Crystal Structure of the A-DNA Decamer d(CCIGGCCm^{5}CGG) at 1.6 Å Showing the Unexpected Wobble I·m^{5}C Base Pair

B. Ramakrishnan and M. Sundaralingam
Laboratory of Biological Macromolecular Structure, Departments of Chemistry and Biochemistry, The Ohio State University, Columbus, Ohio 43210 USA

ABSTRACT The crystal structure of the self-complementary decamer d(CCIGGCCm^{5}CGG), where I and m^{5}C replace A and T, respectively, in the Watson-Crick B-DNA decamer d(CCAGGCCTGG) is in the A-DNA conformation. Furthermore, the A-DNA duplex exhibits the unexpected wobble I·m^{5}C$^+$ base pairs with N3 of 5-methylcytosine protonated. The crystals belong to the orthorhombic system, space group $P2_12_12_1$, with a = 25.02, b = 44.95 and c = 47.62 Å with one DNA duplex in the asymmetric unit. Intensity data were collected on our Siemens area detector to 1.6 Å resolution. The structure was solved by the molecular placement method starting with a model from an isomorphous structure. The refinement gave a final R value of 16.3% for 404 DNA atoms and 104 water molecules using 5119 reflections. The hydration of the I·m^{5}C$^+$ wobble base pairs in the major groove stabilizes them as in the G·T/G·U wobbles. A comparison with the Watson-Crick base pairs of another isomorphous structure d(GCGGGCCCGC) reveals that the wobble base pairs are better described by a rotation of the individual nucleotide units around their centers of gravity. This is in contrast to the earlier description of translation of the bases into the grooves. The exposed N4 amino group of 5-methylcytosine in the wobble base pair provides a rationale for its deamination to thymine.

INTRODUCTION

Single crystal structures of the same DNA oligomer in multiple conformations, A-, B-, and Z-DNA, are not known. However, fiber diffraction studies have revealed the A↔B conformational interconversion of random DNA base sequences under different relative humidities (Leslie et al., 1980). Also, circular dichroism studies on alternating purine-pyrimidine GC sequences have shown the conformational transition from the right-handed B-DNA to the left-handed Z-DNA by increasing the salt concentration (Pohl and Jovin, 1972). However, single crystal structure analysis of oligonucleotides approximately 4–12 base pairs long give somewhat different results; the left-handed Z-DNA (Wang et al., 1979) is preferentially obtained when the oligomer starts with a 5′-pyrimidine and ends with a 3′-purine, and the right-handed A/B-DNA is preferentially obtained when the oligomer starts with a 5′-purine and ends with a 3′-pyrimidine (Quadrifoglio et al., 1984; Jain et al., 1987). To study the effect of the base sequence on DNA type, we had replaced the purine·pyrimidine A·T base pairs in the B-DNA decamer d(CCAGGCCTGG) (Heinemann and Alings, 1989) by the pyrimidine·purine C·G base pairs. This replacement transformed the structure into an A-DNA (Ramakrishnan and Sundaralingam, 1993a). Prompted by this observation we minimized the base stacking perturbation by substituting the purine·pyrimidine I·m^{5}C base pairs for A·T pairs, which keeps the minor groove solvent environment the same. The results of this study are reported here.

