scientific reports

Check for updates

OPEN

Reproducible Ala-Gly oligomerization catalyzed by the natural Borate colemanite in prebiotic conditions

Isabella Rimoldi[1], Giorgio Facchetti[1], Lucia Feni[1], Giacomo Diego Gatta[2], Paolo Lotti[2], Silvia Rizzato[3,4], Michael S. Christodoulou[5], Lucia Silvestrini[6,7] & Leonardo Lo Presti[3]✉

The abiogenesis of complex peptides is a yet unsolved problem concerning the origin of life, as it is unclear how specific amino acid sequences could be formed in the absence of a regulation mechanism. Crystalline minerals could have provided template scaffolds to sustain replicable oligomerization processes. We demonstrate that the natural organophilic borate colemanite, $CaB_3O_4(OH)_3{\cdot}H_2O$, fosters consistent Gly-Ala oligomerization into AG_8, G_7AG and G_7GA strands with well-defined primary sequence (8%, 47% and 45% respectively) from activated glycine and alanine as N-carboxyanhydrides. These Gly-rich sequences are mechanically flexible and may have played a pivotal structural role in driving amino acid loops in ancient polypeptides. In fact, they occur repeatedly in proteins that exploit either structural or transport functions, which belong to ancestral organisms, including *Cyanobacteria* and *Thermoproteota*. Our results demonstrate that the consistent synthesis of short well-defined amino acid sequences is possible in a few days of incubation under mild conditions in a prebiotically plausible environment.

Life as we know cannot exist without amino acids (aas) and proteins, and in fact there is a broad consensus on the occurrence of aas since the earliest times of Earth. The abiotic synthesis of biomolecules from elementary building blocks is possible[1] and, since Miller's pioneering work based on a reducing atmosphere, huge efforts were carried out to dissect various physicochemical scenarios, in less reducing conditions[2,3] or even in completely different environments[4]. Aas were also found in carbonaceous chondrites[5–8], suggesting that chemical processes leading to their synthesis could be ubiquitous, and implying that some aas could have been brought on Earth since the late heavy bombardment era, around 4 billion years (By) ago[9].

The next step of chemical evolution was the assembling of aa chains, able to participate in auto-catalytic protometabolic pathways[10]. If the broad picture is mostly accepted, our understanding of abiotic polymerization is still fragmentary[11]. The condensation of a peptide bond is endergonic by 2–4 kcal·mol^{-1}[12–14], but unshielded UV rays could have easily destroyed complex organic molecules on Earth's surface. Thus, abiotic polymerization required (i) a sheltered environment, (ii) enough gathering of reagents to provide an effective kinetic drive[15] and (iii) energy, preferably in the form of heat. These constraints make open surface sea waters unlikely to serve as chemical incubators for proto biopolymers. An alternative scenario that is invoked to explain also self-assembling of other proto biopolymers, like nitrogenous bases[16], involves hydrothermal ponds[17,18], like coastal tidal pools or crater lakes. These could have provided a mineral-rich environment to concentrate aas under mild solar and/or geothermal heating, where transition metals promoted the oligomerization of simple aas[19] and the synthesis of proto-nucleic acid strands[20], in what could have been a first attempt of abiotic metabolism[21,22]. It has been suggested[23,24] that iron ions in crystalline clays could have acted as catalytic sites, fostering the synthesis of aas and nucleobases, and subsequently their polymerization directly onto the mineral surface[25]. Also computational[26] evidence support this view. Ab initio Molecular Dynamics simulations suggested that amorphous silica surfaces could speed up the condensation of glycine[27]. On the experimental side, it has been

[1]Department of Pharmaceutical Sciences, Università degli Studi di Milano, Via Venezian 21, 20133 Milano, Italy. [2]Department of Earth Sciences "Ardito Desio", Università degli Studi di Milano, Via Botticelli 23, 20133 Milan, Italy. [3]Department of Chemistry, Università degli Studi di Milano, Via Golgi 19, 20133 Milan, Italy. [4]National Institute for Nuclear Physics (INFN), National Laboratories of Frascati, Frascati (Roma), Italy. [5]Department of Food, Environmental and Nutritional Sciences, Università degli Studi di Milano, Via Celoria 2, 20133 Milan, Italy. [6]DG for Life, Lita-Ghirlanda square, 20013 Magenta, MI, Italy. [7]Department of Microbiology and Virology, University "Vita-Salute" San Raffaele, Via Olgettina 58, 20132 Milan, Italy. ✉email: leonardo.lopresti@unimi.it

shown that mild hydrothermal conditions (p ~ 500 MPa and T ~ 150 °C), like those to which deep sediments are subjected, might promote glycine oligomerization up to 10 units[28]. Glutamic acid chains up to 55 units long were obtained on illite by applying a feeding protocol of repeated incubation of activated aa at different concentrations[29]. Similar results were obtained for glycine on clay[30]. In an excellent review[22], Ashkenasy & Leman also pointed out the importance of condensing agents to drive uphill the endergonic aa polymerization. However, it is still unclear how such agents could have worked effectively in a water-rich environment. Alternative mechanisms were proposed, like for example oxidative acylation by aminoacyl thioacids[31]. Another general problem is that the condensation of different aas is uncommonly addressed and is often limited to the study of very short oligopeptides[22,32].

In this Article, we investigate how he natural organophilic borate colemanite, $CaB_3O_4(OH)_3{\cdot}H_2O$, might have fostered Gly-Ala oligomerization in a water-rich environment, by releasing borate species as effective catalytic agents.

Results

Synthesis of N-carboxyanhydrides and polymerization setup

We started from high-energy N-carboxyanhydrides (NCAs), which are universal activated forms of amino acids and were proposed as possible intermediates in various prebiotic chemistry scenarios[22,33–35]. Owing their high electrophilic reactivity, NCAs may pay the endergonic cost for peptide bond formation, and it has been proposed that they could be spontaneously produced by natural processes on early Earth[36–39].As concerns the present experiments, if non-activated NCA-free aas are employed, no polymerization is observed at all (Table S1 in Supporting Information, SI, entries 1, 4 and 31).

The synthesis of the anhydride was carried out according to literature protocols (Fig. 1A1)[40]. Amino acid was incubated with dry THF at 50 °C for 5 h under inert atmosphere, with triphosgene, obtaining the pure starting material in a good yield (70%). Then, NCA-amino-acids were incubated in a glass reactor on a reciprocal shaker (150 rpm) for 72 h under continuous stirring and natural light irradiation, in the presence of one mineral catalyst among heulandite ($(Ca,Na,K)_5(Si_{27}Al_9)O_{72}{\cdot}26H_2O$), quartz ($SiO_2$) and colemanite ($CaB_3O_4(OH)_3{\cdot}H_2O$) (Section S1 SI). The outcomes were evaluated as a function of several physicochemical parameters, including temperature, reactant concentrations, catalyst/substrate ratio, pH, ionic strength, oxidant (air) or inert (N_2) atmosphere (Section S2 SI, Table S1). All the experiments were carried out in triplicate at least (see Methods).

Catalytic formation of homopeptides

First, alanine N-carboxyanhydride (Ala-NCA) was reacted with L- alanine (Ala) (Fig. 1A2). When the reaction was carried out in a Schlenk tube, extensive Ala polymerization resulted in the precipitation of micrometer-size highly hydrophobic poly-Ala fibrils (Section S3 SI, Figures S3-S5), which occurred in a higher amount when any mineral was present. The oligomerization proceeded in solution, as evidenced by HPLC-mass (panels A and B of Figure S6 SI).

When the reaction was carried out with a reciprocal shaker (Fig. 1, upper left inset and Figure S2 SI), the fibrils did not appear. In the absence of catalyst, a maximum of 6 Ala was bonded together, and the reaction was more effective under inert N_2 atmosphere (Table S1 SI, entries 1–3).

With heulandite and quartz catalysts, random and poorly reproducible results in terms of length of the poly-Ala chains were obtained (max 6–8 consecutive Ala: Table S1 SI, entries 6–23). When O_2 was present, the results were even less reproducible: only variable mixtures of short oligopeptides were obtained (Table S1 SI, entries 2, 14–16 3 and 22). On the contrary, when colemanite was used as catalyst, the reproducibility in terms of chain length was respected (up to Ala_{24}, Fig. 1 and Table S1 SI, entries 24–27, panel C of Figure S6 SI). By "reproducibility", we intend that experiments made in triplicate provided invariably the same result (see Methods). This is true for colemanite, not for heulandite and quartz. To confirm the obtained sequences and assigned peaks, Ala_5, purified by preparative HPLC, was characterized by NMR spectroscopy and mass spectrometry (Section S4 SI, Figures S7–S8).

