Supporting Information

Ester-Mediated Amide Bond Formation Driven by Wet–Dry Cycles: A Possible Path to Polypeptides on the Prebiotic Earth**

Jay G. Forsythe, Sheng-Sheng Yu, Irena Mamajanov, Martha A. Grover, Ramanarayanan Krishnamurthy,* Facundo M. Fernández,* and Nicholas V. Hud*

anie_201503792_sm_miscellaneous_information.pdf
I. Methods

Materials
Glycine, L-lactic acid, glycolic acid, L-alanine, D-alanine, L-leucine, L-serine, L-phenylalanine, triglycine, hydrochloric acid, sodium hydroxide, ammonium hydroxide, triethylamine, and deuterium oxide (99.9 mol %) were obtained from Sigma-Aldrich. Potassium hydrogen phthalate was from Fisher Chemical. Ultrapure water (18.2 MΩ/cm) was from a Barnstead Nanopure Diamond system (Van Nuys, CA).

Day/night cycling reactions
Reactions were started with 200 µL aqueous solution of L-lactic acid (100 mM) and of amino acid (100 mM), with the exception of the data presented below in Figure SE (which contained monomers at an initial concentration of 1 mM). In the warm/day stage, the solution was allowed to dry at 85°C for 18 hours. In the following cool/night stage, the dried samples were rehydrated with 200 µL of deionized water for 30 min. at room temperature and capped at 65°C for 5.5 hours, resulting in one 24-hour cycle. This process was repeated for the reported number of cycles. The only exceptions to this are shown in Figure SD along with their corresponding details. Before MS and/or NMR analysis, dried samples obtained after the warm/dry stage were rehydrated and diluted to the desired concentrations in either ultrapure water (MS) or D$_2$O (NMR), except for the lactic acid control sample (lactic acid oligomers; Figure 1c) which was dissolved in 90% acetonitrile and 10% water. No filtration was used for any sample. Before IR analysis, samples were not rehydrated after the final cycle as Attenuated Total Reflectance FTIR was used.

For the glycine-only control experiment (Figure 1b), solution pH was acidified using 0.5 M HCl in order to mimic the acidic environment of lactic acid. For pH-adjusted experiments, the pH of 100 mM L-lactic acid and 100 mM L-alanine aqueous solution was raised using 1 M sodium hydroxide, 1 M ammonium hydroxide, or 0.2 M triethylamine to pH 5, 7, or 9. After the drying process, reaction mixtures were rehydrated and the pH of the solution was adjusted to the initial value.

Mass spectrometry
MS analysis was performed on Waters Synapt G2 HDMS traveling-wave ion mobility and Waters Xevo G2 mass spectrometers, both equipped with a quadrupole/time-of-flight (qTOF) mass analyzer (Resolution mode; m/Δm = 20,000 FWHM). Samples were diluted in ultrapure deionized water to concentrations ranging from 0.013 to 0.044 mg mL$^{-1}$ (of starting monomers, concentrations of individual oligomers are lower) and directly infused into the mass spectrometer by electrospray ionization (10 µL min$^{-1}$ for 2 min) in negative-ion mode. For all samples, the capillary was set at 2.00 kV, the sampling cone voltage was 30, the extraction cone voltage was 3.0, the source temperature was 90°C, the cone gas flow was 20 L hr$^{-1}$, the desolvation gas flow was 650 L hr$^{-1}$, and the desolvation temperature was 250°C. The TOF was calibrated daily using sodium formate clusters from m/z 50 to 1200 with a threshold of 1 ppm. For accurate mass data (<10 ppm), peaks were centered based upon area and LockMass correction was applied using leucine enkephalin ([M-H]$^-$ = m/z 554.2615). Tandem MS was performed using collision-induced dissociation (CID) using argon gas in the transfer collision cell (Synapt G2). All specified collision energies are given in electron volts (eV) and are lab-frame. When ion mobility separation was used, the wave velocity ranged from 550 - 900 m s$^{-1}$, the wave amplitude was set at 40 V, and the mobility gas flow was 90 mL min$^{-1}$ N$_2$ (corresponding to a pressure of 3.67 mbar in the cell). Ion mobility data extraction was performed using Driftscope v2.1 software (Waters). Spectra were plotted in OriginPro 9.0 software.

Unless otherwise noted, all depsipeptide peak assignments correspond to [M-H]$^-$ ions. In traditional proteomic analysis, proteolytic peptides are commonly analyzed using positive-ion mode. However, because depsipeptides studied here contain an -OH instead of an -NH$_2$ on the “N” terminus, we found the sensitivity to be lower in positive mode. Additionally, tandem MS sequencing was more effective in negative-ion mode as essentially all fragment ions contained the C-terminus. This is consistent with literature on negative-ion mode peptide fragmentation (“α-cleavages”). Several DKP-like cyclic byproducts could only be detected in positive-ion mode and not in negative-ion mode (Figure SL).

For the data given in Figure 2b, $^{13}$C isotopic deconvolution was performed manually in order to determine the relative amounts of alanine and lactic acid within each oligomer. The mass difference between lactic
acid and alanine residues is 0.9840 Da, and thus X+1 isotopic signal from one combination of lactic acid and alanine (e.g., 1LA+2A) often overlapped with subsequent combinations of lactic acid and alanine (e.g., 2LA+1A). In order to quantitatively correct for this effect, signals were integrated and peak areas were obtained for each combination of lactic acid and alanine. Based upon the number of carbons, the relative contribution of $^{13}$C in the mass spectrum was determined (1.109% per carbon) and subtracted out, resulting in the net peak area from each unique combination of lactic acid and alanine. These areas were summed for each oligomer (e.g., $n=3$) and the relative amount of each combination was calculated. The average amounts of alanine and lactic acid were determined by multiplying each percentage by the number of residues within the depsipeptide. This process does assume that differences in ionization efficiency between combinations of lactic acid and alanine of the same size (e.g., all $n=3$ LA+A depsipeptides) are negligible. If positive-ion mode had been used, this assumption may not be valid, as backbone nitrogen atoms can be protonated. However, in negative-ion mode, we find it unlikely for deprotonation to persist anywhere other than the C-terminus (that is, without collisional activation), especially considering the side chains of lactic acid, alanine, and glycine. An example isotopic correction calculation is below:

**Determining relative amounts of LA+A trimer ($n=3$) for 12 cycles.**

$m/z$ #1 (corresponding to 1LA+2A; $C_9H_{15}N_2O_5$) = 231.10; peak area = 1.37E6. $^{13}$C contribution = 0; net area = 1.37E6.

