The structure of an oligo(dA)·oligo(dT) tract and its biological implications

Hillary C. M. Nelson, John T. Finch, Bonaventura F. Luisi & Aaron Klug

Medical Research Council Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Poly(dA)·poly(dT) has unusual properties in that it cannot associate into nucleosomes and short, phased runs of it cause DNA bending. The crystal structure of a B-type DNA dodecamer containing a homopolymeric run of six A·T base pairs shows that this region possesses special structural features, including a system of bifurcated hydrogen bonds, which explains some of the properties of this simple homopolymer.

THE homopolymer poly(dA)·poly(dT) has many unusual and distinct structural features that differentiate it from other B-type DNA sequences. The helical repeat in solution was determined to be 10.0 base pairs (bp) per turn, in contrast to 10.5 bp per turn found for random DNA and alternating poly(dA)·poly(dT)[1,2], and the axial rise of the helix is less than that found for other B-type DNA fibres (3.2 versus 3.4 Å)[3]. Fibres of the homopolymer are not affected by the environmental changes such as humidity, cation, or salt concentration, which induce fibres of other sequences to undergo the B- to A-type transition[4]. Neither poly(dA)·poly(dT) nor poly(dG)·poly(dC) can be reassociated into nucleosomes, but the latter case is explained by the rigidity conferred by the three hydrogen bonds per base pair[5,6]. What is it about poly(dA)·poly(dT) that produces the same apparent rigidity, as well as the resistance to environmental changes?

Because of its distinctive properties, poly(dA)·poly(dT) is important in dictating both the translational and rotational positioning of DNA on nucleosomes. Lengths of 80–100 bp are sufficient to be at least partially excluded from nucleosomes[7,8], whereas runs of about 20 bp are preferentially positioned at the ends of the nucleosome core[8]. An instance when the translational positioning of nucleosomes may be of functional significance is found in yeast: oligo(dA)·oligo(dT) tracts act as upstream promoter elements that constitutively activate transcription, perhaps due to the exclusion of nucleosomes from that region of the chromosome[9]. Shorter lengths can be incorporated into nucleosomal DNA, but even then, runs more than 4 or 5 bp long have preferential rotational settings on the nucleosome supercoil at positions of minimum curvature[10,11].

A further example of the use of the poly(dA)·poly(dT) structure in biological recognition and control is the ability of short, repeated runs of the polymer to influence the configuration of long runs of free DNA. Runs of about 5 bp of the homopolymer phased every 10 to 11 bp (every helical repeat) cause an overall bend to the DNA configuration, as seen by abnormally slow gel electrophoretic mobility[12–14], relaxation times in electric dichroism[15,16], and electron microscopy[17]. The relationship between the altered DNA configuration and the repeating sequence pattern of oligo(dA)·oligo(dT) was first observed in trypanosome kinetoplast minicircles[18,19], and has now been shown to occur in eukaryotic and prokaryotic replication origins and in transcriptional regulatory regions[20–23]. The bent configuration is required for activity at some of these sites: for example, SV40 T-antigen and DNA interaction is dependent on the altered DNA configuration[24].

The questions we have addressed are: what is the structure of poly(dA)·poly(dT), and can this structure explain the unusual physical properties and hence its biological function? We have solved by x-ray analysis the crystal structure of a DNA dodecamer whose sequence includes six bps of oligo(dA)·oligo(dT). We show that the oligo(dA)·oligo(dT) tract is essentially straight, the base pairs being parallel to each other and perpendicular to the helix axis. The base pairs do have an unusual structure, due to the high propeller twist (see below) at each A·T base pair. This results in maximal overlap of the bases on each strand and the formation of a run of additional, non-Watson–Crick, cross-strand hydrogen bonds. These distinctive structural features explain what differentiates poly(dA)·poly(dT) from other DNA sequences, why poly(dA)·poly(dT) seems conformationally rigid, and suggest how runs of oligo(dA)·oligo(dT) are involved in DNA bending.

```
C1  G2  C3  A4  A5  A6  A7  A8  A9  G10 C11 G12
|   |   |   |   |   |   |   |   |   |   |   |
G24 C23 G22 T21 T20 T19 T18 T17 T16 C15 G14 C13
```

Fig. 1 Sequence used for crystallization. The base pairs are C1–G24, G2–C23, and so on and the base steps defined as C1, G2–C23, G24. The single strands (dCGCAAAAAAGCG) and (dCGCTTTTTTGCG) were synthesised on an Applied Biosystems DNA synthesizer, and purified individually by column and gel chromatography as an ammonium salt. The strands were then mixed together at equimolar concentrations; more than 99% of the DNA was double-stranded as estimated by autoradiography of a gel-electrophoresed sample. The needle-shaped crystals (average size $0.16 \times 0.22 \times 0.4$ mm) were grown from solutions containing 0.2 mM DNA, 10.0 mM magnesium acetate, 0.5 mM spermine-HCl, and 5% (v/v) 2-methyl-2,4-pentanediol (MPD), vapour diffused at 4 °C against 30 to 45% MPD. X-ray data were collected on a Watts–Hilger four-circle diffractometer for three crystals at 4 °C. Over 5,000 reflections were collected in the resolution range of 20–2.5 Å, and the intensities corrected for Lorentz polarization, absorption and time-dependent decay. Symmetry-equivalent reflections were averaged; and the data from the three crystals were merged with an R-merge in intensity of 10%. Of the final 2,572 unique reflections in the 20–2.5 Å resolution range, 2,093 reflections had amplitudes greater than 0.5σ; data below 0.5σ were excluded from the refinement. In the 3–2.5 Å resolution range, two thirds of the reflections were used in the final data set.