These results are very interesting, as all replicated experiments confirm that only colemanite produced concordant and well-defined results, while the other two minerals invariably gave rise to short aa oligomers of variable length (≤ 8 units). A possible reason is that colemanite has a measurable solubility in water (~ 1 $g{\cdot}L^{-1}$ in mild alkaline conditions)[41], while heulandite and quartz are completely insoluble. Moreover, the solid–liquid equilibrium is set up in a few hours[42]. Thus, 72 h long incubation ensures full saturation. Even though the exact catalytic mechanism remains to be disclosed (*vide infra*), the interplay of boric ions in solution and the mineral surface appears to be a crucial requirement for effective oligomerization, which implies that colemanite-mediated catalysis occurs at least partly in the homogeneous phase[41–44]. This hypothesis is consistent with very recent findings on the ability of boric acid to assist the condensation of poly-Gly units in a nearly neutral abiotic environment[45].

Catalytic formation of heteropeptides

The same protocol was applied starting from commercially available Gly-NCA in presence of different amount of L-Ala, in absence (Table S1 SI, entries 4–5 and 34) or in presence of catalyst (panel B of Fig. 1 and Table S1 SI, entries 36–43).

The effect of colemanite is evident: all reaction mixtures that lacked this mineral produced just short homopolymers with random length, never longer than 5–6 Ala or Gly units. As colemanite is the only mineral with a measurable solubility (see above), it is reasonable that borate species could play a catalytic role (see also the Discussion below). However, it is necessary that they are released slowly. If boric acid (H_3BO_3) is directly added in solution, only a short glycine oligomer is formed (≤ 5 Gly, Table S1 SI, entry 43), despite Ala being present in the reaction mixture. This result demonstrates that boric acid as a reactant does not allow fine control

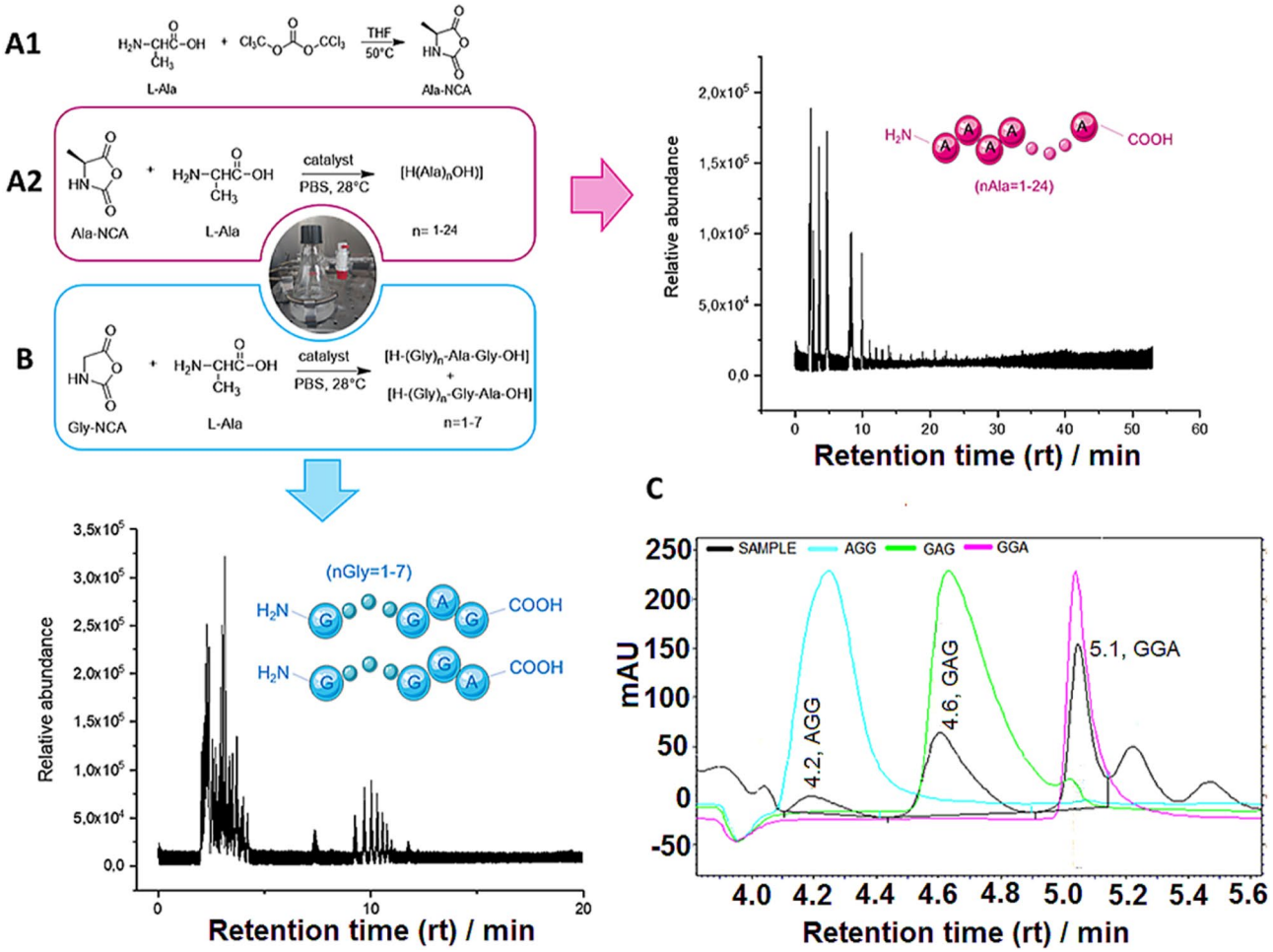

Fig. 1. Mineral-mediated synthesis of amino acid strands with well-defined primary sequence: procedure and main results. (**A1**): Synthesis of high-energy N-carboxyanhydrides (NCAs). Triphosgene is just a high-energy reactant to induce cyclization, and we do not claim that it was present on early Earth; this is why panel A1 is not highlighted. (**A2**): Model for prebiotoic synthesis of poly-Ala stretches. The optimized conditions were: colemanite/Ala-NCA (10% w/w/), PBS buffer (pH = 7.4), ratio Ala-NCA/Ala = 1/1, [Ala-NCA] = 60 mM and 28 °C in inert atmosphere. A mixture of homopeptides, up to A_{24}, was obtained, as evidenced by the mass chromatogram (right). Inset: chemical reactor. More details in Figures S1-S3 SI. (**B**): Model for prebiotoic synthesis of Gly-Ala stretches. Same as A2, starting with commercial Gly-NCA with 0.5 eq of Ala. Bottom and right: mass chromatograms, with the interpretation of the peaks from ESI mass spectrometry. (**C**): overlap of HPLC spectra of the three standard peptides (AGG rt 4.2 min, GAG rt 4.6 min and GGA rt 5.1 min) with purified experimental sample.

of polymerization. Boric acid might promote decarboxylation of Gly-NCA, producing activated Gly-$OB(OH)_2$ species that in turn might directly react with Gly-NCA units. This hypothesis is consistent with prediction of thermodynamic parameters by quantum simulation (*vide infra* and see Section S5 SI, panel D of Scheme S1 for full details). In this scenario, the reaction of non-activated Ala to form Ala-$OB(OH)_2$ boric esters is slower than the analogue reaction with activated Gly-NCA, thus preventing the inclusion of Ala into the growing strand. The easier availability of borate species prevents the reaction from producing long polypeptides: short poly-Gly strands are formed just at the beginning of the reaction and cannot grow further due to depletion of anhydrides available to decarboxylation. The outcome is a population of very short and Ala-free oligomers, as we observe.

