# NMR Study of the Conformation of the 2-Aminopurine:Cytosine Mismatch in DNA†

Patricia A. Fagan,‡ Carme Fàbrega,§ Ramon Eritja,§ Myron F. Goodman,|| and David E. Wemmer*,‡

*Department of Chemistry, University of California at Berkeley and Structural Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, European Molecular Biology Laboratory, D-69117 Heidelberg, Germany, and Department of Biological Sciences, Hedco Molecular Biology Laboratories, University of Southern California, Los Angeles, California 90089-1340*

*Received November 7, 1995; Revised Manuscript Received January 22, 1996*⊗

ABSTRACT: DNA polymerase makes errors by misincorporating natural DNA bases and base analogs. Because of the wide variety of possible mismatches and the varying efficiency with which they are repaired, structural studies are necessary to understand in detail how these mispairs differ and can be distinguished from standard Watson−Crick base pairs. 2-Aminopurine (AP) is a highly mutagenic base analog. The objective of this study was to determine the geometry of the AP•C mispair in DNA at neutral pH. Although several studies have focused on the AP•C mispair in DNA, there is not as of yet consensus on its structure. At least four models have been proposed for this mispair. Through the use of NMR spectroscopy with selective $^{15}$N-labeling of exocyclic amino nitrogens on bases of interest, we are able to resolve ambiguities in previous studies. We find here that, in two different DNA sequences, the AP•C mispair at neutral and high pH is in a wobble geometry. The structure and stability of this base mispair is dependent upon the local base sequence.

Fidelity in replication is of the utmost importance in maintaining an error-free genome. Overall mutation rates are typically only $10^{-9}$−$10^{-11}$ (Drake et al., 1969; Saenger, 1983) per base replicated. DNA polymerase makes errors by misincorporating natural DNA bases and base analogs. Polymerase proofreading and mismatch repair machinery are needed to recognize and repair these mismatched base pairs. It is not clear by what mechanism the recognition of non-Watson−Crick base pairs occurs, but it has been suggested that DNA polymerase screens for the base-pairing free energy of the pair and the precise geometry of the Watson−Crick base pair (Kornberg, 1980; Echols & Goodman, 1991). Because of the wide variety of possible mismatches and the varying efficiency with which they are repaired, structural studies can contribute to understanding in detail how these mispairs differ and can be distinguished from standard Watson−Crick base pairs.

Normal replication is impaired when base analogs are incorporated into DNA, allowing mutations to arise through subsequent replication steps or repair processes. 2-Aminopurine (AP),[1] a highly mutagenic base analog, has been studied for decades [for a review, see Ronen (1979)]. AP is known to cause increased frequency of G•C → A•T and A•T → G•C transitions. The proposed pathway for such transitions requires that AP mispair with both thymine and cytosine (Freese, 1959; Ronen, 1979). As a result, much discussion about the cause of AP mutagenicity has focused on its base pairing interactions with pyrimidines (Watanabe & Goodman, 1982).

AP preferentially base pairs with thymine during replication (Bessman et al., 1974; Watanabe & Goodman, 1981). AP has been shown by NMR spectroscopy to form a stable mispair with thymine in a DNA oligomer, stabilized by two hydrogen bonds in a Watson−Crick geometry (Sowers et al., 1986). AP incorporation opposite thymine is favored by approximately 25-fold over cytosine (Watanabe & Goodman, 1981, 1982). Although several studies have focused on the AP•C mispair in DNA, there is not as of yet consensus on its structure. At least four models have been proposed for this interaction: a Watson−Crick geometry involving one hydrogen bond [Figure 1a (Ronen, 1979)], the imino tautomer of either C or AP, paired with the amino tautomer of the other base [Figure 1b (Janion & Shugar, 1973; Goodman & Ratliff, 1983)], a protonated cytosine resulting in Watson−Crick geometry, involving two hydrogen bonds [at neutral pH, Figure 1c (Sowers et al., 1986)], and a wobble geometry [Figure 1d (Sowers et al., 1989)].

The rare imino tautomer form of bases has long been used to explain spontaneous and base analog-induced mutation (Watson & Crick, 1953a,b; Freese, 1959; Topal & Fresco, 1976; Saenger, 1983). However, calculated equilibrium

† This work was supported in part by the National Institutes of Health Grant GM-43219 (D.E.W.) and through instrumentation grants from the US Department of Energy, DE FG05-86ER75281 and the National Science Foundation, DMB 86-09305 and BBS 87-20134, and by the Director, Office of Biological and Environmental Research, General Sciences Division of the US Department of Energy under Contract No. DE-AC03-76SF00098 (D.E.W.). M.F.G. and R.E. are grateful for support from National Institutes of Health Grants GM21422 and GM42554 (M.F.G.) and North Atlantic Treaty Organization Collaborative Research Grant 900554.

* Author to whom correspondence should be addressed.

‡ Structural Biology Division, Lawrence Berkeley National Laboratory, and Department of Chemistry, University of California, Berkeley, CA 94720.

§ European Molecular Biology Laboratory, D-69117 Heidelberg, Germany.

|| Department of Biological Sciences, Hedco Molecular Biology Laboratories, University of Southern California, Los Angeles, CA 90089-1340.

⊗ Abstract published in *Advance ACS Abstracts,* March 15, 1996.

[1] Abbreviations: AP, 2-aminopurine; NMR, nuclear magnetic resonance spectroscopy; 1D, one-dimensional; 2D, two-dimensional; NOESY, nuclear Overhauser effect spectroscopy; EDTA, ethylenediaminetetraacetic acid.

