Structure Factor Calculations of Various DNA Duplexes

MANJU BANSAL AND GOUTAM GUPTA
Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India

Abstract

Based upon a stereochemical guideline, both right- and left-handed duplexes were generated for *A*, *B*, and *D* forms of DNA using a mononucleotide as the repeat. Structure factor amplitudes were computed following two methods: (i) one employed an infinite helix as the motif and (ii) the other had an exact crystallographic repeat (e.g., 10 b.p. for B-DNA) as the motif. Both the procedures showed that DNA in either handedness for *A*, *B*, and *D* forms of DNA are consistent with the observed fiber data. This observation is attributed to the fact that fibre pattern (resolved only upto 3 Å) only gives the image of the gross structure of the molecule. Thus, when the gross structure is suitably fitted to match the observed pattern, it is immaterial as to what the precise stereochemistry of the repeating unit (fine structure) and handedness of a model are. Sequence specific helical conformations were obtained using dinucleotide as the repeating unit. Helices fell into two classes: (i) right-handed uniform (RU) and (ii) left-handed zig-zag (LZ) structures. A few aspects concerning the scattering profile of these helices are discussed.

Introduction

Structural features of helical DNA duplexes can be classified into two parts: (i) gross and (ii) fine structure. Gross structure involves the overall dimension of the phosphate cylinder, separation between the two phosphate chains across minor/major groove and the position and the orientation of the base plane with reference to the helix axis. Fine structure relates to the precise stereochemistry of the repeating unit. It turns out that the fiber-diffraction data (resolved only upto 3 Å) only give the image of the gross structure and defy any direct information about the fine structure. However, single crystal structure of nucleic acid components provide necessary details about the precise stereochemistry of the repeating unit. In fact, these data when aided by conformational analyses, enabled us to formulate a stereochemical guideline for molecular model building which led to both right- and left-handed duplexes for various polymorphic DNA's [1–3]. It was interesting to note that both right- and left-handed duplexes for a given form of DNA retained the same gross structure of the molecule. Hence, as expected, both right- and left-handed duplexes of a given polymorphic DNA gave equally good agreement with the observed fiber data [2,3]. It was also observed that more than one conformationally distinct repeating units could lead to nearly the same gross structure of a polymorphic DNA with a given handedness.

International Journal of Quantum Chemistry, Vol. XX, 407–417 (1981)

CCC 0161-3642/81/080407-11$01.10

Method of Calculation

Stereochemical possibility of both right- and left-handed DNA duplexes was shown for *A*, *B*, and *D* forms of DNA by molecular model building using a stereochemical guideline [1–3]. This was followed by calculation of structure factor amplitudes of the models to test the agreement with the observed data. The molecular model building was done in two steps: (i) First, a mononucleotide was chosen as the repeating unit such that all the nucleotides had the same conformation irrespective of the bases attached to them. (ii) Next, a dinucleotide was chosen as the repeating unit in order to invoke sequence specific molecular conformation wherein purine nucleotides had a conformation different from that of the pyrimidines. Calculation of the structure factor amplitudes was done in two ways: (i) one in which an infinite helix was chosen as the motif [4] and (ii) in another the exact crystallographic repeat (e.g., 10 b.p. for *B*-DNA) was regarded as the motif.

Results

A. Gross Structure and Its Relation to the Fiber Pattern

As mentioned in the previous section, fiber pattern of a DNA duplex is chiefly determined by the gross structure of the molecule. This can be readily seen from the simple arguments given below.

