amino acid and oligopeptide fragments during the experimental cycles.

A Chemputer based system for prebiotic experiments was developed, the CSE, and this platform enabled the automated investigation of amino acid oligomerisation reactions in distinct mineral environments. Furthermore, an advanced analytical workflow was developed to show selective effects in amino acid systems by leveraging the CSE platform and using tandem MS and peptide sequencing through Python code. The initial screening experiment revealed trends in the behaviour of the chemical systems in the presence of different mineral environments. The findings suggest that the type of mineral present in the reactor influences the outcome of the experiment in terms of product space. A prolonged automated study of the four most distinct mineral environments revealed that general trends, such as the tendency of serpentine to enable the formation of longer sequences, remained present over an extended experimental period. The favoured production of specific fragment sequences in certain mineral environments was demonstrated by comparing the fragment distribution pattern of the samples. The initial hypothesis was confirmed by demonstrating that not only specific fragments were favourably produced but also that some other fragments, which did not lead to complex oligomer species, were less favoured. This behaviour strongly suggests selective processes on a molecular scale in the serpentine experiment. Estimating the copy number of the products showed that staurolite made the highest number of products, while serpentine showed only a high intensity in a later cycle. The initially low product peak intensity of the serpentine environment implies that recursion may have enabled the emergence of selection in this system. Also, the increase in complexity in the serpentine experiment appears not to be solely based on an increase in high product intensity and further strengthens the selectivity argument.

Serpentine as the mineral environment showing this phenomenon is specifically interesting, as this mineral is the product of the serpentinization processes. In this process, seawater reacts with minerals of the Earth's crust, producing H_2 and serpentine minerals.[35] It has been observed to occur close to hydrothermal

vents, which have been repeatedly named as a possible geographical location for the emergence of life.[36] While the mineral serpentine is usually not considered since it is buried in the sediment and not exposed in the actual vent, this result does raise the question if this mineral should be further looked into. Also, while they were not the focus of this work, a detailed analysis of the mineral surfaces taking part in oligomerisation reactions, and which product species remained adhered to them, will constitute part of a future study.

This project set out to investigate selective effects in chemical systems, and the presented workflow for the model chemical system demonstrated that selection can indeed emerge on a molecular level in the presence of a mineral environment. Selection can lead to a complexification of simple building blocks towards a highly functional system and is a critical process before the emergence of Darwinian evolution. This study represents only the beginning of investigating selectivity in amino acid reactions and there is a variety of amino acids that would be further interesting targets to investigate as well as a wide amount of different minerals that could have extended this study. A high-throughput automated platform instead of an automated batch system could open the door to a much wider investigation of these and similar chemical systems. Furthermore, we propose that the diversity of the product space might be increased further by combining certain mineral environments that are seen to increase the chemical system's complexity. Observing selection in a chemical system leads us one step further into understanding evolutionary processes on a molecular level and the transition from chemistry to biology.

Experimental Methods

All materials and solvents were purchased from commercial sources (Merck or Fischer Scientific and Alfa Aesar) unless otherwise stated and used without further purification.
Unless otherwise stated, bench samples were prepared in triplicate, and CSE experiments were performed in duplicates.

Stock solutions: Amino acid stock solutions were prepared in bulk and refreshed weekly. Stock solutions were made in 1L HPLC-grade water, and amino acids were dissolved by sonication.

L-phenylalanine (4.13g, 25 mM), Glycine (1.88, 25 mM), L-leucine (3.28, 25 mM), L-histidine (3.88, 25 mM)

CDI was made on demand or prepared daily as CSE stock. For 100 mL, 1.01g (62.5 mM) were dissolved through short sonication in HPLC-grade acetonitrile.

Mineral preparation: Ulexite, olivine, axinite, almandine, staurolite, albite, vesuvianite, serpentine, chalcopyrite, pyrite and quartz were obtained from Richard Tayler Minerals, Cobham, Surrey, England.

All minerals listed above were crushed in a Mad Mining Rock Crusher with a Solid Steel Frit, and sieved to a size between 2 - 4.75 mm. The minerals were boiled and stirred in HPLC-grade water for 2 hours and continuously rinsed with fresh HPLC-grade water until the solution in touch with the minerals remained clear. After that, the minerals were dried and directly transferred to the reactor.

Single mineral screening: For the single mineral experiment, 14 mL glass vials were charged with 1 g of a mineral each. The reaction volume was 8 mL, and 6 mL were recursively replenished (75 % recursion).

Single mineral input amounts. For the initial cycle, the amounts of stock solutions added were 1.6 mL amino acids (each), 0.8 mL CDI, and 0.8 mL water. For all subsequent recursive cycles, the amounts of stock solutions added were 1.2 mL amino acids (each), 0.6 mL CDI, and 0.6 mL water.

The experiments were heated at 45°C and stirred at 300 rpm, the cycle time was 24 hours.

CSE single mineral experiment: The experiments were prepared in 100 mL 3-neck round bottom flasks. The reaction volume was 40 mL and the recursive volume 30 mL (75 % recursion). 20 Cycles of 4hrs were performed.

