Chemical Determinants of DNA Bending at Adenine-Thymine Tracts†

Hyeon-Sook Koo and Donald M. Crothers*

Department of Chemistry, Yale University, New Haven, Connecticut 06511

Received October 24, 1986; Revised Manuscript Received February 9, 1987

ABSTRACT: DNA fragments having homopolymeric adenine-thymine tracts phased with the helix screw are known to be bent. According to our working model, adenine-thymine tracts adopt a polymorphic structure (H-DNA), and juxtaposition of H-DNA with B-DNA results in bending at the junction between the two structures. We incorporated different base analogues in addition to the four ordinary bases into oligonucleotides; ligated multimers of oligonucleotide duplexes were run on polyacrylamide gels. By comparison of gel mobility data for different sequences, we identified factors both necessary and irrelevant for bending, corresponding to the formation of H-DNA. The 5-methyl group on pyrimidines is not essential, and the 2-amino group on purines interferes with the formation of H-DNA, either because it provides a third H bond between the bases or because it alters water structure in the minor groove. The strong base stacking of A may be an important contributing factor to stabilization of the anomalous DNA structure responsible for bending.

DNA bending by homopolymeric adenine-thymine tracts has been confirmed by a variety of experiments on DNA fragments containing tracts of appropriate length and phasing. Such fragments show slower mobilities on polyacrylamide gels (Marini et al., 1982; Wu & Crothers, 1984; Diekmann, 1986; Koo et al., 1986), faster rotational relaxation times (Marini et al., 1982; Hagerman, 1984; Levene et al., 1986), and higher circularization probabilities (Ulanovsky et al., 1986) than expected. In addition, circular forms have been observed in the electron microscope (Griffith et al., 1986). A·T tracts are found in such biologically important sequences as the upstream scaffold-attached regions of several *Drosophila* genes (Gasser & Laemmli, 1986), yeast centromeres (Bloom et al., 1984), and the replication origin of bacteriophage λ (Zahn & Blattner, 1985). Also, an A_6 tract between two directly repeated T_{Ag} binding sites at the SV40 origin I is critical for optimum binding of T_{Ag} (Ryder et al., 1986). In our previous paper, we proposed that bending occurs at the junction between A tracts and adjacent B-DNA owing to the polymorphic structure of A tracts (Koo et al., 1986). In order to adopt the polymorphic form [heteronomous DNA (H-DNA)], A tracts need to be at least 4 base pairs (bp) long, and their repetition in phase with the helix screw optimizes macroscopic bending.

Here we seek to find the features in A·T base pairs that drive A tracts into the H-form, whose juxtaposition with B-DNA results in bending, according to our model. We approached the problem by the same method as described in our previous paper (Koo et al., 1986), but we incorporated different base analogues such as uridine, 5-methylcytidine, and inosine, in addition to the four ordinary bases, in oligonucleotides. Multimers of 10 bp long synthetic oligonucleotide duplexes were run on polyacrylamide gels, and the gel mobilities of the multimers were compared with those of size-marker DNA. Using the resulting gel mobility data collected for different sequences, we identified factors both necessary and irrelevant for the formation of H-DNA.

†Supported by Grant DMB-8405494 from the National Science Foundation and by Project Grant for Parasitology and Tropical Medicine from the MacArthur Foundation. The DNA synthesizer was purchased with the help of NSF Grant PCM 8413761.

Table I: Sequences Examined[a]

name	sequence
A_5N_5	(5′)GGCAAAAACG(3′) GTTTTTGCCC
U_5N_5	AAAAA UUUUU
I_5N_5	IIIII CCCCC
iG·C	AAGAA TTCTT
iG·C^{5m}	AAG AA TTC^{5m}TT
iI·C	AAIAA TTCTT
iI·A	AAIAA TTATT
iG·A	AAGAA TTATT
iI·T	AAIAA TTTTT

[a] All the sequences have the same flanking sequences as shown around $(A·T)_5$ in A_5N_5 and thus are 10 bp long. i(G·C) is the same sequence as IAG in the previous paper (Koo et al., 1986); the symbol "i" means interrupting A tracts in the middle.

Materials and Methods

Synthesis and Purification of Oligonucleotides. Phosphoramidite reagents of the ordinary four bases and base analogues such as uridine, 5-methylcytidine, and inosine were purchased from Applied Biosystems. The oligonucleotides in Table I were made on a DNA synthesizer (Applied Biosystems) and purified as described previously (Koo et al., 1986).

