J A C S
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Subscriber access provided by ISTANBUL TEKNIK UNIV

Article

# Stabilization by Extra-Helical Thymines of a DNA Duplex with Hoogsteen Base Pairs

Joan Pous, Lourdes Urpi#, Juan A. Subirana, Catherine Gouyette, Jorge Navaza, and J. Lourdes Campos

*J. Am. Chem. Soc.*, **2008**, 130 (21), 6755-6760 • DOI: 10.1021/ja078022+ • Publication Date (Web): 01 May 2008

**Downloaded from http://pubs.acs.org on February 8, 2009**

## More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

ACS Publications
High quality. High impact.

Journal of the American Chemical Society is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

J|A|C|S
ARTICLES
Published on Web 05/01/2008

# Stabilization by Extra-Helical Thymines of a DNA Duplex with Hoogsteen Base Pairs

Joan Pous,[†] Lourdes Urpí,[‡] Juan A. Subirana,[‡,*] Catherine Gouyette,[§] Jorge Navaza,[#] and J. Lourdes Campos*[,‡]

*Departament d'Enginyeria Química, Universitat Politècnica de Catalunya, E-08028 Barcelona, Spain, Plataforma Automatitzada de Cristallografia, PCB-CSIC, E-08028 Barcelona, Spain, Plate-forme de synthèse d'oligonucléotides, Institut Pasteur, 75724 Paris, France, and Institut de Biologie Structurale, 38027 Grenoble, France*

Received October 25, 2007; E-mail: juan.a.subirana@upc.edu; lourdes.campos@upc.edu

***Abstract:*** We present the crystal structure of the DNA duplex formed by d(ATATATCT). The crystals contain seven stacked antiparallel duplexes in the asymmetric unit with A•T Hoogsteen base pairs. The terminal CT sequences bend over so that the thymines enter the minor groove and form a hydrogen bond with thymine 2 of the complementary strand in the Hoogsteen duplex. Cytosines occupy extra-helical positions; they contribute to the crystal lattice through various kinds of interactions, including a unique CAA triplet. The presence of thymine in the minor groove apparently contributes to the stability of the DNA duplex in the Hoogsteen conformation. These observations open the way toward finding under what conditions the Hoogsteen duplex may be stabilized in vivo. The present crystal structure also confirms the tendency of A•T-rich oligonucleotides to crystallize as long helical stacks of duplexes.

## Introduction

Some time ago, we reported the structure of $[d(ATATAT)]_2$ duplexes in two crystal forms.[1,2] The oligonucleotides form antiparallel duplexes with Hoogsteen base pairs instead of the standard Watson−Crick base pairs. A peculiar feature of the reported structures was the presence of some extra-helical thymines which entered into the minor groove of a neighbor Hoogsteen duplex. It was not clear whether such behavior was a crystallographic curiosity or an essential feature in order to stabilize Hoogsteen pairing. Therefore, we designed the oligonucleotide sequence d(ATATATCT), which might form an $(AT)_3$ duplex with CT free tails at both ends, with the expectation that they would fold back and introduce the terminal thymines into the minor groove. This is indeed what we found. A scheme of the structure is presented in Figure 1. An intermediate cytosine was added in order to provide flexibility to the terminal tail of the oligonucleotide. It is known that cytosine does not favor the B-form of DNA.[3] In fact, it is frequently found that terminal cytosines are disordered and not visible in oligonucleotide crystals (for example in ref 4). However, in our case, most of the cytosines are visible in the electron density map; they contribute to establish the crystal lattice. The TCTC sequence has also been used as a flexible turn region in hairpin duplexes.[5]

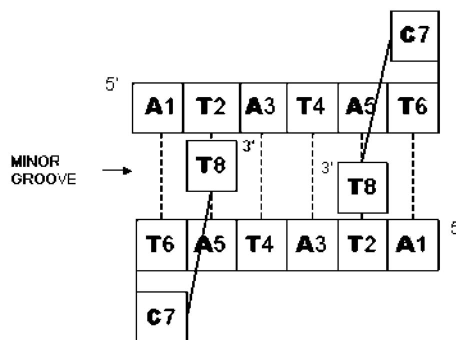


***Figure 1.*** Numbering scheme of duplexes. Cytosines are extra-helical, whereas most terminal thymines fold back and enter the minor groove. They form a hydrogen bond with thymine 2 in the other strand of the duplex.

