Crystallographic Analysis of C-C-A-A-G-C-T-T-G-G and Its Implications for Bending in B-DNA†

Kazimierz Grzeskowiak, David S. Goodsell, Maria Kaczor-Grzeskowiak, Duilio Cascio, and Richard E. Dickerson*

Molecular Biology Institute, Department of Chemistry and Biochemistry, and Institute of Geophysics and Planetary Physics, University of California, Los Angeles, California 90024

Received April 14, 1993; Revised Manuscript Received June 2, 1993

ABSTRACT: Stacked B-DNA double helices of sequence C-C-A-A-G-C-T-T-G-G exhibit the same 23° bend at -T-G-G C-C-A- across the nonbonded junction between helices that is observed in the middle of the decamer helix of sequence C-A-T-G-G-C-C-A-T-G, even though the space group (hexagonal vs orthorhombic), crystal packing, and connectedness at the center of the bent segment are quite different. An identical bend occurs across the interhelix junction of every monoclinic crystal structure of sequence C-C-A-x-x-x-x-T-G-G, suggesting that T-G-G-C-C-A constitutes a natural bending element in B-DNA. The bend occurs by rolling stacked base pairs about their long axes; there is no "tilt" component. Of the three possible models for A-tract bending—bent-A-tract, junction bends, or bent-non-A—which cannot be distinguished by solution measurements, all crystallographic evidence over the past 10 years unanimously supports the non-A regions as the actual bending loci.

Intrinsic bending in DNA has been studied extensively by biochemical and solution methods: by anomalous gel migration (Hagerman, 1986; Diekmann, 1992; Crothers & Drak, 1992), by kinetics of circularization (Koo et al., 1990; Crothers et al., 1992), and by phasing of DNA in nucleosomes (Drew & Travers, 1985; Sachwell et al., 1986). From these studies, the repeating sequence motif N_nA_{10-n} has been shown to induce a rigid bend in free DNA, where N typically is guanine or cytosine and, more specifically, where N_n contains the sequence G-G-C. As has been generally recognized (Crothers et al., 1990; Koo et al., 1990), although a bend results from the introduction of poly(A) tracts, these solution methods do not allow one to assign the actual site of bending unequivocally to the poly(A) tract itself, to the non-A region, or to junctions between the two regions.

All crystallographic results to date, including C-G-C-G-A-A-T-T-C-G-C-G (Wing et al., 1980; Drew et al., 1981), C-G-C-A-A-A-A-A-A-G-C-G (Nelson et al., 1987), C-G-C-A-A-A-T-T-T-G-C-G (Coll et al., 1987), and C-G-C-A-A-A-A-A-T-G-C-G (DiGabriele et al., 1989), overwhelmingly favor a model in which the A-tracts are straight and unbent. Where bending has been observed at junctions between GC and AT regions, it occurs at 90° to that which would be demanded by a junction–bend model in order to account for solution and gel measurements (Koo et al., 1986). In contrast, the recent structure of C-A-T-G-G-C-C-A-T-G by Goodsell et al. (1993) reveals a pronounced 23° bend in the central G-G-C-C region. In that structure, as in the GC/AT junction bends observed in the dodecamer crystal structure analyses, bending always occurs by the rolling of base pairs about their long axes. Indeed, Zhurkin et al. (1979) demonstrated the extraordinary energy cost entailed by a tilt wedge that lifts apart the stacked base pairs at one end. There is no shred of crystallographic evidence to suggest that the energetically unfavorable "tilt wedge bend" ever actually occurs in nature.

B-DNA decamers whose structures have been solved to date generally have fallen into one of two classes: a monoclinic *C*2 family with type sequence C-C-A-R-x-x-Y-T-G-G (R = purine, Y = pyrimidine) and an orthorhombic $P2_12_12_1$ family with type sequence C-G-A-T-x-x-A-T-C-G (Grzeskowiak et al., 1991; Privé et al., 1991). In this report we describe the single crystal structure analysis of the B-DNA decamer C-C-A-A-G-C-T-T-G-G, with the sequence matching the earlier monoclinic C-C-A-A-C-G-T-T-G-G of Privé et al. (1991) except for reversal of the two central base pairs.

In hexagonal space group *P*6, the new helix is isomorphous with the methylated decamer C-C-A-G-G-C-m^{5}C-T-G-G (Heinemann & Alings, 1991; Heinemann & Hahn, 1992), whose unmethylated parent sequence (Heinemann & Alings, 1989) falls in the monoclinic family. Hence the familial relationships of this new sequence are no surprise. What is surprising, however, is that the present decamer shows two distinct forms of sequence-directed bending. The helix has a writhed bend over its central six base pairs. In addition, the sequence -T-G-G C-C-A- across the unbonded junction between one stacked helix and the next exhibits a bend identical to that found at the central six base pairs of C-A-T-G-G-C-C-A-T-G, reinforcing (a) the concept of T-G-G-C-C-A as an intrinsically bent structural unit in B-DNA and (b) the idea that base stacking is more important than backbone connections in determining local helix structure (Privé et al., 1991; Diekmann et al., 1992).

METHODS

Crystallization and Data Collection. Crystals were grown by vapor diffusion from standing drops at 5 °C. The droplet solution contained 0.37 mM DNA decamer double helix, 127 mM calcium acetate, and 7.5% v/v MPD (2-methyl-2,4-pentanediol), pH 7.5. Crystals in the shape of hexagonal rods, 0.5 mm long and 0.2–0.3 mm across, appeared after equilibration with a reservoir concentration of 45% v/v MPD.

From survey precession photographs taken at 6 °C, hexagonal space group *P*6 was assigned, with cell dimensions $a = b = 53.08$ Å, $c = 34.32$ Å, and $\gamma = 120°$. These dimensions are quite close to those observed with C-C-A-G-G-C-m^{5}C-T-G-G in the same space group: $a = b = 53.77$ Å and $c =$

† This work was supported by NSF Grant MCB-8916261 and NIH Program Project Grant GM-31299.

* Address correspondence to this author.

0006-2960/93/0432-8923$04.00/0

34.35 Å. With a unit cell volume of 83 741 Å^3 and an assumed one decamer double helix per asymmetric unit, the crystals have 1396 Å^3 per base pair, a figure quite in line with other B-DNA crystals (Dickerson, 1991). Three-dimensional intensity data were collected from one large crystal on an R-Axis II image plate at 5 °C and processed with the accompanying software. The data are of excellent quality, with 4165 structure factor observations to 1.9-Å resolution, of which 4095, or 89% of the possible observable reflections, have a magnitude greater than 2σ in F. The overall R_{merge} is 5.61%.

Molecular Replacement. Strong axial reflections at 3.4 Å along the crystallographic c^* axis, and the 34.32-Å length of that axis, suggested that the helix was B-DNA and was oriented along the real-space c axis. Because of the similar cell dimensions of C-C-A-A-G-C-T-T-G-G and C-C-A-G-G-C-m^{5}C-T-G-G, we began with the assumption that the two structures were isomorphous. Indeed, use of coordinates from the latter structure as obtained from the Brookhaven Protein Data Bank yielded a residual error or R factor of only 33.7% with 8.0–5.0-Å data for the new structure.

