Insights into the DNA stabilizing contributions of a bicyclic cytosine analogue: crystal structures of DNA duplexes containing 7,8-dihydropyrido [2,3-*d*]pyrimidin-2-one

Ella Czarina Magat Juan[1], Satoru Shimizu[1], Xiao Ma[1], Taizo Kurose[1], Tsuyoshi Haraguchi[1], Fang Zhang[2], Masaru Tsunoda[2], Akihiro Ohkubo[1], Mitsuo Sekine[1], Takayuki Shibata[3], Christopher L. Millington[4], David M. Williams[4] and Akio Takénaka[1,2,*]

[1]Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Yokohama 226-8501, [2]Faculty of Pharmacy, Iwaki-Meisei University, Iwaki 970-8551, [3]Graduate School of Biomedical Sciences, Nagasaki University, Nagasaki 852-8523, Japan and [4]Centre for Chemical Biology, Krebs Institute, University of Sheffield, Sheffield S3 7HF, UK

ABSTRACT

The incorporation of the bicyclic cytosine analogue 7,8-dihydropyrido[2,3-*d*]pyrimidin-2-one (X) into DNA duplexes results in a significant enhancement of their stability (3–4 K per modification). To establish the effects of X on the local hydrogen-bonding and base stacking interactions and the overall DNA conformation, and to obtain insights into the correlation between the structure and stability of X-containing DNA duplexes, the crystal structures of [d(CGCGAATT-X-GCG)]$_2$ and [d(CGCGAAT-X-CGCG)]$_2$ have been determined at 1.9–2.9 Å resolutions. In all of the structures, the analogue X base pairs with the purine bases on the opposite strands through Watson–Crick and/or wobble type hydrogen bonds. The additional ring of the X base is stacked on the thymine bases at the 5′-side and overall exhibits greatly enhanced stacking interactions suggesting that this is a major contribution to duplex stabilization.

INTRODUCTION

Chemically modified nucleic acids are being evaluated for use in many applications in biotechnology such as probes or primers (1), in DNA microarrays (2–5) and as therapeutic agents (6–10). One important consideration to be made in analysing the potential use of modified nucleic acids is whether they are able to increase the duplex stability when they hybridize to the defined target sequences on DNA or RNA. For this purpose, we have been exploring effective modifications (11) and found that incorporation of the bicyclic cytosine analogue 7,8-dihydropyrido[2,3-*d*]pyrimidin-2-one (X, Figure 1a), in which the N4 and C5 atoms are linked via a *cis*-propenyl unit, results in an increase in the melting temperature (T_m) of DNA duplexes of 3–4 K per modification compared to the unmodified duplexes (12). A DNA duplex containing an X:A mismatch also showed a 3 K increase in T_m relative to the unmodified mismatched duplex, but the stability of this duplex was significantly less than that containing a T:A base pair. These results suggest that X stabilizes DNA duplexes by forming base pairs with either guanine or adenine, but that the analogue shows a notable preference for pairing with guanine.

In order to examine the interaction of the modified base X, shown in both tautomeric forms in Figure 1a, we performed X-ray analyses on DNA duplexes containing this analogue. We chose to introduce X into the self-complementary Dickerson–Drew dodecamer sequence, as shown in Figure 1b, since this sequence is easily crystallized. The X base is paired with guanine or
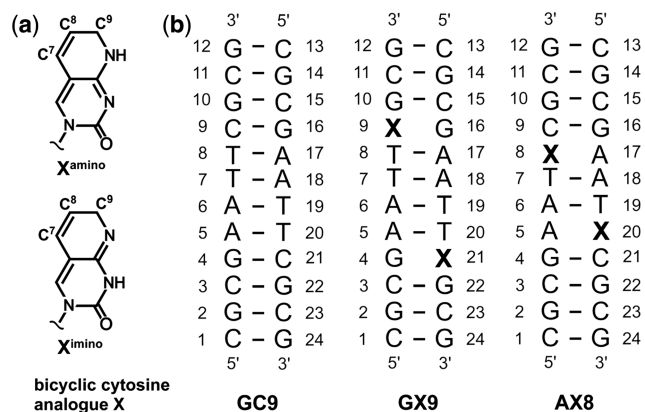
Figure 1. Chemical structure of X (**a**) and the sequences and numbering schemes of the unmodified and X-containing DNA duplexes (**b**). When the two strands are crystallographically identical, the residue numbers 13–24 are replaced by 1–12 with a symbolic mark, e.g. X21 = X9* and G16 = G4*.

adenine at two sites in the GX9 and AX8 duplexes, respectively. Two different GX9 crystals and one AX8 crystal were obtained under different conditions. Their crystal structures have been determined at resolutions ranging from 1.9 to 2.9 Å. In this article, we describe the structures of the base pairs formed between X and G and between X and A. The stacking interactions of X with bases above and below it will be discussed to explain the stabilization of duplexes upon introduction of this bicyclic cytosine analogue.

MATERIALS AND METHODS

Oligodeoxyribonucleotide synthesis

The phosphoramidite of base analogue X was prepared as described (12) and oligodeoxyribonucleotides (ODNs) GX9 and AX8 (Figure 1b) were synthesized on an Applied Biosystems 394 using base-labile phosphoramidites and solid support (Tac-dA, Tac-dG, Tac-dC and dT from ProOligo), as described earlier (12). The products were deprotected with concentrated aqueous ammonia solution at room temperature overnight, and purified by reversed phase HPLC (ODS C18, 300×4.6 mm^2; Altech) using a flow rate of 1 ml/min with a 5–50% CH_3CN gradient in 0.1 M triethylammonium acetate (pH 7.0) over 30 min. The purified samples were then de-tritylated at room temperature for 1 h using 20% AcOH, and re-purified by reversed phase HPLC as above. The oligonucleotides were characterized using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry.

