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†

Andrei Lipanov,‡ Mary L. Kopka, Maria Kaczor-Grzeskowiak, Jordi Quintana, and Richard E. Dickerson*

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

Received August 7, 1992; Revised Manuscript Received November 5, 1992

ABSTRACT: For the first time, the same B-DNA oligomer has been crystallized and its structure solved in two different space groups. Crystallization of C-C-A-A-C-I-T-T-G-G with Ca^{2+} yields monoclinic space group $C2$ with a = 31.87 Å, b = 25.69 Å, c = 34.21 Å, β = 114.1°, and five base pairs per asymmetric unit. The 5026 2σ data to 1.3 Å refine to R = 0.152 with 72 waters, one heptavalent hydrated calcium complex, and one cacodylate ion per asymmetric unit. In contrast, crystallization with Mg^{2+} yields trigonal space group $P3_221$ with $a = b$ = 33.23 Å, c = 94.77 Å, γ = 120°, and 10 base pairs per asymmetric unit. The 1725 2σ data to 2.2 Å refine to R = 0.164 with 36 water molecules and one octahedral magnesium complex per asymmetric unit. The monoclinic form is virtually isostructural with previously solved monoclinic decamers, including twist angles of ca. 50° at C-A and T-G steps. In contrast, the trigonal structure has quite different local helix parameters, with twist angles of ca. 36° at the corresponding steps. These local parameter differences can only be attributed to crystal packing, suggesting that certain sequences of B-DNA are more flexible and influenced by their surroundings than had previously been thought. Such deformability may be important for interaction of B-DNA with control proteins, where both static structure and dynamic deformability comprise components of the recognition process. The crossing of two helices at an angle of 120° in the trigonal cell is a model for an antiparallel, uncrossed Holliday junction, as has been noted earlier by Timsit and Moras [Timsit, Y., & Moras, D. (1991) *J. Mol. Biol. 221*, 919–940] from a rhombohedral DNA dodecamer structure analysis.

Synthetic DNA oligomers in the physiologically relevant B conformation have been observed in several crystal forms, primarily orthorhombic dodecamers and monoclinic and orthorhombic decamers (Dickerson, 1990, 1992). No dependable correlation has been found between simple two-base-pair sequences and local helix parameters, and it has been suggested that at least a tetrad of four base pairs is necessary to define local conformation (Yanagi et al., 1991). This tetrad hypothesis is complicated by the fact that only ca. 36 of the 136 possible tetrads have yet been examined in crystal structures and by the possibility that local helix structure may be affected by crystal packing forces. Unscrambling of these effects requires examination both of different sequences in the same crystal environment and of the same sequence in different environments.

Distinct sequence preferences or "signatures" have been observed in all three of the principal oligomer crystal forms (Baikalov et al., 1993, Table I). To date, dodecamers have been found in the orthorhombic $P2_12_12_1$ space group only for sequences of the type C-G-Y-x-x-x-x-x-R-C-G (Y = pyrimidine; R = purine), where the outermost two base pairs at each end provide the overlap that stabilizes columns of helices through the crystal (Dickerson et al., 1987). Among decamers, monoclinic $C2$ occurs only for sequences of the type C-C-A-R-x-x-Y-T-G-G, with a large twist, roll, and slide at the penultimate C-A (= T-G) steps (Privé et al., 1987, 1991; Heinemann & Alings, 1989). In contrast, orthorhombic $P2_12_12_1$ is encountered only for sequences of the type C-G-A-T-x-x-A-T-C-G, all of which have a normal twist at the penultimate G-A (= T-C) steps (Grzeskowiak et al., 1991; Quintana et al., 1992; Yuan et al., 1992).

Minor groove width also correlates with crystal type: the orthorhombic dodecamers and monoclinic decamers all have narrow minor grooves in the middle of the helix and wider grooves toward the ends, whereas the minor groove in all of the orthorhombic decamers is wide at the ends, narrows farther in from each end, and widens again at the center. Repeated crystallization attempts have indicated that the preferences of sequences for one crystal packing mode or the other—monoclinic vs orthorhombic—is strong and sequence-dependent. It has not been possible to induce one and the same sequence to occupy both monoclinic and orthorhombic forms by varying crystallization conditions.

Heinemann and co-workers have succeeded in crystallizing C-C-A-G-G-C-C-T-G-G and its methylated analogue C-C-A-G-G-C^{5me}C-T-G-G in two different space groups: the unmethylated in the familiar monoclinic $C2$ and the methylated species in hexagonal space group $P6$ (Heinemann & Aligns, 1989, 1991; Heinemann & Hahn, 1992). They find virtually identical local helix parameters in the two completely different environments and therefore argue for the unimportance of crystal packing effects. In contrast, Baikalov et al. (1993) have observed that the methylated decamer C-G-A-T-C-G-^{6me}A-T-C-G in trigonal space group $P3_221$ has quite different local helix parameters than the unmethylated sequence in orthorhombic $P2_12_12_1$ (Grzeskowiak et al., 1991).

Clearly the issue of relative importance of base sequence vs helix packing in determining detailed helix structure is not a simple one. Some DNA sequences are "softer" than others and are more susceptible to modification. But it is also true that some changes in local crystal environment are more drastic than others. In the monoclinic, hexagonal, and orthorhombic

† This work was performed with the support of NSF Grant MCB-8916261 and NIH Program Project Grant GM-31299.

* To whom correspondence should be addressed, at the Molecular Biology Institute.

‡ Permanent address: Institute of Molecular Genetics, Russian Academy of Sciences, Moscow, Russian Federation.

0006-2960/93/0432-1373$04.00/0

crystals just mentioned, decamer double helices are stacked atop one another to form semicontinuous helical rods, which then are packed in parallel to build the crystal lattice. Interhelix contacts, although not identical, might be expected to be similar, involving primarily the close approach of two phosphate backbones.

In contrast, in these trigonal crystals, semicontinuous helical rods cross over one another, interlocking their phosphate backbones and major grooves in a manner recalling that encountered in rhombohedral space group *R*3 for the G/C decamer C-C-G-G-C-G-C-C-G-G (Heinemann et al., 1992) and the *Nar*I dodecamer A-C-C-G-G-C-G-C-C-A-C-A (Timsit et al., 1989, 1991; Timsit & Moras, 1991). Now the most intimate contacts are between the phosphate backbone of one helix and the floor of the major groove in the other helix and are quite different from interactions encountered previously.

In both of the above methylation comparisons, adding methyl groups along the floor of the major groove introduced added complications in the form of hydrophobicity and steric hindrance. This paper describes the structure of the decamer C-C-A-A-C-I-T-T-G-G, which crystallizes in two different space groups, monoclinic *C*2 or trigonal $P3_221$, depending on whether the cation used is calcium or magnesium. The present sequence was conceived as one of the smallest possible variants on a previously known structure, monoclinic C-C-A-A-C-G-T-T-G-G, deleting one minor groove guanine amine group from each of the two strands. This deletion leaves the two central C·I base pairs with only two hydrogen bonds instead of three, and one might predict that the C·I base pairs would behave more like T·A than C·G pairs. But the structure analysis to be described below did not show T·A-like behavior in such matters, for example, as larger propeller.

Local helix parameters for this trigonal structure are quite different from those of the monoclinic form with the same sequence and are equally different from those of the other known trigonal decamer, C-G-A-T-C-G^{6me}A-T-C-G. As noted by Baikalov et al. (1993), the trigonal packing is more open and could therefore be more permissive, allowing the B-DNA oligomers to approximate more closely the conformations that they would adopt in solution. And as Timsit and co-workers have observed with their somewhat different rhombohedral *Nar*I dodecamer (Timsit & Moras, 1991; Timsit et al., 1989, 1991), the 120° crossing of helices in the trigonal cell provides an excellent molecular model for a Holliday junction (Holliday, 1964; Broker & Lehman, 1971; Sigal & Alberts, 1972).

For brevity of reference in what follows, DNA oligomers will be cited using the following shorthand terminology: CG = monoclinic C-C-A-A-C-G-T-T-G-G; CICa = monoclinic C-C-A-A-C-I-T-T-G-G crystallized with Ca^{2+} ions; CIMg = trigonal C-C-A-A-C-I-T-T-G-G crystallized with Mg^{2+} ions; KK = orthorhombic C-G-A-T-C-G-A-T-C-G; KKMe = trigonal C-G-A-T-C-G^{6me}A-T-C-G; G/C = rhombohedral C-C-G-G-C-G-C-C-G-G; *Nar*I = rhombohedral A-C-C-G-G-C-G-C-C-A-C-A//T-G-T-G-G-C-G-C-C-G-G-T.

MATERIALS AND METHODS

Monoclinic Form, with Calcium Ions. Crystals of CICa were grown from a solution containing 0.53 mM DNA decamer double helix, 21.7 mM calcium acetate, 21.7 mM sodium cacodylate (pH 7.0), 0.43 mM spermidine (pH 7.1), and 9% MPD (2,4-methylpentanediol) by sitting drop vapor diffusion at 4 °C against a more concentrated MPD reservoir, with a final MPD concentration of 40%. The space group is monoclinic *C*2, with $a = 31.87$ Å, $b = 25.69$ Å, $c = 34.21$ Å, $\beta = 114.1°$, and five base pairs or one-half helix per asymmetric unit. These crystals have a volume of 1278 Å^3 per base pair and diffract to 1.3 Å resolution.

X-ray intensity data were collected at −25° on a Rigaku AFC-5R diffractometer equipped with a graphite monochromator, using Cu Kα radiation from a rotating anode operated at 50 kV and 170 mA. The crystal was ca. 0.3 × 0.2 × 0.6 mm. Intensities were corrected for Lorentz and polarization factors, time-dependent radiation damage, and empirical absorption factors. Of 5848 reflections measured in the range of 8–1.3 Å, 5026 2σ reflections were considered as statistically significant and used for refinement, making this the highest resolution B-DNA oligomer data set thus far collected.

Because of the isomorphism of these crystals with those of the other monoclinic decamers, an idealized, Arnott fiber-derived B helix of the correct sequence was placed in the unit cell in the position found with C-C-A-A-G-A-T-T-G-G (Privé et al., 1987, 1991). Rigid body refinement was carried out with X-PLOR (Brünger et al., 1987), using a gradually increasing number of rigid bodies as resolution was increased from 3 to 1.3 Å. This was followed by positional refinement with X-PLOR and NUCLSQ (Hendrickson & Konnert, 1980; Westhof et al., 1985) and *B* factor refinement with NUCLSQ. The initial *R* factor of 0.54 for the unrefined Arnott model fell to 0.37 under positional refinement and to 0.29 after *B* factor refinement.

Solvent peaks were identified as ordered water molecules if they were present simultaneously at the 3σ level in an ($F_o - F_c$) difference map and at the 1σ level in a ($2F_o - F_c$) map. One hydrated calcium ion with seven water ligands was clearly visible within the minor groove from the beginning of the water localization process. Seventy-two other water molecules were located, along with an unusual fragment of density that has been interpreted as a partially ordered cacodylate ion. Refinement converged at $R = 0.152$ for the 5026 reflections above 2σ in the range 8–1.3 Å.

Trigonal Form, with Magnesium Ions. Quite different trigonal CIMg crystals were obtained from a solution containing 0.53 mM DNA decamer double helix, 21.7 mM magnesium acetate, 21.7 mM sodium cacodylate (pH 7.0), 0.32 mM spermine, and 9% MPD, by sitting drop vapor diffusion at 4 °C against a MPD reservoir, with crystals appearing at 35% MPD. From survey precession photographs, the space group could be either $P3_121$ or $P3_221$, with $a = b = 33.23$ Å, $c = 94.77$ Å, and $\gamma = 120°$. (Subsequent analysis showed $P3_221$ to be correct.) The asymmetric unit now contains an entire double helix of 10 base pairs. These crystals are more open than the monoclinic variant, with 1510 Å^3 per base pair, and diffract only to 2.2 Å resolution. X-ray intensity data were collected as described for the monoclinic form, from a crystal ca. 0.3 × 0.2 × 0.3 mm. Of 3202 reflections measured in the range of 8–2.2 Å, 1725 were above the 2σ level and were used in structure analysis.