We should stress, as previously discussed,[1,2,6] that AT-rich sequences are very abundant in noncoding regions of the genome.[7] The idea that such sequences are "junk" DNA is presently being abandoned. It is clear that they play an important role in genome function. For example, it has been shown[8] that $(AT)_n$ sequences are involved in recurrent chromosomal translocations, which are related with many genetic disorders. Isolated

† PCB-CSIC.
‡ Universitat Politècnica de Catalunya.
§ Institut Pasteur.
# Institut de Biologie Structurale.

(1) Abrescia, N. G. A.; Thompson, A.; Huynh-Dinh, T.; Subirana, J. A. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2806–2811.
(2) Abrescia, N. G. A.; González, C.; Gouyette, C.; Subirana, J. A. *Biochemistry* **2004**, *43*, 4092–4100.
(3) Foloppe, N; MacKerell, A. D., Jr. *J. Phys. Chem. B* **1999**, *103*, 10955–10964.
(4) Tereshko, V.; Urpí, L.; Malinina, L.; Huynh-Dinh, T.; Subirana, J. A. *Biochemistry* **1996**, *35*, 11589–11595.
(5) Freyer, M. W.; Buscaglia, R.; Cashman, D.; Hyslop, S.; Wilson, W. D.; Chaires, J. B.; Lewis, E. A. *Biophys. Chem.* **2006**, *126*, 186–196.
(6) Campos, L.; Valls, N.; Urpí, L.; Gouyette, C.; Sanmartín, T.; Richter, M.; Alechaga, E.; Santaolalla, A.; Baldini, R.; Creixell, M.; Ciurans, R.; Skokan, P.; Pous, J.; Subirana, J. A. *Biophys. J.* **2006**, *91*, 892–903.
(7) Lander, E. S. *Nature* **2001**, *409*, 860–921.
(8) Gotter, A. L.; Nimmakayalu, M. A.; Jalali, G. R.; Hacker, A. M.; Vorstman, J.; Duffy, D. C.; Medne, L.; Emanuel, B. S. *Genome Res.* **2007**, *17*, 470–481.

10.1021/ja078022+ CCC: $40.75 © 2008 American Chemical Society

Hoogsteen base pairs have been found in many situations, as reviewed elsewhere.[2] A particular evidence for their use in vivo is provided by human polymerase-ι, which replicates DNA by Hoogsteen base pairing.[9,10] However, it has not been established under what conditions DNA may adopt such conformation in AT-rich regions of the genome.

Understanding how the Hoogsteen duplex structure can be stabilized is crucial to determine its eventual role in a biological environment. Unfortunately, there is no assay presently available in order to determine whether Hoogsteen duplexes are found in vivo. No protein/DNA complex has been crystallized with all-AT sequences. On the other hand, the abundance of such sequences in eukaryotic genomes suggests that Hoogsteen duplexes might be present in the cell under some circumstances. The results presented here open the way to develop thymine analogues which might either recognize or induce the formation of Hoogsteen duplexes. Any drug/protein able to change the conformation of an AT sequence from Watson−Crick to Hoogsteen would certainly have a strong biological effect.

## Experimental Procedures

**Synthesis and Crystallization.** The deoxyoctanucleotide d(ApTpApTpApTpCpT) was synthesized on an automatic synthesizer by the phosphoramidite method and purified by gel filtration and reverse phase HPLC. The ammonium salt of this octamer was prepared by ion exchange chromatography. The oligonucleotide was crystallized at 4 °C using the hanging-drop vapor diffusion method. The best crystals were obtained from one drop containing 1.0 mM single-strand octanucleotide, 22 mM sodium cacodylate buffer pH 6.5, 10 mM magnesium chloride, 1.5 mM spermine tetrachloride, and 10% MPD, which was equilibrated against a reservoir with 25% MPD. The MPD concentration of the reservoir was gradually increased up to 45%, at which conditions crystals suitable for diffraction were collected.

**Data Collection and Analysis.** Crystals were flash-cooled in liquid nitrogen at 113 K. Diffraction data were collected up to 3.0 Å resolution at the ESRF beamline BM16 on a MAR CCD detector at $\lambda = 0.9794$ Å. Data were integrated and scaled with the Denzo and Scalepack programs.[11] Because of a high level of mosaicism, the higher resolution data (3.0−3.08 Å) were rejected, as they showed poor statistics. Further details are given in Table 1.

