

# Molecular Mechanics Studies on Poly(Purine) · Poly(Pyrimidine) Sequences in DNA: Polymorphism and Local Variability

MANJU BANSAL, *Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560012, India, and School of Pharmacy, University of California, San Francisco, California 94143, USA*; and  
N. PATTABIRAMAN,\* *School of Pharmacy, University of California, San Francisco, California 94143, USA*

## Synopsis

Energy minimization has been carried out on three poly(purine) · poly(pyrimidine) sequences— $d(G)_{10} \cdot d(C)_{10}$ ,  $d(A)_{10} \cdot d(T)_{10}$ , and  $d(AG)_5 \cdot d(CT)_5$ —using the molecular mechanics program AMBER (Assisted Model Building and Energy Refinement). In order to extensively scan the conformational space available, five different helical models were studied, three of them being right-handed helices while the other two were left helical. For all three sequences the right-handed A- and B-type helices are energetically slightly preferred over the left helices, but the energy difference between the various right-handed helices is only marginal. A detailed analysis has been carried out to characterize the local structural variability in the refined structures, both in terms of torsion angles as well as other parameters such as base-pair tilt, wedge roll, and wedge tilt, etc. All three sequences exhibit similar structural features for a particular form, but both the forms A and B show significant deviations from fiber models. In particular, the A-form structures have higher unit rise (2.7 Å), and lower unit twist (31°) and base-pair tilt (12°), compared to the fiber model, which has corresponding values of 2.56 Å, 32.7°, and 20°, respectively. All these changes indicate that the refined models are closer to the A-form structure observed in crystals of oligonucleotides. In the refined B-form models, the helical parameters are close to the fiber B-form, although the torsion angles show considerable variations. None of the three sequences examined, including the  $d(A)_n \cdot d(T)_n$  sequence, show any pronounced curvature for the B-form structure.

## INTRODUCTION

The occurrence of various polymorphic forms of DNA, which are readily interconvertible among themselves, has been attributed to several different factors. The driving force for the conformational transitions is generally believed to be the amount of water activity around the DNA molecule. Presence of increased water of hydration is seen to favor the B-form while a reduction of water activity on addition of salts or organic solvents is accompanied by a transition to the A-form or the left-handed Z-form.<sup>1,2</sup> Another hypothesis proposed recently suggests that the various base-pair steps intrinsically favor a “bistable” stacking arrangement, leading to the two commonly observed helical structures, viz. A and B.<sup>3</sup> Earlier molecular mechanical calculations have been carried out for hexanucleotide fragments with the

\*Present address: Code 6030, Laboratory for Structure of Matter, Naval Research Laboratory, Washington, D.C. 20357-5000, USA.

standard A-, B- and Z-form models of DNA, using the united atom approximation.<sup>4,5</sup> Energy minimization in terms of torsion angles and using a mononucleotide repeating unit have also been reported for a few sequences.<sup>6-8</sup> Recently a detailed molecular mechanics study has been carried out for alternating purine-pyrimidine sequences, using several different starting models,<sup>9</sup> while energy values for all possible sequences in the best right- and left-handed structures proposed from fiber diffraction studies have also been reported.<sup>10</sup> However, there is yet no detailed study of the inherent forces that stabilize (or destabilize) the two common helical structures for the different nucleotide sequences and of the minor variations likely to occur within the two families of structures.

In this paper we report the results of detailed molecular mechanical studies on three homo(purine) · homo(pyrimidine) sequences, using the all-atom version of the molecular mechanics program AMBER (Assisted Model Building and Energy Refinement).<sup>11,12</sup> The homo(purine) · homo(pyrimidine) sequences have been implicated in the formation of the "curved DNA" as well as several other functional roles,<sup>13,14</sup> and a detailed understanding of their conformational flexibility is therefore essential. The calculations have been carried out for three different sequences—d(G)<sub>10</sub> · d(C)<sub>10</sub>, d(A)<sub>10</sub> · d(T)<sub>10</sub>, and d(AG)<sub>5</sub> · d(CT)<sub>5</sub>—and with five different starting models—the standard right-handed A-form model and four B-form models, two each of right and left handedness. These five models cover a wide spectrum of possible conformations for the DNA molecule and a detailed analysis of all ten possible sequences in a base-paired dinucleoside monophosphate fragment indicates that all five models, with some structural modifications, are energetically equally favorable.<sup>15</sup> However, there are subtle differences in the interactions stabilizing the different models. Hence the calculations reported here on larger fragments help in determining whether longer range interactions reinforce these differences or reduce them. In addition to comparing these models on the basis of their minimum energy, a detailed examination has been carried out of the local structural variations in the helical and base-pair parameters, like wedge roll, wedge tilt, etc., which are believed to be responsible for the reported curvature of DNA in solution as well as in crystals.<sup>3,16,17</sup>

## METHOD

The molecular mechanics calculations have been carried out using the all atom force field and the computer program AMBER.<sup>11,12</sup> Since the minimization procedures only lead to a local minimum, within the helical domain of the starting models, we have tried to circumvent the problem by refining five different starting models. These correspond to

- (i) the right-handed A-DNA (henceforth referred to as RA) model of Arnott and Hukins<sup>18</sup>,
- (ii) the right-handed stereochemically favorable model for B-DNA with best fit to x-ray data (RB2)<sup>19,20</sup>,
- (iii) the right-handed B-DNA model with C3'-*endo* sugar pucker (RB3) and other torsion angles also close to the values in RA<sup>19,20</sup>,
- (iv) the left-handed B-DNA model (LB1) with C2'-*endo* sugar pucker, the torsion angles about the phosphodiester bonds O3'-P and P-O5' being *trans*

and *gauche*<sup>-</sup>, respectively, and the orientation about the C3'—O3' bond being *trans*, as in all the three right-handed structures listed above<sup>19,20</sup>; and

(v) left-handed B-DNA model (LB2) with C2'-*endo* sugar pucker, the P—O torsion angles being *trans* and *gauche*<sup>+</sup>, and the torsion angle about the C3'—O3' bond having the *gauche*<sup>-</sup> orientation, rather than the commonly occurring *trans* conformation.<sup>19,20</sup>

These five models will be referred to henceforth by their sequence, GG, AA, and AG, followed by RA (for right-handed A-form) RB2 (model ii), RB3 (model iii), LB1, and LB2 (left-handed B-form models iv and v described above). Thus the right-handed C2'-*endo* model for d(AG) · d(CT) is referred to as AGRB2 while the left-handed model (v) above is referred to as AGLB2. The atomic coordinates were generated using the computer program NUCGEN,<sup>21</sup> for decamer fragments in each of the above five conformations, with the three homo(purine) · homo(pyrimidine) sequences d(G)<sub>10</sub> · d(C)<sub>10</sub>, d(A)<sub>10</sub> · d(T)<sub>10</sub>, and d(AG)<sub>5</sub> · d(CT)<sub>5</sub>. Both polar and nonpolar hydrogen atoms were fixed using the program FIXHYD.<sup>21</sup> All 15 structures so generated were refined until the root mean square gradient was less than 0.1 Kcal/mole/Å. A distance-dependent dielectric constant ( $\epsilon = R_{ij}$ ) was used to compute the electrostatic interactions since the solvent and ion atoms were not considered explicitly. Such an approximation has been found to be most appropriate to mimic the solvent contribution.<sup>11,12,22</sup> Preliminary calculations with  $\epsilon = 4R_{ij}$  indicated that the structures refined to almost identical conformation as with  $\epsilon = R_{ij}$ , although the absolute values for the electrostatic energy were much reduced. Hence all the minimizations were carried out with  $\epsilon = R_{ij}$  and only for the refined models were the electrostatic energy calculated with  $\epsilon = 4R_{ij}$ , as well as the value of  $\epsilon$  according to the formula of Hingerty et al. [23] in which  $\epsilon \approx 2R_{ij}$  for small values of  $R$ , but increases exponentially so as to reach a value of 80 for  $R_{ij} \approx 15$  Å. The 1–4 nonbonded and 1–4 electrostatic energy contributions were scaled to one-half as suggested by Wiener et al.<sup>12</sup>

