Nucleic Acids

Assembly of the Complete Eight-Base Artificial Genetic Helix, xDNA, and Its Interaction with the Natural Genetic System**

*Jianmin Gao, Haibo Liu, and Eric T. Kool**

Much recent research has focused on the design of new base pairs that function in natural DNA.[1–11] Beyond this, a few recent studies have begun to explore the possibility of replacing *all* the base pairs of DNA, an approach which may eventually lead to entirely new genetic systems.[12–22] In addition to testing the limits of human-designed biological systems, such work might shed new light on the influence of base pairs on the storage and transfer of genetic information in DNA. Moreover, the resulting DNA-like molecules may be useful as experimental probes of physical interactions in enzyme active sites and at protein–DNA interfaces.

Toward these ends, a recent report described the pairing of benzo-homologated forms of adenine (xA) and thymine (xT) with natural thymine and adenine, respectively, to form a right-handed helical complex with a diameter approximately 2.4 Å larger than that of natural DNA.[12] The study revealed that in some cases these expanded DNA (xDNA) duplexes are more stable than the analogous sequence of natural DNA. Another report described the preparation and properties of the analogously expanded forms of cytosine and guanosine (xC and xG) as well.[16] All four expanded nucleobases were found to be inherently fluorescent, which makes them potentially applicable as biophysical tools.

[*] J. Gao,[+] H. Liu,[+] Prof. Dr. E. T. Kool
Department of Chemistry, Stanford University
Stanford, CA 94305-5080 (USA)
Fax: (+1) 650-725-0259
E-mail: kool@leland.stanford.edu

[+] These persons contributed equally to this work.

[**] This work was supported by the US National Institutes of Health (GM63587). J.G. and H.L. acknowledge Stanford Graduate Fellowships.

Supporting information for this article is available on the WWW under http://www.angewandte.org or from the author.

The combination of the four natural bases with their four expanded analogues—potentially yielding an eight-base artificial pairing system—has not been tested; this is of interest in part because of the large information storage and processing potential of such a molecular system.[23–28] Furthermore, the concept of interfacing the artificial xDNA genetic system with the natural one is intriguing, as the larger helices may have enhanced stability and inherent fluorescence. Targeting natural DNA or RNA sequences would require combination of the full set of four xDNA bases on one strand, which has not yet been explored. With these possibilities in mind, we describe herein the first studies of the full set of xDNA genetic building blocks.

We employed previously described methods to prepare the four xDNA deoxyribosides (Figure 1) as suitably protected phosphoramidite derivatives.[12,16,29] These were incorporated into a number of oligomeric sequences for testing helical assembly properties. Sequences contained the eight bases xC, xG, xA, xT, C, G, A, and T in various combinations (Table 1 and Supporting Information). The stability of the resulting helices was evaluated by optically monitored thermal denaturation; curve fitting and concentration studies were used to derive estimates of free energy (Supporting Information). Buffers contained Na·PIPES (10 mM, pH 7.0), Na^+ (100 mM) and Mg^{2+} (10 mM). CD spectra of putative helices (and DNA controls) were also recorded (Supporting Information).

The initial results with four different eight-base contexts revealed stable self-assembly into bimolecular complexes. Table 1 shows data for these first sequences. The thermal melting experiments revealed apparent all-or-none melting behavior and sigmoidal transitions that are similar in shape to those of the corresponding DNA control helices, but with generally higher transition temperatures (T_m) (Figure 2). The three strongest xDNA complexes gave increases in T_m (12–21 °C) as well as in estimated free energy (2.3–3.0 kcal mol^{-1} at 37 °C) relative to the native DNA helices. These three sequences include two cases in which xDNA bases were scattered on both strands, and one in which they were present on one strand only. The remaining sequence context (Figure 2d) also restricted the xDNA bases to one strand, but was further specialized, as purine and pyrimidine bases were segregated to opposite ends of each strand. Thermal stability in this case was slightly (2.9 °C) lower than that of the corresponding native DNA sequence, and the estimated free energy fell within experimental error of the control. The stoichiometry of the complexes was tested with mixing experiments in which the molar ratios of strand pairs were varied in regular increments. Plots for three of the sequence contexts are available in the Supporting Information; changes in absorbance for the fourth sequence were too small to yield reliable data. The results for the three plots gave strand stoichiometry values of 0.50:0.50, 0.45:0.55, and 0.49:0.51. Thus the data were most consistent with a double-stranded stoichiometry in all three complexes. However, as the hyper- and hypochromicity values were small, the stoichiometric assignments should be considered preliminary.

