

X-Ray analysis of d(CGCGAATTXGCG)₂ containing a 2'-deoxy-*N*⁴-methoxycytosine residue at X: a characteristic pattern of sugar puckers in the crystalline state of the Dickerson–Drew type DNA dodecamers

M. Tofazzal Hossain^a, Jiro Kondo^a, Yoshihito Ueno^{b,1}, Akira Matsuda^b, Akio Takénaka^{a,*}

^aGraduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, Nagatsuta, Midori-ku, Yokohama 226-8501, Japan

^bGraduate School of Pharmaceutical Sciences, Hokkaido University, Nishi 6, Kita 12, Kita-ku, Sapporo 060-0812, Japan

Received 1 October 2001; received in revised form 16 November 2001; accepted 26 November 2001

Abstract

In a series of structural studies on damaged DNA, a modified Dickerson–Drew dodecamer with the sequence d(CGCGAATTmo⁴CGCG), where mo⁴C is 2'-deoxy-*N*⁴-methoxycytidine, was synthesized and its structure in a new crystal form has been determined by the X-ray diffraction method. The two dodecamers form a *B*-form duplex, in which the two mo⁴C residues, respectively, form a wobble pair and a Watson–Crick type pair with the guanine residues of the opposite strand. A comparison of the sugar conformations with those of the other related Dickerson–Drew dodecamers indicates a common feature of their puckering patterns. The sugar pucker of the third residue always adopts an intermediate state (C4'-exo ~ O4'-endo) between the *A*-form and *B*-form. This deviation is ascribed to the stacking interaction of the ribose ring at the third residue with the guanine base at the 12th residue, which is brought about by an extra G12:G2 interaction between two duplexes related by a crystallographic 2₁ symmetry. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: X-Ray structure; Damaged DNA; *N*⁴-methoxycytosine; Sugar pucker; DNA dodecamer

1. Introduction

Five-membered ribose rings of nucleotides are not planar so that one or two atoms of each ring shift up or down from the plane of the other atoms.

* Corresponding author. Tel.: +81-45-924-5709; fax: +81-45-924-5748.

E-mail address: atakenak@bio.titech.ac.jp (A. Takénaka).

¹ Present address: Faculty of Engineering, Gifu University, Yanagido, Gifu 501-1193, Japan.

Such a puckering can move from one atom to another in the ring cyclically. Therefore different types of puckers are possible. Extensive crystallographic studies of oligonucleotides revealed that the ribose ring predominantly adopts one of two conformers, the C2'-endo and C3'-endo forms [1]. Local changes from one puckering to another can lead to a drastic change of the whole structure. Thus, every nucleotide in *B*-form DNA has a C2'-

endo sugar pucker [2], while, every nucleotide in A-form DNA has a C3'-endo sugar pucker, similar to that of RNA [3,4]. Recent high-resolution X-ray analyses have made it possible to understand the detailed structures of nucleic acids, and small changes of the local structures have been described with a large set of standard helical parameters [5,6]. However, sugar puckering is also sensitive to such changes, and can perhaps serve as a more convenient parameter to describe the flexibility of nucleic acid structures. To express the puckering mode of the ribose ring, the most definite parameter is the pseudorotation phase angle of the ribose ring [7], the value of which is 18° for the A-form and 162° for the B-form.

In a series of structural studies on damaged DNA, we reported the crystal structures of Dickerson–Drew type DNA dodecamers containing a methoxylated adenine or a methoxylated cytosine residue [8–10]. Recently we reported the crystal structure of a DNA dodecamer (mo⁴C:G-1) with the sequence d(CGCGAATTmo⁴C:GCG), where mo⁴C is 2'-deoxy-*N*⁴-methoxycytidine [11]. In the present study, a second crystal form (mo⁴C:G-2) has been obtained under slightly different crystallization conditions. To investigate the conformational differences due to crystal packing, the crystal structure of mo⁴C:G-2 has been determined. The detailed conformational variations in DNA structures have been recently reviewed from X-ray [12] and NMR [13] studies, and discussed in relation to the transition from the A-form to B-form [14]. Therefore a comparison of the present structure with those of other related DNAs will give some insight into the flexibility and conformational transitions of DNA, which are important not only for understanding DNA–protein recognition, but also for designing specific functional DNAs such as deoxyribozymes or aptamers.

