The EMBO Journal vol.5 no.1 pp.191–196, 1986

# The structure of cruciforms in supercoiled DNA: probing the single-stranded character of nucleotide bases with bisulphite

**Gerald W.Gough, Karen M.Sullivan and David M.J.Lilley**

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

Communicated by D.M.J.Lilley

**The single-stranded character of cytosine bases in three cruciform structures has been assessed by an examination of reactivity towards sodium bisulphite. Unpaired cytosine residues undergo deamination at C4 to give deoxyuracil, and propagation in an *ung Escherichia coli* host results in C-G → T-A transition mutations, detectable by restriction cleavage or sequence analysis. Very high frequencies of such mutations have been found at cruciform loops, confirming their unpaired character, with almost zero background mutation frequencies elsewhere. A low level of modification was observed at the four-way junction of a cruciform. The results indicate that the optimal cruciform loop size is four bases, with loose 'breathing' at the first base pair at the top of the cruciform stem at 37°C, and little or no opening of base pairs at the four-way junction.**

*Key words:* cruciforms/cytosine/single-stranded character/sodium bisulphite

## Introduction

DNA sequences which possess 2-fold symmetry may re-organise their base pairing to form cruciform structures, in which there is local intra-strand hydrogen bonding. Conceptually this idea is fairly old (Platt, 1955; Gierer, 1966), although experimental evidence has become available only recently (Gellert *et al.*, 1979; Lilley, 1980; Panayotatos and Wells, 1981), since this required an appreciation of the role of DNA supercoiling in stabilising structural polymorphs.

There has been little attempt to analyse the molecular geometry of cruciform structures. This is due partly to the requirement for negative supercoiling, i.e., relatively large circular DNA molecules, which do not lend themselves readily to diffraction or spectroscopy. Analysis of cruciform structure divides into two problems – the structure of the unpaired loop, and that of the four-way junction. We have recently approached the latter by an examination of pseudo-cruciform fragments (Gough and Lilley, 1985). These constructs exhibit pronounced retardation of gel electrophoretic migration which is maximal when the junction is central, consistent with a geometry in which there is a pronounced bend at the cruciform junction.

Here we address a different aspect of the structural problem – the degree of unpairing of bases in the cruciform structure. There are two reasons for expecting some loss of base pairing in the loop of a cruciform. First, it is not possible to construct stereochemically acceptable models in which the base pairing is not disrupted, and such attempts suggest that at least four bases must be unpaired. Second, cruciform structures are cleaved by single strand-specific nucleases (Lilley, 1980; Panayotatos and Wells, 1981) and modified by single strand-selective chemicals (Lilley, 1983; Lilley and Paleček, 1984) at or near their symmetric centres. We may ask how many bases are required to form a cruciform loop. In the case of some inverted repeats this is effectively pre-determined by the sequence. For example, the ColE1 cruciform (Lilley, 1980) has a central AAATG sequence which is not part of the symmetry element, and which must, therefore, be unpaired in the cruciform. Even in this case, however, we cannot exclude the possibility that an extra base pair will be opened, to generate a loop of seven nucleotides. In the case of an inverted repeat which is perfect, the loop size cannot be predicted. It will form a loop of sufficient size, and no greater, to minimise the free energy of the structure. In this work, we aim to identify this optimum size. Whilst the cruciform loop is certainly unpaired to some extent, may the same be true of the four-way junction? This is much harder to predict with any certainty, and models may be constructed which entail disruption to hydrogen bonding, while others maintain full base pairing.

We have, therefore, set out to analyse the single strand character of nucleotide bases in a number of cruciform structures, with a resolution at the level of a single base. What is the probe of choice for such an investigation? We rejected nuclease probes mainly for reasons of steric interference. Enzymes are macromolecules, and it seems highly probable that bases close to the cruciform stem may be prevented from entry into the active site despite being unpaired. To support this contention, we note that single base resolution of S1 nuclease cleavage of the ColE1 cruciform detected three cleaved phosphodiester bonds (Lilley, 1981), whilst we know that at least five bases must be unpaired. Small molecule probes seem preferable, therefore. However, the chemical probes which we have employed hitherto, viz bromoacetaldehyde, osmium tetroxide and glyoxal (Lilley, 1985), suffer from two collective drawbacks. First, adduct formation has been detected by nuclease cleavage, reducing resolution. Second, chemical reaction at a particular base predisposes its neighbours to further reaction, and a 'bubble' of modification expands rapidly, driven by the resulting relaxation of supercoiling.