Kinasing, Ligation, and Electrophoresis. The oligonucleotides, each hybridized to its complementary strand, were kinased with [γ-^{32}P]ATP, and the hybridized duplexes were self-ligated to make multimers of the oligonucleotide (Koo et al., 1986). The ligated products were run on 8% polyacrylamide gels at either room temperature or 4 °C. Ligated products of 10 bp *Bam*HI linkers (New England Biolabs) were used as size standards. The apparent length of each multimer is defined as the length of the marker DNA having the same mobility. The ratio R_L of apparent length to real length

0006-2960/87/0426-3745$01.50/0

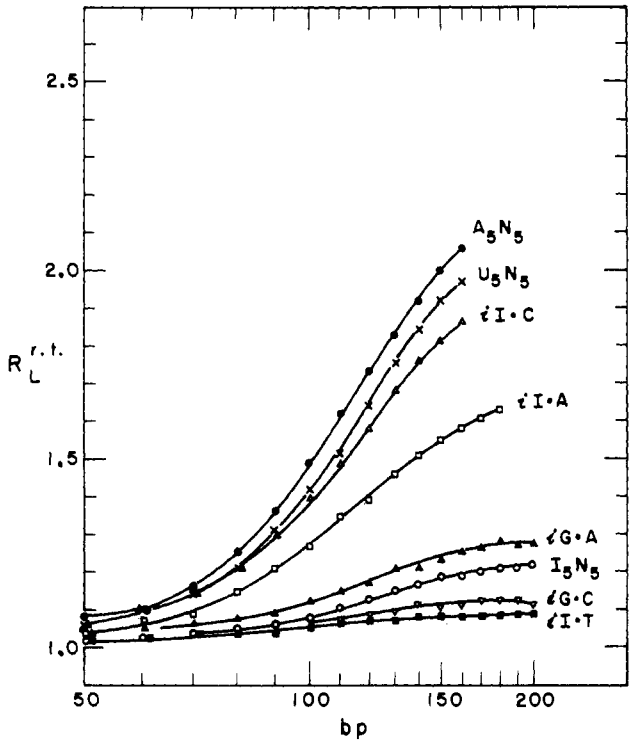

FIGURE 1: R_L values at room temperature (r.t.) vs. actual chain length. The ratio R_L of apparent multimer size, determined on polyacrylamide gels, to the actual chain length is an increasing function of the extent of bending. Sequences of repeating units in the multimers are illustrated in Table I. The electrophoresis conditions are described under Materials and Methods and in the previous paper (Koo et al., 1986). $(iG{\cdot}C^{5m})_n$ have the same R_L^{rt} values as $(iG{\cdot}C)_n$.

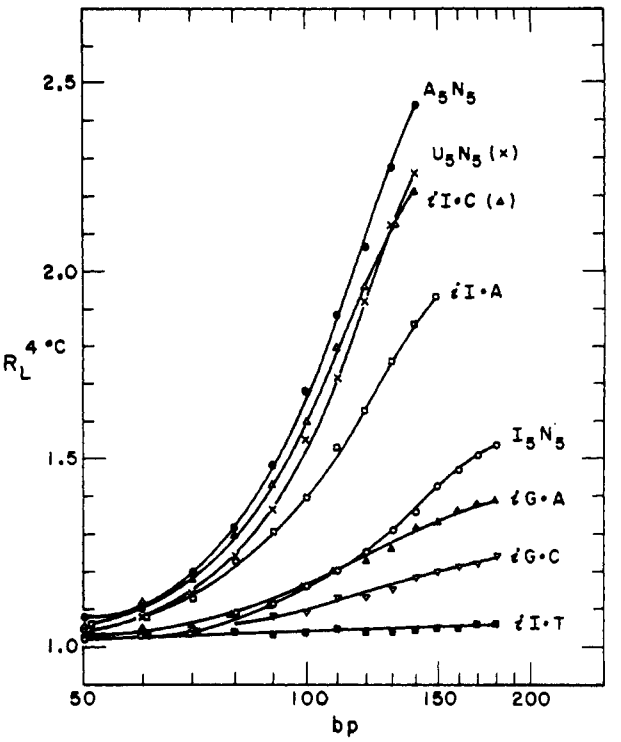

FIGURE 2: R_L values at 4 °C vs. actual chain length.

indicates the extent of anomaly in gel mobility, which is taken as a first-order approximation to be an increasing function of the extent of bending.