0006-2960/96/0435-4026$12.00/0 © 1996 American Chemical Society

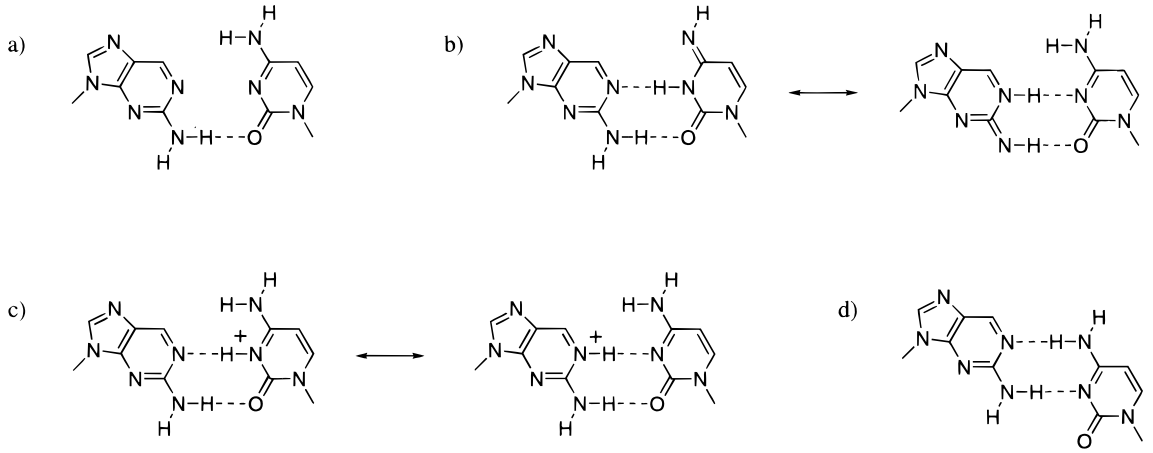

FIGURE 1: Possible structures for the 2-aminopurine·cytosine mispair: (a) Watson−Crick, (b) imino tautomer of either AP or C paired with the amino tautomer of the other base to form a Watson−Crick geometry, (c) protonated Watson−Crick, and (d) wobble.

constants of imino/amino tautomeric equilibria do not explain observed mutational rates (Ronen, 1979), and structural studies to date have found no evidence of the imino tautomer form of cytosine or 2-aminopurine in DNA oligomers. NMR evidence has been presented for an AP·C mispair which is protonated at the cytosine N3 position at neutral pH (Sowers et al., 1986). At high pH, $^{15}$N multiplicity and chemical shift data show normal amino tautomers in both bases and suggest that the AP·C mispair is in a wobble geometry (Sowers et al., 1989).

Owing to the highly fluorescent nature of AP (Ward et al., 1969), there has been a recent resurgence of interest in utilizing this analog as a "real-time" probe to study relaxation dynamics of DNA (Guest et al., 1991) and the pre-steady-state kinetics of DNA synthesis (Bloom et al., 1993; Frey et al., 1995) and proofreading (Bloom et al., 1994). In this study we address the question of the geometry of the AP·C mispair in DNA over a wide range of pH values and in the context of nearest neighboring base pairs. Through the use of selective $^{15}$N-labeling of exocyclic amino nitrogens on bases of interest, we are able to resolve ambiguities in previous studies. We find here that, in two different DNA sequences, the AP·C mispair at neutral and high pH is in a wobble geometry. The structure and stability of this base mispair is dependent upon the local base sequence.

## MATERIALS AND METHODS

*Synthesis and Purification of Oligonucleotides.* The sequences of the two DNA duplexes in this study are as follows:

Duplex I:

5'- $G_1$ $A_2$ $T_3$ $G_4$ $AP_5$ $G_6$ $T_7$ $A_8$ $C_9$ -3'

3'- $C_{18}$ $T_{17}$ $A_{16}$ $C_{15}$ $C_{14}$ $C_{13}$ $A_{12}$ $T_{11}$ $G_{10}$ -5'

Duplex II:

5'- $G_1$ $A_2$ $T_3$ $C_4$ $AP_5$ $C_6$ $T_7$ $A_8$ $C_9$ -3'

3'- $C_{18}$ $T_{17}$ $A_{16}$ $G_{15}$ $C_{14}$ $G_{13}$ $A_{12}$ $T_{11}$ $G_{10}$ -5'

where AP = 2-aminopurine.

In samples of duplexes I and II, $C_{14}$ was $^{15}$N-labeled at the exocyclic N4 position. A sample of duplex II was also synthesized which contained $AP_5$ $^{15}$N-labeled at the exocyclic N2 position and $C_{14}$ $^{15}$N-labeled at the exocyclic N4 position, by methods as published elsewhere (Acedo et al., 1994).

The oligonucleotides were purified by standard reverse-phase HPLC methods.

*Sample Preparation.* The concentrations of the single-stranded oligonucleotides were determined by UV absorbance. The extinction coefficients at 260 nm were calculated to be $0.92 \times 10^5$ $M^{-1}cm^{-1}$ and $0.84 \times 10^5$ $M^{-1}cm^{-1}$ for d(G-A-T-G-AP-G-T-A-C), assuming AP = G, and d(G-T-A-C-C-C-A-T-C), respectively. For duplex II, calculated extinction coefficients were $\epsilon_{260} = 0.85 \times 10^5$ $M^{-1}$ $cm^{-1}$ for d(G-A-T-C-AP-C-T-A-C) assuming AP = G, and $\epsilon_{260} = 0.90 \times 10^5$ $M^{-1}$ $cm^{-1}$ for d(G-T-A-G-C-G-A-T-C). Equimolar amounts of the complementary strands were mixed. Duplexed oligonucleotides were dialyzed using 1 kDa molecular mass cutoff cellulose tubing, against 300 mM NaCl, 30 mM NaCl, and then deionized water for 24 h each.

Oligonucleotide duplex samples were dissolved in 0.5 mL of 10 mM potassium phosphate buffer, 100 mM NaCl (pH 7.0), and 0.1 mM EDTA, lyophilized, and redissolved in 0.5 mL of 99.96% $D_2O$ or 90% $H_2O$/10% $D_2O$. The pH of the samples was adjusted using 0.1 M or 0.01 M HCl and 0.1 M or 0.01 M NaOH. The concentration of the duplex I NMR sample was 1.0 mM and that of the 4-$^{15}NH_2$-$C_{14}$ containing duplex II NMR sample was 1.8 mM. A sample of duplex II which contained 4-$^{15}NH_2$-$C_{14}$ and 2-$^{15}NH_2$-$AP_5$ was 1.0 mM in concentration.

