The contrasting structures of mismatched DNA sequences containing looped-out bases (bulges) and multiple mismatches (bubbles)

Anamitra Bhattacharyya and David M.J.Lilley*

Department of Biochemistry, The University, Dundee DD1 4HN, UK

Received July 3, 1989; Revised and Accepted August 8, 1989

ABSTRACT

We have studied the structure and reactivities of two kinds of mismatched DNA sequences – unopposed bases, or bulges, and multiple mismatched pairs of bases. These were generated in a constant sequence environment, in relatively long DNA fragments, using a technique based on heteroduplex formation between sequences cloned into single-stranded M13 phage. The mismatched sequences were studied from two points of view, viz

1. The mobility of the fragments on gel electrophoresis in polyacrylamide was studied in order to examine possible bending of the DNA due to the presence of the mismatch defect. Such bending would constitute a global effect on the conformation of the molecule.
2. Sequences in and around the mismatches were studied using enzyme and chemical probes of DNA structure. This would reveal more local structural effects of the mismatched sequences.

We observed that the structures of the bulges and the multiple mismatches appear to be fundamentally different. The bulged sequences exhibited a large gel retardation, consistent with a significant bending of the DNA at the bulge, and whose magnitude depends on the number of mismatched bases. The larger bulges were sensitive to cleavage by single-strand specific nucleases, and modified by diethyl pyrocarbonate (adenines) or osmium tetroxide (thymines) in a non-uniform way, suggesting that the bulges have a precise structure that leads to exposure of some, but not all, of the bases. In contrast the multiple mismatches ('bubbles') cause very much less bending of the DNA fragment in which they occur, and uniform patterns of chemical reactivity along the length of the mismatched sequences, suggesting a less well defined, and possibly flexible, structure. The precise structure of the bulges suggests that such features may be especially significant for recognition by proteins.

INTRODUCTION

Mismatched sequences can arise in DNA as a result of errors in replication, recombination between sequences that are not fully homologous, or mutagenesis events [1]. In addition, mismatches are a common feature in the secondary structure of single-stranded nucleic acids, such as most RNA molecules and some viral genomes. Base mismatches can take a number of forms. Single mismatches occur where two non-Watson-Crick bases are in opposition. When a number of such mismatches occur consecutively, this constitutes a multiple mismatch, and we will refer to this situation as a 'bubble'. A further type of mismatch may arise, where there is an additional base or bases on one side of the double helix, which we call a 'bulge'. Such bulges are believed to be intermediates in frameshift mutagenesis [2]. Furthermore, they may be substrates for certain structure-selective nucleases such as T4 endonuclease VII [3], and could therefore lead to illegitimate recombination events. Bulges in RNA molecules may act as specific protein binding sites, exemplified by the binding site for the coat protein of phage R17 [4].

Base mismatches in DNA have been the subject of considerable structural investigation. Single base mismatches have been studied principally by X-ray crystallographic [5, and references therein] and NMR [6, and references therein] methods, with the general result that mismatching appears not to be very disruptive—all the mismatches studied are accommodated within the double helix with only very local perturbation of structure. Multiple mismatches have been relatively neglected from a structural point of view, while studies of base bulges have been largely restricted to single extra bases. Two kinds of structures have been deduced for these bulges. Woodson and Crothers [7] have studied oligonucleotides containing an extra guanine base using NMR spectroscopy and molecular mechanics, and deduced that the extra purine is inserted into the helix, where it creates a kink in the helix axis of about 20°. NMR studies of sequences containing additional adenine bases [8–10] have reached similar conclusions, but crystallographic studies of DNA oligonucleotides containing single adenine bulges led Sussman and co-workers [11,12] to conclude that the adenine was extrahelical. An extrahelical structure was also proposed for a sequence containing a single cytosine bulge by Morden et al [13] on the basis of NMR investigations—in general it is believed that an extra pyrimidine base can be either intra- or extrahelical, depending on the temperature.

We have recently described a technique for generating defined mismatches in a constant sequence environment, in relatively long heteroduplexed DNA fragments [14]. We used these constructs to study single base mismatches from two points of view. First, we have measured the mobility of these fragments by polyacrylamide gel electrophoresis. Studies on sequence-directed curvature in DNA [15–19] have shown that this method is very sensitive to distortion of the helix axis, which is rather difficult to demonstrate using crystallographic methods or NMR spectroscopy. Second, we examined the sensitivity of the bases in and around the mismatches to enzyme and chemical probes of DNA structure. We showed that none of the single mismatches caused measurable distortion of the helix axis, but that some mismatches (notably those involving the formation of wobble basepairs) were associated with specific chemical reactivity.

In the present study we have applied these methods to the analysis of base bulges and multiple mismatches. We conclude that bulges and bubbles are of very different character. Base bulges are associated with a precise bending of the helix axis, and generate specific patterns of chemical reactivity, while bubbles appear to cause very much less bending, and a more uniform chemical reactivity.

EXPERIMENTAL

Construction and preparation of plasmids

Oligonucleotides were synthesised on an Applied Biosystems 381A DNA synthesizer, using β-cyanoethyl phosphoramidite chemistry [20,21]. Fully deprotected oligonucleotides were purified on 20% denaturing polyacrylamide gels, and DNA recovered by electroelution and ethanol precipitation.

Oligonucleotides possessing *Bam*HI cohesive end, were cloned into the *Bam*HI site of pAT153 [22]. Recombinants were screened for the presence of monomer inserts having the correct sequence and orientation, using chemical sequencing techniques [23].
Supercoiled plasmid DNA was prepared from *E. coli* grown for 16 h in the presence of 150 μg/ml chloramphenicol, by lysozyme, EDTA, SDS lysis and isopycnic centrifugation in CsCl/ethidium bromide gradients. DNA was recovered by side puncture, ethidium

Table 1.

A. Plasmids and phages.

oligonucleotide duplexes	plasmids	M13 phages + strand	M13 phages - strand
GATCCAGGTCG GTCCAGCCTAG	pMM31	M13mp8-31	M13mp9-31
GATCCAGGAAATCG GTCCTTTAGCCTAG	pLPA3	M13mp8A3	M13mp9T3
GATCCAGGTTTTCG GTCCAAAAGCCTAG	pLPT3	M13mp8T3	M13mp9A3
GATCCAGGAAAAATCG GTCCTTTTTAGCCTAG	pLPA5	M13mp8A5	M13mp9T5
GATCCAGGTTTTTTTCG GTCCAAAAAAAGCCTAG	pLPT5	M13mp8T5	M13mp9A5

B. Heteroduplex constructions.

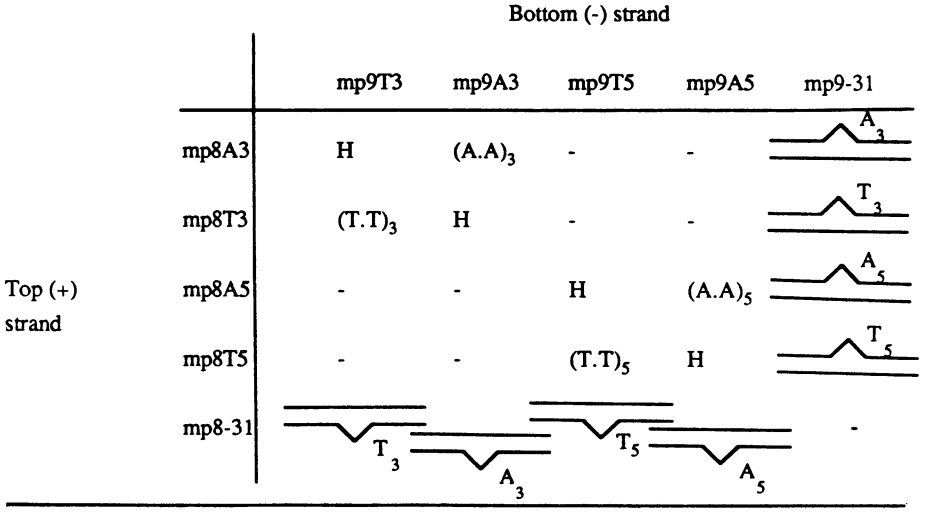


Table 1. Summary of the plasmids and phages constructed for these studies, and the possible species generated by hybridisation between M13mp8 and M13mp9 phages. **A.** Plasmids and phages used. The oligonucleotides shown were cloned into the *Bam*HI site of pAT153, generating the plasmids clones indicated. The *Eco*RI to *Sal*I fragments of each plasmid were excised and cloned separately into the phages M13mp8 and M13mp9, the polylinkers of which are in opposite orientation. Thus different strands of the cloned sequences occur in the encapsulated single-stranded DNA of the phage. **B.** Summarises the different species that can be made by hybridisation of appropriate combinations of mp8 and mp9 clones. H indicates homoduplexes.