2. Materials and methods

2.1. Preparation and data collection

A Dickerson–Drew type DNA dodecamer containing 2'-deoxy-*N*⁴-methoxycytidine at the ninth position was synthesized by the reported method [15]. The mo⁴C:G-2 crystals were grown at 4 °C

Table 1

Statistics of data collection and crystal data

Space group	P2 ₁ 2 ₁ 2 ₁
Unit cell (Å)	<i>a</i> = 25.3, <i>b</i> = 41.4, <i>c</i> = 64.3
Asymmetric unit (duplexes)	4
Resolution (Å)	100–2.1
Measured reflection	22 624
Unique reflections	3959
Completeness (%)	93.8
In the outer shell (%)	85.7 (2.21–2.1 Å)
<i>R</i> _{merge} ^a (%)	2.7

$$^a R_{\text{merge}} = 100 \times \sum_{\text{hklj}} |I_{\text{hklj}} - \langle I_{\text{hkl}} \rangle| / \sum_{\text{hklj}} \langle I_{\text{hkl}} \rangle.$$

by the hanging drop vapor diffusion method, under conditions very similar to the condition for the first crystal (mo⁴C:G-1) [11], except that the 2-methyl-2,4-pentanediol (MPD) concentration was 25% in the reservoir. Suitable crystals were transferred into a cryoprotectant solution containing 40% (v/v) MPD, and then picked up and flash frozen in liquid N₂.

X-Ray diffraction data were collected at 100 K on the Sakabe–Weissenberg camera [16] with synchrotron radiation ($\lambda = 1.00$ Å) at the Photon Factory (BL-6B) in Tsukuba. The crystal diffracted at a 2.1-Å resolution. Diffraction patterns were processed with the program DENZO [17]. Intensity data were scaled and merged into a set of inde-

Table 2

Statistics of structure refinement

Resolution range (Å)	10–2.1
Used reflection	3912
<i>R</i> -factor ^a (%)	21.1
<i>R</i> _{free} ^b (%)	25.2
Number of DNA atoms	490
Number of waters	80
Number of magnesium atoms	1
R.m.s. deviation from ideal geometry	
Bond lengths (Å)	0.004
Bond angles (°)	0.9
Improper angles (°)	1.3
Average <i>B</i> -factors (Å ²)	
DNA	26.8
Waters	36.8
Magnesium atom	29.8

^a *R*-factor = $100 \times \sum \|F_o\| - |F_c| / \sum |F_o|$, where $|F_o|$ and $|F_c|$ are the observed and calculated structure factor amplitudes, respectively.

^b Calculated using a random set containing 10% of observations that were not included during refinement [23].

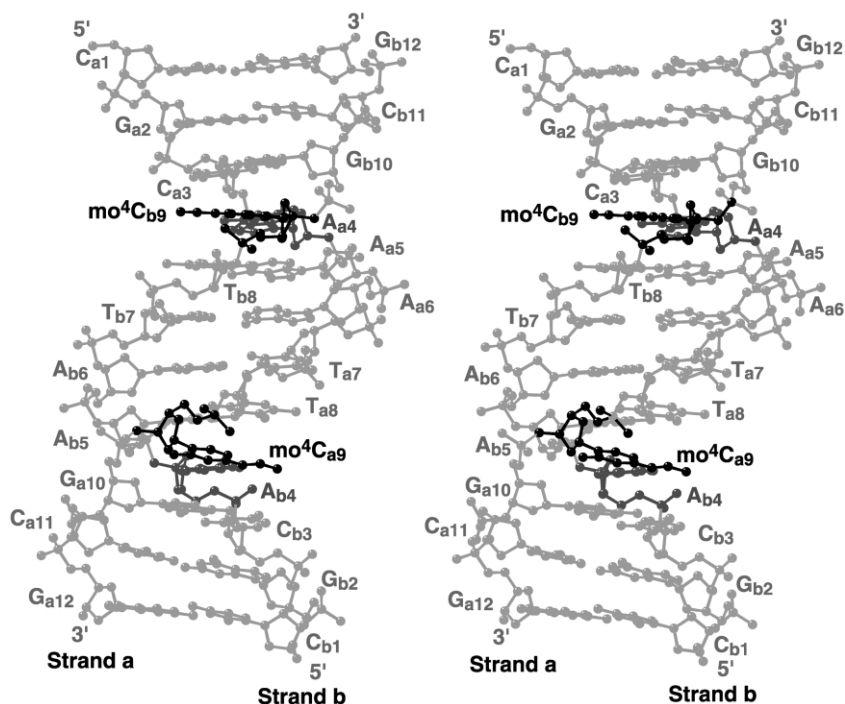


Fig. 1. A stereo view of the present DNA dodecamer $\text{mo}^4\text{C}:\text{G}-2$ with the sequence $\text{d}(\text{CGCGAATT}\text{mo}^4\text{CGCG})$. This diagram was drawn with the program MOLSCRIPT [24]. The nucleotides are numbered from the 5' end independently in the two strands a and b.

