The pH dependent configurations of the C.A mispair in DNA

Y.Boulard, J.A.H.Cognet¹, J.Gabarro-Arpa¹, M.Le Bret¹, L.C.Sowers² and G.V.Fazakerley*

Service de Biochimie et de Génétique Moléculaire, Département de Biologie Cellulaire et Moléculaire, Centre d'Etudes de Saclay, 91191 Gif-sur-Yvette Cedex, ¹Laboratoire de Physicochimie Macromoléculaire, Institut Gustave Roussy, F-94800 Villejuif, France and ²Division of Pediatrics, City of Hope National Medical Center, Duarte, CA 91010, USA

Received January 20, 1992; Revised and Accepted March 20, 1992

ABSTRACT

The structure of the cytosine-adenine mispair in a 7 base pair duplex has been investigated by proton NMR spectroscopy. At low pH, the predominant structure is protonated on the A residue and assumes a wobble conformation consistent with previous findings. The C residue of the mispair is found in a C2'–C3' *endo* equilibrium. This is confirmed by molecular dynamics calculations which suggest that the conformation of the protonated wobble is flexible and not as rigid as a normal base pair. As the solution pH is raised, a structural transition is observed with an apparent pK of 7.54 at 23°C. At higher pH the predominant structure is one in which both the C and A residues are intrahelical. Evidence is presented that this structure corresponds to a reverse wobble in which the two bases are held together by one hydrogen bond. This structure is much less stable than the protonated form and even at low temperature single strands are observed in slow exchange with the neutral duplex form.

INTRODUCTION

The adenosine-cytosine mispair, when formed in DNA, would serve as an intermediate in establishing transition mutations *in vivo*. Over the years, many studies have been conducted to identify the structure of the C.A mispair. Although a variety of techniques have been employed, elucidation of the C.A structure has proven difficult, elusive and controversial.

Originally, it was proposed that the C.A mispair could form in Watson–Crick geometry if either A or C formed a rare tautomeric structure (1,2). The tautomer model historically has been preferred, however, few experimental studies have supported this model. Since 1959, a total of at least nine other structural alternatives have been suggested.

Patel et al. (3) were the first to examine the C.A mispair in DNA using high field NMR spectroscopy. Their results, for the mispair studied at low pH, were inconsistent with imino tautomers

in Watson–Crick geometry, however, there was not sufficient information to unambiguously define the structure. Hunter et al. (4) established that the C.A mispair in a DNA duplex crystal obtained at acidic pH was in a wobble geometry. Crystallography could not, however, determine whether the structure was protonated or involved tautomeric forms.

We have been interested in the phenomena of ionized base pairs due to the lower energy required for ionization versus tautomer formation at physiological pH (5,6). We demonstrated that ionization and tautomer formation induced changes in proton chemical shifts of similar magnitude but opposite direction. Using this information, we examined the C.A data of Patel et al. Our results were consistent with protonation of the A residue and inconsistent with formation of imino tautomeric forms by either the A or C residues. Recently, protonation of the A residue in an C.A mispair at acidic pH has been unambiguously confirmed (7) by NMR spectroscopy of N-15 enriched oligonucleotides.

The presence of ionized bases in mispairs implies that the structure of the mispair may vary as a function of solvent pH. Transitions between structurally distinct base pair configurations, induced by changes in solvent pH, have been demonstrated for fluorouracil-guanine (8), bromouracil-guanine (9) and guanine-adenine (10,11) base mispairs. It has been shown for oligonucleotides containing C.A mispairs that the melting transition temperature is highly pH dependent (12) and the N-15 NMR study (7) indicated that the protonated C.A mispair loses a proton with increasing solvent pH. There have been no reported studies on the nature of the C.A mispair under neutral or alkaline conditions.

In order to provide a more complete understanding of the C.A mispair, we have conducted NMR studies on a model DNA duplex containing an C.A mispair as a function of pH.

MATERIALS AND METHODS

The duplex studied here was formed from the oligonucleotides 5'd(C1 A2 G3 C4 G5 G6 C7).(G8 C9 C10 A11 C12 T13 G14). The duplex was 4mM in strand concentration in 10 mM phosphate buffer, 150 mM NaCl and 0.2 mM EDTA.

* To whom correspondence should be addressed

NMR spectra were recorded on either AMX500 or AMX600 Bruker spectrometers. Details of the NMR experiments and data treatment have been given previously (13).

Molecular mechanics calculations. Energy minimizations and molecular dynamics were carried out using the program AMBER (14–16) on IBM 530 RISC 6000 computers. The parameters used are as described previously (13). The charges for the protonated adenosine base were calculated on a VAX 8530 with the program QUEST of the AMBER package (15–19) in accordance with the AMBER database. Comparison between the NMR data and model data, function F, is as described previously (13).

Molecular dynamics (MD). The molecular dynamics computations were all performed as described previously (20). The preparation procedure was developed for a system with a very sensitive dynamic probe. NMR distance measurements were used as restraints in the MD. The force constant used was 10 kcal/mol.Å² (MDI). In a second run, MDII, the restraint was solely on the torsion angle δ of all residues except C4 and A11. These were restrained to $\delta_0 = 144^\circ$. The force constant used was 10 kcal/mol. These two MD runs are compared. Kinetic, potential and total energy of the molecule remained at equilibrium values during the entire 200 ps runs without drifting.

The molecular structures were displayed on a Silicon Graphics IRIS 4D70GT using the programs, MORCAD and OCL (21,22).