MATERIAL AND METHODS

The DNA decamer d(CCIGGCCm^{5}CGG) was synthesized with our in-house Applied Biosystem DNA synthesizer-381 (Foster City, CA). The phosphoramidites of inosine and 5-methylcytosine monomers were purchased from Biogenex USA. The crude oligomer was base deprotected and precipitated by ethanol at −25°C in the presence of 2.5 M ammonium acetate. The precipitate was lyophilized and used for crystallization without further purification. Crystallization with the conditions of the B-DNA decamer d(CCAGGCCTGG) (Heinemann and Alings, 1989) did not yield any crystals. However, crystals were readily grown by using conditions very similar to those employed for the A-DNA decamers in our laboratory (Ramakrishnan and Sundaralingam, 1993a,b). The best crystals were grown by the hanging drop method from a drop containing 2 mM DNA (single strand), 40 mM sodium cacodylate buffer, pH 7.0, and 5 mM spermine tetrachloride against 50% (v/v) 2-methyl-2,4-pentanediol in water. The crystals belong to the orthorhombic system, space group $P2_12_12_1$, with a = 25.02, b = 44.95, and c = 47.62 Å with one DNA duplex in the asymmetric unit. For x-ray data collection a crystal of size 0.3 × 0.3 × 0.15 mm was mounted in a thin-walled glass capillary with the mother liquor at one end. Three-dimensional, 1.6-Å resolution intensity data were collected at room temperature (290 K) on our Siemens four-circle area detector with a MacScience rotating anode x-ray source operating at 40 kV and 100 mA and a graphite monochromated CuKα (λ = 1.5418 Å) radiation. The crystal-to-detector distance was 12.0 cm. One ϕ scan (range 180°) and three ω scans (range 55° each) were performed at 0.25° steps, with each frame exposed for 90 s. The frames were processed with XENGEN 2.0 (Howard et al., 1985). Of a total of 21,000 reflections collected, 6,143 were unique (86% of possible reflections) with an R_{sym} of 2.1%. The crystal unit cell dimensions are quite similar to the all G·C/C·G A-DNA decamers d(CCCGGCCGGG) and d(GCGGGCCCGC), indicating that it is isomorphous to them. The refinement of the structure was carried out with X-PLOR (Brunger, 1990). A rigid body refinement starting with the coordinates of the isomorphous decamer d(CCCGGCCGGG) did not show any significant root mean square (rms) deviations between the starting and final positions of the model and gave an initial R value of 34% for 326 reflections between 10 and 5 Å resolution. Further positional and

Received for publication 3 April 1995 and in final form 16 May 1995.

Address reprint requests to Dr. Muttaiya Sundaralingam, Department of Chemistry/Biochemistry, Ohio State University, 012 Rightmire Hall, 1060 Carmack Rd., Columbus, OH 43210-1170. Tel.: 614–292-2999; Fax: 614–292-2524; E-mail: sunda%biot@mps.ohio-state.edu.

0006-3495/95/08/553/06 $2.00

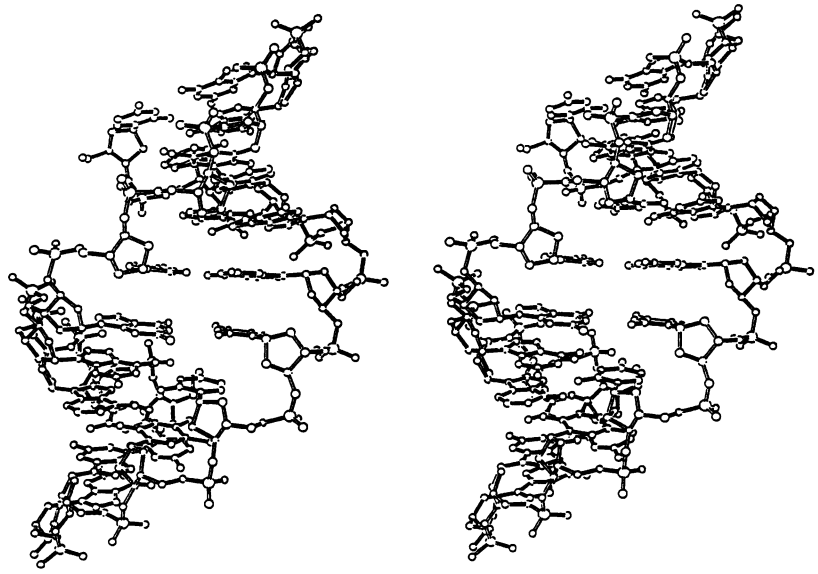


FIGURE 1 The A-DNA decamer d(CCIGGCCm^{5}CGG). The two I·m^{5}C$^+$ wobble base pairs are shown with filled bonds.

thermal parameter refinement using 3015 reflections between 6 and 2 Å resolution dropped the R value to 23.5%. At this stage, "omit" electron density maps showed the base changes, I at positions 3 and 13 and m^{5}C at 8 and 18. The maps also clearly revealed that the I·m^{5}C base pairs were engaged in a wobble base pairing scheme rather than the expected Watson-Crick pairing scheme. The model was corrected to reflect the wobble base pairs and subjected to a simulated annealing refinement by heating the structure at 4000 K and slowly cooling to room temperature. This further dropped the R value to 19.5%. At this stage, 40 water molecules were located from F_o-F_c difference Fourier maps. Further refinement by expanding the data in two steps, to 1.8 Å and 1.6 Å resolution, and adding more water molecules from difference Fourier maps gave a final R value of 16.3% for 5119 reflections with $F > 3\sigma(F)$, between 10 and 1.6 Å resolution. The final model contained 404 DNA atoms and 104 water molecules. The R value for all 5667 reflections was 18.1%. When the water molecules were removed, the R value increased from 16.3% to 21.9%. In the final model, the rms deviations from the ideal geometry (nucleic acid library file PARAM11.DNA) for bond lengths, angles, torsion angles, and improper deviations are 0.013 Å, 3.6°, 31.9°, and 3.1°, respectively. The final atomic coordinates and the structure factors have been deposited with the Protein Data Bank (Bernstein et al., 1977).

