GDCh

Communications

Angewandte
Chemie
International Edition
www.angewandte.org

VIP | **Origin of Life** *Very Important Paper*

Enzyme-Free Copying of 12 Bases of RNA with Dinucleotides

Gabrielle Leveau[+], Daniel Pfeffer[+], Bernhard Altaner, Eric Kervio, Franziska Welsch, Ulrich Gerland, and Clemens Richert*

Abstract: The synthesis of complementary strands is the reaction underlying the replication of genetic information. It is likely that the earliest self-replicating systems used RNA as genetic material. How RNA was copied in the absence of enzymes and what sequences were most likely to have supported replication is not clear. Here we show that mixtures of dinucleotides with C and G as bases copy an RNA sequence of up to 12 nucleotides in dilute aqueous solution. Successful enzyme-free copying occurred with in situ activation at 4 °C and pH 6.0. Dimers were incorporated in favor of monomers when both competed as reactants, and little misincorporation was detectable in mass spectra. Simulations using experimental rate constants confirmed that mixed C/G sequences are good candidates for successful replication with dimers. Because dimers are intermediates in the synthesis of longer strands, our results support evolutionary scenarios encompassing formation and copying of RNA strands in enzyme-free fashion.

Nature uses semiconservative replication to pass genetic information from one generation to the next. The molecular basis of replication is the synthesis of complementary strands, directed by template strands. Genetic copying of extant biology relies on polymerase-catalyzed formation of phosphodiester bonds via the nucleophilic attack of the 3′-terminus of the growing strand on a nucleoside triphosphate engaging in Watson–Crick base pairing with the template. Genetic copying is best known for DNA,[1] but, to this day, RNA viruses use RNA-directed RNA polymerization when replicating their genome,[2] and so may have the earliest self-replicating systems of evolution.[3] The latter probably did so

without the help of enzymes, using activated nucleotides with organic leaving groups.[4,5]

While, conceptually, enzyme-free copying of RNA is straightforward, even copying of sequences just long enough to encode a very short ribozyme[6] remains challenging experimentally. One reason for this is the sequence dependence of the enzyme-free reaction. Early reports used homopolymers as templates,[7] with the most successful example being poly(C), on which oligoguanylates form.[8] Unfortunately, homopolymer sequences are not useful as genes. When mixed sequences were studied, sequences not dominated by cytidylic acid were found to be poor templates.[9,10] To this day, even extension of a primer by four nucleotides succeeds only in rare cases,[11] or for immobilized templates with repeated replenishing of the monomers.[12] A combination of preactivated monomers and trimers, furnished with the best organic leaving groups known, still did not give "reads" longer than seven nucleotides.[13] Also, preactivated trimers were found to be poor building blocks.[14] Enzyme-free ligation of longer RNA is an unlikely alternative, as this reaction is known to be less efficient for RNA than for DNA.[15] Further, the best sequences remain to be identified. Systematic searches for sequences that favor copying have been proposed,[16] and performed,[11] but have not yielded a system that undergoes copying for a stretch of at least 10 bases.

Naïvely, one may consider primer extension just a bimolecular reaction. If so, an increase in concentration should increase the rate of the reaction and should lead to high yields. This was found not to be the case. Rather, as the concentration increases, spent monomers, i.e. the hydrolysis products of activated monomers, increasingly inhibit incorporation by blocking the extension site.[17] This problem can be reduced by periodically replacing spent monomers[12] or by re-activating free nucleotides in situ.[18] Further, a poor template effect, caused by weak base pairing, can be reduced by employing strands that provide additional stacking interactions to incoming building blocks.[19] Finally, misincorporations that cause stalling,[20] may be avoided by relying on high fidelity copying with bases that pair well.

