Proc. Natl. Acad. Sci. USA
Vol. 78, No. 12, pp. 7318–7322, December 1981
Biochemistry

Kinematic model for B-DNA*

(DNA sequence/DNA helix structure/torsion angles in DNA/steric constraints in DNA)

RICHARD E. DICKERSON AND HORACE R. DREW

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

Communicated by P. D. Boyer, August 10, 1981

ABSTRACT **X-ray structure analysis of B-DNA double helix with sequence C-G-C-G-A-A-T-T-C-G-C-G has revealed several sequence-dependent structural features. Four of these are shown in this paper to be related to one another by simple structural or kinematic principles: (*i*) the correlation between glycosyl torsion angle χ and main chain C4′—C3′ torsion angle δ, (*ii*) the observations that purines prefer larger χ and δ angles than do pyrimidines, (*iii*) the anticorrelation of χ or of δ angles between sugars associated with one base pair, and (*iv*) the observation that successive base planes in purine–pyrimidine steps open up the angle between them toward the major groove, whereas pyrimidine–purine steps open toward the minor groove. These features offer the beginning of an understanding of the way in which specific base sequences can perturb the structure of a B-DNA double helix so as to be "read" by intercalating drugs, repressors, and other recognition proteins.**

The structure of a double-helical dodecamer of B-DNA with the sequence C-G-C-G-A-A-T-T-C-G-C-G has recently been solved by single-crystal x-ray analysis (1–4). Examination of this molecule has revealed several structural features that appear to be related to base sequence, among them the following.

(*i*) The C1′—N glycosyl torsion angle about the bond connecting sugar and base, χ, and the main chain C5′—C4′—C3′—O3′ torsion angle, δ, are strongly correlated. As χ ranges from $-140°$ to $-90°$, δ varies between 80° and 160°.

(*ii*) Purines exhibit a systematic preference for larger χ and δ values than pyrimidines do.

(*iii*) The χ or δ values for sugars associated with paired bases are *anticorrelated*: for any one base pair they tend to occur at equal distances to either side of a common midpoint located at $\chi = -117°$ or $\delta = 123°$.

(*iv*) Two successive base pairs with the sequence 5′ (pyrimidine)-*P*-(purine) 3′ tend to roll about their long axes in such a manner as to open the angle between them toward the minor groove; base pairs of sequence 5′ (purine)-*P*-(pyrimidine) 3′ tend to open toward the major groove.

The purpose of this communication is to demonstrate that all four of these observations are interrelated and have a simple structural explanation.

Correlation of torsion angles χ and δ

The seven torsion angles required to define a polynucleotide structure are shown in Fig. 1, in the conformation normally encountered in a B helix. The χ and δ angles for C-G-C-G-A-A-T-T-C-G-C-G are plotted against one another in Fig. 2, and a relationship between them is immediately apparent. This correlation has already been remarked upon (2), without structural explanation.

The reason for the correlation can be understood from Fig. 3, which depicts two different sugar ring orientations as viewed in a direction along the helix axis. The phosphorus atom separations along a given helix strand are quite uniform in C-G-C-G-A-A-T-T-C-G-C-G, with a mean (±SD) of 6.68 ± 0.23 Å. Furthermore, linear regression analysis shows that this phosphate separation is not significantly correlated with any of the seven torsion angles, the highest value of any regression coefficient being only 0.53 with angle γ. The phosphate backbone in this B helix is quite insensitive to local perturbations of base and sugar geometry.

In such a regularly spaced helix backbone, the relationship between χ and δ arises because of the necessity of spanning a

