The EMBO Journal vol.5 no.13 pp.3697–3703, 1986

Structures of mismatched base pairs in DNA and their recognition by the *Escherichia coli* mismatch repair system

G.V.Fazakerley, E.Quignard, A.Woisard, W.Guschlbauer, G.A.van der Marel[1], J.H.van Boom[1], M.Jones[2] and M.Radman[2]

Service de Biochimie, Bât. 142, Département de Biologie, Centre d'Etudes Nucléaires de Saclay, 91191 Gif-sur-Yvette, Cédex, France, [1]Gorlaeus Laboratoria, University of Leiden, 2300A-Leiden, The Netherlands, and [2]Laboratoire de Mutagénèse, Institut J.Monod, Centre National de la Recherche Scientifique, Université Paris 7, Tour 43, 75251 Paris, Cédex 05, France

Communicated by M.Radman

The *Escherichia coli* mismatch repair system does not recognize and/or repair all mismatched base pairs with equal efficiency: whereas transition mismatches (G·T and A·C) are well repaired, the repair of some transversion mismatches (e.g. A·G or C·T) appears to depend on their position in heteroduplex DNA of phage λ. Undecamers were synthesized and annealed to form heteroduplexes with a single base-pair mismatch in the centre and with the five base pairs flanking each side corresponding to either repaired or unrepaired heteroduplexes of λ DNA. Nuclear magnetic resonance (n.m.r.) studies show that a G·A mismatch gives rise to an equilibrium between fully helical and a looped-out structure. In the unrepaired G·A mismatch duplex the latter predominates, while the helical structure is predominant in the case of repaired G·A and G·T mismatches. It appears that the *E. coli* mismatch repair enzymes recognize and repair intrahelical mismatched bases, but not the extrahelical bases in the looped-out structures.

Key words: heteroduplex/oligonucleotides/nuclear magnetic resonance/DNA loops/mutagenesis

Introduction

DNA base pair mismatches can occur *in vivo* as a consequence of (i) replication errors, (ii) heteroduplex formation in the course of genetic recombination between homologous, but not identical sequences and (iii) deamination of 5-methyl cytosine to thymine giving rise to a G·T mismatch (Radman and Wagner, 1984). In *Escherichia coli* mismatch repair depends on several genes, e.g. *mutH, mutL, mutS, mutU* and *ssb* (Radman and Wagner, 1984; Wagner *et al.*, 1984; Lu *et al.*, 1984). Replication errors are corrected by the localized excision of the newly synthesized strand and resynthesis using the parental strand as a template (Radman and Wagner, 1984; Wagner *et al.*, 1984; Lu *et al.*, 1983, 1984). Strand discrimination for the mismatch repair system in *E. coli* is provided by adenine methylation of GATC sequences (Radman *et al.*, 1980; Pukkila *et al.*, 1983). Mismatch repair occurs preferentially on unmethylated strands of hemimethylated duplexes (Wagner *et al.*, 1984; Lu *et al.*, 1983, 1984; Radman *et al.*, 1980; Pukkila *et al.*, 1983) and is greatly reduced in DNA regions in which GATC sequences are fully methylated. Newly synthesized strands are transiently undermethylated, because *dam* methylation lags somewhat behind replication (Lyons and Schendel, 1984). It is therefore believed that mismatch repair in *E. coli* occurs primarily on newly synthesized strands immediately behind the replication fork (Wagner and Meselson, 1976), and that it eliminates over 99% of replication errors (Glickman and Radman, 1980).

The repair of all classes of non-complementarities between two strands was tested in heteroduplex DNAs of phage λ transfecting *E. coli* (Wagner *et al.*, 1984). Transition mismatches (G·T and A·C) (Wagner *et al.*, 1984; Lu *et al.*, 1983; Dohet *et al.*, 1985) and single unpaired bases (Dohet *et al.*, 1986) are well repaired on either of the strands of unmethylated λ heteroduplex. Three of the transversion mismatches (G·A, C·T, C·C), however, are not well repaired (Wagner *et al.*, 1984; Dohet *et al.*, 1985; Kramer *et al.*, 1984). Repair of G·A and C·T mismatches depends on the environment of the mismatch and is proportional to the G·C content of the (minimum) 4 bp on each side of the mismatch (M.J., R.Wagner and M.R., unpublished results).

To elucidate the molecular basis of mismatch recognition by the repair system, we have studied the structure of a repaired and an unrepaired G·A mismatch, as well as the structure of a repaired G·T mismatch (at the same site). Undecamers, corresponding to the exact sequence of phage λ DNA in which the mismatch repair system was tested, were synthesized and studied by proton n.m.r. We have measured the inter-imino and imino to amino and adenosine H^2 nuclear Overhauser effects (NOE) to elucidate the conformation of the dominant species in solution. Saturation transfer studies demonstrated exchange between the major and minor species in the G·A mismatch duplexes and permitted identification of the minor species. We also report some preliminary two dimensional nuclear Overhauser effect (NOESY) spectra on one of the duplexes. We find that both repairable mismatches (G·T and G·A) form an intrahelical wobble pair, whereas the unrepaired G·A mismatch provokes a looped-out structure.

