Structure of an alternating-B DNA helix and its relationship to A-tract DNA

[DNA groove width/poly(dA)-poly(dT)/DNA bending/propeller-twisted base pairs/DNA groove hydrogen bonds]

CHUN YOON, GILBERT G. PRIVÉ, DAVID S. GOODSELL, AND RICHARD E. DICKERSON

Molecular Biology Institute, University of California at Los Angeles, Los Angeles, CA 90024

Contributed by Richard E. Dickerson, May 27, 1988

ABSTRACT The crystal structure of the synthetic DNA dodecamer CGCATATATTCG has been solved at 2.2-Å resolution. Its central 6 base pairs adopt the alternating-B-DNA helix structure proposed nearly a decade ago. This alternating poly(AT) structure contrasts with the four known examples of what can be termed a poly(A) subfamily of B-DNA structures: CGCGCAGTCGC, CGCGAATTGTGC, CGCGAAATTGCG, and CGCGAATTGCG, their defining characteristic being a succession of two or more adenines along one strand, in a region of 4 or more A-T base pairs. All five helices show a characteristic major groove in their AT centers, but the mean propeller twist at A-T base pairs is lower in the alternating poly(AT) helix than in the poly(A) subfamily of helices. Three general principles emerge from x-ray analyses of B-DNA oligonucleotides: (i) GC and mixed-sequence B-DNA have a wide minor groove, whereas the minor groove is narrow in heteropolymer or homopolymer AT sequences. (ii) G-C base pairs have low propeller twist; A-T pairs can adopt a high propeller twist but need not do so. A high propeller twist can be stabilized by cross-strand hydrogen bonds in the major or minor groove, examples being the minor groove bonds seen in CCAGATTGG and the major groove bonds that can accompany AA sequences in the poly(A) family. (iii) Homopolymer poly(A) tracts may be stiffer than are alternating AT or general-sequence DNA because of these cross-strand major groove hydrogen bonds. Poly(A) tracts appear internally unbent, but bends may occur at junctions with mixed-sequence DNA because of differences in propeller twist, base pair inclination, and base stacking on the two sides of the junction. Bending occurs most easily via base roll, favoring compression of the broad major groove.

For more than a decade B-DNA containing only A-T base pairs has been suspected to differ in important structural respects from the ideal B-DNA model obtained from fiber diffraction (1, 2). At one extreme, the homopolymer poly(dA)-poly(dT), to be identified hereafter merely as poly(A), exhibits 10.1 ± 0.1 bp per turn, close to the 10.0 bp expected from fiber analyses, but poly(A) is unique in failing to wind around nucleosome cores during reconstitution experiments (3–6). In contrast, the heteropolymer poly(dA-dT)-poly(dA-dT), or more simply poly(AT), can be wound around nucleosome cores like general-sequence B-DNA. Surprisingly, both poly(AT) and general-sequence DNA are less tightly wound than poly(A), with 10.6 ± 0.1 bp per turn (3–8).

Both poly(AT) and poly(A) have been proposed, at one time or another, to have an unusual or unorthodox B-DNA structure. A heteronymous structure was suggested for poly(A) from fiber diffraction studies, with different backbone conformations on adenine and thymine strands (9). But this structure has been challenged both by NMR experiments (10, 11) and by later fiber diffraction work (12, 13). Alternatively, the poly(AT) helix has been proposed to adopt a characteristic alternating-B structure, in which the helical twist angle is smaller at A-T steps than at T-A steps (14). The issue of B-DNA conformation in AT regions became particularly important when gel migration and other studies of kinetoplast fragments and synthetic oligomers showed that short runs of poly(A) appeared to produce helical bends, especially when these poly(A) runs synchronize with the ~10 bp per turn periodicity of the helix itself. (For current summaries by most involved principals, see the papers in ref. 15.) Two models were developed to explain this bending. The wedge model proposes that successive A-T base pairs in poly(A) tracts are not perfectly parallel, so that a sequence of adenines will lead to curvature in the poly(A) region itself (16–18). The junction model, in contrast, assumes that the poly(A) region is unbent, but that bends occur at the ends of the poly(A) segment where it joins regions of general-sequence B-DNA (19, 20). The problem of the conformation of AT tracts of B-DNA has been addressed by several single-crystal x-ray structure analyses of DNA sequence CGCGNNGNNGCG, with variable bases N. The first of these sequences, CGCGAATTGC-G (21, 22), showed a high propeller twist at A-T base pairs, a pronounced narrowing of the minor groove in the AT zone, and a bend in the helix axis at the upper junction between GC and AT tracts. The bend was produced by a rolling of adjacent base pairs along their long axes in a direction that compressed the wide major groove (Fig. 1a). The absence of an equivalent bend at the lower AT-to-GC junction in the helix and the removal of the upper bend under slightly different crystal packing conditions in CGCGAATTTCGGC (23) showed that the junction was not a static kink but a flexible hinge, possibly with a preferred direction of flexure compressing the major groove. These first two structures also were significant in demonstrating that an AT region as short as 4 successive base pairs was sufficient for adoption of this unusual helix geometry.