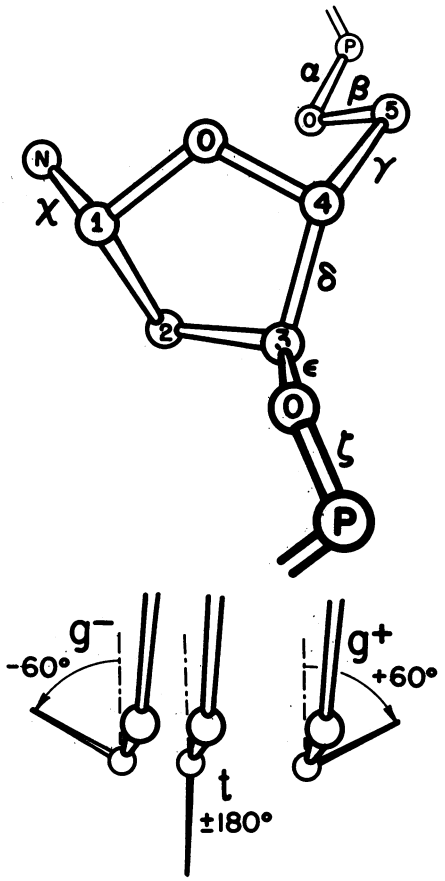

FIG. 1. (*Upper*) Definitions of main chain (α-ζ) and glycosyl (χ) torsion angles. Deoxyribose atoms C1′ through C5′ are numbered 1–5, and atoms O1′, O3′, and O5′ are attached to carbon atoms of the same number. N is the attached atom of a purine or pyrimidine base. χ is defined by O1′—C1′—N1—C2 (pyrimidines) or O1′—C1′—N9—C4 (purines). (*Lower*) Torsion angle sign conventions and definition of the three staggered bond configurations: *gauche*$^-$ (g^-), *trans* (t), and *gauche*$^+$ (g^+). The main chain configuration in *Upper* is that of the B-DNA helix g^-, t, g^+, t, t, g^-, and the sugar conformation is C2′-*endo*.

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

* This is paper no. 4 of a series; papers nos. 1–3 are refs. 2–4.

