N.m.r. determination of the solution conformation and dynamics of the A·G mismatch in the d(CGCAAATTGGCG)₂ dodecamer

Andrew N. LANE,*§ Terence C. JENKINS,†§ Dorcas J. S. BROWN‡ and Tom BROWN‡

*Laboratory of Molecular Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, U.K.,

†Cancer Research Campaign Biomolecular Structure Unit, Institute of Cancer Research, Sutton, Surrey SM2 5NG, U.K., and

‡Department of Chemistry, University of Edinburgh, West Mains Road, Edinburgh EH9 3LL, U.K.

A·G base-paired mismatches that occur during replication are among the most difficult to detect by repair enzymes. Such purine·purine mispairs can exist in two conformations, one of which is stabilized by protons [Gao & Patel (1988) *J. Am. Chem. Soc.* **110**, 5178–5182]. We have undertaken a ¹H-n.m.r. and ³¹P-n.m.r. study of the mismatched dodecamer d(CGCAAATTGGCG)₂ as a function of both temperature and pH to determine the conformational features of the A·G mismatch. At pH > 7 the mispaired bases are each in the *anti* conformation and are stacked in the B-like helix. As the pH is decreased, a second conformation becomes populated (apparent pK_a approx. 5.9) with concomitant changes in the chemical shifts of protons of the mispaired bases and their nearest neighbours. Data from two-dimensional nuclear-Overhauser-enhancement spectroscopy show unequivocally that, at low pH, the dominant conformation is one in which the mismatched G residues are in the *syn* conformation and are hydrogen-bonded to the A residues that remain in the *anti* conformation. Residues not adjacent to the A·G sites are almost unaffected by the transition or the mispairing, suggesting considerable local flexibility of the unconstrained duplexes. Despite the bulging of the mispaired bases, the conformation of the A(*anti*)·G(*anti*) duplex is very similar to the native dodecamer, whereas the AH⁺(*anti*)·G(*syn*) duplex shows a greater variation in the backbone conformation at the mismatched site. According to the chemical shifts, the duplex retains twofold symmetry in solution. The equilibrium between the *syn* and *anti* conformations of G9/G21 is strongly dependent on pH, but only weakly dependent on temperature (Δ*H* approx. 16 kJ·mol⁻¹). The first-order rate constant for the transition is approx. 9 s⁻¹ at 283 K and approx. 60 s⁻¹ at 298 K, with an activation enthalpy of approx. 100 kJ·mol⁻¹. The stabilization of the A(*anti*)·G(*syn*) conformation by protons is consistent with models invoking N1 protonation of adenine. Using the derived glycosidic torsion angles we have used restrained molecular dynamics to build models of the neutral and protonated d(CGCAAATTGGCG)₂ oligomers. The results confirm that the A(*anti*)·G(*anti*) and AH⁺(*anti*)·G(*syn*) conformations are favoured at high pH and low pH respectively, in accord with n.m.r. and single-crystal X-ray data.

INTRODUCTION

Base-pair mismatches occur relatively frequently during replication, and must therefore be recognized and excised by error-checking apparatus to decrease the rate of point-mutation to an acceptable level (approx. 10⁻⁸). One method of error checking is the 3′-5′-exonuclease activity of DNA polymerases, which must not only recognize that an incorrect base-pair has been formed, but also excise the incorrect nucleotide. It seems likely that the efficiency of this error correction will depend on the nature of the mismatched pair, and possibly also on the neighbouring sequence. Further, the A·G mismatch has the highest frequency of escaping detection by error-correcting processes (Fersht *et al.*, 1982), presumably because it does not significantly disrupt the conformation of the helix at the local level.

Crystallographic studies have shown that the A·G mispair can exist as the A(*anti*)·G(*syn*) state (Brown *et al.*, 1989). However, spectrophotometric thermal denaturation ('melting') measurements as a function of pH suggest that in solution there may be a pH-dependent dynamic equilibrium between this state and one in which both nucleotides adopt the *anti* conformation (Brown *et al.*, 1990) (see Fig. 1 for structures). Indeed, Gao & Patel (1988) showed by n.m.r. that this equilibrium does exist in solution for both the d(CGAGAATTCACG)₂ and d(CGAGAATTGGCG)₂ duplexes, with the A(*anti*)·G(*syn*) state stabilized by protonation.

In the present paper we describe the pH-dependent conformational properties in solution of the A·G mispair in a related dodecamer sequence, namely d(CGCAAATTGGCG)₂,

Fig. 1. Structure of the A·G wobble base-pairs

(a) A(*anti*)·G(*anti*); (b) AH⁺(*anti*)·G(*syn*).

Abbreviations used: NOE, nuclear Overhauser enhancement; NOESY, two-dimensional nuclear-Overhauser-enhancement spectroscopy; HOHAHA, homonuclear Hartmann–Hahn spectroscopy; c.p.u., central processor unit.

§ To whom correspondence should be addressed (either author).

and the determination of the thermodynamics and kinetics of the $G(\text{syn}) \rightleftharpoons G(\text{anti})$ transition.

MATERIALS AND METHODS

Materials

The oligodeoxyribonucleotide d(CGCAAATTGGCG) was synthesized by means of phosphoramidite chemistry and purified by ion-exchange h.p.l.c., reverse-phase h.p.l.c., Sephadex filtration and finally Dowex 50 (Na^+ form) filtration. For n.m.r. studies, the annealed duplex was dialysed versus aqueous 10 mM-sodium phosphate buffer, pH 7, containing 0.1 mM-KCl, freeze-dried and redissolved in 100% $^2\text{H}_2\text{O}$ (Aldrich Chemical Co.) containing 0.1 mM-4,4-dimethylsilapentane-1-sulphonate as internal reference. The pH was adjusted by addition of ^2HCl or NaO^2H as required; in $^2\text{H}_2\text{O}$ solutions the uncorrected meter reading was used, and is referred to as pD*.

N.m.r. spectroscopy

^1H -n.m.r. spectra were recorded at 500 MHz over a spectral width of 5 kHz with 16384 points in $^2\text{H}_2\text{O}$. One-dimensional spectra in $^1\text{H}_2\text{O}$ solutions were recorded over a spectral width of 12 kHz with the use of the 1331 sequence of Hore (1983) for solvent suppression. A relaxation delay of 2.4 s (recycle time 4 s) was used for quantitative work.

^{31}P -n.m.r. spectra were recorded at 81 MHz over a spectral width of 600 Hz with 2048 complex data points. Protons were decoupled at all times by using Waltz-16. Heteronuclear P-H correlation was performed with 40 ms delays for refocusing, with 1024 points in F2 and 128 points in F1 and with spectral widths of 600 Hz and 1500 Hz in F2 and F1 respectively. The data matrix was zero-filled to 2048 by 1024 points before Fourier transformation.

Phase-sensitive NOESY and HOHAHA spectra were recorded by using the time-proportional phase incrementation scheme (Marion & Wüthrich, 1983) with 2048 points in F2 and 512 points in F1. MLEV-17 was used for the spin-lock field (Bax & Davis, 1985), which was 7.9 kHz. Data matrices were zero-filled to 2048 by 2048 points and apodized by using 90°-shifted sine-squared functions in both dimensions before Fourier transformation. The receiver phase was adjusted to detect the sine-modulated data, and the deadtime was adjusted to ensure zero signal for the first point (Frenkiel *et al.*, 1990). The NOESY spectrum in $^1\text{H}_2\text{O}$ was recorded with presaturation, as the exchange of the imino protons with water protons is slow at low buffer concentration and low pH (Rajagopal *et al.*, 1988).

One-dimensional truncated NOEs were recorded by using the method of Wagner & Wüthrich (1979), and correlation times and conformations were determined from the time-dependent NOE data as previously described (Lane *et al.*, 1986; Lane, 1990). Saturation transfer measurements were made by using the truncated NOE experiment, and the rate constants for exchange (k) were recovered by non-linear regression to the equation:

$$\Delta M / M^0 = [k / (\rho + k)] \{1 - \exp [-(\rho + k)t]\} \quad (1)$$

where ΔM is the change in magnetization of the observed resonance, M^0 is its equilibrium magnetization and ρ is the intrinsic spin-lattice relaxation rate constant.