**Structure Determination.** The structure was determined by molecular replacement with the AMoRe program.[12] We used two models of the hexamer d(ATATAT): the standard B form constructed with the TURBO program[13] and the duplex d(ATATAT) already solved with Hoogsteen base pairs (ref 2, entry UD0049 of the NDB). We did not introduce at this stage the CT part of the sequence. The routine procedure for solving the structure did not succeed with any of both models. Since the correlation factors were significantly better with the Hoogsteen model, we used it in the following calculations. Then, by combination of the self-rotation and translation functions, three duplexes that form a tube could be placed in the asymmetric unit. The progress of the molecular replacement search is given in detail as Supporting Information. Seven duplexes were finally located in the asymmetric unit. They form a tube of stacked duplexes.

**Refinement.** The model from molecular replacement with seven hexamer duplexes containing only the AT portion of the structure was refined using REFMAC5 version 5.3.0022.[14] This version adds some new commands for distance and NCS restraints. The resolution range spanned all the observed reflections, from 24 to 3.1 Å. Although the last shell of resolution (Table 1) has poor statistics, it was found to be essential to use these data since they contain information about the stacking and orientation of the bases. Initial maps showed very clearly that the double helical parts present a Hoogsteen base pairing. The Watson−Crick model was tested again at this stage but resulted in an unrefinable structure. Extra density was found at the 3′-end of the helical ATATAT sequences, which corresponds to the CT terminal part. All of the sugars were set to the C2′-endo conformation after manual testing,[15] but during refinement, they tend to move to the C1′-exo region and beyond. The nonhelical parts (terminal CTs) were constructed after the refinement of the helical ATs was firmly settled. The length of the Hoogsteen-related distances, taken from the original Hoogsteen structure,[16] was restrained during refinement. The $R$ factors and overall geometry of the structure improved when such restraints were used. NCS (noncrystallographic symmetry) restraints were also applied between all of the seven duplexes.

**Table 1.** Crystal Data and Refinement Statistics

| | |
|---|---|
| wavelength (Å) | 0.9794 |
| temperature (K) | 113 |
| space group | $C2$ |
| cell dimensions | $a = 120.836$ Å, $b = 39.540$ Å, $c = 71.753$ Å, $\beta = 95.08°$ |
| resolution range (last shell) (Å) | 24−3.08 (3.16−3.08) |
| unique reflections (last shell) | 5188 (251) |
| free $R$ factor reflections (last shell) | 555 (31) |
| completeness (last shell) (%) | 88.9 (58.0) |
| redundancy factor | 2.96 |
| $I/\sigma I$ (last shell) | 13.5 (3.6) |
| $R_{merge}$ (last shell) | 0.077 (0.222) |
| asymmetric unit contents | 7 DNA duplexes, 40 $H_2O$ Total: 2221 non-hydrogen atoms |
| $R_{work}$ [a] | 0.21 (0.26) |
| $R_{free}$ [b] | 0.29 (0.49) |
| mean $B$ factor | 65.9 Å$^2$ |
| rms | bonds 0.012 Å, angles 1.2° |

[a] $\sum_{hkl}|F_o(hkl) - KF_c(hkl)|/\sum_{hkl}F_o(hkl)$. [b] $R$ factor of reflections used for cross validation in the refinement.

At the expected region of the minor grooves, a clear electron density blob was apparent. Additional electron density was visible in the external regions of the duplex. We placed the extra-helical cytosines in the latter regions and thymines in the minor groove. In chain F, no electron density was visible in the minor groove, so that no thymine was placed in that case. The position of extra-helical thymines in the minor groove was restricted by imposing a hydrogen bond between O2 in thymine 2 and N3 in thymine 8 of the complementary strand. In one case (chain I), thymine could not be placed correctly in the minor groove. It was then substituted by cytosine which formed a correct hydrogen bond between its N4 atom and O2 in thymine 2 of the complementary strand, with an adequate electron density map. In all other cases, the extra-helical thymines formed correct hydrogen bonds, so it is unlikely that they could be replaced by cytosines.

Placing cytosines was not straightforward. Only the cytosine in chain H was very clear in early electron density maps. This cytosine forms a CAA triplet to be described below. The rest of cytosines appeared as regions with clear electron density but ill-defined shape. In fact, it is likely that cytosines have multiple conformations. We chose an optimal position, which was adjusted in such a way that the electron density map and $R$ factors improved.