In addition to comparing the various models in terms of their energy and conformational parameters, a detailed analysis has been carried out to compare their structural parameters. For this purpose a program similar to that described by Fratini et al.<sup>24</sup> has been developed.<sup>25</sup> The neighboring units have been compared in absolute space, without involving a global helical axis. Such an approach is useful when examining local variations in curved molecules, particularly when comparing fragments of different lengths. As expected, the program gives exactly similar values for the ideal fiber models. However, for purposes of calculating unit rise ( $h$ ) and base pair tilt ( $\Theta_T$ ), an average helix axis, obtained for the central tetrameric unit was used. The propeller twist ( $\Theta_P$ ) is the angle between the two base normals in the hydrogen-bonded pair and contains some contribution due to buckling of the bases, in addition to twisting about the long axis of the base pair, defined by the vector joining the atoms C8 of purine and C6 of pyrimidine. In order to calculate the unit-twist, wedge-roll, and wedge-tilt parameters between neighboring base pairs, a different set of axes were defined. The average of the two base normals was taken as the  $z$  axis, the above-mentioned long axis of the base pair as the  $y$  axis, and a vector perpendicular to both these as the  $x$  axis. The angle between the  $x$  axis of  $i$ th unit and that ( $i + 1$ )th unit defines the helical twist. In order to define the wedge-roll and the wedge-tilt parameters independently

of the helix axis, they were calculated in terms of the angle between the  $z$  axis (base normal) of the  $i$ th unit and the vectors along the  $y$  and  $x$  axes of the  $(i + 1)$ th unit, respectively.

## RESULTS AND DISCUSSION

### Energetics of Various Models

The total energy for the decamers are listed first in Tables I–III for the sequences  $d(G)_{10} \cdot d(C)_{10}$ ,  $d(A)_{10} \cdot d(T)_{10}$ , and  $d(AG)_5 \cdot d(CT)_5$ , respectively, and then the various components are given in each case. The minimized structures for  $d(G)_{10} \cdot d(C)_{10}$  are shown alongside the starting conformations in Fig. 1(a–d) for the three right-handed models (RA, RB2, RB3) and one of the left-handed models (LB2). Only the four energy-minimized structures are shown for  $d(A)_{10} \cdot d(T)_{10}$  and  $d(AG)_5 \cdot d(CT)_5$  in Figs. 2(a–d) and 3(a–d), respectively. In all cases the root mean square deviations between starting and refined models are about 1 Å. From Tables I–III it is clearly seen that the total energy of all five refined models is within 10% of the most minimum refined value, and in general the difference is less than 5.0 Kcal/mole base pair for all models for the three sequences studied. However, an examination of the various energy components reveals that when compared to the right-handed helices, the left-handed structures are destabilized due to slightly unfavorable van der Waals interactions. This is compensated to some extent by favorable electrostatic interactions, as reported earlier for alternating purine–pyrimidine sequences.<sup>9</sup> These results are also similar to those obtained very recently for  $d(A)_{12} \cdot d(T)_{12}$ , using a more sophisticated electrostatic function.<sup>26</sup> In fact, the total energy obtained for the left-handed model by these workers was even lower than ours, presumably due to their using the united atom version of the AMBER program and a left-handed C-DNA (with unit twist of  $-40^\circ$ ) as the starting structure.<sup>27</sup> Thus while the values reported here for the electrostatic energy may not be absolutely accurate due to approximation of the solvent effects, they are apparently a good indication of the relative stability of various models. This was confirmed by calculations of the electrostatic energies for the refined models using dielectric constant  $\epsilon = 4R_{ij}$ , as well as the formula of Hingerty et al.<sup>23</sup> As expected, the absolute values as well as the differences between the electrostatic energies of the various models reduced considerably, particularly for the longer fragments. Therefore, for each of the sequences considered here, the three right-handed models now have very similar values for the electrostatic energy and this is comparable to the values for the left-handed structures. This leads to the left-handed models being predicted to be even more unfavorable energetically when compared to the right-handed structures, since the left-handed models have less favorable van der Waals energies.

Since there was a tendency for the terminal base pairs to be frayed and to show large variations in their orientation, a detailed analysis was carried out on the central tetramer fragment. This fragment will also be free of any artifacts due to long-range electrostatic effects, which are most difficult to model. As expected, the energy differences are smaller for the tetranucleotide fragments when compared to the decamer, with all five models being within

TABLE I  
Energy Component Analysis<sup>a</sup> of the Deca-, Tetra-, and Monomer units of the Refined Models of  $d(G)_n \cdot d(C)_n$

	RA	RB2	RB3	LB1	LB2
$d(G)_{10} \cdot d(C)_{10}$					
$E_{\text{tot}}$	-814.2	-804.6	-763.9	-764.5	-808.3
$E_{\text{vdw}}$	-223.6	-229.2	-228.4	-212.3	-203.5
$E_{\text{es}}$	-863.8	-841.7	-814.1	-885.3	-912.2
$E_{\text{hb}} (10-12)$	-2.4	-4.9	-8.1	-3.6	-2.6
$E_{\text{tor}}$	182.7	182.2	210.9	222.9	200.9
$E_{\text{ba}}$	84.8	81.2	67.8	105.1	99.9
$E_{\text{bl}}$	7.9	7.7	8.1	8.7	9.2
$d(G)_4 \cdot d(C)_4$					
$E_{\text{tot}}$	-324.3	-323.9	-312.6	-308.2	-313.2
$E_{\text{vdw}}$	-71.8	-69.6	-72.3	-63.8	-55.8
$E_{\text{es}}$	-356.5	-358.1	-350.4	-375.2	-374.7
$E_{\text{others}}$	104.0	103.8	110.1	130.8	117.3
Average base-stacking energy					
$E_{\text{vdw}}$	-17.9	-18.1	-17.4	-15.2	-17.2
$E_{\text{es}}$	+0.3	+0.5	-0.2	-0.6	-0.3
Average nucleoside (base + sugar) stacking energy					
$E_{\text{vdw}}$	-23.4	-23.3	-22.8	-21.9	-22.0
$E_{\text{es}}$	+1.2	+0.4	+0.9	-0.2	-0.3
Average internal energy of nucleoside unit (N)					
Gua	-20.5	-20.0	-19.2	-20.0	-19.6
Cyt	-19.1	-19.2	-18.2	-18.0	-17.5
(5'P...N)	-4.7/-4.8	-4.9/-5.3	-4.5/-5.3	-5.1/-3.1	-4.4/-4.4
(N...3'P)	0.0/-0.1	-0.1/-0.2	-0.9/-1.1	-2.6/-2.9	-2.3/-2.8
(P...P)	0.7/1.1	0.6/0.6	0.8/0.8	0.4/0.4	0.4/0.3

<sup>a</sup>The energy components (in kcal/mole) are defined as in AMBER.<sup>12</sup>  $E_{\text{tot}}$ , total energy;  $E_{\text{vdw}}$ , van der Waals energy calculated using the Lennard-Jones (6-12) potential;  $E_{\text{es}}$ , electrostatic energy;  $E_{\text{hb}}$ , hydrogen-bond energy calculated using a 10-12 potential function;  $E_{\text{tor}}$ , torsional energy;  $E_{\text{ba}}$ , bond angle distortion energy;  $E_{\text{bl}}$ , bond length distortion energy.

<sup>b</sup>The 5' phosphate-nucleoside, nucleoside-3' phosphate, and neighboring phosphate-phosphate interaction energy have been listed separately for each strand (purine/pyrimidine).

