Prebiotic Chemistry Hot Paper

International Edition: DOI: 10.1002/anie.201605321
German Edition: DOI: 10.1002/ange.201605321

Amyloid Aggregates Arise from Amino Acid Condensations under Prebiotic Conditions

Jason Greenwald, Michael P. Friedmann, and Roland Riek**

Abstract: *Current theories on the origin of life reveal significant gaps in our understanding of the mechanisms that allowed simple chemical precursors to coalesce into the complex polymers that are needed to sustain life. The volcanic gas carbonyl sulfide (COS) is known to catalyze the condensation of amino acids under aqueous conditions, but the reported di-, tri-, and tetra-peptides are too short to support a regular tertiary structure. Here, we demonstrate that alanine and valine, two of the proteinogenic amino acids believed to have been among the most abundant on a prebiotic earth, can polymerize into peptides and subsequently assemble into ordered amyloid fibers comprising a cross-β-sheet quaternary structure following COS-activated continuous polymerization of as little as 1 mM amino acid. Furthermore, this spontaneous assembly is not limited to pure amino acids, since mixtures of glycine, alanine, aspartate, and valine yield similar structures.*

The lack of a genetic record beyond that of a hypothetical last common ancestor means that little can be known about the prebiotic peptides that preceded the proteins of life. We can safely assume that they must have existed and that they were in many ways simpler than those that existed in the first living systems. However, it is still not clear how even simple peptides could have accumulated to a significant extent on the early earth and by what mechanisms they could have attained the complexity, including the formation of tertiary and quaternary structures, that is required to support the functions on which life depends. This open question has led us and others to suggest that amyloids may have played an important role in the early evolution of proteins.[1–4] This so-called "amyloid world" hypothesis has several interesting implications for the stability, activity, and replicative potential of short polypeptides. However, there is no experimental evidence to date that amyloids can be a significant outcome of condensation of amino acids under prebiotic conditions. We therefore set out to test the plausibility of the amyloid as a prebiotic entity.

Alanine was selected for the initial investigations because it is the simplest of the chiral α-amino acids and thus arguably the most likely to have been present in the prebiotic world. The first challenge was to create polymers of sufficient length to induce β-aggregation. Of the several proposed prebiotic syntheses of peptides from amino acids,[5–9] we chose to work with the volcanic gas carbonyl sulfide (COS). Previous work has revealed that the rate of the COS-induced condensation of amino acids is greatly enhanced by acylating and oxidizing agents. This enhancement was explained by the fact that these molecules can modify the sulfhydryl group of the thiocarbamate, thereby creating a better leaving group for the rate-limiting cyclization to the *N*-carboxyanhydride.[6] For this reason, we used an excess of potassium ferricyanide $K_3Fe(CN)_6$ in all of the COS-mediated condensation reactions.

Whereas the reported yields of di- and tri-peptides from the COS-activated polymerization of 50 mM phenylalanine are promising, our aim of generating aggregated polyalanine peptides necessitated the production of significantly longer peptides.[10] While simply increasing the amino acid concentration may have increased the yield of longer peptides, we reasoned that a more realistic prebiotic scenario would be a steady supply of a low concentration of an amino acid thiocarbamate. To emulate such a continuous low-concentration reaction, we performed the polymerization of L-alanine (Ala) by treating a solution of amino acid in borate buffer at pH 9.2 with an excess of COS gas and then adding the COS-treated solution, either stepwise every few minutes or continuously, into a separate reaction containing an excess of potassium ferricyanide. The peptide products formed this way are longer and occur in higher yield than from a stochiometrically equivalent reaction carried out with a single mixing of reactants. The increase in the concentration of soluble polymers (length = 2–6) upon each 1 mL addition in a 10 mM Ala polymerization reaction was monitored by reverse-phase chromatography of the reaction supernatant (Figure 1a,b) and the length of the polymers in the insoluble fraction was monitored by matrix-assisted laser desorption/ionization Fourier-transform ion cyclotron resonance mass spectrometry (MALDI-FT-ICR MS; Figure 1c,d).