These considerations led to our current study, which employs in situ activation, strongly pairing building blocks, and sequences related to those known to support enzyme-free replication.[21–23] To be plausible, we also chose conditions that support strand formation *and* copying,[18] as depicted in Figure 1. Critically, we opted for both mono- and dinucleotides[24] to overcome the poor performance of known systems. Here we report that dimer-containing mixtures lead to extension of a primer by up to 12

[*] G. Leveau,[+] Dr. D. Pfeffer,[+] Dr. E. Kervio, F. Welsch, Prof. C. Richert
Institute of Organic Chemistry, University of Stuttgart
70569 Stuttgart (Germany)
E-mail: lehrstuhl-2@oc.uni-stuttgart.de

Dr. B. Altaner, Prof. Dr. U. Gerland
Physics of Complex Biosystems, Technical University Munich
85748 Garching (Germany)

[+] These authors contributed equally to this work.
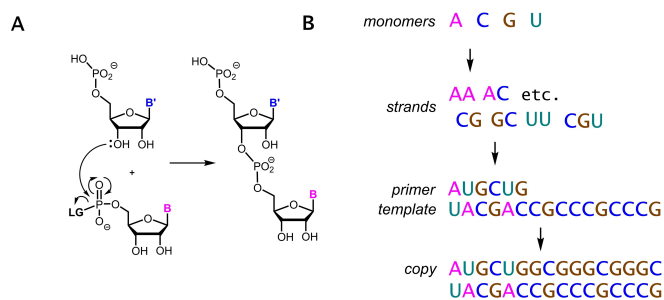
Figure 1. A) Phosphodiester formation as the molecular basis of strand formation and genetic copying. B) Putative steps of molecular evolution from mononucleotides to building blocks for copying, and copying itself.

nucleotides in a row, and that the sequences identified have the potential to support replication, as suggested by the results of an in silico study.

Figure 2 shows the RNA sequences employed in our study. All assays were ′single-run′ experiments, in which the RNA strands and unactivated mono- or dinucleotides were dissolved in homogeneous aqueous condensation buffer and allowed to react at 4°C without feeding fresh reagents or starting materials. The buffer contained 0.4 M 1-ethyl-3-(3-

dimethylaminopropyl) carbodiimide (EDC) as condensation agent, an equimolar amount of 1-ethylimidazole as organocatalyst, and 0.08 M magnesium chloride as the only other salt. Aliquots drawn after stated intervals were analyzed by MALDI-TOF mass spectrometry under conditions that allow for quantitative detection.[25] Calibration experiments were performed for all major products to correct for differences in desorption and ionization, using synthetic RNA strands, and the presence of the main extension products was confirmed by HPLC for a typical product mixture (see Chapter 4 of the Supporting Information).