Results

Variable repair efficiencies of the G·A and G·T mismatches in bacteriophage λ DNA

It has been possible to form heteroduplex species with a single defined mismatched bp by annealing purified separated strands of DNA from wild-type bacteriophage λ with those from phages carrying a sequenced mutation (Dohet *et al.*, 1985). The analysis of the progeny of individual heteroduplex molecules, following transfection and plating, allows an estimate of the mismatch repair efficiency. High fraction of mixed (wild-type/mutant, i.e. c^+/c) infective centres reflects the absence of mismatch repair (Dohet *et al.*, 1985). Table I shows repair efficiencies for the G·A and G·T mismatches at positions 43 and 208 in the cI gene. At each position, the G·A mismatch is repaired less efficiently than the G·T mismatch. Both mismatches are repaired less efficiently in position 208 than in position 43. This particular example is typical for the general pattern of mismatch repair efficiency which appears to be proportional to the G·C content of the mismatch neighbourhood (a minimum of 4 bp on each side of the mismatch

Table I. Repair efficiencies for G·A and G·T mismatches in two different nucleotide sequence contexts in bacteriophage λ DNA heteroduplexes

Heteroduplex[a]	% Mixed (c/c^+) progeny		Approximate repair efficiency[b]
	Mut$^+$	*MutL*	
(1) 5′ ^{1}AAATTATCAAA^{11} (c) / ^{22}TTTAAGAGTTT^{12} (c^+)	83	78	0%
(2) 5′ ^{1}AAATTTTCAAA^{11} (c) / ^{22}TTTAAGAGTTT^{12} (c^+)	38	85	55%
(3) 5′ ^{1}TTGAGGACGCA^{11} (c^+) / ^{22}AACTCATGCGT^{12} (c)	38	66	42%
(4) 5′ TTGAGGACGCA (c^+) / AACTCTTGCGT (c)	5	75	93%

[a]The numbers above the mismatch correspond to the sequence position from the amino terminus of the cI gene product. (c^+) is the wild-type strand sequence, (c) is the strand sequence from λ cI mutants: (1) C·G→A·T transversion SP40, (2) C·G→T·A transition SP44, (3) G·C→T·A transversion LP206 and (4) G·C→A·T transition UV23, respectively. Heteroduplexes (1), (2) and (3) (numbering as used in text) were chemically synthesized and analyzed by n.m.r.

[b]This is approximate *mutL*-dependent repair efficiency, calculated by taking the percent mixed progeny from *mutL* transfections as 100%, see Materials and methods.

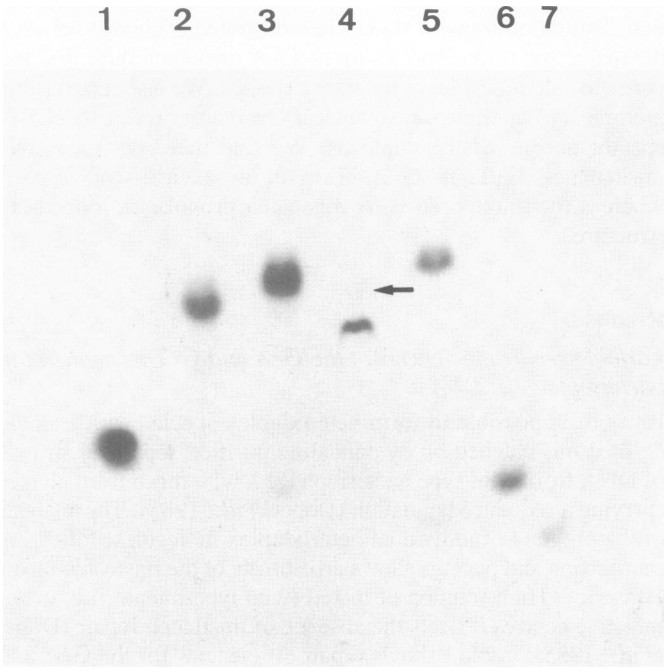


Fig. 1. Electrophoresis in 20% polyacrylamide gel. The oligonucleotides were end labelled with ^{32}P and the gel run at 10°C in 0.089 M Tris, 0.09 M boric acid, 3 mM EDTA, 15 mM $MgCl_2$, pH 8.3. **Lanes 6** and **7** are the separate strands of heteroduplex (1). **Lanes 5, 2** and **1** are 11, 10 and 8 bp oligonucleotide markers containing no mismatches. In **lane 4** is heteroduplex (1) and in **lane 3** heteroduplex (3). The arrow indicates the position of a minor species visible after longer exposure.

