

Prebiotic Chemistry

International Edition: DOI: 10.1002/anie.201903400
German Edition: DOI: 10.1002/ange.201903400

Direct Prebiotic Pathway to DNA Nucleosides

Jennifer S. Teichert, Florian M. Kruse, and Oliver Trapp*

Abstract: It is assumed that RNA played a key role in the origin of life, and the transition to more complex but more stable DNA for continuous information storage and replication requires the development of a ribonucleotide reductase to obtain the deoxyribonucleotides from ribonucleotides. This step, as well as an alternative path from abiotic molecules to DNA-based life is completely unknown. Shown here is the formation of deoxyribonucleosides under relevant prebiotic conditions in water in high regio- and stereoselectivity, from all canonical purine and pyrimidine bases, by condensation with acetaldehyde and sugar-forming precursors. Thus, a continuous path to deoxyribonucleosides, starting from simple, prebiotically available molecules has been discovered. Furthermore, the deoxyapionucleosides (DApiNA) were identified as a potential DNA progenitor. The results suggest that the DNA world evolved much earlier than previously assumed.

The question of how to create finely balanced living biological systems from a pool of small prebiotic molecules is undoubtedly one of the most fascinating scientific challenges within living memory. It is assumed that complex organic molecules, such as nucleobases (**1**; see Figure 1), amino acids, and sugars, evolved from prebiotic feedstock molecules.^[1,2] Today, these molecules can still be detected on extraterrestrial material^[3] and they were part of the early Earth.^[4,5] The RNA world^[6–8] hypothesis postulates that ribonucleosides (**2**) were initially formed from these molecular building blocks,^[9] which later aggregated into the information-storing RNA polymer. This path, accepted by large parts of the scientific community, combines catalytic function, inherent properties of replication, mutation, and selection and is thus able to herald a chemical evolution.^[10–14]

Influenced by Emil Fischer's historical approach to nucleosides by glycosidation of nucleobases,^[15] accessing RNA nucleosides (**2**) and nucleotides seems to be easier than the access to the DNA nucleosides **3**, since the former consist of the sugar component D-ribose (**4**) and the respective nucleobase **1** (Orgel pathway; Figure 1 a).^[16–22] For the purine nucleosides, a modified pathway has recently been proposed by Carell and co-workers,^[23,24] based on the formamidopyr-

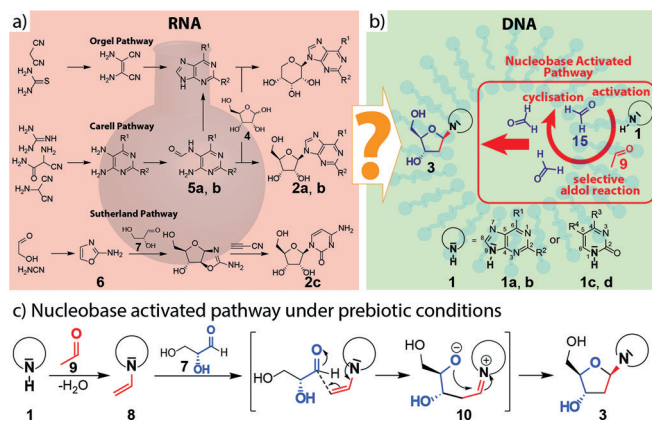


Figure 1. Prebiotic pathways to the ribonucleosides **2** and the so far unknown transition into the deoxyribonucleosides **3** in comparison to the direct access of **3**. a) Orgel and Carell pathways utilizing either purines or formamidopyrimidines (FaPys; **5a,b**) in a reaction with D-ribose (**4**). The Sutherland pathway via 2-aminooxazole (**6**) as building block for assembly of the sugar and pyrimidine base. b) The prebiotic pathway to **3** by nucleobase activation of acetaldehyde (**9**) in combination with sugar-forming aldol reactions. c) Stereoselectively controlled mechanism leading to deoxyribonucleosides (dA **3a**, dG **3b**, dC **3c**, dT **3d**) via **8** and reaction with D-glyceraldehyde (**7**). a: R¹ = NH₂, R² = H; b: R¹ = OH, R² = NH₂; c: R³ = NH₂, R⁴ = H; d: R³ = OH, R⁴ = CH₃.

imidines **5** (FaPys) as purine precursors, which are easier to condense with aldoses.