The peptide yield is further enhanced in a more continuous reaction through the dropwise addition of COS-treated Ala. In such a reaction with 20 mM Ala, a steady-state concentration of short, soluble polymers is quickly established. That is, within the first 1 mL (ca. 20 drops) of reaction volume, the rate that new polymers appear, the rate at which they elongate, and the rate at which they finally precipitate are approximately equal. The yield of insoluble peptides increases quickly at first, then levels off as the reaction volume increases (Figure 1e). The yield of Ala incorporation into longer, insoluble polymers (6–15-mers) reaches 34% of the total amino acid. In contrast, the yield of peptides in the precipitate of the single-step reaction is below the detection limit for the assay, equivalent to less than 1%. Throughout the

[*] Dr. J. Greenwald, M. P. Friedmann, Prof. R. Riek
Laboratory of Physical Chemistry, D-CHAB, ETH Zürich
Vladimir-Prelog-Weg 2, 8093 Zürich (Switzerland)
E-mail: jason.greenwald@phys.chem.ethz.ch
roland.riek@phys.chem.ethz.ch

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: http://dx.doi.org/10.1002/anie.201605321.

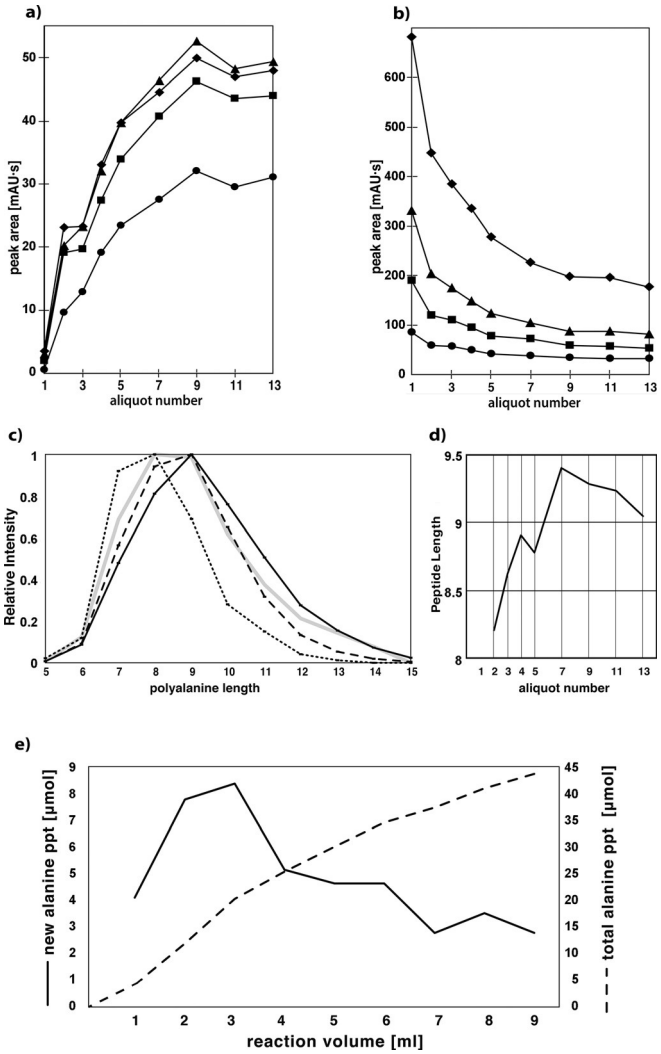

Figure 1. Analysis of the products of the COS-based condensation of poly-Ala. a, b) The reverse-phase HPLC peak area of poly-Ala peptides (a) and their hydantoin analogues (b) from the reaction supernatant of a polymerization performed with the stepwise addition of thirteen 500 µL aliquots of 10 mM COS-activated Ala. The symbols ◆, ▲, ■, and ● represent the peptides comprised of 3, 4, 5 and 6 Ala residues and the hydantoins comprised of 2, 3, 4 and 5 Ala residues, respectively. c) The relative intensity of the poly-Ala ($[M+Na]^+$ adduct) signals in the mass spectra of the solubilized precipitates from aliquots 3 (dotted line), 5 (dashed line), 7 (solid line), and 13 (gray line) indicate an increase in the relative abundance of longer peptides with the number of reaction steps followed by a small decrease. d) The change in the average peptide length in the precipitates as determined by MS signal intensities as a function of aliquot number. e) In a polymerization reaction with the continuous addition of 20 mM COS-activated Ala, the yield of insoluble peptide products increases as a function of the reaction volume. Ala in the precipitate (ppt) was quantitated at 1 mL intervals through a ninhydrin assay of the total acid hydrolysis of the precipitate. The solid line (left axis) shows the amount of Ala newly incorporated into the precipitate since the previous measurement. The dashed line (right axis) is the total Ala incorporated into the precipitate.

