A Novel DNA Duplex. A Parallel-Stranded DNA Helix with Hoogsteen Base Pairing

Keliang Liu,* H. Todd Miles,* Joe Frazier, and V. Sasisekharan*

Laboratory of Molecular Biology, NIDDK, National Institutes of Health, Bethesda, Maryland 20892

Received June 22, 1993; Revised Manuscript Received August 17, 1993®

ABSTRACT: We show here for the first time that a stable parallel double helix with Hoogsteen pairing can exist independently of the triple helix of which it is a component part. The experiments employ DNA oligonucleotides with mixed sequences of normal bases. These duplexes are distinct from previously reported ribopolynucleotide helices containing bulky substituents which prevent Watson–Crick pairing as well as from parallel duplexes with Donohue, or reversed Watson–Crick, pairing. Stoichiometry is established by mixing curves and gel electrophoresis. T_m depends linearly upon pH, increasing with acidity because of the need to protonate N3 of C. The T_m of the 20-mer studied here is 52 °C at pH 5.2 and 0.1 M NaCl. At pH above 6, the molecule rearranges to form an antiparallel duplex with imperfect Watson–Crick pairing and loops, and the T_m is then independent of pH. The CD spectrum of the parallel duplex is very similar to that of the corresponding triple helix but quite different from that of the Watson–Crick helix. The infrared spectrum in the double bond region closely resembles that of the triple helix but, as with the CD, is quite different from that of the Watson–Crick duplex. The infrared spectra of the duplex and triple helix are also nearly identical in the region from 800 to 1000 cm^{-1}, which is sensitive to backbone conformation. The only symmetry element present is a pseudorotational axis coincident with the helix axis of the parallel duplex as well as with the axis of the corresponding triple helix. We suggest that structural and spectroscopic properties of the triple helix are determined by the Hoogsteen-paired strands and not by the Watson–Crick-paired component strands.

A third polynucleotide strand can be bound to the all-purine strand of the Watson–Crick helical duplex by the hydrogen-bonding scheme observed by Hoogsteen (1963) in monomeric crystals [cf. Felsenfeld et al. (1957), Miles and Frazier (1964), Inman and Baldwin (1964), Howard et al. (1964), Blake and Fresco (1966), Riley et al. (1966), Hattori et al. (1976), and Howard and Miles (1984)].[1] The pyrimidine third strand is parallel to the purine strand and antiparallel to the pyrimidine strand of the Watson–Crick-paired duplex. In order for GC^+ Hoogsteen pairs to be formed, the C must be protonated to provide a hydrogen bond donor for bonding to N7 of G [cf. Howard et al. (1964)]. This protonation serves the further functions of blocking Watson–Crick pairing and stabilizing the Hoogsteen GC^+ pairs by electrostatic interaction of the C^+ with the phosphates.

We show here for the first time that stable parallel double helices with Hoogsteen base pairing can exist independently of triple helices. The experiments employ DNA oligonucleotides with mixed sequences of normal bases. The helices are unconstrained in the sense that no sterically blocking groups are present which could prevent formation of Watson–Crick-paired helices (see below). The two strands are parallel, and the transition temperatures exhibit a marked pH dependence, reflecting protonation of C residues in the mixed C,T sequence. When the pH is raised to approximately 6.2, the structure rearranges to an antiparallel Watson–Crick-paired helix with mispairs, and T_m then becomes independent of pH. Infrared spectra in D_2O solution in the conformationally sensitive region between 800 and 1000 cm^{-1} have bands characteristic of B-form geometry and C2′-endo sugar pucker. These spectra are closely similar to those of the corresponding triple helices, suggesting a similarity of backbone conformations in the double and triple helices.

Ribopolynucleotide AU parallel double helices with Hoogsteen pairing have been reported previously with sterically blocking groups at C2 of A which prevent Watson–Crick base pairing (Ishikawa et al., 1972; Ikehara et al., 1972; Hattori et al., 1974). [A purine–purine duplex with parallel strands between poly(I) and poly($2NH_2A$) with bonding to N7 of $2NH_2A$ has also been reported and discussed by Howard et al. (1977).] In contrast to the above sterically blocked polynucleotides, the novel parallel DNA helices discussed in this report are formed from normal bases capable of either Watson–Crick or Hoogsteen pairing. They may thus be described as unconstrained parallel helices with Hoogsteen pairing, as mentioned above.

We note that Wilkins et al. (1970) have stated that they had not been able to build Hoogsteen models with good stereochemistry and that they also had no success in building stereochemically reasonable models with Donohue pairs (see Donohue (1956); these pairs are also sometimes referred to as reversed Watson–Crick). In this paper we present experimental evidence that a double helix with Hoogsteen pairing can readily be formed, and a parallel modeling study (G. Raghunathan, H. T. Miles, and V. Sasisekharan, manuscript in preparation) shows that a stereochemically satisfactory structure can be built with standard torsional angles, acceptable hydrogen bonds, and no short contacts. In the case of parallel helices with Donohue pairs, Pattabiraman (1986) has shown earlier that a stereochemically satisfactory model can be built, and experimental evidence for the existence of such helices has been reported in a number of papers [cf. German et al. (1987), and van de Sande et al. (1988)].

® Abstract published in *Advance ACS Abstracts*, October 1, 1993.

[1] Helical complexes formed by synthetic oligonucleotides designated Py20, Pu20, and RPy20 are described. The standard center dot notation is used to indicate a paired complex. Where needed for clarity or emphasis, the words Watson–Crick or Hoogsteen or their abbreviations WC or H may be placed in parentheses to indicate the hydrogen-bonding scheme.

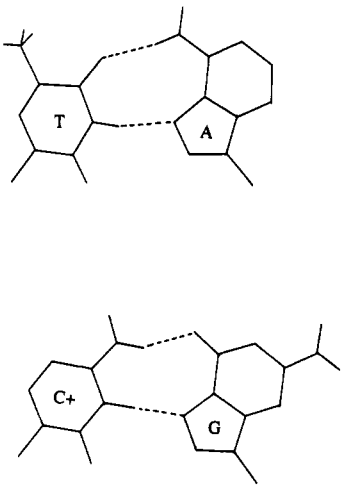


FIGURE 1: Hoogsteen AT and GC^+ base pairs.

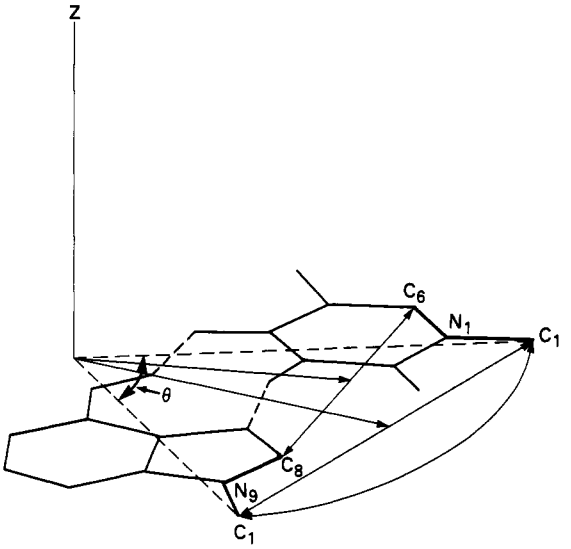


FIGURE 2: Pseudorotational symmetry of the Hoogsteen AT pair. The rotational symmetry axis (z) is at the intersection of the perpendicular bisectors of the lines C1′(A)–C1′(T) and C8(A)–C6-(T). The C1′N1C6 part of T is identical to the C1′N9C8 part of A. θ is the angle of rotation about the z-axis which brings these sets of atoms into coincidence and brings the backbones of A and T into coincidence.