The first experiments employed template **1** and primer **2**, and the template region to be copied was ten nucleotides in length. When mononucleotides CMP and GMP were used as building blocks, little extension to products **3**–**12** was detected, with singly extended **3** being the main product (Figure 3A and Figure S7a in the Supporting Information). When dimers CG and GG were employed instead, extension products up to the full ten nucleotides of the templating region appeared in the MALDI spectrum (Figure 3B). The peak for compound **12** indicated an overall yield of 18% for the five-step sequence. The corresponding assay with a mixture containing both monomers and dimers yielded a similar result (Figure 3C), with minor peaks for the
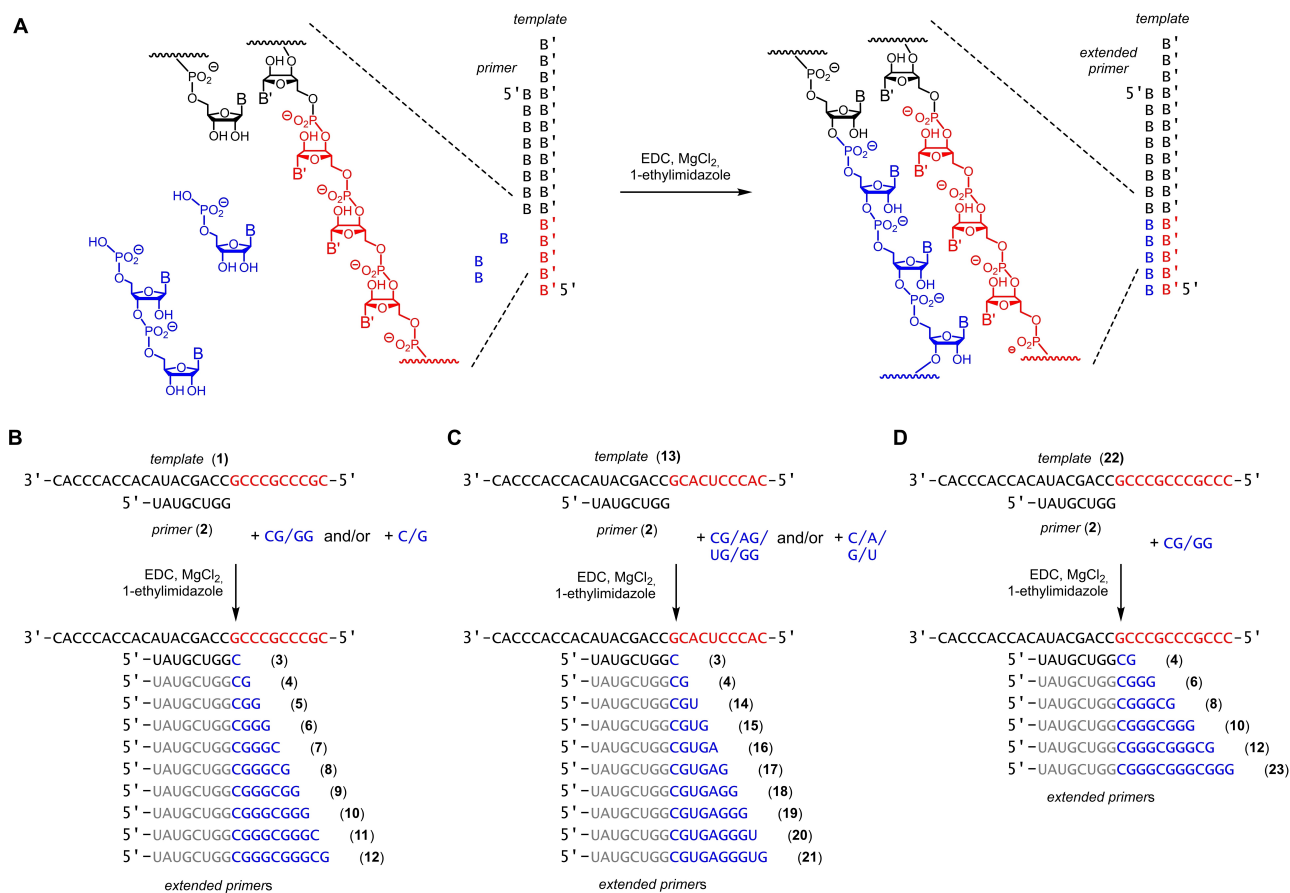
Figure 2. Genetic copying reactions studied, with regions to be copied in red and regions of products that are copies in blue. Conditions: 45 μM primer, 60 μM template, monomers and/or dimers 0.5–2 mM each, 0.4 M EDC, 0.4 M 1-ethylimidazole, 0.08 M MgCl₂, pH 6 and 4°C.
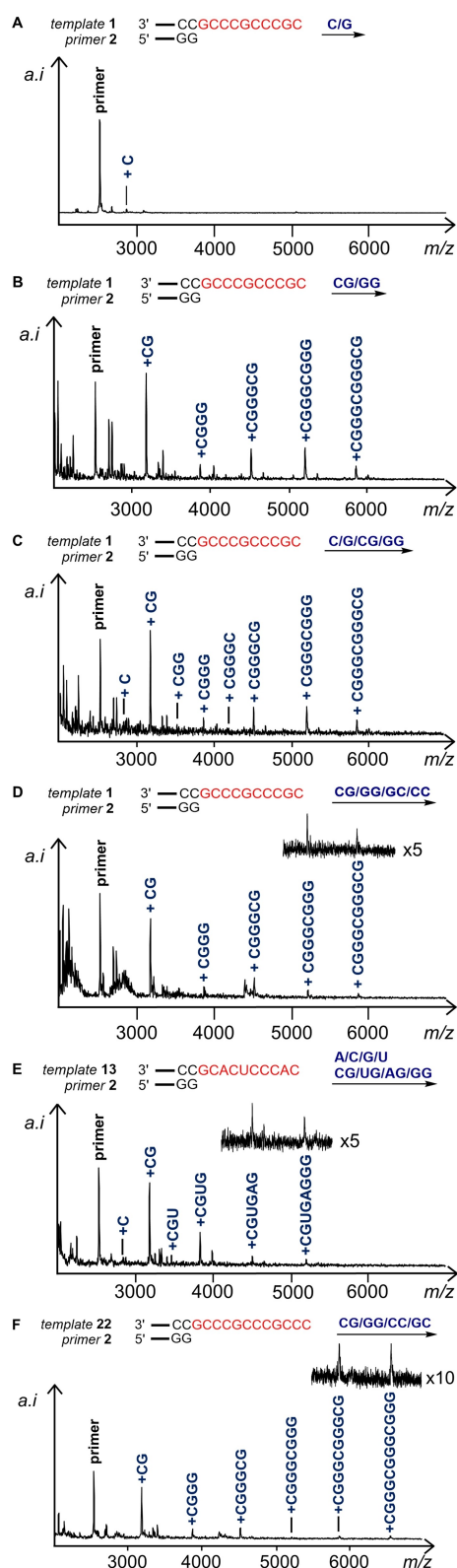
Figure 3. Genetic copying with the primer/template combinations shown in the upper part of each panel, as detected by MALDI-TOF MS. Building blocks were 2 mM each, except for E (0.5 mM each). The building blocks were: A) C and G; B) CG and GG; C) C, G, CG, and GG; D) CG, GG, GC, and CC; E) A, C, G, U, AG, CG, GG, and UG; F) CG, GG, GC, and CC. Conditions: 45 µM primer, 60 µM template, 0.4 M EDC, 0.4 M 1-EtIm, 0.08 M MgCl$_2$, pH 6, 13 d (A–E) or 18 d (F) at 4°C.

