An Aminoisoxazole-Based Proto-RNA

Felix Xu[+],[a] Stefan Wiedemann[+],[a] Jonas Feldmann,[a, b] Sidney Becker,[a, c] and Thomas Carell*[a]

The RNA world hypothesis predicts that life started with the development of replicating and catalytically active RNA, which evolved in a process of molecular evolution to increasingly complex chemical structures. RNA is, however, so complex that it has most likely formed from a precursor (proto-RNA) that was more easily accessible in a prebiotic world. Recently, 3-amino-isoxazoles (IO3) were identified as building blocks that can form under prebiotic conditions and can rearrange to give the nucleoside cytidine (C). The present study shows that the constitutional isomer 5-aminoisoxazole (IO5) can undergo the same reaction to give uridine (U). Both compounds (IO3 and IO5), if embedded in RNA, react selectively to C and U, which are the main pyrimidine nucleosides of the genetic system. Importantly, the stereochemical outcome of the IO5 reaction in RNA depends on the neighboring bases. If they are β-configured RNA nucleosides, the reaction proceeds with high selectivity to give exclusively the β-configured U RNA base (anomeric control).

Introduction

RNA plays a central role in our concepts of how life has started because the molecule is capable of catalyzing reactions, and has the potential for self-replication.[1–7] These two characteristics likely served as the foundation for the development of increasingly complex molecules and chemical systems through chemical evolution, ultimately paving the way for the emergence of life as we know it.[1,8] The difficulty with the concept is that the prebiotically plausible reaction pathways that have been postulated to give the canonical RNA nucleosides, as the key building blocks for RNA, are complex and it is hard to conceive how RNA could have formed in the first place.[9] It is consequently a major scientific goal to find molecular proto-RNA structures that could have preceded RNA and to investigate how such structures could have evolved into RNA.[10–15]

We recently introduced the concept of urea-RNA, in which ribofuranose with a triuret, connected to the C1-position of ribose, acts as a proto-RNA. In these nucleosides, the triuret unit was able to establish productive H-bonds with the counter base guanin.[16] In addition, we showed that 3-aminoisoxazoles (IO3), connected via a urea linkage to the C1-atom of ribose are highly efficient precursors of cytidine (C). The needed N-isoxazol-3-yl-urea was shown to be readily accessible under prebiotically plausible conditions.[17] The compound embedded as a riboside in RNA was able to perform the cascade reaction to C (Figure 1A).[18] Although this chemistry provides access to C, the second important pyrimidine nucleoside of the genetic code U is in this scenario only available via the hydrolysis of C. This, however, is a slow reaction particularly if the C is embedded in RNA.[19,20]

In order to solve this problem and to find a common prebiotically plausible route to both pyrimidines C and U, we investigated the starting point of the IO-based chemistry, namely the reaction of cyanoacetylene (**1**) with hydroxylamine (**2**) in greater detail. As previously shown, 3-aminoisoxazole could have formed by reacting cyanoacetylene (**1**) with hydroxylamine (**2**) under strongly basic (pH ≥ 14) conditions or

[a] F. Xu,[+] Dr. S. Wiedemann,[+] Dr. J. Feldmann, Dr. S. Becker, Prof. Dr. T. Carell
Department of Chemistry
Ludwig-Maximilians-Universität (LMU) München
Butenandtstrasse 5–13, 81377 Munich (Germany)
E-mail: thomas.carell@lmu.de

[b] Dr. J. Feldmann
Current address:
Department of Chemistry
University of Cambridge
Lensfield Road, Cambridge CB2 1EW (UK)

[c] Dr. S. Becker
Current address:
Department of Chemical Biology
Max Planck Institute of Molecular Physiology
Otto-Hahn-Strasse 11, 44227 Dortmund (Germany)

[+] These authors contributed equally to this work.

