Chemistry–A European Journal

Communication
doi.org/10.1002/chem.202002929

Chemistry
Europe
European Chemical
Societies Publishing

▌ **Oligonucleotides**

Amino Acid Modified RNA Bases as Building Blocks of an Early Earth RNA-Peptide World

Milda Nainytė,[a] Felix Müller,[a] Giacomo Ganazzoli,[a] Chun-Yin Chan,[a] Antony Crisp,[a] Daniel Globisch,[b] and Thomas Carell*[a]

Abstract: Fossils of extinct species allow us to reconstruct the process of Darwinian evolution that led to the species diversity we see on Earth today. The origin of the first functional molecules able to undergo molecular evolution and thus eventually able to create life, are largely unknown. The most prominent idea in the field posits that biology was preceded by an era of molecular evolution, in which RNA molecules encoded information and catalysed their own replication. This RNA world concept stands against other hypotheses, that argue for example that life may have begun with catalytic peptides and primitive metabolic cycles. The question whether RNA or peptides were first is addressed by the RNA-peptide world concept, which postulates a parallel existence of both molecular species. A plausible experimental model of how such an RNA-peptide world may have looked like, however, is absent. Here we report the synthesis and physicochemical evaluation of amino acid containing adenosine bases, which are closely related to molecules that are found today in the anticodon stem-loop of tRNAs from all three kingdoms of life. We show that these adenosines lose their base pairing properties, which allow them to equip RNA with amino acids independent of the sequence context. As such we may consider them to be living molecular fossils of an extinct molecular RNA-peptide world.

The RNA-peptide co-evolution hypothesis describes the emergence of self-replicating molecules that contained amino acids and RNA.[1] At the macromolecular level, this tight coexistence of peptides and RNA is established in the ribosome, where encoding and catalytic RNA is supported by proteins.[2] Although

we cannot delineate how such an early RNA-peptide world may have looked like, it seems not too implausible to assume that some of the molecular components may have survived until today as vestiges of this extinct world.[3] tRNAs derived from all three kingdoms of life contain a large number of modified bases,[4] and some of them are indeed modified with amino acids.[3] The most wide spread amino acid modified bases are adenosine nucleosides, in which the amino acid is linked via urea connector to the N^6-amino group of the heterocycle as depicted in Figure 1a. Particularly ubiquitous are adenosine modifications containing the amino acids threonine (t⁶A)[5–7] and glycine (g⁶A),[8] together with hn⁶A.[9, 10] Based upon recent phylogenetic analyses and the fact that t⁶A is found in all three kingdoms of life, it has been suggested that such amino acid modified bases were already present in the last universal common ancestor (LUCA), from which all life
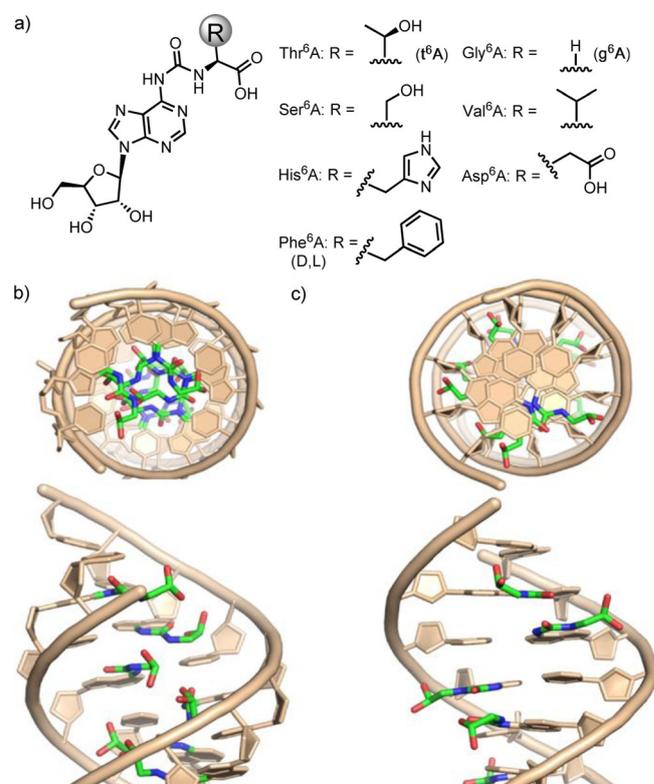
[a] M. Nainytė, F. Müller, G. Ganazzoli, C.-Y. Chan, A. Crisp, Prof. Dr. T. Carell
Department of Chemistry, LMU München
Butenandtstr. 5–13, 81377 München (Germany)
E-mail: thomas.carell@lmu.de