To avoid any bias from past determinations, an ideal B-DNA helix of the proper sequence was fitted to the C-C-A-G-G-C-m^{5}C-T-G-G coordinates in order to position the new helix in the unit cell, and this ideal structure became the starting model for refinement. Rigid body refinement in X-PLOR (Brünger et al., 1987) with 1, 2, 3, 10, and 20 rigid bodies in succession led to a model with R = 40.0% for data between 8.0 and 3.0 Å. Data then were added to 1.9-Å resolution in 10 equal-volume shells, with positional refinement via NUCLSQ (Hendrickson & Konnert, 1980; Westhof et al., 1985) at each step, yielding a model with R = 33.9%. A final round of positional and temperature factor (B) refinement with all data out to 1.9-Å resolution reduced this to R = 30.0%.

Water molecules were added in ten steps by inspection of $(2F_o - F_c)$ and $(F_o - F_c)$ maps, which were quite clear. At each step, six to ten waters were added at positions having $(2F_o - F_c)$ density greater than 1σ, $(F_o - F_c)$ density greater than 3σ, and reasonable hydrogen-bonding geometry. Two peaks were recognized immediately as calcium ions, both because of their strength and because of the geometry of surrounding water peaks. These peaks were interpreted as water oxygens at step 1 of the search. At the stage with 27 added water molecules and R = 23.9%, strong positive $(F_o - F_c)$ peaks appeared at the two calcium sites, indicating the need for a heavier atom than oxygen. At this point, the fourth round of water addition, the two calcium sites were identified as such. As more water positions were added, the full heptavalent coordination of these ions became apparent. (One axial water coordination site around calcium complex Ca2 appeared as only a small peak in difference maps, and behaved poorly upon refinement. Its position accordingly was not included in the atom list, although the locations of the other six waters in that complex leave no doubt that the troublesome water actually is present.) After the tenth round of water peak searching from recalculated density and difference density maps, the decamer double helix was accompanied by 79 water molecules and two calcium ions, for a residual error of R = 17.9%.

At this point an $(F_o - F_c)$ omit map was calculated, omitting all waters and ions. Four water peaks failed to return with appreciable density and were deleted. The final structure contains 404 DNA atoms, 2 calcium ions, and 75 water molecules, with R = 18.1% for the 4095 structure factor observations greater than $2\sigma(F)$, between 8.0- and 1.9-Å resolution. Root mean square (rms) deviations from ideal geometry, as obtained from NUCLSQ, are listed in Table I. Both the F_o intensity data and the atomic coordinates have been deposited with the Brookhaven Protein Data Bank, for immediate release.

Table I: Root Mean Square Deviations of Final Structure from Ideal Geometry and Standard Deviations

	rms	σ
bonds (Å)	0.016	0.030
distance across angles (deg)	0.034	0.040
P bonds (Å)	0.024	0.025
distance across P angles (deg)	0.042	0.050
planes (Å)	0.010	0.020
chirals	0.100	0.150

Table II: Identification Key to Decamer Helices Mentioned Frequently

name	sequence	crystal form
CG	C-C-A-A-C-G-T-T-G-G	monoclinic *C*2
GC	C-C-A-A-G-C-T-T-G-G	hexagonal *P*6
HA	C-C-A-G-G-C-C-T-G-G	monoclinic *C*2
HAMe	C-C-A-G-G-C-m^{5}C-T-G-G	hexagonal *P*6
CAT	C-A-T-G-G-C-C-A-T-G	orthorhombic $P2_12_12_1$
KK	C-G-A-T-C-G-A-T-C-G	orthorhombic $P2_12_12_1$

RESULTS AND DISCUSSION

For ease of reference in what follows, the sequences discussed will be identified by abbreviated names as listed in Table II. These names, although containing an element of arbitrariness and artificiality, are more easily recognized as labels than are endless repetitions of full sequences.

Crystal Packing and Minor Groove Hydration. The GC (or C-C-A-A-G-C-T-T-G-G) decamer is held in its *P*6 lattice by three distinct types of lattice contacts. The most intimate of the three is the stacking of helices to form pseudocontinuous columns parallel to the crystallographic c axis. The base step from the top of one helix to the bottom of the next is normal in every respect, except for the absence of connecting phosphate backbones. Helical twist at the junction step is 41.3°, roll is +4.7°, and rise is 2.66 Å, all within the ranges seen in normal base steps within a helix. [The rise can be less than the 3.4-Å thickness of one base pair, of course, because rise is measured from Cl′ atoms and because the two base pairs are buckled slightly so as to exhibit a value for cup of 16.5°. See Figure 6 of Yanagi et al. (1991).]

In the second type of lattice contact, six columns of helices pack in a hexagonal array around the 6-fold rotation axis to form a hollow tube, as seen end-on in Figure 1 and in a side view in Figure 2. Close interhelix contacts of this sort are listed in Table IIIA,B. Phosphates P5 and P6, flanking base G5, contact phosphates P16 and P15 of a neighboring helix around the hexagonal tube (middle contacts in Figure 2). These contacts would serve to narrow the width of the minor groove at the center of the helix, and indeed this packing may be responsible for the small relative narrowing of groove width in hexagonal HAMe (or C-C-A-G-G-C-m^{5}C-T-G-G) by comparison with monoclinic HA (or C-C-A-G-G-C-C-T-G-G).

Five base pairs up or down from this contact, at the junction between stacked helices, more elaborate interactions involve bridging via two minor groove heptaaquocalcium complexes (Figure 3). Calcium water oxygens O3, O4, and O6 interact with guanine N2 and N3, cytosine O2, and deoxyribose O4′ atoms of the helix in whose minor groove the complex sits.