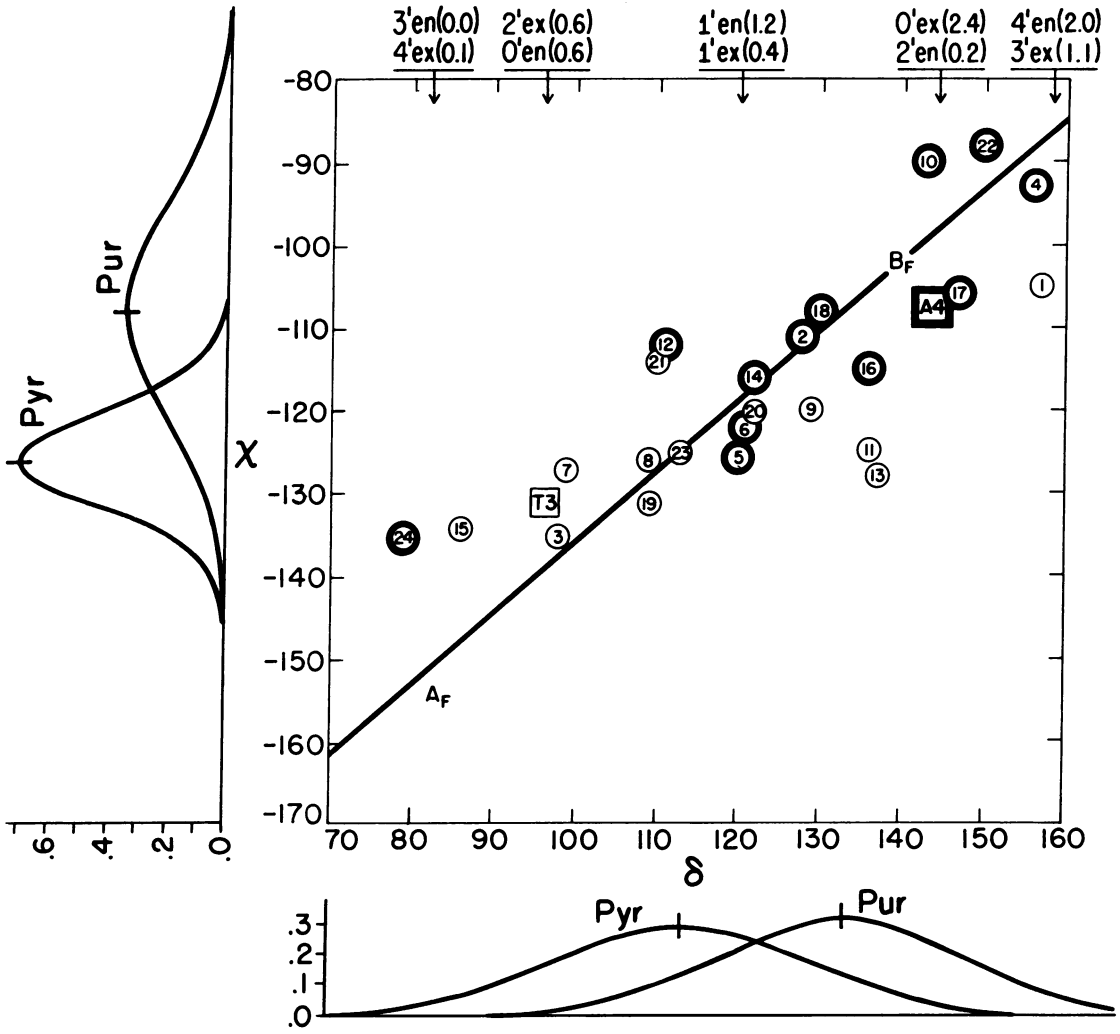

FIG. 2. Correlation diagram of glycosyl torsion angle χ and main chain torsion angle δ for the 24 independent bases of C-G-C-G-A-A-T-T-C-G-C-G (circles) and the two independent and nonintercalative bases of the daunomycin complex of C-G-T-A-C-G (squares). Purines are given heavy black borders to distinguish them from pyrimidines. A_F and B_F are the classical A- and B-DNA conformations as deduced from fiber diffraction data. The diagonal line is the best linear regression fit to the C-G-C-G-A-A-T-T-C-G-C-G data, with a regression coefficient of 0.78. Normal distribution curves on the same relative scale (i.e., with the same area under the curve) have been fitted to the χ and δ data for purines and pyrimidines separately at left and bottom. Means ± SD are given in the text. (Base plane 1/24 has been omitted from averaging, as representing an end effect from crystal packing.) C-G-C-G-A-A-T-T-C-G-C-G data are from refs. 2 and 3; daunomycin complex data are from ref. 5.

roughly constant distance between atoms O3′ and C5′ as shown in Fig. 3. If the sugar ring is oriented perpendicular to the plane of bases (Fig. 3*a*), then the vector C3′—C4′ is perpendicular also and makes no contribution to bridging the gap. Hence, torsion angle δ must be large. In contrast, if the sugar ring and its C3′—C4′ vector are tilted as in Fig. 3*b*, then torsion angle δ need not open as far in order to reach from O3′ to C5′, and the observed correlation between glycosyl angle χ and main chain angle δ is accounted for.

Preference of purines for larger χ and δ angles than pyrimidines

A prominent feature of Fig. 2 is the systematically higher (less negative) values of χ and δ adopted by purines in comparison with pyrimidines. Purines cluster around δ values close to the value associated with the C2′-*endo* sugar conformation, whereas the pyrimidine distribution is centered between C1′-*exo* and O1′-*endo*. This too can be given a structural explanation, in terms of steric clashes between atoms in the neighborhood of the pyrimidine glycosyl bond (Fig. 4). At a χ value of $-120°$ as shown, the C1′ hydrogen atom on the sugar is in an eclipsing position relative to the pyrimidine O2 and is only 2.19 Å from it. The generally accepted minimal contact distance between nonbonded H and O is 2.4 Å, and an extreme lower limit is 2.2 Å (6, 7). This tight contact cannot be eliminated by rotating χ to more positive values because this brings a hydrogen atom on C2′ into even closer contact with the hydrogen atom of C6 on the pyrimidine ring. The overlap can be relieved, however, by rotating χ to a slightly more negative value. In contrast, the larger C1′—N—C bond angle associated with the five-membered ring of a purine (dotted outline in Fig. 4) means that neither the purine N3 nor its C8 hydrogen is in any danger of close contact with the sugar, so χ is free to adopt the value favored by a classical C2′-*endo* B-DNA helix. Hence, in Fig. 2, the purine conformations cluster about $\chi = -108° \pm 13°$ and $\delta = 133° \pm 14°$, whereas pyrimidines are distributed around $\chi = -126° \pm 6°$ and $\delta = 113° \pm 16°$ (mean ± SD omitting base pair 1/24). The narrowness of the χ distribution for pyrimidines may simply reflect the fact that χ is trapped between the steric constraints just mentioned and the extreme lower limit compatible with a B-DNA backbone.