An initial attempt at solution was made using a conventional rotation–translation search. An *hk*0 precession survey photograph revealed strong clusters of reflections at 3.4 Å spacing along the (*h*,*h*,0) row and every 60° around the photo, indicating that rods of stacked B-DNA decamers were oriented along the *a* and *b* real space axes [and hence along the −(*a*+*b*) direction also]. This fixed two of the three Euler rotation angles, leaving only the rotation of the helix about its own axis. Figure 1 shows the result of this rotation search using the known C-C-A-A-C-G-T-T-G-G structure (Privé et al., 1991) as a starting model: five peaks spaced 36° apart, representing the screw symmetry of the B-DNA helix. A translation search using the best of the five rotation choices led to a large number of nearly equivalent peaks in translation

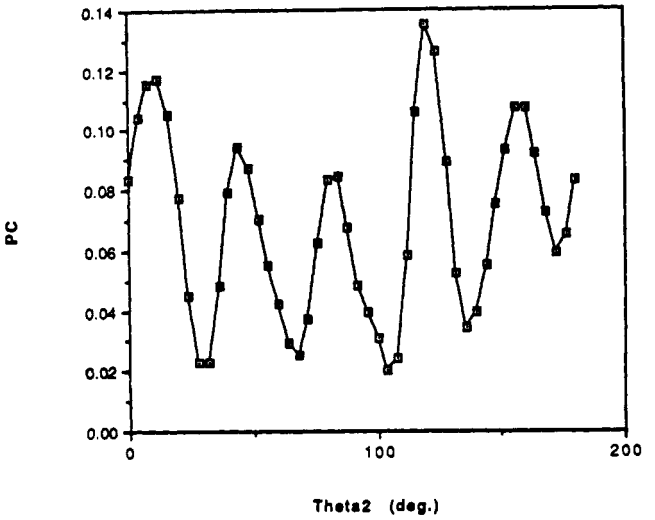

FIGURE 1: One-dimensional rotation function search for the trigonal form of C-C-A-A-C-I-T-T-G-G. The two Euler angles defining the orientation of the helix axis are determined a priori by the knowledge, derived from strong clusters of 3.4-Å reflections in the diffraction pattern, that helices are oriented along a, $-(a+b)$, and b. This remaining angle describes rotation of a helix about its axis. The five solutions, separated by 36°, reflect the 10-fold symmetry of the helix.

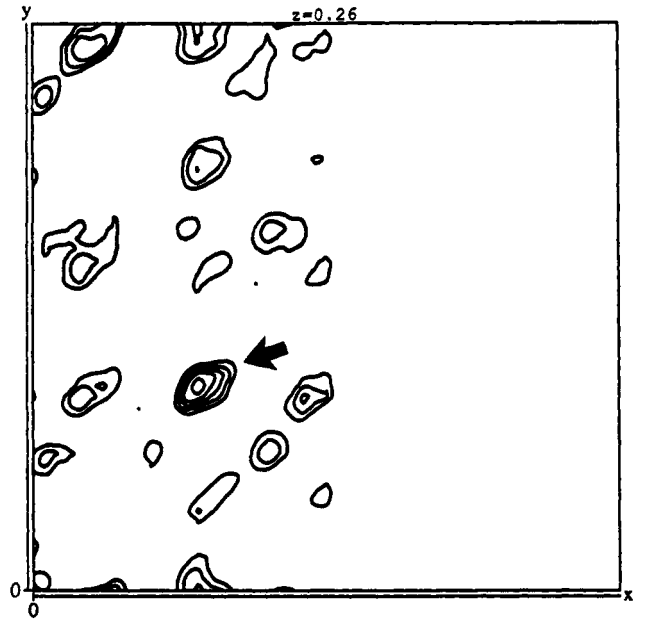

FIGURE 2: Section through the three-dimensional translation function for trigonal C-C-A-A-C-I-T-T-G-G. The section is at $z = 0.26$, with x running horizontally from 0.0 to 0.5 and y running vertically for one entire unit cell. The largest peak in the full three-dimensional function is marked by an arrow.

space, without a clear-cut answer. The PC refinement program from X-PLOR (Brünger et al., 1987) then was tried. This carries out rigid body refinement of the model against the standard correlation coefficient prior to a translation search. The model resulting from PC refinement yielded a unique best translation solution in space group $P3_221$ (Figure 2) but for space group $P3_121$ gave only a collection of lower peaks, no one of which stood out. Rigid body refinement of this $P3_221$ solution in X-PLOR, followed by positional refinement with NUCLSQ (Hendrickson & Konnert, 1980; Westhof et al., 1985), led to an R factor of 0.26.

At this point the CIMg structure was compared with the recently solved KKMe structure (Baiklov et al., 1993) and was observed to differ from it by one screw axis step, that is, by 36° rotation and 3.34-Å translation. Considering the stacked decamers helices as a long helical rod, the break

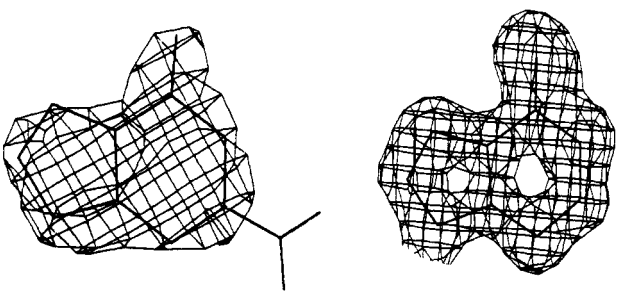

FIGURE 3: Difference electron density maps for the inosine base, after refinement with this ring omitted. (a) Trigonal CIMg structure at 2.2-Å resolution. (b) Monoclinic CICa structure at 1.3-Å resolution. The added N2 amine group on the superimposed purine skeleton at left illustrates that this base is clearly inosine and not guanine. Note the improved definition of the purine double ring at higher resolution.

between helices had been shifted by one base pair down the rod. The KKMe-derived solution was better behaved under refinement. The R factor, beginning at 0.51 for an Arnott ideal B helix of the proper sequence, fell to 0.28 after XPLOR rigid body and positional refinement and to 0.22 after NUCLSQ B factor refinement. Systematic deletion of all other pairs of phosphates along the helical rod and subsequent refinement gave no better solution, and so that taken from the KKMe structure was considered proven for CIMg as well. After 36 water molecules and one hexahydrated magnesium ion were added, the final R factor was 0.162.

Because this trigonal space group was so different than what had been expected, or encountered in the past, extra steps were taken to verify its correctness. The Patterson search strategy that proved so succesful with KKMe (Baikalov et al., 1993) was tried with the CIMg data and yielded the same solution as just refined. As an additional check, this solution was refined with deletion of the inosine bases, and the inosine image then was examined in a difference map. The image (Figure 3a) clearly shows the absence of an N2 amine group even at 2.2-Å resolution, supporting this location for I·C base pairs along the helical rod. In a final check of the independence of local helix parameters on starting model, refinement was repeated beginning with coordinates of the CG helix (Privé et al., 1991) rather than GA, positioned correctly in the trigonal cell. The refined models starting with this helix and with the Arnott ideal helix had rms differences of only 0.37 Å in atomic positions, prior to addition of waters, and very similar local helix parameters.

Both X-ray intensity data and final refined atomic coordinates for the CICa and CIMg structures have been deposited with the Brookhaven Protein Data Bank and are available for immediate distribution.

RESULTS

Crystal Packing: Monoclinic CICa. Crystals of the CICa and CG helices are isomorphous. Crystal packing in the CG decamer is shown in Figures 1 and 2 of Privé et al. (1991), which could also serve perfectly well for the present CICa structure. Endless columns of stacked decamer helices along the c axis direction of the monoclinic cell are packed side-by-side in a c-centered manner. As Figure 3b of Grzeskowiak et al. (1991) shows, this corresponds to a roughly square packing of cylindrical columns, with a separation of 19.5 Å between centers. The monoclinic angle is produced by a lateral displacement of neighboring columns, which permits the phosphate backbones of one column of helices to fit into the major and minor grooves of its neighbors. Missing phosphates at the gaps between stacked helices occur at especially close

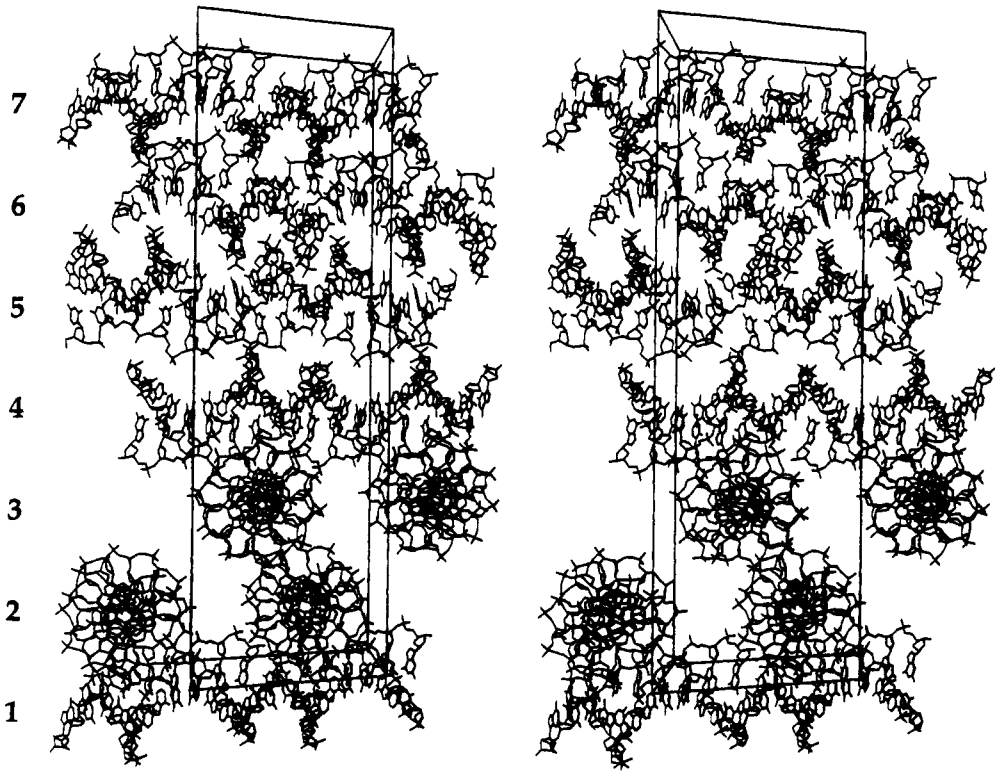

FIGURE 4: Helix packing in the trigonal unit cell of CIMg. Origin at left bottom rear. The *c* axis is vertical, *a* rises out of the page toward the viewer, and *b* sinks 30° into the page from left to right. Compare the KKMe decamer in Figure 3 of Baikalov et al. (1993). As in that structure, the crystal lattice is built from six layers of pseudo-continuous helical rods parallel to the *ab* plane (numbered 1 to 7, with layer 7 being identical to layer 1 except for a unit cell translation along *c*), each rod being constructed from decamer double helices stacked end to end. Helices are parallel to *a* in layers 2 and 3, to the *ab* face diagonal in layers 4 and 5, and to *b* in layers 6 and 7.

approaches of phosphate backbones of adjacent columns, permitting closer approach of columns than would otherwise be possible.

Crystal Packing: Trigonal CIMg. Packing in the trigonal CIMg crystals is quite different from this and in fact is identical to that seen with KKMe (Baikalov et al., 1993). Six layers of helical rods parallel to the *ab* plane are stacked up the trigonal *c* axis, as in Figure 4. Layers 2 and 3 in Figure 4 are parallel to *a*, layers 4 and 5 are 120° away along the *ab* diagonal, and layers 6 and 7 (= 1) are rotated yet another 120° to lie along *b*. As with KKMe, the fundamental structural unit of crystal packing in this trigonal cell can be regarded as a lattice built from two layers of rods oriented 120° apart, with interlocking of phosphate backbones and major and minor grooves where the helices cross. Layers 1 and 2 in Figure 4 comprise one such lattice unit, as do layers 3 and 4 and layers 5 and 6. The top helices of one bilayer are parallel to the bottom helices of the next bilayer up the *c* axis, and their interhelix contacts are not unlike those seen in the monoclinic, orthorhombic, and hexagonal decamers, involving primarily the sugar–phosphate backbone. But interactions within one lattice bilayer are a different story and are more like that seen with rhombohedral G/C and *Nar*I.

The crossing of helices is shown from three points of view in Figure 5: a "top" view down the *c* axis in Figure 5a, a view directly from the right in Figure 5c, and an intermediate 45° view in Figure 5b. The two crossed helices are related by a crystallographic 2-fold axis, which is horizontal in the plane of the page in Figure 5a and perpendicular to the page through the center of Figure 5c. The crossing of helices is depicted schematically but accurately in Figure 6. The primary interaction occurs where each helix inserts one of its sugar–phosphate backbone chains into the major groove of the other helix. More peripherally, each helix also nests part of its other backbone into the minor groove of its neighbor. A 2-fold symmetry axis runs horizontally in the plane of the diagram in Figure 6. Insertion of the backbone into the major groove is unsymmetrical: the inserted chain is farther from that wall which must be reached by crossing over the 2-fold axis than it is from the wall that is reached by moving away from the axis. This feature, which also is clearly visible in Figure 5c, will assume more importance later, in the comparison with the rhombohedral *Nar*I structure.