Structure determination and refinement

The sequence used is shown in Fig. 1, and the crystallization methods are described in the legend. The crystals diffract to 2.5 Å, but only weakly beyond 3 Å, and have unit cell dimensions of $a = 25.4$, $b = 40.7$, $c = 65.8$ Å, with space group $P2_12_12_1$. Our approach was to use a sequence related to the self-complementary CGCGAATTCGCG dodecamer, whose structure had already been solved[25,26]. This was justified by the similarities in the patterns of X-ray intensities, indicating the same crystal packing. It made possible a solution of the structure by using the sugar–phosphate chains of the CGCGAATTCGCG dodecamer as a starting model and manually substituting our sequence for that of the other dodecamer. However, it also meant that much care was needed to efface the original structure during the refinement (details in legend to Fig. 2).

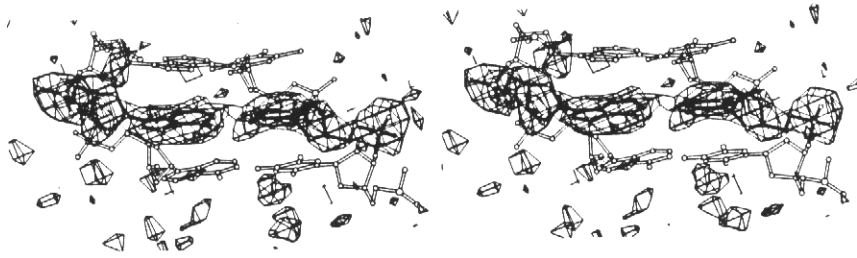


Fig. 2 Stereo view of 'partial' difference map for base pair A6–T7 in the final refined structure. This map is of the kind used in the refinement. They are generated by omitting one nucleotide pair from the calculation of structure factors, and using these phases to calculate a difference $|F_{obs}|-|F_{calc}|$ map. The density at the position of the omitted nucleotide pair is then entirely due to its contribution to the X-ray amplitudes. The map shown is contoured at 0.17 e/Å^3, twice the value of the r.m.s. density. Because our sequence is not self-complementary, the structure is polar and can be introduced into the unit cell either way up. The two possible models, differing i polarity, were refined against the X-ray data, using the rigid body refinement program CORELS (ref. 41). At first, the entire model was defined a one rigid body, varying the resolution in successively higher resolution shells until reaching the 10–3 Å range. At this point, the individual nucleotid were treated as rigid bodies, and then after R-factor convergence, the constituent sugars, bases and phosphate groups were allowed position freedom. The resolution shell was extended to the 8–2.5 Å range, and the refinement continued. The $2|F_{obs}|-|F_{calc}|$ and $|F_{obs}|-|F_{calc}|$ electron densi maps were examined for satisfactory convergence at each stage of the refinement, using an Evans and Sutherland graphics machine with the progra FRODO (ref. 42) (as modified by P. Evans). When the R-factor fell from 40% to about 30%, the two models were compared. One model consistent had a R-factor that was lower by 1–2% than the model in the opposite orientation. A comparison of these partial difference maps, generated described above, showed that with the better choice of orientation, the purine-rich and pyrimidine-rich strands could be discriminated, where this was not so in the opposite orientation. The refinement was continued with only the better model. In order to minimize bias in the phases fro the starting model, extensive use was made of partial maps and recalculation of phases. Partial difference maps were generated omitting each ba pair in turn and these were used as a guide, along with difference maps calculated with all atoms present, to adjust by computer graphics a change in the position of the base pairs. After two rounds of this procedure, with the phases regenerated after each step, the model was refin using the Hendrickson–Konnert refinement program NUCLSQ, in which positions and isotropic temperature factors of individual atoms are refin by least squares, subject to a set of distance and general geometry restraints[43,44]. No restraints were applied to hydrogen bond distances for t central eight base pairs or to the sugar-phosphate torsion angles. Four such cycles of generating partial maps, fitting the base pairs into the electr density, and refining the fitted structure were carried out. After the first cycle, solvent molecules were located where stereochemically feasible the peaks were present in the $|F_{obs}|-|F_{calc}|$ and $2|F_{obs}|-|F_{calc}|$ maps before and after density fitting and phase recalculations. All peaks were assign to water molecules on the basis of their refined temperature factors and proximity to polar groups; however, at this resolution, we cannot distingui between oxygen and other light ions. A total of 27 water molecules were included in the final model. There were no unassigned peaks in t $|F_{obs}|-|F_{calc}|$ map that were greater than 0.23 e/Å^3, which is four times the value of the r.m.s. density.