Crystallization and data collection

Prior to crystallization, the oligonucleotides were electrophoresed on 20% poly-acrylamide gels containing 8 M urea, eluted from excised gel slices, and then purified by ion exchange column chromatography (Toyopearl SP-650C). Initial screenings of crystallization conditions were performed using the hanging drop vapor diffusion

method, equilibrating 2 µl droplets against 1 ml of the reservoir solution. The optimized conditions for growing the two different crystals of GX9 (GX9^2 and GX9^3) and the AX8 crystal were as follows. For GX9^2, a droplet of 20 mM sodium cacodylate buffer solution (pH 7.0) containing 0.4 mM DNA, 50 mM sodium chloride, 6 mM spermine tetrahydrochloride and 5% (v/v) 2-methyl-2,4-pentanediol (MPD) was equilibrated against 35% (v/v) MPD at 277 K. For GX9^3, a droplet of 20 mM sodium cacodylate buffer solution (pH 7.0) containing 0.6 mM DNA, 40 mM potassium chloride, 6 mM spermine tetrahydrochloride, 0.2% 3-[(3-cholamidopropyl)-dimethylammonio]-2-hydroxy-1-propanesulfonate, 0.3 mM 4′,6-diamidino-2-phenylindole (DAPI, Supplementary Figure S1a) and 5% (v/v) MPD was equilibrated against 30% (v/v) MPD at 277 K. For AX8, a droplet of 20 mM sodium cacodylate buffer solution (pH 7.0) containing 0.6 mM DNA, 6 mM sodium chloride, 40 mM potassium chloride, 6 mM spermine tetrahydrochloride, 0.2% 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), 0.3 mM 2′-(4-hydroxyphenyl)-6-(4-methyl-1-piperazinyl)-2,6′-bi-*1H*-benzimidazole (Hoechst 33258, Supplementary Figure S1b) and 5% (v/v) MPD was equilibrated against 45% (v/v) MPD at 277 K.

Crystals suitable for X-ray data collections were picked up from their droplets with a nylon loop (Hampton Research) and transferred into liquid nitrogen (100 K). X-Ray experiments for the GX9^2, GX9^3 and AX8 crystals were performed with synchrotron radiation at BL17a, AR-NW12a and BL6a, respectively, of the Photon Factory in Tsukuba ($\lambda = 1.00$ Å). Diffraction patterns with 1° oscillation (a total 180 frames for GX9^3, and 360 frames for GX9^2 and for AX8) were collected. A second data set was taken for each crystal by changing exposure time to compensate overloaded reflections. The diffraction patterns of the three crystals were processed subsequently using the program 'HKL2000' (13). The crystal data and the statistics of data collection are summarized in Table 1.

Structure determination and refinement

Initial phases were derived by molecular replacement with the program 'AMoRe' (14) using the atomic coordinates of the corresponding unmodified DNA duplexes [PDB ID 1EHV for GX9^2, (15); PDB ID 355D for GX9^3 and AX8, (16)] as structural probes. The molecular structures were constructed and modified on a graphic workstation with the program 'QUANTA' (Accelrys Inc.). The atomic parameters were refined with the program 'CNS' (17) through a combination of rigid-body, crystallographic conjugate gradient minimization refinement and *B*-factor refinements, followed by interpretation of an omit map at every nucleotide residue. Newly defined patches for the modified residue were used. The statistics of structure refinements are summarized in Table 1. All global and local helical parameters, as well as the torsion angles and pseudorotation phase angles of sugar rings, were calculated using the program '3DNA' (18).

Coordinates

Coordinates and structure factors have been deposited in the Protein Data Bank with accession codes 3N4N, 3GJH and 3N4O for GX9[2], GX9[3] and AX8, respectively.

Table 1. Crystal data and statistics of data collection and structure refinement

Crystal code	GX9[2]	GX9[3]	AX8
Crystal data			
Space group	$P3_212$	$P2_12_12_1$	$P2_12_12_1$
Unit cell (Å)			
a	26.5	25.2	25.0
b	26.5	41.4	41.7
c	99.0	64.9	64.6
Z[a]	1	2	2
Data collection			
Resolution range (Å)	50–1.92	50–2.9	50–2.9
Outer shell (Å)	1.99–1.92	3.00–2.90	3.00–2.90
Observed reflections	66 898	10 371	22 423
Unique reflections	3233	1708	1714
Completeness (%)	99.4	98.9	99.8
In the outer shell (%)	100	94.6	100
R_{merge} (%)[b]	6.4	6.4	3.5
In the outer shell (%)	29.1	32.2	29.8
I/σ	91.6	38.7	69.6
In the outer shell	15.9	2.5	14.3
Redundancy[c]	20.7	6.1	13.1
In the outer shell	21.4	4.2	13.5
Structure refinement			
Resolution range (Å)	11.7–1.92	10–2.9	9.9–2.9
R-factor (%)[d]	25.8	25.9	23.2
R_{free} (%)[e]	26.5	30.6	29.9
RMSD			
Bond distances (Å)	0.011	0.008	0.009
Bond lengths (°)	2.0	0.9	1.1
No. of additive molecules	–	1 DAPI	1 Hoechst 33 258
No. of water molecules	49	16	30

[a]Number of DNA strands in the asymmetric unit.
[b]$R_{merge} = 100 \times \Sigma_{hj} |I_{hj} - <I_h>| / \Sigma_{hj} I_{hj}$, where I_{hj} is the *j*-th measurement of the intensity of reflection h and $<I_h>$ is its mean value.
[c]Diffraction patterns of 1° oscillation ranges were collected in total 180 frames of GX9[3], and 360 frames of GX9[2] and those of AX8. In the same ways, the second data sets were taken for each crystal with a short exposure time to compensate overloaded reflections.
[d]R-factor $= 100 \times \Sigma ||F_o| - |F_c|| / \Sigma |F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively.
[e]Calculated using a random set containing 10% of observations that were not included throughout refinement (17).

