Biochemistry

 Volume 33, Number 38 September 27, 1994

Perspectives in Biochemistry

DNA Triplexes: Solution Structures, Hydration Sites, Energetics, Interactions, and Function[†,Δ]

Ishwar Radhakrishnan[‡] and Dinshaw J. Patel[*,‡,§]

Department of Biochemistry and Molecular Biophysics, College of Physicians and Surgeons, Columbia University, New York, New York 10032, and Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, New York, New York 10021

Received July 6, 1994

Historical Perspective. The formation of a specific complex between two strands of poly(uridylic acid) with one strand of poly(adenylic acid) in the presence of divalent cations was first described by Felsenfeld *et al.* (1957). Subsequently, triple-stranded complexes containing other combinations of polynucleotide strands were described including those belonging to the deoxyribose series [reviewed in Felsenfeld and Miles (1967), Wells *et al.* (1988), Cheng and Pettitt (1992b), and Sun and Hélène (1993)]. These studies culminated in low-resolution, X-ray fiber-diffraction models for the $(A)_n$·$2(U)_n$, $(A)_n$·$2(I)_n$, and $(A)_n$·$2(T)_n$ combinations (Arnott & Selsing, 1974, and references cited therein). Further, a potential biological function for these structures was identified when they were found to inhibit RNA[1] polymerase-mediated transcription *in vitro* (Morgan & Wells, 1968).

Interest in triplexes revived following the discovery of single-strand-specific S1 endonuclease hypersensitive sites in the upstream regions of several eukaryotic genes (Larsen & Weintraub, 1982). These sites were mapped to $(R)_n$·$(Y)_n$ sequences which, when subcloned into supercoiled plasmids, exhibited the same sensitivity toward S1 nuclease [reviewed in Wells *et al.* (1988) and Htun and Dahlberg (1989)]. After a brief controversy, the observed hypersensitivity was assigned to the single-stranded regions arising from a duplex → intramolecular triplex (+ single strand) transition (Lee *et al.*, 1984; Christophe *et al.*, 1985; Lyamichev *et al.*, 1986).

Site-specific binding of an oligonucleotide to the major groove of a DNA duplex via intermolecular triple helix formation was demonstrated by Moser and Dervan (1987). Simultaneously, Hélène and co-workers reported the formation of a triple-stranded complex between an α-oligonucleotide and its cognate DNA duplex sequence (Le Doan *et al.*, 1987). Both studies demonstrated a new strategy for recognition of specific double-stranded DNA sequences with far-reaching implications in the field of genetics, biochemistry, and medicine.

Triplex-Based Applications. The power of triplex-based approaches lies mainly in the extreme specificity of the interaction between third strand oligonucleotides and their cognate duplex DNA. Equipping third strand oligonucleotides with cleaving agents such as Fe(II)-EDTA (Moser & Dervan, 1987) or copper(II) phenanthroline (François *et al.*, 1989), for instance, allows them to be used as "rare-cutting" artificial endonucleases. A vivid demonstration of this approach is provided by the cleavage (with more than 80% yield) of a unique site 16 base pairs in length within 10 gigabase pairs of genomic DNA (Strobel & Dervan, 1991; Strobel *et al.*, 1991). This result has implications in genome mapping because it may facilitate the dissection of chromosomes into megabase-sized fragments which can then be analyzed separately.

The therapeutic implications of site-specific triplex formation are also of interest. Several studies have shown that triplex-forming oligonucleotides when targeted to the promoter

[†] This work was supported under NIH Grant GM 34504 to D.J.P.

[Δ] Coordinates for the Y·RY triplexes containing G·TA (Accession No. 149D) and T·CG (Accession No. 177D) triples and the R·RY triplex (Accession Nos. 134D, 135D, 136D) have been deposited with the Protein Data Bank, Brookhaven National Laboratory, Upton, NY 11973, from whom copies can be obtained.

* Address correspondence to this author at Memorial Sloan-Kettering Cancer Center.

[‡] Columbia University.

[§] Memorial Sloan-Kettering Cancer Center.

[1] Abbreviations: DNA, deoxyribonucleic acid; RNA, ribonucleic acid; G, guanine; A, adenine; T, thymine; C, cytosine; I, inosine; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; RMD, restrained molecular dynamics.

0006-2960/94/0433-11405$04.50/0

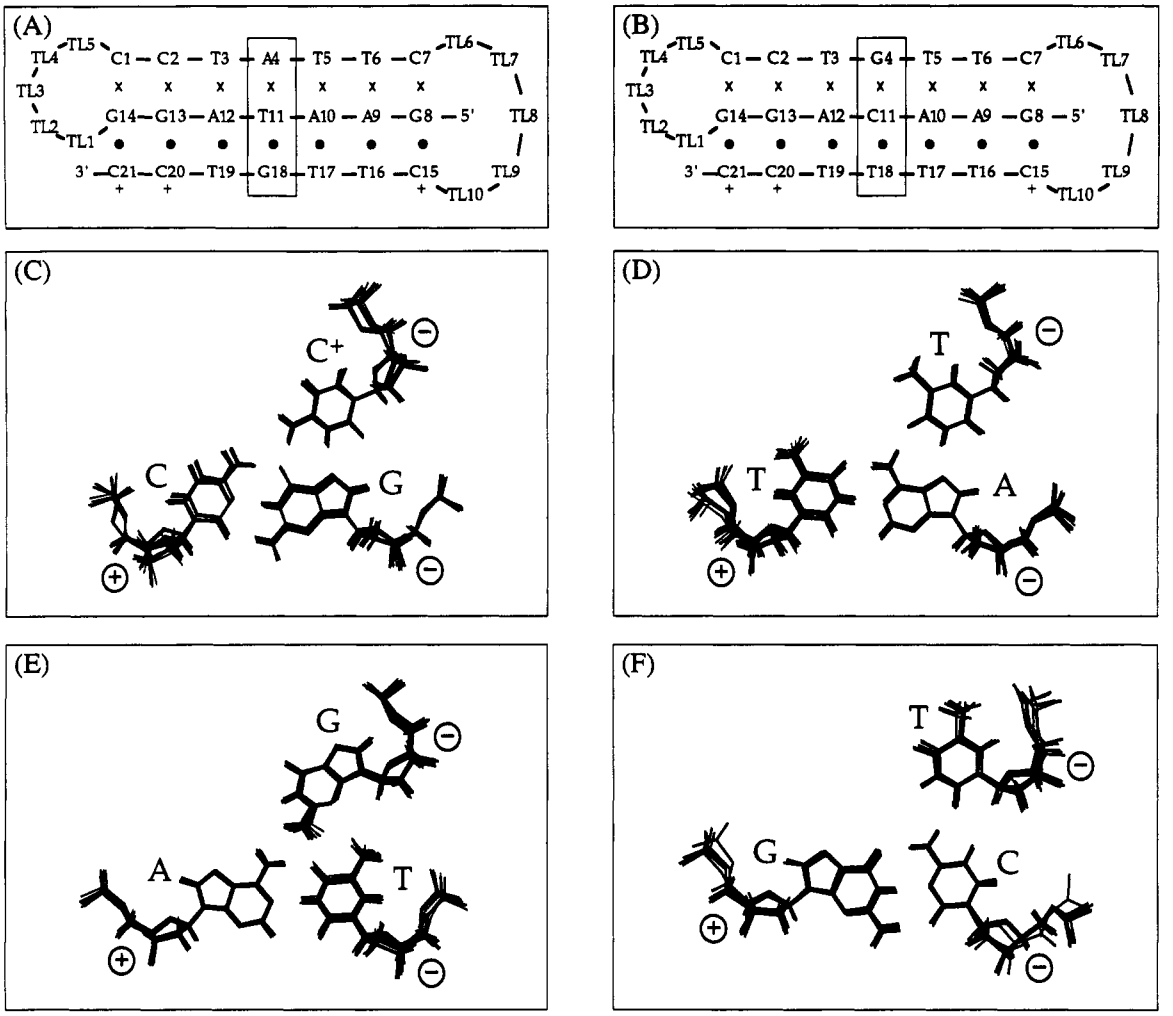

FIGURE 1: Oligonucleotide sequences of intramolecular Y·RY DNA triplexes containing (A) a G·TA triple and (B) a T·CG triple that were employed for NMR studies. Pairing alignments of (C) C⁺·GC, (D) T·AT, (E) G·TA, and (F) T·CG triples were taken from the solution structures of Y·RY triplexes (Radhakrishnan & Patel, 1994a,c). The (+) and (–) indicate the relative strand orientations.

and coding regions of specific genes can alter the expression patterns presumably by inhibiting the initiation and elongation of mRNA synthesis by RNA polymerase (Cooney *et al.*, 1988; Maher *et al.*, 1990, 1992; Durland *et al.*, 1991; Grigoriev *et al.*, 1992; Young *et al.*, 1991; Duval-Valentin *et al.*, 1992). This opens the possibility of selectively controlling the expression of specific genes, which could be of great value in the treatment of genetic disorders.

Considerable progress in understanding the structure and energetics of triple-stranded structures has been achieved in recent years. In this review, we shall endeavor to reconcile the conclusions drawn from a variety of biochemical and biophysical studies, within a broad structural framework.

Solution Structure Determination

Structural Motif Definitions. The current definition of a triple helix in nucleic acids is similar in spirit to what was originally proposed (Felsenfeld *et al.*, 1957). It results from specific, Hoogsteen or reversed-Hoogsteen-type hydrogen-bonding interactions (Hoogsteen, 1963) between bases in a homopurine strand of a Watson–Crick duplex and an additional oligonucleotide strand (henceforth designated strand III; for brevity, the Watson–Crick paired homopyrimidine and homopurine strands will be designated strands I and II, respectively).

At least two structural motifs in DNA triplexes have been characterized. In the so-called "pyrimidine motif" or pyrimidine–purine–pyrimidine (Y·RY) triplexes, a pyrimidine-rich third strand is aligned in parallel with the purine strand of the Watson–Crick duplex to form T·AT and C⁺·GC base triple combinations.[2] In the "purine motif" or purine–purine–pyrimidine (R·RY) triplexes, a purine-rich third strand is aligned antiparallel to the purine strand, yielding G·GC, A·AT, and T·AT base triple combinations.