products with an odd number of newly added nucleotides resulting from NMP incorporation.

When a mixture of all four possible dinucleotides in a C/G-based genetic system was employed, full-length product was again detectable by MALDI mass spectrometry (Figure 3D). This was an important result because such a mix is more plausible, assuming statistical oligomerization as the source for strands. Further, self-pairing between dimers could have suppressed template binding and thus successful copying of the desired sequence.

Next, we asked whether the dimer-based copying reaction tolerates a modest level of weakly pairing bases (A and U). For this, template **13** was employed, which features a templating region with all four canonical nucleobases, three of which are either A or U. Again, copying with mononucleotides was largely unsuccessful, with less than 30% conversion to any products. Strands **3** and **4** gave the only significant peaks in a spectrum acquired after 13 d reaction time (Figure S7b, Supporting Information). When dimers AG, CG, GG and UG were present in the reaction mixture, strands up to **19** were detectable in the spectrum after the same time span (Figure 3E), even though eight components were used and the concentration of dimers was only 500 µM each. While these results show that a few weakly pairing bases are tolerated, our study on the individual incorporation of all possible dinucleotide sequences gave yields as low as 3% for some dimers consisting of A and U only, and just 56% for the most favorable dimer (AU),[24] making it unlikely that A/U-rich sequences will be copied successfully under the current assay conditions.