*NMR Experiments and Resonance Assignments.* NMR experiments were performed on a 500 MHz GE Omega spectrometer or on a Bruker AMX-600 spectrometer. $^{15}$N spectra were acquired at 60 MHz and 25 °C using a broad band probe, with a spectral width of 200 ppm and 15 000−23 000 scans, using a relaxation delay of 2 s. $^{15}$N spectra were referenced downfield of neat $^{15}NH_3$ (Live et al., 1984). $^1$H spectra in $H_2O$ were acquired with a spectral width of 20 ppm and 256 scans, using a relaxation delay of 2 s. One-dimensional $^{15}$N-filtered $^1$H spectra in $H_2O$ were acquired with or without $^{15}$N decoupling during acquisition (Sklenar & Bax, 1987). 1D NOE experiments were performed on the AMX-600, using duplex II in $H_2O$ at 5 °C and pH 8.7, with a spectral width of 22 ppm and 1024−2048 scans each for the reference and saturated spectra. pH titrations were performed on samples of duplexes I and II in 150 mM NaCl and monitored by 1D $^1$H NMR at 15 °C. Typically, the

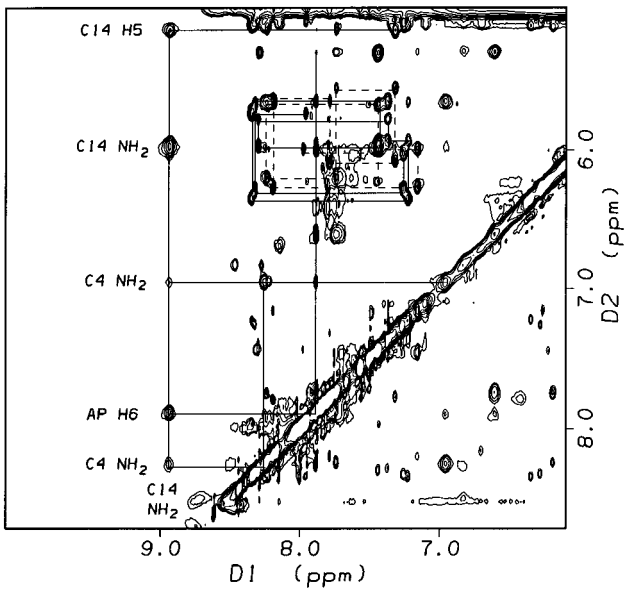


FIGURE 2: Expansion of the amino and aromatic to deoxyribose H1′ region of a $^{15}$N-decoupled $^{1}$H−$^{1}$H NOESY spectrum of $^{15}NH_2$-$C_{14}$ labeled duplex II (in $H_2O$ solution, 15 °C, pH 7.7, 175 mM NaCl, $\tau_M$ = 200 ms). Various NOE cross peaks to the mispaired $C_{14}$ amino group are labeled.

intense solvent resonance was suppressed using a 1:1 read pulse (Kime & Moore, 1983).

2D NOESY spectra of duplex I in $H_2O$ were acquired at 20 °C (pH 6.5) and 26 °C (pH 7.0), with the mixing time $\tau_M$ = 200 ms. A 2D NOESY spectrum in $D_2O$ was acquired at 20 °C (pH 6.5) with $\tau_M$ = 200 ms. 2D NOESY spectra of duplex II in $H_2O$ were acquired at 15 °C and 20 °C (pH 7.7). A $^{15}$N-decoupled 2D NOESY of duplex II in $H_2O$ was acquired at 10 °C with $\tau_M$ = 150 ms. Spectral widths were typically 14 000 Hz in $H_2O$ and 6000 Hz in $D_2O$ solution. Between 450 and 600 $t_1$ increments were acquired in the NOESY experiments.

All spectra were acquired using time-proportional phase incrementation (TPPI) mode (Marion & Wüthrich, 1983). The data were processed with FELIX v. 2.30 (Biosym, San Diego) on a Silicon Graphics workstation. Skewed sine bell functions were used for apodization of the free induction decays. A convolution method was applied to remove the intense $H_2O$ resonance from the frequency spectrum. Assignments of the DNA resonances were made using standard sequential methods (Hare et al., 1983; Wüthrich, 1986).

## RESULTS

*Nonexchangeable Proton Resonance Assignments.* Part of a NOESY spectrum of duplex II in $H_2O$ is shown in Figure 2. The chemical shifts of the DNA base and sugar H1′ resonances are listed in Table 1. Intraresidue deoxyribose NOE cross peak patterns and intensities and interresidue NOE cross peak patterns indicate that most of the DNA is near the standard B-form. The mispaired bases stack in the helix with some minor distortion of the DNA helix in the region of the mispair.

*Resonance Assignment of Mismatch Protons.* The $^{15}$N-bound protons could be identified in the 1D $^{1}$H spectrum because they appear as doublets split by 92 Hz, the $^{15}$N−$^{1}$H one bond scalar coupling constant. These proton resonances could be more clearly seen in $^{15}$N-filtered 1D experiments, in which only resonances of $^{15}$N-bound protons are observed.

In 4-$^{15}NH_2$-$C_{14}$ labeled duplex I, two proton resonances are observed in $^{15}$N-filtered 1D experiments at 9.16 and 6.26 ppm and in labeled duplex II at 8.94 and 5.98 ppm. The line widths of these resonances are dependent on both temperature and pH. In 2-$^{15}NH_2$-$AP_5$ labeled duplex II, two AP proton resonances are observed in the $^{15}$N-filtered 1D spectrum at 8.47 and 6.83 ppm. These resonances are broadened into the baseline at 25 °C but sharpen with decreasing temperature. The line widths of the AP 2-amino proton resonances do not appear to be pH-dependent. No evidence is seen for any extra imino-type protons, which would suggest that one of the bases has a significant population of imino tautomeric form. Because labeled $^{15}NH_2$-$AP_5$ duplex I was not available, the $AP_5$ $NH_2$ resonances were not assigned in duplex I.

The specific assignment of the two 4-amino protons of $C_{14}$ is determined by the relative intensities of the NOE between the amino protons and the $C_{14}$ H5 proton. In both duplexes, the stronger NOE is observed between the H5 and the upfield amino proton resonance (near 6 ppm in both duplexes). This information dictates that the upfield amino resonance be assigned to the proton nearest to the H5, and the downfield amino resonance (near 9 ppm in both duplexes) be assigned to the amino proton farthest from the H5, on the base pairing edge of the base. The $C_{14}$ amino resonance near 9 ppm is downfield relative to single-stranded cytosine amino protons (7.0−8.0 ppm) and is even downfield relative to the hydrogen bonded cytosine amino protons in standard Watson−Crick base pairs (8.0−8.6 ppm).

