

# Origins of life and molecular information: selectivity in mineral surface induced prebiotic amino acids polymerization

Lise Bedoin, Sandra Alves\*, Jean-François Lambert\*

Amino acids, peptides, origins of life, information, mass spectrometry, FT-ICR.

---

**ABSTRACT:** In current living matter, biopolymers follow specific sequences that give them special properties, such as the sequence of amino acids in proteins and peptides. A major challenge for the elucidation of the origins of life lies in understanding how e.g. non-random polypeptides have been selected among all the possible ones. While many investigations established plausible prebiotic polymerization pathways, surprisingly, only a few attempted to study the selectivity of these processes. We studied a mineral surface polymerization scenario based on moderate thermal activation of leucine + glutamic acid mixtures on silica. Oligopeptides up to octamers were quantitatively formed in a “clean” prebiotic reaction and analyzed by high-resolution mass spectrometry, using FT-ICR spectrometry for unambiguous molecular assignments. Non-trivial oligomerization selectivities are evidenced in both stoichiometric compositions and AA sequence, while comparable selectivities are not observed in other polymerization scenarios. They must therefore be due to specific catalytic reaction pathways occurring on the SiO<sub>2</sub> surface. A statistical measure of information contained in oligopeptide distributions is proposed. It could be used to follow the evolution of potentially meaningful complexity in biopolymers from the mineral to the biochemical world.

---

## 1- Introduction

The study of the origins of life on the primitive Earth is a complex puzzle, involving several successive steps of matter organization. The emergence of simple biomolecules has been demonstrated over 60 years ago, when the famous Miller experiments<sup>1</sup>, recently reanalyzed,<sup>2</sup> showed that amino acids (AAs) could emerge through non-biological chemical reactions. Alternatively, amino-acid blocks could also have been synthesized in interplanetary space and brought to Earth by meteorites<sup>3</sup>. The next step in the emergence of biochemistry-like complexity would have been their polymerization into oligopeptides, small-scale versions of proteins that can exhibit chemical functions, such as enzyme-like catalysis or self-replication,<sup>4,5,6</sup> even for relatively small sized oligomers.<sup>7</sup> Amino acid polymerization involves a condensation reaction between the amino and carboxylic groups of two AAs, forming a peptide bond. This reaction is endergonic, and thus disfavored in water solution. Several alternative scenarios have been proposed for peptide bond formation in primitive Earth conditions, including the use of plausible sacrificial activators, allowing polymerization through N-carboxyanhydrides chemistry.<sup>8</sup> The simplest scenarios, however, involve drying steps: removal of water, a reaction product,

makes AA polymerization more favorable according to basic thermodynamic principles. Indeed, simple drying of aqueous solutions with moderate temperature activation has been shown to induce the formation of potentially useful peptides when applied to pure amino acids or mixtures of them.<sup>9,10</sup> However, in realistic geochemical settings, it is likely that drying would have occurred in the presence of minerals rather than in a bulk phase. Around 1950, Bernal<sup>11</sup> had already proposed the implication of mineral surfaces, presumably present on the primitive Earth, in AA polymerization. In addition to theoretical discussions,<sup>12,13</sup> experimental evidence of amino acid oligomerization on different surfaces was presented by many investigators,<sup>14,15,16,17,18</sup> including our previous works on high-surface fumed silica.<sup>19,20,21,22</sup> In the latter case, in addition to its thermodynamic effect on the polymerization reaction by providing a drying platform, the silica surface also has a kinetic effect as it helps AA condensation by mild acid catalysis, allowing the reaction to proceed at lower temperatures than in the bulk.<sup>19</sup>

Thus, surface AA polymerization is well established, but strangely enough some key questions relevant to the prebiotic potential of this scenario have hardly been investigated. Among them is the question of selectivity. Most of surface polymerization studies concern single amino acids.

Not only is this situation unlikely to have happened in realistic settings where amino acid synthesis pathways generally produce a mixture of several species; even more importantly, if the next step is to be the development of function in oligopeptides, then their amino acid sequences are crucial in conferring on them whatever special properties they may have. In extant organisms, protein sequences are precisely coded in the genetic blueprint, with little tolerance for substitution. Even though there may have been a somewhat higher tolerance for variation in primitive functional oligopeptides, the probability of forming the “right” sequence in a random polymerization falls to astronomically low values very quickly when the oligomer length increases. As a consequence, the expected concentration of functional peptides would be very low, compromising the possibility of starting any “interesting” chemistry.

In the literature, the issues of polymerization selectivity in prebiotic oligopeptide synthesis and of the existence of non-random peptide sequences were only addressed in a few cases (somewhat more developed reflections are available for nucleotides polymerization<sup>23</sup>). In early studies, a non-random peptide oligomerization has been postulated under thermal activation of a mixed Glu + Asp solid phase.<sup>24,25</sup> Nakashima *et al.*<sup>26</sup> copolymerized a mixture of glycine, glutamic acid and tyrosine and reported that among the eight possible tripeptide sequences, two (PyroGlu-Tyr-Gly and PyroGlu-Gly-Tyr) were largely predominant. In order to explain these experimental observations, Hartmann *et al.*<sup>27</sup> proposed a polymerization mechanism attempting to rationalize di- and tripeptides abundances in the system. These authors assumed that pyroglutamic acid reacts with cyclic dimers of the other two AAs (substituted diketopiperazines or DKPs for short) to specifically form two tripeptides. Mosqueira *et al.*<sup>28</sup> further elaborated on these results with a probabilistic model but did not provide new experimental data. Later on, Bujdák and Rode<sup>29</sup> studied surface condensation products from neutral hydrophobic amino acid couples. They observed a degree of polymerization selectivity, as Gly-X sequences were preferentially formed over X-Gly sequences when X = Leu or Val, whereas the reverse was observed when X = Pro or Ala. The authors explained such a non-random peptide synthesis by inductive effects of the amino acid side chains. Thus, while not absent from the literature and sometimes tantalizing, studies of AA polymerization selectivity remain few and often anecdotal. This overview would not be complete without mentioning the recent works by Krishnamurthy *et al.*<sup>30,31,32</sup> on depsipeptides, the products of co-condensation of  $\alpha$ -hydroxy acids with amino acids on SiO<sub>2</sub> nanoparticles surfaces. In these more complicated systems, the authors established interesting selectivity patterns, including a preference for proteinaceous over non-proteinaceous amino acids.

We undertook the present work in order to obtain hard data on the occurrence or not of selectivity in surface mineral-mediated AA polymerization. Non-porous fumed silica (SiO<sub>2</sub>) nanoparticles were chosen as the mineral phase because silica is known to constitute an efficient platform for simple heating-induced AA condensation. The simplest possible systems to study selectivity consist in a pair of different amino acids: here a mixture of leucine (Leu or L) and glutamic acid (Glu or E) was chosen. The choice of these amino acids was dictated by several considerations. First, it is quite logical to start with two AAs with different intrinsic properties: Glu is hydrophilic and acidic, whereas Leu is hydrophobic. This has practical consequences on peptide structuring as alternating Glu-Leu peptides are known to produce well-defined peptide secondary structures.<sup>33,34</sup> Second, upon single-step thermal activation, many amino acids<sup>18,19,22,35</sup> or couples of amino acids,<sup>27</sup> will overwhelmingly produce cyclic dimers (DKPs), which are not at first sight a significant step forward towards complex biomolecules. It has been shown by several authors that cyclic dimers may be “recycled” to longer polymers by using wetting-and-drying cycles,<sup>36,37</sup> and these scenarios are highly interesting as pathways to complexity. However, they involve successive treatments that are rather difficult to analyze and, as a first step towards a mechanistic understanding of peptides formation, we focused the present study on a simple thermal activation procedure.

Obtaining long peptides instead of cyclic dipeptides involves a first level of reaction selectivity. In the literature, the formation of polypeptides up to 16-mers was observed on silica, using a vapor phase deposition method,<sup>38</sup> but such a scenario is less relevant for prebiotic chemistry than adsorption from an aqueous solution. Using the latter method, Sakhno *et al.* reported the production of linear oligopeptides up to the hexamers in a single thermal activation step from two amino acid couples deposited on silica,<sup>39</sup> i.e. (Glu+Leu, and Asp+Val) pairs, each system involving one trifunctional AA. However, the products of (Asp+Val) gave rise to practical difficulties in the analysis procedure (see Materials and Methods). Finally, in an origin-of-life perspective, both glutamic acid and leucine are reasonable choices. Even if they do not belong to the “short list” singled out in the “GAVD hypothesis”,<sup>40,41</sup> they are part of a longer list of ten or so amino acids that are synthesized by prebiotic reactions and/or found in meteorites.<sup>42</sup>

In the present work, analysis of AA condensation products was performed on desorption solutions using mass spectrometry (MS). Commonly used in proteomics, MS, in particular Fourier Transform Ion Cyclotron Resonance (FT-ICR) spectrometry, appears as the most suitable analytical technique thanks to its selectivity, sensitivity and specificity, to characterize mixtures as complex as our condensation products. Especially, FT-ICR is an extremely powerful MS device providing high analytical performances in terms of mass resolution, mass measurement accuracy and wide dynamic range.<sup>43</sup> Thus, the present study

also aims to demonstrate the potential of an FT-ICR instrument equipped with the dynamically harmonized cell for a prebiotic chemistry application. The use of accurate mass measurements demonstrates the formation of non-random oligopeptide sequences.

## 2- Materials and Methods

The mineral support used for the adsorption of amino acids was the fumed silica ( $\text{SiO}_2$ ) Aerosil 380 provided by Evonik Industries (Hanau, Germany), with BET surface area of  $380 \text{ m}^2/\text{g}$ . L-Leucine and L-Glutamic acid were purchased from Sigma Aldrich (Steinheim, Germany). The tripeptides LEL, ELL, ELE, LEE and LLE were synthesized by the Institut de Biologie Paris Seine (Sorbonne Université, Paris) and LLL, EEE and EEL were purchased from Bachem (Bubendorf, Switzerland).

Single amino acids and/or mixtures of both were deposited on silica from aqueous solutions using the “incipient wetness impregnation” (IWI) method, derived from the field of supported catalysts synthesis: the required amounts of the AAs were dissolved in water and the resulting solution was then added to silica in a ratio of 10 mL for 1 g of silica. This resulted in a slurry without a separate liquid phase, which was then dried under air flow at room temperature overnight. A reference sample was prepared with 1.25 % Leu and 1.4 % Glu (all loadings are in weight percent), corresponding to a Leu:Glu molar ratio of 1:1. In this communication, “Leu+Glu/ $\text{SiO}_2$ ”, without further specification, refers to this reference sample. The reference loading corresponds to 0.15 Leu and 0.15 Glu molecules per  $\text{nm}^2$ . This would be about one tenth of a physical monolayer based on the sizes of the two amino acids: the amino acids are then relatively “scarce” on the surface.

Additionally, 2.5 % Leu+ 1.4 % Glu/ $\text{SiO}_2$  and 6.25 % Leu+ 1.4 % Glu/ $\text{SiO}_2$ , corresponding to 2:1 and 5:1 Leu:Glu ratios, and single amino acids 1.25 % Leu/ $\text{SiO}_2$  and 1.4 % Glu/ $\text{SiO}_2$  samples were also prepared. They will be denoted as “Leu+Glu/ $\text{SiO}_2$  2:1” and “Leu+Glu/ $\text{SiO}_2$  5:1”.

Thermal activation of the AA(s) / silica systems was carried out in a U-shaped cell under argon flow, placed in a tubular oven controlled by a temperature programmer. A  $1^\circ\text{C}/\text{min}$  ramp was used to reach a final value of  $160^\circ\text{C}$ , where a plateau was maintained for 30 min. In order to obtain reproducible assays, the linearity of the temperature ramp must be carefully controlled. The resulting powders were stored in a dessicator.

The Leu+Glu/ $\text{SiO}_2$  samples were analyzed by different methods:

Thermogravimetric analyses (TGA) of freshly dried samples (not thermally activated) were recorded using a TA instrument with a STD Q600 analyzer. Analyses were performed with a heating rate of  $1^\circ\text{C}/\text{min}$  up to  $200^\circ\text{C}$ , under a  $100 \text{ mL}/\text{min}$  dry air flow. The use of air rather than Ar (as in the thermal activation procedure) is common practice

in the TG study of AA/ $\text{SiO}_2$  systems; it has generally been observed that the reactivity of  $\text{O}_2$  does not influence the AA polymerization reactions that are the focus of the present study (although  $\text{O}_2$  definitely influences the decomposition reactions at higher temperatures).

Infrared spectroscopic characterization was carried out on a Nicolet Magna-IR 550 spectrometer equipped with a MCT detector, with nominal  $4 \text{ cm}^{-1}$  resolution. Self-supporting pellets were prepared by pressing the sample in powder form under a pressure of  $8 \text{ tons}/\text{cm}^2$ . The pellets were then placed in the FTIR sample compartment under  $\text{N}_2$  flow. 128 scans were accumulated in the  $4000\text{--}400 \text{ cm}^{-1}$  spectral domain. IR spectra were normalized to the  $1870 \text{ cm}^{-1}$  silica band. This method was applied both before and after thermal treatment.

