ARTICLE

Simple prebiotic synthesis of high diversity dynamic combinatorial polyester libraries

Kuhan Chandru[1], Nicholas Guttenberg[1], Chaitanya Giri[1], Yayoi Hongo[1], Christopher Butch[1], Irena Mamajanov[1] & H. James Cleaves II[1,2,3,4]

It is widely believed that the origin of life depended on environmentally driven complexification of abiotically produced organic compounds. Polymerization is one type of such complexification, and it may be important that many diverse polymer sequences be produced for the sake of selection. Not all compound classes are easily polymerized under the environmental conditions present on primitive planets, and it is possible that life's origin was aided by other monomers besides those used in contemporary biochemistry. Here we show that alpha-hydroxy acids, which are plausibly abundant prebiotic monomers, can be oligomerized to generate vast, likely sequence-complete libraries, which are also stable for significant amounts of time. This occurs over a variety of reaction conditions (temperature, concentration, salinity, and presence of congeners) compatible with geochemical settings on the primitive Earth and other solar system environments. The high-sequence heterogeneity achievable with these compounds may be useful for scaffolding the origin of life.

[1] Earth-Life Science Institute, Tokyo Institute of Technology, Tokyo 152-8551, Japan. [2] Blue Marble Space Institute of Science, Seattle 98154 WA, USA. [3] Institute for Advanced Study, Princeton, NJ 08540, USA. [4] Center for Chemical Evolution, Georgia Institute of Technology, Atlanta, GA 30332, USA. Correspondence and requests for materials should be addressed to H.J.C.II. (email: cleaves@elsi.jp)

The generation and selection of molecules with activities including catalysis and replication from random complex mixtures remains a major challenge in origin of life research[1–3]. This problem is intertwined with the question of the minimum polymer length necessary for catalytic activity, especially a replicase activity[4, 5] and what the frequency of catalysts within the sequence space of various types of polymer systems is[6].

Dynamic combinatorial chemistry (DCC) is a powerful method for generating novel molecules through the reversible reaction of building blocks. The term DCC usually refers to systems under thermodynamic control; however, systems under kinetic control generating diverse libraries have also been reported[7]. The resulting products are known as dynamic combinatorial libraries (DCLs)[7–10]. Biology produces sets of molecules resembling DCLs, perhaps the most obvious examples are organismal proteomes, in which amino acids are recombined into polypeptides via translation and proteolysis, or the antibodies of higher organisms[11], in which encoded genes can be expressed multiple ways to construct peptides from shared pools of amino acids. In both of these systems, the DCL's composition is genetically controlled; however, strong control of biopolymer structure may not have been a feature of the earliest evolvable chemical systems.

Biological catalysis is largely enabled by polymers, due to their ability to fold into discrete three-dimensional shapes that present functional groups in defined substrate-binding sites[12]. However, in the context of the origin of life, the abiotic synthesis of combinatorial libraries of oligomers sufficiently long to fold, without the intervention of enzymes, remains a longstanding challenge[13, 14].

Here we show that extremely complex DCLs spontaneously form from mixtures of presumably abundant prebiotic alpha-hydroxy acids (αHAs) under conditions likely to be present on many primitive planetary and pre-planetary bodies. Electrospray ionization–mass spectrometric (ESI-MS) analysis indicates that αHA DCLs can cover the entirety of sequence space up to remarkable linear and cyclic oligomer lengths, here measured up to m/z 2500, corresponding in some cases to 48-mers, but likely extending to longer species below the detection limit of the analysis conducted here. These polymers, though randomly and reversibly synthesized, can persist for significant periods depending on ambient conditions. The facile unactivated synthesis of these DCLs suggests that prebiotic and early biological systems may have had access to large swathes of catalytic organic polymer space. These results may be extendable to other monomer types. This work further expands previous investigations into polyester[15] and depsipeptide[16] formation from monomeric αHAs or mixtures of αHAs with α-amino acids (αAAs), and suggests

mild and acidic conditions, such as those which might occur in and around subaerial hot spring environments[17, 18], are conducive to combinatorial polymer formation on planetary surfaces, which may be a critical step for the origin of life.

Results

Oligomerization of glycolic acid and lactic acid. We used the αHAs shown in Fig. 1 for this study. These were chosen based on their simplicity, structural diversity, and commercial availability.

The oligomerization of 1 M total glycolic acid (GA) or DL-lactic acid (LA) alone or combined at various pH values (between 3 and 11) under drying conditions between 60 and 120 °C was first investigated. Results from these initial studies are shown in Fig. 2.

In Fig. 2, labeled peaks were detected as MNa^+ ions. Asterisks represent peaks showing water loss, indicating cyclic and/or acrylic acid-terminated oligomers. All reactions were initially 1 M total concentration, pH 3, and dried at 80 °C.

GA and LA oligomerized to a significant extent at all measured temperatures at their natural pH, and their products were easily identifiable within 3–5 ppm mass accuracy and intensity above of 5000. Multiple types of ionized adduct species were identified in positive ionization mode, including MH^+, MNa^+, MK^+, and MNH_4^+ in positive mode (see Supplementary Fig. 1). For GA oligomerization, this is evidenced by the Δ58.005 Da mass ladder (see Fig. 2b), which pervades the spectra (from m/z 50–1500), corresponding to the repeating $-(OCH_2CO)-$ unit, for example, up to 737.065 Da in the 100 °C spectrum, explicable as the MNa^+ 12-mer. The degree of polymerization increases with increasing temperature, though significant degradation is apparent above 100 °C, and quite pronounced at 120 °C, as evidenced by numerous unassignable ancillary peaks. Samples showed evidence for linear oligomers of GA up to the 16-mer (as its M+2Na-H adduct, m/z 991.071 Da) at 80 °C, and 18-mer (as its M+2K-H adduct, m/z 1139.030) at 60 °C.

The corresponding spectra for LA oligomerized at pH 3 and 100 °C (Fig. 2b) showed similar ladders with a Δ72.021 Da repeat, corresponding to the repeating $-(OCH(CH_3)CO)-$ unit, and extending to a peak at 1337.380 Da explicable as the MNa^+ linear 18-mer. Weaker intensity peaks assignable to higher oligomers in this series were also detected.

Also evident in the spectra collected from reactions of both GA and LA are series of peaks attributable to monodehydrated and monosodiated-monodehydrated oligomers, corresponding either to cyclic oligomers in the case of GA or cyclic and/or acrylic acid-terminated oligomers in the case of LA (Fig. 2a, b).