RESULTS AND DISCUSSION

Quality of X-ray analyses

The trigonal GX9[2] crystal diffracted well up to ~2Å. All the residues were traced on the electron density map, except for the first residue of C1 (see Figure 1 for the residue numbering). The electron density corresponding to C1 was protruded and broadened in the solvent region, suggesting that the C1 residue is flipped out from the stacked column of duplexes and is disordered in its conformation. Therefore, it was positioned temporally along with the density shape for further structural refinements. The AX8 crystal was obtained after many trials under conditions similar to those for GX9[2], but it was too small for X-ray diffraction experiments. Co-crystallization with several dyes was then attempted with the hope of stabilizing duplex formation. This approach proved successful, and large crystals of AX8 and a second crystal of GX9 (GX9[3]) were grown. The new crystals were in the orthorhombic form, possibly due to the addition of the duplex-stabilizing dyes. The DAPI and Hoechst 33258 dyes were bound in the central region of the minor grooves of GX9[3] and AX8, respectively (see the details in Supplementary Figure S1).

The X:A base pairs in the AX8 duplex could either be Watson–Crick or wobble type base pairs (Figure 2). At the present resolution, it was difficult to distinguish between these two possible configurations by looking at the initial $|F_o| - |F_c|$ map, in which A5, X8, A17 and X20 residues were omitted. In order to resolve this issue, a disorder model, wherein the Watson–Crick and wobble types were assigned half occupancies, was subjected to least-squares refinement. The resulting $|F_o| - |F_c|$ map, calculated by omitting the base moieties of the four relevant residues, suggested that the refined Watson–Crick types fit better onto the omit map than the refined wobble types which protruded partly from the electron density cages, as shown in Figure 3. The final $|F_o| - |F_c|$ maps of GX9s, in which the X bases were omitted, are also depicted in Figure 3.

Overall structures

The unmodified DNA duplex d(CGCGAATTCGCG)$_2$, well known as the Dickerson–Drew duplex, was originally crystallized in the orthorhombic space group $P2_12_12_1$ with the asymmetric unit containing two DNA strands of the duplex (GC9-$P2_12_12_1$) (19). More recent studies, however, showed that the Dickerson–Drew dodecamer could also
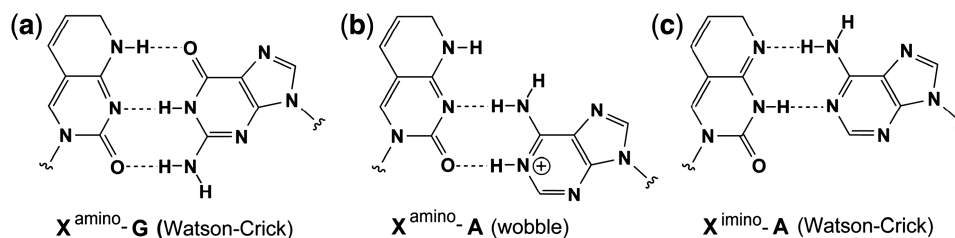

Figure 2. Possible chemical structures and their hydrogen bonding schemes (**a**) for the Watson–Crick type X:G pair, (**b**) for a Watson–Crick type X:A pair, and (**c**) for a wobble type X:A pair.
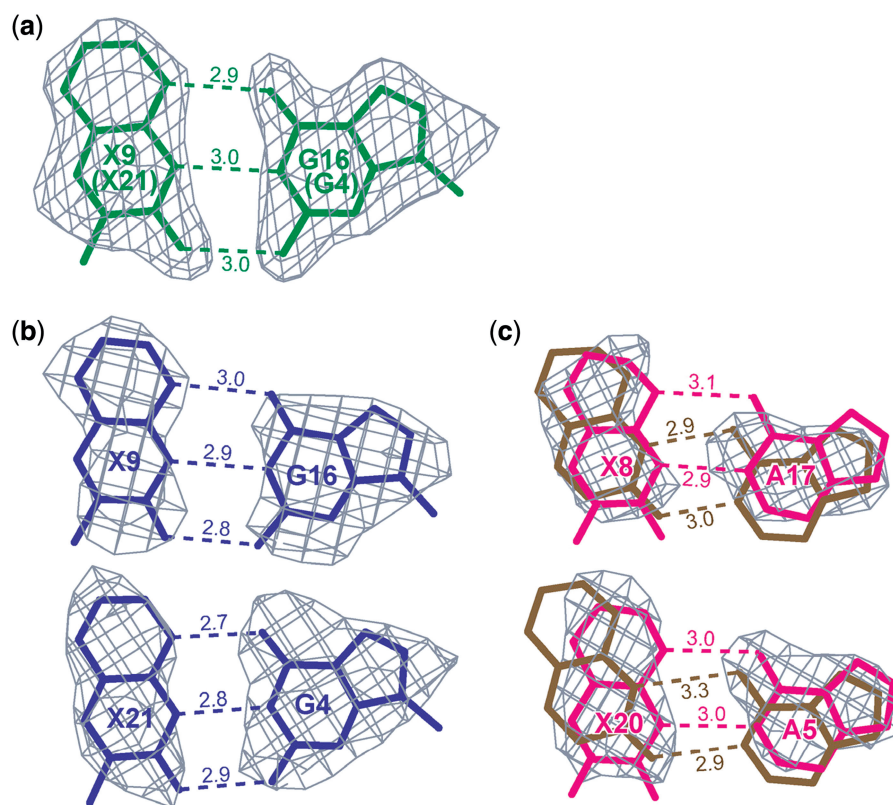
Figure 3. Final $|F_o| - |F_c|$ omit maps of the X:G pairs in GX9[2] (**a**) and GX9[3] (**b**), and of the X:A pairs in AX8 (**c**). The Watson–Crick and wobble type X:A pairs in (**c**) are colored magenta and brown, respectively. The maps were calculated by omitting only the base moieties of the respective pairs, and were contoured at 2σ level. The values indicated are hydrogen bond distances in angstroms (Å).

be packed in the trigonal space groups, R3 [GC9-R3; (20,21)] and P3₂12 [GC9-P3₂12; (15)]. In the present study, GX9 has been crystallized in the P3₂12 (GX9[2]) and P2₁2₁2₁ (GX9[3]) space groups. The AX8 crystal also belongs to the P2₁2₁2₁ space group and its unit-cell parameters are similar to those of the unmodified GC9-P2₁2₁2₁ and the GX9[3] crystals, suggesting that they are isomorphous to each other. In the trigonal GX9[2] crystal, the two strands of the duplex are related by a crystallographic 2-fold symmetry, the axis of which passes between the two central base pairs, A6:T7* and A6*:T7 (the asterisks indicate residues in the symmetry-related strand.). On the other hand, in the orthorhombic GX9[3] crystal, the asymmetric unit consists of a duplex, the two strands of which are not related by a crystallographic 2-fold symmetry.

