Base Pairing

DOI: 10.1002/anie.201305555

A Nucleobase Analogue that Pairs Strongly with Adenine**

*Marco Minuth and Clemens Richert**

Base pairing is a pivotal process in nature. Genetic information is passed on by the formation of pairs between complementary nucleobases.[1,2] The detection of DNA or RNA sequences through hybridization also relies on base pairing, with multiple base pairs being formed simultaneously.[3–5] The two canonical base pairs (A:T and G:C) are of different strength, though, so that the affinity of strands for their target sequence depends strongly on the G/C content. Differences in duplex stability between one sequence and another are desirable for genetic material, for example to facilitate strand separation at a TATA box, but are undesirable for detecting nucleic acid sequences in parallel or in standardized fashion.[6] For example, the failure of A/T-rich sequences to hybridize under conditions appropriate for mixed sequences causes false negative signals in diagnostic tests, and the ability of G/C-rich sequences to form duplexes with mismatches can lead to false positive calls.[7]

Nature tunes base pairing strength, where needed, either through varying G/C content or through chemical modification, as best exemplified by modifications in tRNAs.[8] Synthetic chemists have also developed nucleobase analogues whose affinity for complementary bases exceeds that of their natural counterparts,[9–11] including bases with alkynyl substituents.[12] Among the two natural bases that pair weakly (A and T/U), adenine is the one more readily converted into a high affinity analogue. A well-established high affinity derivative of adenine is 2,6-diaminopurine, which forms three hydrogen bonds with thymine, resulting in more stable duplexes.[13] Binding to A with a high-affinity analogue of T is more difficult, though, as adenine does not offer a third functional group for hydrogen bonding on its Watson–Crick face, making it necessary to reach into the major or minor groove to engage more distant hydrogen bond acceptors or to provide other interactions. To the best of our knowledge, no nucleobase analogue that pairs with A with greater affinity and greater selectivity than thymine, when incorporated into DNA strands, is known.

Starting from thymidine as the lead, one possible way to engage in additional molecular interactions with adenine is to place an alkynyl group at the 2-position. Unfortunately, this leads to a pyrimidinium species when maintaining the aromatic character of the nucleobase (**1**, Figure 1). The instability of such species can be avoided when replacing a natural deoxynucleoside with a C-nucleoside, so that a neutral and more stable glycosidic bond arrangement results. We have previously reported a 2-ethynyl derivative of difluorotoluene (**2**), an isostere of thymidine,[14] in the form of ethynylfluorobenzene nucleoside **3**, and we were able to show that it pairs more strongly with adenine when incorporated in an oligonucleotide than its fluoro counterpart **2**.[15] Still, the overall base pairing strength of **3** was considerably weaker than that of natural thymidine, as evidenced by a depression in the UV-melting point of a dodecamer duplex by 12.7 °C, compared with the natural T-containing duplex. Herein, we present 6-ethynylpyridone C-nucleoside **4**, which combines the hydrogen bonding capabilities of T at the 3- and 4-positions with the ability of an ethynyl substituent to engage in stacking and van der Waals interactions. In duplexes, **E** pairs more strongly and more selectively with adenine than thymidine. In fact, **4** pairs with almost equal strength as a C does when it binds to G, as detected in UV-melting experiments.