The $AP_5$ H6 resonance was assigned in both duplexes I and II (see Table 1). This proton was assigned through the use of spectra acquired in $H_2O$, since there are no other nonexchangeable protons nearby. In both duplexes I and II, the H6 has an clear, medium intensity NOE cross peak to the downfield $C_{14}$ amino proton, a weaker NOE cross peak to the upfield $C_{14}$ amino proton, and an even weaker NOE to the $C_{14}$ H5. This pattern of NOE intensities to the two $C_{14}$ $NH_2$ protons is consistent with the assignment of $AP_5$ H6, since it is the only unassigned nonexchangeable proton nearby, and it is closer to the $C_{14}$ $NH_2$ proton on the base pairing edge of the cytosine base than to the upfield (non-hydrogen-bonded) $C_{14}$ $NH_2$ proton.

In both duplexes, it is evident that the intensities of the $C_{14}$ H6−H1′ and $AP_5$ H8−H1′ cross peaks are weak and of similar intensity as the other aromatic−H1′ cross peaks, suggesting that $AP_5$ and $C_{14}$ are *anti*, which excludes unusual geometries such as those found in Hoogsteen or reverse-Hoogsteen base pairs.

*Comparison of NOE Patterns in the Mismatch Region.* At neutral and high pH, the 2D NOESY spectrum of duplex I contains weak cross peaks from $AP_5$ H8 to $G_4$ H1′ and $AP_5$ H1′, and the $AP_5$ H8 resonance is sharp. In duplex II, at neutral pH the $AP_5$ H8 cross peaks to $C_4$ H1′ and $AP_5$ H1′ are both very weak in intensity, and the $AP_5$ H8 and H6 resonances have somewhat broader linewidths. Also, all cross peaks to the $C_4$ H6 resonance (from $T_3$ H1′, $C_4$ H1′, and $C_4$ H5) are broad, suggesting some conformational dynamics are occurring in the region of the mispair. The $G_{15}$ H8−$C_{14}$ H1′ and $G_{15}$ H8−$G_{15}$ H1′ cross peaks are both very weak. At pH 9.0, these resonances are sharp and give strong cross peaks in the NOESY spectrum.

*$^{15}$N NMR Resonance Assignments.* The directly detected $^{15}$N spectra of 4-$^{15}NH_2$-$C_{14}$ labeled duplexes I (at pH 6, Figure

Table 1: Chemical Shift Assignments of d(GATGPGTAC):d(GTACCCATC), Duplex I, and d(GATCPCTAC):d(GTAGCGATC), Duplex II[a]

| | H6/H8 | | | H1′ | | | A H2/AP H6/T CH3 | | |
|---|---|---|---|---|---|---|---|---|---|
| | GPG | CPC | Δδ | GPG | CPC | Δδ | GPG | CPC | Δδ |
| strand 1 | | | | | | | | | |
| G1 | 7.95 | 8.01 | +0.06 | 5.71 | 5.76 | +0.05 | | | |
| A2 | 8.34 | 8.43 | +0.09 | 6.32 | 6.41 | +0.09 | 7.92 | 8.06 | +0.14 |
| T3 | 7.15 | 7.27 | +0.12 | 5.73 | 5.98 | +0.25 | 1.42 | 1.43 | +0.01 |
| G4/C4 | 7.72 | 7.42 | −0.30 | 5.48 | 5.82 | +0.34 | | | |
| AP5 | 8.14 | 8.34 | −0.20 | 5.78 | 6.00 | +0.22 | 7.45 | 7.94 | +0.49 |
| G6/C6 | 7.55 | 7.48 | −0.07 | 5.94 | 5.96 | +0.02 | | | |
| T7 | 7.22 | 7.50 | +0.28 | 5.74 | 5.74 | 0.00 | 1.35 | 1.68 | +0.33 |
| A8 | 8.28 | 8.39 | +0.11 | 6.30 | 6.33 | +0.03 | 7.64 | 7.63 | −0.01 |
| C9 | 7.28 | 7.43 | +0.15 | 6.06 | 6.12 | +0.06 | | | |
| strand 2 | | | | | | | | | |
| G10 | 7.97 | 8.05 | +0.08 | 5.99 | 6.06 | +0.07 | | | |
| T11 | 7.44 | 7.53 | +0.09 | 5.78 | 5.75 | −0.03 | 1.42 | 1.46 | +0.04 |
| A12 | 8.36 | 8.27 | −0.09 | 6.31 | 6.22 | −0.09 | 7.57 | 7.52 | −0.05 |
| C13/G13 | 7.40 | 7.75 | +0.35 | 5.78 | 5.61 | −0.17 | | | |
| C14 | 7.49 | 7.35 | −0.14 | 6.18 | 6.09 | −0.09 | | | |
| C15/G15 | 7.45 | 7.82 | +0.37 | 5.58 | 5.68 | +0.10 | | | |
| A16 | 8.41 | 8.22 | −0.19 | 6.32 | 6.31 | −0.01 | 7.69 | 7.86 | +0.17 |
| T17 | 7.17 | 7.28 | +0.11 | 5.96 | 6.06 | +0.10 | 1.50 | 1.41 | −0.09 |
| C18 | 7.48 | 7.67 | +0.19 | 6.22 | 6.34 | +0.22 | | | |

[a] Chemical shifts are given in ppm (±0.01 ppm). The residual HDO resonance is referenced to 4.85 ppm (20 °C) for duplex I and to 5.01 ppm (8 °C) for duplex II.

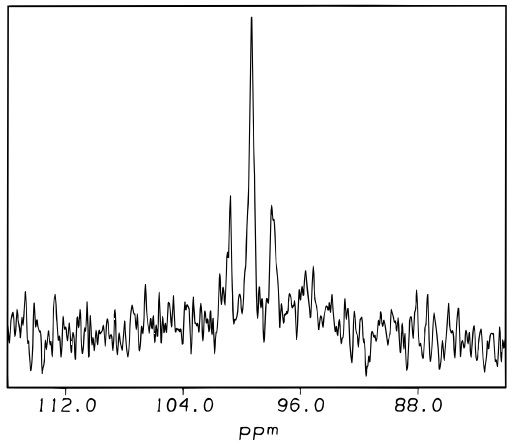


FIGURE 3: $^{15}NH_2$-$C_{14}$ labeled duplex I, at pH 6 and 10 °C.

3) and II (at pH 9) consist of one triplet, indicating that there are two protons bound on the cytosine N4. The 4-$^{15}NH_2$-$C_{14}$ nitrogen resonates at 100.6 ppm in duplex I and at 99.4 ppm in duplex II, referenced to $^{15}NH_3$. These values are consistent with the value of 99.6 ppm measured previously at high pH (Sowers et al., 1989), originally referenced to [$^{15}$N]aniline, which is downfield of $^{15}NH_3$ by 59.6 ppm (Witanowski et al., 1986). The chemical shift of the 2-$^{15}NH_2$-$AP_5$ nitrogen resonance was not measured.