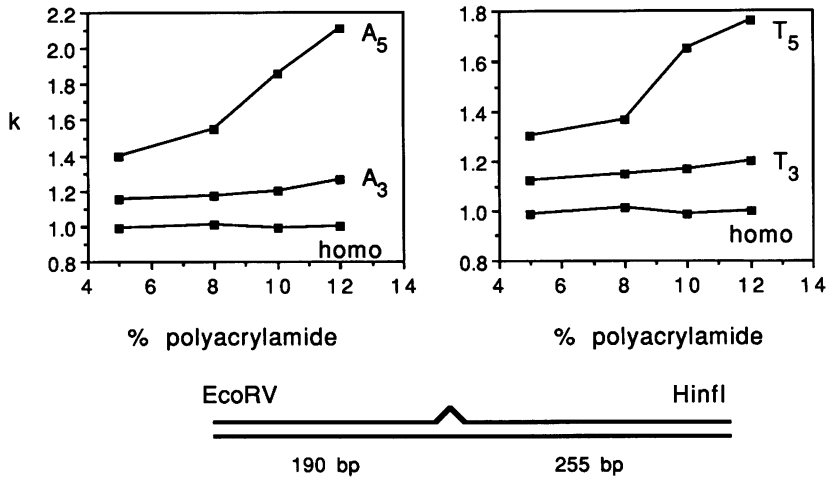


Figure 1. Base bulges cause retardation of mobility in polyacrylamide gels. Variation in gel mobility of DNA fragments containing base bulges, as a function of gel concentration. k-values were measured for the *Eco*RV-*Hin*fI fragment (shown below) containing either no bulge (homo), three or five bulged bases, either adenine or thymine.

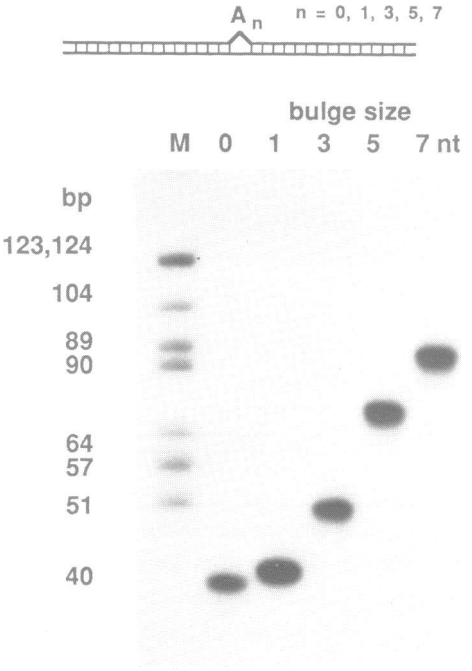


Figure 2. Electrophoretic migration of bulged oligonucleotides in polyacrylamide gels depends on bulge size. 40 bp oligonucleotides containing central bulges of 0, 1, 3, 5 or 7 adenine bases were electrophoresed in a 15% polyacrylamide gel. Marker DNA fragments (pAT153 digested by *Hae*III) were electrophoresed in an adjacent lane (M) to enable the calculation of k-values.

bromide removed by repeated extraction in butan-1-ol, followed by extensive dialysis against cold 10 mM Tris.HCl (pH 7.5), 1 mM EDTA (TE) buffer.

Construction and preparation of M13 clones

The 660 bp *Eco*RI-*Sal*I fragments from each of the recombinant plasmids were cloned into the polylinkers of phages M13mp8 and M13mp9 [24], and propagated in *E. coli* JM103. Single- stranded phage DNA was prepared by precipitation of cellular supernatants using 20% polyethylene glycol/2.5 M NaCl, extraction with phenol/chloroform, and DNA recovered by ethanol precipitation.The relevant sequences of all M13 phage clones were verified by dideoxy procedures.

Hybridisation of phage DNA

Equal quantities (approximately 20 μg) of phage DNA from M13mp8 and mp9 clones were hybridised together in 0.15 M Na citrate, 0.015 M NaCl for 3 h at 65°C. Heteroduplex fragments were isolated as described by Gough and Lilley [25], by restriction cleavage of the heteroduplex, and polyacrylamide gel electrophoresis.

Enzymes

S1 nuclease (Bethesda Research Labs) digestions were performed (using <1 μg of DNA) in 50 mM Na acetate (pH 4.6), 50 mM NaCl, 1 mM $ZnCl_2$ on ice for 2 h, using 5−10 units of S1 nuclease.

Micrococcal nuclease (Sigma) cleavage reactions were performed in 15 mM Tris.HCl (pH 7.4), 75 mM NaCl, 1 mM $MgCl_2$, 2 mM $CaCl_2$, 0.2 mM EDTA, 15 mM β-mercaptoethanol on ice for 15−20 min, using 1.7×10^{-4} units of micrococcal nuclease. Nuclease reactions were stopped by ethanol precipitation of the DNA.

Restriction enzymes were purchased from either Bethesda Research Laboratories, Amersham or New England Biolabs, and used as directed by the manufacturer.

5′-termini of DNA fragments were radioactively labelled by the exchange technique [26] using [γ-^{32}P] ATP and T4 polynucleotide kinase (Amersham). 3′-termini were labelled using [α-^{32}P] dCTP and Klenow DNA polymerase I (Amersham).

Chemical modifications

Osmium tetroxide (Sigma) was dissolved in distilled water as a 20 mM stock solution (stored as aliquots at −70°C). Modification reactions [27−29] were performed in 5 mM Tris.HCl (pH 7.5), 1 mM EDTA, 1% (v/v) pyridine (Aldrich), using 1 mM osmium tetroxide on ice or at 25°C for 15 min. Reactions were terminated by ethanol precipitation of DNA.

Diethylpyrocarbonate (DEP; Sigma) reactions [30−32] were carried out in 50 mM Na cacodylate (pH 7.1), 1 mM EDTA to which 3 μl of DEP was added, and incubated at 20°C for 15 min. The reaction was stopped by precipitation of the DNA with ethanol.

In both cases, chemically modified DNA was incubated with *Taq*I, radioactively labelled, and asymmetrically end-labelled fragments isolated by restriction cleavage with *Sph*I, followed by electrophoresis in 5% polyacrylamide. Modified DNA fragments were recovered by electroelution and ethanol precipitation. The DNA backbone was cleaved at the modified base(s) by treatment with 1.0 M piperidine, at 90°C for 30 min, followed by extensive lyophilisation.