$m/z$ #2 (corresponding to 2LA+1A; $C_9H_{14}NO_6$) = 232.10; peak area = 1.46E5. $^{13}$C contribution = (# carbon atoms)*($^{13}$C relative isotopic abundance)*area = (9)*(0.01109)*(1.37E6) = 1.37E5; net area = 1.46E5 - 1.37E5 = 9.26E3.

Sum area (all $n=3$ oligomers) = 1.37E6 + 9.26E3 = 1.38E6.

$1LA+2A = 1.37E6/1.38E6 = 99.3\%$

$2LA+1A = 9.26E3/1.38E6 = 0.7\%$

$%A = 0.993*(2/3\ residues) + 0.007*(1/3\ residues) = 66\%$

$%LA = 0.993*(1/3\ residues) + 0.007*(2/3\ residues) = 34\%$

Therefore, for $n=3$ depsipeptides containing alanine and lactic acid, the relative abundance of alanine is 66% and the relative abundance of lactic acid is 34%.

**IR spectroscopy**

IR data was obtained on a Thermo Nicolet 4700 FTIR Spectrometer. Samples were analyzed dry in an Attenuated Total Reflectance (ATR) sample chamber. Spectra were background-subtracted from 400 to 4000 cm$^{-1}$ and signal-averaged (16 scans per spectrum). Data processing (normalization and subtraction) was performed using OriginPro 9.0 software.

**NMR spectroscopy**

The NMR spectra were recorded by a Bruker Avance II-500 spectrometer. Dried samples obtained after the last warm/dry period were rehydrated in 600 µL $D_2$O. For quantitative $^1$H NMR analysis, a 1 mm capillary with potassium hydrogen phthalate (KHP) $D_2$O solution was used as external standard in the 5 mm diameter NMR tube, and data were collected using 30 degree pulse program with 30 seconds relaxation delay. HSQC and HMBC data were obtained by standard pulse program, hsqcetgpsi2 and hmbcgplpndqf, with 1.5 s relaxation delay.$[^3]$ Spectra width of $^1$H is 3501.401 Hz. For $^{13}$C, the spectra width is 25153.822 Hz. All spectra were processed and plotted by MestReNova 9.1.

Heteronuclear single quantum coherence (HSQC) and heteronuclear multiple bond coherence (HMBC) two-dimensional NMR analysis techniques also confirmed amide bond formation. For these analyses, we used the 16-cycle glycine and lactic acid sample with the replenishment of lactic acid at the 8th cycle as a representative sample. The $^1$H resonances of glycine and lactic acid residues were assigned using HSQC (Figure S7). The correlations between methylene and methyl groups were used to confirm the location of lactic acid residues in depsipeptides.

In the carbonyl region of HMBC spectra, most of the glycine residues correlate to two carbonyl carbons, with one of the carbonyl carbons correlated to the methyl groups of lactic acid residues. Because HMBC
is most sensitive for correlations spanning 2 or 3 covalent bonds, these correlations were used to assign LA and G residues linked amide bonds. Using triglycine as a standard, we confirmed that the C-terminal G-G amide bond is also detected in the HMBC spectrum (Figure Sγ). Due to spectral overlap, we are not able to differentiate internal G-G amide linkages from internal LA-G amide linkages at this time. Nevertheless, internal G-G amide linkages are observed by tandem MS sequencing (Figure 3b). Overlapping internal LA-G and internal G-G amide signals were integrated together in our yield calculations (combined with C-terminal G-G amide signals).

The chemical shifts of resonances in the above discussion are provided in Table S1. HSQC and HMBC spectra of the G/LA mixture with different cycles are shown in Figures Sη and Sθ, which show similar 1H, 13C chemical shifts/correlations.

Quantification of free glycine monomer and reacted glycine (incorporated into depsipeptides) was accomplished using 1H NMR. Peaks corresponding to amide bond-linked glycine residues were identified by two-dimensional NMR analyses as discussed in the previous section. A 30 sec relaxation delay was used to ensure full relaxation for accurate integration. Each data point was repeated three times.

As indicated in Eq. 1, the integrals of different species in the NMR spectrum are proportional to the products of molar concentration ($M_x$) and number of protons ($N_x$).

$$\frac{I_x}{I_y} = \frac{M_x N_x}{M_y N_y}$$

In this study, potassium hydrogen phthalate (KHP) was used as an external standard because its signals do not overlap with those of interest. The effective concentration of KHP ($M_{KHP}$) was calibrated by 75 mM glycine standard solution. The integral ratio between KHP and the glycine standard solution can be expressed as Eq. 2, which is rearranged (Eq. 3) to provide $M_{KHP}$.

$$\frac{I_{KHP}}{I_{Gly, std}} = \frac{M_{KHP} N_{KHP}}{M_{Gly, std} N_{Gly}}$$

$$M_{KHP} = \frac{I_{KHP} M_{Gly, std} N_{Gly}}{I_{Gly, std} N_{KHP}}$$

The concentration of a compound of interest ($M_x$) is calculated using Eq. 4 and previously determined $M_{KHP}$.

$$M_x = \frac{I_x M_{KHP} N_{KHP}}{I_{KHP} N_x}$$

Because every sample was diluted three-fold for replicate NMR experiments, the concentration of a compound of interest in the original sample is three times larger than the value calculated using Eq. 4. As a result, the concentration of glycine and lactic acid residues were calculated using Eqs. 5 and 6. All resonances from the methyl group of lactic acid units, including monomer and oligomers, were considered in our calculations.

$$M_{Gly} = \frac{I_{Gly} M_{KHP} N_{KHP}}{I_{KHP} N_{Gly}} \cdot 3$$
The yields of glycine residues with LA-G amide linkage ($Y_{LA-G}$) and C-terminal G-G linkage ($Y_{G-G}$) were calculated by using Eqs. 7 and 8. The volume of the sample solution ($V$) and the initial moles of glycine monomer ($n_{Gly\ initial}$) were considered to calculate the amount of amide-linked glycine. The ratio between the amount of amide-linked glycine and the initial amount of glycine monomer gives the yield of amide bonds. It was not possible to differentiate internal G-G amide linkages from internal LA-G amide linkages. All signals in the region of LA-G linkages were included in the yield calculation. Both $Y_{LA-G}$ and $Y_{G-G}$ were included in the calculation of overall amide bond yield ($Y_{amide\ bond}$) as shown in Eq. 9.

\[
Y_{LA-G} = \frac{I_{LA-G} M_{KHP} N_{KHP}}{I_{KHP} N_{LA-G}} \cdot V \cdot \frac{100\%}{n_{Gly\ initial}}
\]

(7)

\[
Y_{G-G} = \frac{I_{G-G} M_{KHP} N_{KHP}}{I_{KHP} N_{G-G}} \cdot V \cdot \frac{100\%}{n_{Gly\ initial}}
\]

(8)

\[
Y_{amide\ bond} = Y_{LA-G} + Y_{G-G}
\]

(9)

Example Calculation, amount of amide bonds in the 16-cycle LA-Gly sample:

The locations and integrals of peaks used for quantitation are shown in Figure Sπ and listed in Table S2. Other parameters used for calculation were given by Table S3.