Besides base pairing strength, there is an effect of concentration on the yield of copying. Exploratory experiments with a model system, using just two equivalents of dimer CG and ten-fold dilution of the reaction mixture, resulted in a roughly ten-fold decrease in conversion to product, as shown in Figure S8 of the Supporting Information.

In the last assay of the experimental part of our study, we used template **22**, featuring a templating stretch of a dozen bases. All four C/G-containing dimers were added as building blocks, and the reaction was allowed to proceed for 18 d in the cold. Even though the full length product was ever more difficult to separate from the template strand (and thus to detect), with a calculated UV melting point of 94°C at the Mg^{2+} concentration chosen, a clearly visible peak, corresponding to 11% overall conversion, was detected for the desired product (**23**), extended by twelve bases total (Figure 3F).

To gain insights into how unique the sequences employed are in their ability to support copying, we performed an exploratory study on how the template sequence affects enzyme-free copying with dimers by a computational approach. For this, we first extracted effective rate constants for the ligations of CG, GG, GC, and CC from kinetic data for single-step primer extension experiments with those dimers (see Supporting Information). Based on these rate constants, we inferred the times that would be required to copy each of the 1024 possible template sequences of ten bases consisting only of C/G nucleotides. Although cycliza-

Angew. Chem. Int. Ed. **2022**, *61*, e202203067 (3 of 5)

© 2022 The Authors. Angewandte Chemie International Edition published by Wiley-VCH GmbH

tion of dinucleotides can be a significant side reaction,[26,27,24] we did not correct for this phenomenon, as a previous study had shown similar levels of cyclization for all four C/G-containing dinucleotides ($\leq$ 50–70 %, even after 20 d).[24] Figure 4A shows the distribution of the calculated copying times, with the sequences plotted in 'classes' containing the same number of G nucleotides.

As the plot of this exploratory data shows, experimental template **1** ranks best among the decamer sequences containing three G nucleotides, together with nine other such sequences with equal predicted copying time. For these ten sequences, the calculated copying time is only about 20 % longer than for the fastest-copying poly(C) sequence. Moreover, the copy of the experimental sequence, i.e., its reverse complement, also appears to be well suited for copying (in Figure 4A, the circles containing sequence **1** and its reverse complement are marked in red). In order to rank sequences according to their "replicability", we also computed the replication time for each sequence, i.e., the sum of the copying times for the sequence and for its reverse complement, see Figure 4B. Here, the replication time of the experimental template **1** (red circle) is within 45 % of the sequence $(GC)_5$, which marks the global optimum for enzyme-free replication with dimers. In contrast to the periodic optimal sequence $(GC)_5$, which is not useful as a gene, the experimental sequence **1**, together with its 9 equivalently fast replicating sequences (all consisting of three GC and two CC dimers), provides an example of a prebiotically plausible set of sequences that can encode information and, at the same time, are kinetically well suited for enzyme-free replication. As seen in Figure 4B, this set could be further extended without increasing replication

times, by including the five best-replicating sequences containing four G nucleotides, and the 20 second-best replicating sequences containing five G nucleotides.

Taken together, our data provides a fascinating glimpse at what RNA sequences are most likely to undergo enzyme-free replication. Because in situ activation was employed that allows for both strand formation and copying,[18] a "holistic" scenario now seems realistic that starts from ribonucleotides and evolves into a primitive genetic system with traits that make it fit for replication. Several challenges remain, however, before replication can be demonstrated. Among them is the strand separation problem caused by product strands pairing strongly with the template, inhibiting the next phase of genetic copying. Possible solutions to this are thermal cycling[28] or the use of organic salts to slow down re-annealing of strands.[29] Another challenge that remains to be addressed is the regioselectivity of phosphodiester formation. Template-directed reactions are known to be more regioselective than oligomerizations,[30,31] but a yet to be determined percentage of 2′-5′-linkages may still be formed in genetic copying with dimers, and future studies should tackle this issue.