MATERIALS AND METHODS

Oligonucleotides were synthesized with Applied Biosystems Model 380A DNA synthesizer using a solid-phase cyanoethylphosphoramidite method in 10 μM scale. Oligonucleotides were purified on 15% or 20% denaturing polyacrylamide/Bis (19:1) gel with 7.6 M urea, 0.09 M Tris/boric acid buffer, pH 8.3, 0.02 M EDTA. Molar extinction coefficients were measured by phosphate analysis, as described previously (Muraoka et al., 1980). The molar extinction coefficients for the oligonucleotides are 8138 (Py20), 8638 (Pu20), and 8196 (RPy20).

Analytical gel electrophoresis was run on 15% polyacrylamide/Bis (19:1) nondenaturing gel under the following conditions: (a) 0.2 M Na_2HPO_4/citric acid buffer without extra salt, pH 5.5, 35 °C, 280 V; (b) 0.09 M Tris/boric acid buffer, pH 8.3, 0.02 M EDTA, 0.02 M $MgCl_2$, 380 V, 35 °C. Each gel sample (5×10^{-6} M in each strand) was dissolved in 10 μL of 30% glycerol/H_2O with the same buffer and salt as the gel, heated to 80 °C, and then cooled slowly to room temperature.

CD spectra were measured with a J-500A spectropolarimeter controlled by an LDACS system (Powell et al., 1980), and the data were processed by a Vax 11/84 computer. Each spectrum is the average of eight scans. The sample concentration was 9×10^{-5} M in base. The samples were heated to 80 °C for 3 min and annealed by slow cooling to room temperature.

UV melting curves were measured automatically with a Carry 210 spectrophotometer controlled by an LDACS computer system as above. UV spectra for mixing curves were measured with a Carry 118 spectrophotometer also controlled by an LDACS system.

The IR spectra were measured with a Perkin-Elmer 500B spectrometer controlled by an LDACS system, and the data were processed by a Vax 11/84 computer. The spectra were measured in D_2O solution using CaF_2 windows in the double-bond region and ZnSe windows at low frequency [cf. Miles (1971, 1981), Govil et al. (1981), and Howard et al. (1992)].

RESULTS AND DISCUSSION

Design of the Oligonucleotides. The Hoogsteen hydrogen-bonding schemes are shown in Figure 1. Assuming an anti conformation of the bases, it is necessary that the strands be parallel and that one contains only purines and the other only pyrimidines. The AT and GC^+ pairs in Figure 1 are isostructural in that they can be superimposed with approximate coincidence of the corresponding atoms of the bases. The parallel ribose–phosphate backbones, moreover, can be brought into exact coincidence by rotation about a pseudorotational axis coincident with the helix axis (Figure 2). As a consequence, the AT and GC^+ pairs can, in principle, be interchanged in any order, and the helix might have any base composition. In practice, however, the chemistry of the bases and the existence of competing reactions place limitations on the sequences and compositions which can be expected to form parallel duplexes with Hoogsteen pairing. At the extreme of 100% AT composition, it is known that poly(dA) and poly-(dT) spontaneously form helices with Watson–Crick rather than Hoogsteen pairing. At the other extreme, 100% GC^+ composition cannot be employed either. We have found that the stable hemiprotonated self-structures formed by poly-(dC) and oligo(dC) [cf. Inman (1964), and Gray et al. (1980)] prevent formation of CGC^+ triple helices or GC^+ duplexes (F. B. Howard, J. Frazier, and H. T. Miles, unpublished).[2] Oligomers with runs of dC were therefore avoided. Within the composition range in which hetero interaction does occur, the possibility still exits of a polarity reversal with formation of a Watson–Crick-paired duplex with mispairs or loops. For this reason, sequences are chosen to be perfectly matched for Hoogsteen but not for Watson–Crick pairing. As we show below, a polarity reversal does occur even with sequences perfectly matched for Hoogsteen pairing as the pH increases to 6 and above. Thus, to achieve such a parallel duplex, a judicious choice of composition and sequences must be made to ensure that polarity reversal would be accompanied by extensive mismatching and that runs of dC are avoided. With these considerations in view, the following sequences were synthesized:

[2] Though it is generally believed that a protonated deoxyCGC$^+$ triple helix is stable, we have found using CD and IR spectroscopy that this triple helix is not formed at a series of acid pH values and salt concentrations. Poly(dC) forms, instead, the hemiprotonated CC$^+$ acid self-structure to the exclusion of the CGC$^+$ triple helix. The same result was obtained when oligo(dC) of length 6 or 20 was used. With poly(rC) and rGMP, dGMP (Howard et al., 1964; Miles & Frazier, 1982) and poly(rG) (F. B. Howard, J. Frazier, and H. T. Miles, unpublished), IR spectra characteristic of the triple helix are observed. These are quite different from spectra of the hemiprotonated CC$^+$ helix as well as from those of the Watson–Crick duplex.

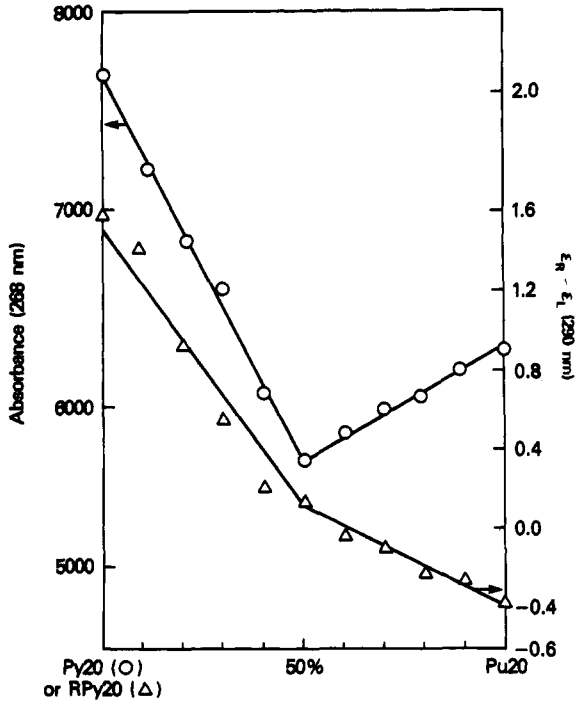


FIGURE 3: Mixing curves of Watson–Crick- and Hoogsteen-paired duplexes. O, UV mixing curve of Py20·Pu20(WC), 0.002 M sodium pyrophosphate buffer, pH 7.5, 0.1 M sodium chloride, 20 °C. Δ, CD mixing curve of Pu20·Rpy20(H), 0.02 M sodium cacodylate buffer, pH 5.1, 0.1 M sodium chloride, 20 °C.

