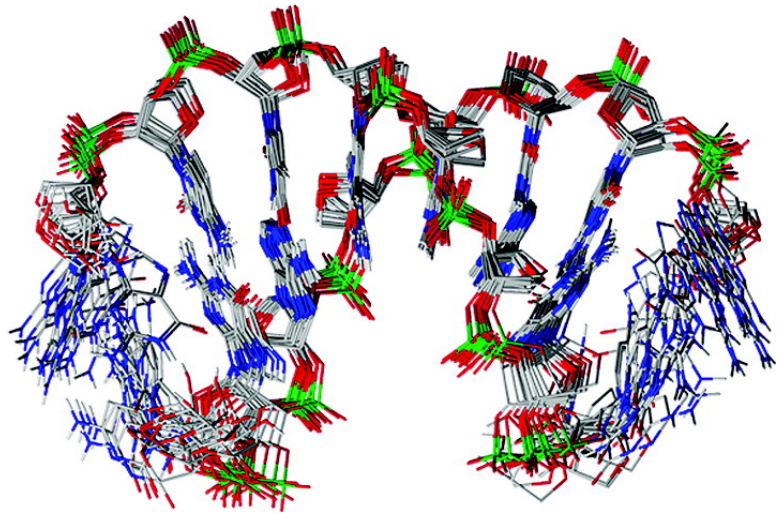
Article

The Structure of a TNA#TNA Complex in Solution: NMR Study of the Octamer Duplex Derived from #-(l)-Threofuranosyl-(3#-2#)-CGAATTCG

Marc-Olivier Ebert, Christian Mang, Ramanarayanan Krishnamurthy, Albert Eschenmoser, and Bernhard Jaun

J. Am. Chem. Soc., **2008**, 130 (45), 15105-15115 • DOI: 10.1021/ja8041959 • Publication Date (Web): 18 October 2008



J|A|C|S
ARTICLES

Published on Web 10/18/2008

The Structure of a TNA−TNA Complex in Solution: NMR Study of the Octamer Duplex Derived from α-(L)-Threofuranosyl-(3′-2′)-CGAATTCG

Marc-Olivier Ebert,[†] Christian Mang,[†,‡] Ramanarayanan Krishnamurthy,[‡] Albert Eschenmoser,[†,‡] and Bernhard Jaun*[,†]

Laboratory of Organic Chemistry, ETH Zurich, Wolfgang-Pauli-Strasse 10, CH-8093 Zürich, Switzerland, and The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92307

Received June 3, 2008; E-mail: jaun@org.chem.ethz.ch

Abstract: TNA (α-(L)-threofuranosyl-(3′-2′) nucleic acid) is a nucleic acid in which the ribofuranose building block of the natural nucleic acid RNA is replaced by the tetrofuranose α-(L)-threose. This shortens the repetitive unit of the backbone by one bond as compared to the natural systems. Among the alternative nucleic acid structures studied so far in our laboratories in the etiological context, TNA is the only one that exhibits Watson−Crick pairing not only with itself but also with DNA and, even more strongly, with RNA. Using NMR spectroscopy, we have determined the structure of a duplex consisting entirely of TNA nucleotides. The TNA octamer (3′-2′)-CGAATTCG forms a right-handed double helix with antiparallel strands paired according to the Watson−Crick mode. The dominant conformation of the sugar units has the 2′- and 3′-phosphodiester substituents in quasi-diaxial position and corresponds to a 4′-exo puckering. With 5.85 Å, the average sequential P_i-P_{i+1} distances of TNA are shorter than for A-type DNA (6.2 Å). The helix parameters, in particular the slide and *x*-displacement, as well as the shallow and wide minor groove, place the TNA duplex in the structural vicinity of A-type DNA and RNA.

Introduction

TNA (α-(L)-threofuranosyl-(3′-2′) nucleic acid) is a nucleic acid in which the pentose sugar building block of the natural nucleic acid RNA is replaced by the tetrose sugar α-(L)-threose.[1,2] The phosphodiester bridges in the TNA backbone connect the sugar units between the 3′- and 2′-positions of the threose (instead of the ribose 5′- and 3′-positions in the natural series). Thereby, the backbone displays a constitutional periodicity[3,4] of only five atoms (or bonds) as compared to the canonical six in DNA and RNA. Despite this rather drastic structural modification, the system undergoes efficient and selective Watson−Crick base-pairing with itself as well as with RNA and DNA. TNA-oligonucleotides were first synthesized and studied in the context of an experimental program[5] that has its roots in the question as to what extent the constitutional and configurational structure type of the canonical nucleic acids is functionally unique with respect to the property that constitutes the chemical basis of the nucleic acid's biological function, the capability of Watson−Crick base-pairing.[6,7] TNA is one of several alternative nucleic acid systems that were explored in our laboratories in this context: these are, besides homo-DNA,[8] three diastereomeric hexopyranosyl-(6′-4′)-oligonucleotides[9−12] and four diastereomeric β-pentopyranosyl-(4′-2′)-oligonucleotides,[13] including the pyranosyl form of RNA.[14,15] The backbone structures of all of these nucleic acid alternatives were (except the model system homo-DNA[8]) taken from the close structural neighborhood of RNA. The criteria for their selection were: first, a structure must be derivable from the corresponding sugar unit by the same *type* of chemistry that allows the structure of RNA to be derived from ribose;[5] second, a structure should be capable of informational Watson-Crick base-pairing. Among the modified oligonucleotide systems investigated, some were observed to be altogether incapable of Watson−Crick

[†] ETH Zurich.
[‡] The Scripps Research Institute.

(1) Schöning, K.-U.; Scholz, P.; Wu, X.; Guntha, S.; Delgado, G.; Krishnamurthy, R.; Eschenmoser, A. *Helv. Chim. Acta* **2002**, *85*, 4111–4153.
(2) Schöning, K. U.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. *Science* **2000**, *290*, 1347–1351.
(3) Moser, H.; Fliri, A.; Steiger, A.; Costello, G.; Schreiber, J.; Eschenmoser, A. *Helv. Chim. Acta* **1986**, *69*, 1224–62.
(4) Wippo, H.; Reck, F.; Kudick, R.; Ramaseshan, M.; Ceulemans, G.; Bolli, M.; Krishnamurthy, R.; Eschenmoser, A. *Bioorg. Med. Chem.* **2001**, *9*, 2411–2428.
(5) Eschenmoser, A. *Science* **1999**, *284*, 2118–2124.
(6) Eschenmoser, A. *Nachr. Chem., Tech. Lab.* **1991**, *39*, 795-8–804-7.
(7) Eschenmoser, A.; Dobler, M. *Helv. Chim. Acta* **1992**, *75*, 218–259.
(8) Hunziker, J.; Roth, H. J.; Boehringer, M.; Giger, A.; Diederichsen, U.; Goebel, M.; Krishnan, R.; Jaun, B.; Leumann, C.; Eschenmoser, A. *Helv. Chim. Acta* **1993**, *76*, 259–352.
(9) Diedrichsen, U. Ph.D. Thesis, ETH, Zurich, Nr. 10122, 1993.
(10) Fischer, R. Ph.D. Thesis, ETH, Zurich, NR. 9971, 1992.
(11) Groebke, K. Ph.D. Thesis, ETH, Zurich, Nr. 10149, 1993.
(12) Helg, A. G. Ph.D. Thesis, ETH, Zurich, Nr. 10464, 1994.
(13) Beier, M.; Reck, F.; Wagner, T.; Krishnamurthy, R.; Eschenmoser, A. *Science* **1999**, *283*, 699–703.
(14) Pitsch, S.; Wendeborn, S.; Jaun, B.; Eschenmoser, A. *Helv. Chim. Acta* **1993**, *76*, 2161–83.
(15) Pitsch, S.; Wendeborn, S.; Krishnamurthy, R.; Holzner, A.; Minton, M.; Bolli, M.; Miculka, C.; Windhab, N.; Micura, R.; Stanek, M.; Jaun, B.; Eschenmoser, A. *Helv. Chim. Acta* **2003**, *86*, 4270–4363.

base-pairing (e.g., the hexopyranosyl-(6′-4′)-oligonucleotides[9−12]), whereas others (e.g., the family of pentopyranosyl-(4′-2′)-oligonucleotides)[13,16−18] were found to undergo efficient and selective Watson−Crick pairing even more strongly than the natural isomers. These studies also revealed the Watson−Crick base-pairing mode's capacity to mediate more than one "base-pairing language" in the sense that, whereas none of the members of the family of diastereomeric pentopyranosyl-(4′-2′)-systems undergoes intersystem cross-pairing with the natural nucleic acids, they are perfectly fit for intrasystem base-pairing as well as intersystem cross-pairing within the family.[19] NMR-structure determinations in our laboratories of duplexes derived from the self-complementary base sequences AAAAATTTTT in the homo-DNA-series,[20] CGAATTCG in the ribopyranosyl-,[21] and arabinopyranosyl-RNA[22] series, as well as most recently an X-ray structure analysis of a homo-DNA duplex with the same sequence (CGAATTCG) by Egli and collaborators,[23] revealed the major structural factors that prevent these modified Watson−Crick base-pairing systems from cross-pairing with DNA and RNA. These are the preference for ladder-like pairing conformations and a large inclination[23] between the averaged backbone and base-pair axes, implying that these systems cannot adapt to the helical pairing conformations of the natural systems. The remarkable capacity of TNA to undergo not only intrasystem base-pairing but, in contrast to the other nucleic acid alternatives investigated in our laboratories, also cross-pairing with DNA and, especially, RNA received a first structural interpretation by Egli and collaborators.[24,25] They carried out X-ray structure analyses of A- and B-type DNA-duplexes of self-complementary base sequences containing a single α-(L)-threose residue. Here, we report the results of a NMR analysis of the solution structure of a full TNA duplex derived from the same self-complementary base sequence CGAATTCG as we already had studied in duplexes of the ribopyranosyl- and arabinopyranosyl-oligonucleotide series.[21,22]