XRD analysis was used to confirm that amino acids are well dispersed on silica surface: we mostly checked that the diffractogram only showed the broad diffusion signal due to amorphous silica, without the sharp peaks that would indicate the precipitation of bulk amino acids crystals (see **Figure S1a**). Diffractograms were recorded on a Bruker D8 diffractometer using the Cu  $K\alpha$  radiation. An amorphous plastic sample holder was used. The  $2\theta$  range was set from  $10^\circ$  to  $90^\circ$ , with a number of steps of 3810 and an accumulation time of 1 s.

In order to carry out MS analyses, thermally activated samples were submitted to a desorption procedure. 10 mg of the solid samples were dispersed in 1 mL of MilliQ water, followed by centrifugation during 5 min at 6000 rpm. The study of three successive desorption steps showed that there are no significant differences between the compositions of the first and successive desorption solutions, only a global dilution. The analysis of complex polypeptide mixtures can hardly be accomplished *in situ* in the adsorbed phase, so that efficient desorption is a necessary requirement for quantitative analysis of the reaction products. In the related system Asp+Val/ $\text{SiO}_2$  system, also studied by Sakhno et al.<sup>39</sup> as already mentioned, some of the reaction products had a very high affinity for the surface and hardly desorbed at all, so that the composition of the desorption solution did not faithfully reflect that of the condensation products. The importance of peptide desorption was underlined, from another point of view, by de Duve and Miller.<sup>12</sup>

In order to better evidence the specificities of the silica surface scenario with respect to other AA polymerization methods, peptide condensation was also performed in solution using a sacrificial activator. More precisely, amide bond formation was carried out using carbonyldiimidazole (CDI). 1.5 mM L-Leu, 1.5 mM L-Glu and 10 mM KCl were dissolved in 10 mL of water. 6 mM carbonyldiimidazole was then added and the solution was incubated at room temperature for 24 h. The mechanism of CDI-induced polymerization of amino acids was described by Ehler.<sup>44</sup>

The products of both silica surface and CDI condensation scenarios were analyzed by mass spectrometry. Direct infusion mass spectrometry experiments were performed in the positive electrospray ionization mode on a QhFT-ICR 7 T instrument (Solarix<sup>XR</sup>, BrukerDaltonics, Germany) equipped with a dynamically harmonized cell. All mass spectra were acquired over a mass range of  $m/z$  100–2000. Each sample was diluted 50 times in 50:50 (v/v) water/acetonitrile (containing 0.1 % v/v of  $\text{CH}_3\text{COOH}$ ) to reach a total concentration of approximately  $5 \mu\text{g}\cdot\text{mL}^{-1}$  (supposing that the desorption step is quantitative). A preliminary study was carried out to determine the optimal sample dilution. A dilution factor of 50 allows sufficient peak intensities while avoiding space charge phenomena inside the trapping device, which may cause mass measurement shifts.<sup>45</sup> The resulting desorption solution was directly infused at a flow rate of  $180 \mu\text{L}\cdot\text{min}^{-1}$  into the ESI source. The following conditions were used: time of flight was set at 1.2 ms, ion accumulation time at 0.2 s, drying gas flow  $2 \text{ L}\cdot\text{min}^{-1}$  at  $200^\circ\text{C}$ , nebulizer gas pressure of 1.5 bar. A capillary voltage of 4.5 kV was applied. The method was calibrated using a solution of acetate and sodium formate clusters. MS/MS experiments were performed using CID activation in the hexapole collision cell in order to study the fragmentation patterns of condensation products i.e., synthesized oligopeptides. The isolation window of precursor ions was about 1  $m/z$  unit and the collisional energy was optimized depending on precursor ion relative intensity (6–12 V).

### 3- Results

#### 3-1 Evidences of peptide bond formation under thermal activation

First, direct analysis of solid Leu+Glu/ $\text{SiO}_2$  system was performed using thermogravimetric analysis (TGA) in order to observe water emission events that may be assigned to peptide bond formation under thermal activation conditions (**Figure S1b in supplementary information**). The TGA traces of Leu+Glu/ $\text{SiO}_2$  systems show two endothermic events at  $115^\circ\text{C}$  and  $148^\circ\text{C}$ . As previously demonstrated in studies of AA/ $\text{SiO}_2$  systems,<sup>21,22,46</sup> peaks in this range can be assigned to condensation reaction between amine and carboxylic acid functions, forming amide bonds and accompanied by  $\text{H}_2\text{O}$  loss. For most of amino acids, a single amide condensation event is observed in this region. The existence of two separate events is typical of systems containing some trifunctional amino acids such as glutamic acid, since in addition to peptide condensation they can also give rise to intramolecular condensation with formation of a lactam group;<sup>21,39</sup> in the case of glutamic acid, this leads to the formation of pyroglutamic residues. The TG results are the basis for our selection of an activation temperature of  $160^\circ\text{C}$ , since the water evolution events are completed at this temperature.

IR spectroscopy was used to follow peptide bond formation. **Figure 1** compares the IR spectra of Leu+Glu/ $\text{SiO}_2$  system before and after thermal activation, and finally after the desorption step applied for amino acid and oligomer extraction prior to MS analysis. The observed bands for the unactivated samples are in keeping with previously published data, assuming the existence on the surface of amino acids in the form of monomeric zwitterions (with the side chain  $\text{COOH}$  of Glu unionized). Band assignments are given in Table S1i. Changes occur between Leu+Glu/ $\text{SiO}_2$  IR spectra before and after thermal activation with the appearance of additional bands. The two bands at  $1667 \text{ cm}^{-1}$ , assigned to amide I (mostly due to  $\text{C}=\text{O}$  stretching vibration), and  $1520 \text{ cm}^{-1}$  (corresponding to amide II vibration) reveal peptide bond formation. Moreover, the presence of the amide II band suggests the formation of linear peptides,<sup>39</sup> since this band is not apparent in cyclic dipeptides (DKPs) that are the chief polymerization products in many AA/ $\text{SiO}_2$  systems.<sup>22</sup>

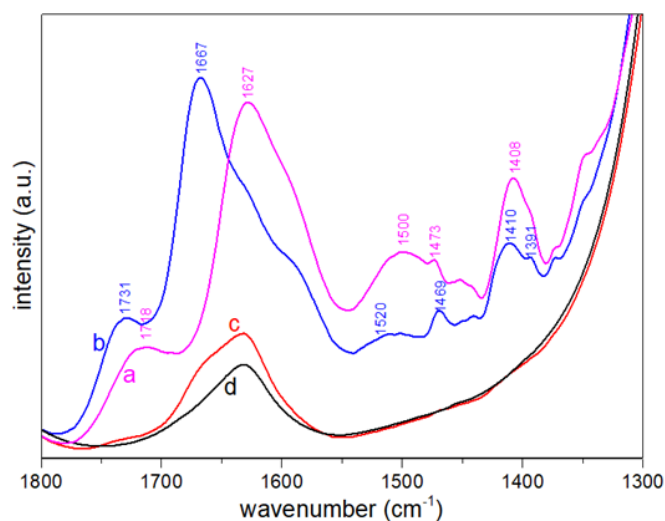


Figure 1. Normalized IR spectra in the  $1800\text{--}1300 \text{ cm}^{-1}$  range of Leu+Glu/ $\text{SiO}_2$  system (a) before and (b) after thermal activation, (c) of the remaining solid material after water desorption and (d) of the raw  $\text{SiO}_2$  support.

In addition, IR results gives an estimation of the desorption step efficiency.<sup>35</sup> The integrated intensity of the bands due to the organic molecules on the surface can indeed be estimated by subtracting the spectrum of raw silica support from that of the sample, after proper normalization (see experimental part). Applying this procedure to the IR spectra in Figure 1b and 1c reveals that the desorbed sample (spectrum c) retains less than 5% of the intensity of “organic bands” as compared to the same sample before desorption (spectrum b). Since thermal reaction products were almost quantitatively desorbed, their characterization could be undertaken by high resolution mass spectrometry (HRMS) analysis of desorption solutions. Based

on ppm mass accuracy measurements, oligopeptide-derived ions were identified, providing unambiguous molecular formulas of singly-charged oligopeptides.

We studied the condensation products of single amino acids on silica, as references for comparison with the binary system Leu+Glu/SiO<sub>2</sub>. The ESI mass spectra from desorption solution of thermally activated Leu/SiO<sub>2</sub> and Glu/SiO<sub>2</sub> are reported in **Figure S2** (supplementary information). There is little or no degradation to non-peptidic molecules and only little formation of long polymers. From Leu/SiO<sub>2</sub> analysis, the predominant peak, corresponding to [L<sub>2</sub>-H<sub>2</sub>O]<sup>+</sup> ions, can be assigned to a cyclic dimer cyclo(L-L), in other words, a substituted diketopiperazine (DKP), in accordance with previous work.<sup>37</sup> At first sight, smaller amounts of trimers and tetramers could be present, but relevant peaks could as well be assigned to non-covalent complexes.

As regards Glu/SiO<sub>2</sub>, no untransformed monomer is left after thermal activation but the product of internal condensation, PyroGlu, is detected. In addition, several Glu-containing oligopeptides seem to be present in accordance with previous work.<sup>21</sup> However, the same ambiguity as for Leu/SiO<sub>2</sub> is encountered since they could also be attributed to non-covalent complexes. In this case, we performed specific experiments to discriminate between these two possibilities (see **Figure S3** and **S4**), finding that non-covalent complex formation in the conditions of analysis accounts well for the observations. In summary, neither Glu/SiO<sub>2</sub> nor Leu/SiO<sub>2</sub> gives rise to the formation of long peptides.

Previous spectroscopic and TGA experiments<sup>19,21,22,47</sup> suggested that thermal activation of a single amino acid on a silica surface does not usually produce linear oligopeptides when deposition is carried out from an aqueous solution.

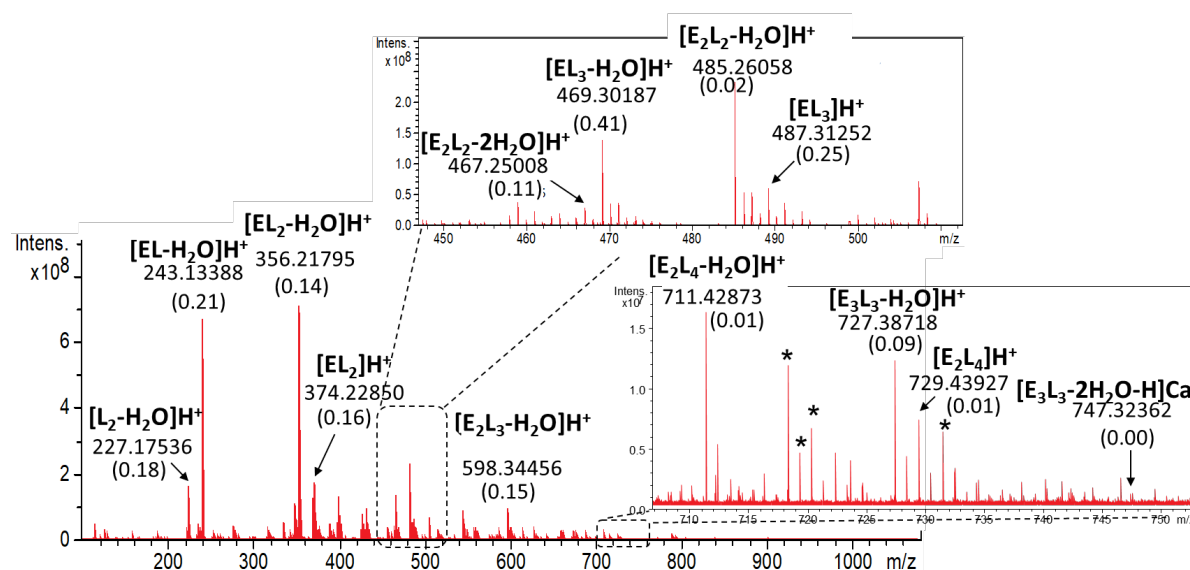


Figure 2. Positive ESI mass spectrum of the desorption solution obtained from thermally activated Glu+Leu/SiO<sub>2</sub> system using HRMS detection. The ppm accuracy of m/z value measurements are shown in brackets. Peaks marked with a star correspond to unidentified compounds.