Experimental NMR Design. In recent years, solution-state nuclear magnetic resonance (NMR) spectroscopy has provided structural information on nucleic acids at levels comparable to those obtained routinely from X-ray crystallographic analyses. This inspired analogous NMR approaches on single-stranded oligonucleotide sequences that form intramolecular triple-stranded structures with defined strand orientations (Häner & Dervan, 1990; Sklenár & Feigon, 1990; Radhakrishnan *et al.*, 1991a). In order to optimize the thermal stability of these triplexes without substantially increasing the complexity of NMR spectra, the stem length and loop sizes were generally restricted to seven base triples and five thymine residues, respectively (Radhakrishnan *et al.*, 1991a,b). Y·RY triplexes containing either a G·TA or a T·CG triple (see below) in addition to canonical T·AT and C⁺·GC triples (Figure 1) were characterized with Na^+ as the counterion at high resolution (Radhakrishnan & Patel, 1994a,c). The incor-

[2] In this notation, the third strand base is written first and the dot (·) signifies the Hoogsteen-type interaction.

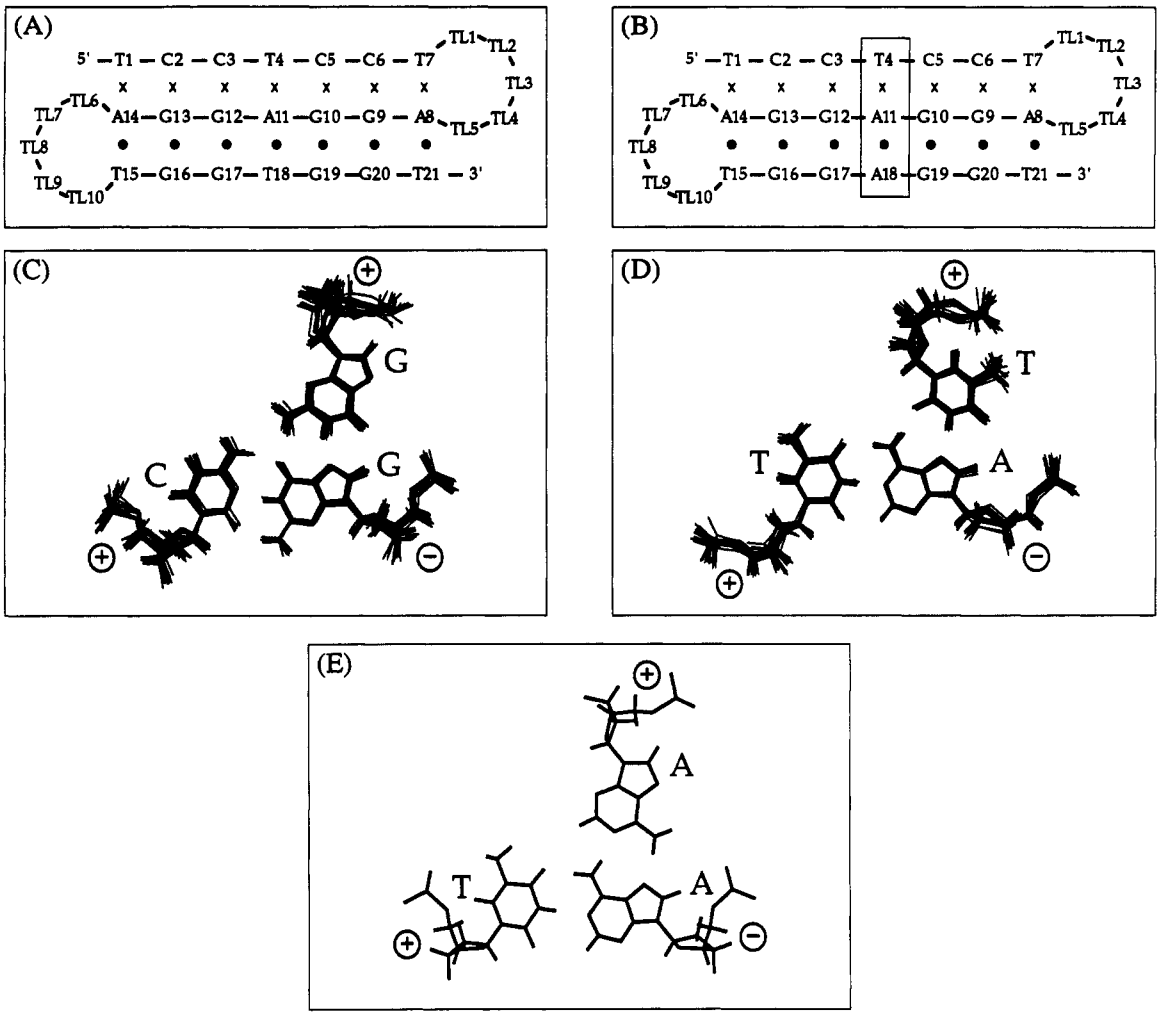

FIGURE 2: Oligonucleotide sequences of intramolecular R·RY DNA triplexes containing (A) G·GC and T·AT triples and (B) G·GC, T·AT, and A·AT triples that were employed for NMR studies. Pairing alignments of (C) G·GC and (D) T·AT triples were taken from the solution structures of an R·RY triplex (Radhakrishnan & Patel, 1993). (E) Pairing alignment of A·AT triples in R·RY triplexes was taken from a hand-built model that is consistent with experimental data (Radhakrishnan *et al.*, 1993b). The (+) and (–) indicate the relative strand orientations.

poration of noncanonical triples resulted in increased dispersion of the NMR signals, allowing a detailed exploration of the conformational features of Y·RY triplexes in general and the unusual triples in particular. Likewise, the nonisomorphic nature of G·GC and T·AT triples (see below) permitted the structure of an R·RY triplex (Figure 2) to be determined at high resolution (Radhakrishnan & Patel, 1993). The tendency of G-rich sequences to form higher order aggregates in the presence of Na^+ counterions warranted the use of Li^+ as the counterion for R·RY triplexes. The high molecular weight (*ca.* 10 000) of such oligonucleotide sequences sometimes necessitated homonuclear three-dimensional NMR approaches in D_2O solution to resolve ambiguities and maximize the spectral information content involving nonexchangeable protons (Radhakrishnan *et al.*, 1992; Radhakrishnan & Patel, 1994a,c).

Structure Determinations. Structures were determined by incorporating the NMR data initially in the form of distance restraints (~16–19 exchangeable and nonexchangeable proton restraints per residue per structure), with the distance restraints involving nonexchangeable protons subsequently replaced by intensity restraints (~12–14 nonexchangeable proton restraints per residue per mixing time per structure) in restrained molecular dynamics (RMD) simulations. These calculations were conducted either in the presence of explicit solvent and counterions as in the case of Y·RY triplexes (Radhakrishnan & Patel, 1994a,c) or *in vacuo* with a distance-dependent dielectric to mimic the effect of solvent as in the case of the R·RY triplex (Radhakrishnan & Patel, 1993). Conformational searches were conducted at high temperature (400 K) typically for 10 or 20 ps and at room temperature for about 10 ps. Convergence to morphologically similar structures with atomic root-mean-square deviations of less than 1 Å (measure of precision) was usually achieved from dissimilar starting structures and initial conditions. The low *R*-factors (measure of accuracy) and consistency with nonquantifiable NMR data (*e.g.*, chemical shifts) confirmed the validity of the final structures.

Y·RY Triplexes

Base Triple Combinations, Pairing Alignments, and Strand Orientation. Analogous to U·AU and C^+·GC combinations in ribonucleotides (Felsenfeld *et al.*, 1957; Lipsett, 1964), T·AT and C^+·GC combinations in deoxyribonucleotides (Riley *et al.*, 1966; Lee *et al.*, 1979) were identified by early studies on triplex polymers. Recent studies have revealed at least two additional base triple combinations within this motif, including G·TA (Griffin & Dervan, 1989) and T·CG (Yoon *et al.*, 1992). These triples differ from canonical T·AT and C^+·GC triples in that they result from specific interactions between third strand bases and *pyrimidine* residues in a purine-rich strand.

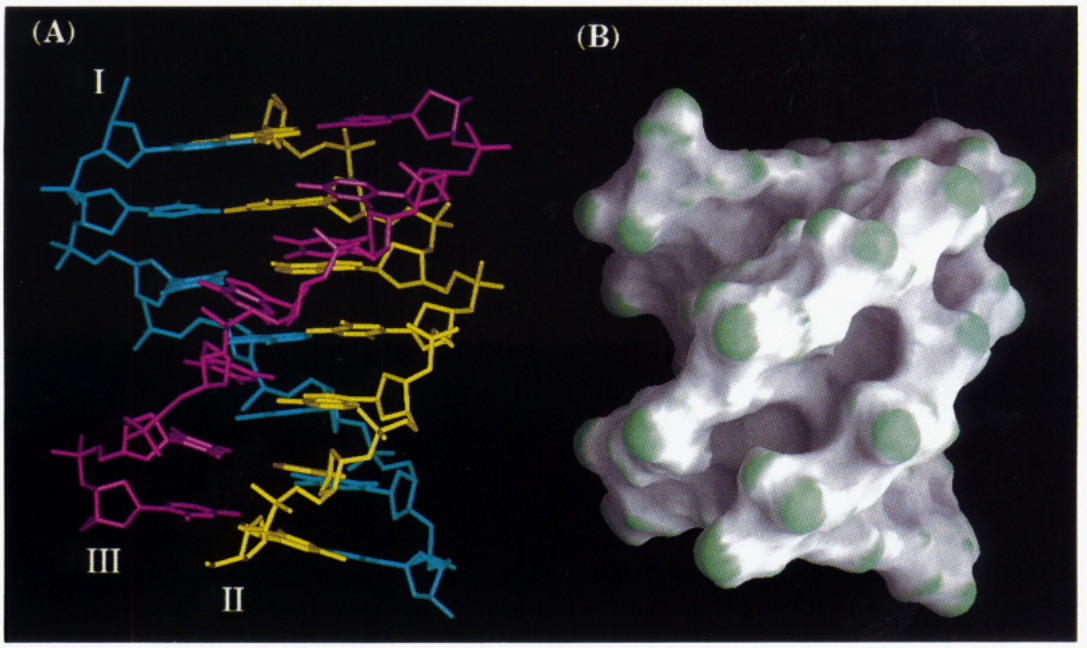

FIGURE 3: (A) Triple-stranded region in the solution structure of a Y·RY triplex containing a G·TA triple (Figure 1A). Labels I, II, and III identify the 5′-ends of strands I, II, and III, respectively. (B) Molecular surface view of the same triplex emphasizing the nature of the Crick–Hoogsteen groove. The molecular surface was mapped using a spherical probe of radius 1.4 Å in the GRASP program. The most convex, concave, and planar regions of the surface are coded green, gray, and white, respectively; colors are linearly interpolated for intermediate values.

