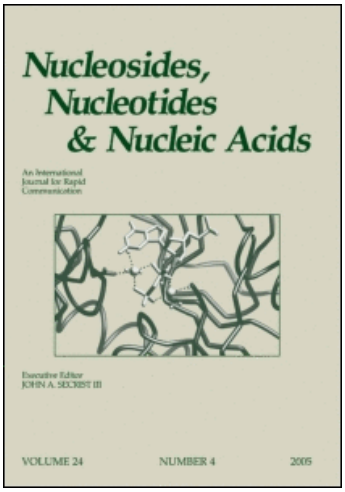

Nucleosides, Nucleotides and Nucleic Acids

Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597286

Polymorphism, Packing, Resolution, and Reliability in Single-Crystal DNA Oligomer Analyses

Richard E. Dickerson[a]; Kazimierz Grzeskowiak[a]; Maria Grzeskowiak[a]; Mary L. Kopka[a]; Teresa Larsen[a]; Andrei Lipanov[a]; Gilbert G. Privé[b]; Jordi Quintana[b]; Peter Schultze[b]; Kazunori Yanagi[b]; Hanna Yuan[b]; Hyo-Chun Yoon[b]

[a] Molecular Biology Institute, University of California, Los Angeles, CA, U. S. A. [b] Laboratory of Chemical Biodynamics, Lawrence Berkeley Laboratory, Building 3, University of California, Berkeley, CA

NUCLEOSIDES & NUCLEOTIDES, 10(1-3), 3-24 (1991)

POLYMORPHISM, PACKING, RESOLUTION, AND RELIABILITY IN SINGLE-CRYSTAL DNA OLIGOMER ANALYSES

Richard E. Dickerson*, Kazimierz Grzeskowiak, Maria Grzeskowiak, Mary L. Kopka, Teresa Larsen, Andrei Lipanov, Gilbert G. Privé†, Jordi Quintana, Peter Schultze, Kazunori Yanagi, Hanna Yuan, and Hyo-Chun Yoon

Molecular Biology Institute, University of California, Los Angeles CA 90024, U. S. A.
†Present address: Laboratory of Chemical Biodynamics, Lawrence Berkeley Laboratory, Building 3, University of California, Berkeley CA 94720

Abstract: Synthetic double-helical B-DNA oligonucleotides crystallize in orthorhombic, monoclinic, and trigonal patterns that are determined by fine details of intermolecular contacts. Crystal packing apparently has relatively little influence on local helix structure, and noncrystallographic symmetry differences can be used to evaluate the quality of analyses.

Guardian Angel of Molecular Biologists: "I have some good news and some bad news for you."

Molecular Biologist: "What's the good news?"

G.A.M.B.: "You will have some crystals to work with."

M.B.: "So what's the bad news?"

G.A.M.B: "You will have to work with some crystals."

To a molecular biologist who wants to learn the structure of a macromolecule, crystals are both a blessing and a curse. By ensuring that several trillion molecules scatter X-radiation cooperatively, the crystal amplifies the scattering signal to the point where it is experimentally measurable. To date no other physical method can yield the same profusion of precise molecular information as does x-ray diffraction. But a crystal of macromolecules can only be built up via touching contacts between these molecules. If the contacts are loose, the accuracy of orientation of the trillions of molecules in the crystal is poor, and the diffraction pattern consequently has low resolution. In contrast, if the contacts are tight enough to ensure an orderly crystal lattice and a high resolution diffraction pattern, then these tight contacts have

a significant probability of deforming the macromolecule, giving it structural features that it does not have when free in solution. A macromolecular crystallographer must always ask which features of the final structure are inherent to the free molecule, and which features may have been introduced by crystal packing.

The structure that results at the end of an x-ray crystal analysis has three components:

1. Structural features inherent to the free macromolecule as it would exist in solution.
2. Local changes in structure caused by intermolecular packing interactions in the crystal.
3. Experimental errors arising from limitations on the data and on the refinement process.

We ideally would like to study #1, control #2, and minimize if not eliminate #3. #2 can give us useful information about the deformability of the DNA helix, of a type that may be relevant to the binding and recognition of DNA by proteins. How do we evaluate #3 and #2, in order to focus on #1?

I. Polymorphism in B-DNA Oligomer Crystals

For nearly a decade after single-crystal structure analyses of synthetic double-helical DNA oligomers became possible in 1979, the only examples of the biologically significant B-DNA were the "Drew dodecamer", C-G-C-G-A-A-T-T-C-G-C-G, and its sequence variants, including mismatches and drug complexes[1]. All of these various dodecamers were isomorphous, crystallizing in the same orthorhombic space group $P2_12_12_1$, with cell dimensions ca. 25 x 41 x 66 Å, and with one dodecamer helix or 12 base pairs per asymmetric unit. Although nearly all of the base sequences synthesized were self-complementary, so that the dodecamer helix when free in solution would have a molecular dyad or twofold axis perpendicular to the helix axis through the center of the molecule, these internal molecular dyads were never used as an element of crystal symmetry. The dodecamer helix crystallized as an asymmetric object, even in most cases exhibiting a gentle bend at one end but not the other. These analyses were only of intermediate resolution, 1.9 Å at best to 2.6 Å at worst, and all experienced identical crystal packing effects.

Then in 1987 a new class of B-DNA structures appeared: the high-resolution decamers. The improved diffraction patterns of these analyses, 1.3 to 1.6 Å, arose because of the orderly way in which helices were stacked within the crystals. Both a monoclinic and an orthorhombic crystal packing mode have been observed. More recently, other decamers have been found to adopt a trigonal packing, but these crystals have lower resolution even than those of the dodecamers, typically ca. 3 Å. This paper is concerned with the packing modes that DNA oligomers can adopt, their effects on helix symmetry, and limitations on our knowledge of the DNA structure as it would exist free in solution.

The four decamer structures that have been solved to date, and a representative selection of nine dodecamers for comparison purposes, are listed in TABLE 1. For brevity, these

TABLE 1. Decamer and Dodecamer Structure Analyses Discussed in This Paper

I. Decamers

Code	Sequence	Space Group	Resolu.	Reference
CG	C-C-A-A-C-G-T-T-G-G	C2	1.4	Privé et al. 1990[2,3]
HA	C-C-A-G-G-C-C-T-G-G	C2	1.6	Heinemann and Alings 1989[4]
GA	C-C-A-A-G-A-T-T-G-G	C2	1.3	Privé et al. 1987[5,6]
KK	C-G-A-T-C-G-A-T-C-G	$P2_12_12_1$	1.5	Grzeskowiak *et al.* 1990[7]

II. Dodecamers (all space group $P2_12_12_1$)

Code	Sequence	Resolu.	Reference
CO	C-G-C-A-A-A-T-T-T-G-C-G (+ Distamycin)	2.2	Coll *et al.* 1987[8]
DR	C-G-C-G-A-A-T-T-C-G-C-G	1.9	Drew *et al.* 1981[9]
DW	C-G-C-G-A-A-T-T-C-G-C-G	1.9	Westhof *et al.* 1985[10]
LA	C-G-T-G-A-A-T-T-C-A-C-G	2.5	Larsen *et al.* 1990[11] Narendra *et al.* 1990[12]
M7	C-G-C-G-A-A-T-T-^{5me}C-G-C-G	2.3	Fratini *et al.* 1982[13]
MD	C-G-C-G-A-^{6me}A-T-T-C-G-C-G	2.0	Frederick *et al.* 1988[14]
N1	C-G-C-G-A-T-A-T-C-G-C-G (+ Netropsin)	2.4	Coll *et al.* 1989[15]
NE	C-G-C-A-A-A-A-A-A-G-C-G	2.5	Nelson *et al.* 1987[16]
YO	C-G-C-A-T-A-T-A-T-G-C-G	2.2	Yoon *et al.* 1988[17]

Resolu. = Nominal resolution reported for the structure analysis, in Å.

thirteen helices usually will be identified by a two-letter code that appears at the left of the table. DR and DW are two independent refinements of the same data set by Jack-Levitt and Hendrickson-Konnert methods, respectively. CO and N1 are DNA/drug complexes, in which the drug molecule stabilizes the DNA and improves the resolution. N1 is orientation No. 1 of two simultaneous disordered drug molecule orientations that the investigators reported. NE is the nonself-complementary sequence paired with C-G-C-T-T-T-T-T-T-G-C-G.