pendent reflections by the programs SCALA and AGROVATA, and finally converted to structure factors using the program TRUNCATE from the CCP4 suite [18]. In total, 3959 unique reflections with

R_{merge} 2.7% were obtained from 22 624 observed reflections. Completeness of the data was 93.8% in the 100–2.1-Å resolution range and 85.7% in the highest resolution shell (2.21–2.1 Å). Statistics

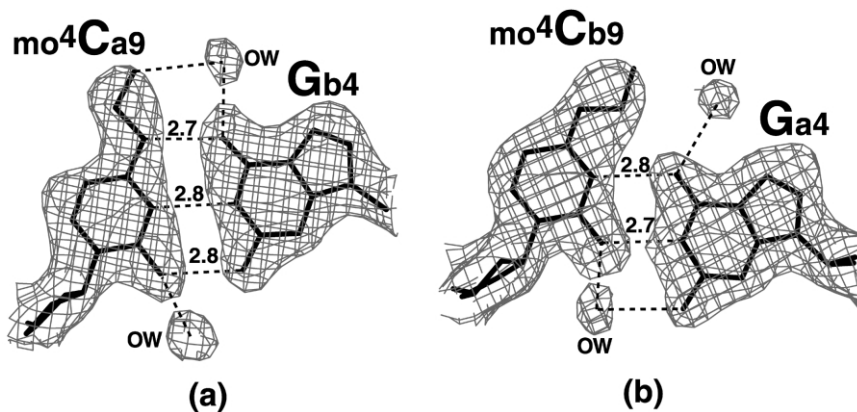


Fig. 2. $2|F_o| - |F_c|$ electron density maps for the $\text{mo}^4\text{Ca}_9:\text{Gb}_4$ (a) and $\text{mo}^4\text{Cb}_9:\text{Ga}_4$ (b) base pairs of the $\text{mo}^4\text{C}:\text{G}-2$ crystal. Broken lines indicate possible hydrogen bonds. Maps are contoured at the 1.0σ level by the program O [25].

of data collection and crystal data are given in Table 1.

2.2. Structure determination

Initial phases were derived by molecular replacement with the program AmoRe [19] using the atomic coordinates of the original DNA dodecamer d(CGCGAATTCGCG) [20] as a probe. The molecular structure was constructed and modified on a graphics workstation by inspecting $|F_o| - |F_c|$ omit maps at every nucleotide residue with the program QUANTA (Molecular simulation Inc.). According to the hydrogen-bonding scheme observed in the $|F_o| - |F_c|$ maps, one of the two mo⁴C residues of the duplex was assumed to adopt an amino form with the methoxy group in the *anti* conformation and the other mo⁴C residue was assigned as an imino form with a *syn* methoxy group. The stereochemical parameters of the two forms of mo⁴C were taken from the file used in the previous X-ray analysis of the mo⁴C:G-1 crystal [11]. The structure was refined with the program CNS [21] through a combination of rigid body, simulated annealing, crystallographic conjugate gradient minimization refinements and *B*-factor refinement, followed by interpretation of omit maps at every nucleotide residue. During refinement no restraints were applied between paired nucleotides. At the beginning of refinement,

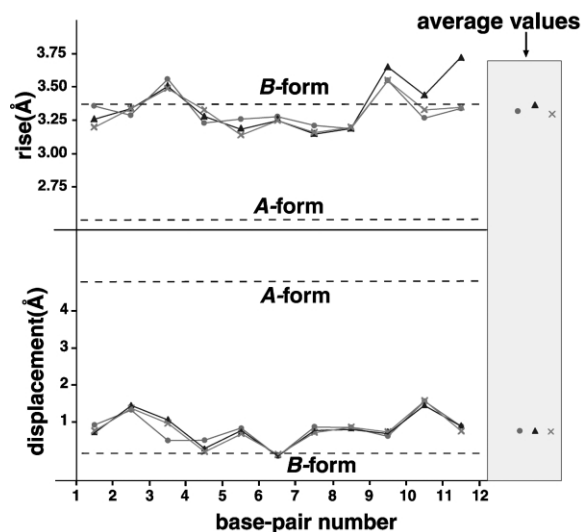


Fig. 4. A comparison of two local helical parameters of mo⁴C:G-2 (crosses), mo⁴C:G-1 (triangles) [11] and the original dodecamer (circles) [22]. The helical rises (top) and displacements (bottom) are plotted along the nucleotide sequences. The average values are indicated in the shaded column, and the ideal forms are indicated with broken lines.