TNA's potential significance for the etiology of nucleic acid structure[5] and the concept of a pre-RNA world[26,27] has recently been substantiated by chemical studies on the nonenzymatic template directed formation of RNA base sequences on TNA-templates,[28] and by the remarkable finding that certain DNA polymerases and reverse transcriptases are capable of catalyzing both a faithful transcription of TNA into DNA base sequences[29] as well as, importantly, one from DNA templates into TNA-sequences.[30−32]

Experimental Section

Synthesis of the t-(3′CGTTAACG2′) Sequence (1). The sequence was prepared from the corresponding A, T, G, and C TNA-phosphoramidites following a published procedure[1] with the following modifications to the *Pharmacia gene assembler* protocol: The detritylation time was changed from 6.7 to 15 min, and the coupling time of the monomer to the solid support was changed from 7.25 to 15.2 min. These modifications allowed us to synthesize the octamer on a 10 μmol scale with an average coupling efficiency of about 95% per step. The deprotection of the phosphate- and nucleobase-protecting groups, the detachment from the solid support, and HPLC purification of the octamer were accomplished using the previously described protocols.[1] The desalting of the oligomer (to remove NaCl) was achieved using a sephadex G-10 column. The product containing fractions (monitored by UV at 260 nm) were collected and concentrated by lyophilization in a speed-vac centrifuge. HPLC and mass spectral data (MALDI-TOF-MS) (see Figure S1a and S1b in the Supporting Information) document the degree of purity of the TNA octamer isolated and used for the NMR study. After it had turned out that the amount of the obtained material was not enough for an NMR investigation (70 O.D. instead of the planned 200 O.D.), the oligomer synthesis was repeated under the same conditions. The purification of this second batch was carried out with the help of Stefan Pitsch. The melting point of the duplex as determined by UV spectroscopy was 29.8 °C (conditions: concentration of single strand 10 μM, 1 M NaCl, 10 mM Na_2HPO_4, pH 7.0, 0.1 mM Na_2EDTA).

Sample Preparation. The oligonucleotide **1** (sodium salt, 100 O.D., 1.3 μmol) was dissolved in 0.25 mL of sodium arsenate buffer in D_2O (50 mM, pH 7.0). The resulting 5.2 mM solution was centrifuged and decanted into a NMR tube (Shigemi Inc.). For experiments involving exchangeable protons, the sample was lyophilized and redissolved in 0.25 mL of H_2O/D_2O 9:1.

NMR Spectra. Spectra were recorded on a Bruker DRX-500 spectrometer at 20 °C. Quadrature detection in the indirect dimension was achieved using TPPI. For measurements in H_2O/D_2O, the solvent signal was suppressed using excitation sculpting.[33] The spectra were processed on a SGI Octane2 workstation with the software XWIN-NMR 2.6[34] and analyzed with SPARKY 3.1.[35]

1D Spectra. ^{1}H spectrum in D_2O: spectral width (SW) 3125 Hz, 53 956 data points, 128 scans. ^{1}H spectra in H_2O/D_2O (T = 15.1, 19.5, 27.2 °C): SW 12 019 Hz, 88k data points, 128 scans. Broadband ^{1}H-decoupled ^{31}P spectrum: SW 2441 Hz, 15 756 data points, 64 scans. Broadband ^{1}H-decoupled ^{13}C spectrum: SW 12 019 Hz, 88k data points, 116 871 scans. These spectra were processed without zero filling and convoluted by multiplication of the fid with an exponential function (line broadening: ^{1}H spectra 0.1 Hz, ^{31}P spectra 1 Hz, ^{13}C spectra 2 Hz).

2D Spectra. NOESY. Spectra in D_2O were recorded with mixing times of 50, 100, and 150 ms, spectra in H_2O/D_2O with 100 ms. For the assignment of the ^{1}H signals, an additional spectrum in

(16) Jungmann, O.; Beier, M.; Luther, A.; Huynh, H. K.; Ebert, M.-O.; Jaun, B.; Krishnamurthy, R.; Eschenmoser, A. *Helv. Chim. Acta* **2003**, *86*, 1259–1308.
(17) Reck, F.; Wippo, H.; Kudick, R.; Krishnamurthy, R.; Eschenmoser, A. *Helv. Chim. Acta* **2001**, *84*, 1778–1804.
(18) Wagner, T.; Huynh, H. K.; Krishnamurthy, R.; Eschenmoser, A. *Helv. Chim. Acta* **2002**, *85*, 399–416.
(19) Jungmann, O.; Wippo, H.; Stanek, M.; Huynh, H. K.; Krishnamurthy, R.; Eschenmoser, A. *Org. Lett.* **1999**, *1*, 1527–1530.
(20) Otting, G.; Billeter, M.; Wuethrich, K.; Roth, H. J.; Leumann, C.; Eschenmoser, A. *Helv. Chim. Acta* **1993**, *76*, 2701–56.
(21) Schlönvogt, I.; Pitsch, S.; Lesueur, C.; Eschenmoser, A.; Jaun, B.; Wolf, R. M. *Helv. Chim. Acta* **1996**, *79*, 2316–2345.
(22) Ebert, M.-O.; Luther, A.; Huynh, H. K.; Krishnamurthy, R.; Eschenmoser, A.; Jaun, B. *Helv. Chim. Acta* **2002**, *85*, 4055–4073.
(23) Egli, M.; Pallan, P. S.; Pattanayek, R.; Wilds, C. J.; Lubini, P.; Minasov, G.; Dobler, M.; Leumann, C. J.; Eschenmoser, A. *J. Am. Chem. Soc.* **2006**, *128*, 10847–10856.
(24) Pallan, P. S.; Wilds, C. J.; Wawrzak, Z.; Krishnamurthy, R.; Eschenmoser, A.; Egli, M. *Angew. Chem., Int. Ed.* **2003**, *42*, 5893–5895.
(25) Wilds, C. J.; Wawrzak, Z.; Krishnamurthy, R.; Eschenmoser, A.; Egli, M. *J. Am. Chem. Soc.* **2002**, *124*, 13716–13721.
(26) Herdewijn, P. *Angew. Chem., Int. Ed.* **2001**, *40*, 2249–2251.
(27) Orgel, L. *Science* **2000**, *290*, 1306–1307.
(28) Heuberger, B. D.; Switzer, C. *Org. Lett.* **2006**, *8*, 5809–5811.
(29) Chaput, J. C.; Ichida, J. K.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 856–857.
(30) Kempeneers, V.; Vastmans, K.; Rozenski, J.; Herdewijn, P. *Nucleic Acids Res.* **2003**, *31*, 6221–6226.
(31) Chaput, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 9274–9275.
(32) Ichida, J. K.; Zou, K.; Horhota, A.; Yu, B.; McLaughlin, L. W.; Szostak, J. W. *J. Am. Chem. Soc.* **2005**, *127*, 2802–2803.
(33) Hwang, T.-L.; Shaka, A. J. *J. Magn. Reson., Ser. A* **1995**, *112*, 275–279.
(34) *XWIN-NMR Software Manual*; Bruker Analytik GmbH: Rheinstetten, 1997.
(35) Goddard, T. D.; Kneller, D. G. *SPARKY 3*; University of California: San Francisco, CA, 2002.

D_2O with a mixing time of 300 ms was recorded. 2D NMR parameters: see Table S2 in the Supporting Information.

Simulated H,P-COSY Spectrum. The spectrum was simulated using NMR-SIM2.7.[36] H,P coupling constants that resulted in the best agreement between experimental and simulated spectra were: ^3J(H2′,PO2′), 10 Hz (C1−C7); ^3J(H3′,PO3′), 1.5 Hz (T6, C7), 2 Hz (A4, T5), 3 Hz (A3), 5 Hz (G2, G8). T_2 relaxation time for all nuclei was set to 0.04 s. The same pulse sequence as for the experiment was used (spectral width in ω_2, 6.25 ppm; time increment in the indirect dimension, 0.2048 ms; total number of data points, 2k/512; number of accumulated scans for each increment, 16; recycle delay, 75 ms; a Gaussian function was used for apodization in both dimensions).

Generation of Restraints Based on the Two-Spin Approximation. Nonexchangeable Protons. For each of the NOESY spectra with 50, 100, and 150 ms mixing time, the coefficients a and b in eq 1 were determined independently using linear regression[37] (r, proton−proton distance; V, NOESY cross-peak volume). The following cross-peaks were used for calibration: H1′/H2′ (G2,A3), H4′1/H4′2 (A3, A4, T6), and H5/H6 (C7) with known distances of 2.85 ± 0.15, 1.8 ± 0.01, and 2.5 ± 0.01 Å. Among the three calibration distances, $d_{H1'H2'}$ is the only one with a weak dependence on the actual sugar conformation. This was taken into account by assigning a larger error limit to it. Integration of the NOESY spectra was done in SPARKY[35] by fitting the peaks with an appropriate set of gaussians.

$$r = a + b * V^{-1/6} \quad (1)$$

The calibration points were weighted by one over the square of their distance uncertainties. The coefficients a and b were then used to generate a distance restraint for every unambiguously assigned NOESY cross-peak. The average cross-peak volume over both sides of the diagonal was used for this. The volumes of cross-peaks involving a methyl group were divided by three. Three sets of distance restraints were obtained by this procedure. Of the maximum three distances for a given NOE correlation, the largest one was used for structure calculation. 20% of their value was added for the upper limits of the distance restraints. The lower limits were set to zero.

Exchangeable Protons. The same procedure as described above was used for the NOESY spectrum in H_2O/D_2O. The following cross-peaks were used for calibration: H4′1/H4′2 (T6) and H1′/H3 (T5, T6) with known distances of 1.8 ± 0.01 and 4.85 ± 0.15 Å. Distance $d_{H1'H3}$ depends on the dihedral angle χ only. Between $\chi = 180°$ and $\chi = -60°$, this distance varies by ca. 5%, which was taken into account by assigning a larger error limit to it. Only distances derived from cross-peaks to imino protons were considered for structure calculation.