At the crossover between helices, each helix extends its P16–P17–P18 backbone of strand 2 (residues C11–G20) down into the major groove of the other helix, with its O1P oxygen of P17 hydrogen-bonded to the N4 amine of cytosine C1 of the neighbor, and its O2P bonded to the N4 of cytosine C2 (Figure 7). As is nearly always the case in other packing modes, the close approach of two rods of stacked helices occurs near the interhelix junction in one rod, where the absence of phosphates permits a closer contact. Moreover, close inspection of Figure 5a reveals that the intruded phosphate backbone is not parallel to the walls of the major groove but is inclined

FIGURE 5: Interlocking backbone-and-groove contacts between two helices in a lattice such as is formed by layers 1 and 2 in Figure 4, or alternatively layers 3 and 4, or 5 and 6. One and a half helices are drawn in each case. (a, top) "Top" view down the trigonal *c* axis. In the heavy-line helix nearer the viewer, base pair G10·C11 is at upper left. The helix continues down toward the lower right to base pair C1·G20, followed by the first five base pairs of the next helix in the stack. The light-line helix begins with base pair G10·C11 at lower left. The two helices are related by a crystallographic 2-fold axis horizontal in the plane of the page. (b, middle) Rotation of 45° about a vertical axis. (c, bottom) After 90° rotation, viewing panel a from the right. The crystallographic 2-fold axis is perpendicular to the page through the center of this view. To this view have been added two symmetry-related octahedral $Mg(H_2O)_6^{2+}$ complexes that help hold the interlocked helices together.

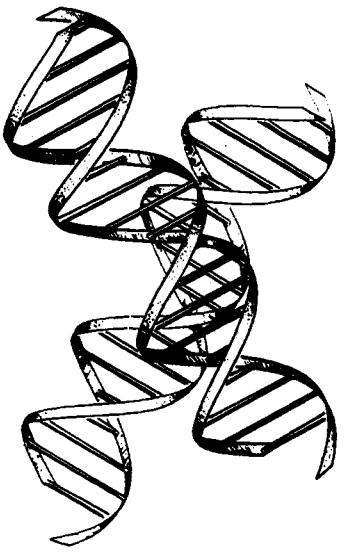

FIGURE 6: Interlocking of backbone and grooves in the CIMg structure. A horizontal 2-fold axis through the middle of the diagram relates the top and bottom helices. Each helix inserts a portion of its backbone chain into the major groove of the other and wraps its adjacent minor groove around a short "elbow" of backbone from the other helix. Hydrogen bonding involves only the major groove contacts, not the minor.

roughly 14° to them. At no other region of the contact zone between crossed helices do groups come close enough together for hydrogen bonds; the insertion of backbone into minor groove above and below the primary contact region is merely a passive van der Waals nesting.

The two phosphate junctions in Figure 5c are given added stability by the bridging of octahedrally hydrated magnesium cations. This bridging is shown in closeup in Figure 7, and potential hydrogen-bond distances of 3.5 Å or less are listed in Table I. In essence, N4 amines of cytosines bind directly to the phosphate oxygens, whereas water molecules of the magnesium complex bridge from phosphate P17 to guanine O6 and N7 of three adjacent base pairs: the first two base pairs of one helix of the neighboring stack and the final base pair of the helix just above it. This bridging is different from that in KKMe, where the magnesium complex uses one phosphate oxygen as a ligand and where bonds from the other magnesium oxygens to the second helix appear to be more significant than are direct helix–helix hydrogen bonds.

Table I: Hydrogen Bond Distances between Two Interlocked Helices and Their Bridging Hydrated Magnesium Cation

		magnesium complex				phosphate P17	
base	atom	W1	W3	W4	W5	O1P	O2P
G10	N7		3.29				
	O6				3.41		
C1	N4	3.12				2.89	
G20	O6				2.43		
	N7						
C2	N4						2.98
G19	O6	2.91					
	N7			2.85			
P17	O1P	3.22	2.80				
	O2P	2.78					

A very similar crossing of helices has been reported for the G/C decamer C-C-G-G-C-G-C-C-G-G (Heinemann et al., 1992) and for the non-self-complementary *Nar*I dodecamer (Timsit et al., 1989, 1991; Timsit & Moras, 1991), and these will be compared more extensively in the Discussion. Interlocked helices in the trigonal cell cross at 120°. But the rhombohedral *R*3 space group and cell dimensions of the G/C decamer and *Nar*I dodecamer mean that their helices cross at an angle of 120° only in projection onto the trigonal *ab* basal plane. In three dimensions, helices of the G/C decamer cross at an angle of 102.9°, and those of the *Nar*I dodecamer cross at 106.4°. This has the effect of actually improving interlocking of helices, by making the intruded phosphate backbone lie parallel to the walls of the major groove into which it fits [see Figure 2a of Timsit and Moras (1991)].

Contacts between layers of parallel stacked helices, such as 2 vs 3, 4 vs 5, and 6 vs 7 in Figure 4, are sparse. The "notch" between phosphates P5 and P6 along one chain is fitted against the gap between base G20 of one helix and C11 of the following helix along a rod, where phosphate P11 would be if the rods were continuous. The three closest contacts at this locus are (a) G6 O1P to C11 O5′, 3.09 Å; (b) G6 O2P to G20 O3′, 2.82 Å; and (c) C5 O5′ to G10 $O3'_3$, 3.40 Å.

Ions and Hydration: Monoclinic CICa. Ordered water molecules around the CICa molecules in their monoclinic cell are much like those in the CG decamer: a regular spine of hydration down the narrow minor groove in the center of the helix, terminated at each end by a hydrated cation where the groove begins to widen (Figure 8a). One difference is that in CICa the cation is sited one bond length or one-half base

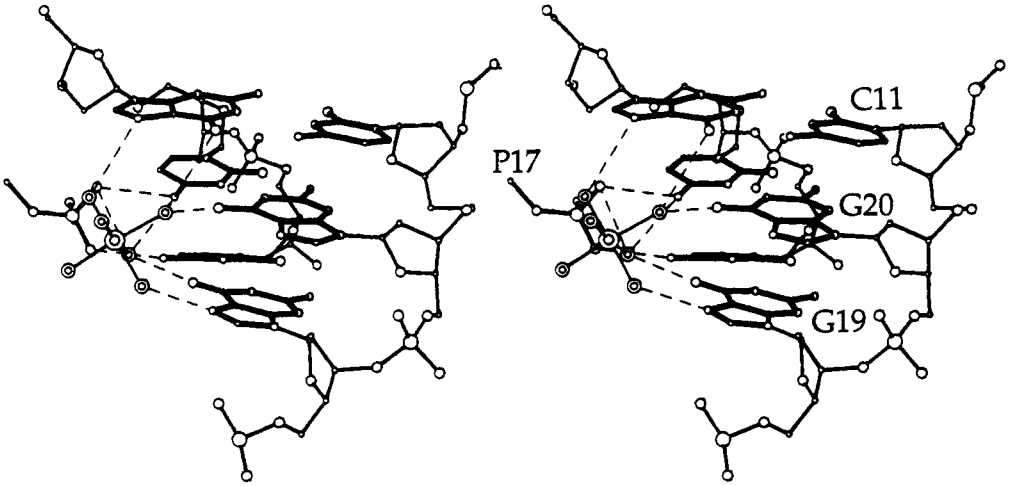

FIGURE 7: Closeup of the interaction between interlocked CIMg helices, bridged by a hydrated magnesium cation. The view is that of the upper contact in Figure 5c. The three base pairs drawn, from bottom to top, are C2·G19 and C1·G20 from one helix and G10·C11 from the helix stacked atop it. A third helix at left is represented here only by its P17 phosphate group. Direct hydrogen bonds extend from the two phosphate oxygens to the N4 amine groups of C1 and C2. The octahedral magnesium complex provides additional hydrogen bond bridging from phosphate P17 to N7 and O8 of guanines. See Table I for bond distances.

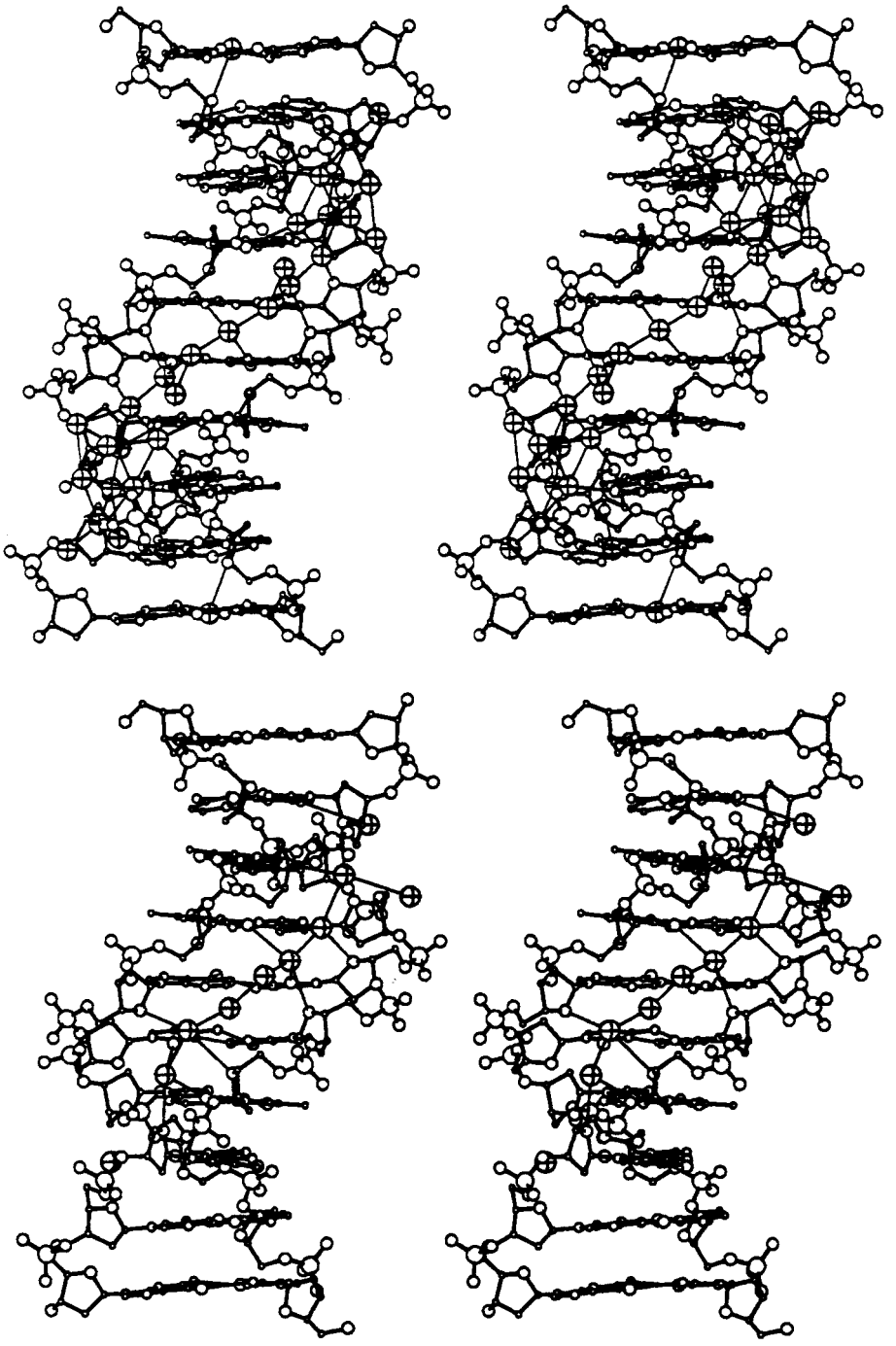

FIGURE 8: Minor groove hydration in the CI helices. (a, top) Monoclinic CICa. Note heptahydrated calcium ions at upper right and lower left in minor groove. (b, bottom) Trigonal CIMg, with base pair C1·G20 at top and G10·C11 at bottom. No hydrated cation is present in this minor groove; it occurs between helices in the major groove instead.

step farther toward each end of the helix, even though the minor groove width is identical in CG and CICa helices. As a consequence, the room available toward each end of the helix for double ribbons in the wider portions of the minor groove is minimal, and the ribbons, although present, are rudimentary. This difference in cation location probably reflects the larger size of the heptacoordinated $Ca(H_2O)_7^{2+}$ ion of CICa, by comparison with the hexacoordinated $Mg(H_2O)_6^{2+}$ of CG. In both cases the cation occupies a "pocket" between adjacent helices in the crystal, with water bridges to P2 phosphate oxygens on a neighboring molecule [see Figure 2 of Privé et al. (1991)].