every sugar pucker was assumed to be C2'-endo as in *B*-form DNA, but in the final refinements, the puckering restraint was released. A magnesium cation, which is octahedrally surrounded by six water molecules, was found. A total of 80 peaks were assigned as water molecules. These atoms

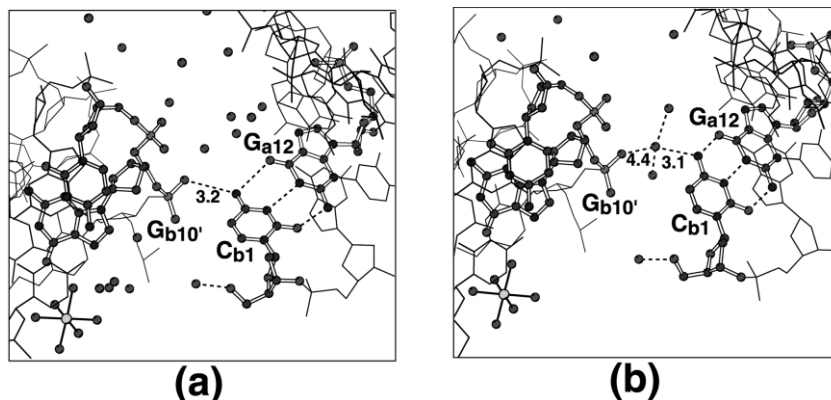


Fig. 3. Different molecular packing of duplexes. The terminal cytosine residue of strand b directly forms a hydrogen bond with the phosphate oxygen atom of the adjacent duplex in the mo⁴C:G-1 crystal (a), while a water molecule mediates between them in mo⁴C:G-2 crystal (b). Broken lines with distances (Å) show possible interactions.