The success of the natural DNA genetic system rests not only in its ability to form double helices, but importantly, in its

  DOI: 10.1002/ange.200500069

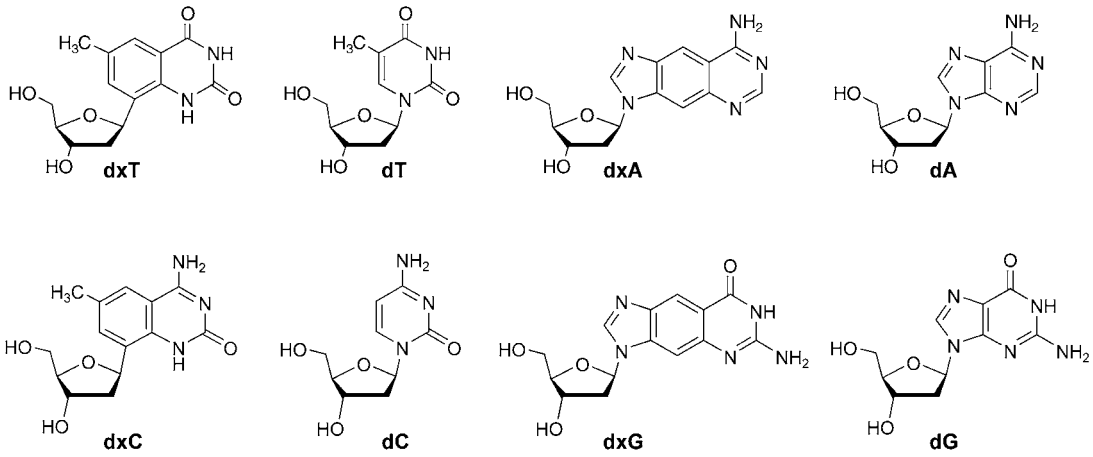

Figure 1. Structures of the eight nucleoside components of expanded DNA (xDNA).

Table 1: Thermal melting data and estimated free energies for decamer xDNA duplexes containing eight different bases.[a]

Entry	Sequence	T_m[b] [°C]	ΔG°_{37}[c] [kcal mol^{-1}]
1	5′-**A T** C A **C** T **G T** G C 3′-T A **G T** G **A** C A **C G**	59.9 ± 0.5	−12.8 ± 0.3
	5′-A T C A C T G T G C 3′-T A G T G A C A C G	44.6 ± 0.5	−10.0 ± 0.1
2	5′-**A C G T T A G T C G** 3′-T G C A A T C A G C	53.1 ± 0.5	−11.5 ± 0.2
	5′-A C G T T A G T C G 3′-T G C A A T C A G C	41.5 ± 0.5	−9.2 ± 0.1
3	5′-**T** G **T** A **C G** C **A G** T 3′-A **C** A **T** G C **G** T C **A**	63.4 ± 0.5	−11.9 ± 0.2
	5′-T G T A C G C A G T 3′-A C A T G C G T C A	40.3 ± 0.5	−8.9 ± 0.1
4	5′-**T C T T C G G A A G** 3′-A G A A G C C T T C	36.1 ± 0.5	−8.3 ± 0.1
	5′-T C T T C G G A A G 3′-A G A A G C C T T C	39.0 ± 0.5	−8.7 ± 0.1
5	5′-**A C G T T A G T C G** 3′-*U G C A A U C A G C*	47.8 ± 0.5	−9.4 ± 0.2
	5′-A C G T T A G T C G 3′-*U G C A A U C A G C*	42.6 ± 0.5	−9.5 ± 0.1
6	5′-**T C T T C G G A A G** 3′-*A G A A G C C U U C*	43.0 ± 0.5	−9.8 ± 0.2
	5′-T C T T C G G A A G 3′-*A G A A G C C U U C*	43.2 ± 0.5	−9.6 ± 0.1