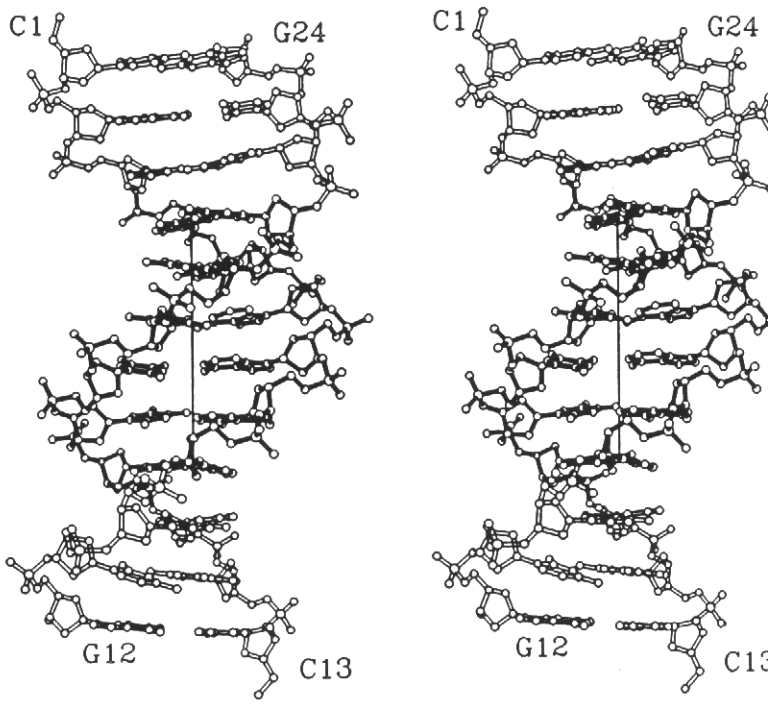


Fig. 3 Stereo view of the structure, the A·T base pairs being shown with black bonds, and the G·C base pairs with white bonds. The best helix axis for the oligo(dA)·oligo(dT) tract was determined using the program HELIX (refs 27, 29), and is shown in the figure. The helical repeat averages 10.0 bp per turn. Although the helical structure of this region is similar to B-type DNA, there are several differences from the standard fibre structure which are a consequence of its high propeller twisted structure. These differences include the decreased axial rise (3.1 ± 0.1 Å compared with 3.4 Å for fibre B-DNA) and a decreased minor groove width, as narrow as 9.1 Å at the centre of the AT region, compared with 12 Å for fibre B-DNA. Note that the major groove width is only slightly narrower than the average B-DNA values (17.2 Å at the narrowest versus an average of 17.5 Å for fibre B-DNA). As has been found in other DNA crystal studies, the sugar–phosphate torsion angles can vary greatly, suggesting an inherent flexibility that will accommodate structural changes[27–30]. The distribution of angles of the structure presented here fall within the range observed for other DNA crystals. It is interesting to note that all the average torsion angles on the thymine strand are somewhat different and vary more widely than those on the adenine strand. Although the AT region shows some heteronomity, the differences are not as extreme as that modelled for the heteronomous fibre form of poly(dA)·poly(dT) (ref. 45). The average δ torsion angles, defined as C5′–C4′–C3′–O3′, are 128° (±16) and 116° (±29) respectively for the adenine and thymine strands; this suggests sugar puckers around C1′-exo and C2′-endo, similar to what has been found for other B-t crystal structures. The χ torsion angles, defined as O4′–C1′–N1–C2 for pyrimidines and as O4′–C1′–N9–C4 for purines, are −101° (±8) and − (±20) respectively for the adenine and thymine strands, showing that the bases have similar orientations around their glycosidic bonds.

Table 1 Helical parameters of a DNA dodecamer with a run of six A·T base pairs

Base step	Roll (deg.)	Tilt (deg.)	Slide (Å)	Twist (deg.)	Rise (Å)	Pr tw (deg.)	Buckle (deg.)
C1-G24						19.0	2.4
G2-C23	−1.1	−2.2	−0.2	42.2	2.9	12.3	−1.1
C3-G22	−2.8	3.0	−0.4	38.1	3.0	7.7	−1.1
A4-T21	10.3	1.7	−0.9	29.4	4.2	14.6	0.4
A5-T20	1.0	−1.2	0.1	35.6	3.2	23.0	−2.0
A6-T19	−0.8	−0.7	0.1	35.4	3.2	25.5	−0.8
A7-T18	1.5	0.8	0.7	38.5	2.9	22.5	5.5
A8-T17	−0.1	0.2	−0.2	33.2	3.2	17.7	−2.9
A9-T16	−5.1	−1.9	0.2	38.0	3.0	19.2	5.0
G10-C15	3.9	0.2	−0.9	34.6	3.9	10.8	−3.7
C11-G14	−10.0	−4.1	−0.7	36.6	3.1	15.2	−5.2
G12-C13	8.8	2.6	−0.6	42.0	3.6	6.2	−0.1

Pr tw, propeller twist. The helical parameters are calculated using the programs BROLL and CYLIN (refs 27, 29). Because the two bases of a base pair are often non-coplanar, the mean plane of the base pair is used when appropriate for the calculations. Roll, tilt and slide are defined here for one base pair relative to a successive base pair, so they refer to the geometry of the base step. Roll is the rotation of the long axis of a base pair in the direction of the major or minor groove (by convention positive when rotating away from the minor groove). Tilt is the rotation around the short axis towards either sugar phosphate backbone (by convention here positive if rotating towards the adenine strand). Slide is the relative displacement of the C6-C8 long axis of one base pair with respect to the successive one (by convention positive when towards the major groove). Twist and rise are defined with respect to the helix axis, although they are relatively insensitive to the choice of a best-helix axis. Twist is the rotation between successive base pairs about the best-helix axis. Rise is the average axial distance between C1′ atoms on successive base pairs. Propeller twist and buckle are defined for each base pair as described in the text. By convention, propeller twist is positive for the counter-clockwise rotation of the distant base when viewed from the near base along the sugar-sugar vector. By convention here, buckle is positive when opening towards the 3′ end of the adenine strand.

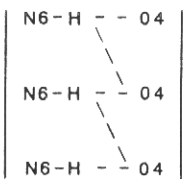
The final model, shown in Fig. 3, has an R factor of 20% for the 2,003 reflections greater than 0.5 σ in intensity in the 8 to 2.5 Å resolution range. The r.m.s. deviation in atomic positions between the starting model and the final model is 2.3 Å, and indicates that the starting model had been effectively effaced. The thermal parameters are within observed ranges for other DNA crystal structures, implying that there is no extra disorder within the crystals[27,28]. Indeed, the average temperature factors for the adenines and thymines were actually lower by several Å^2 than for the guanines and cytosines. The individual isotropic temperature factors range from 13 to 57 Å^2, with averages over atoms within groups of 39 for the phosphate, 33 for the sugars, and 25 for the bases. The solvents had individual isotropic temperature factors ranging from 28 to 71, with an average of 49 Å^2.