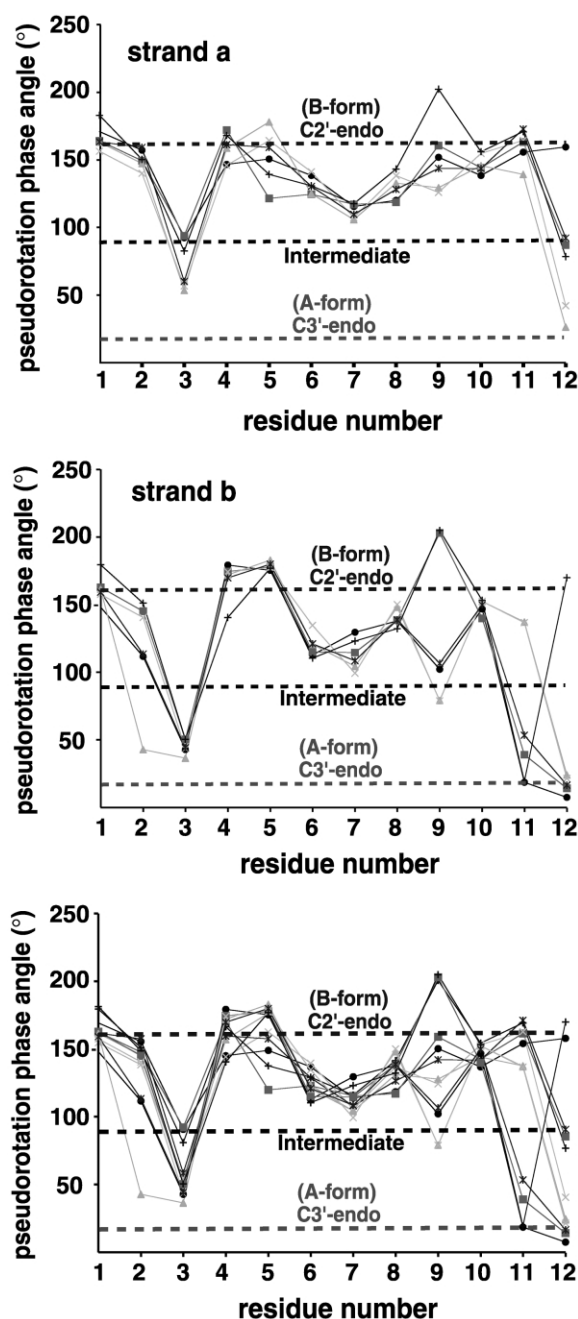


Fig. 5. Comparison of sugar pucker in $\text{mo}^4\text{C}:\text{G}-2$ (crosses) with those in $\text{mo}^4\text{C}:\text{G}-1$ (triangles), $\text{mo}^4\text{C}:\text{A}$ (squares), $\text{mo}^6\text{A}:\text{T}$ (pluses), $\text{mo}^6\text{A}:\text{C}$ (stars) and the original dodecamer (circles). The top and middle are for strands a and b, respectively. In the bottom, the two diagrams are superimposed. In each diagram, the three broken lines indicate the sugar pucker of *B*-form, *A*-form and their intermediate state, respectively.

were included in the refinements. Statistics of the structure determination are summarized in Table 2. The atomic coordinates and structure factors have been deposited in the Protein Data Bank² (PDB) with the code number 1I47.

For comparison of the conformational differences, the atomic coordinates of four types of Dickerson–Drew type DNA dodecamers containing a methoxylated base, $\text{d}(\text{CGCAAATTmo}^4\text{CGCG})$ (designated as $\text{mo}^4\text{C}:\text{A}$, PDB-ID 1J8L) [10], $\text{d}(\text{CGCGAATTmo}^4\text{CGCG})$ ($\text{mo}^4\text{C}:\text{G}-1$, PDB-ID 1I3T) [11], $\text{d}(\text{CGCGmo}^6\text{AATTCGCG})$ (designated as $\text{mo}^6\text{A}:\text{T}$, PDB-ID 1EDR) [8], $\text{d}(\text{CGCGmo}^6\text{AATCCGCG})$ (designated as $\text{mo}^6\text{A}:\text{C}$, PDB-ID 456D) [9], and the original dodecamer, $\text{d}(\text{CGCGAATTCGCG})$ (PDB-ID 1FQ2) [22], have been retrieved from Nucleic acid Data Base (NDB). Their local helical parameters and the pseudorotation phase angles of the individual residues have been calculated with the program NUPARAM [5].

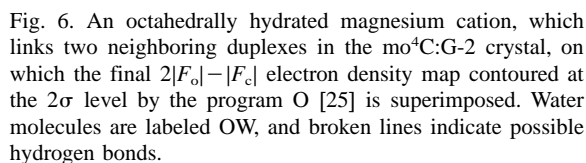
3. Results and discussion

The $\text{mo}^4\text{C}:\text{G}-1$ and $\text{mo}^4\text{C}:\text{G}-2$ crystals, obtained under slightly different conditions, seem to be isomorphous, belonging to the same space group $\text{P}2_12_12_1$ with similar unit-cell dimensions. But when their X-ray intensity data are compared to each other, the discrepancy ($R=0.35$ between them) is large, suggesting some structural differences.

The DNA dodecamers (strands a and b) of the $\text{mo}^4\text{C}:\text{G}-2$ crystal form a duplex in *B*-form conformation, as shown in Fig. 1. The electron-density maps of the two *N*⁴-methoxycytidine residues, $\text{mo}^4\text{C}_{\text{a}9}$ and $\text{mo}^4\text{C}_{\text{b}9}$, are shown in Fig. 2 (see the numbering system in Fig. 1). The $\text{mo}^4\text{C}_{\text{a}9}$ residue forms a Watson–Crick type pair with the $\text{G}_{\text{b}4}$ residue of the opposite strand, while the $\text{mo}^4\text{C}_{\text{b}9}$ residue forms a wobble pair with $\text{G}_{\text{a}4}$. These structural features are the same as those found in the $\text{mo}^4\text{C}:\text{G}-1$ crystal [11].

A significant difference is observed in the crystal packing of the duplexes. Fig. 3 shows the terminal cytosine residues at the first position of

² PDB Reference: DNA dodecamer duplex, 1I47.