RESULTS AND DISCUSSION

In the 1D spectrum of the oligonucleotide at pH 7 certain of the aromatic resonances displayed large line widths indicating the presence of an exchange phenomenon. Spectra were thus recorded as a function of pH and subsequently of temperature to determine the optimum experimental conditions. As the pH was progressively lowered the broad resonances became sharper and others showed changes in their chemical shift. Below pH 5.8 no further changes were observed in the spectrum. On raising the pH, at 23°C, certain resonances in the aromatic region, for which exchange was in the fast exchange region, could be continuously followed. For example, the resonance subsequently identified as A11 H2, showed a normal titration curve and an apparent pK of 7.54. While at pH 5.8 the spectrum indicates the presence of only one species this was not the case at pH 8.9 where at least two species were observed. The relative concentration of the second species could be followed by integration of a well resolved resonance at 7.08 ppm. The chemical shifts show that this second species is not the low pH species. We observed that the relative concentration of the second species is strongly temperature dependent, decreasing to ca. 10% at 1°C.

By following the proton chemical shifts as a function of temperature at pH 5.8 we observe a T_m of ca. 39°C for the duplex.

NOESY spectra in D₂O, pH 5.8 at 23°C. Assignment of the nonexchangeable proton resonances

To assign the base, H1', H2', H2'' and H3' protons a NOESY spectrum was recorded with a 400-ms mixing time. One region of this spectrum is shown in Figure 1. The sequential connectivities can be followed for both strands, no break is observed at the mismatch site. The cross peak corresponding to the intrasite interaction C10 H6-H1' appears weak and not well aligned but this is an artefact introduced by the sine bell filtering to improve the resolution which produces negative wings on the H5-H6 cross peak of this residue. When the data are

multiplied by a sine bell shifted by $\pi/2$ before Fourier transformation this cross peak has an intensity similar to all other such intrasite cross peaks, Figure 1 (top).

Four interbase cross peaks, A–D, are observed corresponding to H8/H6-CH5 interactions for G6-C7, G8-C9, C9-C10 and A11-C12 respectively. A fifth one might be expected between the G3-C4 residues but this overlaps with the G6 H8-G5 H1' cross peak. A resonance at 7.75 ppm gives two overlapping cross peaks with the H1' resonances of A2 and G3, peak E, and another with that of G14, peak F, identifying this resonance as A2 H2. Similarly a resonance at 8.10 ppm gives cross peaks with the H1' resonances of A11, C12 and G5, peaks H–J, and thus can be assigned to A11 H2.

The above assignments are confirmed by analysis of the H8/H6-H2'/H2''/CH₃ region (not shown). The interactions observed in both regions show that the C and A residues of the mismatch stack into the helix and that the duplex adopts a globally B DNA structure. Although the interactions are strongly influenced by spin diffusion it is apparent that no major deviation from a B form structure takes place at the mismatch site. Assignment of the H3' and H4' resonances was obtained from analysis of TOCSY (80-ms mixing time) and phase sensitive COSY spectra (not shown). In order to probe in more detail the structure of the duplex, NOESY spectra were recorded with mixing times of 40, 50, 60 and 80-ms. The initial build up rates

Figure 1. Part of the NOESY spectrum recorded with a 400 ms mixing time at 23°C and pH 5.8.

of the NOEs were measured and converted into interproton distances (23) for model building studies. From the magnitude of the H8/H6-H1' NOEs all residues must be in an *anti* conformation. These spectra also gave the relative assignment of the H2'/H2'' resonances and were used to determine the major sugar pucker from the base proton-H2'/H3' NOEs (24). These results were confirmed by analysis of the phase sensitive COSY spectra. Excepting the terminal residues and C4 all sugars show interactions corresponding to a predominantly C2' *endo* conformation. For C4 we observe NOEs of similar intensity for the interactions C4 H6-H2' and H3' indicating approximately equal populations of C2' and C3' *endo* conformations. This is confirmed by the measurement of the sum of JH1'-H2' and JH1'-H2'' found to be 11.8 Hz, significantly lower than for all other non-terminal residues.

Figure 2. A) 1D spectrum recorded in H₂O at 1°C, pH 5.8. B) as A but at 30°C. C) 1D difference spectrum following presaturation for 0.5 s at 20°C of the resonance at 9.5 ppm.

1D and NOESY spectra in H₂O

Part of the 1D spectrum of the duplex recorded at 1°C is shown in Figure 2A. We observe five resonances in the region for G imino protons and one resonance corresponding to a T imino proton. Additionally we observe a broad resonance around 9.5 ppm. Lowering the pH to 5 has no effect upon the linewidth of the resonance at 9.5 ppm indicating that it is not broadened by exchange with the solvent. When the temperature is lowered this resonance broadens into the base line, on the other hand it narrows as the temperature is raised. Its linewidth is ca. 400 Hz at 15°C, 270 Hz at 20°C and 180 Hz at 30°C. The spectrum at 30°C is shown in Figure 2B. Integration of this peak relative to that at 8.75 ppm (which corresponds to one proton, see below) shows that it corresponds to two protons.

