Oligonucleotides

Deutsche Ausgabe: DOI: 10.1002/ange.201803074
Internationale Ausgabe: DOI: 10.1002/anie.201803074

Enzyme-Free Replication with Two or Four Bases

*Elena Hänle and Clemens Richert**

Abstract: *All known forms of life encode their genetic information in a sequence of bases of a genetic polymer and produce copies through replication. How this process started before polymerase enzymes had evolved is unclear. Enzyme-free copying of short stretches of DNA or RNA has been demonstrated using activated nucleotides, but not replication. We have developed a method for enzyme-free replication. It involves extension with reversible termination, enzyme-free ligation, and strand capture. We monitored nucleotide incorporation for a full helical turn of DNA, during both a first and a second round of copying, by using mass spectrometry. With all four bases (A/C/G/T), an "error catastrophe" occurred, with the correct sequence being "overwhelmed" by incorrect ones. When only C and G were used, approximately half of the daughter strands had the mass of the correct sequence after 20 copying steps. We conclude that enzyme-free replication is more likely to be successful with just the two strongly pairing bases than with all four bases of the genetic alphabet.*

Replication of genetic information is critical for life. Cells must replicate their DNA prior to cell division to pass on their genome to daughter cells. Without replication, there is also no Darwinian evolution. All known cells use enzymes for replication. How replication started before enzymes were available is unclear.[1,2] A simpler form of producing copies must have existed that relied on molecular recognition and chemical reactivity only.[3] Enzyme-free copying reactions based on nucleotides are known that produce or extend a strand complementary to a template,[4–6] but replication requires two copying phases and strand separation to produce a replica of the original sequence. Replication from nucleotides in the absence of enzymes or ribozymes has never before been demonstrated experimentally. In 2012, Szostak wrote: "even given prebiotically generated RNA templates and abundant ribonucleotides, we do not understand how cycles of template-directed RNA replication could occur."[7] Among the problems noted are low fidelity, slow and incomplete reactions, and difficulties in separating the strands. Without an experimental model, it is difficult to assess the feasibility of enzyme-free replication because processes such as stalling after misincorporation, error propagation, and secondary-structure formation may have severe effects on overall fidelity. These effects cannot be readily extrapolated from assays involving single extension steps, and only full replication shows that both the initial sequence and its complementary strand (the copy) successfully act as templates.

Most experimental systems for copying without polymerases or ribozymes[8] use activated nucleotide monomers with an organic leaving group, such as 2-methylimidazole[9] or oxyazabenzotriazole (OAt).[10] Early studies focused on template-directed oligomerization,[4,11,12] but recent work has mostly focused on the extension of a primer strand or hairpin, terminating in a ribonucleoside[13–16] or an aminonucleoside.[17–19] Incomplete conversion severely limits ribonucleoside-based copying,[20] since inhibition by hydrolyzed monomers slows down reactions that are already inefficient in aqueous solution.[21] Inhibition can be overcome through immobilization of the template[21] or in situ (re)activation of monomers,[22] but solely RNA-based assays are too slow to be practicable for studies on multiple extensions, and only short stretches of RNA have been copied thus far.[23]

Unless fidelity is sufficiently high, the information stored in genes will be lost upon replication.[24] By using reactive 3′-amino-2′,3′-dideoxynucleoside termini, the fidelity of enzyme-free copying of DNA has been studied. With deoxynucleotide monomers, extension stops after a single incorporation, facilitating product analysis by mass spectrometry.[6,10] Error rates of up to 28% per extension step were found when using this method.[6] This is too poor to replicate even the very shortest of genes, such as the one coding for a pentamer RNA enzyme.[25] Kinetic phenomena, such as stalling after misincorporation,[26,27] can improve fidelity, but they also lower the yield of copying and thus replication.

Because of the lack of an experimental system for replication, the question of "which and how many nucleobases a self-replicating system must have" has remained open. This question has been the subject of considerable debate. The number of four bases in today's genetic code has been described as accidental, rather than as an evolutionary optimum,[28] and nucleic acid precursors containing only two bases have been proposed.[29–31] Furthermore, a ribozyme containing only two different nucleotides that shows ligase activity has been described,[32] thus allaying concerns that such a simple nucleic acid would be without function.