While the nucleobases **1** are easily accessible as stable building blocks under prebiotic conditions,^[25–29] there is no such simple pathway to **4**^[30,31] for the formation of **2**. Notably, **4** is subject to rapid decomposition in aqueous solution^[32] and, even more crucial, the selectivity of the formose reaction^[33] is negligibly low with respect to **4**. Sutherland has elegantly solved this challenge for the preparation of pyrimidine ribonucleosides by introducing 2-aminooxazole (**6**) as the connecting building block between sugar and pyrimidine base.^[34] The ribose unit is assembled by reaction with D-glyceraldehyde (**7**).

The prebiotic access to the DNA building blocks **3** seems to be even more difficult since the Fischer–Orgel–Carell strategy would require D-deoxyribose, which is not readily accessible under prebiotic conditions.^[35] Furthermore, there is no uniform pathway including both purine and pyrimidine nucleosides. The RNA world hypothesis therefore postulates the evolutionary development of a ribonucleotide reductase^[36] to create a significantly more stable DNA, which is capable of storing information over a prolonged period of time.^[37] A chemical prebiotic pathway by photoreduction has been recently elaborated by Krishnamurthy and Sutherland et al.^[38] The possibility that **3** could have developed directly from their easily accessible nucleobases and simple prebiotically present aldehydes has not yet been considered. Based on

[*] J. S. Teichert, F. M. Kruse, Prof. Dr. O. Trapp
Department of Chemistry, Ludwig-Maximilians-University Munich
Butenandtstr. 5–13, 81377 Munich (Germany)
E-mail: oliver.trapp@cup.uni-muenchen.de

J. S. Teichert, Prof. Dr. O. Trapp
Max-Planck-Institute for Astronomy
Königstuhl 17, 69117 Heidelberg (Germany)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
<https://doi.org/10.1002/anie.201903400>.

this idea, we performed a retrosynthetic analysis of prebiotic reaction networks and investigated possible reaction pathways activated by nucleobases in aqueous solution (Figure 1b).

The nucleophilic characters at N9 in the purines **1a** (pK_a 9.80) and **1b** (pK_a 9.21), and at N1 in the pyrimidine bases **1c** (pK_a 12.2) and **1d** (pK_a 9.94; Figure 1b), compared to the primary amino groups, have been neglected so far and have been only recently quantitatively characterized by Mayr et al.^[39] In analogy to aminocatalysis,^[40,41] the corresponding enamines, the N9-vinylpurine bases **8a,b** and N1-vinylpyrimidine bases **8c,d**, are obtained with the prebiotically available acetaldehyde^[42] (**9**; 4.0 mmol) and **1** (0.1 mmol) in water (1 mL). This reaction mixture is sealed in a glass ampoule, stirred at 50°C for four days and analyzed after opening the ampoule. We verified the formation of **8a,b** and **8c,d**, by HPLC-UV-detection, UHPLC—QTOF MS using co-injections with reference samples (see the Supporting Information), and high-resolution Orbitrap MS (Figures 2a and b; see Table S1 in the Supporting Information). It is important to note that the reaction mixture has to be sealed to avoid evaporation of volatile **9**. Furthermore the heated ampoule simulates the elevated atmospheric pressure on early Earth.^[43]

The obtained vinyl nucleobases **8a–d** (0.1 mmol) react stereoselectively with D-glyceraldehyde (**7**; 0.15 mmol) or even with formaldehyde (**15**; 37 wt. % in water, 4 mL) to give the deoxyribonucleosides dA **3a** (0.68 %), dG **3b** (0.45 %), dC **3c** (2.49 %), and dT **3d** (0.33 %), respectively (Figures 1b and c; see Table S2). The reaction mixtures were stirred and heated in a sealed glass ampoule at 50°C for 4 days. We verified the formation of **3** by HPLC-UV-detection, UHPLC—QTOF MS using co-injections with reference samples, and high-resolution Orbitrap MS (Figures 2c–e).