90 min course of the continuous reaction, there is a gradual decrease in the yield of precipitated peptides. This decrease corresponds to the rate of hydrolysis of the thiocarbamate of Ala, which we quantitated by ^{1}H-NMR measurements at various times after treatment with COS. In the borate buffer used for the polymerization reactions, the thiocarbmamate of Ala has a relatively short half-life of approximately 1 h compared to the approximately 10 h reported for phenylalanine.[6] Another outcome of a spatiotemporal separation of the COS activation and polymerization steps is that the relative yield of peptides over their urea derivatives and, in particular, their hydantoin derivatives increases as the reaction proceeds (Figure 1 a,b). This observation may be due to a decrease in dissolved COS in the later aliquots, since COS will lead to the formation of peptide thiocarbamates that may cyclize to their respective hydantoins, possibly via an isothiocyanate intermediate.[11] Therefore, the sequential addition of activated amino acid to the polymerization reaction results in unmodified peptides as the major product. For example, MALDI-FT-ICR MS suggests that the final precipitate from a 20 mM Ala polymerization reaction performed in a stepwise manner (13 total additions of 500 µL each) is approximately 90% in the form of peptides with an N-terminal amine, 10% peptides with an N-terminal hydantoin, and trace (<1%) amounts of peptide-urea derivatives (Figure S1 in the Supporting Information).

We also tested the polymerization reaction at lower amino acid concentrations and we were able to detect poly-Ala products by HPLC from a continuous reaction of 1 mM COS-activated Ala (Figure S2). However, the 1 mM Ala reaction did not produce a visible peptide precipitate. While the mechanism behind the increase in peptide yield and polymer length under continuous reaction conditions should be independent of amino acid concentration, the accumulation of β-structured aggregates is a concentration-dependent phenomenon. Therefore, we dried the final 1 mM reaction under vacuum to a final volume 100× less than the original. In this smaller volume, there was a precipitate that could be collected by centrifugation and had a CD spectrum typical for a β-sheet (Figure S3).

The stepwise and continuous polymerization reactions of 10 or 20 mM Ala produce a significant amount of insoluble peptides that begin to precipitate from solution within minutes. Similar reactions with L-valine (Val) instead of Ala produce even more precipitated peptide, which can be attributed to the greater hydrophobicity of Val. The precipitates were collected by centrifugation at $25000 \times g$ and washed with water for subsequent analyses. The circular dichroism (CD) and attenuated total reflectance Fourier-transform infrared spectroscopy (ATR-FTIR) spectra of the poly-Ala and poly-Val precipitates are typical of highly β-structured peptides (Figure 2 a,b). Both ATR-FTIR spectra show a narrow maximum at 1625 cm^{-1} (Ala) or 1626 cm^{-1} (Val) and a smaller peak at 1696 cm^{-1} (Ala) or 1688 cm^{-1} (Val), which is consistent with an antiparallel β-sheet arrangement of the peptides in the aggregates.[12] Negatively stained electron micrographs of the precipitates reveal the presence of long fibril-like structures (Figure 2 c,d).