A parallel orientation for the third strand relative to the purine strand in Y·RY triplexes was established by affinity cleavage and chemical modification experiments in intermolecular (Moser & Dervan, 1987) and in intramolecular triplexes [reviewed in Wells *et al.* (1988)], respectively. The same result was obtained for triplexes containing third strand α-oligodeoxyribonucleotides (Le Doan *et al.*, 1987). Direct evidence for the pairing alignments in T·AT and C^+·GC triples was first provided by high-resolution NMR spectra of intermolecular triplexes (De los Santos *et al.*, 1989; Rajagopal & Feigon, 1989). Both T·AT and C^+·GC triples were found to be stabilized by two Hoogsteen hydrogen bonds between bases on strands II and III without affecting the Watson–Crick interactions between bases on strands I and II (Figure 1C,D).

Global Structural Features. Detailed insight into the conformational features of Y·RY triplexes has been provided by solution structures (Radhakrishnan & Patel, 1994a,c) that were determined using restraints derived from NMR data (Radhakrishnan *et al.*, 1991a, 1992) for the sequences shown in panels A and B of Figure 1. A representative structure determined for the triplex containing a noncanonical G·TA triple is shown in Figure 3A. The structure calculations, undertaken using RMD methods in the presence of explicit solvent and counterions, were greatly aided by the high quality of the NMR spectra and the large number of interproton NOEs, particularly those between residues in strands II and III.

Analyses of the helicoidal parameters of these solution structures indicate that the binding of the third strand induces global conformational changes in the Watson–Crick-paired duplex region of the triplex. This is most dramatically illustrated by the *x*-displacement values (~−2 Å) for the duplex region in the two solution structures (Radhakrishnan & Patel, 1994a,c), which is significantly different from what is normally observed for canonical B-form helices in crystal structures (<−1 Å). The negative *x*-displacement values imply that the Watson–Crick base pairs move toward the minor groove, presumably to optimize nonbonded interactions with the incoming third strand in the major groove. These changes are accompanied by a slight unwinding of the helix, which is reflected in the average value for the helical twist in these structures (~31°). The glycosidic torsion angles for the various residues in the solution structures are mostly restricted to the *anti* range while most of the sugars adopt S-type (in the C2′-*endo* range) conformations not unlike those found in B-form helices. However, the interstrand and intrastrand overlap patterns exhibited by aromatic heterocycles in the Watson–Crick duplex segment are similar, although not identical, to those found in A-form helices. These results indicate that Y·RY triplexes adopt a conformation that is distinct from both canonical A-form *and* B-form helices (Radhakrishnan & Patel, 1994a,c).

For almost two decades, the fiber-diffraction model proposed by Arnott and co-workers for a $d(T)_n$·$d(A)_n$$d(T)_n$ triplex (Arnott & Selsing, 1974) served as the working model for studies of Y·RY triplexes. This model shares structural features with A-form helices including N-type (C3′-*endo*) sugars and large negative *x*-displacement (−3.2 Å) values. The NOE patterns (Radhakrishnan *et al.*, 1991a, 1992; Wang *et al.*, 1992), the measured values for the proton–proton scalar coupling constants (Macaya *et al.*, 1992), and the solution structures (Radhakrishnan & Patel, 1994a,c) are inconsistent with this assumption of an N-type conformation in Y·RY triplexes. Marker bands and lines characteristic of S-type sugar conformation have also been observed in the infrared and Raman vibrational spectra of DNA triplexes (Thomas & Peticolas, 1991; Howard *et al.*, 1992; Ouali *et al.*, 1993). Deviations from Arnott's model have also been reported for structures of Y·RY triplexes probed by molecular dynamics simulations (Laughton & Neidle, 1992a) and molecular modeling studies (Raghunathan *et al.*, 1993).

Structural Perturbations at Noncanonical Triples. Besides canonical triples T·AT and C^+·GC, at least two additional base triple combinations, namely, G·TA and T·CG, are supported within the Y·RY motif. These noncanonical triples are interesting in two respects: (i) they extend the Y·RY triplex "code" to all four naturally occurring base pairs, which

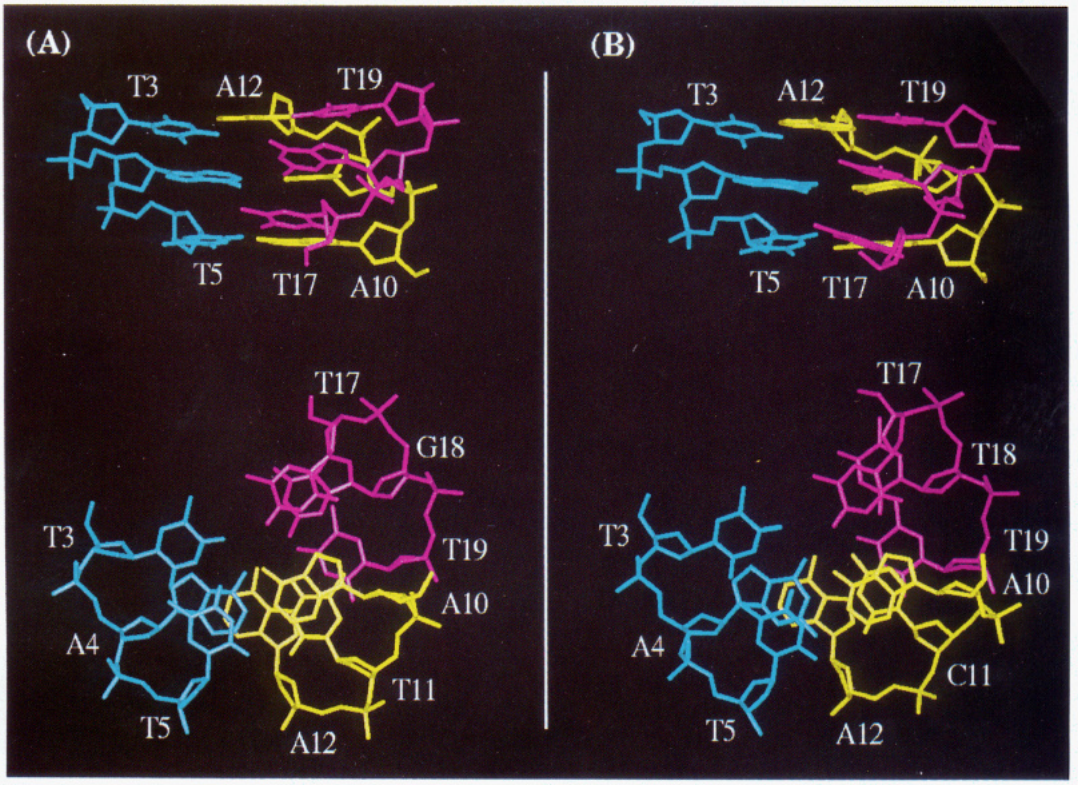

FIGURE 4: Side-by-side views of (A) the (T3-A4-T5)·(A10-T11-A12)·(T17-G18-T19) segment containing the G·TA triple and (B) the (T3-G4-T5)·(A10-C11-A12)·(T17-T18-T19) segment containing the T·CG triple taken from the solution structures of the corresponding Y·RY triplexes (panels A and B of Figure 1, respectively). These views are approximately perpendicular (top panels) and parallel (bottom panels) to the helix axis.

has implications in various triplex-based applications, and (ii) they bear little resemblance with the canonical triples and thus are expected to provide new insights into molecular recognition. This prediction is borne out in the solution structures of Y·RY triplexes containing G·TA (Figure 1A; Radhakrishnan & Patel, 1994a) and T·CG triples (Figure 1B; Radhakrishnan & Patel, 1994c).

The guanine in the G·TA triple is associated with two hydrogen-bonding interactions, each involving a different amino proton (Radhakrishnan & Patel, 1994a). Covalent constraints presumably dictate the choice of the amino proton used for the hydrogen-bonding interaction with the thymine residue of the same triple (Figure 1E). Steric factors, including a potential clash with the bulky methyl group of the thymine residue, force the guanine base out of the average plane formed by the TA base pair, leading to a second, albeit weak, hydrogen-bonding interaction between the other amino proton and a thymine residue of an adjacent triple (Figure 4A, top). The third strand helix undergoes a dramatic overwinding at the TpG step (twist angle, 72°) and an equally dramatic unwinding at the GpT step (twist angle, –20°), which leads to somewhat "compressed" and "extended" backbone conformations at the respective dinucleotide steps. These changes are accompanied by an extensive overlap between base moieties at the TpG step but a remarkably poor overlap at the GpT step (Figure 4A, bottom). Another noteworthy feature in this segment is the N-type sugar conformation of the guanine residue; other residues in the triplex exhibit mostly S-type sugar conformations.

A single hydrogen bond between thymine and cytosine residues stabilizes the T·CG interaction (Figure 1F) (Radhakrishnan & Patel, 1994c). At the dinucleotide steps involving the thymine residue, the third strand helix undergoes overwinding and unwinding almost to the same extent as in the G·TA-containing Y·RY triplex. The backbone conformations in these regions are qualitatively similar in both triplexes, and not unlike the guanine residue in the G·TA triple, the thymine residue in the T·CG triple also adopts an N-type sugar pucker. This remarkable conservation of structural features is perhaps explained by the fact that both guanine and thymine base moieties utilize similarly located hydrogen-bonding donor and acceptor groups in their interactions with TA and CG base pairs (compare panels E and F of Figure 1). Unlike the guanine in the G·TA triple however, the base moiety of thymine in the T·CG triple approaches a coplanar alignment with the CG base pair (Figure 4B, top) and overlaps only partially with its 5′-linked thymine (Figure 4B, bottom). Structural distortions are not restricted to the third strand of the helix, however. Localized structural perturbations occur near the T·CG triple in the purine strand of the triplex. β and γ backbone torsion angles generally found in the *trans*, +*gauche* ranges in helical regions are found in the +*gauche*, *trans* domains at the ApC step (Figure 1B). In apparent correlation with these changes, an extensive overlap between the adenine and cytosine base moieties is detected at the ApC step.