The average local helical parameters for the modified and unmodified duplexes, as well as for the high-resolution A- and B-form DNA duplexes (22), are listed in Table 2. These parameters show that all the GX9 and the AX8 duplexes adopt the B-form conformation. Superimpositions of the present structures onto the unmodified duplex structures are shown in Figure 4. Excluding the terminal residues at both ends of the duplexes, superimpositions of GX9[2] onto GC9-P3₂12 yielded RMSD values of 0.5Å. Likewise, superimpositions of GX9[3] and AX8, with the exception of the disordered X bases and their partners, onto GC9-P2₁2₁2₁ resulted in

Table 2. Average local helical parameters[a]

	x-Displacement (Å)	Inclination (°)	Helical twist (°)	Helical rise (Å)
GC9-P3₂12[b]	−0.1	1	35	3.2
GX9[2]	−0.04	0.2	35	3.2
GC9 − P2₁2₁2₁[c]	−0.2	2	36	3.3
GX9[3]	0.1	1	36	3.3
AX8 (Watson and Crick)	0.3	2	36	3.3
AX8 (wobble)	0.2	2	36	3.3
B-DNA[d]	0.05	2.1	36.5	3.29
A-DNA[d]	−4.17	14.7	32.5	2.83

[a]Calculated with the program 3DNA (18).
[b]Ref. 15, [c]Ref. 19 and [d]Ref. 22.

RMSD values of 0.7 and 0.8Å, respectively. Closer inspection of the superimposed structures also shows no drastic differences between GX9[2] and the unmodified GC9-P3₂12 duplex. However, the differences between GX9[3] or AX8 and GC9-P2₁2₁2₁ are relatively more pronounced around the A6 and T7 residues. Plots of the minor groove widths (Supplementary Figure S2) indicate that GX9[3] and AX8 are wider at the center compared with those of the other DNA duplexes. These changes in the DNA conformation in GX9[3] and AX8 are again presumably due to the binding of the DAPI and Hoechst 33258 dyes rather than the X substitutions.
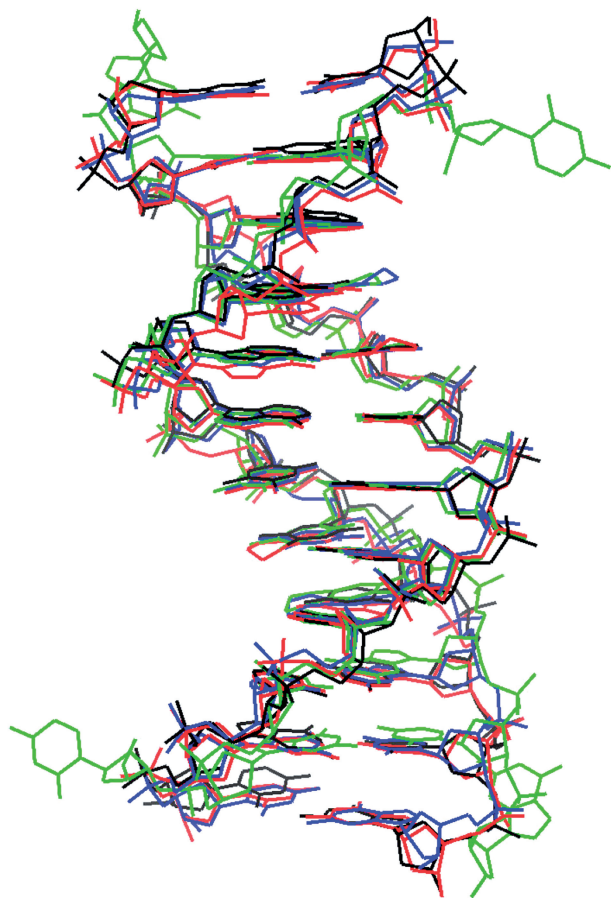
Figure 4. Superimposition of the GX9^2 (green), GX9^3 (blue) and AX8 (red, Watson–Crick type) duplexes onto the GC9-P2$_1$2$_1$2$_1$ duplex (black).

Effects of X substitutions on base stacking and hydrophobicity

The overall stability of nucleic acids is considered a sum of contributions of (i) ionic interactions, (ii) hydrophilic interactions or hydrogen bonds, and (iii) hydrophobic effects or van der Waals interactions. The phosphate backbone is highly ionic and the deoxyribose units are hydrophobic except for their oxygen atoms. The bases have two properties different between two directions: in-plane and out-of-plane. The edge atoms including the carbon atoms are hydrophilic and have strong directionality for interactions. They prefer to form hydrogen bonds in the directions of their valence orbitals including lone-pair electrons (23). In contrast, the other property is hydrophobic for van der Waals interactions in both directions above or below the base plane (24). In general, purines stack more strongly than pyrimidines possibly due to their larger surface area and greater polarizability (25). In hydrophilic environments such as physiological conditions, the hydrophobic parts are excluded from water and brought into close proximity to reduce their exposed surface. In DNA duplexes, the helical nature of the base pairs results in a part of the base surface being exposed at each step. Decreasing as much as

possible the exposure of the bases results in the stabilization of the duplex structure.