The only chemical difference between CI and CG helices is the presence of an N2 amine on the central guanine in the minor groove of the CG helix and its absence in CI. Molecular dynamics simulation (Chuprina et al., 1991) suggests that these guanine N2 amines should disturb the spine of hydration, in particular the central first-shell water that bridges N3 atoms of purines G6 and G16, causing that water molecule to favor one N3 atom or the other. The symmetrical hydration observed in CG by Privé et al. (1991) may represent an averaging of two slightly displaced waters. These perturbing N2 amine groups are absent in the CI helix, so a bridge of hydration between N3 atoms of residues I6 and I16 can be formed without hindrance.

Ions and Hydration: Trigonal CIMg. At a lower resolution of 2.2 Å, hydration of the trigonal CIMg structure is less well defined. A spine of hydration still is traced up the narrow central region of the minor groove (Figure 8b), but without the terminating cation clusters. Only a single hexahydrated magnesium cation is sufficiently localized to be visible in the CIMg structure analysis, but it lies in the major groove rather

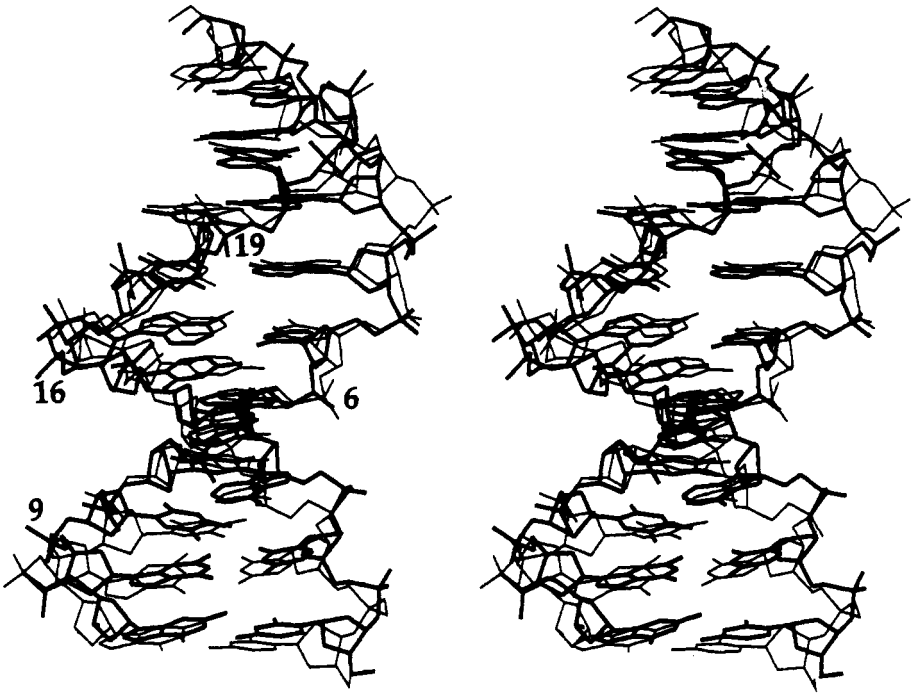

FIGURE 9: Overlap stereodrawing of the trigonal CIMg helix (dark bonds) and the monoclinic CICa (light bonds). In CIMg, base pair C1·G20 is at top and G10·C11 at bottom. This orientation of the helix provides a side view of the outer two base pairs at bottom and an end view of the first two base pairs at top. Phosphates 6, 9, 16, and 19 are marked. The two ends of the CICa helix are identical; those of the CIMg helix are not.

than the minor. Its role in bridging crossed helices has already been discussed.

Comparison of CI and CG Structures: Local Helix Parameters. The two trigonal structures, CIMg and KKMe, are compared in detail by Baikalov et al. (1993), and that comparison will not be repeated here. Suffice it to say that local helix parameters are not identical in the two trigonal structures, meaning that the trigonal packing, for all of its apparent radical nature, imposes milder constraints on the helices that adopt it than is the case for either monoclinic or orthorhombic packing.

The CI sequence, C-C-A-A-C-I-G-G-T-T, differs from the CG sequence, C-C-A-A-C-G-T-T-G-G, only in elimination of the minor groove N2 amine group from guanine at the seventh position along each strand. Hence the CI vs CG comparison potentially can yield information about the importance of that guanine amine in determining local helix structure. In contrast, the comparison of two identical CI sequences in different crystal environments allows one to evaluate the influence of environment on structure. All of the monoclinic *C*2 structures share the same tight crystal packing. This close packing yields a high-resolution diffraction pattern but substantially restricts the local structure of the helix and hence also restricts the sequences that will adopt this particular packing mode. In contrast, the more open trigonal packing imposes fewer structural constraints on the helix and, as will be seen, yields local helix parameters that are less extreme and are more typical of an isolated B-DNA helix in solution.

Local helix parameters were calculated using the program NEWHEL92 (available from the Brookhaven Protein Data Bank or from R.E.D. at the E-mail address RED@UCLAUE). These values are listed in Appendix Tables III–VIII, in what has become a standard format for structures from this laboratory.

The monoclinic and trigonal forms of the CI decamer are superimposed in Figure 9, in a view that displays the outermost base pairs end-on at the top and in side view at the bottom. The most striking difference between CICa and CIMg

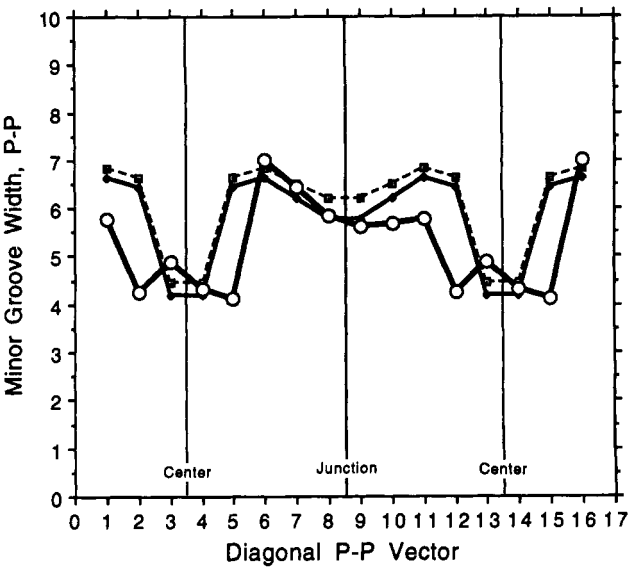

FIGURE 10: Minor groove width in the CG (dashed line), CICa (light solid line), and CIMg (heavy line) decamers. Distances are P–P separations across the groove less 5.8 Å for two phosphate group radii. Phosphate numbers are given for reference below the plot. Missing phosphate distances at interhelix junctions are interpolated from adjacent P–P separations.

structures is in minor groove width, monitored in Figure 10. Minor grooves are equally narrow at the center of the CICa and CIMg helices and equally wide toward the ends. But in CIMg the narrow region is extended toward either end by one base pair. Figure 9 shows that this narrowing is caused primarily by a moving inward of phosphates P16 and P9 at the lower end and P6 and P19 at the upper end. Such a narrowing would make the minor groove completely inhospitable for a hydrated cation, and, indeed, no cation is found within the minor groove in the CIMg structure.

One characteristic of all of the monoclinic decamer DNA structures is a very large helical twist of ca. 50° at the second step in from either end of the helix, C-A (=T-G), with a less

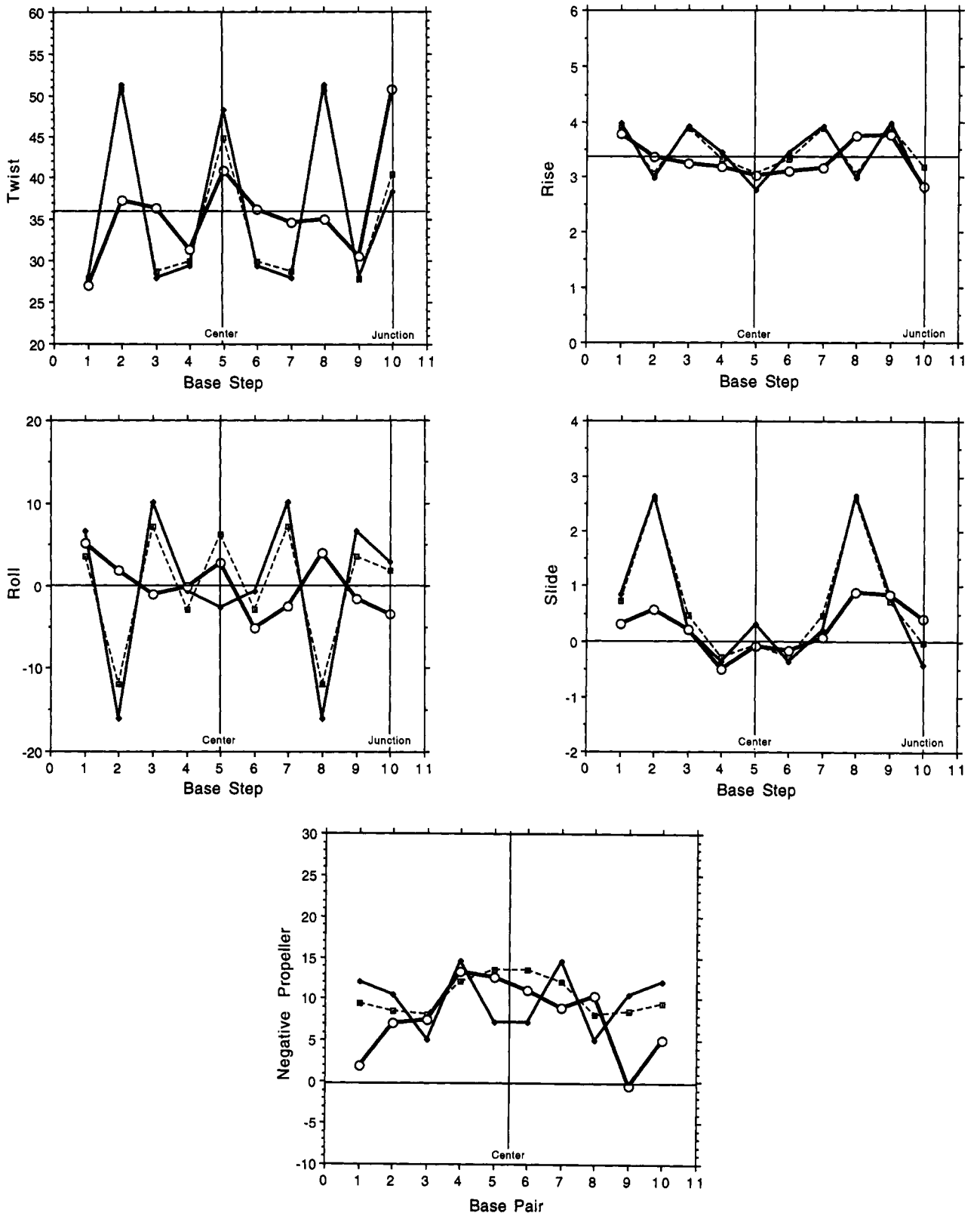

FIGURE 11: Local helical parameters in the CG (dashed line), CICa (light line), and CIMg (heavy line) decamers. (a, top left) Twist. (b, top right) Rise. (c, middle left) Roll. (d, middle right) Slide. (e, bottom) Propeller (with reversed sign). Standard ranges have been used for comparison with other papers from this laboratory: 40° for angles and 6 Å for distances.

extreme maximum at the center of the helix. Such a large twist is not encountered at C-A steps in the various dodecamers. Yanagi et al. (1991) have proposed that the large twist of a C-A step in the monoclinic decamers might be dictated by its flanking sequences: high twist in the setting Y-C-A-R and normal twist otherwise. The present study (Figure 11a) indicates instead that the C-C-A-A tetrad is at least bi-stable: capable of adopting either a high or a low twist at its center, depending on its environment. Other sequences may not be capable of such a twist distortion, and indeed this probably is why they choose other space groups and packing modes than monoclinic *C*2.

Twist angles are much less varied in the trigonal form of CI than in the monoclinic and are somewhat lower overall. Indeed, the nine internal twist angles for CIMg average 34.37°, which corresponds to 10.5 base pairs per turn. As discussed

by Baikalov et al. (1993), this agrees with solution measurements of DNA periodicity and suggests that the looser trigonal crystals might be better models for free DNA than any of the previous monoclinic or orthorhombic forms, including dodecamers. As a consequence, the unbonded step between helices must have a large 51° twist in order to make up 360° over one unit cell repeat ("Junction" in Figure 11a).

Homopolymer G-G and A-A steps tend to have low to medium twist because of the efficiency of their base stacking (Yanagi et al., 1991; Grzeskowiak et al., 1991), and this trend is continued in both CICa and CIMg structures (steps 1, 3, 7, and 9 in Figure 11a).