We also analyzed the precipitates by X-ray diffraction of samples that were aligned by slowly drying an aqueous suspension of the precipitate between two glass rods. The diffraction images of the poly-Ala and poly-Val aggregates

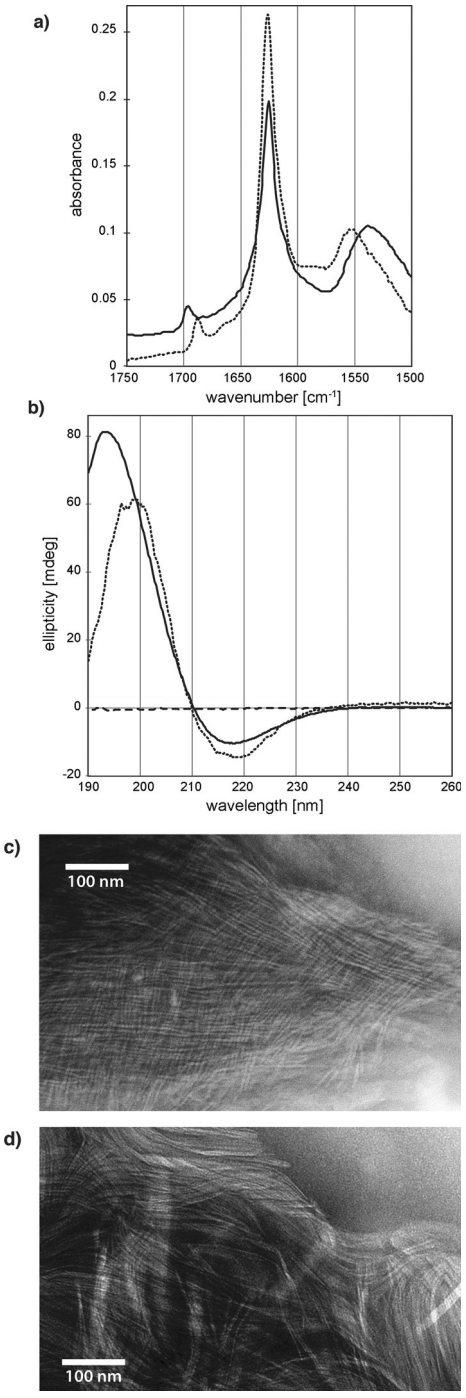

Figure 2. Biophysical characterization of the poly-Ala and poly-Val precipitates indicates an amyloid-like structure. The precipitates from the condensation reactions were washed with water in several rounds of centrifugation at 25 000 × *g* and then resuspended in water and dried on a diamond ATR cell for FTIR spectroscopy (a) or measured directly in water for CD spectroscopy (b). The spectra of the poly-Ala (solid line) and poly-Val (dotted line) precipitates are both indicative of highly β-sheet-structured aggregates. Negatively stained EM images of the poly-Ala (c) and poly-Val (d) precipitates show the fibril-like nature of the precipitates.