Figure 5 shows the overlap areas between the additional rings in the X bases and the bases of the subsequent residues in both GX9^3 and AX8. In order to examine the stacking effects, we estimated the differences in water-accessible surface areas with and without the additional ring of X, using the program 'Naccess' (26). The values (summarized in Table 3) indicate that the overlap areas of most base pair steps containing X substitutions are significantly larger compared to those in unmodified duplexes. Furthermore, the values show that the area changes largely depending on whether the subsequent residue is positioned on the 5'-side or on the 3'-side of the X residue. The overlap area is maximized when a pyrimidine residue occupies the 5'-side of X. The change in the overlap area is more remarkable in the case of the standard B-form DNA structure, because the value is easily affected by the surrounding interactions in crystalline states.

In addition, the second ring in X could block approaching water molecules. This hydrophobic effect of X, although possibly weak, could contribute to stabilization of the duplex form. A similar example is that of a thymine base stabilizing the duplex form more strongly than uracil (27,28) because the methyl group on the C^5 position of uracil impedes water molecules from approaching the base pair.

Geometry of the G:X and A:X base pairs

The final $|F_o| - |F_c|$ maps in which the base moieties of the X:G pairs have been omitted are shown in Figure 3. In the GX9^2 and GX9^3 duplexes, the X:G base pairs have configurations similar to the canonical C:G Watson–Crick base pair, as shown in Figures 2a. All of the N^4(X)...O^6(G), N^3(X)...N^1(G) and O^2(X)...N^2(G) distances are within the range of allowed hydrogen bond distances for base pair formation.

The omit map of the AX8 duplex suggests a preference of the X:A pairs for the Watson–Crick type over the wobble type conformation. Furthermore, in the wobble type, the N^1 atom of the adenine moiety has to be protonated for such a pairing, but it would be difficult for that to occur at neutral pH. In the case of the Watson–Crick type, the modified cytosine base has to adopt the 'imino' tautomer to form the pair with adenine, in which the donor and acceptor sites for hydrogen bonding mimic those of thymine base, as shown in Figure 2c. The T_m values of DNA duplexes containing X show that at neutral pH the A:X pair is more stable than the A:C mismatch pair, while the A:X pair is less stable than the A:T pair (12). This might be ascribed to the low basicity of the X$_{imino}$ base. Observation of pyrimidine bases in their minor tautomeric forms within crystal structures is extremely rare and restricted to analogues in which electronegative substituents are attached to the N^4 amino group of cytosine. Previously we have found that the cytosine analogue, N^4-methoxycytosine, adopts the 'imino' form against an adenine base in a B-form DNA duplex structure (29).
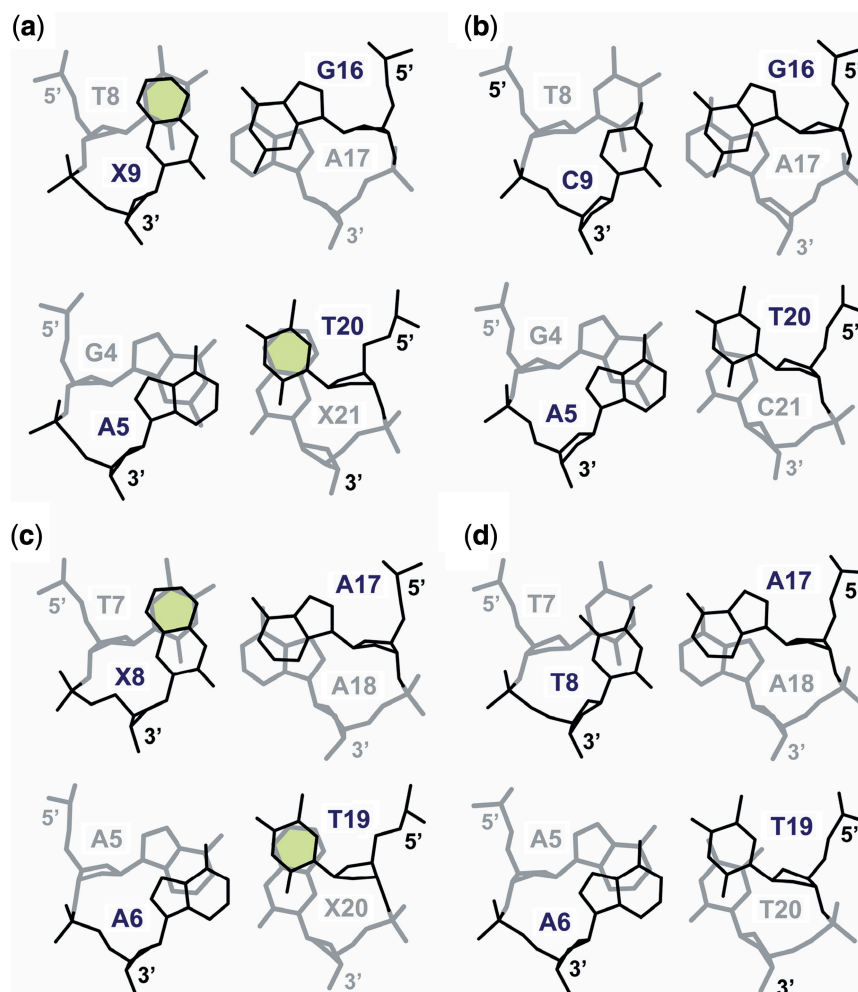
Figure 5. Improvement of the stacking interactions as a result of the X substitutions (highlighted in light green). The stacked base pairs in GX9[3], X9:G16 on T8:A17 and T20:A5 on X21:G4 (**a**), and the corresponding base pairs in GC9-P2$_1$2$_1$2$_1$ (**b**). The stacked base pairs in AX8 (Watson–Crick type), X8:A17 on T7:A18 and T19:A6 on X20:A5 (**c**), and the corresponding base pairs in GC9-P3$_2$12 (**d**). The stacked base pair in GX9[2] is omitted because they closely resemble those in GX9[3]. The figures were generated with the program '3DNA' (18).