[a] Benzo-homologated bases are represented in boldface; native DNA controls are shown for each sequence; entries 5 and 6: xDNA–RNA duplexes in which the RNA strand bases are italicized. Conditions: NaCl (100 mM), $MgCl_2$ (10 mM), and Na·PIPES (10 mM, pH 7.0). All DNA and xDNA sequences are 3′-phosphorylated (omitted for clarity). [b] Oligonucleotide concentration = 5.0 μM. [c] Averages of values from van't Hoff and curve-fitting methods.

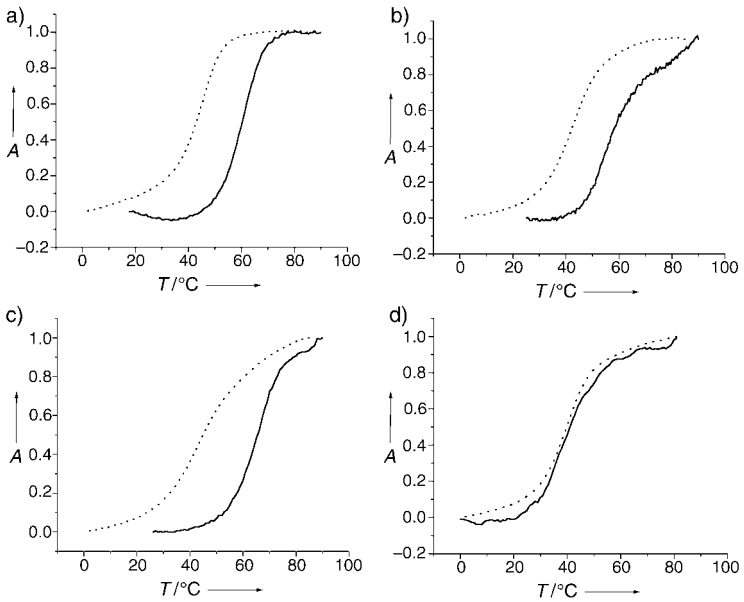

Figure 2. Normalized thermal denaturation plots for four different expanded DNA duplexes: a) d(xAxTCAxCTxGxTGCp)·d(xGxCACxAGxTxGATp) versus d(ATCACTGTGCp)·d(GCACAGTGATp); b) d(xAxCxGxTxTxAxGxTxCxGp)·d(CGACTAACGTp) versus d(ACGTTAGTCGp)·d(CGACTAACGTp); c) d(xTGxTAxCxGCxAxGTp)·d(xACTxGCGxTAxCAp) versus d(TGTACGCAGTp)·d(ACTGCGTACAp); d) d(xTxCxTxTxCxGxGxAxAxGp)·d(CTTCCGAAGAp) versus d(TCTTCGGAAGp)·d(CTTCCGAAGAp) for which the xDNA data were smoothed from five adjacent points to minimize noise from the small change in absorbance. Experimental conditions for data from both xDNA (——) and natural DNA (----): oligonucleotide (5 μM), NaCl (100 mM), $MgCl_2$ (10 mM), Na·PIPES buffer (10 mM, pH 7.0).

ability to do so with high specificity in sequence recognition. Thus it was of interest to test whether expanded-diameter xDNA helices could also exhibit sequence discrimination. This was done by evaluating the thermal stability of all possible natural-base mismatches of the four expanded bases. Figure 3 shows data from the analysis of all four matched pairs and the twelve possible mismatches. The same set of measurements were taken with native DNA as a control, so that a direct comparison of base-mismatch selectivities could