For base pairs 2−7, hydrogen-bond restraints were implemented as distance restraints between the heavy atoms involved (3.1 ± 0.4 Å). An additional distance restraint (2.2 ± 0.2 Å) between the hydrogen atom and the acceptor was used to enforce the linearity of the hydrogen bonds.[38] A mild planarity restraint ($k_{plan} = 100$ kcal/mol) was used to prevent excessive staggering and buckling of the bases.[39] Thereby, the atoms N1, C6, and C2 of the purine bases were forced into the same plane as the atom N3 of the pyrimidine bases, and the atoms N3, C2, and C4 of the pyrimidine bases were forced into the same plane as the atom N1 of the purine bases.

Generation of Restraints Based on the MARDIGRAS Full Relaxation Matrix Approach. Nonexchangeable Protons. The restraint files used for the calculation of structures according to the two-spin approximation were modified, and every restraint was replaced by a very conservative distance range of 0−7 or 0−9 Å for cross-peaks involving a methyl group. From the converged (see below) structures in a preliminary first round of simulated annealing using this modified restraint file, the 15 lowest in energy were selected. Each of these structures served as input for a series of MARDIGRAS[40] calculations.

A RANDMARDI[41] procedure over 20 rounds was used. The relative error E of a peak volume was calculated using eq 2, where V is the peak volume and V_{av} is the corresponding average peak volume over both sides of the diagonal.

$$E = |V_{av} - V|/V_{av} \quad (2)$$

Average peak volumes were used in the MARDIGRAS calculations, and also the corresponding relative errors were averaged over the diagonal. The minimal relative error was set to 20%. Volumes of peaks that could only be integrated on one side of the diagonal were assigned a relative error of 100%. For the noise, an absolute unnormalized value of 50 000 was used. In subsequent rounds, the isotropic correlation time was set to 2.0, 2.5, 3.0, 3.5, and 4.0 ns. All volumes were used for normalization. Remaining parameters: FREQUENCY 500.03 Hz; MININT 2, MAXINT 20, DELTA6D 0.0005, R6THD 0.0005, METHYL JUMP 3.

The whole procedure was performed separately with volumes from the NOESY spectra in D_2O with mixing times of 50, 100, and 150 ms. Only cross-peaks that could be assigned in all three spectra were considered. Thus, 225 (3 sets of peak volumes × 5 correlation times × 15 input structures) sets of distance restraints were generated. For each restraint, the lowest and the highest value from this distribution were used as restraints in a second round of simulated annealing. From the converged structures, the 15 lowest in energy were again selected, and the MARDIGRAS procedure described above was repeated. The values for the lower and upper distance limits obtained in this second round were used for the final structure calculation that led to the structures in Figure 6b. Distances that were included in the calculation of structures in Figure 6a but were missing here were appended to the restraints file. Restraints identical to those in the calculation that led to the structures of Figure 6a were used for the exchangeable protons. Hydrogen bond and planarity restraints were also the same.

Structure Calculation. Structure calculation was done with the programs XPLOR 98.1[42,43] and XPLOR-NIH 2.0.4.[44] Topology and parameter files were adapted to TNA residues (the equilibrium values for bond lengths and angles were taken from an X-ray structure of a TNA nucleoside). The force constant for the stretching of a bond was 1000 kcal/mol, and the ones for angles and impropers were set to 500 kcal/mol. The simulated annealing procedure was the same as that proposed by Stein et al. for a DNA duplex.[45] Additionally, we added a short MD sequence of 3 ps before the final minimization (3000 steps). The calculated structures were defined to have converged if they showed no violation of a NOE restraint >0.2 Å, no violation of a dihedral angle restraint >5°, and no deviation from the equilibrium values for bond lengths, bond angles, and impropers >0.05 Å, > 5°, and >1.5°, respectively.

Results

The sequence-specific assignment of nonexchangeable protons was accomplished using information from DQF-COSY and H,P-

(36) Kessler, P. *NMR-SIM Software Manual. Version 2.7*; Bruker Analytik GmbH: Rheinstetten, 1998.
(37) Wijmenga, S. S.; van Buuren, B. N. M. *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 287–387.
(38) Varani, G.; Aboulela, F.; Allain, F. H. T. *Prog. Nucl. Magn. Reson. Spectrosc.* **1996**, *29*, 51–127.
(39) Kuszewski, J.; Schwieters, C.; Clore, G. M. *J. Am. Chem. Soc.* **2001**, *123*, 3903–3918.
(40) Borgias, B. A.; James, T. L. *J. Magn. Reson.* **1990**, *87*, 475–487.
(41) Liu, H.; Spielmann, H. P.; Ulyanov, N. B.; Wemmer, D. E.; James, T. L. *J. Biomol. NMR* **1995**, *6*, 390–402.
(42) Badger, J.; Kumar, R. A.; Yip, P.; Szalma, S. *Proteins: Struct., Funct., Genet.* **1999**, *35*, 25–33.
(43) Brünger, A. T. *X-PLOR. A System for X-Ray Crystallography and NMR*; Yale University Press: New Haven, CT, 1992.
(44) Clore, G. M.; Schwieters, C. D.; Kuszewski, J. J.; Tjandra, N. *J. Magn. Reson.* **2003**, *160*, 66–74.
(45) Stein, E. G.; Rice, L. M.; Brünger, A. T. *J. Magn. Reson.* **1997**, *124*, 154–164.

Table 1. ^{1}H and ^{31}P Chemical Shifts [ppm] of α-L-Threofuranosyl-(3′→2′)-(CGAATTCG)$_2$ at 20 °C

	H1′	H2′	H3′	H4′1	H4′2	H6[a]
C1	5.69	4.48	4.12	4.46	4.46	7.59
G2	5.73	5.09	4.84	4.18	4.18	7.67
A3	5.90	5.04	4.84	4.44	4.23	7.82
A4	5.89	4.64	4.72	4.61	4.17	7.73
T5	5.69	4.29	4.55	4.51	3.90	7.03
T6	5.82	4.66	4.55	4.41	3.93	7.41
C7	5.71	4.66	4.51	4.49	3.83	7.35
G8	5.43	4.31	4.53	4.16	3.92	7.30

	H5[b]	H2	H1[c]	NH_2[d]	^{31}P[e]
C1	5.74			—[f]	
G2			12.21	6.13	−2.77
A3		7.29		—	−3.43
A4		7.70		6.40	−3.85
T5	1.26		13.55		−4.35
T6	1.49		13.61		−4.52
C7	5.57			6.77, 8.30*	−3.69
G8			—	—	−2.79

[a] Pyrimidines; corresponding proton in purines: H8. [b] Cytidines; corresponding proton in thymidines: Me-C5. [c] Purines; corresponding proton in pyrimidines: H3. [d] NH_2 protons in H-bonds are denoted with an *. [e] ^{31}P resonances are listed in 3′ to 2′ direction; chemical shifts in ppm vs external 85% H_3PO_4. [f] No assignment possible.

Table 2. ^{13}C Chemical Shifts [ppm] of α-L-Threofuranosyl-(3′→2′)-(CGAATTCG)$_2$ at 20 °C

	C1′	C2′	C3′	C4′	C5	C6
C1	91.06	83.26	74.69	74.32	96.10	142.00
G2	86.12	82.83	79.97	72.60	—[a]	—
A3	86.35	82.70	79.97	72.57	—	—
A4	86.35	83.38	79.56	71.88	—	—
T5	88.62	82.18	80.13	70.29	—	137.20
T6	89.08	81.41	80.13	70.82	—	138.70
C7	89.96	81.40	80.11	70.26	97.39	141.20
G8	87.95	80.42	80.52	71.65	—	—

	C8	C2	Me
C1	—	—	
G2	136.40	—	
A3	138.80	151.50	
A4	138.80	152.10	
T5	—	—	11.52
T6	—	—	11.34
C7	—	—	
G8	136.30	—	

[a] Not determined.

COSY spectra and a NOESY spectrum with 300 ms mixing time of duplex **1** measured in D_2O, 50 mM sodium arsenate buffer at pH 7. The assignment of the geminal protons at C4′ could be verified independently by inspection of the ^{1}H,^{13}C-HSQC spectrum. Tables 1 and 2 list the resonances of the ^{1}H and ^{31}P nuclei as well as the resonances of the proton-bearing carbons.

Base-Pairing Mode. At 19.5 °C, the ^{1}H spectrum of **1** (5.2 mM in H_2O/D_2O 9:1, 50 mM sodium arsenate buffer at pH 7) shows three sharp signals in the region of the imino protons (Figure 2). Based on a simple model of an antiparallel TNA duplex in Watson−Crick pairing mode, interstrand NOE correlations between the imino protons of T5 and T6, and the H2 protons of the adenines in the partner strand are expected. Indeed, strong NOESY cross-peaks between the H2 resonances and the two imino resonances at lower field (Figure 3) were observed, whereas, in the same region, the third imino signal

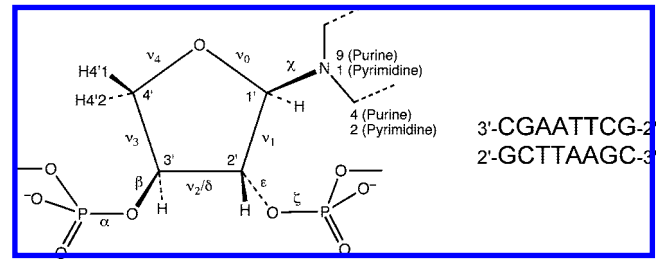


Figure 1. Chemical structure of the (3′-2′)-α-L-threofuranosyl nucleic acid repeating unit and base sequence of the duplex investigated. The definition of the backbone dihedral angles of TNA has been chosen in analogy to the natural systems[64] and is shown in the structural formula. Backbone atoms: $O2'_{i-1}-P_i-O3'_i-C3'_i-C2'_i-O2'_i-P_{i+1}$. χ-Purine: C4−N9−C1′−O4′. χ-Pyrimidine: C2−N1−C1′−O4′. Endocyclic sugar torsion angles: ν_0, C4′−O4′−C1′−C2′; ν_1,O4′−C1′−C2′−C3′; ν_2,C1′−C2′−C3′−C4′; ν_3,C2′−C3′−C4′−O4′; ν_4, C3′−C4′−O4′−C1′. Note that because of the shorter backbone TNA possesses only four exocyclic torsion angles as compared to five in RNA and DNA. To stay as close as possible to the conventional naming scheme for natural nucleic acids, we chose to leave away γ rather than ζ. Although not used in our analysis, δ is defined with respect to the flanking backbone atoms (also in analogy to RNA and DNA): O3′−C3′−C2′−O2′.