RESULTS AND DISCUSSION

Overall DNA conformation

The DNA decamer is in the A-DNA conformation with the following average helix parameters (R. E. Dickerson, personal communication): twist angle 32(4)°, rise per base pair 2.6(3) Å, base pair slide −2(2) Å and base pair inclination 16(3)° (numbers in parentheses are rms values). The average P··P separation in the broad minor groove is 9.9(6) Å and in the narrow major groove is 4.9(4) Å (Fig. 1). The averages of the six backbone torsion angles fall well within the ranges of values seen in other A-DNA crystal structures. The torsions α (P-O5′) and γ (C5′-C4′) of the fifth guanine residue in strand 1 are in the elongated *trans-trans* conformation instead of the conventional *gauche*$^-$-*gauche*$^+$. The assumption of this conformation has been attributed to the intermolecular hydrogen bonding interactions between the fourth guanine and the terminal base

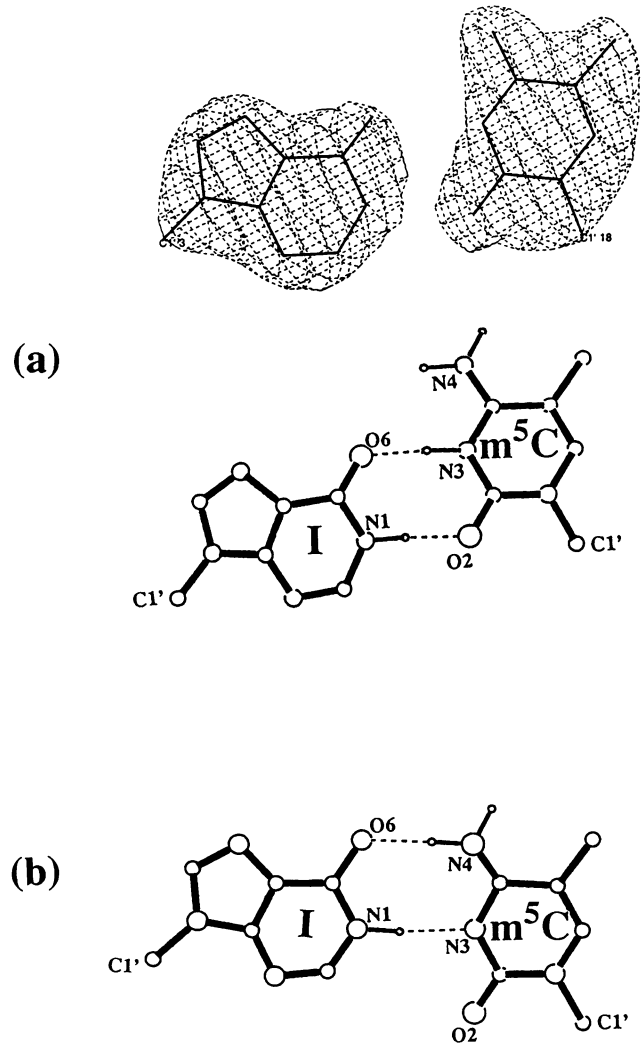


FIGURE 2 (*a*) Difference electron density map for the I(3)·m^{5}C(18) base pair at 1.5 σ and the corresponding wobble base pair. (*b*) For comparison, the Watson-Crick base pair I·m^{5}C is shown.

