# Minor Groove Functional Groups Are Critical for the B-Form Conformation of Duplex DNA†

Tao Lan and Larry W. McLaughlin*

*Department of Chemistry, Merkert Chemistry Center, Boston College, Chestnut Hill, Massachusetts 02167*

*Received July 5, 2000; Revised Manuscript Received July 13, 2000*

ABSTRACT: Two analogue bases are described: 3-deazaadenine is a derivative of adenine from which N3 has been deleted and 3-methyl-2-pyridone is a C-nucleoside that mimics thymine but lacks the O2 carbonyl. The dc$^3$A-dm$^3$2P base pair is similar to dA-dT but eliminates the polar functional groups in the minor groove. The presence of this base pair in dA-dT rich sequences results in destabilized duplexes or conformational preferences for monomolecular hairpins rather than bimolecular duplexes. When present in dG-dC rich sequences, no significant differences in helix stability are observed. These differences are explained on the basis of hydration effects, most notably, the elimination of the minor groove spine of hydration normally present in dA-dT rich sequences. CD spectra suggest that sequences with a fully modified core (four analogue base pairs) are more A-like helices than B-like helices. Sequences containing two analogue base pairs can be transformed to A-like helices under conditions of high salt, or 65% trifluoroethanol. These conformational changes are also explained in terms of a loss of hydration in the minor groove that normally stabilizes the B-form conformation. In the absence of such hydration, the helices are conformationally mobile and adopt a more A-like helix form.

Early crystal structures defined two right-handed helical conformations for DNA designated as the A-form and B-form (*1*). Interconversion between the two helical forms is generally observed to be cooperative and has been described as being driven by base stacking effects (*2*), although a recent structure has been determined that appears to be intermediate between A- and B-forms (*3*). A-form helices require dG-dC rich sequences and low water activity; the latter form is typically obtained using salts (*4*−*6*) or organic solvents (*7*−*9*). Transitions from the B-form helix to the A-form helix are promoted by d(CC)·d(GG) sequence steps.

The stabilities of both the B-form and A-form helices, and their interconversion, are generally explained in terms of hydration effects (*10*). In the B-form helix, there appears to be relatively little specific hydration of the internucleotide phosphates, but a well-ordered spine of hydration is present in the minor groove of dA-dT rich sequences. This spine of hydration is formed from a primary layer of water molecules that bridges the base pairs by binding to the O2 carbonyls of the dT residues and the N3 atoms of the dA residues (*10*). This interaction is not within a single base pair, but bridges residues from adjacent base pairs. A second layer of water molecules is piggybacked onto the primary layer to create a continuous chain of water molecules within the minor groove. Recent high-resolution structures have confirmed the nature of this spine of hydration (*11*−*13*), and two structures have indicated that additional water molecules can bind to form hexagonal complexes (*11*, *12*). In general, when DNA and RNA duplexes are compared, B-form helices appear to be more hydrated than A-form helices on the basis of changes in molar volumes (*14*−*16*). It has also been reported that monovalent ions can be sequestered within the minor groove spine of hydration (*11*, *17*−*19*), although this electrostatic view of the minor groove is disputed (*13*). This water structure has been used to account for the difficulty in observing the B → A transitions in sequences rich in dA-dT base pairs, and the fact that poly(dA)·poly(dT) sequences have never been reported to adopt a stable A-form conformation.

A-form duplexes do not exhibit a minor groove spine of hydration. In contrast to the B-form conformation, the phosphates of an A-type helix display well-ordered hydration, and along the phosphate chain, individual phosphates are bridged by one to three water molecules (*10*). The major groove in the A-form helix is very narrow, and the conformational stability may be facilitated by water molecules that bridge the phosphates across the major groove. The transition from a B-form helix to an A-form structure is thought to require the destabilization of the ordered minor groove spine of hydration present in the former (*10*). Destabilization of the B-form hydration in the minor groove can be accomplished in part by increasing the dG-dC content of the sequence; the N2-amino groups of the guanine residues protrude into the minor groove and interfere with the ordered minor groove spine of hydration. Reducing the water activity of the solvent [using high concentrations of salt (*4*−*6*) or organic solvent (*7*−*9*)] then facilitates the cooperative transition.

The orientation and the inter-residue stacking of the nucleobases differ in the B-form and A-form helices, and these variations in base stacking effects can be readily detected by circular dichroism spectropolarimetry (CD).[1] In

† This work was supported by a grant from the NSF (MCB 0077667).

* To whom correspondence should be addressed. Telephone: (617) 552-3622. Fax: (617) 552-2705. E-mail: larry.mclaughlin@bc.edu.

10.1021/bi0015212 CCC: $20.00 
Published on Web 01/06/2001

addition to changes in base stacking, the sugar puckering in B-form helices favors the C2′-endo conformation while in A-form helices the sugar puckering tends to be C3′-endo, but an early (*2*) and still common view is that sugar puckering is a consequence of the B to A transitions rather than being a cause.

Some modified DNA duplexes will adopt A-form duplexes even under conditions of high water activity. For example, sequences containing 3′-N−P-5′ phosphoramidate linkages form A-type duplexes (*20*); model studies with simple dimers suggest that with these amidate linkages the C3′-endo sugar conformation is preferred. Crystallographic analysis has confirmed the A-form conformation of this sequence (*21*). In the study presented here, we have examined the properties of duplexes containing one or more modified dA-dT base pairs in which the functional groups normally present in the minor groove (N3 of dA and the O2 carbonyl of dT), those responsible for minor groove hydration, have been removed without otherwise altering base pairing properties. The importance of this functionality and its role in helix stability and duplex conformation has been examined.

## EXPERIMENTAL PROCEDURES

### *Materials*

Oligonucleotides were synthesized on an Applied Biosystems 381A DNA synthesizer. 2′-Deoxynucleotide phosphoramidites and 3′-terminal nucleoside-controlled pore glass support (CPG) were purchased from Glenn Research (Sterling, VA). High-performance liquid chromatography (HPLC) was performed on a Beckman HPLC system using C-18 reversed-phase columns (ODS-Hypersil, 5 mm particle size, 120 Å pore) with detection at 260 nm. Oligonucleotides were desalted with Econo-Pac 10DG disposable chromatography columns (Bio-Rad, Hercules, CA). $T_M$ measurements were taken on an AVIV spectrophotometer (model 14DS UV/Vis).

### *Methods*

*DNA Synthesis.* The native and modified 12-mers were prepared and deprotected by solid-phase DNA synthesis using standard protocols (*22*). A slightly different methodology was applied when thymidine analogue dm$^3$2P was involved. The *p*-nitrophenylethyl protecting group for dm$^3$2P was deprotected as the first step by treating the solid support with a solution of 40% TEA and pyridine for 2 h and then 0.5 M 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU) in anhydrous pyridine for 8−48 h and finally by washing with acetonitrile three times. After evaporation of the volatiles, ammonia treatment was performed in the conventional manner.