Two representative local helical parameters, rise and displacement, plotted along the nucleotide

In the second strand b, it seems that the puckering behaviors are very similar to those of strand a (see Fig. 5, middle). This similarity may be due to the molecular twofold symmetry in the palindromic sequence. Some differences of sugar pucker between the two strands a and b are found at the ninth and 11th residues. At the ninth residue, the fluctuation is enlarged. The sugar pucker of

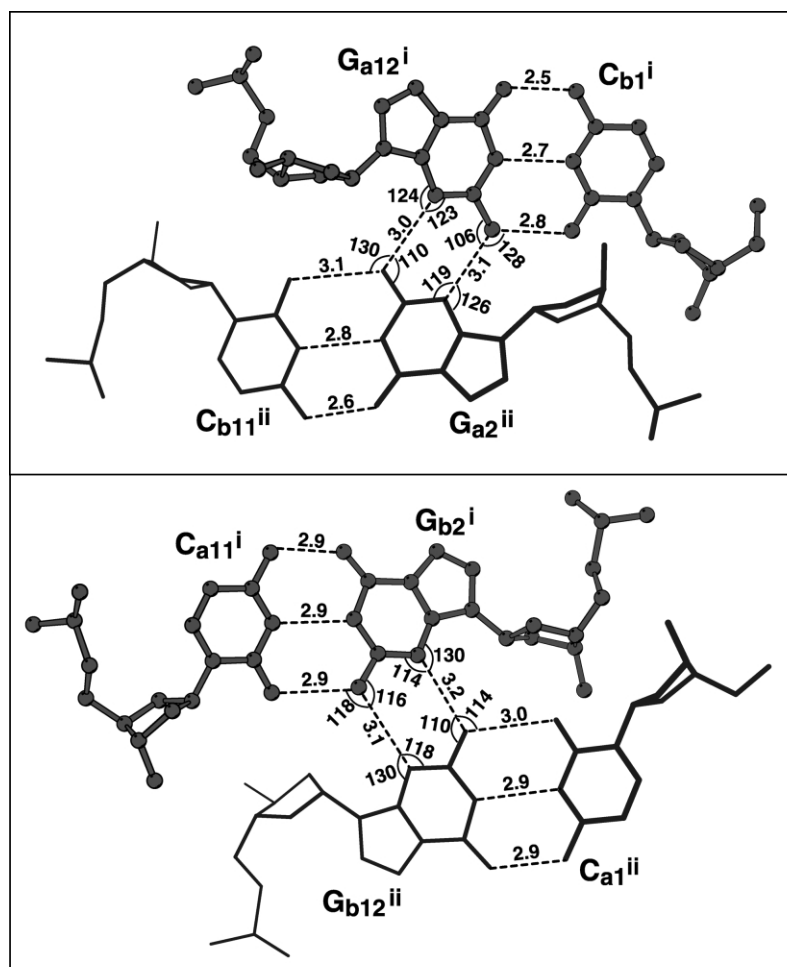


Fig. 7. The guanine–guanine interaction between the two duplexes related by the crystallographic 2_1 symmetry along the c -axis (at the symmetry codes $i:x,y,z$ and $ii:1/2-x,-y,1/2+z$). This interaction occurs at the G_{a12} and G_{b2} residues of one end of the duplex and at the G_{a2} and G_{b12} residues of the other end. Broken lines with distances (Å) indicate possible hydrogen bonds, and values with three digits are angles.

$\text{mo}^4\text{C:A}$ at this residue is C2'-endo or *B*-form. This difference may be due to the surrounding solvent structure. At the 11th residue, the sugar pucker of $\text{mo}^6\text{A:T}$ is the *A*-form, and those of $\text{mo}^6\text{A:C}$ and $\text{mo}^4\text{C:A}$ are close to it. These differences may be ascribed to the environmental difference surrounding the duplexes in the $\text{P2}_12_12_1$ crystal packing. One reason is that a hexagonally hydrated magnesium cation, which is bound strongly to the 10th guanine residue, may affect the ninth residue's conformation as shown in Fig.

6. Another reason for local fluctuations may be that these crystals contain several different ions depending on their crystallization conditions. Although other cations and spermine were not identified in the electron density map due to disordering, they must be involved in neutralization of the phosphate negative charges.

The puckering patterns of the two strands are very similar to each other, representing a common feature of the Dickerson–Drew dodecamers. When they are superimposed, the characteristic pattern

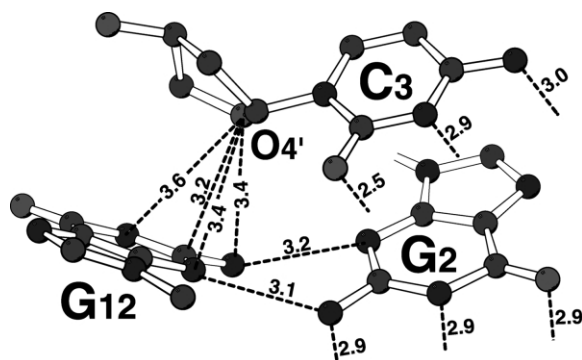


Fig. 8. A side view of guanine–guanine interaction. To form the G:G pair, the guanine base of the 12th residue is very close the ribose ring of the third residue, as if their stacking on each other stabilizes the intermediate sugar pucker. Distances are in Å.