The concentrations of remaining glycine monomer and lactic acid residues were calculated by Eqs. 5 and 6.

\[
M_{Gly} = \frac{0.83 \times 7.34 \times 4}{4 \times 2} \times 3 = 9.13 \text{ mM}
\]

\[
M_{LA} = \frac{7.99 \times 7.34 \times 4}{4 \times 3} \times 3 = 58.60 \text{ mM}
\]

The yield of amide-linked glycine was calculated by using Eqs. 7, 8 and 9. For the calculation of $Y_{LA-G}$, the integral values of peak 2 and 3 in Figure Sπ were used.

\[
Y_{LA-G} = \frac{(0.52 + 1.53) \times 7.34 \times 4}{4 \times 2} \times 600 \times 10^{-6} \times \frac{100\%}{0.02} = 22.56\%
\]

\[
Y_{G-G} = \frac{1.46 \times 7.34 \times 4}{4 \times 2} \times 600 \times 10^{-6} \times \frac{100\%}{0.02} = 16.06\%
\]

\[
Y_{amide\ bond} = 22.56\% + 16.06\% = 38.62\%
\]
II. Compounds in this Study

Main text and SI:

L-lactic acid  
Formula: C₇H₆O₄  
Neutral mass: 90.0317 Da

Glycine  
Formula: C₆H₁₄N₂O₂  
Neutral mass: 75.0320 Da

L-alanine  
Formula: C₅H₁₁NO₂  
Neutral mass: 89.0477 Da

SI only:

D-alanine  
Formula: C₅H₁₂N₂O₂  
Neutral mass: 89.0477 Da

L-leucine  
Formula: C₁₅H₂₃NO₂  
Neutral mass: 131.0946 Da

L-serine  
Formula: C₅H₁₁NO₃  
Neutral mass: 105.0426 Da

Glycolic acid  
Formula: C₃H₄O₂  
Neutral mass: 76.0160 Da
III. Mechanism

Two Possible Pathways

Pathway A:

1. HOOC(CH₃)CH(OH)COOH
2. HOOC(R₁)N₂H

3. HOOC(CH₃)CH(OH)COOH ↔ cyclic lactide
4. HOOC(R₁)NHCOOH ↔ cyclic lactone-amide

5. HOOC(R₁)NHCOO(CH₃)COOH → ester elongation
6. HOOC(R₁)NHCOO(CH₃)COOH

7. HOOC(R₁)N₂H

Pathway B:

1. HOOC(R₁)N₂H

8. HOOC(R₁)NHCOOH

9. HOOC(R₁)NHCOO(CH₃)COOH

10. HOOC(R₁)NHCOO(CH₃)COOH

11. HOOC(R₁)NHCOO(CH₃)COOH

12. HOOC(R₁)NHCOO(CH₃)COO(CH₃)COO(CH₃)COO(CH₃)COO(CH₃)COO(CH₃)COO(CH₃)COO(CH₃)COO(CH₃)COO(CH₃)

Detected by MS:

- Pathway A:
  - 4. cyclic lactone-amide

- Pathway B:
  - 8. not detectable by MS

Not prevalent:

- Pathway A:
  - 5. ester elongation

- Pathway B:
  - 9. HOOC(R₁)NHCOO(CH₃)COOH

Further activation and elongation:

- Pathway A:
  - 6. HOOC(R₁)NHCOO(CH₃)COOH
  - Oligopeptides with lactic acid residues at both ends, especially N-terminus
  - Experimental data supports pathway A to be predominant

- Pathway B:
  - 10. HOOC(R₁)NHCOO(CH₃)COO(CH₃)COO(CH₃)COO(CH₃)COO(CH₃)COO(CH₃)COO(CH₃)COO(CH₃)COO(CH₃)
  - Oligopeptides with lactic acid residue at C-terminus
  - Pathway B seems to be minor

Diketopiperazine (hinders further oligomerization)
We propose that two major pathways, labeled pathway A and pathway B, through which depsipeptide formation and amino acid enrichment may occur.
Description of proposed mechanisms: Lactic acid (LA, 1) and amino acid (AA, 2) are initially mixed together and dried down. As water evaporates, pathway A favors the formation of ester bonds between lactic acid residues to form a lactic acid dimer (3). This linear dimer is likely in equilibrium with the cyclic lactide, the ester equivalent of a diketopiperazine or DKP. The lactide dimer is less stable than an amide-linked DKP, however, and is expected to reversibly interchange with the linear form. The amine of an amino acid monomer attacks the ester carbonyl carbon in 3 to produce a heterodimer (4) containing one lactic acid residue and one amino acid residue linked with an amide bond. This species may cyclize as well. We do tentatively detect the 2LA and 1LA+1AA cyclic dimers in our samples (Figure SL). It is currently unclear whether the linear or cyclic forms are the active species; however, we note that depsipeptide formation does appear to be much less efficient starting from DKP than from the glycine monomer (Figure SD). Upon repeated cycling, the amide-linked heterodimer 4 continues to grow to 5 and, eventually, to 12. Spontaneous amide-ester exchange occurs on 12 to form 6, a linear depsipeptide with a lactic acid residue at the "N"-terminus and amino acid residues elsewhere. Further esterification leads to lactic acid residues being found at both termini and amino acid sequences being enriched internally. It is likely that not only monomers but also other short depsipeptides (e.g., the heterodimer 4) may add onto 6, resulting in the introduction of an internal lactic acid residue (13). Indeed, we do observe depsipeptides with a lactic acid residue between internal amino acids and C-terminal amino acid (Figures SR and ST). However, with increasing cycles we expect these species to convert further to pure peptide internal sequences (14). Pathway A appears to be most consistent with our data; therefore, we propose this is the main pathway for depsipeptide formation and the resulting transition towards peptide sequences.