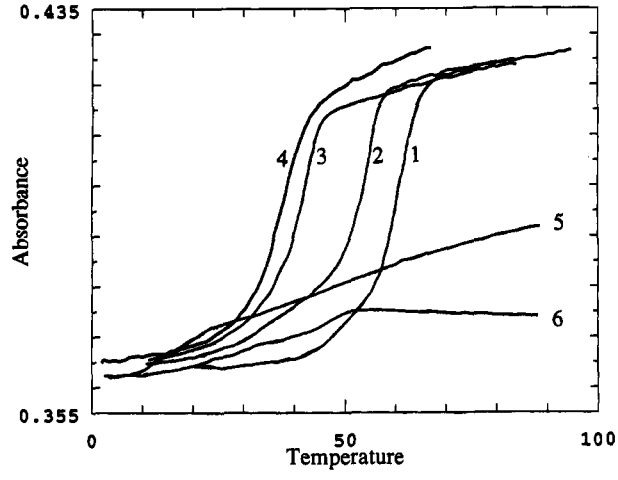


FIGURE 4: UV melting curves, 0.02 M sodium cacodylate buffer, 0.1 M sodium chloride, 255 nm. Curve 1: Pu20·Py20(WC), pH 6.9. Curves 2–4: Pu20·RPy20(H), 2, pH 5.2; 3, pH 5.7; 4, pH 8.5. Curve 5: Pu20 alone, pH 5.2. Curve 6: RPy20 alone, pH 5.2. In curve 4 at pH 8.5, the structure has rearranged to a Watson–Crick duplex with loops (see text).

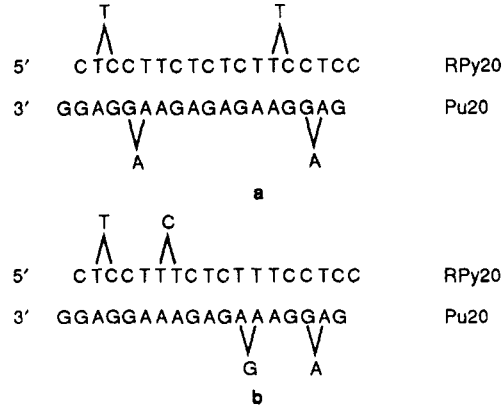
Py20	3′	C T T C C T T C T C T C T T T C C T C C
Pu20	5′	G A A G G A A G A G A G A A A G G A G G
RPy20	5′	C T T C C T T C T C T C T T T C C T C C

The homopurine strand Pu20 can form a perfectly paired parallel duplex (H) with the homopyrimidine strand RPy20 and a perfectly paired antiparallel duplex (WC) with Py20 of opposite polarity. The latter complex is used for comparison with the Hoogsteen-paired parallel duplex and the looped or mispaired antiparallel helix (WC) formed by strand reversal of Pu20·RPy20 at pH above 6.

Stoichiometry. A CD mixing curve of Pu20 and RPy20 at pH 5.1 shows two straight lines intersecting at 50% Pu20, demonstrating formation of a 1:1 parallel complex. Pu20 and Py20 form a 1:1 antiparallel Watson–Crick-paired duplex (Figure 3).

Thermal Transitions. Melting curves of the parallel duplex Pu20·RPy20 (H) at pH 5.2 and 5.7 are shown in curves 2 and 3, Figure 4. The curves are cooperative and monophasic with T_m values of 52 and 40 °C, respectively, in 0.1 M NaCl. Curve 1, shown for comparison, is that of the corresponding Watson–Crick helix Pu20·Py20 and has a T_m of 60 °C. Curve 4 falls in a pH-independent range and represents an antiparallel (Watson–Crick) helix formed by rearrangement of Pu20·RPy20 at pH 6.9 (see below). Temperature profiles of the separate component strands are shown in curves 5 (Pu20, pH 5.2) and 6 (RPy20, pH 5.2). In each case, small rises in absorbance indicate a minor amount of self-structure but clearly differ too much in position and magnitude to account for the large transitions seen in the other curves.

pH Dependence of T_m. We expect from the requirement of protonation at C3 that T_m of the parallel duplex (Hoogsteen) will vary inversely with pH. We see that the dependence is linear over the range of pH 4.6–6 (Figure 5). Above pH 6, there is an abrupt change, and the curve becomes independent of pH. The fully paired helix Pu20·Py20 (WC) above pH 6 exhibits a parallel line at 21 °C and higher temperature (Figure 5, triangles). We consider the mixing curve and the observed pH dependence below pH 6 to be strong evidence that the structure in this range is a duplex with Hoogsteen pairing and that the abrupt transition to pH independence above pH 6 reflects a rearrangement to an imperfectly paired antiparallel (Watson–Crick) helix. Though other structures are possible in principle, the structures of the rearranged product which we consider most likely are shown here.



Both structures have 17 base pairs, four looped out (or mispaired) bases, and one unpaired base at each end. Structure **a** with nine GC and eight AT pairs should be more stable than **b** with eight GC and nine AT pairs. Structures involving a larger number of loops or unpaired bases are considered unlikely. By forming an antiparallel (WC) duplex with mispairs, Pu20·RPy20 has a T_m reduced by 21 °C compared to the perfectly paired antiparallel (WC) duplex. (cf. Figure 5). Lomant and Fresco (1975) estimated that for an average long RNA helix of 50% GC content, a 1% mismatch reduces the T_m by about 1 °C. If this relation is applied in the present case to mispaired Pu20·RPy20 (WC), a depression of 23.5 °C is obtained, compared to the observed 21 °C. The result is consistent, though in view of the difference in chain length and the presence of deoxyribose rather than ribose, the closeness of agreement may be in part fortuitous.

Salt Dependence. The transition temperature of the parallel duplex Pu20·RPy20 (H) is almost independent of salt concentration, with $dT_m/d\log[Na^+] = 2.2°$ at pH 5.1 (Figure 6). The perfectly paired duplex (WC) Pu20·Py20, in contrast, has a slope of 16.9°, in the normal range for such complexes.