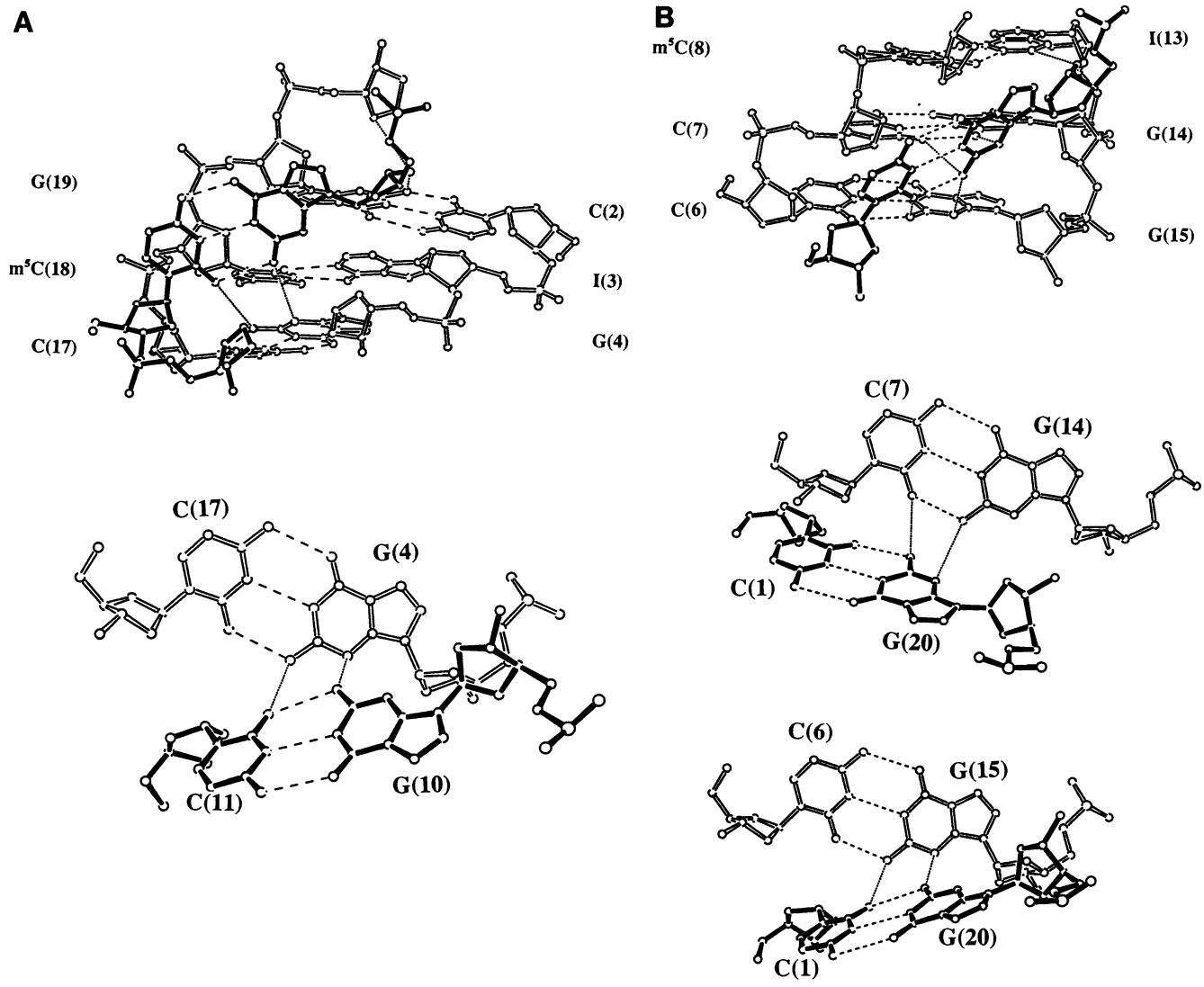


FIGURE 3 Minor groove intermolecular base multiplex interactions in the (*a*) top half and (*b*) bottom half of the duplex. The reference molecule is shown in open bonds and the terminal base pairs of symmetry-related molecules in dark bonds. (*a*) As in the isomorphous crystal structures, the intermolecular base triplet involving the C(11)·G(10) base pair is formed by the hydrogen bonding of the N2 and N3 atoms of guanine G(4) with the O2 atom of cytosine and the N2 atom of guanine. In addition, the N2 atom of G(4) hydrogen bonds to the symmetry-related O4′ sugar ring oxygen atom of C(12) as found in the second isomorphous decamer d(GCGGGCCCGC) (Ramakrishnan and Sundaralingam, 1993b). The terminal O3′ hydroxyl group of G(10) makes hydrogen bonds with the N2 atom of G(19) and the O4′ oxygen of G(20). (*b*) The intermolecular C(1)·G(20) base pair in the bottom half of the duplex makes two base triplets with somewhat weaker hydrogen bonds. The N2 and N3 atoms of G(20) form hydrogen bonds with the O2 atom of cytosine C(7) and N2 atom of G(14) of the C(7)·G(14) base pair. Similarly, the N2 and N3 atoms of G(15) hydrogen bonds to O2 of C(1) and N2 of G(20) in the symmetry-related C(1)·G(20) base pair. The terminal hydroxyl group of the G(20) is engaged in a possible C-H··O interaction with C2 of I(13).