Molecular modelling and structure refinement

The initial co-ordinates of the d(5'-HO-CGCAAATTGGCG-OH-3')₂ dodecamer duplex were those of idealized canonical B-DNA (Arnott *et al.*, 1983) and were generated by using the program GENHELIX2 (T. C. Jenkins, unpublished work). Inter-active molecular modelling was performed by using the GEMINI

1.03 graphics package (C.R.C. Biomolecular Structure Unit, Institute of Cancer Research, Sutton, Surrey, U.K.) with all visualizations on a Silicon Graphics Iris 4D-20G workstation. All calculations were carried out on an Alliant FX40/3 computer.

Initial models for each of the four possible *anti-anti*, *anti-syn*, *syn-anti* and *syn-syn* A·G-base-paired duplexes were prepared by rotation of the appropriate C1'(sugar)-N(base) glycosidic torsion angles, with χ -values altered in tandem for corresponding nucleosides in each strand of the duplex (i.e. A4 $\equiv$ A16 and G9 $\equiv$ G21). Glycosidic torsion angles (χ) are defined by O4'-C1'-N1-C2 for pyrimidines and O4'-C1'-N9-C4 for purines. Equivalent models were also generated in which the N1 purine ring positions of A4 and A16 were simultaneously protonated; standard bond lengths and angles were used for the construction. Thus a total of eight dodecanucleotide duplex models were generated for subsequent structure refinement and energy minimization.

The energy was minimized at the all atom level by using the X-PLOR V1.5 program (Brünger, 1988), adapted (J. Jiang & R. E. Hubbard, University of York, U.K., 1989) for double-precision use with the Alliant FX40 array processor, and CHARMM force-field parameterization (Brooks *et al.*, 1983). Force-field parameters for the N1-protonated adenosine (the ADH residue) were interpolated from those for the purine nucleotides. Atom-centred charges for the individual 2'-deoxynucleotides were derived from the semi-empirical MNDO wavefunctions calculated for the corresponding 3',5'-(CH₃OPO₂O)₂- or 5'(3')-hydroxy-3'(5')-(CH₃OPO₂O)-substituted deoxynucleosides by using the AMPAC molecular orbital package (QCPE Program 506). Such fragments were constructed by using idealized B-DNA co-ordinates, except in the case of ADH, where the initial coordinates of the purine and C1'(sugar) atoms were subjected to prior quantum-mechanical (MNDO) minimization. In this case satisfactory SCF convergence was achieved after 5660 c.p.u.·s to a final energy gradient of $\leq 1.2 \text{ kJ} \cdot \text{mol}^{-1}$ (i.e. $\Delta\Delta H, \leq 0.03 \text{ kJ} \cdot \text{mol}^{-1}$). Electrostatic charges were computed according to the procedure suggested by Singh & Kollman (1984) by using the program MEP, with quantum-mechanical and monopole-monopole molecular electrostatic potentials (MEPs) calculated at points on two Connolly surface layers by using the procedure of Orozco & Luque (1990a,b). A surface density of 0.75 point/unit area was employed. Partial charges were fitted to the final MEPs for each nucleotide by using MEPFIT, a Levenberg-Marquardt non-linear optimization procedure. Typically, each calculation of molecular electrostatic potential required approx. 12 600 c.p.u.·s to achieve satisfactory fits.

The X-PLOR protocol used for structure refinement is that recommended by Brünger (1990), with the use stepwise of (i) 200 steps of conjugate-gradient minimization, (ii) molecular dynamics with heating to 300 K followed by 5 ps dynamics at constant temperature, and (iii) conjugate-gradient (molecular-mechanics) minimization to an ultimate root-mean-square gradient of $\leq 4 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-1}$ (i.e. $\leq 0.10 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-1}$). Glycosidic torsion-angle constraints alone were applied throughout the three-stage refinement. Typically each minimization required approx. 15 700 c.p.u.·s on the Alliant FX40/3.

Solvent water and counterions were not explicitly included; instead their effects were simulated by using a distance-dependent dielectric constant with $\epsilon = 4r_{ij}$ for conjugate-gradient minimization steps and $\epsilon = r_{ij}$ for the dynamics calculations. Electrostatic damping in this manner has recently been tested in a model nucleic acid system (Orozco *et al.*, 1990). No attempt was made to decrease or mask the net charge on the phosphate residue (Tidor *et al.*, 1983). The hydrogen-bonded interactions were switched on for heavy-atom-donor-acceptor distances in

Fig. 2. Effect of pH on the n.m.r. spectra of d(CGCAAATTGGCG)₂.

¹H-n.m.r. spectra were recorded at 400 MHz, and ³¹P-n.m.r. spectra were recorded at 81 MHz with Waltz-16 proton decoupling. (a) ¹H-n.m.r. spectra of base protons. Spectra were recorded at 298 K with 16384 complex points over a spectral width of 4 kHz and a recycle time of 3 s. Free induction decays were apodized with an exponential function (line broadening = 1 Hz) before Fourier transformation. Upper spectrum, pD* 5.1; lower spectrum, pD* 7.45. (b) ¹H-n.m.r. spectra of imino protons. Spectra were recorded with 16384 points, zero-filled to 32768 points over a spectral width of 10 kHz with the use of the 133I sequence (Hore, 1983) at 283 K. Free induction decays were apodized with an exponential function (line broadening = 2 Hz) before Fourier transformation. Upper spectrum, pH 5.6; lower spectrum, pH 7.5. (c) ³¹P-n.m.r. spectra. Spectra were recorded at 298 K with 2048 points zero-filled to 8192 points over a spectral width of 600 Hz. Free induction decays were apodized with an exponential function (line broadening = 0.5 Hz) before Fourier transformation. Upper spectrum, pD* 5.3; lower spectrum, pD* 7.45.

the range 0.55–0.65 nm, and terms up to 0.75 nm were included. Non-bonded energy terms were included up to 1.15 nm, with switching between 0.95 and 1.05 nm, and a factor of 0.4 was used to damp 1,4 electrostatic interactions. Explicit distance restraints corresponding to Watson–Crick base-paired geometry were not employed.

RESULTS AND DISCUSSION

Effects of temperature and pH on the n.m.r. spectra

According to Brown *et al.* (1989, 1990) the conformation of the A·G mispair is A(*anti*)·G(*syn*) at low pH in the crystal state, and the conformation changes as the pH is raised from 5 to > 7. We have recorded n.m.r. spectra at different pH values to characterize the nature of this conformation change in greater detail. One-dimensional n.m.r. spectra of the mismatched dodecamer at two pD* values are shown in Fig. 2(a). The spectra

in the aromatic region are significantly different. In particular, a broad peak is present at 8.4 p.p.m. at low pH, which moves upfield in the high-pH spectrum. The area of this peak is equivalent to about 0.6 proton under these conditions, and probably represents an exchange process between two conformations that is moderately slow on the chemical-shift timescale (and see below).

The spectra of the imino protons are shown in Fig. 2(b). At high pH there are only peaks expected for G·C and A·T base-pairs, which are only slightly attenuated by saturation of the water resonance. In contrast, at low pH there are additional peaks near 10 p.p.m. that are readily bleached by saturation of the water resonance, implying that these protons are only weakly hydrogen-bonded. Clearly, different conformations are present at the two pH values.

³¹P-n.m.r. spectra at 81 MHz are shown in Fig. 2(c). At high pH there are sharp resolved resonances whose chemical shifts are very similar to those observed in the native dodecamer (Ott &

Fig. 3. NOESY spectra of d(CGCAAATTGGCG)₂ at pD* 5.6

The phase-sensitive NOESY spectrum was recorded at 500 MHz and 283 K as described in the Materials and methods section. The data matrix consisted of 2048 points in F2 and 512 in F1, zero-filled to 2048 points over a spectral width of 5 kHz. The free induction decays were apodized with a 90°-shifted sine-squared function in both dimensions before Fourier transformation. (a) Base proton to sugar H1'; (b) base proton to sugar H2' and H2''.

Table 1. Proton assignments for d(CGCAAATTGGCG)₂ at 283 and 298 K and pD* 5.7 and 7.45

Proton resonances were assigned from NOESY and HOHAHA spectra as described in the text. The first and second entries refer to 283 and 298 K respectively. Chemical shifts (δ) are referenced to internal 4,4-dimethylsilapentane-1-sulphonate. Abbreviation: N.D., determined.