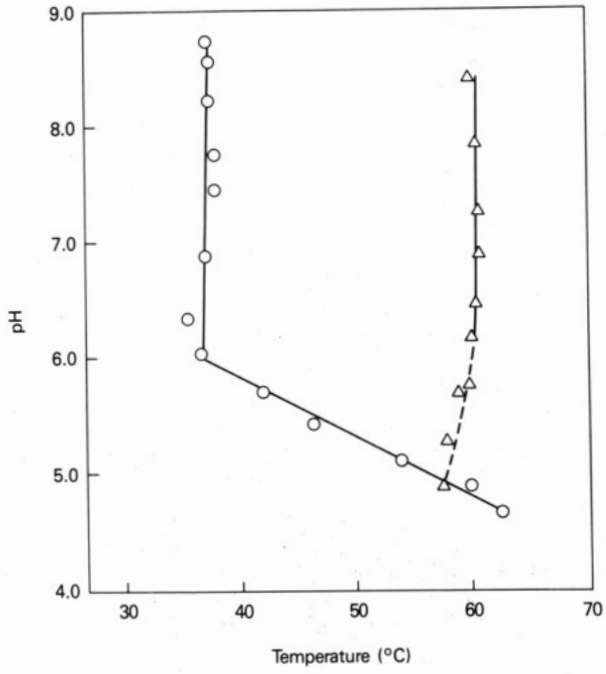


FIGURE 5: Dependence of T_m on pH in 0.02 M sodium cacodylate and 0.1 M sodium chloride. ○, Pu20·RPy20; this duplex has Hoogsteen pairing below pH 6 and Watson–Crick pairing with loops above pH 6 (see text). △, Pu20·Py20 Watson–Crick duplex.

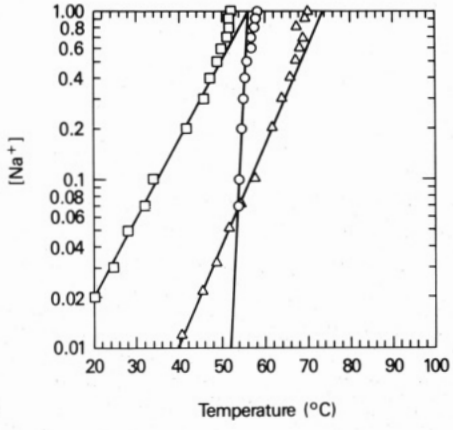


FIGURE 6: Salt dependence of melting temperature. ○, Pu20·RPy20-(H), 0.02 M sodium cacodylate, pH 5.1, $dT_m/d \log[Na^+] = 2.2°$; □, Pu20·RPy20, 0.02 M sodium cacodylate, pH 7.5, Watson–Crick duplex with loops, $dT_m/d \log[Na^+] = 21.2°$; △, Pu20·Py20, 0.02 M sodium cacodylate, pH 7.0, perfectly paired Watson–Crick duplex, $dT_m/d \log[Na^+] = 16.9°$.

The mispaired or looped Pu20·RPy20 (WC) formed by rearrangement of the Hoogsteen-paired complex near neutral pH has a slope of 21.2°, 4.3° higher than that of the fully paired Watson–Crick duplex. The increase in slope of the salt dependence curve with extent of mispairing of Watson–Crick helices is consistent with the results of other studies [cf., for example, Lomant and Fresco (1975)].

The marked salt dependence of T_m of antiparallel helices with Watson–Crick pairing is due to screening of electrostatic repulsion of phosphates of the two strands by counterions. In the Hoogsteen-paired duplex, this relationship is fundamentally altered by the presence of positive charges on the C residues. This positive charge on CH^+ is delocalized over the N3, C4, and N6 atoms and is polarizable by nearby negative charges. In the present case we found (modeling studies, unpublished) that the distance from the positive charge in one C residue to the nearest phosphate on the opposite strand is roughly equal to the distance from the positive charge to the nearest phosphate of the same strand as the C residue. The positive charge may thus contribute to partial internal neutralization of one negative charge as well as to stabilization of the complex by electrostatic attraction for the opposite strand.

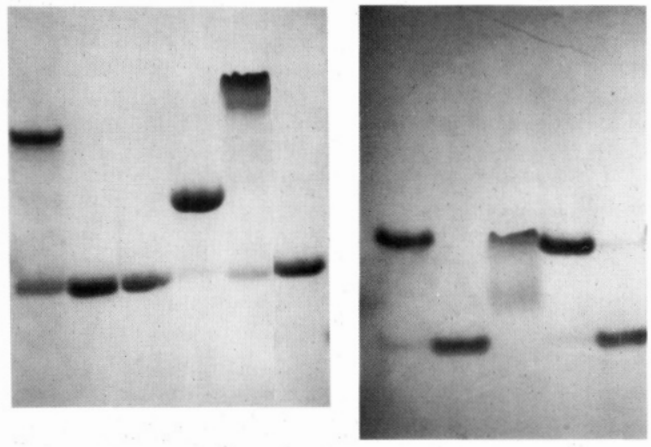


FIGURE 7: Electrophoresis of oligonucleotide helices on a nondenaturing gel [15% polyacrylamide/bis (19:1)]. Left panel: 0.2 M Na_2HPO_4/citric acid, pH 5.5, $[Na^+] = 0.12$ M, 280 V, 8.5 h at 35 °C. Lanes are numbered from left to right. Lane 1, Pu20·RPy20 (H); lane 2, RPy20; lane 3, Pu20; lane 4, Py20·Pu20 (WC); lane 5, Py20·Pu20·RPy20 (1:1:1 triplex); lane 6, Py20. Right panel: 0.09 M Tris/boric acid, pH 8.3, 0.02 M EDTA, 0.02 M $MgCl_2$, 380 V, 8 h, 35 °C. Lane 1, Pu20·RPy20 (WC with loops; see text); lane 2, RPy20; lane 3, Pu20; lane 4, Pu20·Py20 (WC); lane 5, Py20.

Gel Electrophoresis. At pH 5.5, the parallel duplex Pu20·RPy20 (H) has a much lower mobility than the Watson–Crick duplex at the same pH because the positive charges of the C^+ residues in the former reduce the net negative charge on the helix (Figure 7A, bands 1 and 4; see previous discussion). The parallel duplex (H) is also distinguished from the triple helix Py20·Pu20·RPy20, which travels more slowly because of its higher molecular weight (Figure 7A, band 5). At higher pH, the Pu20·RPy20 complex rearranges to form an antiparallel Watson–Crick-paired duplex with loops. Its mobility is almost the same as that of the perfectly paired Watson–Crick duplex Pu20·Py20 (Figure 7B, lanes 1 and 4). In both nondenaturing gels at pH 5.5 and 8.3, Py20 and RPy20 show sharp bands (Figure 7A, lanes 2 and 6, and Figure 7B, lanes 2 and 5), but Pu20 at higher pH in the presence of Mg^{2+} shows a smeared band, presumably as a result of self-structure formation (Figure 7B, lane 3). In denaturing gel, all of the separate components travel as sharp bands (not shown).