A NOESY spectrum was recorded at 1°C. Two regions of this spectrum are shown in Figure 3. The spectrum was assigned in the usual way. We note that the G5 imino proton gives cross peaks with the amino protons of C10 but only very weak cross peaks with those of C9, visible only in the matrix column. Further it shows strong cross peaks with a pair of amino protons at 7.95 and 5.38 ppm, a strong cross peak is observed between these two resonances, which must belong to the mismatch pair or the G10.C5 pair. In general, when the H5 resonance is not too close to the solvent, cross peaks are observed from the hydrogen bonded C amino resonance through the non-hydrogen bonded resonance to the H5 resonance. This is the case here for all G.C pairs with the exception of C7 and C9 for which the H5 resonances are under the solvent resonance. From the resonance

Figure 3. Part of the NOESY spectrum recorded in H₂O at 1°C, pH 5.8 with a mixing time of 200 ms.

at 7.95 ppm we do not observe a cross peak with a CH5 resonance excluding that this pair of protons arises from C4, for which the H5 resonance is well removed from the solvent frequency. They must therefore arise from the amino group of G5 or A11. At this stage of the assignment three observations suggest that these resonances arise from the amino group of G5. The intensity of the NOEs from the imino proton of G5 are similar to those observed between imino protons and the C amino protons of the same base pair thus showing spatial proximity to the imino proton of G5. Secondly, we might expect to observe the same pair of NOEs from the imino proton of G3 if these resonances arise from A11, which, at least for the resonance at 5.38 ppm, we do not. Thirdly, the chemical shift of this resonance at 5.38 ppm, would be very unusual for an A residue. While the G amino protons of non-terminal G.C base pairs normally give rise to very broad and unobservable resonances at low temperature this is not necessarily the case for base pairs adjacent to a mismatch site (25) where, for reasons of steric hindrance, amino group rotation can be slowed down into the slow exchange region.

All the amino protons other than those of the G residues (except tentatively G5), C4 and of A11 have been identified as indicated in Figure 3. Two other cross peaks in Figure 3 with a resonance at 6.00 ppm can be assigned to the pair of amino protons of G8 which, via fraying, are in rapid rotation on a proton time scale as previously observed (25).

The 9.5 ppm resonance is far too broad to give any interactions in NOESY spectra at any temperature. We have carried out a 1D difference spectrum with presaturation for 0.5 s at 20°C, Figure 2C. A very large NOE is observed at 6.90 ppm, the

intensity of which indicates close spatial proximity of these protons. NOEs are observed to the hydrogen bonded and coincident non-hydrogen bonded amino protons of C10 and C12 and to the imino protons of G3 and G5. When the 1D difference spectrum was recorded at 30°C (not shown) we observe that the 6.9 ppm resonance, like that at 9.5 ppm sharpens with increasing temperature. This suggests that the protons corresponding to both these resonances undergo a similar proton exchange phenomenon which influences their linewidths. Although we cannot integrate the resonance at 6.9 ppm it is very probable that it corresponds to two protons. We have attempted to carry out difference spectra upon preirradiation of the resonance at 6.9 ppm. We observe NOEs (not shown) to the CH5 protons of C12, C4 and C10. As selective presaturation is not possible these results cannot be unambiguously interpreted.

The observation of a structural change occurring as a function of pH for a duplex containing only normal bases indicates that protonation is occurring at low pH. In a different sequence containing a G.A mismatch we have been able to observe, at 16 ppm, the additional proton in the structure G.⁺A, protonated on N1 at pH 5 (11). We have titrated the C.A duplex down to pH 4 but we do not observe a resonance corresponding to an additional proton. This may be due to greater solvent accessibility or more frequent base pair opening in the case of a purine-pyrimidine mismatch relative to a purine-purine mismatch.

Table 1. Chemical shifts of non-exchangeable protons at 23°C and of exchangeable protons at 1°C. The second line (b) for the central dinucleotide gives the chemical shifts observed for the duplex at pH 8.9 and 12°C.

	H8/H6	H5/H2 CH3	H1'	H2'	H2''	H3'	NH	NH2
C1	7.60	5.87	5.52	1.85	2.42	4.67		8.30,6.95
A2	8.28	7.75	5.92	2.79	2.86	5.02		7.95,6.60
G3	7.69		5.91	2.47	2.71	4.96	12.69	
G3b	7.74		5.71	2.50	2.52	4.92		
C4	7.29	5.60	5.69	1.51	2.15	4.75		6.90
C4b	7.28	5.20	6.02	2.01	2.46	4.81		
G5	7.89		5.59	2.72	2.74	4.98	13.15	7.95,5.38
G5b	7.81		5.52	2.55	2.64	4.80		
G6	7.69		5.92	2.49	2.66	4.96	13.08	
C7	7.24	5.06	6.08	2.19	2.19	4.48		8.15,6.40
G8	7.89		5.92	2.63	2.73	4.78	12.89	6.00
C9	7.48	5.29	6.05	2.18	2.52	4.86		8.37,6.41
C10	7.64	5.67	5.61	2.29	2.44	4.88		8.75,7.16
C10b	7.35	5.58	5.46	1.70	2.13	4.77		
A11	8.51	8.10	6.38	2.66	3.00	5.04		9.50
A11b	8.31	8.05	6.19	2.71	2.84			
C12	7.35	5.40	5.81	1.85	2.43	5.04		8.32,7.20
C12b	7.41	5.37	5.79	1.95	2.41	4.62		
T13	7.27	1.67	5.76	1.84	2.31	4.85	14.18	
G14	7.89		6.12	2.37	2.63	4.68	12.91	

Table 2. NMR fit, F, for different sugar pucker combinations of C4 and the protonated A11.

conformations	A11 C3' <i>endo</i>	A11 C2' <i>endo</i>
C4 C3' <i>endo</i>	0.5	1.1
C4 C2' <i>endo</i>	1.7	2.4

Figure 4. The protonated C.A mismatch viewed from the top of the oligonucleotide showing the three types of hydrogen bonding. T) classical hydrogen bonding O2-H1 and N3-HN6B. Models P and X result from sliding one base relative to the other perpendicular to the helix axis resulting in partial hydrogen bonding.