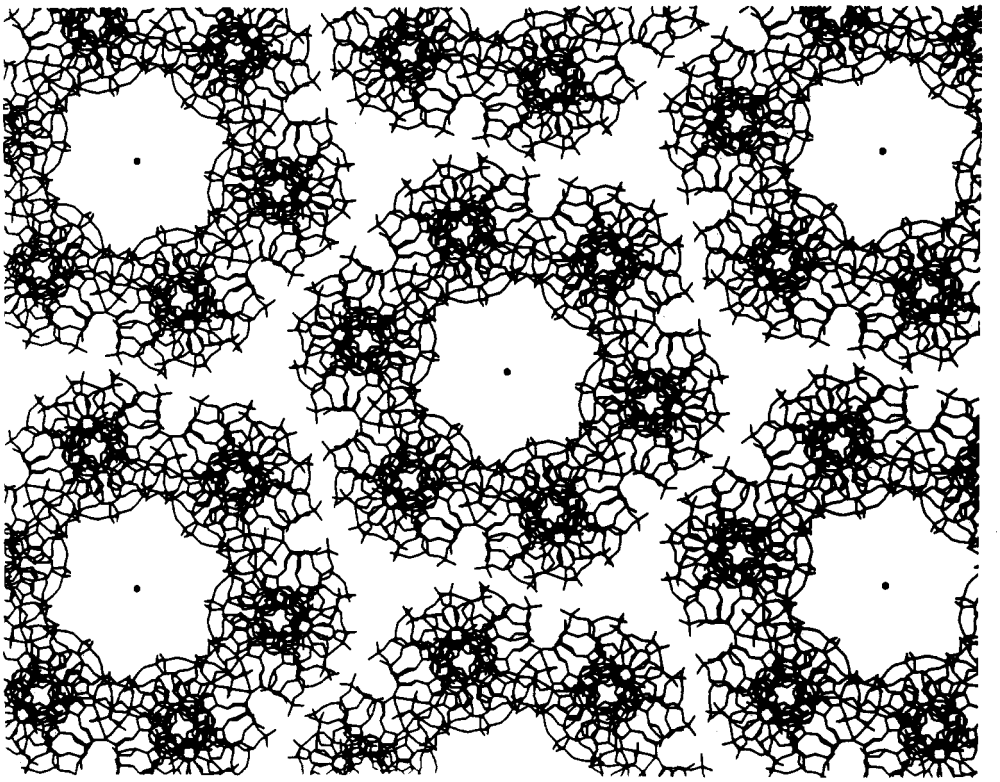

FIGURE 1: Packing of columns of stacked decamer helices, viewed down the *c* axis, the axis of columns. One ring of six columns, in isolation, builds a hollow tube that exhibits almost ideal 622 symmetry. These tubes then are packed in a triangular lattice to build the crystal.

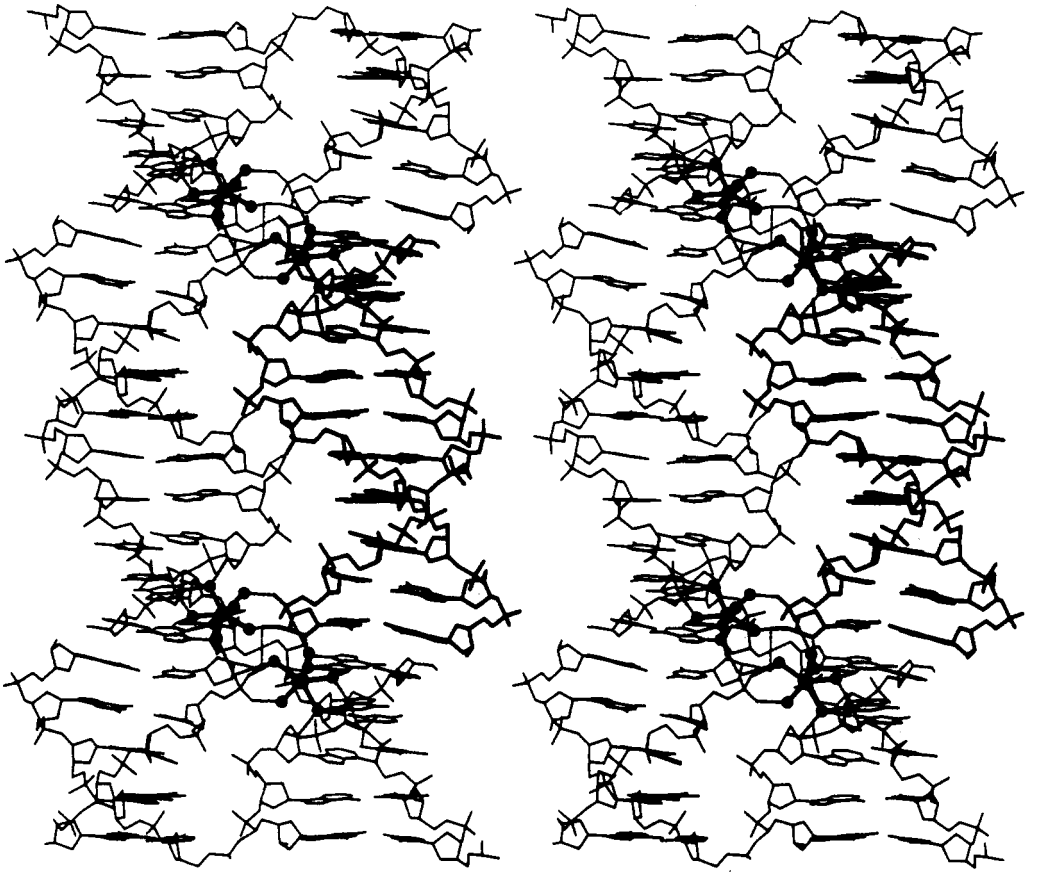

FIGURE 2: Stereoview of lateral interhelix contacts of 3.5 Å or less, as listed in Table III. Contacts are between neighbors around a 6-fold tube, including bridging by calcium complexes. Complex Ca1 is to the upper left of C2 in each pair.

Water oxygens O1 and O2 reach across to a 5′-terminal –OH and phosphate oxygen of the next helix around the 6-fold axis, and oxygen O5 makes a bond across the interhelix junction between two stacked helices. As with the monoclinic, orthorhombic, and trigonal decamers, backbone chains in two adjacent helical columns come into closest contact precisely

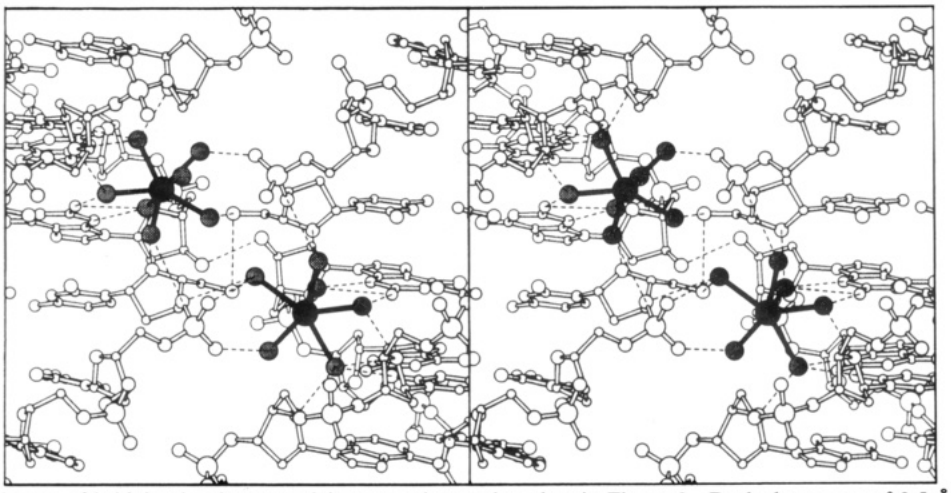

FIGURE 3: Stereo closeup of bridging by the two calcium complexes, viewed as in Figure 2. Dashed contacts of 3.5 Å or less are listed in Table IIIA. Complex Ca1 is to the upper left of Ca2. The five coplanar oxygens of complex Ca1 are numbered 1–5 in a counterclockwise direction starting from the lower right, and equivalent oxygens in Ca2 are numbered counterclockwise from the upper left of that complex. In each cluster, axial ligand O6 is behind the plane of the five. Axial ligand O7 is shown in complex Ca1 but is not drawn in complex Ca2, even though the ligand must be present. A small peak was observed in Fourier and difference Fourier maps at the appropriate position, but it refined poorly and was omitted from the atom list.

