High Fidelity Enzyme-Free Primer Extension with an Ethynylpyridone Thymidine Analog

Jianyang Han,[a] Eric Kervio,[a] and Clemens Richert*[a]

Abstract: High fidelity base pairing is important for the transmission of genetic information. Weak base pairs can lower fidelity, complicating sequencing, amplification and replication of DNA. Thymidine 5′-monophosphate (TMP) is the most weakly pairing nucleotide among the canonical deoxynucleotides, causing high errors rates in enzyme-free primer extension. Here we report the synthesis of an ethynylpyridone *C*-nucleoside analog of 3′-amino-2′,3′-dideoxythymidine monophosphate and its incorporation in a growing strand by enzyme-free primer extension. The ethynylpyridone *C*-nucleotide accelerates extension more than five-fold, reduces misincorporation and readily displaces TMP in competition experiments. The results bode well for the use of the *C*-nucleoside as replacements for thymidine in practical applications.

There is a growing interest in the structures and reactions underlying the replication of genetic information.[1] The extension of a primer, directed by a template, is the molecular basis of replication and transcription.[2] Primer extension is used to amplify genetic information in the polymerase chain reaction (PCR),[3] to sequence genetic material through Sanger sequencing, and to study processes that may have led to first self-replicating systems.[4,5] Primer extension assays are also used when evaluating new polymerases[6] or new inhibitors as potential therapeutics.[7] The results of such assays can lead to new antivirals that target the replication step of viral life cycles.[8–10]

The best-known form of primer extension is the one employing a polymerase,[11] but enzyme-free versions of the reaction also exist. Both RNA[12] and DNA template sequences[13] have been copied in reactions driven solely by chemical reactivity and molecular recognition.[5,14] Imidazolides[15] or oxyazabenzotriazolides of ribonucleotide[16] can act as monomers in such reactions, replacing the triphosphates of enzymatically catalyzed forms.[17] With the chemically activated forms of ribonucleotides, insights can be gained into plausible prebiotic pathways of genetic copying,[18,19] but the same insights are also useful for practical applications.[20]

Most prebiotic chemistry studies focus on RNA, but DNA is the predominant carrier of genetic information in extant biology. For DNA, the reactivity of the primers with 2′-deoxynucleoside termini is too low to study enzyme-free primer extension,[21] and oligodeoxynucleotides with 3′-amino-terminal residues are commonly used as isoelectronic replacements.[13,22,23] With enzyme-free copying systems, in-depth data on the strength of the template effect,[24] mechanistic details,[25] and the sequence dependence of the yield and fidelity of primer extension have been obtained. Using these insights, the replication of a sequence of ten nucleotides with C and G as nucleobases was recently reported,[26] but nucleotide-based replication attempts with all four canonical bases have failed thus far.

A key difficulty encountered when trying to replicate DNA sequences in the absence of polymerases is low fidelity. Misincorporation rates in enzyme-free primer extension assays can exceed 25 %.[24,27] Thymine and uracil show particularly poor fidelity in enzyme-free copying.[28] Weak Watson-Crick pairing, with just two hydrogen bonds, resulting in K_d values in the range of 38 mM to $>$500 mM,[25,28] as well as wobble base pairing with guanine as an alternative pairing mode, contribute to the poor performance of these pyrimidines. In our study on enzyme-free replication of DNA sequences, we came to the conclusion that with T:A base pairs, transmission of genetic information was unlikely to be successful.[26]

Low fidelity copying with uracil may be overcome using replacements with more favorable pairing properties. For example, the activated form of 2-thiouridine has been found to have favorable properties for copying RNA sequences,[29,30] and so do 3′-amino-2′,3′-dideoxynucleotides with 2-thiothymidine as base.[31,32] Recent studies on hybridization probes showed that ethynylpyridone *C*-nucleosides can pair even more strongly and selectively than 2-thiothymidine[33] does with adenine in DNA target strands,[34,35] and the pairing is confirmed in theoretical studies.[36,37] Further, an ethynylpyridone analog of azidothymidine (AZT) has recently shown detectable antiviral activity.[38] This prompted us to ask whether nucleotide monomers with the ethynylpyridone base would be successfully incorporated in primer extension assays.