DNA sequencing and gel electrophoresis

End-labelled DNA fragments were sequenced by the chemical degradation method [23]. After piperidine degradation and lyophilisation, samples were redissolved in formamide and analysed in 8.0% polyacrylamide gels containing 7 M urea, using 90 mM Tris.borate (pH 8.3), 10 mM EDTA (TBE) running buffer. Electrophoresis was carried out so that the plates were hot to the touch. Gels were dried on Whatman 3MM paper, and

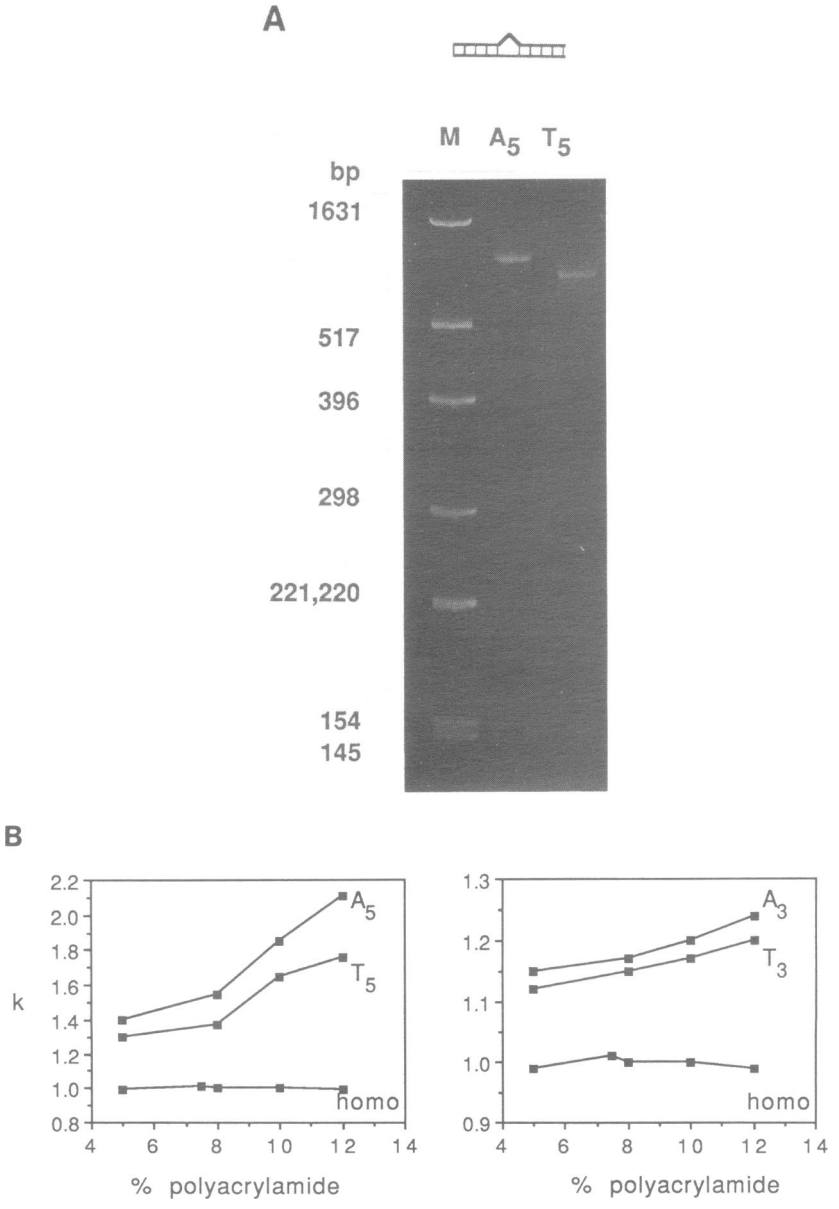


Figure 3. Electrophoretic migration of bulge-containing DNA fragments depends on the sequence of the loop. **A.** 5% polyacrylamide gel electrophoresis of the *Eco*RV-*Hin*fI fragment containing either A_5 or T_5 bulges, compared with a *Hin*fI digest of pAT153 (M). **B.** Plots of k-values for 5-base and 3-base bulges against percentage of polyacrylamide. The perfectly matched sequences are indicated (homo).

Table 2.

Bulge	Loop-out on mp8(+) strand	Loop-out on mp9(−) strand
A_5	1.57	1.80
T_5	1.38	1.38
A_3	1.26	1.35
T_3	1.20	1.20

Table 2. Gel mobility of *Eco*RV to *Hin*fI fragments containing 5 and 3 base bulges as a function of sequence context. Mobilities are expressed as k-values, the apparent size of the fragment (interpolated from the mobility of normal DNA restriction fragments) divided by the true size. The sets of k-values for 5 and 3 base bulges are not strictly comparable, as they were measured in 8 and 12% polyacrylamide gels respectively.

autoradiographed with Fuji-RX X-ray film at −70°C using Ilford fast-tungstate intensifier screens. M13 clones were sequenced using polymerase chain termination procedures [33].

Studies of gel electrophoretic migration were performed on 2.0% agarose and various concentrations (between 5% and 15%) of native polyacrylamide gels. Samples were

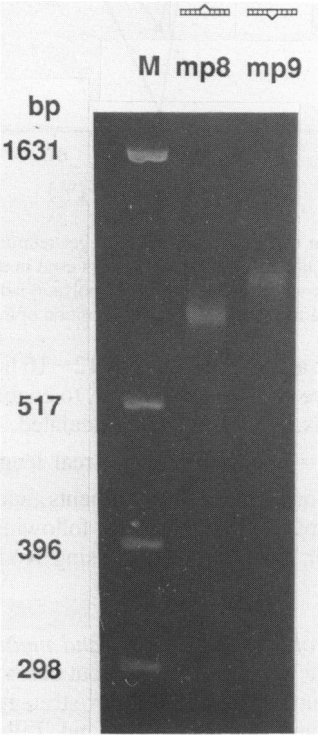


Figure 4. Sequence context may affect gel retardation due to base bulges. 5% polyacrylamide gel electrophoresis of the *Eco*RV-*Hin*fI fragment containing an A_5 bulge on either the top (mp8) or bottom (mp9) strand, compared with a *Hin*fI digest of pAT153 (M).

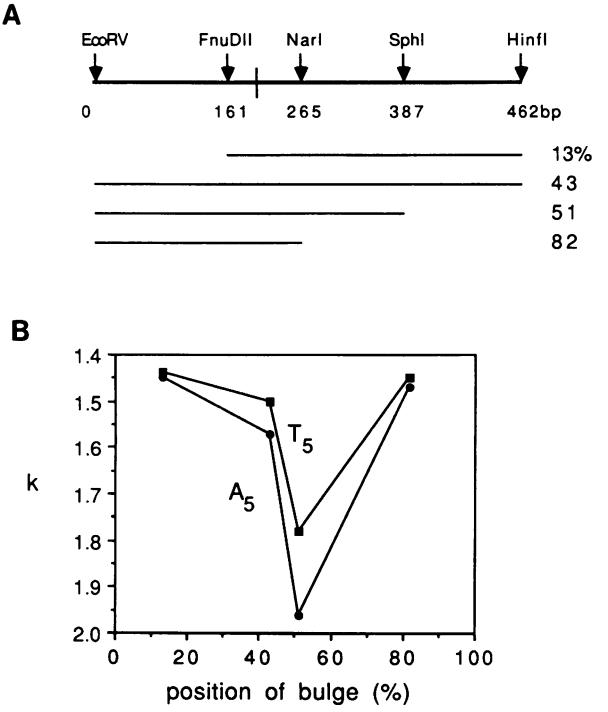


Figure 5. The bending is localised at the bulges. Variation in gel retardation as a function of the position of the bulge. **A.** Schematic showing the bulge-containing fragments used in the analysis. The vertical line denotes the position of the bulge. **B.** Plots of k-value (derived from 5% polyacrylamide gel electrophoresis) of fragments containing either A_5 or T_5 bulges, as a function of the percent position of the bulge relative to the fragment ends.

electrophoresed in TBE buffer at 20 ± 0.1°C, for 12−16 hours. Typically, several DNA molecular weight markers were additionally loaded, to enable apparent size determination of DNA fragments. From this, k values were calculated, where

$$k = \text{apparent length / real length}$$

in order to facilitate comparisons between experiments, and with other work. DNA was visualised by staining in 1 μg/ml ethidium bromide; following destaining in distilled water, gels were photographed under UV illumination using Kodak Tri-X Pan film.