We observe one amino group at 9.5 ppm and very probably another amino pair at 6.9 ppm. This latter conclusion is based upon the temperature dependence of the line width and that, from presaturation at 9.5 ppm we do not observe, by spin diffusion, an NOE to another exchangeable proton. It could, however, be under the solvent resonance. We have previously shown (5) that protonation on N1 of adenosine shifts the H8 proton significantly downfield. The chemical shift observed here relative to the neutral species (see below) is entirely consistent with protonation of A in the C.A mismatch. It has also been observed (26) that base protonation results in large downfield shifts for the amino protons of both AMP and CMP. Similar results have been observed for the amino protons of protonated bases in triple stranded structures (27–30). This leads to the conclusion that the 9.5 ppm resonance corresponds to the amino group of A11. This is further supported by comparison of spectra of an C.A and A.O4meT mismatch (31). The 6.9 ppm resonance must therefore be assigned to the C4 amino protons. In the structure of a protonated C.A mismatch the C amino protons are not engaged in hydrogen bonding. While in G.C base pairs the amino protons are not broadened by rotation about the C-N bond at 30°C (32) this is not the case for CMP (33) where at this temperature the two resonances are collapsed. On the other hand we have generally been able to observe separate resonances for A amino protons at low temperature. That rotation about the C-N bond is faster, now being in the intermediate exchange region, could be due either to weakening of the C-N bond because of base protonation or due to the structure of the base pair or both. The proton chemical shifts are given in Table 1.

Molecular modelling

The starting models were constructed from canonical B-DNA, with the minimum number of changes necessary to accommodate the protonated C.A base pair in the double helix. Four variants of this molecular model were examined. These were the four permutations of C2' *endo* and C3' *endo* sugar conformations for the central base pair. The fit to the NMR data, F values, after energy minimization when these conformations were fixed is shown in Table 2. When any other sugar pucker was forced to C3' *endo* the fit is poorer. The model with C4 in C3' *endo*, A11 C2' *endo*, hereafter called model A, appears to fit best. However, the NMR data indicate that there is a conformational equilibrium for the C4 deoxyribose.

The C.AH+ mismatch is characterized by the presence of two hydrogen bond acceptors on C4 and two hydrogen bond donors on AH+. This explains why we observe during the energy minimization runs that the C.AH+ base pairing oscillates between three patterns of hydrogen bonding as shown in Figure 4 thereby generating a large number of base pairing conformations. These can be described as P) partial hydrogen bonding A H1–C O2 and N3, also A HN6B–C N3. T) normal hydrogen bonding A H1–C O2 and A HN6B–C N3. X) partial hydrogen bonding A H1–C O2 and also A HN6B–C N3 and C O2. Energy minimization tends to prefer conformation P but the NMR data do not distinguish well among the three. This base oscillation is much more weakly observed for the Watson Crick base pairs.

Model A with standard hydrogen bonding was used as the starting structure for the MD runs. The first MD run (MD1) was

Table 3. Interproton distances determined by NOE measurements and from the 200 ps run MDII. The first entry is from the NMR data and the second is the % difference between the model distance and the NMR data, (model–NMR)/NMR.

H8/H6 base i	H1'intra i	H1'inter i–1	H2'intra i	H2'inter i–1	H5inter i+1	CH3 i+1	H3'intra i	H2'inter i–1
C 1	3.6 3		2.3 1					
A 2	3.6 8	3.7 –6	2.2 6	2.4 1				
G 3			2.4 –1	2.3 7				
C 4	4.1 –10*	3.6 –3*	2.7 –2*	2.1 10*			3.0 –13*	
G 5	3.8 3		2.2 8	2.8 –9*				2.5 –5*
G 6	3.6 9	3.1 9	2.2 11	2.2 6	3.8 –1			
C 7	3.6 3	3.2 0	2.3 2	2.1 10				
G 8			2.3 7					
C 9	3.8 –2	3.5 –1	2.2 2	2.4 –5	3.8 –2			
C10		3.7 –10	2.3 –3	2.3 0				
A11	4.1 –5	3.2 6	2.6 –8	2.8 –15	3.8 –4			
C12	3.8 –2	3.4 2	2.2 2	2.1 12		3.5 –2		
T13	3.9 –4	3.6 3	2.1 8	2.2 5				
G14			2.3 9					

Table 4. Torsion angles and standard deviations for the 200 ps MDI run with all NMR distance constraints except those involving C4 (marked with * in Table 3).