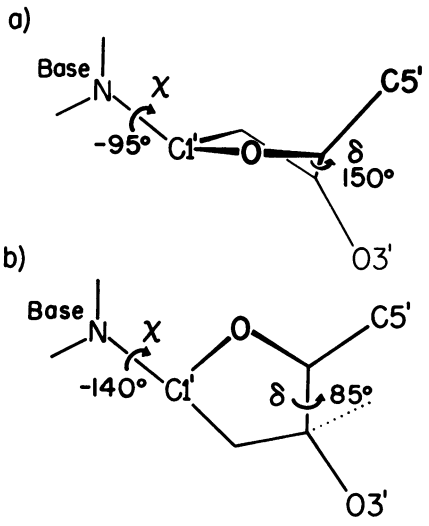

FIG. 3. Structural basis for the correlation of torsion angles χ and δ. If the spacing between phosphates along one helix strand (or the O3' . . .C5' separation as drawn here) is roughly constant, then tilting the sugar ring in such a manner as to decrease χ from $-95°$ to $-140°$ means that δ need not open as wide to span the gap between O3' and C5' and hence can close down from 150° to 85°. The view in these drawings is directly down the helix axis, perpendicular to the plane of the bases.

Anticorrelation of torsion angles in paired bases

Not only do purines and pyrimidines occur on opposite sides of the center of Fig. 2 but also the points for paired bases are found at roughly the same distance to either side of the midpoint, either in χ or δ. This has been termed the principle of anticorrelation (2). As examples, paired bases 15 and 10 have conformations that bracket the midpoint far to either side, as do bases 3 and 22. Bases 24 and 1 also display a large degree of anticomplementarity even though their purine–pyrimidine conformational behavior is reversed. (This is an end effect attributable to crystal packing.) In contrast, the angular spread for base pairs 2/23 and 5/20 is less but the points still cluster symmetrically about the center of the diagram. When individual midpoints are calculated for all 12 base pairs, they have mean values of: $\chi_{mid} = -117.5° \pm 5.8°$ and $\delta_{mid} = 122.8° \pm 6.7°$. For comparison with these standard deviations, the mean of the total spread, or difference in torsion angles between paired bases, is 19.2° in χ and 31.5° in δ.

Both the correlation of χ and δ and the anticorrelation across paired bases are encountered in the single unperturbed base step of the other currently available B helix structure, the complex of daunomycin with C-G-T-A-C-G (5). Both of its C-G ends are perturbed by intercalation of daunomycin between C and G. But the central T-A step is shielded from intercalators above and below and is a normal B-DNA step. The χ and δ values for bases T3 and A4 have been added to Fig. 2 as squares. (Only one point exists for each because the helix has a crystallographic 2-fold axis, making both ends identical.) From these two examples, C-G-C-G-A-A-T-T-C-G-C-G and C-G-T-A-C-G, one might expect the principle of anticorrelation to be a general feature of base pairing in B-DNA.

The stereo drawings of two base pair steps from C-G-C-G-A-A-T-T-C-G-C-G in Fig. 5 and the schematic in Fig. 6 illustrate the explanation of anticorrelation: anticorrelation between glycosyl torsion angles of paired bases simply means that, when the planes of the sugar rings tilt away from perpendicularity to the plane normal to the helix axis (the plane of the drawing in Fig. 6), they do so in tandem and to the same extent as viewed from outside the helix. For example, sugar rings 2 and 23 in Fig. 5 are perpendicular to the plane of the page as drawn in Fig. 6*a*, whereas both rings 3 and 22 are tilted so their nearer edge moves toward the top of the drawing, as in Fig. 6*b*. Rings 4 and 21 once again are perpendicular to the page. Further examples can be found in figure 7 of ref. 3. Among these, sugar rings 9 and 16 are normal to the page, the rings of 10 and 15 are tilted so their nearer edge moves toward the bottom of the drawing (the reverse of the motions in Fig. 6*b*), and 11 and 14 are nearly perpendicular again. Because the two backbone chains of the helix run in opposite directions, tilting the sugar rings in parallel means giving them glycosyl rotations that are opposite in sign. Furthermore, tilting them by equal amounts means that the two glycosyl rotations are of the same magnitude, although of opposite sign. This is just the principle of anticomplementarity as observed in C-G-C-G-A-A-T-T-C-G-C-G and C-G-T-A-C-G (Fig. 2).