Base	δ (p.p.m.)						
	H8/6	H5/2/Me	H1'	H2'	H2''	H3'	H4'
Assignments at pD* 5.7							
C1	7.61/62	5.82/87	5.68/72	1.98	2.39	4.68/68	4.05
G2	7.95/96	—	5.90/98	2.62/63	2.73/72	4.97/96	4.33
C3	7.38/37	5.44/43	5.30/N.D.	2.05/05	2.32/32	4.82/80	4.10
A4	8.44/43	7.57/7.60	5.94/92	2.80/79	3.00/2.95	5.08/N.D.	4.44
A5	8.21/18	7.02/7.02	5.93/90	2.70/69	2.93/91	5.07/05	4.44
A6	8.15/11	7.42/7.45	6.12/12	2.55/57	2.93/90	5.02/01	4.47
T7	7.12/09	1.23/25	5.88/86	2.00/1.97	2.52/50	4.84/82	4.18
T8	7.24/22	1.54/55	5.98/96	2.05/06	2.82/75	4.87/86	4.21
G9	7.67/70	—	5.71/70	2.60/58	3.18/15	N.D./4.68	N.D.
G10	7.90/83	—	5.67/70	2.54/62	2.62/70	4.98/97	4.39
C11	7.35/32	5.38/39	5.77/76	1.90/88	2.33/32	4.81/80	4.18
G12	7.91/90	—	6.09/10	—	2.6/2.6	4.67/65	4.15
Assignments at pD* 7.45							
C1	7.61/62	5.83/87	5.67/73	1.91/91	2.38/37	4.67/67	4.05
G2	7.96/94	—	5.81/82	N.D./2.36	2.63/65	4.96/95	4.32
C3	7.17/17	5.37/39	5.64/65	1.39/39	1.98/99	4.76/N.D.	4.06
A4	8.07/05	7.68/7.70	5.7/70	2.75/74	2.85/84	4.99/98	4.31
A5	8.15/13	7.02/05	5.78/79	2.7/68	N.D.	5.03/03	4.42
A6	8.17/14	7.52/7.52	6.13/13	2.6/2.58	2.88/88	4.82/81	4.45
T7	7.14/10	1.28/29	5.84/83	1.93/1.92	2.46/43	4.82/81	4.13
T8	7.22/21	1.59/60	5.74/74	1.76/74	2.17/14	4.84/83	N.D.
G9	7.98/96	—	5.69/67	N.D./2.75	2.81/80	5.00/4.99	4.38
G10	7.77/77	—	5.91/91	2.57/56	2.72/69	4.98/98	4.40
C11	7.32/31	5.38/40	5.70/74	1.91/90	2.32/32	4.83/82	4.15
G12	7.92/92	—	6.10/13	—	2.6/2.6	4.68/66	4.15

Eckstein, 1985). At low pH there are many fewer sharp resolved peaks, and some rather broad resonances, again indicating the presence of exchange among two or more conformations that involve the phosphate backbone. Further, the chemical-shift dispersion at low pH (Δδ 0.84 p.p.m.) is considerably greater than at high pH (Δδ 0.52 p.p.m.), which can be taken as evidence

of a wider variety of phosphate torsion angles present in the low-pH conformation. The chemical-shift dispersion found at high pH is similar to that commonly observed in normal B-like DNA (Gorenstein *et al.*, 1988; Ott & Eckstein, 1985).

Hence it appears that there is a single dominant conformation present at high pH which resembles that of the native dodecamer,

Fig. 4. Changes in chemical shift in NOESY spectra of d(CGCAAATTGGCG)₂ compared with the native dodecamer

Chemical shifts of the native dodecamer are taken from Lane *et al.* (1991), and the differences for the spectra of d(CGCAAATTGGCG)₂ at pD* 7.45 (▨) and pD* 5.3 (□) were calculated. Differences for the mismatched bases (A4/G9) are not included. (a) Chemical-shift differences for the H8/H6 resonances; (b) chemical-shift differences for the H2' resonances.

Table 2. Assignments of exchangeable protons in the A·G-mispaired dodecamer

Proton assignments are given at 283 K and were determined from temperature- and pH-dependent one-dimensional spectra in ²H₂O and NOESY at pH 5.3 and 278 K.

Resonance	δ (p.p.m.)
G2·C11 NH	13.05
C3 N4NH ₂	8.45, 6.80
A4 NH ₂	9.95, 8.45 (<i>syn</i>)
A5·T8 NH	13.55
A5·T8 NH ₂	6.25
A6·T7 NH	13.60
A6·T7 NH ₂	6.05
G9 N1H	11.0 (<i>syn</i>)
	12.2 (<i>anti</i>)
G10 N1H	12.85 (<i>anti</i>)
	12.75 (<i>syn</i>)
C11 N4NH ₂	8.27, 6.65

whereas at low pH there are at least two conformations, at least one of which is substantially different from that present at high pH. The dependence of the populations and the rate of interconversion on pH and temperature are discussed below.

¹H-n.m.r. assignments

Non-exchangeable proton resonances were assigned primarily by recording NOESY spectra at pD* 5.7 and 7.45 at both 283 and 298 K, with some assistance from HOHAHA spectra.

NOESY spectra at both pD* values were recorded with the use of mixing times of 150 ms at 298 K and 90 ms at 283 K. These mixing times are inversely proportional to the correlation times at the two temperatures, when the relative NOE intensities should not differ if the conformations are independent of the temperature (Lane, 1991).

Portions of the lower-temperature NOESY spectra at low pD* are shown in Fig. 3. Because some peaks are quite broad at low pH, some cross-peaks are weak. However, all the connectivities could be found in the usual way at the relatively long mixing times used. The ¹H assignments at the two pH values are given in Table 1. As the data in Table 1 show, there are substantial differences between shifts of those bases involved in the mismatch and the nearest neighbours, but not elsewhere in the sequence. Further, the chemical shifts of the resonances are for the most part very similar to those in the native dodecamer (Hare *et al.*, 1983; Lane *et al.*, 1991), except for the immediate neighbours of the mismatched bases, as is shown in Fig. 4. This shows that there are differences in the conformations at the two pH values, but only around the site of the mismatch; elsewhere in the sequence the conformation appears to be very similar to the native non-mismatched dodecamer (and see below).

The exchangeable protons and adenine H2 resonances were assigned by using NOESY and one-dimensional NOE experiments in ¹H₂O (cf. Fig. 2b), at low temperature (278 and 283 K), and from the temperature-dependence of the resonance widths. The sharp resonances at approx. 13.5 p.p.m. are readily assigned to the NH group of A5·T8 and A6·T7, and the sharp resonances near 13.0 p.p.m. to the NH group of G2·C11 and C3·G10. The remaining broader resonances exchange relatively readily, as shown by increasing the temperature to 298 K. Cytosine and thymidine amino resonances were assigned from NOESY spectra (results not shown), which show strong cross-peaks both to the imino proton and to each other. Although we observed no NOE from the peak at 11 p.p.m. (low pH), all of the other resonances in the imino proton region are accounted for. We therefore tentatively assign the resonance at 11 p.p.m. (low pH, Fig. 2b) to the N1H atom of G9 in the *syn* conformation, which presumably moves downfield to 12.2 p.p.m. in the *anti* conformation. The chemical shifts of the exchangeable protons are given in Table 2.

Fig. 5 shows a portion of the low-pH NOESY spectrum, in which there are exchange cross-peaks between the base proton resonances corresponding to the two conformations for C3, A4, A5, G9 and G10. The chemical shifts of the two states can therefore be simply read off from the spectrum, and it is found that in all cases one of them corresponds to the chemical shift found at high pH, where no such exchange cross-peaks were observed. Exchange cross-peaks were also observed for the H1' resonances of C3, A4, T8 and G9 (results not shown). Hence the simplest model consistent with the data is one in which there are two conformations, one of which is stabilized by protons. The rate of exchange between these two conformations must also be of the order 1/*t_m* for exchange peaks to be observed (i.e. 5–10 s⁻¹).

We note also that at high pH, where there is a single conformation, there is only one set of signals in both the ¹H-n.m.r. and the ³¹P-n.m.r. spectra, implying that the two strands are magnetically equivalent. This observation is true also of the low-pH spectra when due allowance is made for the exchange between the high-pH and low-pH states.

Conformation

The correlation time for overall rotation of the molecule was determined as 3.3 ± 0.3 ns at 298 K from the cross-relaxation-rate constant for the H6–H5 vectors of the three cytosine residues. This value is similar to that determined for the native dodecamer

Fig. 5. Chemical exchange in $d(CGCAAATTGGCG)_2$ at $pD^* 5.6$

The NOESY spectrum is the same as in Fig. 3, showing exchange peaks for the base protons. The relative intensities of the exchange cross-peaks show that the rate constant for interconversion is approx. 5 s^{-1} .