We have developed a method for template-directed primer extension on immobilized templates.[33] This method uses "reversible termination" and mass spectrometric analysis as read-out for each step. Our previous work involved primer extension with the correct nucleotide only,[33] not genetic copying with incorporation of nucleotides from a mixture of competing monomers. Strand separation at the end of copying, capture of the copy, and a second copying phase, were also missing. Here, we present an experimental system for the enzyme-free replication of genetic sequences. Replication of a sequence of the length of one helical turn (ten bases) was achieved by using the four nucleobases of natural DNA. Our

[*] M. Sc. E. Hänle, Prof. C. Richert
Institut für Organische Chemie, Universität Stuttgart
70569 Stuttgart (Germany)
E-mail: lehrstuhl-2@oc.uni-stuttgart.de

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.201803074.

results suggest that replication with all four bases (A/C/G/T) is too error-prone to be useful in replicating even short genes in the absence of enzymes or ribozymes, whereas replication with the two strongly pairing bases (C/G) appears more feasible.

Figure 1 shows the monomers and principal steps of our assay. As in other experimental systems,[17–19,27,35,36] an amino group replaces the 3′-hydroxy group at the primer terminus, and an OAt group replaces the pyrophosphate of natural dNTPs to avoid overly long reaction times. Furthermore, a small protecting group (Azoc) is used to avoid polymerization and thus allow for read-out after each extension step.[33] The core structure responsible for pairing with the template is unchanged compared to natural monomers, the nucleophilic group at the terminus of the primer is isoelectronic with that

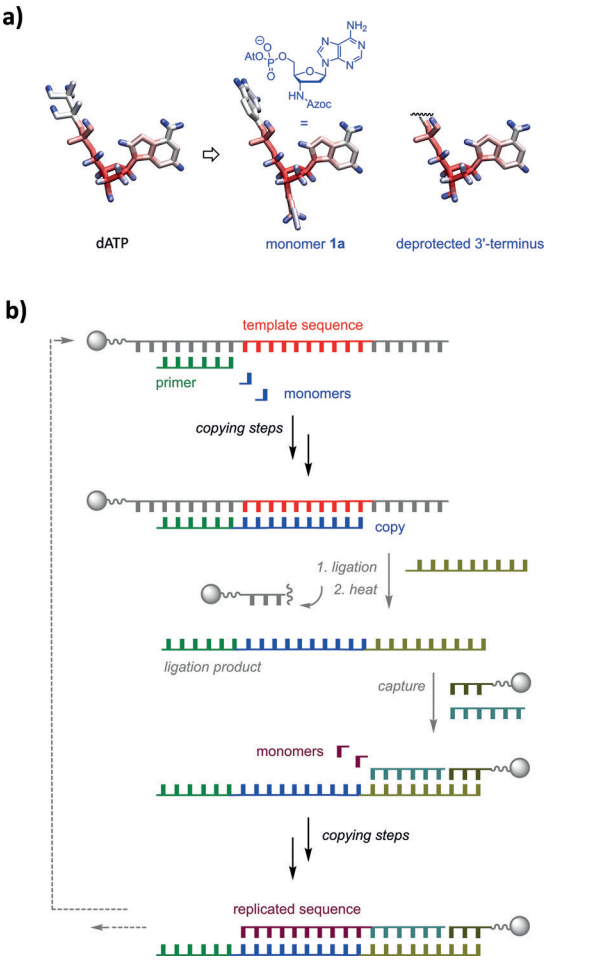

Figure 1. System for enzyme-free replication. a) Structure of a monomer of enzymatic replication (dATP) and of the enzyme-free replication performed here (**1a**), together with the structure of the 3′-terminus resulting from incorporation of **1a**. Structures are those of free acids, with coordinates generated in Chem3D Pro, 14.0 visualized in VMD. b) Assay format: the template sequence (red) is a stretch of a DNA strand, immobilized on a support (gray), to which a primer (green) is bound. Step-wise extension of the primer through reversible termination with monomers (blue) produces the first copy (blue), which is ligated to a capture strand and then used as the template for the second copying phase to produce the original sequence (purple), completing replication.

of deoxynucleotides, and all reactions occur in aqueous buffer, including the deblocking after read-out, which is induced under native conditions and does not affect the fidelity of primer extension. After a copying run, a non-enzymatic ligation is followed by heat denaturation and capture of the product on a new support.