The effects of incorporating X·GC (X = A, G, T) mismatches within an intramolecular Y·RY triplex have been assessed by NMR spectroscopy (Macaya *et al.*, 1991). From a qualitative analysis of the data, it appears that each of these mismatches are stabilized by one or two hydrogen bonds between bases on strands II and III.

Hydration Sites. The Watson–Crick (strands I and II), Crick–Hoogsteen (strands II and III), and Watson–Hoogsteen (strands I and III) grooves in Y·RY triplexes represent potential sites of interaction with ligands. Among the three grooves, the Crick–Hoogsteen groove is quite narrow (~2–3 Å) and is marked by the presence of pockets that are

approximately the size of phosphate groups (Figure 3B). A remarkable feature is the consistent detection of long-lived water molecules with lifetimes ≥1 ns in this groove (Radhakrishnan & Patel, 1994b,c). It is conceivable that these water molecules form an ordered, interconnected network analogous to those found in several DNA duplexes (Schneider *et al.*, 1993). Such water molecules can contribute to the overall stability not only by solvating the macromolecule but also by screening repulsive electrostatic interactions between phosphate groups that line the narrow Crick–Hoogsteen groove.

Despite its width (>7 Å) and its somewhat hydrophobic character conferred by the methyl groups of thymine residues, the Watson–Hoogsteen groove harbors several long-lived water molecules at low temperatures (Radhakrishnan & Patel, 1994b,c). These water molecules are typically found near methyl groups and may bridge the O^4-carbonyl oxygen atoms of thymine residues within the same triple.

The Watson–Crick groove (width ~6–7 Å) is vastly different from the minor groove of A-form helices but shows some similarities with the minor groove of B-form helices. Hydration sites are detected near adenine H2 protons in this groove, but the NOE patterns are somewhat different from those observed for AT-rich duplex sequences which are generally associated with a "spine" of hydration (Kubinec & Wemmer, 1992; Liepinsh *et al.*, 1992). The hydration patterns in all three grooves appear to be relatively indifferent to changes in the nature of the counterion from Na^+ to Li^+ (Radhakrishnan & Patel, 1994b).

Effects of pH and Temperature on Triplex Stability. Besides intrinsic factors such as length and sequence composition, extrinsic factors such as pH and temperature influence the stability of Y·RY triplexes. In general, the stability of Y·RY triplexes containing cytosine residues in the third strand decreases as the pH of the solution is raised from acidic to alkaline values (Lee *et al.*, 1979; Moser & Dervan, 1987). Substitution of third strand cytosines with novel synthetic heterocycles such as 3-methyl-5-amino-1*H*-pyrazolo[4,3-*d*]-pyrimidin-7-one (designated P1) (Koh & Dervan, 1992) permits formation of Y·RY triplexes at neutral pH. NMR studies of the P1·GC interaction (Radhakrishnan *et al.*, 1993a) suggest that these heterocycles effectively mimic the Hoogsteen hydrogen-bonding interactions of protonated cytosines in Y·RY triplexes.

The thermodynamic stability of several inter- and intramolecular Y·RY triplex oligomers has been assessed via calorimetry (Plum *et al.*, 1990; Manzini *et al.*, 1990; Pilch *et al.*, 1990; Roberts & Crothers, 1991; Völker *et al.*, 1993) and affinity cleavage (Singleton & Dervan, 1992a,b) experiments. From calorimetric measurements conducted over a pH range of 4.5–6.5, the free energy change associated with Hoogsteen base pair formation appears to be less than that associated with Watson–Crick base pair formation. The corresponding enthalpy change has been estimated at –2.0 kcal/mol of base triples formed (Plum *et al.*, 1990) though larger values (–6.6 kcal/mol of base triples) have also been reported (Manzini *et al.*, 1990). Ignoring nearest neighbor effects, the free energy contributions made by C^+·GC and T·AT triples appear to be equivalent (Pilch *et al.*, 1990). The presence of monovalent or divalent cations enhances the thermal stability of triplexes as measured by the change in melting temperature with cation concentration, $\partial(T_m)/\partial(\log[Na^+]) \sim 8$–12 (Plum *et al.*, 1990; Völker *et al.*, 1993). 5-Methylcytosines, when substituted for cytosines in the third strand, enhance the thermal stability of triplexes via an entropic effect (Plum *et al.*, 1990; Xodo *et al.*, 1991).

Single base bulges or single "mismatches" (depending on the type of mismatch) in the third strand destabilize the triplex by 2.5–4.0 kcal/mol (Roberts & Crothers, 1991; Singleton & Dervan 1992a); the destabilization is mainly enthalpic in origin and is in the same range observed for DNA duplexes. The stability of mismatches and noncanonical triples are often subject to positional and flanking sequence effects (Mergny *et al.*, 1991; Kiessling *et al.*, 1992).

R·RY Triplexes

Base Triple Combinations, Pairing Alignments, and Strand Orientation. Early studies exploring possible base triple combinations in this motif utilized optical spectroscopy and ultracentrifugation measurements which led to the identification of G·GC and I·IC base triple combinations (Lipsett, 1964; Inman, 1964; Marck & Thiele, 1978). More recently, however, Fresco and co-workers scanned the full spectrum of base triple combinations using affinity chromatography and identified additional interactions including A·AU/T, I·AU/T, and I·GC triples in homopolymers (Letai *et al.*, 1988). Subsequent studies confirmed the occurrence of G·GC (Kohwi & Kohwi-Shigematsu, 1988; Beal & Dervan, 1991) and A·AT and T·AT (Beal & Dervan, 1991) base triple combinations within the R·RY motif of DNA triplexes, although weaker interactions were detected for C·AT, A·GC, and T·CG combinations (Beal & Dervan, 1992a).

The relative orientation of the purine strands in R·RY triplexes (Cooney *et al.*, 1988) was initially a matter of debate with an antiparallel orientation established by the pattern of protection and cleavage in chemical probing (Kohwi & Kohwi-Shigematsu, 1988) and affinity cleavage experiments (Beal & Dervan, 1991), respectively. In the same study, Beal and Dervan (1991) proposed two plausible models for the pairing alignments in G·GC, T·AT, and A·AT triples. The models differed in the conformation of the glycosidic torsion angles in the third strand and in the relative positions of the various hydrogen-bonding donor and acceptor groups with consequent differences in the locations of the third strand sugar–phosphate backbone. The various pairing alignments proposed by Beal and Dervan (1991) were also analyzed by theoretical approaches (Cheng & Pettitt, 1992a; Laughton & Neidle, 1992b). In these studies, within the constraints of their adopted structural models, comparable energies were computed for the alternative alignments of a given triple. The actual pairing alignments were deduced from NOE and chemical shift patterns in high-resolution proton NMR spectra of R·RY triplexes (Radhakrishnan *et al.*, 1991b, 1993b). All three triples (G·GC, T·AT, and A·AT) are stabilized by two reversed-Hoogsteen-type hydrogen bonds between bases in the third strand and the purine strand of the Watson–Crick duplex with the glycosidic torsion angles restricted to the *anti* domain (Figure 2C–E).

Global and Local Structural Features. The solution structure (Figure 5A) provides detailed insight into the conformational features of an R·RY triplex containing G·GC and T·AT triples (Radhakrishnan & Patel, 1993). The sequence employed for the NMR studies (Figure 2A) yielded high-quality NMR data, and the deduced restraints were incorporated in RMD structure calculations. The large number of interproton NOEs, particularly those between residues on strands II and III, was pivotal in defining the conformation of the molecule.

The triple-stranded region of this intramolecular triplex contains four G·GC and three T·AT triples. It is apparent

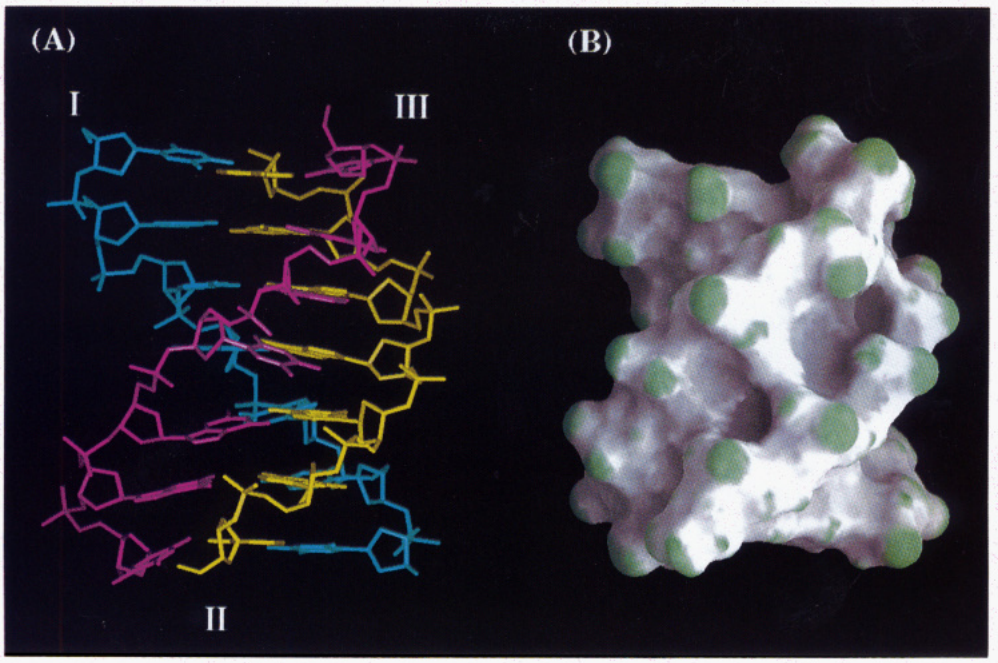

FIGURE 5: (A) Triple-stranded region in the solution structure of an R·RY triplex (Figure 2A). Labels I, II, and III identify the 5′-ends of strands I, II, and III, respectively. (B) Molecular surface view of the same triplex emphasizing the nature of the Crick–Hoogsteen groove. The molecular surface was mapped using a spherical probe of radius 1.4 Å in the GRASP program. The most convex, concave and planar regions of the surface are coded green, gray, and white, respectively; colors are linearly interpolated for intermediate values.

from the pairing alignments shown in Figure 2 that the G·GC, T·AT, and A·AT triples in R·RY triplexes are nonisomorphic (*i.e.*, the C1′ atoms of third strand residues in each of the three triples are located at remarkably different positions when the C1′ atoms of the Watson–Crick base pairs are superimposed). Not unexpectedly, structural perturbations are found at the junctions between G·GC and T·AT triples. Surprisingly, however, these perturbations are mostly confined to the 5′-d(TpG)-3′ steps and, to a lesser extent, at the 5′-d(GpT)-3′ steps in the third strand segment. When compared with the 5′-d(GpG)-3′ steps, the 5′-d(TpG)-3′ steps are underwound and in a somewhat extended conformation (Figure 6A). An opposite trend (albeit to a lesser extent) partially compensating these effects is observed at the 5′-d(GpT)-3′ steps. These features along with what are thought to be high-energy (*gauche*, *trans*) states for the phosphodiester torsion angles at the 5′-d(GpG)-3′ steps suggest a novel conformation for the entire third strand segment. A distinctive feature of this conformation is the poor overlap between neighboring base moieties on the third strand (Figure 6B); instead, the bases generally pack against the sugar moieties of their respective 5′-linked residues as observed for 5′-d(TpG)-3′ (Figure 6C) and 5′-d(GpG)-3′ (Figure 6D) steps. Deviations from S-type sugar puckers for thymine residues and the observation of *high anti*-glycosidic conformations for guanine residues further characterize the unique conformation of the third strand segment.