🖥 Supporting information for this article is available on the WWW under https://doi.org/10.1002/ceur.202300057
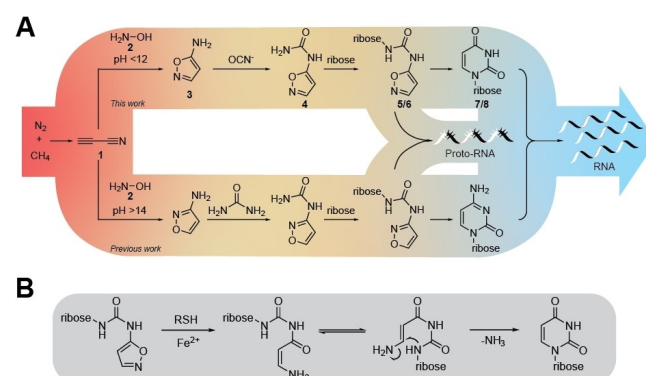
Figure 1. A) Prebiotically plausible formation of cytidine (C) and uridine (U) following the reaction of cyanoacetylene (**1**) with hydroxylamine (**2**) by reaction with the O-nucleophile of hydroxylamine (bottom)[17] and by reaction with the N-nucleophile of hydroxylamine (top). B) Putative cascade reaction that allows IO5 to react to uridine.

with hydroxylurea under slightly basic conditions (pH~10).[17] The reaction of **1** (considered to be a prebiotically plausible starting material[21–24]) with **2**[17,25] under strongly basic conditions, favours the reaction of the O-atom of **2** with **1**, which then starts the chemical trajectory to C via 3-aminoisoxazole (Figure 1A). If, however, the reaction is performed at ambient pH-values, we reasoned that it might allow the N-atom of **2** to react first which could open a pathway to U (Figure 1A) by a cascade reaction shown in Figure 1B.

The beauty of this concept lies in the possibility that both canonical pyrimidine nucleosides C and U could be traced back to the same starting materials, with just the pH-value of the first reaction dictating whether C or U is formed. This would provide conceptually ideal conditions for the prebiotic origin of the genetic code with a focus on the pyrimidine nucleosides.