RESULTS

Strategy for the construction of sequence bulges and multiple mismatches.

We constructed DNA fragments containing looped-out bases and multiple-base mismatches in a constant sequence environment using a similar strategy to that previously described by Gough and Lilley [25] and Bhattacharyya and Lilley [14]. A set of synthetic oligonucleotides was cloned into pAT153 generating the set of plasmids shown in Table 1. The 660bp *Eco*RI-*Sal*I fragment from each plasmid was cloned into the single-stranded phages M13mp8 and M13mp9.

Three types of hybrid duplex species could be created by hybridisation of the appropriate single-stranded phage DNA, as shown in Table 1:

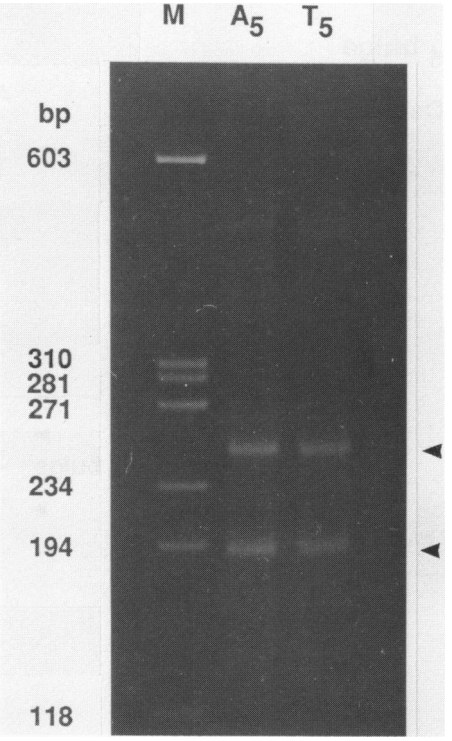


Figure 6. Cleavage of base bulges by a single-strand specific enzyme. *Eco*RV-*Hin*fI fragments containing either A_5 or T_5 bulges were digested with micrococcal nuclease at 0°C, and the products analysed by electrophoresis in a 5% polyacrylamide gel. Cleavage at the bulge site results in the generation of two fragments of 260 and 190 bp (indicated arrowed on the right). The marker DNA fragments (M) were a *Hae*III digest of phage φX174.

1. Heteroduplexes containing looped-out bases (DNA bulges), on either the top (mp8) or bottom (mp9) strand. Bulges containing three or five bases, either adenines or thymines, were constructed.
2. Heteroduplexes containing $(A.A)_n$ or $(T.T)_n$ base mismatches, where n = 3 or 5.
3. Normal Watson-Crick homoduplexes.

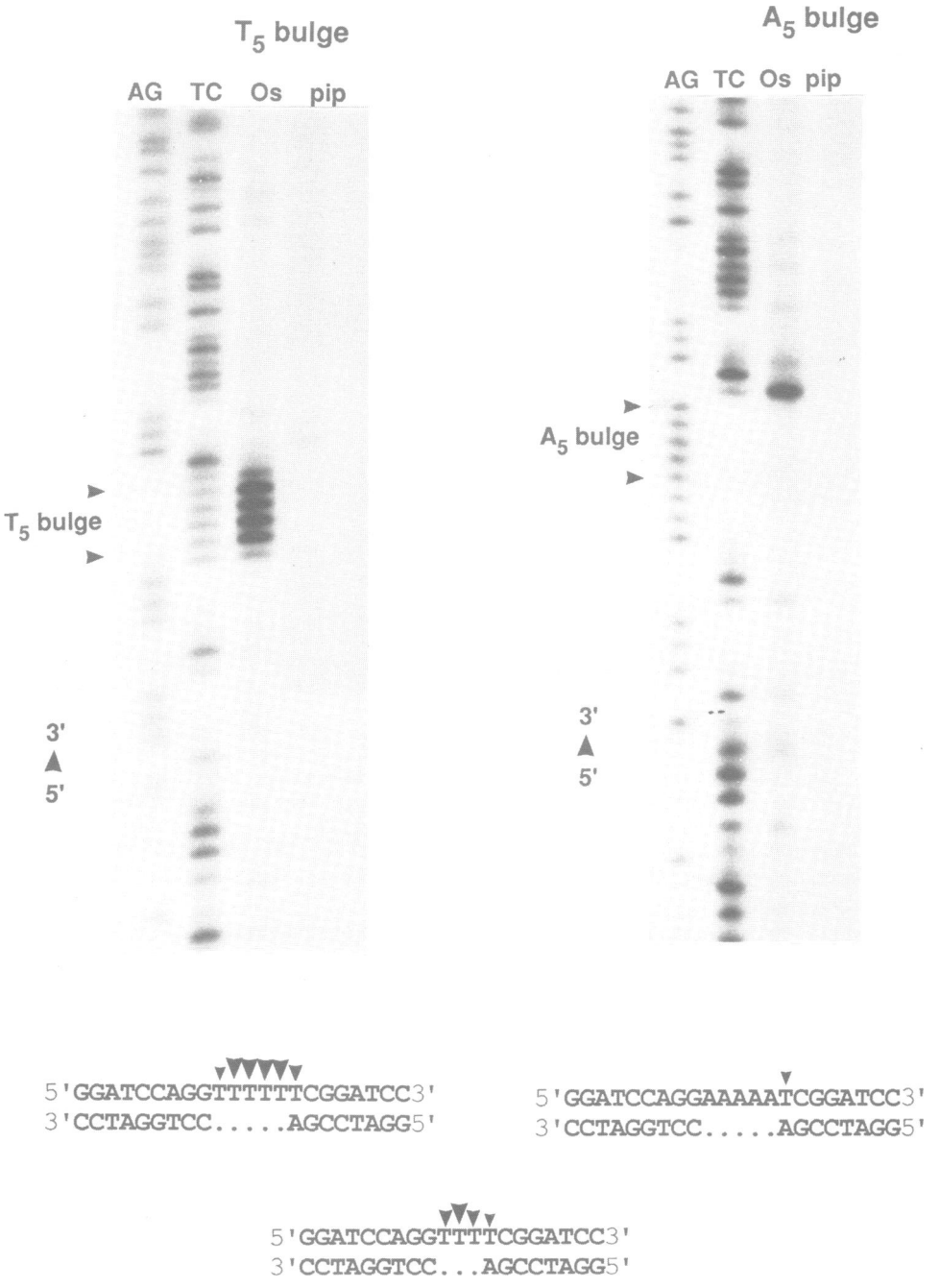
The structural properties DNA base bulges

Our preliminary studies exploited the general observation that distorted (eg curved) DNA molecules may exhibit significant retardation when electrophoresed through polyacrylamide

Table 3.

Bulge	S1 nuclease	Micrococcal nuclease
A_5	+	+
T_5	+	+
A_3	−	−
T_3	−	−

Table 3. Summary of results of nuclease digestion of 5 and 3 base bulges. + indicates cleavage, − indicates very weak or no cleavage.


T5 bulge
AG TC Os pip
T5 bulge
3'
5'
A5 bulge
AG TC Os pip
A5 bulge
3'
5'
5'GGATCCAGGTTTTTTCGGATCC3'
3'CCTAGGTCC.....AGCCTAGG5'
5'GGATCCAGGAAAAATCGGATCC3'
3'CCTAGGTCC.....AGCCTAGG5'
5'GGATCCAGGTTTTCGGATCC3'
3'CCTAGGTCC...AGCCTAGG5'

gels [15–19, 25]. We applied this approach to investigate possible global effects associated with introducing DNA base bulges into a DNA helix.

(a) Anomalous gel mobility caused by bulges.