More recently, structure analyses of CGCGAAAAAGC-G/CGCTTTTTTTTCG (24) and CGCGAATTGTGCG (25) have confirmed the narrow minor groove and high propeller twist of AT tracts and the permissive bend in helix axis at a GC-to-AT junction. These analyses also have emphasized a feature potentially present in the earlier structures but which had been overlooked: the possibility of a cross-strand hydrogen bond in the major groove between an adenine N-6 amine and the thymine O-4 on the opposite strand of helix and the adjacent base pair (see Fig. 7). These cross-strand hydrogen bonds have been proposed (24, 25) as a means of stiffening the poly(A) helix to that point where it would resist winding around a nucleosome core and as a way to lock down the high propeller twist seen in A-T base pairs.

We have solved the crystal structure of the alternating-A T dodecamer CGCATATATGC-G as a control and counter-example to the foregoing four structures, which because of

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

6332
their shared A–A steps and other physical similarities can be regarded as members of a common poly(A) subfamily of B-DNA. The alternating poly(AT) structure shows both similarities and differences when compared with the poly(A) group and confirms the Klug alternating-B helix.

EXPERIMENTAL METHODS

The synthesis, crystallization, structure analysis, and refinement of CGCATATGGCG are discussed at length elsewhere (G.G.P., C.Y., and R.E.D., unpublished work) and will only be summarized here. The dodecamer was synthesized by the solid-support phosphoramidite method and crystallized by vapor diffusion from 0.5 mM DNA/22 mM Mg(OAc)2/0.4 mM spermine/10% (vol/vol) 2-methyl-2,4-pentanediol (MPD) solution–crystals occurring at 40% MPD. The P2₁,2,2₁ unit cell, 23.54 Å × 38.86 Å × 66.57 Å, is nearly isomorphous with that of the Drew dodecamer (22). A total of 2994 independent reflections were collected at 6°C on an Area Detector Systems (San Diego, CA) multi-wire detector to 2.2-Å resolution, of which 1915 reflections were above the 1-σ (F) level.

The structure was solved by molecular replacement by the use of the MERLOT program library (26) to position and orient three trial models: the Drew helix (22), the Fratini MPD7 helix (23), and an idealized Arnott helix (1). The Drew and Arnott models were selected for further Konnert–Hendrickson refinement. These two models converged separately and independently to residual errors of R = 0.254 and 0.271, respectively, with an rms difference of 0.718 Å. During this refinement the minor groove of the Arnott model narrowed to match that of the Drew model, and the helical twist angles came into good agreement. Angles that depend on orientation of base and base pair normal vectors, such as base roll and propeller twist, converged less well in the two refinements, undoubtedly as a consequence of the limited resolution.

Both models were then put through the X-PLOR molecular dynamics refinement programs (27), heating them to a nominal 1000 K, and cooling again. This treatment produced convergence, after which the two models were given brief final Konnert–Hendrickson refinement. Final R factors for the 1915 1-σ reflections were 0.217 for the Arnott-based model and 0.239 for the Drew-based structure, with rms differences in atomic positions of 0.37 Å. The Arnott-based model was selected for further discussion because of its presumed greater objectivity.

Difference maps were used to locate 43 water molecules, the incorporation of which reduced the R factor to 0.187. Fragments of a spine of hydration were visible in the minor groove: the first hydration shell bridges from A-6 to A-20 and T-9 to T-17, with two tetrahedrally coordinated second-shell waters in the latter case. No spine had been visible at all after NUCLSQ refinement and before X-PLOR refinement. The failure to detect localized water peaks within the minor groove after NUCLSQ refinement probably results from the low resolution of the x-ray data, and the subsequent appearance of a partial spine may reflect the better phasing power of X-PLOR refinement. Details will be given elsewhere (G.G.P., C.Y., and R.E.D., unpublished work).

RESULTS AND DISCUSSION

At first glance the structure of CGCATATATACCG (Fig. 1b) is scarcely distinguishable from that of CGCGAAATTCGCG (Fig. 1a), in spite of the fact that the latter structure was not used as a starting point in refinement of the new helix as discussed hereafter. The same narrow minor groove is seen and the same tipping to the left of the top of the helix, although to a slightly lesser extent. Significant differences show up, however, when one examines the details of helical structure.