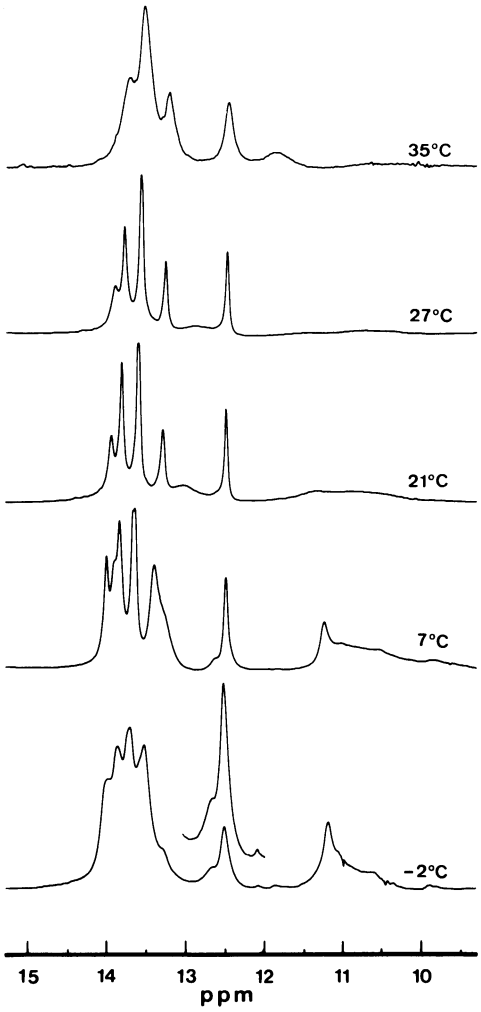


Fig. 2. Temperature dependence of the 500 MHz imino proton n.m.r. spectra of heteroduplex (1) in 10 mM phosphate, 150 mM NaCl, 0.2 mM EDTA, 9:1 H_2O-D_2O, pH 7.3.

is crucial) (M.J., R.Wagner and M.R., unpublished). These results prompted us to study the structure of the undecamer oligonucleotide heteroduplexes (1), (2) and (3) of Table I in order to get an insight into the molecular basis of mismatch recognition by the *E. coli* mismatch repair enzymes.

Polyacrylamide gel electrophoresis

Changes in helix length or major changes in conformation should influence the mobility of oligonucleotides during polyacrylamide gel electrophoresis carried out under non-denaturing conditions. Figure 1 shows the relative mobility of three undecamer duplexes. In lane 5 is the fully matched duplex which has the same sequence as heteroduplex (1) except for the central base pair which is a G·C. Lane 3 shows that the mobility of heteroduplex (3) (the G·A mismatch in the G·C rich environment) is almost identical to that of the duplex marker. We have noted small changes in mobility as a function of the G·C content for oligonucleotides of given length. By contrast heteroduplex (1) (the G·A mismatch in the A·T rich environment) shows quite different behaviour. The major band in lane 4 shows greater mobility than the decamer marker, lane 2. A minor species, more visible after longer exposure, migrates between the 10 and 11 bp markers.

Because heteroduplex (1) is not stable thermally, the electrophoresis was carried out under conditions in which it remains

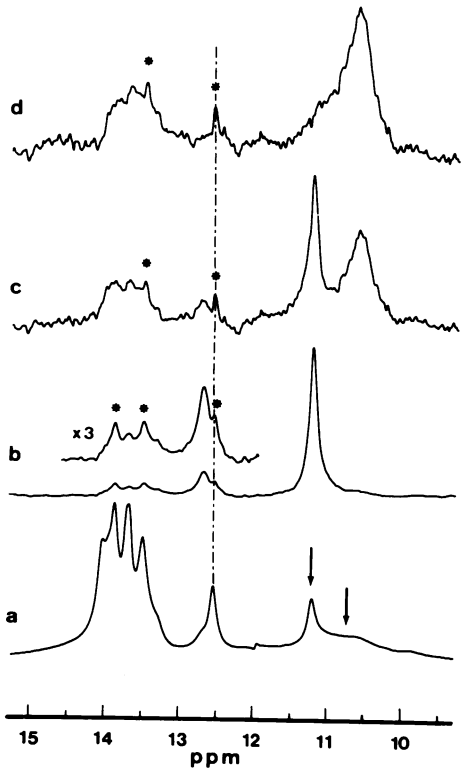


Fig. 3. (a) 500 MHz proton spectrum (imino region) of heteroduplex (1) at 2°C. (b) and (c) NOE difference spectra following 50 ms pre-irradiation of the resonance at 11.2 p.p.m. (b) and 10.7 p.p.m. (c), respectively. (d) Difference between spectra (b) and (c).

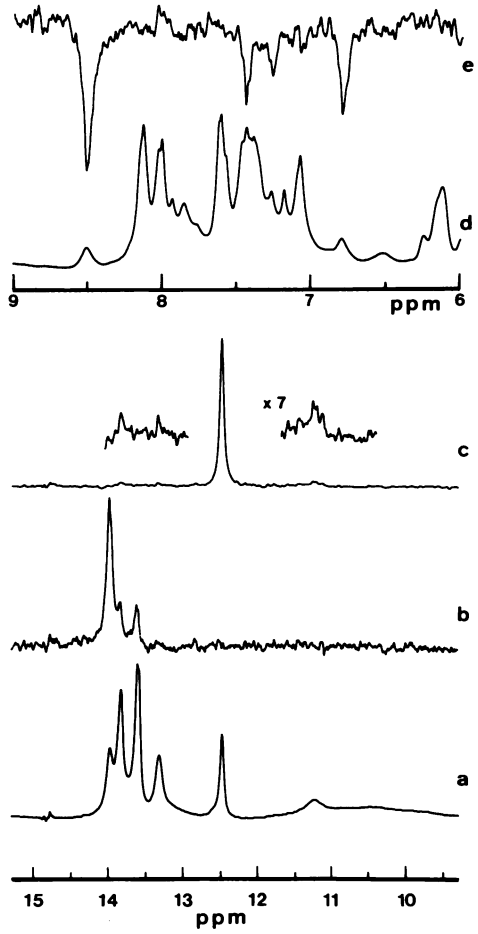


Fig. 4. 500 MHz proton spectra of heteroduplex (1) at 14°C: (a) imino region, (d) aromatic region. NOE difference spectra following 1 s pre-irradiation of the resonance at 14 p.p.m. (b), at 12.51 p.p.m. (c, imino region) and (e, aromatic region).

in a double-stranded form, as shown by the absence of bands migrating like the separate strands in lanes 6 and 7. These results indicate a major conformational change in heteroduplex (1) relative to the same sequence without a mismatch and relative to the G·A mismatch in the G·C rich sequence of heteroduplex (3).