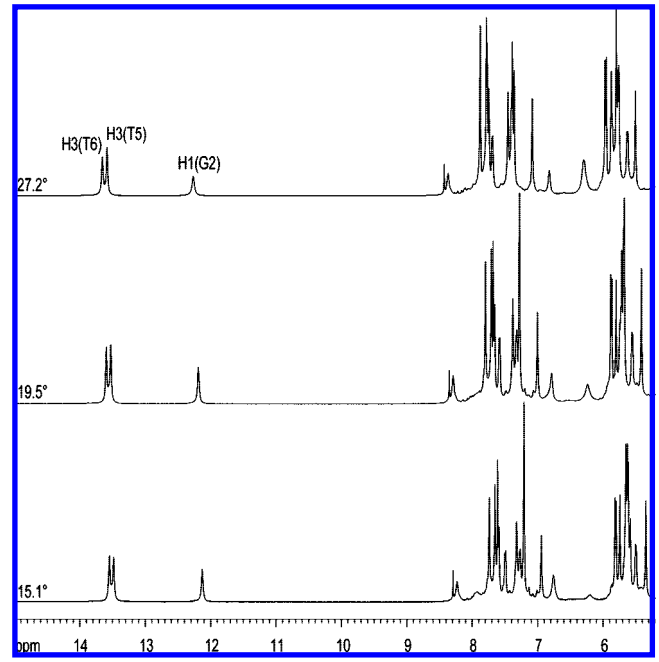


Figure 2. Detail of the ^{1}H NMR spectrum of α-L-threofuranosyl-(3′→2′)-(CGAATTCG)$_2$ in H_2O/D_2O 9:1, 50 mM sodium arsenate, pH 7.0, at different temperatures. The imino protons are sharpest at 19.5 °C.

showed only a weak correlation. With additional confirmation from inter- and intraresidual cross-peaks to the methyl groups of T5 and T6, we therefore assign the signals at 13.55 and 13.61 ppm to the H3 protons of T5 and T6.

The third imino signal was expected to belong to H1(G2). This was confirmed by a NOE to the amino group of the prospective pairing partner C7. This NOE is characteristic for CG base pairs in Watson−Crick pairing mode.[38] The assignment of the amino resonances of C7 themselves was obtained through NOEs to the H5 proton of the same residue and correlations to the methyl group of T6. The assignment of the amino group of A4 was also possible via its pairing partner. The assignment of the amino signal of G2 was based on the intense correlation with the already identified imino proton of the same base.

Similar to the pattern typical for RNA duplexes,[38] H1(G2) and the amino group of C7 both showed NOEs to the imino

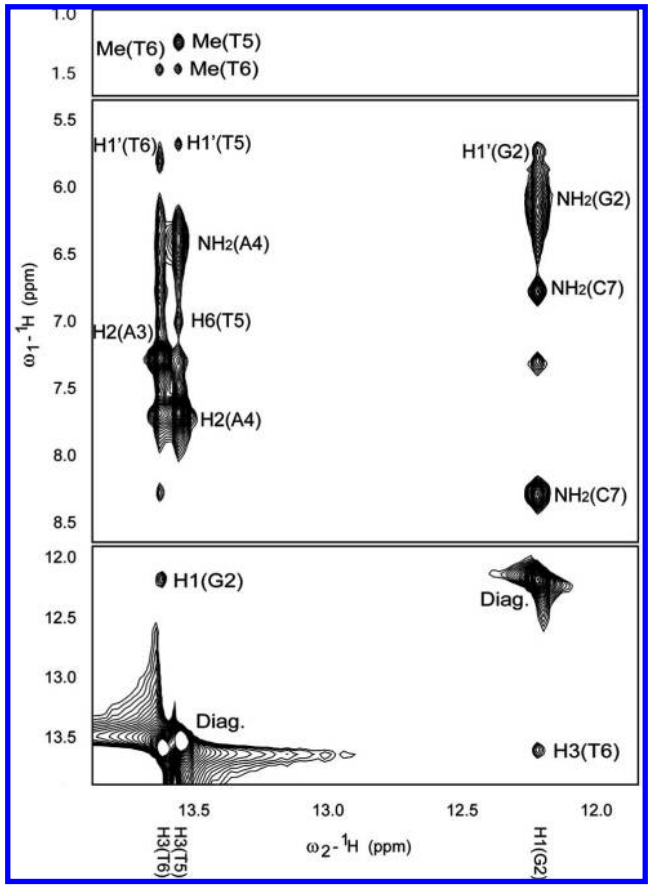


Figure 3. Detail of the NOESY spectrum (mixing time 100 ms) of α-L-threofuranosyl-(3′→2′)-(CGAATTCG)$_2$ in H_2O/D_2O 9:1, 50 mM sodium arsenate, pH 7.0, at 20 °C. Cross-peaks between imino and amino/aromatic protons are shown.

proton H3(T6) of the neighboring base pair. The remaining amino protons (A3, C1, G8) as well as H1(G8) could not be assigned. The missing signals of the outermost base pairs, however, are expected to be severely broadened under experimental conditions because of the shorter lifetime of the terminal base pairs. Taken together, the experimental evidence from the NOESY spectra demonstrates the formation of an antiparallel duplex in the Watson−Crick pairing mode.

Conformation of the Backbone As Deduced from Vicinal Coupling Constants. Sugar Ring Conformation. For all residues except A3, we were able to fit a set of gaussians to the F2-traces of the H1′,H2′ cross-peaks in the DQF-COSY spectrum, giving values for $^3J(1',2')$ of 2.0−2.4 Hz. Because of heavy spectral overlap and the low signal intensity of the corresponding cross-peaks, it was not possible to determine $^3J(2',3')$, $^3J(3',4'1)$, and $^3J(3',4'2)$ in the same way. For T6 and C7, there was no noticeable cross-peak between H2′ and H3′. For A3, T6, and G8, cross-peaks between H3′ and H4′1 were also absent. For A3, we found a weak cross-peak between H3′ and H4′2.

The description of furanose ring conformation in terms of pseudorotation parameters P (phase angle) and τ_m (puckering amplitude)[46,47] is not restricted to ribose. Empirical correlations between ribose H−H torsion angles and P and τ_m in ribose[48,49] were modified by Altona and co-workers to accommodate also diastereoisomeric furanose sugars.[50] H−H torsion angles (θ_{HH}) and the corresponding $^3J_{HH}$ coupling constants in the sugar ring are related via the Haasnoot equation, an empirical generalization of the Karplus equation.[51]

The relative configuration between stereogenic centers at C1′, C2′, and C3′ in α-L-threofuranose and β-D-xylofuranose nucleosides is identical. Consequently, the equations relating the corresponding H−H torsion angles and the pseudorotation parameters in xylose[50] were used for the analysis of the ring conformation in TNA. It is assumed that the local geometry at C4′ in TNA does not differ significantly from the one in β-D-xylofuranose nucleosides. Therefore, the relation between the pseudorotation parameters and $\theta_{H3'H4'2}$ in TNA was also taken from the literature.[50] Based on the same assumption, the remaining equation for $\theta_{H3'H4'1}$ is obtained by a 120° phase shift. This set of equations together with the Haasnoot equation allows the conformational analysis of the furanose ring in TNA in terms of P and τ_m based on H−H coupling constants.

Following these lines of argument, the values found for $^3J(1',2')$ correspond to a phase angle of ca. 40° (36−45° depending on the assumed puckering amplitude) on the pseudorotation circle and thus to a 4′-exo conformation[46,47] of the sugar ring. The predicted values for TNA in this conformational range are 0.6−1.2 Hz for $^3J(2',3')$, 1.5−2.0 Hz for $^3J(3',4'1)$, and 1.7−2.8 Hz for $^3J(3',4'2)$. The latter values explain the low intensity or absence of the corresponding cross-peaks. We have tacitly assumed that the ring stays predominantly in one conformation. The number of coupling constants available is, however, not sufficient to verify this assumption. Consequently, we did not include any restraints on the furanose ring conformation in our structure calculation (see below).

Backbone Angle ε. The resonances of the H2′-protons of G2 and A3 are shifted toward lower field (detail of ^{1}H spectrum in Figure 4). We iteratively fitted a set of Lorentzians to these signals in the ^{1}H spectrum (^{31}P coupled) starting with values of 2 Hz for $^3J(1',2')$ and approximately 1 Hz for $^3J(2',3')$. The final values for ^3J(H2′, PO2′) obtained by this procedure were 10 ± 1.5 Hz for both residues. This simple approach was not possible for other residues.

Inspection of the ^{1}H,^{31}P-COSY spectrum (Figure 4) showed, however, that all cross-peaks between PO2′ and H2′ have approximately the same intensity. This led to the conclusion that ^3J(H2′,PO2′) lies close to 10 Hz for every residue from C1 to C7. This value corresponds to a dihedral angle ε of ca. −120° (cf., the Karplus curve shown in Figure 5). Two additional regions of the Karplus curve near 0° and 120° would also match the experimental coupling constants but are not consistent with the formation of an antiparallel Watson−Crick duplex.