RESULTS

5-Methyl Group of Pyrimidines. As shown in Figures 1 and 2, the multimers $(U_5N_5)_n$, where adenosines are base paired with uridines instead of thymidines, show a large extent of bending, close to that of $(A_5N_5)_n$. The absence of the 5-methyl group in U compared with T reduces the apparent extent of bending slightly, but most of the bending in $(A_5N_5)_n$ is retained in $(U_5N_5)_n$. The sequences represented by $(iG{\cdot}C)_n$, in which the A tract is interrupted by G at the middle, were earlier shown to have only a very small amount of bending (Koo et al., 1986). The sequences $(iG{\cdot}C^{5m})_n$, which have 5-methylcytidine instead of C in the middle of the A·T tract, show exactly the same gel mobilities as $(iG{\cdot}C)_n$; the presence of the 5-methyl group in all five pyrimidines in a row does not increase the extent of bending in $(iG{\cdot}C^{5m})_n$. These two comparisons confirm that the pyrimidine 5-methyl group is not essential for the formation of bends which result from conversion to the putative H-DNA polymorphic structure.

2-Amino Group of Purines. Inosine resembles G but lacks the 2-amino group and can be base paired with C through two hydrogen bonds. When A tracts are interrupted by I·C as in $(iI{\cdot}C)_n$, the extent of gel anomaly is reduced by only 20%. This is surprising, considering the fact that the interruption by G·C almost completely abolishes bending by A tracts in $(iG{\cdot}C)_n$. The significant difference in the extent of bending between $(iI{\cdot}C)_n$ and $(iG{\cdot}C)_n$ indicates that the 2-amino group on purines destabilizes H-DNA. Whether the destabilization is caused by an additional hydrogen bonding or some other mechanism involving the amino group will be discussed below.

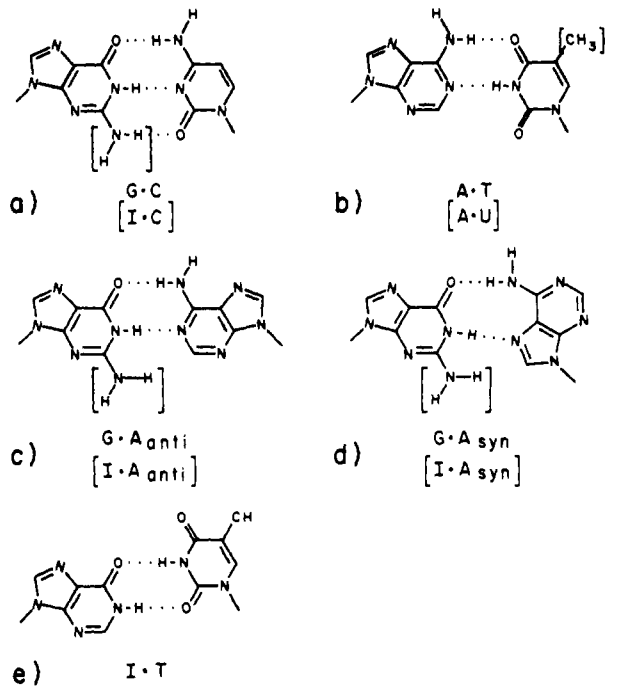

FIGURE 3: Schematic diagrams of the base pairs which occur in the sequences in Table I; adapted from Figure 2 in the paper by Kennard (1985). Base pairing modes of the base pairs in brackets are those lacking the functional groups in the corresponding brackets.

Strength of Base Stacking. It is possible that I·C base pairs, with only two hydrogen bonds as in A·T, may form a structure similar to H-DNA when stacked serially. However, as shown in Figure 1, the multimers $(I_5N_5)_n$ have only a small amount of bending compared to $(A_5N_5)_n$. This substantial difference between $(I_5N_5)_n$ and $(A_5N_5)_n$ in the extent of bending may be due to differences in the strength of stacking interactions of the bases; the strong tendency of A to stack may be an important factor for the formation of H-DNA.

Miscellaneous Interrupted Sequences. G·A and I·A have two hydrogen bonds of the same type, as shown in Figure 3

for both syn and anti configurations of A, but $(iG{\cdot}A)_n$ and $(iI{\cdot}A)_n$ have significant differences in their gel mobilities, which may originate from the presence or absence of the 2-amino group on G or I, respectively. The molecules $(iI{\cdot}T)_n$ show nearly normal gel mobilities, meaning that I·T has destroyed the H-DNA form completely. The I·T pair has no 2-amino group on the purine, which the earlier results showed to be favorable for the formation of H-DNA, bu the I and T bases are dislocated into the minor and major grooves, respectively (Kennard, 1985; Patel, 1982). The shift of bases into opposite grooves may be the cause of the disruption of the H form.