The mobility of heteroduplexes containing bulges of 3- or 5-purine (adenine) or pyrimidine (thymine) bases in polyacrylamide gels was studied. The variation in the k-factor (apparent size divided by the true size of the fragment) as a function of gel concentration is summarised in Figure 1. These bulge-containing fragments exhibited anomalously slow migration in polyacrylamide gels, which was greater for the larger bulges. A similar effect has been recently noted by Hsieh and Griffith [34], using a linker ligation procedure.

We investigated the size dependence of the retardation further using 40-mer oligonucleotides containing an additional 0 (homoduplex), 1, 3, 5 or 7 extra adenine residues on one strand. The sequence environment around the additional adenines, was the same as that present in the phage heteroduplex species. The migration of these species in 15% polyacrylamide is shown in Figure 2. The migration of the fragments becomes progressively slower as additional bases are added to the unpaired tract. We interpret these results in terms of a local bending or kinking brought about by the presence of the bulge, where the extent of the distortion of the path of the helix axis depends on the number of bases present in the bulge.

(b) The effect of bulge sequence on the gel anomaly

The sequence-dependence of the gel retardation due to bulges was examined using the M13 phage DNA heteroduplexes containing 5-base bulges comprising either A_n or T_n. Figure 3 shows that there is a marked difference in the gel anomaly of the A_5 and T_5 bulges. This was also true for the three-base bulges. Our results suggest that there is an effect of bulge sequence on the distortion of the DNA kinking, as also noted by Kline *et al* [35]. In general, adenine base bulges produce a greater distortion than thymine bulges, possibly due to the greater size of the purine bases. The extent of retardation due to the bulges is considerable, and is comparable to that due to multiple phased runs of adenines [16–18].

(c) The role of flanking DNA sequences on the gel electrophoretic behaviour

We examined the effect of sequence context on the gel anomaly due to bulges. In all the experiments described so far, the looped-out bases were present on the top (mp8) strand of the heteroduplex, giving a constant sequence environment, namely 5′-AGG(A/T)$_n$TCG-3′. However, by using the phage heteroduplex species in which the extra bases were looped-out on the lower (mp9) strand, the flanking sequence was changed to 5′-CGA(A/T)$_n$CCT-3′. The results are summarised in Table 2. For T_n bulged heteroduplexes (n = 3 or 5) changes in the flanking sequence failed to alter the size of the gel anomaly. However, the situation was quite different for the adenine bulges. When we moved the A_5 bulge from one side of the helix to the other, a significant difference in gel anomaly resulted, shown in Figure 4. The A_5 bulge in mp9 sequence context, shows an increased k-factor relative to the value measured for the mp8 sequence environment. A similar result was observed for the A_3 bulge (data not shown).

Figure 7. Thymine-containing bulges are reactive to osmium tetroxide. Asymmetrically [^{32}P] labelled *Taq*I-*Sph*I fragments containing T_5 or A_5 bulges were modified with 1 mM osmium tetroxide, followed by piperidine cleavage. These were electrophoresed alongside sequencing markers (AG and TC) derived from the equivalent fragments of pLPT5 and pLPA5, and a control (pip) in which the osmium tetroxide was omitted. The results are summarised below the autoradiographs, together with the results of equivalent experiments on the T_3 bulge.

Table 4.

Bubble	k-value
$(A.A)_5$	1.04
$(T.T)_5$	1.04
$(A.A)_3$	1.02
$(T.T)_3$	1.02
$(A.T)_3$	1.00

Table 4. Gel mobilities of bubble-containing DNA fragments. *Eco*RV to *Hin*fI fragments were electrophoresed in a 5% polyacrylamide gel, and apparent sizes measured by interpolation from normal DNA restriction fragments.

(d) Localisation of the bend centre

The gel retardation of bulge-containing fragments suggests that the extra bases on one strand might introduce a kink into the linear DNA molecule. In an attempt to localise the effect, we employed a procedure related to circular permutation analysis [18]. A set of A_5 and T_5 heteroduplexed fragments (shown in Figure 5) were prepared in which the position of the bulged structure was varied relative to the fragment ends, by choosing different pairs of restriction enzymes to release the fragment from the hybridised phage.

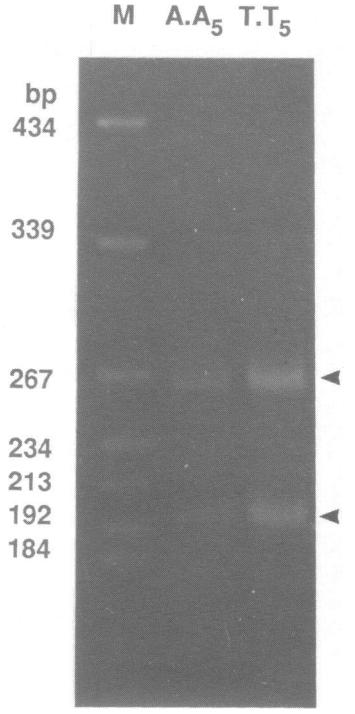


Figure 8. Multiple mismatches are cleaved by a single-strand specific nuclease. *Eco*RV-*Hin*fI fragments containing either $(A.A)_5$ or $(T.T)_5$ bubbles were digested with micrococcal nuclease at 0°C, and the products electrophoresed on a 5% polyacrylamide gel, alongside marker fragments (M) derived by *Hae*III cleavage of pAT153. The arrows on the right show the expected products of cleavage of the fragment at the bubble.

Table 5.

Mismatch	S1 nuclease	Micrococcal nuclease
$(A.A)_5$	+	+
$(T.T)_5$	+	+
$(A.A)_3$	−	+
$(T.T)_3$	−	+

Table 5. Summary of results of nuclease digestion of multiple mismatches. + indicates cleavage, − indicates very weak or no cleavage.

This procedure was used previously to localise the bending in a cruciform-containing DNA fragment [25]—structural features that bend or kink the DNA should exert their maximal effect when placed at the centre of the DNA fragment. The bulge-containing fragments were electrophoresed in polyacrylamide, beside normal DNA molecular weight markers, in order to calculate k-values. Figure 6 shows the results for the A_5-bulged heteroduplex (similar results were obtained for the T_5 bulge; data not shown). A plot of k-value against bulge position shows that the fragments exhibited maximal retardation when the bulge was centrally positioned, implying that the distortion is centred on the bulge itself.