Results and Discussion

Formation of uridine from 5-aminoisoxazole

In order to investigate the concept, we reacted cyanoacetylene (**1**) with hydroxylamine (**2**) at pH-values below 14, particularly at slightly basic pH values of around pH = 9–10. Indeed, under these conditions we observed exclusively the nucleophilic attack of the amino group of **2** with **1**. The reaction was performed in water (1 eq. NaHCO$_3$, 25 °C, 24 h). Upon dry down precipitation of the pure 5-aminoisoxazole (**3**) may occur, because it is a solid material in contrast to 3-aminoisoxazole, which was found to be a liquid. Washing of the remaining material with water provided NMR-pure **3** (Figure 2A). This reaction is easily conceivable in a connected pond model with one wet-dry cycle and rain showers,[26,27] which would have allowed the washing away of residual **1** and **2** along with potential side products (Figure 2B). Next, we treated the precipitated material **3** with cyanate,[28–30] which might have flooded into the pond from a second pond. We now detected a clean and exclusive formation of N-isoxazol-5-yl-urea (**4**), which again precipitated upon drying down. Another rain shower may have again washed away residual cyanate. Inflow of an aqueous solution of ribose[31–33] from a third pond into the reaction pond followed by another dry-down step leads to the reaction of **4** with ribose. When we performed the reaction, we indeed detected the clean formation of all four expected reaction products namely the two α- and β-ribopyranosides (α-/β-**5**) and the two α- and β-ribofuranosides (α-/β-**6**) as a mixture. Under our conditions the α-furanoside α-**6** was formed as the dominant reaction product with a yield of 39%. When we started the envisioned cascade cyclization reaction (Figure 1B) by addition of a thiol (DTT) and catalytic amounts of Fe^{2+} to the obtained α- and β-**6**, we indeed detected efficient reductive N–O-bond cleavage followed by a clean cyclization[34] to give the U-ribopyranosides (α-/β-**7**) and the U-ribofuranosides (α-/β-**8**, β-**8** = U). To our surprise, we noted that now the β-furanoside was by far the dominating product. The yield of the reaction is with 31% for all ribosides and with 25% for the β-furanoside U (β-**7**) astonishingly high, given the simplicity of the reaction
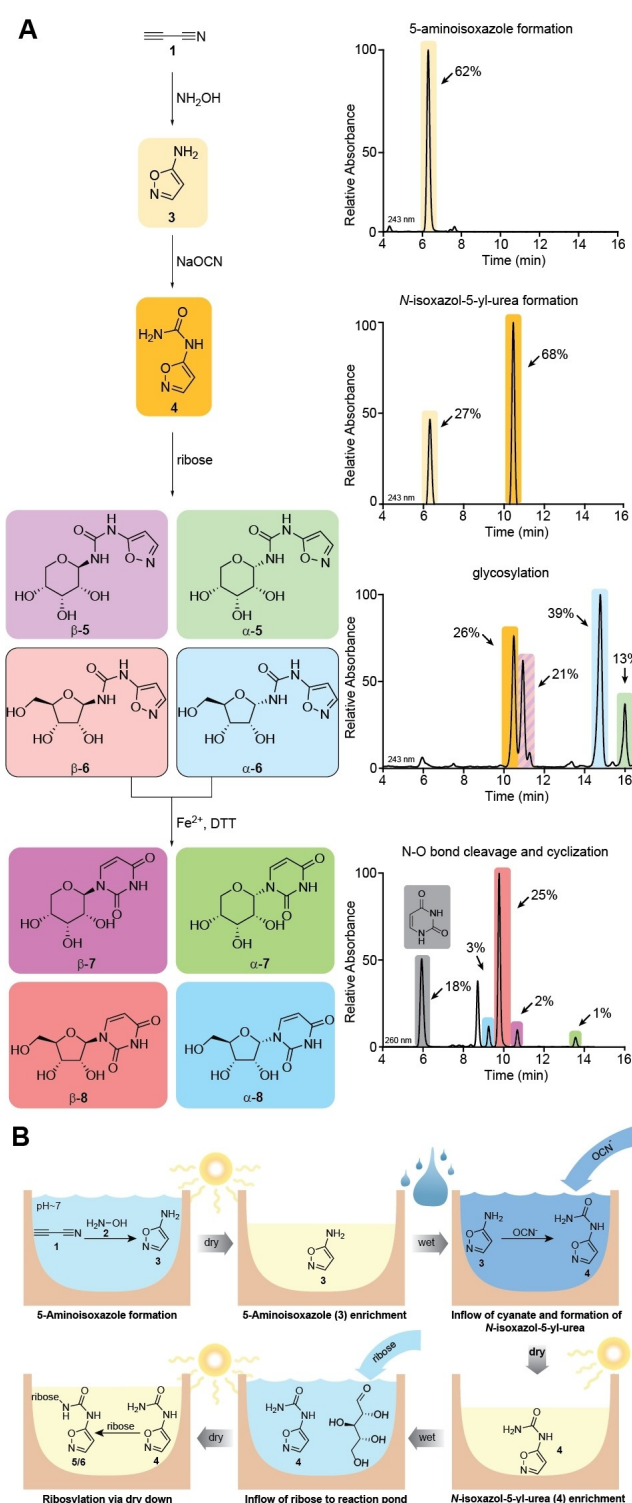

Figure 2. A) The reaction sequence that generates the pyrimidine U via the IO pathway with the corresponding HPL-chromatographic analyses of the reaction outcome showing clean conversions. B) Depiction of the dry-down connected pond model for the putative prebiotic formation of the IO5-ribosides.

sequence. The high yield of the β-furanoside U together with the remarkable stereoselectivity of the reaction make us believe that the IO-pathway to C and U provides a good model of how

the pyrimidine nucleoside could have formed under Early Earth's condition. As hypothesized, in this scenario, it is the pH-value of the initial step that determines whether the pathway delivers the C- or the U-nucleoside.