Table III: Close Contacts of 3.50 Å or Less between Columns of Helices[a]

(A) Helix N and O Atoms Bridged via Calcium Complexes (See Figure 3)

Ca no.	O no.	sym	base	atom	dist (Å)	Ca no.	O no.	sym	base	atom	dist (Å)
Ca1	O1	*n*	C1	O5′	2.95	Ca2	O1	*n*	C11	O5′	2.92
						Ca2	O1	*n*	C11	O3′	3.41
						Ca2	O1	*n*	C12	O2P	3.42
Ca1	O2	*n*	C2	O1P	2.88	Ca2	O2	*n*	C12	O1P	2.74
Ca1	O3	*p*	C2	O2	2.83	Ca2	O3	*p*	C12	O2	2.68
Ca1	O3	*p*	C2	O4′	3.48	Ca2	O3	*p*	C12	O4′	3.31
Ca1	O3	*p*	A3	O4′	2.72	Ca2	O3	*p*	A13	O4′	2.84
Ca1	O4	*p*	C2	O4′	2.82	Ca2	O4	*p*	C12	O4′	3.07
Ca1	O4	*p*	G20	N2	3.28	Ca2	O4	*p*	G10	N2	3.20
Ca1	O5	+*c*	C11	O3′	3.41	Ca2	O5	−*c*	C1	O3′	3.24
Ca1	O6	+*c*	C11	O4′	2.74	Ca2	O6	−*c*	C1	O4′	2.57
Ca1	O6	*p*	G20	N3	3.18	Ca2	O6	*p*	G10	N3	3.21
Ca1	O6	*p*	G20	N2	3.12	Ca2	O6	*p*	G10	N2	3.31

(B) Direct Oxygen Bridging between DNA Helices Related by the 6-fold Symmetry Axis[a]

base	atom	sym	base	atom	dist (Å)	best illustrated
C1	O5′	+*c*	C11	O5′	3.44	Figure 3
G10	O3′	−*c*	G20	O3′	2.88	Figure 3
G5	O5′	*n*	C16	O1P	3.34	Figure 2, center
C6	O1P	*n*	G15	O5′	3.62	Figure 2, center

(C) Direct Oxygen Bridging between DNA Helices Related by the 2-fold Symmetry Axis

base	atom	base	atom	dist (Å)	best illustrated
G9	O1P	G10	O2P	3.44	Figure 4
G9	O1P	G9	O3′	3.36	Figure 4
G9	O1P	G9	O1P	4.03	Figure 4

[a] In the symmetry column, *p* designates a particular helix, *n* is one of its neighbors to either side around the 6-fold axis, and +*c* and −*c* indicate unit cell displacements up or down the *c* axis from the *p* helix. Quasi-equivalent interactions involving the two different calcium clusters are on the same row. Oxygen atoms O6 and O7 are axial to the plane of the other five calcium ligands, O1–O5. Atom O7 is uncoordinated in both calcium complexes.

where phosphates are absent at interhelix junctions. Hence the gaps between helices themselves contribute to closer packing. At each junction, visible in the center of Figure 3, two 5′-terminal –OH are hydrogen-bonded to one another, as are two 3′-terminal –OH.

The third, and most tenuous, type of interhelix contact is formed when the hexagonal tubes of six helices are arranged into the triangular crystal lattice, as seen in Figures 1 and 4. Each helix makes a single contact across a crystallographic 2-fold axis, involving only phosphates P9 and P10 on both of the contacting helices.

The packing into 6-fold tubes preserves an intrinsic 2-fold symmetry perpendicular to the 6-fold axis through the center of the decamer that is not demanded by the *P6* space group. This noncrystallographic symmetry is broken only by contacts across a 2-fold axis when the 6-fold tubes are packed into a crystal lattice (Figure 4). Those contacts are so weak that the rms difference in atom positions between two ends of the helix is only 0.28 Å, approaching the accuracy of the structure analysis itself. As expected, most of the asymmetry occurs at phosphate positions P9 and P10, flanking base G9. The two calcium complexes near the interhelix junctions occupy quasi-symmetrical positions, and even the water structure exhibits strong 2-fold symmetry.

Two interpretations of cause and effect might be invoked here. One might argue that the decamer helices are themselves rigidly symmetrical end for end and must therefore pack into the crystal lattice in a manner that preserves this symmetry approximately. Alternatively, one could maintain that the packing constraints of the 6-fold rotation axis enforce an end-for-end symmetry. But the observation that helices pack in a manner that aligns their molecular dyads to form approximate 622-point symmetry, instead of a different rotation of the helix around its local axis, argues for an intrinsic 2-fold symmetry of the helix. A flexible and easily deformed helix could have selected any of 10 different local axis rotations, aligning phosphates in similar positions but perhaps finding an orientation with more suitable packing. The highly localized consequences of the 2-fold contacts between helices—affecting only the two phosphate groups but not the sugars or bases attached to them—also argues for an intrinsically rigid helix.

It is interesting to note that, among all the available self-complementary B-DNA crystal structure analyses, those crystals exhibiting highest resolution are the ones that minimally perturb the intrinsic 2-fold internal symmetry of the molecule. Highest resolution, 1.3–1.6 Å, is obtained in monoclinic space group *C*2, where the molecular symmetry