N.m.r. study of synthetic undecamer heteroduplexes

The low field region of the proton spectra of the heteroduplex (1) (see Table I) in H_2O as a function of temperature is shown in Figure 2. The spectrum at −2°C shows some important differences compared with the many published spectra of B form oligodeoxyribonucleotides. We observe a fairly sharp resonance at 11.19 p.p.m. and one broad resonance to even higher field. In addition to the resonance observed in the G·C region of the spectrum at 12.51 p.p.m., we observe a broad resonance at 12.67 p.p.m. If the resonance at 12.51 p.p.m. corresponds to one proton, the shoulder at 12.67 p.p.m. integrates for about 0.15 proton, the resonance at 11.19 for about one proton and the whole of the region 10.4−11.6 p.p.m. for about 2−2.5 protons. On raising the temperature to 21°C the high field region of the spectrum becomes very broad and the resonance at 12.67 p.p.m. disappears. At 27°C the spectrum integrates for six protons in the A·T region and one in the G·C region. At 35°C all the remaining resonances broaden due to exchange with solvent but curiously a resonance at 11.9 p.p.m. appears. At 40°C (not shown) the imino region shows only broad humps.

Imino resonances at 10−12 p.p.m. have been observed either for a G·T wobble pair (Patel *et al.*, 1982; Tibanyenda *et al.*, 1984) or for imino protons non-hydrogen bonded in loop structures (Haasnoot *et al.*, 1980, 1983; Patel *et al.*, 1985). To establish the nature of the resonances at high field, we have carried out saturation transfer experiments at 2°C. Figure 3a shows the spectrum at this temperature and pre-irradiation for 50 ms at the positions indicated give the difference spectra (Figure 3b and c). Under these conditions NOE effects should be very small, but specific irradiation is impossible as high power had to be used to give an acceptable signal-to-noise ratio. Irradiation at 11.2 p.p.m., Figure 3b, predominantly gives an effect at 12.67 p.p.m. and these two resonances must correspond to a proton in two different environments. Irradiation at 10.7 p.p.m. predominantly gives a broad hump in the A·T region. This is more clearly seen in the difference of Figure 3b and c, weighted to eliminate the 11.2 p.p.m. resonance. The resonances marked with an asterisk arise from NOEs and are absent in spectra obtained with a 20 ms pre-irradiation time under which conditions transfer of saturation is ~60%, whereas at 50 ms it is complete. At 10°C we observed that transfer of saturation was complete with a pre-irradiation time of 20 ms.

Having established that two species are present in a ratio of ~1:6, NOE measurements at a variety of temperatures were carried out to try to establish the nature of the predominant species. Irradiation at 14°C of the lowest field resonance at 14.00 p.p.m. gives rise to two NOEs at 13.63 p.p.m. and 13.86 p.p.m., Figure 4b, showing that its two neighbours are both A·T base pairs. By contrast, pre-irradiation of the resonance at 12.51 p.p.m., Figure 4c, gives two very weak NOEs in the A·T region and another at 11.2 p.p.m. The magnitude of the NOEs is 5−8%