Another route, pathway B, is also possible. In pathway B, a reversible ester-linked heterodimer (7) is formed as water is either removed or added. This heterodimer is observed in mass spectrum shown in Figure SE; however, it disappears as cycling increases. Subsequent attack on the ester carbonyl carbon by the amino acid leads to irreversible formation of dipeptide (8). Upon cycling, lactic acid adds to the C-terminus of the peptide to form 9, followed by amide-ester exchange to form the tripeptide 10. A likely side-product in pathway B is a DKP (11).[6-8] Although we tentatively identify mixed ester-amide DKP-like cyclic species, pure amino acid DKPs have not been detected to date (Figure SL). Moreover, we do not observe pure peptides in our samples without containing at least one lactic acid residue, and depsipeptides with lactic acid residues solely on C-termini are extremely rare. Therefore, at this time, we propose that pathway B provides only a minor contribution, if any.
### IV. Accurate Mass Tables

| Oligomer (n=2 thru 8) | Exact mass  | Theo. [M-H]^- | Meas. [M-H]^- | Error (ppm) |
|----------------------|-------------|---------------|---------------|-------------|
| **Lactic Acid + Glycine** |             |               |               |             |
| 1 LA + 1 G           | 147.0532    | 146.0454      | 146.0455      | 0.9         |
| 1 LA + 2 G           | 204.0747    | 203.0669      | 203.0671      | 1.1         |
| 1 LA + 3 G           | 261.0962    | 260.0884      | 260.0876      | -3.0        |
| 1 LA + 4 G           | 318.1177    | 317.1099      | 317.1097      | -0.5        |
| 2 LA + 1 G           | 219.0743    | 218.0665      | 218.0664      | -0.3        |
| 2 LA + 2 G           | 276.0958    | 275.0880      | 275.0882      | 0.8         |
| 2 LA + 3 G           | 333.1173    | 332.1095      | 332.1097      | 0.7         |
| 2 LA + 4 G           | 390.1388    | 389.1310      | 389.1311      | 0.3         |
| 2 LA + 5 G           | 447.1603    | 446.1525      | 446.1508      | -3.7        |
| 2 LA + 6 G           | 504.1818    | 503.1740      | 503.1712      | -5.5        |
| 3 LA + 1 G           | 291.0954    | 290.0876      | 290.0875      | -0.2        |
| 3 LA + 2 G           | 348.1169    | 347.1091      | 347.1091      | 0.0         |
| 3 LA + 3 G           | 405.1384    | 404.1306      | 404.1308      | 0.6         |
| 3 LA + 4 G           | 462.1599    | 461.1521      | 461.1518      | -0.6        |
| 3 LA + 5 G           | 519.1814    | 518.1736      | 518.1728      | -1.5        |
| 4 LA + 1 G           | 363.1165    | 362.1087      | 362.1087      | 0.0         |
| 4 LA + 2 G           | 420.1380    | 419.1302      | 419.1305      | 0.8         |
| 4 LA + 3 G           | 477.1595    | 476.1517      | 476.1512      | -1.0        |
| 4 LA + 4 G           | 534.1810    | 533.1732      | 533.1719      | -2.4        |
| 5 LA + 1 G           | 435.1376    | 434.1298      | 434.1299      | 0.3         |
| 5 LA + 2 G           | 492.1591    | 491.1513      | 491.1508      | -1.0        |
| 5 LA + 3 G           | 549.1806    | 548.1728      | 548.1722      | -1.0        |
| 6 LA + 1 G           | 507.1587    | 506.1509      | 506.1516      | 1.4         |
| 6 LA + 2 G           | 564.1802    | 563.1724      | 563.1710      | -2.4        |
| 7 LA + 1 G           | 579.1798    | 578.1720      | 578.1715      | -0.8        |
| Oligomer \((n=2\text{ thru } 8)\) | Exact mass | Theo. [M-H] \(^{-}\) | Meas. [M-H] \(^{-}\) | Error (ppm) |
|-----------------|------------|-----------------|-----------------|-------------|
| **Lactic Acid + Alanine** | | | | |
| 1 LA + 1 A | 161.0688 | 160.0610 | 160.0611 | 0.8 |
| 1 LA + 2 A | 232.1059 | 231.0981 | 231.0981 | 0.0 |
| 1 LA + 3 A | 303.1430 | 302.1352 | 302.1338 | -4.5 |
| 1 LA + 4 A | 374.1801 | 373.1723 | 373.1711 | -3.1 |
| 2 LA + 1 A | 233.0899 | 232.0821 | 232.0826 | 2.3 |
| 2 LA + 2 A | 304.1270 | 303.1192 | 303.1194 | 0.8 |
| 2 LA + 3 A | 375.1641 | 374.1563 | 374.1565 | 0.6 |
| 2 LA + 4 A | 446.2012 | 445.1934 | 445.1930 | -0.8 |
| 2 LA + 5 A | 517.2383 | 516.2305 | 516.2281 | -4.6 |
| 2 LA + 6 A | 588.2754 | 587.2676 | 587.2625 | -5.0 |
| 3 LA + 1 A | 305.1110 | 304.1032 | 304.1039 | 2.4 |
| 3 LA + 2 A | 376.1481 | 375.1403 | 375.1404 | 0.3 |
| 3 LA + 3 A | 447.1852 | 446.1774 | 446.1771 | -0.6 |
| 3 LA + 4 A | 518.2223 | 517.2145 | 517.2142 | -0.5 |
| 3 LA + 5 A | 589.2594 | 588.2516 | 588.2506 | -0.6 |
| 4 LA + 1 A | 377.1321 | 376.1243 | 376.1257 | 3.8 |
| 4 LA + 2 A | 448.1692 | 447.1614 | 447.1619 | 1.2 |
| 4 LA + 3 A | 519.2063 | 518.1985 | 518.1988 | 0.6 |
| 4 LA + 4 A | 590.2434 | 589.2356 | 589.2375 | 3.3 |
| 5 LA + 1 A | 449.1532 | 448.1454 | 448.1479 | 5.6 |
| 5 LA + 2 A | 520.1903 | 519.1825 | 519.1836 | 2.2 |
| 5 LA + 3 A | 591.2274 | 590.2196 | 590.2213 | 2.9 |
| 6 LA + 1 A | 521.1743 | 520.1665 | 520.1695 | 5.8 |
| 6 LA + 2 A | 592.2114 | 591.2036 | 591.2045 | 1.6 |
| 7 LA + 1 A | 593.1954 | 592.1876 | 592.1918 | 7.1 |
| Oligomer \((n=2\text{ thru }8)\) | Exact mass | Theo. \([\text{M-H}]^-\) | Meas. \([\text{M-H}]^-\) | Error (ppm) |
|---------------------------------|------------|-----------------|-----------------|------------|
| **Lactic Acid + Leucine**       |            |                 |                 |            |
| 1 LA + 1 L                      | 203.1158   | 202.1080        | 202.1082        | 1.1        |
| 1 LA + 2 L                      | 316.1999   | 315.1921        | 315.1918        | -0.9       |
| 1 LA + 3 L                      | 429.2840   | 428.2762        | 428.2803        | 9.6        |
| 2 LA + 1 L                      | 275.1369   | 274.1291        | 274.1277        | -5.0       |
| 2 LA + 2 L                      | 388.2210   | 387.2132        | 387.2136        | 1.1        |
| 2 LA + 3 L                      | 501.3051   | 500.2973        | 500.2983        | 2.1        |
| 3 LA + 1 L                      | 347.1580   | 346.1502        | 346.1499        | -0.8       |
| 3 LA + 2 L                      | 460.2421   | 459.2343        | 459.2343        | 0.0        |
| 3 LA + 3 L                      | 573.3262   | 572.3184        | 572.3165        | -3.3       |
| 4 LA + 1 L                      | 419.1791   | 418.1713        | 418.1706        | -1.6       |
| 4 LA + 2 L                      | 532.2632   | 531.2554        | 531.2543        | -2.0       |
| 5 LA + 1 L                      | 491.2002   | 490.1924        | 490.1912        | -2.4       |
| 5 LA + 2 L                      | 604.2843   | 603.2765        | 603.2722        | -7.1       |
| **Lactic Acid + Serine**        |            |                 |                 |            |
| 1 LA + 1 S                       | 177.0637   | 176.0559        | 176.0561        | 1.3        |
| 1 LA + 2 S                       | 264.0957   | 263.0879        | 263.0872        | -2.5       |
| 1 LA + 3 S                       | 351.1277   | 350.1199        | 350.1176        | -6.5       |
| 2 LA + 1 S                       | 249.0848   | 248.0770        | 248.0758        | -4.7       |