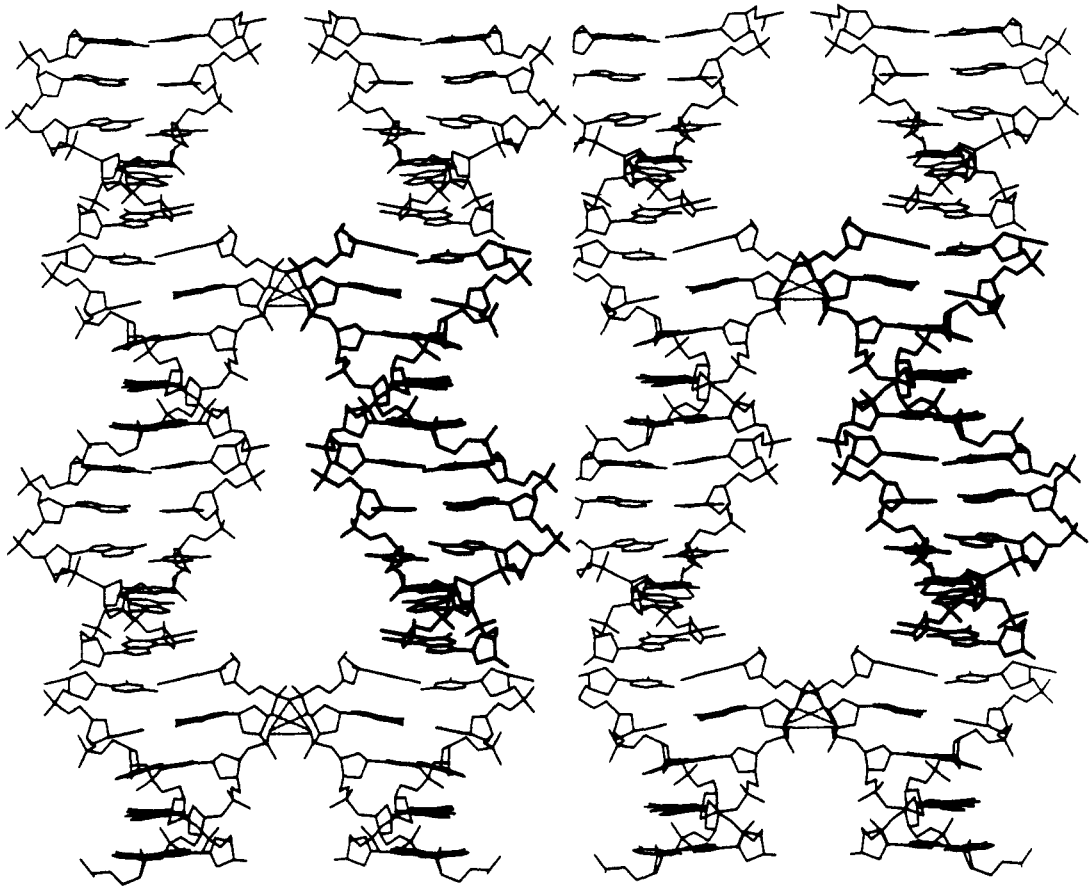

FIGURE 4: Stereoview of contacts of 3.5 Å or less across two helices related by a vertical 2-fold axis. The *c* axis is vertical, and one decamer is emphasized with dark bonds. Base pairs from the top of each column as drawn here are G5·C16, A4·T17, A3·T18, C2·G19, C1·G20, interhelix break, G10·C11, G9·C12, ..., etc.

coincides with a true crystallographic 2-fold axis. The hexagonal *P*6 and orthorhombic $P2_12_12_1$ decamers possess a small number of asymmetric packing contacts, and their crystals diffract somewhat less strongly, 1.5–2.0 Å. The trigonal decamers and orthorhombic dodecamers possess extensive asymmetric packing contacts, show significant deviations from internal 2-fold symmetry, and diffract least well, 1.9–2.6 Å. The DNA has a strong tendency to exhibit the end-for-end symmetry inherent in its base sequence. Where this is forbidden by anisotropic crystal packing, a price is paid in terms of increased disorder within the crystal structure and lower resolution.

Hydration in the new structure (Figure 5) generally resembles that encountered in all other decamers, no matter what their space group or crystal packing mode: a monolayer of water molecules hydrogen-bonded to polar groups along the phosphate backbone and within the major groove, without apparent superstructure, and more structured hydration within the minor groove. In the narrow minor groove at the middle of the helix is found a single zigzag spine of hydration familiar from CG (or C-C-A-A-C-G-T-T-G-G) and other decamers; in the wider regions of the minor groove toward the ends of the molecule the spine breaks up into two irregular ribbons of hydration along the two walls of the groove. Such behavior by now is so familiar that it can be considered as a characteristic property of B-DNA, to be expected whenever the resolution of the structure analysis is sufficient to permit accurate localization of water positions.

Comparison of GC and HAMe Decamer Structures. The GC and HAMe decamers are essentially identical, both in chemical properties and in crystal structure. In the major groove, both exhibit a 5-methyl group on the pyrimidine at position 7 (5-methylcytosine vs thymine); only the direction of the N–H···O hydrogen bond is reversed. Local patterns of hydrophobicity and hydrophilicity are the same. Hence the adoption of the same crystal form by the two sequences is not surprising. Crystallographically, the two independently crystallized and studied oligomers exhibit a rms difference in atomic positions of only 0.44 Å (Figure 6).

The two structures, GC and HAMe, differ mainly in the location of ions. HAMe was crystallized with magnesium ions, which locate more toward the center of the helix, linking base edges of A3 and A4 to phosphate P17 on the neighboring helix around the 6-fold axis. In contrast, the calcium complexes in GC link base edges of G20 and C2 to phosphate P2 and the nearby 5′-terminal –OH. The large calcium complex occupies a wider region of the minor groove than does the smaller magnesium complex, exactly as was observed by Lipanov et al. (1993) in the two crystal forms of C-C-A-A-C-I-T-T-G-G. In spite of this difference in bridging ion locations between GC and HAMe, no difference is seen in the local helix parameters of the bases at each position. Obviously, the ions serve more as mortar to fill gaps between helices (at least in this space group) than as active agents shaping local helix parameters and forcing crystallization in a given space group.

Bending. Like the earlier HA and HAMe structures, the present GC helix shows a substantial bend across its center, produced by rolling base pairs about their long axes in a direction that compresses the broad major groove. This is seen by comparing inclinations of base pairs 3 and 8 in Figure 6, and by observing the motion of points 3–4–5–6–7–8 across

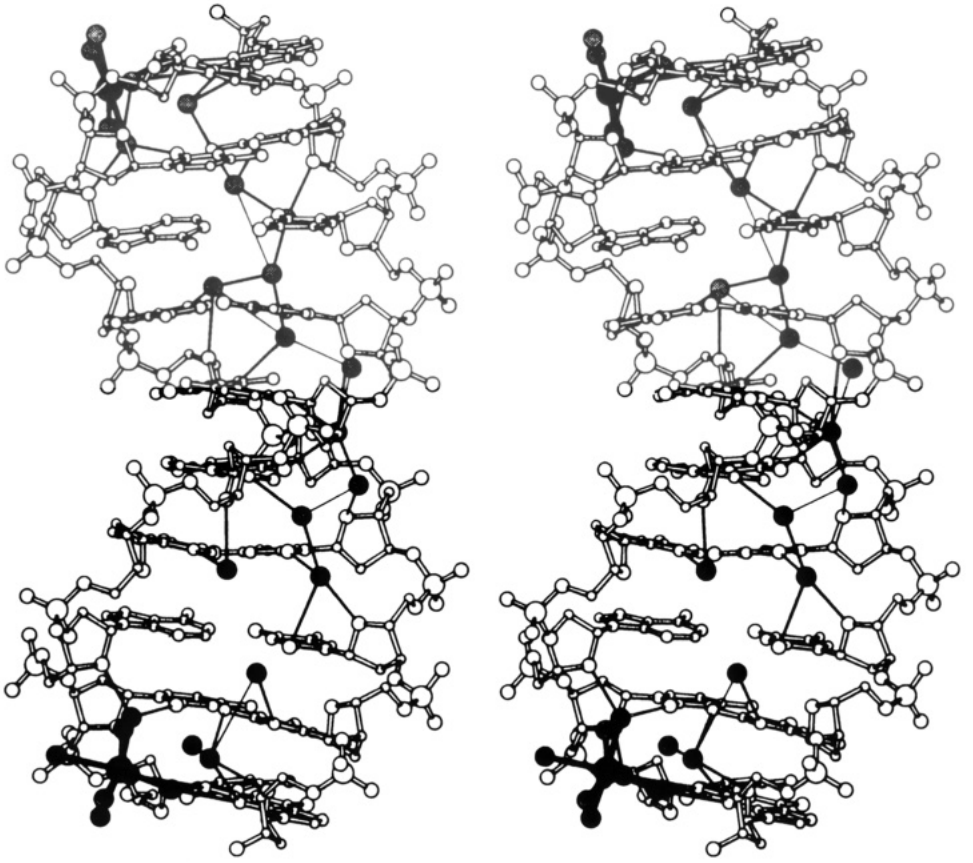

FIGURE 5: Minor groove hydration in C-C-A-A-G-C-T-T-G-G. Base pair C1·G20 is at the bottom and C11·G12 is at the top. Water molecules within the minor groove are represented by dark spheres connected by single lines: thick for distances less than 3.5 Å and thin for distances between 3.5 and 4.0 Å. Notice the symmetry of the water structure around the center of the helix and the single spine of hydration there, fraying into a double strand at both ends of the helix.