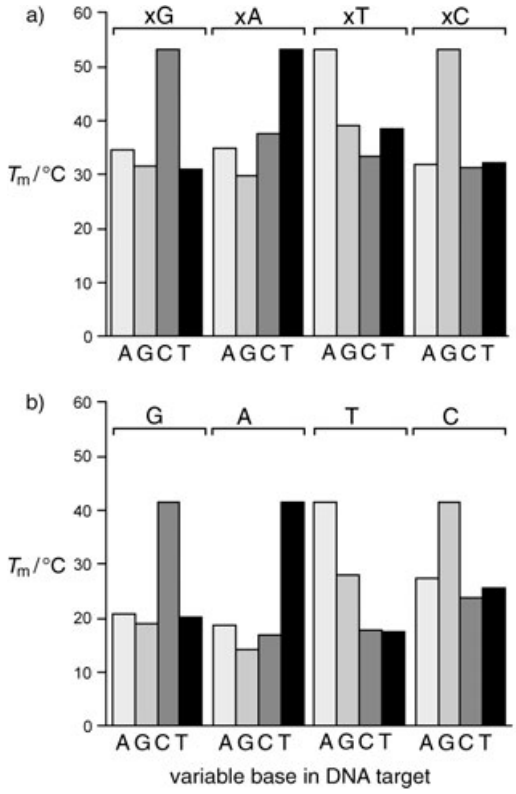

Figure 3. Histograms showing similar sequence selectivity for a) xDNA and b) native DNA of analogous sequence. Thermal melting temperatures are shown for singly mismatched and matched sequences in the context of a decamer duplex. (Details and conditions provided in Supporting Information).

be made (numerical data available in Supporting Information).

The results show that the four xDNA bases have surprisingly similar mismatch selectivity as natural DNA. Remarkably, the overall profiles of relative selectivities of xDNA had almost the same shape as those of native DNA double helices (Figure 3), which indicates that the two have similar mismatch preferences. The data show T_m values that are decreased by 13.9–23.4 °C for single xDNA mismatches, which are nearly the same as the respective decreased T_m range of 13.6–24.5 °C for mismatched bases in natural DNA. Therefore, oligonucleotide strands composed entirely of the four xDNA bases can recognize naturally structured DNA with high sequence selectivity and with high affinity.

In a search for evidence of helicity, CD was performed on all four double-stranded xDNA complexes (sequences shown in Table 1) in addition to the analogous DNA sequences as a comparison (Supporting Information). The xDNA spectra were clearly different from those of the control DNA, and also varied from sequence to sequence. Because the benzo-expanded DNA bases are electronically different from natural nucleobases, the excitonic interactions with stacked neighbors are different. However, CD spectra did reveal induced chirality in the xDNA that is consistent with helical complexes. An early structural study of a two-base xDNA system showed right-handed helicity,[17] which suggests that the present complexes also likely adopt a right-handed form. The variability in the xDNA spectra may arise from the fact that the sequences are complex: there are 64 possible nearest neighbors in this eight-base system (compared with only 16 in DNA). Thus these short xDNA sequences vary much more than native DNA in nearest-neighbor composition, which results in different electronic interactions.

Finally, full xDNA strands were tested for their ability to bind complementary RNA sequences. This was tested preliminarily in two all-xDNA strand contexts (Table 1, entries 5 and 6). Note that incorporation of the expanded bases into both strands of the duplexes was not possible, as "xRNA" nucleosides are not yet available. Representative denaturation curves for the xDNA–RNA complexes are given in the Supporting Information. The data showed clearly that both xDNA strands formed the predicted complexes with the decamer RNA target strands. In one case the xDNA–RNA complex was slightly more stable than the analogous DNA–RNA control, with an increase in T_m of 5.2 °C. In the other context, the thermal stabilities of the xDNA–RNA and DNA–RNA complexes were equal. Overall, the xDNA–RNA complexes were less stable than xDNA–DNA complexes (Table 1), which suggests that the expanded helices may prefer a B-like over an A-like helical conformation.