The behavior of helical rise (Figure 11b) in CICa is the inverse of that of twist, as was noted first by Yanagi et al. (1991). The rise in CIMg is remarkably regular and uniform, matching the relatively small excursions of twist from the mean. The very great twist dislocation at the second step from either end of the helix in all of the monoclinic decamers is accompanied by a large negative roll (opening toward the major groove, Figure 11c) and positive slide (Figure 11d). This roll is especially visible near the top of Figure 9, and slide is visible in the motion of bases G9 and G10 at bottom. No such behavior is found in the CIMg structure, again emphasizing that the twist, roll, and slide behavior found in the monoclinic structures is a consequence of that particular mode of packing of molecules and is not inevitable from sequence.

To this point the discussion of local parameters has contrasted the two forms of CI helix, the monoclinic CICa and CG structures being considered as essentially identical. But the behavior of propeller in the two monoclinic helices (thin line and dashed line in Figure 11e) illustrates the chemical difference between guanine and inosine and also the influence of bound cations. Propeller is variable at the ends of each helix, and tends toward an intermediate magnitude of ca. 8° farther in from each end and a larger ca. 12° at the center of the CG helix (dashed line). But the two C·I base pairs at the center of the monoclinic CICa structure are twisted only half as much. This is surprising, as one might have expected the loss of one of the three C·G hydrogen bonds in C·I to have encouraged an even larger propeller. But any tendency in this direction probably is counterbalanced by the difference in interactions with hydrated ions. In the CG structure [Figure 7 of Privé et al. (1991)], the O2 carbonyl groups of bases C5 and C15 are rotated slightly toward the ends of the helix because their oxygen atoms are hydrogen-bonded directly to water molecules in the hydrated magnesium complexes. The zig-zag spine of hydration is only three water oxygens long, not counting waters in the two terminating cation clusters. In contrast, displacement of the larger hydrated calcium complexes toward the two ends of the helix in CICa yields a spine five atoms long, so that the O2 atoms of the central C5 and C15 form hydrogen bonds, not with waters of the cation clusters themselves but with waters in the spine that sit very nearly in the plane of the base pairs. Hence good hydrogen bonds can be formed without requiring a significant base pair propeller twist.

Base Pair Stacking and Phosphate Conformation in CICa and CIMg. Figure 12 compares the stacking of base pairs in CIMg and CICa helices. The 16-part figure is laid out so that corresponding base pair steps in the CICa and second half of the CIMg helix can be compared horizontally across a two-page display: Figure 12e–j for CIMg and Figure 12k–p for CICa. The first half of CIMg, in Figure 12a–d, also is to be compared with CICa: panel a with panel o, panel b with panel n, panel c with panel m, and panel d with panel l. These diagrams reinforce the view of Y-R as being an especially variable step (Yanagi et al., 1991; Grzeskowiak et al., 1991) and the importance, to Y-R stacking, of overlap between pyrimidine O2 atoms and the five-membered or six-membered ring of the adjacent purine (Heinemann & Alings, 1989). This variability is evidenced in comparing Figure 12 panels b, h, and n, the first two with near-canonical twist in CIMg, the last example being a HTP step [high twist profile; see Yanagi et al. (1991)] from the same sequence position in CICa. In examples with lower twist angle, the pyrimidine O2 is stacked upon the adjacent purine five-membered ring (Figure 12b,h). With a higher twist, the O2 begins to move off the edge of the five-membered ring (Figure 12k), potentially creating an unfavorable stacking situation. This can be remedied by introducing a positive slide, so that each O2 now sits squarely over the six-membered ring of the adjacent purine, as in Figure 12n. Hence the structural function of slide, in this setting, is to allow the maintaining of stacking in the face of an unusually large twist.

The B_{II} phosphate conformation can be seen as another adaptation to unusually large twist, one that permits greater extension of the backbone chain. At the 51° twist step in CICa (Figure 12n), connecting phosphates on both strands are B_{II}, whereas, in the corresponding lower-twist sites in CIMg (Figure 12b,h), both phosphates are B_I. In a similar manner, at the 41° twist central C-I step of CIMg (Figure 12e) only B_I phosphates are present, whereas enlargement of twist to 48° in CICa (Figure 12k) leads to B_{II} phosphate conformations. All other phosphates in CICa, and indeed all phosphates in CIMg, are B_I.

The R-R steps, G-G = C-C (Figure 12a,i,o) and A-A = T-T (Figure 12c,g,m), all are consistent in having low or intermediate twist, 27–36°. As with previous B-DNA structures, the base pairs appear to pivot about well-stacked purines, allowing pyrimidines to separate until the minor groove O2 of one pyrimidine rests atop the six-membered ring of the other. This is a further example of Heinemann and Alings' (1989) principle of the importance of stacking of polar side groups over aromatic rings.

R-Y steps in general exhibit ITP (intermediate twist profile; Yanagi et al., 1991). In the present analyses these steps (Figure 12d,f,l) show good ring-upon-ring stacking and twists of 29–36°. They lack the O2 atoms in pyrimidines at the first position, which would permit opening up the twist angle wider via O2-on-ring stacking.

The distribution of B_{II} phosphate conformations in the known DNA decamer structures mirrors this twist behavior. Of the 23 Y-R steps in the eight decamers listed in Table II, more than half, or 13 of 23, are B_{II}. The B_{II} conformation is much less common at R-R steps, occuring only 4 in 17 times. It is still rarer at Y-Y steps, 1 out of 17, and never occurs at all at any of the 15 R-Y steps. Hence Y-R steps seem to be bimodal in both twist behavior and phosphate conformation, whereas other steps tend to be unimodal and conservative in both aspects.

In summary, in these structures as in those that have come before them, R-Y steps tend to show low to moderate twist because they are limited to ring-upon-ring stacking. R-R = Y-Y steps also exhibit low twist because of strong purine–purine stacking, although the pyrmidines rotate to achieve O2-upon-ring stacking at the other end of the base pair. Y-R steps are variable, perhaps even bimodal, because their O2-upon-ring stacking can be achieved with either ring of the adjacent purine. Neither of the HTP steps of previous structures, G-C or G-A (= T-C), is present in this CI sequence. It now is apparent that, although the C-A (=T-G) step is allowed to have high twist, it is not required to be so. The

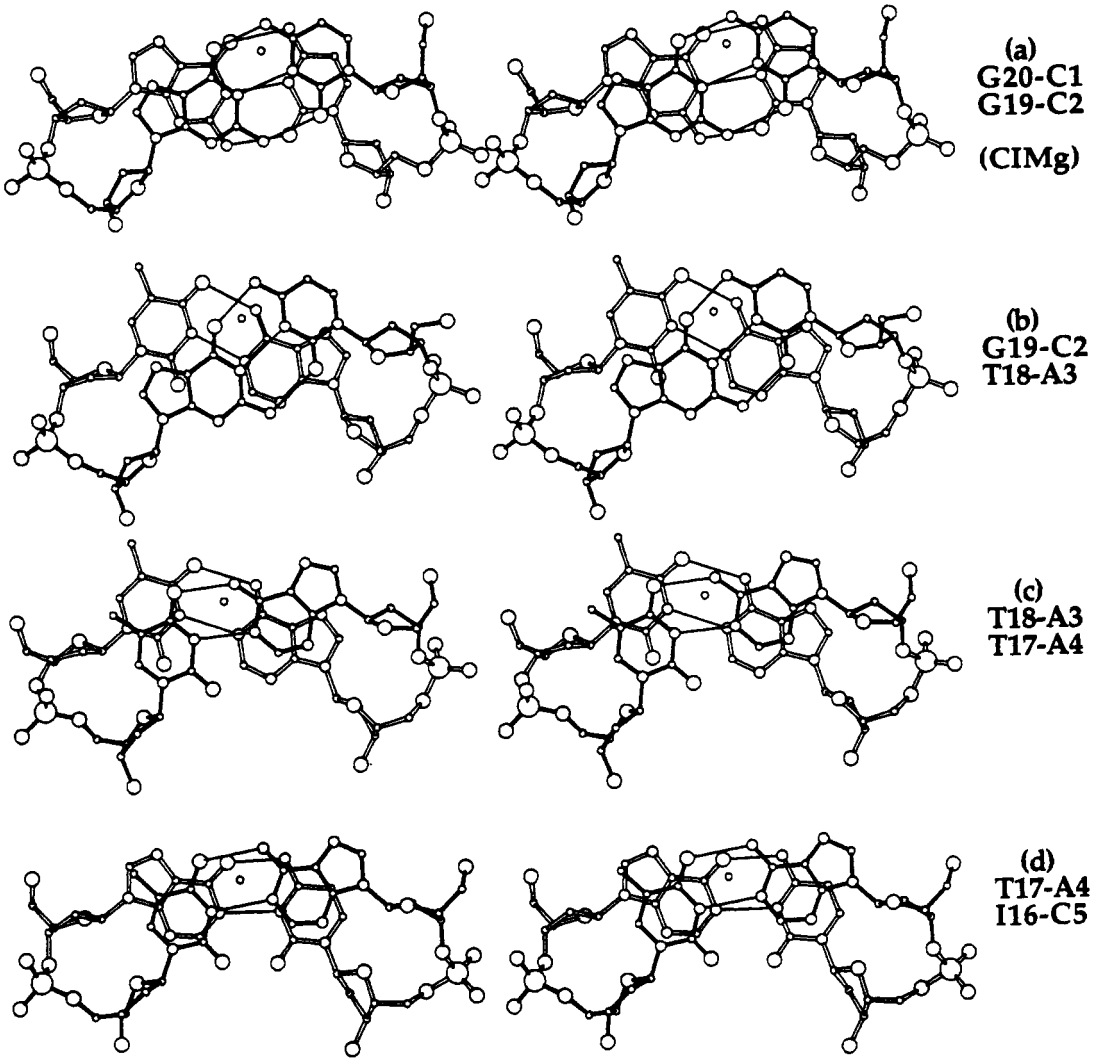

FIGURE 12: Two base pair stereos, viewed down the helix axis, which is marked by a small open circle. (a–j) The ten base pair steps of the trigonal CIMg decamer, as identified by labels to the right of each stereo. (k–p) The six steps of the monoclinic CICa decamer, as identified to right. Only six steps need be shown because of the 2-fold symmetry relating the ends of the CICa helix. In comparing the two decamers, steps e–j of CIMg correspond individually to steps k–p of CICa, and steps a, b, c, and d of CIMg correspond to steps o, n, m, and l of CICa.

choice of twist from the repertoire of this particular base step is made by the local environment within the crystal. Only those sequences for which the penultimate step has such variability are capable of crystallizing in this monoclinic *C*2 form; less compatible sequences will select another form, such as the orthorhombic, trigonal, or others not yet encountered. To paraphrase an old aphorism, "Sequence proposes; setting disposes", if by setting one means the local macromolecular environment in which the DNA finds itself.

DISCUSSION

Comparison with the NarI Dodecamer and G/C Decamer. The crossing of layers of helices in the CIMg structure resembles that found for the *Nar*I dodecamer, A-C-C-G-G-C-G-C-C-A-C-A, by Timsit et al. (1989, 1991) and Timsit and Moras (1991), and the G/C decamer, C-C-G-G-C-G-C-C-G-G, by Heinemann et al. (1992), but differs from both in significant details. The packing of stacked semicontinuous rods of *Nar*I dodecamers is shown in Figure 13, and that for G/C decamers would be similar. The space group is rhombohedral (trigonal) *R*3, with nine double-helical dodecamers in a trigonal cell of dimensions $a = b = 65.9$ Å, $c = 47.1$ Å, and $\gamma = 120°$. But the structure is more easily understood with reference to the corresponding rhombohedral cell, which has dimensions $a = b = c = 41.2$ Å, $\alpha = \beta = \gamma = 106.4°$, and only three dodecamers per cell. The length of the rhombohedral cell repeat exactly matches that expected for a B-DNA dodecamer with a rise of 3.43 Å between base pairs. One semicontinuous column of stacked dodecamers runs along each axis of the rhombohedral cell, producing an angle of 106.4° between crossed helices rather than 120° as with the trigonal CIMg decamer. (The *Nar*I structure is described as having an angle of 120° between crossed helices, but this is true only in projection onto the trigonal *ab* plane.) As mentioned earlier, this 106.4° angle means that the intruded phosphate backbone lies more nearly parallel to the walls of the groove [Figure 2a of Timsit and Moras (1991)] than is the case for the CIMg decamer (Figure 5a). For the G/C decamer, trigonal cell dimensions of $a = b = 54.07$ Å, $c = 44.59$ Å, and $\gamma = 120°$ correspond to a rhombohedral cell of $a = b = c = 34.58$ Å and $\alpha = \beta = \gamma = 102.9°$. The rhombohedral cell edge delineates one 10-base-pair helix, with helices crossing at an angle of 103° (not 120°).