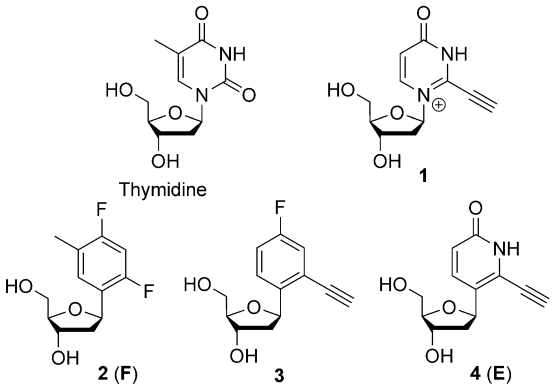


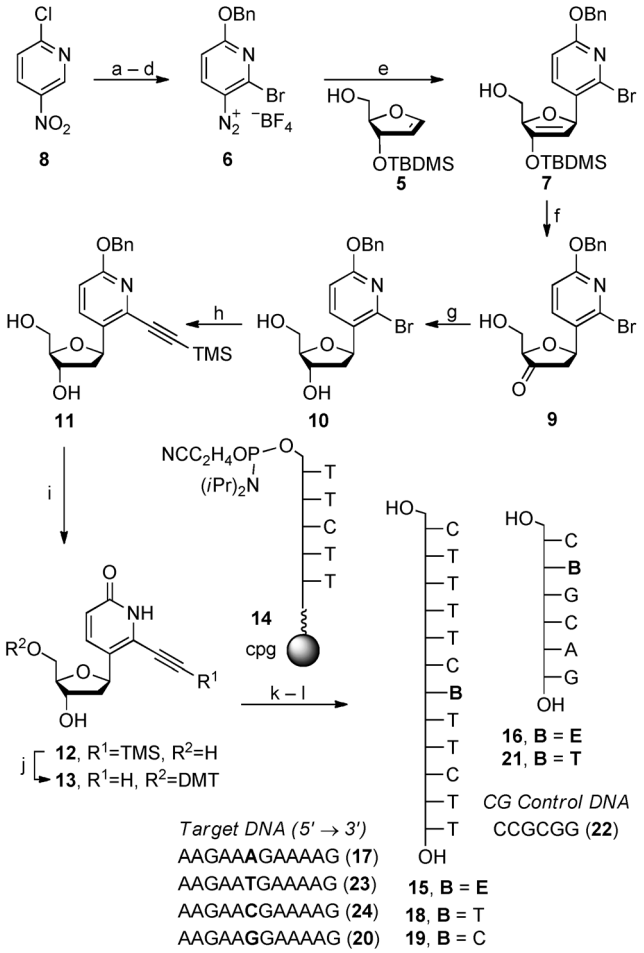
Figure 1. Structure of thymidine and thymidine analogues.

Several routes to a building block containing **4** were tested. Key intermediates of the preferred, but not yet optimized route are shown in Scheme 1. Details of the synthesis and spectroscopic data are given in the Supporting Information. Glycal **5**[16] was prepared from deoxyribose in four steps and 50 % overall yield, by a route that is shorter and higher yielding than earlier routes.[17,18] A Heck coupling was used to couple **5** to diazonium salt **6** in diastereoselective fashion, in an extension of elegant work by Hocek et al. on pyridyl C-nucleosides that employs less reactive aglycons,[19,20] producing enol ether **7**. Compound **6** was accessible from 2-chloro-4-nitropyridine **8** in four steps and 59 % overall yield. Desilylation of **7** to ketone **9** was followed by reduction with $NaBH(OAc)_3$ to give C-nucleoside **10**, whose stereochemical

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

[**] This work was supported by the DFG (RI 1063/9-1 and RI 1063/13-1 to C.R.) The authors thank Dr. N. Griesang and Dr. K. Müller for sharing results, D. Göhringer for skilled technical assistance, H. Griesser for a review of the manuscript, Dr. W. Frey for X-ray crystallography, and C. Kröner, S. Vollmer, and A. Göckel for help with acquiring spectra.

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201305555.



Scheme 1. Synthesis of C-nucleoside **13** and sequences of DNA strands. a) NaH, BnOH, THF, 83%; b) Fe, NH_4Cl, EtOH, 50°C, 96%; c) NBS, CH_3CN, 0°C, 83%; d) $BF_3{\cdot}OEt_2$, *t*$BuNO_2$, THF, −10°C, 95%; e) **5**, $Pd(OAc)_2$, Ag_2CO_3, CH_3CN; f) 3 HF·NEt_3, THF; g) $NaBH(OAc)_3$, CH_3CN, AcOH, 0°C, 33%; h) $[Pd(PPh_3)_2Cl_2]$, CuI, NEt_3, ethynyltrimethylsilane, 80°C, 63%; i) NaI, TMS-Cl, CH_3CN, 82%; j) 1) DMT-Cl, pyridine, 2) TBAF, THF, 29%; k) 1) **14**, tetrazole, CH_3CN, 2) I_2, pyridine, water, THF; l) DNA synthesis cycles, NH_4OH, 55°C. Bn = benzyl, cpg = controlled pore glass, DMT = 4,4′-dimethoxytrityl, NBS = *N*-bromosuccinimide, TBAF = tetrabutylammonium fluoride, TBDMS = *tert*-butyldimethylsilyl.

configuration was confirmed by X-ray crystallography (Supporting Information, Figure S24). A Sonogashira coupling at elevated temperature then gave **11**, which was debenzylated to produce pyridone **12**. Protection of the 5′-hydroxy group and removal of the trimethylsilyl (TMS) group led to **13** in modest yield. Because at this point **13** was deemed more valuable than a short standard oligonucleotide on cpg, we opted for phosphitylating the latter to produce **14**, followed by immediate on-support coupling to **13** under conditions previously reported.[21] After detritylation, standard coupling cycles completed the chain assembly, which was followed by conventional deprotection and HPLC purification, yielding oligodeoxynucleotide **15**, whose sequence was chosen to allow for comparison with other C-nucleosides.[15,22] Additionally, we prepared self-complementary oligonucleotide **16**, whose duplex $(\mathbf{16})_2$ contains two residues of the ethynylpyridone (**E**).