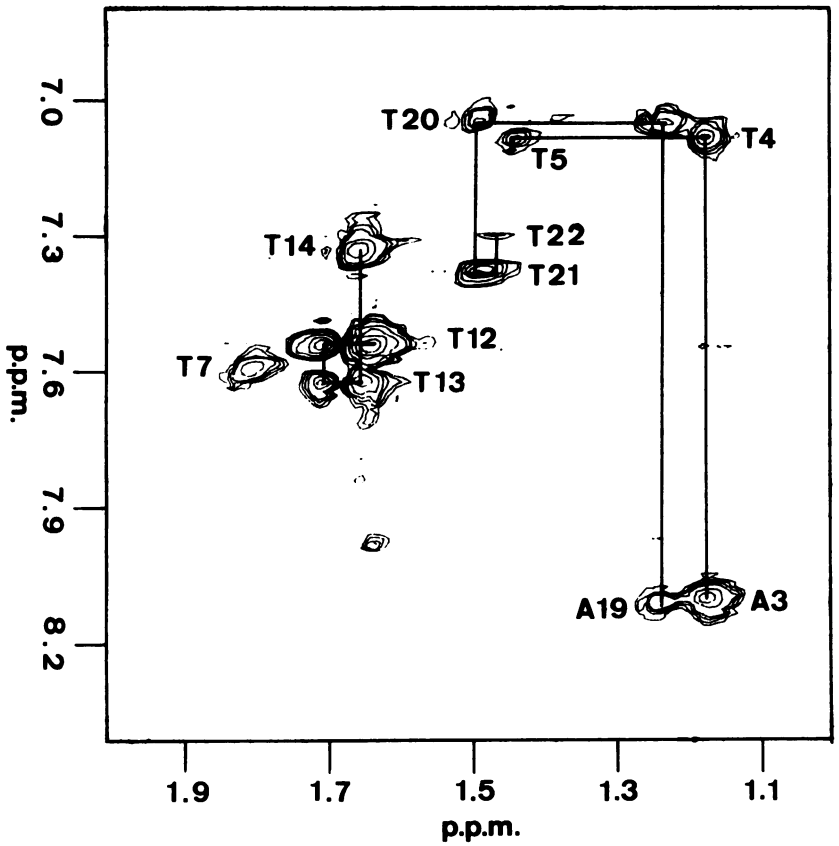


Fig. 5. NOESY spectrum of heteroduplex (1). Only the region of cross peaks between aromatic protons and methyl groups is shown. The spectrum was recorded at 12°C with a mixing time of 300 ms.

of those observed in Figure 3b. The resonance at 12.51 p.p.m. is that of a guanosine imino proton from its chemical shift and thus either a G·C or G·A base pair. Examination of the aromatic region resolves this ambiguity. Figure 4e shows the observed NOEs, the two most important ones are found at 8.50 and 6.79 p.p.m. and these two resonances are both exchangeable protons. In this temperature range no sharp resonances are found for guanosine or adenosine amino protons (Fazakerley *et al.*, 1985), but strong NOEs to each of the cytidine amino protons are observed (Fazakerley *et al.*, 1984). Thus the resonance at 12.51 p.p.m. arises from a G·C base pair. The extremely small interimino and imino to H^2 NOEs show that the adjacent bp are abnormally far away, the magnitude of the interimino NOEs is consistent with an internuclear distance of ~0.5 nm. The apparent NOE at 11.2 p.p.m. is the reverse magnetization transfer to that shown in Figure 3a.

Imino protons in loop structures are usually more accessible to the solvent than hydrogen bonded ones. Spectra of heteroduplex (1) at −2°C in the absence of phosphate buffer and at pH 5.8 (not shown) are identical to those observed in the presence of phosphate. The line widths of the resonances at high field must be determined by exchange between different species and not by exchange with solvent. This is consistent with the appearance at 35°C (Figure 2) of a resonance at 11.9 p.p.m. This probably corresponds to a proton (or protons) of the different species now in fast exchange on an n.m.r. time scale.

We have also measured 2D NOESY spectra of the three duplexes and also of the fully complementary A·T rich sequence. For the latter duplex normal base−base interactions H^8 or H^6 to TCH_3 or H^5 link $^3AH^8$ to $^8CH^5$ on one strand and $^{12}TH^6$ to

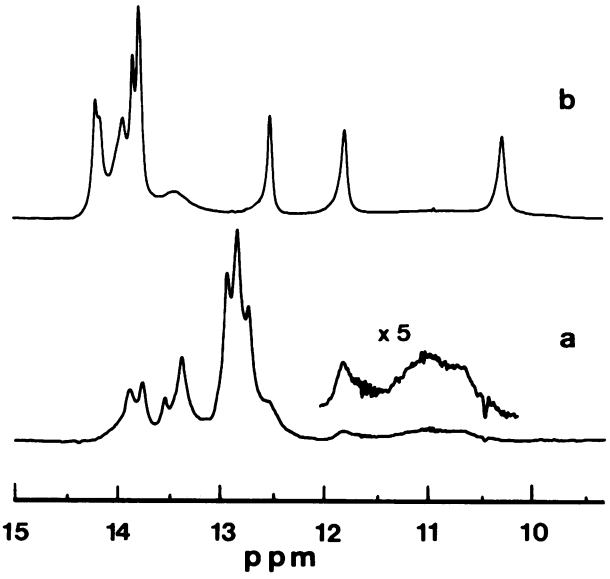


Fig. 6. 500 MHz proton spectra (imino region) of **(a)** heteroduplex (3) at −2°C and **(b)** heteroduplex (2) at 0°C.

$^{14}TCH_3$ and $^{19}AH^8$ to $^{22}TCH_3$ on the other strand (not shown). Among these a normal strong interaction $^6CH^6 \cdot {}^7TCH_3$ is observed. If ^6C is replaced by A, heteroduplex (1), we would expect a cross peak for $^6AH^8 \cdot {}^7TCH_3$ for a Watson−Crick G·A pair or a strong cross peak $^6AH^2 \cdot {}^7TCH_3$, if ^6A were *syn*, as observed in a recent crystal structure (Kennard, 1985). The region of the

NOESY spectrum corresponding to interactions between the aromatic and methyl region is shown in Figure 5. The only major difference between this and the fully complementary duplex is that $^{7}TCH_3$ is shifted 0.2 p.p.m. downfield and that there is no interaction with any ^{6}A base proton. We have examined the matrix column through the cross peak assigned $^{7}TH^{6}\cdot{}^{7}TCH_3$ and find no other interaction with $^{7}TCH_3$. Further, the row through this cross peak shows a resonance corresponding to a single proton on the diagonal eliminating the possibility of coincident cross peaks. No strong NOE in the aromatic to anomeric region is observed which would account for a *syn* conformation, as required for a Hoogsteen pair (Kennard, 1985; Topal and Fresco, 1980).