($3.0 \pm 0.2\text{ ns}$ at 303 K ; Lane *et al.*, 1991), indicating that the duplex behaves hydrodynamically as a typical rod-like helix. Further, as the linewidths of base protons not involved in chemical exchange scale according to the viscosity, we conclude that the double-helical character is maintained over the temperature range $278\text{--}303\text{ K}$.

Fig. 6 shows cross-sections from a NOESY spectrum through the H8 of A4, A5 and G9 at $pD^* 5.6$ and 298 K , all plotted on the same scale. Both A4 and A5 show intra-residue relative intensities in the order $H2' > H2'' > H1'$, which is characteristic of the *anti* conformation, whereas the G9 cross-section shows relative intensities in the order $H1' > H2' \approx H2''$, which is characteristic of the *syn* conformation. At high pD^* values all cross-sections through the H8 show weak NOE intensities to $H1'$ and $H2''$ and strong intensities to $H2'$ (results not shown). Further, the NOE intensities to the nearest neighbours of A5 are typical of B-like DNA, whereas that of G10 at low pH is consistent with a conformation in which the base is aligned *syn* to the deoxyribose (Table 3). Finally, there is an NOE from A4 H2 to G10 H8, which can only occur if one of the bases adopts the *syn* conformation. Hence we conclude that, at high pH, the mismatch is $A(anti) \cdot G(anti)$, whereas at low pH it is predominantly $A(anti) \cdot G(syn)$.

These conclusions are supported by calculations of the expected NOESY intensities for the mixing times used and the measured correlation time of the duplex. We have used the program NUCFIT (Lane, 1990) to calculate the expected NOE intensities for the $A(anti) \cdot G(syn)$ and the $A(anti) \cdot G(anti)$ mismatches in the $d(CAA) \cdot d(TGG)$ segment. The cross-peak intensities, normalized to the cytosine H6–H5 cross-peak intensity, are given in Table 3 for a correlation time of 5 ns and a mixing time of 90 ms (corresponding to the experiments performed at 283 K), along with the experimental values. Clearly, at low pH the data are consistent with the $A(anti) \cdot G(syn)$ conformation but not the $A(anti) \cdot G(anti)$ conformation. The inter-nucleotide NOE intensities also indicate that the mismatched bases are stacked

Fig. 6. Cross-sections from a NOESY spectrum of $d(CGCAAATTGGCG)_2$ at $pD^* 5.6$ and 283 K

The NOESY spectrum was recorded and transformed as described in the text, with zero-filling to 4096 points in F2 and 2048 in F1. Cross-sections parallel to F2 were taken through diagonal peaks corresponding to (a) A4 H8, (b) A5 H8 and (c) G9 H8 to show the relative intensities of the $H1'$, $H2'$ and $H2''$ cross-peaks.

inside the helix (on a time-averaged basis) in both conformations. Although the intranucleotide NOE intensities match those for the B-DNA conformation (except G9), there are significant deviations for the inter-nucleotide intensities, indicating the presence of slight variations from the canonical B-type structure. This is a common finding for DNA duplexes in solution (Lefèvre *et al.*, 1987; Nerdal *et al.*, 1988).

One-dimensional NOE time courses were measured at $pD^* 7.45$ and 298 K for irradiation of the H8/H6 of A5, A6, T7 and T8; the nucleotide conformations were determined by using the program NUCFIT. Fig. 7 shows a typical time course and

Table 3. NOE intensities in the G(anti) and G(syn) conformations

NOESY intensities were calculated by using the program NUCFIT (Lane, 1990) for a 3 bp segment d(CAA)·d(TGG) for a 5 ns correlation time and for mixing times up to 200 ms. The pseudo-rotation phase angles were all set at 162°, and the glycosidic torsion angles were all −110°, except for the G(syn) conformation, where it was 250°. the helical twist and rise were 36° and 0.34 nm respectively. The slide, shift, tilt and roll angles were all set to zero. NOE intensities at 90 ms were normalized to that for the cytosine H6–H5 vector.

Nucleotide	H8(<i>i</i>)–H _j (<i>i</i>)				H8(<i>i</i>)–H _j (<i>i</i> –1)		
	H1′	H2′	H2″	H3′	H1′	H2′	H2″
Calculated intensities							
A4	0.1	0.83	0.27	0.11	0.39	0.86	2.5
A5	0.1	0.83	0.28	0.11	0.39	0.87	2.5
G9 (<i>syn</i>)	0.9	0.06	0.09	0.01	< 0.01	< 0.01	< 0.01
G10	0.1	0.97	0.36	0.12	0.29	0.44	1.3
Observed intensities							
A4	0.15	1.0	0.4	0.2	0.1	0.1	1.0
A5	–	1.3	–	0.26	–	–	1.0
G9	0.85	0.1	0.1	–	–	< 0.1	< 0.1
G10	–	–	–	–	0.1	–	1.0

Fig. 7. Confirmation of T7 of d(CGCAAATTGGCG)₂ at pD* 7.4 and 298 K

One-dimensional NOEs to T7 H1′, H2′, H2″ and H3′ were obtained for irradiation of T7 H6 as described in the text. The lines are the simultaneous best-fit curves to the data obtained by using NUCFIT (Lane, 1990), assuming $P_s = 144^\circ$. ●, H3′; ○, H1′; □, H2″; ■, H2′.

the associated best fits. The conformational parameters are given in Table 4, where they are compared with those obtained for the native dodecamer (Lane *et al.*, 1991). It is immediately apparent that the conformations of the nucleotide residues outside the mismatch site are not perturbed by the A·G base-pairing, indicating that the local conformation is sufficiently flexible to accommodate significant conformational changes. We have also determined the cross-relaxation-rate constant for the A5 H2–A4 H2 and A5 H2–A6 H2 vectors. These vectors are nearly parallel to the helix axis, so that the appropriate correlation time for the cross-relaxation is τ_L , which can be calculated from the cytosine correlation time assuming an axial ratio of 1.7 for the dodecamer (Birchall & Lane, 1990). We obtain a value for τ_L of 4.6 ns at 298 K. Using this correlation time, we obtain an apparent distance of 0.35 nm for both the A5–A4 and the A5–A6 steps. The distance for the A5–A6 vector is identical within experimental error with the distance determined for the native dodecamer. We note that in the NOESY spectra at low pH there is a weak NOE between A5 H2 and A6 H1′, consistent with a significant negative base-pair roll or propeller twist at this step. There is also a stronger cross-strand NOE between A5 H2 and G9 ($\equiv$ G21) H1′. Hence the helical conformation of the central region of the molecule is not significantly perturbed by the A(anti)·G(anti) mismatch sites. This conclusion is supported by the ³¹P-n.m.r.

Table 4. Conformation of the A·G-mispaired dodecamer at pD* 7.45 and 298 K

Conformations were determined from one-dimensional NOE time courses at 298 K and 400 MHz. Values for the native d(CG-CGAATTCGCG)₂ duplex are taken from Lane *et al.* (1991). A4* refers to the conformation determined from NOEs at pD* 5.7 and 283 K. The conformation of A5 is based on an incomplete data set, and the fraction C2′(endo) is assumed to be unity. *R* is the mean fractional error defined as:

$$R = (1/N)\Sigma|(NOE_{obs.} - NOE_{calc.})/NOE_{obs.}|.$$

Residue	A·G-mispaired			Native
	χ (°)	f_s	<i>R</i>	χ (°)
C3	−117	1.0	0.12	−115
A4*	−101	1.0	0.06	–
A5	(−106)	(1.0)	–	−110
T7	−114	0.98	0.16	−115
T8	−114	0.85	0.10	−117
G10	−104	1.0	0.19	−104
C11	−113	0.9	0.17	−117

spectrum, which at high pH is very similar to that of the native dodecamer (see Fig. 2c), suggesting that the conformation of the phosphate backbone is not greatly perturbed by the A(anti)·G(anti) base-pairing, despite the increased separation of the C1′ atoms that is produced by the A·G wobble base-pair (cf. Fig. 1).