| 2 LA + 2 S                       | 336.1168   | 335.1090        | 335.1071        | -5.6       |
| 2 LA + 3 S                       | 423.1488   | 422.1410        | 422.1380        | -7.0       |
| 3 LA + 1 S                       | 321.1059   | 320.0981        | 320.0962        | -5.8       |
| 3 LA + 2 S                       | 408.1379   | 407.1301        | 407.1273        | -6.8       |
| 3 LA + 3 S                       | 495.1699   | 494.1621        | 494.1576        | -9.0       |
| 4 LA + 1 S                       | 393.1270   | 392.1192        | 392.1170        | -5.5       |
| 4 LA + 2 S                       | 480.1590   | 479.1512        | 479.1484        | -5.8       |
| 4 LA + 3 S                       | 567.1910   | 566.1832        | 566.1826        | -1.0       |
| Oligomer \((n=2\text{ thru }8)\) | Exact mass  | Theo. \([\text{M-H}]^-\) | Meas. \([\text{M-H}]^-\) | Error (ppm) |
|-----------------------------|-------------|----------------------------|-----------------------------|-------------|
| **Lactic Acid + Gly + Ala** |             |                            |                             |             |
| 1 LA + 1 A + 1 G            | 218.0903    | 217.0825                   | 217.0836                    | 5.2         |
| 2 LA + 1 A + 1 G            | 290.1114    | 289.1036                   | 289.1044                    | 2.9         |
| 3 LA + 1 A + 1 G            | 362.1325    | 361.1247                   | 361.1263                    | 4.5         |
| 4 LA + 1 A + 1 G            | 434.1536    | 433.1458                   | 433.1471                    | 3.1         |
| 1 LA + 2 A + 1 G            | 289.1274    | 288.1196                   | 288.1189                    | -2.3        |
| 1 LA + 1 A + 2 G            | 275.1118    | 274.1040                   | 274.1040                    | 0.0         |
| 2 LA + 2 A + 1 G            | 361.1485    | 360.1407                   | 360.1406                    | -0.2        |
| 2 LA + 1 A + 2 G            | 347.1329    | 346.1251                   | 346.1252                    | 0.4         |
| 2 LA + 2 A + 2 G            | 418.1700    | 417.1622                   | 417.1623                    | 0.3         |
| 2 LA + 1 A + 3 G            | 404.1544    | 403.1466                   | 403.1461                    | -1.2        |
| 2 LA + 3 A + 1 G            | 432.1856    | 431.1778                   | 431.1754                    | -5.5        |
| 2 LA + 2 A + 3 G            | 475.1915    | 474.1837                   | 474.1811                    | -5.4        |
| 3 LA + 1 A + 1 G            | 362.1325    | 361.1247                   | 361.1263                    | 4.5         |
| 3 LA + 2 A + 1 G            | 433.1696    | 432.1618                   | 432.1625                    | 1.7         |
| 3 LA + 1 A + 2 G            | 419.1540    | 418.1462                   | 418.1481                    | 4.6         |
| 3 LA + 2 A + 2 G            | 490.1911    | 489.1833                   | 489.1828                    | -1.0        |
Figure S1. Zoomed-in mass spectrum of a glycine control sample (pH 3; 4 cycles). Glycine monomer is present but no glycine peptides are detected at m/z 131.0458 (di-Gly), 188.0673 (tri-Gly), or 245.0888 (tetra-Gly). The majority of signals in this m/z range correspond to trace contaminants found in the solvent blank, including: m/z 157.1242, 171.1389, 185.1572, 194.0496, 199.1718, 213.1869, 227.2039, 255.2346, 283.2603, and 297.1497. For example, we tentatively identify m/z 255.2346 as palmitic acid, a common ESI contaminant with a theoretical [M-H]⁻ of 255.2330 Da (~6 ppm error). However, it is also possible environmental cycling side products are present in low abundance as well.
Figure S2. Depsipeptide formation at a lower dry-state temperature of 65°C. Tandem MS sequencing of 1LA+1G depsipeptide formed using dry and wet temperatures of 65°C for two cycles. Both the ester-linked and amide-linked depsipeptide are observed. CID collision energy = 10 eV (lab-frame). While this is the lowest temperature that we observe an amide bond, oligoesters are observed with a dry-down temperature of 55°C (data not shown).
Figure S3. Depsipeptide formation occurs starting with cyclic-diGly (diketopiperazine), but is less efficient than starting with Gly monomer. Tandem MS sequencing of 1LA+1G depsipeptide found in LA + Gly and LA + Gly-Gly DKP samples. Both the ester-linked and amide-linked depsipeptide are observed. CID collision energy = 8 eV (lab-frame). Signal for the LA-G-G trimer (m/z 203.07) was too low to isolate and fragment in the LA + GlyGly DKP sample, although it was tentatively identified.
Figure S4. Investigation on the formation of diketopiperazines (DKPs) in the alanine-lactic acid system after 4 cycles by positive-ion mode MS (2.50 kV capillary voltage, 40 V sampling cone, 4 V extraction cone). In the above mass spectrum, 25 µM alanine DKP is spiked in the sample. Alanine DKP is not observed in the unspiked sample; however, 1LA+1A cyclic lactone amide and 2LA cyclic lactide are tentatively identified. Nevertheless, signal intensities for 1LA+1A and 2LA DKPs are too low to confirm by tandem MS fragmentation.
Figure S5. Mass spectra of 1:1 mol lactic acid and alanine after 1 cycle, 4 cycles, 8 cycles, and 12 cycles (at standard reaction conditions of 85°C for 18 hours and 65°C for 5.5 hours). As the number of cycles increases, the relative abundance of alanine in each oligomer increases. Notably, however, the overall length of the depsipeptides does not appear to change significantly without further addition of lactic acid monomer.
Figure S6. Tandem MS sequencing of lactic acid-alanine heterodimer (parent ion [M-H]^- = 160.06 Da) after 1 cycle, 4 cycles, 8 cycles, and 12 cycles. As the number of cycles increases, the relative abundance of ester bond (Ala-LA) decreases and the relative abundance of amide or peptide bond (LA-Ala) increases. After 8 cycles, essentially all heterodimer is LA-Ala. CID collision energy = 15 eV (lab-frame).
Figure S7. FTIR spectra of (top) glycine, (middle) diglycine, and (bottom) triglycine standards. The Amide I and II bands are clearly observed in the tripeptide (~1640 cm\(^{-1}\) and ~1520 cm\(^{-1}\), respectively).
Figure S8. FTIR spectra of (top) lactic acid monomer and (bottom) polylactic acid standards. The C=O band shifts from 1714 cm\(^{-1}\) (free acid) to 1757 cm\(^{-1}\) (ester). No bands are observed in the Amide I and II regions.
Figure S9. FTIR spectra of 1:1 mol glycine and lactic acid mixture after (top) 1 cycle, (middle) 8 cycles, and (bottom) 16 cycles where lactic acid was re-added after the 8th cycle due to depletion. Growth of the Amide I and II bands (1641 cm$^{-1}$ and 1531 cm$^{-1}$) matches those from the triglycine standard sample.
Table S1. Chemical shifts of assigned resonances observed in HSQC and HMBC spectra of the glycine-lactic acid mixture after 16 cycles. Initial amount of lactic acid (0.02 mmol) was added at 8th cycle to reactivate the reaction.