Purification of the oligonucleotides was accomplished by fast flow HPLC (4.6 mm × 120 mm, reverse-phase C18 column, trityl on), starting with 100% A for 1 min, using a linear gradient from 0 to 50% B over the course of 4.5 min [A being 50 mM triethylammonium acetate (pH 7.0) and B being 50 mM triethylammonium acetate in 70% acetonitrile]. The DMT-protected 12-mers had retention times of about 5.0 min. The collected oligonucleotides were reduced in volume and detritylated with 80% aqueous acetic acid (60 min, on ice). The resulting oligonucleotides were then desalted (Sephadex G-10) and stored at −20 °C. The unnatural nucleotides could be incorporated into the DNA strands with essentially the same coupling efficiency as the common nucleotides.

MALDI-Voyager MS for 5′-d(CGCGaattCGCG)-3′ (a is dc$^3$A and t dm$^3$2P): calcd for 12-merH$^+$ 3606, found 3600.

*Nucleoside Analysis.* Small amounts of oligonucleotides containing modified bases were digested with snake venom phosphodiesterase and calf intestinal alkaline phosphatase into monomeric units: a 50 $\mu$L reaction mixture containing 0.5 $A_{260}$ unit of oligonucleotides was incubated overnight at 37 °C with 1 unit of snake venom phosphodiesterase and 1 unit of calf intestinal alkaline phosphatase. An aliquot of this mixture was analyzed by HPLC [4.6 mm × 250 mm column of ODS-Hypersil, 50 mM potassium phosphate (pH 5.5) and a gradient of 0 to 70% methanol over the course of 60 min]. The results of this analysis indicated that no undigested material remained, and that all of the nucleoside constituents had been fully deprotected. On the other hand, both dG and dc$^3$A coeluted, and dm$^3$2P was only partially resolved from that mixture. To confirm the presence of all four nucleosides, an aliquot of the digestion mixture was reanalyzed using 0.1 M triethylamine/acetic acid (pH 8.0) and 5% acetonitrile (isocratic) which resulted in the following elution times: 3.7 min for dC, 5.9 min for dG, 6.6 min for dm$^3$2P, and 8.0 min for dc$^3$A. The dm$^3$2P was still difficult to observe because of its low extinction coefficient at 260 nm, but reanalysis with the detection at 297 nm confirmed its presence. The extinction coefficients of nucleoside constituents at pH 5.5 and 260 nm are as follows: 11 700 L M$^{-1}$ cm$^{-1}$ for dG, 7300 L M$^{-1}$ cm$^{-1}$ for dC, 10 200 L M$^{-1}$ cm$^{-1}$ for dc$^3$A, and 1400 L M$^{-1}$ cm$^{-1}$ for dm$^3$2P.

*Thermal Denaturation Studies.* Thermal denaturation studies were performed in solutions of 20 mM $NaH_2PO_4$ (pH 7.0) and varying the concentrations of NaCl. Absorbance and temperature values were measured with an AVIV 14DS UV−visible spectrophotometer equipped with digital temperature control. The temperature of the cell compartment was increased in 1.0 °C steps (from 10 to 95 °C), and when equilibrium was reached, temperature and absorbance data were collected. The $T_M$ values were determined manually from the absorbance versus temperature plots.

*Electrophoresis.* PAGE was performed using gels prepared from solutions of 12% acrylamide, 0.6% bis(acrylamide), 10 mM $Mg(OAc)_2$, 50 mM Tris-HOAc, and 0.1% ammonium persulfate (pH 7.0) (30 cm × 15 cm × 0.1 cm). A 7 $\mu$L aliquot of 2.5 $\mu$M oligonucleotide in 10 mM $Mg(OAc)_2$, 50 mM Tris-HOAc buffer was heated at 95 °C for 10 min, cooled to ambient temperature, and equilibrated at 4.0 °C overnight. To this solution were added 1 $\mu$L of [$\gamma$-$^{32}$P]ATP, 1 $\mu$L of T4 polynucleotide kinase buffer (10×), and 1 $\mu$L of T4 PNK, and the resulting mixture was incubated at 37 °C for 30 min. Samples were mixed with 10 mM $Mg(OAc)_2$, 50 mM Tris-HOAc (pH 7.0), 40% sucrose, and 0.05% bromophenol blue and 0.05% xylene cyanol as dye markers for loading. Electrophoresis was performed at 4.0 °C. The gel was run at a constant voltage of 200 V, until the xylene cyanol dye had migrated approximately 20 cm (~10 h). The gels were visualized using a phosphorimager.

[1] Abbreviations: dc$^3$A, 2′-deoxy-3-deazaadenosine (also abbreviated in sequences as "a"); dm$^3$2P, 5-(1,2-dideoxy-$\beta$-D-*erythro*-pentofuranosyl)-3-methyl-2-pyridone (also abbreviated in sequences as "t"); CD, circular dichroism; $T_M$, thermal melting point.

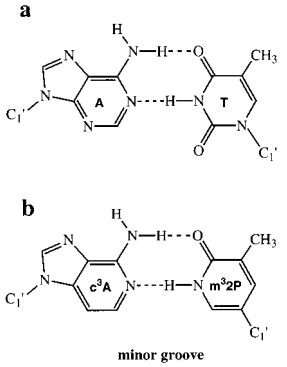


FIGURE 1: Illustrations of (a) a native dA-dT base pair and (b) an analogue dc$^3$A-dm$^3$2P base pair.

*Circular Dichroism Studies.* CD spectra were obtained at 3 or 6 $\mu$M duplex in 20 mM sodium phosphate buffer (pH 7.0) with varying concentrations of sodium chloride and/or trifluoroethanol usually at 5 °C. The spectrum of the buffer was obtained separately in all cases and then subtracted from the spectrum obtained with the present sample.

## RESULTS

The DNA duplexes examined in this study contained one or more modified dA-dT base pairs in which the minor groove functional groups were absent. This effect was achieved by using nucleosides with a 3-deazaadenine base residue (dc$^3$A) in place of dA and a 3-methyl-2-pyridone heterocycle (dm$^3$2P) in place of dT (Figure 1). Even with the formation of a conventional W-C base pair, the absence of purine N3 and pyrimidine O2 functional groups will eliminate ordered hydration in the minor groove as well as the possibility of localized metal ions.

The phosphoramidite derivatives of the two modified nucleosides could be incorporated into the growing oligonucleotide strand with coupling efficiencies that were essentially indistinguishable from those of the common nucleoside phosphoramidites. Deprotection of the dm$^3$2P-containing strands began with a treatment with DBU/pyridine to remove the *p*-nitrophenylethyl group and unmask the carbonyl. The length of time necessary for complete deprotection of dm$^3$2P-containing sequences varied with sequence composition, but a reaction time of 48 h ensured that the protecting group had been quantitatively eliminated. Subsequently, a conventional ammonia treatment completed the removal of protecting groups and released the sequences from the solid support. Purification was accomplished by reversed-phase HPLC.