Enzyme-free primer extension allows one to separate the nucleic acid level from the idiosyncrasies of specific enzymes

[a] J. Han, Dr. E. Kervio, Prof. C. Richert
Institut für Organische Chemie,
Universität Stuttgart
70569 Stuttgart (Germany)
E-mail: lehrstuhl-2@oc.uni-stuttgart.de

Supporting information for this article is available on the WWW under https://doi.org/10.1002/chem.202102996

and can thus provide important insights for the design of antivirals. Here we report the synthesis of a building block for enzyme-free copying, and results from primer extension assays with and without activated thymidine monophosphate as competitor.

Figure 1 shows the synthesis of the activated deoxynucleotide (**1w**) and the primer extension system employed in our study. The latter relied on DNA template **2**, immobilized on magnetic beads, and a primer with a 3′-terminal aminodideoxynucleoside (**3**). The primer is allowed to react with 3′-Azoc protected[39] nucleotide building blocks, activated as oxyazabenzotriazolide (OAt) esters,[13] to give extension products **4** or **5**. Each extension step is followed by reductive deblocking with TCEP-containing buffer under non-denaturing conditions. The reversible termination with the Azoc group provides control over each extension step, while leaving the nucleoside part of the structure that is responsible for base pairing unencumbered by a bulky or lipophilic substituent.[26] A series of three extension and deprotection steps gives **6**. Each extended primer can be thermally released from the template and analyzed by MALDI-TOF mass spectrometry under conditions that allow for quantitative detection.[40] Besides pyridone *C*-nucleotide **1w**, the OAt esters of the four building blocks with canonical bases (**1a**–**t**) were employed.[23,26]