The HRMS analysis of the desorption solution from activated Leu+Glu/SiO<sub>2</sub> shows a starkly different picture. **Figure 2** presents the whole ESI mass spectrum of the desorption solution from activated Leu+Glu/SiO<sub>2</sub>, with close-up views of the tetrapeptides and hexapeptides mass regions shown in insets. The vast majority of detected species may indeed be attributed to peptides. Abundant protonated and cationized ions from homo- and hetero-oligopeptides are detected, which proves the occurrence of oligopeptide formation (further confirmed by fragmentation patterns using MS/MS experiments, see below). Oligomer ions up to octamers have been detected with a mass accuracy better than 1 ppm. Thus, as previously reported,<sup>39</sup> the copolymerization of amino acid pair (E and L) on silica leads to extensive formation of long oligopeptides. On the contrary

this was not significantly achieved for either AA when deposited alone. The four predominant peaks were assigned to stoichiometries [EL-H<sub>2</sub>O]<sup>+</sup>, [EL<sub>2</sub>-H<sub>2</sub>O]<sup>+</sup>, [EL<sub>2</sub>]<sup>+</sup> and [E<sub>2</sub>L<sub>2</sub>-H<sub>2</sub>O]<sup>+</sup>, respectively. Other oligopeptide species including [L<sub>2</sub>-H<sub>2</sub>O]<sup>+</sup>, [E<sub>2</sub>L-H<sub>2</sub>O]<sup>+</sup> and [E<sub>2</sub>L<sub>3</sub>-H<sub>2</sub>O]<sup>+</sup> are also detected with quite high abundances. Interestingly, the most abundant oligopeptide species correspond to dehydrated forms (denoted as “-H<sub>2</sub>O”, i.e., -18.01056 uma with respect to the calculated formula for a linear peptide), with the exception of the EL<sub>2</sub> tripeptide.

In order to assess the reproducibility and check the possible effect of sample aging, several HRMS analyses were performed on solutions desorbed from activated Leu+Glu/SiO<sub>2</sub> samples after variable solid sample storage durations, either immediately after desorption or after different freezing-thawing cycles. Similar MS profiles, with

comparable relative intensities, were obtained (data not shown), indicating that the conditions of sample storage do not modify the composition of the oligopeptide mixture.

### 3.2 Polymerization selectivity: peptide stoichiometry

The main goal of our study was to investigate the possible existence of polymerization selectivity during peptide formation in the scenario of thermal activation on a silica surface. Caution is needed however, since discrimination effects exist in mass spectrometry, which strongly impacts the quantification of analytes from MS. These effects stem from phenomena occurring in different parts of the MS instrumentation. In addition to ionization discrimination effects (see following paragraph), mass discrimination phenomena exist in the FTICR mass analyzer device characterized by a pseudo potential well.<sup>43</sup> So, relative abundances of detected oligomer ions using a broad 100-1000  $m/z$  detection range cannot be used for a quantitative evaluation of all oligomers: only comparisons within a limited  $m/z$  range can be meaningful.

The first aspect of the selectivity question is whether some stoichiometries are preferred among polypeptides of a given length. Taking tripeptides as an example, the possible stoichiometries are  $L_3$ ,  $EL_2$ ,  $E_2L$  and  $E_3$ . Assuming a random polymerization, with two amino acids present in equimolar amounts, there would be a probability of 0.5 for each position in the chain to be occupied by Leu and Glu, respectively, so that the respective molar ratios of these different peptides would be 1:3:3:1 (the binomial coefficients), or 1/3 : 1 : 1 : 1/3, normalizing the highest amounts to 1. **Figure 3** compares the theoretically expected tripeptide abundances to the experimentally observed peak intensities, for our reference sample (L:E molar ratio of 1:1). It must be underlined that to obtain the experimental values, the peak intensities of all the oligopeptide ions observed for a given stoichiometry (protonated or cationized species, from dehydrated or fully hydrated forms) have been summed. Average values from seven repeated analyses are plotted together with error bars corresponding to standard deviations.

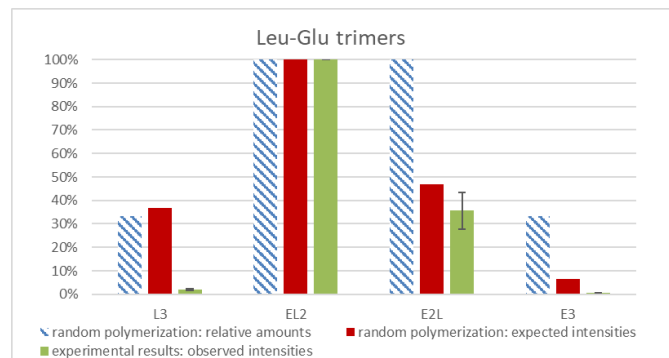


Figure 3. Comparison of the theoretically expected peak intensities for tripeptides in a random polymerization to the experimentally observed peak intensities. Relative amounts in the random scenario are corrected for ionization efficiencies to give “expected intensities”. All the distributions are normalized with respect to  $EL_2$ , arbitrarily set at 1.

At first sight, the experimental data would seem to be strongly different from the expected amounts for random polymerization, suggesting a definite selectivity. However, discrimination effects arise at the ionization step, particularly when using atmospheric pressure ionization (API) methods.<sup>48</sup> Variable ionization efficiencies are expected for different oligopeptide sequences. It is not straightforward to evaluate these MS discrimination effects quantitatively; typically, quantitative studies involve the use of reference labeled compounds and developed methods are generally compound- and compound family dependent.<sup>49</sup> To estimate such phenomena in the present work, reference tripeptides were used. ESI mass spectra of equimolar mixtures containing three reference tripeptides ( $LLL$  and  $EEE$  together with one of the six other hetero-tripeptides) were analyzed using the same procedure; the results are reported in **Figure S5**. In general, leucine-rich peptides display a higher ionization efficiency; in the most extreme case, the peak intensity of  $[L_3]H^+$  is higher than that of  $[E_3]H^+$  by almost an order of magnitude. This was expected, as analytes with hydrophobic moieties should have a higher ESI response than more polar ones due to their high affinity for electrospray droplet surfaces.<sup>48</sup> They tend to have a greater charge excess which may explain their greater ionization yields. Thus, the ionization efficiencies follow a relative order that depends on the global stoichiometry:  $L_3 > EL_2 > E_2L > E_3$ . Tripeptides of the same stoichiometry but with different sequences exhibit close MS response. Based on these results, we calculated an average MS response factor for the two stoichiometries having several possible sequences (i.e.  $EL_2$  and  $E_2L$ ), and used them to evaluate the expected peak intensities that would correspond to random polymerization (**Figure 3**). These calculations remain approximate because they do not take into account possible differences in respective ionization efficiency between protonated and cationized species, nor between fully hydrated and dehydrated peptides from the same peptide sequence, but they should represent an improvement over raw peak intensities.

The discrepancy between (corrected) random and experimental distributions is reduced, but not eliminated, as regards the relative amounts of peptides with  $EL_2$  and  $E_2L$  stoichiometries – there is about 90% chance that the predominance of  $EL_2$  over  $E_2L$  is significant. In contrast, the quasi absence of species with stoichiometry  $L_3$  is made more conspicuous, because they have a high response factor and their signal should be exalted if they were present.  $E_3$  species also appear definitely under-represented. Comparisons between random model and experimental data

are reported for other peptide lengths, and also for other E/L molar ratios in **Figures S6 to S8**. The lack of suitable reference peptides for chains longer than tripeptides did not permit even the rough quantifications that we attempted for tripeptides – not to mention the exponentially more complicated difficulty of the task when peptide length increases. Nevertheless, experimental distributions do seem to be significantly different from random ones. Especially, homopolymers seem to be very scarce. The verification of this observation was one of the reasons to investigate AA mixtures of different stoichiometries (see below). Note that the latter results are somewhat different from previous ESI results on the same system, where EL<sub>4</sub> and EL<sub>5</sub> species were the only detected stoichiometries for pentamers and hexamers.<sup>39</sup> In the latter study, however, less care was taken to control the residual pressure of water during thermal treatment: the effect of water activity on the polymerization process is a direction for further research.

### 3.3 Polymerization selectivity: peptide sequence determination

The next question that must be raised relates to the existence of sequence selectivity for a given peptide stoichiometry. Peptides with the same AA composition but different sequences give peaks at the same *m/z* values in MS, but tandem mass spectrometry (MS/MS) should provide access to a unique peptide sequence through the detection of sequence-specific series of fragment ions, and more specifically the so-called *b/y* fragments.<sup>50,51</sup> The protonated molecular ion for a given peptide stoichiometry is selected as precursor ion and undergoes CID (Collision Induced Dissociation) fragmentations. **Figure 4** presents the CID spectra of the protonated ELL, LEL and LLE reference tripeptides and the one of the peak detected experimentally at *m/z* 374.228 from the desorbed solution from thermally activated Leu+Glu/SiO<sub>2</sub>. The MS/MS spectrum of the precursor ion from the desorption solution displays only two major product ions at *m/z* 243.13348 (assigned to a *b*<sub>2</sub> fragment) and at 132.10175 (assigned to a *y*<sub>1</sub> fragment). This fragmentation pattern suggests the predominant presence of the ELL sequence among tripeptides with the EL<sub>2</sub> stoichiometry, thus demonstrating a selectivity in the sequence of the oligopeptides formed.

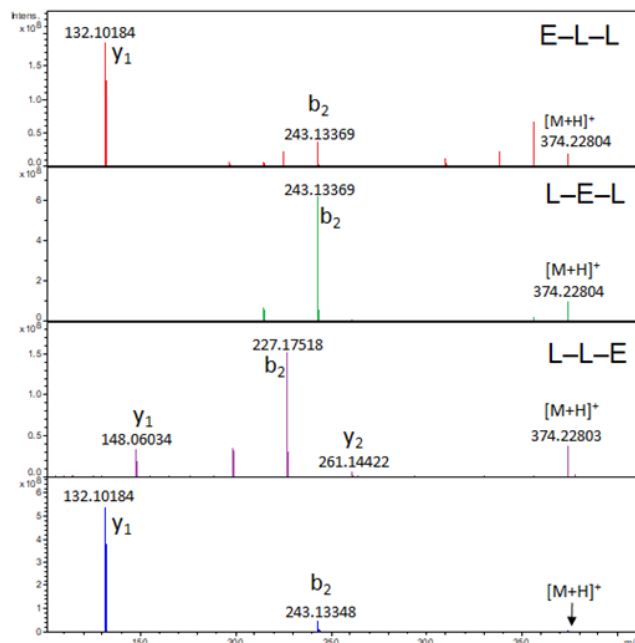


Figure 4. MS/MS spectra of the protonated molecule  $[M+H]^+$  at *m/z* 374.2280 for reference tripeptides ELL, LEL and LLE sequences, and for the peak at *m/z* 374.2280 detected from the thermally activated Leu+Glu/SiO<sub>2</sub> system (bottom spectrum).

MS/MS experiments were also performed for other detected linear and dehydrated oligopeptides ions (see **Figure S9 in SI**). Conclusive interpretations are hard to make because of the lack of reference peptides to obtain characteristic fragmentation patterns of genuine compounds, and because of the very similar fragmentation pathways for alternating Leu/Glu heteropeptides, but especially in the case of tetrapeptides there are definite hints of sequence selectivity. To go further, a MS<sup>n</sup> study is under investigation.

### 3.4 Effect of AA molar ratio on polymerization

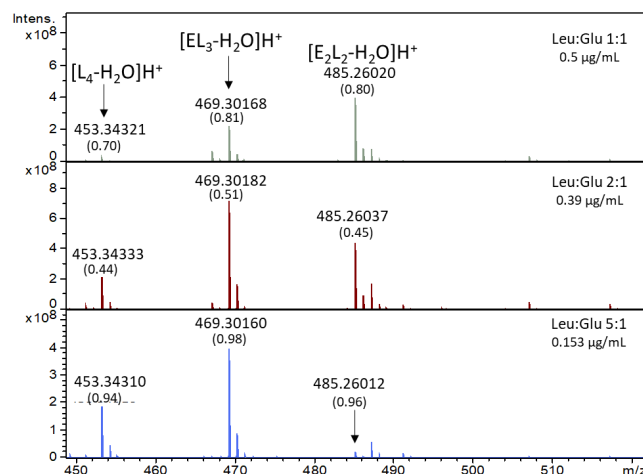


Figure 5. Positive ESI mass spectra in the tetrapeptide ( $m/z$  450-520) mass range of thermally activated Leu+Glu/SiO<sub>2</sub> systems prepared with different Leu:Glu ratios. Note that the sample concentrations were optimized to obtain similar MS response and avoid saturation.

The effect of molar ratio between two amino acids on the oligomer production was studied using the same HRMS conditions. **Figure 5** shows a zoomed  $m/z$  range corresponding to the tetrapeptide ions for the three different Leu/Glu ratios investigated. Comparisons of theoretically expected tripeptide abundances to the experimentally observed peak intensities for the different oligomer lengths and three different L/E molar ratios are reported in **Figure S6 to S8**. As could be expected, leucine-rich peptides are formed in higher amounts when starting from mixtures containing higher proportion of leucine. For example, the  $[L_4-H_2O]^+H^+$  ion relative intensity ( $m/z$  453.34353) increases for the 2:1 and 5:1 ratios. Similarly, the  $[E_2L_2-H_2O]^+H^+$  ion relative intensity, ( $m/z$  485.26059) as compared to the  $[EL_3-H_2O]^+H^+$  ion decreases, whereas the  $[E_4-H_2O]^+H^+$  peptide is never detected. However, some polymerization selectivity is still observed. In particular, heteropeptide ions are always predominant, even when an excess of Leu should favor the  $(Leu)_n$  homopolymers if the polymerization were random. Regarding the tripeptides, for which a semi-quantitative comparison is possible, the trends discussed above for the 1:1 ratio are confirmed: in addition to the very small amounts of  $L_3$  even for L:E ratio

of 5:1, the predominance of  $EL_2$  over  $E_2L$  is confirmed (**Figure S6 in SI**), and indeed more conspicuous than for the 1:1 ratio. If we look at the data for the three ratios, we see that the following oligomers seem overrepresented, with respect to the random scenario:  $EL_3$  and  $E_2L_2$  for the tetramers,  $E_2L_3$  for the pentamers,  $E_2L_4$  for the hexamers,  $E_3L_5$  and  $E_3L_4$  for the heptamers, and  $E_3L_5$  for the octamers (**Figures S6 to S8**).