Nucleoside analyses helped to confirm the presence of the analogue residues in the final purified products. We typically performed two analyses of the oligonucleotides, both by reversed-phase HPLC after snake venom phosphodiesterase−alkaline phosphatase digestions. The first at pH 5.5 (phosphate buffer and elution with methanol) did not effectively resolve the two analogue residues from dG (data not shown), but this analysis did confirm that the *p*-nitrophenylethyl group(s) had been effectively removed from the dm$^3$2P residue(s). A second analysis using the same column but eluted isocratically in 0.1 M triethylamine/acetic acid (pH 8.0) and 5% acetonitrile provided better separation of all four nucleosides, but the dm$^3$2P residue was not as prevalent as the other residues (Figure 2, bottom trace). Reanalysis of the mixture with detection at 297 nm ($\lambda_{max}$ for dm$^3$2P) clearly showed the presence of the analogue residue (Figure 2, top trace).

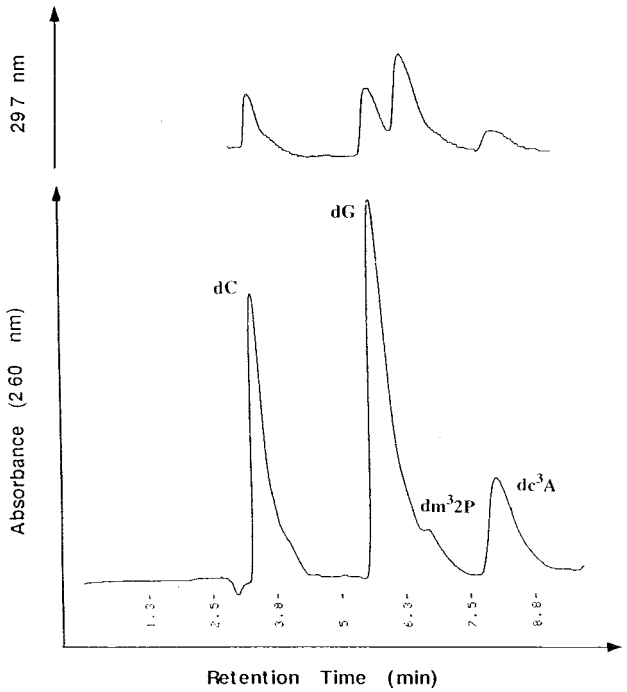


FIGURE 2: HPLC analysis of a small quantity of the sequence 5′-d[CGCG (c$^3$A)(c$^3$A)(m$^3$2P)(m$^3$2P) CGCG] after treatment with snake venom phosphodiesterase and alkaline phosphatase. HPLC conditions: buffer A [0.1 M triethylammonium acetate (pH 8.0)], 5% acetonitrile, and isocratic elution. Detection at 260 (bottom trace) and 297 nm (top trace).

Table 1: Sequences and $T_m$ Values for Self-Complementary Sequences Containing dc$^3$A and dm$^3$2P Substitution[a]

5′ d(C G C G **W X Y Z** C G C G)
3′ d(G C G C **W′X′Y′Z′**G C G C)

| | sequence[b] | | | | |
|---|---|---|---|---|---|
| entry | W-W′ | X-X′ | Y-Y′ | Z-Z′ | $T_M$ (°C) |
| 1 | A-T | A-T | T-A | T-A | 67 |
| 2 | A-T | A-**t** | **t**-A | T-A | 56 |
| 3 | A-**t** | A-**t** | **t**-A | **t**-A | hp[c] |
| 4 | **a**-**t** | **a**-**t** | **t**-**a** | **t**-**a** | hp |
| 5 | T-T | T-T | T-T | T-T | hp |
| 6 | A-T | G-C | C-G | T-A | 72 |
| 7 | A-**t** | G-C | C-G | **t**-A | 71 |

[a] $T_m$ values were obtained at a concentration of 3 $\mu$M in 20 mM $NaH_2PO_4$, 1 M NaCl, pH 7.0 buffer using an AVIV 14DS spectrophotometer. Samples were heated at a rate of 1.0 °C/min. [b] **a** is dc$^3$A and **t** dm$^3$2P. [c] hp means the sequence exists predominantly in the monomolecular hairpin conformation.

*Thermally Induced Transitions.* We initiated our study of duplexes containing the analogue base pairs with the preparation of the Dickerson dodecamer with the central core sequence d(AATT)$_2$ replaced with four dc$^3$A-dm$^3$2P analogue base pairs d(aatt)$_2$ (Table 1, entry 4). At a relatively high concentration (16 $\mu$M) of dodecamer, two absorbance versus temperature transitions were observed (Figure 3a). The early transition occurred near 20 °C and the latter near 70 °C. At a lower concentration (3 $\mu$M), the early transition disappeared and only a single transition was observed near 70 °C. This behavior is similar to that reported for sequences with a