For reactions of GA or LA alone, the sparseness of the spectra allowed for easy MS/MS fragmentation of individual peaks, despite the relatively broad 3 Da isolation window. MS/MS
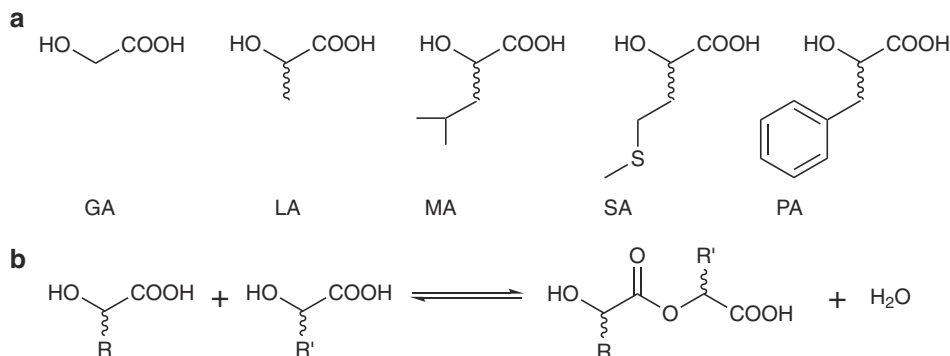

Fig. 1 αHAs used in this study and the general condensation reactions of polyesters. **a** Glycolic acid (GA), DL-lactic acid (LA), DL-2-hydroxy-4-methylpentanoic acid (MA), DL-2-hydroxy-4-(methylsulfanyl)butanoic acid (SA), and DL-3-phenyllactic acid (PA) and the αHA analogs of glycine, DL-alanine, DL-leucine, DL-methionine, and DL-phenylalanine, respectively. **b** General scheme for the condensation reaction of αHAs to give polyesters
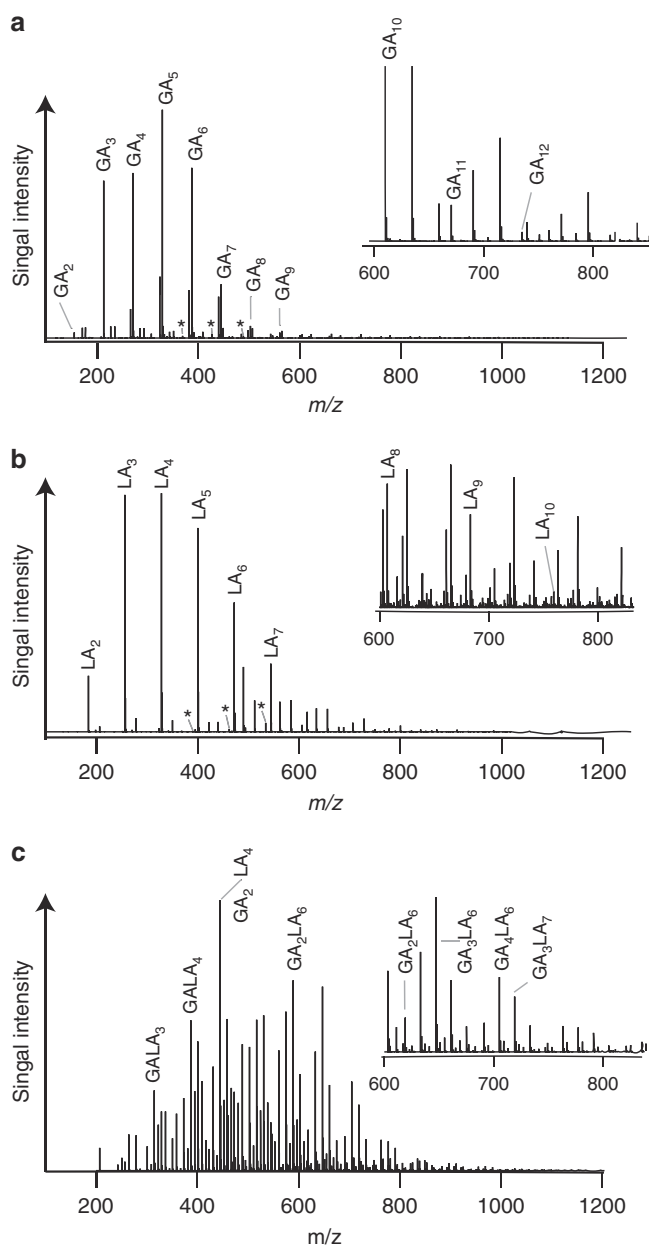
Fig. 2 Oligomerization of glycolic and lactic acid solutions. Positive mode ESI-QToF mass spectra of oligomeric products of **a** GA, **b** LA, and **c** GA +LA. Asterisks represent peaks showing water loss

analysis of isolated GA (data not shown) or LA oligomer peaks gave identical fragmentation patterns regardless of the precursor ion mass (e.g., a regular ~72 Da spacing, see Fig. 3a), suggesting that each is a regular linear polyester of GA or LA.

The mass of an internal residue of LA is 72.021 Da (indicated by long horizontal arrows in Fig. 3) and that of an internal residue of GA is 58.005 Da (indicated by short horizontal arrows). For Fig. 3a, since each identified oligomer fragment is composed of only LA monomers, only end-group mass losses are proposed. For Fig. 3b, each mass could represent multiple linear sequences, which could be the result of end-group loss in multiple ways. All reactions were initially 1 M in total concentration, pH 3, and dried at 80 °C in a single drying cycle.

GA+LA dried together yielded peaks consistent with being assignable to isobaric mixtures of variable sequence. Figure 3b shows the fragmentation spectrum of the peak assignable to

$GA_2LA_4Na^+$. Since it has been established that polyesters tend to fragment from their termini[19], this offers a way to establish the sequences of these oligomers. If the isolated peaks were composed of a mix of sequences, one would expect a complex fragmentation spectrum giving all possible fragments, which is expected and observed (see for example Supplementary Fig. 2).

Having explored these reactions under drying conditions down to 60 °C, the robustness of this chemistry at still lower temperatures was studied. Oligomers were formed even at temperatures as low as 30 °C, the lowest temperature explored, under drying conditions. Drying reactions were rehydrated every 24 h, but for reactions between 30 and 50 °C, there was little distinguishable difference between spectra measured after one drying cycle and those measured after two, or three where tested, suggesting that equilibrium is effectively reached during the first cycle in this temperature range (Supplementary Fig. 3).