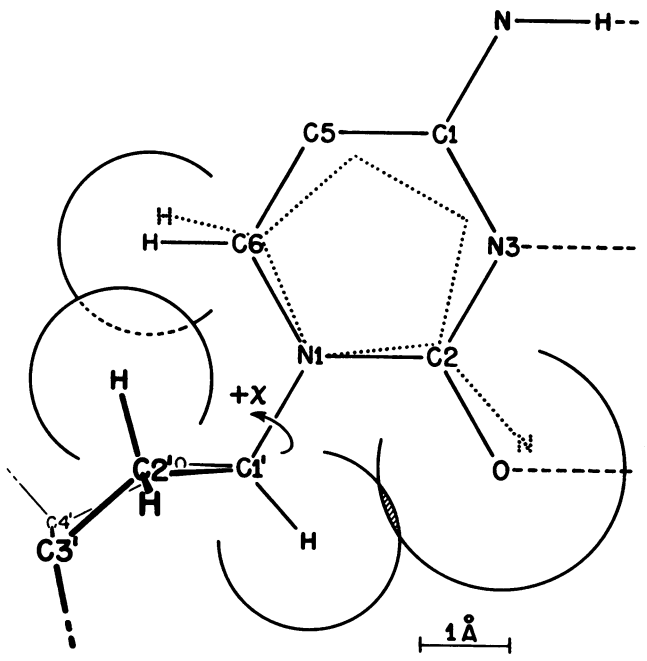

FIG. 4. Steric interference leading to low χ values for pyrimidines. At $\chi = -120°$, as drawn here, a small overlap occurs between H and O, shown by the hatched region. Rotation of χ to a less negative value to relieve this overlap brings the H of C2' down into contact with the H of C6, seen below it on this $\chi = 120°$ drawing. But rotation of χ to a more negative value relieves all of the tight contacts. None of these steric clashes is a problem with purines (dotted five-membered ring). Hence purines are free to adopt values broadly distributed about $\chi = -108°$, whereas the pyrimidine distribution is more tightly clustered around $\chi = -126°$.

Relationship between glycosyl torsion angles and base plane orientation

The rotation of glycosyl angles depicted in Fig. 6 has consequences beyond simple anticorrelation. The C1'—N glycosyl bond does not lie in the approximate plane of the sugar ring; it is bent sharply away from it by the tetrahedral bond geometry at C1'. Hence, when the sugar rings are rocked as observed at base pair C3/G22 in Fig. 5 (and sketched in Fig. 6*b*), the N ends of both C1'—N glycosyl bonds are pushed down into the page. This means that the entire base pair dips down at its top edge (the major groove). This edge moves farther away from base pair G2/C23 (which would lie between the viewer and the base pair depicted in Fig. 6*b*), hence opening the G2–C3 step toward the major groove. At the same time, the top edge of C3/G22 moves closer to base pair G4/C21 (which would be below the page in Fig. 6*b*), opening the C3–G4 step toward the minor groove.

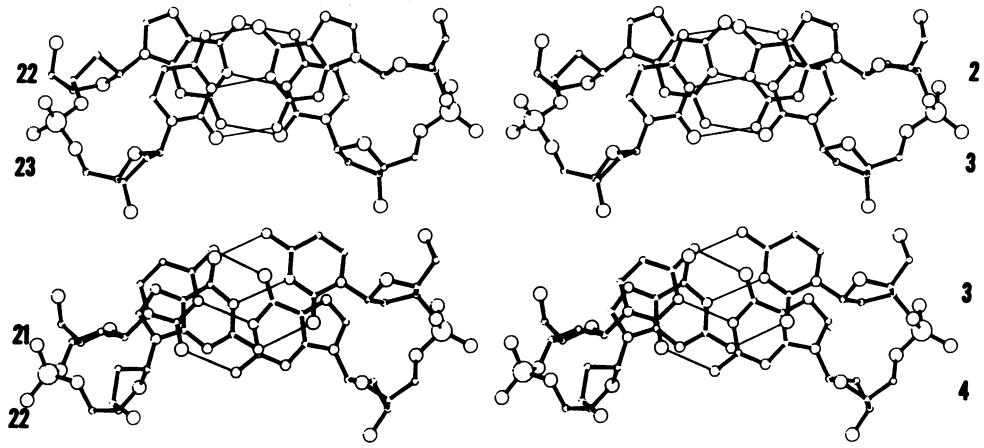

FIG. 5. Stereo drawings of 2 of the 11 base pair steps in C-G-C-G-A-A-T-T-C-G-C-G, viewed down the helix axis as in Fig. 3. The major groove of the helix is at the upper edge of each stereo pair, and the minor groove is at the lower edge. Base numbering is at left and right. The complete set of stereos of base pair steps appears in ref. 3.