Figure 2 shows the compounds employed. The 3′-amino-nucleotide monomers **1a**–**t** were utilized under conditions for extension and deblocking previously described for assays with individual, preselected monomers.[33] We chose template sequences of ten bases, that is, twice the length of the short ribozyme mentioned above,[25] which is one helical turn, and thus only slightly shorter than the longest stretches of sequence copied with highly reactive 2′-aminonucleotides.[19] We included templates with just two bases (**2**, **5**, and **7**) and the four bases (**4** and **9**) in our study. All had flanking regions for binding the primer (**3** or **6**) and downstream-binding strands, as well as a T_{11} linker to the beads. After each extension, samples were analyzed by MALDI-TOF mass spectrometry, using a matrix/co-matrix mixture known to allow quantitative detection.[34] Mass spectrometry gives more detailed information on base composition than HPLC[4] or gel electrophoresis.[23,35] Controlled, stepwise extension and the monitoring technique allowed us to gather information on fidelity for each nucleotide incorporation.

Our experimental system enabled replication because, after the first round of copying, the ten-fold extended primer was ligated to a 5′-phosphorylated strand dubbed "ligator" (**16** or **17**, Figure 2). This step was performed using in situ activation to avoid the handling of labile active esters. Enzyme-free ligations have been employed in studies on the origins of replication in the past,[36,37] and so have capture steps after denaturation.[38] The captured ligation products containing the copied sequence were then treated with primers **18** or **20**, which were subsequently extended in the second round of enzyme-free copying to obtain a replica of the original template sequence.

Figure 3 shows mass spectra after two extension steps. Significant peaks for misincorporation products are discernible, particularly for sequence **4**, the copying of which was discontinued after the fourth step due to broadening of the peak cluster for the extension products. Data for individual misincorporations during early copying steps can be found in Table S2 of the Supporting Information. Template **2** was copied further, and the fidelity after each step was analyzed (three spectra per sample). Figure 4a shows a plot of the fidelity, as expressed by the relative intensity of the peak of the correct mass over the sum of the peak intensities for all side products of the same length. The standard deviation of this ratio, determined from three separate spectra, was less than or equal to 5% in all cases. Numerical data can be found in Tables S3–S8. For template **2**, the fidelity varied strongly with the position in the template. The accumulation of incorrectly extended products reduced the number of fully correct copies to 35% or less after ten extension cycles.

We suspected that the variability in fidelity was due to folding of template **2** into secondary structures that reduce the accessibility of templating bases to incoming monomers (uppermost entry in Figure 4b). A UV melting curve of **2**