*1D NMR pH Titrations.* The results from the one-dimensional $^1$H NMR pH titrations of duplexes I and II are shown in Figure 4. Titrations were performed at identical salt conditions and temperature. In both duplexes, the appearance of a very broad resonance at 11.0 ppm is concurrent with the broadening and disappearance of the doublet assigned to the hydrogen-bonded 4-amino proton of $C_{14}$. In the pH titration of duplex I, the broad resonance at 11.0 ppm is detectable at pH 6.0 and below, and in duplex II at pH 6.4 and below. In the 9.5−11.5 ppm region of the NMR spectrum, broad resonances typically appear from aromatic ring nitrogen-bound protons which are protected from exchange with bulk solvent but are not hydrogen-bonded (i.e., for imino protons of thymines in hairpin loops). Either the AP N1, with a p$K_a$ of 4.8 in a DNA polymer (Janion & Shugar, 1973), or cytosine N3, with a p$K_a$ of 4.6 (Fasman, 1975), are likely sites for protonation at low pH.

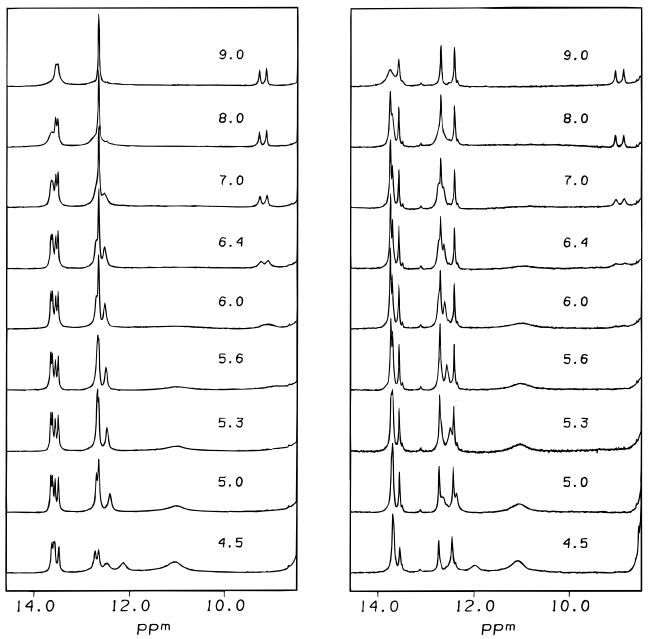


FIGURE 4: $^1$H NMR spectra acquired at 15 °C during a pH titration of (A, left) DNA duplex I and (B, right) DNA duplex II. Samples were 1.0 mM in duplex/150 mM NaCl.

Broadening of the $C_{14}$ amino proton doublet is evident at pH values lower than 6.4 for duplex I, and at pH values lower than 7.0 for duplex II. The broadening of the $C_{14}$ amino resonance may be due to increased exchange with solvent, or to exchange with another structural form on the intermediate NMR time scale, or both.

The imino resonances are essentially unperturbed from pH 9.0 down to pH 5.0, with only minor changes in chemical shift and no apparent broadening. Below 5.0, the iminos shift and broaden. This indicates that the standard Watson−Crick base pairs in both duplexes are stable down to pH 5.0.

*1D NOE Measurements.* To identify the base pairing in the AP mismatches, duplex II was annealed using strands containing 4-$^{15}NH_2$-$C_{14}$, and 2-$^{15}NH_2$-$AP_5$. In a wobble conformation, the two hydrogen-bonded amino protons on the labeled nitrogens are within 4.0 Å and therefore should give rise to an NOE. When the downfield (hydrogen-bonded) amino proton resonance on $AP_5$ is saturated, a weak NOE is observed on the downfield (hydrogen bonded) $C_{14}$ amino proton resonance (data not shown). However, when the reverse experiment is done, an NOE is not observed. A possible reason for this asymmetry in the NOE is that the two protons undergo different rates of exchange with solvent. Such a difference in rates might explain the differing pH and temperature dependences observed in the line widths of the $C_{14}$ and $AP_5$ amino proton resonances (data not shown).

## DISCUSSION

### *Structure of the AP·C Mispair*

*Imino vs Amino Tautomers.* The imino tautomeric form of DNA bases has long been used to explain spontaneous and base analog-induced mutation (Watson & Crick, 1953a,b; Freese, 1959; Topal & Fresco, 1976; Saenger, 1983) (Figure 1b). $^{15}N$-detected and $^{15}N$-filtered $^1H$ NMR experiments show clear evidence that, in 4-$^{15}NH_2$-$C_{14}$ labeled duplexes I and II and 2-$^{15}NH_2$-$AP_5$ labeled duplex I, two protons are bound to the labeled nitrogens in the AP·C mispair. No evidence is seen for any extra imino-type protons in the downfield region of the $^1H$ NMR spectrum, which if observed would suggest that one of the bases has a significant population of imino tautomeric form.

Moreover, in direct-detected $^{15}N$ spectra of 4-$^{15}NH_2$-$C_{14}$ labeled duplexes I (Figure 3) and II, one triplet is observed, indicating that one tautomeric form is detectable in which the nitrogen resonance is split by two directly attached protons. The chemical shift of the cytosine $^{15}NH_2$ resonance also provides information about the tautomeric state of the base. The amino−imino tautomerization of 2-aminopyridine results in an estimated 115 ppm upfield shift of the exocyclic $^{15}N$ resonance (Witanowski et al., 1986). A similar shift can be expected for the tautomerization of cytosine. Therefore, the measured chemical shift is sensitive to the presence of even a small (i.e., 1%) population of imino tautomer in the fast exchange limit. Since the $^{15}NH_2$-$C_{14}$ chemical shift in both AP-containing duplexes (100.6 ppm in duplex 1 and 99.4 ppm in duplex II) is within 2 ppm of that found in duplex DNA (101.3 ppm),[2] the AP·C mispair does not seem to shift the equilibrium at all toward the imino tautomer compared to a G·C pair.