pair of the symmetry-related molecule, as found in the isomorphous crystal structure (Ramakrishnan and Sundaralingam, 1993b). The conformation of the fifth residue is the same in the other isomorphous A-DNA decamer crystal structures with differing base sequences (Ramakrishnan and Sundaralingam, 1993a,b). All of the sugars and the glycosyl torsion angles are in the C3′-*endo* (N-type) puckering and *anti* conformation, respectively.

The wobble I·m^{5}C base pair

It is known that the complementary bases inosine (I) and cytosine (C) form the standard Watson-Crick hydrogen bonds ((I)N1-H··N3(C) and (I)O6··H-N4(C)) in crystals (Xuan and Weber, 1992; Kumar and Weber, 1993; Lipanov et al., 1993). Besides, in crystals, several DNA oligomers containing 5-methylcytosine and guanine also form the standard Watson-Crick (m^{5}C·G) base pair. In both of these cases, the cytosine or 5-methylcytosine is not protonated. In the present crystal structure, we have observed for the first time the I·m^{5}C wobble base pair between these bases instead of the standard Watson-Crick base pair, with the (I)N1··O2(C) and (I)O6··N3(C) hydrogen bonds (Fig. 2 *a*) and N3 of 5-methylcytosine protonated. It is not clear how the m^{5}Cs got protonated at pH 7, although the base sequence may have played a role.

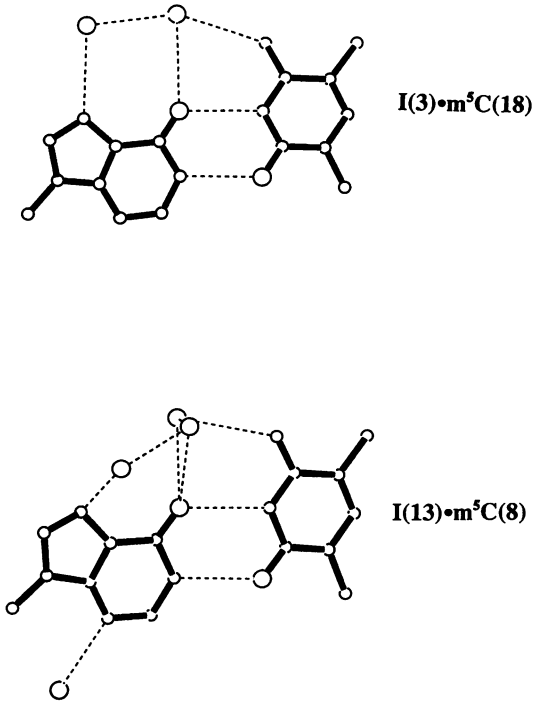


FIGURE 4 Hydration of the two wobble base pairs.

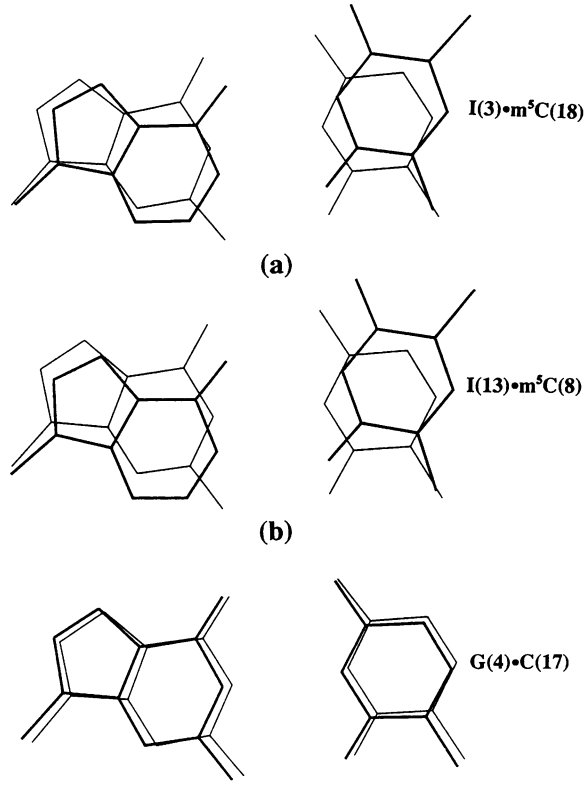


FIGURE 5 A superposition of the wobble (thick lines) base pair on the Watson-Crick (thin line) base pair of the isomorphous decamer structure showing the displacements, purines toward the minor groove and pyrimidines toward the major groove.