[b] Dr. D. Globisch
Department of Medicinal Chemistry, Uppsala University
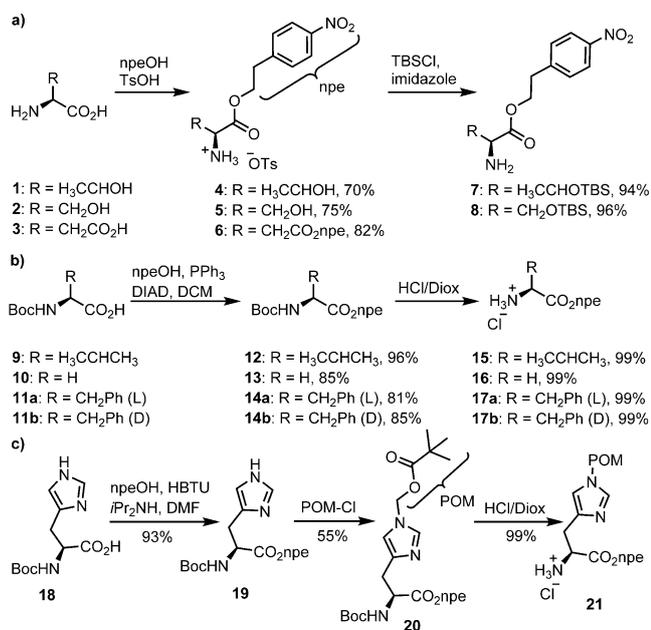Husargatan 3, 75123 Uppsala (Sweden)

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under:
https://doi.org/10.1002/chem.202002929.

Figure 1. (a) Depiction of the amino acid modified A-bases (aa⁶A) together with computer visualizations that show how such bases may reside in an (b) A-form RNA duplex and a (c) B-form DNA duplex. The sequence used for the visualization is: 5′-CAUAUAUAUAUG-3′ with A = g⁶A.