Circular Dichroism. The parallel duplex Pu20·RPy20 (H) at pH 5.5 has a positive band of moderate intensity at 281 nm, weak negative maxima at 255 and 230 nm, a minimum at 213 nm, and a maximum at 201 nm (Figure 8A; Table I). At the same pH, the duplex Pu20·Py20 (WC), in contrast, has a typical B DNA spectrum with a conservative pair of bands at 272 and 242 nm, equally displaced from the crossover at 255 nm, a maximum at 217 nm, and a minimum at 212 nm. This spectrum is essentially unchanged from that observed at neutral pH, indicating that the Watson–Crick structure is preserved in this mild acid solution. Pu20·RPy20 rearranged to an antiparallel mispaired helix at pH 7.2 has a spectrum closely similar to that of the duplex with perfect Watson–Crick pairing at a pH 5.1 or 7.2. If the pH is lowered to 5.1, the structure Pu20·RPy20 reverts to the Hoogsteen-paired duplex and the spectrum to curve 1, Figure 8A.

The spectra of the parallel duplex Pu20·RPy20 (H) and the triple helix Py20·Pu20·RPy20, both at pH 5.1, are compared in Figure 8B (curves 1 and 2, respectively). The spectra exhibit a remarkable similarity, differing only in intensity. The spectrum of the triple helix is evidently determined by its

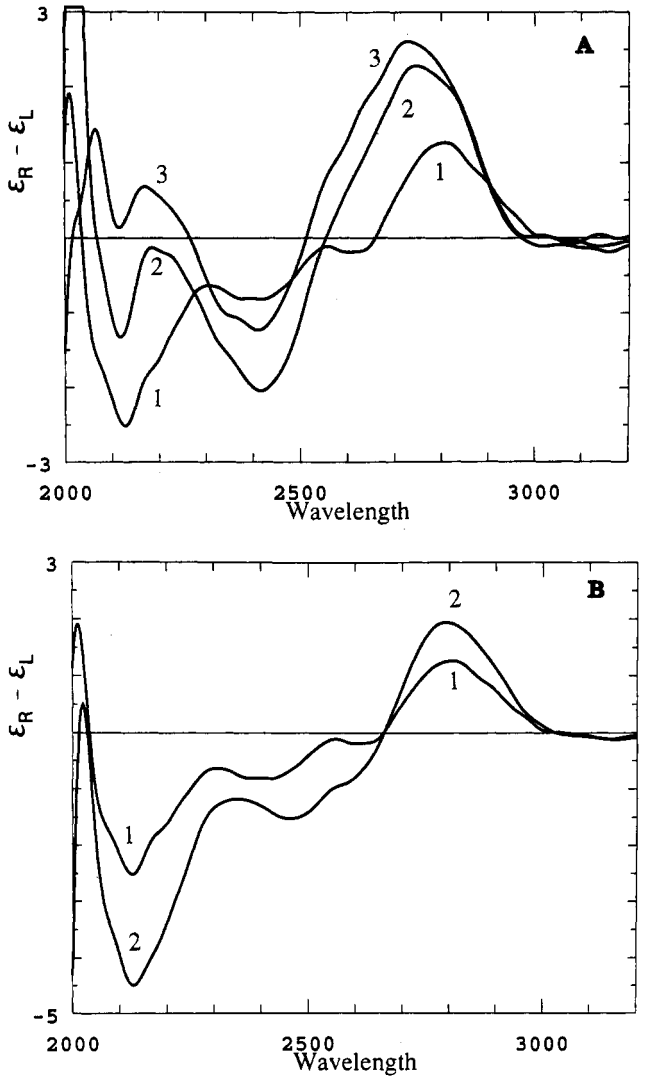


FIGURE 8: (A) CD spectra, 0.02 M sodium cacodylate, 0.1 M NaCl, 20 °C. Curve 1, Pu20·RPy20 (H), pH 5.1; curve 2, Pu20·Py20-(WC), pH 5.1; curve 3, Pu20·RPy20 (WC with loops, see text), pH 7.2. (B) CD spectra, 0.02 M sodium cacodylate, pH 5.1, 0.1 M NaCl, 20 °C. Curve 1, Pu20·RPy20 (H); curve 2, Py20·Pu20·RPy20 (1:1:1) triplex.

Hoogsteen component and shows little effect of the Watson–Crick component.

A further comparison of interest is that of the CD spectra of the double and triple helices above with the spectrum of the triple helix formed by regularly alternating copolymers, poly[d(C^+-T)d(A-G)d(C-T)] (Antao et al., 1988). Despite the differences in sequence, the spectra are quite similar, with a characteristic moderately strong minimum at 215 nm, a weak minimum at 263 nm, and a maximum at 282 nm. In view of the above results, it appears that the CD spectrum of the regularly repeating high-polymer triplex is also determined by its Hoogsteen-paired component.

As with the IR spectra (see below), the similarity of the CD spectra of the parallel duplex and the triple helix and their differences from independent Watson–Crick duplexes can be explained by the structure and symmetry of the three helices. The parallel duplex and triple helix have a common helix axis and axis of pseudorotational symmetry and the same number of residues per turn (G. Raghunathan, H. T. Miles, and V. Sasisekharan, manuscript in preparation). The independent Watson–Crick duplex, in contrast, has an axis closer to the bases (D = 0–1 Å), no rotational symmetry, and 10 rather than 12 residues per turn. These structural properties control the disposition of the bases in three dimensions and hence the

Table I: Summary of CD and IR Spectral Data

material	pH	Circular Dichroism λ_{max} (nm)	$\epsilon_L - \epsilon_R$	λ_{min} (nm)	$\epsilon_L - \epsilon_R$
Pu20·RPy20	5.1	281	1.25	260	–0.25
		255	–0.10	243	–0.85
		230	–0.62	237	–0.85
		201	1.90	213	–2.50
	7.2	272	2.60	242	–1.25
		217	0.70	211	0.15
		206	1.40		
Pu20·Py20	5.1	272	2.25	242	–2.05
		217	–0.10	212	–1.30
Py20·P20·RPy20	5.1	278	1.95	247	–1.50
		234	–1.20	213	–4.50
		202	1.95		

material	Infrared ν_{max} (cm^{-1})	ν_{max} (cm^{-1})
Pu20·RPy20 (H)	1702 (s)	1580 (s)
	1656 (s)	~1565 (m; sh)
	1625 (s)	1500 (m)
Py20·Pu20·RPy20	1702 (s)	~1564 (w; sh)
	1656 (s)	1500 (m)
	1627 (m; sh)	
	1586 (w)	
Py20·Pu20 (WC)	~1696 (s; sh)	1580 (m)
	1677 (s)	1562 (w; sh)
	1650 (s)	~1522 (w; sh)
	1622 (m)	1501 (m)
$d(G)_n$·$d(C)_n$	1687.5 (s)	~1565 (w; sh)
	1653 (s)	1530 (w)
	~1625 (w; sh)	1503.5 (m)
$d(T)_n$·$d(A)_n$·$d(T)_n$	1695 (s)	
	1658 (s)	
	1630 (m; sh)	
$d(A)_n$·$d(T)_n$	1695.5 (s)	1621 (m)
	1662.5 (m)	1574.5 (w)
	1641.5 (m)	