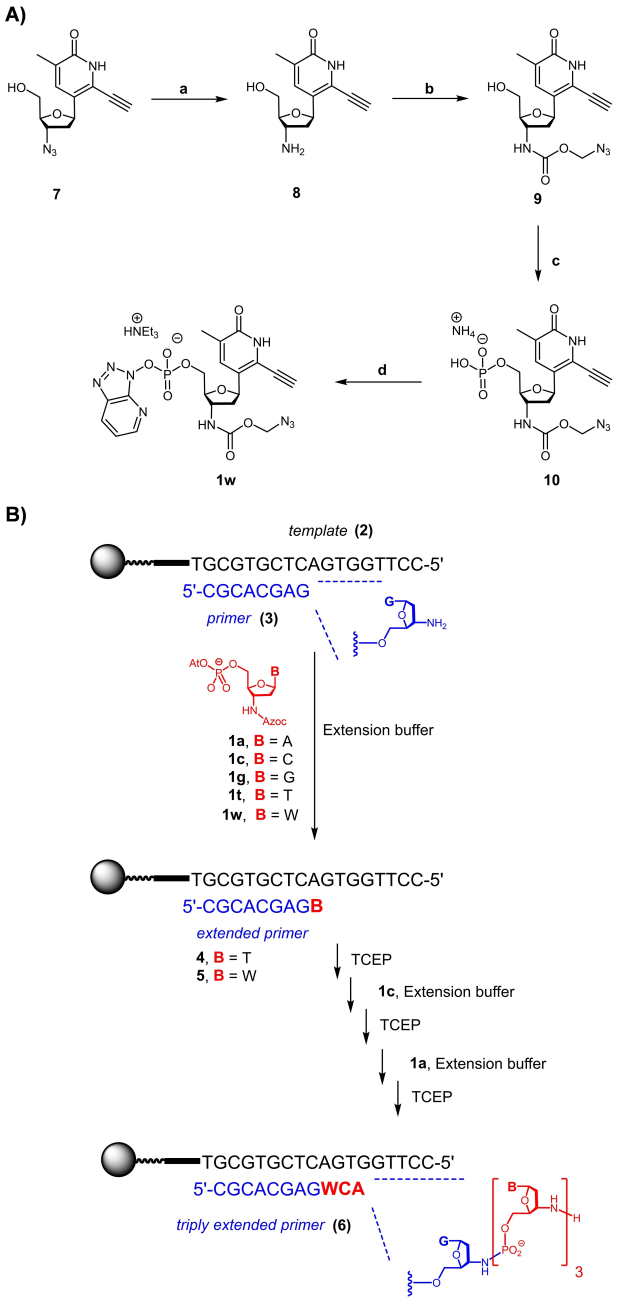

Figure 1. Building block synthesis and primer extension. A) Synthesis of **1w** from **7**;[38] a) PPh_3, THF/H_2O, 45 °C; b) azidomethyl(4-nitrophenyl) carbonate, DIPEA, DMF, 72% over two steps; c) $POCl_3$, pyridine, H_2O, CH_3CN, −10 °C, 56%; d) HOAt, EDC, H_2O, pH 5, 68%; B) enzyme-free primer extension with template **2** on magnetic beads, extension buffer = 0.2 M HEPBS 0.4 M NaCl, 0.08 M $MgCl_2$, pH 8.9; deprotection with 0.3 M tris(2-carboxyethyl)-phosphine (TCEP) in buffer at 0 °C for 60 min.

First, we established that **1w** is successfully incorporated. Figure 2A shows that complete conversion does indeed occur with the new building block, even at 4 °C. When the same assay was performed with thymine (**1t**), extension was much slower (Figure 2B). A monoexponential fit gave second order rate constants of $k' = 1.75\ h^{-1}M^{-1}$ for extension with **1t** and $9.25\ h^{-1}M^{-1}$ for extension with **1w**. The 5.3-fold faster extension lowers the half-conversion time ($t_{1/2}$) from 9.9 h to 1.9 h. The acceleration is similar to that observed when switching from T to G as the nucleotide to be incorporated.[25] We then asked whether incorporation of the ethynylpyridone at the primer terminus interferes with further extension of the chain. For this, three-fold extension was performed to give **6**. Figure 2(C) shows MALDI spectra after the first and second elongation step, and at the end of the assay. Full conversion after 20 h at 20 °C was found for each step.

Next, we studied fidelity by employing mixtures of monomers, again using the conditions of earlier experiments on base pairing selectivity, which were performed at 20 °C.[26] When the mixture **1a**/**1c**/**1g**/**1t** (1:1:1:1) was used on template **2a**, which is known to give poor fidelity at the templating base,[26] the MALDI spectrum of Figure 3A was obtained. The primer extended by T made up only 46% of the products, with prominent peaks for incorporation of A and C. In contrast, when **1w** replaced monomer **1t** in the mixture, the otherwise identical assay gave the correct extension product with drastically improved fidelity. Extension by A, which would manifest itself by a shoulder in the MALDI peak for W, was undetectable with our method, and the misincorporation of C and G was just 7% and 2%, respectively (Figure 3B). When the