These motions are represented schematically in the diagram of the unwound helix (Fig. 7*a*). A similar situation exists in figure 7 *i* and *j* of ref. 3: rocking the sugar rings of base pair G10/C15 in the manner shown (the opposite of Fig. 6*b*) tilts the N ends of the C1'—N glycosyl bonds up and out of the page, bringing the upper edge of the base pair toward the viewer. This opens step C9–G10 toward the minor groove and G10–C11 toward the major groove.

In both of these examples, the consequence of base plane rolling is an opening of CpG steps toward the minor groove and of GpC steps in the opposite direction. This was found to be a general principle in C-G-C-G-A-A-T-T-C-G-C-G (3): pyrimidine–purine steps open by a few degrees toward the minor groove, purine–pyrimidine steps open toward the major groove, and homopolymer steps (Pur–Pur) and (Pyr–Pyr) tend not to roll in either direction (Fig. 7*d*). Now we see that this is a natural consequence of the fact that purines prefer higher (less negative) glycosyl χ angles than do pyrimidines.

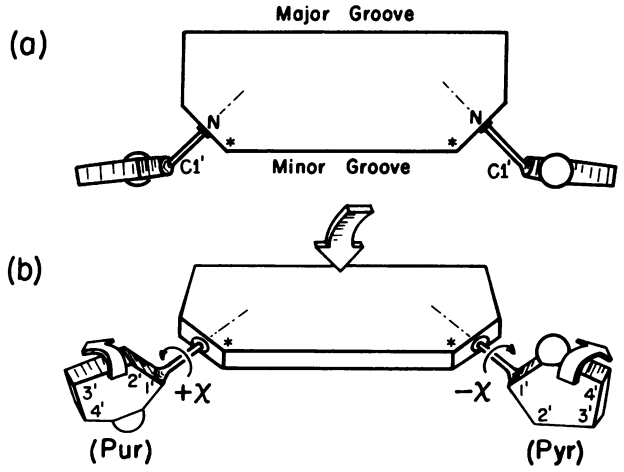

FIG. 6. Illustration of the origin of anticomplementarity and of the connection between glycosyl angle χ and the rolling of base pair planes about their long axis. If the two sugar rings of a base pair are tipped by the same amount as viewed from outside the helix, then their glycosyl angles change by equal amounts in opposite directions, which is just the principle of anticomplementarity (*a*). Rotation of the sugar rings as shown in (*b*) also tips the upper or major groove edge of the base pair into the plane of the page. Such a rotation is expected in general if the base at the left is a purine (increased χ) and that at the right is a pyrimidine (decreased χ). The opposite rotation is expected with pyrimidine at the left and purine at the right, lifting the major groove edge of the base pair out of the plane of the diagram. χ = 0° when the sugar oxygen (round ball) is aligned with the asterisk on the base plane.

The expected base plane roll for several double helical examples is shown in Fig. 7: *a* and *d* are as observed in C-G-C-G-A-A-T-T-C-G-C-G, and *b* and *c* illustrate alternating roll in poly(T-A) and the parallel stacking of homopolymer sequences. The tendency of a pyrimidine–purine step to open toward the minor groove may assist the entry of intercalators such as ethidium from that side (Fig. 7*e*). Sobell and coworkers reported (8, 9) that, in ethidium complexes with r(^{I}CpG) and r(^{I}UpA), the base planes are opened by 8° toward the minor groove and the phenyl side chain of the ethidium ring makes it virtually certain that this is the direction of entry of the intercalator. By our analysis, the C3'-*endo*(3'-5')C2'-*endo* mixed sugar pucker that is encountered in many dinucleotide intercalation complexes would lead to just such an opening of the minor groove, and this is visible in the stereo drawings of complexes of r(^{I}CpG) with 9-aminoacridine (10), acridine orange (11), ellipticine (12), and 3,5,6,8-tetramethyl-*N*-methylphenanthrolinium (12).