(9) Nair, D. T.; Johnson, R. E.; Prakash, S.; Prakash, L.; Aggarwal, A. K. *Nature* **2004**, *430*, 377–380.
(10) Ghosal, G.; Muniyappa, K. *Biochem. Biophys. Res. Commun.* **2006**, *343*, 1–7.
(11) Otwinowsky, Z.; Minor, W. *Methods Enzymol.* **1997**, *276*, 307–326.
(12) Navaza, J *Acta Crystallogr.* **1994**, *D50*, 157–163.
(13) Roussel, A. Legaigneur, P. Inisan, A. G.; Cambillau, C. www.afmb.univ-mrs.fr/-TURBO.
(14) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr.* **1997**, *D53*, 240–255.
(15) Parkinson, G. P.; Vojtechovsky, J.; Clowney, L.; Brunger, A. T.; Berman, H. M. *Acta Crystallogr.* **1996**, *D52*, 57–64.
(16) Hoogsteen, K. *Acta Crystallogr.* **1959**, *12*, 822–823.

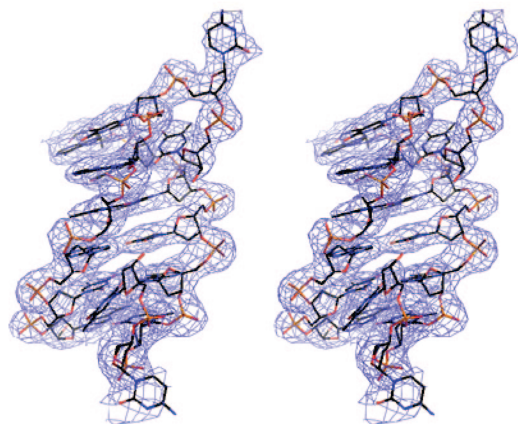

***Figure 2.*** Stereoview of the $2F_o - F_c$ electron density map (at $1\sigma$ level) of the GH duplex. The electron density map of the other duplexes is similar. The electron density of the terminal cytosine of chain G (bottom end) is less well-defined than the cytosine of chain H (top end). The latter forms a CAA triplet as it is shown in Figure 7.

Manual cycles of refinement were carried out using TURBO,[13] Pymol,[17] and Coot.[18] Maps were computed using FFT and Mapmask from the CCP4[19] suite. In the final stage, 40 water molecules were added manually and only the ones having hydrogen bonds and adequate density were kept in the structure. TLS[20] refinement as it is present on REFMAC5 was added in the last cycle of refinement, starting at a fixed set of $B$ factors of 40 Å$^2$ and only one rigid-body composed of the full structure.

A stereoview of the resulting map clearly shows the base pairs and the phosphodiester chain (Figure 2). An omit map is presented in the Supporting Information. The maps are somewhat noisy due to the limited resolution of our data. Some blobs of density were left empty. We attribute them to double conformations of low occupancy in the structure and to disordered counterions. The resulting structure contains the full sequence of the seven hexamers except thymidine 8 in chains I and F, and cytosine 7 in chain M, for which no clear electron density was found. See Table 1 for refinement statistics. Oligonucleotide structural parameters have been calculated with 3DNA[21] from the positions of the C1′ atoms. Drawings were made with the CERIUS2 (Accelrys Inc.) and Pymol[17] programs. Final coordinates have been deposited in the Nucleic Acid Database (NDB code UD0077).

## Results

**Packing Features.** The oligonucleotide duplexes pack end-to-end and form helical columns of infinite length in the crystal. A similar organization is found in many all-AT oligonucleotides crystallized in the B form.[6] A stereoview of the asymmetric unit is shown in Figure 3. Each duplex is rotated with respect to its neighbors in the column, so that a pseudocontinuous helical structure with seven duplexes in four turns is generated. The average rotation between neighbor duplexes is thus $360° \times 4/7 = 205.7°$, and the average base-step twist angle is $34.3° = 205.7°/6$. End-to-end stacking between neighbor duplexes is mainly due to the adenine bases. As a result, the axes of individual duplexes are appreciably displaced from the overall axis of the helix formed by the seven duplexes in the asymmetric unit, as it is clearly apparent in Figure 3.

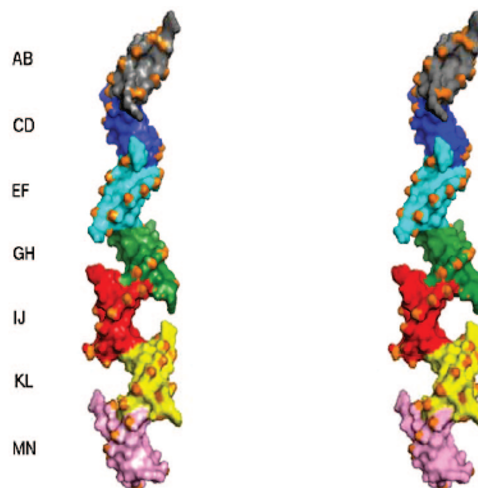


***Figure 3.*** Stereoview of one column of seven duplexes (asymmetric unit) in the crystal structure. Each independent duplex is shown in a different color. The names of the chains in the PDB file are also indicated. Phosphate groups are represented as brown spheres. The duplexes are organized in an approximate helical fashion, with seven duplexes in the helical repeat. The average rotation of each duplex with respect to its neighbor in the column is 205.7°. Neighbor duplexes are adequately stacked, but the presence of the extra-helical cytosines gives the wrong impression that they are displaced in a lateral direction.