TABLE II  
Energy Component Analysis as in Table I, but for the Sequence  $d(A)_n \cdot d(T)_n$

	RA	RB2	RB3	LB1	LB2
$d(A)_{10} \cdot d(T)_{10}$					
$E_{tot}$	-578.9	-574.3	-551.9	-538.2	-555.3
$E_{vdw}$	-204.2	-212.5	-217.8	-196.6	-174.7
$E_{es}$	-643.2	-625.4	-612.4	-676.7	-678.1
$E_{hb}$ (10-12)	-2.1	-2.1	-5.0	-2.2	-2.6
$E_{tor}$	186.5	186.4	206.9	232.4	191.8
$E_{ba}$	77.6	72.1	68.7	95.5	99.2
$E_{bl}$	7.5	7.0	7.8	8.3	8.6
$d(A)_4 \cdot d(T)_4$					
$E_{tot}$	-230.9	-235.7	-230.0	-221.4	-222.6
$E_{vdw}$	-63.9	-63.8	-68.3	-57.6	-46.4
$E_{es}$	-269.2	-274.3	-272.2	-292.1	-290.6
$E_{others}$	101.2	102.4	110.5	128.3	114.4
Average base-stacking energy					
$E_{vdw}$	-17.1	-17.6	-17.3	-15.4	-15.5
$E_{es}$	-0.3	-0.3	-0.3	-0.4	-0.8
Average nucleoside (base + sugar) stacking energy					
$E_{vdw}$	-22.8	-23.9	-22.4	-22.0	-19.4
$E_{es}$	+0.8	+0.6	+0.5	+0.9	-1.4
Average internal energy of nucleoside unit (N)					
Ade	-16.7	-17.4	-16.3	-17.2	-16.2
Thy	-9.8	-9.4	-9.1	-9.2	-7.8
5'P...N	-4.7/-4.9	-5.0/-5.2	-4.6/-5.5	-3.0/-4.2	-4.4/-4.5
N...3'P	-0.2/-0.8	-0.1/-0.1	-1.0/-1.1	-3.1/-3.0	-2.7/-2.8
P...P	0.8/1.1	0.7/0.6	0.8/0.9	0.3/0.4	0.4/0.4

TABLE III  
Energy Component Analysis as in Table I, but for the Sequence  $d(AG)_n \cdot d(CT)_n$

	RA	RB2	RB3	LB1	LB2
$d(AG)_6 \cdot d(CT)_6$					
$E_{\text{tot}}$	-695.8	-688.7	-658.5	-651.5	-685.1
$E_{\text{vdw}}$	-209.5	-219.2	-222.6	-200.8	-188.7
$E_{\text{es}}$	-757.0	-734.8	-718.2	-782.7	-798.5
$E_{\text{hb}} (10-12)$	-3.0	-2.7	-4.6	-2.8	-3.8
$E_{\text{cor}}$	189.6	184.8	211.4	221.8	196.4
$E_{\text{ba}}$	76.6	75.9	67.8	104.5	100.3
$E_{\text{bl}}$	7.3	7.3	7.7	8.5	9.1
$d(GA)_2 \cdot d(CT)_2$					
$E_{\text{tot}}$	-277.7	-279.8	-270.0	-259.6	-268.5
$E_{\text{vdw}}$	-65.6	-65.5	-69.5	-58.0	-50.5
$E_{\text{es}}$	-314.1	-318.2	-313.2	-332.7	-334.4
$E_{\text{others}}$	102.0	103.9	112.7	131.1	116.4
Base-stacking energy (A-G step/G-A step)					
$E_{\text{vdw}}$	-17.4/-16.8	-17.2/-18.1	-16.6/-17.1	-14.4/-13.3	-16.0/-16.1
$E_{\text{es}}$	-0.3/-1.3	-0.7/-0.3	-1.2/+0.7	-0.4/-0.1	-0.5/-0.7
Average internal energy of nucleoside unit (N)					
Gua/Ade	-20.2/-16.8	-20.4/-17.0	-19.2/-16.6	-19.8/-15.5	-19.7/-16.1
Cyt/Thy	-19.1/-9.6	-18.3/-10.0	-18.7/-8.5	-18.6/-8.7	-17.3/-7.5
5'P...N	-4.7/-5.0	-4.9/-5.5	-4.5/-5.4	-4.7/-5.0	-4.3/-4.5
N...3'P	-0.1/-0.7	-0.1/-0.4	-0.9/-1.1	-3.0/-2.9	-2.7/-2.8
P...P	0.8/1.1	0.7/0.6	0.8/0.8	0.2/0.1	0.3/0.4

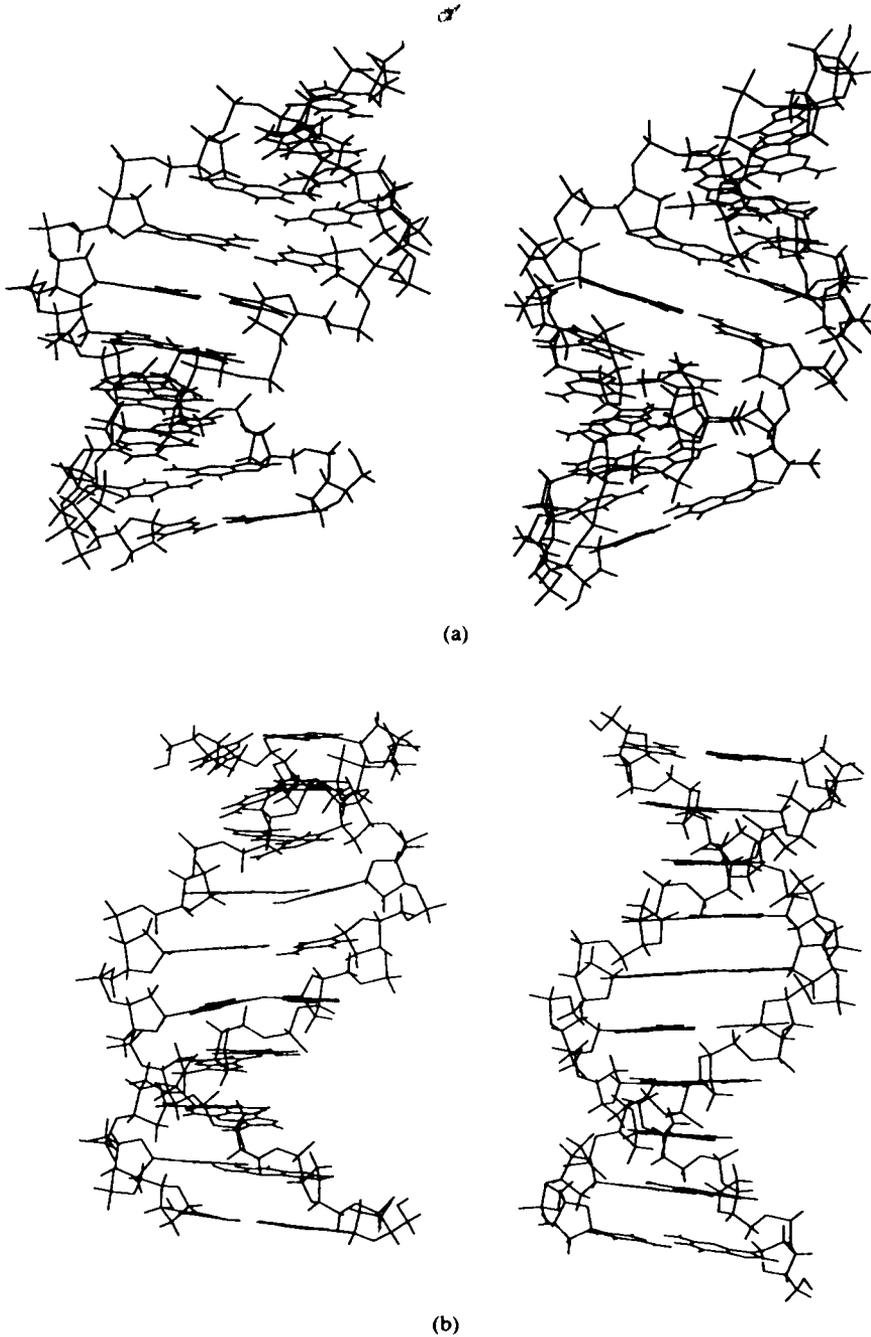


Fig. 1(a-d). The energy-refined models RA, RB2, RB3, and LB2 for  $d(G)_{10} \cdot d(C)_{10}$  are shown on the left, while the starting models obtained from x-ray fiber diffraction analysis are shown alongside, on the right. The slight buckling of the base pairs is clearly seen in all the refined models, as also the reduction in base-pair tilt for the RA model.