| Residues                          | HSQC          | HMBC          |
|-----------------------------------|---------------|---------------|
|                                   | \( \delta^1 \) (ppm) | \( \delta^{13} \) (ppm) | \( \delta^1 \) (ppm) | \( \delta^{13} \) (ppm) |
| Glycine monomer (C-H\(_2\))       | 3.45          | 41.29         | 3.45              | 172.57               |
| G-G, C-terminus (C-H\(_2\))      | 3.67          | 42.88         | 3.67              | 171.15,176.53        |
| LA-G (C-H\(_2\))                 | 3.83-3.96     | 39.17-44.59   | 3.83-3.96         | 167.51-179.69        |
| LA (C-H\(_2\))                   | 4.04          | 40.84         | 4.04              | 170.63,178.45        |
| LA (C-H\(_2\))                   | 4.00          | 68.32         | 1.29              | 178.7                |
| LA (C-H\(_2\))                   | 4.22          | 67.57         | 1.33              | 178.45               |
| LA (C-H\(_3\))                   | 4.78          | 72.75         | 1.28              | 67.62                |
| LA (C-H\(_3\))                   | 5.11          | 72.19         |                   |                     |
| LA                                 | 1.28          | 19.28         |                   |                     |
|                                   | 1.40          | 16.61         |                   |                     |
Figure S10. $^1$H-$^{13}$C HSQC (green) and HMBC (red) spectra of the 16-cycle glycine-lactic acid sample with replenishment of lactic acid at the 8th cycle. (a) Full spectra. (b-e) Enlarged spectral regions of interest. Red lines indicate the correlation of methylene or carbonyl carbon to the methyl protons of lactic acid. Blue lines show the correlation of methylene protons to two different carbonyl carbons.
Figure S11. Confirmation of C-terminal amide bond. (a) $^1$H-$^{13}$C HMBC spectrum of the triglycine standard. Five peaks were assigned based on the correlation between methylene protons and carbonyl carbons. (b) HMBC spectra overlap of the 16 cycles sample (green) and the triglycine standard (red). C-terminus correlations (peaks 4 and 5) were also observed in the 16-cycle mixture suggested the presence of C-terminal amide bond.
Figure S12. $^1$H-$^{13}$C HSQC (green) and HMBC (red) spectra of the 1-cycle glycine-lactic acid mixture.
Figure S13. $^1H-^{13}C$ HSQC (green) and HMBC (red) spectra of the 8 cycle glycine-lactic acid mixture.
Figure S14. $^1$H NMR spectrum of glycine-lactic acid mixture after 16 cycles. Lactic acid was replenished at the 8th cycle.

| Peak Number | Residues                  | Range (ppm) | Integral value ($I_x$) | Number of protons ($N_x$) |
|-------------|---------------------------|-------------|------------------------|---------------------------|
| 1           | KHP                       | 7.20-7.45   | 4.00                   | 4                         |
| 2           | LA-G (CH$_2$)             | 4.02-4.06   | 0.52                   | 2                         |
| 3           | LA-G (CH$_2$)             | 3.84-3.95   | 1.53                   | 2                         |
| 4           | G-G (CH$_2$)              | 3.64-3.68   | 1.46                   | 2                         |
| 5           | Glycine monomer (CH$_2$)  | 3.32-3.55   | 0.83                   | 2                         |
| 6           | LA (CH$_3$)               | 1.15-1.53   | 7.99                   | 3                         |
Table S3. List of parameters used in the yield calculation.

| Name                                      | Notation | Value             |
|-------------------------------------------|----------|-------------------|
| The volume of NMR samples                 | $V$      | $600 \times 10^{-6}$ L |
| The concentration of glycine standard     | $M_{\text{Gly, std}}$ | 75 mM             |
| solution                                  |          |                   |
| The concentration of KHP                  | $M_{\text{KHP}}$ | 7.34 mM           |
| The initial amount of glycine             | $n_{\text{Gly, initial}}$ | 0.02 mmol         |