The central TpA step of the C-G-T-A-C-G·daunomycin complex clearly opens toward the minor groove as would be expected from the positions of T3 and A4 on the χ/δ plot in Fig. 2 (see figure 3 of ref. 5). Even in the regions of C-G-T-A-C-G where base stacking has been perturbed by the intercalator, the connection between tilt of individual bases (not base pairs) and glycosyl angle χ diagrammed in Fig. 6 can be discerned. But daunomycin is an exception to the previously mentioned intercalators in that a long and narrow planar ring system intercalates its long axis through the helix from the minor groove and actually extends out on the other side into the major groove. This may explain why the base pairs stacked against it to either side appear to be nearly parallel rather than opened toward the minor groove and why the intercalation site consequently need not exhibit mixed sugar pucker. A diagram of base geometry in this complex, as judged from figure 3 of ref. 5, is shown in Fig. 7*f*. Both the roll and anticomplementarity of the central TpA step, and the lack of either base roll or mixed sugar pucker at the intercalation sites, are accounted for by our kinematic model.

The nucleoside pair as a semirigid unit

The relative orientation of adjacent base pairs in a B-DNA helix is seen to depend on the tilting of glycosyl bond vectors, which

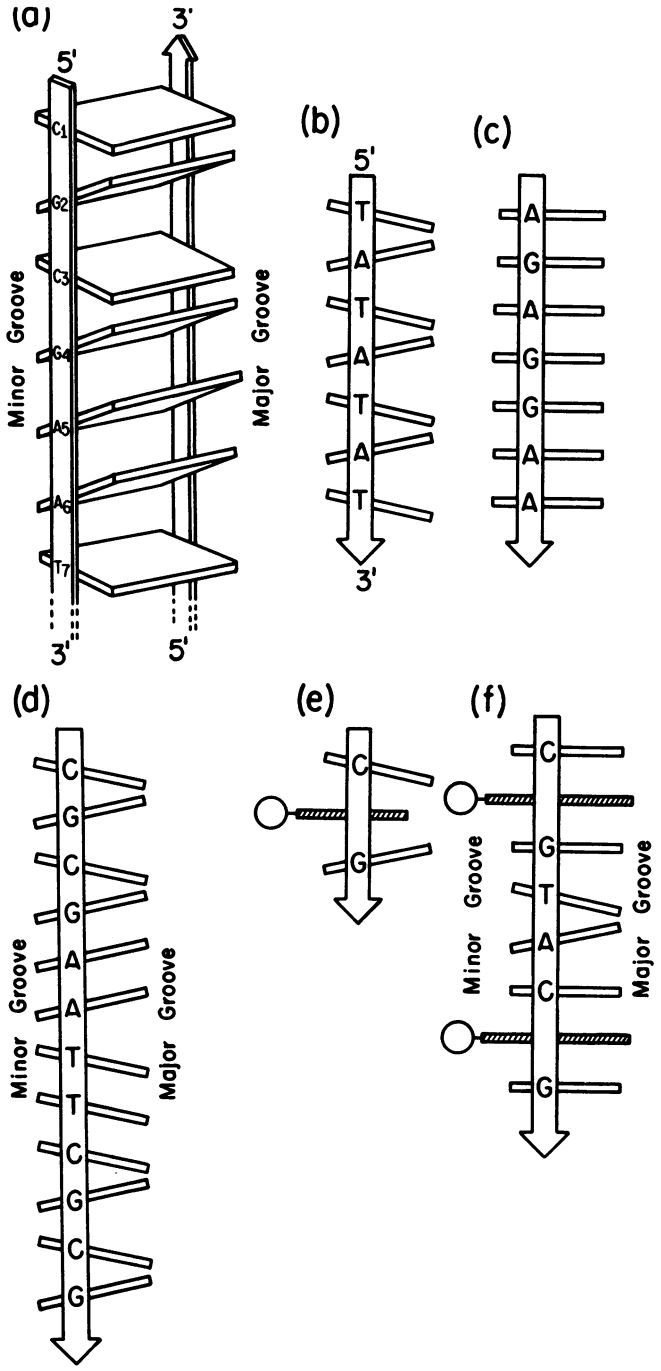

FIG. 7. Unwound-helix or Venetian blind diagrams of DNA, showing the observed or expected base plane roll (greatly exaggerated for visibility) for several double-helix sequences. (*a*) The first seven base pairs, C-G-C-G-A-A-T. . ., of the dodecamer. (*b*) Alternating poly(T-A) sequence, with alternating base plane roll angles. (*c*) Parallel stacking of homopolymers. (*d*) The full C-G-C-G-A-A-T-T-C-G-C-G dodecamer. Actual roll angles are in ref. 3. (*e*) The manner in which the observed 8° opening of r(CpG) toward the minor groove (8, 9) facilitates entry of an intercalator such as ethidium (hatched). Circle is an attached nonplanar group. (*f*) Observed base roll in the complex of C-G-T-A-C-G with two daunomycin intercalators (hatched).

in turn is a function of the value of the glycosyl torsion angle χ. Base tilt is observed to show a sequence dependence because purines and pyrimidines themselves have different preferences for χ values, and this ultimately arises from differences in steric constraints about the bonds to five-membered or six-membered rings. Such an explanation considers the sugar-base-base-sugar assembly as a functional unit, with swivels at the glycosyl bonds but with little other internal freedom. This simple model of course will be perturbed by other degrees of freedom such as propellor twist, and a careful examination of stereo diagrams such as figure 7 of ref. 3 reveals a few examples in C-G-C-G-A-A-T-T-C-G-C-G of what could be termed second-order perturbations in structure. But the simple model does a surprisingly good job of accounting for observations. The correlation between torsion angles χ and δ, of course, is a consequence not of the semirigid sugar-base-base-sugar unit but rather of the requirement that this unit be fitted into a regular helix backbone with approximately constant phosphorus–phosphorus spacing.