Chemistry–A European Journal

Communication
doi.org/10.1002/chem.202002929

Chemistry Europe
European Chemical
Societies Publishing

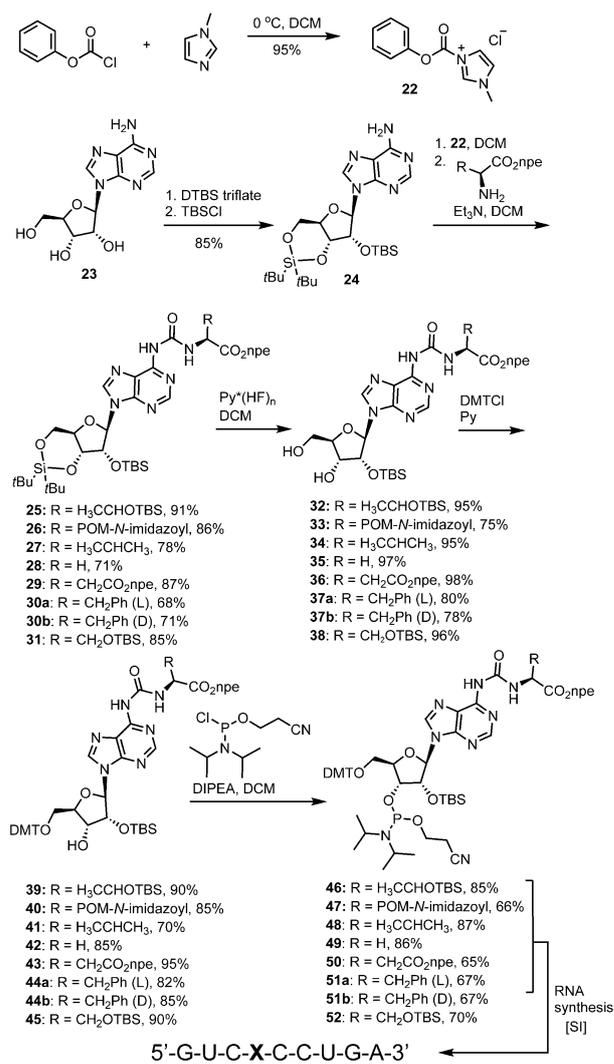forms descended.[11–14] t6A is for example today found in nearly all ANN decoding tRNAs.[15] We recently reported a plausible prebiotic route to some of these amino acid modified A-bases, which strengthens the idea that they could indeed be living chemical fossils of the extinct RNA-peptide world.[16] Despite the interesting philosophical genotype-phenotype dualism that characterizes these structures and their contemporary importance for the faithful decoding of genetic information, a general synthesis of aa6A modified bases (Figure 1a) and a systematic study of their properties is lacking.

Here we report the synthesis of a variety of aa6A nucleosides with canonical amino acids (aa = Asp, Gly, His, Phe, Thr,[17] Ser, Val), their incorporation into DNA and RNA and an investigation of how they influence the physicochemical properties of oligonucleotides. We were particularly interested to study how they might affect the stability of RNA and DNA. The computer visualization shows that in A-form RNA (Figure 1b), the amino acid part of the aa6A base would need to reside inside the helix, shielded from the outside. In the B-form DNA one could imagine a decoration of the major groove with the amino acid side chains as depicted in Figure 1c.

In the Schemes 1 and 2 we show the synthesis of the different urea linked amino acid A-derivatives (aa6A). We first prepared the amino acid components for the coupling to the A-nucleoside (Scheme 1). Our starting points for Thr6A, Ser6A and Asp6A were the free amino acids 1–3, in which we first transformed all carboxylic acids into the p-nitrophenylethyl esters (npe, 4–6).[17] The hydroxy groups of the Thr and Ser compounds were finally protected as TBS-ethers to give the final products 7 and 8 (Scheme 1a). For Val, Gly and Phe we started with the Boc-protected amino acids 9–11, which we also converted into the npe-esters 12–14 using Mitsunobu type

Scheme 2. Synthesis of phosphoramidite building blocks of Thr6A, Ser6A, Asp6A, Val6A, Gly6A, Phe6A and His6A and their incorporation into RNA.

chemistry[18] followed by acidic (4 M HCl in dioxane) Boc-deprotection to give the amino acid products 15–17 (Scheme 1b).[19] For His6A, we again started with the Boc-protected amino acid 18 (Scheme 1c) and used HBTU activation to generate the npe ester 19. Protection of the imidazole N^τ with POM-chloride followed again by Boc-deprotection furnished the ready to couple amino acid 21.

The connection of the amino acid with the A-nucleoside via the urea moiety was next carried out as depicted in Scheme 2. We first treated phenyl chloroformate with N-methylimidazole to obtain the 1-N-methyl-3-phenoxycarbonyl-imidazolium chloride (22).[20] Adenosine was converted in parallel into the cyclic 3′,5′-silyl protected nucleoside, followed by conversion of the 2′-OH group into the TBS-ether.[21] The reaction of compound 24 with the activated carbonate and the corresponding amino acid, provided in all cases the amino acid coupled products 25–31 in good to excellent yields. Subsequent cleavage of the

Scheme 1. Synthesis of the amino acid building blocks as needed for the coupling to the nucleoside A to give Thr6A, Ser6A, Asp6A, Val6A, Gly6A, Phe6A and His6A.