Stacking of B-DNA decamers to build a semicontinuous helical rod is a simple matter, because one decamer contains exactly one turn of helix. This is true for both the trigonal decamers of this study and the rhombohedral G/C decamer. But the stacking of *Nar*I dodecamers requires a twist adjustment where each helix meets its neighbors, because the dodecamer helix is more than one turn long. The sum of twist angles at the 11 internal steps of the *Nar*I helix is 386°, so

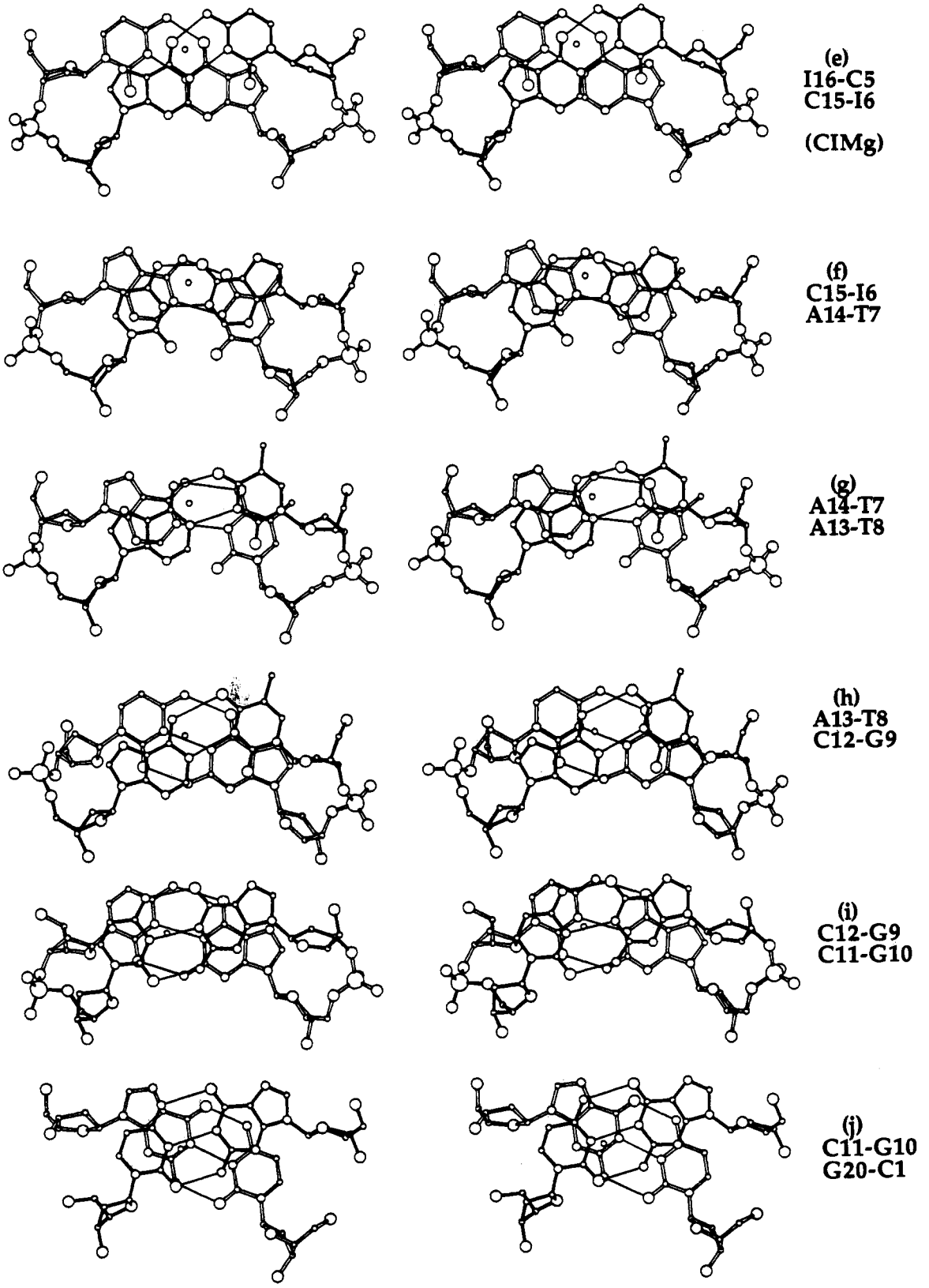

FIGURE 12e–j

the step between the helices must have a reverse twist of –26°. This is never mentioned specifically but can be seen from Figure 2 of Timsit et al. (1991). Each terminal base pair sits astride the end base pair of a neighboring helix, with its 5′ and 3′ chain termini hanging into the major and minor grooves of that helix. In contrast, for a 10-base-pair helix, no such dislocation adjustment is required, and the nonbonded step between helices is in all respects an ordinary helical step.

Hydrogen-bonded contacts for the CIMg decamer and *NarI* dodecamer at the crossing of two helices are contrasted in Figure 14. In each case the backbone chain bearing the accepting phosphates, P17 or P20, can be designated as strand 2 (Timsit and Moras' "T strand"). Two cytosines from adjacent base pairs of the opposite helix use their N4 amines to donate hydrogen bonds to the phosphate acceptor, in two bonding clusters above and below a real or approximate 2-fold

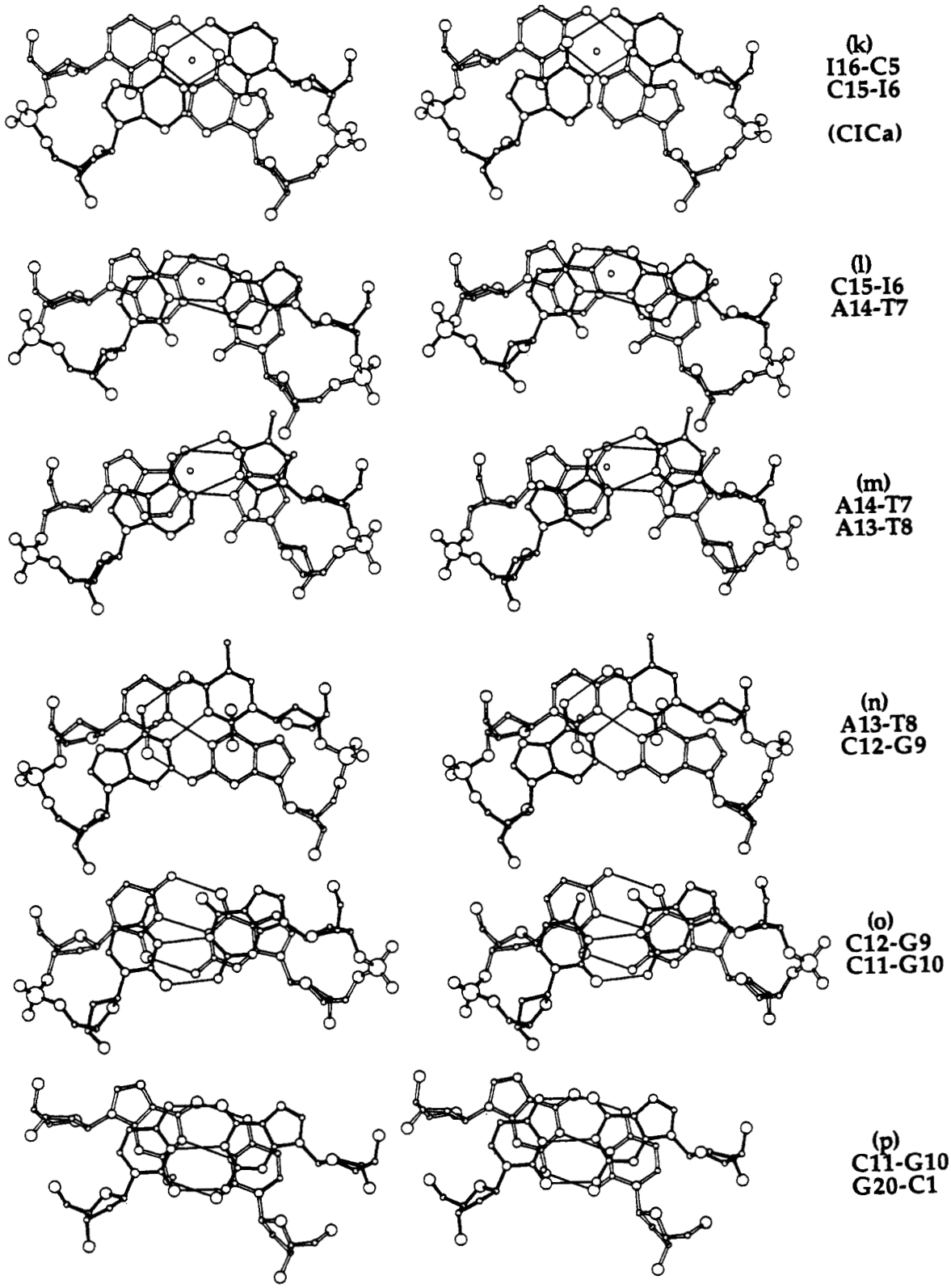

FIGURE 12k–p

symmetry axis. With CIMg (Figure 14a), the donating cytosines in the upper cluster occur three and four base pairs away from the symmetrical accepting phosphate on the same helix as the lower cluster. Because of rotation of the helix, the nearer end of each base pair to its receiving phosphate occurs on strand 1 of the helix.

In contrast, for the *Nar*I dodecamer (Figure 14b), the bridging cytosines are found only two and three base pairs away from the accepting phosphate of the other cluster. For the third base pair, the end nearest the accepting phosphate again occurs on strand 1 (Timsit and Moras' "A strand"). But the second base pair has rotated less, so its strand 2 end is closer to the phosphate. Hence for the spacing between bonding clusters shown in Figure 14a, the appropriate sequence for two cytosine hydrogen-bond donors is C-C, whereas, for the smaller spacing of Figure 14b, the appropriate sequence is C-G (not G-C). As a consequence of this one base pair difference in bonding geometry, for the *Nar*I dodecamer a given phosphate backbone lies closer to that wall of the enclosing major groove that is reached by crossing the central

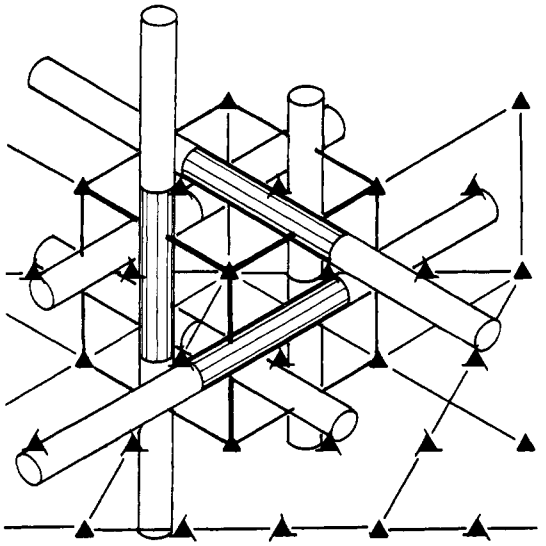

FIGURE 13: Packing of *Nar*I dodecamers in rhombohedral (trigonal) space group *R*3, as found by Timsit et al. (1989, 1991) and Timsit and Moras (1991). The view is down the *c* axis of the trigonal cell or along the body diagonal of the rhombohedral cell. The diameter of each B-DNA cylinder has been drawn at one-third its actual value for clarity. If true diameters were drawn, the three shaded helices, with an effective diameter of 18.5 Å, would be seen to be in touching contact around the central 3-fold axis. The three shaded helices are to be associated formally with this rhombohedral cell; all other unshaded helices belong to neighboring cells. A packing diagram for the rhombohedral G/C decamer (Heinemann et al., 1992) would be very similar. Because the *Nar*I dodecamer involves more than one turn of helix, a reverse twist of –26° is necessary at each interhelix junction along a rod. But the G/C decamer contains just one turn, so its interhelix junction twist angle is a normal 37°.

Table II: Frequency of Occurrence of B_{II} vs B_I Phosphate Conformations in DNA Decamer Crystal Structures

code	SpGp	sequence[a]
CIMg	t	C°C°A°A–C°I–T°T°G°G
CICa	m	C°C*A°A–C*I–T°T*G°G
CG	m	C°C*A°A–C°G–T°T*G°G
HA	m	C°C*A°G*G–C°C°T*G°G
KK	o	C*G°A–T°C°G*A–T°C°G
TA	o	C*G°A–T°T°A*A–T°C°G
AT	o	C*G°A–T°A–T*A–T°C°G
HAMe	h	C*C*A°G*G–C^{me}C°T*G°G

YR frequency		RR frequency		YY frequency	
CA	4/5	GG	2/7	CC	1/7
TG	4/5	AA	1/4	CT	0/2
CI	1/2	GA	1/4	TT	0/4
CG	3/8	AG	0/2	TC	0/4
TA	1/3				
all YR	13/23	all RR	4/17	all YY	1/17
RY frequency	0/15				

[a] (*) YR or RR step with B_{II} phosphate conformation; (°) YR or RR step with B_I phosphate conformation; (–) RY step, always B_I. SpGp, space group type: t = trigonal; m = monoclinic; o = orthorhombic; h = hexagonal.

dyad axis [see Figure 1a of Timsit and Moras (1991)], whereas for CIMg it is closer to the other, "outer" wall (Figure 5c).