The packing observed in all of the dodecamers is shown in FIGS. 1 and 2. Columns of double helices overlap their final two G·C base pair by turning their minor grooves toward one another. At each of the two base pair contacts of an overlap region, one hydrogen bond leads from a guanine N2 amine on the first helix to a guanine N3 acceptor atom on the second helix, and another bond leads back from the second helix's N2 to the N3 of the first helix. (See FIG. 2 of REF. 18.) As FIG. 1 shows, the base pairs on the two helices are twisted at the contact point, so that each base pair fits down into the minor groove of the other helix. This twisted contact

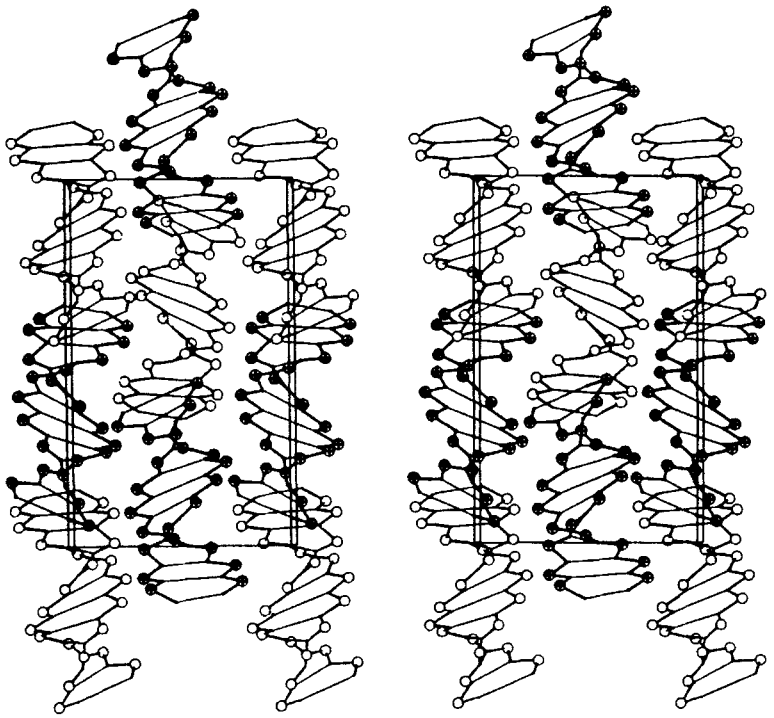

FIG. 1. Packing of DNA dodecamer double helices in the orthorhombic cell. Endless vertical columns of helices overlap their outermost two base pairs at each end, minor groove to minor groove. Noncrystallographic molecular dyads are approximately perpendicular to the page.

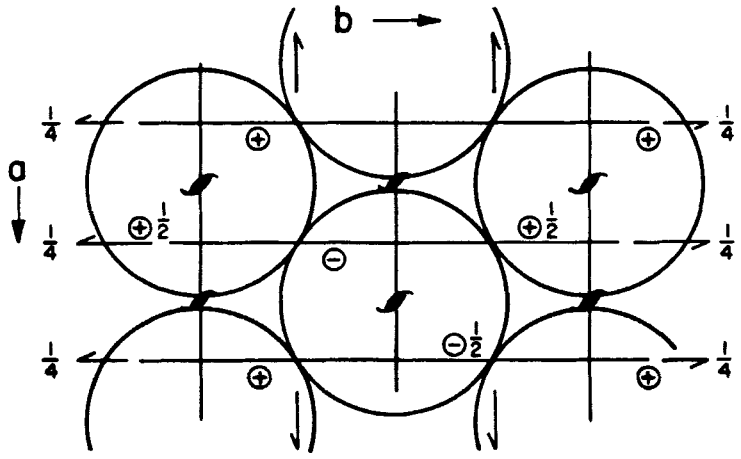

FIG. 2. End view of the orthorhombic cell, showing how the columns are in close-packed array along the c axis, with 24 Å between centers of columns.

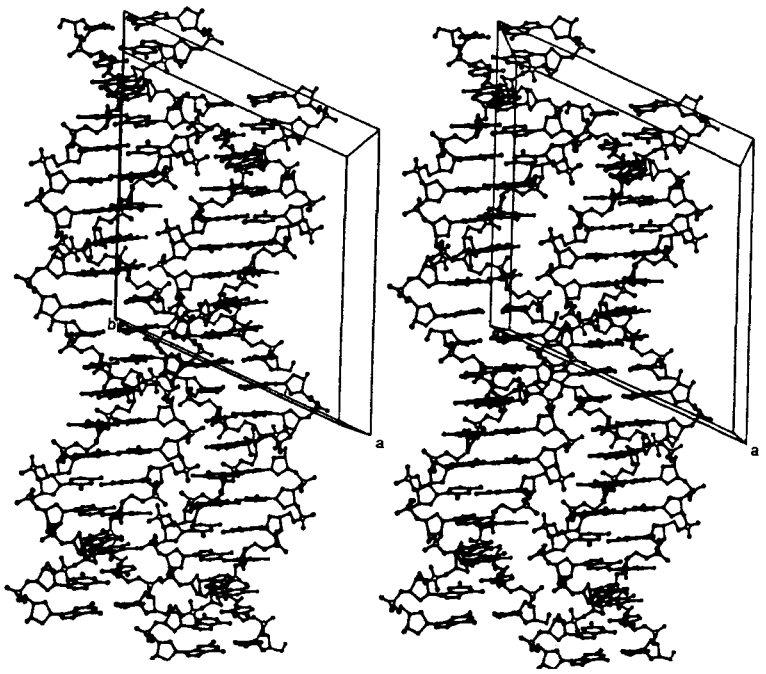

FIG. 3. Packing of DNA decamers in the monoclinic cell. Endless vertical columns of helices are stacked one atop another without overlap, and in a manner that closely simulates an infinite helix. Crystallographic dyads are perpendicular to the page. The columns are more orderly than in the dodecamers, and resolution is higher.

undoubtedly is the reason for the bend in helix axis that is exhibited by every dodecamer but M7. Such a twisted intermeshing of minor grooves gives the crystals sufficient stability to survive, but also introduces an element of disorder that limits these crystals to intermediate resolution. The twisted columns of dodecamers then are bundled side by side in a close-packed manner as in FIG. 2.

Packing in the monoclinic CG, HA and GA decamers is simpler and more orderly, and leads to a more orderly diffraction pattern. Ten base pairs constitute the natural helical repeat in B-DNA. As FIG. 3 shows, successive decamers stack atop one another in endless columns. (Two decamers are drawn in each column.) The nonbonded step from one helix to the next is scarcely different in any of its parameters from any of the internal steps of the helix[2-7], and indeed it takes a sharp eye to find the breaks between the two helices in either column of FIG. 3. The columns pack against one another in a square array as in FIG. 4. They slide past one another until the phosphate backbone chains of one helix interlock into the minor groove of the other, as one can see near the center of FIG. 3. The closest approach of the backbone chain of one column

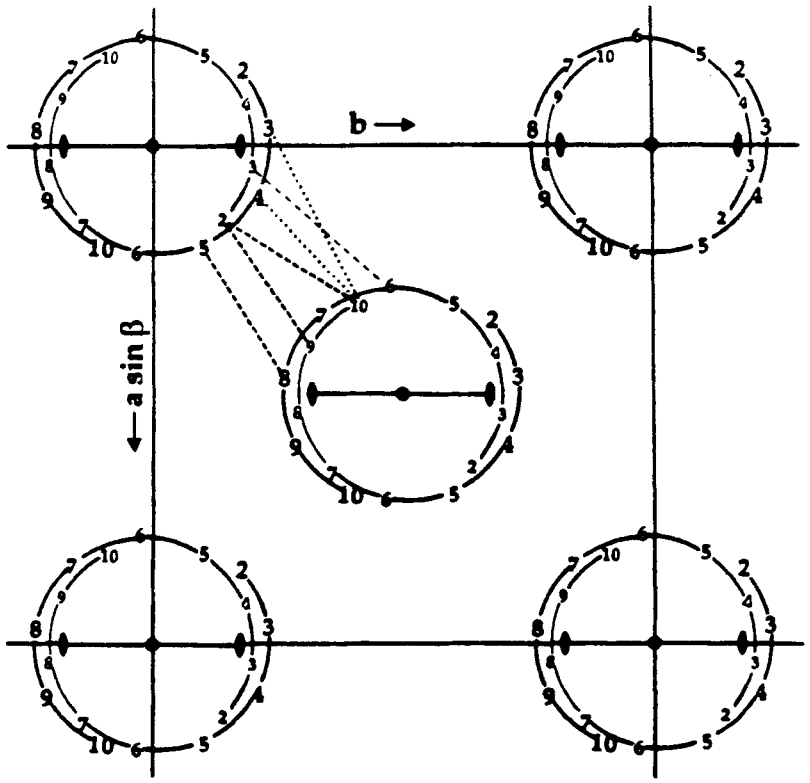

FIG. 4. End view of the monoclinic cell, showing how columns are in square array with 19.5 Å between centers along the cell diagonal, rather than being close-packed. Crystallographic dyads are horizontal, and inter-column contacts are at ±45° to the dyad direction.

occurs exactly where phosphates are absent at the nonbonded break between helices in the other column. This meshing and interlocking leads to axial displacement of neighboring columns, and to the observed monoclinic angle[2]. The internal dyad of each decamer double helix is used as a twofold symmetry axis of the crystal. As a consequence, the unique or asymmetric unit is only half a helix, or five base pairs. The two ends of the CG, HA and GA helices are required to be absolutely identical in the crystal. This is a quite different situation than was found in the dodecamers. The regular stacking of helices causes the diffraction patterns to show clusters of strong intensities around 3.4 Å in the ±*c* directions, which establishes the orientation of helix axes as being along *c*. With the dodecamers, the twisted overlaps of helices created enough variation in axis directions that the clusters of strong 3.4 Å reflections were less striking.