The ability of the nucleotide residue of **E** to pair with adenine was tested in UV-melting experiments. Because the absorbance maximum of **E** is shifted to 320 nm, the hyperchromicity accompanying duplex melting for strands containing **E** is smaller than that of duplexes made up of natural deoxynucleotides only. Figure 2 shows UV-melting curves of

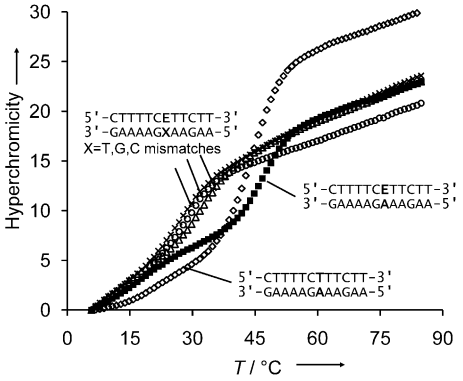


Figure 2. UV-Melting curves of dodecamer duplexes featuring a T:A, **E**:A, or **E**:C, **E**:G, or **E**:T base combination at the center of the helix. Conditions: strands (1.6 μM), phosphate buffer (10 mM, pH 7), NaCl (1 M), and $MgCl_2$ (10 mM). See Tables 1 and 2 for melting points.

the duplex between **E**-containing **15** with fully complementary DNA strand **17**, together with the corresponding curves of control duplex **18**:**17**, and duplexes of **15** with target strands containing mismatched nucleobases opposite **E**. It can be discerned that a significant increase in the melting point occurs when **E** replaces T in the duplex ($\Delta T_m = +2.6$°C; Table 1). This increase rivals that induced by replacing the

Table 1: UV-melting points [°C] of DNA:DNA or DNA:RNA duplexes.

Sequence	T_m[a]	$T_m(+Mg^{2+})$[a]	ΔT_m[b]
3′-GAAAAG**A**AAGAA-5′ (**17**)			
5′-CTTTTC**T**TTCTT-3′ (**18**)	43.8	44.2	–
5′-CTTTTC**E**TTCTT-3′ (**15**)	46.4	46.8	2.6
3′-GAAAAG**G**AAGAA-5′ (**20**)			
5′-CTTTTC**C**TTCTT-3′ (**19**)	46.8	47.0	2.8
r(3′-GAAAAG**A**AAGAA-5′) (**25**)			
5′-CTTTTC**T**TTCTT-3′ (**18**)	46.2	45.9	–
5′-CTTTTC**E**TTCTT-3′ (**15**)	48.2	48.0	2.1
(5′-C**T**GC**A**G-3′)$_2$ $(\mathbf{16})_2$[c]	25.2	25.8	–
(5′-C**E**GC**A**G-3′)$_2$ $(\mathbf{21})_2$[c]	37.3	37.4	11.6
(5′-C**C**GC**G**G-3′)$_2$ $(\mathbf{22})_2$[c]	40.8	41.0	15.2

[a] Average of two curves, either with or without $MgCl_2$; conditions as given in the legend to Figure 2. [b] ΔT_m to T:A-containing control at 10 mM $MgCl_2$. [c] Strand concentration = 6.4 μM.

T:A base pair in this duplex with a C:G base pair (duplex **19**:**20**; $\Delta T_m = +2.8$°C). For **15**:**17**, there is also a discernable transition at lower temperature, as expected for a sequence that can form triplexes, suggesting that the ethynyl group of **E** does not obstruct the major groove, where triplex formation

occurs. This finding is consistent with the acetylene of **E** pointing into the minor groove, as planned when designing **E**. Further, a very significant drop in melting point was observed for mismatch-containing duplexes. In fact, for any of the three possible mismatches, this drop was greater than in the natural duplexes, where a T rather than an **E** is facing the mismatched bases (Table 2).