The crystal packing is similar to that found for the native dodecamer, brominated derivatives, and mismatch variants, in which the minor grooves of symmetry-related molecules in the unit cell overlap by three base pairs at their termini[27-29]. There are strong contacts between the bases of G2 and G12 in symmetry-related molecules, and weaker contacts between the bases of G14 and G24. The other contacts involve sugar-phosphate atoms, and presumably have a minor effect in determining base position, given the inherent flexibility of the sugar-phosphate chain. Thus, the last two base pairs are the most affected by crystal packing forces, whereas the oligo(dA)·oligo(dT) region is relatively unaffected by crystal packing forces.

Propeller twist

One of the most obvious features of the oligo(dA)·oligo(dT) region of the structure is that the bases within a base pair are non-coplanar, as seen in Figs 3 and 4. This feature was recognized in the electron density maps in the early stages of refinement; therefore specific Watson-Crick hydrogen bond restraints were not employed in this region in refining the structure. Despite their omission from the refinement, they emerged in the final model with average distances well within the standard values of 2.7 to 3.0 Å (see legend to Fig. 4). The non-coplanarity has two principal components: 'propeller twist' and 'buckle'[27,29]. Propeller twist is defined as the rotation of the bases along their longitudinal axis, and buckle is the dihedral angle of the bases along their short axis after the propeller twist has been flattened out by rotating to zero. As seen in Table 1, the buckle is variable, ranging from −2.9 to 5.5°, but averages to an almost planar 0.9°. The propeller twist is large, ranging from 15 to 25° with an average of 20°, even higher than the 17° propeller twist seen for the AT base pairs in the CGCGAATTCGCG structure[30].

The high degree of propeller twist has two effects on the structure of oligo(dA)·oligo(dT), both of which improve the overall stability of the helix: it helps maximize purine-purine stacking interactions and creates a potential system of additional hydrogen bonds. The propeller twist causes the major groove side of each base to point towards the 3′ end of its strand. As seen in Fig. 4, this pushes the N6 atom of adenine, which is in the major groove, towards the O4 atom of thymine on the 3′ side of it, making feasible a non-Watson-Crick hydrogen bond diagonally across the major groove, in addition to the normal Watson-Crick hydrogen bonds. The same proton from the amino group would thus be close to two hydrogen bond acceptor atoms (carbonyl oxygens), therefore allowing the formation of a three-centred bifurcated hydrogen bond[31,32]. As carbonyl oxygens can receive two hydrogen bonds as well, the overall effect on the oligo(dA)·oligo(dT) region is to produce a zig-zag system of hydrogen bonds down the major groove.



It would also seem to follow that a minimum of three A·T base pairs is needed to stabilize the interaction, by allowing the middle base pair to make two of these three-centred bifurcated hydrogen bonds. The cross-strand diagonal hydrogen bonds are long, but the geometry of the interaction, with the combination of long and short bonds, is typical for these types of bifurcated hydrogen bonds, which have been located in proteins[33], as well as in tRNA (ref. 34). In DNA, from geometrical and chemical restraints, a system of cross-strand diagonal hydrogen bonds in the major groove can form only in certain homopolymer stretches of B-type DNA, where the hydrogen bonds of the base pair straddle the helix axis. Isolated instances of cross-strand diagonal hydrogen bonds have been found in the minor groove at the G·A mismatch positions of a decamer structure[35], and can be found in the major groove at the AA steps of some of the brominated derivatives of the CGCGAATTCGCG dodecamer.

Base-stacking, specifically purine-purine stacking, is a very significant force in stabilizing homopolymer structures, and poly(dA)·poly(dT) maximizes this through the high degree of propeller twist which rotates the bases around their longitudinal axis, allowing more interactions between neighbouring bases. As seen in Fig. 5, the adenines stack so as to overlap their six-membered rings and the thymines stack such that the C5-methyl group makes weak carbon-carbon van der Waals' contacts with C6 atoms of the pyrimidine ring below. This, together with the additional hydrogen bonding system, causes an increase in the helical twist and a decrease in the helical rise (see Table 1). We cannot estimate the relative contributions of the base stacking and additional hydrogen bonds to the overall stability