### 3.5 Origin of polymerization selectivity: comparison with another condensation method

The observed selectivity could have a thermodynamic origin, meaning that some oligomers are formed in higher amounts because they are thermodynamically more stable, or a kinetic origin, meaning that the particular reaction mechanism on the silica surface specifically opens certain transformation pathways by lowering their activation energy (heterogeneous catalysis). In order to shed some light on this question, a homogeneous condensation procedure using a sacrificial activator i.e., carbonyldiimidazole (CDI), was applied to an equimolar Leu+Glu mixture. The ESI mass spectrum of the condensation products is reported in **Figure 6**. Oligopeptides up to pentamers are detected, but the results diverge from those obtained from surface polymerization, in two important respects.

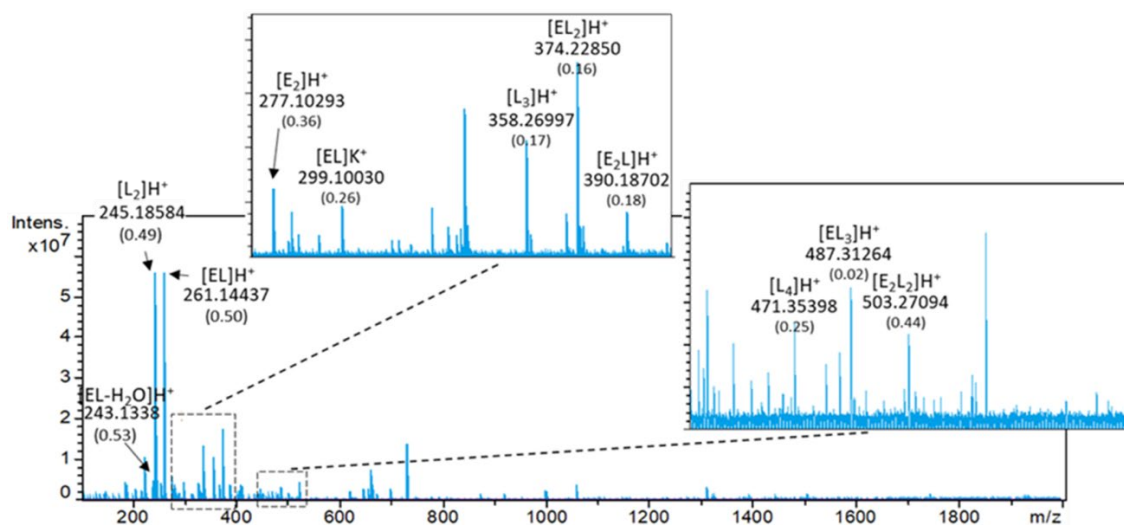


Figure 6. Positive ESI mass spectrum of the condensation products obtained in solution using carbonyldiimidazole as a sacrificial activator.

First, little to no stoichiometry selectivity is obvious here. Indeed, the experimentally observed distributions are rather close to the random oligomerization once the relative ionization yields between Leu- and Glu-rich peptides are taken into account (see the di-, tri- and tetra-peptide mass ranges in **Figure 6**), strongly suggesting that the

CDI method is less selective and closer to random polymer formation. These observations underline the key role of the silica surface in polymerization selectivity. As a matter of fact, they do not definitely prove that the selectivity observed in the silica surface scenario has a kinetic origin, since it may be argued that the various oligomers have different stabilities on the silica surface due to different

amino-acid/surface interaction energies. We will examine this question in more detail in a forthcoming publication dealing with reaction mechanisms and kinetics. Suffice it to say here that the peptide distributions discussed above may not correspond to an equilibrium situation, and that they vary significantly with the duration of the thermal treatment.

Second, no dehydrated peptide ions are detected when using the CDI procedure, even for Glu-containing oligopeptides. The absence of dehydrated peptidic forms when using a sacrificial activator in contrast to surface activation is of major interest.

#### 4- Discussion

Let us recapitulate the different aspects of condensation reaction selectivity that are relevant for the origin-of-life significance of amino acid polymerization on mineral surfaces.

First, there is the overall yield of peptide formation. From a thermodynamic point of view, the amino acid condensation equilibrium can be displaced as much to the right as desired, for all practical purposes close to 100%, simply by decreasing the water activity (harsh drying conditions). From a kinetic point of view, many early studies<sup>52</sup> were carried out in temperature conditions where the reaction is rather slow and takes weeks to be completed. Although the yields reported in those low T studies may be only a few percent, TG results show that quantitative condensation can be achieved in a matter of minutes at 160°C – although the peptide bonds will keep rearranging with time. Then, HRMS analysis shows that most observable peaks are attributable to singly charged oligopeptides. Actually, from previous studies, we know that various peptide degradation reactions may be induced by heating at higher temperatures, but on the silica surface they do not become significant before at least 250°C.<sup>19</sup> Therefore, amino acid condensation on mineral surfaces is a robust scenario, where close to 100% selectivity to the “prebiotically interesting” compounds is achieved and maintained over a large range of experimental conditions. This deserves to be underlined since many reported prebiotic pathways do form molecules of biological significance, but only in a few % yields and/or in conditions where they are quickly degraded.

Second, there is the question of selectivity for rather long oligopeptides over cyclic dipeptides (DKPs). This point was mentioned in the introduction, as one of the reasons for choosing the (Leu+Glu) AA pair, in view of previous work<sup>39</sup> where this system can indeed display this type of selectivity. We have confirmed that peptides containing three or more amino acid residues are indeed predominant, while neither single AA system (Leu/SiO<sub>2</sub> or Glu/SiO<sub>2</sub>) forms

large amounts of them. Actually, it had been shown by Fox *et al*<sup>9</sup> that mixed systems containing Glu could form long polypeptide chains, although this was for a bulk polymerization scenario (no surface involved) that is less likely prebiotically. Later on, the same authors established that the necessary condition for making long polypeptides in their scenario was the presence in sufficient amounts of tri-functional amino acids having either a –COOH (Glu, Asp) or a –NH<sub>2</sub> group (Lys) in their side chain.<sup>24,25</sup> In the present study, the most abundant oligopeptide ions were tri- and tetramers, with significant presence of higher polymers up to the octamers. This may seem small in comparison to the bulk polymerization scenario where peptides containing up to 100 residues were present. However, the length of chains formed on surfaces may increase with reaction time, as will be studied in a forthcoming publication. Here, we may already underline the fact that the observed trends imply a significant mobility of the AAs and peptides on the surface, since the 2-D AA density was only 1/10 of a physical monolayer: in other words, if the AAs are homogeneously distributed, they are not in close contact with their neighbors.

Third, and most central to the usefulness for origin of life scenarios, comes the question of stoichiometry and sequence selectivity. The interpretation of experimental data is not straightforward due to the well-known difficulty of quantifying MS results. We have attempted to evaluate the ionization yields through the analysis of reference tripeptides. This process is quite cumbersome and semi-quantitative at best. However, the example of tripeptides, which was worked out in detail, as well as the less detailed analysis of other formed polypeptides, definitely indicates that the mineral surface scenario is selective, in the sense that some reaction products are favored at the expense of others, most probably due to the selective opening of some reaction pathways by catalytic effects. The preferential formation of heteropeptides rather than homopeptides is a general trend; the preference of EL<sub>2</sub> over E<sub>2</sub>L stoichiometry is less marked but still seems significant. Going from stoichiometry to sequences proper is even harder as it requires MS/MS experiments, but can be quite rewarding, showing for instance that the EL<sub>2</sub> stoichiometry predominantly corresponds to the ELL sequence, rather than the other two theoretically possible ones. Partial data on other oligopeptides also point at significant sequence selectivities in general.

The definition of selectivity generally used in heterogeneous/industrial catalysis, namely the amount of starting reagent transformed to the product of interest, expressed as a percentage of total reagent transformation, is insufficient to evaluate the prebiotic interest of this, or any other polymerization scenario. Instead, as a first step towards evaluating to what degree a particular scenario can preferentially lead to a particular peptide, or a particular category

of peptides, one would like to have a quantitative measure of how much the observed distribution deviates from a “non-discriminating one”. The problem is made more complex because for a complete treatment, we would need to address both the distribution among various polypeptide lengths, and, for a given length, the distribution among possible sequences/stoichiometries. Let us consider the second question, and more specifically, let us take as a specific example the distribution among possible tripeptides, for which we have the most experimental data.

It turns out that this problem is akin to one that was often addressed in theoretical biochemistry, namely, the assessment of genetic heterogeneity among a population of organisms. Among other existing practical measures, Adami<sup>33,54</sup> proposed to evaluate the average physical complexity of an ensemble of genetic sequences  $X$ , according to a formula based on Shannon’s entropy in information theory.<sup>55</sup> In their case,  $X$  was an ensemble of known sequences, i.e. of DNA chains with potentially very high lengths ( $L$ , the length of the sequence, can reach up to 2 million bases even for a single gene). In our case,  $X$  would be an ensemble of oligopeptides, e.g. of tripeptides, with  $L = 3$ . The entropy  $H^*$  is calculated along

$$H(X) = - \sum_{i=1} p_i \log p_i$$

where the summation is over all possible sequences  $i$  in ensemble  $X$ , and  $p_i$  are their respective probabilities. The number of possible sequences depends on the sequences length and on the number of monomer types,  $D$  (the “size of the alphabet”). In the case of DNA sequences,  $D = 4$  (there are 4 possible nucleobases), while in our case  $D = 2$  (the monomers can be either L or E). The number of possible sequences is then  $N = D^L$ ; in our “world” of E+L polymerization,  $N = 2^3 = 8$  (LLL, LLE ... EEE).

Now if all sequences were equally likely, the probability of each one would be  $1/N$ , so that:

$$H_{max}(X) = - \sum_{i=1}^N \frac{1}{N} \log \left( \frac{1}{N} \right) = \log N,$$

providing a maximum entropy situation. Genomes in a real environment,  $E$ , are submitted to (natural) selection

pressures so that the observed sequence probabilities  $p_i$  are greatly different from this equal probability situation. The conditional entropy of the ensemble of sequences  $X$  in environment  $E$  will then be different from  $H_{max}$ , and denoted  $H(X|E)$ .  $H(X|E)$  is indeed lower than  $H_{max}$  because the sequence distribution in the real environment is “more ordered” – it is easy to realize that in the limiting case where only one sequence would be expressed in the population,  $H(X|E)$  would be zero. The difference  $H_{max} - H(X|E)$  is denoted  $I(X : E)$ , and may be considered as an information content. Adami indeed equates it to “the information that  $X$  has about  $E$ ”, the information that the population of genomes has about its environment.

It is easy to apply these notions to the world of oligopeptide synthesis. In our case,  $p_i$  may be equated to the relative abundances of the trimers in the analytical mixture (estimated by correcting the peak intensities for ionization efficiencies as explained above), and  $E$  will be the “silica surface environment”, or in more common parlance the silica surface AA polymerization scenario, so we will write  $SiO_2$  instead of  $E$ .  $I(X : SiO_2)$  will then constitute a quantitative evaluation of the “non-randomness” of the observed oligopeptide distribution, due to the fact that it is established in the  $SiO_2$  scenario. This extension would be rather straightforward if we had precise data on the experimental abundances of each trimeric sequence. Establishing oligopeptide sequences requires a lot of experimental work (in the results discussed above, the predominant sequence was unequivocally established only for the  $EL_2$  stoichiometry), so that we will couch the calculation in terms of experimentally observed stoichiometries instead. Then, as already stated, the maximum entropy (= random scenario) will correspond to relative intensities 1:3:3:1, which may be converted to probabilities by normalizing to 1. The steps of the calculation of  $I(X : SiO_2)$  are shown in **Table 1**, and this quantity is found equal to  $1.256 - 0.788 = 0.468$  entropy units. This approach may be used to compare the two polymerization scenarios, on silica and in solution with the CDI activator. It is found that  $I(X : CDI)$ , it is only 0.222 entropy units.

**Table 1.** Calculation of the information content of the experimental tripeptide distribution in the silica surface and the CDI activator scenarios, as compared to a random distribution. Natural logarithms were used; conversion to base 2 log may be obtained multiplying by 1.4427.

---

\* The choice of this notation for the entropy is unfortunate since it may bring about a confusion with the enthalpy in general thermodynamics, but it is already well-established.