These reactions were also examined as a function of pH under drying conditions. Experiments conducted at pH 5 and higher showed similar spectra, which merely suggested the formation of sodiated clusters (data not shown), although it was possible to confirm the formation of covalently linked oligomers at pH 3 using liquid chromatography–MS (Supplementary Fig. 4), consistent with ester formation being predominantly an acid-catalyzed process. Since the pK_a's of GA and LA are 3.6 and 3.8, respectively[20], deprotonation of these species at higher pH values makes them less electrophilic.

Though most of these experiments were examined synthesis under drying conditions, oligomerization of LA in solution at concentrations ranging from 10^{-3} to 10^{-1} M at 80 °C was also examined. Surprisingly, oligomers up to octamers ($m/z =$ 612.211 Da, (MNH_4^+)) were still measurable at 10^{-3} M after 5 h, though in evidently lower yield (Supplementary Fig. 5). This behavior is, to our knowledge, not observable for αAAs, for which the equilibrium for dimerization and elongation is not high under these conditions, e.g., a solution of glycine reacted and quasi-equilibrated at any pH and a 10^{-3} M dilution will not yield much dimer or cyclic dimer. After 72 h, however, the detected polymers converted into unidentifiable compounds.

While there are an unambiguous number of oligomers of GA, LA contains a stereocenter. Though ESI-MS can indicate oligomer chirality[21], it seems likely that mixtures of stereo-isomers are generated, though we did not attempt to measure chiral selection. There are 2^n possible stereoisomers for each oligomer of LA of length n, all of which are isobaric. For copolymers of GA and LA, there are 2^n possible sequences and 3^n possible stereoisomeric sequences.

MS (see Fig. 2c) and MS/MS (see Fig. 3b and Supplementary Fig. 6) analyses strongly suggest complete sequence coverage of the mixed GA+LA oligomer space up to ~1100 Da, which represents 155 unique singly charged (MH^+) parent mass species or 3,473,404 possible sequences. Scrutiny of the spectra argue for significant coverage of sequence space well beyond this mass range, for example, oligomers attributable to GA_3LA_{12} (MNa^+, m/z 1079.270. Eight of the possible 20 eicosamers of GA and LA were also easily detectable, including $GA_{20}NH_4^+$ (m/z 1196.161 measured, 1196.155 calculated), at 5 ppm error. Eventually detection becomes difficult due to the mass resolution and detection limit of the instrument.

MS/MS analysis of select peaks suggests that these are composed of the entire cohort of possible sequence permutations up to the resolution of analysis (see Fig. 3b). A fragmentation scheme accounting for the mass losses shown in Fig. 3b for the parent mass assigned to a $GA_2LA_4Na^+$ oligomer adduct is shown in Supplementary Fig. 6. There are 15 sequence permutations for GA_2LA_4 oligomers, all of which are isobaric and thus the isolated peak could represent one or more, and possibly all, of these
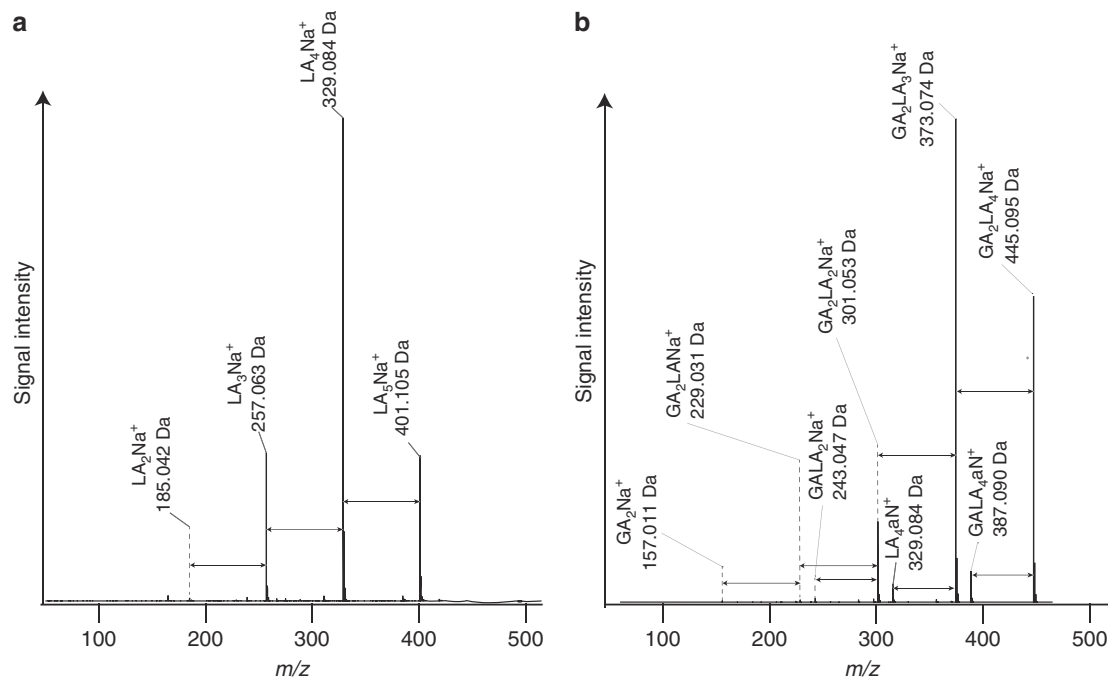
a

b

Fig. 3 Fragmentation analysis. Positive-mode ESI-QToF-MS/MS fragmentation spectra of selected oligomer peaks from the drying reaction of **a** LA and **b** GA+LA
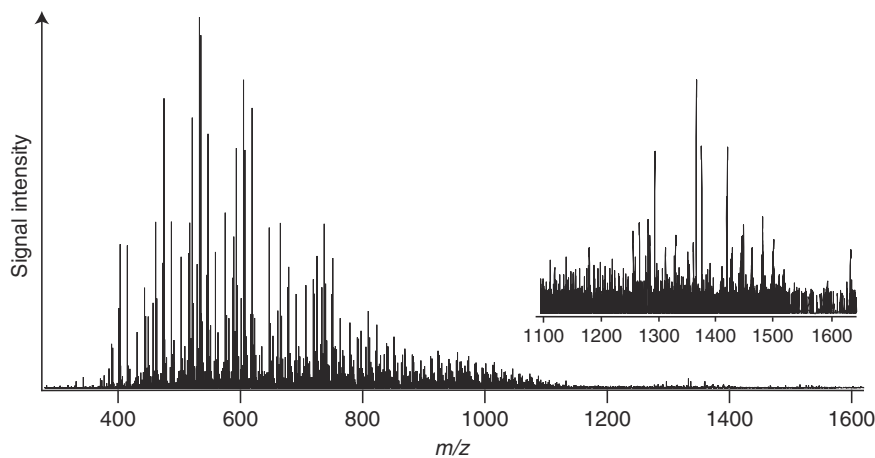