Sodium bisulphite suffers from neither of these problems. Cytosine bases of single strand character are deaminated to deoxyuracil via bisulphite addition across the C5-C6 double bond (Shapiro *et al.*, 1970, 1973; Hayatsu *et al.*, 1970; Goddard and Shulman, 1972; Hayatsu, 1976). The reaction is not self-propagating, and modification of a single base in one molecule is readily detected. Bisulphite has been used extensively as a probe of secondary structure in RNA molecules (Furuichi *et al.*, 1970; Kućan *et al.*, 1971; Chambers *et al.*, 1973; Singhal, 1974; Lowdon and Goddard, 1976; Goddard and Lowdon, 1978), but application to DNA structure has not been tried before to our knowledge. However, the method is similar in principle, if not in purpose, to that of D-loop and gap mutagenesis (Shortle and Nathans, 1978; Peden and Nathans, 1982). In our approach, supercoiled DNA is reacted with sodium bisulphite, and the modified DNA transformed into a repair-deficient *Escherichia coli* strain. Thus,

© IRL Press Limited, Oxford, England

the net effect of a base modification is a C→T transition. DNA from individual colonies of transformants may be analysed by restriction enzyme cleavage or DNA sequencing, and mutation events scored. In this way the single-stranded character of selected cytosine bases may be assessed in a precise manner.

The conclusions which we draw from these studies are that there is a very high frequency of modification of cytosine bases in the loop regions of a number of cruciforms. This is greatest at the central four nucleotides, is substantially lower for the next pair, and zero beyond this. Modification of cytosines at the four-way junction occurs at very low frequency – only two examples have been found.

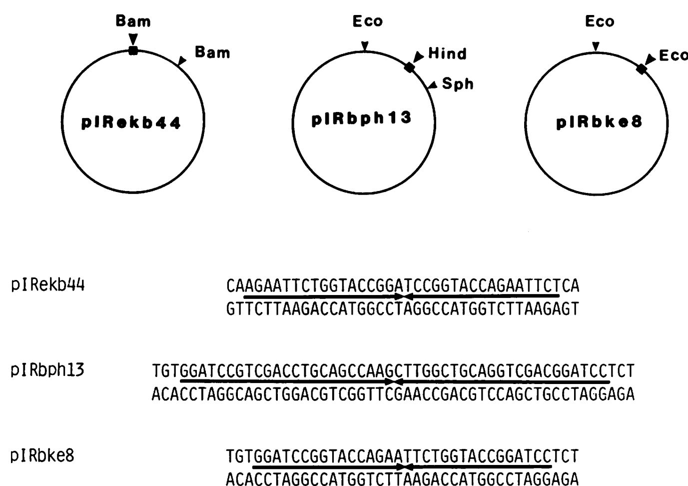


**Fig. 1.** Plasmid maps and inverted repeat base sequences of pIRekb44, pIRbph13 and pIRbke8. **Upper:** restriction maps of the plasmids employed in these studies. The locations of inverted repeats are indicated by filled boxes, whilst restriction enzyme target sequences are shown by arrowheads. **Lower:** the sequences of the inverted repeats present in these plasmids. The extent of 2-fold symmetry in each repeat is shown by opposing arrows.

## Results

### *Bisulphite mutagenesis selects a cruciform loop*

The plasmids and their inverted repeat sequences employed in these studies are shown in Figure 1. The plasmid used initially was pIRekb44, in which an inverted repeat was created at the *Eco*RI site of pAT153 by insertion of two synthetic oligonucleotide fragments. This inverted repeat is completely symmetrical, having a *Bam*HI target site at its centre, i.e., GGATCC. If this sequence adopts cruciform geometry it is highly probable that unpairing of bases within this recognition sequence will occur. pIRekb44 contains a second *Bam*HI target site, 375 bp removed from the potential cruciform, which should exist as DNA of unperturbed geometry.

Supercoiled pIRekb44 was incubated with bisulphite for 24 h at 37°C. Competent *E. coli* BD1528 cells were transformed with this DNA and spread on agar to give colonies. BD1528 cells are *ung* (Duncan *et al.*, 1978), and are unable to repair deoxyuracil which is the product of cytosine deamination. On replication of the plasmid in the repair-deficient host strain there is therefore a good probability that a C-G base pair which has been attacked by bisulphite will become mutated to T-A. Of course, if any of the four C-G base pairs of the *Bam*HI site becomes mutated in this way, the sequence will no longer be recognised and cleaved by the restriction enzyme *Bam*HI, and this provides a rather quick and easy screen for such modification reactions.