Minor Groove Width. The most striking deviation from a classical B-DNA model in all five of these A+T-rich helices is the narrowing of the minor groove. This narrowing is shown in Fig. 2 for the pure alternating CGCATATACCG, the pure homopolymer CGGAAAAAGCG, and CGCGAAATTCGCG. The width of the minor groove opening expected in B-DNA from fiber diffraction is 5.7 Å (Minimum phosphate–phosphate distance of ≈11.5 Å and an assumed phosphate group radius of 2.9 Å.) This width is approached at the bottom of the helix (right in Fig. 2), but the minor groove opening narrows to 3.5 Å in the AT region, a feature that may be of considerable importance in the binding to DNA of drug molecules, such as netropsin, that are built from flat organic rings (28, 29). The narrow minor groove appears to be one of the most fundamental properties of A+T-rich regions of B-DNA. The extra widening of the minor groove at the top of each helix (left in Fig. 2) occurs because of the bend in the helix, which compresses the major groove and opens the minor one.

Helix Twist Angles. The behavior of helical twist angles in CGCATATATACCG is compared in Fig. 3 with that of CGC-
AAAAGCG. In a perfectly regular Arnott B-DNA helix of alternating adenine and thymine bases, with a uniform helical twist of 36°, the overlap between base pairs at a 5'-Tpa-3' step is intrinsically greater than that at the subsequent 5'-Tpa-3' step. One might have expected, in an alternating poly(AT) helix, that the more fully overlapped step would relax somewhat to improve overlap at the poorer step, approaching an equipartition of base pair stacking energies.

The structure of pATAT (30), noteworthy as the first single crystal x-ray analysis of any synthetic DNA oligomer longer than a dimer, revealed a quite different behavior. The efficiently stacked A-T steps did simulate a helix, but the intrinsically less efficient T-A step was completely disrupted. The first and second A-T steps of a given tetramer strand actually made Watson-Crick base pairs with different neighboring molecules in the crystal lattice.

Klug and coworkers (14) generalized these observations into an alternating-B model for poly(AT), with 36° helical rotation at A-T steps and 39° rotation at T-A steps. The already efficient A-T steps increase their overlap even more by diminishing the local twist angle, whereas the poorer T-A steps become yet poorer by an increase in helical twist. Observed values at successive A-T and T-A steps of our CGCATATGCGG are 33°, 39°, 36°, 430, and 33°. One example of each type of step from this structure is shown in Fig. 4, and these should be compared with fig. 2 of ref. 14.

**Roll Angles.** The most striking feature of roll angles, in Fig. 5, is the regular alternation of roll along the CGCATATGhelix, with pyrimidine-purine steps opening their roll angles toward the minor groove (positive), whereas purine-pyrimidine steps open toward the major groove (negative). This effect was first noted by Dickerson and Drew (22) in connection with CGCGATTCGCG. It is consistent with the expectations of the Calladine model for local helix deformations based on cross-chain purine-purine clash (32, 33), although it does not prove that such clash is the origin of the effect. The same roll angle behavior is implicit in the published model coordinates of the alternating-B helix of Klug and coworkers (14). Analysis of those coordinates with HELIB library (23, 34) shows a roll of +12° at T-A steps and -13° at A-T steps—larger in magnitude but identical in sign to that seen in our alternating-AT dodecamer.

CGCAAAAAAGCG shows the same roll angle behavior at its two ends where purines and pyrimidines alternate. But in the AAAAAA center the roll angles are nearly flat, presumably because A-T base pairs stack in parallel atop one another efficiently, without any "wedge" effect. The alternation of roll angles is greater at the right (lower) end of each helix than at the left because the upper end of each helix is held tightly between neighboring molecules, whereas the lower end is less constrained (31).**

**Mean Propeller Twist.** Mean propeller twist values for this structure and other B-DNA helices are compared in Fig. 6. It was first noticed in CGCGATTCGCG that A-T base pairs had systematically higher propeller twists than G-C pairs by ~8°, and this behavior was confirmed in CGCGATCGCG, CGCAAAAAAGCG, and CGCGATTCGCG. We originally proposed that this higher propeller twist might arise because A-T base pairs are connected only by two hydrogen bonds, whereas G-C pairs have three bonds (22), and that the narrowness of the minor groove in AT regions might be a consequence of high propeller twist, causing the walls of the groove to move close together (23).

Fig. 6 shows the spread in mean propeller twist between G-C and A-T base pairs in CGCATATGCGG as determined independently from the Drew-derived and Arnott-derived starting models. The Drew-based model began at a mean propeller twist of 15.9° for A-T base pairs, and the Arnott-
based model started from only 1.2°. After NUCLSQ and X-PLOR refinement they converged at 15.6° and 16.1°, respectively. In both cases the refined propeller twist at G-C base pairs is comparable with that seen in the four poly(A) structures and in the phosphorothioate analogue of GCGGC as well (35). The refined mean propeller twist at A-T base pairs in CGCATATGC(G) again is greater than at G-C pairs, but only by = 4–5°. Large propeller twist evidently is possible with A-T base pairs but is not mandatory.