Fig. 4 Positive-mode FT-ICR mass spectrum of five αHA mix. The inset shows magnification of the *m/z* 1100–1650 range, indicating the density of detected masses

sequences. However, some of these have a GA residue at their COOH termini and others have an LA residue, thus initial end-terminal fragmentation should lead to a precursor ion-minus-58 Da peak or a precursor ion-minus-72 Da peak, and this pattern should propagate during fragmentation to give an increasingly complex fragmentation spectrum (as opposed to the fairly regular fragmentation observed in Fig. 3a), which is observed.

Oligomerization of five αHA mix. This concept was extended even further to examine a system containing all five αHAs (see Fig. 1). Isobaric αHA's or those with multiple condensable functional groups were not employed to simplify analysis and avoid forming branched oligomers. In this system, an enormous variety of mixed oligomer sequences are possible and high-resolution Fourier transform ion cyclotron resonance MS (FT-ICR-MS) (Fig. 4) and MS/MS (Fig. 5) analysis supports the notion that a large proportion, if not all, possible sequences are formed (~4300 unique mass peaks between *m/z* 210 and 1400

were detected, a range which theoretically includes 10,450 unique linear 2–20-mer permutations, with a very small duplication of exact mass), until the physical capacity of the system to generate them and the ability of our analytical methods to detect them are exhausted.

Figure 4 shows a mass spectrum of the products of a mixture of 0.15 M each GA, LA, DL-2-hydroxy-4-methylpentanoic acid (MA), dl-2-hydroxy-4-(methylsulfanyl)butanoic acid (SA) and dl-3-phenyllactic acid (PA) dried at 80 °C for 48 h.

It was often possible to make more than one permutation assignment within a 1 ppm mass accuracy limit. However, it was possible to isolate several peaks for MS/MS analysis, which helped suggest assignments (Fig. 5).

MS/MS experiments allow for some inquiry into the composition of the most intense isolated peaks. For example, MS/MS analysis was conducted on a major peak detected in the positive ESI spectrum at *m/z* ~809.2 by FT-ICR-MS. Allowing for an ~1 Da isolation window (*m/z* ~808.7519–809.7519), there are 19 measured peaks above the noise threshold that could potentially
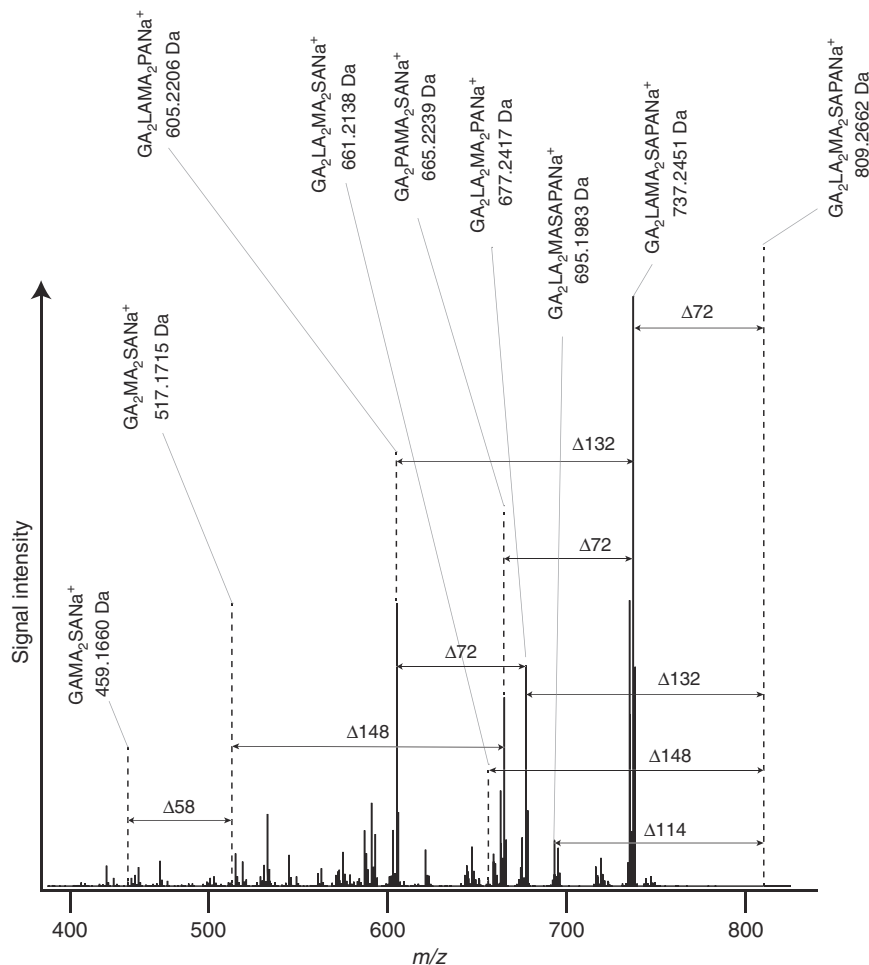
Fig. 5 Assignment of 809.2662 Da peak. Positive-mode FT-ICR-MS/MS analysis of the 809.2662 Da peak detected from the oligomers produced from the five αHA mix

contribute to the measured fragmentation spectrum, though 6 peaks contribute >55% of the peak intensity in this window.