The Watson–Crick duplex region of the triplex appears to be relatively impervious to perturbations in the third strand segment although the calculated helical twist and *x*-displacement values indicate a conformation that is different from canonical A- and B-form DNA. The base pairs are displaced toward the minor groove (~−1.9 Å), presumably to accommodate the third strand in the major groove. The base-stacking pattern in the duplex segment is reminiscent of A-form DNA. Unlike A-form helices, however, an S-type sugar pucker is retained by most of the residues in the duplex region. The glycosidic torsion angle ranges in the two strands are generally confined to the *anti* domain.

The structural details of an R·RY triplex containing an A·AT triple (Figure 2B) have been assessed qualitatively by NMR studies (Radhakrishnan *et al.*, 1993b). The adenine residue is readily accommodated within the third strand helix and forms two $N^6H\cdots N$ reversed-Hoogsteen-type hydrogen bonds with the adenine residue in the AT base pair (Figure 2E). The hydrogen-bonding interactions of G·GC triples flanking the A·AT triple remain unperturbed. The A·AT pairing alignment determined by NMR is supported by the results obtained from chemical modification experiments (Jayasena & Johnston, 1992). NMR studies of an R·RY triplex containing a T·CG triple (Dittrich *et al.*, 1994) suggest structural perturbations that extend up to two base triples on either side of the T·CG triple. These studies, which are of a qualitative nature to date, also implicate a potential hydrogen-bonding interaction between the thymine and cytosine residues in the T·CG triple.

Hydration Sites. The purine-rich strands in R·RY triplexes define a narrow (~3 Å) but somewhat shallow Crick–Hoogsteen groove (Figure 6B). The detection of long-lived water molecules at several locations in this groove suggests extensive hydration of this groove (Radhakrishnan & Patel, 1994b). The Watson–Hoogsteen groove, despite its width (~8 Å), is also characterized by long-lived water molecules. Remarkably, these water molecules are positioned along the inner edges of this groove, possibly belonging to the esterified oxygen atoms of phosphate groups. These experimental results thus suggest that hydration could be an important determinant of stability of R·RY triplexes. Theoretical studies also predict such a role. For example, a spine of water molecules bridging the amino groups of cytosine and guanine residues in the Watson–Hoogsteen groove was identified in molecular dynamics simulations of an R·RY triplex containing a continuous stretch of G·GC triples (Mohan *et al.*, 1993).

Effects of Sequence Composition and Temperature on Triplex Stablity. Unlike Y·RY triplexes, the stability of R·RY triplexes is relatively independent of pH. A survey of the sequences employed in the studies of R·RY triplexes suggests that guanine-rich sequences may be an implicit requirement

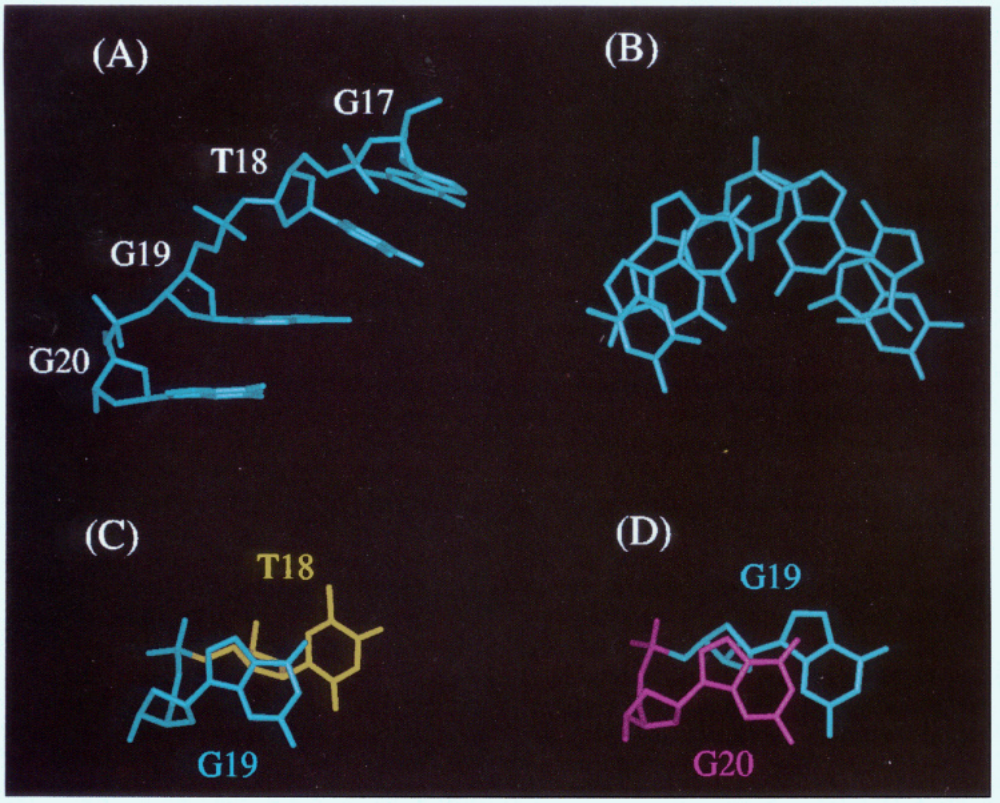

FIGURE 6: (A) View of the G17-T18-G19-G20 third strand segment shown approximately perpendicular to the helix axis in the solution structure of the R·RY triplex (Figure 2A). (B) View of the stacking arrangement between bases in the third strand segment (residues 15–21) in the same structure. The 5′-terminus of this segment is on the right. (C) View down the helix axis for the T18-G19 step. (D) View down the helix axis for the G19-G20 step.

for such triplexes to form. The results from a study comparing relative stabilities of T·AT and A·AT triples within such G·GC-rich triplexes indicate that these triplexes tolerate interruptions by T·AT triples better than A·AT triples (Beal & Dervan, 1991).

A calorimetric study of an intermolecular R·RY triplex containing G·GC and A·AT triples (Pilch *et al.*, 1991) determined the free energy change associated with triplex formation (−2.6 kcal/mol per base triple) to be approximately twice that of duplex formation (−1.3 kcal/mol per base pair), suggesting comparable stabilities for the Watson–Crick and Hoogsteen interactions. This result would predict higher stabilities for R·RY triplexes than for the corresponding Y·RY triplexes (*i.e.*, for the same Watson–Crick duplex sequence).

Common Features of Y·RY and R·RY Triplexes. Despite fundamental differences in pairing alignments, base composition, and orientation of the third strand, Y·RY and R·RY triplexes show a striking similarity in their global structural features especially in the duplex regions. Most remarkably, the third strand sugar–phosphate backbones in both motifs are situated virtually at the same position relative to their respective Watson–Crick paired duplexes (Radhakrishnan & Patel, 1994a). The key distinguishing feature between the two motifs is the relative locations of the base moieties, which is obviously a consequence of the different polarities of the third strand.

The global conformational similarities of purine and pyrimidine motifs might explain the success of alternate strand approaches which make use of both motifs to recognize mixed DNA sequences (see below). The conformational differences relative to canonical duplexes on the other hand probably account for the observed hypersensitivity to chemical agents in the duplex regions located near triplex–duplex junctions (Hartman *et al.*, 1991; Malkov *et al.*, 1993).

Ligand–Triplex Complexes

There are three grooves in triple-helical DNA which can be targeted by ligands that bind either noncovalently or covalently to the base pair edges. Since the third strand aligns with the major groove of the Watson–Crick duplex, the unoccupied minor groove has been the principal target for binding ligands that had been previously targeted to the same groove in duplex DNA. There is also the interesting question as to whether planar chromophores can unwind the triple-helical DNA and intercalate between base triples similar to intercalation into duplex DNA.

Intercalation. Hélène and co-workers have studied the binding of a benzo[*e*]pyridinoindole derivative (BEPI) to Y·RY triplexes in an attempt to establish whether this ligand containing four aromatic rings and two positive charges could intercalate between base triples (Mergny *et al.*, 1992; Pilch *et al.*, 1993b). A combination of spectroscopic, fluorescence, and hydrodynamics measurements on the complex established unequivocally that BEPI intercalates into Y·RY DNA triplexes with intercalation favored between uncharged T·AT triples relative to charged C^+·GC triples. Further, BEPI binds preferentially to Y·RY triplexes relative to their duplex counterparts and can be covalently linked to the DNA triplex following photochemical irradiation of the complex. This group also showed that the sequence-specific inhibition of transcription initiation through triplex formation is strongly enhanced by stabilizing the triplex through formation of the BEPI intercalation complex (Mergny *et al.*, 1992). Two other charged heterocyclic aromatic ring systems, coralyne (Lee *et al.*, 1993) and a naphthyl-substituted quinoline (Wilson *et al.*, 1993), have also been shown recently to intercalate into