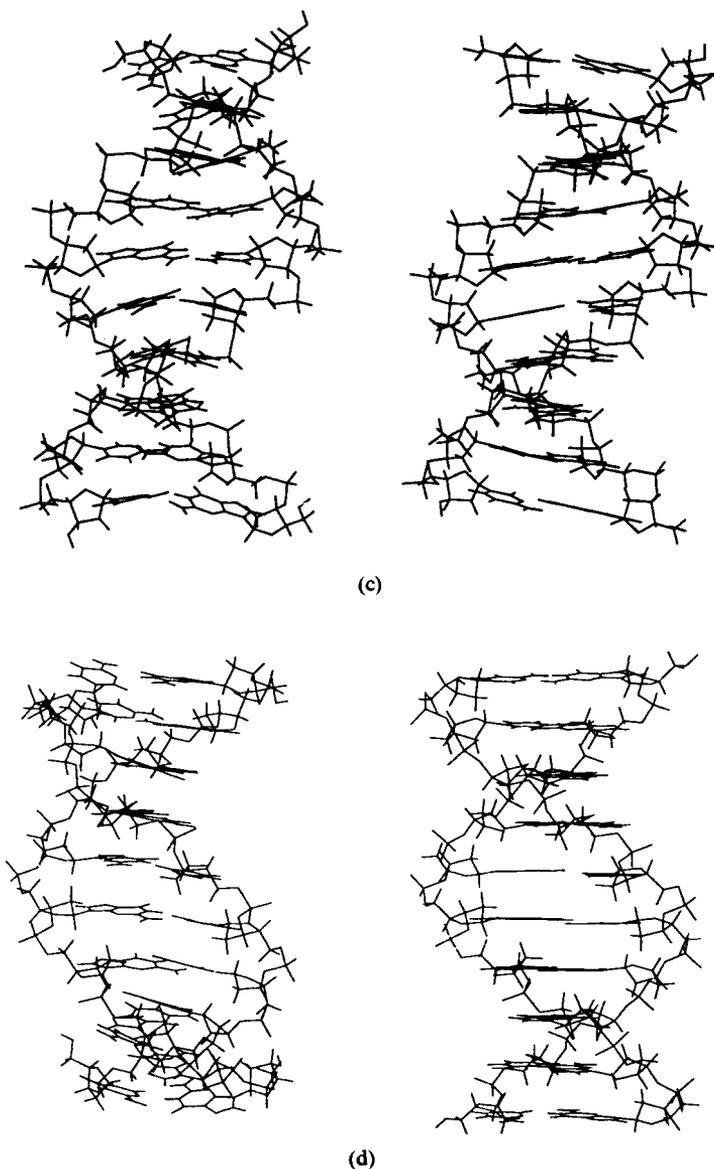


Fig. 1. (Continued from the previous page.)

5% of each other. A detailed analysis of the interactions between the various groups in DNA was also carried out for the tetramer fragments, and the results are shown in Tables I-III, as are also the breakup of stacking and monomer energy components. It can be seen from these tables that the characteristic difference of favorable nonbonded and less favorable electrostatic interactions between right- and left-handed models is seen for the stacking interactions between the bases as well as for the nucleosides. The higher internal energy for the pyrimidine nucleosides in the LB1 and LB2 models is due to the glycosidic torsion angle  $\chi$  [O4'-C1'-N9(N1)-C8(C6)] being close to  $0^\circ$ . Another feature that leads to unfavorable overall energy for the

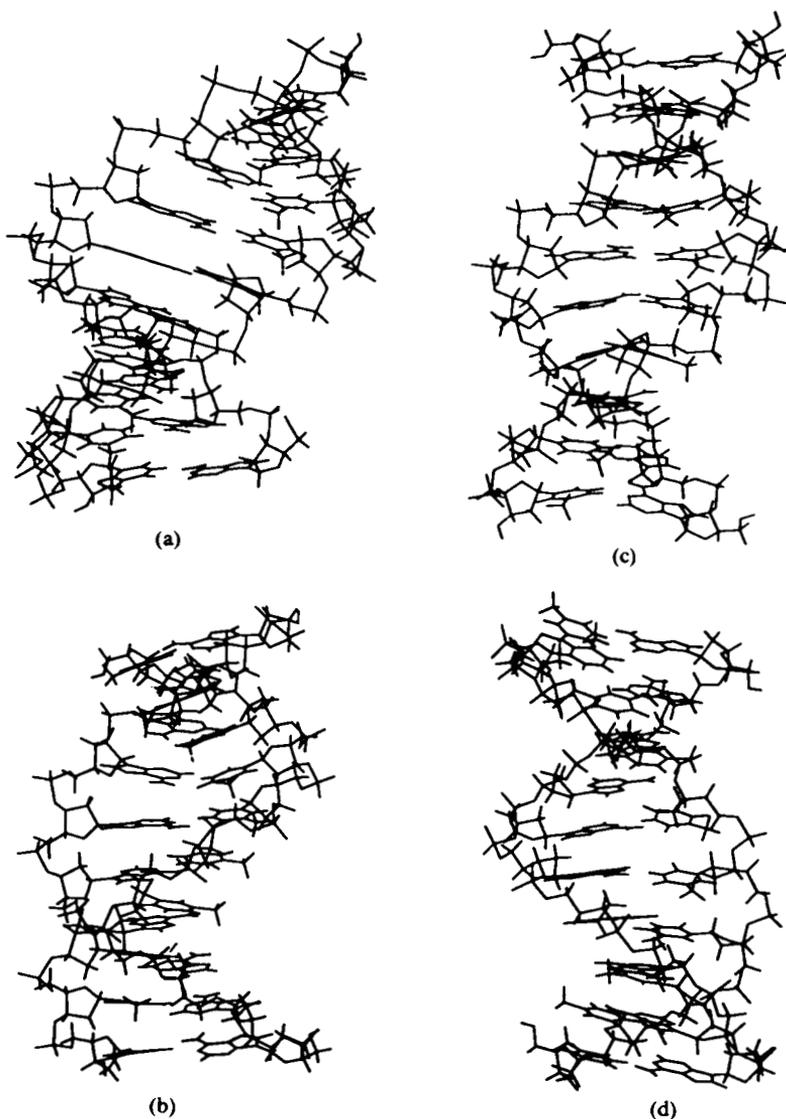


Fig. 2(a-d). The energy-refined models RA, RB2, RB3, and LB2 for  $d(A)_{10} \cdot d(T)_{10}$ .

left-handed models is the high bond angle distortion energy at the sugar ring atoms C3' and C4' and the bond angle at phosphate being  $100^\circ$  (equilibrium value =  $103^\circ$ ).

The phosphate group orientation at the 5' end is almost equally favorable for all the models; however, the orientation at the 3' end is much more favorable in the left-handed models, which have the torsion angle  $\epsilon(C4'-C3'-O3'-P)$  in the *gauche*<sup>-</sup> orientation for both the refined models, while it is *trans* for all three right-handed models.