When colemanite is added to the reaction mixture, in presence of 0.5 eq of L-Ala, two separate families of compounds are invariably obtained (panel B of Fig. 1 and Figures S9–S11 SI). In the first one (rt between 2.7 and 4.2 min), Gly-NCA works as a building block for the conjugation to one monomer of alanine followed by other glycine residues obtaining the sequence derivatives [H-$(Gly)_{n1}$-Ala-$(Gly)_{n2}$-OH], where the maximal value is $n1 + n2 = 8$ and $0 \leq n2 \leq 2$ (hereinafter, GAGs) (Figure S10 SI). By analytical experiments using C18 column, we managed to isolate the peak to $[M + H]^+ = 204.0$ corresponding to rt 2.86 min (see panel A of Figure S12, SI). The comparison between the spectra of synthesized GAGs isomers and the sample confirmed the correct sequence assignation for each isomer. In particular, the GAG strand is the most present in comparison with the other possible structural isomers (GGA and AGG) in a ratio AGG: GAG: GGA = 8:47:45. (Fig. 1, C and Figure S12 SI) Considering that the colemanite influenced the formation of GAG as the major product and that the proposed

mechanism of chain elongation (Fig. 2) is the one described in literature[46,47], G_7AG (route B) is probably the major polypeptide identified, together with G_7GA (route A).

The second group of peptides (rt between 9.2 and 10.9 min), less significant for importance and abundance, refers to the urea derivatives of lower affinity for the mobile phase and thus with longer retention times in comparison with GAGs polypeptides group (Figure S11 SI). These byproducts stem from the nucleophilic attack of the nitrogen at the 2-carbonyl of the oxazolidine-2,5-dione ring[48].

Finally, when the reaction in the presence of colemanite was performed using Ala-NCA with Gly, or Gly-NCA with Gly, very short homopeptides (max 8-unit long alanine or glycine homopeptides) are obtained (Table S1 SI, entries 28–30 and 35). The shortness of the chain is due again to the competition with the formation of urea derivatives.

As for heulandite and quartz, they led to the formation of some oligomers, the longest ones consisting of 8 Ala units (Table S1 SI, entries 10–22 and 28–29). However, experiments conducted in triples (at least) failed to produce consistently the same polymerization product(s), as oligopeptides of different lengths were always obtained (see footnotes of Table S1 SI). In some cases, no polymerization products were even obtained at all (Table S1 SI, entries 6–9 and 23).

Most important, these performances are very similar to those obtained in the absence of any catalyst, where no reproducibility was observed in the chain length, and the longest oligomers consisted of 6 Ala units at most (Table S1 SI, entries 1–5). If these minerals have some effects, these should be related to adhesive and support function of the inert surfaces, rather than to a true catalytic function[49].

In conclusion, the relatively high solubility of colemanite, not in the sense of immediate availability, but over the full reaction time (4 days), seems to offer the right balance between the support effect and the borate-catalyzed effect, with borate being gradually released into the reaction system. Thus, only with this material the same sequence of Ala-Gly strands is obtained, indicating a level of reproducibility almost comparable to that of a classical chemical catalyst.

Computational chemistry simulations

DFT quantum simulations were carried out on isolated reactants and products to gain first principle insights on the thermodynamics of the first steps of the aa polymerization. The purpose was to provide a tentative explanation of the observed preference for the GAG and GGA carboxy ends over the AGG one. Results *in vacuo* and in water are reported in the SI (Section S5). The main outcome is that chemical activation of aas is a mandatory prerequisite to promote exergonic polymerization routes toward specific sequences. The most reactive species in our system is the N-carboxy anhydride, which in fact determines the major aa content of the resulting oligopeptide. Starting from Ala-NCA, only short poly-Ala strands are obtained, while Gly-NCA produces Gly-rich polypeptides (Table S1 SI, entries 30 and 38ff.). The latter are the most interesting, as their sequence invariably includes a single Ala unit, whose distance from the carboxy end depends on the reaction route (Fig. 2). Thus, our simulations focused on GGA, GAG and AGG tripeptides, as the corresponding aa chains grow by further Gly-NCA-mediated addition of Gly units at their amino end.

The formation of the GGA-OH ending can be explained by Ala-promoted decarboxylation of Gly-NCA, followed by repeated addition of up to 6 Gly units (route A). In the presence of colemanite, an alternative route can be hypothesized, where Gly-NCA reacts with boric species in solution to form high-energy Gly-$OB(OH)_2$ boric ester (Section S5, Tables S2-S3 and panel D of Scheme S1 SI). This reaction is weakly endergonic in water by + 3 kcal·mol^{-1} (Table S4 SI), but the subsequent addition of free Ala at the carboxy end to produce a Gly-Ala dipeptide is favored by -11 kcal·mol^{-1} (Table S4 SI).

However, neither of these mechanisms can explain the occurrence of the GAG-OH ending, which is the other major constitutional isomer (≅47%, Fig. 1C). GAG can result from the activation of Ala units by condensation with boric acid to give the Ala-$OB(OH)_2$ boric ester intermediate (route B in Fig. 2). This reaction step is endergonic by some + 5–8 kcal·mol^{-1} (Table S4 and Scheme S1 SI), but the energy cost is entirely repaid by the next addition of Gly to produce a Ala-Gly dimer, which is exergonic by 15–17 kcal·mol^{-1} (Table S4 SI). Overall, the Gibbs free energy balance of formation of the GGA and GAG tripeptides is identical (≅ -16 kcal·mol^{-1} in water).

Route C (Fig. 2) accounts for the minority polymerization product, H_2N-G_nAGG-OH (≅8%, Fig. 1C). The addition of the Ala-$OB(OH)_2$ intermediate to form the AGG tripeptide has a comparable Gibbs free energy balance with respect to the A and B routes (≅ -14 kcal·mol^{-1} in water, Table S4 SI). However, this reaction step must wait for the formation of a Gly-Gly adduct. Therefore, process C is significantly slower than direct condensation of Ala-$OB(OH)_2$ with a single Gly unit (or *vice-versa*), thus explaining the much lower yield in this peptide (Fig. 1C).

It is worth noting that the driving force of the whole process is the release of CO_2 in almost all steps of the polymerization, which is always accompanied by the release of the strain energy of the 5-membered Gly-NCA ring. Both H_2O and H_3BO_3 are continuously recycled (panels B-D in Scheme S1 SI). Boric acid produces catalytically the high-energy Ala-$OB(OH)_2$ intermediates.

This scenario is consistent with the known role of boric acid in promoting amide bond condensation[50], even in an abiogenic perspective[47]. However, it cannot be excluded that more complex boron intermediates could be also involved[51]. Further experimental evidence, including characterization of reaction intermediates by negative ion mass spectrometry or X-ray crystallography, is needed to verify the hypothesis sketched in Fig. 2. The boric esters intermediates proposed here are just a reasonable working hypothesis, which is based on the present computational evidence, as well as on available literature data[47,52,53].