Figure S15. The amounts of glycine monomer and total lactic acid (monomer + residues in products) across different cycles. $M_{\text{Gly}} = \text{concentration of unreacted glycine monomer}$. $M_{\text{LA}} = \text{concentration of total lactic acid}$. Initial amount of lactic acid (0.02 mmol) was added at 8th cycle to reanimate the reaction. As the amount of total lactic acid increases after the 8th cycle, the reaction rate appears to increase as well, resulting in less free glycine monomer. All values came from the calculation using Eqs. 5 and 6.
Figure S16. (top) Mass spectrum of alanine-lactic acid mixture after 4 cycles, adjusted to pH 7 with triethylamine. (middle-bottom) Enrichment of alanine-containing depsipeptides from 1 to 4 cycles. The rate of enrichment may be lower than that of unbuffered solutions (also see Figs. 2a, 4a in main text); we look to explore this in future experiments. Signals with m/z 378.11 correspond to $^{13}$C-containing 5LA oligoesters.
Figure S17. Mass spectrum of alanine-lactic acid mixture after 1 cycle, adjusted to pH 5, 7, and 9 with ammonia. Peaks labeled with asterisks correspond to sodiated monomer cluster species and not oligomers (e.g., m/z 201.0388 corresponds to \([2\text{Lactic acid} + 1\text{Na} - 2\text{H}]^+\); theo. m/z = 201.0376, C\(_6\)H\(_{10}\)O\(_6\)Na). NH\(_4\)OH was added to adjust pH, forming an ammonium lactate buffer solution. Depsipeptide formation was most efficient at pH 9 but also occurred at pH 5 and pH 7.
Figure S18. Mass spectrum of alanine-lactic acid mixture after 1 cycle, adjusted to pH 7 with sodium hydroxide. Peaks labeled with asterisks correspond to sodiated monomer cluster signals and not oligomers. For example, m/z 201.0394 corresponds to [2Lactic acid+1Na-2H] and m/z 200.0555 corresponds to [1Lactic acid+1Alanine-2H+Na]. (inset) The 1LA+1A heterodimer is observed but is the only depsipeptide detected in this sample (that is, after 1 cycle). Samples pH-adjusted with ammonia and triethylamine generate significantly more depsipeptide signal. We hypothesize this is due to the fact that sodium hydroxide cannot provide a proton to make the carboxylate a good leaving group in the ester-amide exchange reaction. Other factors include that ammonia and trimethylamine are more volatile, and the ionization efficiency in MS analysis is likely reduced in NaOH due to sodium-induced ion suppression.
Figure S19. Ion mobility (IM) separation of alanine-lactic acid depsipeptides based on oligomer size. Depsipeptides were formed from initial ratios of 30% alanine and 70% lactic acid and subjected to 8 cycles. Ion mobility traveling wave velocity = 650 m s$^{-1}$. Traveling wave height = 40 V.
Figure S20. The largest oligomers observed using electrospray IM-MS to date ($n=12$-$14$) are alanine-lactic acid depsipeptides. These depsipeptides are made using 30% alanine and 70% lactic acid after 8 environmental cycles. IM was used prior to mass analysis to spread the signal into a second dimension and thus decrease overlap and improve S/N for larger oligomers. Ion mobility traveling wave velocity = 900 m s$^{-1}$. Traveling wave height = 40 V.
Figure S21. Depsipeptide formation at a lower initial concentration (10 mM monomers; 10x dilution of stock solution used in other experiments) after 4 cycles. MS analysis was performed on Agilent 6130 single quadrupole mass spectrometers, samples were diluted in ultrapure deionized water to 0.0825 mg mL\(^{-1}\) and directly injected into the mass spectrometer by electrospray ionization in negative-ion mode.
Figure S22. Mass spectra of alanine-lactic acid mixtures after 8 cycles with varying mole ratios of starting monomer. (top) 90% alanine and 10% lactic acid; (middle) 30% alanine and 70% lactic acid; (bottom) 10% alanine and 90% lactic acid. Depsipeptide formation and alanine incorporation is optimal when a higher ratio of lactic acid is present. Sodiated alanine monomer clusters, marked with asterisks, are observed in the 90% alanine sample and not the others (e.g., m/z 199.0676 corresponds to [1Alanine+1Alanine+Na-2H]; theo. m/z = 199.0696, C₆H₁₂O₄N₂Na), reflecting a higher amount of unreacted alanine monomer. Previous results from Cronin and Moore[9] and Peltzer and Bada[10] show the amounts of alanine and lactic acid in the Murchison meteorite to be 3.5 µg g⁻¹ and 5.9 µg g⁻¹, respectively. This mass ratio corresponds to a relative mole ratio of approximately 40% alanine and 60% lactic acid starting monomers. Notably, we observe depsipeptides form readily from similar starting ratios (e.g., 50% alanine and 50% lactic acid, as shown in the main text, and 30% alanine and 70% lactic acid, as shown here).
Figure S23. Mass spectra for the 1:1 mol leucine and lactic acid system after 1 cycle, 4 cycles, and 8 cycles. Leucine forms depsipeptides with lactic acid like glycine and alanine.
Figure S24. Tandem MS sequencing of leucine-lactic acid depsipeptides. (top) Leucine-lactic acid heterodimer sequence after 8 cycles is LA-Leu. Collision energy = 15 eV. (middle) Two 2LA+2L sequences are observed after 8 cycles, LA-Leu-Leu-LA and LA-Leu-LA-Leu. Collision energy = 15 eV. (bottom) Two 2LA+3L sequences are observed after 8 cycles, LA-Leu-Leu-Leu-LA and LA-Leu-Leu-LA-Leu. Collision energy = 18 eV. The presence of the LA-Leu sequence on the C-terminus suggests the LA-Leu heterodimer is adding to the larger depsipeptide in addition to LA monomer.
Figure S25. Mass spectra of lactic acid, glycine, alanine, and leucine (3:1:1:1 mol ratio) after 1, 4, and 8 cycles. Depsipeptides containing various mixtures of the three amino acids are observed, and depsipeptides have higher amino acid content as cycling increases.
Figure S26. Tandem MS sequencing of isobaric 2LA+1G+1A+1L depsipeptides (theo. parent ion [M-H]⁻ = 402.1877 Da). Formed after 8 cycles. Collision energy = 15 eV. One lactic acid residue is almost exclusively at the N-terminus, and the other lactic acid residue is either at the C-terminus or the residue adjacent to the C-terminus. The presence of LA-AA heterodimer sequences at the C-terminus is consistent with that in Figure SR and discussed in Section III (Proposed Mechanism). As a result, we conclude that up to 12 isobaric sequences could be present: (1) LA-Ala-Gly-Leu-LA, (2) LA-Ala-Gly-LA-Leu, (3) LA-Ala-Leu-Gly-LA, (4) LA-Ala-Leu-LA-Gly, (5) LA-Gly-Ala-Leu-LA, (6) LA-Gly-Ala-LA-Leu, (7) LA-Gly-Leu-Ala-LA, (8) LA-Gly-Leu-LA-Ala, (9) LA-Leu-Ala-Gly-LA, (10) LA-Leu-Ala-LA-Gly, (11) LA-Leu-Gly-Ala-LA, and (12) LA-Leu-Gly-LA-Ala. All potential combinations are shown.
Figure S27. Mass spectra of serine and lactic acid system after 1 and 8 cycles. # corresponds to loss of CH₂O from serine side chain in serine-lactic acid dimer, such that serine is converted to glycine in the gas-phase (i.e., this m/z value corresponds to 1LA+1G). This is a common fragmentation pathway for serine in negative-ion mode⁶ and typically corresponds to the base peak in tandem mass spectra of serine-lactic acid depsipeptides. & corresponds to trace background contaminant from previous experiments. After one cycle, serine-lactic acid depsipeptides are observed. However, after increased cycling, these products disappear. There are several potential explanations for this result, which are likely related to the potentially-reactive hydroxyl group on the side chain of serine. One potential explanation is that cyclic products form which stunt depsipeptide formation and amino acid enrichment. We do observe a signal with the same mass as 1LA+1A; it is possible this is a byproduct of larger cyclic species which are decomposing, but this has not yet been confirmed. Future studies will focus on serine as well as other polar side chain amino acids such as aspartic acid.
Figure S28. Tandem MS of serine-lactic acid depsipeptides after 1 cycle. The primary sequences observed are Ser-LA/LA-Ser and Ser-LA-Ser. Collision energy = 20 eV. Several notable differences are observed between tandem MS data of serine-lactic acid depsipeptides and those using other amino acids. One is the intense loss of the serine side chain, which is unsurprising (see Figure SU). Another, more surprising, observation is that serine is frequently present on the N-terminus of these depsipeptides. Typically, we observe lactic acid on the N-terminus of depsipeptides. This suggests a different mechanism to form polar-side-chain depsipeptides vs. nonpolar-side-chain depsipeptides. This is also consistent with the full mass spectra obtained, which show depsipeptides are not present after repeated cycling. We hypothesize that unstable ester linkages are formed with the serine side chain; however, further studies are needed to confirm this.
Figure S29. Mass spectra of 1:1 mol glycine-glycolic acid depsipeptides after 1 cycle and after 8 cycles. After 1 cycle, the majority of depsipeptide signals correspond to water-loss species, suggesting they may be cyclic instead of linear. However, after 8 cycles, it appears the depsipeptides preferentially adopt a linear structure over the water-loss species, similar to those formed using lactic acid. Glycolic acid is the simplest α-hydroxy acid with an amino acid analog and is found at similar abundance in prebiotic reaction mixtures\textsuperscript{11} and meteorites\textsuperscript{10} as lactic acid, glycine, and alanine. Based on this data, it appears amino acid incorporation may be somewhat more efficient using glycolic acid than lactic acid.
**Figure S30.** Tandem MS sequencing of glycine-glycolic acid heterodimer after 1 and 8 cycles (theo. parent ion [M-H] = 132.0298 Da). Collision energy = 8 eV. After 1 cycle, both the ester and amide-linked heterodimer are present. After 8 cycles, only the amide-linked heterodimer remains.
Figure S31. Depsipeptide formation does not appear to show a stereochemical preference. (top) Zoomed-in mass spectrum of L-alanine and L-lactic acid pentamers after 1 cycle; (middle) zoomed-in mass spectrum of D-alanine and L-lactic acid pentamers after 1 cycle; (bottom) zoomed-in mass spectrum of D-alanine and L-lactic acid pentamers after 4 cycles. As with L-alanine and L-lactic acid, depsipeptides are readily formed and are enriched with peptide sequences. The m/z 377.13 signal in the bottom spectrum corresponds to $^{13}$C-containing depsipeptides and not the 5LA oligoester.
VI. Supplemental References