The orthorhombic KK crystals are similar to the monoclinic form in principle but quite different in detail. Again one encounters endless columns of regularly stacked helices (FIG. 5),

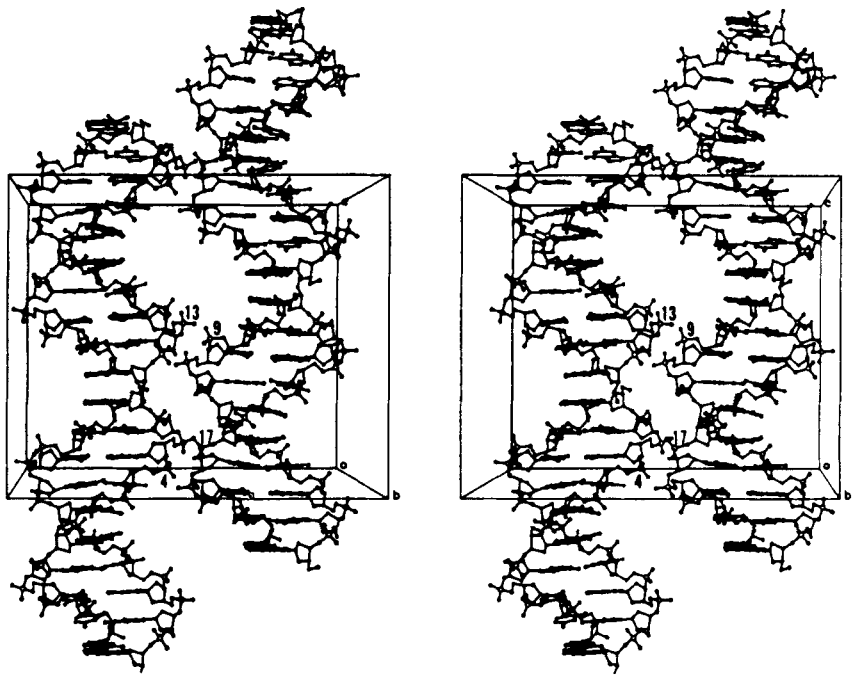

FIG. 5. Packing of DNA decamers in their orthorhombic cell. Endless vertical columns of stacked helices again simulate an infinite helix in an extremely orderly manner. Noncrystallographic molecular dyads in this view are horizontal.

and the orientation of the helices is obvious from strong 3.4 Å reflections in the $\pm c$ directions. But the columns are displaced differently. (Two decamers again are drawn in each of the two columns, but the gap between decamers is no easier to find than in FIG. 3.) The same kind of backbone-into-groove interlocking occurs, but the contact points are different. The internal helix dyad is horizontal in both FIGS. 4 and 6. In the orthorhombic KK helix (FIG. 6) the contact zones between columns occur along the internal dyad axis direction and at right angles to it. In contrast, in the monoclinic packing (FIG. 4), the contact zones along the columns are 45° to either side of the internal dyad. This leads to a different degree of sliding of the columns over one another to find the backbone-into-groove fit, and to the difference between monoclinic and orthorhombic cells. Both packing schemes, however, have comparable orderliness, and give rise to the same high 1.3 - 1.6 Å resolution. But the internal helix dyad is not used in the orthorhombic crystal form, so the asymmetric or unique unit is the entire 10 base pairs. Any observed experimental asymmetry between ends of the KK helix, or between ends of any of the dodecamers, cannot be a consequence of base sequence, as the two chains are identical and self-

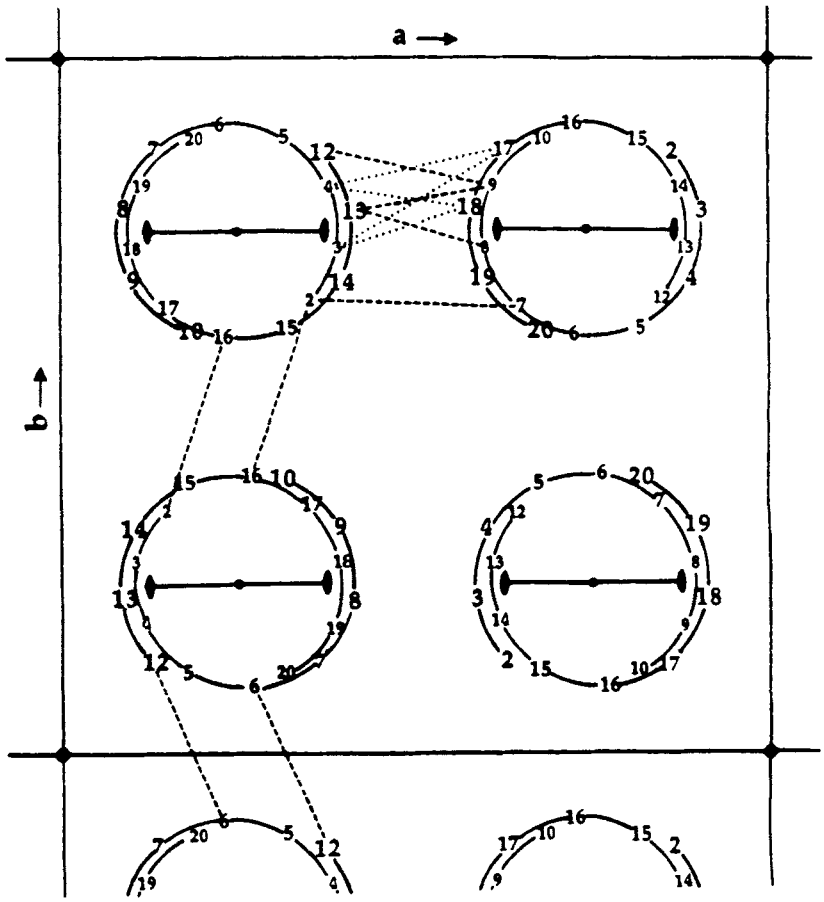

FIG. 6. End view of the orthorhombic decamer cell, showing how columns are in square array with 19.7 Å between centers. Noncrystallographic dyads are horizontal, and inter-column contacts are along and at right angles to the dyad axis direction.

complementary (with the exception of the NE helix). Hence one or both of the other two factors invoked earlier must be responsible: crystal packing, or errors in the data and refinement. We will come back to this idea at the end of the paper.