Table 2: UV-melting points [°C] of DNA duplexes with a mismatched nucleobase opposite a T or **E** residue.

Sequence	Bases	T_m[a]	T_m (+ Mg^{2+})[a]	ΔT_m[b]
5′-CTTTTC**T**TTCTT-3′(**18**)				
3′-GAAAAG**T**AAGAA-5′(**23**)	T:T	30.9	30.9	−13.3
3′-GAAAAG**C**AAGAA-5′(**24**)	T:C	30.0	29.9	−14.3
3′-GAAAAG**G**AAGAA-5′(**20**)	T:G	32.3	32.5	−11.7
5′-CTTTTC**E**TTCTT-3′(**15**)				
3′-GAAAAG**T**AAGAA-5′(**23**)	**E**:T	32.3	32.3	−14.5
3′-GAAAAG**C**AAGAA-5′(**24**)	**E**:C	28.4	28.7	−18.1
3′-GAAAAG**G**AAGAA-5′(**20**)	**E**:G	26.6	27.0	−19.8

[a] Average of two curves; conditions as given in the legend to Figure 2. [b] ΔT_m to perfectly matched duplex in the presence of Mg^{2+}, listed in Table 1.

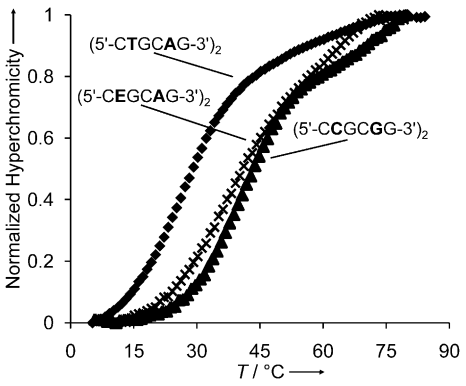


Figure 3. UV-Melting curves of duplexes of self-complementary sequences **16**, **21**, or **22**. Conditions: strands (6.4 μM), phosphate buffer (10 mM, pH 7), NaCl (1 M), and $MgCl_2$ (10 mM). For melting curves without normalization of hyperchromicity, see Figure S23 in the Supporting Information.

That the stability of an **E**:A base pair rivals that of a C:G base pair was confirmed in UV-melting experiments with self-complementary duplex $(\mathbf{16})_2$ and control duplexes $(\mathbf{21})_2$ and $(\mathbf{22})_2$ (Figure 3). The latter two duplexes feature two T:A or two C:G base pairs instead of the **E**:A pairs in $(\mathbf{16})_2$. The melting point increase for the **E**:A-containing duplex was 11.6 °C, whereas two C:G pairs replacing the T:A pairs of $(\mathbf{21})_2$ gave 15.2 °C. Further, the residue of **E** was also found to stabilize a duplex with RNA as the target strand, with a melting point increase of 2.1 °C for the duplex of dodecamer **15** with oligoribonucleotide **25** as target strand, as compared with the duplex of T-containing **18** with the same target (Table 1). Finally, modeling confirmed that there should be no steric conflict when the ethynyl group of **E** interacts with A in a canonical duplex (Figure 4).

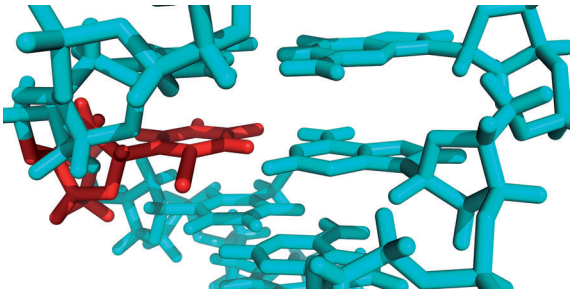


Figure 4. Modeling of a duplex where **E** pairs with deoxyadenosine in the interior of a helix. The coordinates were produced in Maestro 7.5 with B-form duplex **18**:**17** as the starting structure. The residue of **E** is shown in red, and the remainder of the sequence is in light blue.

The melting point increase observed for **E**:A over T:A is larger than that induced by propargyl groups at the 5-position.[12a] This suggests that there are more than just stacking interactions with neighboring nucleobases in the duplex. Further, there is a very significant increase in base-pairing selectivity (Table 2). Either observation is consistent with the idea that shape complementarity plays an important role in the pairing. The complementarity to adenine should be greater for **E** than for T. Further, alternative base pairs, such as wobble base pairs, should be more difficult to form with **E** because the ethynyl group can cause steric conflicts in non-canonical pairing arrangements.