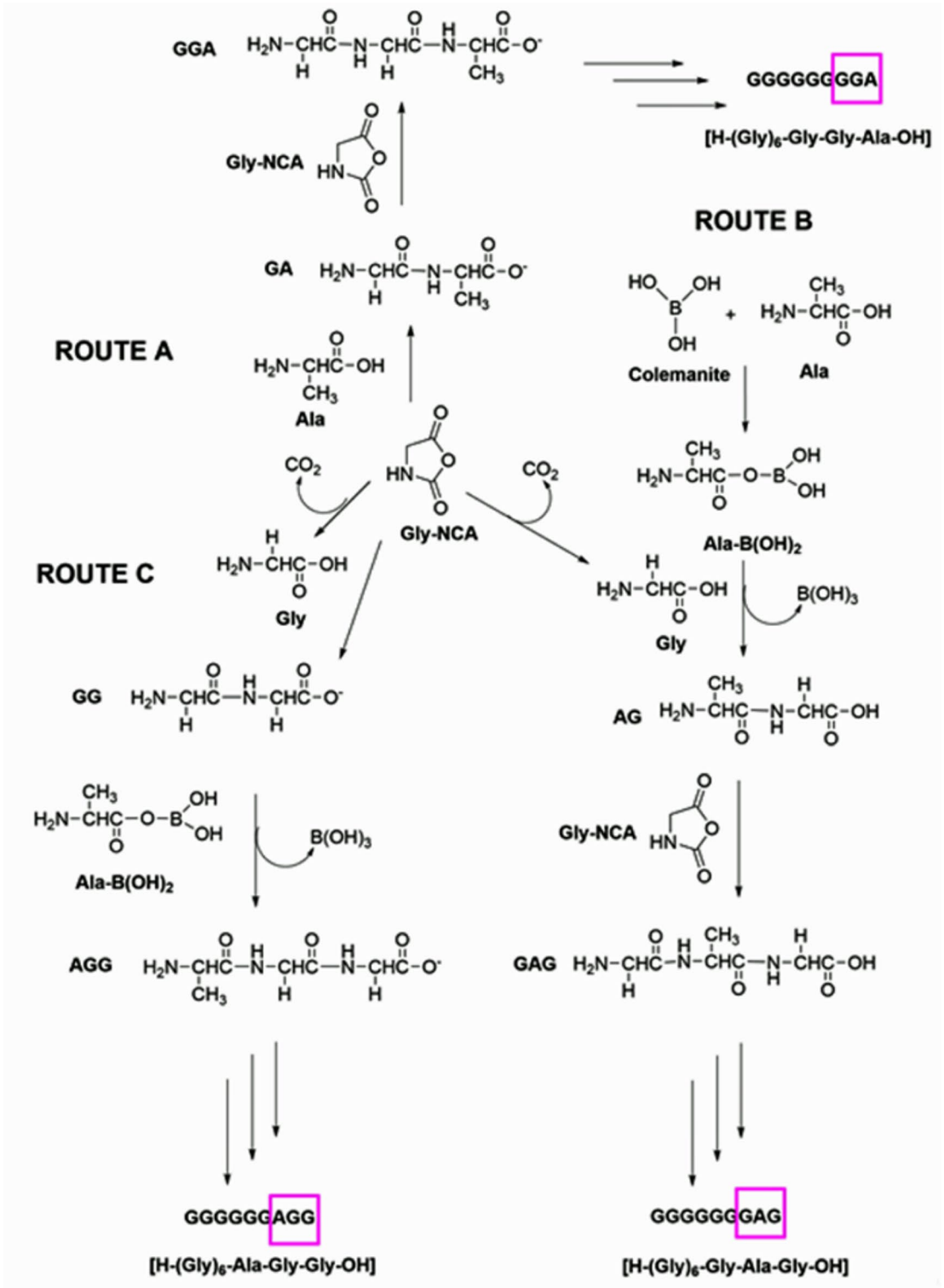

Fig. 2. Mechanisms proposed for the formation of GAG isomers. Route A: Direct addition of Ala to Gly-NCA, followed by Gly elongation, to form the A-terminal G_nA peptide.[39] Route B: Reverse Gly-Ala condensation to produce the observed GAG terminal sequence. Here Ala is activated by boric species released in the presence of Colemanite, which allow the subsequent sum of free Gly. Route C: Formation of the minoritarian AGG terminal sequence. This route is unfavored as activated Ala addition must wait for preliminary condensation of two Gly units.

Role of the abiogenic G_7AG and G_7GA peptides in protein evolution

The occurrence of G_7AG and G_7GA stretches in real proteins is crucial to understand what roles they could have played during the development of early (proto-)metabolic pathways, especially if they could be found in early proteinaceous polypeptides. To this end, we employed the NCBI Blastp tool ClusteredNR database to align the G_7AG and G_7GA amino acid sequences vs. both *Cyanobacteria* (taxid:1117) and *Thermoproteota* (*Xenarchaeota* taxid:28,889) databases.[54] The alignment, limited to 100 sequences for each considered kingdom in order to keep the highest score frequencies, was constrained to (i) *Cyanobacteria*, as primeval lifeforms thriving in anaerobic environments but able to produce oxygen and (ii) *Thermoproteota* (formerly *Crenarchaeota*), as phylum belonging to *Archaea* domain (synonim of *Korarchaeota* or *Xenarchaeota* phylum) as the closest archeal relatives to LUCA and first forms of life observed in extreme environments. In particular, *Thermoproteota* are extremophiles sulfur-dependent, thermophilic or hyperthermophilic microbes able to grow up to 113°C[52]. These microbes can colonize both extreme (thermophilic, hyperacidophilic, hyperhalophilic and/or metal-resistant) and non-extreme environments[53]. They are often considered key to understanding the environmental and evolutionary context of the Last Universal Common Ancestor (LUCA)[55]. Genomic and biochemical reconstructions of ancestral proteins and ribosomal RNA suggest that LUCA thrived in moderately to highly elevated temperatures, potentially inhabiting hydrothermal environments[56,57].

The aminoacidic sequences G_7GA and G_7AG are both present in *Thermoproteota* and in *Cyanobacteria*, even if in the latter those show a wider distribution. In *Cyanobacteria* the G_7GA sequence shows higher frequency than G_7AG, both concerning the number of organisms and the number of protein hits. In *Thermoproteota*, the frequency of hits for G_7AG and G_7GA stretches is also significant, as these are harbored in several proteins even if less than those found in *Cyanobacteria* (Fig. 3).

In *Cyanobacteria*, *Cyanophyceae* and unclassified *Cyanobacteriota* were the two main groups identified for both G_7AG and G_7GA fragments (Fig. 4). Among *Cyanophyceae*, *Oscillatoriophycidae* is the major order harboring the G_7AG while the G_7GA sequence has a comparable distribution between *Oscillatoriophycidae* and *Nostocales* orders. *Nostocales* is the order of *Cyanobacteria* containing most of its forms. Their shape is quite various due to their filamentous morphology which could be simple or branched[58]. This complexity requires a strong adaptation, which most probably taking advantage to the protein structural flexibility. Interestingly, cyanobacteria species like *Nostoc* require boron to grow efficiently in a range of 0.1–1 ppm. *Nostoc* cultures without boron develop chlorosis, a strong acidification which leads to the lacking of the typical green color[59].

A broader analysis concerning *Cyanobacteria* and the *Xenarchaeota* shows how some G_7GA and G_7AG aminoacidic stretches dropped, for example, into DUF (Domain of Undefined Function) and in several hypothetical proteins of both *phyla* (Tables S5–S8 SI). Interestingly, lectin-like proteins have been detected only in *Cyanobacteria* for G_7GA while only one for G_7AG fragment (Tables S7 and S8 SI). In general, lectin-like proteins recognize complex structures (usually carbohydrates) at the cell surface.

In *Thermoproteota*, two proteins related to ATP pathway and harboring the G_7GA sequence have been found. In particulr, proteins related to ATP synthase are considered ancient proteins involved in the mechanism of energy conservation. Similar proteins have not been found for the G_7AG sequence. Several proteins involved in sulfur metabolim have also been detected, harboring both G_7AG and G_7GA stretches. Enzymes involved in sulfur metabolism represent a core part of the early Earth biochemistry.G_7AG sequence has also been found in Methanogens of *Thermoproteota* phylum (*Candidatus Methanodesulfokora*, Tables S5 and S6 SI), often containing the VWA (von Willebrand Factor type A) domain, thought to be ancient and present in LUCA. In methanogens, VWA domain appears in proteins involved in cell surface interactions, cell adhesion, S-layer proteins and in those involved in protein stability in extreme environments. The presence of proteins with a VWA domain may reflect the early adaptation of proteins for cell surface interactions and environmental sensing, especially in extreme anaerobic conditions. A cupredoxin-domain containing protein has also been found in *Thermoproteota*. In methanogens, this type of domain participates in electron transfer chains during methanogenesis.

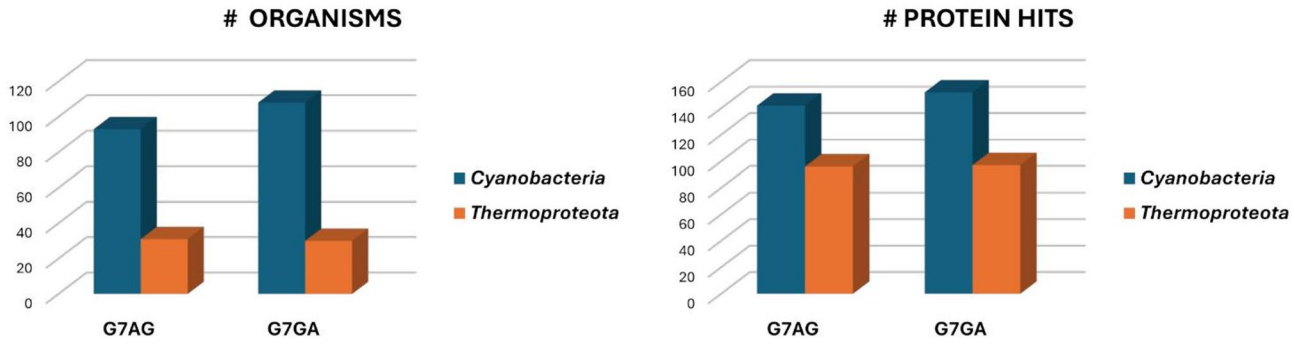

Fig. 3. Distribution of major condensation products, G_7GA and G_7AG peptides, in Cyanobacteria and *Thermoproteota* (BLAST name *Crenarchaeota*). In the first graph, blocks represent the number of organisms both in *Cyanobacteria* (blue blocks) and *Thermoproteota* (orange blocks) where G_7AG and G_7GA stretches have been found. In the second graph, blocks show the number of protein hits harboring the G_7AG and G_7GA amino acidic sequences both in *phyla*. Frequencies were computed by BLASTp tool ClusteredNR Database.