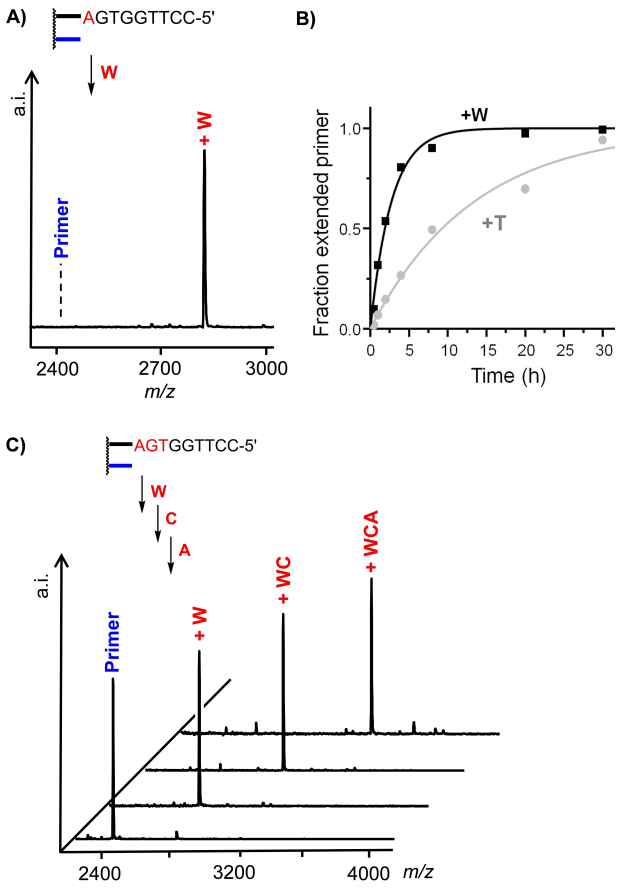

Figure 2. Primer extension with nucleotide monomers. A) MALDI mass spectrum for extension of **3** with **1 w**, recorded after 30 h at 4 °C. B) Kinetics of primer extension, with either **1 w** (black) or **1 t** (gray). Lines are monoexponential fits. Conditions: 40 mM monomer, 0.2 M HEPBS, 0.4 M NaCl, 0.08 M $MgCl_2$, pH = 8.9, at 4 °C. C) High-yielding extension after incorporation of W, as determined by three-fold elongation of primer **3**. Conditions: 40 mM monomers in extension buffer at 20 °C for 20 h, followed by deprotection with 0.3 M TCEP for 1 h at 0 °C. See Figure 1B for full sequences.

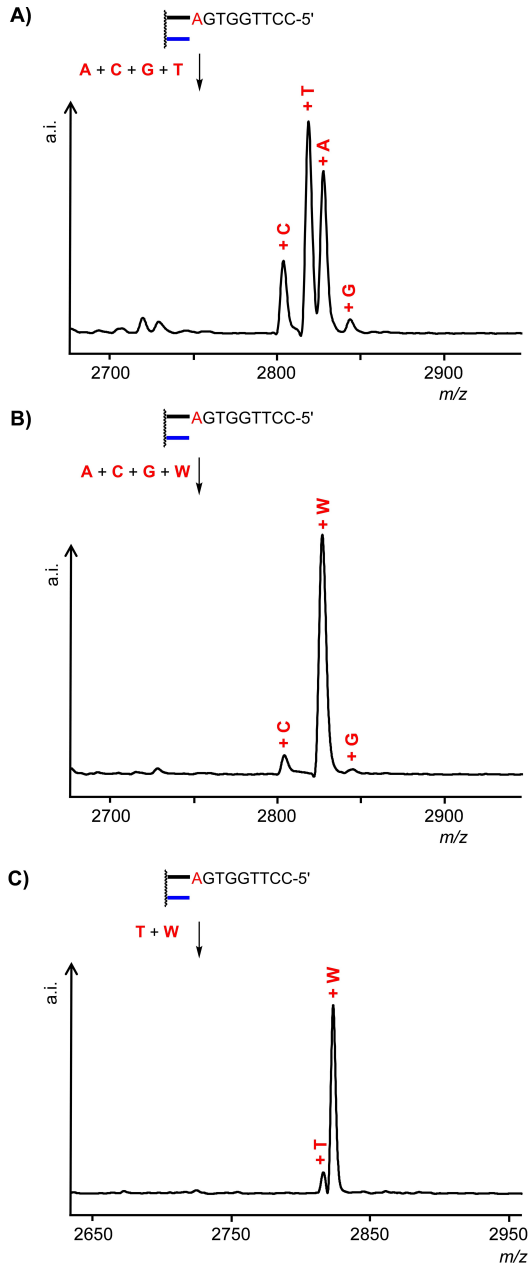

Figure 3. Fidelity assays and competition of W and T. A) MALDI spectrum for extension of **3** on template **2** with an equimolar mixture of **1 a**, **1 c**, **1 g** and **1 t** (10 mM each), in 0.2 M HEPBS, 0.4 M NaCl, 0.08 M $MgCl_2$, pH = 8.9 acquired after 20 h at 20 °C; B) same as A) except that **1 w** was used instead of **1 t**. C) MALDI spectrum from competition experiment with **1 t** and **1 w** (20 mM each) in the same extension buffer, recorded after 20 h at 4 °C.

3′-azido analog of **1 w** was used in a competition assay, the extension product for incorporation of A gave the least intense of the four peaks observable, with 2% relative intensity (Figure S17 of the Supporting Information), confirming high fidelity incorporation of W.