Table 3. The stacked/covered area (Å^2) increased by substitution of C with bicyclic C

	GX9^2 $S(S')$	GX9^3 $S(S')$	AX8$^{\text{Watson–Crick}}$ $S(S')$	AX8$^{\text{wobble}}$ $S(S')$	Increase $<\Delta S>$	B-DNA $S(S')$	B-DNA ΔS
TX	72.0 (60.5)	70.8 (56.7)	71.9 (65.3)	71.5 (65.3)	9.0	72.2 (59.9)	12.3
		73.3 (62.0)	70.3 (63.4)	70.1 (63.4)			
CX	–	–	–	–	–	70.1 (60.4)	9.7
GX	–	–	–	–	–	76.1 (68.0)	8.1
AX	–	–	–	–	–	74.8 (66.6)	8.2
XT	–	–	–	–	–	63.8 (61.6)	2.2
XC	–	–	62.1 (56.7)	62.3 (56.7)	5.0	62.0 (60.3)	1.7
			66.5 (62.0)	66.4 (62.0)			
XG	68.1 (64.2)	68.7 (65.9)	–	–	3.7	61.7 (59.6)	2.1
		68.0 (63.6)					
XA	–	–	–	–	–	64.5 (62.7)	1.8

The left-most column indicates the local sequence from 5' to 3'. Values of S and S' are accessible surface areas (Å^2) of the two stacked bases with X = bicyclic C and X = C (in parentheses). Their difference ΔS ($\Delta S = S - S'$) is defined as the increase in the stacked area of X. $<\Delta S>$ is an average of the values of ΔS. The S' values were calculated for the corresponding bases extracted from the unmodified GC9-P3$_2$12 [PDB-ID 1EHV, (19)] and GC9-P2$_1$2$_1$2$_1$ [PDB-ID 1FQ2, (15)] structures. The standard B-form DNA structure was constructed by QUANTA (Accelrys Inc.) and refined by CNS (17). In calculations of S and S', the program Naccess (26) was used. The values of S were calculated as the difference between the two states (stacked and separated).

The crystal structures of duplexes containing the P base, a bicyclic analogue of N^4-methoxycytosine, also reveal that P appears to pair with adenine using its 'imino' tautomer via a protonated wobble pair (30). However, ^{1}H NMR reveals a Watson–Crick base pair for P:A in which P exists as its 'imino' tautomer (31). Further studies such as NMR are clearly required to ascertain the significance of the current findings with respect to the nature of the X:A base pair in DNA duplexes in solution.

The unmodified cytosine residue, however, is expected to predominantly exist in the 'amino' form (32). For A:X to form a wobble pair, the modified cytosine should be in the 'amino' form, and the opposing adenine base must be protonated at the N^1 position. Hydrogen bonds can be formed between N^1-H of adenine and O^2 of X, and between N^6-H of adenine and N^3 of X. Previous studies have suggested that similar base pairing occurs between adenine and cytosine in the structures of unmodified DNA duplexes (33,34). Moreover, thermodynamic studies have shown that the A:C mismatch pair is stabilized by protonation of the adenine base, and that the degree of stabilization is pH dependent (35). The A^+:C pair is most stable near pH 5.3. The present AX8 crystal was, however, obtained at pH 7.0. It might be expected that if AX8 were crystallized at more acidic conditions, A:X may form a wobble pair. We are currently engaged in such a study and will report our findings elsewhere.

As mentioned earlier, the incorporation of the X analogue in place of C in a standard Watson–Crick base pair with G leads to a 3–4 K increase in T_m. A similar increase in T_m is seen when the analogue replaces C in a C:A mismatch i.e. a 3-K increase in T_m relative to the mismatch. In both situations stacking interactions are likely to be the main contributions toward stability. While tautomerization may not directly lead to stabilization, it is possible that X may adopt an 'imino' tautomer in the X:A pair such that the standard Watson–Crick base pair shape can occur which is the optimum arrangement for enhanced stacking.

Local structural effects of X substitutions

The geometric parameters (shear, stretch, stagger, buckle, propeller, opening) for selected base pairs in the X-modified and unmodified duplexes are summarized in Supplementary Table S1. Detailed discussion is difficult at the present resolutions of X-ray analyses. Compared with the corresponding base pairs in the unmodified GC9-P$3_2$12 and GC9-P$2_12_12_1$ duplexes, however, it seems that the geometric values fluctuate within their tolerant ranges, suggesting that there are no remarkable changes even after the X substitutions. Combinations of small changes in all local base pair parameters could allow the extra ring of the X residue to be accommodated in all of the duplexes.

Dye binding

The binding interactions of DAPI and Hoechst 33258 are shown in the Supplementary Figure S2a–g. DAPI is bound in the minor groove of GX9^3 with its indole NH group forming bifurcated hydrogen bonds with the O^2

atoms of the two central thymine bases. These interactions are similar to those found in other structures containing DAPI [PDB ID 1D30 and 432D; (36,37)]. Hoechst 33258 is also bound in the central region of the minor groove of AX8. One of the two benzimidazole groups of Hoechst 33258 donates its NH group to form bifurcated hydrogen bonds similar to the case of DAPI. However, the NH group of the other benzimidazole group forms a hydrogen bond with the N^3 atom of the A6 base. The bulky piperazine ring pushes away the ribose ring of the A5 residue. These interactions are well established in the crystal structures of other DNA duplexes containing Hoechst 33258 [PDB ID 1DNH, 1D43, 1D44, 1D45, 1D46, 127D, 128D, 269D and 303D; (38–41)]. The piperazine moiety, however, has no direct interaction with the A5 base.