There were two masses that could be assigned to the major peak within experimental error, 809.25186 and 809.26622. These were potentially assignable to either the octamers $GA_2LA_2MA_2$-SAPA or the heptamers $LA_2MA_2SA_3$. MS/MS fragmentation, however, suggests the dominant presence of the octamers by the ability to assign >99% of the possible fragment products out to four sequential single end-terminal residue losses and major assignable losses of single residues of GA and PA. The repeating nature of the oligomers is also supported by modified Kendrick mass defect plots in positive ionization mode (Supplementary Fig. 7).

Supplementary Table 1 shows the composition of the most intense 217 MNa^+ ions identified. Notably, SA and PA appear under represented. This may be because these monomers are incorporated less efficiently or because oligomers containing them volatilize or ionize less efficiently, especially as singly charged sodiated adducts. Numerous peaks assignable as oligomers containing multiple residues of SA and/or PA as multiply charged species were observed.

Discussion

It is often proposed that the origin of life depended on oligomer-mediated replication[1, 2, 4–6], and these oligomers were shorter than modern biopolymers (for example, average bacterial-coded proteins are estimated to be on the order of 200–270 amino acid residues[22, 23]). The types of chemistry that could have facilitated early replication and translation-like activities are unknown, but it seems likely the first replicating systems had weaker control over the composition of the polymers they were capable of producing and were more error prone. The first nucleic acid-based replicases were only barely able to surpass the Eigen error threshold[24], and the first translation systems only produced something like "statistical proteins"[25] until the translation apparatus became streamlined and integrated. Polyesters based on αHAs may have preceded this integration and that proto-organisms capable of producing or sequestering αHAs from the environment automatically gained possession of a reversible suite of polyesters of considerable length, possibly invested with catalytic potential.

αHAs are likely to be common reactants in many primitive solar system environments, formed from the same mechanisms which form gave rise to αAAs, for example, the Strecker-cyanohydrin synthesis[26, 27]. Importantly, depending on pH and the concentration of free ammonia, αHAs are produced more efficiently than αAAs[26]. Measurements from a variety of prebiotic simulations and natural sources such as meteorites strongly suggest that αHAs and αAAs are concomitantly produced in abiotic contexts and that the lowest molecular weight species are generally the most abundant[27–29], suggesting glycine and GA would tend to be the predominant reactants, followed by lesser amounts of higher homologs[30]. Furthermore, αHAs are likely more stable than β-hydroxy acids, which readily eliminate water to yield reactive α,β-unsaturated carbonyl compounds, suggesting

that some further sources of molecular selection may have operated in the environment.

Although the synthesis of the libraries presented here appears to require relatively low pH, considerable uncertainty still exists regarding the conditions that prevailed on the prebiotic Earth[31], and various microenvironments may have existed, and that certain steps in the origin of life may have occurred off-Earth[32]. Polyesters are reasonably stable in the acid-to-neutral pH region[33], though they may show marked degradation at higher pH values[34]. The conditions explored here are remarkably consistent with those suggested to be compatible with the oligomerization of nucleic acid monomers in the presence of lipids in hot spring-like environments[17].

Even best-case estimates for endogenous organic synthesis or exogenous organic delivery to the early Earth would have produced fairly dilute (mM or lower) bulk ocean concentrations of organic compounds[35, 36]. In principle, dehydration could begin from any arbitrary dilution, though it is likely the temperature, congeners, pH, and rate of dehydration have some bearing on the overall yield. It is shown here that this is the case for αHAs but that the oligomerization chemistry is remarkably robust.

A variety of studies have examined the polycondensation of amino acids under dry-heating conditions and in solution[14, 37–39]. In fact, long oligomers can be produced, especially at conditions well above the boiling point of water. In contrast with αAAs, αHAs yield remarkably longer oligomers in higher yield at much lower temperatures. This is explicable by the significantly lower ΔG of formation of an ester vs. a peptide bond (~0 kcal mol^{-1} vs +3.5 kcal mol^{-1}, corresponding to dimerization equilibria of ~1 for esterification and only 2.7×10^{-3} for peptide bond formation at 25 °C[34, 38]). αHAs are unlike AAs in a number of ways. For example, their racemization kinetics, stability of polymer bonds, and ability to form regular H-bond-stabilized secondary structural motifs, such as α-helices and β-sheets[40, 41].

Polyester and polypeptide degradation rates depend on many parameters, including stereochemistry, polymer length, composition, the presence of co-solutes, pH, and temperature. Hydrolysis kinetics for peptides and polyesters are incompletely measured for simple comparison across all possibly relevant conditions, and there may be multiple mechanisms that govern the equilibria for polymer formation and degradation, which differ for polypeptides and poly-αHAs under specific conditions.

The half-life of diglycine in water has been reported as approximately 350 years at pH 7 and 25 °C[42] (alternative values of ~7 years have also been reported[43]). The hydrolysis kinetics of DL-LA 4-18-mers have been studied[44], though not over the range of conditions reported for peptide dimers related above. Nevertheless, extrapolation of the reported hydrolysis rates at 25 °C suggests the half-life would be ~3.5 years at pH 7.4, and ~345 years at pH 4.5 (which was found to be the region of their greatest stability). Peptide hydrolysis rates are approximately equal between pH 4.5 and 10 at 37 °C[45]; for a dipeptide at pH 4.5 and 37 °C, a hydrolysis rate of ~10^{-10} s^{-1} was measured, corresponding to a half-life of ~22 years. Thus there could be conditions of temperature and pH, among other parameters, under which one or the other type of polymer is more stable. The two polymer systems could be comparably stable under certain conditions, but it is unlikely they have similar degrees of polymerization under many conditions when polymerization is not driven by disequilibrium and that elongation is more robust for polyesters at lower pH and higher for polypeptides near neutral pH.

Monomer kinetic and thermodynamic polymerizability depends on numerous equilibria, including the simple equilibrium between monomer, dimer, and cyclic dimer, which may bottleneck elongation in the case of αAAs but does not appear to hamper αHA elongation under the conditions explored here. Peptides show complex dimerization and cyclization behaviors depending on concentration, temperature, and pH (see, for example, refs. [14, 38, 39]); the same is true for polyesters, though this has not been well explored from the standpoint of abiotic oligomer synthesis. These simple reaction equilibria are of extreme interest to astrobiology and warrant more detailed investigation. We hope this study will motivate exploration of this space so that direct comparison will be possible.