The scattering amplitude of a single helical molecule is given as [4]

$$A(R,l) = \sum_{j,n} f_j J_n (2\pi Rr_j) \exp i \left(\frac{2\pi l z_j}{C} - n\phi_j\right), \tag{1}$$

where (r_j, ϕ_j, z_j) are the cylindrical polar coordinates of the jth atom of the repeating unit (e.g., a mononucleotide); f_j is the corresponding atomic scattering factor corrected for the presence of water as suggested by Arnott and Hukins [5]; R is the radial coordinate in reciprocal space; C is the fiber repeat; l is the layer line; n = Bessel order for a given l, where

$$n = l - Nm \ (m = 0, \pm 1, \pm 2, \ldots) \quad \text{for an } N\text{-fold right-handed helix;}$$
$$n = -l + Nm \quad \text{for an } N\text{-fold left-handed helix.}$$

It turns out that for *A* and *B* forms of DNA (when $l \leqslant N$), the phosphate groups give major contribution to the scattering profile of lower layer lines ($0 \leqslant l \leqslant 5$) [4]. In such a case, only $n = l$ gives significant contribution for a right-handed helix while $n = -l$ contributes for a left-handed helix. Therefore, in one case (right-handed helix) the exponent in expression (1) becomes multiple of $i\,(2\pi z_j/C - \phi_j)$, while in the other it becomes a multiple of $i\,(2\pi z_j/C + \phi_j)$. However, in either case, the exponent (without i) defines half of the chain separation across the minor groove. Thus, for an N-fold helix, the scattering amplitude $A(R,l)$ on lower layer lines ($0 \leqslant l \leqslant 5$) due to right- and left-handed helix would be similar as long as r_j (phosphate radius) and exponent in expression (1) (i.e., the phosphate chain separation across the minor groove) remain nearly

the same. For higher layer lines ($l \geq 6$), bases as well play an important role. The scattering profile due to bases, is dependent upon the base parameters [5], viz., displacement D, tilt θ_x, and twist θ_y. Inasmuch as these parameters are linked to phosphate radius and chain separation, it is possible to simultaneously achieve good agreement of a model on lower and higher layer lines by judiciously adjusting the parameters of the gross structure. Therefore, when the gross structure is suitably fitted to match the observed pattern it hardly matters what the fine structure and handedness of a model are. Calculations of Fourier transform and structure factor amplitude of various models of a polymorphic DNA suggested that agreement with the observed fibre data, in each case, was possible for a limited range of gross structure. For example, agreement with the fiber data of *B*-DNA required that phosphate radius should be within 8.8 to 9.3 Å, chain separation across the minor groove should be 11 to 15 Å and displacement D within -1 to 0.5 Å and tilt θ_x within $-5°$ to $10°$. Molecular model building of different polymorphic DNAs was constrained to yield gross structures which would fall within the range required to match the observed pattern.

B. Molecular Model Building Using Mononucleotide as the Repeating Unit: RU and LU helices

In a mononucleotide, phosphodiester (P—O) links are absent. Thus, one cannot study the effect of two major degrees of freedom, i.e., (β,γ) torsions, in relation to molecular model building. Also, stacking interactions, a vital stabilizing factor in a polynucleotide, are absent in a mononucleotide. For these reasons, a base-paired dinucleoside monophosphate was chosen as the basic unit which embodies all the essential attributes of a polymeric DNA. In such a unit,

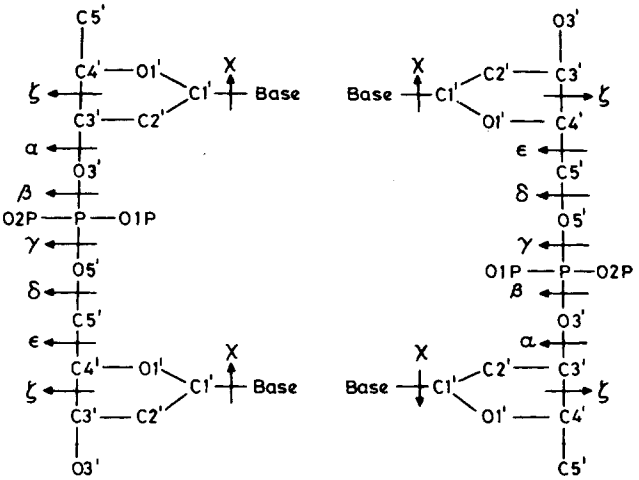

Figure 1. Base-paired dinucleoside monophosphate. The alphabetical nomenclature of the torsion angles was taken from Ref. 6. Note that the exact repeating unit is a mononucleotide.