Chemistry–A European Journal

Communication
doi.org/10.1002/chem.202002929

Chemistry Europe
European Chemical
Societies Publishing

cyclic silylether with HF·pyridine complex,[22,23] protection of the 5′-OH group with dimethoxytritylchloride (DMTCl)[24] allowed the final conversion of the compounds into the corresponding phosphoramidites **46–52**. Standard solid phase RNA chemistry[25–31] was subsequently employed to prepare RNA strands containing the individual aa⁶A nucleosides stably embedded. The standard RNA synthesis protocol did not require any adjustment. In all cases we observed fair coupling of the aa⁶A phosphoramidites and no decomposition during deprotection. Deprotection required three steps. First, with DBU in THF at r.t. for 2 h we cleaved the npe-protecting group. Second, we deprotected the bases and cleaved from the solid support with aqueous NH₃/MeNH₂. Finally, we removed the 2′-silyl group with HF in NEt₃.

In order to investigate how aa⁶A bases would affect the stability of DNA duplexes we also prepared as a representative molecule t⁶dA as depicted in Scheme 3. To this end we first acetyl-protected dA **53**,[32] performed the coupling of the protected threonine with the activated carbonate **22**, cleaved the acetyl groups and converted the nucleoside subsequently into the 5′-DMT protected phosphoramidite **57**. The purification of compound **57** was quite difficult due to its high polarity. We needed to use rather polar mixture of EtOAc/Hex (2/1) as the mobile phase for the chromatographic separation. This provided the phosphoramidite **57**, however the material had a lower purity in comparison to the RNA phosphoramidites. Nevertheless, solid phase DNA synthesis and deprotection of the DNA strand **ODN1** proceeded again smoothly and in high yields.
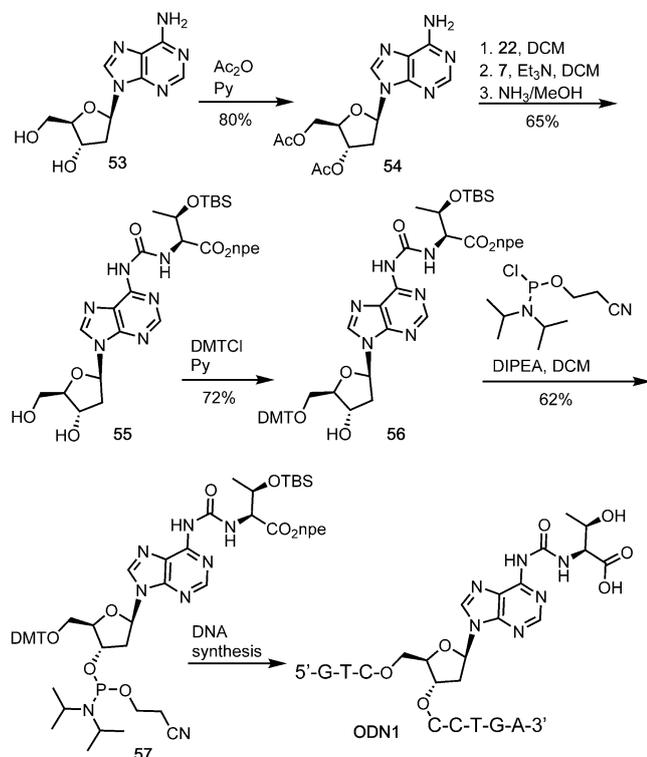
Figure 2a shows as an example the raw HPL-chromatograms of **ON1** (RNA strand with embedded t⁶A) and the corresponding chromatogram after purification (inset) together with the obtained MALDI-TOF mass spectrum (Figure 2b). The chromatograms of the raw material show a good quality of the obtained RNA material. The analytical chromatogram after purification and the MALDI-TOF data prove the purity of the finally obtained RNA oligonucleotide and the integrity of the t⁶A-containing RNA strand.