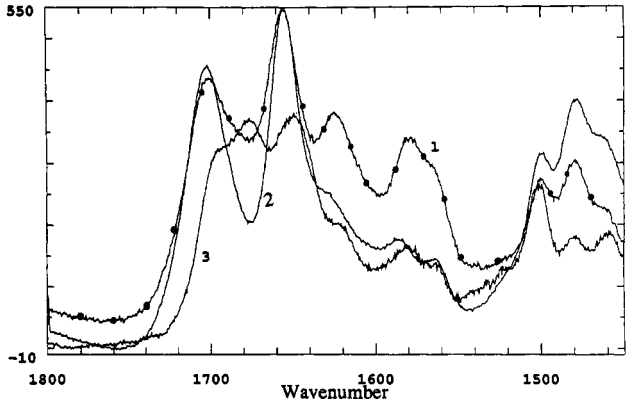


FIGURE 9: IR spectra in double-bond region. DNA concentration: 2.6×10^{-3} M helix for 1 and 2 and 1.7×10^{-3} M for 3, 0.1 M NaCl, D_2O. Spectrum 1, Pu20·RPy20 (Hoogsteen-paired duplex), 0.005 M sodium phosphate buffer, pH 5.0, 32 °C; spectrum 2, Py20·Pu20·RPy20 (1:1:1 triplex), 0.01 M sodium cacodylate, pH 4.8, 9.5 °C; spectrum 3, Py20·Pu20 (Watson–Crick duplex), 0.01 M sodium cacodylate, pH 7.0, 15 °C. Ordinate, molar absorbance on the basis of monomer residues.

electronic interactions among the bases which produce the CD spectra.

Infrared Spectra in the Double-Bond Stretching Region. Infrared spectra of the parallel duplex with Hoogsteen pairing, Pu20·RPy20, the triple helix, Py20·Pu20·RPy20, of which the parallel duplex is a component, and the Watson–Crick duplex, Py20·Pu20, are shown in Figure 9, and the band frequencies are listed in Table I. In later paragraphs, all of these bands will be assigned to specific bases and atomic groupings within bases and will be used to reach important structural conclusions. These assignments are based on studies

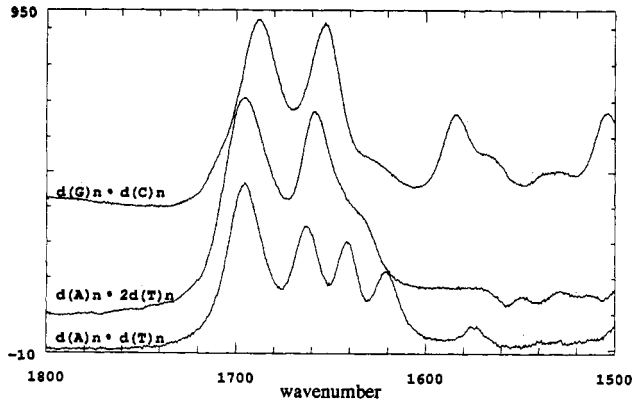


FIGURE 10: IR spectra of the double-bond region of (top) $d(G)_n$·$d(C)_n$ 0.10 M in polymer phosphate, 0.19 M NaCl, 0.02 M Na_2HPO_4, pH 7.5, 4.2 °C; (center) $d(T)_n$·$d(A)_n$·$d(T)_n$, 0.011 M in polymer phosphate of $d(A)_n$ and 0.022 M in polymer phosphate of $d(T)_n$, 0.28 M NaCl, 0.011 M sodium cacodylate, pH 7.3, 5.1 °C; (bottom) $d(A)_n$·$d(T)_n$, 0.02 M in polymer phosphate, 0.06 M NaCl, 0.01 M Na_2HPO_4, pH 7, 29 °C. Ordinate units are molar absorbance, and the index marks represent 100 units.

of simpler homopolymer systems as outlined in the following paragraphs. In relevant cases, the assignments are based on specific isotopic substitutions, described in the cited papers.

The triple helix formed between GMP and poly(rC) in slightly acidic solution has bands at 1708 (s), 1687 (m), 1656 (vs), 1622 (m), 1580 (s), 1563 (s), 1529 (w-m), and 1504 cm^{-1} (s) (Howard et al., 1964; Miles & Frazier, 1982). These can be reliably assigned in order to the protonated amidinium ring system of C, the carbonyl stretching vibration of G [this is the only band in the spectrum that shows an isotopic shift when the complex is prepared with 6-^{18}O-5′GMP; cf. Miles (1968)], the carbonyl band of C, ring vibration of C, two ring vibrations of G, and two ring vibrations of C.

The Watson–Crick double helix dG·dC has spectra with bands at 1688, 1653, 1635 (sh), 1584, 1568 (sh), 1529, and 1504 cm^{-1} (Figure 9). The first two bands are assigned to coupled G and C carbonyl vibrations [cf. Howard et al. (1969)], the next to a C ring vibration, the next two to G ring vibrations, and the last two to C ring vibrations (Howard et al., 1964, 1969; Miles & Frazier, 1982). Low-frequency IR spectra show that the dG·dC helix has B-form geometry under these conditions (total $[Na^+]$ 0.32 M, 0.01 M phosphate buffer, pH 7.5; Howard et al., 1992).

The double helix dA·dT has well-resolved bands at 1696 (s) 1664 (s), 1643 (m), and 1623 cm^{-1} (w-m) (Figure 10). The first three are from the thymidine residues and the last from an adenine ring vibration. By analogy with uridine (Miles, 1964) and with 1-methyluracil, for which extensive data from isotopic substitution are available (Lewis et al., 1984), the 1696-cm^{-1} band is assigned to the 2-carbonyl stretching vibration, that at 1664 cm^{-1} to the 4-carbonyl mixed with ring double-bond vibrations, and that at 1643 cm^{-1} to a T ring vibration with some 4CO character. The triple helix dT·dA·dT has resolved bands at 1696 and 1658 cm^{-1} and a shoulder at 1638 cm^{-1} (Figure 10), and the same assignments apply.

The spectrum of the Watson–Crick helix Pu20·Py20 (Figure 9) can be readily interpreted in terms of the spectra cited above. The strong band at 1696 cm^{-1} (sh) is assigned to the C2 carbonyl stretch of T. The strong band at 1678 cm^{-1} occurs at the mean of the relevant T (1664 cm^{-1}) and G (1687 cm^{-1}) carbonyl vibrations and is the sum of these bands. The band at 1650 cm^{-1} is due to the coupled C carbonyl stretching vibration, the band at 1622 cm^{-1} to overlapped A and C ring vibrations, those at 1580 and 1563 cm^{-1} to G ring vibrations, and those at 1523 and 1500 cm^{-1} to C ring vibrations.