In the broadband-decoupled ^{13}C spectrum, we observed a splitting of the C1′-resonances of C1, T5, and T6 into a doublet. By fitting two Lorentzians to these signals, we obtained a value of 10 ± 1 Hz for ^3J(C1′,PO2′). This agrees with values of ε around −30° or −90°. Because the Karplus curve for ^3J(H2′, PO2′) shows a minimum at −30°, this region can safely be rejected on the basis of the ^3J(H,P) coupling constants discussed above. The other isolated C1′ signals showed no noticeable

(46) Altona, C.; Sundaralingam, M. *J. Am. Chem. Soc.* **1972**, *94*, 8205–8212.
(47) Sundaralingam, M. *J. Am. Chem. Soc.* **1971**, *93*, 6644–6647.
(48) De Leeuw, H. P. M.; Haasnoot, C. A. G.; Altona, C. *Isr. J. Chem.* **1980**, *20*, 108–126.
(49) Haasnoot, C. A. G.; De Leeuw, F. A. A. M.; De Leeuw, H. P. M.; Altona, C. *Org. Magn. Reson.* **1981**, *15*, 43–52.
(50) De Leeuw, F. A. A. M.; Altona, C. *J. Chem. Soc., Perkins Trans. 2* **1982**, 375–384.
(51) Haasnoot, C. A. G.; De Leeuw, F. A. A. M.; Altona, C. *Tetrahedron* **1980**, *36*, 2783–2792.

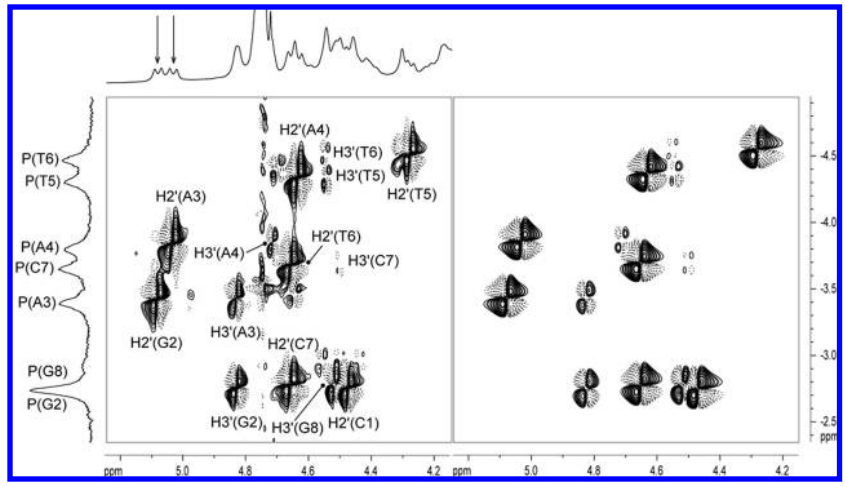


Figure 4. Left: H,P-COSY spectrum of α-L-threofuranosyl-(3′-2′)-(CGAATTCG)$_2$ in D_2O, 50 mM sodium arsenate, pH 7.0, at 20 °C. Arrows denote the positions of the H2′ signals of G2 and A3 in the ^{1}H spectrum. The resonances are split by ^3J(H2′,PO2′). Right: Simulated H,P-COSY spectrum (see Experimental Section).

splitting and are therefore indicative for values of ^3J(C1′,PO2′) < 10 Hz. Taken together, these findings corroborate the previously postulated region of ε close to −120°. For ^3J(H2′,PO2′), this value of ε lies close to a maximum on the Karplus curve. Hence, the coupling constant cannot be an average value resulting from large conformational fluctuations, and the vicinal coupling constants indicate that backbone motion around ε is severely restricted.

Backbone Angle *β*. Qualitative conformational analysis of the pairing system predicted a dihedral angle β in the trans-region of the Karplus curve, close to 180°. A simulated H,P-COSY spectrum with ^3J(H2′,PO2′) fixed at 10 Hz for residues C1−C7 agreed best with the experimental spectrum with values for ^3J(H3′,PO3′) of ca. 2 Hz for weak cross-peaks (A4, T5, T6, C7) and 3 Hz (A3) as well as 5 Hz (G2, G8) for the more intense cross-peaks (Figure 4). Inspection of the corresponding Karplus curve (Figure 5) shows that these coupling constants are consistent with the predicted value of β.

Derivation of Distance and Exocyclic Dihedral Restraints from NMR Data. All unambiguously assigned cross-peaks in the NOESY spectra recorded in D_2O were used for the generation of distance restraints (see Supporting Information Table S3).

Intrastrand correlations between the methyl groups of T5 and T6 and H2′, H2 and H8 of A3, and H1′ and H8 of A4, respectively (calculated distance restraints of 6.5 and 7 Å), contributed significantly to the definition of the global structure. From the cross-peaks in the NOESY spectrum recorded in H_2O/D_2O 9:1, only correlations to the imino protons were considered for the derivation of distance restraints (see Supporting Information Table S4). Because the observed three imino resonances as well as their NOE patterns prove the existence of six Watson−Crick base pairs, corresponding hydrogen-bond restraints were included into the calculation and treated like NOE restraints.

Based on the experimentally determined coupling constants, the allowed region for ε was restrained to −120 ± 30°. Because we could not determine a value for β directly from experimental data, this dihedral angle was only loosely restrained to a region of 180 ± 60°. The conformation of the furanose ring was not restricted through dihedral restraints, and α, ζ, and χ were also left unrestrained.

Table 3. Statistics for the Converged Structures in the Preliminary (48 out of 48) and Final (37 out of 48) Simulated Annealing Calculations

	final calculation (full relaxation matrix) Figure 6b	preliminary calculation (two-spin approximation) Figure 6a
number of NOE restraints[a]		
intraresidual	54	52
interresidual	50	48
interstrand	9	9
H-bonds	14	14
dihedral angle restraints[a,b]	12	12
rmsd from experimental restraints		
distance restraints [Å]	0.033 ± 0.002	0.012 ± 0.002
dihedral angle restraints [deg]	0.06 ± 0.09	0.03 ± 0.04
rmsd from holonomic restraints (force field)		
bonds [Å]	0.0048 ± 0.0001	0.0040 ± 0.0001
bond angles [deg]	0.808 ± 0.009	0.777 ± 0.004
impropers [deg]	0.10 ± 0.03	0.05 ± 0.01
atomic rmsd [Å]		
residues 2−7[c]	0.49	0.81
residues 3−6	0.47	0.61

[a] Because of C_2 symmetry of the duplex, each entry corresponds to a pair of restraints. [b] Only angles β and ε in residues 2−7 were restrained. [c] Superposition was carried out over backbone atoms O2′, C2′, C3′, O3′, and P. PO3′ of the first selected residue and PO2′ of the last selected residue were excluded from the superposition.

Structure Calculation. The distance and dihedral angle restraints derived from the NMR data were used to generate structural bundles by repeated simulated annealing calculations. In a preliminary simulation, distance restraints calculated from NOESY data using a modified two-spin approximation (Experimental Section) were used. For the final structure calculation, distance restraints derived with the program MARDIGRAS[40] from NOESY spectra with different mixing times were used. This approach allows one to treat spin diffusion explicitly by taking into account the complete relaxation matrix. Therefore, the error limits in the derived distances, especially for distances resulting from intense cross-peaks, could be set much narrower than this can safely be done under the two-spin approximation (for details, see Experimental Section). The statistics for both the preliminary and the final calculations are summarized in Table 3.

Figure 6 shows a bundle of the 15 converged structures with lowest energy calculated with the two-spin approximation

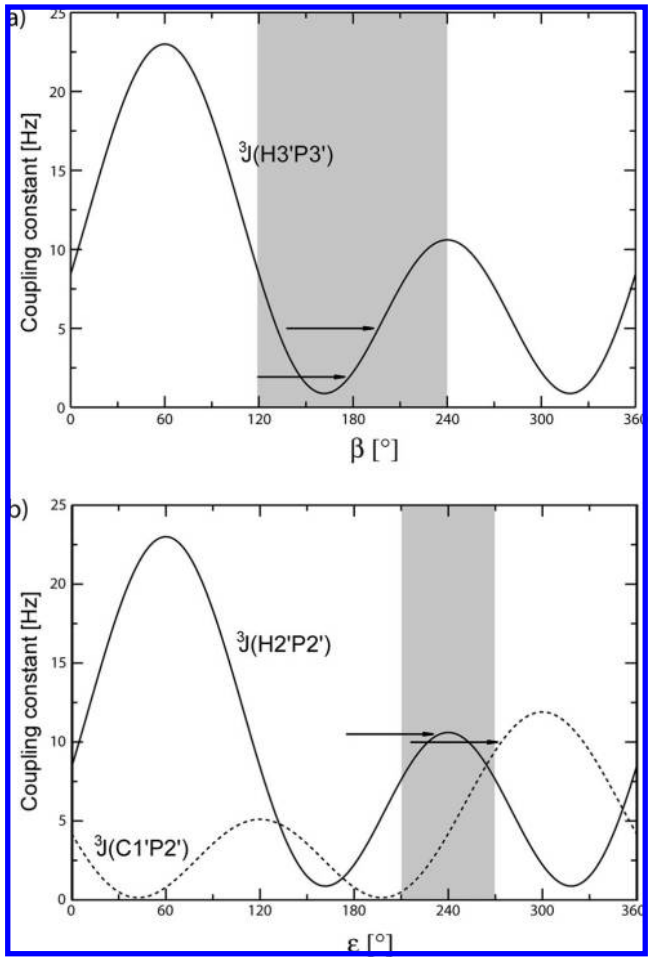


Figure 5. Karplus curves used for the derivation of the dihedral angles ε and β from ^3J(H,P) and ^3J(C,P).[65] Arrows denote typical experimental values found for TNA. Shaded areas indicate the range to which the angles were restrained during structure calculation.