The nucleophilic attack of **8** onto the carbonyl group of **7** is stereoelectronically controlled in an unambiguous manner^[44] (Figure 1c). The formed open-chain deoxyribose system **10** undergoes kinetically favored 5-*exo-trig* cyclization^[45] to the furanose form to deliver the β -deoxyribonucleoside **3** with defined stereochemistry (see Figures S14–S16). In contrast to this pathway, if the reaction takes place at the primary amine group, an easily cleavable iminium ion is formed and therefore it can be considered as a natural protecting group.

The herein described deoxyribonucleoside-forming reaction proceeds selectively and under strictly prebiotic and mild reaction conditions in water between 40 and 70°C with 50°C being optimal. Additives, that is, sodium acetate, disodium hydrogen phosphate, are not required, but they enhance the rate of formation (see Table S7). This reaction is very robust and proceeds in a neutral to alkaline pH range. Furthermore, prebiotically difficult explainable modulations of the pH are not necessary.

The excellent selectivity and the slow but continuous formation of **3** in this way is of great advantage. Usually rapid reactions with high yields lead from one equilibrium to another, and thus do not favor self-developing reaction networks, which are essential for the formation and maintenance of life.

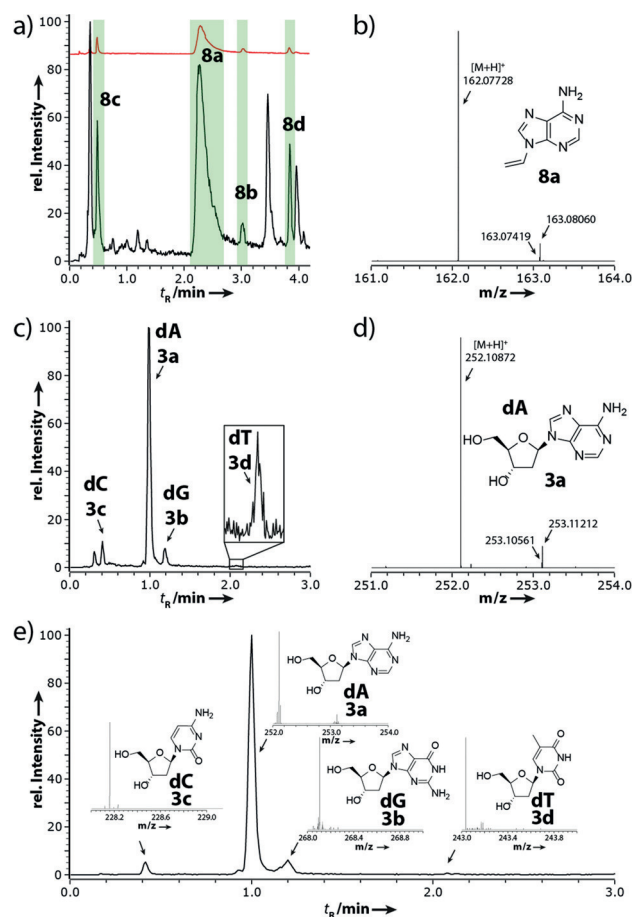


Figure 2. Formation of **3**. a) Depiction of **8** obtained by one-pot reaction of all four nucleobases **1** and **9** in water, as analyzed by UHPLC-QTOF MS (mass-filter m/z 138.064–138.068, 153.063–153.067, 162.075–162.079, 178.070–178.074, black curve) with co-injection of reference compounds (red curve). b) High-resolution Orbitrap mass spectrum of the vinyl adenine **8a** (0.1 mmol adenine **1a**, 1 mmol **9** in 1 mL H₂O for 4 days at 50°C). c) Depiction of **3** formed by one-pot reaction of all four nucleobases **1**, **9**, and **7** in H₂O, as analyzed by UHPLC-QTOF MS (mass-filter m/z 228.095–228.099, 243.095–243.099, 252.107–252.110, 268.102–268.106). d) High-resolution Orbitrap mass spectrum of **3a** (0.1 mmol **1a**, 1 mmol **9**, 0.15 mmol, **7** in 1 mL H₂O for 4 days at 50°C). e) Depiction of **3** formed by one-pot reaction of all four vinyl nucleobases **8** and **15** in H₂O (0.1 mmol **8** each, 1 mL formaldehyde solution (37 wt. % in H₂O) for 4 days at 50°C), as analyzed by UHPLC-QTOF MS (mass-filter m/z 228.095–228.099, 243.095–243.099, 252.107–252.110, 268.102–268.106). See the Supporting Information for experimental details of the syntheses and analyses.