Taken together, these data indicate that both $C_{14}$ and $AP_5$ are amino tautomers, since two protons are observed on the exocyclic amino nitrogens and no extra imino protons were observed in the NMR spectrum. We have determined that the amino tautomer is prevalent by at least 10−20-fold relative to any other tautomeric form, with the lower bound set by the limited sensitivity of the NMR experiment. To date, no structural study has produced evidence for any substantial population of the rare imino tautomeric form of bases, consistent with our findings here.

[2] Originally referenced to 2 M $^{15}NH_4Cl$ in 2 M HCl, 29.4 ppm downfield of neat $^{15}NH_3$ (James et al., 1981).

*Watson−Crick Geometry.* A standard Watson−Crick geometry pairing between AP and C would result in a mispair stabilized by one hydrogen bond (Figure 1a). Both $^1H$ and $^{15}N$ chemical shift data argue against this pairing geometry. Since amino protons of cytosines involved in Watson−Crick base pairing with guanine typically resonate between 8.0 and 8.6 ppm, the downfield chemical shift (9.16 ppm in duplex I and 8.94 ppm in duplex II) of the 4-amino proton of $C_{14}$ is highly suggestive that it is hydrogen bonded to the ring nitrogen. The $C_{14}$-amino to AP-H6 NOE is weaker than would be expected for the very short distance in the Watson−Crick geometry and is more consistent with a wobble. Additionally, the $^{15}N$ chemical shift of the $C_{14}$ N4 (100.6 ppm in duplex I and 99.4 ppm in duplex II, referenced to $^{15}NH_3$) is consistent with a hydrogen bond at the $C_{14}$ 4-amino group (Sowers et al., 1989). However, in the standard Watson−Crick geometry for the AP·C mispair the $C_{14}$ 4-amino group is not hydrogen bonded. This geometry is also inconsistent with previous work by Goodman and Ratliff (1983), which showed that the UV absorption spectra of polynucleotides containing 2-aminopurine were characteristically red-shifted in the presence of cytosine, which suggests that AP·C base pairs contain a hydrogen bond at the N1 position of AP.

Protonation at N1 of $AP_5$ or at N3 of $C_{14}$ would result in a charged base pair with Watson−Crick geometry stabilized by two hydrogen bonds (Figure 1c). A proton at one of these positions would likely resonate in the downfield region of the exchangeable proton NMR spectrum (>10 ppm) due both to its involvement in a hydrogen bond and its positioning in between and in the plane of two aromatic bases, which would further deshield the proton through ring current shift. At pH below 7, the hydrogen-bonded $C_{14}$ amino proton resonance is broad and a resonance is visible at about 11 ppm, the region expected for base protonation. However, at neutral and high pH there is no longer a resonance at 11 ppm, and all resonances in the downfield region of the NMR spectrum have been otherwise assigned, indicating that protonation is not occurring. The proton NMR spectrum remains essentially unchanged from pH 7 to 9, indicating that no significant structural rearrangements or changes in charge state are occurring in this range. Therefore, we observe no evidence for protonation of $AP_5$ at the N1 position or of $C_{14}$ at the N3 position in either duplex at neutral pH.

Previous NMR studies of a DNA heptamer duplex containing an AP·C mispair suggested that, at neutral pH, the cytosine paired with AP in the local sequence GAPG:CCC is protonated at the N3 position (Sowers et al., 1986). With the use of selective $^{15}N$ labeling, we are able to make unambiguous assignments of the exchangeable protons directly involved in the AP·C mispair which contradict this earlier assignment. Direct evidence from 1D $^{15}N$-filtered and 2D $^1H$ NOESY spectra in this study shows that the resonance assigned to a cytosine N3 proton in previous work (at 9.37 ppm) (Sowers et al., 1986) is actually a downfield-shifted cytosine $NH_2$ proton (at 9.16 ppm in duplex I in this study). Additionally, a resonance previously assigned to an AP amino proton (at 6.60 ppm) is properly assigned to the upfield cytosine $NH_2$ proton in the mispair (at 6.26 ppm in duplex I in this study).

These corrected assignments are consistent both with our current observations and with the observed NOE patterns in the previous NMR study. The assignments qualitatively are

more reasonable since, in the previous NMR study, no assignments were made for the amino protons of the cytosine in the AP·C mispair. Cytosine amino proton resonances, in both double- and single-stranded DNA, typically are sharper and hence more readily observed by NMR than guanine or adenine amino protons. Guanine and adenine amino groups often rotate at or near intermediate exchange on the NMR time scale, causing severe broadening of the amino proton resonances and making assignment of these protons difficult. In the study by Sowers and co-workers, although assignments were not made for the two cytosine amino protons in the AP·C mispair, assignments were made, however, for the 2-amino protons on 2-aminopurine, which are more likely to behave similarly to guanine and adenine amino protons in the NMR spectrum. Therefore, we conclude that these previous assignments were in error and led to the incorrect conclusion that the AP·C base pair is protonated and in Watson−Crick geometry.

*Wobble Base Pair.* The presence of stable wobble pairs in DNA and RNA has been established by NMR spectroscopy (Patel et al., 1982; Hare et al., 1986) and X-ray crystallography (Brown et al., 1985; Kneale et al., 1985). A wobble geometry for the AP·C mispair in DNA (Figure 1d) was first suggested by Janion and Shugar (1973), and evidence for this geometry was observed at high pH (Sowers et al., 1989; Guest et al., 1991). The studies of Fazakerley et al. (1987) determined the geometry of the AP·C base pair to be wobble at high pH by utilizing $^{15}N$ chemical shift data to identify base nitrogens which are directly involved in hydrogen-bonding interactions.

Several lines of evidence in the present study lead to the conclusion that the AP·C base pair is in a wobble geometry at neutral as well as high pH. $^{1}H$ NMR spectra of the two duplexes are essentially identical at pH 9.0 and 7.0, with some resonances in the mismatch broadening somewhat at neutral pH. Therefore, it appears that the structure of the AP·C base pair present at pH 9.0 remains dominant at neutral pH. The structure at lower pH is less clear.

The 1D NOE experiments taken at high pH show direct evidence that the cytosine amino group is within 5.0 Å of the 2-aminopurine amino group. This requires that the AP·C base pair be in a wobble geometry. The NOE is weak, however, probably because of the large line width of the AP and C amino protons.