Figure 2 shows *Bam*HI restriction cleavage analysis of DNA mini-preparations from single colonies after bisulphite treatment. Loss of a single *Bam*HI target site is easily recognised as this leads to the formation of a full-length linear pIRekb44 molecule, while cleavage at both sites gives a molecule which is shorter by 388 bp. It is readily seen that approximately half of the colonies contain plasmid DNA which has lost one *Bam*HI target sequence. Analysis of 31 colonies from such experiments revealed a mutation frequency within the *Bam*HI recognition sequence of

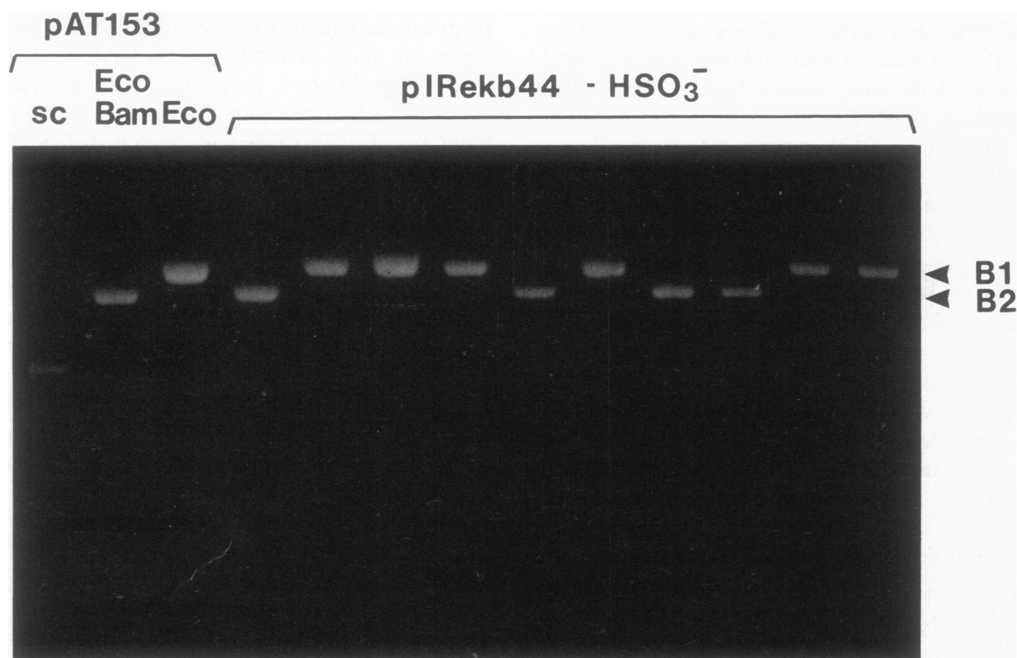


**Fig. 2.** Restriction enzyme cleavage analysis of plasmid DNA prepared from colonies of *E. coli* transformed with bisulphite-reacted pIRekb44. Plasmid DNA was prepared by standard mini-preparation methods, and cleaved to completion with *Bam*HI. If the plasmid contains a single *Bam*HI target, a full-length linear species of 3683 bp is produced (band B1), whereas if two *Bam*HI sites are retained this is shortened by 388 bp to generate a 3295-bp fragment (band B2). The first three tracks contain pAT153 digests for comparison of fragment lengths – sc, supercoiled pAT153, and pAT153 digested with *Eco*RI + *Bam*HI (3282 bp) or *Eco*RI alone (3657 bp).

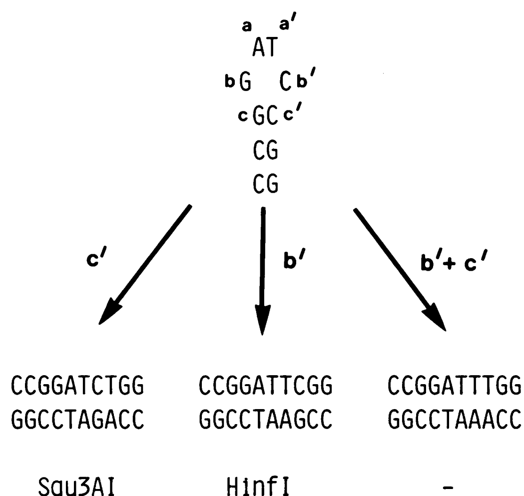


**Fig. 3.** Mutation of cytosine residues at the loop of the pIRekb44 cruciform may lead to a new restriction enzyme site. We have employed the notation shown for a symmetrical cruciform loop, where the inner bases are denoted a and a′, the next bases b and b′, and so on outwards. A single b′ mutation leads to the formation of a *Hinf*I site, whilst a double b′ + c′ mutation renders the sequence refractory to all restriction enzymes. A single c′ mutation retains the central GATC sequence, which is therefore cleavable by *Sau*3AI. Note that none of these sequences can be cleaved by *Bam*HI.