(Figure 3) reveal a clear macroscopic alignment of the fibrillar precipitates, with distinct equatorial and meridional reflections. It is worth noting that the poly-Ala sample does not display an amyloid-typical fiber diffraction since it lacks a meridional reflection at 4.7 Å. Most amyloid fibers give rise to a prominent reflection near 4.7 Å because the distance between the hydrogen-bonded β-strands in a β-sheet is constrained by geometry to be close to 4.7 Å. However, the poly-Ala fibers have a strong meridional reflection at 4.35 Å, a spacing consistent with previously reported β-sheet structures of poly-Ala.[13,14] The Arnott model has a $C222_1$ symmetry, for which the reflection conditions are $h+k=2n$. Thus, the 100 reflection, which corresponds to a 4.73 Å spacing, is systematically absent, while the 110 reflection near the meridian with a 4.35 Å spacing is observed. The report from Asakura et al.[13] comes to a similar conclusion for the poly-Ala model, and their powder diffraction pattern aligns well with the data from the poly-Ala precipitate of this study (Figure S4). In contrast to these previous reports, the fibrous precipitate studied here gives rise to additional lower-resolution equatorial reflections representing larger spacings. This suggests another level of organization that is absent in the reported powder diffraction (Figure S4). The lowest-resolution reflection is broad, with a maximum at 47 Å, similar to the fibril width measured by electron microscopy (EM). Taking the $C_{\alpha(i)}$–$C_{\alpha(i+2)}$ distance to be 6.9 Å for antiparallel β-strands, the 47 Å measured suggests the longest strand length to be 13 amino acid residues, which is within the range of polymer lengths detected in the precipitate by MALDI-FT-ICR (Figure 1c). One unexpected feature of the diffraction is that the 5.27 Å reflection (index 020 in the Arnott model) has, in addition to its expected equatorial intensity, a significant meridional component. Since this reflection arises from the spacing between the sheets, the meridional component could be explained by a second structural species in the precipitate in which the β-sheets lie perpendicular to the fiber axis. Taken together, the data indicate that Ala polymers of length 7–13 that are formed during the COS-activated polymerization of Ala are assembled into fibrils composed of the amyloid cross-β-sheet motif.

In contrast to the atypical features of the poly-Ala fiber diffraction, the diffraction from the aligned poly-Val precipitate was characteristic of the cross-β-sheet pattern, with a clear 4.63 Å meridional reflection. The additional meridional reflection at 9.3 Å indicates that the structural repeat occurs every other strand along the axis of the fiber, which is consistent with an antiparallel arrangement of the β-strands, as also suggested by the FTIR data. There are several weak low-resolution equatorial reflections (Figure 3b,d), however it appears that the spacing between the sheets is actually 9.3 Å, since there is also an equatorial component to the diffracted intensity at this resolution. In fact, there is a subtle difference in 2θ for the maximum intensity of the equatorial and meridional component of this reflection (imperceptible in the image but visible in a comparison of the meridional and equatorial cross-sections of the diffracted intensity as shown in Figure S5).

The polymerization reactions described thus far are not particularly representative of what would have occurred in the prebiotic soup since they involve isolated amino acids. In order to explore the robustness of the polymerization and self-assembly, we chose a mixture of glycine (Gly), Ala,