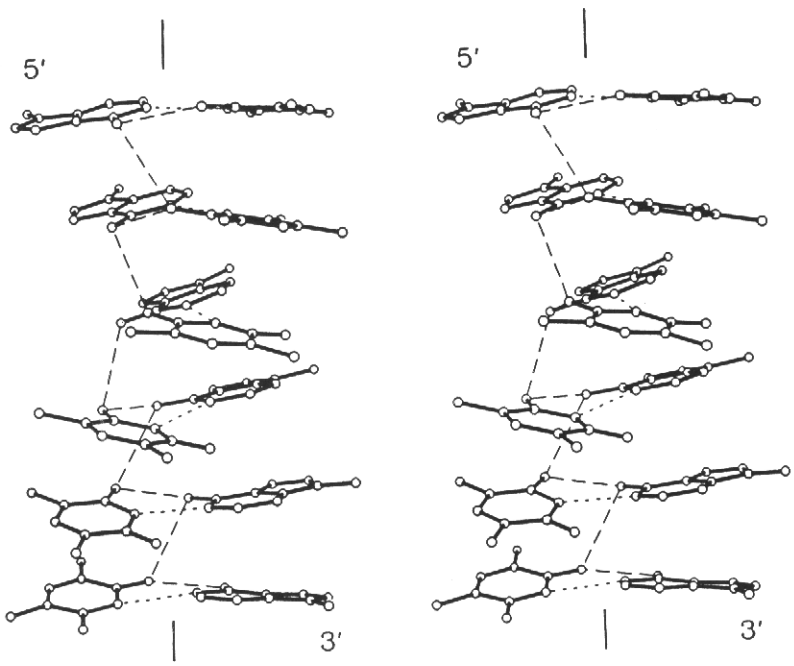


Fig. 4 The oligo(dA)·oligo(dT) stretch, illust bases and best helix axis. The 5′ and 3′ ends of t strand are labelled accordingly. The Watson- adenine/N3-thymine distances are: A4 to T21, to T20, 2.8 Å; A6 to T19, 2.7 Å; A7 to T18, 2. T17, 2.7 Å; and A9 to T16, 2.7 Å. The Watson- adenine/O4-thymine distances are: A4 to T21, 2 T20, 3.1 Å; A6 to T19, 3.1 Å; A7 to T18, 2.7 Å; 2.8 Å; and A9 to T16, 2.7 Å. The cross-stran N6-adenine/O4-thymine distances are: A4 to A5 to T19, 3.1 Å; A6 to T18, 3.4 Å; A7 to T17, A8 to T16, 3.3 Å. Although the positions of th atoms cannot be located at the resolution of study, when they are put into idealised positions ing cross-strand diagonal bonds fit the criteria component of the bifurcated hydrogen bonds a by Taylor *et al.*[32]. At the present time, we have tion for the one longish bond in the middle (A ought to be remembered that the energy of th bond, being largely electrostatic in nature, does sharply with distance. Indeed, energy calcul predicted cross-strand diagonal interactions (M sonal communication).

of the structure. However, the effect of this compact stacking arrangement on the sugar-phosphate chains is to narrow the width of the minor groove (see Fig. 3). If measured by the distance between phosphorus atoms on opposite strands staggered by three base pairs, the minor groove in the A·T tract is on average 9.5 ± 0.4 Å wide.

It has been suggested that a high degree of propeller twist (and consequent narrowing of the minor groove) might be stabilized by a minor groove spine of hydration, whereby solvent molecules are highly ordered by tetrahedral coordination between cross-strand pyrimidine O2 and purine N3 atoms one base pair apart[27,29]. Although we find solvents in the major groove and near the sugar-phosphate chains, we have not located ordered solvent atoms in the minor groove. But, owing to the limited resolution of this study, in which only 27 solvents have been unambiguously located, we cannot comment definitively on the state of minor groove solvation.

Roll, tilt and junctions

The path of a helix can be described by calculating the orientation of the best mean plane of the base pairs, either with respect to a successive base pair (defining the orientation of a base step), or with respect to a helix axis. The latter calculation is heavily dependent on the choice of the best helix axis, which can be particularly difficult if the path of the helix is changing. Table 1 shows the helical parameters describing the orientation of the base pairs; the definitions of these parameters are in the table legend. With the exception of the last base step (A8, A9-T16, T17), discussed in the legend to Fig. 6, the roll and tilt angles for the AA-TT base steps are close to zero. The extreme variation of the base steps at the ends of the structur minimal variation within the AA-TT base steps striking. The near-zero values for the roll, tilt and s AA-TT base steps indicate that each A·T base changing position with respect to its neighbours, the being parallel to each other.

Because of the near-uniform geometry of the A best-helix axis was calculated for this region alone a a reference to examine the path of the rest of the h 6 shows the direction of the normals to the base pa as projections of vectors relative to a common features about the direction of the helix are obviou diagram. First, the projection of vectors for the A·T (especially for the vectors 4 to 8, corresponding to A·T base pairs) are short and clustered together abou indicating that the base pairs are not only parallel to but perpendicular to the helix axis. Second, the over of the vectors changes from right to left in the diagram an overall change in the helix direction. These ch most abruptly at two places: between vector 3 and v is, between base pairs C3-G22 and A4-T21) and bet 10 and vector 11 (between base pairs G10-C15 and Thus, the overall picture from the vector diagram bit of oligo(dA)·oligo(dT) with two junctions hav effects on the change in the helix direction.

The nature of these particular junctions are wort in light of the junction-based theories for DNA be can be explained simply as a combination of base- propeller twist. In B-DNA, at a 5′-pyrimidine-pu step, the propeller twist and larger size of purine

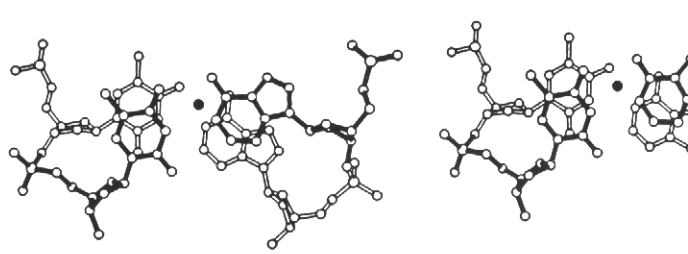


Fig. 5 Stereo views of a typical AA-TT base step. The two base pairs, A5-T20 (in black) and A6-T19 (in white) are viewed down the helix axis (illustrated by a black circle). Compare this base stacking arrangement to that found in poly(dG)·poly(dC), where the guanines stack in a staggered fashion (six-membered ring over five-membered ring), and the cytosines show no overlap[46].