We are also investigating this base pair in other DNA sequences. As a control, we have tried to crystallize the nonmethylated decamer d(CCIGGCCCGG) but have not yet obtained crystals. It is noteworthy that the observed wobble pairing is similar to the other known wobbles I·T, G·T, G·U, and A^+·C (Cruse et al. 1989; Hunter et al., 1986, 1987b; Kennard, 1985).

Crystal packing, interduplex base pairing, and hydration

The packing interactions with four surrounding symmetry-related duplexes is very similar to the isomorphous crystals (Ramakrishnan and Sundaralingam, 1993a,b). Each duplex interacts with two symmetry-related duplexes at the termini and two at the minor groove. G(4) forms a base triplet with the terminal C(11)·G(10) base pair (Fig. 3 *a*). Similar but somewhat weaker triplets are formed between the C(7)·G(14) and C(8)·G(13) base pairs in the lower half of the duplex with the terminal C(1)·G(20) base pair (Fig. 3 *b*). A weak C-H··O interaction (3.4 Å) is observed between the terminal O3′ hydroxyl group of G(10) and the C2-H atom of inosine I(13).

Of the 104 water molecules found in the crystal structure, 62 are directly bound to the oligomer in the first coordination sphere. As is found in the other A-DNA decamer crystals, fewer water molecules are found in the minor groove compared with the major groove. The hydration environment of the two I·m^5C^+ wobble base pairs in the crystal are slightly different. The wobble in the top half of the duplex is hydrated by two water molecules interacting

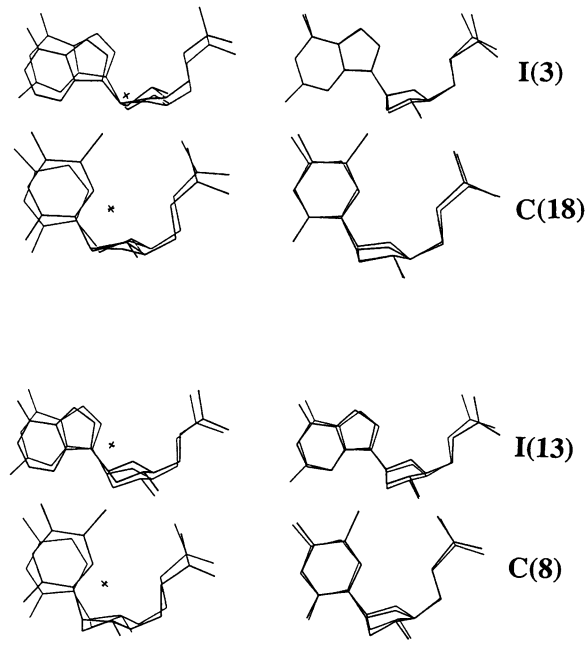
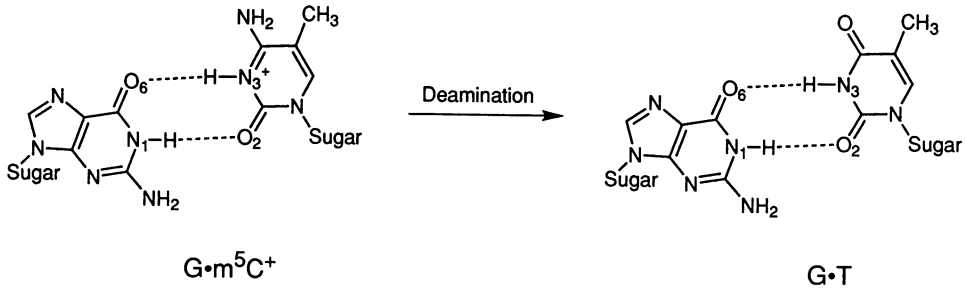


FIGURE 6 The geometrical relationship between the wobble base pair and the Watson-Crick base pair. The inosine nucleotides (thick line) rotate anticlockwise by ~5° and cytosine nucleotides (thick line) rotate clockwise by ~12° along an axis nearly perpendicular to the plane of the paper passing through their centers of gravity shown as a star. The superposition of the base pairs after applying the rotations is shown on the right.