(e) DNA bulges are sensitive to single-strand specific nucleases

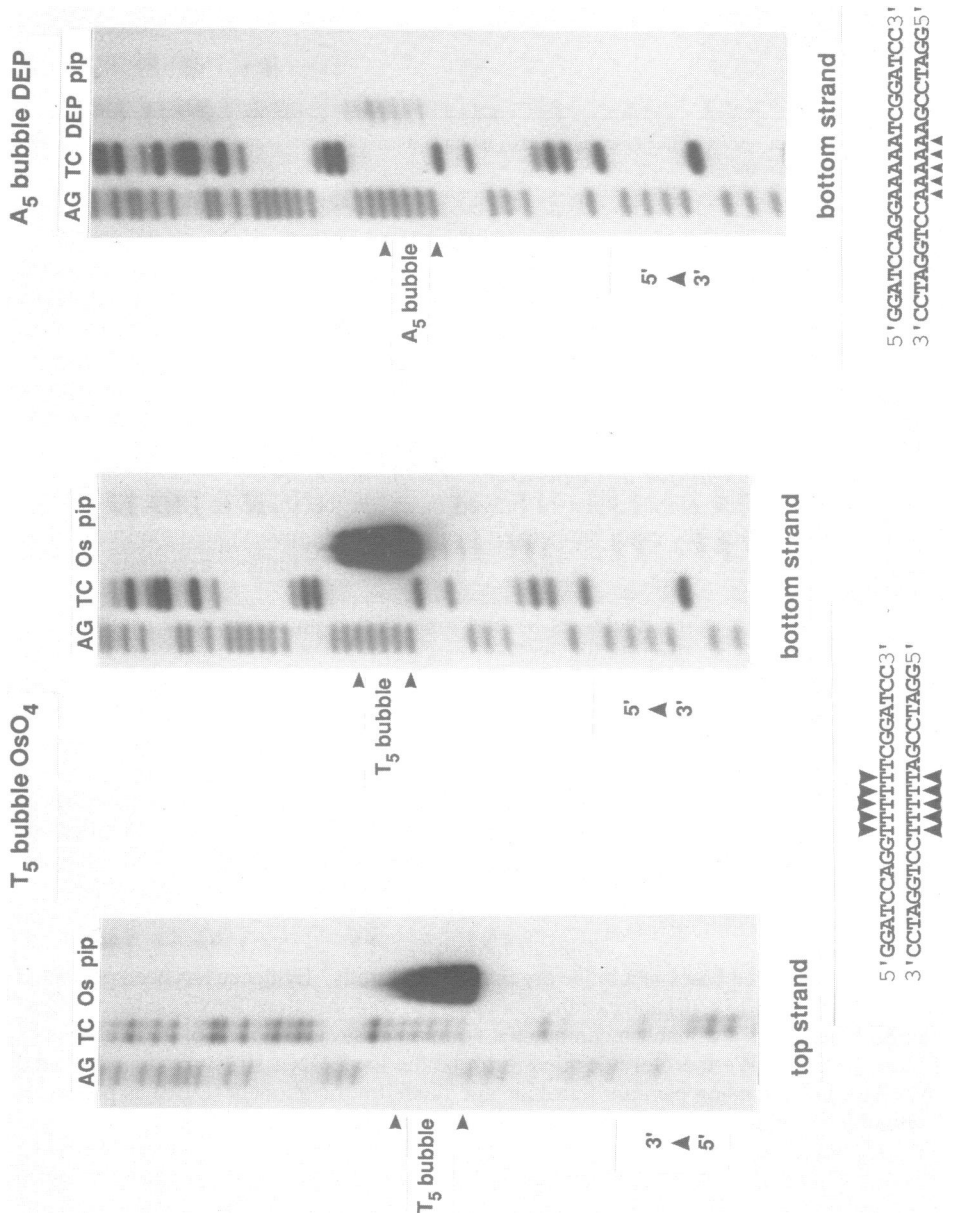
We have described the global effect of bulged bases on the shape of DNA molecules, without suggesting how the extra bases exert their effect. We investigated the local DNA structure and stereochemistry of the bases in and around the bulges, using enzyme and chemical probes.

The single-strand specific nucleases S1 and micrococcal nucleases were used for this purpose. S1 nuclease has been shown to cleave cruciform loops [36,37], and other unusual DNA structures [38−40]. Micrococcal nuclease also cleaves cruciform loops [41,42]. A_n and T_n bulges (with looped-out bases on the mp8 strand) were digested with both micrococcal and S1 nuclease. If cleavage were to occur at the bulge site, then we should expect to see the two smaller fragments of approximately 260 bp and 190 bp. Figure 6 shows that micrococcal nuclease cuts both the A5 and T_5 bulges. S1 nuclease also cleaves these structures (data not shown), and the data are summarised in Table 3. Whilst both the nucleases cut the 5-base bulges, they do not cut either of the smaller 3-base bulges. This suggests that the molecular geometry of the 3-base bulges is much tighter than that of their 5-base counterparts.

(f) DNA bulges are reactive to chemical attack.

How are the unpaired bases of a bulge organised structurally? Evidence has been presented for both intrahelical [7−10] and extrahelical [11−13] conformations for single base bulges. Chemical probing is a valuable technique for examining the accessibility of particular types of bases in DNA. We therefore decided to probe the local DNA structure in and around DNA base bulges, using osmium tetroxide and diethylpyrocarbonate (DEP) as probes of thymine and adenine accessibility in the T_n and A_n bulges, respectively.

The T_5 bulge (with the loop-out on the mp8 strand) was reacted with osmium tetroxide, followed by piperidine cleavage of thymine-osmate ester adducts, and the products examined on a sequencing gel. The autoradiograph is shown in Figure 7. The thymines in the bulge were reactive to osmium, but there was differential reactivity between the different thymines within the bulge. In particular, the first 5′T in the bulge (T1) exhibited the lowest reactivity of all. The remaining thymines of the bulge T2−5 were all reactive, and to the same degree.


T5 bubble OsO4
AG TC Os pip
T5 bubble
3'
5'
top strand
5'GGATCCAGGTTTTTTTCGGATCC3'
3'CCTAGGTCCTTTTTAGCCTAGG5'
AG TC Os pip
T5 bubble
5'
3'
bottom strand
A5 bubble DEP
AG TC DEP pip
A5 bubble
5'
3'
bottom strand
5'GGATCCAGGAAAAATCGGATCC3'
3'CCTAGGTCCAAAAAAGCCTAGG5'

This suggests that these bases are predominantly extrahelical, and may be responsible for conferring sensitivity to single-strand specific nucleases. In addition to the thymines of the bulge, the first base to the 3′ side is also a thymine, which should be base-paired. However, this thymine was also slightly reactive. Reactivity of the 3′ flanking thymine was also observed in the A_5 bulge (Figure 7), which has an identical sequence context. Several NMR studies of single-base bulged duplexes have detected structural perturbations flanking the bulge site [10, 13, 45].

We also examined reactivity to osmium tetroxide in the T_3-bulge. The observed pattern of reactivity was: T2 (central) > 5′-T1 = 3′-T3, see Figure 7. The overall reactivity was lower than found in the T_5 bulge, requiring a higher concentration of pyridine to observe modification (3% in place of 1%). This implies that the geometry of the T_3 bulge is more constrained than that of the T_5 bulge, which is also consistent with the results obtained using the single-strand specific nucleases.

DEP modifies adenines (and to a lesser extent guanines) by carbethoxylation at the N7 position of the imidazole ring [46, 47]. It has been shown to modify purines in cruciform structures [31, 32], and we have noted a differential reactivity within cruciform loops, suggesting that the reagent is sensitive to local structure in the loop. The A_5-bulged heteroduplex (loop-out the mp8 strand) was reacted with DEP, followed by piperidine cleavage (data not shown). Bases A2, A3, A4 and A5 in the bulge, were equally reactive, producing a uniform pattern of modification. However, the first 5′-adenine (A1) in the bulge was completely unreactive to DEP, indicating that N7 was inaccessible to attack. The other bases in the bulge (A2−A5) are therefore probably extrahelical and hence susceptible to nuclease attack.

The structural properties of multiple-base mismatches

We decided to compare the properties of bulged sequences, that have additional bases on one strand, with multiple mismatches, where bases are present on both strands, but such that Watson-Crick basepairing is not possible. Our phage DNA sequences allowed us to construct DNA heteroduplexes containing 3 or 5 consecutive A.A or T.T mismatches, in a constant sequence context. These could be compared with single mismatches and bulged sequences, from the points of view of DNA bending and chemical reactivity.

(a) Multiple-base mismatches have a minor effect on helical trajectory

We performed gel electrophoresis assays on heteroduplex fragments containing $(A.A)_n$ and $(T.T)_n$ mismatches (n = 3 or 5), in polyacrylamide gels of concentrations from 5% to 12%, see Table 4. For both sets of mismatched sequences there was a small retardation in the gel relative to homoduplexed DNA (that contained an $(A.T)_3$ sequence in place of the mismatches) but the effects were very much smaller than for the bulged sequences. Moreover, the distortion was not typical of DNA bending, in that the k values were

Figure 9. Bubbles are uniformly modified by osmium tetroxide and diethyl pyrocarbonate. The fragment containing the $(T.T)_5$ bubble was reacted with 1 mM osmium tetroxide; studied on both strands independently. These were electrophoresed alongside sequencing markers (AG and TC) derived from pLPT5 (note; for the bottom strand this results in an A-run at the position of the bubble). The fragment containing an A_5 bubble was reacted with DEP, cleaved with piperidine and electrophoresed on a sequencing gel alongside sequencing markers derived from pLPT5. In all cases, the right-most lane contains the result of a control experiment where the chemical probe was omitted (pip). The results of chemical modifications are summarised below the autoradiographs.

independent of gel concentration (data not shown). Thus, the effects may have an origin in flexibility rather than permanent bending induced by the mismatches.