1. Bowie, J. H., Brinkworth, C. S. & Dua, S. Collision-induced fragmentations of the (M-H)-- parent anions of underivatized peptides: An aid to structure determination and some unusual negative ion cleavages. *Mass Spectrom. Rev.* **2002**, 21, 87-107.

2. Bokatzian-Johnson, S. S., Stover, M. L., Dixon, D. A. & Cassady, C. J. Gas-phase deprotonation of the peptide backbone for tripeptides and their methyl esters with hydrogen and methyl side chains. *J. Phys. Chem. B* **2012**, 116, 14844-14858.

3. Parella, T. Pulse Program Catalogue, NMR Guide v4.0. *Bruker Biospin* 2004.

4. Orgel, L. E. The origin of polynucleotide-directed protein synthesis. *J. Mol. Evol.* **1989**, 29, 465-474.

5. Cleaves, H. J., Aubrey, A. D. & Bada, J. L. An evaluation of the critical parameters for abiotic peptide synthesis in submarine hydrothermal systems. *Orig. Life Evol. Biosph.* **2009**, 39, 109-126.

6. Danger, G., Plasson, R. & Pascal, R. Pathways for the formation and evolution of peptides in prebiotic environments. *Chem. Soc. Rev.* **2012**, 41, 5416-5429.

7. Kitadai, N. Thermodynamic prediction of glycine polymerization as a function of temperature and pH consistent with experimentally obtained results. *J. Mol. Evol.* **2014**, 78, 171-187.

8. Parker, E. T. *et al.* A plausible simultaneous synthesis of amino acids and simple peptides on the primordial earth. *Angew. Chem. Int. Ed.* **2014**, 53, 8132-8136.

9. Cronin, J. R. & Moore, C. B. Amino acid analyses of Murchison, Murray, and Allende carbonaceous chondrites. *Science* **1971**, 172, 1327-1329.

10. Peltzer, E. T. & Bada, J. L. Alpha-hydroxycarboxylic acids in Murchison meteorite. *Nature* **1978**, 272, 443-444.

11. Miller, S. L. & Urey, H. C. Organic compound synthesis on the primitive Earth. *Science* **1959**, 130, 245-251.