The small energy differences between the various refined models clearly indicate that the DNA structure is intrinsically polymorphic, with several local energy minima. Hence to comprehensively scan the conformational space

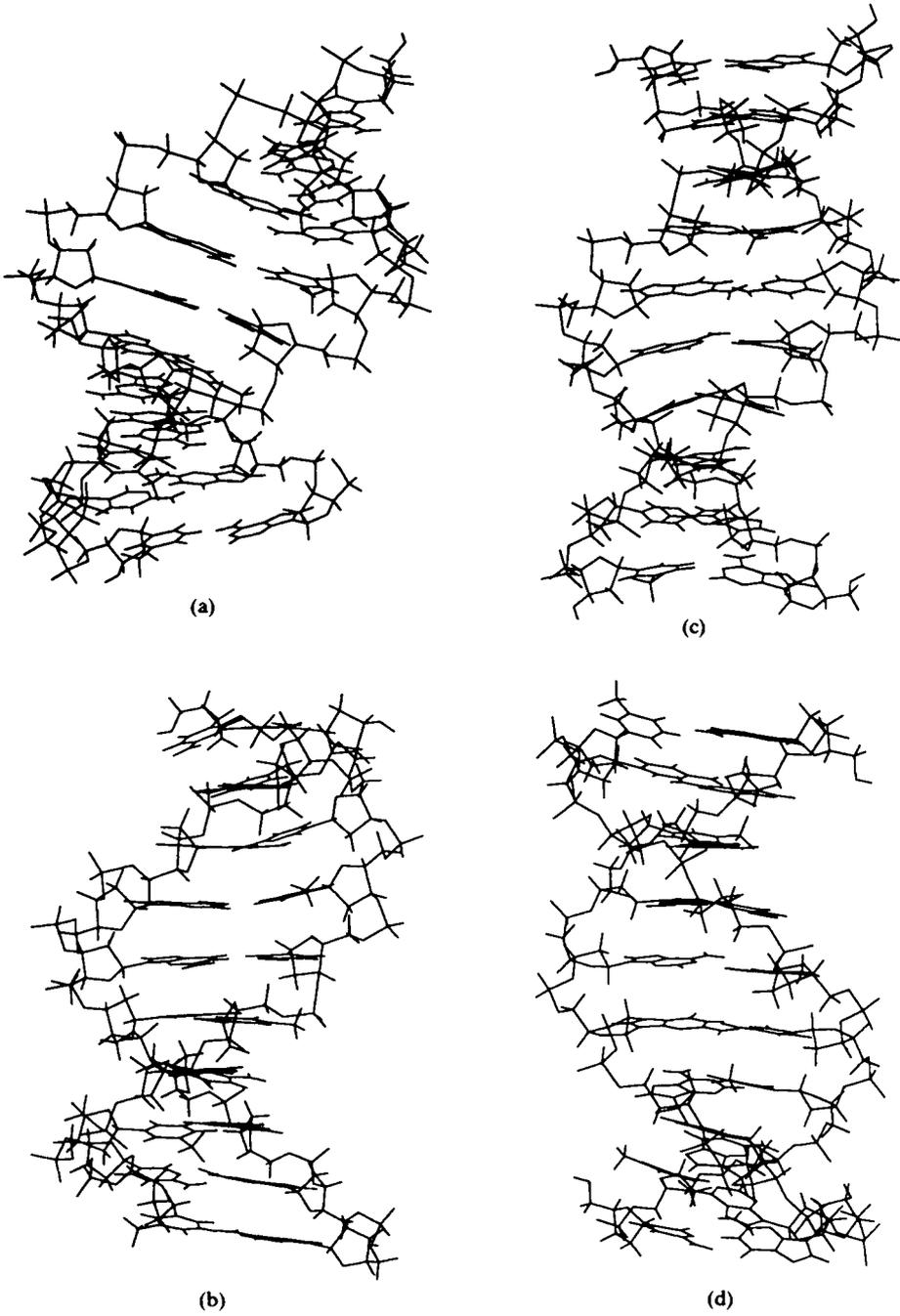


Fig. 3(a-d). Same as Fig. 2, but for  $d(\text{AG})_5 \cdot d(\text{CT})_5$ .

of DNA structure, an even larger number of starting models should be examined.

### CONFORMATIONAL AND STRUCTURAL DETAILS OF THE REFINED MODELS

The deviations of the average values of backbone torsion angles in the central tetranucleotide, in the refined models, from the starting fiber models are shown in Fig. 4. The values for the purine and pyrimidine strands are plotted separately, and can be clearly seen that all the models show similar differences between the two strands, while there are less marked differences between the various base sequences A-A, G-G, and A-G-A. This would seem to indicate that all poly(purine) · poly(pyrimidine) tracts will show similar features, like curvature, etc., although they will be length dependent, since the terminal residues generally have slightly different values for the variable torsion angles. All the torsion angles tend to approach the ideal staggered orientations (i.e., values close to  $180^\circ$  or  $\pm 60^\circ$ ) while the starting models, which have been built to fit the x-ray data, have considerable deviations from staggeredness. Another interesting feature that can be seen from Fig. 4 is the convergence of all three right-handed models (RA, RB2, and RB3) to very similar low-energy conformations. The torsion angles  $\alpha$ (P-O5'),  $\beta$ (O5'-C5'),  $\gamma$ (C5'-C4'), and  $\epsilon$ (C3'-O3') are very similar for all sequences in the three refined models of right handedness. Only the sugar torsion  $\delta$ (C4'-C3'), the phosphodiester torsion angle  $\zeta$ (O3'-P), and glycosidic torsion  $\chi$  show some differences between purines and pyrimidines, and also between the three models. A comparison of the various models shows that, as predicted by earlier theoretical studies,<sup>7,20,29</sup>  $\delta$  and  $\chi$  show a positive correlation while  $\zeta$  shows an anticorrelation with both. The differences in the torsion angles between the refined RA, RB2, and RB3 models are considerably smaller than in the starting models. In particular, some of the sugar torsion angles minimize in the C1'-*exo*, O4'-*endo* region, for both RA and RB2 models. The refined RB3 model (C3'-*endo*, B-form) has torsion angles that are closest to the starting fiber model for A-form (RA). Thus it is seen that an analysis of torsion angles alone does not give a clear picture of the details of the refined structures, although the two strands are apparently nonequivalent.

In the case of the left-handed model LB2, the structure minimizes close to the starting model, indicating that the conformational space for left-handed models is rather restricted when compared to the right-handed region, which encompasses the RA, RB2, and RB3 models. It is also seen from Fig. 4 that the left-handed LB1 model, which had the torsion angle  $\epsilon$  in the *trans* orientation, has minimized, with  $\epsilon$  going to the *gauche*<sup>-</sup> region as in the LB2 model, indicating that this is a more preferred conformation.

The variations seen in the central tetramers are similar to those reported in some of the structures of the larger oligonucleotides.<sup>17,30-32</sup> Since the torsion angles show a wide variation, depending on the base sequence and considerable deviations from the starting models, a detailed calculation was carried out to obtain other structural parameters such as unit rise and twist, base tilt, as well as the base-pair wedge parameters, roll and tilt. These structural parameters calculated for all three sequences are also listed in Table IV.