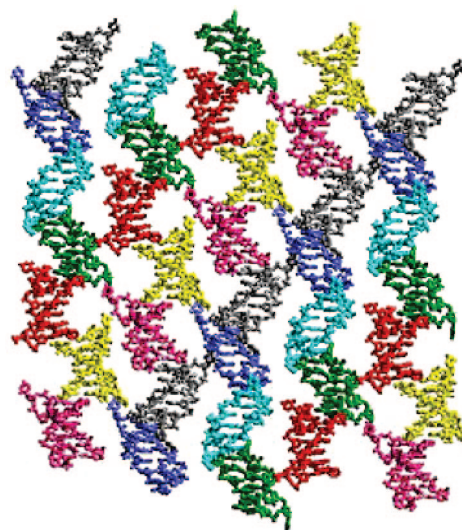

***Figure 4.*** Organization of one layer of parallel columns of duplexes in the crystal structure. Crystallographically independent duplexes are indicated with the same colors as in Figure 3. Neighbor layers above and below the one shown are identical, but the orientation of their columns forms a 95° angle with those shown in the figure. A schematic view is presented in Figure 5.

The duplex columns shown in Figure 3 are placed side by side in planes, as it is represented in Figure 4. In the crystal, such planes are stacked in such a way that neighbor planes cross in space at an angle of 95°. Thus the duplexes shown in Figure 4 are sandwiched between identical sets of columns, but with a different orientation, as it is schematically shown in Figure 5.

The crystal structure appears to be stabilized mainly by interactions of the extra-helical cytosines. No clear counterions or spermine molecules could be localized, although they certainly play a role in stabilizing the structure by neutralizing the phosphate groups. Most cytosines are involved in interactions with neighbor duplexes, such as cytosine−cytosine stacking or hydrogen bonding, as shown in Figures 6 and 7. Exceptions

(17) DeLano, W. L. *Pymol*; DeLano Scientific: Palo Alto, CA, 2002.
(18) Emsley, P.; Cowtan, K *Acta Crystallogr.* **2004**, *D60*, 2126–2132.
(19) The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr.* **1994**, *D50*, 760−763.
(20) Winn, M.; Isupov, M.; Murshudov, G. N. *Acta Crystallogr.* **2001**, *D57*, 122–133.
(21) Lu, X. J.; Olson, W. K. *Nucleic Acids Res.* **2003**, *31*, 5108–5121.