	Pucker	Amp.	α	β	γ	δ	ϵ	ζ	χ
G3	142 13	43 5	–69 12	176 10	57 9	131 10	–178 8	–95 11	–114 10
C4	79 37	41 6	–74 13	169 10	61 9	94 20	–176 9	–84 18	–139 13
G5	149 15	42 5	–69 12	176 10	61 10	135 10	–179 11	–114 15	–108 9
C10	147 12	44 5	–71 12	178 10	57 9	137 9	–164 26	–133 26	–110 10
A11	150 18	40 6	–74 18	172 13	58 10	136 11	179 8	–98 11	–116 9
C12	135 13	42 5	–69 11	177 9	58 9	126 10	–178 12	–103 14	–121 9
Averages and standard deviations over all bases except the terminal and central bases.									
	143 14	42 5	–69 12	176 10	58 9	132 10	–176 11	–105 15	–116 10

performed with all NMR distance restraints given in Table 3 except those involving C4 (marked with an asterisk) in order to allow conformational flexibility for this residue. The resulting structure is described in Table 4 in terms of average torsion angles. All sugar puckers are found in the C2' *endo* conformation except that of C4 which oscillates between C3' and C2' *endo*. This is clearly seen in terms of the interproton distances and δ torsion angle for C4 in Figure 5, A–C. A11 remains close to C2' *endo* but this residue has the distance constraints included.

A second MD run (MD2) was carried out with solely a mild constraint on the the torsion angles δ for all sugars except C4 and A11 (see Materials and Methods). In this run we observe, Figure 5, D–F, from the interproton distances and torsion angle δ for A11 that it rarely departs from a C2' *endo* conformation. The results obtained from MD1 and MD2 show no significant differences in terms of interproton distances and torsion angles. A11 remains C2' *endo* and C4 oscillates. The agreement with the NMR data is shown in Table 3.

The lengths of the hydrogen bonds O2–H1 and N3–HN6B observed during MD1 as a function of time are shown in Figure 6, A, B. The length of these hydrogen bonds fluctuate, $2.1 \pm 0.3 \text{ \AA}$ and $2.5 \pm 0.4 \text{ \AA}$ respectively. This compares with a value of $2.1 \pm 0.2 \text{ \AA}$ for normal non-terminal base pairs. The origin of the fluctuations observed, in particular for N3–HN6B lies in the observation of partially hydrogen bonded structures in both MD runs. Occasional excursions bring O2 and HN6B within hydrogen bonding distance, structure X, Figure 6, C but very frequently N3 and H1 are found close, structure P, Figure 6, D. In general we observe populations in the order $P > T > X$. This is further visualized by monitoring the base pair shear (34) during MD1, Figure 6E, bottom. On the other hand the base pair stretch remains constant, Figure 6E, top. Figure 6F shows a scatter plot

Figure 5. Interproton distances (A) H6–H2', (B) H6–H3' and (C) torsion angle δ for the C4 residue as a function of time for MD1. (D)–(F), the same parameters for the residue A11 in MD2.

of the interaction energy between C4 and A11 versus the shear distance during MD1. These results are very similar to those observed in the molecular mechanics study.

The description of characteristic helical parameters during MD1 is given in Table 5.

NOESY spectra in D₂O, pH 8.9 at 12°C. Assignment of the nonexchangeable proton resonances

A region of the NOESY spectrum recorded with a 400 ms mixing time is shown in Figure 7. At 12°C the relative concentration of the second species is ca. 20%. Although this diminishes on further lowering the temperature this is accompanied by significant broadening of the resonances and loss of spectral resolution. The chain of connectivities can be followed without interruption for both strands of the major species. Five interresidue cross peaks, H8/H6–CH5, peaks A–E, corresponding to the steps G3–C4, G6–C7, G8–C9, C9–C10 and A11–C12 characteristic of a B DNA are observed. We can therefore conclude that both C4 and A11 are stacked into the helix.

Figure 6. Hydrogen bond donor to acceptor distances observed during MD1. (A) O2–H1, (B) N3–HN6B, (C) O2–HN6B and (D) N3–H1, (E) time evolution of the stretch (top) and shear (bottom) parameters (31). (F) scatter plot of the interaction energy between the C and A bases vs. the shear. Results are for MD1 but those of MD2 are very similar. The letters P, T and X refer to the models shown in Figure 4.

Table 5. Characteristic helical parameters followed by standard deviations for the 200 ps MD1 run (in degrees). Definition of helical parameters as in (31).

Base pair	Incl	Tip	Tilt	Roll	Twist	Open.	Shear
G3-C12	5 3	–7 5	2 2	–1 4	31 3	0 4	–0.7 0.5
C4-A11	–1 4	8 6	4 4	11 8	40 5	9 6	–1.2 0.7
G5-C10	3 5	9 7	–2 2	2 4	35 4	0 5	–0.7 0.5

We are able to follow another set of connectivities for the minor species. For the first strand we observe different chemical shifts for the base H8/H6 protons for all the residues. On the other strand separate resonances are only observed for C12. Their assignment is confirmed by exchange cross peaks observed around the aromatic region of the ROESY spectrum (not shown). As we are able to follow the connectivities through C4 and A11 for the second species we can exclude the possibility of an extrahelical structure.

The neutral C.A mismatch can exist in three structures (see below) but it would be surprising that a change of conformation for the central base pair would strongly influence the chemical shift of the terminal base protons. Rather, the temperature dependence of the concentration of the second species could suggest that it corresponds to single strands. However exchange between a duplex in the presence of an excess of one strand is generally in the fast proton exchange region. We have recorded NOESY spectra on the first strand under otherwise identical conditions. We are able to follow the connectivities for the single strand exactly as for the duplex (not shown). This single strand must have a right handed helical structure. The formation of a such a structure for a single strand has previously been observed (35). All the observed chemical shifts correspond exactly with those of the minor species. The two strands were titrated against each other until integrations of resolved resonances of each strand were equal to better than 5%. The observed resonances of the first strand in the single stranded form do not arise from an excess of this strand. It is not clear why there is an apparent difference in the behaviour of the two single strands. Strand 1 may have a more defined structure than strand 2 or the difference in chemical shifts between the single and double stranded form for strand 2 could be smaller resulting in fast exchange except for C12 H6. Clearly, on raising the pH we have significantly destabilized the duplex. Following the T13 H6 resonance as a function of temperature we observe a T_m for the duplex of ca. 29°C or ca. 10°C below that of the low pH structure. We did

not observe separate resonances for the double and single strand species when measuring the T_m of the low pH species.