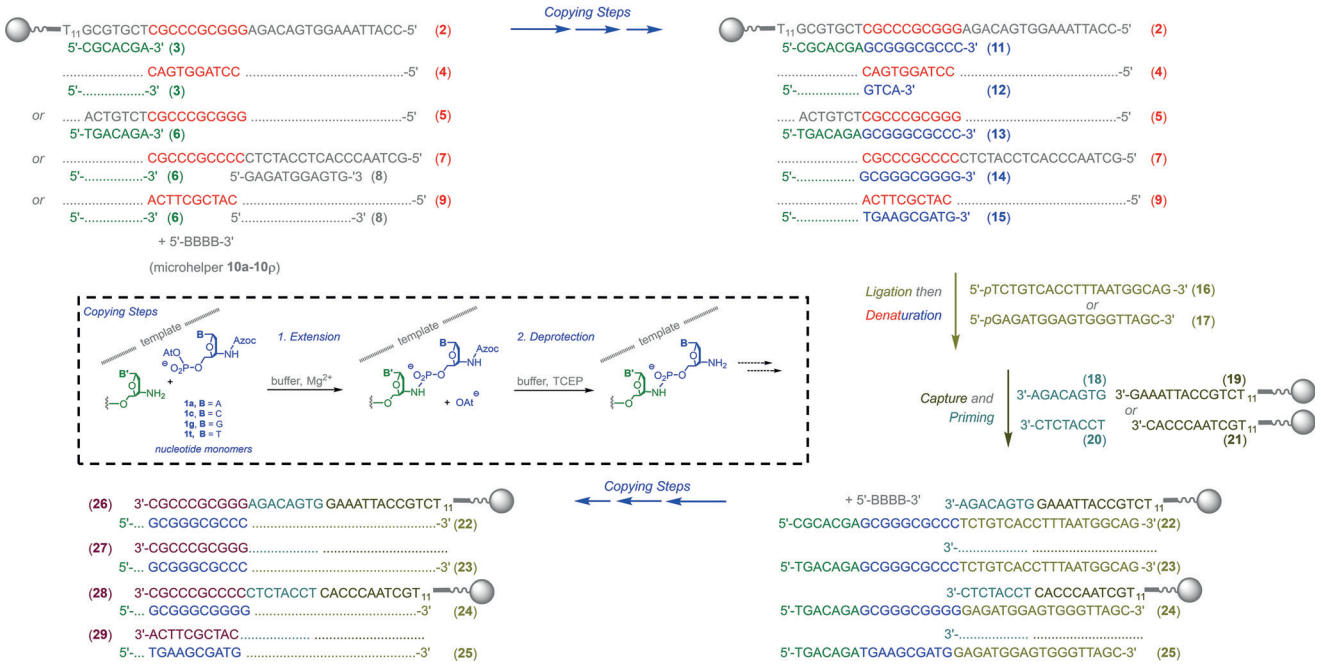

Figure 2. Sequences and reactions of enzyme-free replication. The inset shows the two reactions of copying with reversible termination. Red: "gene" sequences, green: aminoterminal primers, blue: the copy of the gene sequence with a phosphoramidate backbone, purple: the replicated sequences. Regions not involved in copying are shown as dots; *p* denotes a 5′-phosphate. Primer extensions used monomer mixtures **1a–t** (40 mM total nucleotide concentration) in 0.2 M HEPBS, pH 8.9, 400 mM NaCl, and 80 mM $MgCl_2$, with reaction times of 4 h for templates **2**, **5** and **7** and 12 h for templates **4** and **9** at 22 °C. Deblocking was carried out using 100 mM aqueous TCEP in 0.2 M HEPBS, pH 8.9, 400 mM NaCl, and 80 mM $MgCl_2$ for 30 min at 0 °C. Azoc = azidomethoxycarbonyl, B/B′ = nucleobases, OAt = 1-oxy-7-azabenzotriazole, gray spheres are beads.

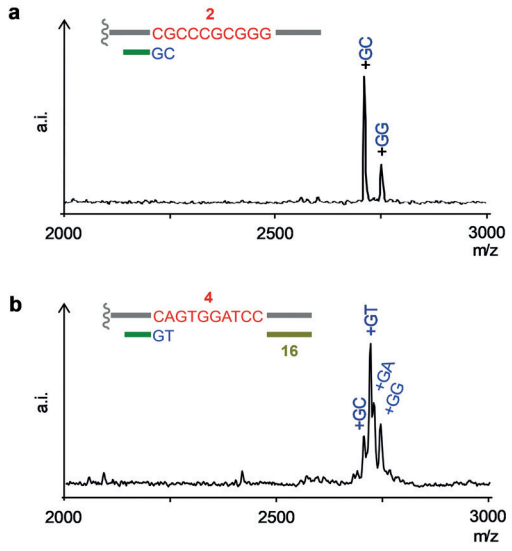

Figure 3. High error levels of copying after two extension cycles, as determined by MALDI-TOF MS. a) Template **2**, copied with just two monomers (C and G). b) Template **4**, copied using all four bases (A, C, G, and T). The most prominent peak has the mass of the correct copy, smaller peaks are from "mutants" resulting from misincorporation.