TABLE I. Conformational parameters and related agreement indices of *B*-DNA models using mononucleotide as the repeat.

	Right-handed models		Left-handed model
	(C3′-*endo*, g^-g^-) conformation	(C2′-*endo*, tg^-) conformation	(C2′-*endo*, tg^-) conformation
α	184	228	241
β	292	202	204
γ	269	302	270
δ	179	144	135
ϵ	75	41	36
ζ	97	149	137
χ	34	74	−3
(R, R'')[a]	I (0.36, 0.39)	(0.35, 0.38)	(0.37, 0.40)
	II (0.35, 0.39)	(0.36, 0.39)	(0.38, 0.41)

[a] I refers to agreement indices obtained employing an infinite helix as the motif, while II refers to the values obtained with an exact crystallographic repeat as the motif.

sugars at 5′ and 3′ ends had the same puckering and bases attached to them had the same glycosyl torsion. Thus, the true repeat was a mononucleotide.

A base-paired dinucleoside monophosphate is shown in Figure 1. Different torsion angles are also indicated following Seeman et al. [6]. Based upon a stereochemical guideline, a set of gross structures were obtained for *A*, *B*, and *D* forms of DNA which fell within the range stipulated for matching the observed fiber data [5,7]. For a particular polymorphic DNA, the models which met the following four criteria were described as probable structures: (i) Watson–Crick base-pairing scheme and energetically favorable stacking overlap; (ii) allowed stereochemistry of the backbone; (iii) satisfactory packing arrangement in the unit cell; and (iv) agreement with the fiber diffraction data.

Conformational parameters of three *B*-DNA models, finally chosen, are given in Table I. Related agreement indices (R and R'') [5] are also indicated. It is seen that two of the models are right handed: one with (C3′-*endo*, g^-g^-) conformation and the other with (C2′-*endo*, tg^-) conformation. The remaining one is left handed and has (C2′-*endo*, tg^-) conformation. It is to be noted that the "best *B*-DNA model" by Arnott and Hukins [5] is a right-handed duplex but has unorthodox (C2′-*endo*, g^-g^-) conformation and hence stereochemically unacceptable [2,3,8]. The agreement indices of all the models have similar values and are comparable with those of the "best *B*-DNA model" no matter what procedure was adopted to compute the structure factor calculations.

Following the same procedure two models of *D*-DNA were obtained. The conformational parameters and the related agreement indices are given in Table II. Both the duplexes have (C2′-*endo*, tg^-) conformation: one is right handed and the other is left handed. It may be mentioned that the *D* pattern is only obtained for polynucleotides with alternating purine–pyrimidine sequence (PAPP, for short). It is also known that wet fibers of PAPP gives characteristic *B* pattern

TABLE II. Conformational parameters and related agreement indices of *D*-DNA models using mononucleotide as the repeat.

	Right-handed model (C2′-*endo*, tg^-) conformation	Left-handed model (C2′-*endo*, tg^-) conformation
α	208	225
β	209	215
γ	302	263
δ	147	156
ϵ	61	32
ζ	154	152
χ	72	−3
(R, R'')	(0.37, 0.40)	(0.36, 0.40)

[9] which on drying goes over to the *D* pattern. Such a smooth transition is compatible with our observation that both *B* and *D* DNA can exist in either handedness.

A stereochemically allowed model of left-handed *A*-DNA was also obtained. The model has (C2′-*endo*, tg^-) conformation. The model shows general agreement with the observed data and is at present under a process of refinement The conformational parameters, at the present stage of refinement, are given in Table III. This is the first time, the stereochemical possibility of a left-handed *A*-DNA is established which leads to the conclusion that a left-handed model for DNA cannot be ruled out on the basis of the existence of $A \rightleftharpoons B$ transitions.