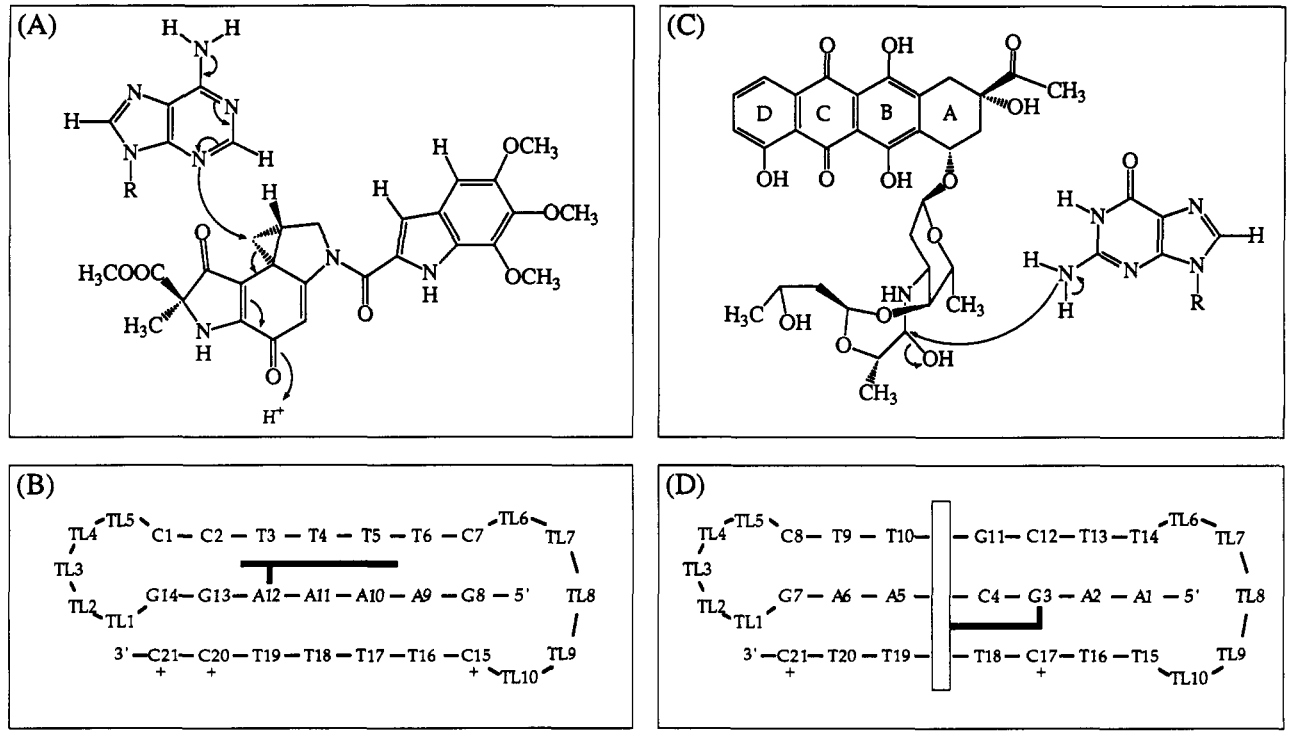

FIGURE 7: (A) Covalent adduct formation between duocarmycin A and adenine. (B) Oligonucleotide sequence of a Y·RY triplex used to generate a site-specific adduct between duocarmycin A and adenine residue A12. (C) Covalent adduct formation between SN-07 and guanine. (D) Oligonucleotide sequence of a Y·RY triplex used to generate a site-specific adduct between SN-07 and guanine residue G3. The site of intercalation by the aglycon moiety of the chromophore is indicated by the rectangular box.

Y·RY DNA triplexes establishing the generality of this approach.

Noncovalent Groove Binding. The previous research on the binding of the minor groove-specific ligand netropsin to duplex DNA [reviewed in Zimmer and Wahnert 1986)] has now been extended to its complexes with Y·RY DNA triplexes (Park & Breslauer, 1992; Durand *et al.*, 1992). A combined spectroscopic and calorimetric investigation by Park and Breslauer (1992) established that the binding of netropsin in the minor groove of the Y·RY triplex does not displace the third strand from the major groove of the triplex. Further, netropsin binding thermally destabilizes the triplex to duplex transition in contrast to stabilization of the duplex to single-strand transition. There is also a decrease in the cooperativity of the triplex to duplex transition on netropsin binding. These results establish major groove–minor groove "cross talk" where the drug, which is bound in the minor groove of the Y·RY triplex, affects the properties of the host triplex as related to the equilibrium in which the third strand is expelled from the major groove (Park & Breslauer, 1992).

Site-Specific Covalent Drug–Triplex Complexes. Our own laboratory has approached the drug–triplex complex problem by focusing on adducts formed by drugs bound covalently and site-specifically at a single defined residue on the intramolecular triplex. This approach overcomes the complications observed with multiple and competing drug binding sites associated with noncovalent drug–nucleic acid complexes.

Initially, we targeted the minor groove of an intramolecular Y·RY triplex using the drug duocarmycin A (Figure 7A) (Takashi *et al.*, 1988), which is known to bind covalently to the minor groove N^3 position of adenine and target the 3′-adenine in an 5′-A-A-A-A-3′ segment (Boger *et al.*, 1990). The intramolecular Y·RY triplex containing the A_4 stretch on strand II was synthesized; the duocarmycin covalent adduct was generated site-specifically at A12 (Figure 7B) and was subsequently purified and characterized by NMR techniques (Lin & Patel, 1992). The intermolecular NOEs established that the adduct covalently linked at A12 in the sequence was positioned in the minor groove and directed toward the 5′-end of the A-A-A-A stretch. Formation of the duocarmycin adduct in the minor groove did not displace the third strand from the major groove but rather affected the pH-dependent triplex–duplex equilibrium whose pK_a was lowered by 1.8 pH units in the complex (Lin & Patel, 1992). Current efforts are focused on defining the conformational perturbations associated with the cross talk between the drug bound in the minor groove and the third strand bound in the major groove.

The anthracycline antitumor antibiotics contain a planar aglycon ring that intercalates into the DNA helix and an attached amino sugar that is positioned in and interacts with the minor groove (Quigley *et al.*, 1980). We have generated and purified a site-specific covalent adduct between the anthracycline SN-07 chromophore (Figure 7C) (Kikuchi *et al.*, 1985) and the exocyclic amino group of guanine at the G3-C4-A5 segment on strand II of a T·CG-containing intramolecular Y·RY triplex (Figure 7D). The covalent complex yielded well-resolved NMR spectra whose characterization established that the SN-07 chromophore was covalently bound site-specifically at G3, the amino sugar and its attached eight-membered ring was positioned in the minor groove, and the aglycon was intercalated between triples at the C4-A5 step (Ye *et al.*, 1993). Further, the minor groove binding and intercalation associated with SN-07 complex formation was achieved without disruption of the third strand in the major groove. Our structural approach has definitively established that the planar anthracycline aglycon can site-specifically intercalate between triples in the interior of the triplex while retaining its attached amino sugar in the minor groove and the third strand in the major groove.

The initial efforts on ligand–triplex complexes have focused on the Y·RY family of triplexes. The greater challenge in the longer term will be to design experiments that will address

structural and energetics issues related to complexes formed between ligands and R·RY triplexes.

Future Prospects for Structural Studies

There are several future challenges in triplex research that can be approached on the basis of an improved understanding of the structural bases of triplex formation. These include (1) extending oligonucleotide-mediated recognition to any random sequence of double-stranded DNA, (2) overcoming the current limitations associated with the acidic pH requirement for formation of Y·RY triplexes and nonisomorphic positioning of triples in R·RY triplexes, (3) characterizing triplex–duplex junctions and their stabilization through intercalative binding of attached aromatic ligands, (4) extending the structural research on DNA triplexes to their RNA and hybrid counterparts, and (5) characterizing base triple alignments associated with triple-stranded intermediates in genetic recombination. These issues are outlined in some detail in the following paragraphs.

Alternate Strand Recognition. In this approach, two pyrimidine-rich oligonucleotides coupled 3′–3′ or 5′–5′ recognize purine residues located on alternate strands in duplex sequences of the type 5′-$(R)_m$-$(Y)_n$-3′ and 5′-$(Y)_m$-$(R)_n$-3′ (Horne & Dervan, 1991). An elegant extension of this approach circumvents the need for a 3′–3′ or 5′–5′ covalent linkage by employing an oligonucleotide with natural 3′–5′ phosphodiesters which binds to the purine tracts on alternate strands forming contiguous Y·RY and R·RY triplexes (Jayasena & Johnston, 1992, 1993; Beal & Dervan, 1992b). Although 5′-$(R)_m$-$(Y)_n$-3′ sequences can be targeted using this approach, recognition of 5′-$(Y)_m$-$(R)_n$-3′ sequences requires at least two additional nucleotides on the third strand at the crossover site between the triplex domains (Beal & Dervan, 1992b). Structural studies of alternate strand triplexes formed through third strand targeting of 5′-$(Y)_m$-$(R)_n$-3′ and 5′-$(R)_m$-$(Y)_n$-3′ sequences might explain some of the intriguing features of triplex–triplex and triplex–duplex–triplex junctions.

Third Strand Oligonucleotides Containing Synthetic Heterocycles. Synthetic approaches have been undertaken either to extend the recognition to pyrimidine residues or to improve the specificity of interaction with purine residues. Novel heterocycles, 4-(3-benzamidophenyl)imidazole (designated D_3) and 2-deoxynebularine (designated dN), have been designed and synthesized by Dervan and co-workers to recognize CG base pairs within Y·RY and R·RY triplex motifs. From affinity cleavage it appears that D_3 interacts specifically with not only CG base pairs but also TA base pairs as well (Griffin *et al.*, 1992). Preliminary data for dN also indicate that it recognizes AT base pairs in addition to CG base pairs (Stilz & Dervan, 1993). Thus, although these synthetic approaches represent an important first step in formulating and extending the triplex code to all base pairs in DNA, structural studies might be useful in gaining a better understanding of the current limitations and suggest approaches for further improvements. Preliminary NMR studies of a D_3·TA-containing Y·RY triplex established that the D_3 heterocycle participates in the recognition process through intercalation into the helix (Koshlap *et al.*, 1993).

Novel heterocycles including pseudoisocytosine (Ono *et al.*, 1991) and 8-oxoadenosine and its derivatives (Miller *et al.*, 1992; Krawczyk *et al.*, 1992) have been synthesized and shown to mimic protonated cytosine residues effectively in Y·RY triplexes, thereby eliminating the acidic pH requirement for triplex stability. An analogous approach has been undertaken to stabilize R·RY triplexes containing guanines and thymines in the third strand. The heterocycle 7-deaza-2′-deoxyxanthosine (designated dzaX) when incorporated in third strand oligonucleotides opposite adenines in AT base pairs has been shown to dramatically enhance the stability of R·RY triplexes compared to thymines at the same position (Milligan *et al.*, 1993). The structural basis for the success of these approaches could provide useful leads in the design of new heterocycles for the recognition of other bases.