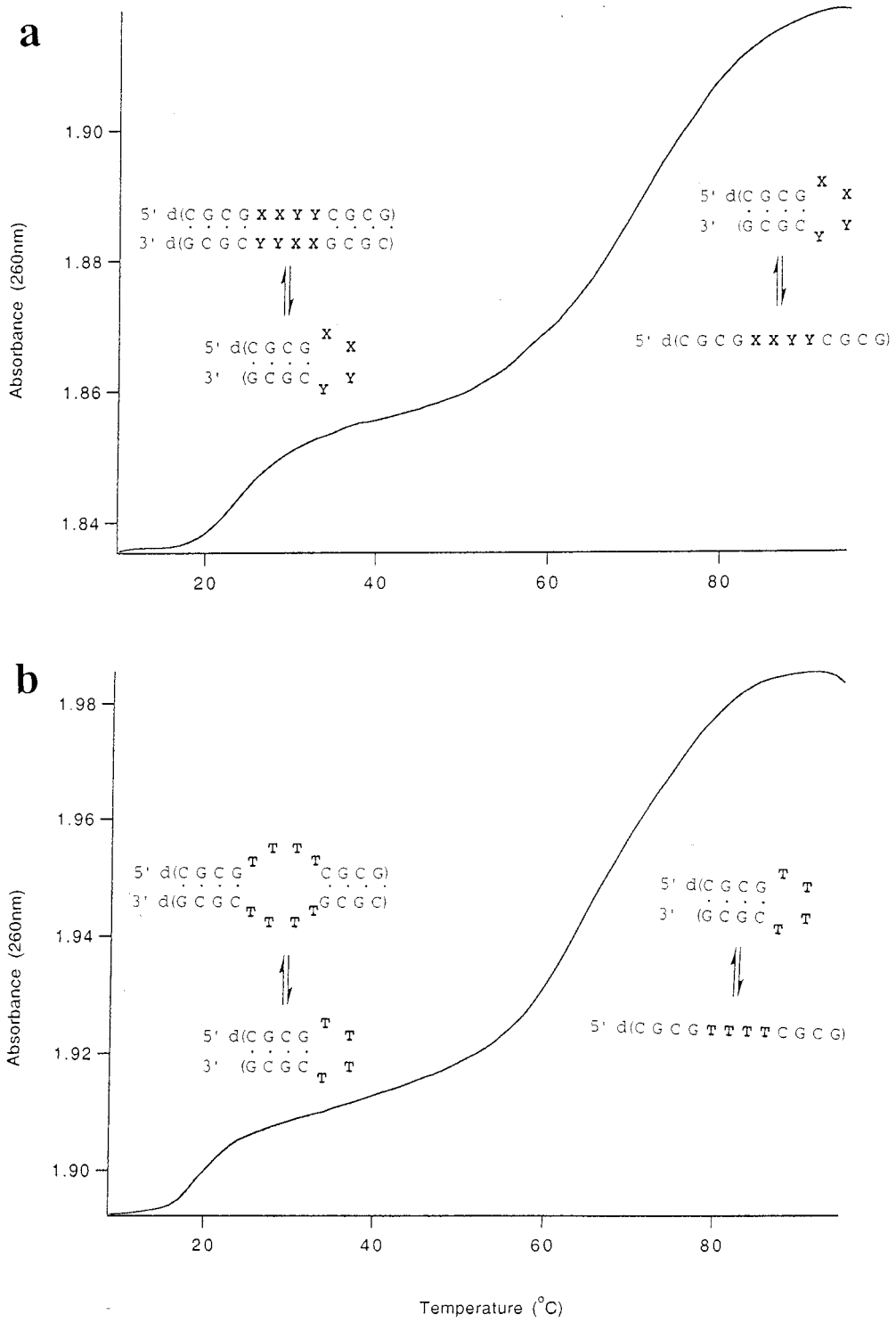


FIGURE 3: Thermal denaturation ($T_M$) analyses at a total strand concentration of 16 $\mu$M in 20 mM $NaH_2PO_4$ (pH 7.0) and 1 M NaCl: (a) 5′-d[CGCG ($c^3$A)($c^3$A)($m^3$2P)($m^3$2P) CGCG] and (b) 5′-d(CGCG TTTT CGCG).

central core of non-Watson−Crick base pairs (*28*), and in fact, similar behavior was observed (Figure 3b) for a self-complementary dodecamer with a core of dT residues (Table 1, entry 5). These $T_M$ experiments suggest that in the presence of a core of four dc$^3$A-dm$^3$2P analogue base pairs, the self-complementary Dickerson dodecamer prefers the monomolecular hairpin complex rather than the bimolecular simple duplex. Evidence for this observation was obtained from nondenaturing PAGE analyses under conditions similar to those of the $T_M$ study (Figure 4). The analogue sequence (lane 6) has a gel mobility that is essentially identical with that of the hairpin sequence (containing the dT loop, LTlane 5) and has much greater mobility than the native Dickerson dodecamer (lane 3).

One possible explanation for the observed preference for the hairpin conformation with a core sequence of analogue base pairs is that despite the careful design of the two analogue nucleosides, base pairing and base stacking effects are dramatically affected such that a non-base-paired core sequence results. We have examined computationally (*23*) both interstrand hydrogen bonding and changes in dipole moment as indicators of significant changes in base pairing

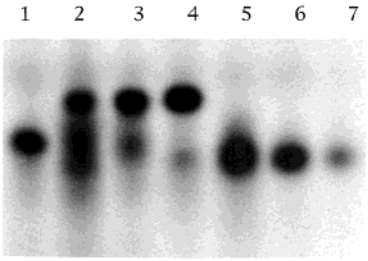


FIGURE 4: Nondenaturing PAGE analysis, with 50 mM Tris-HOAc (pH 7.0) and 10 mM $Mg(OAc)_2$: lane 1, 5′-d(CCGG AATT CGCC); lane 2, 5′-d(CCGG AATT CGCC)·5′-d(GGCG AATT CCGG); lane 3, 5′-d[CCGG A($c^3$A)TT CGCC]·5′-d[GGCG AA-($m^3$2P)T CCGG]; lane 4, 5′-d(CGCG AATT $CGCG)_2$; lane 5, 5′-d(CGCG TTTT CGCG); lane 6, 5′-d[CGCG ($c^3$A)($c^3$A)($m^3$2P)-($m^3$2P) CGCG]; and lane 7, sample from lane 6 diluted with 10× 5′-d(CGCG AATT CGCG).

Table 2: Sequences and $T_m$ Values for Non-Self-Complementary Sequences Containing $dc^3A$ and $dm^32P$ Substitution[a]

5′ d(C C G G **W** **X** **Y** **Z** C G C C)
3′ d(G G C C **W′** **X′** **Y′** **Z′** G C G G)

| | sequence[b] | | | | |
|---|---|---|---|---|---|
| entry | W-W′ | X-X′ | Y-Y′ | Z-Z′ | $T_M$ (°C) |
| 1 | A-T | A-T | T-A | T-A | 63 |
| 2 | A-T | **a**-T | T-A | T-A | 62 |
| 3 | A-T | A-**t** | T-A | T-A | 62 |
| 4 | A-T | **a**-**t** | T-A | T-A | 60 |
| 5 | A-T | **a**-**t** | **t**-**a** | T-A | 53 |
| 6 | **a**-**t** | **a**-**t** | **t**-**a** | **t**-**a** | 37 |
| 7 | A-**t** | A-**t** | T-**a** | T-**a** | 52 |
| 8 | A-T | A-**t** | **t**-A | T-A | 56 |

[a] $T_m$ values were obtained at a concentration of 3 $\mu$M in 20 mM $NaH_2PO_4$, 1 M NaCl, pH 7.0 buffer using an AVIV 14DS spectrophotometer. Samples were heated at a rate of 1.0 °C/min. [b] **a** is $dc^3$A and **t** $dm^32$P.

and stacking properties of the analogues, but found no dramatic differences. As a control experiment, we prepared a second pair of native and/or analogue sequences. The native sequence contained most of the elements of the Dickerson dodecamer, but the central dA-dT and dT-dA base pairs were replaced with dG-dC and dC-dG, respectively. The final two dA-dT base pairs are now located within a dG-dC rich sequence (Table 1, entry 6). The analogue sequence replaces the two dT residues with $dm^32$P (dt) residues (Table 1, entry 7). $T_M$ measurements (Table 1) and thermodynamic parameters (*23*) indicate that both the native and analogue sequences have thermal stabilities that are essentially identical. With the same two dt residues present within the core $d(AATT)_2$ sequence of the Dickerson dodecamer (Table 1, entry 2), a single transition resulted at all the concentrations that were examined, suggesting that the sequence exists as the duplex rather than the monomolecular hairpin, but with a $T_M$ reduced by 11 °C. Further substitutions of the dt residue for the remaining two dT residues (Table 1, entry 3) again resulted in a complex that prefers the monomolecular hairpin rather than the duplex.