Stoichiometry $i$	$L_3$	$EL_2$	$E_2L$	$E_3$	
$p_i$ (random)	0.125	0.375	0.375	0.125	
$p_i \log(p_i)$ (random)	-0.260	-0.368	-0.368	-0.260	$H_{max}(X) = -\sum_{i=1} p_i \log p_i$ (rnd) = 1.256
$p_i$ (exp, $SiO_2$ )	0.010	0.570	0.407	0.013	
$p_i \log(p_i)$ (exp, $SiO_2$ )	-0.045	-0.321	-0.366	-0.057	$H(X SiO_2) =$ $-\sum_{i=1} p_i \log p_i$ (obs) = 0.788
$p_i$ (exp, CDI)	0.300	0.495	0.205	0.000	
$p_i \log(p_i)$ (exp, CDI)	-0.361	-0.348	-0.325	0.000	$H(X CDI) =$ $-\sum_{i=1} p_i \log p_i$ (obs) = 1.034

This example shows that it is possible to quantify the information content of an experimentally observed distribution of polymers obtained in a particular polymerization scenario. A generalization accounting also for the distribution between different polymer lengths could be devised, although practical measurements involve some experimental difficulties. The quantity  $I(X : \text{SiO}_2)$  calculated in this way has tantalizing implications in research on the origins of life. It was originally defined for populations of genomes, and in this context, it may be intuitively understood as the “amount of information that a population stores about the environment in which it lives”.<sup>54</sup> Nobody would of course contend that the mixed oligopeptide systems formed by surface AA polymerization are alive – to the best of our knowledge, they do not do anything useful from a systems chemistry perspective, while genomes are functional macromolecules. Still, it may be interesting to have a common measure of complexity that can be applied at such widely separated steps on the road from chemistry to life.

The course we have followed in the preceding developments is to evaluate the non-random character of a polypeptide distribution in order to determine whether it is promising in an OoL scenario. Thus, we took the polypeptide distribution as a piece of data to be interpreted, without trying to establish how exactly it was produced. The latter question is of course a legitimate one. Specialists of surface science and catalysis may think it is the central one, and indeed unraveling the surface reaction mechanisms would be very helpful to determine how polymerization selectivity can be improved, other than by blind variation of the reaction parameters. This endeavor is not without precedent in the literature, in fact essentially for one system similar to ours.

Dose *et al* have studied the copolymerization of Glu with several AAs with hydrophobic residues (thus similar to Leu), denoted X, both in the bulk polymerization<sup>27</sup> and carbodiimide<sup>56</sup> (a solution scenario with sacrificial activator, akin to our CDI) scenarios. They found that the trimers formed mostly consisted in PyroGlu-X-X (pE-X-X), similar to the predominance of E-L-L that we observed. They explained this by hypothesizing that Glu first undergoes internal cyclization to pyroglutamic acid (“self-protecting” the NH<sub>2</sub> terminal of Glu) while the amino acids X condense to cyclic dimers, followed by the reaction  $\text{pE} + \text{cyclo}(\text{X-X}) = \text{pE-L-L}$ . It is tempting to transpose their mechanism to our systems; it would be compatible with the observed reactivity of the single amino acid systems, and with the existence of large amounts of dehydrated peptides, in our experiments. However, our initial results (data not shown) indicate that this mechanism probably does not take place in our case, and we will leave a full mechanistic discussion for a further publication.

## 5- Conclusion

In this study, close to quantitative formation of oligopeptides up to octamers has been demonstrated for Leu+Glu/SiO<sub>2</sub> systems after a single thermal activation step, a realistic scenario in origins of life studies. Unambiguous peptide assignments with ppm mass accuracy measurements were obtained by using high resolution mass spectrometry (HRMS) detection. These, and MS/MS experiments, demonstrate non-trivial distributions of oligomers between different stoichiometries and possible amino acid sequences, which probably have a kinetic origin – a memory of heterogeneous catalytic pathways opened up on the inorganic surface. Polymerization in aqueous solution through the use of sacrificial activators does not result in the same oligopeptide distributions, being closer to random. In order to quantify the information contained in experimentally observed distributions, we propose to use a metric derived from genomics analysis. However, it should not be overinterpreted. It provides an interesting measure to evaluate the otherwise ill-defined notion of selectivity in the progressive emergence of biopolymers. Further investigations will be performed into the polymerization process on surfaces, especially to follow polymerization kinetics and thus establish the likely mechanism, as well as investigations of the polymerization selectivity with other characterization methods (hyphenated methods as Liquid Chromatography/Mass Spectrometry (LC/MS) or Ion Mobility-Mass Spectrometry (IMMS) to identify peptide sequences). More complex amino acid mixtures will also be investigated in order to get closer to the likely complexity of prebiotic environments.

## ASSOCIATED CONTENT

Supporting information: XRD and DTA characterizations, additional ESI and MS/MS spectra, peptide distribution histograms, IR bands assignments

This material is available free of charge at ...

## AUTHOR INFORMATION

### Corresponding Authors

Sandra Alves, Institut Parisien de Chimie Moléculaire (UMR 8232 CNRS), Sorbonne Université, Paris, France orcid.org/0000-0003-0063-8760

Email : [sandra.alves@sorbonne-universite.fr](mailto:sandra.alves@sorbonne-universite.fr)

Jean-François Lambert, Laboratoire de Réactivité de Surface (UMR 7197 CNRS), Sorbonne Université, Paris, France orcid.org/0000-0002-8124-5709

Email : [jean-francois.lambert@sorbonne-universite.fr](mailto:jean-francois.lambert@sorbonne-universite.fr)

### Authors

Lise Bedoin, Laboratoire de Réactivité de Surface (UMR 7197 CNRS) and Institut Parisien de Chimie Moléculaire (UMR 8232 CNRS), Sorbonne Université, Paris, France.

Email : [lise.bedoin@sorbonne-universite.fr](mailto:lise.bedoin@sorbonne-universite.fr)

### Author Contributions

The manuscript was written through contributions of all authors.

## ACKNOWLEDGMENT

This program was supported by a Ph.D. grant from the Paris Ile-de-France region awarded in the frame of the DIM program ACAV+ under the project name EmBioPol. A financial contribution of the CNRS MIT programme Origines 2019 is acknowledged. The National FT-ICR network (FR 3624 CNRS) providing FT-ICR facilities for conducting the research is also gratefully acknowledged.

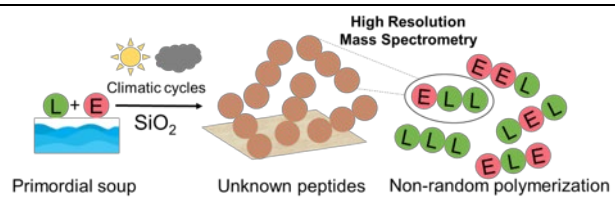
## REFERENCES

- (1) Miller, S. L. A Production of Amino Acids under Possible Primitive Earth Conditions. *Sci. New Ser.* **1953**, *117* (3046), 528–529.
- (2) Parker, E. T.; Cleaves, H. J.; Dworkin, J. P.; Glavin, D. P.; Callahan, M.; Aubrey, A.; Lazcano, A.; Bada, J. L. Primordial Synthesis of Amines and Amino Acids in a 1958 Miller H<sub>2</sub>S-Rich Spark Discharge Experiment. *Proc. Natl. Acad. Sci.* **2011**, *108* (14), 5526–5531. <https://doi.org/10.1073/pnas.1019191108>.
- (3) Meierhenrich, U. J.; Caro, G. M. M.; Bredehöft, J. H.; Jessberger, E. K.; Thiemann, W. H.-P. Identification of Diamino Acids in the Murchison Meteorite. *Proc. Natl. Acad. Sci.* **2004**, *101* (25), 9182–9186. <https://doi.org/10.1073/pnas.0403043101>.
- (4) Brack, A. From Interstellar Amino Acids to Prebiotic Catalytic Peptides: A Review. *Chem. Biodivers.* **2007**, *4* (4), 665–679. <https://doi.org/10.1002/cbdv.200790057>.
- (5) Maury, C. P. J. Origin of Life. Primordial Genetics: Information Transfer in a Pre-RNA World Based on Self-Replicating Beta-Sheet Amyloid Conformers. *J. Theor. Biol.* **2015**, *382*, 292–297. <https://doi.org/10.1016/j.jtbi.2015.07.008>.
- (6) Issac, R.; Chmielewski, J. Approaching Exponential Growth with a Self-Replicating Peptide. *J. Am. Chem. Soc.* **2002**, *124* (24), 6808–6809. <https://doi.org/10.1021/ja026024i>.
- (7) Gorlero, M.; Wieczorek, R.; Adamala, K.; Giorgi, A.; Schininà, M. E.; Stano, P.; Luisi, P. L. Ser-His Catalyses the Formation of Peptides and PNAs. *FEBS Lett.* **2009**, *583* (1), 153–156. <https://doi.org/10.1016/j.febslet.2008.11.052>.
- (8) Danger, G.; Boiteau, L.; Cottet, H.; Pascal, R. The Peptide Formation Mediated by Cyanate Revisited. N-Carboxyanhydrides as Accessible Intermediates in the Decomposition of N-Carbamoylamino Acids. *J. Am. Chem. Soc.* **2006**, *128* (23), 7412–7413. <https://doi.org/10.1021/ja061339+>.
- (9) Harada, K.; Fox, S. W. The Thermal Condensation of Glutamic Acid and Glycine to Linear Peptides<sup>1</sup>. *J. Am. Chem. Soc.* **1958**, *80* (11), 2694–2697. <https://doi.org/10.1021/ja01544a027>.
- (10) Rodríguez-García, M.; Surman, A. J.; Cooper, G. J. T.; Suárez-Marina, I.; Hosni, Z.; Lee, M. P.; Cronin, L. Formation of Oligopeptides in High Yield under Simple Programmable Conditions. *Nat. Comm.* **2015**, *6* (1), Art. #8385. <https://doi.org/10.1038/ncomms9385>.
- (11) Bernal, J. D. *The Physical Basis of Life*; Routledge and Kegan Paul: Londres, 1951.
- (12) de Duve, C.; Miller, S. L. Two-Dimensional Life? *Proc. Natl. Acad. Sci.* **1991**, *88* (22), 10014–10017. <https://doi.org/10.1073/pnas.88.22.10014>.
- (13) Smith, J. V. Biochemical Evolution. I. Polymerization on Internal, Organophilic Silica Surfaces of Dealuminated Zeolites and Feldspars. *Proc. Natl. Acad. Sci.* **1998**, *95* (7), 3370–3375. <https://doi.org/10.1073/pnas.95.7.3370>.
- (14) Bujdák, J.; Rode, B. M. Silica, Alumina, and Clay-Catalyzed Alanine Peptide Bond Formation. *J. Mol. Evol.* **1997**, *45* (5), 457–466. <https://doi.org/10.1007/PL00006250>.
- (15) Lahav, N.; Chang, S. The Possible Role of Solid Surface Area in Condensation Reactions during Chemical Evolution: Reevaluation. *J. Mol. Evol.* **1976**, *8* (4), 357–380. <https://doi.org/10.1007/BF01739261>.
- (16) Bujdák, J.; Slosiarikova, H.; Texler, N.; Schwendinger, M.; Rode, B. M. On the Possible Role of Montmorillonites in Prebiotic Peptide Formation. *Monatshefte Für Chem. Chem. Mon.* **1994**, *125* (10), 1033–1039. <https://doi.org/10.1007/BF00811510>.
- (17) Basiuk, V. A.; Gromovoy, T. Y.; Glukhoy, A. M.; Golovaty, V. G. Chemical Transformations of Proteinogenic Amino Acids during Their Sublimation in the Presence of Silica. *Orig. Life Evol. Biosph.* **1991**, *21* (3), 129–144. <https://doi.org/10.1007/BF01809442>.
- (18) Guo, C.; Holland, G. P. Alanine Adsorption and Thermal Condensation at the Interface of Fumed Silica Nanoparticles: A Solid-State NMR Investigation. *J. Phys. Chem. C* **2015**, *119* (45), 25663–25672. <https://doi.org/10.1021/acs.jpcc.5b10236>.
- (19) Lambert, J.-F.; Jaber, M.; Georgelin, T.; Stievano, L. A Comparative Study of the Catalysis of Peptide Bond Formation by Oxide Surfaces. *Phys. Chem. Chem. Phys.* **2013**, *15* (32), 13371–13380. <https://doi.org/10.1039/c3cp51282g>.
- (20) Georgelin, T.; Jaber, M.; Bazzi, H.; Lambert, J.-F. Formation of Activated Biomolecules by Condensation on Mineral Surfaces – A Comparison of Peptide Bond Formation and Phosphate Condensation. *Orig. Life Evol. Biospheres* **2013**, *43* (4–5), 429–443. <https://doi.org/10.1007/s1084-013-9345-2>.
- (21) Bouchoucha, M.; Jaber, M.; Onfroy, T.; Lambert, J.-F.; Xue, B. Glutamic Acid Adsorption and Transformations on Silica. *J. Phys. Chem. C* **2011**, *115* (44), 21813–21825. <https://doi.org/10.1021/jp206967b>.
- (22) Meng, M.; Stievano, L.; Lambert, J.-F. Adsorption and Thermal Condensation Mechanisms of Amino Acids on Oxide Supports. 1. Glycine on Silica. *Langmuir* **2004**, *20* (3), 914–923. <https://doi.org/10.1021/la035336b>.
- (23) Miyakawa, S.; Ferris, J. P. Sequence- and Regioselectivity in the Montmorillonite-Catalyzed Synthesis of RNA. *J. Am. Chem. Soc.* **2003**, *125* (27), 8202–8208. <https://doi.org/10.1021/ja034328e>.
- (24) Fox, S. W.; Harada, K. The Thermal Copolymerization of Amino Acids Common to Protein<sup>1</sup>. *J. Am. Chem. Soc.* **1960**, *82* (14), 3745–3751. <https://doi.org/10.1021/ja01499a069>.
- (25) Fox, S. W.; Harada, K. Thermal Copolymerization of Amino Acids in the Presence of Phosphoric Acid. *Arch. Biochem. Biophys.* **1960**, *86*, 281–285.
- (26) Nakashima, T.; Jungck, J. R.; Fox, S. W.; Lederer, E.; Das, B. C. A Test for Randomness in Peptides Isolated from a Thermal Polyamino Acid. *Int. J. Quantum Chem.* **1977**, *12* (S4), 65–72. <https://doi.org/10.1002/qua.560120708>.
- (27) Hartmann, J.; Christel Brand, M.; Dose, K. Formation of Specific Amino Acid Sequences during Thermal Polymerization of Amino Acids. *Biosystems* **1981**, *13* (3), 141–147. [https://doi.org/10.1016/0303-2647\(81\)90055-1](https://doi.org/10.1016/0303-2647(81)90055-1).
- (28) Mosqueira, F. G.; Ramos-Bernal, S.; Negrón-Mendoza, A. A Simple Model of the Thermal Prebiotic Oligomerization of Amino Acids. *Biosystems* **2000**, *57* (2), 67–73. [https://doi.org/10.1016/S0303-2647\(00\)00089-7](https://doi.org/10.1016/S0303-2647(00)00089-7).
- (29) Bujdák, J.; Rode, B. M. Preferential Amino Acid Sequences in Alumina-Catalyzed Peptide Bond Formation. *J. Inorg. Biochem.* **2002**, *90* (1–2), 1–7. [https://doi.org/10.1016/S0162-0134\(02\)00395-1](https://doi.org/10.1016/S0162-0134(02)00395-1).
- (30) McKee, A. D.; Solano, M.; Saydjari, A.; Bennett, C. J.; Hud, N. V.; Orlando, T. M. A Possible Path to Prebiotic Peptides Involving Silica and Hydroxy Acid-Mediated Amide Bond Formation. *ChemBioChem* **2018**, *19* (18), 1913–1917. <https://doi.org/10.1002/cbic.201800217>.