indicated that secondary structures do indeed form (Figure S16). To suppress this, ligator strand **16** was hybridized to **2**. A modest increase in fidelity and a more even error frequency at the individual positions resulted. Next, we introduced short downstream-binding strands (tetramers or trimers) as microhelpers[6,16] that are washed off at the end of an extension and that further suppress folding of the template. Again, a significant improvement in fidelity resulted (third entry in Figure 4b). Because sequence analysis indicated that primer **3** had a tendency to dimerize, possibly leading to off-template extensions, the system was changed to a less self-complementary version (primer **6**, template **5**). Now, the product of the correct mass gave more than or equal to 50% of the peak intensity after ten extension cycles.

We noted that ligator **16** with its 5′-phosphate slowly reacted during copying, most probably through formation of pyrophosphate-linked products. To prevent this, it was replaced with an unphosphorylated "blocking strand" (**8**), which is later displaced. Furthermore, a "gene" with a template sequence was picked that had a lower propensity to fold, as predicted computationally.[39] On this template (**7**), the accumulated error level at the end of the assay was lower. The high fidelity at position 5 suggested that stalling after misincorporation[26,27] contributed to faithful copying (penultimate entry in Figure 4b). Finally, we adjusted the ratio of the monomers[6,40] to give the more readily incorporated G less of an advantage, resulting in an improved overall fidelity ($\leq$74% after ten steps). A control experiment showed that in this system, blocking strand **8** has no measurable influence on overall fidelity (73% ±1 after ten copying steps if **8** was omitted), but the yield of the ligation product, as determined by gel electrophoresis, was slightly higher (Figure S15).

With the optimized replication system, we then revisited the question of whether a two- or four-base system may have been successful in an early form of replication. For this, we changed the all-G/C "gene" of template **7** to a sequence with

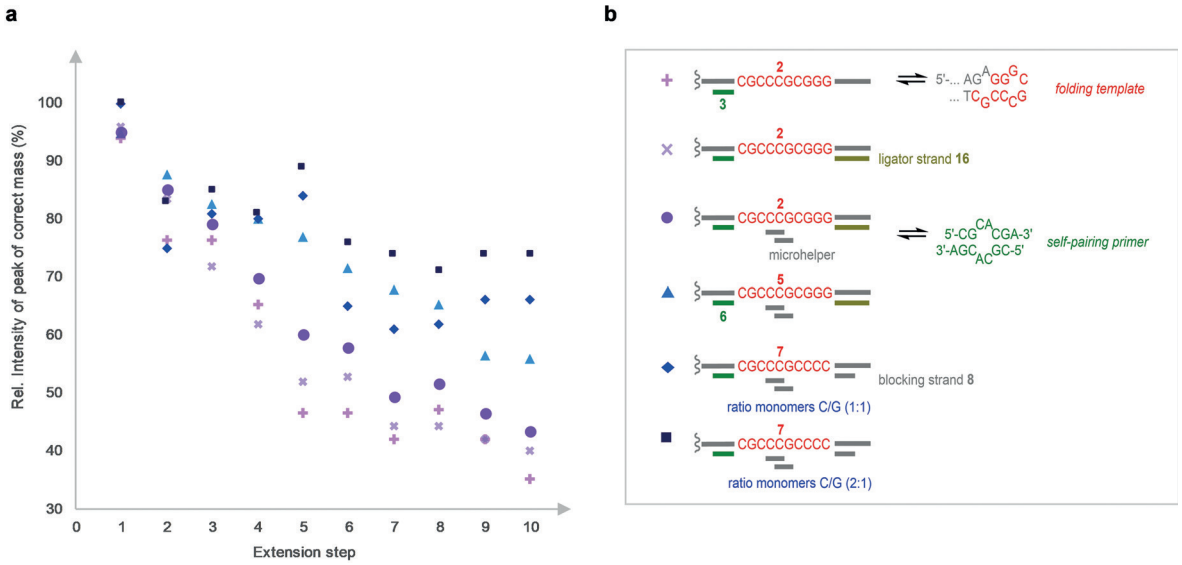

Figure 4. Effect of sequence and strand composition on fidelity. a) The relative intensity of the peak for the correct extension product is plotted for each extension step for the templates shown in part (b), in the absence or presence of a ligator or blocking strand and short microhelper tetramers, added together with the monomers. b) Template sequences and ancillary strands; issues likely to lower fidelity are highlighted..