The synthesis of **3** succeeds not only in a one-pot reaction with the respective nucleobase **1**, but also out of the mixture **1a–d**. This result underlines the plausibility of the reaction that we have found.

The question arose as to which deoxynucleosides are obtained when other sugar precursors and **15** are used in the reaction with **8**. Therefore, we investigated the reaction of **1** with **9**, and subsequent reactions with the C2-precursor glycolaldehyde **13** and **15**, thus yielding the deoxythreonucleoside **16** (Figure 3, red pathway; see Table S3). This deoxynucleoside is complementary to Eschenmoser's threonucleosides (TNA) as an RNA progenitor.^[46] The compound

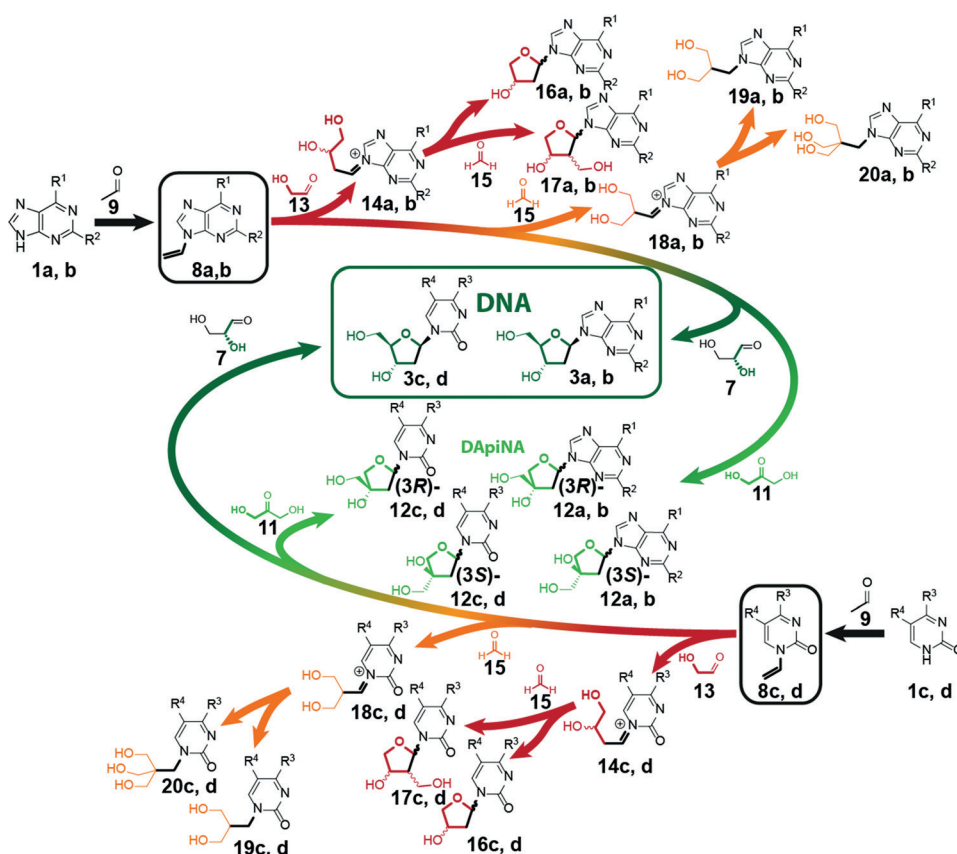


Figure 3. Experimentally verified selective pathways to all deoxysugar nucleosides formed by the nucleobase activated pathway. The initial step is formation of **8** from the **1** and **9** (black pathway). Reaction with **13** and **15** leads to **16** (dTNA) and **17** (HMENA), respectively, which are unable to polymerize into a defined structure (red pathway). In a side-reaction under Cannizzaro conditions with **15**, two deoxynucleosides are formed, **19** and **20** (orange pathway). Reaction of **8** with **11** results in the DNA progenitor **12** (DApiNA nucleosides, light-green pathway), and with **7**, **3** is obtained (dark-green pathway). a: $R^1 = \text{NH}_2$, $R^2 = \text{H}$; b: $R^1 = \text{OH}$, $R^2 = \text{NH}_2$; c: $R^3 = \text{NH}_2$, $R^4 = \text{H}$; d: $R^3 = \text{OH}$, $R^4 = \text{CH}_3$.