We observe significant chemical shift changes for the mismatched pair between the low and high pH structures. On raising the pH we find upfield shifts for A11 H8, A11 H1' and C4 H6 of 0.18, 0.2 and 0.4 ppm respectively. The C4 H1' resonance moves 0.33 ppm downfield. The residues in the 5', but much less so in the 3' direction are also influenced. The G3 H1' and C10 H6 resonances show large upfield shifts.

The region of the NOESY spectrum corresponding to the interactions H8/H6-H2'/H2''/CH₃ confirms the assignments given above and that the C.A bases are intrahelical.

Short mixing time NOESY spectra have also been recorded. These provide the relative assignment of the H2' and H2'' protons. The intensities of the H8/6-H1' cross peaks show that all the bases, except that of C4, are in a normal anti conformation. These cross peaks are ca. 10% of the volume of the CH6-H5 cross peaks. For C4 the cross peak is ca. 50% of that of the average CH6-H5 cross peak volumes. This is considerably smaller than that which would be observed for a syn structure and indicates rather a high anti conformation for this residue. Whereas the C4 sugar showed a significant contribution from a C3' *endo* conformation at low pH we find at high pH it reverts to a normal C2' *endo* conformation.

We have not attempted to determine interproton distances from the NOE build up curves. For certain protons the exchange between single strand and double strand is in the fast exchange region and for others it is in the slow exchange region and this exchange may affect the cross peak volumes.

We have attempted to record spectra in H₂O but at this high pH the imino protons are very broad due to exchange with the solvent. We have not been able to obtain any structural information from these spectra.

DISCUSSION

Low pH structure

The data from this, as well as other studies, establish that below physiological pH, the adenine-cytosine mispair is in a wobble conformation. Two hydrogen bonds are formed due to protonation of the adenine residue.

In the system studied here, the interproton connectivities establish that the A and C residues are intrahelical. At low pH, the A11 H8 proton is 0.20 ppm downfield of the A11 H8 proton of an A.T pair in the same sequence (25). The chemical shift of the C4 H6 proton is essentially the same as the C4 H6 proton of a C.G pair within the same sequence (25). These data are consistent with protonation of the A residue as discussed previously (5).

Table 6. Chemical shifts of pyrimidine H5 protons in different environments. $\Delta\delta$ refers to the difference between the monomer chemical shift and the chemical shift of the H5 proton in the indicated base pair. The indicated base pair is the central base pair in the 7 base pair oligonucleotide duplex studied in the paper.

	δ ppm	$\Delta\delta$		δ ppm	$\Delta\delta$
U monomer	5.89		C monomer	6.05	
U.A	5.08	0.81	C.G	5.25	0.80
U.G	5.44	0.45	C.AH+	5.60	0.45
			C.A	5.20	0.85

Figure 7. Part of the NOESY spectrum recorded at pH 8.9, 12°C with a 400 ms mixing time. On the top axis are marked the chemical shifts of the identified single strand species.

In order to form the protonated wobble pair, the C residue moves towards the major groove and the A residue towards the minor groove. Table 6 lists proton chemical shifts for pyrimidine H5 resonances in several different environments. Chemical shifts of the CH5 of the C.A mispair are compared to the U.A and U.G pairs. Upon formation of a U.A base pair, the chemical shift of the UH5 proton moves upfield by 0.81 ppm. In the U.G wobble pair, the magnitude of the stacking induced upfield shift for the H5 proton is reduced to 0.45 ppm.

In the same neighbouring sequence, the CH5 of the C.G pair experiences a stacking induced shift of 0.80 ppm. In the C.A mispair at low pH, the CH5 proton resonates 0.45 ppm upfield from the corresponding chemical shift of the monomer. These data suggest that the geometry and position of the C in the protonated C.A pair is very similar to the position of the U in the U.G wobble pair.

However, the NMR and MD results show that the protonated structure does not correspond to a unique conformation. The NMR data show clearly a conformational equilibrium for the deoxyribose of C4. This is very well reproduced in the MD calculations, as we have previously seen in other studies (20,24). Agreement between the NMR data and computations with sole restraints on the torsion angles δ , except at the mismatch site, are excellent. This lead us to investigate the hydrogen bonding pattern at the mismatch site during these MD runs. The hydrogen bonding situation is unusual in that there are two donors on one base and two acceptors on the other (with another potential hydrogen bond donor but without any possibility to participate in hydrogen bonding). The NMR data suggest close proximity between the amino groups of A11 and C4 but this cannot be quantified. However, in both MD runs this is the preferred orientation of the mismatched bases, that is structure P is most frequently formed.

2. Structural transition with increase in solvent pH. Structural transitions have been observed previously for duplex oligonucleotides containing the C.A mispair. Recently Jones et al. (7) measured an apparent pK of 6.6 for a 12 base pair duplex containing two C.A mispairs. In this study, we have one C.A mispair in a seven base pair duplex and measured an apparent pK of 7.54 at 23°C. The appearance of a pH dependent structural

transition further confirms that the C.A mispair is protonated at acidic pH.