C. Molecular Model Building with Dinucleotide as the Repeating Unit: RU and LZ Helices

As stated earlier, a dinucleotide was chosen as the repeating unit to incorporate sequence specific molecular conformation. For this purpose, PAPP happens to be an ideal system wherein a purine nucleotide has a conformation different from the pyrimidine. Figure 2 shows three basic units for PAPP. The basic unit, shown in Figure 2(a), leads to a right-handed uniform (RU) helix. It has purine nucleotides in (C3′-*endo*, g^-g^-) conformation with glycosyl torsion (χ) in low *anti* region while pyrimidines have (C2′-*endo*, tg^-) conformation with χ in *anti* region. In a RU helix, two neighboring phosphate groups always have nearly same helical twist (t) and rise per residue (h). The conformational parameters

TABLE III. Conformational parameters of left-handed *A*-DNA using mononucleotide as the repeating unit.

Conformation	α	β	γ	δ	ϵ	ζ	χ
(C2′-*endo*, tg^-)	204	214	312	131	32	120	−12

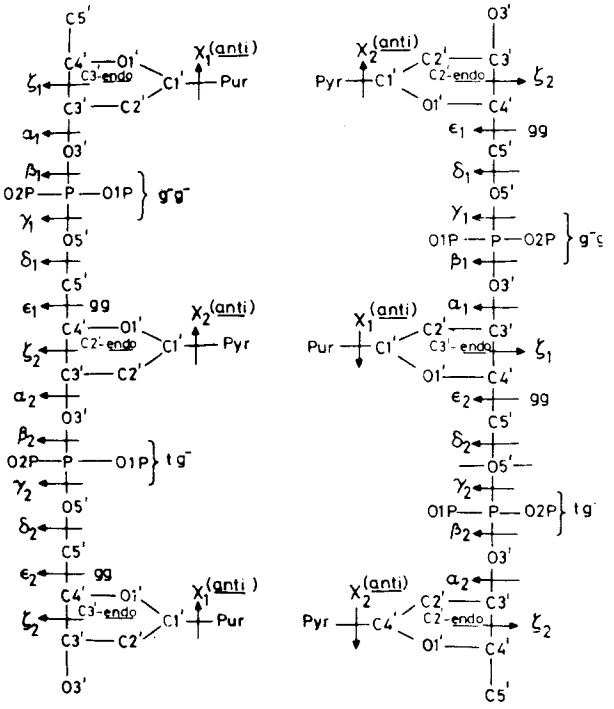

Figure 2. Three conformations of trinucleoside diphosphate as basic units which lead to RU and LZ helices. (a) Trinucleoside diphosphate with (C3′ *endo*, g^-g^--C2′-*endo*, tg^--C3′-*endo*) conformation which leads to RU helix. (b) Trinucleoside diphosphate with (C3′-*endo*, g^-t-C2′-*endo*, g^+g^+-C3′-*endo*) conformation which leads to a left-handed zig-zag (LZ1) helix. (c) Trinucleoside diphosphate with (C3′-*endo*, g^+g^+-C2′-*endo*, g^+g^+-C3′-*endo*) conformation which leads to a left-handed zig-zag (LZ2) helix.

of the RU helix in *B* form using dinucleotide as the repeat are given in Table IV. This model gave good agreement with the fiber data of calf-thymus *B* DNA.

Figure 2(b) gives the basic unit which leads to a left-handed zig-zag (LZ1) helix. This has purine nucleotides in (C3′-*endo*, g^-t) conformation (a helical domain) and pyrimidine nucleotides in (C2′-*endo*, g^+g^+) conformation (a nonhelical domain) [1,10,11]. Figure 2(c) gives the basic unit of another left-handed zig-zag (LZ2) helix. This structure has purine nucleotides in (C3′-*endo*, g^+g^+) conformation and pyrimidines in (C2′-*endo*, g^+g^+) conformation (both being nonhelical domains) [12]. However, the LZ1 and LZ2 helices have the following conformational features in common: (i) C3′-*endo* sugar have *gt* conformation around C4′—C5′ bond while C2′-*endo* sugars have *gg* conformation around the same bond; (ii) purines are attached to C3′-*endo* sugars and have *syn* conformation of the bases while pyrimidines are attached C2′-*endo* sugars and have *anti* conformation of the bases.