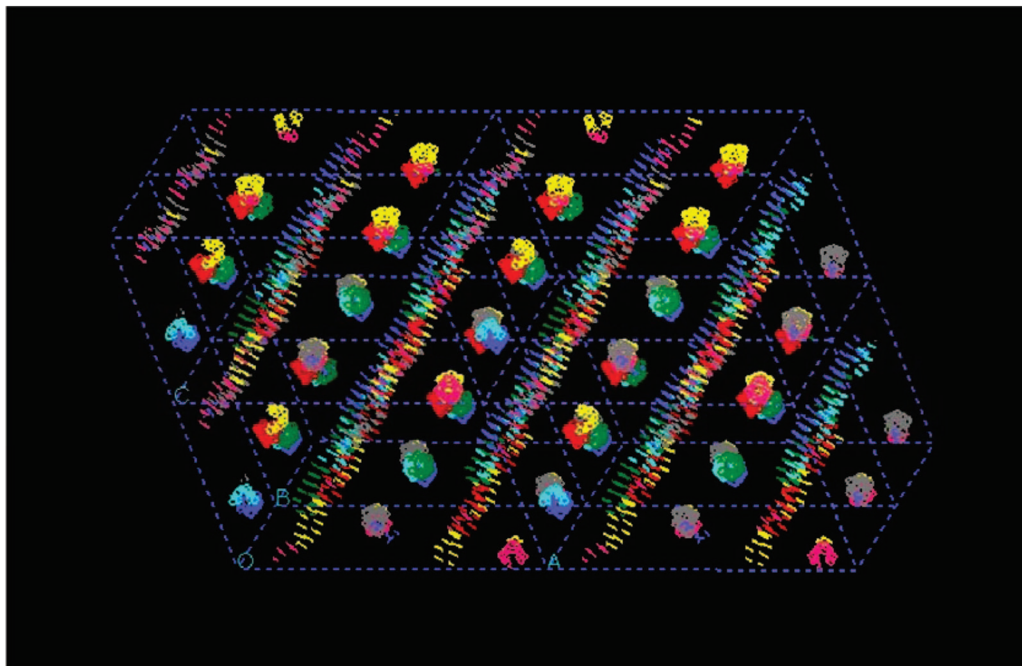


***Figure 5.*** View of eight unit cells of the crystal structure. For clarity, only the Hoogsteen AT bases are shown, the phosphodiester chain has also been omitted. The figure demonstrates how neighbor layers of duplex columns are organized in approximate perpendicular directions. Individual duplexes are color coded as in Figures 3 and 4.

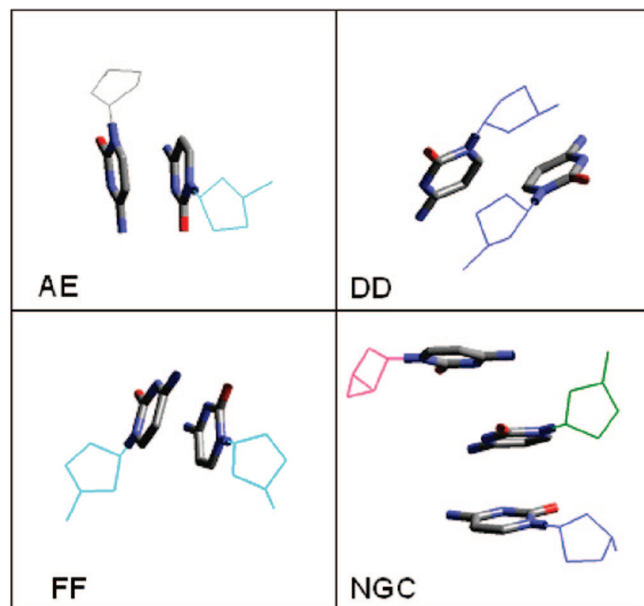


***Figure 6.*** Stacking interactions between extra-helical cytosines in the crystal structure. The letters correspond to the chain to which each cytosine belongs. The attached sugar rings have been color coded as in Figures 3 and 4. In the case of DD, stacking is only partial.

are the cytosines in chains J, L, and M. Cytosines J (red) and L (yellow) lie in an empty space in the structure, as it is apparent in Figure 4. Cytosine J is close to the AE stack (Figure 6) but does not appear to interact with them. Cytosines B and K are close to phosphate groups, with which they might have an electrostatic interaction through the exocyclic N4 amino group. No electron density was apparent for the cytosine base in the M oligonucleotide, and it appears to be disordered.

A striking case is the cytosine in chain H, which forms a triplet structure with two terminal adenines from the IJ duplex. If we take into account the thymine in the minor groove, a complex interaction involving six bases is found, as shown in Figure 7. The triplet interaction is unique for Hoogsteen duplexes since it involves the N1 atom of one of the adenines. It should be noted that this cytosine appeared very clearly in the electron density map at the early stages of refinement, which indicates that this is a very stable feature in the crystal structure. A similar triplet structure of an extra-helical cytosine with two adenines has also been reported in a normal B form duplex.[22] However, in that case, the N7 of adenine is involved instead of N1.

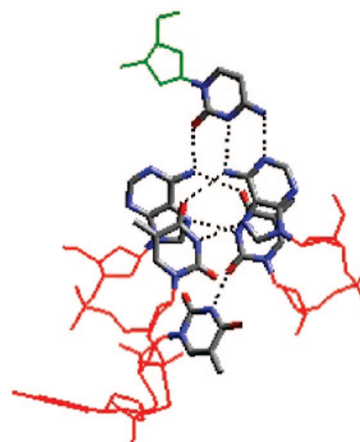

***Figure 7.*** Triplet interaction between the H cytosine (at the top) and the major groove of the IJ duplex. Two base pairs of the latter duplex are shown. Cytosine interacts through hydrogen bonding with the N6 atom of adenine 5 in chain J and with the N1 and N6 atoms of adenine 1 in chain I. The Hoogsteen hydrogen bonds are also indicated, as well as the one formed by thymine 8 (at the bottom) in the minor groove.