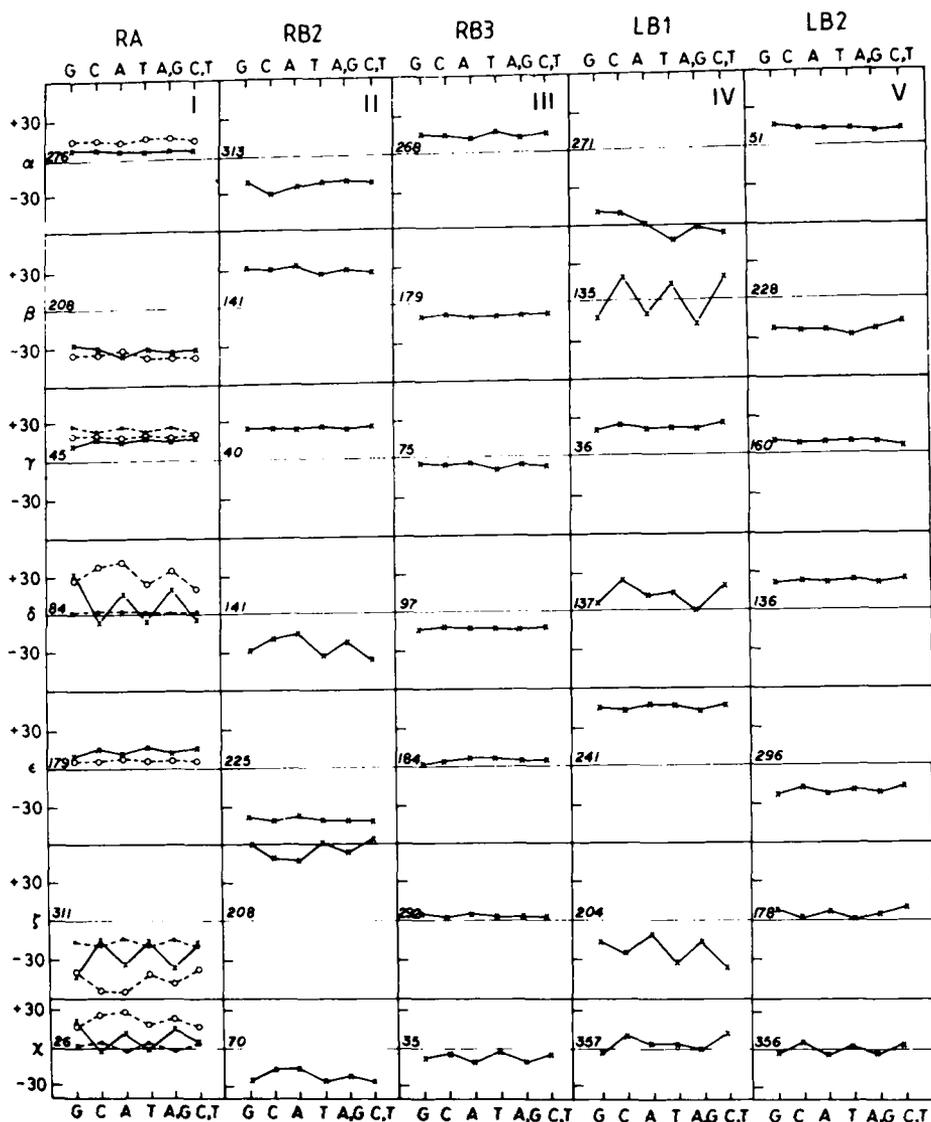


Fig. 4. Plots showing the differences (in degrees) between the five starting and refined models (in panels I–V), for the six backbone torsion angles<sup>28</sup>:  $\alpha$ (P-O5'),  $\beta$ (O5'-C5'),  $\gamma$ (C5'-C4'),  $\delta$ (C4'-C3'),  $\epsilon$ (C3'-O3'),  $\zeta$ (O3'-P), and the glycosidic torsion angle  $\chi$ (O4'-C1'-N9(N1)-C8(C6)). The values of the torsion angles in the five starting structures (RA, RB2, RB3, LB1, and LB2) are the baselines in panels I–V and are shown along the  $x$  axes, in each panel. The average value of each torsion angle was obtained for the central four nucleotides in each strand and its deviation from the corresponding starting model is shown by a ( $\times$ ). The purine nucleotides are represented by G, A, A, G and the pyrimidines by C, T, C, T in the sequences  $d(G)_n \cdot d(C)_n$ ,  $d(A)_n \cdot d(T)_n$ , and  $d(AG)_n \cdot d(CT)_n$ , respectively. In panel I, in addition to the deviations for the RA model ( $\times-\times$ ), the deviations in torsion angles for refined RB2 and RB3 models, from the fiber RA model, are also shown by ( $\circ-\circ$ ) and ( $\bullet-\bullet$ ). The torsion angles  $\alpha$ ,  $\beta$ , and  $\epsilon$  for the refined RB3 model are almost identical to those of refined RA and hence cannot be seen in panel I. Three other torsion angles  $\delta$ ,  $\zeta$ , and  $\chi$  in refined RB3 are closer to the starting RA values (baseline in panel I) than the refined RA model itself. For the refined RB2 model, torsion angles  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\epsilon$  are close to those of refined RA, while  $\delta$  and  $\chi$  show significant differences.

TABLE IV  
Structural Parameters for Three Homo(Purine) · Homo(Pyrimidine) Sequences after AMBER Refinement for Five Different Starting Models<sup>a</sup>

	RA	RB2	RB3	LBI	LB2
$d(G)_n \cdot d(C)_n$					
t (°)	30.7 (0.7)	34.6 (2.0)	35.1 (0.2)	-38.9 (0.7)	-37.4 (1.5)
h (Å)	2.75 (0.1)	3.46 (0.1)	3.47 (0.1)	3.89 (0.1)	3.45 (0.2)
$\Theta_P$ (°)	16.4 (1.0)	16.2 (3.5)	12.5 (6.3)	18.0 (7.0)	14.6 (9.7)
$\Theta_T$ (°)	9.7 (1.2)	-7.8 (2.8)	0.2 (0.1)	13.9 (1.3)	7.0 (2.3)
D (Å)	4.1 (0.1)	0.6 (0.2)	0.4 (0.1)	-2.1 (0.3)	-0.3 (0.1)
$W_r$ (°)	13.6 (1.3)	0.5 (1.9)	0.9 (1.5)	4.5 (2.2)	-2.6 (1.1)
$W_t$ (°)	3.9 (0.3)	3.3 (1.3)	0.1 (0.9)	6.6 (1.5)	-0.3 (2.7)
$d(AG)_n \cdot d(T)_n$					
t (°)	31.2 (1.1)	35.9 (2.2)	35.6 (0.1)	-38.4 (4.5)	-37.8 (1.0)
h (Å)	2.73 (0.2)	3.45 (0.2)	3.32 (0.0)	4.13 (0.0)	3.71 (0.1)
$\Theta_P$ (°)	17.1 (0.8)	14.7 (5.7)	10.2 (4.2)	13.2 (5.9)	16.1 (9.6)
$\Theta_T$ (°)	12.6 (0.8)	-6.2 (2.1)	-3.1 (0.2)	18.0 (0.8)	0.4 (0.8)
D (Å)	3.9 (0.1)	0.4 (0.2)	0.4 (0.0)	-1.8 (0.4)	-0.2 (0.1)
$W_r$ (°)	12.5 (3.9)	2.0 (1.1)	1.3 (1.2)	8.6 (3.0)	-1.7 (1.5)
$W_t$ (°)	4.6 (0.3)	5.0 (0.2)	2.4 (0.3)	6.4 (2.1)	1.6 (0.8)
$d(AG)_n \cdot d(CT)_n$					
t (°)	31.2 (1.5)	34.7 (1.1)	35.2 (0.3)	-38.2 (3.5)	-37.6 (1.3)
h (Å)	2.74 (0.2)	3.51 (0.1)	3.42 (0.0)	3.82 (0.2)	3.64 (0.1)
$\Theta_P$ (°)	16.1 (4.4)	15.2 (3.4)	11.1 (6.3)	16.7 (8.0)	16.1 (9.2)
$\Theta_T$ (°)	15.1 (1.5)	-7.5 (1.1)	-2.6 (0.3)	13.4 (2.1)	4.5 (1.2)
D (Å)	3.9 (0.1)	0.3 (0.2)	0.4 (0.1)	-1.6 (0.2)	-0.3 (0.1)
$W_r$ (°)	14.4 (4.5)	1.1 (2.0)	1.6 (1.1)	7.7 (4.6)	-2.3 (3.1)
$W_t$ (°)	4.0 (1.2)	5.6 (1.2)	2.1 (1.2)	5.1 (3.2)	0.8 (2.5)

<sup>a</sup>The values given correspond to averages for the central tetramer fragment and the standard deviation for each is given in parentheses. The nomenclature used is as follows:

t and h: the unit twist and rise between successive base pairs.