Figure 2c and 2d show the same data set for the t⁶dA containing DNA oligonucleotide (**ODN1**), proving again the successful synthesis of t⁶dA containing oligonucleotide. The aa⁶(d)A nucleosides can exist in two different conformations.[33] The first, s-trans, maintains the Watson–Crick hydrogen bonding capabilities with the urea amino acid oriented towards the imidazole ring system (Figure 3a). This allows formation of a Hoogsteen type 7-membered ring H-bond with the N^7. In the corresponding s-cis-conformation, the urea amino acid orients towards the Watson–Crick side thereby establishing a typically strong intramolecular 6-membered H-bond with N^1 (Figure 3b). In order to investigate if the embedding of the amino acid would enforce s-trans-conformation and hence Watson–Crick H-bonding, we measured melting points of all aa⁶A containing RNA strands and of the t⁶dA containing DNA strand hybridized to the corresponding counter strands (Figure 3). In the RNA:RNA situation we noted for all aa⁶A strands that we investigated, a single clear melting point, showing that only one conformer of the aa⁶A base likely exists in the RNA:RNA duplexes. In situation where the aa⁶A base exists in two different stable conformations, one would expect a more complex melting behaviour. In all cases we saw that the melting point is strongly reduced by 10–15 °C. When we embedded two aa⁶A building blocks into a short RNA strand no duplex formation
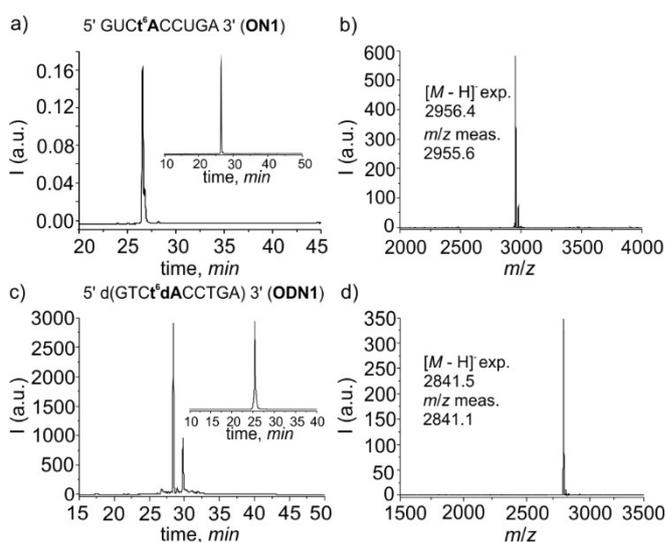

Scheme 3. Synthesis of t⁶dA phosphoramidite and its incorporation into DNA.

Figure 2. (a) Raw-HPL chromatogram of **ON1**, with the inset showing the HPL-chromatogram of purified **ON1**; (b) MALDI-TOF mass spectrum of **ON1** after purification; (c) raw-HPL chromatogram of **ODN1**, with the inset showing the HPL chromatogram of purified **ODN1**; (d) MALDI-TOF mass spectrum of **ODN1** after purification.
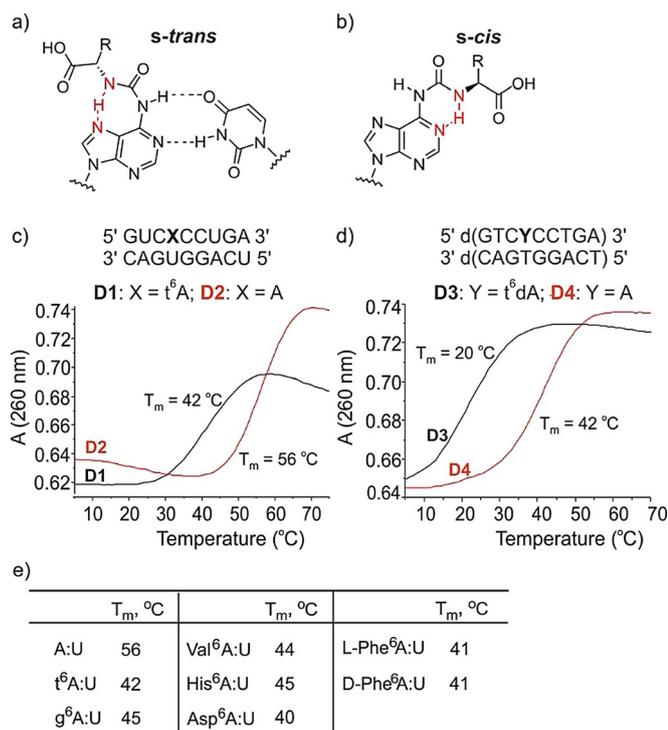
Chemistry–A European Journal