3. Conformation of the high pH structure. Because the apparent pK of the C.A mispair is near physiological pH, the protonated wobble structure described above is only a partial description of the C.A mispair in DNA. It is important to note that the protonated form was described first in the literature because 1) it is more stable, 2) it forms crystals and 3) the proton resonances are more easily observed at low pH. A study of the C.A mispair at high pH is inherently more difficult due to the relative instability of the neutral structure and problems with observing exchangeable proton resonances in alkaline solution.

As with the low pH form, interproton connectivities establish that the helix containing the C.A pair is a B-DNA and that the A and C residues are intrahelical. The observed pH transition may correspond to the loss of one of three protons: the N1 proton of the protonated A residue, or one of the amino protons from either the C or A residues. Loss of the N1 A proton would generate the neutral wobble structure which could be paired by a single hydrogen bond. Loss of an amino proton would generate rare tautomeric forms of C or A, both in wobble geometry but with two possible hydrogen bonds.

With increasing solvent pH, the chemical shift of the A11 H8 proton shifts upfield by 0.2 ppm, consistent with formation of the neutral, amino form of adenine. The chemical shift of the C4 H6 proton changes very little, 0.01 ppm, during the titration. Based upon our previous results (5), these data demonstrate that structures which invoke rare tautomeric forms are not significant contributors to the neutral C.A structure.

The neutral C.A wobble would then appear to be a plausible candidate for the structure. However significant changes are observed at the neighbouring base pairs which argue in favour of an alternative configuration.

Referring back to Table 6, the stacking induced change in the chemical shift of C4 H5 in the protonated structure compares well with the corresponding shift for U4 H5 in the U.G wobble structure. With increasing pH, the chemical shift of the C4 H5 undergoes a significant, 0.40 ppm, upfield change. Indeed, the stacking induced change, 0.85 ppm, now exceeds that observed for C4 H5 in a C.G pair at the same site. It would appear that the C4 H5 has rotated towards the helix axis and in the neutral C.A species, is closer to the helix axis than in a normal C.G pair. A structure consistent with this change is the reverse wobble shown in Figure 8, bottom. The CH6 and AH8 protons are points of rotation as the A residue rotates towards the major groove and the C residue towards the minor groove.

Figure 8. Structure of the protonated C.A pair, top and structure for the neutral C.A pair bottom.

Table 7. Chemical shifts (ppm) of the CH6 proton of the C.G base pairs adjacent to the mispair site. $\Delta\delta$ refers to the difference in the chemical shift of the H6 proton between the reference (Watson-Crick) base pair and the observed wobble base pair.

Base pair	C10 H6	$\Delta\delta$	C12 H6	$\Delta\delta$
C.G	7.43		7.39	
U.G	7.53	+0.10	7.36	-0.03
C.I	7.47		7.38	
T.I	7.58	+0.11	7.35	-0.02
U.A	7.48		7.30	
C.AH+	7.64	+0.16	7.35	-0.04
C.A	7.35	-0.13	7.41	+0.02

Additional support for the reverse wobble comes from the chemical shifts of the C residues above and below the mispaired A residue in the helix. Previously, we have studied guanine and inosine wobble pairs in the same sequence. We found that the H6 protons of C10 and C12 are sensitive to changes in the central base pair of the helix. When comparing the C.G versus U.G wobble and C.I versus T.I wobble, the chemical shift of C10 H6 is observed to move downfield 0.10 and 0.11 ppm respectively and the C12 H6 moves upfield 0.03 and 0.02 ppm respectively, Table 7. In this sequence, the purine in a wobble configuration moves towards the minor groove and the centre of the purine ring moves closer to C12 H6 and further from C10 H6.

The relative position of the A residue of the C.A pair may then be examined by comparing the relative chemical shifts of C10 H6 and C12 H6 with the sequence containing a central U.A pair. As shown in Table 7, in the protonated C.A structure, C10 H6 has moved downfield 0.16 ppm and the C12 H6 upfield, 0.04 ppm, consistent with movement of the A residue towards the minor groove with formation of the protonated C.A wobble.

With increasing pH, significant changes are observed in the chemical shifts of the neighbouring C10 H6 and C12 H6 resonances. Relative to U.A, the changes in the proton chemical shifts are -0.13 and $+0.02$ ppm respectively for the neutral C.A pair. The magnitude of these changes are similar to those observed when comparing U.A with the protonated C.A but are of opposite sign. These changes are consistent with movement of the A residue towards the major groove for the neutral structure with formation of the reversed wobble structure. We conclude that the configuration of a C.A mismatch in DNA may be described best as a composite of structures, the most predominant of which are the protonated wobble and the neutral reverse wobble. The relative contributions of each would be strongly dependent upon temperature, pH and base sequence. The implications of the equilibrium configurations described here for the fidelity of DNA replication or mismatch repair are as yet unknown.

ACKNOWLEDGEMENTS

We are most grateful to M. Leminor (INSERM) for access to the VAX 8530 computer. M. L. B. is a recipient of grants from the Ligue Nationale contre le Cancer, l'Association pour la Recherche contre le Cancer and from l'Université Pierre et Marie Curie. L.C.S. acknowledges grant GM41336 from the National Institutes of Health.