NOE time courses were also generated at pD* 5.3 and 283 K, where the mispair is largely A(anti)·G(syn). Irradiation of A4 H8 gave NOEs to the H1′, H2′, H2″ and H3′ of A4 and weak NOEs to G3 sugar protons. The conformation of A4 under these conditions is found to be C2′(endo) with a glycosidic torsion angle of −101°, typical for nucleotides in the B-DNA family of conformations. However, the relatively weak NOEs to the preceding base suggests that there is a non-standard B-like helical step due, for example, to a significant slide or shift to accommodate its partner G9 in the *syn* conformation. This is also consistent with the ³¹P-n.m.r. spectrum, which indicates a larger spread of backbone torsion angles than either in the native dodecamer or in the mismatched dodecamer at high pH (see above).

Fig. 8. Kinetics of the $A4(anti) \cdot G9(anti) \rightleftharpoons A4(anti) \cdot G9(syn)$ transition in $d(CGAAATTGGCG)_2$.

Rates were determined by saturation transfer and from the temperature-dependence of the linewidth of A4 H8, as described in the text. (a) Saturation transfer time course at pD* 5.3 and 283 K: ■, $anti \rightarrow syn$; □, $syn \rightarrow anti$. Lines are best-fit non-linear-regression curves. (b) Eyring plot for the $syn \rightarrow anti$ interconversion process. The activation energy is approx. $102 \text{ kJ} \cdot \text{mol}^{-1}$, and the correlation coefficient is 0.95 ($n = 5$).

Thermodynamics and kinetics of the $syn-anti$ equilibrium

We have recorded ^1H -n.m.r. spectra of the mispaired duplex at several pH values and temperatures. At pD* 5.1, the intensity of the H8 resonance of A4 (at 8.44 p.p.m.) is at least 90% that for the other base proton resonances. At 283 K the linewidth of the resonance is essentially equal to that of the other base protons, indicating that at pD* 5.1 and 283 K the mismatch is predominantly $A(anti) \cdot G(syn)$. As the pH is raised the intensity of the resonance at 8.45 p.p.m. decreases, and a new resonance at 8.07 p.p.m. appears. Further, the linewidth of the resonance at 8.45 p.p.m. reaches a maximum near pD* 6, and its relative intensity approaches 0.5 in the 5.7–6.1 pD* range. At pD* 7.45, the intensity of the peak at 8.45 p.p.m. is negligible, whereas the intensity of the resonance at 8.07 p.p.m. is essentially equal to that of one proton. Hence we conclude that the $A(anti) \cdot G(syn)$ state is stabilized by protons, and the prototropic transition occurs with an apparent pK_a value of approx. 5.9.

From the intensity of the NOESY cross-peaks connecting the resonances at 8.45 and 8.07 p.p.m. in the spectra recorded at pD* 5.7 and 283 K it is possible to estimate the rate constant for the exchange process to be about 5 s^{-1} . In addition, we have used the one-dimensional saturation transfer technique to measure the rate constants for the forward and reverse reactions at pD* 6.1 and 283 K. The time-dependence of the magnetization transfer is shown in Fig. 8(a), with the corresponding least-squares fit to the data. The rate constants under these conditions were determined to be 4.5 ± 0.3 and $5.4 \pm 0.2 \text{ s}^{-1}$ for the $syn \rightarrow anti$ (k_{+1}) and $anti \rightarrow syn$ (k_{-1}) transitions respectively. The ratio of the rate constants ($k_{+1}/k_{-1} = 0.83$) is very similar to that estimated from the population of the syn state (see above). The rate

Table 5. Equilibrium between G9 (syn) and G9 ($anti$)

The $anti/syn$ population ratio was estimated from the relative areas of the H8 of A4 in the syn state to that of the non-exchanging H8 resonances (cf. Fig. 2a). Rate constants (k_{+1} , k_{-1}) were determined from NOESY cross-peak intensities, linewidths (LW) and by saturation transfer (ST).

pH	Temperature (K)	$anti/syn$ ratio	k_{+1} (s^{-1})	k_{-1} (s^{-1})	Method used
5.6	283	0.6	3	5	NOESY
5.7	283	0.7	9	(13)	LW
	298	0.9	105	(117)	LW
6.1	283	1.6	8.4	(5.3)	LW
	298	1.6	122	(76)	LW
	283	1.2	5.4	4.5	ST

constants are also in reasonable agreement with those estimated from the NOESY spectra at pD* 5.6.

We have recorded one-dimensional n.m.r. spectra at pD* 5.7 as a function of temperature. The population of the $A(anti) \cdot G(syn)$ state was determined by the relative intensity of the resonance at 8.45 p.p.m., and was found to depend only weakly on the temperature. As the determination of the equilibrium constant is relatively imprecise, and the temperature range studied is quite small (25 K), we cannot determine the van't Hoff enthalpy accurately, but only place limits of $18 \pm 6 \text{ kJ} \cdot \text{mol}^{-1}$. To supplement the direct measurement of the rate constant, we have determined the temperature-dependence of the linewidths of resolved resonances of base protons. Whereas the linewidths of the resonances of G2 H8, G12 H8 and A6 H8 decrease monotonically with increasing temperature, with an apparent activation energy of $19 \text{ kJ} \cdot \text{mol}^{-1}$, the linewidth of A4 H8 increases dramatically with increasing temperature, from 17.5 Hz at 278 K to 61.5 Hz at 303 K, indicating a high activation energy for the transition. Because separate resonances are observed for the $G(syn)$ and $G(anti)$ states, and the chemical shifts are essentially independent of temperature, the exchange must be slow on the chemical-shift timescale ($\Delta\delta$ 0.38 p.p.m. or 190 Hz at 11.75 T). Further, the rate constant for the $A(anti) \cdot G(syn) \rightarrow A(anti) \cdot G(anti)$ transition is only 4.5 s^{-1} at 283 K, which is equivalent to an exchange broadening contribution of only 1.4 Hz in the slow exchange limit; this contribution will be even smaller at 278 K. Hence the rate constant $k(T)$ for exchange in the $syn \rightarrow anti$ direction can be calculated from the linewidths by using the equations for slow exchange (Sandström, 1982):

$$k(T) = \pi[LW(T) - LW_0(T)] \quad (2)$$

where $LW(T)$ and $LW_0(T)$ are the linewidths at temperature T in the presence and in the absence of exchange respectively. The value of $LW_0(T)$ was calculated assuming that the value of the linewidth at 278 K contains an insignificant exchange contribution (see above), and the viscosity law, in which:

$$LW_0(T) = LW_0(T_0) \cdot \exp[(E/R)(1/T - 1/T_0)] \quad (3)$$

where E is the activation energy for the viscosity of $^2\text{H}_2\text{O}$ ($20 \text{ kJ} \cdot \text{mol}^{-1}$; Wilbur *et al.*, 1976). Fig. 8(b) shows an Eyring plot for the derived rate constants, from which an activation enthalpy of $102 \pm 5 \text{ kJ} \cdot \text{mol}^{-1}$ was determined; the kinetic parameters are shown in Table 5.

As predicted from thermal denaturation studies (Brown *et al.*,

1990; Leonard *et al.*, 1990) the A·G mismatch in this dodecamer undergoes a conformational transition from A(*anti*)·G(*syn*) at low pH to A(*anti*)·G(*anti*) at high pH, with a mid-point for the transition near pH 6. The effects of the transition are sensed only by the nearest neighbouring bases. Interestingly, the equilibrium constant for this transition seems to be affected mainly by pH; the small van't Hoff enthalpy is consistent with similar stacking energies for the two conformations.

In contrast, the activation energy for the transition is quite large, as might be expected for such a large conformational transition (i.e. a 180° rotation of the base about the glycosidic bond). Although the energy barrier to rotation about the glycosidic bond in free nucleotides is very low (Saenger, 1984), neighbouring base-pairs in the DNA duplex must impose severe restrictions on the rotation. Either the neighbouring base-pairs have to move apart to permit the rotation to occur while the base is stacked, or the base-pairing must break and one or both bases swing out into solution where free rotation about the glycosidic bond becomes possible. The former mechanism is analogous to rotation of aromatic rings about the C_(β)–C_(γ) bond in proteins, which occurs on the 1 μs–1 ms timescale (Wagner, 1983) and with an activation energy of the order of 200 kJ·mol⁻¹. It is possible that the activation energy for large-scale 'breathing' in DNA would be smaller than in a protein. Nevertheless the rate constant for the transition would be of the order of 200 s⁻¹ at physiological temperatures. We note that the timescale of the rotation is similar to that for exchange of imino protons from duplex DNA, which is also a single-base-pair event (Kochoyan *et al.*, 1987). However, the activation energy for base-pair opening is in the range 30–60 kJ·mol⁻¹, which is significantly smaller than determined for the *syn* → *anti* transition. This may reflect the greater re-arrangement required to effect the rotation about the glycosidic bond.