αHA dimers can cyclize; however, owing to various steric and energetic factors, 2,5-diketo-1,4-dioxanes are much more reactive than diketopiperazines (DKPs) to oligomerization under the appropriate conditions. Cyclic αHA dimers are known to form polyesters[46], and indeed these form the starting point for industrial αHA polymer synthesis. In contrast, temperatures above ~110–120 °C are often needed to drive peptide synthesis from DKPs[38, 39]. It is noteworthy that Fig. 2 and Supplementary Figs. 1, 3 and 5 do not exhibit major peaks from glycolide (nominal neutral mass 116), 2,4-diketo-3-methyl-1,4-dioxane (nominal neutral mass 130), or lactide (nominal neutral mass 144), suggesting that these are not significant products. In contrast, DKPs are often a significant fraction of thermally oligomerized αAA products[16, 38, 39].

A complete comparison of where one or the other is "more oligomerizable" under a given set of conditions is beyond the scope of this manuscript, but evidence indicates that the αHA cyclic dimers are more reactive than DKPs.

As αHA oligomerization is thermodynamically and kinetically favorable under these conditions and does not appear to highly discriminate against the incorporation of different side-chains (though the apparent dearth of sequences with more than one residue of SA or PA is notable, which may also be due to volatilization or ionization problems), enormously diverse libraries can form very rapidly. For the five αHA mix experiments, there are 5^{20} (~9.5×10^{13}) possible eicosameric sequences and the same number of unique mass oligomers, ignoring stereoisomers, or ~1.2×10^{19} oligomer enantiomers. Notably, previous research has shown that racemic poly-LA hydrolyzes faster than its enantiopure counterparts at neutral pH and 37 °C[47], which could suggest a simple physical mechanism for the preferential persistence of enantiopure oligomers.

MS evidence suggests that the entire suite of 20-mer sequences is likely represented in one of these samples, along with the concomitant suite of shorter oligomers, as well as an undetermined but likely very high coverage of longer species. Although not fully characterized, this would make this one of the largest combinatorial libraries created to date (see, for example, ref. [48]). Though quantification of these oligomers is not possible using this technique, the fact that they are detectable implies that the yield for each oligomerization step is quite high. The apparent fine-scale variation in MS signal intensity may imply that stochastic effects come into play in determining the ultimate distribution of oligomers.

The extreme sequence diversity generated in these αHA DCLs begs three questions: (1) Do these experiments generate the entire possible sequence diversity of these libraries? (2) Can the sequence diversity generated be detected using these analytical methods? and (3) To what extent would uncontrolled oligomerization allow sequence exploration in a protocellular cytosolic environment? A typical *Escherichia coli* cell contains individual αHAs (e.g., GA, LA, and malic acid) at a concentration of perhaps 1–2 mM, though these levels are tightly regulated by their flux through various enzymatically controlled metabolic pathways[49]. A simple model is constructed here to examine how much oligomeric polyester diversity could be generated in a proto-cell-like compartment given similar monomer concentrations. Supplementary

Table 2 shows the total number of sequences possible for each library (e.g., GA, LA, GA+LA, and the five αHA mix) and the extent of sequence coverage attained in a model 1-μm internal diameter spherical proto-cellular space ($\sim 5 \times 10^{-16}$ L) at equilibrium, assuming 10 mM total starting concentrations of αHAs and a Flory distribution of oligomers at equilibrium at 25 °C. At equilibrium, oligomers as long as 22-mers are present. For systems containing only one type of αHA, complete sequence coverage is attained, but this is only true up to 11-mers for two-αHA systems and only for pentamers for five αHA systems. Beyond these lengths, the sequence diversity is only stochastically populated.

Given the remarkable facility with which αHAs oligomerize, the fact that they have been measured to be indigenous in numerous carbonaceous chondrites, generally along with their cognate αAAs[50], suggests that they may have been efficiently delivered to the primitive Earth's surface.

We also note a possible implication of these results for biochemical evolution. It remains highly contentious when life arose on Earth, and how the pH and temperature of the oceans changed as a function of time during Earth's earliest evolution[31]. Beyond this, it is uncertain whether life arose in rare niche environments with parameters well outside of those estimated for global averages[17]. Thus it is possible that αHA oligomers preceded αAA oligomers as compounds that enabled the chemical complexification that led to life, especially if some stages of chemical evolution occurred under relatively acid conditions, due to both their higher propensity of αHAs to oligomerize and their relatively higher stability under acidic conditions.

In summary, αHAs polymerize easily to give high molecular weight polymers under simple low temperature drying and aqueous conditions, and near complete sequence libraries appear to be obtainable. While these polyesters cannot form the regular secondary structures of polypeptides due their inability to form intra-molecular H-bonds[40], they may nevertheless form unique folded structures. Furthermore, it may be that a variety of other side-chain functional groups could be incorporated into αHAs, including those mimicking nucleic acids, allowing the generation of novel high-diversity libraries that are capable of being sequenced. This chemistry is presently being studied.

Methods

Materials. GA, LA, MA, SA, and DL-3-phenyllactic acid (PA), (Fig. 1) were purchased from Sigma.

Experimental procedures. All glassware was pre-ashed at 500 °C for 3 h to eliminate organic contamination. All water was from a Milli-Q *Integral 3* Water Purification system and was of 18.2 MΩ conductivity at 25 °C and contained 3 ppb TOC.

Experiments were conducted in open borosilicate test tubes or sealed borosilicate ampoules, depending on the nature of the experiment, and under air in both cases. pH was adjusted using 1 N NaOH and measured using Sigma-Aldrich Hydrion Brilliant disposable pH sticks, which are accurate to ±0.5 pH unit. Reactions were held at constant temperature (±0.1 °C) using *Sahara 310* dry heating baths as monitored by conventional liquid thermometers.

Five αHA mix reactions were initially 100 μL in volume and 0.15 M in each αHA and were adjusted to pH 3–11 using dilute aqueous HCl or NaOH. These were allowed to dry at 60–100 °C and rehydrated by addition of the appropriate amount of water every 24 h.

Electrospray ionization mass spectrometry. Samples were diluted 1000× with water prior to MS analysis. ESI quadrupole time-of-flight MS (ESI-QToF-MS) analysis was carried out by direct infusion at a flow rate of 0.4 mL min^{-1} using a Waters Xevo G2-XS QToF-MS operated in positive or negative mode. Source settings were as follows: positive and negative-mode: ion source temperature 150 °C, desolvation gas temperature 550 °C, cone voltage 20 V, cone gas flow rate 50 L h^{-1}, and desolvation gas flow rate 1000 L h^{-1}. The capillary voltage was 1.2 kV in negative mode and 1.0 kV in positive mode.