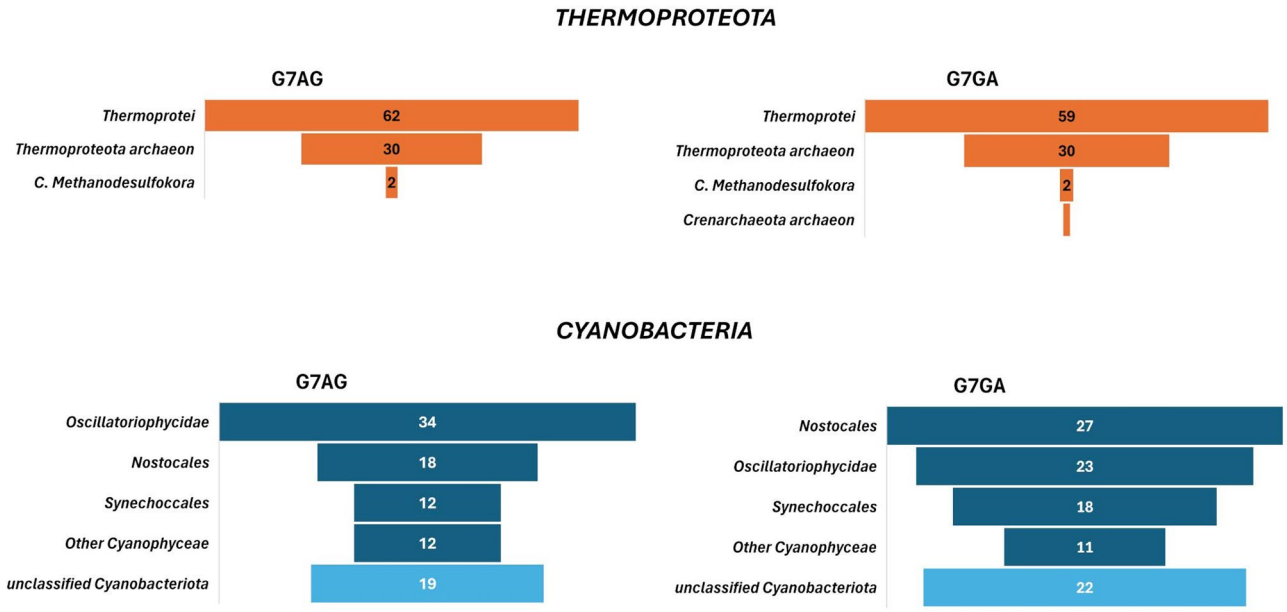

Fig. 4. Distribution of G_7AG and G_7GA short sequences in classes and orders of both *Thermoproteota* and *Cyanobacteria*. *Thermoproteota*, a phylum of the domain *Archaea*, shows a comparable number of protein hits harboring both aminoacidic stretches under investigation. *Cyanophyceae* (dark blue bars) include several groups where G_7AG prevails in *Oscillatoriophycidae* on G_7GA amino acid sequence for which *Nostocales* and *Oscillatoriophycidae* have a comparable distribution. *Nostocales*, an order of *Cyanobacteria* containing most of its forms, includes several filamentous shapes, both simple and branched. In *Cyanobacteria* panel, the light blue indicates the protein hits belonging to the group of unclassified *Cyanobacteria*.

Discussion

This investigation steps on the path traced by the Bernal's intuition[60] on the ability of some minerals to drive a spontaneous abiotic polymerization. We here provide a possible model for aa polymerization, where the chemical energy required to form the peptide bond comes from N-carboxy anhydride. The latter was synthesized here from triphosgene for convenience, but several natural synthetic routes were proposed, which could have produced NCA-activated aas on early Earth[38–41]. The main step forward is the preferential selective formation of GAG and GGA-type aas sequences, which could have been served as multi-purpose structural building blocks of primeval biopolymers. We demonstrate that the selective synthesis of strands with well-defined primary sequence is feasible in a few days at room conditions (ambient pressure and T ~ 295–310 K) under an inert atmosphere. We focused on glycine (Gly, G) and alanine (Ala, A), as their prebiotic synthesis requires the least Gibbs free energy expense[61]. As catalysts, we selected three natural compounds to span a reasonably wide range of crystallochemical conditions, namely a zeolite (heulandite), a borate (colemanite) and the simplest silicate (quartz), all compatible with water ponds related to hydrothermal and volcanic systems (Section S1 SI). Heulandite is one of the most common natural zeolites, a group of open-framework silicates with remarkable cation-exchange capacity, reversible dehydration, and catalytic properties[62]. Colemanite is the most common constituent in borate deposits formed in alkaline lacustrine environments, typically under warm conditions, as well as one of the major mineral commodities of boron[63]. Finally, quartz is one of the most diffused rock-forming silicates, occurring in a series of geological environments.

Our chemical model is consistent with the alleged composition of archaic ponds, where minerals could easily have been in equilibrium with warm water rich in organic substances. Noteworthy, the only artificial sources of energy we used were continuous shaking and moderate heating, up to 28 or 35 °C (Table S1 SI). According to the known composition of primeval rocks[64], the atmosphere of the late Hadean eon should have lacked O_2[65], and in fact the experiments carried out in the presence of air failed to produce specific aa sequences. The reason is probably the occurrence of reactive oxygen species (ROS), like OH radicals and singlet O_2, at the water–air interface[66,67]. ROS may initiate fast side reaction and prevent the aa condensation or may favor the breaking of peptide bonds[68].

While quartz and heulandite show only a modest catalytic effect, colemanite produces spontaneously GGA- and GAG-type sequences, plus shorter intermediates, without the need of artificial feeding with fresh reagents following pre-determined sequences[31]. Likely, the peptide bond condensation occurs in homogeneous phase with the aid of catalytic borate ions[47]. Quantum chemical simulations suggest that preference for the H_2N-G_nGAG-OH sequence might be due to the addition of activated Ala-$OB(OH)_2$ boric esters at the beginning of the polymerization process. The other main isomer, H_2N-G_nGGA-OH, can be produced via either direct addition of Gly units to free alanine, or condensation of alanine and high-energy Gly-$OB(OH)_2$ esters. In both cases, the driving force for successive polymerization is provided by high energy Gly-NCA intermediates. This mechanism is supported by both experimental evidence and quantum chemical simulations and highlights the

central role of colemanite-derived boric species in solution. This puts in the background the structure of the mineral surface, even though the latter could still play a role in concentrating aa monomers from the aqueous environment.

These findings raise the question of whether colemanite occurred on the primitive Earth, in sufficient quantities to promote oligomerization processes in prebiotic conditions. The formation of playa-lake borates *ca.* 2.4 By ago is plausible[69], but they occurrence back to *ca.* 4.3 By ago cannot be geologically confirmed neither denied[70,74]. More likely, in late Hadean eon, boron was concentrated in other minerals as a secondary element, for example with concentration up to 5000 wt ppm replacing aluminum in feldspars. Recently, Sumie et al.[47] discussed that the occurrence of tourmaline in 3.8 By-old metasediments in Greenland suggests that boron-rich fluids could have occurred in isolated basins on early Archean era, and proposed that these could have promoted oligomerization of glycine.

The present results are also in line with recent outcomes by Stimmer et al.[71], who showed that alkaline evaporative environments on early Earth may have favored the formation of abiotic peptides containing Gly and Ala. Interestingly, they obtained mostly Gly-rich sequences, despite different experimental conditions (absence of catalysts and higher temperature), which might suggest that multiple sources of short Gly-Ala oligomers were present in pre-biotic environments. The abundance of Gly in such aa strands likely reflects the relative abundance of this aa on early Earth. In addition, our work suggests that mineral-mediated mechanisms exist, which could make specific sequences less frequent (AGG *vs.* GGA and GAG).