TABLE 2 shows that the orthorhombic packing of the KK helix also is adopted by other sequences, whose structures are under way but not complete as of the time of writing. The JTA and JAT sequences are variants of KK. JTA simply replaces the central pyrimidine-purine, C-G, with the other pyrimidine-purine, T-A. JAT reverses the central pair and yields a six base alternation, A-T-A-T-A-T, that is reminiscent of the YO helix. Crystals of both sequences grown in the presence of magnesium ions are isomorphous with KK, and both structures are currently

TABLE 2. Crystal Data for DNA Decamers

I. Monoclinic, Space group C2

	Sequence	*a*	*b*	*c*	*β*	Volume	V/bp	Bp/au	Res
CG*	C-C-A-A-C-G-T-T-G-G	32.25	25.53	34.38	113.4°	25,978	1299	5	1.4
GA*	C-C-A-A-G-A-T-T-G-G	32.52	26.17	34.30	118.9°	25,556	1278	5	1.3
HA*	C-C-A-G-G-C-C-T-G-G	32.15	25.49	34.82	116.7°	25,492	1275	5	1.6
Y7	C-G-C-T-T-A-A-G-C-G	77.27	25.97	79.17	104.5°	153,810	1282?	30?	1.5

II. Orthorhombic, Space group $P2_12_12_1$

	Sequence	*a*	*b*	*c*	Volume	V/bp	Bp/au	Res
KK*	C-G-A-T-C-G-A-T-C-G	38.93	39.63	33.30	51,375	1284	10	1.5
JTA	C-G-A-T-T-A-A-T-C-G, Mg^{2+}	38.72	39.24	33.23	50,488	1262	10	1.55
JAT	C-G-A-T-A-T-A-T-C-G, Mg^{2+}	38.64	39.54	33.67	51,442	1286	10	2.6
JTA	C-G-A-T-T-A-A-T-C-G, Ca^{2+}	36.04	39.09	33.04	46,547	1164	10	ca. 3
JAT	C-G-A-T-A-T-A-T-C-G, Ca^{2+}	35.35	38.46	33.36	45,355	1134	10	ca. 3

III. Trigonal, Space group $P3_121$ or $P3_221$, Helix axis in *ab* plane

	Sequence	*a*	*b*	*c*	Volume	V/bp	Bp/au	Res
^{me}KK	C-G-A-T-C-G-^{6me}A-T-C-G	33.3	33.3	97.8	93,900	1565	10	2.6
An1	CG + Anthramycin 1	33	33	103.7	97,900	1630	10	3.0
Str	CG + Streptonigrin	33.2	33.2	96.3	91,900	1531	10	ca.3
TA	C-C-A-A-T-A-T-T-G-G	37	37	57	67,600	1127	10	ca.3

IV. Trigonal, Space group $P3_121$ or $P3_221$, Helix axis inclined to *c*.

	Sequence	*a*	*b*	*c*	Volume	V/bp	Bp/au	Res
An2	CG + Anthramycin 2	24	24	89	38,450	1281	5	3.0
TL5	C-G-T-G-A-T-C-A-C-G	36.7	36.7	82.5	96,200	1603	10	2.9

V. Trigonal/Hexagonal, uncharacterized

Sequence	*a*	*b*	*c*	Volume	V/bp	Bp/au	Res
C-G-A-T-G-C-A-T-C-G + Actinomycin D	25	25	180	97,400	1623	10?	ca.3

* = Crystal structure has been solved. All others are in analysis or in preparation for analysis. V/bp = Volume per base pair. Bp/au = Number of base pairs per asymmetric unit.

being refined. Crystallization with calcium ions, in contrast, causes shrinkage of the *a* axis and changes in the intensity pattern as observed in survey precession photographs. In all cases the direction of the helix axes is clear from clusters of 3.4 Å reflections. Another rather unusual monoclinic cell has been found with the Y7 decamer, which apparently has three full helices per asymmetric unit!

But the monoclinic and orthorhombic decamers are not the entire story. A third very populous family has been found recently, that uses trigonal space group $P3_121$ or $P3_221$. (None of

these structures has yet been solved, so the sense of the threefold screw axis remains uncertain.) Merely methylating one adenine of the KK sequence in the major groove is sufficient to induce the decamer to adopt an entirely new packing scheme. Cocrystallization of the CG sequence with either of two drugs, anthramycin or streptonigrin, induces the same crystal form. The TA helix, which is to the CG helix as JTA is to KK, uses the same trigonal space group but is not isomorphous with the others, having quite different cell dimensions and molecular packing. All of these lack the high resolution of the monoclinic and orthorhombic decamers, but all of them show six diffuse 3.4 Å reflections in hk0 survey photographs that prove that helices are stacked in an orderly manner with their axes lying in the *ab* plane. The most probable model is one which has six layers of close-packed helix cylinders stacked up the *c* axis in pairs, with a reorientation of the helix axes by 120° from one pair of layers to the next.

In the anthramycin/CG complex just mentioned, anthramycin is covalently bound to the DNA via a minor groove N2 amine of a guanine. Two covalent fractions resulted from the reaction, and both have been purified and crystallized. The second fraction, An2, crystallizes in trigonal space group $P3_121$ (or $P3_221$), but the diffuse 3.4 Å smears 20° to either side of the *c* axis indicate that the helices are aligned around the threefold screw axis. The TL5 helix behaves similarly, with a broad cluster of 3.4 Å reflections resembling those of the dodecamers. A rotation function search suggests that the TL5 axes are tipped ca. 25° away from *c*.

Hence the B-DNA decamers have proven to be polymorphic, but within a rather limited range of crystal packing. The customary volume per base pair in crystals of B-DNA dodecamers and decamers has been 1200 to 1400 $Å^3$ /bp. We at first thought that the lower resolution of the trigonal crystals arose from their looser packing, with ca. 1600 $Å^3$/bp. But this cannot be the entire explanation; the orthorhombic JAT and JTA crystals grown in the presence of calcium ions, the TA helix, and An2 all have 1200 $Å^3$/bp but to date have shown no better than 3 Å resolution.

II. Accuracy of Noncrystallographic Symmetry in B-DNA Crystals

The extent of non-identity of the two ends of the KK decamer helix or of any of the dodecamers is a measure of the importance of crystal packing and/or errors in the structure analysis. TABLES 3 and 4 evaluate the degree of internal, noncrystallographic symmetry in the published structures. Coordinates for all save LA and NE are in the Brookhaven Data Bank, and the NE coordinates are in the Cambridge Data File. Local parameters were calculated for all 13 helices using NEWHEL90, and the parameters to be studied were extracted from the longer NEWHEL90 output files by means of the utility programs EXSTEP and EXBP. The symmetry analysis leading to TABLES 3 and 4 was done by the program SYMTEST. All of these programs are available upon request from the first author. Helix parameters will not be defined here, but are described in references 2, 3 and 19. In TABLES 3 and 4, AV is the mean value of the

TABLE 3. Noncrystallographic Symmetry in Local Helix Parameters for the KK Decamer and for Nine Isomorphous Dodecamers

Sym	Res		Roll	Cup	Twist	Rise	D_{xy}	Incl	Prop	Buckle
KK	1.6	AV	0.32	1.65	37.14	3.25	3.70	0.70	-11.48	-0.62
KK	1.6	SD	4.04	8.69	4.11	0.47	0.51	1.24	5.24	6.97
KK	1.6	DI	2.20	3.77	1.45	0.19	0.20	1.08	3.55	3.17
DR	1.9	AV	0.32	-0.37	35.86	3.34	3.56	0.72	-12.14	-0.35
DR	1.9	SD	5.13	7.58	4.09	0.45	0.64	5.86	5.72	5.27
DR	1.9	DI	4.36	5.57	2.46	0.16	0.44	9.36	4.12	5.23
DW	1.9	AV	0.81	-0.42	35.74	3.35	3.56	0.69	-11.65	-0.72
DW	1.9	SD	5.72	8.53	3.29	0.46	0.61	6.44	4.96	5.56
DW	1.9	DI	4.12	9.47	3.42	0.18	0.48	10.56	2.44	4.88
M7	2.3	AV	-2.29	-1.80	36.01	3.37	3.55	-2.31	-15.00	-2.95
M7	2.3	SD	5.77	14.26	4.67	0.55	0.66	3.73	7.85	10.90
M7	2.3	DI	5.02	6.47	3.36	0.11	0.28	4.31	4.79	6.37
LA	2.5	AV	0.61	0.12	35.98	3.33	3.53	0.68	-14.16	-0.42
LA	2.5	SD	4.37	6.75	4.28	0.41	0.61	5.70	4.82	4.90
LA	2.5	DI	3.47	5.56	3.96	0.20	0.54	9.27	5.14	4.81
NE	2.5	AV	0.48	1.66	36.72	3.30	3.59	0.12	-14.83	-0.61
NE	2.5	SD	5.79	9.28	3.71	0.40	0.74	4.67	7.12	7.12
NE	2.5	DI	6.07	6.78	2.42	0.27	0.50	6.85	6.26	5.14
MD	2.0	AV	0.01	-0.52	35.80	3.39	3.53	0.00	-10.79	-2.70
MD	2.0	SD	5.71	7.97	4.80	0.43	0.70	6.05	5.92	4.83
MD	2.0	DI	4.06	5.20	2.26	0.18	0.39	8.46	7.47	5.92
N1	2.4	AV	1.24	0.83	36.16	3.35	3.58	1.35	-9.43	-0.29
N1	2.4	SD	6.25	14.14	6.45	0.63	0.71	4.42	4.87	8.17
N1	2.4	DI	5.79	6.70	5.24	0.27	0.63	6.58	3.25	5.30
YO	2.2	AV	-0.54	1.12	36.10	3.32	3.51	-0.26	-13.75	0.26
YO	2.2	SD	9.57	11.07	4.29	0.36	0.55	6.15	6.60	8.59
YO	2.2	DI	10.77	11.28	2.76	0.33	0.54	9.03	10.34	9.20
CO	2.2	AV	0.18	0.85	36.60	3.30	3.67	0.32	-14.72	-1.28
CO	2.2	SD	5.56	13.40	5.19	0.45	0.63	4.64	5.64	9.65
CO	2.2	DI	6.37	12.43	4.50	0.34	0.49	6.87	5.72	8.24

The smallest two DI values per column are double underlined, and the largest two are italicised.

given parameter along the helix, and SD is the standard deviation. DI is the mean difference between parameter values that would be identical if the helix possessed a true internal symmetry dyad.