Formation of C and U from isoxazoles directly in RNA

We recently showed that the nucleoside C can form directly in RNA from embedded IO3 ribofuranosides, which allowed us to show that IO3-precursor RNAs are an extant model for a proto-RNA that could have preceded canonical RNA.[18] In order to investigate if this direct "in-RNA" transformation is also possible with IO5 to generate U and to study if C and U can form simultaneously from IO3 and IO5-containing proto-RNA, we prepared the IO3-phosphoramidite building blocks for solid phase RNA synthesis. This was achieved following a synthetic pathway outlined in Scheme 1A. The starting point of the reaction was the well-known 3′-5′-protected riboside-azide **9**, which we reacted with 5-aminoisoxazole (**3**) in a Pd-catalyzed CO insertion reaction to directly give the IO5 riboside **10** as a mixture of the α- and the β-anomers. The anomers were subsequently separated by flash column chromatography. Next, we proceeded separately with the α- and the β-isomers. Deprotection of the 5′-3′ silyl protecting group with HF in pyridine gave **11**, followed by dimethoxy-tritylation of the 5′OH group to **12**. Subsequent conversion of **12** into the phosphoramidite **13** furnished the needed building block for solid phase

oligonucleotide chemistry. Because we performed the reactions with the two anomers in parallel, we obtained both phosphoramidites β-**13** and α-**13**.

We subsequently used the building blocks β- and α-**13** to perform and optimize the RNA synthesis. We found that both building blocks (0.1 M in THF) are best incorporated into RNA using an elongated coupling time (600 s) and shortened detritylation time (40 s). Deprotection was first achieved with ammonium hydroxide (55 °C, 5 h) followed by a second deprotection step with TEA·3HF (65 °C, 2.5 h). The result of the synthesis of an RNA strand having β-**13** incorporatetd is depicted in Scheme 1B. The HPL-chromatogram of the purified strand 5′-CUUAC-β5IO-CUGA-3′ and the corresponding MALDI-TOF mass spectrum are depicted in Scheme 1B. The data proves the structural integrity of the RNA strand with the embedded IO5 nucleoside.

To study the "in-RNA" N—O-bond cleavage and cyclization of IO5 to U and to learn about the stereochemical outcome of the reaction, we prepared the RNA strand 5′-CUAAC-βIO5-CUGA-3′ using β-**13** (Figure 3A). This oligonucleotide was digested to the nucleoside level and the obtained nucleoside composition was analyzed by HPLC-MS which revealed the presence of all expected nucleosides β-C, β-U, β-G, β-A and of the additional unit β-IO5. Next, we treated the oligonucleotide with DTT and catalytic amounts of Fe^{2+}. This led to the formation of a new oligonucleotide with a slightly changed retention time. Digestion of this oligonucleotide provided the expected nucleoside mixture, but now without the β-IO5 nucleoside, arguing that the conversion of β-IO5 to uridine had happened as expected in the RNA strand.

We next investigated if two β-IO5 nucleosides would also be converted to U and in addition we wanted to get better proof for the conversion of β-IO5 to U and about the stereoselectivity. To answer these questions, we prepared the U-and IO5-only RNA strand using a mixture of α-**13** and β-**13** and we isolated the RNA strand with exactly one β-IO5 and one α-IO5 (Figure 3B). Digestion of this RNA strand provided as expected one signal for β-U and one each for the α- and β-IO5 nucleoside. Upon treatment with DTT and Fe^{2+}, a new strand with a shifted retention time was again formed in a clean reaction. Upon digestion, this strand gave only one signal for β-U, showing that both compounds α- and β-IO5 had reacted to the corresponding β-configured U nucleoside. The stereochemical information of the RNA strand obviously directs the outcome of the reaction, like we have previously shown for the IO3 nucleoside. This leads to a conversion of also the α-configured IO5 to β-configured U.