The G·A mismatch in heteroduplex (1) is poorly repaired whereas the G·A mismatch in heteroduplex (3) is well repaired (Table I). The proton spectrum of the imino region of heteroduplex (3) is shown in Figure 6a. In the region 10−12 p.p.m. resonances are observed but these correspond to a minor species (~15%, if the resonance at 11.82 p.p.m. corresponds to one proton of this species), as opposed to the spectrum of heteroduplex (1), Figure 2a. The major species give resonances in the usual A·T and G·C regions and from interpretation of interimino NOEs (not shown) is the fully helical form. The G·A imino proton is observed at 12.82 p.p.m. The identity of the minor species has not yet been established. If it is a looped-out form similar to scheme II the slipped G·C pair would probably be between ^{6}G and ^{18}C.

We have also studied the A·T rich sequence but with a G·T mismatch in place of the G·A mismatch [heteroduplex (2) in Table I]. The proton spectrum of the undecamer heteroduplex (2) is shown in Figure 6b. As expected the two imino protons of the G·T wobble pair are found in the high field region. No minor species have been observed for this DNA at any temperature.

Discussion

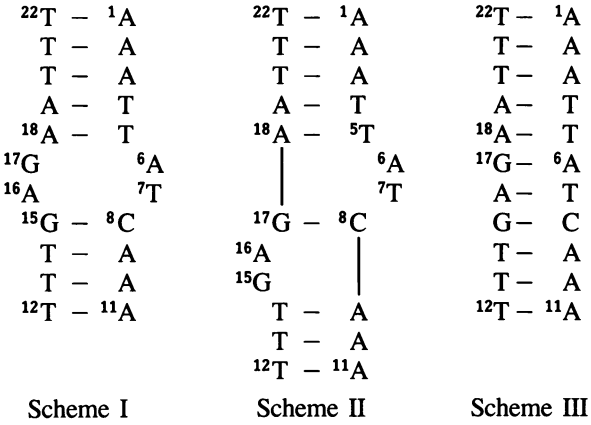
Table I shows that the mismatch repair efficiency in *E. coli* depends on the nature of the mismatch as well as on the nature of the base pairs surrounding the mismatch: the transition mismatch (G·T) is repaired more efficiently than the transversion mismatch (G·A) and both are repaired more efficiently in the G·C rich environment than in the A·T rich environment. This 'context effect' is more striking for the poorly repaired G·A mismatch than for the well repaired G·T mismatch. We have determined (M.J., R.Wagner and M.R., unpublished) that the repair efficiency is directly proportional to the G·C content of the (minimum) 4 bp on each side of the mismatch.

We have examined by different techniques the structure of the undecanucleotide heteroduplexes with the mismatch in the middle corresponding to the λ DNA heteroduplexes, the repair of which was studied *in vivo* (Table I). While none of these techniques provides sufficiently precise insight into the structural reasons for the high and low repair efficiency of these defined mismatches, together they suggest a major structural difference between an unrepaired mismatch and three well repaired mismatches.

In the gel electrophoresis experiment we observe what must be a major conformational change for the (unrepaired) heteroduplex (1) relative to the (repaired) heteroduplex (3) and to the fully matched duplex. The observed mobility of heteroduplex (1) corresponds to a helix of ~9 bp, suggesting that several base pairs must be involved. The simplest model would involve two nucleotides on either strand extrahelical and the remaining bases pairing and stacking more or less normally.

The n.m.r. data provide considerable information on the location of the helix perturbation. The NOESY spectrum, Figure 5, shows the normal connectivities for ^{3}A to ^{5}T, ^{19}A to ^{22}T and ^{14}T to ^{12}T. The 5 A·T base pair sequence observed by sequential interimino NOEs, with the above NOESY connectivities, shows that from ^{1}A to ^{5}T the bases are paired and the helix shows no major conformational change. Similarly at the other end of the helix we observed two adjacent A·T pairs. At 14°C we would expect that one or both of the terminal A·T base pairs fray and that the imino resonance would be very broad as previously observed for sequences terminating with A·T base pairs (Feigon *et al.*, 1983). This, with the NOESY spectrum, shows that from ^{12}T to ^{14}T the bases must be paired.

There remain three bases on each strand to which we observe no NOESY connectivities. For this section we observe, however, all three imino protons. A guanosine imino proton of a G·C base pair is observed in a normal chemical shift range. A second G and a T imino proton are observed to high field, in a region characteristic for imino protons in non-hydrogen bonded loop structures. This suggests that we have a normal $^{15}G\cdot{}^{8}C$ base pair with $^{17}G\cdot{}^{6}A$ and $^{16}A\cdot{}^{7}T$ looped out, scheme I:



Scheme I Scheme II Scheme III

This would also explain the absence of a NOESY connectivity between ^{6}A and ^{7}T.

Two observations, however, contradict this interpretation. Firstly, we do not observe a normal interimino NOE from the G·C base pair to $^{14}T\cdot{}^{9}A$. Rather we observe an extremely small NOE to the imino proton of this base pair and a similar small NOE to the imino proton of $^{18}A\cdot{}^{5}T$. The G·C base pair is thus approximately equidistant from these two base pairs. Secondly, the imino resonances at high field do not exchange with solvent, this requires that the imino protons are protected by the conformation (loop structures) since an extrahelical base would exhibit rapid imino exchange.