We prepared a second series of non-self-complementary sequences that contain the core $d(AATT)_2$ sequence (Table 2). While the presence of a single $dc^3$A or $dm^32$P residue in the core sequence did not have any significant effect on the observed $T_M$ values (Table 2, entries 2 and 3), the presence of a $dc^3$A-$dm^32$P base pair (Table 2, entry 4) resulted in a moderate decrease in $T_M$ ($\Delta T_M = -3$ °C). If the duplex contained two adjacent analogue base pairs (Table 2, entry 5) the $T_M$ value was reduced by 10 °C, and with the entire $d(AATT)_2$ core sequence replaced with analogue base pairs (Table 2, entry 6) the $T_M$ value dropped to 37 °C ($\Delta T_M = -26$ °C). This latter effect was much more dramatic than if simply one strand in the core sequence was replaced with the analogue residues (Table 2, entry 7). The substitution of two dt residues into the non-self-complementary sequence reduced the $T_M$ value about the same amount as that observed for the self-complementary sequence (compare entry 8 in Table 2 with entry 2 in Table 1).

Since the dt residue exhibits a maximum absorbance at 297 nm ($\lambda_{min} = 255$ nm), it was advantageous to monitor thermal melting effects for the fully substituted dodecamer (Table 2, entry 6) at 260 and 297 nm (Figure 5). The 7 °C difference in observed $T_M$ values suggests that unstacking of the $dm^32$P residues, and presumably the disruption of the central base pairs, occurs as a relatively early event along the denaturation pathway. By comparison, no differences were noted for $T_M$ values obtained at the two different wavelengths for the dG-dC rich sequence (Table 1, entry 7).

*Circular Dichroism.* We have examined the circular dichroism spectra for both native and analogue sequences to determine whether the presence of the $dc^3$A-$dm^32$P base pairs results in any change in helix conformation. The native dodecamer sequence (Table 2, entry 1) exhibits a conventional B-type (i, Figure 6a). The spectra of all sequences containing $dm^32$P residue(s) exhibited some complexity due to the fact that the absorbance maximum for this residue is centered at 297 nm (and $\lambda_{min} = 255$ nm). With the presence of one analogue base pair (Table 2, entry 4), the spectrum remains largely unchanged, but with two $dc^3$A-$dm^32$P base pair substitutions (Table 2, entry 5), the negative CD transition loses intensity and there are two positive transitions present (ii, Figure 6a). The blue shift of the stronger positive transition to 262 nm and the decrease in intensity of the negative transition near 250 nm are more typical of A-form helices than corresponding spectral characteristics of B-form helices. The smaller positive transition (285 nm) as well as the new negative transition near 300 nm may reflect ellipticity effects resulting from the presence of two $dm^32$P residues. With the incorporation of four analogue base pairs into the core sequence (Table 2, entry 6), the positive transition at 262 nm increases further (iii, Figure 6a) while the second positive transition has largely disappeared. The spectra of this complex are distinctly different from the classic B-form CD spectra, and more like those of an A-form helix.

*Salt Effects.* A number of recent studies have indicated that the minor groove functional groups or the ordered water molecules at these sites are coordination sites for monovalent metal ions, and that these localized metal ions in the minor groove may be critical for stability and conformational effects in A-tract sequences (*24*−*26*). We have examined the relationship of duplex stability and sodium ion concentrations for a series of sequences, including the native dodecamer (Table 2, entry 1), a sequence with a single da-dt base pair (Table 2, entry 4), a sequence containing two dt residues (Table 2, entry 8), a sequence containing two analogue da-dt base pairs (Table 2, entry 5), and a sequence with the

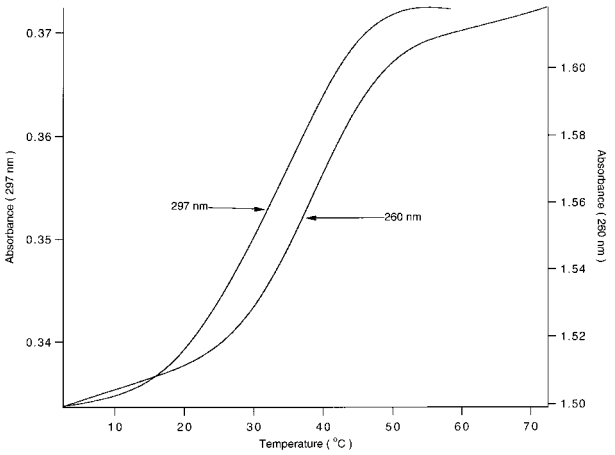


FIGURE 5: Thermal denaturation ($T_M$) analyses at a total strand concentration of 3 $\mu$M in 20 mM $NaH_2PO_4$ (pH 7.0) and 1 M NaCl for the sequence containing four analogue dc$^3$A-dm$^3$2P base pairs (Table 2, entry 5) with detection at 260 and 297 nm.

fully modified core sequence (Table 2, entry 6). The first three sequences (Table 2, entries 1, 4, and 8) all exhibited $T_M$ versus $Na^+$ concentration plots that are parallel (Figure 7), suggesting that the sequences all have linear relationships between $T_M$ and salt concentration. A similar relationship was also observed for the sequence containing two da-dt base pairs at sodium concentrations of $\leq$220 mM [Figure 7 (△)]. At >220 mM, the relationship became biphasic with $T_M$ values exhibiting a reduced sodium ion dependence and suggesting a conformational change in the complex. Similarly, the sequence with four da-dt base pairs [Figure 7 (■)] exhibited a relationship that was similar to that observed for the sequence with two da-dt base pairs at sodium concentrations of >220 mM. To confirm the nature of this conformational change, we examined the CD spectra of sequence 5 (Table 2) (that exhibiting a biphasic salt dependence) at both low and high salt concentrations (Figure 6b). Both spectra exhibit two positive transitions, but the intensities of the transitions differed, suggesting that the increased salt concentrations resulted in a significant conformational change in the duplex.

*Studies under Low Relative Humidity.* We have examined the CD spectra of the sequence with two analogue base pairs (Table 2, entry 5) at pH 7.0 and 100 mM sodium chloride under conditions of varied trifluoroethanol concentrations. Under conditions of 0% trifluoroethanol, the observed CD spectrum exhibits the two positive transitions characteristic for this sequence (i, Figure 6c); the primary positive transition is at 262 nm and the secondary transition at 285 nm. As the alcohol concentration is increased, the primary transition increases in intensity, while the secondary transition is slightly red shifted and decreases in intensity until at 65% alcohol there is but a single transition (iv, Figure 6c). This latter spectrum is very similar (see the inset of Figure 6c) to that obtained for the sequence containing the fully modified da-dt core (Table 2, entry 6). Further increases in the content of trifluoroethanol resulted in denaturation of the duplex.