Res is the nominal resolution of the crystal structure analysis as reported by the original investigators. It is defined as the outer limit in the diffraction pattern of data used in the analysis, and defines the probable level of detail that one could expect from the analysis. But

TABLE 4. Noncrystallographic Symmetry in Torsion Angles for the KK Decamer and for Nine Isomorphous Dodecamers

Sym		α	β	γ	δ	ε	ζ	χ	ε-ζ	Pseud
KK	AV	296.3	169.4	54.4	128.0	191.9	251.6	252.8	-59.7	144.8
KK	SD	12.7	20.0	15.1	20.8	24.3	34.5	18.1	56.4	32.5
KK	DI	13.3	11.8	14.0	15.7	8.9	13.7	12.6	14.7	25.0
DR	AV	297.2	171.3	59.4	122.7	191.0	251.5	242.7	-60.5	127.5
DR	SD	8.3	14.2	25.5	20.7	25.2	34.5	13.5	58.7	31.3
DR	DI	6.5	9.6	18.4	16.5	10.2	12.8	10.3	20.0	24.8
DW	AV	293.6	170.4	64.6	126.8	194.4	251.0	246.3	-56.6	138.4
DW	SD	19.6	13.6	42.1	18.0	33.0	39.3	13.0	71.3	29.7
DW	DI	18.7	12.9	16.4	10.7	13.4	20.2	10.3	29.4	22.6
M7	AV	297.9	169.6	59.1	119.7	190.8	250.9	239.8	-60.0	122.3
M7	SD	10.4	17.7	23.8	27.3	32.2	42.4	18.3	73.0	37.7
M7	DI	10.7	16.1	19.1	19.3	22.4	27.6	13.5	42.5	25.0
LA	AV	285.8	175.1	75.2	127.3	198.6	247.0	251.1	-48.4	137.8
LA	SD	27.0	19.5	68.5	29.2	29.7	41.1	21.4	69.9	47.6
LA	DI	28.1	16.7	52.6	23.9	20.6	25.7	16.5	43.8	45.9
NE	AV	294.3	167.9	75.5	120.6	195.7	233.7	254.9	-37.9	146.7
NE	SD	21.0	27.1	67.9	30.7	40.5	68.6	23.9	87.7	56.1
NE	DI	32.0	26.0	38.4	24.9	31.1	47.8	15.0	62.2	45.3
MD	AV	177.8	159.0	184.2	131.7	194.2	241.4	248.4	-47.1	135.3
MD	SD	147.8	26.5	138.4	29.0	22.7	36.1	23.9	55.7	68.1
MD	DI	*214.6*	32.0	*167.1*	28.8	12.7	28.6	27.2	36.6	*62.4*
N1	AV	274.3	175.2	92.9	134.8	198.5	248.3	240.3	-49.7	158.2
N1	SD	40.8	52.6	80.8	35.2	60.5	72.0	36.4	99.8	58.2
N1	DI	55.7	*51.3*	91.1	*35.7*	*68.4*	*77.0*	*27.8*	*124.9*	*62.8*
YO	AV	260.5	176.7	59.9	128.8	182.7	218.3	250.6	-35.5	139.4
YO	SD	87.0	32.2	54.3	28.3	57.5	76.1	19.4	88.4	46.8
YO	DI	79.7	25.6	49.9	20.6	*54.0*	*57.2*	20.9	*86.5*	32.8
CO	AV	270.8	165.5	144.5	137.1	199.8	243.8	247.0	-44.0	142.0
CO	SD	89.2	34.2	127.1	36.7	23.4	35.7	25.7	54.2	67.3
CO	DI	*87.6*	*40.3*	*168.5*	*36.2*	28.8	42.0	*28.1*	66.0	38.2

The smallest two DI values per column are double underlined, and the largest two are italicised.

nominal resolution is a somewhat subjective quantity, in view of the fact that diffraction data fade rapidly beyond ca. 3 Å, and the percent of possible intensities actually observed falls with increasing resolution. Two different investigators might decide that useful data for the same data set came to an end at different nominal resolution values. Nevertheless, the resolution is a good approximate guide to the detail that one might expect from an analysis.

If the mean difference values listed under DI arose primarily from crystal packing, then one would expect their values to be similar for most of the nine dodecamers, since all of these

crystal structures are isomorphous. Instead, the DI values of the dodecamers vary extensively, rising in general with poorer resolution (larger Res values). The meaning of larger DI values can be appreciated from FIG. 7, in which the actual twist values are plotted in a forward and a reverse direction for the KK, DR and N1 structures, chosen respectively as the best-determined structure, the standard dodecamer, and the helix with the greatest internal lack of symmetry. The DR helix places an effective limit on the role of crystal packing in the N1 helix asymmetry. Intermolecular contacts are the same for DR and N1, so any extra asymmetry of the N1 structure must come from component #3 of the list at the beginning of this paper: inadequacies in data and/or refinement, including a lack of resolution.

A careful analysis of the KK helix also tends to downplay the influence of crystal packing on helix structure[7]. The Y displacement of base pairs along their long axes is asymmetric in the KK helix, but the difference is never more than 0.5 Å. Furthermore, the observed Y displacements are not correlated at all with those that would be predicted by repulsions arising from interhelix contacts. If an external variable such as base pair Y displacement is unaffected by intermolecular contacts, then it seems even less likely that internal variables such as twist, roll or cup would be packing-dependent. Hence it is a reasonable assumption that most if not all of the DI differences in TABLES 3 and 4 are a measure of experimental errors and the absolute accuracy of the respective structure analyses.

In TABLES 3 and 4, the two smallest values of the mean symmetry difference DI in each column are marked by double underlining, and the two largest DI values are italicised and given dotted underlining. One can "score" each structure analysis in a rough fashion by assigning +1 to every double underlined DI and -1 to every italicised value, and summing in rows across the two tables. These scores in fact have been used to determine the order of presentation of helices in TABLES 3 and 4, from the most internally symmetric structures at the top (KK, DR and DW), to the least symmetric at the bottom (N1, YO and CO). As can be seen from TABLE 3, the correlation with nominal resolution is rough but far from precise.

These two tables constitute a research tool with which one can study and compare the various B-DNA helices. Mean values (AV) of twist, rise and propeller are remarkably constant among all the structures, the first two being close to their expectation values from fiber diffraction. The standard deviations (SD) indicate significant local structural variation about the mean values. End-for-end helix asymmetry, as measured by DI, is smaller than SD for the best-determined structures, but rises to match or even surpass the SD values as experimental errors come to dominate the structure analysis. When the differences between sequentially identical base pairs or steps at the two ends of the helix become as large as those between sequentially different pairs or steps along the helix, then our confidence in the reality of the sequence-dependent differences is lost.