Finally, we wanted to examine if a proto-RNA strand containing both IO3 and IO5 nucleosides would react upon treatment with DTT to give C and U simultaneously. For this experiment, we prepared the oligonucleotide 5′-CC-βIO3-CU-βIO5-UU-3′ using the corresponding β-configured phosphoramidites (Figure 3C). Digestion of the strand generated signals for β-C, β-U and as expected for β-IO3 and β-IO5. Again, treatment of the strand with DTT and Fe^{2+} cleanly generated a new strand with a different retention time. Upon digestion only two signals were obtained: One for β-C and one for β-U,
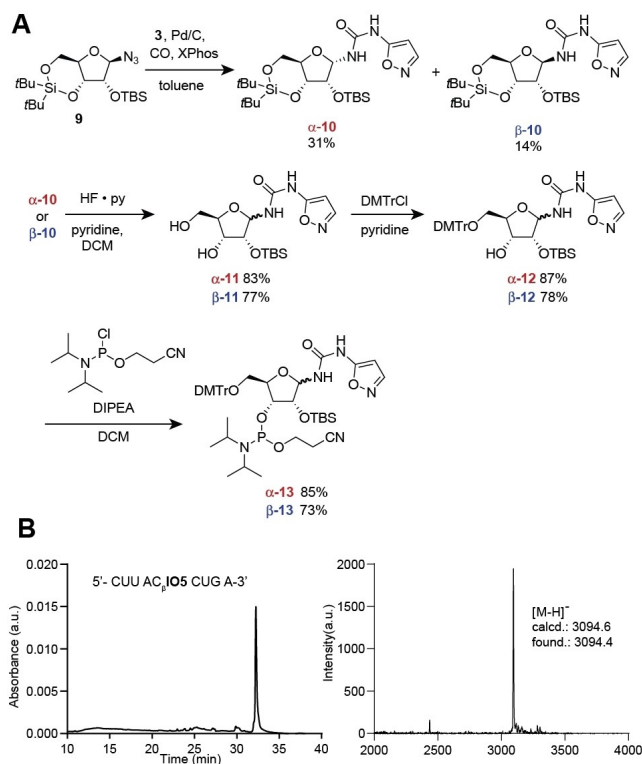
Scheme 1. A) Synthesis of the IO5 phosphoramidite and conditions B) HPL-chromatogram and MALDI-TOF mass spectra of an oligonucleotide synthesized with the IO5 phosphoramidite β-**13** shown as an example.
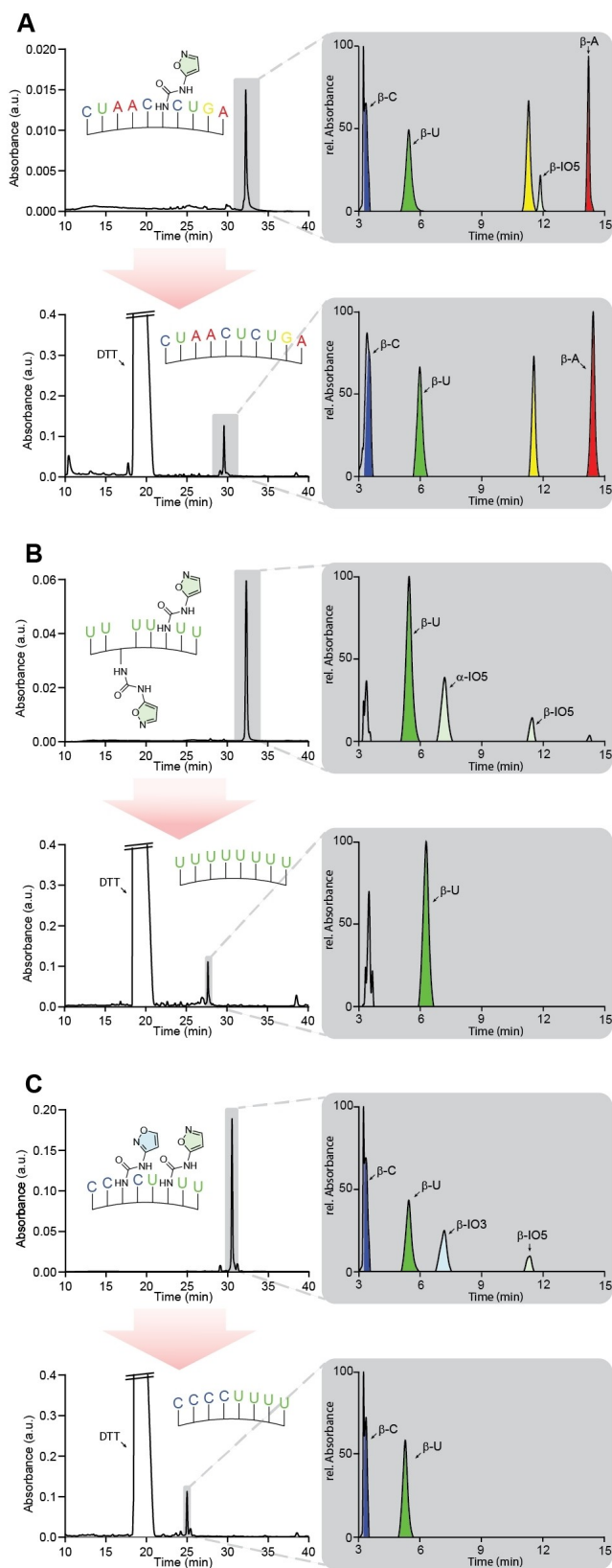
Figure 3. HPLC, MALDI, HPLC-MS (at 243 nm; grey brackets) analyses of the IO5 containing Strands before the reaction with DTT, Fe²⁺ (top) and after N–O cleavage and cyclization reaction (bottom). A) 5′-CUAAC-$_\beta$IO5-CUGA-3′ B) 5′-UU-$_\alpha$IO5-UU-$_\beta$IO5-GA-3′ C) 5′-CC-$_\beta$IO3-CU-$_\beta$IO5-UU-3′. A = adenosine, C = cytidine, G = guanosine, IO3 = 3-aminoisoxazole, IO5 = 5-aminoisoxazole, U = uridine.