16 is unable to polymerize into a polynucleic acid because of the lack of a second hydroxy group, and thus it has to be assumed that this path is an evolutionary dead end. Interestingly, the 2C-hydroxymethylerythrionucleoside **17** (HMENA) is observed (Figure 3, red pathway; see Table S4). This path also has to be regarded as an evolutionary dead end because of the formation of eight possible stereoisomers.

The side-reaction of **8** with **15** gives two nucleosides, the 2-hydroxymethyl-1,3-propanediol **19** and pentaerythrol (PentaNA) **20** (Figure 3, orange pathway; see Table S5). The Cannizzaro reaction,^[47] that is, the reduction by hydride migration from formaldehyde, is the key step in both reactions. Potentially both compounds could form a polynucleic acid, but such a structure would be difficult to replicate because of branching and ambiguous sequencing. In addition, through polymerization the central carbon atom of the sugar-type moiety becomes stereogenic, leading to enantiomers and diastereomers in the polymer.

In the absence of the Cannizzaro reaction^[47] and saturation of **18** to give **19**^[48] and **20**, the cleavage of the sugar-type moiety is favored because of the impossibility of cyclization. Therefore **1** is available again as a precursor to activate **9**. It is known that sugars can decompose to **15** by irradiation with

UV light,^[49] yielding **9**, and thus both building blocks are available for reaction with **1**. The statistical formation of the reaction products eventually leads to the deoxyribo-nucleosides **3**.

The C3-ketose dihydroxyacetone **11**, which is isomeric to **7**, reacts with **8** to give the deoxyapionucleosides **12** (Figure 3, light-green pathway; see Table S6). The unambiguous assignment was made by high-resolution Orbitrap mass spectrometry and UHPLC-QTOF-MS. For **12**, two peaks are observed in the HPLC chromatograms and result from the formation of the α - and β -deoxyapionucleosides, which result in the formation of diastereoisomers. Remarkably, the carbohydrate apiose occurs in pectins in the cell wall of plants and plays an important role in the glycosylation of secondary metabolites.^[50,51] It is noteworthy that **12** is also capable of polymerizing into a polynucleic acid and can therefore be regarded as a previously undiscovered DNA progenitor. While the D-configured **3** result in a defined manner from **7** (see Figure 1c), two additional configurational isomers of **12** are possible, summing up to a total of four epimers.

These results not only prove the highly selective formation of β -deoxynucleosides by activation of **9** and an in situ sugar-forming cascade (Figure 3), but also in particular the stereochemical course, which has led to the preferred formation of deoxyribose over other deoxysugars. The molecular evolutionary selection principle is clearly the minimization of possible stereoisomers. This selection corresponds to a minimization of entropy and, by definition, to a shift from equilibrium, which can be understood as a criterion for life on a molecular basis.

The RNA world hypothesis is the central consensus in the origins of life research, although many questions arising from this hypothesis have not yet been answered. Among these are the transition from RNA to DNA and the pre-eminence of **4** in all coding polymers of life. Our results finally provide a rational and provable mechanism for the formation of **3**. In a nucleobase-activated aldol reaction, a transition from all the canonical nucleobases **1a–d**, via the vinyl nucleobase intermediates **8a–d**, to the deoxyribonucleosides **3a–d** occurs. The reaction proceeds under constant, ambient conditions, in water, and basic mineral additives enhance the reaction. When using either **13**, **15**, or **11**, instead of **7**, deoxynucleoside derivatives were also detected. However, these potential DNA progenitors are ruled out because of either their

inability to polymerize or their stereochemical variability. With **12**, a potential DNA progenitor was found. Likewise, **12** could not persist, and proves the ingenuity of nature to minimize the number of possible stereoisomers (see Table S8). This work explains why D-deoxyribose is favored over all other sugars and thus opens a non-enzymatic pathway towards **3**. We therefore conclude that the DNA world did evolve much earlier than previously proposed.