G·m^{5}C$^+$ G·T

FIGURE 7 Schematic diagram showing that the deaminated product of G·m^{5}C$^+$ is G·T.

with N7, O6, and N4 atoms, similar to those found in the I·T/G·T wobble base pairs (Cruse et al., 1989; Hunter et al., 1986) (Fig. 4), whereas the wobble in the lower half is hydrated by three water molecules in the major groove (Fig. 4). On the other hand, in the minor groove the wobble in the top half is devoid of hydration whereas that in the lower half has a water molecule hydrogen bonded to N3 of inosine.

Relationship between the wobble and the Watson-Crick base pairs

It is known that in A-, B-, and Z-DNAs the wobble bases are displaced, the pyrimidine base into the major groove and the purine base into the minor groove, relative to the standard Watson-Crick base pairs (Kennard, 1985). To study the nature of this displacement we overlaid the decamer on the isomorphous Watson-Crick structure, d(GCGGGCCCGC) (Ramakrishnan and Sundaralingam, 1993b) by omitting the terminal base pairs to minimize end effects. This process eliminated to a large extent the effects of crystal packing and revealed mainly the wobble-induced structural changes. The wobble base pairs I(3)·m^{5}C(18) and I(13)·m^{5}C(8) in the 3rd and 8th position showed rms deviations of 0.46 and 0.57 Å with the corresponding Watson-Crick base pairs and, as expected, the inosine base (purine) is displaced toward the minor groove and the 5-methylcytosine base (pyrimidine) toward the major groove (Fig. 5). This displacement can be more accurately described by rotating the individual nucleotide units of the wobble around an axis through their centers of gravity and perpendicular to the base planes. The result is that the inosine nucleotide is rotated by ~5° anticlockwise from the Watson-Crick purine base and the 5-methylcytosine nucleotide by ~12° clockwise from the Watson-Crick pyrimidine base (Fig. 6). A similar description can be given for the G·T wobble (Hunter et al., 1987a, b) in B-DNA (Wing et al., 1980; Westhof, 1987).

Biological implications of the I·m^{5}C$^+$ wobble

It is easier to protonate N3 of 5-methylcytosine (m^{5}C) than cytosine because of its higher pK (Saenger, 1988), which promotes the formation of the wobble base pair. Solvent accessibility calculations (Richards, 1985) indicate that the N4 amino group in the major groove of the wobble pair is twice as exposed as the Watson-Crick base pairs in A- or B-DNA, where N4 is buried by the complementary hydrogen bonding and base stacking. The exposed N4 amino group can undergo deamination more readily, thus mutating m^{5}C to thymine in the DNA double helix. Indeed, it is known that 5-methylcytosine undergoes enhanced mutation (to T) by a factor of 10 compared with cytosine (to U) (Sinden and Wells, 1992). In naturally occurring DNA, guanine (G) occurs instead of inosine (I), which can be expected to form the G·m^{5}C$^+$ wobble base pair that can be deaminated to the G·T wobble base pair (Fig. 7). Thus, the wobble provides a possible mechanism for the deamination of 5-methylcytosine/cytosine.

We gratefully thank the National Institutes of Health for supporting this work through grant GM 17378. M. S. also wishes to thank the Ohio Regents and the University for partial support of this work through an endowed Ohio Eminent Scholar Chair.

REFERENCES

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. B. Rogers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.

Brunger, A. T. 1990. XPLOR Manual, Version 2.1. Yale University, New Haven, CT.

Cruse, W. B. T., J. Aymani, O. Kennard, T. Brown, A. G. C. Jack, and G. A. Leonard. 1989. Refined crystal structure of an octanucleotide duplex with I·T mismatched base pairs. *Nucleic Acids Res.* 17:55–72.

Heinemann, U. and C. Alings. 1989. Crystallographic study of one turn of G/C-rich B-DNA. *J. Mol. Biol.* 210:369–381.

Howard, A. J., C. Nielsen, and N. H. Xuang. 1985. Software for a diffractometer with multiwire area detector. *Methods Enzymol.* 144: 211–237.