- (31) Frenkel-Pinter, M.; Haynes, J. W.; C, M.; Petrov, A. S.; Burcar, B. T.; Krishnamurthy, R.; Hud, N. V.; Leman, L. J.; Williams, L. D. Selective Incorporation of Proteinaceous over Nonproteinaceous Cationic Amino Acids in Model Prebiotic Oligomerization Reactions. *Proc. Natl. Acad. Sci.* **2019**, *116* (33), 16338–16346. <https://doi.org/10.1073/pnas.1904849116>.
- (32) Forsythe, J. G.; Yu, S.-S.; Mamajanov, I.; Grover, M. A.; Krishnamurthy, R.; Fernández, F. M.; Hud, N. V. Ester-Mediated Amide Bond Formation Driven by Wet-Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth. *Angew. Chem. Int. Ed.* **2015**, *54* (34), 9871–9875. <https://doi.org/10.1002/anie.201503792>.
- (33) Xiong, H.; Buckwalter, B. L.; Shieh, H. M.; Hecht, M. H. Periodicity of Polar and Nonpolar Amino Acids Is the Major Determinant of Secondary Structure in Self-Assembling Oligomeric Peptides. *Proc. Natl. Acad. Sci.* **1995**, *92* (14), 6349–6353. <https://doi.org/10.1073/pnas.92.14.6349>.
- (34) Bertrand, M.; Brack, A. Conformational Variety of Poly-anionic Peptides At Low Salt Concentrations. *Orig. Life Evol. Biosph.* **1997**, *27* (5), 585–595. <https://doi.org/10.1023/A:1006585925390>.
- (35) Lambert, J.-F. Adsorption and Polymerization of Amino Acids on Mineral Surfaces: A Review. *Orig. Life Evol. Biospheres* **2008**, *38* (3), 211–242. <https://doi.org/10.1007/s11084-008-9128-3>.
- (36) Zamaraev, K. I.; Romannikov, V. N.; Salganik, R. I.; Wlasoff, W. A.; Khramtsov, V. V. Modelling of the Prebiotic Synthesis of Oligopeptides: Silicate Catalysts Help to Overcome the Critical Stage. *Orig. Life Evol. Biosph.* **1997**, *27* (4), 325–337. <https://doi.org/10.1023/A:1006571800690>.
- (37) Sun, Y.; Frenkel-Pinter, M.; Liotta, C. L.; Grover, M. A. The PH Dependent Mechanisms of Non-Enzymatic Peptide Bond Cleavage Reactions. *Phys. Chem. Chem. Phys.* **2020**, *22* (1), 107–113. <https://doi.org/10.1039/C9CP05240B>.
- (38) Martra, G.; Deiana, C.; Sakhno, Y.; Barberis, I.; Fabbiani, M.; Pazzi, M.; Vincenti, M. The Formation and Self-Assembly of Long Prebiotic Oligomers Produced by the Condensation of Unactivated Amino Acids on Oxide Surfaces. *Angew. Chem. Int. Ed.* **2014**, *53* (18), 4671–4674. <https://doi.org/10.1002/anie.201311089>.
- (39) Sakhno, Y.; Battistella, A.; Mezzetti, A.; Jaber, M.; Georgelin, T.; Michot, L.; Lambert, J.-F. One Step up the Ladder of Prebiotic Complexity: Formation of Nonrandom Linear Polypeptides from Binary Systems of Amino Acids on Silica. *Chem. - Eur. J.* **2019**, *25* (5), 1275–1285. <https://doi.org/10.1002/chem.201803845>.
- (40) Ikehara, K. Possible Steps to the Emergence of Life: The [GADV]-Protein World Hypothesis. *Chem. Rec.* **2005**, *5* (2), 107–118. <https://doi.org/10.1002/tcr.20037>.
- (41) van der Gulik, P.; Massar, S.; Gilis, D.; Buhrman, H.; Rooman, M. The First Peptides: The Evolutionary Transition between Prebiotic Amino Acids and Early Proteins. *J. Theor. Biol.* **2009**, *261*, 531–539. <https://doi.org/10.1016/j.jtbi.2009.09.004>.
- (42) Longo, L. M.; Lee, J.; Blaber, M. Simplified Protein Design Biased for Prebiotic Amino Acids Yields a Foldable, Halophilic Protein. *Proc. Natl. Acad. Sci.* **2013**, *110* (6), 2135–2139. <https://doi.org/10.1073/pnas.1219530110>.
- (43) Gordon, E. F.; Muddiman, D. C. Impact of Ion Cloud Densities on the Measurement of Relative Ion Abundances in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: Experimental Observations of Coulombically Induced Cyclotron Radius Perturbations and Ion Cloud Dephasing Rates. *J. Mass Spectrom.* **2001**, *36* (2), 195–203. <https://doi.org/10.1002/jms.121>.
- (44) Ehler, K. W.; Orgel, L. E. N,N'-Carbonyldiimidazole-Induced Peptide Formation in Aqueous Solution. *Biochim. Biophys. Acta BBA - Protein Struct.* **1976**, *434* (1), 233–243. [https://doi.org/10.1016/0005-2795\(76\)90055-6](https://doi.org/10.1016/0005-2795(76)90055-6).
- (45) Marshall, A. G.; Hendrickson, C. L.; Jackson, G. S. Fourier Transform Ion Cyclotron Resonance Mass Spectrometry: A Primer. *Mass Spectrom. Rev.* **1998**, *17*, 1–35.
- (46) Nunes, R. S.; Cavaleiro, É. T. G. Thermalbehavior of Glutamic Acid and Its Sodium, Lithium and Ammonium Salts. *J. Therm. Anal. Calorim.* **2007**, *87* (3), 627–630. <https://doi.org/10.1007/s10973-006-7788-7>.
- (47) Jaber, M.; Georgelin, T.; Bazzi, H.; Costa-Torro, F.; Lambert, J.-F.; Bolbach, G.; Clodic, G. Selectivities in Adsorption and Peptidic Condensation in the (Arginine and Glutamic Acid)/Montmorillonite Clay System. *J. Phys. Chem. C* **2014**, *118* (44), 25447–25455. <https://doi.org/10.1021/jp507335e>.
- (48) Cech, N. B.; Enke, C. G. Practical Implications of Some Recent Studies in Electrospray Ionization Fundamentals. *Mass Spectrom. Rev.* **2001**, *20* (6), 362–387. <https://doi.org/10.1002/mas.10008>.
- (49) Lill, J. Proteomic Tools for Quantitation by Mass Spectrometry. *Mass Spectrom. Rev.* **2003**, *22* (3), 182–194. <https://doi.org/10.1002/mas.10048>.
- (50) Roepstorff, P.; Fohlman, J. Proposal for a Common Nomenclature for Sequence Ions in Mass Spectra of Peptides. *Biol. Mass Spectrom.* **1984**, *11* (11), 601–601. <https://doi.org/10.1002/bms.120011109>.
- (51) Downard, K. M.; Biemann, K. The Effect of Charge State and the Localization of Charge on the Collision-Induced Dissociation of Peptide Ions. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 966–975. [https://doi.org/10.1016/1044-0305\(94\)80015-4](https://doi.org/10.1016/1044-0305(94)80015-4).
- (52) Lahav, N.; White, D.; Chang, S. Peptide Formation in the Prebiotic Era: Thermal Condensation of Glycine in Fluctuating Clay Environments. *Science* **1978**, *201* (4350), 67–69. <https://doi.org/10.1126/science.663639>.
- (53) Adami, C.; Cerf, N. J. Physical Complexity of Symbolic Sequences. *Phys. Nonlinear Phenom.* **2000**, *137* (1), 62–69. [https://doi.org/10.1016/S0167-2789\(99\)00179-7](https://doi.org/10.1016/S0167-2789(99)00179-7).
- (54) Adami, C. What Is Complexity? *BioEssays* **2002**, *24* (12), 1085–1094. <https://doi.org/10.1002/bies.10192>.
- (55) Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27* (3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.
- (56) Dose, K.; Hartmann, J.; Christel Brand, M. Formation of Specific Amino Acid Sequences during Carbodiimide-Mediated Condensation of Amino Acids in Aqueous Solution. *Biosystems* **1982**, *15* (3), 195–200. [https://doi.org/10.1016/0303-2647\(82\)90004-1](https://doi.org/10.1016/0303-2647(82)90004-1).