The conservation of G_7AG and G_7GA sequences from very ancient lineages suggests they might represent an early evolved, functionally important motif. The presence of these sequences in proteins associated with organisms related to LUCA is consistent with (but not conclusive proof of) an ancient origin, potentially even pre-dating LUCA, and warrants future investigation into whether they represent vestiges of prebiotic chemistry or very early evolutionary innovation. G_7GA and G_7AG strands were found in very different protein families, as part either of random coil regions, or as flexible loops. Gly-rich stretches are important structural elements that confer high flexibility and elasticity, which allow the proper protein folding. Disordered Gly-rich stretches are also, for example, involved in intermolecular aggregation mechanisms. Intriguingly, most of the bacteria species showing one or more matches with G_7GA and G_7AG can grow in extreme environments characterized by high salinity, high temperature and/or anaerobiosis (Tables S5-S8 SI). By considering the distribution of both sequences, one can hypothesize that G_7GA and G_7AG could represent advantageous aminoacidic fragments to be included in protein structures for their evolution to more complex conformations (Fig. 4). G_7GA and G_7AG strands may have fostered chemical adaptation mechanisms, for example by incorporation into proto biopolymers that later could have been specialized to exploit more specific functions. Within a protein structure, those Gly-rich sequences could work as flexible bridges between two domains or as hinge regions. Some proteins contain Gly-rich motifs at sites where they need to undergo conformational changes. The wide variety of protein domains containing G_7GA and G_7AG aa stretches in primeval organisms like *Archaea* and *Cyanobacteria* could represent a model of advantageous diffusion of poly-amino acid short stretches. In particular, the presence of poly-Gly stretches coupled to a different amino acid, like Ala in our study, could play the role of a conserved system incorporated in more evolved and complex protein structures. This strategy aimed at conferring the typical flexibility of protein loops, usually characterized by poly-glycine sequences. This hypothesis pioneers how abiogenic self-assembled aa fragments could have marked the birth of advantageous key-blocks, evolved in primitive proteinaceous forms. It is mandatory to carefully consider this preposition as a hypothesis due to the complexity of discovering hidden primeval traces within aminoacidic sequences data of nowadays proteins. These primordial proteins run into continuous adaptation for functional improvement parallel to life development.

Conclusions

An integrated crystal chemical, spectroscopic and computational approach was exploited to gain insights into the abiogenesis of polypeptides with well-defined primary sequences in the primeval Earth. In the presence of the hydrous Ca-borate colemanite, a specific mixture of $G_{n1}AG_{n2}$ and $G_{ntot}A$ ($n_{tot} = n_1 + n_2$ up to 8) oligopeptides form from NCA-activated glycine and alanine spontaneously in aqueous solution under an inert atmosphere. This demonstrates the importance of a source of boron in aqueous solution to promote reproducible Ala-Gly oligomerization in prebiotic conditions, which is not potentially restricted to a unique mineral species. A few days of incubation are enough to obtain a detectable quantity of Gly-Ala oligopeptides, without the need for enzymatic catalysts, recombinant DNA or harsh chemical conditions. This evidence is robust within the framework of the present chemical model, as the results here discussed are obtained from three to five replicas of the proposed experiments and are consistent with literature evidence on possible occurrence of Gly-rich strands in early Earth scenarios. Even though it remains to be seen whether and to what extent this model is directly transferable to real-life systems, the obtained Gly-rich sequences belong to mechanically flexible regions of proteins, which were found in ancestral extremophiles. The present findings represent a first step toward a tentative molecular phylogenesis of prebiotic polypeptides, with applications that could range from the search for extraterrestrial lifeforms to the development of cheap chemical protocols for the synthesis of biocompatible fibrous materials.

Materials and methods

Experimental procedures

Synthesis of alanine N-carboxyanhydride (Ala-NCA)

In 60 mL of dry THF at 50 °C under inert atmosphere, 3.11 g of triphosgene (10.5 mmol, 0.5 eq.) was added to 1.87 g (21.0 mmol, 1.0 eq.) of L-alanine and stirred for 5 h. The solvent was reduced and the solid was washed

and dried with ethyl acetate (3 × 50 mL) in order to remove the excess of triphosgene. The residue was purified by crystallization from a mixture of ethyl acetate/hexane. White solid, 70% yield, $C_4H_5NO_3$ (MW = 115.03); ^{1}H-NMR (300 MHz, dmso-d_6) δ 8.99 (s, NH), 4.46 (q, *J* = 6.1 Hz, 1H), 1.31 (d, *J* = 7.0 Hz, 3H). All reagents were purchased from Sigma-Aldrich unless otherwise specified.

Synthesis of alanine-glycine-glycine (AGG), glycine-alanine-glycine (GAG) and glycine-glycine- alanine standard
Peptides were synthesized by microwave-assisted automated Fmoc/*tBu*-based solid phase peptide synthesis (MW-SPPS) using Liberty Blue synthesizer (CEM Corporation). Wang resin was used as solid support and preloaded with Fmoc-Ala-OH or Fmoc-Gly-OH obtaining a loading of 0.6–0.9 mmol/g. The synthesis was carried out on a 0.1 mmol scale. Coupling was performed using 5 eq. of the Fmoc-protected aa, previously dissolved in DMF to obtain a 0.2 M solution. As coupling reagents 5 eq. of N,N′-Diisopropylcarbodiimide (DIC, 0.5 M in DMF) and 5 eq of.

Oxyma Pure (1 M in DMF) were used. To deprotect the Fmoc group, a solution of piperidine in DMF (20% (v/v) was applied. The coupling reaction was accomplished at 25 °C for 120 s, followed by 480 s at 50 °C and 35W. The cleavage of the peptides was performed using the scavenger cocktail thioanisole/water/TFA (2.5:2.5:95) for 180 min. After the cleavage, the peptides were precipitated and washed using ice-cold *ter*t-butyl methyl ether/hexane (2:1). The peptides were purified by RP-HPLC as described after. The purified peptides was freeze-dried and stored at 0 °C. ESI–MS for $C_7H_{13}N_3O_4$: calcd 203.09; found for GAG 204.0 $[M+H]^+$; 406.9 $[2\ M]^+$ and for AGG and GGA 226.1 $[M+Na]^+$; 248.1 $[M+2Na]^+$[72].

General procedure for catalytic formation of polypeptides
In a Schlenk flask under inert atmosphere, 1 eq. of the corresponding anhydride was added to 1 eq. of aminoacid (L-Ala or L-Gly) in aqueous media and to the appropriate catalyst (10–30% w/w). The mixture was stirred on a reciprocal shaker (150 rpm) for 72 h at 28 °C. To remove the minerals suspended into solution, the mixture was centrifuged (6000 rpm, 5 min) and the supernatant was analyzed by MS-HPLC analysis using UHPLC. Each experiment was carried out in triplicate.

Optimized procedure for preparation of polypeptides using colemanite as a catalyst
In a Schlenk flask under inert atmosphere, 1 eq. of the corresponding Ala-NCA or Gly-NCA was added to a 1 eq. of amino acid (L-Ala or L-Gly) in PBS buffer pH 7.4 and to colemanite catalyst (10% w/w). The mixture was stirred on a reciprocal shaker (150 rpm) for 72 h at 28 °C. To remove the minerals suspended into solution, the mixture was centrifuged (6000 rpm, 5 min) and the supernatant was analyzed by MS-HPLC analysis. Each experiment was carried out 5 times.

Preparative HPLC
The purification of the peptide/the lyophilized mixture was carried out by preparative reverse phase HPLC. The elution was performed using Adamas® C18-AQ column (5u, 250 mm × 21.2 mm) with a linear gradient of 1–30% B in A in 30 min with a flow rate of 20 ml/min (A: 0.1%TFA in H_2O; B: 0.1% TFA in ACN). The UV absorbance was detected at 220 nm and the peptide containing fractions were collected in plastic tubes. Afterwards, the solution was concentrated and then lyophilized.

Analytical HPLC
The characterization of the peptides formed during the reaction was carried out by means of reverse phase high-performance liquid chromatography electrospray ionization mass spectrometry (HPLC–ESI–MS) using an Ultimate 3000 (Thermo Scientific) UHPLC system connected to a LCQ Fleet (Thermo Scientific) mass spectrometer. The linear gradient used was typically 1–60% B in A in 15 min with a flow rate of 1 ml / min (A: 0.1% FA in water, B: 0.1% FA in ACN) with a Zorbax-SB C18 column (5 um, 4.6 × 250 mm) and isocratic 2% B in A (A: 0.1% TFA in water, B: 0.1% TFA in ACN) with a YMC Carotenoid column (5 µm, 4.6 × 250 mm). The eluent was ionized in the mass spectrometer and the m/z values of the pseudo-molecular ions were detected. Control of HPLC and ESI–MS systems as well as the evaluation of the mass spectra were performed with the Software Xcalibur (Version 2.2, Thermo Scientific). Images of mass spectra were processed with Origin.