## DISCUSSION

Probing the nature of the minor groove functional groups by examining sequences in which the functional groups of interest have been removed requires the design and preparation of analogues that function in general like native residues, but lack the functional groups of interest. In the study presented here, we have chosen dc$^3$A for the dA analogue and the C-nucleoside dm$^3$2P as the dT analogue. The dc$^3$A residue is a ring-fused derivative of 2-aminopyridine. Spectral studies (*27*) and acid dissociation (*28*) constants of 2-aminopyridines have established that these derivatives prefer the amino tautomeric form. The dm$^3$2P residue is a 2-pyridone. UV studies have shown (*29*) that the keto rather than the enol form is preferred. On the basis of the preferred tautomeric structures, this pair of base residues should exhibit

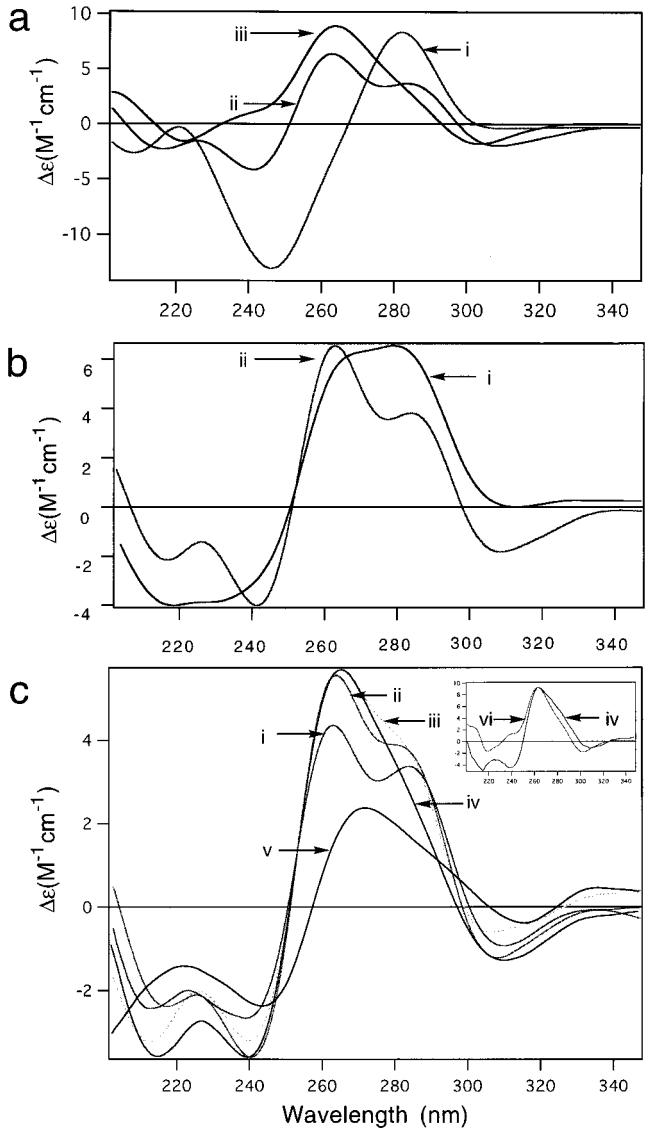


FIGURE 6: (a) Circular dichroism spectra at a total strand concentration of 6 $\mu$M in 20 mM $NaH_2PO_4$ (pH 7.0) and 1 M NaCl: (i) the native sequence (Table 2, entry 1), (ii) the sequence containing two dc$^3$A-dm$^3$2P base pairs (Table 2, entry 5), and (iii) the sequence containing four dc$^3$A-dm$^3$2P base pairs (Table 2, entry 6). (b) Circular dichroism spectra at a total strand concentration of 6 $\mu$M in 20 mM $NaH_2PO_4$ (pH 7.0) for the sequence containing two dc$^3$A-dm$^3$2P base pairs (Table 2, entry 5) in (i) 0.01 M NaCl and (ii) 1.0 M NaCl. (c) Circular dichroism spectra at a total strand concentration of 3 $\mu$M in 20 mM $NaH_2PO_4$ (pH 7.0) and 100 mM NaCl for the sequence containing two dc$^3$A-dm$^3$2P base pairs (Table 2, entry 5) (i), and additionally containing (ii) 40% trifluoroethanol, (iii) 60% trifluoroethanol, (iv) 65% trifluoroethanol, and (v) 70% trifluoroethanol. The inset shows a comparison of the spectrum with 65% trifluouroethanol (iv) with that of the fully modified 12-mer in the absence of trifluoroethanol (vi).

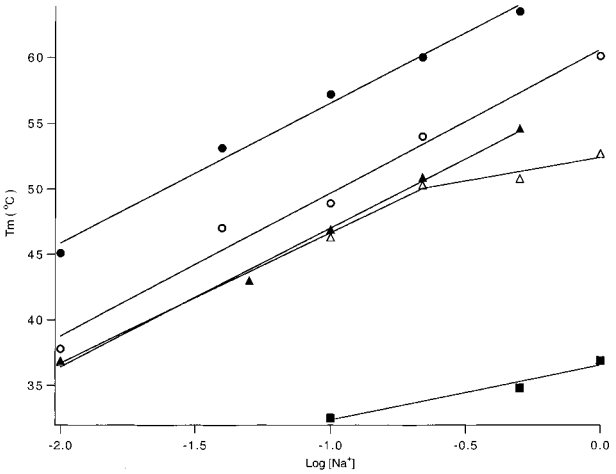


FIGURE 7: Plots of $T_M$ vs sodium ion concentration at a total strand concentration of 3 $\mu$M in 20 mM $NaH_2PO_4$ (pH 7.0): (●) the native sequence (Table 2, entry 1), (○) the sequence containing one dc$^3$A-dm$^3$2P base pair (Table 2, entry 4), (▲) the sequence containing two dm$^3$2P residues (Table 2, entry 8), (△) the sequence containing two dc$^3$A-dm$^3$2P base pairs (Table 2, entry 5), and (■) the sequence containing four dc$^3$A-dm$^3$2P base pairs (Table 2, entry 6).

complementary hydrogen bonding interactions and mimic the dA-dT base pair as illustrated in Figure 1. Interactions that interfere with complementary hydrogen bonding usually lead to dramatic duplex destabilization (*30*, *31*). In addition to hydrogen bonding interactions, inter-residue base stacking effects are critical to effective duplex formation. Base stacking interactions for adjacent heterocycles are more difficult to quantify. It appears that heteroaromatics exhibit better stacking interactions than simple aromatic compounds, but the reasons for these differences are unclear. Both dipole−dipole interactions (*32*) and the hydrophobic effect (*33*) probably contribute to the parameters defining base stacking in aqueous solutions. We have performed calculations on dipole moments for both analogues, and while some changes are present, the average for the base pair remains largely unchanged (*23*).