It must be emphasized that under the conditions of these experiments heteroduplex (1) exhibits exchange in the imino proton spectra between three species: the two observed duplex forms and the single strand species, especially at 20°C. We do not propose a 3D model for the major species but rather we have determined the region of major conformational change in the heteroduplex.

Scheme II fits our data much better than any of the others we have considered. A CPK model of scheme II shows the feasibility of this pairing. The $^{17}G\cdot{}^{8}C$ pair is displaced towards the small groove because of the two loops in the 5′ direction. Additional

stabilizations through Hoogsteen pairing of ^{15}G and ^{7}T with A·T base pairs may occur and would account for these two observed imino resonances. The minor form observed is certainly that of the fully hydrogen bonded helix (scheme III) with base pairing between $^{17}G \cdot {}^{6}A$.

This study shows that the local structure of DNA at and around the site of a mismatch is determined both by the nature of the mismatched bases and by the surrounding base pairs. The G·T mismatch in the A·T rich region and the G·A mismatch in the G·C rich region are repaired with similar efficiencies (Table I) and were found in regular helical structures by n.m.r.

For heteroduplex (1) corresponding to the unrepaired G·A mismatch in the A·T rich region (Table I), we propose that the major species observed has a structure with two loops which are not directly opposite each other, but are separated by a G·C base pair, scheme II. We have not been able to fit our data to any fully helical conformation. The very small NOEs observed from the G·C imino proton to two A·T imino protons must reflect real long imino−imino proton internuclear distances. While the separation between these two A·T base pairs cannot be accurately calculated scheme II predicts that the helix length would be shortened by 0.3 to 0.5 nm, relative to a fully base paired form. This prediction is in agreement with the gel electrophoresis experiment (Figure 1).

The major species is in fairly rapid equilibrium with one or more minor forms which are very likely fully base-paired. On increasing the G·C content [heteroduplex (3), Figure 6a], the equilibrium is strongly shifted towards the fully base-paired form, but a minor species with a different looped-out structure is still observed.

The nature of the base pairs adjacent to the G·A mismatch plays a major role. Two previous n.m.r. studies on the G·A mismatch (Kan *et al.*, 1983; Patel *et al.*, 1984) both on duplexes with 50% G·C pairs found G·A pairing with both bases *anti*. Interestingly, a minor species giving an imino resonance at 12.11 p.p.m. was observed in one of these studies (Patel *et al.*, 1984). We also observed other resonances to even higher field consistent with loop formation rather than a different G·A base pair form.

As we observed only the fully base-paired form for the well-repaired G·T mismatch, we can conclude that the extent of G·A mismatch repair is directly related to the loop forming capacity and thus to the A·T content of the sequence. More detail of these structures will be obtained from the full interpretation of the 2D NOESY spectra, and from chemical modification experiments which are in progress.

Not all other unrepaired mismatches need to show the same looped-out structure (scheme II) because of the apparent sequence dependence. Possibly any extrahelical bases and/or loops may remain unrecognized by the *E. coli* mismatch repair system. It should be noted that large single-stranded loops in heteroduplex DNA are not repaired by this *E. coli* repair system (Wagner *et al.*, 1984).

Because of the high mutator effect in mismatch repair deficient *E. coli mut* mutants, it was suggested that unrepaired mismatches must be very rare errors in DNA replication (Wagner *et al.*, 1984; Dohet *et al.*, 1985; Kramer *et al.*, 1984). The proposed looped-out structure suggests a reason: when such mismatches occur during DNA synthesis they are expected to be eliminated by the single-strand specific 3′→5′ 'proof-reading' exonuclease more efficiently than the well-fitted intrahelical mismatches. Even initial misincorporation resulting in looped-out mismatches may be less likely than for other mismatches: T_4 phage DNA polymerase makes less mistakes in the A·T rich regions (prone to yield unrepaired looped-out mismatches) than in the G·C rich regions (Petruska and Goodman, 1985).

Materials and methods

Genetic experiments with heteroduplex DNA

λ phages with sequenced mutations in the cI gene (referred to in Table I) were obtained from Dr F.Hutchinson (Yale University). Strand preparation and annealing to form defined heteroduplex DNA, transfection conditions and scoring procedures for the analysis of the phage progeny derived from individual heteroduplex molecules have been described (Dohet *et al.*, 1985). Unmethylated DNA was prepared from phages grown in *dam*$^-$ bacteria GM33 (Marinus and Morris, 1973). The percentage of infective centers from *E. coli* transfected with a single heteroduplex molecule was determined by analyzing at least 220 individual infective centers for each heteroduplex. Percent mixed progeny from transfections of mismatch repair deficient *E. coli mutL* (*mut* −211) bacteria (Dohet *et al.*, 1986) was used as standard aginst which the repair efficiency in *mut*$^+$ *E. coli* C600 was calculated.

Syntheses

Three undecamer duplexes (Table I) have been used in this study, corresponding exactly to the λ mismatch sequences studied. The five undecamers were synthesized by the phosphotriester method (van der Marel *et al.*, 1981; Marugg *et al.*, 1984).

N.m.r. spectroscopy

Spectra were recorded at 500 MHz on a Bruker WM-500 spectrometer. The solvent peak was suppressed by a 1-τ-1 hard pulse sequence (Moore and Kim, 1983) and the origin of the spectrum placed in the imino region. When integrals are reported the origin was placed symmetrically between the resonances whose integrals were to be compared. Two dimensional NOESY spectra (Jeener *et al.*, 1979; Macura *et al.*, 1981) were recorded in D_2O with a mixing time of 300 ms.