all four bases (template **9**), again carefully avoiding sequences predicted to fold into stable secondary structures. Figure 5 shows the results of copying assays performed on **7** with C and G as monomers (ratio 2:1) or on template **9** with all four bases (reactivity-adjusted ratio[6,40] T/A/C/G of 5:2:2:1).

Figure 5a shows that for the two-base system, the correct product dominates at the end of the assay, whereas even for the optimized four-base system, the accumulation of mutants is massive after copying one helical turn. Numerical data for the fidelity in representative early copying steps are given in Table S1 of the Supporting Information. For **9**, the most prominent peak after the first round of copying still had the mass of the correctly copied sequence. This changed after ligation and the subsequent second copying round (Figure 6). For the system with all four bases, a broad cluster of peaks was found, and the peak with the correct mass was indistinguishable from those of mutants. In other words, the genetic information encoded in the ten-nucleotide "gene" had disappeared into the "noise". In contrast, even after two rounds of copying, the peak of the correct mass was still the most prominent signal for the all-C/G system. This peak made up 47 % ± 1 of the total signal intensity in the mass region of the fully extended product (Figure 6a).

In conclusion, we show the first enzyme-free replication with mononucleotides in a genetic model system isoelectronic to natural DNA. This system relies on enzyme-free copying, enzyme-free ligation, thermal strand separation, and strand capture by hybridization. By monitoring both phases of replication quantitatively, nucleotide by nucleotide, we were able to provide a glimpse of molecular processes that are important for understanding the ability of nucleic acids to replicate in the absence of enzymes. Some sequences show effects that could not have been extrapolated from single-nucleotide extension reactions or copying of short stretches of a single template. The results of Figure 4 in particular show that fidelity depends on sequence in more than one way. Some sequence combinations are more likely to succeed in the molecular selection of the fittest than others, thus favoring rapid molecular evolution. The successful systems identified thus far use an interplay of upstream- and downstream-

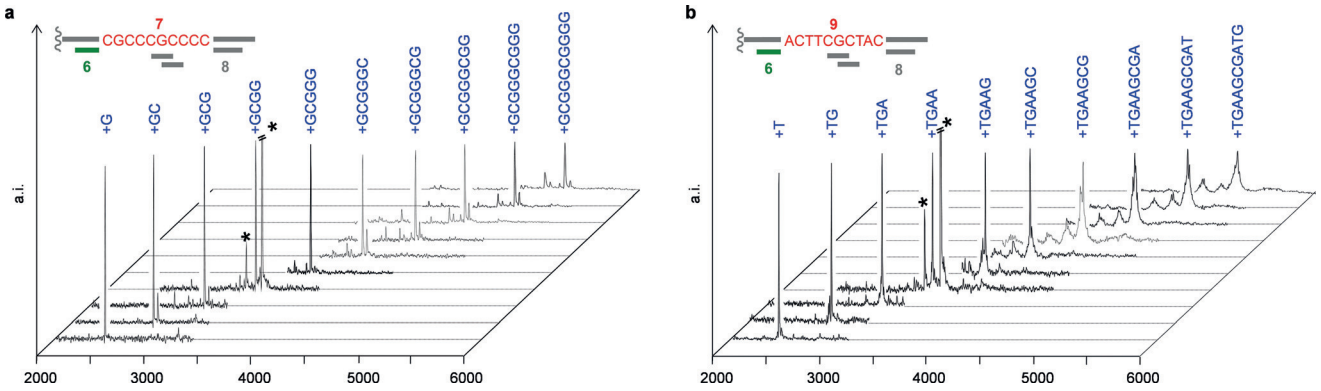

Figure 5. Tenfold primer extension in sequences with minimized possible secondary structures. a) A C/G template leads mostly to the correct copy of the template. Peaks with * are from blocking strand **8**. b) A template sequence with all four bases leads to an "error catastrophe".