To our knowledge, this is the first report of a new genetic form that encodes as many as eight fundamental units of information. Previous studies from other research groups have focused on new hydrogen-bonded base pairs that are designed to function within the context of the natural genetic system.[1,5] Such artificial pairs, of which isoC–isoG and kappa–pi are examples, have been shown to pair selectively in naturally structured oligonucleotides.[30] Recent experiments with DNA polymerases have shown that such pairs may be replicated with sufficient selectivity to function properly in PCR amplifications.[31] The work reported herein is distinct from that precedent in two primary ways: first, all the current size-expanded base pairs adopt a structure that differs from the natural purine–pyrimidine base-pair structure, and second, the system presented herein is designed in its entirety, whereas the previous systems were meant to function in the natural context.

The current experiments illustrate a way for the xDNA artificial genetic system to interface with the natural biological genetic system. In the design of xDNA, the large base homologs can either be distributed among both strands (to yield the eight-base system described above), or they can be segregated into one of the two strands. In the latter scenario, the non-expanded strand is composed of natural DNA or RNA. The present results have confirmed for the first time that xDNA can recognize natural DNA and RNA in a sequence-dependent manner.

Finally, the eight-base xDNA design may lend itself to the storage and transfer of high-density information. Recent experiments have shown that natural DNA can be useful in encoding structural and computational information in designed nanostructures, computing systems, and machines.[23–28,32–39] The sequence complexity of natural DNA scales with 4^n, in which n = sequence length, whereas that of xDNA scales with 8^n. Thus the information density of xDNA is 2^n greater; for example, there are 262144 possible

hexamer sequences of xDNA, and only 4096 hexamer sequences of DNA. This high information density may be useful for data encryption and processing,[23–28] and in assembled structural designs of greater complexity than can be generated with DNA.

Experimental Section

The deoxynucleoside phosphoramidite derivatives of the four benzo-homologated nucleosides were prepared as described previously.[16,29] For synthesis of oligomers, the eight nucleosides described herein (Figure 1) were incorporated into oligonucleotides with automated solid-phase methods. All xDNA and DNA control sequences contained a 3′ phosphate group to avoid the need for eight different support resins, and were prepared with a "3′-phosphate-ON" controlled-pore glass support (Glen Research). Oligonucleotides were synthesized on an Applied Biosystems 392 DNA/RNA synthesizer in the trityl-off mode with standard β-cyanoethylphosphoramidite chemistry. Prolonged coupling times were used, which gave average stepwise yields of $>93\%$ by trityl monitoring. After synthesis, oligonucleotides were deprotected and removed from solid supports in the conventional manner. Oligomer products were purified by HPLC and quantified by UV/Vis absorption. Molar extinction coefficients were calculated by the nearest-neighbor method,[40] and those for oligonucleotides containing artificial residues were calculated by adding the extinction coefficient of the artificial nucleoside to the that of the core duplex. Previous reports have confirmed that the four modified xDNA nucleotides can be incorporated intact into oligonucleotides.[12,14,16] To confirm this for our work, we characterized two sequences by MALDI-TOF mass spectrometry: one with all four modified bases and the other with all eight bases, and the expected masses were observed. RNA oligonucleotides were purchased from Dharmacon. They were deprotected according to the manufacturer's protocol, then purified with denaturing preparative gel electrophoresis.

Thermal denaturation studies were conducted with sample concentrations that ranged between 1 and 40 μM in volumes of 1 mL in NaCl (100 mM), $MgCl_2$ (10 mM), and Na·PIPES (10 mM, pH 7.0). Solutions were then heated at 90 °C for 5 min and annealed by slow cooling to room temperature and then to 0 °C. Melting studies were carried out in teflon-capped quartz cells (path length = 1 cm) under N_2 atmosphere on a UV/Vis spectrophotometer (Varian Cary 1) equipped with a Peltier temperature controller. Absorbance was monitored at $\lambda = 320$ nm (xDNA) or $\lambda = 260$ nm (control DNA) during a temperature increase from 5 to 90 °C at a rate of 0.5 °C min^{-1}.