Terminal interactions for crystal packing

A remarkable difference between the orthorhombic and trigonal forms is that in the former all bases in one strand are paired with those on the opposite strand, while in the latter the two residues at both ends do not form base pairs and are flipped out of the duplex. In the orthorhombic form (P$2_12_12_1$), the G12 residue paired with C13 gets in contact with G2$^\#$ of another duplex related by a 2_1 screw symmetry along the *c*-axis, through the N^2-H(G12)...N^3(G2$^\#$) and N^3 (G12)...H-N^2(G2$^\#$) hydrogen bonds to form a C:G:G:C quartet. This interaction, typically found in Dickerson–Drew type crystals, results in a high-dihedral angle between the two guanine bases (42).

In the trigonal crystals, a double-stranded column, composed of 10 bp, is stacked on another column, related by a crystallographic 3_2 symmetry (Supplementary Figure S3a). The G12 residue is folded back at the end of the strand so that its G base covers the ribose ring above the C2′ atom and the N^1 atom reach the phosphate oxygen atom of the subsequent C11 residue to form a hydrogen bond (Supplementary Figure S4). It is difficult for RNA to adapt this conformation due to the bulky hydroxyl group attached to the C2′ atom. The base group of the unpaired G12 residue interacts with that of G2$^\#$, which is base paired with C11$^\#$, thus forming a G:G$^\#$:C$^\#$ triplet (Supplementary Figure S3b; the number signs indicate residues in the symmetry-related column). On the other hand, the unpaired C1 and C1$^\#$ residues are protruded into the solvent region at both ends of the duplex, and are slightly disordered. This conformational feature is similar to the unmodified GC9-P$3_2$12 duplex (15).

CONCLUSIONS

The present study identified two important features of X. First, the additional ring of X stabilizes duplex formation by stacking on or by covering the hydrophobic surface of the base adjacent to it. Useful clues for developing antigene and antisense nucleic acids, as well as other nucleic acid based technologies, have been found, e.g. incorporating an X residue on the 3′-side of a pyrimidine

would result in stronger base stacking interactions. Second, X can form pairs with both G and A perhaps through tautomerization between the 'amino' and the 'imino' forms. The latter would be clarified by high-resolution X-ray analysis. However, the stacking effects for duplex stabilization are increased in both Watson–Crick and wobble types, as shown in Table 3.

ACCESSION NUMBERS

3N4N, 3GJH, 3N4O.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

FUNDING

REFERENCES

1. Verma,S. and Eckstein,F. (1998) Modified oligonucleotides: synthesis and strategy for users. *Annu. Rev. Biochem.*, **67**, 99–134.
2. Fodor,S.P., Read,J.L., Pirrung,M.C., Stryer,L., Lu,A.T. and Solas,D. (1991) Light-directed, spatially addressable parallel chemical synthesis. *Science*, **251**, 767–773.
3. Case-Green,S.C., Mir,K.U., Pritchard,C.E. and Southern,E.M. (1998) Analysing genetic information with DNA arrays. *Curr. Opin. Chem. Biol.*, **2**, 404–410.
4. Lipshutz,R.J., Fodor,S.P., Gingeras,T.R. and Lockhart,D.J. (1999) High density synthetic oligonucleotide arrays. *Nat. Genet.*, **21**, 20–24.
5. Pirrung,M.C. (2002) How to make a DNA chip. *Angew Chem. Int. Ed. Engl.*, **41**, 276–289.
6. Praseuth,D., Guieysse,A.L. and Hélène,C. (1999) Triple helix formation and the antigene strategy for sequence-specific control of gene expression. *Biochim. Biophys. Acta*, **1498**, 181–206.
7. Herdewijn,P. (2000) Heterocyclic modifications of oligonucleotides and antisense technology. *Antisense Nucleic Acid Drug Dev.*, **10**, 297–310.
8. Kurreck,J. (2003) Antisense technologies. Improvement through novel chemical modifications. *Eur. J. Biochem.*, **270**, 1628–1644.
9. Buchini,S. and Leumann,C.J. (2003) Recent improvements in antigene technology. *Curr. Opin. Chem. Biol.*, **7**, 717–726.
10. Kurreck,J. (2009) RNA interference: from basic research to therapeutic applications. *Angew Chem. Int. Ed. Engl.*, **48**, 1378–1398.
11. Brazier,J.A., Shibata,T., Townsley,J., Taylor,B.F., Frary,E., Williams,N.H. and Williams,D.M. (2005) Amino-functionalized DNA: the properties of C5-amino-alkyl substituted 2′-deoxyuridines and their application in DNA triplex formation. *Nucleic Acids Res.*, **33**, 1362–1371.
12. Shibata,T., Buurma,N.J., Brazier,J.A., Thompson,P., Haq,I. and Williams,D.M. (2006) 7,8-Dihydropyrido[2,3-d]pyrimidin-2-one; a