**Table I.** Analysis of DNA from 12 transformant colonies of bisulphite-reacted pIRekb44, pre-selected to have lost one *Bam*HI site

| Colony | 1 | 2 | 3 | 11 | 15 | 18 | 21 | 23 | 25 | 28 | 29 | 31 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Bam*HI (13) | – | – | – | – | – | – | – | – | – | – | – | – |
| *Bam*HI (388) | + | + | + | + | + | + | + | + | + | + | + | + |
| *Hinf*I | – | – | + | + | + | – | – | + | + | + | – | – |
| *Sau*3AI | – | – | – | – | – | – | – | – | – | – | – | – |

*Bam*HI (13) lies at the centre of the cruciform, whilst *Bam*HI (388) is well removed from the cruciform-forming sequence. Successful restriction enzyme cleavage is indicated +, while failure to cleave, indicating a mutated recognition sequence, is indicated –.

42% (13/31). To find which *Bam*HI site was mutated each DNA preparation was subjected to double digestion with *Bam*HI + *Sal*I, with the result that the *Bam*HI removed by mutation was invariably that at the inverted repeat.

Clearly the *Bam*HI target sequence of the inverted repeat of pIRekb44 shows enhanced reactivity towards bisulphite-induced deamination of cytosine. This is to be expected if this sequence forms the unpaired loops of a cruciform structure extruded from the inverted repeat.

*Modification frequencies at individual cytosines within the inverted repeat of pIRekb44*

For a detailed analysis of the positions of chemical reactions at the inverted repeat of pIRekb44, two approaches have been taken. The easier of the two, although relatively limited in scope, was to use further restriction enzyme analysis, whilst a complete analysis required DNA sequencing.

In the first approach, deductions about which base(s) within the GGATCC sequence had mutated could be made by challenging with additional restriction enzymes as indicated in Figure 3. If one of the inner GC base pairs has mutated, to give GAATCC for example, then this sequence now contains the recognition target for *Hinf*I. If both are mutated the sequence becomes GAATTC, an *Eco*RI target site, although we would not expect