Temperature Dependence. All the sequences in Figure 2, except $(iI{\cdot}T)_n$, show larger bending at lower temperature; according to our model, this results from stabilization of the H form. The observations support the view that bends are caused by the same mechanism in all the sequences.

Discussion

General Considerations. DNA multimers with sequence phasings of 10 and 11 bp show pronounced gel mobility anomalies when the repeated sequence contains an A·T tract, but the anomalies are much smaller when the sequence phasing is 9 or 12 bp (Hagerman, 1985; Koo et al., 1986). Minor variations in phasing relative to the helical screw axis can result if base pair substitutions are made, due to possible alteration of the helix twist angles. However, these changes cannot be large enough to move the phase match with the sequence repeat out of the acceptable range of 10–11 bp per turn for observation of anomalies due to bending. Hence, major changes in gel mobilities must be due to alterations in bending for each repeating unit. At the same time, it should be recognized that minor effects, such as seen upon substitution of U for T (Figure 1), could be due to changes in phasing which result from an altered helical screw.

5-Methyl Group of Pyrimidines. The series $(U_5N_5)_n$, in which uridines have replaced the T's in $(A_5N_5)_n$, shows close similarity to $(A_5N_5)_n$ in apparent extent of bending. We conclude that the methyl group of T makes very little contribution to the formation of H-DNA. The same gel mobilities of $(iG{\cdot}C)_n$ and $(iG{\cdot}C^{5m})_n$ also support this conclusion.

2-Amino Group of Purines: H Bonding and the Influence of Water Structure. The multimers of (iI·C), in which I·C interrupts the A tract, show a much larger degree of bending than seen in the series $(iG{\cdot}C)_n$ which has G·C at the same position. The amino group on G, the only difference between the two sequences, is clearly responsible for the different extent of bending. The NH_2 group brings an additional hydrogen-bonding capability and is also known to hinder the regularity of water bridges in the minor groove of B-DNA (Dickerson et al., 1982). A special role for I in the formation of H-DNA is also indicated by the fiber diffraction pattern of alternating poly[d(A-I)]·poly[d(T-C)], which is found to be similar to that of poly(dA)·poly(dT) (S. Arnott, private communication). At present, we can offer only a conjecture concerning the relative contributions of the hydrogen-bonding and water structure disrupting features of the 2-NH_2 group to the destabilization of the H-form of DNA.

In the structure of poly(dA)·poly(dT) determined by linear dichroism, both bases are more inclined from the plane perpendicular to the helix axis than is the case in B-DNA, and T is inclined more than A (Edmondson & Johnson, 1985). The application of these features to A tracts gave a satisfactory explanation of the A-tract bending mechanism (Koo et al., 1986). It is likely that A·T, with two linkages between paired bases, has greater flexibility in the relative movement of the bases in a pair with respect to each other than is the case in G·C, with three H-bond linkages. The different inclinations of the bases in poly(dA)·poly(dT) are probably allowed by the larger flexibility in the A·T pair. On the assumption that base pairs with two linkages have enough flexibility to be fitted into H-DNA, the different extent of bending between $(iI{\cdot}C)_n$ and $(iG{\cdot}C)_n$ can be explained by assuming that I·C, with two hydrogen bonds, can be stacked in H-DNA, while G·C can not, thus disrupting the continuity of H-DNA.

However, the different extents of bending characterizing $(iI{\cdot}C)_n$ and $(iG{\cdot}C)_n$ can be analyzed from the alternative point of view. In B-DNA, a regular water bridge is formed in the minor groove, but the amino group of G disrupts the regularity (Dickerson et al., 1982). A tracts without the 2-amino group probably have an ordered water spine in the minor groove, which may stabilize the H-form as it does B-DNA. The assumption that the water spine stabilizes H-DNA leads to the following interpretation: Whereas I·C keeps the regularity of the water spine, preserving the H-form, G·C disrupts it, thus destabilizing the H-form. This view is supported by the results obtained on $(iI{\cdot}A)_n$ and $(iG{\cdot}A)_n$. In the G·A mismatch, there are two possible ways of base pairing (Figure 3); the G_{anti}-A_{anti} form is claimed on the basis of NMR data (Kan et al., 1983; Patel et al., 1984) and the G_{anti}-A_{syn} form on the basis of X-ray crystallographic data (Kennard, 1985). No matter which form is correct in solution, both I·A and G·A have two hydrogen bonds, but the regularity of the water spine in $(iG{\cdot}A)_n$ is disrupted by G in the middle of the A run, while it is kept in $(iI{\cdot}A)_n$. This results in larger bending in $(iI{\cdot}A)_n$ than in $(iG{\cdot}A)_n$. The 2-amino group on purines may reduce the extent of bending, or equivalently the stability of H-DNA, by either of the mechanisms described here, or by both. The distinction could be clarified further by investigating sequences having 2-aminopurine, 2-aminoadenine, or adenine with a hydrophobic bulky group at the C_2 position instead of A in A tracts.