*Stability Effects.* The incorporation of two dm$^3$2P residues into a dG-dC rich sequence results in a $T_M$ value that is essentially identical with that of the native dT-containing sequence (Table 1, entries 6 and 7). Thermodynamic parameters determined for both of these sequences were also identical (*23*). This similarity suggests that the dm$^3$2P analogue forms a Watson−Crick base pair and provides inter-residue stacking interactions that are at least similar to those present with the native dT residue. Both characteristics have now been confirmed by X-ray crystal structure analysis (*34*). By comparison, the presence of two dm$^3$2P residues in a dA-dT rich sequence results in a decrease in $T_M$ by some 11 °C (compare entries 1 and 2 in Table 1). The loss of duplex stability when the analogue base pair is present in a dA-dT rich sequence, and the absence of effects when present in a dG-dC rich sequence, suggests that the analogue does not affect the fundamental properties of interstrand hydrogen bonding and base stacking since such effects would destabilize sequences rich in dG-dC base pairs as well as those that are dA-dT rich. Instead, the analogues must disrupt a structure-stabilizing phenomenon that is present in dA-dT rich sequences but not in dG-dC sequences. The most likely parameter that has been altered by the substitution of these particular analogues is the nature of hydration in the minor groove. From the very first crystal structure reported by Dickerson (*1*), it has been suggested that dA-dT rich sequences have a well-ordered spine of hydration in the minor groove. One set of water molecules bridges the N3 atoms of the dA residues and/or the O2 carbonyls of the dT residues, while a second layer of water connects the underlying first layer of hydration forming a continuous spine of hydration. This structure is absent in dG-dC rich sequences in part because the presence of the guanine N2-amino group in the minor groove interferes with the formation of this ordered spine of hydration (*10*). More recently, this hydration structure has been observed at higher resolution in a number of crystal structures (*12*, *13*, *17*, *35*). The results presented here suggest that this water structure within the minor groove is critical to duplex stability. Previous theoretical studies (*36*, *37*) as well as NMR analyses (*38*, *39*) have also suggested that a stable spine of hydration may be necessary for effective B-form duplex formation. The presence of the cross-strand hydrogen bonding by water molecules in the minor groove may stabilize the B-form duplex by inhibiting base pair opening effects as has been suggested on the basis of theoretical calculations (*34*).

A number of analyses by both NMR (*26*) and high-resolution crystallography (*11*, *17*, *18*, *35*) indicate that monovalent cations can substitute for selected water molecules within the hydration structure in the minor groove, but a related structure (*13*) reports the absence of such localized metal ions. Molecular dynamics simulations have also suggested that such cations could partition into the minor groove (*40*−*42*). In such positions, they could conceivably provide some cross-strand phosphate shielding and thereby enhance the stability of the duplex structure. However, the current studies cannot assist in resolving the issue of metal ion binding within the minor groove of the duplex DNA.

The introduction of increasing numbers of da-dt base pairs into a dodecamer sequence results in decreases in $T_M$ values of 3, 10, and 26 °C for one, two, and four analogue base pairs, respectively. Consistent with the suggestion that the analogue base pairs lacking minor groove hydration give rise to the destabilization of the duplexes are the thermal melting profiles for the sequence with four da-dt base pairs obtained at 260 and 297 nm. The $T_M$ obtained at 297 nm ($\Delta T_M = -7$ °C) is quite specific for hypochromicity effects of the dm$^3$-2P residue. When present in a dA-dT rich sequence, the unstacking $T_M$ for the dm$^3$2P residue suggests that the dA-dT core sequence is the site at which thermal disruption of the base pairs begins, consistent with the observation that in the absence of the minor groove functional groups and/or associated hydration or metal ion binding, the duplex structure in dA-dT rich sequence is significantly destabilized. The fact the central core sequence undergoes duplex melting before the ends of the duplex is not unusual. NMR studies based upon monitoring the imino proton resonances have shown that with central G-U mismatches, the center of the duplex melts before the termini (*43*, *44*).

*Conformational Effects.* The destabilization of dA-dT rich sequences lacking minor groove sites for hydration and metal ion binding is supported by the conformational preferences of self-complementary sequences. The Dickerson dodecamer containing a core of analogue residues, or even just the four dm$^3$2P residues, undergoes a conformational transition whereby the monomolecular hairpin conformation is pref-

erentially adopted. Similar transitions between duplex and hairpin structures are known, but shifting the equilibrium toward the hairpin usually requires the introduction of a base pair mismatch into the duplex (*45*, *46*) such that after the transition it is located in the loop section of the hairpin. The preferential formation of the hairpin with these sequences suggests that the core of analogue base pairs is destabilized relative to native dA-dT base pairs, consistent with the reported thermodynamic parameters (*23*) as well as the $T_M$ value obtained at 297 nm.

Hydration of nucleic acids in general has been related to conformational preferences for double-stranded DNA (*10*, *14*−*16*). The explanation of the observation that B → A transitions in helical DNA only occur in sequences rich in dG-dC base pairs, and cannot be facilitated in poly(dA)-poly(dT), has relied upon hydration differences (*10*). Poly-(dA)-poly(dT) maintains an ordered and very stable minor groove spine of hydration, first suggested with the original structural analysis of the Dickerson dodecamer (*1*) and confirmed with recent high-resolution structures (*12*, *13*, *17*). This spine of hydration must be disrupted to facilitate the formation of the A-form conformation with an alternative mode of ordered hydration involving chains of water molecules bridging the major groove and loss of overal hydration (*14*−*16*). Disrupting the minor groove spine of hydration usually involves both increasing the number of dG-dC base pairs and reducing the water activity of the solution. Salt-induced effects are usually the result of multivalent ions such as spermine or cobalt complexes (*5*), but with the analogue sequences presented here, a more facile transition occurs even with increasing concentrations of a simple monovalent cation like sodium. As with organic solvent effects, dG-dC rich sequences often require ethanol or trifluoroethanol concentrations of 75% for a stable A-conformation to result (*7*−*9*), but with the present sequence containing two analogue base pairs, the A-like conformation is obtained at the lower concentration of 65% organic solvent (iv, Figure 6c). At 65% trifluoroethanol, the spectrum of the partially modified sequence $d(GAatTC)_2$ was very similar to that of the sequence containing a fully modified core $d(GaattC)_2$ obtained in the absence of organic solvent (vi, Figure 6c, inset).

The present group of minor-groove modified sequences do not adopt a conventional B-form (Figure 6a). And while the conformations of the analogue sequences are not strictly speaking A-form, with increasing numbers of analogue base pairs the CD spectra resemble A-like conformations much more than B-like conformations (Figure 6a). If the absence of the minor groove functional groups results in a preference for A-like conformations, then analogue sequences with only partial losses of those functional groups might be more easily transformed to A-like structures under conditions or low water activity.

Selected modifications have also facilitated B to A transitions, most notably, the introduction of a 3′-phosphoramidate linkage (*20*, *21*). But in that case, the transformation was likely induced because of the preference for the modified carbohydrate to adopt a C3′-endo conformation typical of A-form structures, rather than the C2′-endo conformation preferred in B-form duplexes. Although in other cases, B → A transitions appear to be possible without corresponding changes in the conformation of the sugar (*9*), suggesting as noted previously (*2*) that the effects of sugar puckering may more often be a consequence rather than a cause of the B → A transition.