Third Strand Oligonucleotides Linked to Intercalating Ligands. Hélène and co-workers have covalently linked third strand oligonucleotides to ligands that intercalated between base pairs at triplex–duplex junctions [Sun *et al.*, 1989; reviewed in Thuong and Hélène (1993)]. The third strand and the ligand act in a synergistic manner—the third strand prevents nonspecific interactions of the ligand while the ligand enhances the stability of the triplex interaction. Intercalation was strongly favored when the ligand was attached to the 5′-terminus of the third strand oligonucleotide and a 5′-YpR-3′ sequence was present at the triplex–duplex junction (Takasugi *et al.*, 1991). Structural studies of triplex–duplex junctions in both in the presence and absence of ligands should provide useful insights.

RNA–DNA Hybrid Triplexes. Intermolecular triplexes containing mixtures of RNA and DNA strands have been studied by several groups (Roberts & Crothers, 1992; Han & Dervan, 1993; Escudé *et al.*, 1993). Dramatic effects on triplex stability were observed for various combinations of DNA and RNA strands in the Y·RY motif. Triplexes formed by third strand RNA oligomers with duplexes containing a DNA purine strand were generally the most stable. The stability of these triplexes is further enhanced when the 2′-OH groups were replaced with 2′-OCH_3 groups (Escudé *et al.*, 1993). Stable triplexes between third strand RNA oligomers and duplexes containing the RNA purine strand were also observed. By contrast, triplexes formed by DNA third strands and duplexes containing the RNA purine strand were the least stable. Thus it appears that recognition of DNA in duplexes can be achieved by both DNA and RNA oligonucleotides, whereas recognition of RNA can be achieved only by RNA oligonucleotides (Han & Dervan, 1993). The structural basis for these dramatic effects of backbone composition on triplex stability needs to be understood. It is also interesting to note that RNA·DNA·DNA hybrids are not observed in the R·RY motif (Escudé *et al.*, 1993).

Stabilization of Hybrid Triplexes through Ligand Binding. In a recent study, the presence of duplex-specific ligands including berenil [1,3-bis(4′-amidinophenyl)triazene], ethidium bromide, DAPI (4′,6-diamidino-2-phenylindole), and/or netroposin was shown to induce *triplex* formation in sequences that would otherwise not form such structures (Pilch & Breslauer, 1994). Specifically, these ligands promoted the formation of triplexes of the type DNA(Y)·RNA(R)·DNA(Y) and RNA(R)·RNA(R)·DNA(Y), which were previously not observed (see above). There are no structural details of hybrid triplex–ligand interactions, and studies in this area should provide new insights.

Base Triple Interactions in Folded RNA. Base triple interactions were first discovered in a biological context in transfer RNA molecules (Kim *et al.*, 1974; Robertus *et al.*, 1974). Bases involved in these tertiary interactions are invariant or semiinvariant in transfer RNA molecules, underscoring the stabilizing influence of these interactions. Base triple interactions analogous to those described in this paper have been proposed for several other RNA molecules by NMR (Puglisi *et al.*, 1992) and phylogenetic covariation

analyses (Michel *et al.*, 1990). Novel base triple alignments including those involving third strand bases located in the minor groove have been described or anticipated (Chastian & Tinoco, 1992; Michel & Westhof, 1990). One can anticipate that future structural studies will identify additional triples that are involved in the maintenance of the tertiary fold of RNA.

Triple-Stranded Intermediates in Genetic Recombination. Genetic recombination in living organisms proceeds through at least two fundamentally different mechanisms. In homologous recombination, the exchange of genetic information between two chromosomes occurs in regions of extensive sequence identity, whereas in site-specific recombination, the exchange occurs at specific sites. Homologous recombination is thought to proceed via a triple-stranded intermediate (Rao *et al.*, 1990; Hsieh *et al.*, 1990). The formation of this intermediate is facilitated by recombinases (*e.g.*, RecA) which presumably lower the activation energy for the structural transition. Not unlike canonical triplexes, it has been postulated that the third strand in this intermediate is located in the major groove and participates in sequence-specific hydrogen-bonding interactions. However, these triplexes differ from canonical triplexes in that the third strand is homologous and oriented parallel to the same strand in the duplex, probably defining a new structural motif (Rao & Radding, 1994; Zhurkin *et al.*, 1994). A major challenge concerns the definitive characterization of the triple pairing alignments and structures of RecA-catalyzed triple-stranded intermediates in genetic recombination.

ACKNOWLEDGMENT

The GRASP program was kindly provided by Dr. Anthony Nicholls, Ranganathan Bharadwaj, and Dr. Barry Honig.

REFERENCES

Arnott, S., & Selsing, E. (1974) *J. Mol. Biol. 88*, 509–521.
Beal, P. A., & Dervan, P. B. (1991) *Science 251*, 1360–1363.
Beal, P. A., & Dervan, P. B. (1992a) *Nucleic Acids Res. 20*, 2773–2776.
Beal, P. A., & Dervan, P. B. (1992b) *J. Am. Chem. Soc. 114*, 4976–4982.
Boger, D. L., Ishizaki, T., Zarrinmaych, H., Munk, S. A., Kitos, P. A., & Suntorwat, O. J. (1990) *J. Am. Chem. Soc. 112*, 8961–8971.
Chastian, M., & Tinoco, I. (1992) *Biochemistry 31*, 12733–12741.
Cheng, Y.-K., & Pettitt, B. M. (1992a) *J. Am. Chem. Soc. 114*, 4465–4474.
Cheng, Y.-K., & Pettitt, B. M. (1992b) *Prog. Biophys. Mol. Biol. 58*, 225–257.
Christophe, D., Cabrer, B., Bacolla, A., Targovnik, H., Pohl, V., & Vassart, G. (1985) *Nucleic Acids Res. 13*, 5127–5144.
Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J., & Hogan, M. E. (1988) *Science 241*, 456–459.
De los Santos, C., Rosen, M., & Patel, D. J. (1989) *Biochemistry 28*, 7282–7289.
Dittrich, K., Gu, J., Tinder, R., Hogan, M., & Gao, X. (1994) *Biochemistry 33*, 4111-4120.
Durand, M., Thuong, N. T., & Maurizot, J. C. (1992) *J. Biol. Chem. 267*, 24394–24399.
Durland, R. H., Kessler, D. J., Gunnell, S., Duvic, M. D., Pettitt, B. M., & Hogan, M. E. (1991) *Biochemistry 30*, 9246–9255.
Duval-Valentin, G., Thuong, N. T., & Hélène, C. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 504–508.
Escudé, C., François, J.-C., Sun, J.-S., Ott, G., Sprinzl, M., Garestier, T., & Hélène, C. (1993) *Nucleic Acids Res. 21*, 5547–5553.
Felsenfeld, G., & Miles, H. T. (1967) *Annu. Rev. Biochem. 36*, 407–448.
Felsenfeld, G., Davies, D., & Rich, A. (1957) *J. Am. Chem. Soc. 79*, 2023–2024.
François, J.-C., Saison-Behmoaras, T., Barbier, C., Chassignol, M., Thuong, N. T., & Hélène, C. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9702–9706.
Griffin, L. C., & Dervan, P. B. (1989) *Science 245*, 967–971.
Griffin, L. C., Kiessling, L. L., Beal, P. A., Gillespie, P., & Dervan, P. B. (1992) *J. Am. Chem. Soc. 114*, 7976–7982.
Grigoriev, M., Preseuth, D., Robin, P., Hemar, A., Saison-Behmoaras, T., Dautry-Varsat, A., Thuong, N. T., Hélène, C., & Harel-Bellan, A. (1992) *J. Biol. Chem. 267*, 3389–3395.
Han, H., & Dervan, P. B. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 3806–3810.
Häner, R., & Dervan, P. B. (1990) *Biochemistry 29*, 9761–9765.
Hartman, D. A., Kuo, S.-R., Broker, T. R., Chow, L. T., & Wells, R. D. (1992) *J. Biol. Chem. 267*, 5488–5494.
Hoogsteen, K. (1963) *Acta Crystallogr. 16*, 907–916.
Horne, D. A., & Dervan, P. B. (1990) *J. Am. Chem. Soc. 112*, 2435–2437.
Howard, F. B., Miles, H. T., Liu, K., Frazier, J., Raghunathan, G., & Sasisekharan, V. (1992) *Biochemistry 31*, 10671–10677.
Hsieh, P., Camerini-Otero, C. S., & Camerini-Otero, R. D. (1990) *Genes Dev. 4*, 1951–1963.
Htun, H., & Dahlberg, J. E. (1989) *Science 243*, 1571–1576.
Inman, R. B. (1964) *J. Mol. Biol. 10*, 137–146.
Jayasena, S. D., & Johnston, B. H. (1992) *Biochemistry 31*, 320–327.
Jayasena, S. D., & Johnston, B. H. (1993) *Biochemistry 32*, 2800–2807.
Kiessling, L. L., Griffin, L. C., & Dervan, P. B. (1992) *Biochemistry 31*, 2829–2834.
Kikuchi, Y., Niwano, M., Yajima, N., Nakamura, G., & Miyata, N. (1985) *J. Antibiot. 38*, 1670–1676.
Kim, S.-H., Quigley, G. J., Suddath, F. L., McPherson, A., Sneden, D., Kim, J. J., Weinzierl, W., & Rich, A. (1973) *Science 179*, 285–288.
Koh, J. S., & Dervan, P. B. (1992) *J. Am. Chem. Soc. 114*, 1470–1478.
Kohwi, Y., & Kohwi-Shigematsu, T. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 3781–3785.
Koshlap, K. M., Gillespie, P., Dervan, P. B., & Feigon, J. (1993) *J. Am. Chem. Soc. 115*, 7908–7909.
Krawczyk, S. H., Milligan, J. F., Wadwani, S., Moulds, C., Froehler, B. C., & Matteucci, M. D. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 3761–3764.
Kubinec, M. G., & Wemmer, D. E. (1992) *J. Am. Chem. Soc. 114*, 8739–8740.
Larsen, A., & Weintraub, H. (1982) *Cell 29*, 609–622.
Laughton, C. A., & Neidle, S. (1992a) *J. Mol. Biol. 223*, 519–529.
Laughton, C. A., & Neidle, S. (1992b) *Nucleic Acids Res. 20*, 6535–6541.
Le Doan, T., Perrouault, L., Praseuth, D., Habhoub, N., Decout, J. L., Thuong, N. T., Lhomme, J., & Hélène, C. (1987) *Nucleic Acids Res. 15*, 7749–7760.
Lee, J. S., Johnson, D. A., & Morgan, A. R. (1979) *Nucleic Acids Res. 6*, 3073–3091.
Lee, J. S., Woodsworth, M. L., Latimer, L. J. P., & Morgan, A. R. (1984) *Nucleic Acids Res. 12*, 6603–6614.
Lee, J. S., Latimer, L. J. P., & Hempel, K. J. (1993) *Biochemistry 32*, 5591–5597.
Letai, A. G., Palladino, M. A., Fromm, E., Rizzo, V., & Fresco, J. R. (1988) *Biochemistry 27*, 9108–9112.
Liepinsh, E., Otting, G., & Wüthrich, K. (1992) *Nucleic Acids Res. 20*, 6549–6553.
Lin, C. H., & Patel, D. J. (1992) *J. Am. Chem. Soc. 114*, 10658–10660.
Lipsett, M. (1964) *J. Biol. Chem.* 239, 1256–1260.
Lyamichev, V. I., Mirkin, S. M., & Frank-Kamenetskii, M. D. (1986) *J. Biomol. Struct. Dyn. 3*, 667–669.