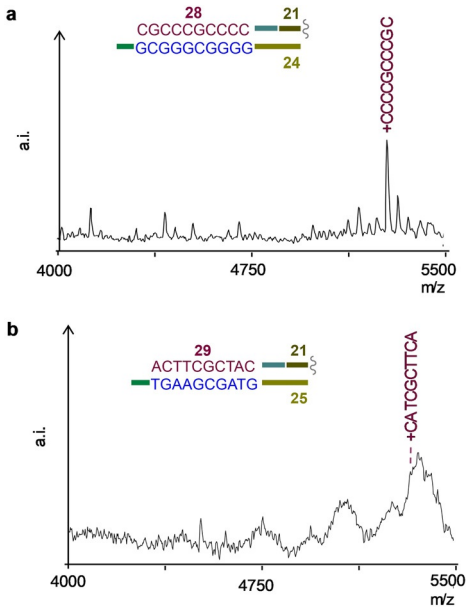

Figure 6. MALDI-TOF mass spectra after replication consisting of 20 primer extension steps. a) Template **7** and ligation product **24**, copied with only C and G. b) Template **9** and product **25**, copied with A, C, G and T. The mass of the correct product is labeled with its sequence (purple).

binding strands and are strongly dependent on "regulation" by secondary structure formation.

We are currently extending this work to RNA and more extensive replication. Because RNA and the phosphoramidate-linked DNA employed here show similar levels of fidelity in enzyme-free copying,[6] it is fair to state even now that two-base systems with just C and G were more likely to have avoided an error catastrophe during enzyme-free replication than present-day four-base systems.

Acknowledgements

We thank D. Pfeffer, A. Kaiser, M. Jauker, and E. Kervio for discussions, and D. Göhringer for technical assistance. This study was supported by DFG grants RI 1063/8-2 and RI 1063/16-1; E.H. was a recipient of a Ph.D. fellowship of the State of Baden-Württemberg.

Conflict of interest

The authors declare no conflict of interest.

Keywords: base pairing · genetic polymers · nucleotides · oligonucleotides · replication

How to cite: *Angew. Chem. Int. Ed.* **2018**, *57*, 8911–8915
Angew. Chem. **2018**, *130*, 9049–9053