Our results confirm and complement the findings reported by Gao & Patel (1988), who previously found that the A(*anti*)·G(*syn*) state is stabilized by protonation, and that there are only minor perturbations beyond the mismatch site itself. We also conclude that the mismatched bases are stacked inside the helix in both conformations. However, we find that the ³¹P chemical-shift dispersion of the A(*anti*)·G(*syn*) state is greater than that of the A(*anti*)·G(*anti*) state, which is opposite to the observation made by Gao & Patel (1988). Further, the NOESY intensities for the low-pH conformation in our sequence imply a significant distortion of the helix at the mismatch sites. Unfortunately, it was not possible to quantify enough NOEs with sufficient accuracy to analyse the conformation in greater detail. However, this conclusion agrees with the recent refinement of the low-pH crystal structure of the mismatched dodecamer (Leonard *et al.*, 1990), which shows some disorder of the sugar-phosphate backbone.

The possible equilibria for A·G mismatches seem to be rather sensitive to both pH and base sequence. Thus the A(*anti*)·G(*anti*) and A(*anti*)·G(*syn*) mismatches have been observed in solution in both d(CGGAATTCACG)₂ (Gao & Patel, 1988) and d(CGCAAATTGGCG)₂ (the present paper), though with apparently different local conformations at the mismatch site for the two sequences, and the A(*syn*)·G(*anti*) conformation has been seen in the crystal state in d(CGCGAATTAGCG)₂ (Brown *et al.*, 1986). The various conformations manifest in the crystal state have been rationalized in terms of flanking sequence (Webster *et al.*, 1990).

Finally, we estimate that, for the sequence examined, approx. 3% of the molecules would be present in the A(*anti*)·G(*syn*) state at physiological pH (approx. 7.5). As this structure is significantly different from that of mean DNA [compared with the A(*anti*)·G(*anti*) state], it may be the one that is recognized by

Table 6. Additional X-PLOR parameters used for N1-protonated adenosine

Atom-types and parameter values correspond to those used within the CHARMM (Brooks *et al.*, 1983) and X-PLOR (Brünger, 1988) all-atom force-fields. The proton attached to N1 of adenine is designated H1.

Atom	Atom-type
H1	H (H1 = donor; N1–H1...X)
N1	NC (N1 ≠ acceptor)
C2	CE
C6	CA
k_{bond} (H–NC) = 197.5 MJ·mol ⁻¹ ·nm ⁻² ; r_0 (H–NC) = 0.101 nm	
k_{angle} (CA–NC–H) = 167.4 kJ·mol ⁻¹ ·degree ⁻¹ ; θ_0 (CA–NC–H) = 120.7°	
k_{angle} (CE–NC–H) = 167.4 kJ·mol ⁻¹ ·degree ⁻¹ ; θ_0 (CD–NC–H) = 120.7°	
k_{improper} (H–X–X–NC) = 167.4 kJ·mol ⁻¹ ; $n = 0$; $\delta = 0^\circ$	

the repair enzymes, and this modest population may partly account for the inefficient or diminished enzymic repair observed for this purine·purine mispairing.

Molecular modelling and energy minimization

Initial models for the four possible *anti*·*anti*, *anti*·*syn*, *syn*·*anti* and *syn*·*syn* A·G and AH⁺·G (where AH⁺ represents the N1-protonated adenosine) base-paired dodecamer duplexes were generated as described in the Materials and methods section. The global B-like DNA conformation was selected on the basis of the derived nucleotide parameters (see above). The two A·G mismatched base-pairs present in each of the duplexes were treated as equivalent for modelling purposes, and no attempt was made to model the possible minor tautomeric forms of either purine involved in the A·G mispair. The resulting eight structural models were each subjected to restrained molecular-dynamics refinement by using the X-PLOR program (Brünger, 1988) and the recommended three-stage procedure (Brünger, 1990). This protocol includes simulated isothermal dynamic annealing (equilibration) at 300 K and final conjugate-gradient optimization to relieve strain and/or bad atomic contacts. The restrained molecular-dynamics X-PLOR strategy employed has been examined (Gronenborn & Clore, 1989) and has recently been used by us to establish the structure of a berenil–EcoRI dodecamer complex in aqueous solution (Lane *et al.*, 1991).

Suitable CHARMM parameterization for the N1-protonated adenosine base was derived by interpolation from available nucleotide data; the additional parameter values and atom-types used are collected in Table 6. Molecular electrostatic potentials were computed from the MNDO wavefunctions by using MEP/MEPFIT as described in the Materials and methods section for (i) each of the four regular nucleotides, (ii) N1-protonated adenosine (AH⁺, residue ADH) and (iii) both the 5'-hydroxylated and the 3'-hydroxylated bases (C5T and G3T respectively) that form the strand termini. The superiority of molecular electrostatic potentials rather than quantum-mechanically derived partial charges in molecular mechanics or dynamics applications has been described (Singh & Kollman, 1984; Orozco & Luque, 1990a,b). The final values for molecular electrostatic potentials shown in Table 7 were rationalized for each of the phosphate residues and charges were averaged for nominally equivalent atoms.

Glycosidic torsion-angle constraints (i.e. $\chi \pm 5^\circ$) for bases not involved in forming the A·G mispair were taken from Table 4 and maintained throughout the refinement procedure. Such

Table 7. Calculated molecular electrostatic potentials for the isolated nucleotides

Molecular electrostatic potentials were calculated for 3',5'-di(CH₃OPO₂)- or 5'(3')-hydroxy-3'(5')-(CH₃OPO₂)-substituted nucleosides from their MNDO wavefunctions by using MEP (Orozco & Luque, 1990a,b) and an adapted version of MEPFIT. Unit scaling was employed. Potentials were rationalized for the phosphate residues alone, with all slack (typically $\leq \pm 0.03$ e) taken equally on O3' and O5' atoms. Molecular-electrostatic-potential (MEP) charges were averaged for nominally equivalent atoms (i.e. O1P and O2P, H21 and H22, H41 and H42, H51-53, H61 and H62). Combined sum charges of C5T (5'-end) and G3T (3'-end) residues = -1.000; ADH is N1-protonated adenosine.

Atom	CYT	C5T	THY	ADE	ADH	GUA	G3T
HO	-	0.326	-	-	-	-	0.326
P	1.161	-	1.161	1.161	1.161	1.161	1.161
O1P	-0.726	-	-0.726	-0.726	-0.726	-0.726	-0.726
O2P	-0.726	-	-0.726	-0.726	-0.726	-0.726	-0.726
C1'	0.276	0.251	0.270	0.529	0.490	0.520	0.318
C2'	-0.278	-0.405	-0.281	-0.351	-0.346	-0.341	-0.445
C3'	0.001	-0.067	-0.030	-0.003	0.017	-0.009	0.065
C4'	0.103	-0.041	0.086	0.066	0.073	0.073	-0.082
C5'	-0.018	-0.024	-0.056	0.007	-0.008	0.004	-0.033
O3'	-0.403	-0.326	-0.405	-0.400	-0.401	-0.402	-0.438
O4'	-0.327	-0.296	-0.317	-0.353	-0.351	-0.354	-0.245
O5'	-0.435	-0.429	-0.423	-0.440	-0.422	-0.444	-0.462
H1'	0.082	0.123	0.075	0.038	0.050	0.040	0.093
H2'	0.092	0.152	0.086	0.131	0.133	0.130	0.182
H2''	0.121	0.182	0.122	0.122	0.127	0.119	0.187
H3'	0.136	0.165	0.136	0.145	0.155	0.146	0.145
H4'	0.051	0.119	0.060	0.057	0.067	0.053	0.083
H5'	0.074	0.084	0.088	0.073	0.091	0.073	0.083
H5''	0.061	0.088	0.073	0.060	0.063	0.061	0.102
N1	-0.579	-0.536	-0.448	-0.689	-0.616	-0.652	-0.648
C2	0.842	0.826	0.696	0.404	0.327	0.708	0.732
N2	-	-	-	-	-	-0.891	-0.921
O2	-0.485	-0.467	-0.456	-	-	-	-
N3	-0.834	-0.808	-0.617	-0.587	-0.422	-0.594	-0.595
C4	0.937	0.927	0.541	0.470	0.485	0.387	0.331
N4	-0.970	-0.968	-	-	-	-	-
O4	-	-	-0.417	-	-	-	-
C5	-0.909	-0.813	-0.372	-0.270	-0.231	-0.213	-0.196
C5A	-	-	-0.336	-	-	-	-
C6	0.553	0.453	0.321	0.688	0.693	0.536	0.535
N6	-	-	-	-0.893	-0.843	-	-
O6	-	-	-	-	-	-0.385	-0.365
N7	-	-	-	-0.452	-0.430	-0.391	-0.366
C8	-	-	-	0.340	0.421	0.263	0.208
N9	-	-	-	-0.490	-0.459	-0.436	-0.323
H1	-	-	-	-	0.414	0.312	0.318
H2	-	-	-	0.106	0.152	-	-
H21	-	-	-	-	-	0.410	0.428
H22	-	-	-	-	-	0.410	0.428
H3	-	-	0.309	-	-	-	-
H41	0.428	0.426	-	-	-	-	-
H42	0.428	0.426	-	-	-	-	-
H5	0.230	0.225	-	-	-	-	-
H51	-	-	0.115	-	-	-	-
H52	-	-	0.115	-	-	-	-
H53	-	-	0.115	-	-	-	-
H6	0.114	0.079	0.181	-	-	-	-
H61	-	-	-	0.418	0.444	-	-
H62	-	-	-	0.418	0.444	-	-
H8	-	-	-	0.147	0.174	0.158	0.174
EMEP	-1.000	-0.328	-1.000	-1.000	0.000	-1.000	-0.672