Nelson et al. (24) and Coll et al. (25) have both proposed that large propeller twists in A-T base pairs can be stabilized by cross-strand hydrogen bonds in the major groove, con-necting the adenine N-6 amine of 1 bp on one strand with a thymine O-4 of the succeeding base pair, on the opposite strand. This can only occur at sequences of two or more successive adenines. Fig. 7 schematizes the major groove hydrogen-bonding potentialities of all dodecamers examined crystallographically thus far. One such cross-strand hydrogen bond is possible at every A-A (= T-T) step. Such cross-chain major-groove hydrogen bonds are totally excluded in the alternating CGCATATGC(G), schematized in Fig. 7d. Each adenine N-6 or thymine O-4 is systematically opposed to another example of its own kind. If cross-strand hydrogen bonds are responsible for stabilizing a large propeller twist, then it is only to be expected that one would find a smaller propeller twist in the alternating-AT helix.

The connection between high propeller twist and narrow minor groove (23) remains valid in these new dodecamer diagrams of CGCAAAAAAGCG (a), CGCAATTTCGG (b), CGCGAATTTCGG (c), and CGCATATATGCG (d) showing the potential cross-chain hydrogen bonds connecting adenine N-6 amine with thymine O-4. One such hydrogen bond (dotted) is possible in principle at each A-A step along the helix yielding five for a, four for b, two for c, and none for d. A-T steps oppose adenine N-6 amine groups across the major groove, and T-A steps oppose thymine O-4 atoms; no hydrogen bonds are possible in either case. A long cytosine-thymine N-4—O-4 bond could be postulated at a C-A step, and indeed, such a bond has been reported for b. A short guanine-adenine O-6—N-6 bond might be imagined at a G-T step, but no such step has yet been encountered in a crystal-structure analysis. The alternating-AT sequence of d is uniquely designed to disallow all such cross-strand hydrogen bonds, and this may be at the heart of the difference in behavior of homopolymer and alternating heteropolymer tracts of A-T base pairs.
CONCLUSIONS

The most fundamental structural division within the B-DNA family of helices appears to be that between the wide minor groove of GC and mixed-sequence DNA, and the narrow groove associated with AT tracts, whether homopolymer [poly(A)] or heteropolymer [poly(AT)]. However, these are only variations on a common B-DNA theme. The differences among these several B structures are nothing as great as between any of them and a different major structural class such as A- or Z-DNA and should not be exaggerated by capricious elaboration of nomenclature.

The preference of most groove-bridging, base-specific antibiotics and antitumor drugs for regions of AT base pairs may arise not only from the extra room generated along the floor of the minor groove by the absence of N-2 amino groups in adenines, but also from the stabilization that results from the quasiintercalation of flat aromatic rings of the drug between the two walls of the narrow minor groove (28, 29, 36). The CGCGAATTCCGG structure shows that as few as 4 successive A-T base pairs are sufficient to induce the narrow-groove geometry. In contrast, CCAACGTTGG with its wide minor groove demonstrates that 2 successive A-T base pairs embedded in a GC tract are insufficient to produce groove narrowing. One cannot predict at present whether 3 consecutive A-T base pairs would suffice.

The comparisons of poly(A) and poly(AT) dodecamer structures, and of the GA and CG decamers, agree in illustrating the ability of cross-strand hydrogen bonds to stabilize large propeller twist in A-T base pairs, whether the stabilizing hydrogen bonds occur in the major or the minor groove. In the absence of such bonds, A-T base pairs in both CGCATATATGCG and CCAACGTTGG adopt smaller propeller twist values, closer to those seen in G-C base pairs. It is attractive to theorize that this difference in propeller twist (and hence in base plane orientations) at junctions between poly(A) and mixed-sequence DNA may contribute to skewed base plane stacking and hence to a bend in helix axis at the junction (24, 25, 38), rather like the model proposed in 1979 for an A-DNA→B-DNA junction (39).

We thank Kazimierz Grzeskowiak for guidance in DNA synthesis, Terri Larsen for crystal-survey photography, and Rajagopal Chattopadhyaya for his involvement in the initial MERLOT fitting of rigid helices. We also thank Hillary C. M. Nelson and Miquel Coll for making available to us information on the groove geometry of their helix structures. This work was supported by the National Science Foundation Grant DMB85-01682. C.Y. also received assistance from the University of California at Los Angeles Medical Scientist Training Program, and G.G.P. was aided by a postgraduate fellowship from the Natural Sciences and Engineering Research Council of Canada.