Acknowledgements

We acknowledge financial support from the Ludwig-Maximilians-University Munich, the Deutsche Forschungsgemeinschaft DFG (INST 86/1807-1 FUGG, SFB 235, the Cluster of Excellence ORIGINS), and the Max-Planck-Society (Max-Planck-Fellow Research Group Origins of Life).

Conflict of interest

The authors declare no conflict of interest.

Keywords: nucleosides · origin of life · organocatalysis · polymers · prebiotic chemistry

How to cite: *Angew. Chem. Int. Ed.* **2019**, *58*, 9944–9947
Angew. Chem. **2019**, *131*, 10049–10052

- [1] J. D. Sutherland, J. N. Whitfield, *Tetrahedron* **1997**, *53*, 11493–11527.
- [2] N. Kitadai, S. Maruyama, *Geosci. Front.* **2018**, *9*, 1117–1153.
- [3] F. Goesmann, H. Rosenbauer, J. H. Bredehöft, M. Cabane, P. Ehrenfreund, T. Gautier, C. Giri, H. Krüger, L. L. Roy, A. J. Macdermott, S. Mckenna-Lawlor, U. J. Meierhenrich, et al., *Science* **2015**, *349*, aab0689.
- [4] A. Butlerow, *Justus Liebigs Ann. Chem.* **1861**, *120*, 295–298.
- [5] S. L. Miller, H. C. Urey, *Science* **1953**, *117*, 245–251.
- [6] A. Rich in *Horizons in Biochemistry* (Eds.: M. Kasha, B. Pullman), Academic Press, New York, **1962**, pp. 103–126.
- [7] W. Gilbert, *Nature* **1986**, *319*, 618.
- [8] G. F. Joyce, *Nature* **2002**, *418*, 214.
- [9] G. F. Joyce, *Nature* **1989**, *338*, 217.
- [10] C. R. Woese, *Proc. Natl. Acad. Sci. USA* **1968**, *59*, 110.
- [11] T. A. Lincoln, G. F. Joyce, *Science* **2009**, *323*, 1229.
- [12] Y. Ura, J. M. Beierle, L. J. Leman, L. E. Orgel, M. R. Ghadiri, *Science* **2009**, *325*, 73.
- [13] N. Hud, B. Cafferty, R. Krishnamurthy, L. Williams, *Chem. Biol.* **2013**, *20*, 466–474.
- [14] S. C. Kim, D. K. O'Flaherty, L. Zhou, V. S. Lelyveld, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 13318.
- [15] E. Fischer, B. Helferich, *Ber. Dtsch. Chem. Ges.* **1914**, *47*, 210–235.
- [16] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Biol.* **1972**, *67*, 25–33.
- [17] W. D. Fuller, R. A. Sanchez, L. E. Orgel, *J. Mol. Evol.* **1972**, *1*, 249–257.
- [18] R. Saladino, E. Carota, G. Botta, M. Kapralov, G. N. Timoshenko, A. Y. Rozanov, E. Krasavin, E. Di Mauro, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, E2746–E2755.
- [19] J. E. Sponer, J. Sponer, O. Novakova, V. Brabec, O. Sedo, Z. Zdrahal, G. Costanzo, S. Pino, R. Saladino, E. Di Mauro, *Chem. Eur. J.* **2016**, *22*, 3572–3586.
- [20] H. Kim, S. A. Benner, *Chem. Eur. J.* **2018**, *24*, 581–584.
- [21] H.-J. Kim, S. A. Benner, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 11315.
- [22] I. Nam, H. G. Nam, R. N. Zare, *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 36–40.
- [23] S. Becker, I. Thoma, A. Deutsch, T. Gehrke, P. Mayer, H. Zipse, T. Carell, *Science* **2016**, *352*, 833–836.
- [24] S. Becker, C. Schneider, H. Okamura, A. Crisp, T. Amatov, M. Dejmek, T. Carell, *Nat. Commun.* **2018**, *9*, 163.
- [25] M. Ferus, D. Nesvorný, J. Šponer, P. Kubelík, R. Michalčíková, V. Shestivská, J. E. Šponer, S. Civiš, *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 657–662.