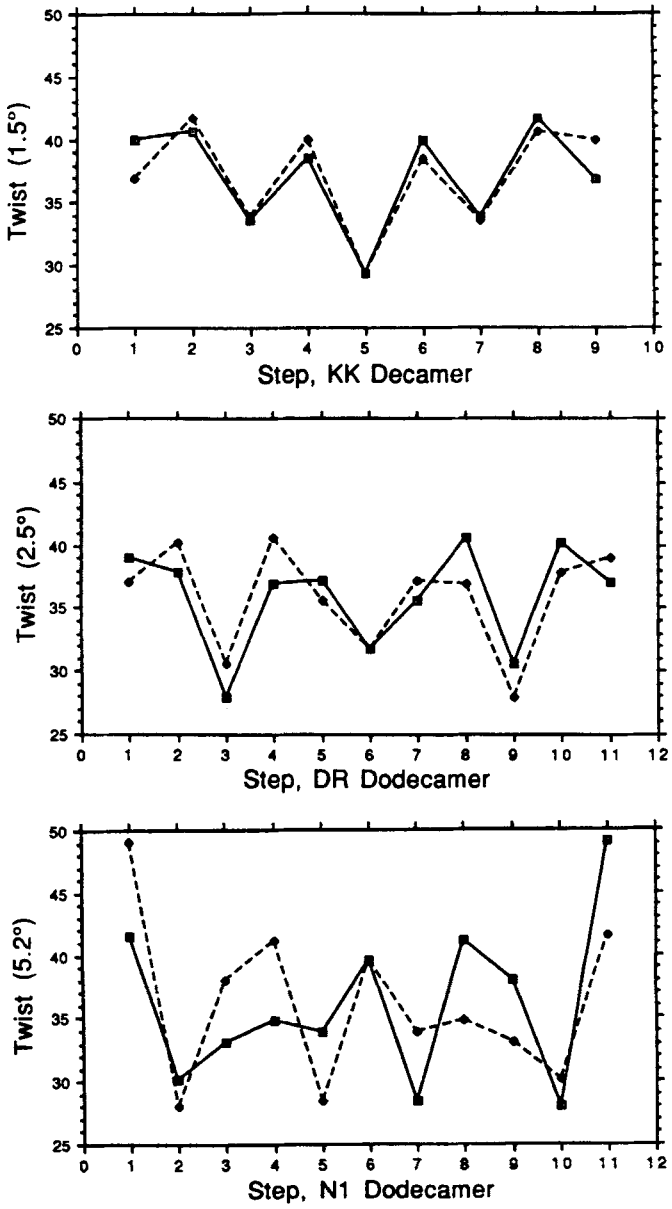

FIG. 7. Symmetry check plots of twist angle in a forward (solid line) and reverse (dashed line) direction down the helix, for the KK decamer and the DR and N1 dodecamers. The mean differences in twist angle between solid and dashed curves are given in parentheses along the left axis, and are listed under "DI" in TABLE 3.

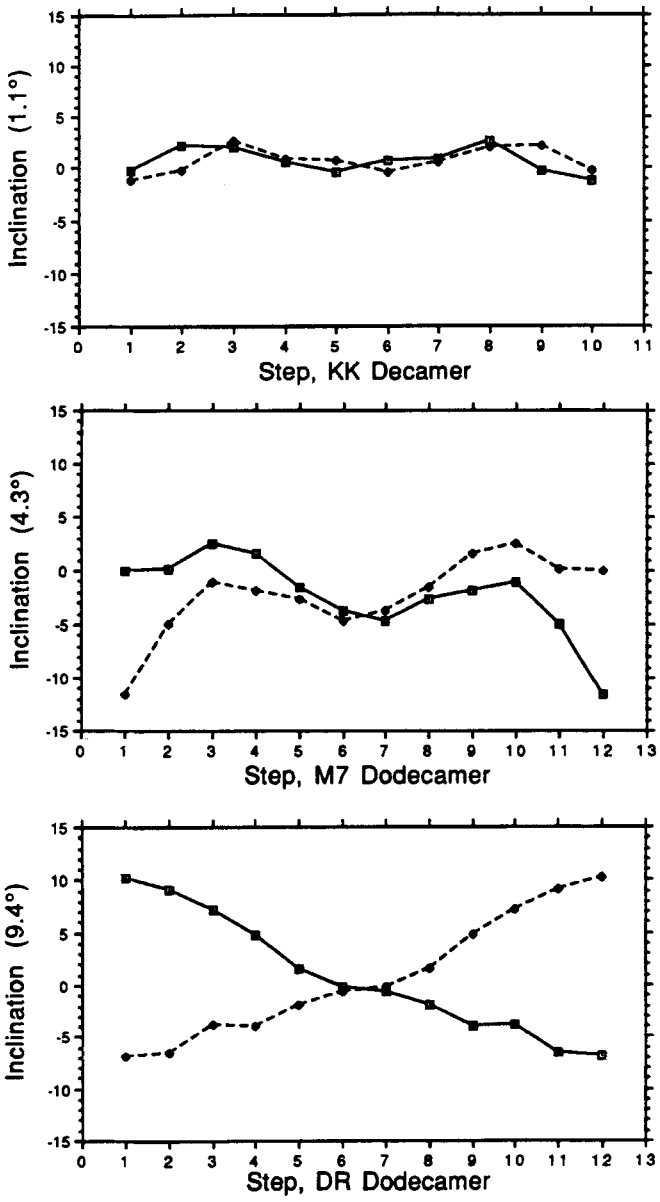

FIG. 8. Symmetry check plots of base pair inclination. Solid line = forward direction. Dashed line = reverse direction. Differences in inclination at the two ends of the DR dodecamer are real, and reflect the axial bend that is present in all of the dodecamer structures except M7, whose plot does not have so great an asymmetry. Mean DI values appear in parentheses at left.

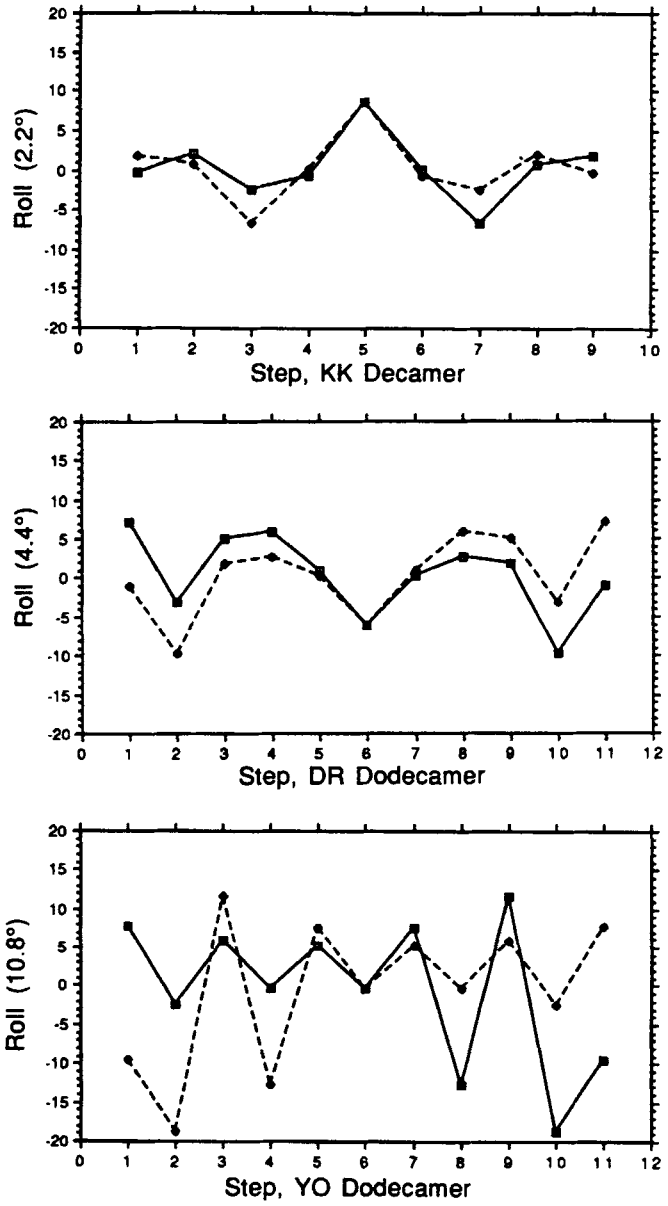

FIG. 9. Symmetry check plots of roll angles between base pairs. The KK decamer is quite symmetric end-for-end. The downward drift to the right in the DR helix again is a consequence of the axial bend. The regular oscillation of roll in the YO helix with its strict alternation of pyrimidine and purine probably is real, but the wild swings at the right end are more doubtful.