becomes clearer as shown in Fig. 5 (bottom). The pucker at the third residue always has an intermediate state (C4'-exo–O4'-endo) between the A-form and B-form, and at the ninth, 11th and 12th residues, the puckering fluctuates. The characteristic pucker at the third residue is ascribed to the extra G:G interaction between the two duplexes related by a crystallographic 2_1 symmetry along the c axis in the crystals (see Fig. 7). The two guanine bases are bound to each other strongly through the two N2–H...N3 hydrogen bonds. To make this interaction, the two guanine bases are not coplanar, but tilted relative to each other. The dihedral angle between the two guanine planes is 38° for G_{a12} and G_{a2} in $\text{mo}^4\text{C}:\text{G}-1$, 40° in $\text{mo}^4\text{C}:\text{G}-2$, 36° in $\text{mo}^4\text{C}:\text{A}$, 38° in $\text{mo}^6\text{A}:\text{T}$ and 40° in $\text{mo}^6\text{A}:\text{C}$. The corresponding angle in strand b is 40° for G_{b2} and G_{b12} in $\text{mo}^4\text{C}:\text{G}-1$, 39° in $\text{mo}^4\text{C}:\text{G}-2$, 34° in $\text{mo}^4\text{C}:\text{A}$, 38° in $\text{mo}^6\text{A}:\text{T}$ and 34° in $\text{mo}^6\text{A}:\text{C}$. Because of these interactions, in all dodecamers, the guanine base of the 12th residue comes closer to the ribose ring of the third residue of another duplex (see Fig. 8). The two planar parts seem to be parallel to each other. Therefore the intermediate pucker of the ribose ring at the third residue may be stabilized through stacking on the guanine base.

The sugar-puckering patterns indicate that even in B-form conformation, the individual residues locally fluctuate between B-form and A-form

through an intermediate state, suggesting a transition state for the passage between them. Such fluctuations are dependent on perturbations by surrounding interactions in the crystals, which may give us useful information for understanding the ability of conformational transitions and/or bending when DNA interacts with other molecules. For a collective conformational transition, the individual fluctuations should be correlated and in phase, keeping base–base stacking interactions as well as hydrophobic effects. These behaviors are biologically important, not only when DNA interacts specifically with proteins, but also when DNA forms a specific structure as a functional molecule such as a deoxyribozyme or aptamer. The conformational variation found in this study is consistent with a trend for such a transition, discussed by Dickerson and Ng [14]. Similar behavior of sugar puckering has been recently reviewed from X-ray [12] and NMR [13] studies.

Acknowledgments

We thank N. Sakabe, M. Suzuki and N. Igarashi for facilities and help during data collection at Photon Factory (Tsukuba), and T. Simonson for proof-reading of the original manuscript. This work was supported in part by a grant for the RFTF (97L00503) from the Japanese Society for the Promotion of Science and by the Structural Biology Sakabe Project and by Grants-in-Aid for Scientific Research (No.12480177) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

References

- [1] M. Sundaralingam, Stereochemistry of nucleic acids and their constituents. IV. Allowed and preferred conformations of nucleosides, nucleoside mono-, di-, tri-, tetraphosphates, nucleic acids and polynucleotides, *Biopolymers* 7 (1969) 821–860.
- [2] R. Langridge, D.A. Marvin, W.E. Seeds, et al., The molecular configuration of deoxyribonucleic acid II. Molecular models and their Fourier transforms, *J. Mol. Biol.* 2 (1960) 38–64.
- [3] S. Arnott, M.H.F. Wilkins, W. Fuller, R. Langridge, Molecular and crystal structures of double-helical RNA III, an 11-fold molecular model and comparison of the agreement between the observed and calculated three-