Strength of Base Stacking. The series $(I_5N_5)_n$, in which I·C base pairs lacking the 2-amino group are stacked in a row, has only a small amount of bending compared to $(A_5N_5)_n$. The stacking tendency of I is likely to be similar to that of G, which is weaker than that of A (Shum & Crothers, 1983). Hence, the smaller amount of bending in $(I_5N_5)_n$ than in $(A_5N_5)_n$ could be ascribed to weaker stacking interactions for I than for A.

Conclusions. According to our model, bending occurs at the junction between H-DNA with B-DNA, and the extent of bending is an increasing function of the stability of H-DNA. Bending in A tracts can also be described in terms of wedge angles between adjacent base pairs (Ulanovsky et al., 1986), with structural consequences closely analogous to the main features of H-DNA (H.-S. Koo and D. M. Crothers, unpublished results). Comparison of gel mobility data for different sequences leads to the following conclusions concerning the origin of the anomalous DNA structure found in homopolymeric A·T tracts: for formation of H-DNA (or its equivalent in terms of the wedge model) and the resultant bending, (i) the 5-methyl group on pyrimidines is not essential; (ii) the 2-amino group on purines is an interfering factor; and (iii) the strong stacking of A may be an important contributing factor.

Registry No. A_5N_5, 108189-53-3; U_5N_5, 108167-85-7; I_5N_5, 108167-88-0; iG·C, 108167-90-4; $iG{\cdot}C^{5m}$, 108167-92-6; iI·C, 108167-94-8; iI·A, 108167-96-0; iG·A, 108167-97-1; iI·T, 108167-99-3; poly(dA)·poly(dT), 24939-09-1; adenine, 73-24-5; thymine, 65-71-4.

References

Bloom, K. S., Amaya, E., Carbon, J., Clarke, L., Hill, A., & Yeh, E. (1984) *J. Cell Biol. 99*, 1559–1568.

Dickerson, R. E., Drew, H. R., Conner, B. N., Wing, R. M., Fratini, A. V., & Kopka, M. L. (1982) *Science (Washington, D.C.) 216*, 475–485.
Diekmann, S. (1986) *FEBS Lett. 195*, 53–56.
Edmondson, S. P., & Johnson, W. C., Jr. (1985) *Biopolymers 24*, 825–841.
Gasser, S. M., & Laemmli, U. K. (1986) *Cell (Cambridge, Mass.) 46*, 521–530.
Griffith, J., Bleyman, M., Rauch, C. A., Kitchin, P. A., & Englund, P. T. (1986) *Cell (Cambridge, Mass.) 46*, 717–724.
Hagerman, P. J. (1984) *Proc. Natl. Acad. Sci. U.S.A. 79*, 4632–4636.
Hagerman, P. J. (1985) *Biochemistry 24*, 7033–7037.
Kan, L., Chandrasegaran, S., Pulford, S. M., & Miller, P. S. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4263–4265.
Kennard, O. (1985) *J. Biomol. Struct. Dyn. 3*, 205–226.
Koo, H., Wu, H-M., & Crothers, D. M. (1986) *Nature (London) 321*, 501–506.
Levene, S. D., Wu, H.-M., & Crothers, D. M. (1986) *Biochemistry 25*, 3988–3995.
Marini, J. C., Levene, S. D., Crothers, D. M., & Englund, P. T. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 7664–7668.
Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Dallas, J., & Itakura, K. (1982) *Biochemistry 21*, 437–444.
Patel, D. J., Kozlowski, S. A., Ikuta, S., & Itakura, K. (1984) *Biochemistry 23*, 3207–3217.
Ryder, K., Silver, S., DeLucia, A. L., Fanning, E., & Tegtmeyer, P. (1986) *Cell (Cambridge, Mass.) 44*, 719–725.
Shum, B., & Crothers, D. M. (1983) *Biopolymers 22*, 919–933.
Ulanovsky, L., Bodner, M., Trifonov, E. N., & Choder, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 862–866.
Wu, H.-M., & Crothers, D. M. (1984) *Nature (London) 308*, 509–513.
Zahn, K., & Blattner, F. R. (1985) *Nature (London) 317*, 451–453.