Returning to the spectrum of the Hoogsteen duplex (Figure 9) and using the assignments outlined above, we assign the strong band at 1701 cm^{-1} to overlapping contributions of the protonated cytosine ring vibration on the high-frequency side and the C2 carbonyl of T on the low-frequency side of the band envelope. The frequency of the band maximum at 1701 cm^{-1} is equal to the mean of indicated strong C^+ and T bands. Present under this band is also a G band of moderate intensity near 1687 cm^{-1}, presumably reflected in the skewed contour of the low-frequency side of the band envelope. The very strong band at 1656 cm^{-1} is a composite of the C carbonyl stretch at about this frequency and the T carbonyl band on the high-frequency side of the band envelope. We note that the symmetrical relationship of the hydrogen-bonded G and C carbonyls present in a WC pair is absent here (cf. Figure 1), evidently eliminating the coupling that forces the carbonyl stretching frequency of G to higher and of C to lower frequencies in Watson–Crick helices [cf. Howard et al. (1969)]. An unresolved T ring vibration with some carbonyl character is presumably present under the band envelope at about 1635 cm^{-1}, and a band at 1625 cm^{-1} is assigned to unresolved A and C ring vibrations, the larger contribution being that of A. The two G ring vibrations are observed at the usual frequencies of 1580 and 1568 cm^{-1} (sh). These bands, like the A and C ring vibrations, are unusual in having much higher relative intensities than in either WC double helices or triple helices. The bands at 1529 and 1504 cm^{-1} are due exclusively to C ring vibrations.

We consider now the anomalously high intensity of the A, C, and G ring vibrations noted in the previous paragraph. One of the most striking changes in infrared spectra of nucleic acids on going from single strands to WC double helices and to triple helices is the large reduction in intensity of the purine ring vibrations [cf. Miles (1971)]. The Hoogsteen duplex provides the first opportunity to observe these bands when the purines are unperturbed by hydrogen bonding to the six-membered ring of the purines. The unusually high intensity of these bands in the new duplex is seen by comparison of the spectra of the parallel (H) and antiparallel (WC) duplexes in Figure 9. This result suggests that hydrogen bonding to the six-membered ring of the purines may be responsible for most of the reduction in intensity of the ring vibrations in WC duplexes and in triple helices. This, in turn, suggests that this hydrogen bonding affects the change of dipole moment during the vibrations responsible for these absorption bands.

Conformationally Sensitive Marker Bands between 800 and 1000 cm^{-1}. Both infrared and Raman spectra of nucleic acids have bands in this region at approximately 812 and 830 cm^{-1} which are correlated with the sugar pucker and backbone conformation of the helix [cf. Tsuboi (1969), Higuchi et al. (1969), Ehrfurt et al. (1972), La Fleur et al. (1972), Brahms et al. (1974), Liquier et al. (1991), Howard et al. (1992)]. The former is characteristic of A-form helical geometry and sugar pucker in the C3′-endo family and the latter of B-form geometry and sugar pucker in the C2′-endo family. Of the additional conformationally sensitive infrared bands in this region [cf. Tsuboi (1969)], we have emphasized those at about 865 (A-form, C3′-endo) and 970 cm^{-1} (B-form, C2′-endo) for the practical reason that they are by far the strongest bands below 1000 cm^{-1}, the most easily observed, and the least variable (Howard et al., 1992).

The spectrum in the 800–1000-cm^{-1} region of the parallel duplex with Hoogsteen pairing (Pu20·RPy20) is shown in Figure 11, together with those of the related triple helix (Py20·Pu20·RPy20) and the antiparallel double helix with Watson–Crick pairing. In each case, the presence of bands

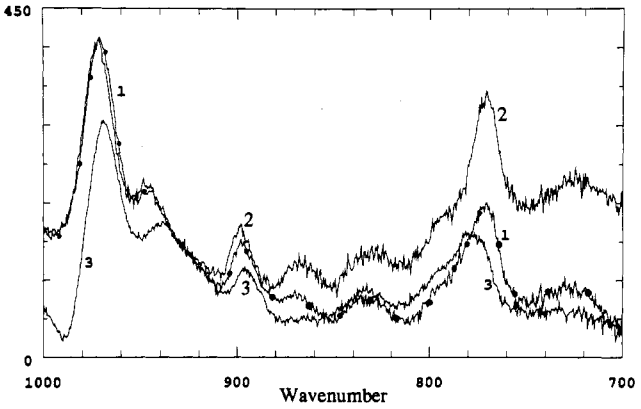


FIGURE 11: IR spectra in low-frequency region. Spectrum 1, Pu20·RPy20 (H), 0.005 M sodium phosphate, pH 5.0, 0.1 M NaCl, 32 °C; spectrum 2, Py20·Pu20·RPy20 (triple helix), 0.005 M sodium phosphate, pH 4.9, 0.1 M NaCl, 31 °C; 3: Pu20·Py20 (WC), 0.02 M sodium phosphate, pH 7.5, 0.1 M NaCl, 33 °C.

at 832 and 970 cm^{-1} and absence of a band near 812 cm^{-1} and of a strong band at 865 cm^{-1} establish that they all have B-form geometry and sugar pucker in the C2′-endo family. Moreover, while spectra of a number of double helices show appreciable variation in frequency and intensity of the bands in this region [cf. Howard et al. (1992)], the parallel duplex and the corresponding triple helix have nearly identical spectra, suggesting that they have the same backbone conformation.

The close similarity of IR spectra in the carbonyl region of the triple helix and the parallel duplex with Hoogsteen pairing has been noted above and explained by assignments of vibrations of component bases. It is important to note that the spectrum of the Watson–Crick-paired portion of the triple helix appears to have little effect on the spectrum of the triple helix. We believe the explanation lies in the structures and symmetries of the three helices [cf. Howard et al. (1992), and Raghunathan et al. (1993)]. In the first place, the triple helix and the independent parallel duplex have coincident helix axes. The parallel duplex and the Hoogsteen-paired strands of the triplex in the model are isostructural. The independent Watson–Crick helix, however, from which the spectrum in Figure 9 was obtained, has a quite different position of the helix axis (D = 0–1 rather than 2.5 Å for the triplex) and number of residues per turn (10) and, therefore, is not isostructural with either of the other helices. Moreover, the two pyrimidine strands in the triple helix are related by a dyad axis, exact for Ts and a pseudodyad for C and C^+, thus accounting for the fact that the Watson–Crick-paired pyrimidine strand in the triple helix does not significantly alter the carbonyl region of the vibrational spectrum. Without these symmetry relationships and with an intrinsically less constrained structure, an independent Watson–Crick helix is relatively free to exhibit spectroscopic divergence from the other helices.

Conformation of A and B Families. The novel parallel duplex described here and our earlier studies on triple helices (Howard et al., 1992; Raghunathan et al., 1993) suggest a rexamination of the current definitions of A and B conformations. Originally these definitions were based on X-ray diffraction patterns of DNA fibers at low and high humidity, respectively (Franklin and Gosling, 1953). As more detailed information became available from spectroscopy, single-crystal X-ray diffraction, and molecular modeling, the definitions were expanded to include more specific structural parameters [for a review, see Saenger (1984)].