(Figure 6a) and with MARDIGRAS (Figure 6b), respectively. As a result of narrower limits for the distance restraints, the atomic positions in Figure 6b are significantly better defined than in the bundle of Figure 6a. Clearly visible in these superpositions is also the fraying out of the structures at both ends of the duplex. On the one hand, this is the direct result of the lower number of restraints per residue for the terminal nucleotides. On the other hand, it is a consequence of the short-range of the employed NOE restraints that leave global structure (curvature of the duplex) partly undefined and, if the central six base pairs are used for the superposition, lead to an apparently less well-defined bundle at the ends. As can be seen by comparing the bundles, this second contribution to the fraying of the termini can be considerably reduced by a procedure that allows one to derive distance restraints more precisely.

Analysis of the Structures. Helix Parameters. As can be seen in Figure 6, the duplex forms a curved right-handed helix. Table 4 shows average helical parameters, calculated with the program 3DNA,[52,53] for the 10 structures with lowest energy in Figure 6b. Parameters describing the relation to and between the terminal base pairs were left out because their position is not well defined by experimental data. The absolute values of shear, stagger, and opening are almost consistently bigger than the corresponding average values found for either A- or B-DNA.[54] The average values for buckle and propeller are also elevated throughout the sequence as compared to the natural systems. Except for the third and sixth base pair that show slightly higher values, the values for stretch correspond closely to values of ca. −0.2 found in canonical A- and B-helices.

It should be emphasized that the distribution of base pair parameters depends heavily on (nonexperimental) hydrogen bond and planarity restraints normally used for the better definition of base pairs. Conservative restraints, as the ones used here, lead to a higher variability of base pair geometry, as made evident by the high standard deviations of some of the calculated base pair parameters.

The prominent roll throughout the sequence leads to a curvature of the entire duplex toward the major groove. The twist between the second and third and between the sixth and seventh base pairs is lowered by almost 10° as compared to the average value of 31° found in high-resolution crystal structures of A-DNA. The twist for the remaining base pair steps, however, matches this value more closely. The highest values for tilt are observed for the base pair steps that also show low twist. It is difficult to judge whether these findings mirror a real unwinding of the helix or are just the consequence of the diminished definition of the terminal residues by experimental restraints.

The average rise for the first and the last of the listed base pair steps is significantly smaller than the expected value of ca. 3.3 Å. This mirrors an extreme compression of bases G2 and A3 that is present in all of the structures. The elevated rise between the two central base pairs corresponds to the high and opposite values for the buckle of the two inner A·T base pairs ("cup"). It is likely that the looser stacking in the center of the helix is an artifact of structure determination. One principal reason for this is the symmetry equivalence of the imino protons in T5 and T13. This leads to ambiguous cross-peaks in the NOESY spectra and precluded the detection of potentially important imino−imino correlations between the central base pairs.

The slide for the listed base pairs lies between −1.1 and −1.5 Å. This parameter is especially interesting because it differs significantly between A- and B-helices of natural nucleic acid duplexes and thus allows us to establish a structural relationship between TNA and its natural counterparts. Values <-0.8 Å are typical for A-DNA, whereas for B-DNA values >0.8 Å are commonly found.[54] As judged from this parameter, the investigated TNA duplex is structurally close to A-DNA or RNA.

The remaining parameters in Table 4 are defined in relation to a local helical axis. The *x*-displacement corresponds to the diameter of the cavity in the center of the helix. It is close to zero for B-DNA but distinctly negative for an A-helix. The average *x*-displacement for the listed TNA base pairs is −4.6 ± 0.6 Å and corroborates the similarity to A-DNA and RNA.

Backbone Dihedral Angles. The global distributions of the backbone dihedral angles (bundle of Figure 6b) are shown in Figure 7. The angles α and β of G2 and G8, ε and ζ of C1 and C7, and χ of C1 and G8 are not included in this picture because of their proximity to the experimentally less precisely defined terminal residues. The angles α and ζ are located in a sector of $-90 \pm 30°$. As would be expected for this pairing mode, values

(52) Lu, X. J.; Shakked, Z.; Olson, W. K. *J. Mol. Biol.* **2000**, *300*, 819–840.
(53) Lu, X. J.; Olson, W. K. *Nucleic Acids Res.* **2003**, *31*, 5108–5121.
(54) Olson, W. K.; Bansal, M.; Burley, S. K.; Dickerson, R. E.; Gerstein, M.; Harvey, S. C.; Heinemann, U.; Lu, X. J.; Neidle, S.; Shakked, Z.; Sklenar, H.; Suzuki, M.; Tung, C. S.; Westhof, E.; Wolberger, C.; Berman, H. M. *J. Mol. Biol.* **2001**, *313*, 229–237.

Table 4. Average Helical Parameters of Base Pairs 2−7 in the 10 Structures Lowest in Energy (Bundle in Figure 6b), Calculated with the Program 3DNA[52,53]

Base-Pair Parameters	shear	stretch	stagger	buckle	propeller	opening
G2·C7	0.0 ± 0.4	−0.2 ± 0.1	0.1 ± 0.1	3.6 ± 2.8	−10.8 ± 3.2	−2.8 ± 2.7
A3·T6	−0.3 ± 0.1	−0.3 ± 0.0	0.3 ± 0.1	6.8 ± 3.8	−15.6 ± 3.2	−3.9 ± 1.7
A4·T5	−0.1 ± 0.7	−0.2 ± 0.1	0.2 ± 0.1	6.7 ± 4.2	−10.7 ± 4.8	5.8 ± 5.8
T5·A4	0.2 ± 0.5	−0.2 ± 0.1	0.3 ± 0.1	−8.7 ± 3.8	−12.8 ± 5.3	6.2 ± 3.8
T6·A3	0.3 ± 0.1	−0.3 ± 0.0	0.3 ± 0.1	−6.8 ± 3.0	−18.7 ± 2.5	−4.6 ± 1.2
C7·G2	−0.1 ± 0.4	−0.2 ± 0.1	0.2 ± 0.1	−4.9 ± 2.1	−14.2 ± 2.1	−5.2 ± 2.3
average	0.0 ± 0.2	−0.2 ± 0.1	0.2 ± 0.1	−0.5 ± 7.1	−13.8 ± 3.1	−0.8 ± 5.3

Base-Pair Step Parameters	shift	slide	rise	tilt	roll	twist
G2A3/T6C7	0.1 ± 0.2	−1.1 ± 0.1	2.8 ± 0.1	−4.0 ± 0.6	11.3 ± 1.0	22.3 ± 1.6
A3A4/T5T6	−0.1 ± 0.3	−1.3 ± 0.2	3.1 ± 0.3	0.1 ± 2.0	13.6 ± 3.3	31.8 ± 3.9
A4T5/A4T5	0.0 ± 0.5	−1.5 ± 0.1	3.8 ± 0.2	0.0 ± 0.6	9.5 ± 2.2	28.0 ± 4.3
T5T6/A3A4	0.1 ± 0.2	−1.2 ± 0.2	3.1 ± 0.2	−0.9 ± 1.4	15.1 ± 2.8	31.8 ± 2.6
T6C7/G2A3	−0.2 ± 0.2	−1.1 ± 0.1	2.8 ± 0.1	3.9 ± 0.6	10.5 ± 1.2	23.0 ± 2.1
average	0.0 ± 0.2	−1.3 ± 0.2	3.1 ± 0.4	−0.2 ± 2.8	12.0 ± 2.3	27.4 ± 4.6

Base-Pair Parameters with Respect to Local Helical Axis	*x*-disp	*y*-disp	rise	incl.	tip	twist
G2A3/T6C7	−5.1 ± 0.5	−1.2 ± 0.5	1.9 ± 0.2	27.0 ± 2.9	9.6 ± 1.9	25.4 ± 1.4
A3A4/T5T6	−4.0 ± 0.5	0.2 ± 0.7	2.3 ± 0.4	23.8 ± 6.7	0.1 ± 3.2	34.8 ± 3.1
A4T5/A4T5	−5.3 ± 1.0	0.0 ± 1.1	3.1 ± 0.3	19.1 ± 4.8	−0.1 ± 1.2	29.6 ± 4.1
T5T6/A3A4	−3.9 ± 0.4	−0.2 ± 0.5	2.3 ± 0.2	25.9 ± 4.8	1.4 ± 2.0	35.2 ± 2.4
T6C7/G2A3	−4.8 ± 0.6	1.4 ± 0.6	2.1 ± 0.2	24.6 ± 3.0	−9.2 ± 2.0	25.6 ± 1.9
average	−4.6 ± 0.6	0.0 ± 1.0	2.3 ± 0.4	24.1 ± 3.0	0.4 ± 6.7	30.1 ± 4.8

Table 5. Statistics of the Backbone Dihedral Angles of the Accepted Structures of α-L-Threofuranosyl-(3′→2′)-(CGAATTCG)$_2$ (Bundle of Figure 6b)[a]

	α	β	ε	ζ	χ	ν_0
C1	—[b]	–	−97 ± 40	−46 ± 51	−97 ± 40	−8 ± 16
G2	−104 ± 32	174 ± 20	−114 ± 1	−78 ± 1	−114 ± 1	−1 ± 10
A3	−77 ± 1	172 ± 1	−95 ± 10	−98 ± 19	−95 ± 10	−21 ± 1
A4	−97 ± 27	−164 ± 16	−135 ± 12	−62 ± 11	−135 ± 12	14 ± 19
T5	−117 ± 15	−143 ± 18	−115 ± 14	−82 ± 11	−115 ± 14	−14 ± 1
T6	−92 ± 16	−170 ± 22	−91 ± 3	−93 ± 3	−91 ± 3	−14 ± 1
C7	−81 ± 3	166 ± 3	−135 ± 8	−50 ± 11	−135 ± 8	−20 ± 2
G8	146 ± 14	162 ± 20	–	–	162 ± 20	−7 ± 13

	ν_1	ν_2	ν_3	ν_4	P[c]	τ_m[d]
C1	−3 ± 9	13 ± 24	−20 ± 32	18 ± 30	47 ± 51	39 ± 1
G2	−20 ± 1	33 ± 10	−36 ± 15	24 ± 16	22 ± 24	39 ± 0
A3	−3 ± 1	25 ± 1	−39 ± 0	38 ± 1	52 ± 2	41 ± 0
A4	−26 ± 3	27 ± 14	−20 ± 25	4 ± 27	47 ± 37	40 ± 1
T5	−11 ± 1	30 ± 1	−40 ± 1	34 ± 1	39 ± 1	39 ± 1
T6	−9 ± 1	27 ± 1	−36 ± 1	32 ± 1	40 ± 2	35 ± 1
C7	−4 ± 2	26 ± 2	−39 ± 1	38 ± 1	51 ± 2	41 ± 1
G8	−14 ± 7	28 ± 9	−33 ± 15	26 ± 17	36 ± 28	37 ± 2

[a] Average and standard deviation. [b] Not defined. [c] Phase angle of pseudorotation. P was calculated from torsion angles ν_0−ν_4 according to eq 3 in ref 46. [d] Puckering amplitude. Once P is known, τ_m can be calculated from each of the endocyclic torsion angles ν_0−ν_4 (see eq 1 in ref 46). The indicated value for τ_m is the average of these five values.

of χ are limited to a region centered at −120°. β is mostly confined to a region close to 180°, but values up to −120° are populated. ε, on the other hand, populates the whole accessible region that is not prohibited by dihedral angle restraints. The dihedral angles in the bundle of Figure 6a show an almost identical global distribution (see Supporting Information Table S5).