Melting temperatures (T_m) were determined by computer fit (Meltwin 3.5) of the first derivative of absorbance with respect to $1/T$. Uncertainty of T_m is estimated at ± 0.5 °C based on repetitions of experiments. Free energy values were estimated with two methods: 1) computer-fitting denaturation data with an algorithm that employs linear sloping baselines by using the two-state approximation for melting, and 2) van't Hoff thermodynamic parameters derived from linear plots of $1/T_m$ as a function of $\ln(C_T)$ by measuring T_m at varied concentrations (C_T = total DNA concentration). Values shown are averages from these two methods. They should be considered preliminary estimates only, as they rely on the assumption of a constant heat capacity (Cp) value,[41] which is not known to be the case for xDNA.

Received: January 7, 2005
Published online: April 18, 2005

Keywords: benzopurines · benzopyrimidines · DNA structures · nucleic acids · structure–activity relationships

[1] S. A. Benner, T. R. Battersby, B. Eschgfaller, D. Hutter, J. T. Kodra, S. Lutz, T. D. K. Baschlin, M. Blattler, M. Egli, C. Hammer, H. A. Held, J. Horlacher, Z. Huang, B. Hyrup, T. F. Jenny, S. C. Jurczyk, M. Konig, U. von Krosigk, M. J. Lutz, L. J. MacPherson, S. E. Moroney, E. Muller, K. P. Nambiar, J. A. Piccirilli, C. Y. Switzer, J. J. Vogel, C. Richert, A. L. Roughton, J. Schmidt, K. C. Schneider, J. Stackhouse, *Pure Appl. Chem.* **1998**, *70*, 263–266.
[2] E. T. Kool, *Acc. Chem. Res.* **2002**, *35*, 936–943.
[3] A. A. Henry, F. E. Romesberg, *Curr. Opin. Chem. Biol.* **2003**, *7*, 727–733.
[4] T. Mitsui, A. Kitamura, M. Kimoto, T. To, A. Sato, I. Hirao, S. Yokoyama, *J. Am. Chem. Soc.* **2003**, *125*, 5298–5307.
[5] H. P. Rappaport, *Nucleic Acids Res.* **1988**, *16*, 7253–7267.
[6] S. Atwell, E. Meggers, G. Spraggon, P. G. Schultz, *J. Am. Chem. Soc.* **2001**, *123*, 12364–12367.
[7] K. Tanaka, Y. Yamada, M. Shionoya, *J. Am. Chem. Soc.* **2002**, *124*, 8802–8803.
[8] H. Weizman, Y. Tor, *J. Am. Chem. Soc.* **2001**, *123*, 3375–3376.
[9] C. Brotschi, C. J. Leumann, *Angew. Chem.* **2003**, *115*, 1694–1697; *Angew. Chem. Int. Ed.* **2003**, *42*, 1655–1658.
[10] J. Parsch, J. W. Engels, *Nucleosides Nucleotides Nucleic Acids* **2001**, *20*, 815–818.
[11] G. Mathis, J. Hunziker, *Angew. Chem.* **2002**, *114*, 3335–3338; *Angew. Chem. Int. Ed.* **2002**, *41*, 3203–3205.
[12] H. Liu, J. Gao, S. R. Lynch, Y. D. Saito, L. Maynard, E. T. Kool, *Science* **2003**, *302*, 868–871.
[13] N. Minakawa, N. Kojima, S. Hikishima, T. Sasaki, A. Kiyosue, N. Atsumi, Y. Ueno, A. Matsuda, *J. Am. Chem. Soc.* **2003**, *125*, 9970–9982.
[14] N. J. Leonard, *Acc. Chem. Res.* **1982**, *15*, 128–135.
[15] H. Liu, J. Gao, E. T. Kool, *J. Am. Chem. Soc.* **2005**, *127*, 1396–1402.
[16] H. Liu, J. Gao, E. T. Kool, *J. Org. Chem.* **2005**, *70*, 639–647.
[17] H. Lu, K. He, E. T. Kool, *Angew. Chem.* **2004**, *116*, 5958–5960; *Angew. Chem. Int. Ed.* **2004**, *43*, 5834–5836.
[18] A. H. F. Lee, E. T. Kool, *J. Org. Chem.* **2005**, *70*, 132–140.
[19] H. Liu, S. R. Lynch, E. T. Kool, *J. Am. Chem. Soc.* **2004**, *126*, 6900–6905.
[20] K. Tanaka, A. Tengeiji, T. Kato, N. Toyama, M. Shionoya, *Science* **2003**, *299*, 1212–1213.
[21] F. Seela, A. Melenewski, C. Wei, *Bioorg. Med. Chem. Lett.* **1997**, *7*, 2173–2176.
[22] K. Groebke, J. Hunziker, W. Fraser, L. Peng, U. Diederichsen, K. Zimmermann, A. Holzner, C. Leumann, A. Eschenmoser, *Helv. Chim. Acta* **1998**, *81*, 375–474.
[23] L. M. Adleman, *Science* **1994**, *266*, 1021–1024.
[24] B. Shimanovsky, J. Feng, M. Potkonjak, *Inform. Hiding Lect. Notes Comput. Sci.* **2003**, *2578*, 373–386.
[25] X. P. Su, L. M. Smith, *Nucleic Acids Res.* **2004**, *32*, 3115–3123.
[26] A. Okamoto, K. Tanaka, I. Saito, *J. Am. Chem. Soc.* **2004**, *126*, 9458–9463.
[27] L. Gao, X. Yang, W. B. Liu, J. Xu, *Prog. Nat. Sci.* **2004**, *14*, 705–709.
[28] K. A. Schmidt, C. V. Henkel, G. Rozenberg, H. P. Spaink, *Nucleic Acids Res.* **2004**, *32*, 4962–4968.
[29] H. Liu, J. Gao, L. Maynard, Y. D. Saito, E. T. Kool, *J. Am. Chem. Soc.* **2004**, *126*, 1102–1109.
[30] C. R. Geyer, T. R. Battersby, S. A. Benner, *Structure* **2003**, *11*, 1485–1498.
[31] A. M. Sismour, S. Lutz, J. H. Park, M. J. Lutz, P. L. Boyer, S. H. Hughes, S. A. Benner, *Nucleic Acids Res.* **2004**, *32*, 728–735.
[32] D. S. Liu, S. Balasubramanian, *Angew. Chem.* **2003**, *115*, 5912–5914; *Angew. Chem. Int. Ed.* **2003**, *42*, 5734–5736.