REFERENCES

- Freese, E. (1959) *J. Mol. Biol.*, **1**, 87–105.
- Topal, M.D. and Fresco, J.R. (1976) *Nature (London)* **263**, 289–293.
- Patel, D.J., Kozlowski, S.A., Ikuta, S. and Itakura, K. (1984) *Biochemistry*, **23**, 3218–3226.
- Hunter, W.N., Brown, T., Anand, N.N. and Kennard, O. (1986) *Nature (London)*, **320**, 552–555.
- Sowers, L.C., Fazakerley, G.V., Kim, H., Dalton, L. and Goodman, M.F. (1986) *Biochemistry*, **25**, 3983–3988.
- Sowers, L.C., Shaw, B.R., Veigl, M.L. and Sedwick, W.D. (1987) *Mutat. Res.* **177**, 201–218.
- Wang, C., Gao, H., Gaffney, B.L. and Jones, R.A. (1991) *J. Amer. Chem. Soc.*, **113**, 5486–5488.
- Sowers, L.C., Goodman, M.F., Eritja, R., Kaplan, B. and Fazakerley, G.V. (1989) *J. Mol. Biol.*, **205**, 437–447.
- Sowers, L.C., Eritja, R., Kaplan, B., Goodman, M.F. and Fazakerley, G.V. (1988) *J. Biol. Chem.*, **263**, 14794–14801.
- Gao, X. and Patel, D.J. (1988) *J. Am. Chem. Soc.*, **110**, 5178–5182.
- Carbonnaux, C., van der Marel, G.A., van Boom, J.H., Guschlbauer, W. and Fazakerley, G.V. (1991) *Biochemistry*, **30**, 5449–5458.
- Brown, T., Leonard, G.A., Booth, E.D. and Kneale, G. (1990) *J. Mol. Biol.*, **212**, 437–440.
- Boulard, Y., Gabarro-Arpa, J., Cognet, J.A.H., Le Bret, M., Guy, A., Téoule, R., Guschlbauer, W. and Fazakerley, G.V. (1991) *Nucl. Acids Res.* **19**, 5159–5167.
- Weiner, P. and Kollman, P.A., J. (1981) *J. Comp. Chem.*, **2**, 287–303.
- Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A. (1986) *J. Comp. Chem.*, **7**, 230–252.
- Singh, U.C., Weiner, P.K., Caldwell, J.W. and Kollman, P.A., (1986) AMBER 3.0 University of California, San Francisco.
- Singh, U.C. and Kollman, P.A. (1984) *J. Comp. Chem.*, **5**, 129–145.
- Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G. and Weiner, P. (1984) *J. Am. Chem. Soc.* **106**, 765–784.
- Bestler, B.H., Merz, K.M. and Kollman, P.A. (1990) *J. Comp. Chem.* **7**, 431–439.
- Cognet, J.A.H., Gabarro-Arpa, J., Cuniasse, Ph., Fazakerley, G.V. and Le Bret, M. (1990) *J. Biomol. Str. Dynam.*, **7**, 1095–1115.
- Gabarro-Arpa, J., Cognet, J.A.H. and Le Bret, M. *J. Mol. Graphics*, (in press).
- Le Bret, M., Gabarro-Arpa, J., Gilbert, J.Ch. and Lemarechal, Cl. *J. Chim. Phys. Phys-Chim. Biol.* (in press).
- Cuniasse, P., Sowers, L.C., Eritja, R., Kaplan, B., Goodman, M.F., Cognet, J.A.H., Le Bret, M., Guschlbauer, W. and Fazakerley, G.V. (1987) *Nucleic Acids Res.*, **15**, 8003–8022.
- Cuniasse, P., Sowers, L.C., Eritja, R., Kaplan, B., Goodman, M.F., Cognet, J.A.H., Le Bret, M., Guschlbauer, W. and Fazakerley, G.V. (1989) *Biochemistry*, **28**, 2018–2026.
- Carbonnaux, C., Fazakerley, G.V. and Sowers, L.C. (1990) *Nucleic Acids Res.* **18**, 4075–4081.
- Raszka, M. (1974) *Biochemistry* **13**, 4616–4622.
- Rajagopal, P. and Feigon, J. (1989) *Nature* **339**, 637–640.
- de los Santos, C., Rosen, M. and Patel, D. (1989) *Biochemistry* **28**, 7282–7289.
- Sklenar, V. and Feigon, J. (1990) *Nature* **345**, 836–838.
- Mooren, M.M.W., Pulleyblank, D.E., Wijmenga, S.S., Bloomers, J.J. and Hilbers, C.W. (1990) *Nucleic Acids Res.*, **18**, 6523–6529.
- Kalnik, M.W., Kouchakdjian, M., Li, B.F.L., Swann, P.F. and Patel, D.J. (1988) *Biochemistry*, **27**, 100–108.
- Fazakerley, G.V., van der Marel, G.A., van Boom, J.H. and Guschlbauer, W. (1984) *Nucleic Acids Res.*, **12**, 8269–8279.
- McConnell, B. and Seawell, P.C. (1973) *Biochemistry* **12**, 4426–4434.
- EMBO Workshop on DNA Curvature and Bending, Churchill College, Cambridge, UK. Sept. 10–15, 1988. Sarma, R.H. (1988) *J. Biomol. Str. Dyn.* **6**, 391–395; (1989) *EMBO J.* **8**, 1–4.
- Clore, G.M. and Gronenborn, A.M. (1984) *Eur. Biophys. J.* **11**, 95–102.