As described in the previous section, the downfield chemical shift of the 4-amino proton of $C_{14}$ suggests that it is hydrogen bonded. A wobble geometry is the only proposed geometry for the AP·C base pair in which the $C_{14}$ 4-amino group is hydrogen bonded. The unusual downfield chemical shift is consistent with a wobble geometry in that the cytosine amino proton is directly hydrogen-bonded to $AP_5$ N1. In this geometry, the hydrogen bonded cytosine amino proton is likely to be 1.7−2.1 Å from the edge of the 2-aminopurine base and may be shifted further downfield than usual due to an aromatic ring current shift. In a G·C base pair, the cytosine amino protons are approximately 2.6 Å from the edge of the guanine base, and consequently should have a smaller ring current shift. Moreover, a wobble geometry is consistent with UV absorption studies which suggest that AP·C base pairs contain a hydrogen bond at the N1 position of AP (Goodman & Ratliff, 1983). This wobble geometry has been proposed by Patel and co-workers for the $O^6$-ethylguanine·cytosine base pair which could pair in a Watson−Crick geometry stabilized by one hydrogen bond but instead forms a wobble pair stabilized by two hydrogen bonds (Kalnick et al., 1989). It appears that the additional hydrogen bond gained in going from a Watson−Crick geometry to a wobble geometry makes up for any loss of interbase stacking interactions.

### *Sequence Dependence of Structure*

Stacking interactions with nearest-neighbor bases affect the local structure and helical stability around a DNA base pair (Werntges et al., 1986; Modrich, 1987; Ke & Wartell, 1993; Doktycz et al., 1995). The sequence dependence on the stability in the region of the AP·C mispair is evidenced in two ways: first, through the pH dependence of the NMR spectra for the two duplexes, and second, through the differences in dynamics of the bases near the mispairs in duplexes I and II.

Below neutral pH (Figure 4), the mispair resonances in both duplexes indicate a structural change which is localized to the mispair, since there are no major chemical shift or line width changes in resonances outside the mispair, and the eight imino protons observed from the eight flanking base pairs are relatively unaffected. The two duplexes exhibit somewhat different pH dependence for this structural change, with the transition occurring approximately 0.5 pH units higher in duplex II than in duplex I.

Duplex II shows more evidence of local dynamics near the mispair at neutral pH than duplex I. Exchangeable and nonexchangeable protons in the $AP_5$·$C_{14}$ mispair and the adjacent $C_4$·$G_{15}$ base pair show evidence of dynamics since NOESY cross peaks involving these protons are weak, due at least in part to broader line widths. No such effect is seen for duplex I. The difference in dynamics observed suggests that the stacking stabilization of the $C_4$-$AP_5$ step in duplex II is smaller than that of the $G_4$-$AP_5$ step in duplex I. These observations are consistent with calculated thermal stabilities of stacked DNA doublets containing AP paired with T, which predict that the doublet C-AP is far less stable than the doublet G-AP (Petruska & Goodman, 1985). It should be noted, however, that the stacking stabilization of these doublets will differ somewhat when AP is paired with C rather than with T.

### *Implications for Misinsertion and Repair of AP*

A model has been developed to describe the method of discrimination by DNA polymerase in selecting the correct base for insertion and in editing misinserted bases. In this model, the polymerase screens according to the base-pairing free energy of the pair and the precise geometry of the Watson−Crick base pair (Kornberg, 1980; Echols & Goodman, 1991), perhaps by measuring bond angles and the cross-strand C1′−C1′ distance which are remarkably similar in A·T and G·C base pairs (Kennard, 1987). This selection process allows the polymerase to distinguish the proper base from all other substrates by a factor of $10^3$−$10^5$, so that at low frequencies other bases are misinserted.

According to this model, the wobble structure of the AP·C mispair may account for the strong preference of DNA polymerase to insert AP opposite T rather than C (Watanabe & Goodman, 1981, 1982). Although both AP·C and AP·T mispairs each contain two hydrogen bonds, the polymerase can distinguish the two mispairs based on their differences

in geometry and inserts the Watson−Crick-like AP·T mispair (Sowers et al., 1986) at a higher frequency than the wobble AP·C mispair (Watanabe & Goodman, 1981).

Likewise, the increased rate of insertion of AP opposite C relative to A (Mhaskar & Goodman, 1984) is consistent with the wobble structure of the AP·C mispair. In both AP·A (Fazakerley et al., 1987) and AP·C mispairs, two hydrogen bonds are formed to generate wobble base pairs, with both bases in the *anti* conformation. However, the AP·A mispair consists of two bulky purines which deviate further from Watson−Crick geometry compared to the AP·C mispair, which consists of one purine and one pyrimidine. Therefore, the AP·C wobble pair may be more easily recognized as a substrate by the polymerase.

The results from this study indicate that the structure and stability of the AP·C base pair is somewhat dependent on local sequence. This has potential implications for the observed dependence of single-site mutation rates with nearest-neighboring bases (Pless & Bessman, 1983; Mendelman et al., 1989; Joyce et al., 1992; Mitra et al., 1993). The mutagenicity of AP has been shown to be strongly dependent on the local DNA sequence at the mutation site (Ronen, 1979). The local DNA sequence, specifically the extent to which the primer-3′-nearest neighbor can stabilize an incoming dNTP substrate at the polymerase active site, has been shown to influence efficiency of AP misinsertion opposite T by DNA polymerase (Pless & Bessman, 1983; Petruska & Goodman, 1985; Bloom et al., 1993). These effects are attributed to base stacking interactions between the incoming AP nucleotide and the 3′-primer terminus. In a similar manner, local DNA sequence has a profound effect on the editing efficiencies of polymerase-associated proofreading exonuclease, which is sensitive to the degree of single-strandedness at a primer-3′-terminus. A striking indication of the effect of local sequence context on the editing of base pairs involving AP was demonstrated in a pre-steady-state kinetic study of exonucleolytic proofreading by bacteriophage T4 DNA polymerase (Bloom et al., 1994). It was found that excision of AP located at a primer-3′-terminus opposite each of the four template bases, within the same surrounding sequence, occurs at rates that are consistent with calculated thermal stabilities of each base pair (Petruska & Goodman, 1985). AP opposite T is excised at a relatively slow rate, while excision of AP opposite A and C occurs more rapidly. However, a Watson−Crick-like AP·T mispair in an A-T-rich environment is removed relatively rapidly compared to an AP·C mispair in a G-C-rich environment, demonstrating the importance of localized sequence context on proofreading specificities.