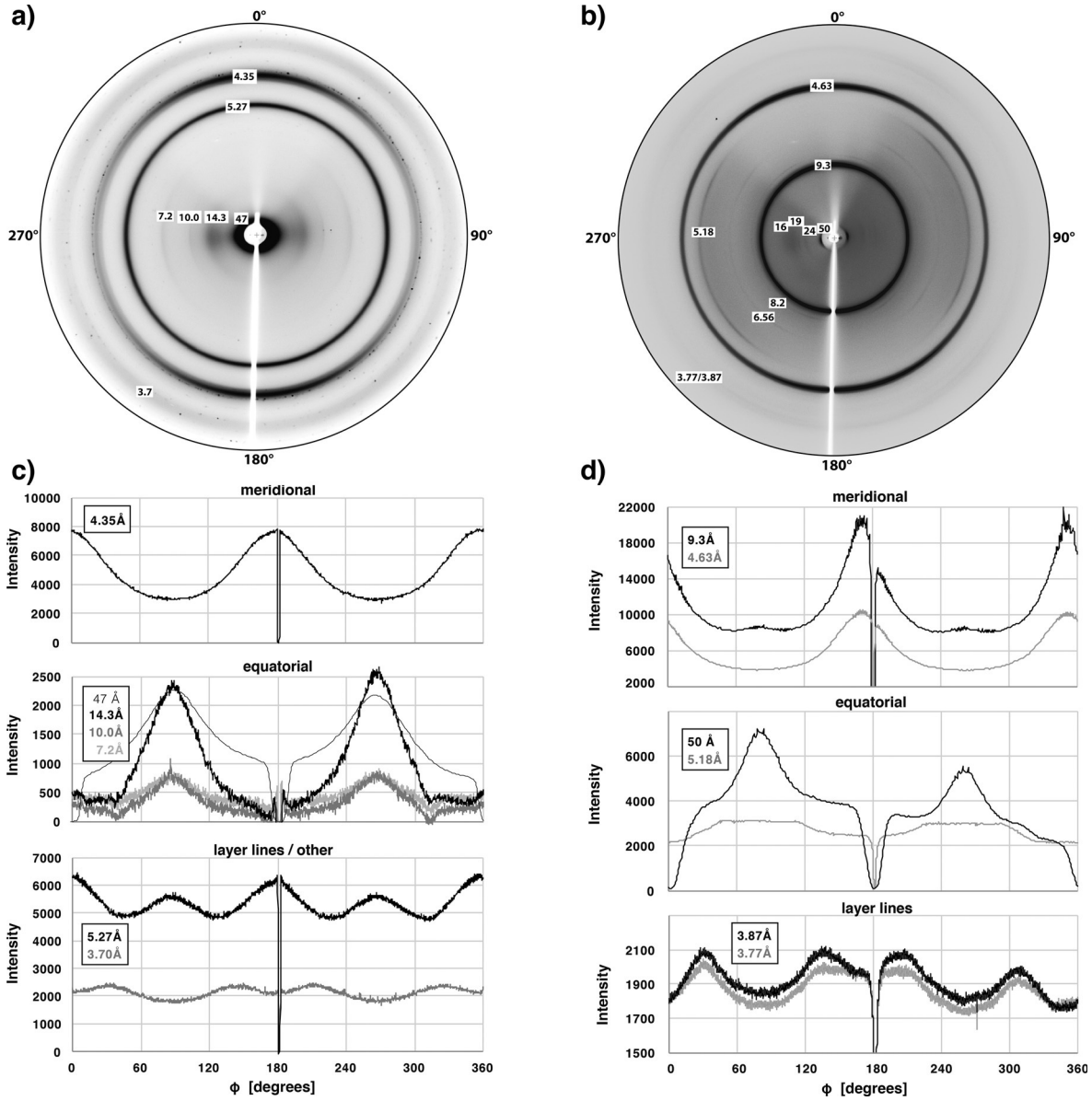

Figure 3. X-ray fiber diffraction of aligned precipitates indicates the presence of cross-β-sheet motifs in the polypeptide precipitates. The diffraction images of dried and aligned poly-Ala (a) and poly-Val (b) samples are shown. The fiber axis is nearly vertical in the plane of the page and the spacing of the reflections in ångström is indicated on the left sides of the images. The radial variation in diffraction intensity is plotted for several resolution rings (indicated with different shades of gray), emphasizing the alignment of the fibers in the poly-Ala (c) and poly-Val (d) samples.