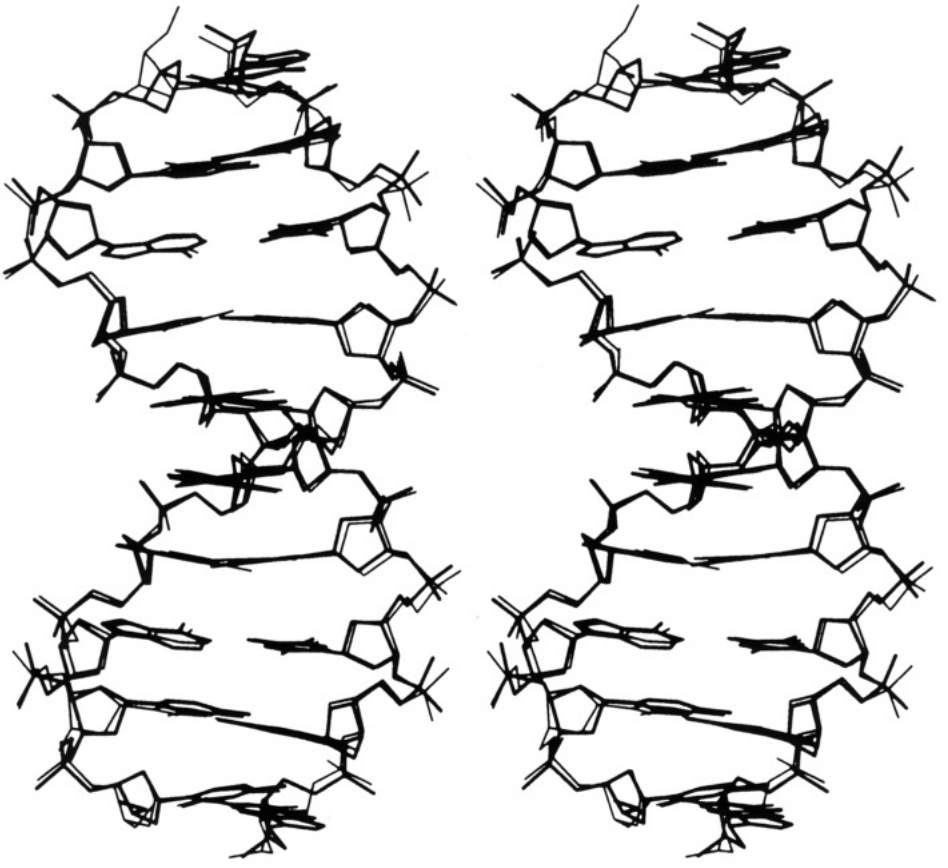

FIGURE 6: Structure comparisons by superposition of C-C-A-A-G-C-T-T-G-G (dark bonds) and C-C-A-G-G-C-m^{5}C-T-G-G (light bonds). The rms difference in the two structures, which have been refined independently in different laboratories, is 0.44 Å.

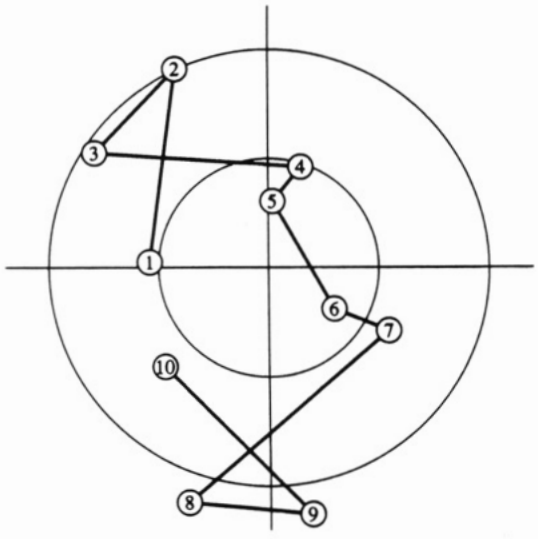

FIGURE 7: Base pair normal vector plot for C-C-A-A-G-C-T-T-G-G. Numbered points 1–10 mark the tips of normal vectors to base pairs 1–10, shifted to a common origin and viewed down the helix axis. A progression of points across the diagram indicates a bend in the helix axis. Inner and outer circles represent 5° and 10° bends, respectively. The GC helix shows a 15° writhing bend over base pairs 3–8, seen here as a clockwise trajectory of points. Even more striking is the nearly planar bend of 23° involving base pairs 9–10–1–2 across the helix junction. This bend across the nonbonded junction also is observed in all of the monoclinic and hexagonal decamers that have been solved to date, all with a cross-junction sequence of -T-G-G C-C-A-.

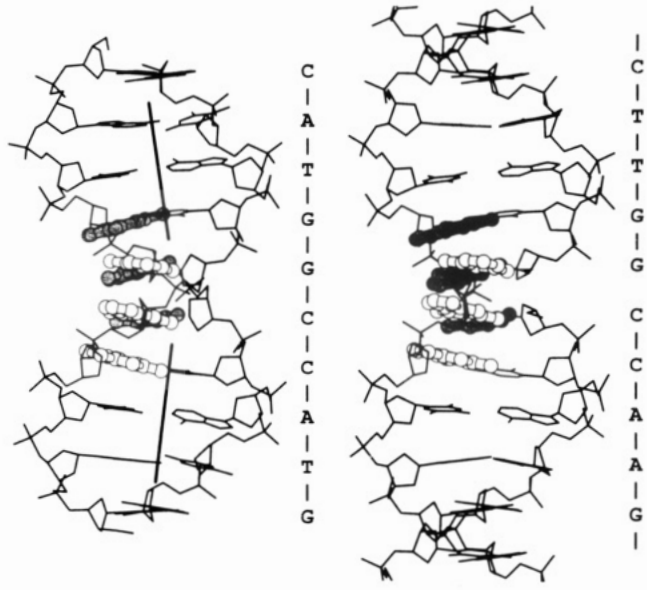

FIGURE 8: Comparison of the bend at -T-G-G-C-C-A- in the center of C-A-T-G-G-C-C-A-T-G (left) and across the interhelix junction of C-C-A-A-G-C-T-T-G-G (right). In the four central base pairs, G-G-C of the nearer strand are shown in open space-filling representation, and G-G-C of the farther strand are in dark space-filling atoms. The bend is essentially identical in the two structures, with a rms difference of only 0.61 Å in atomic position over the central six base pairs. Bending is produced by preferential stacking of guanine bases, as described earlier by Goodsell et al. (1993).

the base pair normal vector plot in Figure 7. The curvature of this latter trajectory indicates that the bend is writhed rather than planar.