Our results can also be discussed in light of the difficulties experienced in earlier studies. Several factors might have contributed to the successful copying of longer stretches of sequence under our conditions. The strongly pairing dimers may have formed a non-covalent helix, in which each dimer aids the binding of a neighboring building block. As a consequence, the residence time of the building blocks probably became long enough to allow for even sluggish reactions, such as the ligation of unmodified RNA strands, to proceed. The well-binding dimers appear to displace monomers, which would otherwise cause significant levels of misincorporation. Overall, dimers appear to possess a favorable combination of binding strength and reactivity that allows for high fidelity copying of longer stretches of RNA.

In conclusion, the copying of RNA sequences up to 12 nucleotides in length was observed, starting from unactivated dinucleotides containing C and G as the only bases, with little misincorporation. To the best of our knowledge, this is the longest "read" of an RNA sequences capable of encoding genetic information by enzyme- or ribozyme-free copying to date. Apparently, dimers, i.e. the very first products of oligomerization reactions, can drive a process that has been difficult to demonstrate using mononucleotides or trimers. The results from the in silico exploration of sequence space indicate that the mixed sequences employed here are suitable for both copying and back-copying, the two steps necessary for replication. We have reason to believe that even longer reads than the twelve-nucleotide stretch of sequence copied here may be achievable using dimers, and we are actively pursuing research in this direction.
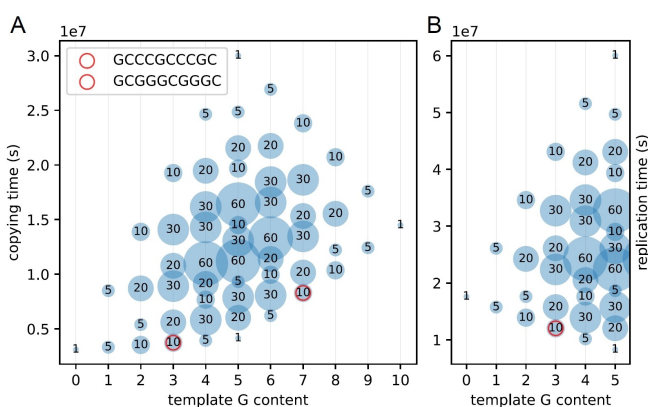
Figure 4. Computational exploration of enzyme-free copying and replication with dimers. A) Copying times for all possible template sequences of length ten consisting only of C and G nucleotides. The timescales were calculated using effective extension rates extracted from kinetic data for the ligation of individual dimers (see Supporting Information). Sequences with the same G content and same predicted copying time are represented as filled circles, with areas proportional to the number of sequences contained (these numbers are also explicitly indicated). The circles containing the experimental template **1** and its reverse complement are marked in red. B) Replication times, i.e., the sums of the copying times of a sequence and its reverse complement, for all sequences from (A). The circle containing the experimental template **1** is marked in red.

[1] A. Kornberg, T. A. Baker, *DNA Replication*, 2nd ed., University Science Books, Mill Valley, **2005**.

[2] E. V. Koonin, V. V. Dolja, M. Krupovic, A. Varsani, Y. I. Wolf, N. Yutin, F. Murilo Zerbini, J. H. Kuhn, *Microbiol. Mol. Biol. Rev.* **2020**, *84*, e00061.

[3] K. Ruiz-Mirazo, C. Briones, A. de la Escosura, *Chem. Rev.* **2014**, *114*, 285–366.

[4] I. A. Kozlov, L. E. Orgel, *Mol. Biol.* **2000**, *34*, 781–789.

[5] S. A. Benner, H.-J. Kim, Z. Yang, *Cold Spring Harbor Perspect. Biol.* **2012**, *4*, a003541.

[6] R. M. Turk, N. V. Chumachenko, M. Yarus, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 4585-4589.

[7] J. Sulston, R. Lohrmann, L. E. Orgel, H. T. Miles, *Proc. Natl. Acad. Sci. USA* **1968**, *59*, 726–733.