- dimensional diffraction data for 10- and 11-fold models, *J. Mol. Biol.* 27 (1967) 535–548.
- [4] S. Arnott, The geometry of nucleic acids, *Prog. Biophys. Mol. Biol.* 21 (1970) 265–319.
- [5] M. Bansal, D. Battacharyya, B. Ravi, NUPARAM and NUCGEN: software for analysis and generation of sequence dependent nucleic acid structures, *Comput. Appl. Biosci.* 13 (1995) 281–287.
- [6] R.E. Dickerson, M. Bansal, C.R. Calladine, et al., Definitions and nomenclature of nucleic acid structure parameters, *J. Mol. Biol.* 26 (1998) 1906–1926.
- [7] C. Altona, M. Sundaralingam, Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation, *J. Am. Chem. Soc.* 94 (1972) 8205–8212.
- [8] T. Chatake, T. Hikima, A. Ono, Y. Ueno, A. Matsuda, A. Takénaka, Crystallographic studies on damaged DNAs II. *N*⁶-methoxyadenine can present two alternate faces for Watson–Crick base-pairing, leading to pyrimidine transition mutagenesis, *J. Mol. Biol.* 294 (1999) 1223–1230.
- [9] T. Chatake, A. Ono, Y. Ueno, A. Matsuda, A. Takénaka, Crystallographic studies on damaged DNAs I. An *N*⁶-methoxyadenine residue forms a Watson–Crick pair with a cytosine residue in a *B*-DNA duplex, *J. Mol. Biol.* 294 (1999) 1215–1222.
- [10] M.T. Hossain, T. Sunami, M. Tsunoda, et al., Crystallographic studies on damaged DNAs, IV. *N*⁴-methoxycytosine shows a second face for Watson–Crick base-pairing, leading to purine transition mutagenesis, *Nucleic Acids Res.* 29 (2001) 3949–3954.
- [11] M.T. Hossain, T. Chatake, T. Hikima, et al., Crystallographic studies on damaged DNAs: III. *N*⁴-methoxycytosine can form both Watson–Crick type and wobbled base pairs in a *B*-form duplex, *J. Biochem.* 130 (2001) 9–12.
- [12] J.M. Vargason, K. Henderson, P.S. Ho, A crystallographic map of the transition from *B*-DNA to *A*-DNA, *Proc. Natl. Acad. Sci. U.S.A.* 98 (2001) 7265–7270.
- [13] L. Trantírek, Štefl, M. Vorlíčková, J. Koča, V. Sklenář, J. Kypr, An *A*-type double helix of DNA having *B*-type puckering of the deoxyribose rings, *J. Mol. Biol.* 297 (2000) 907–922.
- [14] R.E. Dickerson, H.-L. Ng, DNA structure from A to B, *Proc. Natl. Acad. Sci. U.S.A.* 98 (2001) 6986–6988.
- [15] P. Kong Thoo Lin, D.M. Brown, Synthesis and duplex stability of oligonucleotides containing cytosine–thymine analogues, *Nucleic Acids Res.* 17 (1989) 10373–10383.
- [16] N. Sakabe, X-Ray diffraction data collection system for modern protein crystallography with Weissenberg camera and an imaging plate using synchrotron radiation, *Nucl. Instrum. Methods Phys. Res. Ser. A* 303 (1991) 448–463.
- [17] Z. Otwinoski, W. Minor, Processing of X-ray diffraction data collected in oscillation mode: macromolecular crystallography Part A, *Methods Enzymol.* 276 (1997) 307–326.
- [18] Collaborative Computational Project, Number 4, The CCP4 suite: program: programs for protein crystallography, *Acta Crystallog. sect. D* 50 (1994) 760–763.
- [19] J. Navaza, AMoRe: an automated package for molecular replacement, *Acta Crystallog. sect. A* 50 (1994) 157–163.
- [20] X. Shui, L. McFail-Ison, G.G. Hu, L.D. Williams, The *B*-DNA dodecamer at high resolution reveals a spine of water on sodium, *Biochemistry* 37 (1998) 8341–8355.
- [21] A.T. Brünger, P.D. Adams, G.M. Clore, et al., Crystallography & NMR System: a new software suite for macromolecular structure determination, *Acta Crystallog. sect. D* 54 (1998) 905–921.
- [22] C.C. Sines, L. McFail-Isom, S.B. Howerton, D. Vanderveer, L.D. Williams, Cations mediate *B*-DNA conformational heterogeneity, *J. Am. Chem. Soc.* 122 (2000) 11048–11056.
- [23] A.T. Brünger, Free-*R* value: a novel statistical quantity for assessing the accuracy of crystal structures, *Nature* 355 (1992) 472–475.
- [24] P.J. Kraulis, MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures, *J. Appl. Crystallogr.* 24 (1991) 946–950.
- [25] T.A. Jones, J.Y. Zou, S.W. Cowan, M. Kjeldgaard, Improved methods for building protein models in electron density maps and the location of errors in these models, *Acta Crystallogr. Sect. A* 47 (1991) 110–119.