Communication
doi.org/10.1002/chem.202002929

Chemistry Europe
European Chemical
Societies Publishing

Figure 3. (a, b) Possible conformation, base pairing and intramolecular H-bond of aa⁶A; (c, d) melting curves measured for t⁶A containing RNA:RNA duplexes and of a t⁶dA containing DNA:DNA duplex in comparison with the duplexes containing canonical (d)A:(d)T base pairs; (e) table of the determined melting points.
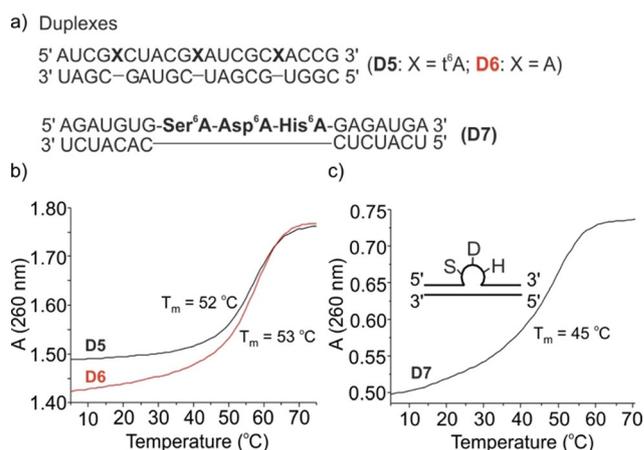

Figure 4. (a) Depiction of the RNA structures containing aa⁶A nucleobases in extrahelical positions forming either three little bulges or assembling a Ser-Asp-His triad known as the catalytic triad in serine proteases; (b, c) depiction of melting curves of duplexes D5, D6, D7; S: serine, D: aspartate, H: histidine.

was obtained. Even stronger reduction of the melting point was observed for the DNA duplex containing one t⁶dA. Here, we also saw just one sharp melting point and a reduction of the T_m by over 20 °C. These data show that the aa⁶A bases and among them t⁶A and g⁶A are unable to base pair. Although we have no direct proof of the structure the data argue for a preferred s-cis-conformation (Figure 3 b) in agreement with the literature.[34]

This conclusion is also supported by the observation that irrespective of the chirality of the attached amino acid (L- versus D-Phe⁶A), we measured the same melting temperature. This would not be expected if the s-trans-conformation and base pairing would be possible. These data suggest that aa⁶A nucleosides within RNA position a given amino acid outside the A-form helix in an unpaired situation and hence independent from the counterbase. As such, multiple aa⁶A containing RNA strands would be structures in which the RNA part is decorated by the amino acid side chains. In order to show that RNA-structures containing multiple amino acids as representatives of an RNA-peptide world can stably form, we prepared two RNA duplexes (Figure 4). In the first (D5), we placed three t⁶A bases as extra bases in an otherwise undisturbed RNA duplex. Indeed, now the stability of this duplex was indistinguishable from the same construct containing just canonical bases (D6). Finally, we prepared an RNA duplex D7, in which we placed the amino acids Ser-Asp-His directly next to each other to simulate what is known in the peptide world as the catalytic triad

present in serine peptidases.[35] Again in this case a stable duplex structure forms with the three aa⁶A bases creating a loop. Although we do not show any catalytic activity here, we believe that it is easily imaginable that if these amino acids are properly positioned in a stably folded RNA the structure could gain catalytic properties.