$\Theta_P$ : propeller twist (taken as angle between the two base normals and hence has a component due to base-pair buckling).

$\Theta_T$  and D: tilt and displacement of base pairs with respect to average helix axis for central tetramer.

$W_r$  and  $W_t$ : wedge roll and wedge tilt, a measure of base pairs opening along the short and long axis, respectively, of two base pairs (calculated as given in the methods section).

The values for the various parameters for the starting RA model are 32.7°, 2.56 Å, 12°, 20°, 4.5 Å, 10.6°, and -2.7°, respectively, while the values of h and t for the starting B-form models are 3.4 Å and ±36°, with all other parameters being close to 0.

The values listed in Table IV for the unit rise ( $h$ ) and twist ( $t$ ) quite clearly lie in two domains—A-region (with  $h \approx 2.7$  Å,  $t \approx 31^\circ$ ) and B-region (with  $h \approx 3.4$  Å,  $t \approx 35^\circ$ ). The left-handed models invariably minimize with slightly larger values of helical twist ( $t \approx -38^\circ$ ), which is closer to the C-DNA structure, but the unit height remains close to 3.4 Å in LB2, which is more B-like. Thus it is clearly seen that though the torsion angles between A- and B-form refined structures are more similar than in the fiber models, the other characteristic features remain distinctly different in these two forms, for all three sequences. However, some of the parameters, such as unit rise and base tilt, are noticeably different from the fiber model for the A-form. The base tilt reduces from  $20^\circ$  to about  $10^\circ$  while the base-pair displacement ( $D$ ) changes from 4.5 to 4.0 Å. Both these changes as well as the increase in unit rise from 2.56 to 2.7 Å and decrease in unit twist from  $32.7^\circ$  to  $31^\circ$  are analogous to those seen in the A-type crystal structures.<sup>30,31</sup> The mononucleotide repeat calculations, even with cations explicitly included, give most of these parameters to be considerably different (unit rise = 2.2 Å, unit twist =  $34.6^\circ$ , tilt =  $26^\circ$  and  $D = 5.6$  Å) from those observed in crystals.<sup>8</sup> The slightly higher values of wedge roll in the refined A-form structures (compared to  $10^\circ$  for the fiber model) may be an indication that the molecules tend to have a slight curvature, a feature observed in some A-form crystal structures.<sup>30</sup> However, it should be noted that although the two strands have slightly different conformations (as indicated by distinctly different values for the torsion angles  $\xi$  and  $\zeta$  in Fig. 4), it leads to the minor groove opening (increase in wedge roll) and not to any significant increase in the wedge tilt.

Among the two B-form models, the RB3 model, which has C3'-*endo* sugar pucker and torsion angles closest to the A-form fiber model, has quite different structural features and has converged to values quite close to those of the RB2 model, with C2'-*endo* sugar. This model has a slightly higher torsional energy than the RA and RB2 models for most sequences, including alternating purine-pyrimidine sequences (M. Bansal, unpublished results). However, it should be noted that while for  $d(G)_n \cdot d(C)_n$  and  $d(AG)_n \cdot d(CT)_n$  sequences the RA and RB2 models have almost identical energies, the RB2 model is slightly preferred over RA for  $d(A)_n \cdot d(T)_n$ . This could explain the observed preference of this sequence for the B-form structure. It may be mentioned that this difference is further enhanced if the sugar puckers are constrained to the C3'-*endo* region during refinement, for the A-form (M. Bansal, unpublished results).

The RB2 models for the three sequences are quite similar, with minor variations in their structural features. The unit rise lies between 3.4 to 3.5 Å and the unit twist varies from  $34.6^\circ$  for  $d(G)_n \cdot d(C)_n$  to  $35.9^\circ$  for  $d(A)_n \cdot d(T)_n$ . This trend is similar to that seen in solution ( $33.6^\circ$  and  $36.3^\circ$ , respectively) for these two polymers,<sup>33,34</sup> and only slightly better than that obtained by mononucleotide repeat calculations.<sup>8</sup> There is not much difference between the three sequences in their wedge roll and wedge tilt, thus not supporting the hypothesis that smooth bending of  $d(A)_n \cdot d(T)_n$  tracts is due to large values of wedge roll and wedge tilt for this sequence.<sup>35</sup> It may be mentioned that we have not considered a starting model with the two strands in distinctly different conformations<sup>36</sup> since recent nmr studies do not support such a model.<sup>37,38</sup> However, the refined RB2 models show some differences in their

purine and pyrimidine strands if the sugars are not constrained to remain in the C2'-*endo* conformation, but this does not per se lead to the molecule being curved. For a molecule to be curved it must have a variation in its base-pair parameters since any values of the wedge parameters if repeated monotonically over a full helical turn will lead to a uniform straight helix (as seen for the fiber model of A-DNA, which has a wedge roll value of  $10.6^\circ$ ,<sup>39,40</sup> as well as the model for polymeric  $d(A)_n \cdot d(T)_n$ <sup>35,36</sup>). Hence a minimization algorithm based on a mono- or dinucleotide repeating unit<sup>8</sup> cannot lead to a curved structure. The molecular mechanics programs EREF,<sup>41</sup> AMBER,<sup>12</sup> or CHARMM<sup>42</sup> should, however, be able to predict curvature in short DNA fragments, even if the starting models are straight, since intrinsic curvature is believed to arise due to near-neighbor interactions.<sup>35</sup>

Since the torsion angles shown in Fig. 4 and the parameters listed in Table IV correspond to the central tetramer in a decamer fragment of  $d(\text{purine})_{10} \cdot d(\text{pyrimidine})_{10}$ , a comparison with B-form crystal structure data is difficult. However, as in the case of the A-form model, the general features are quite close to those observed in the crystal structure of the recently solved dodecamer  $d(\text{CGCAAAAAGCG})$ , which has a stretch of  $d(A)_6 \cdot d(T)_6$  in the middle.<sup>32</sup> It was very gratifying to see that this fragment is quite straight and shows "conformational rigidity," thus supporting the results of our theoretical calculations. It also shows a variation in torsion angles and other local parameters, similar to that obtained by us.

The two left-handed models (LB1 and LB2) show slightly different structural features. While the unit twists are quite similar in the two models ( $-38^\circ$ ), the refined LB1 model shows features closer to a left-handed A-type structure<sup>20</sup> with high base tilt ( $15^\circ$ ), and relatively large displacement and wedge parameters, but with a very large unit rise (3.9 Å). On the other hand, the refined LB2 model is quite close to the fiber models for B- and C-form DNA,<sup>19,20,27</sup> is generally more favorable energetically than the LB1 model, and only marginally higher in energy than the RB2 model (2.5 Kcal/mole base pair).