This difference in contact geometry follows from the difference in base sequence in the two structures. The minimum requirement for stabilizing such a crossover of helices seems to be the availability of hydrogen-bond-donating major groove N4 amines on cytosines of two successive base pairs. If the donor cytosines occur on the same strand, then the symmetrical phosphate acceptor on that same helix ideally should be spaced three steps away as with CIMg; but if the cytosines occur on opposite strands, then the symmetrical

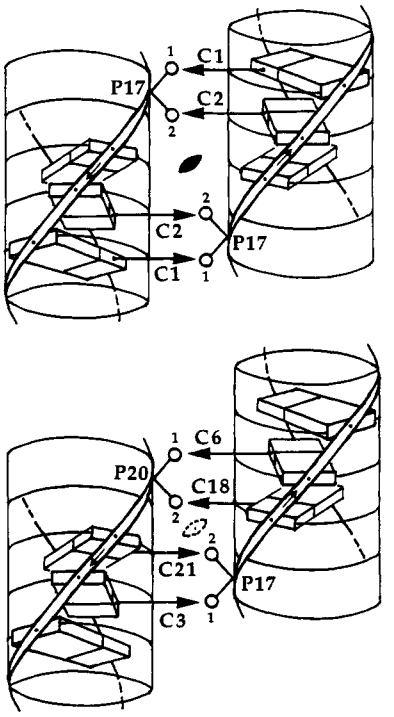

FIGURE 14: Schematic diagrams of the bridging of crossed helices in (a, top) the CIMg decamer and (b, bottom) the *Nar*I dodecamer, viewed down the 2-fold axis as in Figure 5c. Each base pair is drawn as a rectangular slab, divided into two unequal parts to represent pyrimidine and purine bases. Phosphate atoms are numbered 1 for O1P (toward minor groove) and 2 for O2P (toward major groove). Large arrows signify hydrogen bonds to phosphate oxygens from N4 amine groups of cytosines. Small arrows along backbone ribbons indicate 5′-to-3′ chain directions. Phosphates and bridging cytosines within one helix are one base pair farther apart in panel a than in panel b. The 2-fold axis between the helices is exact in panel a but is only approximate in panel b.

phosphate should be only two steps away as in *Nar*I. Either a sequence C-C or C-G can be used for such a helix-crossing junction. From Figure 14b one can deduce that the sequence G-C would be inherently unfavorable, because the bond-donating amines then would occur on the end of each base pair that is more distant from the phosphate acceptor.

The situation for the G/C decamer is intermediate to these cases: One of its two bridging contacts involves cytosines C11 and C12 on the same chain, whereas the other contact involves cytosines C5 and C15 on opposite strands at a C-G sequence step.

A Model for the Holliday Junction. Timsit et al. (1989, 1991) and Timsit and Moras (1991) have observed that the interlocking-groove crossing of two helices seen in their *Nar*I dodecamer provides a structural model for the Holliday junction, a postulated intermediate in genetic recombination (Holliday, 1964; Broker & Lehman, 1971; Sigal & Alberts, 1972). In the formation of such a junction, after two double-helical chromosomal strands are arranged side-by-side with analogous genes aligned, two parallel strands break and reform with one another as shown in Figure 15a. This crossed-chain, parallel-helix form of the Holliday junction can be converted to the topologically equivalent uncrossed, antiparallel form by simple end-for-end rotation of one of the helices (Figure 15b). Lilley and co-workers have established experimentally by a combination of gel electrophoresis and fluorescence spectroscopy that this uncrossed, antiparallel form with helices at an angle of 60° (120° in our reference frame) is the probable structure of a Holliday junction in the presence of metal cations (Duckett et al., 1988; Murchie et al., 1989). The cations shield phosphate–phosphate repulsion at especially

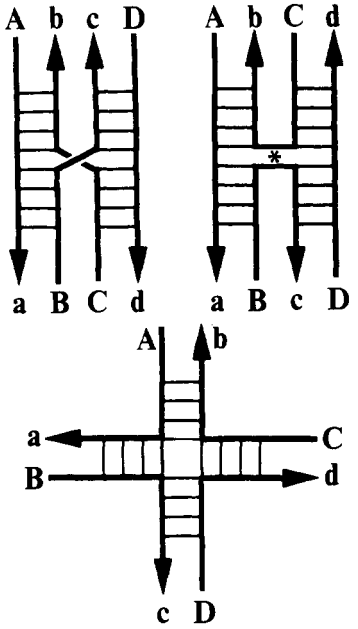

FIGURE 15: Holliday junction. (a, top left) Crossed or parallel-helix form produced by orienting double helices A–a/B–b and C–c/D–d so that corresponding genes are aligned, followed by strand exchange in the middle of strands B–b and C–c. (Strands are marked by a capital letter at the 5′ end and lower case letter at the 3′ end.) (b, top right) Uncrossed or antiparallel-helix form obtained by 180° rotation of helix C–d/D–d. The crossover point now is marked by an asterisk (*). (c, bottom) Open or maximally extended form.

close approaches of backbone. The open, square-planar form of Figure 15c is believed to occur only in the absence of metal cations, when phosphate–phosphate repulsions can exert their maximal effect (Duckett et al., 1990; von Kitzing et al., 1990).

Figure 16a gives a more detailed picture of the uncrossed, antiparallel junction diagrammed in Figure 15b. At first glance this appears to be only an interlocking of two discrete helices like that seen in Figure 6. Close observation is required to detect the critical chain crossing at the center, as marked by an asterisk (*). If the two nearly horizontal chains just above and below the asterisk were to be broken, and chains C–c and B–b were rejoined, as in Figure 16b, then there would result a nested-groove crossing of two discrete helices in what could be termed a "proto-Holliday junction". But this proto-junction is not quite what is seen in either *Nar*I or CIMg, whose nesting structure appears as Figure 16c. In both of these structures, each helix rests an "elbow" of backbone in the minor groove of the other, as is visible in Figure 5c of this paper and Figure 1a of Timsit and Moras (1991). In forming the proto-junction, as the two chains to be crossed are drawn closer together (the two phosphate acceptor chains of Figure 14), each of the two backbone "elbows" is popped out of its minor groove, both walls of which now sits wholly within the major groove of the other helix (Figure 16b). This fitting of the minor groove of one helix into the major groove of a neighbor also is encountered in the orthorhombic decamers [Figure 1 of Grzeskowiak et al. (1991)], although the contact is necessarily less intimate there because the helices are parallel rather than tilted.

Divalent magnesium cations are an important component of this process. They are essential in formation of good crystals of both *Nar*I and CIMg. If too small a cation-to-DNA ratio is used, *Nar*I crystals are of poor quality and do not diffract well (Timsit & Moras, 1991). In the present study, substitution of calcium for magnesium ions causes the CI decamer to crystallize in an entirely different space group.

The hydrated magnesium cations appear to play two distinct roles: (1) neutralization of phosphate–phosphate charge repulsion at the very close contacts needed for this type of helix packing and (2) specific hydrogen bonding across the junction from one helix to the other. Indeed, in the trigonal KKMe structure (Baikalov et al., 1993), one of the magnesium ions actually employs a phosphate oxygen as one of its six octahedral ligands.

Hence a groove-meshing crossover of two helices of the type encountered in CIMg, KKMe, *Nar*I, and G/C can provide a mechanism for formation of a Holliday junction that requires only local chain breaking and displacement. It is, in effect, a nascent junction. Furthermore, as discussed by Timsit and Moras (1991) and by Baikalov et al. (1993), all that is required for branch migration of the junction is counterrotation of the nested helices about their respective long axes. The angle between chains, 120° for CIMg and KKMe, 106.4° for *Nar*I, and 102.9° for G/C, corresponds to the 120° that was expected from solution spectroscopy studies of actual junctions. The role of magnesium ions is understood, in stabilizing both the Holliday junction and the crossed helices from which it

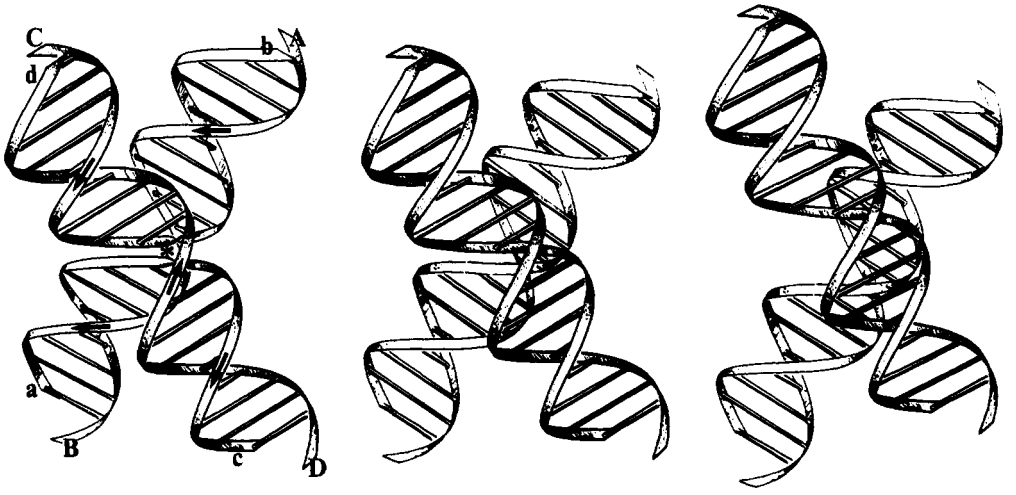

FIGURE 16: Progression from the Holliday junction to two interlocked helices as seen in CIMg. (a, left) Helix diagram of the antiparallel, uncrossed Holliday junction (Figure 15b) as established by Murchie et al. (1989) from fluorescence energy transfer experiments. Adapted from Bhattacharyya et al. (1991). The view is that of Figure 5a. Chain crossover at the asterisk (*) produces a Holliday junction from two crossed helices. (b, middle) Uncrossing of chains at the asterisk to yield a proto-junction of two meshed helices. (c, right) Separation of the formerly crossed chains to positions as observed in the *Nar*I and CIMg structures. Note that this requires the clicking past one another of two backbone "screw threads", so that an elbow of backbone chain from one helix now sits in the minor groove of the other.

Table III: Torsion Angles for the Monoclinic CICa Decamer

	Alpha	Beta	Gamma	Delta	Pseud	Epsil	Zeta	Chi	Ep-Ze
Str.1	---	---	42.7	108.5	99.7	193.6	277.8	231.9	-84.1
	286.4	177.7	46.5	147.8	151.5	257.4	165.0	275.3	92.4
	285.9	149.3	50.7	142.4	173.6	176.6	272.0	271.3	-95.4
	288.7	170.8	53.8	113.3	119.3	173.3	264.4	253.2	-91.1
	305.1	175.2	48.3	146.5	147.3	239.3	182.2	261.4	57.1
	292.5	151.9	44.4	137.2	145.2	184.4	259.3	247.7	-75.0
	297.8	176.2	48.6	112.0	116.1	187.6	273.1	240.1	-85.5
	290.6	183.3	46.8	135.9	137.5	257.5	165.4	267.9	92.1
	284.8	150.4	49.5	143.9	179.5	186.8	272.3	280.7	-85.6
	286.5	163.7	53.1	104.9	96.7	---	---	250.4	---
Str.2	identical to Strand 1								
AV =	290.9	166.5	48.4	129.2	136.6	206.3	236.8	258.0	-30.6
SD =	6.5	12.7	3.4	16.9	27.7	33.7	48.5	15.5	81.6
DI =	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

Table IV: Base Step Parameters for the Monoclinic CICa Decamer

	Roll	Tilt	Cup	Slide	Twist	Rise	Dxy1	Dxy2
C1-C2	6.59	4.10	-7.71	0.86	28.1	3.97	2.87	2.87
C2-A3	-16.02	-0.64	-2.08	2.64	51.3	2.98	4.42	4.42
A3-A4	10.03	-4.50	-0.29	0.21	27.9	3.91	2.88	2.88
A4-C5	-0.70	1.29	-5.48	-0.35	29.4	3.43	2.95	2.95
C5-I6	-2.71	0.00	14.00	0.31	48.2	2.76	4.48	4.48
I6-T7	-0.70	-1.29	-5.48	-0.35	29.4	3.43	3.65	3.65
T7-T8	10.03	4.50	-0.29	0.21	27.9	3.91	2.69	2.69
T8-G9	-16.02	0.64	-2.08	2.65	51.3	2.98	4.73	4.73
G9-G10	6.59	-4.10	-7.71	0.86	28.1	3.97	2.46	2.46
G10-C1	2.92	0.00	17.11	-0.41	38.3	2.77	4.21	4.21
AV =	0.00	0.00	0.00	0.66	36.00	3.41	3.53	
SD =	9.28	2.87	8.44	1.11	10.08	0.50	0.84	
DI =	0.00	0.00	0.00	0.00	0.00	0.00	0.00	

originated. These four oligonucleotide structures are the closest that we have yet come to examining the crystal structure of the elusive Holliday junction.