For MS/MS experiments, collision-induced dissociation energies were set to 6 eV and the isolation window was ±3 Da. Calibration for experiments examining GA or LA separately was conducted using sodium formate over m/z 50–1500. Calibration for experiments examining GA and LA together was conducted using

sodium trifluoroacetate over m/z 200–2500. A water blank was injected every five injections to guard against bleed between injections.

For FT-ICR-MS analysis, samples were analyzed on a custom-built 14.5 T FT-ICR mass spectrometer by positive- and negative-mode ESI. One milliliter of MeOH (HPLC grade, JT Baker) was added to each dried sample vial, sonicated, and centrifuged. Samples were prepared for analysis by negative-mode ESI by a 1:1 dilution with "spray juice". Negative-mode spray juice was a mixture of MeOH with 2% ammonium hydroxide solution (28% in water, Sigma-Aldrich, St. Louis, MO). Positive-mode spray juice was a mixture of MeOH with 2% formic acid (Fluka Analytical).

Data analysis. Data collection was facilitated by a modular ICR data acquisition system (PREDATOR)[51]. One hundred individual time-domain transients were co-added, Hanning-apodized, zero-filled, and fast Fourier transformed prior to frequency conversion to mass-to-charge ratio to obtain the final mass spectrum. Data were analyzed and peak lists were generated with the custom-built MIDAS software[52]. For FT-ICR-MS/MS experiments, collision-induced dissociation energies were set to 45 eV (positive mode) or 35 eV (negative mode) and the isolation window was ±1 Da.

MS spectra were parsed, and peak identifications were made using a custom Python script using the NumPy scientific package. The software ran on a 1.7 GHz Intel Core i7 CPU of a Mac OS El Capitan. Measured peaks were sorted using the script based on a calculated mass table with common adducts[19, 53, 54] and permutation tables generated with Python. QToF-MS data were generally deemed to be accurate within <5 ppm and FT-ICR-MS data within <1 ppm[55].

Data availability. The code, permutation tables, and calculated compound identification spreadsheet for this data analysis are available (Supplementary Data 1). Other relevant source data are available from the corresponding author upon reasonable request.

References

1. Orgel, L. E. Molecular replication. *Nature* **358**, 203–209 (1992).
2. Szathmáry, E. & Smith, J. M. From replicators to reproducers: the first major transitions leading to life. *J. Theor. Biol.* **187**, 555–571 (1997).
3. Ross, D. S. & Deamer, D. Dry/wet cycling and the thermodynamics and kinetics of prebiotic polymer synthesis. *Life* **6**, 28 (2016).
4. Pinheiro, V. B. & Holliger, P. The XNA world: progress towards replication and evolution of synthetic genetic polymers. *Curr. Opin. Chem. Biol.* **16**, 245–252 (2012).
5. Wochner, A., Attwater, J., Coulson, A. & Holliger, P. Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332**, 209–212 (2011).
6. Jiménez, J. I., Xulvi-Brunet, R., Campbell, G. W., Turk-MacLeod, R. & Chen, I. A. Comprehensive experimental fitness landscape and evolutionary network for small RNA. *Proc. Natl. Acad. Sci. USA* **110**, 14984–14989 (2013).
7. Reek, J. H. R. & Otto, S. *Dynamic Combinatorial Chemistry* (Weinheim, Germany, Wiley, 2010).
8. Lehn, J. M. Dynamic combinatorial chemistry and virtual combinatorial libraries. *Chem. Eur. J.* **5**, 2455–2463 (1999).
9. Ji, Q., Lirag, R. C. & Miljanic, O. S. Kinetically controlled phenomena in dynamic combinatorial libraries. *Chem. Soc. Rev.* **43**, 1873–1884 (2014).
10. Corbett, P. T. et al. Dynamic combinatorial chemistry. *Chem. Rev.* **106**, 3652–3711 (2006).
11. Tonegawa, S. Somatic generation of antibody diversity. *Nature* **302**, 575–581 (1983).
12. Miller, B. G. & Wolfenden, R. Catalytic proficiency: the unusual case of OMP decarboxylase. *Ann. Rev. Biochem.* **71**, 847–885 (2002).
13. Benner, S. A., Kim, H.-J. & Carrigan, M. A. Asphalt, water, and the prebiotic synthesis of ribose, ribonucleosides, and RNA. *Acc. Chem. Res.* **45**, 2025–2034 (2012).
14. 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.* **39**, 109–126 (2009).
15. Mamajanov, I. et al. Ester formation and hydrolysis during wet–dry cycles: generation of far-from-equilibrium polymers in a model prebiotic reaction. *Macromolecules* **47**, 1334–1343 (2014).
16. Forsythe, J. G. et al. Ester-mediated amide bond formation driven by wet–dry cycles: a possible path to polypeptides on the prebiotic Earth. *Angew. Chem. Int Ed* **54**, 9871–9875 (2015).
17. Damer, B. & Deamer, D. Coupled phases and combinatorial selection in fluctuating hydrothermal pools: a scenario to guide experimental approaches to the origin of cellular life. *Life* **5**, 872–887 (2015).

18. Mulkidjanian, A. Y., Bychkov, A. Y., Dibrova, D. V., Galperin, M. Y. & Koonin, E. V. Origin of first cells at terrestrial, anoxic geothermal fields. *Proc. Natl. Acad. Sci.* **109**, E821–E830 (2012).

19. Osaka, I., Watanabe, M., Takama, M., Murakami, M. & Arakawa, R. Characterization of linear and cyclic polylactic acids and their solvolysis products by electrospray ionization mass spectrometry. *J. Mass. Spectrom.* **41**, 1369–1377 (2006).

20. Budavari, S., O'Neil, M., Smith, A., Heckelman, P. & Obenchain, J. *The Merck Index, Twelfth Edition.* (Taylor & Francis, London, Washington DC, 1996).

21. Yu, X.-Y. & Yao, Z.-P. Chiral recognition and determination of enantiomeric excess by mass spectrometry: A review. *Anal. Chim. Acta* **968**, 1–20 (2017).

22. Brocchieri, L. & Karlin, S. Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res.* **33**, 3390–3400 (2005).

23. Skovgaard, M., Jensen, L. J., Brunak, S., Ussery, D. & Krogh, A. On the total number of genes and their length distribution in complete microbial genomes. *Trends Genet.* **17**, 425–428 (2001).