(b) Single-strand specific nucleases cleave multiple-base mismatches

S1 and micrococcal nucleases were used to study whether the mismatched bases possessed any single-stranded character. Figure 8 shows the result of incubating the mismatched fragments in the presence of micrococcal nuclease, on ice. The results using single-strand specific nucleases are summarised in Table 5. All the 3 and 5 base $(A.A)_n$ and $(T.T)_n$ mismatches have enough single-stranded character to render them sensitive to micrococcal nuclease. This contrasts with single mismatches, and suggests that these multiple-base mismatches cannot be accommodated into a normal B-DNA helix. Although S1 nuclease was able to cleave both 5-base mismatches, it failed to cut the 3-base counterparts. This reflects the different specificities of these two nucleases. Micrococcal nuclease is,thus better able to recognise small regions of single-stranded DNA than is S1 nuclease.

It has been demonstrated previously that single-base mismatches are resistant to single-strand specific nucleases [14, 43, 44]. In general, it seems to be the case that single non-complementary base pairs are accommodated into a normal B-DNA helix, with minimal distortion of the phosphodiester backbone. In cases where there are several unpaired bases these cannot be so easily accommodated into a conventional B-helix, and so the phosphodiester backbone becomes distorted, rendering it sensitive to nuclease attack.

(c) Chemical hyper-reactivity of multiple-base mismatches

Osmium tetroxide reactions were performed on both strands of the $(T.T)_5$ heteroduplexes, and the results are shown in Figure 9. All thymines present in the five mismatched basepairs, on both strands, were reactive to osmium. The modification pattern was symmetrical, occurring to an equal extent on the two strands. Moreover, the modification was uniform throughout the $(T.T)_n$ tracts. A similar pattern of modification was observed for the $(T.T)_3$ mismatches (data not shown). We have previously shown that a single-base T.T mismatch was unreactive to osmium tetroxide [14]. Several groups have reported that the thymines in a T.T mismatch can stack into the helix [48, 49], whilst Kouchakdjian *et al* [50] have shown evidence for T.T base pair formation. However, these results indicate that the inclusion of several T.T mismatches results in structural deformation of the helix, such that the thymines become reactive to osmium tetroxide.

Analogous chemical modification experiments were then carried out on the $(A.A)_5$ mismatched DNA fragments, using DEP as a probe of adenine accessibility. Probing experiments at single nucleotide resolution were performed on both strands, the results of which are presented in Figure 9. The mismatched adenines on both strands are reactive, indicating steric accessibility at the N7 position. The modification pattern on each strand is uniform throughout the mismatched tract, and the pattern of reactivity is similar to that described for the multiple T.T mismatches. The modification pattern was somewhat asymmetric, in that the lower (mp9) strand was chemically more reactive than the top strand.

Thus the chemical reactivity of the bubbles was characterised by a uniform reactivity along the length of the mismatched tract, but tightly localised to the mismatched bases. The strong, uniform reactivity of the multiple mismatches contrasts with single mismatches, which exhibit variable reactivity depending on the nature of the mismatch, and its context [14, 51].

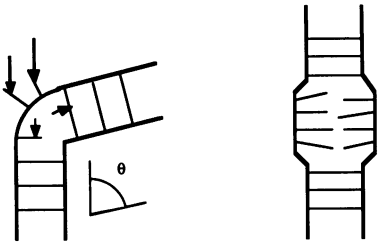


Figure 10. Schematic summary of the deduced structures of bulges and bubbles in DNA. The bulges introduce a marked bend into the DNA molecule, the magnitude of which (θ) depends on the number of bases in the bulge. The bases of the bulge exhibit pronounced chemical reactivity, that is non-uniform (the relative reactivities are indicated by the arrows). The 5′ base of the bulge is protected against chemical attack in the case of adenine and thymine bulges, possibly indicating that it is stacked on the basepair of the adjoining duplex, and the thymine of the first basepair on the opposite duplex is also reactive. By contrast, the bubbles exhibit very much small gel retardation, indicating less bending of the DNA axis, and a uniform chemical reactivity along the multiply mismatched region.

DISCUSSION

We have studied two kinds of local distortion in double stranded DNA, caused by the addition of unpaired bases in the centre of a fully paired DNA fragment. Where the bases are added on one strand only, we call these bulges. If an equal number of non-complementary bases are present on the other strand, this is called a multiple mismatch, or bubble. There are a number of differences between the properties of DNA bulges and bubbles. We have observed that:

1. Base bulges cause a very significant retardation in gel electrophoresis, akin to curved DNA fragments, or those containing a four-way helical junction. The magnitude of the effect depends on the number of unpaired bases in the bulge. By contrast, bubbles of equivalent sizes give rise to very much smaller effects in gel electrophoresis. We conclude that there is a much greater distortion of the helix axis in the case of a bulge, leading to a larger change in helical trajectory of the DNA molecule.
2. Both bulges and bubbles are cleaved by single-strand specific nucleases, and modified by single-strand selective chemical reagents. However, while osmium tetroxide and DEP react with the bases of bubbles in a uniform manner, this is not the case for the bulges. For example, the 5′ base of both A_5 and T_5 bulges exhibit lower levels of reactivity than the remaining unpaired bases.

These differences indicate that bulges and bubbles have contrasting structures in solution. The results may be rationalised in terms of the outline models presented in Figure 10. We suggest that the bulges have a rather precise structure forced upon them by the need to accommodate the extra nucleotides on one strand. This results in a well defined change in the direction of the helix axis, the angle of which depends on the number of extra bases in the loop formed. Thus the gel retardation should increase as a function of the number of bases in this loop, as observed experimentally. Using a linker-ligation assay, Hsieh and Griffith [34] have also noted bulge-size-dependent gel retardation. In the case of single adenine [8−10] and guanine [7] bulges, NMR studies have indicated that the additional

purine stacks into the helix, and causes a local kink in the double helix (estimated to be 18−23° for the guanine bulge, using a combination of NMR and molecular mechanics), although other studies of single-base bulges have found extra-helical conformations [11−13]. The variation in chemical reactivity between different bases in the loop indicates that the single-stranded DNA is at least partially structured, leading to steric protection of some bases against chemical attack. The lower reactivity of bases (both adenine and thymine) at the 5′ end of the loops suggests that these bases may well stack on the end of the duplex. It is interesting to note that thymines of terminal basepairs of the duplex at the 3′ end of the loop are moderately reactive, indicating that these bases are *not* protected against out-of-plane attack by the electrophile, and hence are not involved in stacking interactions with single-stranded bases of the loop. The conformation of the bases in the bulge other than at the 5′ end cannot be deduced from these data, but it seems likely that these will be at least partly structured, and that the detailed three-dimensional structure of the bulge will depend on the sequence and size of the bulge. We note that the extent of gel retardation depends on the nature of the bases in the loop (the A_5 loop appears to cause greater bending than the T_5 loop), the sequence context of the bulge, as well as the number of bases in the loop. Furthermore, we note that the bases of a T_5 bulge are both more reactive to osmium tetroxide and the backbone more readily cleaved by nucleases than for the T_3 bulge, suggesting that the environment of the bases in the bulges varies with size.