The melting data show, that aa⁶A bases alone are unable to establish base pairing, which hinder them to encode sequence information. On the other side, these bases allow the incorporation of amino acids into RNA structures irrespective of the counterbase. Because RNAs are mostly stably folded structures in which many bases are not involved in any base pairing or establish no Watson–Crick interactions the amino acid adenosine nucleosides allow the stable incorporation of amino acid functionality into RNA.

In summary, here we investigated the synthesis and properties of aa⁶A nucleoside-amino acid conjugates, some of which (t⁶A, g⁶A, hn⁶A) are today found as key components in the tRNAs of many species. In these tRNAs the aa⁶A nucleosides reside at the general purine position 37 adjacent to the anticodon loop. They are not involved in base pairing but fine tune the codon-anticodon interaction to enable faithful translation of information into a peptide sequence.[36] Here we show that these bases are indeed unable to base pair. They have to be placed outside the pairing regime that is needed for RNA folding. As such they function as anchors that allow the connection of amino acid to RNA structures independent of the counterbase. The side chains are then available to equip RNA with additional functions that might have been beneficial in an early RNA-peptide world. The fact that aa⁶A nucleosides are stable structures and until today broadly found in today's RNA make them prime candidates to develop idea about the chemical constitution of the vanished RNA-peptide world.

Chemistry–A European Journal

Communication
doi.org/10.1002/chem.202002929

Chemistry Europe
European Chemical
Societies Publishing

Acknowledgements

Conflict of interest

The authors declare no conflict of interest.