24. Eigen, M. Self-organization of matter and the evolution of biological macromolecules. *Naturwiss* **58**, 465–523 (1971).

25. 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.* **261**, 531–539 (2009).

26. Peltzer, E. T., Bada, J. L., Schlesinger, G. & Miller, S. L. The chemical conditions on the parent body of the Murchison meteorite: some conclusions based on amino, hydroxy and dicarboxylic acids. *Adv. Space Res.* **4**, 69–74 (1984).

27. Miller, S. L. The mechanism of synthesis of amino acids by electric discharges. *Biochim. Biophys. Acta* **23**, 480–489 (1957).

28. Wolman, Y., Haverland, W. J. & Miller, S. L. Nonprotein amino acids from spark discharges and their comparison with the Murchison meteorite amino acids. *Proc. Natl. Acad. Sci. USA* **69**, 809–811 (1972).

29. Johnson, A. P. et al. The Miller volcanic spark discharge experiment. *Science* **322**, 404–404 (2008).

30. Parker, E. T., Cleaves, H. J., Bada, J. L. & Fernández, F. M. Quantitation of α-hydroxy acids in complex prebiotic mixtures via liquid chromatography-tandem mass spectrometry. *Rapid Comm. Mass Spectrom.* **30**, 2043–2051 (2016).

31. Cleaves, H. J. Prebiotic chemistry: geochemical context and reaction screening. *Life* **3**, 331–345 (2013).

32. Pizzarello, S. & Shock, E. The organic composition of carbonaceous meteorites: the evolutionary story ahead of biochemistry. *Cold Spring Harb. Perspect. Biol.* **2**, a002105 (2010).

33. Wolfenden, R. Degrees of difficulty of water-consuming reactions in the absence of enzymes. *Chem. Rev.* **106**, 3379–3396 (2006).

34. Compton, R. G., Bamford, C. H. & Tipper†, C. F. H. *Ester Formation and Hydrolysis and Related Reactions* (Elsevier Science, Amsterdam, New York, 1972).

35. Schlesinger, G. & Miller, S. L. Prebiotic synthesis in atmospheres containing CH_4, CO, and CO_2. I. Amino acids. *J. Mol. Evol.* **19**, 376–382 (1983).

36. Chyba, C. & Sagan, C. Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. *Nature* **355**, 125–132 (1992).

37. Fox, S. W. & Harada, K. Thermal copolymerization of amino acids to a product resembling protein. *Science* **128**, 1214 (1958).

38. Meggy, A. B. The free energy of formation of the amide bond in polyamides. *J. Appl. Chem.* **4**, 154–159 (1954).

39. Rodriguez-Garcia, M. et al. Formation of oligopeptides in high yield under simple programmable conditions. *Nat. Commun.* **6**, 8385 (2015).

40. Cleaves, H. J. The origin of the biologically coded amino acids. *J. Theor. Biol.* **263**, 490–498 (2010).

41. Rich, A. in *Chemical Evolution and the Origin of Life* (eds Buvet, R. & Ponnamperuma, C.) 180–196 (North-Holland, Amsterdam, 1971).

42. Radzicka, A. & Wolfenden, R. Rates of uncatalyzed peptide bond hydrolysis in neutral solution and the transition state affinities of proteases. *J. Am. Chem. Soc.* **118**, 6105–6109 (1996).

43. Kahne, D. & Still, M. W. C. Hydrolysis of a peptide bond in neutral water. *J. Am. Chem. Soc.* **110**, 7529–7534 (1988).

44. Schliecker, G., Schmidt, C., Fuchs, S. & Kissel, T. Characterization of a homologous series of D, L-lactic acid oligomers; a mechanistic study on the degradation kinetics in vitro. *Biomaterials* **24**, 3835–3844 (2003).

45. Smith, R. M. & Hansen, D. E. The pH-rate profile for the hydrolysis of a peptide bond. *J. Am. Chem. Soc.* **120**, 8910–8913 (1998).

46. Dechy-Cabaret, O., Martin-Vaca, B. & Bourissou, D. Controlled ring-opening polymerization of lactide and glycolide. *Chem. Rev.* **104**, 6147–6176 (2004).

47. Fukuzaki, H., Yoshida, M., Asano, M. & Kumakura, M. Synthesis of copoly(d, l-lactic acid) with relatively low molecular weight and in vitro degradation. *Eur. Pol. J.* **25**, 1019–1026 (1989).

48. Halford, B. How DNA-encoded libraries are revolutionizing drug discovery. *Chem. Eng. News* **95**, 28–33 (2017).

49. Bennett, B. D. et al. Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* **5**, 593–599 (2009).

50. Pizzarello, S., Wang, Y. & Chaban, G. M. A comparative study of the hydroxy acids from the Murchison, GRA 95229 and LAP 02342 meteorites. *Geochim. Cosmochim. Acta* **74**, 6206–6217 (2010).

51. Blakney, G. T., Hendrickson, C. L. & Marshall, A. G. Predator data station: a fast data acquisition system for advanced FT-ICR MS experiments. *Int. J. Mass. Spectrom.* **306**, 246–252 (2011).

52. Senko, M. W., Canterbury, J. D., Guan, S. & Marshall, A. G. A high-performance modular data system for Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Comm. Mass Spectrom.* **10**, 1839–1844 (1996).

53. Huang, N., Siegel, M. M., Kruppa, G. H. & Laukien, F. H. Automation of a Fourier transform ion cyclotron resonance mass spectrometer for acquisition, analysis, and e-mailing of high-resolution exact-mass electrospray ionization mass spectral data. *J. Am. Soc. Mass. Spectrom.* **10**, 1166–1173 (1999).

54. Keller, B. O., Sui, J., Young, A. B. & Whittal, R. M. Interferences and contaminants encountered in modern mass spectrometry. *Anal. Chim. Acta* **627**, 71–81 (2008).

55. Marshall, A. G. & Hendrickson, C. L. High-resolution mass spectrometers. *Ann. Rev. Anal. Chem.* **1**, 579–599 (2008).

Acknowledgements

Author contributions

K.C. and H.J.C. wrote the manuscript and conducted the experiments. K.C., H.J.C., and Y.H. performed the ToF-MS analysis. All the authors performed the data analysis.

Additional information