[1] K. Ruiz-Mirazo, C. Briones, A. de la Escosura, *Chem. Rev.* **2014**, *114*, 285.
[2] A. Kornberg, T. A. Baker, *DNA Replication*, 2nd ed., University Science Books, Mill Valley, **2005**.
[3] L. E. Orgel, *Nature* **1992**, *358*, 203.
[4] R. Lohrmann, L. E. Orgel, *Nature* **1976**, *261*, 342.
[5] M. Hey, C. Hartel, M. W. Göbel, *Helv. Chim. Acta* **2003**, *86*, 844.
[6] E. Kervio, A. Hochgesand, U. E. Steiner, C. Richert, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 12074.
[7] J. W. Szostak, *J. Syst. Chem.* **2012**, *3*, 1.
[8] J. Attwater, A. Wochner, P. Holliger, *Nat. Chem.* **2013**, *5*, 1011.
[9] R. Lohrmann, L. E. Orgel, *Tetrahedron* **1978**, *34*, 853.
[10] P. Hagenbuch, E. Kervio, A. Hochgesand, U. Plutowski, C. Richert, *Angew. Chem. Int. Ed.* **2005**, *44*, 6588; *Angew. Chem.* **2005**, *117*, 6746.
[11] J. Sulston, R. Lohrmann, L. E. Orgel, H. T. Miles, *Proc. Natl. Acad. Sci. USA* **1968**, *59*, 726.
[12] A. Kanavarioti, C. F. Bernasconi, E. E. Baird, *J. Am. Chem. Soc.* **1998**, *120*, 8575.
[13] T. Wu, L. E. Orgel, *J. Am. Chem. Soc.* **1992**, *114*, 7963.
[14] I. A. Kozlov, L. E. Orgel, *Mol. Biol.* **2000**, *34*, 781.
[15] M. Kurz, K. Göbel, C. Hartel, M. W. Göbel, *Angew. Chem. Int. Ed. Engl.* **1997**, *36*, 842; *Angew. Chem.* **1997**, *109*, 873.
[16] S. R. Vogel, C. Deck, C. Richert, *Chem. Commun.* **2005**, 4922.
[17] J. A. Rojas Stütz, C. Richert, *J. Am. Chem. Soc.* **2001**, *123*, 12718.
[18] S. S. Mansy, J. P. Schrum, M. Krishnamurthy, S. Tobé, D. A. Treco, J. W. Szostak, *Nature* **2008**, *454*, 122.
[19] J. P. Schrum, A. Ricardo, M. Krishnamurthy, J. C. Blain, J. W. Szostak, *J. Am. Chem. Soc.* **2009**, *131*, 14560.
[20] A. R. Hill, L. E. Orgel, T. Wu, *Origins Life Evol. Biospheres* **1993**, *23*, 285.
[21] C. Deck, M. Jauker, C. Richert, *Nat. Chem.* **2011**, *3*, 603.
[22] M. Jauker, H. Griesser, C. Richert, *Angew. Chem. Int. Ed.* **2015**, *54*, 14559; *Angew. Chem.* **2015**, *127*, 14767.
[23] N. Prywes, J. C. Blain, F. Del Frate, J. W. Szostak, *eLife* **2016**, *5*, 1.
[24] M. Eigen, *Naturwissenschaften* **1971**, *58*, 465.
[25] R. M. Turk, M. Illangasekare, M. Yarus, *J. Am. Chem. Soc.* **2011**, *133*, 6044.
[26] S. Rajamani, J. K. Ichida, T. Antal, D. A. Treco, K. Leu, M. A. Nowak, J. W. Szostak, I. A. Chen, *J. Am. Chem. Soc.* **2010**, *132*, 5880.
[27] K. Leu, E. Kervio, B. Obermayer, R. M. Turk-MacLeod, C. Yuan, J.-M. Luevano, E. Chen, U. Gerland, C. Richert, I. A. Chen, *J. Am. Chem. Soc.* **2013**, *135*, 354.
[28] E. Szathmáry, *Proc. R. Soc. London Ser. B* **1991**, *245*, 91.
[29] L. E. Orgel, *J. Mol. Biol.* **1968**, *38*, 381.
[30] F. H. C. Crick, *J. Mol. Biol.* **1968**, *38*, 367.
[31] G. Wächtershäuser, *Proc. Natl. Acad. Sci. USA* **1988**, *85*, 1134.
[32] J. S. Reader, G. F. Joyce, *Lett. Nat.* **2002**, *420*, 841.
[33] A. Kaiser, S. Spies, T. Lommel, C. Richert, *Angew. Chem. Int. Ed.* **2012**, *51*, 8299; *Angew. Chem.* **2012**, *124*, 8424.
[34] D. Sarracino, C. Richert, *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2543.
[35] S. Zhang, J. C. Blain, D. Zielinska, S. M. Gryaznov, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 17732.
[36] D. Sievers, G. von Kiedrowski, *Nature* **1994**, *369*, 221.
[37] T. Li, K. C. Nicolaou, *Nature* **1994**, *369*, 218.
[38] A. Luther, R. Brandsch, G. von Kiedrowski, *Nature* **1998**, *396*, 245.
[39] Predicted by using Unafold, see: M. Zuker, *Nucleic Acids Res.* **2003**, *31*, 3406.
[40] K. Leu, B. Obermayer, S. Rajamani, U. Gerland, I. Chen, *Nucleic Acids Res.* **2011**, *39*, 8135.

Manuscript received: March 13, 2018
Revised manuscript received: April 27, 2018
Accepted manuscript online: May 20, 2018
Version of record online: June 19, 2018