A schematic representation of LZ1 and LZ2 structures are given in Figure 3, where phosphate groups are marked by dark circles. LZ2 structure [Fig. 3(b)]

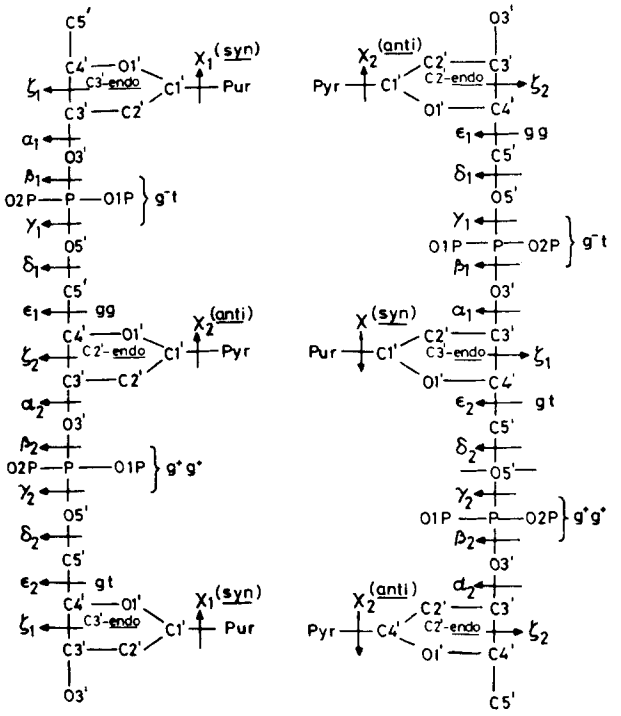

Fig. 2. (*Continued from previous page.*)

has a typical Z character such that two neighboring phosphate groups are either related by $t = -72°$ and $h = 0$ Å or $t = 0$ and $h = 6.8$ Å taking B-DNA as an example. LZ1 helix [Fig. 3(a)] is in between a LU and LZ2 helix such that two neighboring phosphate groups are related either by $t = -49°$ and $h = 1.5$ Å or $t = -23°$ and $h = 5.3$ Å for B-DNA.

Conformational parameters of LZ1 and LZ2 helices in B and Z form are given in Table V. LZ helices in B form suggests that the compression of helix with pitch, $p = 45$ Å (Z form) to $p = 34$ Å (B form) can be effected through slight changes of different torsion angles (see Table V).

LZ helices (both LZ1 and LZ2) are only possible for PAPP because such structures involve alternate bases in *syn* conformation which is stereochemically permitted only for purines and not for pyrimidines [13,14]. Hence, LZ models given in Table V, should only be tried for agreement with the fiber data of PAPP in B form (or Z form). However, such quantitative data are not available for comparison. In absence of that, we computed the structure factor amplitudes of LZ1 and LZ2 helices in B form and compared with the data of calf-thymus DNA in B form. The agreement indices were generally on the higher side when all the 150 reflections of B-DNA were included. On omitting about 30 reflections in the far off region of the reciprocal space the agreement indices dropped to values near 0.40. In view of the fact that LZ1 and LZ2 helices represent really two extremes, it may not be difficult to fit one model or the other to observed