Computational procedures

Quantum chemical simulations
Quantum mechanical geometry optimizations were carried out with Gaussian16[73] at the DFT theory level with the hybrid B3LYP functional, in conjunction with the triple zeta Pople's basis set 6-311G(p,d). The purpose was to extract thermochemical data on the formation of specific bonded aa sequences, starting from the partition functions of isolated building blocks. The solvent (water) was implicitly accounted for with a Polarizable Continuum Model through the Truhlar's SMD variation of the integral equation formalism[74].

Data mining
Peptide alignments were generated by NCBI Blast tool ClusteredNR database. The dataset of relevant Gly-rich peptides was limited to 100 sequences to focus on the highest score frequency due to the very short aminoacidic fragments available. *Cyanobacteria* (taxid: 1117) and *Xenarchaea* (taxid: 28,889) databases[54] were selected as reference set for sequence homology investigations. The alignment was limited to 100 sequences for each taxon. The peptide location within each analyzed protein structure was taken as reference to the hypothetical role of the peptides in protein conformation.

Data availability

All the relevant data associated to this study were deposited as Electronic Supplementary Information (see ESI). Further evidence, e.g. negative results, are available from the Authors upon reasonable request. Gaussian I/O files can be retrieved free of charge from the repository Zenodo (https://doi.org/10.5281/zenodo.13911575). To request the data from this study, kindly contact prof. Leonardo Lo Presti (e-mail: leonardo.lopresti@unimi.it).

Received: 18 November 2024; Accepted: 29 July 2025
Published online: 08 August 2025

References

1. Miller, S. L. A production of amino acids under possible primitive earth conditions. *Science* **117**, 528–529 (1953).
2. Fleury, B., Carrasco, N., Millan, M., Vettier, L. & Szopa, C. Organic chemistry in a CO_2-rich early Earth atmosphere. *Earth Planet. Sci. Lett.* **479**, 34–42 (2017).
3. Jiang, L. et al. Abiotic synthesis of amino acids and self-crystallization under prebiotic conditions. *Sci. Rep.* **4**, 6769 (2014).
4. Ménez, B. et al. Abiotic synthesis of amino acids in the recesses of the oceanic lithosphere. *Nature* **564**, 59–63 (2018).
5. Koga, T. & Naraoka, H. A new family of extraterrestrial amino acids in the Murchison meteorite. *Sci. Rep.* **7**, 636 (2017).
6. Naraoka, H. et al. Soluble organic molecules in samples of the carbonaceous asteroid (162173) Ryugu. *Science* **379**, eabn9033 (2023).
7. Cronin, J. R., Pizzarello, S. & Moore, C. B. Amino acids in antarctic carbonaceous chondrite. *Science* **206**, 335–337 (1979).
8. Glavin, D. P. et al. Extraterrestrial amino acids and L-enantiomeric excesses in the CM2 carbonaceous chondrites Aguas Zarcas and Murchison. *Meteorit. Planet. Sci.* **56**, 148–173 (2021).
9. Abramov, O. & Mojzsis, S. J. Microbial habitability of the Hadean Earth during the late heavy bombardment. *Nature* **459**, 419–422 (2009).
10. Xavier, J. C., Hordijk, W., Kauffman, S., Steel, M. & Martin, W. F. Autocatalytic chemical networks at the origin of metabolism. *Proc. R. Soc. B* **287**, 20192377 (2020).
11. Kitadai, N. & Maruyama, S. Origins of building blocks of life: A review. *Geosci. Front.* **9**, 1117–1153 (2018).
12. Radzicka, A. & Wolfenden, R. Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases. *J. Am. Chem. Soc.* **118**, 6105–6109 (1996).
13. Dobry, A., Fruton, J. S. & Sturtevant, J. M. Thermodynamics of hydrolysis of peptide bonds. *J. Biol. Chem.* **195**, 149–154 (1952).
14. Flegmann, A. W. & Tattersall, R. Energetics of peptide bond formation at elevated temperatures. *J. Mol. Evol.* **12**, 349–355 (1979).
15. Stribling, R. & Miller, S. L. Energy yields for hydrogen cyanide and formaldehyde syntheses: The HCN and amino acid concentrations in the primitive ocean. *Orig. Life Evol. Biosph.* **17**, 261–273 (1987).
16. Song, X., Simonis, P., Deamer, D. & Zare, R. N. Wet–dry cycles cause nucleic acid monomers to polymerize into long chains. *Proc. Natl. Acad. Sci.* **121**, e2412784121 (2024).
17. Damer, B. & Deamer, D. Coupled phases and combinatorial selection in fluctuating hydrothermal pools: A scenario to guide experimental approaches to the origin of cellular life. *Life* **5**, 872–887 (2015).
18. Damer, B. A field trip to the archaean in search of Darwin's warm little pond. *Life* **6**, 21 (2016).
19. Frenkel-Pinter, M., Sargon, A. B., Glass, J. B., Hud, N. V. & Williams, L. D. *RSC Adv.* **11**, 3534–3538 (2021).
20. Fialho, D. M. et al. Depsipeptide nucleic acids: Prebiotic formation, oligomerization, and self-assembly of a new proto-nucleic acid candidate. *J. Am. Chem. Soc.* **143**(34), 13525–13537 (2021).
21. Luisi, P. L. Prebiotic metabolic networks?. *Mol. Syst. Biol.* **10**, 729 (2014).
22. Frenkel-Pinter, M., Samanta, M., Ashkenasy, G. & Leman, L. J. Prebiotic peptides: Molecular hubs in the origin of life. *Chem. Rev.* **120**, 4707–4765 (2020).
23. Cairns-Smith, A. G. The origin of life and the nature of the primitive gene. *J. Theor. Biol.* **10**, 53–88 (1996).
24. Hartman, H. Speculations on the origin and evolution of metabolism. *J. Mol. Evol.* **4**, 359–370 (1975).
25. Lambert, J. F. Adsorption and polymerization of amino acids on mineral surfaces: A review. *Orig. Life Evol. Biosph.* **38**, 211–242 (2008).
26. Erastova, V., Degiacomi, M. T., Fraser, D. G. & Greenwell, H. C. Mineral surface chemistry control for origin of prebiotic peptides. *Nat. Commun.* **8**, 2033 (2017).
27. Brigiano, F. S., Gierada, M., Tielens, F. & Pietrucci, F. Mechanism and free-energy landscape of peptide bond formation at the silica-water interface. *ACS Catal* **12**, 2821–2830 (2022).
28. Ohara, S., Kakegawa, T. & Nakazawa, H. Pressure effects on the abiotic polymerization of glycine. *Orig. Life Evol. Biosph.* **37**, 215–223 (2007).
29. Ferris, J. P., Hill, A. R., Liu, R. & Orgel, L. E. Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381**, 59–61 (1996).
30. 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).
31. Okamoto, R. et al. Regioselective α-peptide bond formation through the oxidation of amino thioacids. *Biochemistry* **58**, 1672–1678 (2019).
32. Bujdák, J. & Rode, B. M. Preferential amino acid sequences in alumina-catalyzed peptide bond formation. *J. Inorg. Biochem.* **90**, 1–7 (2002).
33. Leman, L. J., Huang, Z. Z. & Ghadiri, M. R. Peptide bond formation in water mediated by carbon disulfide. *Astrobiology* **15**, 709–716 (2015).
34. Pascal, R., Boiteau, L. & Commeyras, A. From the prebiotic synthesis of α-amino acids towards a primitive translation apparatus for the synthesis of peptides. *Top. Curr. Chem.* **259**, 69–122 (2005).
35. Kricheldorf, H. R. Polypeptides and 100 years of chemistry of α-Amino acid N-carboxyanhydrides. *Angew. Chem. Int. Ed.* **45**, 5752–5784 (2006).
36. Taillades, J. et al. N-carbamoyl amino acid solid-gas nitrosation by NO/NO_x: A new route to oligopeptides via a-amino acid N-carboxyanhydride. Prebiotic implications. *J. Mol. Evol.* **48**, 638–645 (1999).
37. Taillades, J. et al. N–carbamoyl–α–amino acids rather than free α amino acids formation in the primitive hydrosphere: A novel proposal for the emergence of prebiotic peptides. *Orig. Life Evol. Biosph.* **28**, 61–77 (1998).
38. Collet, H., Bied, C., Mion, L., Taillades, J. & Commeyras, A. A new simple and quantitative synthesis of α–aminoacid–N–carboxyanhydrides (oxazolidines–2,5–dione). *Tetrahedron Lett.* **37**, 9043–9046 (1996).
39. Leman, L., Orgel, L. & Reza Ghadiri, M. Carbonyl sulfide-mediated prebiotic formation of peptides. *Science* **306**, 283–286 (2004).
40. Kramer, J. R., Onoa, B., Bustamante, C. & Bertozzi, C. R. Chemically tunable mucin chimeras assembled on living cells. *Proc. Natl. Acad. Sci. USA* **112**, 12574–12579 (2015).
41. Gülensoy, H. & Gedikbey, T. Solubilities of colemanite, ulexite, pandermite and tunellite minerals in different alkaline solutions. *Bullet. Sch. Eng. Archit. Sakarya* **8**, 37–40 (1981).
42. Karagöz, Ö. & Kuşlu, S. Dissolution kinetics of colemanite in potassium dihydrogen phosphate solution (KH_2PO_4). *Int. J. Hydrog. Energy* **42**, 23250–23259 (2017).