## SUPPORTING INFORMATION AVAILABLE

Synthetic procedures and schemes for the preparation of the $dm^32P$ phosphoramidite. This material is available free of charge via the Internet at http://pubs.acs.org.

## REFERENCES


1. Drew, H. R., and Dickerson, R. E. (1981) *J. Mol. Biol. 151*, 535−556.
2. Calladine, C. R., and Drew, H. R. (1984) *J. Mol. Biol. 178*, 773−782.
3. Ng, H. L., Kopka, M. L., and Dickerson, R. E. (2000) *Proc. Natl. Acad. Sci. U.S.A. 97*, 2035−2039.
4. Minyat, E. E., Ivanov, V. I., Kritzyn, A. M., Minchenkova, L. E., and Schyolkina, A. K. (1978) *J. Mol. Biol. 128*, 397−409.
5. Robinson, H., and Wang, A. H.-J. (1996) *Nucleic Acids Res. 4*, 676−682.
6. Xu, Q., Shoemaker, R. K., and Braunlin, W. H. (1993) *Biophys. J. 73*, 1039−1049.
7. Ivanov, k. I., MInchenkova, L. E., Minyat, E. E., Grank-Kamenitskii, M. D., and Schyolkina, A. K. (1974) *J. Mol. Biol. 97*, 817−833.
8. Zimmermann, S. B., and Pheiffer, B. H. (1979) *J. Mol. Biol. 135*, 1023−1027.
9. Trantirek, L., Stefl, R., Vorlickova, M., Koca, J., Sklenar, V., and Kypr, J. (2000) *J. Mol. Biol. 297*, 907−922.
10. Dickerson, R. E., Drew, H. R., Conner, B. N., Wing, R. M., Fratini, A. V., and Lopka, M. L. (1982) *Science 216*, 475−484.
11. Shui, X. Q., Sines, C. C., McFail-Isom, L., VanDerveer, D., and Williams, L. D. (1998) *Biochemistry 37*, 16877−16887.
12. Tereshko, V., Minasov, G., and Egli, M. (1999) *J. Am. Chem. Soc. 121*, 470−471.
13. Chiu, T. K., Kaczor-Grzeskowiak, M., and Dickerson, R. E. (1999) *J. Mol. Biol. 292*, 589−608.
14. Marky, L. A., and Kupke, D. W. (1989) *Biochemistry 28*, 9982−9988.
15. Rentzeperis, D., Kupke, D. W., and Marky, L. A. (1993) *Biopolymers 33*, 117−125.
16. Kankia, B. I., and Marky, L. A. (1999) *J. Phys. Chem. B 103*, 8759−8767.
17. Shui, X. Q., McFailIsom, L., Hu, G. G., and Williams, L. D. (1998) *Biochemistry 37*, 8341−8355.
18. Tereshko, V., Minasov, G., and Egli, M. (1999) *J. Am. Chem. Soc. 121*, 3590−3595.
19. Woods, K. K., McFail-Isom, L., Sines, C. C., Howerton, S. B., Stephens, R. K., and Williams, L. D. (2000) *J. Am. Chem. Soc. 122*, 1546−1547.
20. Ding, D. Y., Gryaznov, S. M., Lloyd, D. H., Chandrasekaran, S., Yao, S. J., Ratmeyer, L., Pan, Y. Q., and Wilson, W. D. (1996) *Nucleic Acids Res. 24*, 354−360.
21. Tereshko, V., Gryaznov, S., and Egli, M. (1998) *J. Am. Chem. Soc. 120*, 269−283.
22. Matteucci, M. D., and Caruthers, M. H. (1981) *J. Am. Chem. Soc. 103*, 3185−3191.
23. Lan, T., and McLaughlin, L. W. (2000) *J. Am. Chem. Soc. 122*, 6512−6513.
24. Manning, G. S., Ebralidse, K. K., Mirzabedov, A. D., and Rich, A. (1989) *J. Biomol. Struct. Dyn. 6*, 877−889.
25. Strauss-Soukup, J. K., and Maher, L. J. (1998) *Biochemistry 37*, 1060−1066.
26. Hud, N. V., Sklenar, V., and Feigon, J. (1999) *J. Mol. Biol. 286*, 651−660.
27. Anderson, J., and Seeger, I. (1943) *J. Am. Chem. Soc. 65*, 340−345.
28. Angyal, H. (1952) *J. Chem. Soc.*, 1461−1465.
29. Katritzky, A. R., and Lagowski, J. M. (1962) *Adv. Heterocycl. Chem. 1*, 131−141.

30. Rajur, S. B., and McLaughlin, L. W. (1992) *Tetrahedron Lett. 33*, 6081−6084.
31. Searls, T., and McLaughlin, L. W. (1999) *Tetrahedron 55*, 11985−11996.
32. Newcomb, L. F., and Gellman, S. H. (1994) *J. Am. Chem. Soc. 116*, 4993−4994.
33. Friedman, R. A., and Honig, B. (1995) *Biophys. J. 69*, 1528−1535.
34. Williams, L. E. Private communication.
35. Woods, K. K., McFail-Isom, L., Sines, C. C., Howerton, S. B., and Williams, L. D. (2000) *J. Am. Chem. Soc. 122*, 1546−1547.
36. Chen, Y. Z., and Prohofsky, E. W. (1993) *Biophys. J. 64*, 1385−1393.
37. Subramanian, P. S., Swaminathan, S., and Beveridge, D. L. (1990) *J. Biomol. Struct. Dyn. 7*, 1161−1165.
38. Liepinsh, E., Otting, G., and Wuthrich, K. (1992) *Nucleic Acids Res. 20*, 6549−6553.
39. Phan, A. T., Leory, J. L., and Gueron, M. (1999) *J. Mol. Biol. 286*, 505−519.
40. Young, M., and Beveridge, D. L. (1998) *J. Mol. Biol. 281*, 675−687.
41. Young, M. A., Jayaram, B., and Beveridge, D. L. (1997) *J. Am. Chem. Soc. 119*, 59−69.
42. Feig, M., and Pettitt, B. M. (1999) *Biophys. J. 77*, 1769−1781.
43. He, L., Kierzek, R., Santa Lucia, J., Walter, A. E., and Turner, D. H. (1991) *Biochemistry 30*, 11124−11132.
44. Chen, X., McDowell, J. A., Kierzek, R., Krugh, T. R., and Turner, D. H. (2000) *Biochemistry 39*, 8970−8982.
45. Avizonis, D. Z., and Kearns, D. R. (1995) *Biopolymers 35*, 187−200.
46. Chou, S. H., Zhu, L. M., and Reid, B. R. (1996) *J. Mol. Biol. 259*, 445−457.

BI0015212