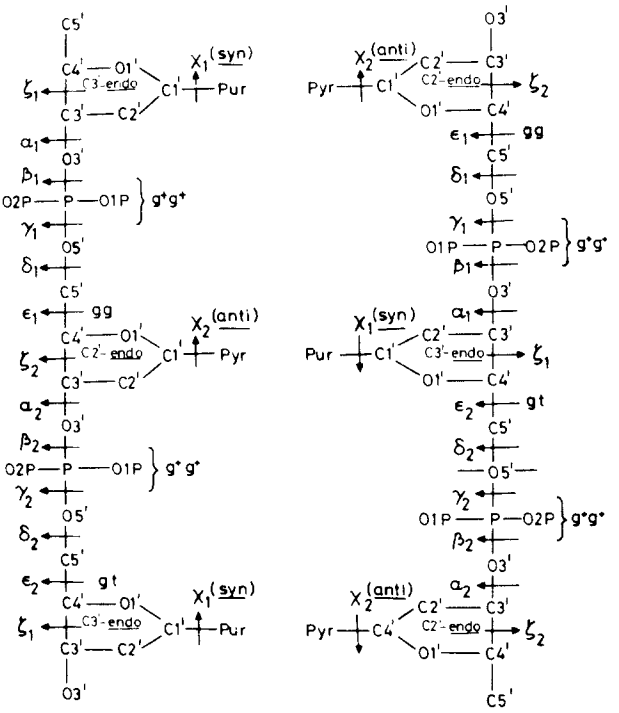

Fig. 2. (*Continued from previous page.*)

TABLE IV. Conformational parameters and related agreement indices of RU helix in *B* form using dinucleotide as the repeating unit

	Pur–Pyr sequence (C3′-*endo*, g^-g^--C2′-*endo*) conformation		Pyr–Pur sequence (C2′-*endo*, tg^--C3′-*endo*) conformation
χ (5′)	24		58
ζ (5′)	89		156
α	183		205
β	292		210
γ	295		286
δ	181		147
ϵ	66		49
ζ (3′)	156		89
χ (3′)	58		24
	I	(0.33, 0.36)	
(R, R'')[a]	II	(0.34, 0.37)	

[a] See footnote to Table I.

data, when available for PAPP, by suitably adjusting the gross structure of the molecule. However, it is to be noted that the scattering profile of LZ1 and LZ2 helices are distinctly different from one another in one respect. In both the

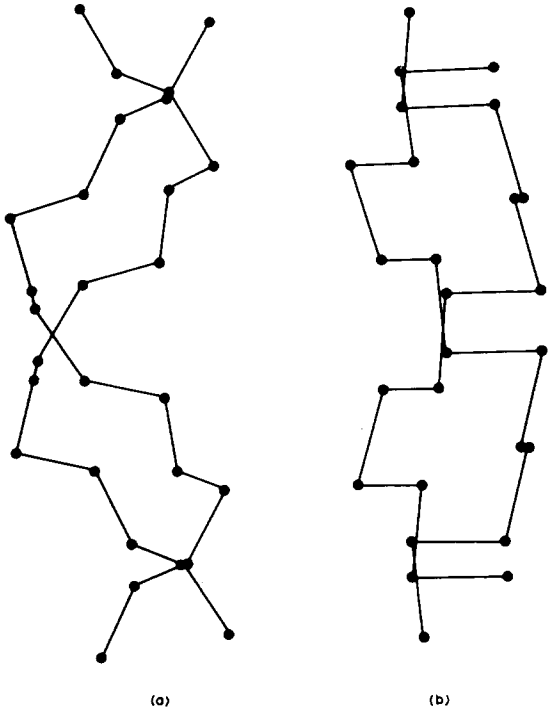

Figure 3. Representation of LZ1 and LZ2 structures. Zig-zag progression of the phosphate groups only is shown.

structures, two neighboring phosphate groups have dissimilar spatial disposition. Thus, for LZ helices in *B* form, a 005 reflection is predicted. However calculations showed that 005 reflection is obtained for LZ1 helix only and for LZ2 helix it is insignificantly small. Therefore, it appears that the spatial interrelation of the neighboring phosphate groups is a more important factor in deciding the occurrence of 005 reflection than the actual spatial positions of the phosphate groups round the helix axis.