aspartic acid (Asp), and Val, since these amino acids are, based on a survey of the literature, four of the most likely amino acids to have existed on the prebiotic Earth.[15,16] Using carbonyldiimidazole (CDI) as a coupling agent, two different compositions of these four amino acids were polymerized: 1 mM Gly, 5 mM Ala, 0.5 mM Asp, 2.5 mM ^{15}N-Val (low-Gly); and 8 mM Gly, 8 mM Ala, 1 mM Asp, 3 mM ^{15}N-Val (high-Gly). While CDI is hardly a prebiotic activating agent, the intermediates on its activation pathway,[7] in particular the *N*-carboxyanhydride common to the COS mechanism,[6] are consistent with prebiotic conditions. The low-Gly polymerization gave rise to a precipitate that formed within hours, while the high-Gly condition yielded a clear solution. After several days, the high-Gly reaction developed a very light precipitate. The precipitates from both reactions were collected by centrifugation at $25000 \times g$, yielding CD spectra typical of β-structured peptides (Figure S6) and EM images with fibrils similar to the poly-Ala aggregates (Figure S7). While we could not determine the sequence composition of these mixed peptide aggregates, FT-ICR-MS analyses give some insight into the types of peptides that were formed and which peptides were more prone to precipitate. The soluble peptides have a composition that closely reflects the initial mixture of amino acids, while the insoluble material has a lower Gly content and is essentially free of Asp. Ala and Val are thus overrepresented in the precipitate compared to the

initial mixture (Figure S8). In neither reaction were pure poly-Ala or poly-Val peptides detected by MS. This latter result is not unexpected considering the multinomial probability of a non-selective polymerization (which in the case of the low-Gly reaction predicts an Ala 9-mer yield of of 0.5% and in the case of high-Gly, an Ala 9-mer yield of 0.02% relative to all possible 9-mers). Taken together, these results show that many different polypeptides in a complex mixture are able to form amyloids under prebiotic conditions. However, the phase separation from soluble peptides to insoluble peptide precipitate does appear to be selective with respect to amino acid composition, displaying an expected tendency towards a higher incorporation of hydrophobic residues.

The self-assembly of peptides, especially those with binary alternating sequences of hydrophobic and hydrophilic residues, is a well-studied phenomenon and has been observed under a variety of settings, including prebiotic conditions.[17–19] However, the robust formation of amyloid-like precipitates from mixtures of amino acids under conditions relevant to a prebiotic earth[20] reveals an unexpected ease of creating a higher-order (tertiary and quaternary) structure in the context of complex mixtures of short peptides. The cross-β-sheet structure is unique amongst protein folds due to its structural repetitiveness in the sub-nm range (4.7 Å for parallel β-sheets and 9.4 Å for antiparallel β-sheets). This repetitiveness creates a local high concentration of identical functional groups that have the potential to bestow an increased binding affinity through cooperativity and avidity effects. The fact that amyloids can grow/replicate by seeded (non-chemical) polymerization, and the recent identification of catalytically active peptide amyloids[21–22] are further indications that amyloids can possess many properties that are required for the first replicative elements in an origin-of-life process. The apparently inherent binding affinities of amyloids for both nucleic acid polymers and membranes[17,23–24] (both attributable to the repetitiveness of the structures), and an evolutionary analysis that predicts that repeats of small β-sheet peptides were among the first secondary structural elements to have interacted with RNA,[25] emphasize the importance of more nuanced models for the origin of life than just the "amyloid world" or "RNA world" hypotheses alone. A comprehensive theory on the origin of life must at some point reconcile the codependence of lipids, proteins, and nucleic acids. Amyloids, with their polyfunctional nature, may allow the bridging of the diverse theories on the origins of biological molecules and, eventually, life.

Acknowledgements

Funding for this work was supported by ETH grant ETH-09 15-1. We also acknowledge the support by the Electron Microscopy Service (EMEZ) and the Mass Spectroscopy Service (LOCMS) of ETH Zurich as well as Dr. Marc Leibundgut and Dr. David Sargent for access to their diffractometer.