In contrast to the bulges, the bubbles a pear to have a more poorly defined, and probably looser, structure. Very much less sign ïcant gel retardation was associated with these species, indicating smaller distortion oí the helix axis. However, the bubbles exhibited sensitivity to single-strand specific enzymes, unlike single base mismatches (which were refractory to S1 and micrococcal nucleases [14], and pronounced chemical reactivity. The uniform chemical reactivity of the bases present in the bubbles suggests that there is a less precise structure, and the best description may be a flexible association of weak base associations. Such structures appear to be related to the DNA conformation adopted in the open complex of RNA polymerase and promoters [52].

The strongly contrasting structures of bulges and bubbles suggests that they may have very different consequences when they occur in single-stranded nucleic acids. The well defined structure of the bulges, in which some bases appear to be well exposed, make them probable targets for recognition by proteins, and thus they may be better repaired when generated in double-stranded DNA, and represent specific protein binding sites in structured single-stranded nucleic acids, especially RNA molecules.

ACKNOWLEDGMENTS

We thank Jack Griffith for provision of a pre-print of his studies of bulged sequences, Alastair Murchie for oligonucleotide synthesis and Alastair Murchie and Fareed Aboul-ela for discussions, and the Medical Research Council and the Welcome Trust for financial support.

*To whom correspondence should be addressed

NOTE ADDED

Since the preparation of this manuscript, two relevent papers concerning the gel

electrophoretic properties of DNA molecules that contain bulges have been published. These are:

Hsieh, C-H and Griffith, JD (1989) *Proc Natl Acad Sci USA* **86**, 4833–4837.
Rice, JA and Crothers, DE (1989) *Biochemistry* **28**, 4512–4516.

REFERENCES

1. Modrich, P (1987) *Ann Rev Biochem* **56**, 435–466.
2. Ripley, LS (1982) *Proc Natl Acad Sci USA* **79**, 4128–4132.
3. Kleff, S and Kemper, B (1988) *EMBO J* **7**, 1527–1535.
4. Romaniuk, PJ, Lowary, P, Wu, H-N, Stormo, G and Uhlenbeck, OC (1987) *Biochemistry* **26**, 1563–1568.
5. Kennard, O (1987) in *Nucleic Acids and Molecular Biology* [Eds Eckstein, F and Lilley, DMJ] Springer-Verlag pp 25–52.
6. Patel, D, Shapiro, L and Hare, D (1987) in *Nucleic Acids and Molecular Biology* [Eds Eckstein, F and Lilley, DMJ] Springer-Verlag pp 70–84.
7. Woodson, SA and Crothers, DE (1988) *Biochemistry* **27**, 3130–3141.
8. Patel, DJ, Kozlowski, SA, Marky, LA, Rice, JA, Broka, C, Itakura, K and Breslauer, KJ (1982) *Biochemistry* **21**, 445–451.
9. Hare, D, Shapiro, L and Patel, DJ (1986) *Biochemistry* **25**, 7456–7464.
10. Roy, S, Sklenar, V, Appella, E and Cohen, JS (1987) *Biopolymers* **26**, 2041–2052.
11. Joshua-Tor, L, Rabinovich, D, Hope, H, Frolow, F, Appella, E and Sussman, JL (1988) *Nature* **334**, 82–84.
12. Miller, M, Harrison, RW, Wlodawa, A, Appella, E and Sussman, JL (1988) *Nature* **334**, 85–86.
13. Morden, KM, Chu, YG, Martin, FH and Tinoco, I (1983) *Biochemistry* **22**, 5557–5563.
14. Bhattacharyya, A and Lilley, DMJ (1989) *J Molec Biol* In the press.
15. Marini, JC, Levene, SD, Crothers, DM and Englund, PT (1982) *Proc Natl Acad Sci USA* **79**, 7664–7668.
16. Diekmann, S and Wang, JC (1985) *J Molec Biol* **186**, 1–11.
17. Hagerman, PJ (1985) *Biochemistry* **24**, 7033–7037.
18. Wu, H-M and Crothers, DM (1984) *Nature* **308**, 509–513.
19. Koo, H-S, Wu, H-M and Crothers, DE (1986) *Nature* **320**, 501–506.
20. Beaucage, SL and Caruthers, MH (1981) *Tetrahedron Lett.* **22**, 1859–1862.
21. Sinha, ND, Biernat, J, McManus, J and Köster, H (1984) *Nucleic Acids Res* **12**, 4539–4557.
22. Twigg, AJ and Sherratt, D (1980) *Nature* **283**, 216–218.
23. Maxam, AM and Gilbert, W (1980) *Meth. Enzymol.* **65**, 499–560.
24. Messing, J and Vieira, J (1982) *Gene* **19**, 269–276.
25. Gough, GW and Lilley, DMJ (1985) *Nature* **313**, 154–156.
26. Berkner, KL and Fold, WR (1977) *J Biol Chem* **252**, 3176–3184.
27. Lilley, DMJ and Palecek, E (1984) *EMBO J.* **3**, 1187–1192.
28. Johnston, BH and Rich, A (1985) *Cell* **42**, 713–724.
29. McClellan, JA, Palecek, E and Lilley, DMJ (1986) *Nucleic Acids Res.* **14**, 9291–9309.
30. Herr, W (1985) *Proc Natl Acad Sci USA* **82**, 8009–8013.
31. Furlong, JC and Lilley, DMJ (1986) *Nucleic Acids Res.* **14**, 3995–4007.
32. Scholten, PM and Nordheim, A (1986) *Nucleic Acids Res* **14**, 3981–3993.
33. Sanger, F, Nicklen, S and Coulson, AR (1977) *Proc Natl Acad Sci USA* **74**, 5463–5467.
34. Hsieh, C-H and Griffith, JD (1989) *Proc Natl Acad Sci USA* In the press.
35. Kline, SW, Yarus, M and Wier, P (1986) *DNA* **5**, 37–51.
36. Lilley,DMJ (1980) *Proc.Natl.Acad.Sci. USA* **77**, 6468–6472
37. Panayotatos,N and Wells,RD (1981) *Nature* **289**, 466–470
38. Mace, HAF, Pelham, HRB and Travers, AA (1983) *Nature* **304**, 555–557.
39. Lyamichev, VI, Mirkin, SM and Frank-Kamenetskii, MD (1985) *J. Biomolec. Structure and Dynamics* **3**, 327–338
40. Pulleyblank, DE, Haniford, DB and Morgan, AR (1985) *Cell* **42**, 271–280
41. Dingwall, C, Lomonosoff, GP and Laskey, RA (1981) *Nucleic Acids Res* **9**, 2659–2673.
42. Lilley, DMJ (1983) *Cold Spring Harbor Symp Quant Biol* **47**, 101–112.
43. Myers, RM, Lumelsky, N, Lerman, LS and Maniatis, T (1985) *Nature* **313**, 495–498.
44. Dodgson, JB and Wells, RD (1977) *Biochemistry* **16**, 2367–2374.
45. Pardi, A, Morden, KM, Patel, D and Tinoco, I (1982) *Biochemistry* **21**, 6567–6574.

46. Leonard, NJ, McDonald, JJ, Henderson, REL and Reichmann, A (1971) *Biochemistry* **10**, 3335–3342.
47. Vincze, A, Henderson, REL, McDonald, JJ and Leonard, NJ (1973) *J Amer Chem Soc* **95**, 2677–2682.
48. Cornelis, AG, Haasnoot, JHJ, den Hartog, JF, de Rooij, M, van Boom, JH and Cornelis, A (1979) *Nature* **281**, 235–236.
49. Arnold, FH, Wolk, S, Cruz, P and Tinoco, I Jr (1987) *Biochemistry* **26**, 4068–4075.
50. Kouchakdjian, M, Li, BFL, Swann, PF and Patel, DJ (1988) *J Molec Biol* **202**, 139–155.
51. Cotton, RGH, Rodrigues, NR and Campbell, RD (1988) *Proc Natl Acad Sci USA* **85**, 4397–4401.
52. Buckle, M and Buc, H (1989) *Biochemistry* **28**, 4388–4396.

This article, submitted on disc, has been automatically converted into this typeset format by the publisher.