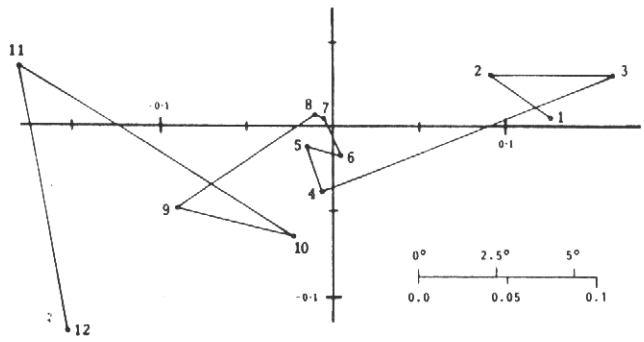


Fig. 6 Polar diagram of normals to the mean plane of the base pairs, projected onto a plane perpendicular to the best helix axis of the AT region. Each point specifies the end of a unit vector drawn perpendicular to the mean plane of the base pair, all vectors being drawn from the same origin. Vector 1 is the normal to the base pair C1-G24; vector 2 is the normal to the base pair G2-C23, and so on. The length of each vector from the origin is the arc sin of the angle between the mean plane of each base pair and the perpendicular to the best helix axis[47]. The distance between any two vector end points is the arc sin of the angle between the mean planes through each base pair (thus, an angle related to a combination of the roll and tilt angles of the base steps). A similar plot constructed for the CGCGAATTCGCG dodecamer shows a similar relationship between the points at the end of the structure; this is due to the similarity in the crystal packing. But the distribution of the vectors in the central region is quite different, indicating a difference in the distribution of bending. The overall bend in the CGCGAATTCGCG dodecamer is delocalized over several base steps, as opposed to the abrupt changes seen for our structure. Although vector 9 (the last A·T base pair, A9-T16) is much larger and separate from the other AT vectors, vector 10 (the G·C base pair at the 3′ end of the adenine stack) is stacked more with the rest of the A·T base pair. This presumably is an end-effect arising because of the conflicting demands of the G·C base pair trying to maximize purine-purine stacking on the A·T base pair, yet unable to undergo the same degree of high propeller twist as the A·T base pairs.

steric clashes between the cross-strand purine bases in the minor groove[36,37]. When the second base pair has a larger degree of propeller twist (effectively pushing the minor groove towards the 5′ side), the normal pyrimidine-purine minor groove steric clash is exaggerated by the difference in the propeller twist causing a larger than normal roll away from the minor groove at that step, as seen by the 10° roll at the C3, A4-T21, G22 base step. At a 5′-purine-pyrimidine-3′ step, the purines now clash in the major groove. Again, the effects of the differential propeller twist will affect this. The situation at the 3′ end of the adenine stack of the structure is made more complicated by the fact that the junction is not a purine-pyrimidine. If, however, one assumes the G10-C15 base pair to be an honorary A·T base pair (as described in the legend to Fig. 6), then the purine-pyrimidine clash/propeller twist clash occurs in the major groove of the G10, C11-G14, C15 base step. Again, there is a large roll angle, but it is in the opposite direction (−10°).

Consequences of structure

The fact that the properties of poly(dA)·poly(dT) in solution and in fibres show a conformational rigidity strongly indicates that the crystal structure of the oligo(dA)·oligo(dT) tract can validly represent a structure of poly(dA)·poly(dT) (see Fig. 7). Indeed, the distinctive features which have emerged (additional hydrogen bonds and good base stacking) are such as would confer a rigidity over and above that expected from base pairs with two hydrogen bonds. It now becomes obvious why the homopolymer will not wrap into nucleosome cores, and why short runs take up positions of minimum curvature: the structure resists bending because it would destroy the good base-stacking and additional hydrogen bonds. The decreased helical repeat, reduced axial rise, and narrow minor groove in relation to standard B-type DNA are just natural consequences of the compact base-stacking arrangement.

How does this structure relate to the bending of DNA by phased runs of oligo(dA)·oligo(dT)? There must be something else about the poly(dA)·poly(dT) structure other than resistance to bending, or else permanent bending could be accomplished by runs of oligo(dG)·oligo(dC). The perpendicularity of the mean plane of the A·T base pairs to the helix axis, with roll and tilt angles near zero, as well as the high degree of propeller twist within the base pairs, must be significant.