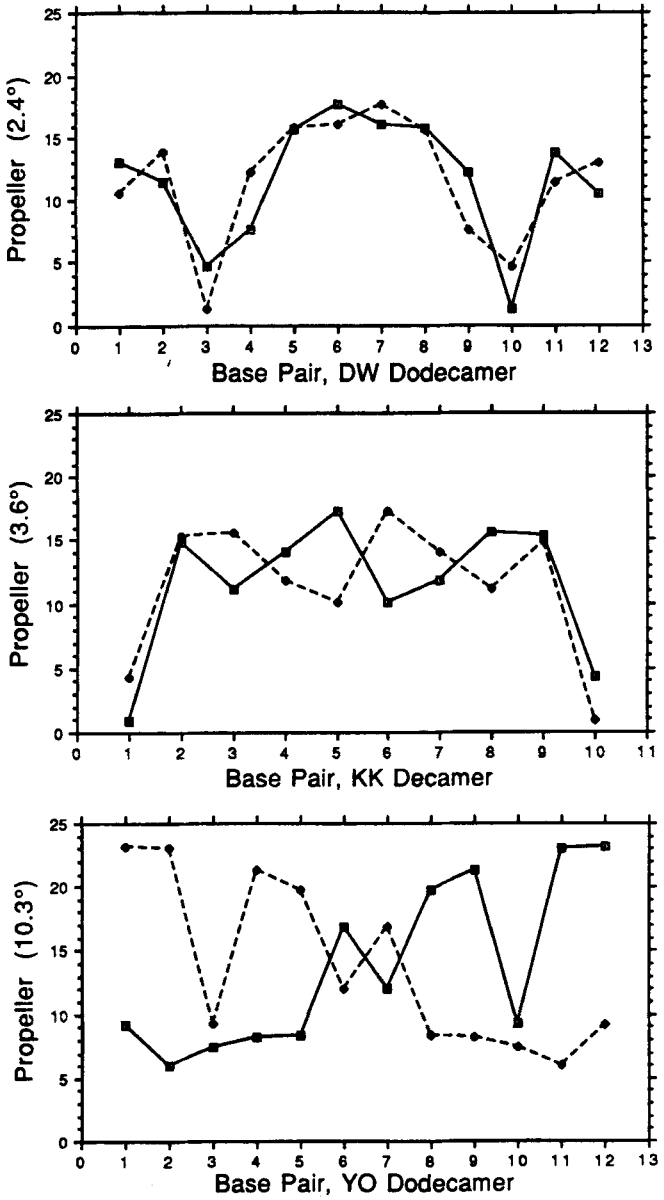

FIG. 10. Symmetry check plots of base pair propeller twist. DW and KK helices are symmetrical and well-behaved, but the wild swings in YO propeller at right may be related to the great swings in roll seen in the previous figure, as both quantities are calculated from base plane normal vector directions.

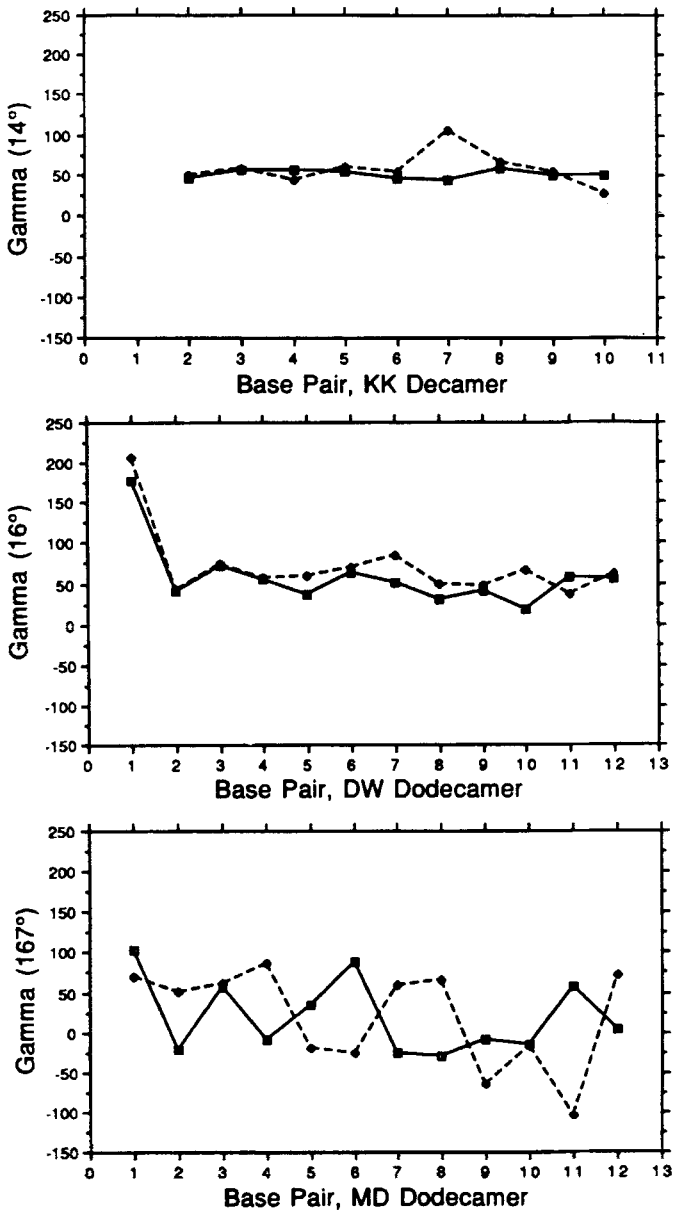

FIG. 11. Symmetry check plots of main chain torsion angle gamma (O5'-C5'-C4'-C3'). For the KK and DW helices, gamma is constant around a *gauche*$^+$ conformation except at the unattached 5' chain ends. The wild swings of gamma in the MD helix are not physically believable, and involve a crankshaft coupling with torsion angle alpha two positions earlier.

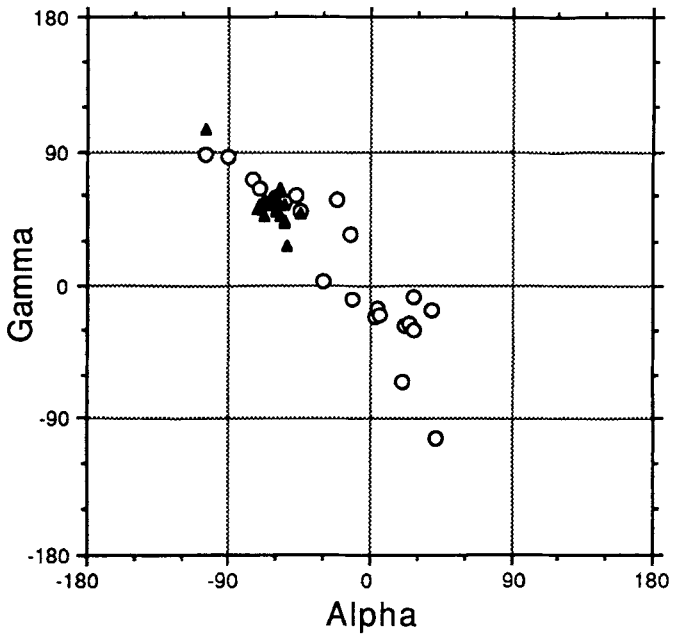

FIG. 12. Plot of main chain torsion angles α and γ for the MD helix (open circles) and the KK helix (filled triangles). The diagonal from upper left to lower right is the locus of angles that satisfy the relationship: α + γ = constant. All points associated with crankshaft rotation about β fall along this line.

Caution must be used in interpreting TABLES 3 and 4. In spite of what has just been said, a large DI does not always imply experimental error. FIG. 8 plots base pair inclination for the KK, M7 and DR helices. The monotonic decrease in inclination along the DR helix is real, and results from the bend in helix axis. Other dodecamers show this same bending behavior, and have appreciable DI values in TABLE 3. Of the dodecamers, only M7 has a straight helix axis, and it also has the smallest DI value for inclination.

The bend in dodecamer axes also shows up in the DR plot in FIG. 9, causing the gradual fall-off of the sinusoidal roll curve from left to right. Roll angles in the KK helix are quite symmetrical, but the YO plot at bottom suggests trouble. The up-down alternation of roll probably is real, resulting from alternation of purines and pyrimidines along the C-G-C-A-T-A-T-A-T-G-C-G sequence. As was first suggested for the DR helix, roll tends to be positive (opening toward minor groove) at Y-R steps and negative at R-Y steps, as observed here. But the exaggerated swings of roll at the right end in FIG. 9 may be real or may reflect only the difficulties in refinement at 2.2 Å resolution. Propeller symmetry check plots in FIG. 10 also indicate potential trouble for the YO structure analysis. The plots for DW and KK helices are