All duplexes in the structure are involved in the interactions depicted in Figures 6 and 7. Some of them occur between

(22) Abrescia, N. G. A.; Malinina, L.; Fernandez, L. G.; Huynh-Dinh, T.; Neidle, S.; Subirana, J. A. *Nucleic Acids Res.* **1999**, *27*, 1593–1599.

***Table 2.*** Geometry of the Hoogsteen Base Pairs[a]

| | |
|---|---|
| hydrogen bond N7(A)−N3(T) | 2.72 (0.16) |
| hydrogen bond N6(A)−O4(T) | 3.01 (0.30) |
| Hydrogen bond O2(T2)−N3 (T8′, extra-helical) | 2.88 (0.26) |
| C1′−C1′ distances | 8.15 (0.24) |
| adenine glycosidic angle (χ) | 46.9° (18.6) |
| thymine glycosidic angle (χ) | −122.8° (13.8) |

[a] Distances are given in angstroms. Standard deviation is in parentheses.

***Table 3.*** Average Twist and Rise of the Seven Independent Hoogsteen Duplexes

| step | twist (°)[a] | rise (Å)[a] |
|---|---|---|
| A1T2 | 33.5 (1.1) | 3.35 (0.32) |
| T2A3 | 35.8 (2.0) | 3.39 (0.10) |
| A3T4 | 28.0 (2.3) | 3.36 (0.21) |
| T4A5 | 35.5 (2.1) | 3.41 (0.20) |
| A5T6 | 32.5 (2.2) | 3.30 (0.39) |
| T6A1[b] | 40.4 (2.2) | 2.28 (0.56) |

[a] Standard deviations are given in parentheses. [b] These values correspond to the virtual base step between two neighbor stacked duplexes. They deviate significantly from the rest.

duplexes in the plane of the columns represented in Figure 4, but most of them involve duplexes from the planes above and below. Thus the crystal structure appears to be stabilized by a complex set of cytosine interactions with different spatial orientations.

**Structural Features of the Hoogsteen Duplex.** The main property of Hoogsteen base pairs is the different pattern of hydrogen bonds in the adenine base, when compared with the usual Watson−Crick base pairs. In the Hoogsteen case, the N7 atom of adenine, instead of N1, is paired with the N3 atom of thymine. Examples are given in Figure 7. For such interaction to take place, the adenine base has to rotate about 180° around the glycosidic bond, so that it is in the *syn* conformation instead of the usual *anti* conformation. At the same time, the C1′−C1′ distance becomes much shorter. A summary of the geometrical features of the Hoogsteen base pairs is given in Table 2. The values found are similar to those previously reported for Hoogsteen duplexes at higher resolution.[1,2]

Hoogsteen base pairs are organized as antiparallel helical duplexes with an overall shape quite similar to the standard B form of DNA. However, the minor groove is now much narrower and only contains a single hydrogen bond acceptor atom in each base pair: the O2 atom of thymine. The N3 atom of adenine has now moved to the major groove side, as discussed in detail elsewhere.[2] The helical parameters are given in Table 3. A moderate alternation of twist is found, which is larger in the TA steps when compared with AT steps. It is interesting to note that such alternation is also found in the B form of DNA,[22] in spite of the different geometry of the individual base pairs. However, in the B form, the alternation is much more noticeable.

A prominent feature of the structure is the presence of extra-helical thymines inside the minor groove, as shown in Figure 8. In the present structure, the thymine bases found in the minor groove belong to the same duplex, whereas, in the previous structures,[1,2] they come from neighbor duplexes in the crystal. They are stabilized by van der Waals interactions and one hydrogen bond between the N3 atom of the extra-helical thymine and the O2 atom of thymine 2 of the other strand in the helical section. In one case (chain I), thymine was found to be replaced by cytosine, as it is also shown in Figure 8. Cytosine N4 is now involved in hydrogen bonding. In a previous structure,[1] it was found that adenine could also occur at this position. It