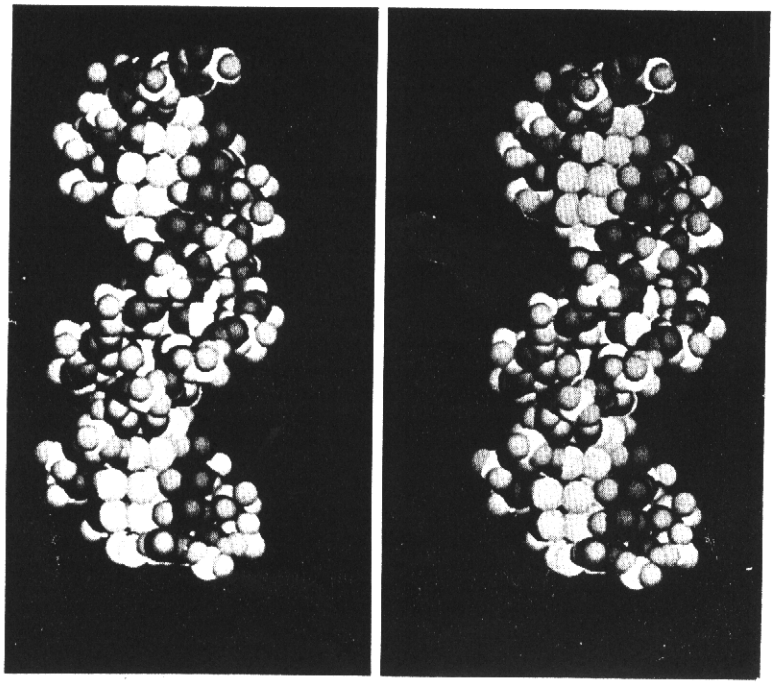


Fig. 7 Stereo view of a CPK model of the homopolymer poly(dA)·poly(dT), created by extension of the double helix from the middle four A·T base pairs. The high degree of propeller twist causes the double helix to have a deep, narrow minor groove and a wide, shallow major groove. This contrasts with the model for homopolymer poly(dG)·poly(dC), built by extension from the crystal structure of the GGGGCCCC octamer, which has a wide, shallow minor groove due to the displacement of the G·C base pairs into the major groove to increase purine-purine stacking[46]. Although the models from the recent solution and fibre studies of poly(dA)·poly(dT) agree with the crystal structure in terms of the gross features of narrow minor groove and non-coplanarity of the base pairs, they disagree in terms of the deconvolution of the non-coplanarity into propeller twist, as well as an absolute tilt of about 6° of the base pair with respect to the helix axis[48-50]. This base pair tilt is not seen in the crystal structure; the average mean plane through an AT base pair is perpendicular to a best helix axis, although the base pairs do exhibit a variable degree of buckle. It is not surprising that these models differ from the crystal structure in this respect due to the inherent difficulties involved in fitting models to cylindrically averaged data at relatively low resolution.

In the literature, there have been several theories of bending which differ in their predictions as to where the actual changes in the direction of the helix axis occur. The wedge model in its latest form proposes that the AT stretch itself is bent due to a combination of roll and tilt at each and every AA-TT step. This large, non-coplanar wedge, estimated to be around 9°, would cause a change in the helix direction with respect to adjoining pieces of DNA (ref. 38). We have shown that the AA-TT steps do not have roll or tilt components, and hence the oligo(dA)·oligo(dT) stretch cannot cause a change in orientation of this sort. This rules out the wedge model and implies that any bending of the helix axis must occur in the regions between the runs of oligo(dA)·oligo(dT).

The junction model suggests that the bending occurs abruptly at the ends of the AT stretch due to some specialised features of that region[12,19]. In fact, we suggest that the larger-than-usual roll angles at the base steps at the ends of the oligo(dA)·oligo(dT) occur partially as a result of the high degree of propeller twist from the structure. Much of the contribution to the overall bending could come from the junctions if, as here, the roll angles at the ends of the A·T stretch had opposite directions with respect to the minor and major grooves, thus causing an additive effect when placed half a helical turn apart. There is a small tilt component to the bending (as seen in Table 1), but it is not as dramatic as the roll component.

We note, however, that it is not in general necessary to have all of the change concentrated at the junctions; it could be more widely distributed over the regions between the AT tract. To give a permament bend, there need only be a net roll accumulated in the intervening regions, as follows from the considerations given by Calladine and Drew[39]. This can be understood most simply by considering two successive half turns of a uniform double helix, which has a fixed roll angle between base pairs (as in Fig. 2 of ref. 40). After one full helical turn, the bas pairs will be parallel to each other. However, each half turn ha a net roll angle between its first and last base pairs (obtaine by vectorial summation of the individual contributions). Becaus of the helical twist between the two halves, they have resultan of opposite sign and so cancel when adjacent; therefore, th helix axis remains unchanged. But if one half turn is replace by an AT stretch of zero net roll, then the net roll contribute by the remaining half is not cancelled, and the direction of th helix axis will change.

Without further direct studies of structures including th intervening regions, it is difficult to decide between an abru change at the junctions, a gradual change over the interveni region, or a combination of the two. Whatever the case, it is t contrast between the special geometry of t oligo(dA)·oligo(dT) tract and that of other DNA sequences th is the basis of the bending.

We thank Horace Drew, Bill Hunter, Michael Levitt, Maxi McCall and Daniela Rhodes for advice and support throughc the whole of this project. In addition, we would like to acknov edge: Terry Smith, for synthesis of oligonucleotides; Dan Brow for help with purification; Phil Evans and Judd Fermi, for he with refinement; and Chris Calladine and Andrew Travers f comments on the manuscript. This paper is dedicated to M Perutz in honour of the 40th anniversary of the MRC Laborato of Molecular Biology. H.C.M.N. is a Fellow of The Jane Cof Childs Memorial Fund for Medical Research. This investigati has been aided by a grant from the Jane Coffin Childs Memor Fund for Medical Research. This work was presented at t Gordon Research Conference on Nucleic Acids on 15 June 19

Note added in proof: Bifurcated hydrogen bonds have now be recognized in the dodecamer d (CGCAAATTTGCG) cryst lized with and without the drug distamycin[51].

Received 30 August; accepted 19 October 1987.