## TOC graphics:

---



## Supporting Information

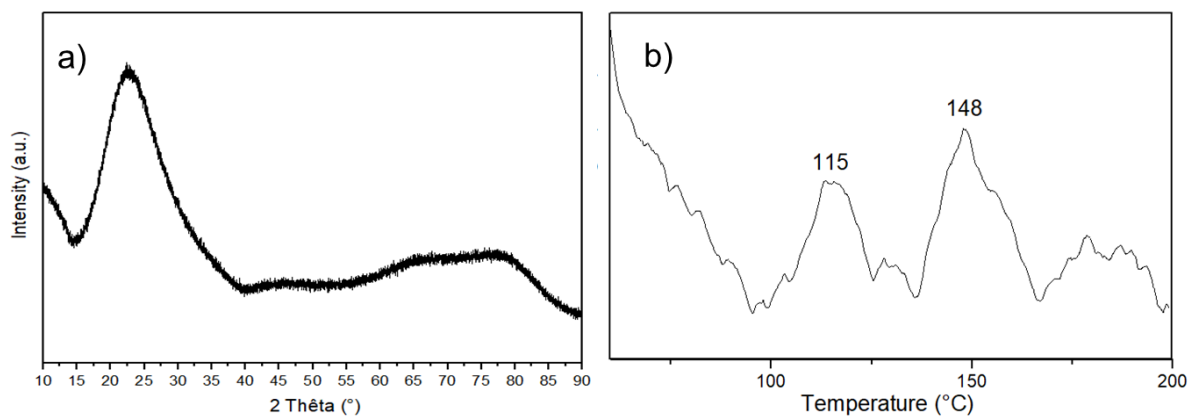
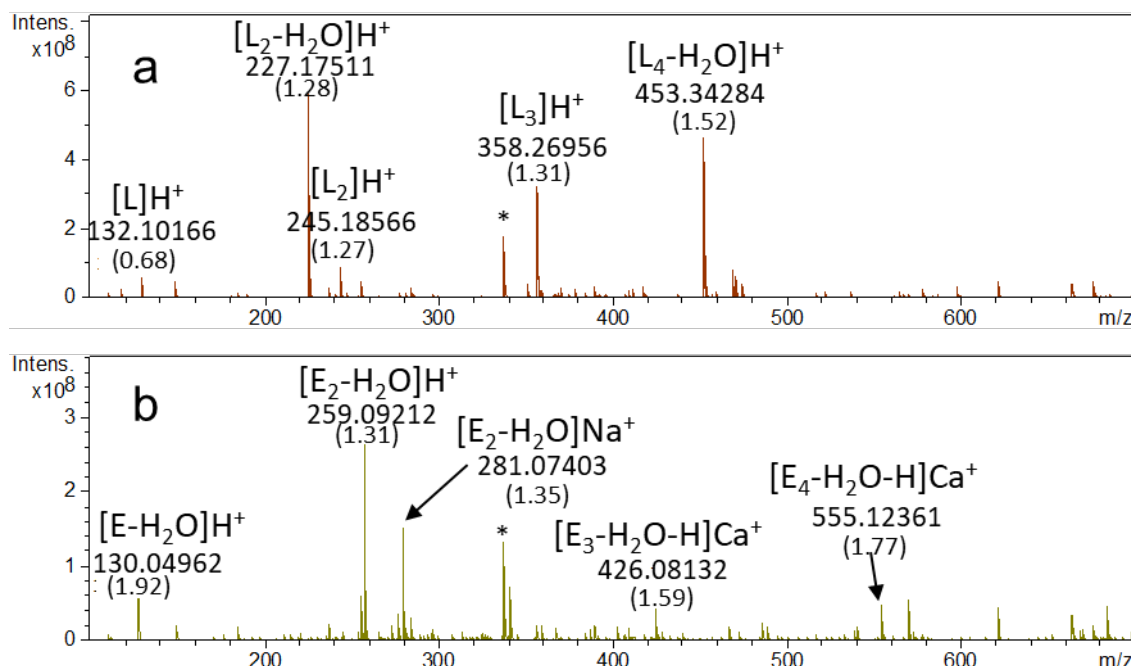


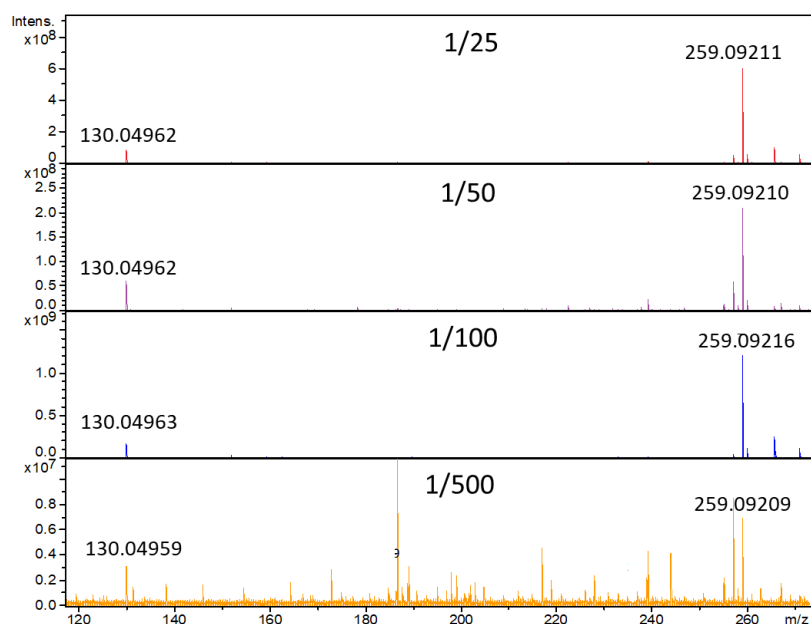
Figure S1. a) XRD and b) TGA spectra of Leu+Glu/SiO<sub>2</sub> system



**Figure S2. Positive ESI mass spectrum of desorption solution from thermally activated Leu/SiO<sub>2</sub> (a) and Glu/SiO<sub>2</sub> (b) systems using high resolution mass spectrometry (HRMS) detection.**

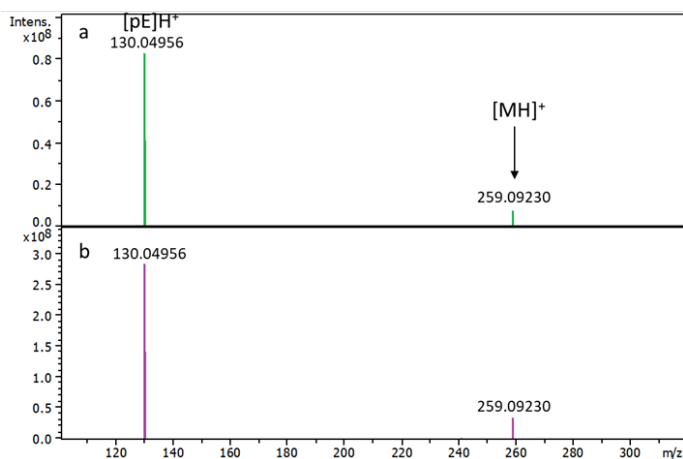
In Figure S2a, the predominant peak at m/z 227.17515 corresponds to stoichiometry [L<sub>2</sub>-H<sub>2</sub>O]H<sup>+</sup>. It can be assigned to a cyclic dimer cyclo(L-L), i.e. a substituted diketopiperazine (DKP), in accordance to previous works on surface catalyzed polymerization. The monomer ion, LH<sup>+</sup>, is still detected at low abundance (m/z 132.10182), meaning that polymerization is not quantitative. Other oligomer compounds are also detected. The linear dipeptide [L<sub>2</sub>]H<sup>+</sup> has an intensity five times smaller than the cyclic [L<sub>2</sub>-H<sub>2</sub>O]H<sup>+</sup>, and its presence is not surprising since it constitutes a likely intermediate in the formation of the latter. Peaks corresponding to the stoichiometries [L<sub>3</sub>]H<sup>+</sup> and [L<sub>4</sub>-H<sub>2</sub>O]H<sup>+</sup> could be due, respectively, to a linear trimer and a cyclic tetramer. However, they could also be assigned to non-covalent multimer ions, respectively between the monomer and DKP (L,cyclo(L-L)H<sup>+</sup>) and between two DKP species (cyclo(L-L),cyclo(L-L)H<sup>+</sup>). Since the constituting elements of these two adducts are indeed present in the solution, this assignment may be more likely than supposing a cyclic tetramer.

In Figure S2b, the desorption solution of activated Glu/SiO<sub>2</sub> does not show any untransformed monomer, but does show [E-H<sub>2</sub>O]H<sup>+</sup>, corresponding to internal cyclization (PyroGlu formation). The most intense peaks nominally correspond to dimers: [E<sub>2</sub>-H<sub>2</sub>O]H<sup>+</sup> and the corresponding sodium form, [E<sub>2</sub>-H<sub>2</sub>O]Na<sup>+</sup>. However, an alternative assignment could be the non-covalent complexes (pE,pEH<sup>+</sup>) and (pE,pENa<sup>+</sup>). This possibility has been specifically addressed by additional experiments (see Figures S3 and S4).



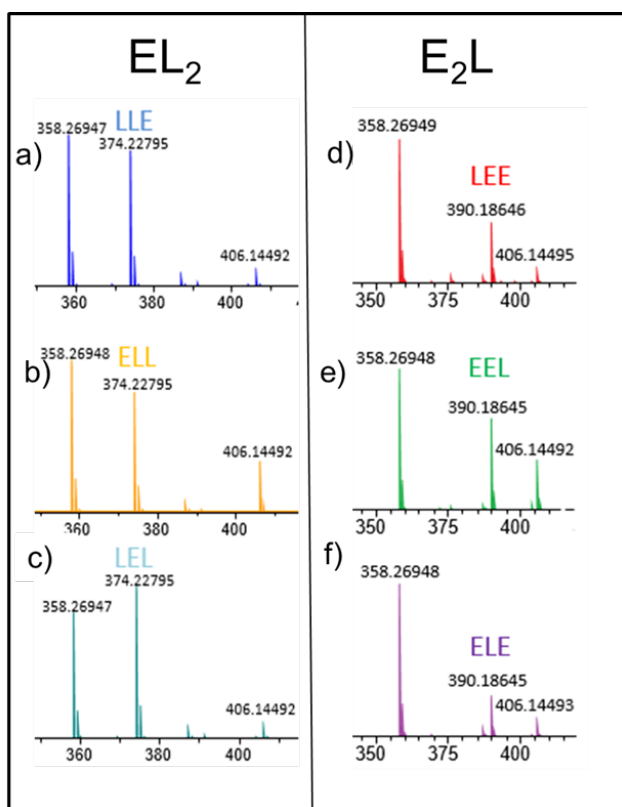
**Figure S3. Positive ESI mass spectrum of desorption solution from thermally activated Glu/SiO<sub>2</sub> at different solution dilutions for HRMS analysis.**

Evidence indicating that the species detected at m/z 259.0921 in the activated Glu/SiO<sub>2</sub> sample correspond to non-covalent multimer ions. The desorption solution was diluted 25, 50, 100 and 500 times: the 259.0921/130.0496 intensity ratio, where the species at m/z 130.0496 is the pE monomer, decreases upon dilution as expected for non-covalent complex species formed during the ESI processes.

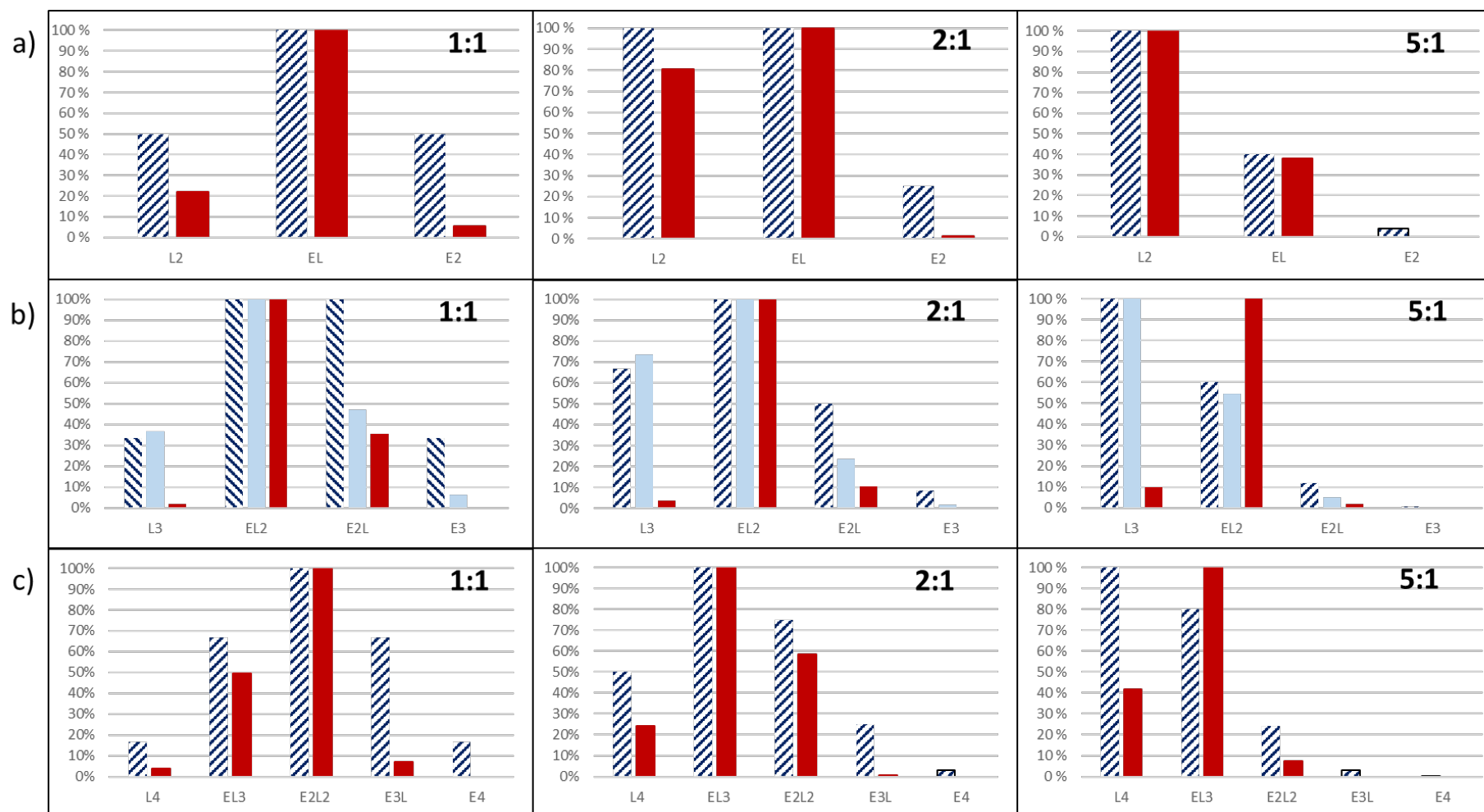


**Figure S4. MS/MS spectra of the  $m/z$  259.09230 precursor ion resulting from (a) E+pE mixture and (b) thermally activated Glu/SiO<sub>2</sub> sample**