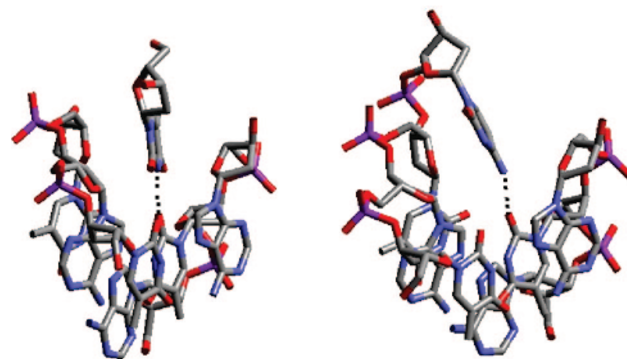

***Figure 8.*** Close-up view of the looped-out bases found in the minor groove. Thymine is shown at the left and cytosine from chain I at the right. In both cases, a nitrogen atom of the external base forms a hydrogen bond (dashed line) with the O2 atom of thymine 2 from the complementary strand of the duplex.

appears that the narrow minor groove of the Hoogsteen duplex is very appropriate to locate an external base, preferably thymine, but cytosine and adenine may also be found in the minor groove.

## Discussion

The discovery by Hoogsteen[16] that adenine and thymine could pair in a way different from that postulated for DNA in the B form casted some doubt on the correctness of the standard model of DNA. However, since then, many structures of oligonucleotides and oligonucleotide−protein complexes have been determined, and it is clear that DNA in general has the standard base pairs postulated by Watson and Crick. On the other hand, the availability of whole genome sequences has shown that AT sequences are very abundant, in particular, in noncoding regions of the genome.[7] Therefore, we decided to study the structure of oligonucleotide sequences which only contain AT base pairs. We found that in general they have the standard B form of DNA,[6] except in the case of alternating AT sequences which may form antiparallel duplexes with Hoogsteen base pairs.[1,2,23] However, in solution, it appears that such sequences tend to adopt the standard B form of DNA,[2] although theoretical calculations show that both forms have similar energies.[24]

Our previous studies[1,2] showed that the Hoogsteen duplexes frequently crystallized with a base (usually thymine) inside the narrow minor groove. The results we present here demonstrate that the presence of thymine in the minor groove is a preferred feature of the Hoogsteen duplex (Figure 8). It appears that the O2 atom of thymine in the minor groove is a strong hydrogen bond acceptor, which may be recognized by proteins which have hydrogen bond donor groups. Our results also suggest that synthetic thymine-like probes could be developed in order to recognize the Hoogsteen conformation.

As mentioned above, it appears that the Hoogsteen conformation is favored in alternating AT sequences. It is not found in most of the all-AT oligonucleotides in the different sequences which we have studied.[6] On the other hand, it has been found in some drug−DNA complexes which contain a TA step.[25] A possible explanation is that the AA/TT base step in the

(23) De Luchi, D.; Tereshko, V.; Gouyette, C.; Subirana, J. A. *Chem. Bio. Chem.* **2006**, *7*, 585–587.
(24) Cubero, E.; Abrescia, N. G. A.; Subirana, J. A.; Luque, F. J.; Orozco, M. *J. Am. Chem. Soc.* **2003**, *125*, 14603–14612.
(25) Ughetto, G.; Wang, A. H.-J.; Quigley, G. J.; van der Marel, G. A.; van Boom, J. H.; Rich, A. *Nucleic Acids Res.* **1985**, *13*, 2305–2323.

Hoogsteen conformation has unfavorable stacking features. If we artificially construct such a step (Supporting Information) inside a Hoogsteen duplex, it turns out that the neighbor thymines stack one on top of the other with close contacts between the methyl groups, which may prevent the formation of the Hoogsteen duplex.

The numerous interactions of cytosines in the crystal lattice (Figures 6 and 7) are an unexpected feature of our results. It appears that extra-helical cytosines may establish well-defined interactions between neighbor DNA duplexes. This is a question which deserves further studies.

## Conclusion

Oligonucleotides with an alternating AT sequence have been found to form an antiparallel double helix in which the standard Watson−Crick base pairing is substituted by Hoogsteen base pairing. Both structures are quite similar, with approximately 10 base pairs in a helical turn. This work demonstrates that Hoogsteen duplexes have a strong tendency to incorporate extra-helical bases in the minor groove, in particular, thymine. This observation opens the way to develop thymine-like drugs which may recognize/induce the DNA Hoogsteen conformation in vivo.

**Acknowledgment.** We thank Ms. T. Sanmartin for her valuable collaboration in the crystallization trials. We also thank the staff of the BM16 line at the European Synchrotron Radiation Facility. This work has been supported by Grant BFU2006-04035 from the Spanish Ministerio de Educación y Ciencia.

**Supporting Information Available:** Structure determination details. One omit map. Model of AAA/TTT Hoogsteen duplex. This material is available free of charge via the Internet at http://pubs.acs.org.

JA078022+