Conclusion

Based upon the results discussed above, the following conclusions can be drawn:

(i) Fiber data of a given polymorphous form of DNA give an image of the gross structure of the molecule and elude any direct information about the fine structure (repeating unit).

(ii) Based on a stereochemical guideline, one can obtain conformationally distinct repeating units (fine structure) which could give rise to roughly the same gross structure of a given polymorphic DNA.

(iii) Models of a given form of DNA, no matter what handedness they have, agree with the fiber data as long as they have similar gross structure.

TABLE V. Conformational parameters of LZ1 and LZ2 helices in *B* and *Z* forms

Helix	Torsion angles	*B* form		*Z* form	
		Pur–Pyr sequence	Pyr–Pur sequence	Pur–Pyr sequence	Pyr–Pur sequence
LZ1	χ (5′)	240	15	220	10
	ζ (5′)	95	120	85	120
	α	237	259	242	262
	β	325	105	320	96
	γ	210	87	212	90
	δ	178	190	182	194
	ϵ	69	160	68	161
	ζ(3′)	120	95	120	85
	χ(3′)	15	240	10	220
LZ2	χ(5′)	210	20	220	28
	ζ(5′)	86	156	96	153
	α	192	235	183	275
	β	98	94	102	85
	γ	108	66	99	67
	δ	148	184	146	168
	ϵ	85	154	75	150
	ζ(3′)	156	86	153	96
	χ(3′)	20	210	28	220

(iv) DNA in either handedness is stereochemically possible for *A*, *B*, and *D* forms of DNA and is consistent with the observed fiber data.

Acknowledgments

This work was supported by the Department of Science and Technology, New Delhi, Grant to Professor V. Sasisekharan. One of us (G.G.) wishes to thank U.G.C., India, for a fellowship.

Intensity data of *A*-DNA were obtained from Professor H. R. Wilson.

Bibliography

[1] V. Sasisekharan and G. Gupta, Curr. Sci. **49,** 43 (1980).
[2] G. Gupta, M. Bansal, and V. Sasisekharan, Int. J. Bio. Macromol. **2,** 368 (1980).
[3] G. Gupta, M. Bansal, and V. Sasisekharan, Proc. Natl. Acad. Sci. USA **77,** 6486 (1980).
[4] R. Langridge, D. A. Marvin, W. E. Seeds, H. R. Wilson, C. W. Hooper, M. H. F. Wilkins, and L. D. Hamilton, J. Mol. Biol. **2,** 19 (1960).
[5] S. Arnott and D. W. L. Hukins, J. Mol. Biol. **81,** 93 (1973).
[6] N. C. Seeman, J. M. Rosenberg, F. L. Suddath, J. J. P. Kim, and A. Rich, J. Mol. Biol. **104,** 145 (1976).
[7] S. Arnott, R. Chandrasekaran, D. W. L. Hukins, P. J. C. Smith, and L. Watts, J. Mol. Biol. **88,** 523 (1974).
[8] N. Pattabiraman, S. N. Rao, and V. Sasisekharan, Nature (London) **284,** 187 (1980).
[9] D. R. Davies and R. L. Baldwin, J. Mol. Biol. **6,** 251 (1963).
[10] V. Sasisekharan, G. Gupta, and M. Bansal, Int. J. Biol. Macromol. **3,** 2 (1981).

[11] G. Gupta, M. Bansal, and V. Sasisekharan, Biochem. Biophys. Res. Commun. **97,** 1258 (1980).
[12] G. Gupta, M. Bansal, and V. Sasisekharan, Biochem. Biophys. Res. Commun. **95,** 728 (1980).
[13] H. Berthod and B. Pullman, FEBS Lett. **30,** 231 (1973).
[14] N. Yathindra and M. Sunderlingam, Biopolymers **12,** 2075 (1973).