Oligonucleotides were 4 mM in strand concentration and dissolved in either D_2O or 90% H_2O/10% D_2O, 150 mM NaCl, 10 mM phosphate, pH 7.2. Chemical shifts are reported relative to internal tetramethylammonium chloride at 3.18 p.p.m.

Acknowledgements

Work in Paris was supported by a fellowship from the Centre National de la Recherche Scientifique (CNRS) to M.J. and by grants to M.R. from CNRS. M.J. and M.R. also acknowledge support from the Association pour la Recherche sur le Cancer.

References

Dohet,C., Wagner,R. and Radman,M. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 503−505.

Dohet,C., Wagner,R. and Radman,M. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3395−3399.

Fazakerley,G.V., van der Marel,G.A., van Boom,J.H. and Guschlbauer,W. (1984) *Nucleic Acids Res.*, **12**, 8269−8279.

Fazakerley,G.V., Téoule,R., Fritzsche,H., Guy,A. and Guschlbauer,W. (1985) *Biochemistry*, **24**, 4540−4546.

Feigon,J., Denny,W.A., Leupin,W. and Kearns,D.R. (1983) *Biochemistry*, **22**, 5930−5942.

Glickman,B. and Radman,M. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 1063−1067.

Haasnoot,C.A.G., den Hartog,J.H.J., de Rooij,J.F.M., van Boom,J.H. and Altona,C. (1980) *Nucleic Acids Res.*, **8**, 169−181.

Haasnoot,C.A.G., de Bruin,S.H., Berendsen,R.G., Janssen,H.G.J.M., Binnedijk, T.J.J., Hilbers,C.W., van der Marel,G.A. and van Boom,J.H. (1983) *J. Biomol. Struct. Dyn.*, **1**, 115−129.

Jeener,J., Meier,B.H., Bachmann,P. and Ernst,R.R. (1979) *J. Chem. Phys.*, **71**, 4546−4553.

Kan,L.S., Chandrasegaran,S., Pulford,S.M. and Miller,P.S. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4263−4265.

Kennard,O. (1985) *J. Biomol. Struct. Dyn.*, **3**, 205−226.

Kramer,B., Kramer,W. and Fritz,H.-J. (1984) *Cell*, **38**, 879−887.

Lu,A.L., Clark,S. and Modrich,P. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4639−4643.

Lu,A.L., Welsh,K., Clark,S., Su,S.S. and Modrich,P. (1984) *Cold Spring Harbor Symp. Quant. Biol.*, **49**, 589−596.

Lyons,S.M. and Schendel,P.F. (1984) *J. Bacteriol*, **159**, 421−423.

Macura,S., Huang,Y., Suter,D. and Ernst,R.R. (1981) *J. Magn. Res.*, **43**, 259−281.

Marinus,M.G. and Morris,N.R. (1973) *J. Bacteriol*, **114**, 1143−1150.

Marugg,J.E., Tromp,M., Ihurani,P., Hoyng,C.F., van der Marel,G.A. and van Boom,J.H. (1984) *Tetrahedron*, **40**, 73–78.
Moore,P. and Kim,P. (1983) *Biochemistry*, **22**, 2615–2622.
Patel,D.J., Kozlowski,S.A., Marky,L.A., Rice,J.A., Broka,C., Dallas,J., Itakura, K. and Breslauer,K.J. (1982) *Biochemistry*, **21**, 437–444.
Patel,D.J., Kozlowski,S.A., Ikuta,S. and Itakura,K. (1984) *Biochemistry*, **23**, 3207–3217.
Patel,D.J., Kozlowski,S.A., Hare,D.R., Reid,B., Ituka,S., Lander,N. and Itakura, K. (1985) *Biochemistry*, **24**, 926–935.
Petruska,J. and Goodman,M.F. (1985) *J. Biol. Chem.*, **260**, 7533–7539.
Pukkila,P.J., Peterson,P., Herman,G., Modrich,P. and Meselson,M. (1983) *Genetics*, **104**, 571–582.
Radman,M. and Wagner,R. (1984) *Curr. Top. Microbiol. Immun.*, **108**, 23–28.
Radman,M., Wagner,R., Glickman,B.W. and Meselson,M. (1980) In Alacevic, M. (ed.), *Progress in Environmental Mutagenesis*. Elsevier, Amsterdam, pp. 121–130.
Tibanyenda,N., de Bruin,S.H., Haasnoot,C.A.G., van der Marel,G.A, van Boom, J.H. and Hilbers,C.W. (1984) *Eur. J. Biochem.*, **139**, 19–27.
Topal,M.D. and Fresco,J.R. (1980) *Nature*, **263**, 285–289.
van der Marel,G.A., van Boeckel,C.A.A., Wille,G. and van Boom,J.H. (1981) *Tetrahedron Lett.*, **22**, 3887–3890.
Wagner,R. and Meselson,M. (1976) *Proc. Natl. Acad. Sci. USA*, **73**, 4135–4139.
Wagner,R., Dohet,C., Jones,M., Doutriaux,M.-P. and Radman,M. (1984) *Cold Spring Harbor Symp. Quant. Biol.*, **49**, 611–615.

Received on 23 June 1986; revised on 7 October 1986