43. Ricardo, A., Carrigan, M. A., Olcott, A. N. & Benner, S. A. Borate minerals stabilize ribose. *Science* **303**(5655), 196 (2004).
44. Saleem, M., Khanif, Y. M., Ishak, Y. M. F. & Samsuri, A. W. Solubility and leaching of boron from borax and colemanite in flooded acidic soils. *Commun. Soil Sci. Plant Anal.* **42**, 293–300 (2011).
45. Sumie, Y., Sato, K., Kakegawa, T. & Furukawa, Y. Boron-assisted abiotic polypeptide synthesis. *Commun. Chem.* **6**, 89 (2023).
46. Denkewalter, R. G. et al. The controlled synthesis of peptides in aqueous Medium. I. The use of α-amino acid N-carboxyanhydrides. *J. Am. Chem. Soc.* **88**, 3163–3164 (1966).
47. Greenwald, J., Friedmann, M. P. & Riek, R. Amyloid aggregates arise from amino acid condensations under prebiotic conditions. *Angew. Chem. Int. Ed.* **55**, 1–6 (2016).
48. Kopple, K. D. The reaction of amines with amino acid N-carboxy anhydrides. *J. Am. Chem. Soc.* **79**, 662–664 (1957).
49. Lackinger, M. Synthesis on inert surfaces. *Dalton Trans.* **50**, 10020–10027 (2021).
50. Mylavarapu, R. K. et al. Boric acid catalyzed amidation in the synthesis of active pharmaceutical ingredients. *Organ. Process Res. Dev.* **11**, 1065–1068 (2007).
51. Arkhipenko, S. et al. Mechanistic insights into boron-catalysed direct amidation reactions. *Chem. Sci.* **9**, 1058–1072 (2018).
52. Hamm, J. N. et al. The parasitic lifestyle of an archaeal symbiont. *Nat. Commun.* **15**, 6449 (2024).
53. Dombrowski, N., Lee, J.-H., Williams, T. A., Offre, P. & Spang, A. Genomic diversity, lifestyles and evolutionary origins of DPANN archaea. *FEMS Microbiol. Lett.* **366**, 2 (2019).
54. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
55. Glansdorff, N., Xu, Y. & Labedan, B. The Last Universal Common Ancestor: emergence, constitution and genetic legacy of an elusive forerunner. *Biol. Direct* **3**, 29 (2008).
56. Mat, W. K., Xue, H. & Wong, J. T. The genomics of LUCA. *Front. Biosci.* **13**, 5605–5613 (2008).
57. Boussau, B. et al. Parallel adaptations to high temperatures in the Archaean eon. *Nature* **456**, 942–945 (2008).
58. Hentschke, G. S. & Gama Jr, W. A. Trends in Cyanobacteria: A contribution to systematics and biodiversity studies. In *The Pharmacological Potential of Cyanobacteria*, Chapter 1, 1–20, Eds G. Lopes, M. Silva, V. Vasconcelos, Academic Press (Elsevier, 2022).
59. Eyster, C. Necessity of boron for *Nostoc muscorum*. *Nature* **170**, 755 (1952).
60. Bernal, J. D. The physical basis of life. *Proc. Phys. Soc. Sect. A* **62**, 537–558 (1949).
61. Higgs, P. G. & Pudritz, R. E. Thermodynamic basis for prebiotic amino acid synthesis and the nature of the first genetic code. *Astrobiology* **9**, 483–490 (2009).
62. Colella, C. & Wise, W. S. The IZA handbook of natural zeolites: A tool of knowledge on the most important family of porous minerals. *Microporous Mesoporous Mater.* **189**, 4–10 (2014).
63. Lotti, P. et al. Crystal chemistry and temperature behavior of the natural hydrous borate colemanite, a mineral commodity of boron. *Phys. Chem. Miner.* **45**, 405–422 (2018).
64. Kasting, J. F. Earth's early atmosphere. *Science* **259**, 920–926 (1993).
65. Halevy, I. & Bachan, A. The geologic history of seawater pH. *Science* **355**, 1069–1071 (2017).
66. Anglada, J. M., Martins-Costa, M. T. C., Francisco, J. S. & Ruiz-López, M. F. Photoinduced oxidation reactions at the air–water interface. *J. Am. Chem. Soc.* **142**, 16140–16155 (2020).
67. Meroni, D. et al. Oxygen vacancies in the spotlight: On the engineering of intrinsic defects in highly defective TiO_2 photocatalysts. *J Photochem. Photobiol. A Chem.* **444**, 114916 (2023).
68. Pattison, D. I., Rahmanto, A. S. & Davies, M. J. Photo-oxidation of proteins. *Photochem. Photobiol. Sci.* **11**, 38–53 (2012).
69. Grew, E. S., Bada, J. L. & Hazen, R. M. Borate minerals and origin of the RNA world. *Orig. Life Evol. Biosph.* **41**, 307–316 (2011).
70. Hazen, R. M. & Morrison, S. M. Mineralogical environments of the Hadean Eon: Rare elements were ubiquitous in surface sites of rock-forming minerals: rare elements ubiquitous on rock-forming mineral surfaces in the Hadean. In *Advances in Astrobiology and Biogeophysics* 43–61 (2021).
71. Stimmer, J., Kakegawa, T. & Furukawa, Y. Abiotic formation of glycine-alanine peptides in alkaline evaporative environments. *Geochem. J.* **58**, 217–226 (2024).
72. Pellegrino, S. et al. Ctr-1 Mets7 motif inspiring new peptide ligands for Cu(i)-catalyzed asymmetric Henry reactions under green conditions. *RSC Adv.* **6**, 71529–71533 (2016).
73. Frisch M. J. *et al*., Gaussian16 revision A.03, Gaussian Inc., (2016).
74. Marenich, A. V., Cramer, C. J. & Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **113**, 6378–6396 (2009).

Acknowledgements

The authors thank Università degli Studi di Milano for funding this research through the "Piano di Sostegno alla Ricerca 2020 grant: Seal of Excellence—SoE SEED 2020" Proj. Code: 1080, titled: "Abiogenesis: a geochemical route for the origin of life- GO_LIFE". The technological platform Unitech COSPECT at Università degli Studi di Milano is warmly acknowledged for the provision of spectroscopic and advanced microscopy tools. Publication costs were covered by Università degli Studi di Milano through the Research Support Plan 2023 ("Piano di sostegno alla Ricerca 2023-LINEA 2") and "Fondo APC of Pharmaceutical Science Department".

Author contributions

Conceptualization was carried out by L.L.P.; Funding acquisition was managed by L. L. P., G. D. G. and I.R. The experimental work was performed by I.R., G. F. and L. F. The computational study, including data mining, was carried out by L. L. P. and L. S. Mineral samples and their characterization were provided and performed by G. D. G. and P. L. All Authors (I.R., G. F., L. F., G. D. G., P. L., S. R., M. S. C., L. S. and L. L. P.) equally contributed to data analysis, interpretation, and curation. The manuscript was written by L. L. P., G. D. G., I. R., G. F. and P. L.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-025-14072-2.

Correspondence and requests for materials should be addressed to L.L.P.