This analysis indicates one way in which base sequence information, expressed as differences in stereochemical behavior of five- and six-membered rings in purines and pyrimidines, can be "amplified" and transmitted to the framework of the helix in a manner that potentially can be recognized by intercalators and by base-specific recognition proteins such as repressors. In the past it has been generally assumed that most of the recognition of DNA base sequence occurs by means of hydrogen bonds to polar N and O groups in the major and minor grooves rather than via helix deformations, perhaps chiefly because information about local helix deformation was not obtainable from fiber diffraction data. The structure analysis of the dodecamer C-G-C-G-A-A-T-T-C-G-C-G reveals that particular base sequences can influence the helix structure in subtle ways. It will be of interest to see whether these influences also play a role in sequence recognition.

We thank Mary Kopka and A. V. Fratini for discussions and criticisms of the manuscript and Lillian Casler for preparation of all of the line drawings. This work was carried out with the aid of National Institutes of Health Grant GM-12121, National Science Foundation Grant PCM79-13959, and a grant from The Upjohn Company. H.R.D. was supported by Fellowship DRG-507 of the Damon Runyon–Walter Winchell Cancer Fund.

1. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1980) *Nature (London)* **287**, 755–758.
2. Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R. E. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 2179–2183.
3. Dickerson, R. E. & Drew, H. R. (1981) *J. Mol. Biol.* **149**, 751–776.
4. Drew, H. R. & Dickerson, R. E. (1981) *J. Mol. Biol.* **151**, 535–556.
5. Quigley, G. J., Wang, A.H.-J., Ughetto, G., Van der Marel, G., Van Boom, J. H. & Rich, A. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7204–7208.
6. Ramachandran, G. N. & Sasisekharan, V. (1968) *Adv. Protein Chem.* **23**, 283–437.
7. Schulz, G. E. & Schirmer, R. H. (1979) *Principles of Protein Structure* (Springer, New York), p. 20.
8. Tsai, C.-C., Jain, S. C. & Sobell, H. M. (1977) *J. Mol. Biol.* **114**, 301–315.
9. Jain, S. C., Tsai, C.-C. & Sobell, H. M. (1977) *J. Mol. Biol.* **114**, 317–331.
10. Sakore, T. D., Reddy, B. S. & Sobell, H. M. (1979) *J. Mol. Biol.* **135**, 763–785.
11. Reddy, B. S., Seshadri, T. P., Sakore, T. D. & Sobell, H. M. (1979) *J. Mol. Biol.* **135**, 787–812.
12. Jain, S. C., Bhandary, K. K. & Sobell, H. M. (1979) *J. Mol. Biol.* **135**, 813–840.