Hunter, W. N., T. Brown, and O. Kennard. 1987a. Structural features and hydration of a dodecamer duplex containing two C·A mispairs. *Nucleic Acids Res.* 15:6589–6606.

Hunter, W. N., T. Brown, G. Kneale, N. N. Anand, D. Rabinovich, and O. Kennard. 1987b. The structure of guanosine-thymidine mismatches in B-DNA at 2.5-Å resolution. *J. Biol. Chem.* 262:9962–70.

Hunter, W. N., T. Brown, D. Rabinovich, and O. Kennard. 1986. Refined crystal structure of an octanucleotide duplex with G·T mismatched base-pairs. *J. Mol. Biol.* 190:605–618.

Jain, S., G. Zon, and M. Sundaralingam. 1987. The potentially Z-DNA-forming sequence d(GTGTACAC) crystallizes as A-DNA. *J. Mol. Biol.* 197:141–145.

Kennard, O. 1985. Structural studies of DNA fragments: the G.T. wobble base pair in A, B and Z-DNA; the G·A base pair in B-DNA. *J. Biomol. Struct. Dyn.* 3:205–226.

Kumar, V. D. and I. T. Weber. 1993. Crystal structure of a Z-DNA hexamer d(CGCICG) at 1.7 Å resolution: inosine·cytidine base-pairing,

and comparison with other Z-DNA structures. *Nucleic Acids Res.* 21: 2201–8.

Leslie, A. G. W., S. Arnott, R. Chandrasekaran, and R. L. Ratliff. 1980. Polymorphism of DNA double helices. *J. Mol. Biol.* 143:49–72.

Lipanov, A., M. L. Kopka, M. Kaczor-Grzeskowiak, J. Quintana, and R. E. Dickerson. 1993. Structure of the B-DNA decamer C-C-A-A-C-I-T-T-G-G in two different space groups: conformational flexibility of B-DNA. *Biochemistry.* 32:1373–89.

Pohl, F. M. and T. M. Jovin. 1972. Salt-induced co-operative conformational change of a synthetic DNA: equilibrium and kinetic studies with polyd(G-C). *J. Mol. Biol.* 67:375–396.

Quadrifoglio, F., G. Manzini, and N. Yathindra. 1984. Short oligodeoxynucleotides with d(G-C)n sequence do not assume left-handed conformation in high salt conditions. *J. Mol. Biol.* 175:419–423.

Ramakrishnan, B. and M. Sundaralingam. 1993a. High resolution crystal structure of the A-DNA decamer d(CCCGGCCGGG): novel intermolecular base-paired G*(G·C) triplets. *J. Mol. Biol.* 231:431–444.

Ramakrishnan, B. and M. Sundaralingam. 1993b. Evidence for crystal environment dominating base sequence effects on DNA conformation: crystal structures of the orthorhombic and hexagonal polymorphs of the A-DNA decamer d(GCGGGCCCGC) and comparison with their isomorphous crystal structures. *Biochemistry.* 32:11460–11468.

Richards, F. M. 1985. Calculation of molecular volumes and area for structures of known geometry. *Methods Enzymol.* 115:440–64.

Saenger, W. 1988. Principles of Nucleic Acid Structure. Springer-Verlag, New York. 105–116.

Sinden, R. R., and R. D. Wells. 1992. DNA structure, mutation, and human genetic disease. *Curr. Opin. Biotechnol.* 3:612–622.

Wang, A. J.-H., G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. A. van der Marel, and A. Rich. 1979. Molecular structure of a left-handed double helical DNA Fragment at atomic resolution. *Nature.* 282:680–686.

Westhof, E. 1987. Re-refinement of the B-dodecamer d(CGCGAATTCGCG) with a comparative analysis of the solvent in it and in the Z-hexamer d(5BrCG5BrCG5BrCG). *J. Biomol. Struct. Dyn.* 5:581–600.

Wing, R., H. Drew, T. Takano, C. Broka, S. Tanaka, K. Itakura, and R. E. Dickerson. 1980. Crystal structure analysis of a complete turn of B-DNA. *Nature.* 287:755–8.

Xuan, J. C. and I. T. Weber. 1992. Crystal structure of a B-DNA dodecamer containing inosine, d(CGCIAATTCGCG), at 2.4 Å resolution and its comparison with other B-DNA dodecamers. *Nucl. Acids Res.* 20:5457–64.