A detailed summary of dihedral angles, sorted according to their sequential position, is compiled in Table 5. The average values of α and β for T5 as well as ε and ζ for A4 are significantly different from their average values in other experimentally well-defined positions. These dihedrals lie close to the center of symmetry of the duplex and are, again, most likely biased by the scarce number of unambiguous distance restraints in the core part of the duplex. In the preliminary calculation (Figure 6a), distance restraints were chosen much looser, and, consequently, such sequence-dependent differences are not observable (see Supporting Information Table S5).

Phase angles P for the bundle in Figure 6b are also shown in Figure 7. Only four structures exhibit a residue among G2−C7 with P close to −90° (always G2). In some of the structures, A4 shows a conformation between O1′-endo and C1′-endo. The average value for the phase angles of the remaining residues lies between 36° and 52°, close to the C4′-exo conformation.

Groove Widths. A comparison between the groove widths of B-DNA, RNA, and the investigated TNA duplex is shown in Figure 8. The minor groove of the short TNA helix is well defined. It is wide and shallow and closely resembles that of

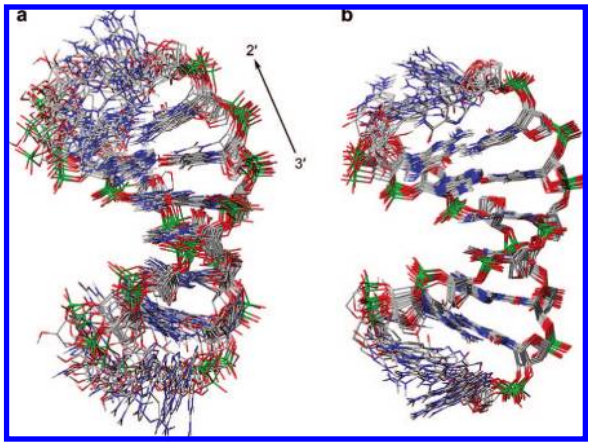


Figure 6. Superposition of the 15 converged structures lowest in energy for (a) distance restraints derived by the two-spin approximation and (b) distance restraints derived using the full relaxation matrix approach (MARDIGRAS). Superposition was carried out over the backbone atoms of residues 2−7 in both strands.

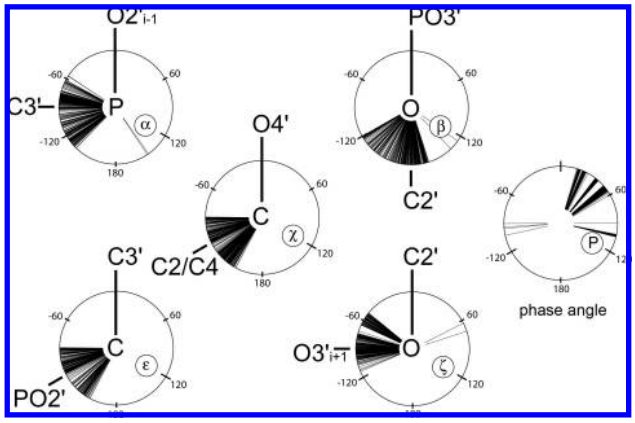


Figure 7. Dihedral angle distribution in the accepted structures (bundle shown in Figure 6b). Data included: α and β of residues 3−7, ε and ζ of residues 2−6, χ of residues 2−7, and pseudorotation phase angle P of the furanose ring (terminal residues are excluded from the picture).

the RNA helix. For the major groove, however, a comparison is difficult because, for such a short sequence, one cannot easily distinguish between the curvature of the helix and groove width. As mentioned above, the structure at hand exhibits significant curvature. The major groove of a longer, straight TNA sequence would likely be at least as wide as shown in Figure 8 but would loose some depth. Such a hypothetical major groove would be shallower than in RNA but wider than that of either a B- or an A-helix.

Discussion

A. Structure. The TNA duplex **1** forms a curved right-handed double helix in the Watson−Crick pairing mode. With the short sequence under study, it is difficult to judge whether global features such as the curvature toward the major groove that is present in all structures of the calculated bundle, or the unwinding of the helix toward the termini, are artifacts of the dynamic averaging of the measured NOEs and coupling constants. The general structural characteristics, however, are well defined.

The significant variation of ^3J(H3′,PO3′) found experimentally cannot readily be explained by a static sequence-dependent

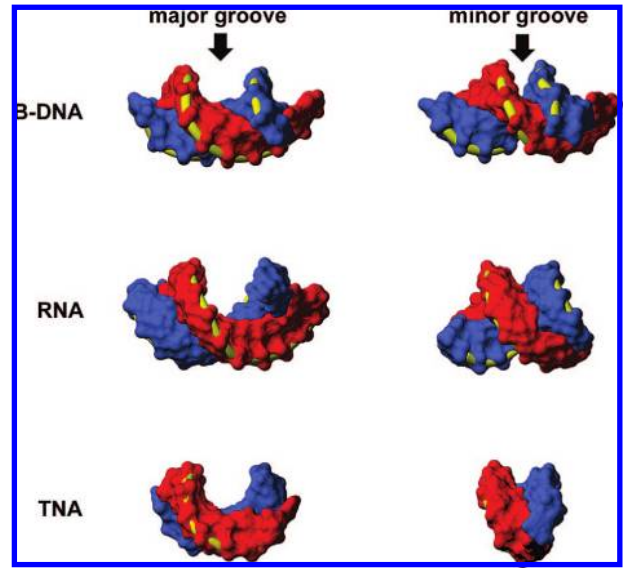


Figure 8. Comparison of groove widths in B-DNA, RNA, and TNA duplexes. pdb accession codes: 1BNA (DNA), 1AL5 (RNA).

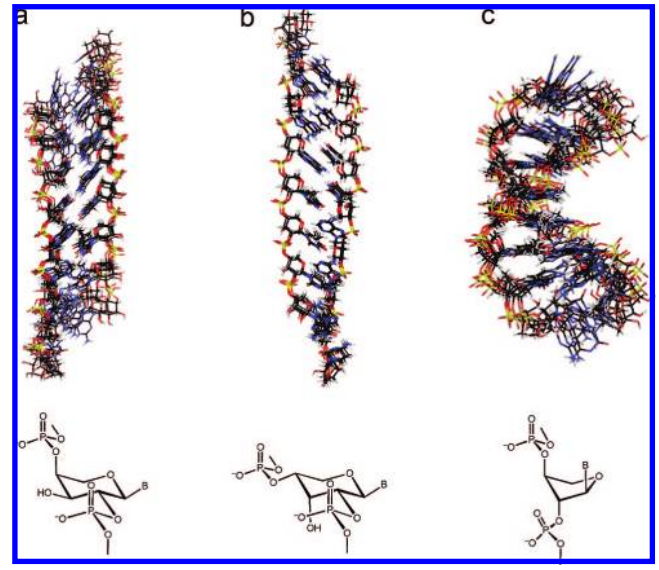


Figure 9. NMR-derived structures of the duplexes formed from the self-complementary sequence CGAATTCG: (a) p-arabino-NA, (b) p-RNA, (c) TNA.

variation of β. More likely, it is the consequence of dynamic averaging that takes place to various extents throughout the sequence.

In the simulated annealing procedure, the experimental values for ^3J(1′,2′) were not used to restrain the conformation of the furanose ring. Nevertheless, the resulting pucker close to 4′-exo matches the measured values of 2 Hz. Because of their different dependence on molecular coordinates, NOEs and coupling constants that result from an ensemble of conformers cannot be satisfied by one average structure. It can therefore be ruled out that this value is the result of dynamic averaging between C3-endo and C2-endo conformations. In the X-ray structures of a B-DNA and an A-DNA duplex, both containing one TNA residue per strand, Egli and co-workers observed a 4′-exo conformation for the TNA residues in both cases.[24,25] Structurally, the all-TNA duplex **1** studied here clearly resembles the A-type helix of RNA as judged from the average values for

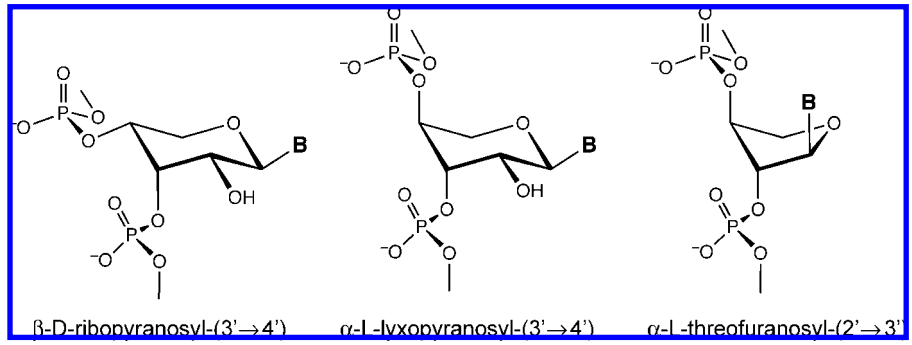


Figure 10. Idealized conformation of the repeating units in (4′-3′)-pRNA, (4′-3′)-lyxopyranosyl NA, and TNA.

slide (−1.3 Å), *x*-displacement (−4.6 Å), and an average $P_i \rightarrow P_{i+1}$ distance of ca. 5.9 Å similar to that shown by A-type DNA duplexes (ca. 6.2 Å).[25] The similarity of TNA and RNA can further be seen in the shallow and wide minor groove that is common to both helices. These findings are consistent with the fact that intersystem cross-pairing between TNA and RNA is stronger than that between TNA and DNA,[1,2] and that RNA is a much better template for ligating TNA-fragments than is DNA.[55]

For TNA, the proton density per residue is even lower than for RNA. With the resulting sparse NOE restraints and using very conservative dihedral angle restraints, it was still possible to obtain NMR structures that are locally and globally well defined. A prerequisite was the use of the complete relaxation matrix for the derivation of distance restraints. This allowed us to narrow the limits for the distance restraints while still avoiding extensive bias due to over-restraining.