By comparison, a decrease in duplex stability has been found for duplexes with a pyridone C-nucleoside lacking the ethynyl group[23] or featuring a fluorine or a methyl group in its place.[24] It is also noteworthy how different the pairing properties of **E** are from those of difluorotoluene (**F**, Figure 1). Whereas **F** causes a drop in the melting point by 18 °C in a dodecamer duplex[15] and a ΔT_m for a mismatched base facing it of less than 3 °C,[15,25] **E** causes a melting point increase of 2.6–5.8 °C per residue in perfectly matched duplexes, and a single mismatch depresses the T_m by 14.5–19.8 °C (Table 2).

In conclusion, we describe a new C-nucleoside that can replace thymidine in DNA. This C-nucleoside engages in high-fidelity base pairing and gives duplex stabilities approaching those of duplexes containing a strong base pair, demonstrating how powerful a combination of hydrogen bonding and shape complementarity is. We expect probes containing **E** instead of T to perform well under universally stringent conditions, helping to close a gap in our ability to form isostable duplexes[7c] when binding multiple sequences in parallel. Hybridization is crucial for many applications of DNA, ranging from PCR to nanostructuring. The ability to pair with A with near-equal strength to the pairs involving C, G, or diaminopurine has the potential to impact any of these applications.

Experimental Section

UV-melting curves were measured on a Lambda 25 spectrometer (PerkinElmer) with a heating or cooling rate of 1 °C min^{-1} and detection at 260 nm. Melting points are the extrema of the first

derivatives. Synthetic protocols and additional data can be found in the Supporting Information.

Please note: Minor changes have been made to this manuscript since its publication in *Angewandte Chemie* Early View. The Editor.