In recent years [cf. Saenger (1984)], a number of defining characteristics have been used for A and B forms, including sugar pucker, tilt of the bases, rise per residue (h), displacement of base pairs from the helix axis (D), residues per turn (n), and the torsional angle about O3′-P (t or g^-). Other structural features such as dimensions of the helical grooves, though important and different in the A and B forms, are not independent but are determined by the foregoing parameters. Values of n from 8 to 10 and D from −2 to +1 have usually been considered characteristic of the B family, though our recent studies show that the triple helix dT·dA·dT has larger values of both n and D [Howard et al. 1992; Raghunathan et al., 1993; cf. theoretical calculations of Zhurkin et al. (1978)]. Further, we also find that the parallel double helix with Hoogsteen pairing has a value for n greater than 10, though it is clearly B-form (G. Raghunathan, H. T. Miles, and V. Sasisekharan, manuscript in preparation). The parameter D is not defined for this structure. In antiparallel duplexes, D has an overlapping middle range of values in both A and B families. Thus, it appears that n and D are not essential defining characteristics of the A and B families.

In view of the foregoing discussion, it appears to us desirable to formulate a set of parameters that are necessary and sufficient to place a right-handed RNA or DNA helix in one or the other of these two families. We suggest that the three parameters reflecting sugar pucker (C2′-endo family for B and C3′-endo family for A), the torsional angle about the O3′–P bond (t for B and g^- for A), and the tilt of the bases (or equivalently the rise per residue) meet the criteria of being necessary and sufficient for assignment of regular helical structures to one of the two families. The first two parameters deal only with the backbone and the last with bases, but all are correlated (Sasisekharan & Pattabiraman 1978; Sasisekharan, 1983).

REFERENCES

Antao, V. P., Gray, D. M., & Ratliff, R. L. (1988) *Nucleic Acids Res. 16*, 719–738.

Arnott, S. (1970) *Prog. Biophys. Mol. Biol. 21*, 153–163.

Blake, R. D., & Fresco, J. R. (1966) *J. Mol. Biol. 19*, 145–160.

Brahms, S., Brahms, J., & Pilet, J. (1974) *Isr. J. Chem. 12*, 153–163.

Donohue, J. (1956) *Proc. Natl. Acad. Sci. U.S.A. 42*, 60–65.

Ehrfurth, S. C., Kiser, E. J., & Petiocolas, W. L. (1972) *Proc. Natl. Acad. Sci. U.S.A. 69*, 938–941.

Felsenfeld, G., Davies, D. R., & Rich, A. (1957) *J. Am. Chem. Soc. 79*, 2023–2024.

Franklin, R. E., & Gosling, R. G. (1953) *Nature 172*, 156–157.

German, M. W., Pon, R. T., & van de Sande, J. H. (1987) *Anal. Biochem. 165*, 399–405.

Govil, G., Fisk, C. L., Howard, F. B., & Miles, H. T. (1981) *Biopolymers 20*, 573–603.

Gray, D. M., & Bollum, F. J. (1974) *Biopolymers 13*, 2087–2102.

Hattori, M., Ikehara, M., & Miles, H. T. (1974) *Biochemistry 13*, 2754–2761.

Hattori, M., Frazier, J., & Miles, H. T. (1976) *Biopolymers 15*, 523–531.

Higuchi, S., Tsuboi, M., & Iitaka, Y. (1969) *Biopolymers 7*, 909–916.

Hoogsteen, K. (1963) *Acta Crystallogr. 16*, 907–916.

Howard, F. B., & Miles, H. T. (1984) *Biochemistry 23*, 6723–6732.

Howard, F. B., Frazier, J., Lipsett, M. N., & Miles, H. T. (1964) *Biophys. Res. Commun. 14*, 21–28.

Howard, F. B., Frazier, J., & Miles, H. T. (1969) *Proc. Natl. Acad. Sci. U.S.A. 64*, 451–458.

Howard, F. B., Hattori, M., Frazier, J., & Miles, H. T. (1977) *Biochemistry 16*, 4637–4646.

Howard, F. B., Miles, H. T., Liu, K., Frazier, J., Raghunathan, G., & Sasisekharan, V. (1992) *Biochemistry 31*, 10671–10677.
Ikehara, M., Hattori, M., & Fukui, T. (1972) *Eur. J. Biochem. 31*, 329–334.
Inman, R. B., & Baldwin, R. L. (1964) *J. Mol. Biol. 8*, 452–469.
Inman, R. B. (1964) *J. Mol. Biol. 9*, 624–637.
Ishikawa, F., Frazier, J., Howard, F. B., & Miles, H. T. (1972) *J. Mol. Biol. 70*, 475–490.
La Fleur, L., Rice, J., & Thomas, G. J., Jr. (1972) *Biopolymers 11*, 2423–2437.
Lewis, T. P., Miles, H. T., & Becker, E. D. (1984) *J. Phys. Chem. 88*, 3253–3260.
Liquier, J., Coffinier, P., Firon, M., & Thaillandier, E. (1991) *Biomol. Struct. Dyn. 9*, 437–445.
Lomant, A. J., & Fresco, J. R. (1975) *Prog. Nucleic Acid Res. Mol. Biol. 15*, 185–218.
Miles, H. T. (1968) *Methods Enzymol. 12B*, 256–267.
Miles, H. T. (1971) *Proced. Nucleic Acid Res. 2*, 205–232.
Miles, H. T. (1964) *Proc. Natl. Acad. Sci. U.S.A. 51*, 1104–1109.
Miles, H. T., & Frazier, J. (1964) *Biochem. Biophys. Res. Commun. 14*, 21–28.
Miles, H. T., & Frazier, J. (1982) *J. Mol. Biol. 162*, 219–230.
Muraoka, M., Miles, H. T., & Howard, F. B. (1980) *Biochemistry 19*, 2429–2439.
Pattabiraman, N. (1986) *Biopolymers 25*, 1603–1606.
Powell, J. I., Fico, R., Jennings, W. H., O'Bryan, E. R., & Schultz, A. R. (1980) *Proc. IEEE Comput. Soc. Int. Conf. 21st*, 185–190.
Raghunathan, G., Miles, H. T., & Sasisekharan, V. (1993) *Biochemistry 32*, 455–462.
Riley, M., Maling, B., & Chamberlin, M. J. (1966) *J. Mol. Biol. 20*, 359–389.
Saenger, W. (1984) in *Principles of Nucleic Acid Structure*, Springer-Verlag, Berlin.
Sasisekharan, V. (1983) *Cold Spring Harbor Symp. Quant. Biol. 47*, 45–52.
Sasisekharan, V., & Pattabiraman, N. (1978) *Nature 275*, 159–162.
Tsuboi, M. (1969) *Appl. Spectrosc. Rev. 3*, 54–55.
van de Sande, J. H., Ramsing, N. B., German, N. W., Elhorst, W., Kalisch, B. W., Kitzing, V. E., Pon, R. T., Clegg, R. C., & Jovin, T. M. (1988) *Science 241*, 551–557.
Wilkins, M. H. F., Arnott, S., Mavin, D. A., & Hamilton, L. D. (1970) *Science 167*, 1693–1694.
Zhurkin, V. B., Lysor, Y. P., & Ivanov, J. I. (1978) *Biopolymers 17*, 337–412.