### *Conclusion*

Previous studies have produced differing models for the structure of the AP·C mispair. We find that the structure of this mispair is wobble and that our results are fully consistent with prior experimental data. Studies which utilize 2-aminopurine to measure rates of DNA base insertion, proofreading, or repair must now consider the AP·C wobble base pairing. The dependence of the stability of the AP·C mispair on the local sequence context highlights the need for consideration of this effect in such studies.

## ACKNOWLEDGMENT

We thank Dr. H. Peter Spielmann for valuable discussions.

## REFERENCES

Acedo, M., Fàbrega, C., Aviño, A., Goodman, M., Fagan, P., Wemmer, D., & Eritja, R. (1994) *Nucleic Acids Res. 22*, 2982−2989.

Bessman, M. J., Muzyczka, N., Goodman, M. F., & Schnaar, R. L. (1974) *J. Mol. Biol. 88*, 409−421.

Bloom, L. B., Otto, M. R., Beechem, J. M., & Goodman, M. F. (1993) *Biochemistry 32*, 11247−11258.

Bloom, L. B., Otto, M. R., Eritja, R., Reha-Krantz, L. J., Goodman, M. F., & Beechem, J. M. (1994) *Biochemistry 33*, 7576−7586.

Brown, T., Kennard, O., Kneale, G., & Rabinovich, D. (1985) *Nature 315*, 604−606.

Doktycz, M. J., Morris, M. D., Dormady, S. J., Beattie, K. L., & Jacobson, K. B. (1995) *J. Biol. Chem. 270*, 8439−8445.

Drake, J. W., Allen, E. F., Forsberg, S. A., Preparata, R.-M., & Greening, E. O. (1969) *Nature 221*, 1128−1132.

Echols, H., & Goodman, M. F. (1991) *Annu. Rev. Biochem. 60*, 477−511.

Fasman, G. D., Ed. (1975) in *Handbook of Biochemistry and Molecular Biology*, Vol. I, CRC Press, Cleveland, OH.

Fazakerley, G. V., Sowers, L. C., Eritja, R., Kaplan, B. E., & Goodman, M. F. (1987) *Biochemistry 26*, 5641−5646.

Freese, E. (1959) *J. Mol. Biol. 1*, 87−105.

Frey, M. W., Sowers, L. C., Millar, D. P., & Benkovic, S. J. (1995) *Biochemistry 34*, 9185−9192.

Goodman, M. F., & Ratliff, R. L. (1983) *J. Biol. Chem. 258*, 12842−12846.

Guest, C. R., Hochstrasser, R. A., Sowers, L. C., & Millar, D. P. (1991) *Biochemistry 30*, 3271−3279.

Hare, D. R., Wemmer, D. E., Chou, S.-H., & Drobny, G. (1983) *J. Mol. Biol. 171*, 319−336.

Hare, D., Shapiro, L., & Patel, D. J. (1986) *Biochemistry 25*, 7445−7456.

James, T. L., James, J. L., & Lapidot, A. (1981) *J. Am. Chem. Soc. 103*, 6748−6750.

Janion, C., & Shugar, D. (1973) *Acta Biochim. Polon. 20*, 271−284.

Joyce, C. M., Chen Sun, X., & Grindley, N. G. (1992) *J. Biol. Chem. 267*, 24485−24500.

Kalnick, M. W., Li, B. F. L., Swann, P. F., & Patel, D. J. (1989) *Biochemistry 28*, 6182−6192.

Ke, S.-H., & Wartell, R. M. (1993) *Nucleic Acids Res. 21*, 5137−5143.

Kennard, O. (1987) *Nucleic Acids Mol. Biol. 1*, 25−52.

Kime, M. J., & Moore, P. B. (1983) *Biochemistry 22*, 2615−2622.

Kneale, G., Brown, T., Kennard, O., & Rabinovich, D. (1985) *J. Mol. Biol. 186*, 805−814.

Kornberg, A. (1980) *DNA Replication*, Freeman, San Francisco.

Live, D. H., Dane, D. G., Agosta, W. C., & Cowburn, D. (1984) *J. Am. Chem. Soc. 106*, 1939−1941.

Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.

Mendelman, L. V., Boosalis, M. S., Petruska, J., & Goodman, M. F. (1989) *J. Biol. Chem. 264*, 14415−14423.

Mhaskar, D. N., & Goodman, M. F. (1984) *J. Biol. Chem. 259*, 11713−11717.

Mitra, R., Pettitt, B. M., Rame, G. L., & Blake, R. D. (1993) *Nucleic Acids Res. 21*, 6028−6037.

Modrich, P. (1987) *Annu. Rev. Biochem. 56*, 435−466.

Patel, D. J., Kozlowski, S. A., Marky, L. A., Rice, J. A., Broka, C., Dallas, J., Itakura, K., & Breslauer, K. J. (1982) *Biochemistry 21*, 437−444.

Petruska, J., & Goodman, M. F. (1985) *J. Biol. Chem. 260*, 7533−7539.

Pless, R. C., & Bessman, M. J. (1983) *Biochemistry 22*, 4905−4915.

Ronen, A. (1979) *Mutat. Res. 75*, 1−47.

Saenger, W. (1983) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

Sklenar, V., & Bax, A. (1987) *J. Magn. Reson. 74*, 469−479.

Sowers, L. C., Fazakerley, G. V., Eritja, R., Kaplan, B. E., & Goodman, M. F. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 5434−5438.

Sowers, L. C., Eritja, R., Chen, F. M., Khwaja, T., Kaplan, B. E., Goodman, M. F., & Fazakerley, G. V. (1989) *Biochem. Biophys. Res. Commun. 165*, 89−92.
Topal, M. D., & Fresco, J. R. (1976) *Nature 263*, 285−289.
Ward, D. C., Reich, E., & Stryer, L. (1969) *J. Biol. Chem. 244*, 1228−1237.
Watanabe, S. M., & Goodman, M. F. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2864−2868.
Watanabe, S. M., & Goodman, M. F. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 6429−6433.
Watson, J. D., & Crick, F. H. C. (1953a) *Nature 171*, 964−967.
Watson, J. D., & Crick, F. H. C. (1953b) *Cold Spring Harbor Symp. Quant. Biol. 18*, 123−131.
Werntges, H., Steger, G., Riesner, D., & Fritz, H.-J. (1986) *Nucleic Acids Res. 14*, 3773−3790.
Witanowski, M., Stefaniak, L., & Webb, G. A., Eds. (1986) in *Annual Reports on NMR Spectroscopy*, Vol. 18, Academic Press, London.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, John Wiley & Sons, New York.

BI952657G