To further demonstrate that there is little unspecific incorporation of **1 w**, and that faithful incorporation of the other three bases (A, C or G) occurs when the template displays T, G or C, additional assays were performed involving more than one elongation step. For this, template **2 b** was employed, which contains all four bases in the first four templating positions (3′-AGTC-5′). The results are shown in Figures S14–S16 of the Supporting Information. In each sequence context, the correct base was incorporated predominantly, despite the presence of **1 w**, and prior incorporation of W did not suppress fidelity in the subsequent step.

Finally, we tested how well ethynylpyridone **1 w** performs in direct competition with **1 t**. The 1:1 mixture of the monomers gave the spectrum of Figure 3C). The primer elongated with W (**5**) was formed 8 times more often than the product **4**. It is therefore reasonable to assume that a modified nucleoside triphosphate containing the ethynylpyridone as base can be an effective competitor for TMP in terms of incorporation in enzymatic primer extension reactions.

In conclusion, we show that with ethynylmethylpyridone as base the processivity and the fidelity problem known for thymine can be all but overcome. The elaboration of *C*-nucleosides is more complicated than that of *N*-nucleosides,

both in nature and in organic synthesis. The assembly of **1w** required 14 steps in our case, including the elaboration of the glycosyl donor and the base,[35,38] and a spontaneous formation of such a structure under prebiotic conditions appears unlikely to us. This may help to explain why this structure is not found in nature. On the other hand, our results show that structure space holds interesting analogs of thymidine in store, and suggests that strongly pairing nucleotides, accessibly by organic synthesis, may be developed into new antivirals.[38]

Acknowledgements

We thank O. Bernhard for technical assistance, D. Pfeffer and G. Leveau for help with assays and J. Eyberg and P. Tremmel for discussions. Supported by DFG, project-ID 364653263 – TRR 235 and Volkswagenstiftung (Az.: 92 768–1). Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: antivirals · *C*-nucleosides · genetic copying · nucleotides · replication