[1] P. T. S. van der Gulik, D. Speijer, *Life* **2015**, *5*, 230–246.
[2] D. L. J. Lafontaine, D. Tollervey, *Nat. Rev. Mol. Cell Biol.* **2001**, *2*, 514–520.
[3] H. Grosjean, E. Westhof, *Nucleic Acids Res.* **2016**, *44*, 8020–8040.
[4] M. A. Machnicka, K. Milanowska, O. O. Oglou, E. Purta, M. Kurkowska, A. Olchowik, W. Januszewski, S. Kalinowski, S. Dunin-Howkawicz, K. M. Rother, M. Helm, J. M. Bujnicki, H. Grosjean, *Nucleic Acids Res.* **2013**, *41*, D262–D267.
[5] G. B. Chheda, R. H. Hall, J. Mozejko, D. I. Magrath, M. P. Schweizer, L. Stasiuk, P. R. Taylor, *Biochemistry* **1969**, *8*, 3278–3282.
[6] S. Takemura, M. Murakami, M. Miyazaki, *J. Biochem.* **1969**, *65*, 489–491.
[7] D. M. Powers, A. Peterkofsky, *Biochem. Biophys. Res. Commun.* **1972**, *46*, 831–838.
[8] M. P. Schweizer, K. McGrath, L. Bacznyskyj, *Biochem. Biophys. Res. Commun.* **1970**, *40*, 1046–1052.
[9] H. Grosjean in *DNA and RNA Modification Enzymes: Structure, Mechanism, Function and Evolution* (Ed.: H. Grosjean), Landes Bioscience, Austin, Texas, pp. 1–18.
[10] D. M. Reddy, P. F. Crain, C. G. Edmonds, R. Gupta, T. Hashizume, K. O. Stetter, F. Widdel, J. A. McCloskey, *Nucleic Acids Res.* **1992**, *20*, 5607–5615.
[11] M. Di Giulio, *J. Theor. Biol.* **1998**, *191*, 191–196.
[12] H. Grosjean, V. de Crécy-Lagard, G. R. Björk, *Trends Biochem. Sci.* **2004**, *29*, 519–522.
[13] E. Szathmáry, *Trends Genet.* **1999**, *15*, 223–229.
[14] L. Perrochia, E. Crozat, A. Hecker, W. Zhang, J. Bareille, B. Collinet, H. van Tilbeurgh, P. Forterre, T. Basta, *Nucleic Acids Res.* **2013**, *41*, 1953–1964.
[15] P. C. Thiaville, D. Iwata-Reuyl, V. de Crécy-Lagard, *RNA Biol.* **2014**, *11*, 1529–1539.
[16] C. Schneider, S. Becker, H. Okamura, A. Crisp, T. Amatov, M. Stadlmeier, T. Carell, *Angew. Chem. Int. Ed.* **2018**, *57*, 5943–5946; *Angew. Chem.* **2018**, *130*, 6050–6054.
[17] V. Boudou, J. Langridge, A. van Aerschot, C. Hendrix, A. Millar, P. Weiss, P. Herdewijn, *Helv. Chim. Acta* **2000**, *83*, 152–161.
[18] O. Mitsunobu, M. Yamada, *Bull. Chem. Soc. Jpn.* **1967**, *40*, 2380–2382.
[19] G. Leszczynska, P. Leonczak, A. Dziergowska, A. Malkiewicz, *Nucleosides Nucleotides Nucleic Acids* **2013**, *32*, 599–616.
[20] F. Himmelsbach, B. S. Schultz, T. Trichtinger, R. Charubala, W. Pfleiderer, *Tetrahedron* **1984**, *40*, 59–72.
[21] V. Serebryany, L. Beigelman, *Tetrahedron Lett.* **2002**, *43*, 1983–1985.
[22] B. M. Trost, C. G. Caldwell, *Tetrahedron Lett.* **1981**, *22*, 4999–5002.
[23] M. Sekine, S. Limura, K. Furusawa, *J. Org. Chem.* **1993**, *58*, 3204–3208.
[24] H. Schaller, G. Weinmann, B. Lerch, H. G. Khorana, *J. Am. Chem. Soc.* **1963**, *85*, 3821–3827.
[25] H. Seliger, D. Zeh, G. Azuru, J. B. Chattopadhyaya, *Chem. Scr.* **1983**, *22*, 95–102.
[26] M. H. Caruthers, D. Dellinger, K. Prosser, A. D. Brone, J. W. Dubendorff, R. Kierzek, M. Rosendahl, *Chem. Scr.* **1986**, *26*, 25–30.
[27] R. Kierzek, M. H. Caruthers, C. E. Longfellow, D. Swinton, D. H. Turner, S. M. Freier, *Biochemistry* **1986**, *25*, 7840–7846.
[28] I. Hirao, M. Ishikawa, K. Miura, *Nucleic Acids Symp. Ser.* **1985**, *16*, 173–176.
[29] H. Tanimura, T. Fukuzawa, M. Sekine, T. Hata, J. W. Efcavitch, G. Zon, *Tetrahedron Lett.* **1988**, *29*, 577–578.
[30] H. Tanimura, M. Sekine, T. Hata, *Nucleosides Nucleotides* **1986**, *5*, 363–383.
[31] H. Tanimura, T. Imada, *Chem. Lett.* **1990**, *19*, 1715–1718.
[32] E. M. van der Wenden, J. K. von Frijtag Drabbe Künzel, R. A. A. Mathot, M. Danhof, A. P. IJzerman, W. Soudijn, *J. Med. Chem.* **1995**, *38*, 4000–4006.
[33] R. Parthasarathy, J. M. Ohrt, G. B. Chheda, *Biochem. Biophys. Res. Commun.* **1974**, *60*, 211–218.
[34] F. V. Murphy IV, V. Ramakrishnan, A. Malkiewicz, P. F. Agris, *Nat. Str. Mol. Biol.* **2004**, *11*, 1186–1191.
[35] L. Polgár, *Cell. Mol. Life Sci.* **2005**, *62*, 2161–2172.
[36] P. C. Thiaville, R. Legendre, D. Rojas-Benítez, A. Baudin-Baillieu, I. Hatin, G. Chalancon, A. Glavic, O. Namy, V. de Crécy-Lagard, *Microb. Cell.* **2016**, *3*, 29–45.