B. Implications. Among the nucleic acid alternatives studied thus far in our laboratories with regard to both base-pairing properties and molecular structure, TNA is the only system that shows not only intrasystem Watson−Crick pairing but also intersystem cross-pairing with RNA and DNA. How this difference in pairing capability reflects itself in a corresponding difference between overall intrasystem duplex structures is demonstrated by Figure 9, which depicts the NMR structures of three duplexes derived from the same self-complementary base sequence CGAATTCG but with backbones that differ as indicated.[21,22] The three duplex structures can be viewed as illustrating what is thought to be the major structural factor deciding whether a given system is able to cross-pair with the natural systems. For efficient intra- or intersystem duplex formation, a given backbone has to adapt to a conformation (pairing conformation) in which the planes of adjacent nucleobases can be in contact with each other within the optimal base-staking distance of about 3.3 Å. There are two major types of conformational adaptations by which a backbone, if its repeating monomer unit has excessive length, can fulfill this requirement while trying to avoid the buildup of steric strain: by either helicalization of the backbone or by adopting a high backbone/base-pair axis inclination either as an intrinsic property of the backbone units' constitution,[4,8,21,23,56] or by tuning nucleosidic torsion angles.[4,19] The capability of intersystem cross-pairing demands mutual compatibility of both sense and degree of such adaptations of partner systems. The adaptation modes available to partner systems can be deduced from the pairing conformation they assume in their intrasystem base-pairing. Characteristic of the entire family of pentopyranosyl oligonucleotide systems is a high backbone/base-pair axis inclination resulting in intrasystem duplexes that are weakly helicalized ladder-like structures (Figure 9).[21,22] In sharp contrast, the intrasystem duplex structure of TNA is strongly helical and, in the right-handed (L)-threose series, remarkably similar to the structure type of RNA and A-type DNA. The capability of the TNA backbone to adapt to the B-type as well as the A-type DNA structure has been already discussed in detail by Egli[24,25] in the context of his two X-ray structure determinations of duplexes derived from self-complementary DNA strands in which one of the deoxyribofuranose units in each strand is replaced by a L-threose unit.

Monomer ring-size as such, 5-memberedness in furanosyl- versus 6-memberedness in pyranosyl-oligonucleotide systems, is by no means a valid criterion for the ability of modified oligonucleotide systems to undergo cross-pairing with the natural nucleic acids. As a matter of fact, it was a discovery made in the context of a systematic study of pairing properties in the pento*pyranosyl*-(4′-2′)-oligonucleotide family that had initiated the project of synthesizing (L)-threo*furanosyl*-(3′-2′)-oligonucleotides for the sake of checking their base-pairing properties.[1,2,4,17] Changing the phosphodiester bridge from the (4′-2′)- to the isomeric (4′-3′)-position (and implicitly shortening the backbone monomer unit) in (D)-ribopyranosyl-(4′-2′)-oligonucleotides ("p-RNA") caused, as expected, loss of the intrasystem base-pairing capability; in contrast, the corresponding constitutional change in the (L)-lyxopyranosyl-(4′-2′)-oligonucleotide series led to a (L)-lyxopyranosyl-(4′-3′) system, which, surprisingly, showed not only weak intrasystem pairing but also (albeit strongly sequence dependent) intersystem cross-pairing with complementary DNA strands.[4] The essential difference between the (4′→3′)-bridged structures of the ribopyranosyl- and lyxopyranosyl series lies in the conformation of the phosphodiester linkage to the pyranosyl chair (4′-equatorial/3′-axial in the ribopyranosyl- and 4′-axial/3′-axial in the lyxopyranosyl series), which results in a tendentially smaller backbone/base-pair axes inclination and larger distance between neighboring phosphorus centers for the oligonucleotide backbone of the latter. Irrespective of the difference in ring size, the spine of the TNA backbone is in this respect similar to that of the (L)-lyxopyranosyl-(4′-3′)-oligonucleotide system (Figure 10).

The fact that 6-membered ring size of the sugar monomer unit of oligonucleotides can be perfectly compatible with intersystem cross pairing with DNA and RNA is amply demonstrated by the important work of Herdewijn[57−60] and collaborators on modified oligonucleotides that display a pyranose substitution pattern of nucleobases and phosphodiester linkages different from the pyranose systems studied in our

(55) Wu, X.; Delgado, G.; Krishnamurthy, R.; Eschenmoser, A. *Org. Lett.* **2002**, *4*, 1283–1286.
(56) Micura, R.; Kudick, R.; Pitsch, S.; Eschenmoser, A. *Angew. Chem., Int. Ed.* **1999**, *38*, 680–683.

laboratories. As can be foreseen by analyzing idealized pairing conformations of such modified 6-membered backbone systems, the backbone/base-pair axes inclination in corresponding oligonucleotide duplexes sensitively depends on such substitution patterns in a (qualitatively) transparent way.[4,60]

We do not consider the quasi-diaxial conformation of the phosphodiester substituents at the 2′- and 3′-positions of TNA's furanose rings as being imposed on the threose rings of the oligonucleotide strand by the structural requirements of duplex formation. All six X-ray structure analyses of monomeric (L)-threofuranosyl-nucleoside derivatives prepared and analyzed in the beginning phase of the TNA project had been found to have the free or functionalized hydroxy substituents at the 2′- and 3′-position in the quasi-diaxial conformation.[1,2] This may be regarded as a steric consequence of the trans-arrangement between the nucleobase substituent and the substituent at the 2′-position. In TNA-duplexes, the adoption of this conformation may be assisted by electrostatic repulsion between the negatively charged phosphodiester oxygens of neighboring repeating units. This latter consideration led to the conjecture[4] that even an acyclic analogue of TNA lacking the endocyclic oxygen center as well as the 5′-methylene group might be capable of intrasystem base-pairing and, consequently, of cross-pairing with the natural systems. The conjecture was not followed up by experiment because such an analogue was not considered to represent a "potentially natural" nucleic acid alternative according to the criteria[5] we consistently followed in the selection of nucleic acid alternatives for experimental study. Fortunately, another research group,[61,62] free of such constraints, showed experimentally that an acyclic analogue of TNA ("glycerol nucleic acid", GNA) is indeed an efficient base-pairing system[63] and capable of cross-pairing with the natural nucleic acids.

Acknowledgment. This work was supported by the ETH Zurich and the Skaggs Foundation at the Scripps Research Institute (La Jolla). We thank Stefan Pitsch for his valuable help with the purification of the oligonucleotide.

Supporting Information Available: HPLC and MALDI-TOF characterization of **1** (Figure S1). NMR parameters for 2D spectra (Table S2). NOEs and derived distance restraints between nonexchangeable protons (Table S3) and NOEs and derived distance restraints involving exchangeable protons (Table S4). Statistics of backbone dihedral angles of the accepted structures for the bundle in Figure 6a, calculated using the two-spin approximation (Table S5). This material is available free of charge via the Internet at http://pubs.acs.org.

JA8041959

(57) Hendrix, C.; Rosemeyer, H.; Verheggen, I.; Seela, F.; Van Aerschot, A.; Herdewijn, P. *Chem.-Eur. J.* **1997**, *3*, 110–120.
(58) Herdewijn, P. *Liebigs Ann. Chem.* **1996**, 1337–1348.
(59) Kerremans, L.; Schepers, G.; Rozenski, J.; Busson, R.; Van Aerschot, A.; Herdewijn, P. *Org. Lett.* **2001**, *3*, 4129–4132.
(60) Lescrinier, E.; Froeyen, M.; Herdewijn, P. *Nucleic Acids Res.* **2003**, *31*, 2975–2989.
(61) Zhang, L.; Peritz, A.; Meggers, E. *J. Am. Chem. Soc.* **2005**, *127*, 4174–4175.
(62) Zhang, L.; Peritz, A. E.; Carroll, P. J.; Meggers, E. *Synthesis* **2006**, 645–653.
(63) Recently, the X-ray structure of a GNA octamer duplex has been published: Schlegel, M. K.; Essen, L. O.; Meggers, E. *J. Am. Chem. Soc.* **2008**, *130*, 8158–9.
(64) Markley, J. L.; Bax, A.; Arata, Y.; Hilbers, C. W.; Kaptein, R.; Sykes, B. D.; Wright, P. E.; Wüthrich, K. *Pure Appl. Chem.* **1998**, *70*, 117–142.
(65) Wijmenga, S.; Mooren, M. M. W.; Hilbers, C. W. In *NMR of Macromolecules. A Practical Approach*; Roberts, G. C. K., Ed.; IRL Press: Oxford, 1993; pp 217−288.