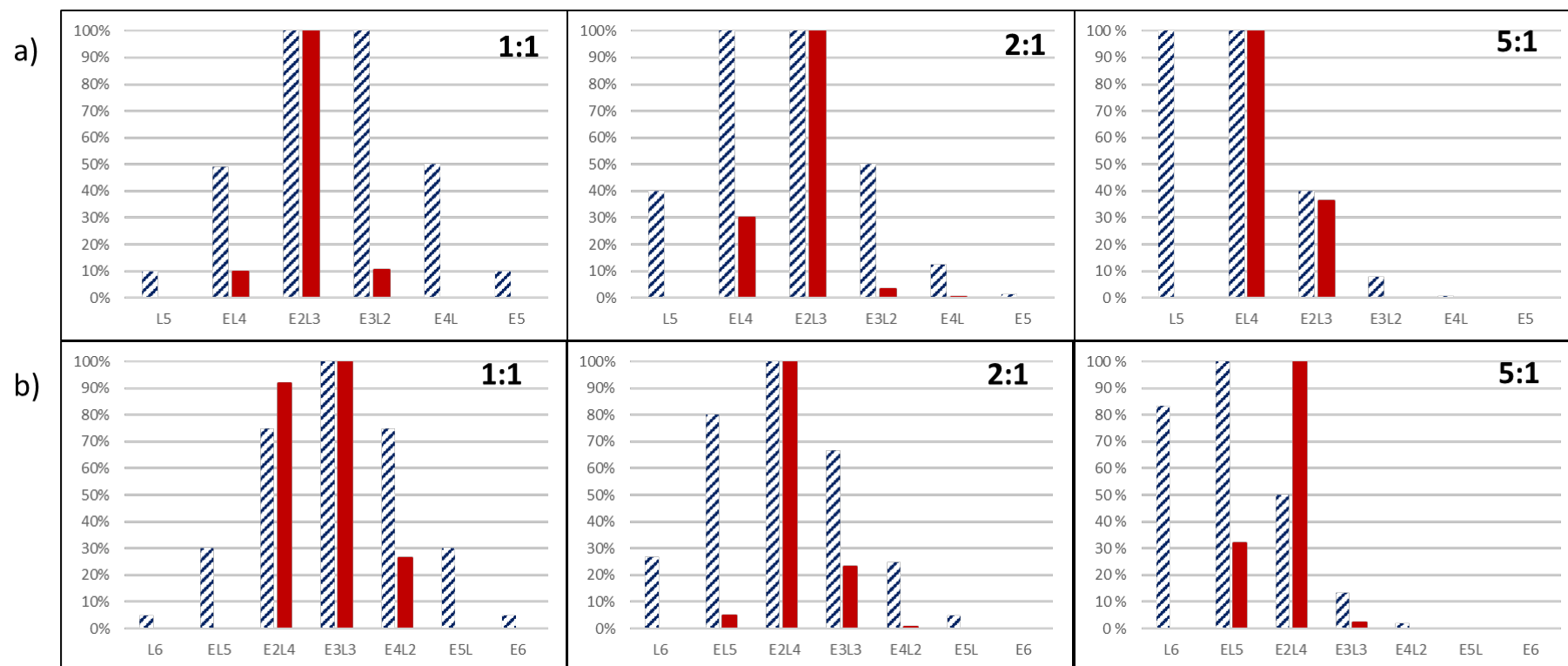
A mixture containing pyroglutamic acid (pE) and glutamic acid was prepared to form non-covalent dimer ions and compare their gas phase stability with supposed non-covalent species in the Glu/SiO<sub>2</sub> sample. Both MS/MS spectra are similar, indicating that  $m/z$  259.09230 ions should correspond to a non-covalent dimer formed by two pyroGlu units.



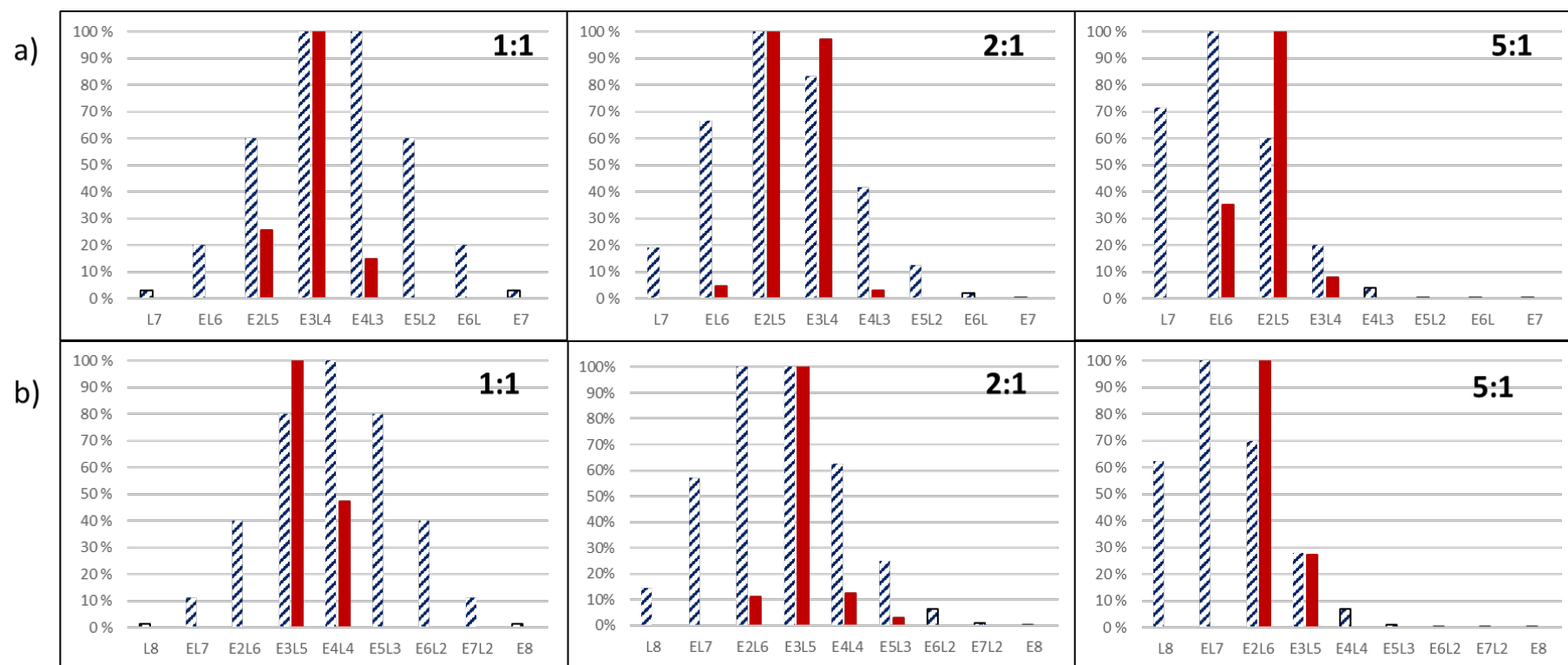
**Figure S5. Positive ESI mass spectra from equimolar mixtures of the reference tripeptides.** Each solution contains an equal amount of **LLL** ( $m/z$  358,26945), **EEE** ( $m/z$  406,14492), and one heterotrimer a) LLE, b) ELL, c) LEL, d) LEE, e) EEL and f) ELE sequences.



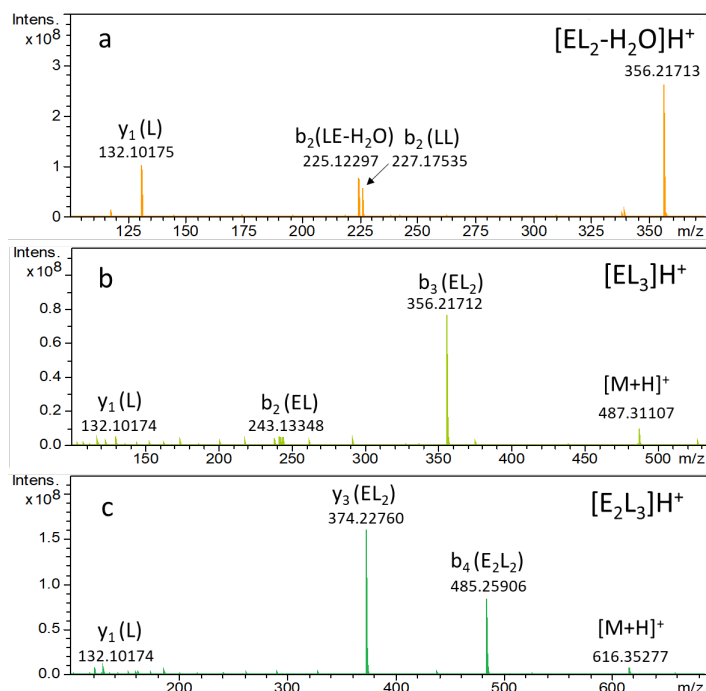
**Figure S6.** Comparison between the experimental peak intensity data (red bars) and expected distributions for random polymerization (hatched bars) for the different possible stoichiometries of a) dipeptides, b) tripeptides, c) tetrapeptides in systems with 1:1, 2:1 and 5:1 L:E molar ratios. For the tripeptides, the grey bars correspond to the expected peak intensities in the random scenario, taking into account the ionization efficiencies (cf. Figure 3).



**Figure S7. Comparison between the experimental peak intensity data (red bars) and expected distributions for random polymerization (hatched bars) for the different possible stoichiometries of a) pentapeptides, b) hexapeptides in systems with 1:1, 2:1 and 5:1 L:E molar ratios.**



**Figure S8. Comparison between the experimental data (red distribution) on a) heptapeptides, b) octapeptides of different stoichiometries and expected values for random polymerization (hatched distribution) of different stoichiometries and expected values for random polymerization for the 1:1, 2:1 and 5:1 L:E molar ratios.**



**Figure S9. MS/MS spectra of precursor ions at (a)  $m/z$  356.21713 corresponding to  $[\text{EL}_2\text{-H}_2\text{O}]\text{H}^+$  tripeptides, (b)  $m/z$  487.31107 corresponding to  $[\text{EL}_3]\text{H}^+$  tetramer, (c)  $m/z$  616.35277 corresponding to  $[\text{E}_2\text{L}_3]\text{H}^+$  pentamer from the thermally activated Leu+Glu/SiO<sub>2</sub> system. Note that putative peptide stoichiometries were obtained from accurate mass measurements.**

The fragmentation patterns detected for each peptide stoichiometry do not allow unambiguous sequence determination but give some clues about partial peptide sequencing.

- Fragments ions were annotated as  $y_1(\text{L})$  at  $m/z$  132 corresponding to a L residue at C-terminal, and two different  $b_2$  corresponding to LE-H<sub>2</sub>O motif (or a EL-H<sub>2</sub>O) and LL-H<sub>2</sub>O at  $m/z$  = 225 and 227 respectively. This fragmentation pattern suggests that we are dealing with is a mixture of two or more possible sequences.
- Fragments ions are annotated  $y_1(\text{L})$  at  $m/z$  132 and  $b_2(\text{EL})$ ,  $b_3(\text{EL}_2)$ . It is not possible to distinguish the ELLL and the LELL possible sequences because of the lack of  $b_1$  and  $y_3$  fragments. One can eliminate LLLLE and LLEEL sequences as no corresponding  $b_2$  or  $b_3$  fragments ions are detected.
- Fragments ions are annotated  $y_1(\text{L})$  at  $m/z$  132,  $y_3(\text{EL}_2)$ , and  $b_4(\text{E}_2\text{L}_2)$ . The absence of  $b_1$ ,  $b_2$  and  $y_2$  does not allow a firm conclusion for a major sequence. Indeed, detected fragment ions could be produced from several sequences and/or a mixture of them.

**Table SI 1: Assignment of main IR bands in non-activated Leu+Glu/SiO<sub>2</sub>**

Wavenumber (cm <sup>-1</sup> )	Assignment
1408	COO <sup>-</sup> symmetric stretching ( $\nu_{\text{sym CO}}$ )
1473	CH <sub>2</sub> asymmetric bending ( $\delta_{\text{as HCH}}$ )
1500	NH <sub>3</sub> <sup>+</sup> symmetric bending ( $\delta_{\text{sym HNH}}$ )
1595 (sh)	Mostly COO <sup>-</sup> asymmetric stretching ( $\nu_{\text{as CO}}$ )
1627	NH <sub>3</sub> <sup>+</sup> asymmetric bending ( $\delta_{\text{as HNH}}$ )
1718	C=O stretching, side chain COOH of Glu.