[1] A. J. Oakley, *Protein Sci.* **2019**, *28*, 990–1004.
[2] M. Nikolausz, A. Chatzinotas, A. Táncsics, G. Imfeld, M. Kästner, *Biochem. Soc. Trans.* **2009**, *37*, 454–459.
[3] M. R. Green, J. Sambrook, *Cold Spring Harb. Protoc.* **2019**, doi: 10.1101/pdb.top095109.
[4] K. Ruiz-Mirazo, C. Briones, Andrés de la Escosura, *Chem. Rev.* **2014**, *114*, 285–366.
[5] M. Sosson, C. Richert, *Beilstein J. Org. Chem.* **2018**, *14*, 603–617.
[6] J. Aschenbrenner, S. Werner, V. Marchand, M. Adam, Y. Motorin, M. Helm, A. Marx, *Angew. Chem. Int. Ed.* **2018**, *57*, 417–421; *Angew. Chem.* **2018**, *130*, 424–428.
[7] G. F. Lu, X. Zhang, W. N. Zheng, J. L. Sun, L. Hua, L. Xu, X. J. Chu, S. Ding, W. Xiong, *Antimicrob. Agents Chemother.* **2021**, *65*, e01508–20.
[8] J. A. Moss, *Radiol Technol.* **2013**, *84*, 247–267.
[9] K. L. Seley-Radtke, M. K. Yates, *Antiviral Res.* **2018**, *154*, 66–86.
[10] M. K. Yates, K. L. Seley-Radtke, *Antiviral Res.* **2019**, *162*, 5–21.
[11] B. Alberts, *Nature* **2003**, *421*, 431–435.
[12] C. Deck, M. Jauker, C. Richert, *Nat. Chem.* **2011**, *3*, 603–608.
[13] P. Hagenbuch, E. Kervio, A. Hochgesand, U. Plutowski, C. Richert, *Angew. Chem. Int. Ed.* **2005**, *44*, 6588–6592; *Angew. Chem.* **2005**, *117*, 6746–6750.
[14] I. A. Kozlov, L. E. Orgel, *Mol. Biol.* **2000**, *34*, 781–789.
[15] R. Lohrmann, L. E. Orgel, *Tetrahedron* **1978**, *34*, 853–855.
[16] S. R. Vogel, C. Deck, C. Richert, *Chem. Commun.* **2005**, *39*, 4922–4924.
[17] J. A. Rojas Stütz, E. Kervio, C. Deck, C. Richert, *Chem. Biodiversity* **2007**, *4*, 784–802.
[18] J. W. Szostak, *J. Syst. Chem.* **2012**, *3*, 2.
[19] J. W. Szostak, *Angew. Chem. Int. Ed.* **2017**, *56*, 11037–11043; *Angew. Chem.* **2017**, *129*, 11182–11189.
[20] N. Griesang, K. Giessler, T. Lommel, C. Richert, *Angew. Chem. Int. Ed.* **2006**, *45*, 6144–6148; *Angew. Chem.* **2006**, *118*, 6290–6294.
[21] H. Vogel, C. Gerlach, C. Richert, *Nucleosides Nucleotides Nucleic Acids* **2013**, *32*, 17–27.
[22] R. Lohrmann, L. E. Orgel, *J. Mol. Biol.* **1977**, *109*, 323–328.
[23] A. Kaiser, S. Spies, T. Lommel, C. Richert, *Angew. Chem. Int. Ed.* **2012**, *51*, 8299–8303; *Angew. Chem.* **2012**, *124*, 8424–8428.
[24] E. Kervio, A. Hochgesand, U. Steiner, C. Richert, *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 12074–12079.
[25] E. Kervio, M. Sosson, C. Richert, *Nucleic Acids Res.* **2016**, *44*, 5504–5514.
[26] E. Hänle, C. Richert, *Angew. Chem. Int. Ed.* **2018**, *57*, 8911–8915; *Angew. Chem.* **2018**, *130*, 9049–9053.
[27] 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.
[28] E. Kervio, B. Claasen, U. E. Steiner, C. Richert, *Nucleic Acids Res.* **2014**, *42*, 7409–7420.
[29] B. D. Heuberger, A. Pal, F. D. Frate, V. V. Topkar, J. W. Szostak, *J. Am. Chem. Soc.* **2015**, *137*, 2769–2775.
[30] J. Sheng, A. Larsen, B. D. Heuberger, J. C. Blain, J. W. Szostak, *J. Am. Chem. Soc.* **2014**, *136*, 13916–13924.
[31] S. Zhang, J. C. Blain, D. Zielinska, S. M. Gryaznov, J. W. Szostak, *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 17732–17737.
[32] E. C. Izgu, S. S. Ohab, J. W. Szostak, *Chem. Commun.* **2016**, *52*, 3684–3686.
[33] H. O. Sintim, E. T. Kool, *J. Am. Chem. Soc.* **2006**, *128*, 396–397.
[34] M. Minuth, C. Richert, *Angew. Chem. Int. Ed.* **2013**, *52*, 10874–10877; *Angew. Chem.* **2013**, *125*, 11074–11077.
[35] T. J. Walter, C. Richert, *Nucleic Acids Res.* **2018**, *46*, 8069–8078.
[36] A. Halder, A. Datta, D. Bhattacharyya, A. Mitra, *J. Phys. Chem. B* **2014**, *118*, 6586–6596.
[37] M. Chawla, S. Gorle, A. R. Shaikh, R. Oliva, L. Cavallo, *Comput. Struct. Biotechnol. J.* **2021**, *19*, 1312–1324.
[38] J. Han, C. Funk, J. Eyberg, S. Bailer, and C. Richert, *Chem. Biodiversity* **2021**, *18*, e2000937.
[39] S. Pothukanuri, N. Winssinger, *Org. Lett.* **2007**, *9*, 2223–2225.
[40] D. Sarracino, C. Richert, *Bioorg. Med. Chem. Lett.* **1996**, *6*, 2543–2548.
[41] J. Stambasky, M. Hocek, P. Kocovsky, *Chem. Rev.* **2009**, *109*, 6729–6764.
[42] K. W. Wellington, S. A. Benner, *Nucleosides Nucleotides Nucleic Acids* **2006**, *25*, 1309–1333.

Manuscript received: August 16, 2021
Accepted manuscript online: September 24, 2021
Version of record online: October 8, 2021