13. bicyclic cytosine analogue capable of enhanced stabilisation of DNA duplexes. *Chem. Commun.*, **33**, 3516–3518.
14. Otwinowski,Z. and Minor,W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.*, **276**, 307–326.
15. Navaza,J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crytallogr. A*, **50**, 157–163.
16. Johansson,E., Parkinson,G. and Neidle,S. (2000) A new crystal form for the dodecamer C-G-C-G-A-A-T-T-C-G-C-G: symmetry effects on sequence-dependent DNA structure. *J. Mol. Biol.*, **300**, 551–561.
17. Shui,X., McFail-Isom,L., Hu,G.G. and Williams,L.D. (1998) The B-DNA dodecamer at high resolution reveals a spine of water on sodium. *Biochemistry*, **37**, 8341–8355.
18. Brünger,A.T., Adams,P.D., Clore,G.M., DeLano,W.L., Gros,P., Grosse-Kunstleve,R.W., Jiang,J.S., Kuszewski,J., Nilges,M., Pannu,N.S. *et al.* (1998) Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.*, **54**, 905–921.
19. Lu,X.J. and Olson,W.K. (2003) 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.*, **31**, 5108–5121.
20. Sines,C.C., McFail-Isom,L., Howerton,S.B., VanDerveer,D. and Williams,L.D. (2000) Cations mediate B-DNA conformational heterogeneity. *J. Am. Chem. Soc.*, **122**, 11048–11056.
21. Liu,J., Malinina,L., Huynh-Dinh,T. and Subirana,J.A. (1998) The structure of the most studied DNA fragment changes under the influence of ions: a new packing of d(CGCGAATTCGCG). *FEBS Lett.*, **438**, 211–214.
22. Liu,J. and Subiran,J.A. (1999) Structure of d(CGCGCAATT CGCG) in the presence of Ca^{2+} ions. *J. Biol. Chem.*, **274**, 24749–24752.
23. 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. *et al.* (2001) A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.*, **313**, 229–237.
24. Sasada,Y. and Takénaka,A. (1989) *Molecular Science in Crystals, Directionality of Hydrogen Bonds*. Kodansha Ltd, Tokyo, pp. 90–104.
25. Petersheim,M. and Turner,D.H. (1983) Base-stacking and base-pairing contributions to helix stability: thermodynamics of double helix formation with CCGG, CCGGp, CCGGAp, ACCG Gp, CCGGUp and ACCGGUp. *Biochemistry*, **22**, 256–263.
26. Bommarito,S., Peyret,N. and Santa Lucia,J. (2000) Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res.*, **28**, 1929–1934.
27. Hubbard,S.J. and Thornton,J.M. (1993) *NACCESS*. Department of Biochemistry and Molecular Biology, University College London.
28. Shugar,D. and Szer,W. (1962) Secondary structure in poly-ribothymidylic acid. *J. Mol. Biol*, **5**, 480–582.
29. Wang,S.H. and Kool,E.T. (1995) Origins of the large differences in stability of DNA and RNA helices; C-5 methyl and 2′-hydroxyl effects. *Biochemistry*, **34**, 4125–4132.
30. Hossain,M.T., Sunami,T., Tsunoda,M., Hikima,T., Chatake,T., Ueno,Y., Matsuda,A. and Takénaka,A. (2001) Crystallographic studies on damaged DNAs IV. N4-methoxytosine shows a second face for Watson–Crick base-pairing, leading to purine transition mutagenesis. *Nucleic Acids Res.*, **29**, 3949–2954.
31. Schuerman,G.S., Van,L., Loakes,D., Brown,D.M., Kong Thoo Lin,P., Moore,M.H. and Salisbury,S.A. (1998) A thymine-like base analogues forms wobble pairs with adenine in a Z-DNA duplex. *J.Mol.Biol*, **282**, 1005–1011.
32. Stone,M.J., Nedderman,A.N.R., Williams,D.H., Kong Thoo Lin,P. and Brown,D.M. (1991) Molecular basis for methoxyamine initiated mutagenesis. *J. Mol. Biol.*, **222**, 711–723.
33. Lee,G.C.Y., Prestegard,J.H. and Chan,S.I. (1972) Tautomerism of nucleic acid bases. I. Cytosine. *J. Am. Chem. Soc.*, **94**, 951–959.
34. Hunter,W.N., Brown,T., Anand,N.N. and Kennard,O. (1986) Structure of an adenine-cytosine base pair in DNA and its implications for mismatch repair. *Nature*, **320**, 552–555.
35. Jang,S.B., Hung,L.W., Chi,Y.I., Holbrook,E.L., Carter,R.J. and Holbrook,S.R. (1998) Structure of an RNA internal loop

consisting of tandem C-A$^+$ base pairs. *Biochemistry*, **37**, 11726–11731.

35. Brown,T., Leonard,G.A., Booth,E.D. and Kneale,G. (1990) Influence of pH on the conformation and stability of mismatch base-pairs in DNA. *J. Mol. Biol.*, **212**, 437–440.

36. Larsen,T.A., Goodsell,D.S., Cascio,D., Grzeskowiak,K. and Dickerson,R.E. (1989) The structure of DAPI bound to DNA. *J. Biomol. Struct. Dyn.*, **7**, 477–491.

37. Vlieghe,D., Sponer,J. and Van Meervelt,L. (1999) Crystal structure of d(GGCCAATTGG) complexed with DAPI reveals novel binding mode. *Biochemistry*, **38**, 16443–16451.

38. Teng,M.K., Usman,N., Frederick,C.A. and Wang,A.H. (1988) The molecular structure of the complex of Hoechst 33258 and the DNA dodecamer d(CGCGAATTCGCG). *Nucleic Acids Res.*, **16**, 2671–2690.

39. Quintana,J.R., Lipanov,A.A. and Dickerson,R.E. (1991) Low-temperature crystallographic analysis of the binding of Hoechst 33258 to the double-helical DNA dodecamer C-G-C-G-A-A-T-T-C-G-C-G. *Biochemistry*, **30**, 10294–10306.

40. Sriram,M., van der Marel,G.A., Roelen,H.L., van Boom,J.H. and Wang,A.H. (1992) Conformation of B-DNA containing O6-ethyl-G-C base pairs stabilized by minor groove binding drugs: molecular structure of d(CGC[e6G]AATTCGCG) complexed with Hoechst 33258 or Hoechst 33342. *EMBO J.*, **11**, 225–232.

41. Clark,G.R., Squire,C.J., Gray,E.J., Leupin,W. and Neidle,S. (1996) Designer DNA-binding drugs: the crystal structure of a metahydroxy analogue of Hoechst 33258 bound to d(CGCGAAT TCGCG)$_2$. *Nucleic Acids Res.*, **24**, 4882–4889.

42. Coll,M., Sherman,S.E., Gibson,D., Lippard,S.J. and Wang,A.H. (1990) Molecular structure of the complex formed between the anticancer drug cisplatin and d(pGpG): C222(1) crystal form. *J. Biomol. Struct. Dyn.*, **8**, 315–330.

43. Jones,T.A., Zou,J.Y., Cowan,S.W. and Kjeldgaard,M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr*, **A47**, 110–111.