showing that in the strand the precursor nucleosides β-IO3 and β-IO5 rearrange simultaneously to give C and U.

Conclusions

RNA is the central molecule within the current concepts of how life could have emerged on the Early Earth because of the molecule's potential to encode information and to catalyze reactions. It therefore effectively combines genotype with phenotype. The RNA molecule, however, is so complex that it most likely formed from a simpler precursor proto-RNA molecule that already had potentially information encoding properties. Current concepts of how RNA may have evolved are based on the formation of canonical nucleosides that got phosphorylated at some point and then oligomerized to yield RNA strands. This, however, requires the prebiotic formation of the canonical nucleosides.

We showed recently that nucleosides such as 3-amino-isoxazole (IO3) are potential precursors for the pyrimidine base C. Cytosine is formed upon simple reductive N–O bond cleavage followed by a cyclization reaction (Figure 1). Here, we show that the constitutional isomer, the 5-aminoisoxazole (IO5) can serve as a precursor for the pyrimidine nucleoside U. Both, the 3- and the 5-aminoisoxazoles form by reacting cyanoacetylene (**1**) with hydroxylamine (**2**). The formation of the different constitutional isomers depends exclusively on the pH-conditions of the reaction. It is prebiotically highly attractive to create the two needed pyrimidine nucleosides from the same chemical precursors (**1** and **2**) by just a change in reaction conditions. In addition, we show that IO5, if embedded in RNA, performs the formation of the nucleoside (U) directly in the RNA strand. This finding allowed us to insert both IO3 and IO5 into a proto-RNA for the simultaneous reaction of the two IO-precursors to C and U by just exposing the strand to thiols and catalytic amounts of Fe²⁺.

Our data show that amino- and subsequently urea-isoxazoles are highly promising prebiotic precursors for the pyrimidine nucleosides. We show that these compounds can lead to the formation of a proto-RNA and through a simple N–O bond cleavage, followed by a cyclization cascade directly within the proto-RNA, canonical RNA can emerge.