But the true significance of this bend in HA and HAMe only became apparent after structure analysis of the CAT helix, in which the central bend is larger, 23°, and is confined to a single bending plane [Figures 1 and 2a of Goodsell et al. (1993)]. In the CAT helix, bending is produced by the motif Y-G-G-C-C-R: the G-G-C-C core induces a bend in order to improve the stacking of G upon G, and the flanking Y-G and C-R steps allow the helix to resume a normal course by breaking the bend via unusually large twist values, 48.8° and 49.6°.

But the central A-G-G-C-C-T in the HA and HAMe helices is a longer R_3Y_3 motif rather than R_2Y_2, and the A-upon-G purine–purine stacking disfavors a break where found in CAT. Instead, the break occurs one step farther toward each end of the helix, at pyrimidine–purine Y-A and T-R steps separated by six base pairs: Y-A-G-G-C-C-T-R. With this longer purine span, the magnitude of the bend must be reduced correspondingly if continuity of the helix is to be preserved. But the ends of the bent region, at steps Y-A and T-R, again are characterized by very high 44°–51° twist angles. The present GC helix also displays the longer pattern Y-R-R-R-Y-Y-Y-R across its center, and the bend described by the central base pairs A4-G5-C6-T7 in Figure 7 is only 5°. The stacking of three successive purines again apparently is sufficiently strong to damp down bending.

A surprise is found, however, at the ends of the helix. Two stacked GC helices exhibit the sequence

C-C-A-A-G-C-T-T-G-G C-C-A-A-G-C-T-T-G-G

The underlined region has a sequence identical to the central six base pair bending motif of the CAT helix, even though the two central base pairs in the motif now are not connected by phosphate groups. And the bend encountered over the central four G-G-C-C base pairs is an identical 23° in CAT and GC structures, corresponding to a 25-Å radius of curvature. For GC this bend is visible as the trajectory of points 9–10–1–2 in Figure 7 and for CAT as points 4–5–6–7 of Figure 2a of Goodsell et al. (1993). Even more remarkably, the T-G-G-C-C-A regions in the two decamers are superimposable with a rms difference of only 0.61 Å (Figure 8). The absence of connecting phosphate backbones in the latter case apparently is irrelevant, by comparison with the identity of base stacking.

Hence in each structure, GC and CAT, columns of stacked helices display two bends—one at the center of a decamer and the other at the interhelix junction. Both the bend and a compensating reverse bend are necessary, of course, if helices are to be stacked vertically in endless columns. Bends at the G-G-C-C of sequence T-G-G-C-C-A are well defined, planar, and identical in the two structures, whereas the compensating reverse bends, at A-A-G-C-T-T in the GC helix and A-T-G C-A-T in CAT, are more diffuse. As an example, the GC helix bend along base pairs 9–10–1–2 in Figure 7 occurs in one plane and over only three base steps. The compensating return bend from base pair 2 back to base pair 9 again takes seven helical steps and includes local writhing motions.

Examination of other decamer structure analyses reveals that this bend at the sequence Y-G-G-C-C-R is one of the most ubiquitous of all features of B-DNA oligomers. The T-G-G C-C-A bend across an interhelix junction is encountered in every decamer that adopts monoclinic space group *C*2 or hexagonal space group *P*6. Table IV compares T-G-G-C-C-A steps in the center of orthorhombic CAT and across the junctions of four of the six decamers that adopt hexagonal or monoclinic crystal packing. A common pattern is observed: low twist, essentially zero slide, and positive roll at the three steps within the G-G-C-C core and sharp breaks in behavior with high twist, large slide, and negative roll at flanking T-G and C-A steps. The combination of high twist and large positive slide completely unstacks bases in the two base pairs, marking the end of the bent region.

Table IV: Twist, Slide, and Roll Profiles of Bent Sequences of the Type Y-R-R-Y-Y-R[a]

CAT (orth)	A	T	G	G	C	C	A	T
twist (deg)	30.7	48.8	31.4	30.8	30.3	49.6	27.7	
slide (Å)	0.30	2.43	0.72	–0.03	0.72	2.72	–0.08	
roll (deg)	–0.62	–5.84	6.23	9.26	7.93	–7.14	2.68	
GC (hex)	T	T	G	G |	C	C	A	A
twist (deg)	26.4	49.2	28.3	41.3	29.7	49.3	24.2	
slide (Å)	0.70	2.40	0.79	–0.23	0.72	2.39	0.96	
roll (deg)	12.06	–5.54	9.20	4.71	8.86	–5.21	9.41	
CG (mono)	T	T	G	G |	C	C	A	A
twist (deg)	28.7	50.8	27.8	40.5	27.8	50.8	28.7	
slide (Å)	0.47	2.59	0.73	–0.01	0.73	2.59	0.47	
roll (deg)	8.80	–6.15	5.13	3.30	5.13	–6.15	8.80	
HAMe (hex)	m^{5}C	T	G	G |	C	C	A	G
twist (deg)	28.4	47.6	29.9	40.5	32.8	44.0	26.2	
slide (Å)	0.74	1.92	0.39	–0.29	0.76	2.07	0.87	
roll (deg)	3.25	–3.76	8.41	3.65	5.04	–2.45	5.88	
HA (mono)	C	T	G	G |	C	C	A	G
twist (deg)	23.8	50.6	28.2	40.2	28.2	50.6	23.8	
slide (Å)	0.94	2.63	0.74	–0.15	0.74	2.64	0.94	
roll (deg)	5.55	–7.04	6.86	3.45	6.86	–7.04	5.55	
KK (orth)	G |	C	G	A	T	C	G	A
twist (deg)	25.8	39.9	40.6	33.6	38.5	29.3	40.0	
slide (Å)	0.12	0.84	–0.36	–0.45	–0.15	0.60	0.57	
roll (deg)	2.04	–0.41	2.09	–2.40	–0.75	8.47	0.14	

[a] Local helix parameters are calculated with the helix analysis program NEWHEL93, which is obtainable upon request either from the author (E-mail address RED@UCLAUE.MBI.UCLA.EDU) or from the Brookhaven Protein Data Bank. The quantity roll is the simple component of the change in base pair normal vector direction in a plane perpendicular to the long base pair axis, rather than the adjusted radj quantity in which the twist angle between base pairs is first rotated back to zero before roll calculation. mono = monoclinic space group $C2$; hex = hexagonal space group $P6$; orth = orthorhombic space group $P2_12_12_1$; | = junction between decamer helices.