CSE single mineral amounts. For the initial cycle, the amounts of stock solutions added were 8 mL amino acids (each), 4 mL CDI, and 4 mL water. For all subsequent recursive cycles, the amounts of stock solutions added were 6 mL amino acids (each), 3 mL CDI, and 3 mL water.

Mass spectrometry via direct injection: The sample was introduced into the Thermo Orbitrap Fusion Lumos Tribrid mass spectrometer by using an Advion Nanomate. 15 µL per sample were injected onto an emitter with 1.2 kV applied and a gas flow of 40 psi. The MS method was in positive mode and 6 minutes long, with a mass range from 100 to 1000 m/z. The isolation window was set at 1 Da, the resolution of the MS1 scan was 240000 and the MS2 scan was 30000. The fragment ions were analysed in the Orbitrap with HCD fragmentation set at 35%. The 15 most intense peaks were selected for MS2 fragmentation. Dynamic exclusion was used for peaks, measured 3 times in 30 seconds for the following 120 seconds.

Supporting Information

The Supplementary Information provides further information on the building block and mineral environment selection and the available chemical space calculation. There is also a detailed description of all steps required to perform the CSE experiment. Finally, there are details about using the OLIGOSS data extraction software and information on the fragment distribution calculation.

Author Contributions

L. C. devised the concept and the initial idea, experiments were designed by S. A., R. W. P. and H. M. M. S. A. and R. W. P. performed test reactions, S. A. build the platform, carried out all experiments, wrote code to calculate the fragment distribution, sequenced the MS data and analysed the data. S.A. wrote the paper with input from G. J. T. C., A. S. and L. C.

Acknowledgements

We would like to thank E. Carrick for help with MS method development. For their insights during manuscript writing, thanks also go out to A. Kahana.

The authors gratefully acknowledge financial support from EPSRC (Grants EP/S019472/1, EP/W001918/1, EP/R01308X/1, EP/S030603/1, EP/S017046/1), ERC (Grant 670467 SMART-POM), NIH (Grant 1UG3TR004136-01) and the John Templeton Foundation (Grant 62231, 61184, 60625).

Keywords: Origin of Life • Automation • Selection • Recursive Cycles • Peptides

[1] M. Preiner, S. Asche, S. Becker, H. C. Betts, A. Boniface, E. Camprubi, K. Chandru, V. Erastova, S. G. Garg, N. Khawaja, G. Kostyrka, R. Machné, G, Moggioli, K. B. Muchowska, S. Neukirchen, B. Peter, E. Pichlhöfer, Á. Radványi, D, Rossetto, A. Salditt, N. M. Schmelling, F. L. Sousa, F. D. K. Tria, D. Vörös, J. C. Xavier, *Life*, **2020**, *10*.

[2] A. Eschenmoser, *Tetrahedron*, **2007**, *63*, 12821.

[3] S. Asche, G. J. T. Cooper, G. Keenan, C. Mathis, L. Cronin, *Nat. Commun.*, **2021**, *12*.

[4] N. Guttenberg, N. Virgo, K. Chandru, C. Scharf, I. Mamajanov, *Philos. Trans. A Math. Phys. Eng. Sci.*, **2017**, *375*.

[5] A. Briš, M. G. Baltussen, G. L. Tripodi, W. T. S. Huck, P. Franceschi, J. Roithová, *Angew. Chem. Int. Ed.*, **2024**, *63*, e202316621.

[6] M. Bechtel, M. Ebeling, L. Huber, O. Trapp, *Acc. Chem. Res.*, **2023**, *56(20)*, 2801.

[7] S. A. Kauffman in *Understanding Origins: Contemporary Views on the Origin of Life, Mind and Society*, (Eds.: F. J. Varela, J.-P. Dupuy), Springer Netherlands, Dordrecht, **1992**, pp. 153-181.

[8] T. Z. Jia, M. Caudan, I. Mamajanov, *Life*, **2021**, 11, 154.

[9] A. Jain, S. A. McPhee, T. Wang, M. N. Nair, D. Krpiss, T. Z. Jia, R. V. Ulijn, *Chem*, **2022**, *8*, 1894.

[10] T. R. Gregory, *Evo. Edu. Outreach*, **2009**, *2*, 156.

[11] M. Frenkel-Pinter, M. Samanta, G. Ashkenasy, L. J. Leman, *Chem. Rev.*, **2020**, *120*, 4707.

[12] O. Taran, C. Chen, T. O. Omosun, M. C. Hsieh, A. Rha, J. T. Goodwin, A. K. Mehta, M. A. Grover, D. G. Lynn, *Philos. Trans. A Math. Phys. Eng. Sci.*, **2017**, *375*, 20160356.

[13] C. C. A. Ng, W. M. Tam, H. Yin, *Nat. Commun.*, **2021**, *12*, 4242.

[14] C. Huber, W. Eisenreich, S Hecht, G. Wächtershäuser, *Science*, **2003**, *301*, 938.