Macaya, R., Gilbert, D. E., Malek, S., Sinsheimer, J. S., & Feigon, J. (1991) *Science 254*, 270–274.
Macaya, R., Wang, E., Schultze, P., Sklenár, V., & Feigon, J. (1992) *J. Mol. Biol. 225*, 755–773.
Maher, L. J., Dervan, P. B., & Wold, B. (1990) *Biochemistry 29*, 8820–8826.
Maher, L. J., Dervan, P. B., & Wold, B. (1992) *Biochemistry 31*, 70–81.
Malkov, V. A., Voloshin, O. N., Soyfer, V. N., & Frank-Kamenetskii, M. D. (1993) *Nucleic Acids Res. 21*, 585–591.
Manzini, G., Gasparotto, D., Xodo, L. E., Quadrifoglio, F., van der Marel, G. A., & van Boom, J. H. (1990) *J. Mol. Biol. 213*, 833–843.
Marck, C., & Thiele, D. (1978) *Nucleic Acids Res. 5*, 1017–1028.
Mergny, J.-L., Sun, J.-S., Rougée, M., Garestier, T. M., Barcelo, F., Chomilier, J., & Hélène, C. (1991) *Biochemistry 30*, 9791–9798.
Mergny, J.-L., Duval-Valentin, G., Nguyen, C. H., Perrouault, L., Faucon, B., Rougée, M., Garestier, T. M., Bisagni, E., & Hélène, C. (1992) *Science 256*, 1681–1684.
Michel, F., & Westhof, E. (1990) *J. Mol. Biol. 216*, 585–610.
Michel, F., Ellington, A. D., Couture, S., & Szostak, J. W. (1990) *Nature (London) 347*, 578–580.
Miller, P. S., Bhan, P., Cushman, C. D., & Trapane, T. L. (1992) *Biochemistry 31*, 6788–6793.
Milligan, J. F., Krawczyk, S. H., Wadwani, S., & Matteucci, M. D. (1993) *Nucleic Acids Res. 21*, 327–333.
Mohan, V., Smith, P. E., & Pettitt, B. M. (1993) *J. Am. Chem. Soc. 115*, 9297–9298.
Morgan, A. R., & Wells, R. D. (1968) *J. Mol. Biol. 37*, 63–80.
Moser, H. E., & Dervan, P. B. (1987) *Science 238*, 645–650.
Ono, A., Ts'o, P. O. P., & Kan, L. S. (1991) *J. Am. Chem. Soc. 113*, 4032–4033.
Ouali, M., Letellier, R., Adnet, F., Liquier, J., Sun, J.-S., Lavery, R., & Taillandier, E. (1993) *Biochemistry 32*, 2098–2103.
Park, Y. W., & Breslauer, K. J. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 6653–6657.
Pilch, D. S., & Breslauer, K. J. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 9332–9336.
Pilch, D. S., Brousseau, R., & Shafer, R. H. (1990) *Nucleic Acids Res. 18*, 5743–5750.
Pilch, D. S., Levenson, C., & Shafer, R. H. (1991) *Biochemistry 30*, 6081–6087.
Pilch, D. S., Martin, M.-T., Nguyen, C.-H., Sun, J.-S., Bisagni, E., Garestier, T., & Hélène, C. (1993a) *J. Am. Chem. Soc. 115*, 9942–9951.
Pilch, D. S., Waring, M. J., Sun, J.-S., Rougée, M., Nguyen, C.-H., Bisagni, E., Garestier, T., & Hélène, C. (1993b) *J. Mol. Biol. 212*, 926–946.
Plum, G. E., Park, Y.-W., Singleton, S. F., Dervan, P. B., & Breslauer, K. J. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 9436–9440.
Puglisi, J. D., Tan, R., Calnan, B. J., Frankel, A. D., & Williamson, J. R. (1992) *Science 257*, 76–80.
Quigley, G. J., Wang, A. H. J., Ughetto, G., van der Marel, G., van Boom, J. H., & Rich, A. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 7204–7208.
Radhakrishnan, I., & Patel, D. J. (1993) *Structure 1*, 135–152.
Radhakrishnan, I., & Patel, D. J. (1994a) *Structure 2*, 17–32.
Radhakrishnan, I., & Patel, D. J. (1994b) *Structure 2*, 395–405.
Radhakrishnan, I., & Patel, D. J. (1994c) *J. Mol. Biol. 241*, 600–619.
Radhakrishnan, I., Gao, X., de los Santos, C., Live, D., & Patel, D. J. (1991a) *Biochemistry 30*, 9022–9030.
Radhakrishnan, I., de los Santos, C., & Patel, D. J. (1991b) *J. Mol. Biol. 221*, 1403–1418.
Radhakrishnan, I., Patel, D. J., & Gao, X. (1992) *Biochemistry 31*, 2514–2523.
Radhakrishnan, I., Patel, D. J., Priestley, E. S., Nash, H. M., & Dervan, P. B. (1993a) *Biochemistry 32*, 11228–11234.
Radhakrishnan, I., de los Santos, C., & Patel, D. J. (1993b) *J. Mol. Biol. 234*, 188–197.
Raghunathan, G., Miles, H. T., & Sasisekharan, V. (1993) *Biochemistry 32*, 455–462.
Rajagopal, P., & Feigon, J. (1989) *Biochemistry 28*, 7859–7870.
Rao, B. J., & Radding, C. M. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 6161–6165.
Rao, B. J., Jwang, B., & Radding, C. M. (1990) *J. Mol. Biol. 213*, 789–809.
Riley, M., Maling, B., & Chamberlin, M. J. (1966) *J. Mol. Biol. 20*, 359–389.
Roberts, R. W., & Crothers, D. M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9397–9401.
Roberts, R. W., & Crothers, D. M. (1992) *Science 258*, 1463–1466.
Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, P., Brown, F. S., Clark, B. F., & Klug, A. (1974) *Nature (London) 250*, 546–551.
Schneider, B., Cohen, D. M., Schleifer, L., Srinivasan, A. R., Olson, W. K., & Berman, H. M. (1993) *Biophys. J. 65*, 2291–2303.
Singleton, S. F., & Dervan, P. B. (1992a) *J. Am. Chem. Soc. 114*, 6957–6965.
Singleton, S. F., & Dervan, P. B. (1992b) *Biochemistry 31*, 10995–11003.
Sklenár, V., & Feigon, J. (1990) *Nature (London) 345*, 836–838.
Stilz, H. U., & Dervan, P. B. (1993) *Biochemistry 32*, 2177–2185.
Strobel, S. A., & Dervan, P. B. (1991) *Nature (London) 350*, 172–174.
Strobel, S. A., Stamm, L. A. D., Riba, L., Housman, D. E., & Dervan P. B. (1991) *Science 254*, 1639–1642.
Sun, J.-S., & Hélène, C. (1993) *Curr. Opin. Struct. Biol. 3*, 345–356.
Sun, J.-S., François, J.-C., Garestier, T. M., Behemoras, T. S., Roig, V., Thuong, N. T., & Hélène, C. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9198–9202.
Takashi, I., Takashi, K. I., Ichimura, M., Morimoto, M., Asano, K., Kawamoto, I., Tomita, F., & Nakano, H. (1988) *J. Antibiot. 41*, 1915–1917.
Takasugi, M., Guendouz, A., Chassignol, M., Decout, J. L., Lhomme, J., Thuong, N. T., & Hélène, C. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 5602–5606.
Thomas, G. A., & Peticolas, W. L. (1991) *J. Biomol. Struct. Dyn. 9*, 437–445.
Thuong, N. T., & Hélène, C. (1993) *Angew. Chem., Int. Ed. Engl. 32*, 666–690.
Völker, J., Botes, D. P., Lindsey, G. G., & Klump, H. H. (1993) *J. Mol. Biol. 230*, 1278–1290.
Wang, E., Malek, S., & Feigon, J. (1992) *Biochemistry 31*, 4838–4846.
Wells, R. D., Collier, D. A., Hanvey, J. C., Shimizu, M., & Wohlrab, F. (1988) *FASEB J. 2*, 2939–2949.
Wilson, W. D., Tanious, F. A., Mizan, S., Yao, S., Kiselyov, A. S., Zon, G., & Strekowski, L. (1993) *Biochemistry 32*, 10614–10621.
Xodo, L. E., Manzini, G., Quadrifoglio, F., van der Marel, G. A., & van Boom, J. H. (1991) *Nucleic Acids Res. 19*, 5625–5631.
Ye, X., Kimura, K., & Patel, D. J. (1993) *J. Am. Chem. Soc. 115*, 9325–9326.
Yoon, K., Hobbs, C. A., Koch, J., Sardaro, M., Kutny, R., & Weis, A. L. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 3840–3844.
Young, S. L., Krawczyk, S. H., Matteucci, M. D., & Toole, J. J. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 10023–10026.
Zhurkin, V. B., Raghunathan, G., Ulyanov, N. B., Camerini-Otero, R. D., & Jernigan, R. L. (1994) *J. Mol. Biol. 239*, 181–200.
Zimmer, C., & Wahnert, U. (1986) *Prog. Biophys. Mol. Biol. 47*, 31–112.