Table V: Twist, Slide, and Roll Profiles of Sequences of the Type R-R-R-Y-Y-Y and R-R-Y-R-Y-Y

GC (hex)[a]	C	A	A	G	C	T	T	G
twist (deg)	49.3	24.2	37.4	37.3	36.9	26.4	49.2	
slide (Å)	2.39	0.96	0.01	–1.03	–0.12	0.70	2.40	
roll (deg)	–5.21	9.41	–0.26	5.61	0.10	12.06	–5.54	
CG (mono)[a]	C	A	A	C	G	T	T	G
twist (deg)	50.8	28.7	29.9	44.8	29.9	28.7	50.8	
slide (Å)	2.59	0.47	–0.28	–0.06	–0.28	0.47	2.59	
roll (deg)	–6.15	8.80	–1.99	6.35	–1.99	8.80	–6.15	
HAMe (hex)	C	A	G	G	C	m^{5}C	T	G
twist (deg)	44.0	26.2	37.3	36.9	36.5	28.4	47.6	
slide (Å)	2.07	0.87	–0.11	–0.79	–0.07	0.74	1.92	
roll (deg)	–2.45	5.88	4.14	4.21	3.67	3.25	–3.76	
HA (mono)	C	A	G	G	C	C	T	G
twist (deg)	50.6	23.8	36.9	40.9	36.9	23.8	50.6	
slide (Å)	2.63	0.94	0.74	0.02	0.74	0.94	2.63	
roll (deg)	–7.04	5.55	4.06	1.44	4.06	5.55	–7.04	

[a] mono = monoclinic space group $C2$; hex = hexagonal space group $P6$.

As a control, the KK helix (Grzekowiak et al., 1991) in Table IV demonstrates that this pattern of behavior is not observed for the core sequence G-A-T-C, even though it also exhibits a Y-R-R-Y-Y-R motif. Apparently the stacking of G's and A's, in G-G-C-C vs G-A-T-C, is quite different. This characteristic bending behavior of the sequence Y-G-G-C-C-R cannot be attributed to crystal packing, because it is observed in three different space groups with quite different local molecular environments.

Table V lists the geometry of the compensating return bends that are necessary to ensure zero net bend in the helix, thus facilitating the stacking of helices in columns. All but one of these sequences are of the general type Y-R-R-R-Y-Y-Y-R, but purines vary between G and A. The Y-R and R-Y flanking steps again represent a high twist, positive slide, negative roll, break in base stacking, but no common or characteristic structural pattern is seen within the R_3Y_3 regions.

The sequence C-C-A-A-C-I-T-T-G-G has been solved in two different crystal environments: monoclinic $C2$ in the presence of calcium ions and trigonal $P3_221$ with magnesium ions (Lipanov et al., 1993). As mentioned earlier, magnesium complexes locate in the narrow central region of the minor groove, whereas calcium complexes bind one base pair farther toward each end of the helix, where the minor groove is wider. The characteristic Y-G-G-C-C-R bend is seen across the interhelix junction in the monoclinic structure but not in the trigonal. This can be explained by the restrictive crystal packing imposed by helix crossing in the trigonal lattice, where

the phosphate backbone of one helix is hydrogen-bonded to the floor of the major groove of another, and the base step from one helix to the next has an anomalously high 51° twist angle.

In summary, the evidence from at least seven decamer crystal structures, in hexagonal, monoclinic, and orthorhombic space groups, suggests that the sequence T-G-G-C-C-A has a pronounced tendency to bend. In contrast, all crystallographic evidence indicates that A-tracts are not intrinsically bent and that bending observed at junctions between GC and AT regions is 90° out of phase from that required for a junction–bend explanation of gel migration and nucleosome phasing experiments. Hence Occam's Razor would suggest that A-tracts induce bending in general-sequence B-DNA, not because the A-tracts themselves are bent but because they remove a certain portion of the natural writhe and flexibility of general-sequence B-DNA, allowing the natural bending tendencies of the remaining GC-rich regions to find macroscopic expression.

ACKNOWLEDGMENT

We thank David Lilley, Andrew Travers, Hilary Nelson, Don Crothers, and Mary Kopka for helpful discussions of B-DNA bending and possible bending models.

REFERENCES

Baikalov, I., Grzeskowiak, K., Yanagi, K., Quintana, J., & Dickerson, R. E. (1993) *J. Mol. Biol. 231*, 768–784.

Brunger, A. T., Kuriyan, J., & Karplus, M. (1987) *Science 235*, 458–460.

Coll, M., Frederick, C. A., Wang, A. H.-J., & Rich, A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8385–8389.

Crothers, D. M., & Drak, J. (1992) *Methods Enzymol. 212*, 46–71.

Crothers, D. M., Haran, T. E., & Nadeau, J. G. (1990) *J. Biol. Chem. 265*, 7093–7096.

Crothers, D. M., Drak, J., Kahn, J. D., & Levene, S. D. (1992) *Methods Enzymol. 212*, 3–29.

Dickerson, R. E. (1991) in *Methods in Enzymology*: *DNA Structures* (Lilley, D. M. J., & Dahlberg, J., Eds.) Vol. 211, pp 67–111, Academic Press, New York.

Diekmann, S. (1992) *Methods Enzymol. 212*, 30–46.

Diekmann, S., Mazzarelli, J. M., McLaughlin, L. W., von Kitzing, E., & Travers, A. A. (1992) *J. Mol. Biol. 225*, 729–738.

DiGabriele, A. D., Sanderson, M. R., & Steitz, T. A. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 1816–1820.

Drew, H. R., & Travers, A. A. (1985) *J. Mol. Biol. 186*, 773–790.

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2179–2183.

Goodsell, D. S., Kopka, M. L., Cascio, D., & Dickerson, R. E. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 2930–2934.

Grzeskowiak, K., Yanagi, K., Privé, G. G., & Dickerson, R. E. (1991) *J. Biol. Chem. 286*, 8861–8883.

Hagerman, P. J. (1986) *Nature 321*, 449–450.

Heinemann U., & Alings C. (1989) *J. Mol. Biol. 210*, 369–381.

Heinemann U., & Alings, C. (1991) *EMBO J. 10*, 35–43.

Heinemann U., & Hahn, M. (1992) *J. Biol. Chem. 267*, 7332–7341.

Hendrickson, W. A., & Konnert, J. H. (1980) in *Computing in Crystallography* (Diamond, R., Ramaseshan, S., & Venkatesan, K., Eds.) pp 13.01–13.23, Indian Academy of Sciences, Bangalore.

Koo, H.-S., Wu, H.-M., & Crothers, D. M. (1986) *Nature 320*, 501–506.

Koo, H.-S., Drak, J., Rice, J. A., & Crothers, D. M. (1990) *Biochemistry 29*, 4227–4234.

Lipanov, A., Kopka, M. L., Kaczor-Grzeskowiak, M., & Dickerson, R. E. (1993) *Biochemistry 32*, 1373–1389.

Nelson, H. C. M., Finch, J. T., Luisi, B. F., & Klug, A. (1987) *Nature 330*, 221–226.

Privé, G. G., Yanagi, K., & Dickerson, R. E. (1991) *J. Mol. Biol. 217*, 177–199.

Sachwell, S. C., Drew, H. R., & Travers, A. A. (1986) *J. Mol. Biol. 191*, 659–675.

Travers, A. A., & Klug, A. (1990) in *DNA Topology and Its Biological Effects* (Cozzarelli, N. R., & Wang, J. C., Eds.) pp 57–106, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 166*, 119–145.

Wing, R. M., Drew, H. R., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980) *Nature 287*, 755–758.

Yanagi, K., Privé, G. G., & Dickerson, R. E. (1991) *J. Mol. Biol. 217*, 201–214.

Zhurkin, V. B., Lysov, Y. P., & Ivanov, V. I. (1979) *Nucleic Acids Res. 6*, 1081–1096.