[15] S. V. Krivovichev, V. G. Krivovichev, R. M. Hazen, S. M. Aksenov, M. S. Avdontceva, A. M. Banaru, L. A. Gorelova, R. M. Ismagilova, I. V. Kornyakov, I. V. Kuporev, S. M. Morrison, T. L. Panikorovskii, G. L. Starova, *Mineralogical Magazine*, **2022**, *86*, 183.

[16] J.-F. Lambert, *BIO Web of Conferences 4*, **2015**, 00012.

[17] H. H. Smith, A. S. Hyde, D. N. Simkus, E. Libby, S. E. Maurer, H. V. Graham, C. P. Kempes, B. Sherwood Lollar, L. Chou, A. D. Ellington, *Life*, **2021**, *11*, 498.

[18] R. M. Hazen, D. A. Sverjensky, in *Mineral Surfaces, Geochemical Complexities, and the Origins of Life. Cold Spring Harbor Perspectives in Biology 2*, **2010**, a002162-a002162.

[19] R. M. Hazen, *Elements*, **2005**, *1*, 135.

[20] A. Brack, in *Developments in Clay Science vol. 1*, (Eds. F. Bergaya, B. K. G. Theng, G. Lagaly), Elsevier, **2006**, pp. 379-391.

[21] J. T. Kloprogge, H. Hartman, *Life*, **2022**, *12(2)*, 259.

[22] C. Sydow, F. Sauer, A. F. Siegle, O. Trapp, *Chem. Sys. Chem.*, **2023**, *5*.

[23] G. Matrajt, D. Blanot, *Amino Acids*, **2004**, *26*, 153.

[24] A. Sharma, D. Czégel, M. Lachmann, C. P. Kempes, S. I. Walker, L. Cronin, *Nature*, **2023**, *622*, 321.

[25] S. Rohrbach, M. Šiaučiulis, G. Chisholm, P.-A. Pirvan, M. Saleeb, S. H. M. Mehr, E. Trushina, A. I. Leonov, G. Keenan, A. Khan, A. Hammer, L. Cronin, *Science*, **2022**, *377*, 172.

[26] S. Steiner, J. Wolf, S. Glatzel, A. Andreou, J. M. Granda, G. Keenan, T. Hinkley, G. Aragon-Camarasa, P. J. Kitson, D. Angelone, L. Cronin, *Science*, **2019**, *363*, 144.

[27] D. Doran, E. Clarke, G. Keenan, E. Carrick, C. Mathis, L. Cronin, *Cell Rep. Phys. Sci.*, **2021**, *2*, 100685.

[28] D. Angelone, A. J. S. Hammer, S. Rohrbach, S. Krambeck, J. M. Granda, J. Wolf, S. Zalesskiy, G. Chisholm, L. Cronin, *Nat. Chem.*, **2021**, *13*, 63.

[29] R. Liu, L. E. Orgel, *Orig. Life Evol. Biosph.*, **1998**, *28*, 47.

[30] J. P. Ferris, A. R. Hill, R. Liu, L. E. Orgel, *Nature*, **1996**, *381*, 59.

[31] J. Bujdák, K. Faybíková, A. Eder, Y. Yongyai, B. M. Rode, *Orig. Life Evol. Biosph.*, **1995**, *25*, 431.

[32] S. Fuchida, H. Masuda, K. Shinoda, *Orig. Life Evol. Biosph.*, **2014**, *44*, 13.

[33] A. J. Surman, M. Rodriguez-Garcia, Y. M. Abul-Haija, G. J. T. Cooper, P. S. Gromski, R. Turk-MacLeod, M. Mullin, C. Mathis, L. Cronin, *Proc. Natl. Acad. Sci. U. S. A.*, **2019**, *116*, 5387.

[34] W. Kew, A. Myers-Pigg, C. Chang, S. Colby, J. Eder, M. Tfaily, J. Hawkes, R. Chu, J. Stegen, *EGUsphere*, **2022**, 1-26.

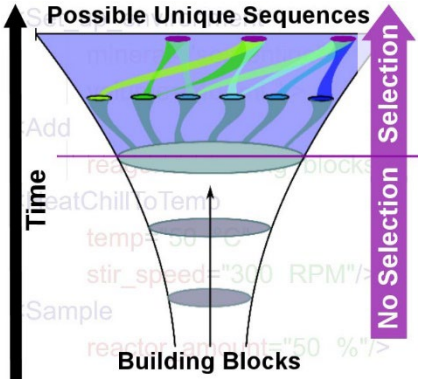
[35] M. Preiner, K. Igarashi, K. B. Muchowska, M. Yu, S. J. Varma, K. Kleinermanns, M. K. Nobu, Y. Kamagata, H. Tüysüz, J. Moran, W. F. Martin, *Nat. Ecol. Evol.*, **2020**, *4*, 534.

[36] V. Sojo, B. Herschy, A. Whicher, E. Camprubí, N. Lane, *Astrobiology*, **2016**, *16*, 181.

WILEY VCH

RESEARCH ARTICLE

Entry for the Table of Contents



An automated platform explores the reactions of amino acids in the presence of mineral environments and searches for evidence of the onset of selection. The outcome of oligomerisation is significantly influenced by the presence of different minerals, with certain environments selecting specific fragments that enable the formation defined sequences as a function of cycle number.