[33] P. Yin, H. Yan, X. G. Daniell, A. J. Turberfield, J. H. Reif, *Angew. Chem.* **2004**, *116*, 5014 – 5019; *Angew. Chem. Int. Ed.* **2004**, *43*, 4906 – 4911.
[34] T. Ye, C. D. Mao, *J. Am. Chem. Soc.* **2004**, *126*, 11410 – 11411.
[35] W. B. Sherman, N. C. Seeman, *Nano Lett.* **2004**, *4*, 1203 – 1207.
[36] W. U. Dittmer, A. Reuter, F. C. Simmel, *Angew. Chem.* **2004**, *116*, 3634 – 3637; *Angew. Chem. Int. Ed.* **2004**, *43*, 3550 – 3553.
[37] Y. Chen, C. D. Mao, *J. Am. Chem. Soc.* **2004**, *126*, 8626 – 8627.
[38] Y. Chen, M. S. Wang, C. D. Mao, *Angew. Chem.* **2004**, *116*, 3638 – 3641; *Angew. Chem. Int. Ed.* **2004**, *43*, 3554 – 3557.
[39] Y. Chen, S. H. Lee, C. Mao, *Angew. Chem.* **2004**, *116*, 5449 – 5452; *Angew. Chem. Int. Ed.* **2004**, *43*, 5335 – 5338.
[40] M. J. Cavaluzzi, P. N. Borer, *Nucleic Acids Res.* **2004**, *32*, e13.
[41] A. Tikhomirova, N. Taulier, T. V. Chalikian, *J. Am. Chem. Soc.* **2004**, *126*, 16387 – 16394.