[8] G. F. Joyce, G. M. Visser, C. A. A. van Boeckel, J. H. van Boom, L. E. Orgel, J. van Westrenen, *Nature* **1984**, *310*, 602–604.

[9] A. R. Hill, L. E. Orgel, T. Wu, *Origins Life Evol. Biospheres* **1993**, *23*, 285–290.

[10] M. Zielinski, I. A. Kozlov, L. E. Orgel, *Helv. Chim. Acta* **2000**, *83*, 1678–1684.

[11] D. Duzdevich, C. E. Carr, D. Ding, S. J. Zhang, T. S. Walton, J. W. Szostak, *Nucleic Acids Res.* **2021**, *49*, 3681–3691.

[12] C. Deck, M. Jauker, C. Richert, *Nat. Chem.* **2011**, *3*, 603–608.

[13] L. Li, N. Prywes, C. P. Tam, D. K. O'Flaherty, V. S. Lelyveld, E. C. Izgu, A. Pal, J. W. Szostak, *J. Am. Chem. Soc.* **2017**, *139*, 1810–1813.

[14] N. Prywes, J. C. Blain, F. Del Frate, J. W. Szostak, *eLife* **2016**, *5*, e17756.

[15] N. G. Dolinnaya, N. I. Sokolova, D. T. Ashirbekova, Z. A. Shabarova, *Nucleic Acids Res.* **1991**, *19*, 3067–3072.

[16] M. Hey, C. Hartel, M. W. Göbel, *Helv. Chim. Acta* **2003**, *86*, 844–854.

[17] E. Kervio, B. Claasen, U. E. Steiner, C. Richert, *Nucleic Acids Res.* **2014**, *42*, 7409–7420.

[18] M. Jauker, H. Griesser, C. Richert, *Angew. Chem. Int. Ed.* **2015**, *54*, 14559–14563; *Angew. Chem.* **2015**, *127*, 14767–14771.

[19] S. R. Vogel, C. Deck, C. Richert, *Chem. Commun.* **2005**, 4922–4924.

[20] K. Leu, E. Kervio, B. Obermayer, R. M. Turk-MacLeod, C. Yuan, J.-M. Luevano Jr, E. Chen, U. Gerland, C. Richert, I. A. Chen, *J. Am. Chem. Soc.* **2013**, *135*, 354–366.

[21] W. S. Zielinski, L. E. Orgel, *Nature* **1987**, *327*, 346–347.

[22] D. Sievers, G. von Kiedrowski, *Nature* **1994**, *369*, 221–224.

[23] E. Hänle, C. Richert, *Angew. Chem. Int. Ed.* **2018**, *57*, 8911–8915; *Angew. Chem.* **2018**, *130*, 9049–9053.

[24] M. Sosson, D. Pfeffer, C. Richert, *Nucleic Acids Res.* **2019**, *47*, 3836–3845.

[25] D. Sarracino, C. Richert, *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2543–2548.

[26] M. Smietana, E. T. Kool, *Angew. Chem. Int. Ed.* **2002**, *41*, 3704–3707; *Angew. Chem.* **2002**, *114*, 3856–3859.

[27] B. L. Gaffney, E. Veliath, J. Zhao, R. A. Jones, *Org. Lett.* **2010**, *12*, 3269–3271.

[28] E. Edeleva, A. Salditt, J. Stamp, P. Schwintek, J. Boekhoven, D. Braun, *Chem. Sci.* **2019**, *10*, 5807–5814.

[29] C. He, I. Gállego, B. Laughlin, M. A. Grover, N. V. Hud, *Nat. Chem.* **2017**, *9*, 318–324.

[30] S. Motsch, P. Tremmel, C. Richert, *Nucleic Acids Res.* **2020**, *48*, 1097–1107.

[31] S. Motsch, D. Pfeffer, C. Richert, *ChemBioChem* **2020**, *21*, 2013–2018.