constraints were applied equally to both DNA strands, assuming direct equivalence for corresponding nucleotides. Further distance or NOE-derived restraints were not used, and no attempt was made to constrain base-pairs, particularly those located at the strand termini. Thus a total of 12 glycosidic dihedral constraints was used to generate the ultimate refined structure in each case, for which root-mean-square gradients of $\leq 4 \text{ kJ} \cdot \text{mol}^{-1} \cdot \text{nm}^{-1}$ (i.e. $\leq 1.0 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-1}$) were achieved.

Table 8 shows the energies and component energy terms

calculated for each of the eight model dodecanucleotides. In the case of the A·G-mispaired duplexes, the A(*anti*)·G(*anti*) is clearly of lower energy than the remaining conformations, largely as a consequence of superior hydrogen-bonded and electrostatic non-bonded terms. It is noteworthy that the A(*syn*)·G(*anti*) base-pair conformation, reported for the X-ray crystal structures of several other A·G-mismatched oligonucleotides (Brown *et al.*, 1986; Hunter *et al.*, 1986; Webster *et al.*, 1990), is less favoured by approx. $22 \text{ kJ} \cdot \text{mol}^{-1}$ for the double mismatch relative to the

Table 8. Calculated energies for A·G- and AH⁺·G-mispaired d(CGCAAATTGGCG)₂ duplexes

Structures for the possible base-pair conformations involving *anti*·*anti*, *anti*·*syn*, *syn*·*anti* and *syn*·*syn* duplexes were refined by using X-PLOR as indicated in the Materials and methods section. The calculated energies are given for the duplexes containing two symmetrically disposed A·G or AH⁺·G base-pairs.

Mispair	E_{vdw} (kJ·mol ⁻¹)	$E_{\text{elec.}}$ (kJ·mol ⁻¹)	$E_{\text{H-bonds}}$ (kJ·mol ⁻¹)	E_{total} (kJ·mol ⁻¹)
A(<i>anti</i>)·G(<i>anti</i>)	-1696	-692	-560	-700
A(<i>anti</i>)·G(<i>syn</i>)	-1703	-685	-501	-638
A(<i>syn</i>)·G(<i>anti</i>)	-1706	-716	-541	-678
A(<i>syn</i>)·G(<i>syn</i>)	-1704	-694	-491	-625
AH ⁺ (<i>anti</i>)·G(<i>anti</i>)	-1632	-824	-506	-718
AH ⁺ (<i>anti</i>)·G(<i>syn</i>)	-1707	-828	-525	-828
AH ⁺ (<i>syn</i>)·G(<i>anti</i>)	-1693	-867	-548	-816
AH ⁺ (<i>syn</i>)·G(<i>syn</i>)	-1752	-813	-479	-799

Fig. 9. Views of the d(CGCAAATTGGCG)₂ duplexes after refinement by using X-PLOR as described in the text

(a) Neutral dodecamer containing two A·G-mispaired sites; (b) protonated dodecamer containing two N1-protonated AH⁺·G-mispaired sites. All hydrogen atoms have been omitted for clarity.

lower-energy A(*anti*)·G(*anti*) configuration. This result contrasts with earlier molecular-mechanics studies, which have suggested (Chuprina & Poltev, 1983; Keepers *et al.*, 1984; Rao *et al.*, 1986) that the enthalpy difference between the different conformers may be only about 5 kJ·mol⁻¹.

In the case of the purine·purine G·A mispairs containing N1-protonated adenosine residues at the mispair sites, the AH⁺(*anti*)·G(*syn*) conformation is calculated to be of lowest energy, primarily as a result of favourable hydrogen-bonded and hydrophobic non-bonded contacts. The structure of next lowest energy, AH⁺(*syn*)·G(*anti*), differs by only about 12 kJ·mol⁻¹. Interestingly, the base-pairing with AH⁺(*anti*)·G(*anti*) represents the least-favoured conformation.

The influence of pH upon the conformation and thermodynamic stability of mismatch base-pairs has been examined by

pH-dependent spectrophotometric studies of thermal denaturation (Brown *et al.*, 1990; Leonard *et al.*, 1990). In particular, it has been shown that the A·G mismatch displays conformational flexibility, with a protonated A(*anti*)·G(*syn*) base-pair occurring in certain base-stacking environments. The computed energies in Table 8 correlate well with both this study and predictions based on recent X-ray-crystallographic and n.m.r. solution data (Gao & Patel, 1988; Brown *et al.*, 1989, 1990; Leonard *et al.*, 1990). It has recently been suggested that A·G mispairs are also likely to be sensitive to the effects of flanking sequence (Leonard *et al.*, 1990; Webster *et al.*, 1990).

The lowest-energy A(*anti*)·G(*anti*) structure determined for the neutral or unprotonated mispair is shown in Fig. 9(a). Helix analysis of the duplex shows an average helical rise of 0.31±0.07 nm and a mean helical twist of 41±11° (8.8 base-

Fig. 10. Stereoviews of the mispaired A4·G21 (top) and A16·G9 (bottom) base-pairs in the refined dodecamer duplexes

(a) A·G base-pairs from the neutral dodecamer; (b) $\text{AH}^+\cdot\text{G}$ base-pairs from the protonated dodecamer containing N1-protonated adenosine residues at both mismatched sites.

pairs/turn), compared with 0.35 ± 0.05 nm and $37 \pm 5^\circ$ (9.7 base-pairs/turn) respectively for the native Dickerson dodecamer (Lane *et al.*, 1991). The duplex clearly adopts a general B-like DNA conformation, but there is considerable induced perturbation, manifest as a general disruption to the parallel base stacking and increased base roll and propeller twist. Further, this helical perturbation is propagated to the strand termini, where effective Watson–Crick base-pairing is decreased.

Examination of the A(*anti*)·G(*anti*) mispair site reveals that the mean planes of A4 and A16 are significantly tilted to facilitate the formation of a network of bifurcated hydrogen bonds involving the exocyclic NH_2 groups of the adenine residues and the N3 atoms of the 5'-side cytosine residues on the same strand. Thus hydrogen bonds are formed in the major groove between H62(A4/A16) and N3(C3/C15) with an $\text{H}\cdots\text{N}$ separation of 0.186 nm and a subtended $\text{N6-H62}\cdots\text{N3}$ angle of 164° . The A·G base-pairs themselves (Fig. 10a) form strong inter-strand hydrogen bonds with $\text{N1(A)}\cdots\text{H1(G)}$ and $\text{H61(A)}\cdots\text{O6(G)}$ separations of 0.188 nm and 0.184 nm respectively and subtended angles of 160° and 165° for $\text{N1}\cdots\text{H1-N1}$ and $\text{N6-H61}\cdots\text{O6}$ respectively. Moreover, there is evidence that both purine N1 atoms may be involved in distinct weak three-centre hydrogen bonds to the exocyclic NH_2 groups on the comple-

mentary base. Thus N1(A4/A16) also forms a hydrogen bond to the NH_2 group of G21/G9, at an $\text{N1}\cdots\text{H22}$ separation of 0.230 nm ($\text{N1}\cdots\text{H22-N2}$ angle 142°), whereas N1(G21/G9) forms an additional hydrogen bond to the NH_2 group of A4/A16 at an $\text{N1}\cdots\text{H61}$ separation of 0.218 nm ($\text{N1}\cdots\text{H61-N6}$ angle 133°). It is clear that this network of hydrogen bonds serves to stabilize the DNA helix.