Keywords: aggregation · amyloids · peptides · prebiotic chemistry · self-assembly

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 11609–11613
Angew. Chem. **2016**, *128*, 11781–11785

[1] J. Greenwald, R. Riek, *J. Mol. Biol.* **2012**, *421*, 417–426.
[2] T. Dale, *J. Theor. Biol.* **2006**, *240*, 337–342.
[3] O. Carny, E. Gazit, *FASEB J.* **2005**, *19*, 1051–1055.
[4] C. P. Maury, *Origins Life Evol. Biospheres* **2009**, *39*, 141–150.
[5] M. G. Schwendinger, B. M. Rode, *Anal. Sci.* **1989**, *5*, 411–414.
[6] L. Leman, L. Orgel, M. R. Ghadiri, *Science* **2004**, *306*, 283–286.
[7] K. W. Ehler, L. E. Orgel, *Biochim. Biophys. Acta Protein Struct.* **1976**, *434*, 233–243.
[8] E. T. Parker, M. Zhou, A. S. Burton, D. P. Glavin, J. P. Dworkin, R. Krishnamurthy, F. M. Fernandez, J. L. Bada, *Angew. Chem. Int. Ed.* **2014**, *53*, 8132–8136; *Angew. Chem.* **2014**, *126*, 8270–8274.
[9] J. G. Forsythe, S. S. Yu, I. Mamajanov, M. A. Grover, R. Krishnamurthy, F. M. Fernandez, N. V. Hud, *Angew. Chem. Int. Ed.* **2015**, *54*, 9871–9875; *Angew. Chem.* **2015**, *127*, 10009–10013.
[10] L. M. Shinchuk, D. Sharma, S. E. Blondelle, N. Reixach, H. Inouye, D. A. Kirschner, *Proteins Struct. Funct. Bioinf.* **2005**, *61*, 579–589.
[11] A. Lindenmann, N. H. Khan, K. Hofmann, *J. Am. Chem. Soc.* **1952**, *74*, 476–480.
[12] C. Toniolo, M. Palumbo, *Biopolymers* **1977**, *16*, 219–224.
[13] T. Asakura, M. Okonogi, K. Horiguchi, A. Aoki, H. Saito, D. P. Knight, M. P. Williamson, *Angew. Chem. Int. Ed.* **2012**, *51*, 1212–1215; *Angew. Chem.* **2012**, *124*, 1238–1241.
[14] S. Arnott, S. D. Dover, A. Elliott, *J. Mol. Biol.* **1967**, *30*, 201–208.
[15] K. Ikehara, *Chem. Rec.* **2005**, *5*, 107–118.
[16] P. G. Higgs, R. E. Pudritz, *Astrobiology* **2009**, *9*, 483–490.
[17] O. Carny, E. Gazit, *Origins Life Evol. Biospheres* **2011**, *41*, 121–132.
[18] R. A. Illos, F. R. Bisogno, G. Clodic, G. Bolbach, I. Weissbuch, M. Lahav, *J. Am. Chem. Soc.* **2008**, *130*, 8651–8659.
[19] A. Brack, L. Orgel, *Nature* **1975**, *256*, 383–387.
[20] S. L. Miller, H. J. Cleaves, *Systems Biology*, Oxford University Press, Oxford, New York, **2007**.
[21] C. M. Rufo, Y. S. Moroz, O. V. Moroz, J. Stohr, T. A. Smith, X. Z. Hu, W. F. DeGrado, I. V. Korendovych, *Nat. Chem.* **2014**, *6*, 303–309.
[22] M. P. Friedmann, V. Torbeev, V. Zelenay, A. Sobol, J. Greenwald, R. Riek, *PLoS One* **2015**, *10*, e0143948.
[23] L. Wang, D. Schubert, M. R. Sawaya, D. Eisenberg, R. Riek, *Angew. Chem. Int. Ed.* **2010**, *49*, 3904–3908; *Angew. Chem.* **2010**, *122*, 3996–4000.
[24] S. Braun, C. Humphreys, E. Fraser, A. Brancale, M. Bochtler, T. C. Dale, *PLoS One* **2011**, *6*, e19125.
[25] V. Alva, J. Soding, A. N. Lupas, *eLife* **2015**, *4*, e09410.

Received: May 31, 2016
Revised: July 14, 2016
Published online: August 11, 2016