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

[1] W. Gilbert, *Nature* **1986**, *319*, 618–618.
[2] F. H. Westheimer, *Nature* **1986**, *319*, 534–536.
[3] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **1983**, *35*, 849–857.
[4] A. J. Zaug, T. R. Cech, *Science* **1986**, *231*, 470–475.
[5] D. R. Mills, R. L. Peterson, S. Spiegelman, *Proc. Natl. Acad. Sci. USA* **1967**, *58*, 217–224.
[6] T. A. Lincoln, G. F. Joyce, *Science* **2009**, *323*, 1229–1232.
[7] L. E. Orgel, *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99–123.
[8] G. F. Joyce, *Angew. Chem. Int. Ed.* **2007**, *46*, 6420–6436.
[9] M. Yadav, R. Kumar, R. Krishnamurthy, *Chem. Rev.* **2020**, *120*, 4766–4805.
[10] G. F. Joyce, *Nature* **2002**, *418*, 214–221.
[11] N. V. Hud, B. J. Cafferty, R. Krishnamurthy, L. D. Williams, *Chem. Biol.* **2013**, *20*, 466–474.
[12] N. V. Hud, *Nat. Commun.* **2018**, *9*, 5171.
[13] A. Eschenmoser, *Tetrahedron* **2007**, *63*, 12821–12844.
[14] G. F. Joyce, A. W. Schwartz, S. L. Miller, L. E. Orgel, *PNAS* **1987**, *84*, 4398–4402.
[15] A. E. Engelhart, N. V. Hud, *Cold Spring Harbor Perspect. Biol.* **2010**, *2*.
[16] H. Okamura, A. Crisp, S. Hübner, S. Becker, P. Rovó, T. Carell, *Angew. Chem. Int. Ed.* **2019**, *58*, 18691–18696.
[17] S. Becker, J. Feldmann, S. Wiedemann, H. Okamura, C. Schneider, K. Iwan, A. Crisp, M. Rossa, T. Amatov, T. Carell, *Science* **2019**, *366*, 76–82.
[18] F. Xu, A. Crisp, T. Schinkel, R. C. A. Dubini, S. Hübner, S. Becker, F. Schelter, P. Rovó, T. Carell, *Angew. Chem. Int. Ed.* **2022**, *61*, e202211945.
[19] E. R. Garrett, J. Tsau, *J. Pharm. Sci.* **1972**, *61*, 1052–1061.
[20] H. Lönnberg, R. Käppi, *Nucleic Acids Res.* **1985**, *13*, 2451–2456.
[21] P. Thaddeus, *Philos. Trans. R. Soc. London Ser. B* **2006**, *361*, 1681–1687.
[22] R. A. Sanchez, J. P. Ferris, L. E. Orgel, *Science* **1966**, *154*, 784–785.
[23] B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy, J. D. Sutherland, *Nat. Chem.* **2015**, *7*, 301–307.
[24] P. B. Rimmer, O. Shorttle, *Life* **2019**, *9*.
[25] K. B. Muchowska, S. J. Varma, J. Moran, *Nature* **2019**, *569*, 104–107.
[26] B. Damer, D. Deamer, *Life* **2015**, *5*, 872–887.
[27] A. Y. Mulkidjanian, A. Y. Bychkov, D. V. Dibrova, M. Y. Galperin, E. V. Koonin, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E821–E830.
[28] Y. Yamagata, *Origins Life Evol. Biospheres* **1999**, *29*, 511–520.
[29] J. P. Ferris, E. A. Williams, D. E. Nicodem, J. S. Hubbard, G. E. Voecks, *Nature* **1974**, *249*, 437–439.
[30] P. A. Gerakines, M. H. Moore, R. L. Hudson, *Icarus* **2004**, *170*, 202–213.
[31] H. J. Cleaves, *Precambrian Res.* **2008**, *164*, 111–118.
[32] A. W. Schwartz, R. M. de Graaf, *J. Mol. Evol.* **1993**, *36*, 101–106.
[33] C. Reid, L. E. Orgel, *Nature* **1967**, *216*, 455–455.
[34] M. Kijima, Y. Nambu, T. Endo, *J. Org. Chem.* **1985**, *50*, 1140–1142.

Version of record online: October 12, 2023