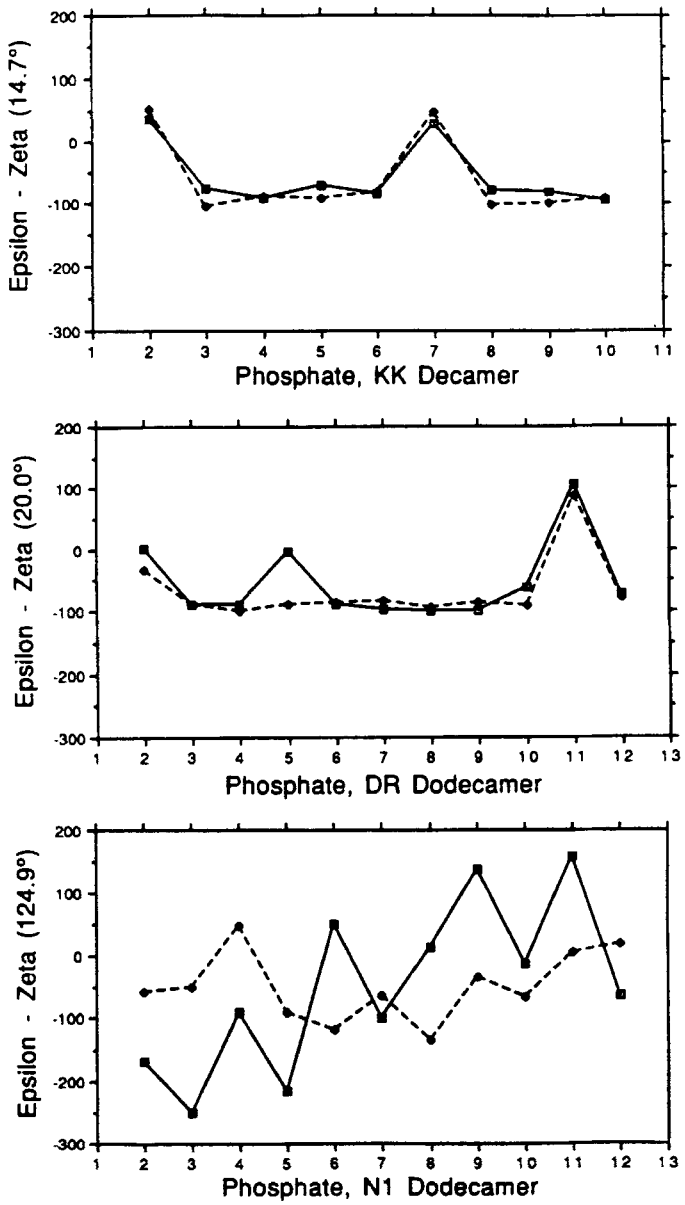

FIG. 13. Symmetry check plots of the difference between torsion angles epsilon and zeta, which is a measure of B_I vs. B_{II} phosphate conformation. (ε-ζ) should be ca. -90° for B_I and +90° for B_{II}. The less frequent B_{II} occur at P2 and P7 for the KK helix, and P11 for most of the dodecamers. The (ε-ζ) plot for the N1 helix essentially contains no physical information.

well-behaved, but the oscillation in roll at the right of the YO plot may again be a consequence of the difficulty in defining base plane orientation accurately at intermediate resolution.

These are examples of the way in which symmetry check plots can be useful tools in screening for errors. FIG. 11 illustrates this process for backbone torsion angle gamma. Gamma ideally should be *gauche*$^+$, in the neighborhood of +60°. Both the KK and DW helices fit this pattern well. But gamma in the MD dodecamer swings wildly over a 200° range, with no agreement at all between the two strands of the helix. This is because γ and α are coupled in a crankshaft motion about the intermediate *trans* β bond, in a way that allows α and γ to vary by an equal amount in opposite directions over a broad range of values without substantially changing the direction of the sugar-phosphate backbone[20]. That is, the backbone chain direction is maintained in the face of alterations in α and γ as long as $\Delta\alpha = -\Delta\gamma$ or $\alpha + \gamma =$ constant. Hence at low resolution, successful fitting of backbone chain to electron density does not necessarily suffice to define the crankshaft configuration of α and γ accurately. FIG. 12 plots the reported values of α and γ for the MD helix, with those for the KK helix added for comparison purposes. Restrained least squares refinement will confine the points to a line with -45° slope in order to preserve the connectivity of the backbone chain. But the x-ray data are not sufficiently powerful in the MD structure analysis to drive the points along this line to the correct α = -60°, γ = +60° locus. The practical resolution limit of this analysis seems to be substantially poorer than the claimed 2.0 Å nominal resolution.

One further important example is shown in FIG. 13. The difference betwen main chain torsion angles ε and ζ is a useful way of discriminating between B_I and B_{II} phosphate conformations[5]. The dodecamers generally exhibit the less common B_{II} conformation at phosphates P11 and P23, symmetrical positions on the two strands. This shows up well in the DR plot, as does the occurrence of B_{II} phosphates at P2, P7, P12 and P17 in the KK decamer. But (ε-ζ) is chaotic for the N1 dodecamer. One finds neither a confinement to ±90° values, nor even a consistency from one helix strand to the other. These torsion angles are essentially noise, and contain no structural information.

In summary, internal noncrystallographic symmetry can be a valuable tool with which to evaluate the quality of a DNA crystal structure analysis. This quality varies in a general way with the nominal resolution, but the latter quantity is so subjective that it is hardly to be trusted. In TABLES 3 and 4, the 2.3 Å M7 and 2.5 Å LA structures would appear to be substantially more reliable than are the supposedly higher resolution 2.2 Å YO and 2.0 Å MD analyses. With care, one can separate and examine all three of the components listed at the beginning of this paper: inherent structural features, crystal packing effects, and errors of analysis. X-ray crystal structure analysis, for all its faults, yields a wealth of useful information. The "streetlight syndrome" (footnote 1 of REF. 1) may be operating, but the light can be very good indeed.

Acknowledgements

This research was carried out with the support of NSF Grants DMB 85-01682 and 89-16261, and NIH Program Project Grant GM 31299.

REFERENCES

1. Dickerson, R. E. (1990). In *Structure and Methods, Vol. 3; DNA and RNA. Proceedings of the Sixth Conversation in Biomolecular Stereodynamics* (R. H. Sarma and M. H. Sarma, eds.), Adenine Press, Schenectady, NY, pp. 1-38.
2. Privé, G. G., Yanagi, K. & Dickerson, R. E. (1990). *J. Mol. Biol.* -- submitted
3. Yanagi, K., Privé, G. G. & Dickerson, R. E. (1990a). *J. Mol. Biol.* -- submitted
4. Heinemann, U. and Alings, C. (1989). *J. Mol. Biol.* **210**, 369-381.
5. Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L. & Dickerson, R. E. (1987). *Science* **238**, 498-504.
6. Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L. & Dickerson, R. E. (1988). In *Structure and Expression* (Olson, W. K., Sarma, M. H., Sarma, R. H. & Sundaralingam, M., eds), vol. 2, pp. 27-47. Adenine Press, Schenectady.
7. Grzeskowiak, K., Yanagi, K., Privé, G. G. and Dickerson, R. E. (1990)--in preparation.
8. Coll, M., Frederick, C. A., Wang, A. H.-J. & Rich, A. (1987). *Proc. Natl. Acad. Sci. USA* **84**, 8385-8389.
9. Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, I. and Dickerson, R. E. (1981). *Proc. Natl. Acad. Sci. USA* **78**, 2179-2183.
10. Westhof, E., Dumas, P. & Moras, D. (1985). *J. Mol. Biol.* **184**, 119-145.
11. Larsen, T. L. *et al.* (1990)--in preparation.
12. Narendra, N., Ginell, S. L., Jones, R. and Berman, H. M. (1990)--in preparation.
13. Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. (1982). *J. Biol. Chem.* **257**, 14686-14707.
14. Frederick , C. A., Quigley, G. J., van der Marel, G. A., van Boom, J. H., Wang, A. H.-J., and Rich, A. (1988). *J. Biol. Chem.* **263**, 17872-17879.
15. Coll, M., Aymami, J., van der Marel, G. A., van Boom, J. H., Rich, A. and Wang, A. H.-J. (1989). *Biochem.* **28**, 310-320.
16. Nelson,H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. (1987). *Nature* **330**, 221-226
17. Yoon, C., Privé, G. G., Goodsell, D. S. & Dickerson, R.E. (1988). *Proc. Natl. Acad. Sci. USA* **85**, 6332-6336.
18. Dickerson, R. E., Goodsell, D. S., Kopka, M. L. & Pjura, P. E. (1987). *J. Biomol. Str. Dynam.* **5**, 557-580.
19. Dickerson, R.E. *et al.* (1989). *EMBO J.* **8**, 1-4; *J. Biomol. Str. Dynam.* **6**, 627-634; *Nucl. Acids Res.* **17**, 1797-1803; *J. Mol. Biol.* **206**, 787-791.
20. Olson, W. K. (1982). *Nucl. Acids Res.* **10**, 777-787.