Received: June 27, 2013
Published online: August 23, 2013

Keywords: base pairing · DNA · nucleobases · oligonucleotides · RNA

[1] J. D. Watson, F. H. Crick, *Nature* **1953**, *171*, 964 – 967.
[2] A. Kornberg, T. A. Baker, *DNA Replication*, 2nd ed., University Science Books, Mill Valley, **2005**, p. 106.
[3] G. Felsenfeld, *Biochim. Biophys. Acta.* **1957**, *26*, 457 – 468.
[4] W. Saenger, *Principles of Nucleic Acid Structure*, Springer, New York, **1983**, p. 141.
[5] G. M. Blackburn, M. J. Gait, D. Loakes, D. M. Williams, *Nucleic Acids in Chemistry and Biology*, 3rd ed., RSC Publishing, Cambridge, **2006**, p. 178.
[6] a) M. C. Pirrung, *Angew. Chem.* **2002**, *114*, 1326 – 1341; *Angew. Chem. Int. Ed.* **2002**, *41*, 1276 – 1289; b) K. Siegmund, C. Ahlborn, C. Richert, *Nucleosides Nucleotides Nucleic Acids* **2008**, *27*, 376 – 388.
[7] a) H.-K. Nguyen, O. Fournier, U. Asseline, N. T. Thuong, D. Dupret, *Nucleic Acids Res.* **1999**, *27*, 1492 – 1498; b) U. Plutowski, C. Richert, *Angew. Chem.* **2005**, *117*, 627 – 631; *Angew. Chem. Int. Ed.* **2005**, *44*, 621 – 625; c) C. Ahlborn, K. Siegmund, C. Richert, *J. Am. Chem. Soc.* **2007**, *129*, 15218 – 15232.
[8] a) P. A. Limbach, P. F. Crain, J. A. McCloskey, *Nucleic Acids Res.* **1994**, *22*, 2183 – 2196; b) Y. Motorin, H. Grosjean in *Modification and Editing of RNA* (Eds.: H. Grosjean, R. Benne) ASM Press, Washington, DC, **1998**, p. 543 – 549.
[9] Selected references: a) K.-Y. Lin, M. D. Matteucci, *J. Am. Chem. Soc.* **1998**, *120*, 8531 – 8532; b) F. Seela, G. Becher, *Nucleic Acids Res.* **2001**, *29*, 2069 – 2078; c) H. Liu, J. Gao, S. R. Lynch, Y. D. Saito, L. Maynard, E. T. Kool, *Science* **2003**, *302*, 868 – 871.
[10] E. T. Kool, J. C. Morales, K. M. Guckian, *Angew. Chem.* **2000**, *112*, 1046 – 1068; *Angew. Chem. Int. Ed.* **2000**, *39*, 990 – 1009.
[11] a) C. R. Geyer, T. R. Battersby, S. A. Benner, *Structure* **2003**, *11*, 1485 – 1498; b) J. Gao, H. Liu, E. T. Kool, *Angew. Chem.* **2005**, *117*, 3178 – 3182; *Angew. Chem. Int. Ed.* **2005**, *44*, 3118 – 3122; c) S. Hikishima, N. Minakawa, K. Kuramoto, Y. Fujisawa, M. Ogawa, A. Matsuda, *Angew. Chem.* **2005**, *117*, 602 – 604; *Angew. Chem. Int. Ed.* **2005**, *44*, 596 – 598; d) Y. Doi, J. Chiba, T. Morikawa, M. Inouye, *J. Am. Chem. Soc.* **2008**, *130*, 8762 – 8768.
[12] Selected references: a) R. W. Wagner, M. D. Matteucci, J. G. Lewis, A. J. Gutierrez, C. Moulds, B. C. Froehler, *Science* **1993**, *260*, 1510 – 1513; b) N. C. Chaudhuri, E. T. Kool, *J. Am. Chem. Soc.* **1995**, *117*, 10434 – 10442; c) R. W. Wagner, M. D. Matteucci, D. Grant, T. Huang, B. C. Froehler, *Nat. Biotechnol.* **1996**, *14*, 840 – 844; d) J. C. Chaput, S. Sinha, C. Switzer, *Chem. Commun.* **2002**, 1568 – 1569; e) J. I. Gyi, D. Gao, G. L. Conn, J. O. Trent, T. Brown, A. N. Lane, *Nucleic Acids Res.* **2003**, *31*, 2683 – 2693; f) T. W. Barnes, D. H. Turner, *J. Am. Chem. Soc.* **2001**, *123*, 4107 – 4118; g) T. Kottysch, C. Ahlborn, F. Brotzel, C. Richert, *Chem. Eur. J.* **2004**, *10*, 4017 – 4028.
[13] K. H. Scheit, H. R. Rackwitz, *Nucleic Acids Res.* **1982**, *10*, 4059 – 4069.
[14] S. Moran, R. X. F. Ren, S. Rumney, E. T. Kool, *J. Am. Chem. Soc.* **1997**, *119*, 2056 – 2057.
[15] N. Griesang, C. Richert, *Tetrahedron Lett.* **2002**, *43*, 8755 – 8758.
[16] M. A. Cameron, S. B. Cush, R. P. Hammer, *J. Org. Chem.* **1997**, *62*, 9065 – 9069.
[17] R. S. Coleman, M. L. Madaras, *J. Org. Chem.* **1998**, *63*, 5700 – 5703.
[18] N. Joubert, R. Pohl, B. Klepetářová, M. Hocek, *J. Org. Chem.* **2007**, *72*, 6797 – 6805.
[19] T. Kubelka, L. Slavětínská, V. Eigner, M. Hocek, *Org. Biomol. Chem.* **2013**, *11*, 4702 – 4718.
[20] J. Štambaský, M. Hocek, P. Kočovský, *Chem. Rev.* **2009**, *109*, 6729 – 6764.
[21] a) M. Meng, C. Ahlborn, M. Bauer, O. Plietzsch, S. A. Soomro, A. Singh, T. Muller, W. Wenzel, S. Bräse, C. Richert, *ChemBioChem* **2009**, *10*, 1335 – 1339; b) A. Singh, M. Tolev, M. Meng, K. Klenin, O. Plietzsch, C. I. Schilling, T. Muller, M. Nieger, S. Braese, W. Wenzel, C. Richert, *Angew. Chem.* **2011**, *123*, 3285 – 3289; *Angew. Chem. Int. Ed.* **2011**, *50*, 3227 – 3231.
[22] B. A. Schweitzer, E. T. Kool, *J. Am. Chem. Soc.* **1995**, *117*, 1863 – 1872.
[23] a) T. Lan, L. W. McLaughlin, *J. Am. Chem. Soc.* **2000**, *122*, 6512 – 6513; b) K. K. Woods, T. Lan, L. W. McLaughlin, L. D. Williams, *Nucleic Acids Res.* **2003**, *31*, 1536 – 1540.
[24] Z. Sun, L. W. McLaughlin, *J. Am. Chem. Soc.* **2007**, *129*, 12531 – 12536.
[25] M. Egli, *Acc. Chem. Res.* **2012**, *45*, 1237 – 1246.