## CONCLUSION

The results of our detailed molecular mechanics calculations clearly indicate that the various homo(purine) · homo(pyrimidine) sequences in DNA are not very different from each other. They also indicate that the DNA molecule is intrinsically polymorphic, with several different conformations occurring within 3 Kcal/mole base pair of the absolute minimum. While the right-handed models are much more favorable on the basis of van der Waals energy, the electrostatic interactions seem to be more favorable in the left-handed models. It would be interesting to see if environmental conditions can be simulated that would make the electrostatic effects more dominant and hence make the left-handed structures more preferable for such uniform homo(purine) · homo(pyrimidine) sequences, similar to the left-handed zigzag structures observed for alternating purine-pyrimidine sequences in a high salt environment.<sup>2</sup>

Another important feature to emerge from these calculations is the absence of any pronounced "wedge effect" in any of the sequences in the B-form,

including  $d(A)_n \cdot d(T)_n$ , which has been postulated to have a highly non-B-type structure.<sup>35</sup> However, the results for both A- and B-form models are in good agreement with the observations from x-ray crystal structure analysis of oligonucleotides with similar sequences and show significant deviations from the starting fiber models, particularly for the A-form. Since the best RA and RB2 models have not only similar total energy, but also comparable components for  $d(G)_n \cdot d(C)_n$  and  $d(AG)_n \cdot d(CT)_n$ , it is concluded that the two structures are intrinsically equally favorable from enthalpy considerations and entropic as well as solvent influences must determine the form observed at any time.<sup>43,44</sup> In the case of  $d(A)_n \cdot d(T)_n$  a structure with sugar pucker in the C2'-endo region and other structural features of the B-form will be slightly more preferable than one of the A-form type.

The use of the facilities at the Computer Graphics Laboratory, University of California, San Francisco (Prof. R. Langridge, director, and supported by NIH RR-1081) is gratefully acknowledged. We also thank Prof. P. A. Kollman for useful discussions. M. B. is grateful to the Council for International Exchange of Scholars, Washington, for the award of a Senior Fulbright Fellowship and to Prof. T. L. James for hospitality while in San Francisco.

### References

1. Langridge, R. L., Wilson, H. R., Hopper, C. W., Wilkins, M. H. F. & Hamilton, L. D. (1960) *J. Mol. Biol.* **2**, 19-37.
2. Rich, A., Nordheim, A. & Wang, A. H.-J. (1984) *Ann. Rev. Biochem.* **53**, 791-846.
3. Calladine, C. R. & Drew, H. R. (1984) *J. Mol. Biol.* **178**, 773-778.
4. Kollman, P. A., Weiner, P. K. & Dearing, A. (1981) *Biopolymers* **20**, 2583-2621.
5. Kollman, P. A., Weiner, P. K., Quigley, G. & Wang, A. H.-H. (1981) *Biopolymers* **21**, 1945-1969.
6. Tumanyan, V. G., Neyfeck, Yu. A. & Ill'icheva, I. A. (1984) *Biopolymers* **23**, 2441-2471.
7. Olson, W. K., Srinivasan, A. R., Cueto, M. A., Torres, R., Maroun, R. C., Cicariello, J. & Nauss, J. L. in *Biomolecular Stereodynamics*, Sarma, R. H. & Sarma, M., Eds., Adenine Press, New York, 1986, pp. 75-99.
8. Lavery, R., Sklenar, H., Zakrzewska, K. & Pullman, B. (1986) *J. Biomol. Struct. Dynam.* **3**, 989-1014.
9. Pattabiraman, N., Rao, S. N., Scott, K., Langridge, R. & Kollman, P. A. (1987) *Biopolymers* **26**, 403-414.
10. Bhattacharyya, D. & Bansal, M. (1987) *J. Biomol. Struct. Dynam.* **4**, 1027-1040.
11. Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., Jr. & Weiner, P. (1984) *J. Am. Chem. Soc.* **106**, 765-784.
12. Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A. (1986) *J. Comp. Chem.* **7**, 230-252.
13. Trifonov, E. N. (1985) *CRC Crit. Rev. Biochem.* **16**, 89-106.
14. Koepsel, R. R. & Khan, S. A. (1986) *Science* **233**, 1316-1318.
15. Bhattacharyya, D. & Bansal, M., manuscript in preparation.
16. Koo, H.-S., Wu, H.-M. & Crothers, D. M. (1986) *Nature* **320**, 501-506.
17. Dickerson, R. E., Kopka, M. L. & Pjura, P. (1985) in *Biological Macromolecules and Assemblies, Vol. 2 Nucleic Acids and Interactive Proteins*, Jurnak, F. A. & McPherson, A., Eds., Wiley-Interscience, New York, 1985, pp. 37-126.
18. Arnott, S. & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun.* **47**, 1504-1509.
19. Sasisekharan, V., Bansal, M. & Gupta, G. (1983) in *Nucleic Acids: The Vectors of Life*, Pullman, B. & Jortner, J., Eds., Reidel Publishing Co., Dordrecht, pp. 101-111.
20. Bansal, M. & Sasisekharan, V. (1986) in *Theoretical Chemistry of Biological Systems*, Naray-Szabo, G., Ed., Elsevier Science, Amsterdam, pp. 127-217.
21. Pattabiraman, N. NUCGEN and FIXHYD, computer programs, University of California, San Francisco.
22. Tidor, B., Brooks, B. & Karplus, M. (1983) *J. Biomol. Struct. Dynam.* **1**, 231-252.

23. Hingerty, B. E., Ritchie, R. H., Ferrell, T. L. & Turner, J. E. (1985) *Biopolymers* **24**, 427-439.
24. Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. (1982) *J. Biol. Chem.* **257**, 14686-14707.
25. Pattabiraman, N., Feuerstein, B. & Bansal, M., manuscript in preparation.
26. von Kitzing, E. & Diekmann, S. (1987) *Eur. Biophys. J.* **15**, 13-26.
27. Premilat, S. & Albiser, G. (1984) *J. Biomol. Struct. Dynam.* **2**, 607-613.
28. IUPAC-IUB Joint Commission on nomenclature for the Description of Conformation of Polynucleotides (1983) *Eur. J. Biochem.* **131**, 9-15.
29. Rao, S. N. & Sasisekharan, V. (1983) *Int. J. Biol. Macromol.* **5**, 83-88.
30. Shakked, Z. & Kennard, O. (1985) in *Biological Macromolecules and Assemblies, Vol. 2: Nucleic Acids and Interactive Proteins*, Jurnak, F. A. & McPherson, A., Eds., Wiley-Interscience, New York, 1985, pp. 1-36.
31. Haran, T. E., Shakked, Z., Wang, A. H.-J. & Rich, A. (1987) *J. Biomol. Struct. Dynam.* **5**, 199-217.
32. Nelson, H. C. M., Finch, J. T., Luisi, B. F. & Klug, A. (1987) *Nature* **330**, 221-226.
33. Peck, L. J. & Wang, J. C. (1981) *Nature* **292**, 375-379.
34. Kabsch, W., Sander, C. & Trifonov, E. R. (1982) *Nucleic Acids Res.* **10**, 1097-1104.
35. Ulanovsky, L. E. & Trifonov, E. N. (1987) *Nature* **326**, 720-722.
36. Arnott, S., Chandrasekharan, R., Hall, H. & Puigjaner, L. C. (1983) *Nucleic Acids Res.* **11**, 4141-4151.
37. Sarma, M. H., Gupta, G. & Sarma, R. H. (1985) *J. Biomol. Struct. Dynam.* **2**, 1057-1084.
38. Behling, R. W. & Kearns, D. R. (1986) *Biochemistry* **25**, 3335-3346.
39. Shakked, Z. & Rabinovitch, D. (1986) *Progr. Biophys. Mol. Biol.* **47**, 159-195.
40. Bhattacharya, D. & Bansal, M. (1988) *J. Biomol. Struct. Dynam.*, **6**, 93-104.
41. Levitt, M. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 640-644.
42. Nilsson, L. & Karplus, M. (1986) *J. Comp. Chem.* **7**, 591-616.
43. Breslauer, K. J., Frank, R., Blöcker, H. & Marky, L. A. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3746-3750.
44. Saenger, W., Hunter, W. N. & Kennard, O. (1986) *Nature* **324**, 385-388.

Received December 28, 1987

Accepted June 7, 1988