Sequence–Structure Variability in B-DNA. This CI sequence, C-C-A-A-C-I-T-T-G-G, has shown that one and the same sequence can exhibit quite different local helix parameters under different circumstances. Yanagi et al. (1991) observed a strong linkage between the four parameters, twist, rise, cup, and roll, and defined two extreme patterns of behavior: (a) high twist profile (HTP), characterized by high twist, low rise, positive cup, and negative roll, and (b) low twist profile (LTP), with low twist, large rise, negative cup, and positive roll. These linkages remain valid for the new structures. R-R steps (except for G-A) generally exhibit low twist behavior, with good base stacking of purines and with the other ends of the base pairs pivoted so that the polar O2 atom of one purine can stack atop the aromatic six-membered ring of its neighbor. R-Y steps (except for G-C) show low to moderate twist behavior because they must rely only on ring-upon-ring stacking. Y-R steps, in contrast, are bimodal, capable of showing either HTP or LTP behavior. In the LTP mode, the pyrimidine O2 of one base pair tends to stack over the five-membered ring of the following purine, whereas in the HTP mode the O2 sits over the six-membered purine ring, and the two base pairs are separated by an appreciable slide. the two steps that in other helices have been found to be consistently HTP, G-C and G-A (= T-C), are not present in this CI sequence.

At the penultimate C–A (= T-G) steps of the CI decamer, the HTP mode is used in the monoclinic crystal form, whereas a lower twist behavior is seen at exactly the same sequence position in trigonal crystals. It is apparent that a sequence that was incapable of HTP behavior could not adopt this particular monoclinic packing scheme, and hence, in a very real sense, individual base sequences "choose" their crystal forms. The choice of monoclinic vs trigonal packing for the present CI decamer sequence probably is to be attributed to the choice of divalent cation present, Ca^{2+} or Mg^{2+}. With calcium, the ions sit within the minor groove as has been seen in other crystal structures; with magnesium, the ions sit in the major groove, in both cases making hydrogen-bonded bridges to the phosphate backbone of neighboring molecules.

Table V: Base Pair Parameters for the Monoclinic CICa Decamer

	Tip	Incl	Prop	Buck	X Dsp	Y Dsp
C1	1.46	2.41	-12.01	8.56	1.72	-3.76
C2	8.00	6.44	-10.48	0.85	-0.39	-3.53
A3	-7.93	5.81	-5.01	-1.23	-1.90	0.03
A4	2.06	1.36	-14.50	-1.51	-1.71	0.92
C5	1.35	2.65	-7.18	-7.00	-0.75	1.06
I6	-1.35	2.65	-7.18	7.00	0.34	1.38
T7	-2.06	1.36	-14.50	1.51	0.45	0.66
T8	7.93	5.81	-5.01	1.23	0.75	0.37
G9	-8.00	6.44	-10.48	-0.85	1.57	1.70
G10	-1.46	2.41	-12.01	-8.56	2.63	1.25
AV =	0.00	3.73	-9.84	0.00	0.27	0.00
SD =	5.33	2.06	3.47	5.17	1.44	1.93
DI =	0.00	0.00	0.00	0.00	1.75	1.75

Table VI: Torsion Angles for the Trigonal CIMg Decamer

	Alpha	Beta	Gamma	Delta	Pseud	Epsil	Zeta	Chi	Ep-Ze
Str.1	---	---	11.3	157.8	207.5	180.4	240.3	255.3	-59.9
	57.6	137.7	292.2	151.6	192.7	168.9	230.3	297.9	-61.4
	324.9	147.5	38.2	145.7	178.5	218.0	189.2	271.2	28.9
	332.6	129.3	34.9	130.2	160.6	206.9	238.8	255.6	-31.9
	315.0	148.2	54.1	102.0	120.8	205.1	251.1	224.4	-46.0
	320.0	163.2	26.9	137.5	151.6	187.4	257.7	246.1	-70.3
	292.8	205.7	27.9	154.5	174.3	187.9	245.2	270.7	-57.3
	317.5	194.4	16.8	140.9	192.2	186.4	266.8	257.3	-80.4
	273.2	196.0	48.9	107.6	114.8	48.6	0.0	264.8	48.6
	159.7	256.5	104.2	102.3	109.7	---	---	246.3	---
Str.2	---	---	338.8	145.5	195.9	132.5	304.3	282.6	-171.8
	266.2	231.0	46.7	147.0	175.7	181.7	201.0	300.7	-19.3
	324.1	148.2	45.0	145.8	182.9	187.0	235.9	258.6	-48.9
	312.4	160.9	47.2	125.9	136.2	184.9	251.2	253.8	-66.3
	318.6	171.2	40.4	130.1	142.4	202.4	252.6	246.2	-50.2
	282.9	163.1	45.4	133.8	155.1	185.9	246.7	257.4	-60.8
	308.1	160.3	42.4	85.4	92.4	118.8	316.4	233.0	-197.6
	265.1	218.0	55.2	112.6	124.7	175.6	254.0	256.9	-78.4
	331.0	160.6	21.9	125.5	160.0	171.6	253.4	271.3	-81.9
	316.7	156.6	24.8	127.7	138.3	---	---	255.3	---
AV =	284.4	174.9	68.2	130.5	155.3	173.9	235.3	260.3	-61.4
SD =	69.7	34.6	87.1	19.8	32.4	39.4	65.6	18.8	57.1
DI =	61.4	39.2	78.4	19.4	30.6	35.6	55.8	13.1	61.4

Table VII: Base Step Parameters for the Trigonal CIMg Decamer

	Roll	Tilt	Cup	Slide	Twist	Rise	Dxy1	Dxy2
C1-C2	5.17	1.35	-17.28	0.32	27.0	3.78	2.55	2.50
C2-A3	1.90	2.00	13.91	0.58	37.3	3.35	3.31	3.60
A3-A4	-1.06	-1.21	-3.55	0.21	36.4	3.23	3.63	3.41
A4-C5	-0.11	-1.87	5.40	-0.49	31.3	3.18	3.55	3.85
C5-I6	2.82	-0.43	6.05	-0.08	40.9	3.02	4.16	3.74
I6-T7	-5.23	1.17	1.16	-0.16	36.2	3.10	3.76	3.63
T7-T8	-2.49	-1.01	-1.73	0.08	34.7	3.16	3.88	3.65
T8-G9	3.94	0.84	-3.44	0.89	35.0	3.73	2.83	3.53
G9-G10	-1.53	0.14	-11.54	0.86	30.6	3.75	3.20	2.94
G10-C1	-3.40	-0.97	11.01	0.41	50.7	2.82	5.50	4.26
AV =	0.00	0.00	0.00	0.26	36.01	3.31	3.57	
SD =	3.28	1.26	9.41	0.43	6.30	0.33	0.66	
DI =	3.06	1.73	5.83	0.26	2.50	0.11	0.38	

Table VIII: Base Pair Parameters for the Trigonal CIMg Decamer

	Tip	Incl	Prop	Buck	X Dsp	Y Dsp
C1	-4.41	2.13	-1.86	3.35	-1.23	-2.08
C2	0.75	3.48	-7.06	-13.94	-1.83	-1.27
A3	2.64	5.48	-7.42	-0.02	-1.98	0.24
A4	1.59	4.27	-13.23	-3.57	-1.54	1.27
C5	1.48	2.40	-12.60	1.83	-0.83	1.26
I6	4.30	1.97	-11.05	7.88	-0.08	1.40
T7	-0.92	3.14	-8.87	9.04	0.72	0.85
T8	-3.41	2.13	-10.33	7.31	0.78	0.27
G9	0.52	2.97	0.45	3.87	1.31	0.34
G10	-1.01	3.10	-5.01	-7.67	1.11	0.32
AV =	0.00	3.11	-7.70	0.00	-0.36	0.00
SD =	2.59	1.07	4.38	7.20	1.24	1.13
DI =	2.78	1.28	3.90	7.37	2.25	1.60

DNA sequence–structure relationships have proven over recent years not to be matters of simple grammar, but to be more akin to poetry, with an underlying form that is hidden from view by circumstance. These subtle relationships demand more from the viewer than would a mere table of grammar, but their rewards are commensurately greater. The investigator is forced to second guess four billion years of oligonucleotide evolution and in amazement to ask with Blake, "What immortal hand or eye, can frame thy fearful symmetry?"

ACKNOWLEDGMENT

We thank Dr. Kazimierz Grzeskowiak for advice and help in DNA synthesis and Prof. David Lilley for helpful conversations regarding conditions for obtaining Holliday junctions.

APPENDIX: HELIX PARAMETER TABLES FOR THE CICA AND CIMG HELICES

Tables III–V and VI–VIII list local helix parameters for the monoclinic CICa and trigonal CIMg structures, respectively. For an explanation of individual parameters, see the appendix to Privé et al. (1991). Parameters are calculated using NEWHEL92, a copy of which may be obtained from the Brookhaven Protein Data Bank or directly from the authors at E-mail address RED@UCLAUE.

REFERENCES

Baikalov, I., Grzeskowiak, K., & Dickerson, R. E. (1983) *J. Mol. Biol.* (in press).

Bhattacharyya, A., Murchie, A. I. H., von Kitzing, E., Diekmann, S., Kemper, B., & Lilley, D. M. J. (1991) *J. Mol. Biol. 221*, 1191–1207.

Broker, T. H., & Lehman, I. R. (1971) *J. Mol. Biol. 60*, 131–149.

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

Chuprina, V. P., Heinemann, U., Nurislamov, A. A., Zielenkiewicz, P., Dickerson, R. E., & Saenger, W. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 593–597.

Dickerson, R. E. (1990) in *Structure and Methods (Proceedings of the Sixth Conversation in Biomolecular Stereodynamics), DNA & RNA* (Sarma, R. H., & Sarma, M. H., Eds.) Vol. 3, pp 1–38, Adenine Press, Schenectady, NY.

Dickerson, R. E. (1992) In *Methods in Enzymology: DNA Structures*, (Lilley, D. M. J., & Dahlberg, J., Eds.) Vol. 211, pp 67–111, Academic Press, New York.

Dickerson, R. E., Goudsell, D. S., Kopka, M. L., & Pjurà, P. E. (1987) *J. Biomol. Struct. Dyn. 5*, 557–580.

Duckett, D. R., Murchie, A. I. H., Diekmann, S., von Kitzing, E., Kemper, B., & Lilley, D. M. J. (1988) *Cell 55*, 79–89.

Duckett, D. R., Murchie, A. I. H., & Lilley, D. M. J. (1990) *EMBO J. 9*, 583–590.

Grzeskowiak, K., Yanagi, K., Privé, G. G., & Dickerson, R. E. (1991) *J. Biol. Chem. 266*, 8861–8883.

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*, 7312–7341.

Heinemann, U., Alings, C., & Bansal, M. (1992) *EMBO J. 11*, 1931–1939.

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, India.

Holliday, R. (1964) *Genet. Res. 5*, 282–304.

Murchie, A. I. H., Clegg, R. M., von Kitzing, E., Duckett, D. R., Diekmann, S., & Lilley, D. M. J. (1989) *Nature 341*, 763–766.

Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L., & Dickerson, R. E. (1987) *Science 238*, 498–504.

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

Quintana, J. R., Grzeskowiak, K., Yanagi, K., & Dickerson, R. E. (1992) *J. Mol. Biol. 255*, 379–395.

Sigal, N., & Alberts, B. (1972) *J. Mol. Biol. 71*, 789–793.

Timsit, Y., & Moras, D. (1991) *J. Mol. Biol. 221*, 919–940.

Timsit, Y., Westhof, E., Fuchs, R. P. P., & Moras, D. (1989) *Nature 341*, 459–482.

Timsit, Y., Vilbois, E., & Moras, D. (1991) *Nature 354*, 167–170.

von Kitzing, E., Lilley, D. M. J., & Diekmann, S. (1990) *Nucleic Acids Res. 18*, 2671–2683.

Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 184*, 119–145.

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

Yuan, H., Quintana, J., & Dickerson, R. E. (1992) *Biochemistry 31*, 8009–8021.