The lowest-energy $\text{AH}^+(\text{anti})\cdot\text{G}(\text{syn})$ structure determined for the protonated mispair is shown in Fig. 9(b). Helix analysis of the duplex shows an average helical rise of 0.31 ± 0.07 nm and a mean helical twist of $40 \pm 8^\circ$ (9.0 base-pairs/turn), which compare favourably with the values determined for the unprotonated mismatch and the native dodecamer. Further, these values are consistent with those reported (0.33 nm and 36° respectively; Leonard *et al.*, 1990) for the X-ray crystal structure of d(CGAAATTGCGC)_2 . The structure of the protonated duplex is clearly of the B-DNA family (Fig. 9b), but there is little evidence of induced conformational changes or propagation of distortion relative to the native oligomer. Indeed, the structure retains parallel stacking of the base-pairs and conformational effects appear to be largely confined to the A·G mispair sites. In contrast with the reported X-ray structure (Leonard *et al.*, 1990), there appears to be no evidence of close contacts between either

N6(A) or O6(G) and either O4(T) or N4(C) on the 3'-side of the A·G mismatch. Thus we are not able to confirm the suggestion that weak bifurcated hydrogen bonds are formed in the major groove of the protonated dodecamer.

Examination of the AH⁺·G base-pairing (Fig. 10b) reveals that two strong inter-strand hydrogen bonds are formed with H1(A4/A16)···N7(G21/G9) and H61(A4/A16)···O6(G21/G9) separations of 0.189 and 0.175 nm respectively, together with subtended N1–H1···N7 and N6–H61···O6 angles of 165° and 176° respectively. This excellent hydrogen-bonded geometry clearly contributes to the favourable A(*anti*)·G(*syn*) conformation adopted by the protonated mismatch.

Note added in proof (received 22 July 1991)

While this paper was in proof an article appeared on the properties of the A·G mismatch in d(GCCACAAGCTC)·d(GAGCTGGTGGC) (Carbonnaux *et al.*, 1991). Our conclusions regarding DNA conformation and protonation of the adenine residues are essentially identical with theirs.

This work was supported by the Medical Research Council, the Science and Engineering Research Council of the U.K. and the Cancer Research Campaign. We thank Dr. K. Borden (National Institute for Medical Research) for helpful comments and Professor S. Neidle and Dr. C. A. Laughton (Institute of Cancer Research) for useful discussions. We are grateful to Dr. M. Orozco and Dr. F. J. Luque (University of Barcelona) for supplying the MEP program used in this work.

REFERENCES

- Arnott, S., Chandrasekaran, R., Hall, I. & Puigjaner, L. C. (1983) *Nucleic Acids Res.* **11**, 4141–4155
- Bax, A. & Davis, D. G. (1985) *J. Magn. Reson.* **65**, 355–360
- Birchall, A. J. & Lane, A. N. (1990) *Eur. Biophys. J.* **19**, 73–78
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comput. Chem.* **4**, 187–217
- Brown, T., Hunter, W. N., Kneale, G. & Kennard, O. (1986) *Proc. Natl. Acad. Sci. U.S.A.* **83**, 2402–2406
- Brown, T., Leonard, G. A., Booth, E. D. & Chambers, J. (1989) *J. Mol. Biol.* **207**, 455–457
- Brown, T., Leonard, G. A., Booth, E. D. & Kneale, G. (1990) *J. Mol. Biol.* **212**, 437–440
- Brünger, A. T. (1988) X-PLOR Manual, version 1.5, Yale University, Newhaven
- Brünger, A. T. (1990) in *Molecular Dynamics: Applications in Molecular Biology* (Goodfellow, J. M., ed.), pp. 137–178, Macmillan Press, London
- Carbonnaux, C., van der Marel, G. A., van Boom, J. H., Guschlbauer, W. & Fazakerly, G. V. (1991) *Biochemistry* **30**, 5449–5458
- Chuprina, V. P. & Poltev, V. I. (1983) *Nucleic Acids Res.* **11**, 5205–5223
- Fersht, A. R., Knill-Jones, J. W. & Tsui, W. C. (1982) *J. Mol. Biol.* **156**, 37–51
- Frenkiel, T. A., Bauer, C. J., Carr, M. D., Birdsall, B. B. & Feeney, J. (1990) *J. Magn. Reson.* **90**, 420–425
- Gao, X. & Patel, D. J. (1988) *J. Am. Chem. Soc.* **110**, 5178–5182
- Gorenstein, D. G., Schroeder, S. A., Fu, J. M., Metz, J. T., Roongta, V. & Jones, C. R. (1988) *Biochemistry* **27**, 7223–7237
- Gronenborn, A. M. & Clore, G. M. (1989) *Biochemistry* **28**, 5978–5984
- Hare, D. R., Wemmer, D. E., Chou, S.-H., Drobny, G. & Reid, B. R. (1983) *J. Mol. Biol.* **171**, 319–336
- Hore, P. J. (1983) *J. Magn. Reson.* **55**, 283–300
- Hunter, W. N., Brown, T. & Kennard, O. (1986) *J. Biomol. Struct. Dyn.* **4**, 173–191
- Keepers, J. W., Schmidt, P., James, T. L. & Kollman, P. A. (1984) *Biopolymers* **23**, 2901–2929
- Kochoyan, M., Leroy, J. L. & Guéron, M. (1987) *J. Mol. Biol.* **196**, 599–609
- Lane, A. N. (1990) *Biochim. Biophys. Acta* **1049**, 189–204
- Lane, A. N. (1991) *Biochem. J.* **273**, 383–391
- Lane, A. N., Lefèvre, J.-F. & Jardetzky, O. (1986) *J. Magn. Reson.* **66**, 201–218
- Lane, A. N., Jenkins, T. C., Brown, T. & Neidle, S. (1991) *Biochemistry* **30**, 1373–1385
- Lefèvre, J.-F., Lane, A. N. & Jardetzky, O. (1987) *Biochemistry* **26**, 5076–5089
- Leonard, G. A., Booth, E. D. & Brown, T. (1990) *Nucleic Acids Res.* **18**, 5617–5623
- Marion, D. & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun.* **124**, 774–783
- Nerdal, W., Hare, D. R. & Reid, B. R. (1988) *J. Mol. Biol.* **201**, 717–739
- Orozco, M. & Luque, F. J. (1990a) *J. Comput. Chem.* **11**, 909–923
- Orozco, M. & Luque, F. J. (1990b) *J. Comput.-Aided Mol. Design* **4**, 411–426
- Orozco, M., Laughton, C. A., Herzyk, P. & Neidle, S. (1990) *J. Biomol. Struct. Dyn.* **8**, 359–373
- Ott, J. & Eckstein, F. (1985) *Biochemistry* **24**, 2530–2535
- Rajagopal, P., Gilbert, D. E., van der Marel, G. A., Boom, J. H. & Feigon, J. (1988) *J. Magn. Reson.* **78**, 526–537
- Rao, S. N., Singh, U. C. & Kollman, P. A. (1986) *Isr. J. Chem.* **27**, 189–197
- Saenger, W. (1984) *Principles of Nucleic Acid Structure*, chapter 4, Springer-Verlag, Berlin
- Sandström, J. (1982) *Dynamic NMR Spectroscopy*, chapter 2, Academic Press, New York
- Singh, U. C. & Kollman, P. A. (1984) *J. Comput. Chem.* **5**, 129–145
- Tidor, B., Irikura, K., Brooks, B. R. & Karplus, M. (1983) *J. Biomol. Struct. Dyn.* **1**, 231–252
- Wagner, G. (1983) *Q. Rev. Biophys.* **16**, 1–57
- Wagner, G. & Wüthrich, K. (1979) *J. Magn. Reson.* **33**, 675–680
- Webster, G. D., Sanderson, M. R., Skelly, J. V., Neidle, S., Swann, P., Li, B. F. & Tickle, I. J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 6693–6697
- Wilbur, D. W., DeFries, T. & Jonas, J. (1976) *J. Chem. Phys.* **65**, 1783–1788

Received 29 January 1991/29 April 1991; accepted 7 May 1991