1. Peck, L. J. & Wang, J. C. *Nature* **292,** 375-378 (1981).
2. Rhodes, D. & Klug, A. *Nature* **292,** 378-380 (1981).
3. Arnott, S. & Selsing, E. *J. molec. Biol.* **88,** 509-521 (1974).
4. Leslie, A. G. W., Arnott, S., Chandrasekaran, R. & Ratliff, R. L. *J. molec. Biol.* **143,** 49-72 (1980).
5. Rhodes, D. *Nucleic Acids Res.* **6,** 1805-1816 (1979).
6. Simpson, R. T. & Kunzler, P. *Nucleic Acids Res.* **6,** 1387-1415 (1979).
7. Kunkel, G. R. & Martinson, H. G. *Nucleic Acids Res.* **9,** 6869-6888 (1981).
8. Prunell, A. *EMBO J.* **1,** 173-179 (1982).
9. Struhl, K. *Proc. natn. Acad. Sci. U.S.A.* **82,** 8419-8432 (1985).
10. Satchwell, S. C., Drew, H. R. & Travers, A. A. *J. molec. Biol.* **191,** 659-675 (1986).
11. Travers, A. & Klug, A. *Phil. Trans. R. Soc.* (in the press).
12. Koo, H.-S., Wu, H.-M. & Crothers, D. M. *Nature* **320,** 501-506 (1986).
13. Hagerman, P. J. *Nature* **321,** 449-450 (1986).
14. Dickmann, S. *Nucleic Acids Res.* **15,** 247-265 (1987).
15. Hagerman, P. J. *Proc. natn. Acad. Sci. U.S.A.* **81,** 4632-4636 (1984).
16. Levene, S. D., Wu, H.-M. & Crothers, D. M. *Biochemistry* **25,** 3988-3995 (1986).
17. Griffith, J., Bleyman, M., Rauch, C. A., Kitchin, P. A. & Englund, P. T. *Cell* **46,** 717-724 (1986).
18. Marini, J. C., Levene, S. D., Crothers, D. M. & Englund, P. T. *Proc. natn. Acad. Sci. U.S.A.* **79,** 7664-7668 (1982).
19. Wu, H.-M. & Crothers, D. M. *Nature* **308,** 509-513 (1984).
20. Bossi, L. & Smith, D. M. *Cell* **39,** 643-652 (1984).
21. Snyder, M., Buchman, A. R. & David, R. W. *Nature* **324,** 87-89 (1986).
22. Zahn, K. & Blattner, F. R. *Science* **236,** 416-427 (1987).
23. Stenzel, T. T., Patel, P. & Bastia, D. *Cell* **49,** 709-717 (1987).
24. Ryder, K., Silver, S., DeLucia, A. L., Fanning, E. & Tegtmeyer, P. *Cell* **44,** 719-725 (1986).
25. Wing, R. *et al. Nature* **287,** 755-758 (1980).
26. Drew, H. R. *et al. Proc. natn. Acad. Sci. U.S.A.* **78,** 2178-2183 (1981).
27. Dickerson, R. E., Kopka, M. L. & Pjura, P. in *Biological Macromolecules and Assemb Nucleic Acids and Interactive Proteins* (eds Jurnak, F. A. & McPherson, A.) **2,** 38 (Wiley, 1984).
28. Hunter, W. N. *et al. J. biol. Chem.* **262,** 9962-9970 (1987).
29. Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. *J. biol. Chem.* **257,** 14686-1
30. Dickerson, R. E. & Drew, H. R. *J. molec. Biol.* **149,** 761-786 (1981).
31. Donohue, J. in *Structural Chemistry and Molecular Biology* (eds Rich, A. & Davidson **1,** 450-456 (Freeman, San Francisco, 1968).
32. Taylor, R., Kennard, O. & Versichel, W. *J. Am. chem. Soc.* **106,** 244-248 (1984).
33. Baker, E. N. & Hubbard, R. E. *Prog. Biophys. molec. Biol.* **44,** 98-179 (1984).
34. Jack, A., Ladner, J. E. & Klug, A. *J. molec. Biol.* **108,** 619-649 (1976).
35. Prive, G. G. *et al. Science* **238,** 498-504 (1987).
36. Calladine, C. R. *J. molec. Biol.* **161,** 343-352 (1982).
37. Dickerson, R. E. *J. molec. Biol.* **166,** 419-441 (1983).
38. Ulanovsky, L. E. & Trifonov, E. N. *Nature* **326,** 720-722 (1987).
39. Calladine, C. R. & Drew, H. R. *J. molec. Biol* **192,** 907-918 (1986).
40. Calladine, C. R. & Drew, H. R. *J. molec. Biol.* **178,** 773-782 (1984).
41. Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S. H. *Acta crystallogr.* sect. 800-804 (1977).
42. Jones, T. A. *J. appl. Crystallogr.* **11,** 268-272 (1978).
43. Hendrickson, W. A. & Konnert, J. H. in *Biomolecular Structure and Conformation, Fur and Evolution* (ed. Srinivasan, R.) **1,** 43-57 (Pergamon, Oxford, 1980).
44. Westhof, E., Dumas, P. & Moras, D. *J. molec. Biol.* **184,** 119-145 (1985).
45. Arnott, S., Chandrasekaran, R., Hall, I. H. & Puigjaner, L. C. *Nucleic Acids Re* 4141-4155 (1983).
46. McCall, M., Brown, T. & Kennard, O. *J. molec. Biol.* **183,** 385-396 (1985).
47. Dickerson, R. E., Kopka, M. L. & Pjura, P. *Proc. natn. Acad. Sci. U.S.A.* **80,** 7099-7103 (1
48. Edmondson, S. P. & Johnson, W. C. Jr *Biopolymers* **24,** 825-843 (1985).
49. Alexeev, D. G., Lipanov, A. A. & Skuratovskii, I. Y. *Nature* **325,** 821-823 (1987).
50. Park, H.-S., Arnott, S., Chandrasekaran, R., Millane, R. P. & Campagnari, F. *J. molec* **197,** 513-523 (1987).
51. Coll, M., Frederick, C. A., Wang, A. H.-J. & Rich, A. *Proc. natn. Acad. Sci. U.S.A.* (i press).