- [26] D. Niether, D. Afanasenkau, J. K. G. Dhont, S. Wiegand, *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 4272.
- [27] J. E. Sponer, R. Szabla, R. W. Gora, A. M. Saitta, F. Pietrucci, F. Saija, E. D. Mauro, R. Saladino, M. Ferus, S. Civiš, J. Sponer, *Phys. Chem. Chem. Phys.* **2016**, *18*, 20047–20066.
- [28] M. Ferus, F. Pietrucci, A. M. Saitta, A. Knížek, P. Kubelík, O. Ivanek, V. Shestivska, S. Civiš, *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 4306–4311.
- [29] H. Okamura, S. Becker, N. Tiede, S. Wiedemann, J. Feldmann, T. Carell, *Chem. Commun.* **2019**, 55, 1939–1942.
- [30] S. Pizzarello, A. L. Weber, *Origins Life Evol. Biospheres* **2010**, *40*, 3–10.
- [31] C. Meinert, I. Myrgorodska, P. D. Marcellus, T. Buhse, L. Nahon, S. V. Hoffmann, L. L. S. D'Hendecourt, U. J. Meierhenrich, *Science* **2016**, *352*, 208–212.
- [32] R. Larralde, M. P. Robertson, S. L. Miller, *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 8158–8160.
- [33] I. V. Delidovich, A. N. Simonov, O. P. Taran, V. N. Parmon, *ChemSusChem* **2014**, *7*, 1833–1846.
- [34] M. W. Powner, B. Gerland, J. D. Sutherland, *Nature* **2009**, *459*, 239–242.
- [35] M. Nuevo, G. Cooper, S. A. Sandford, *Nat. Commun.* **2018**, *9*, 5276.
- [36] A. M. Poole, D. T. Logan, B.-M. Sjöberg, *J. Mol. Evol.* **2002**, *55*, 180–196.
- [37] A. Lazcano, R. Guerrero, L. Margulis, J. Oró, *J. Mol. Evol.* **1988**, *27*, 283–290.
- [38] J. Xu, N. J. Green, C. Gibard, R. Krishnamurthy, J. D. Sutherland, *Nat. Chem.* **2019**, *11*, 457–462.
- [39] M. Breugst, F. Corral-Bautista, H. Mayr, *Chem. Eur. J.* **2012**, *18*, 127–137.
- [40] E. Knoevenagel, *Ber. Dtsch. Chem. Ges.* **1898**, *31*, 738–748.
- [41] B. List, *Angew. Chem. Int. Ed.* **2010**, *49*, 1730–1734; *Angew. Chem.* **2010**, *122*, 1774–1779.
- [42] L. M. Ziurys, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12274.
- [43] H. Lammer, A. L. Zerkle, S. Gebauer, N. Tosi, L. Noack, M. Scherf, E. Pilat-Lohinger, M. Güdel, J. L. Grenfell, M. Godolt, A. Nikolaou, *Astron. Astrophys. Rev.* **2018**, *26*, 2.
- [44] N. T. Anh, O. Eisenstein, *Tetrahedron: Asymmetry* **1976**, *17*, 155–158.
- [45] J. E. Baldwin, *J. Chem. Soc. Chem. Commun.* **1976**, 734–736.
- [46] K.-U. Schöning, P. Scholz, S. Guntha, X. Wu, R. Krishnamurthy, A. Eschenmoser, *Science* **2000**, *290*, 1347.
- [47] S. Cannizzaro, *Justus Liebigs Ann. Chem.* **1853**, *88*, 129–130.
- [48] M. Tohidi, L. E. Orgel, *J. Mol. Evol.* **1989**, *28*, 367–373.
- [49] C. Triantaphylides, H.-P. Schuchmann, C. von Sonntag, *Carbohydr. Res.* **1982**, *100*, 131–141.
- [50] E. Vongerichten, *Justus Liebigs Ann. Chem.* **1902**, *321*, 71–83.
- [51] M. Pičmanová, B. L. Møller, *Glycobiology* **2016**, *26*, 430–442.

Manuscript received: March 19, 2019

Revised manuscript received: May 23, 2019

Accepted manuscript online: May 26, 2019

Version of record online: June 24, 2019