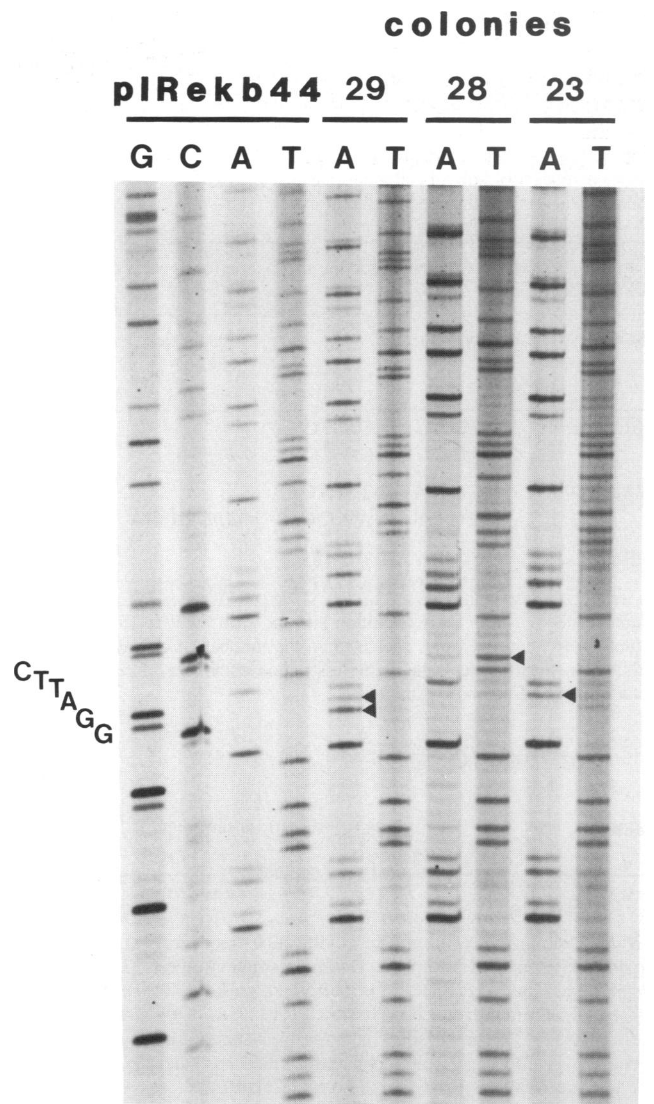


**Fig. 4.** Chain termination sequencing of representative bisulphite-reacted pIRekb44 clones. The left four tracks are the four base reactions on modified pIRekb44, whilst the remaining tracks are A and T reactions on clones 29, 28 and 23, respectively. The bands corresponding to the central *Bam*HI recognition sequence GGATCC are indicated on the left of the autoradiograph. New bands, corresponding to bisulphite-induced C-G → T-A mutations, are shown by the small arrowheads between tracks.

to observe such mutants as this would require mutations on both strands to be propagated, which is not possible. More interestingly, we may ask how frequently the outer GC base pairs become mutated. If one becomes modified, leaving the central tetranucleotide unchanged, then the sequence cannot be cleaved by *Bam*HI, but will remain a target site for *Sau*3AI. The results of screening 12 mutants which had lost the *Bam*HI site are presented in Table I. Half of these mutations resulted in the formation of *Hinf*I cleavage sites, but we observed no cases in which the *Bam*HI sequence was mutated to a *Sau*3AI sequence. We conclude from these results that whilst single mutations may readily occur at the inner cytosines, deamination of the outer bases is either infrequent or is usually accompanied by a reaction at the inner base in addition. This suggests that in the cruciform loop of pIRekb44 the b,b′ positions are more reactive towards bisulphite than the c,c′ positions (see Figure 2).

**Fig. 5.** C-G → T-A mutation event observed by sequencing through the inverted repeats of some bisulphite-modified pIRekb44 clones. Each filled square indicates an observed mutation event.

**Table II.** Mutation frequencies for the three cruciforms with different central (loop) sequences

| Plasmid | Loop sequence | Restriction site | Mutation frequency (%) |
|---|---|---|---|
| pIRekb44 | ACCGGATCCGGT<br>TGGCCTAGGCCA | *Bam*HI | 42 |
| pIRbph13 | GCCAAGCTTGGC<br>CGGTTCGAACCG | *Hin*dIII | 57 |
| pIRbke8 | CCAGAATTCTGG<br>GGTCTTAAGACC | *Eco*RI | 8 |

Each sequence forms the recognition site for a different restriction enzyme. Mutation frequencies were calculated as the number of colonies containing DNA refractory to cleavage by the restriction enzyme as a percentage of the total number of colonies examined.

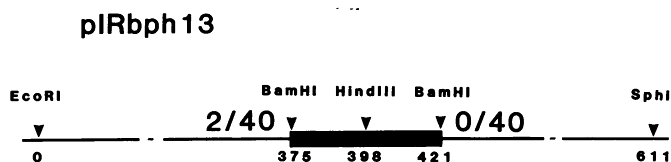


**Fig. 6.** Examination of bisulphite modification frequencies at the four-way cruciform junction of pIRbph13. Double restriction digests (*Eco*RI + *Bam*HI and *Sph*I + *Bam*HI) were performed on bisulphite-reacted pIRbph13 clones to determine if the *Bam*HI sites at the bases of the cruciform arms were mutated. The observation mutation frequencies are shown in large type. The inverted repeat is drawn as the filled box, and the numbers below the line are distances (base pairs) from the *Eco*RI site.

The restriction cleavage data are incomplete in two respects. First, their interpretation cannot be completely free of ambiguity in every case, and second, they can give no indication of whether reaction is possible outside the central hexanucleotide recognition sequence. We therefore chose to derive the complete sequences of the inverted repeat carrying regions of a number of the mutants which had lost the *Bam*HI target site. This was accomplished by designing and synthesising a 15-nucleotide primer which was used to sequence the plasmid DNA directly by the dideoxy method (Sanger *et al.*, 1977; Wallace *et al.*, 1981). A sequencing gel of some representative bisulphite-modified pIRekb44 clones is presented in Figure 4. It is apparent that mutation events may be scored with complete confidence. The results are presented in Figure 5 as a histogram, in which each box represents an observed mutation event. Two points are immediately apparent. First, there are no mutations seen outside the central hexanucleotide. Second, there is a rather greater frequency of modification occurring at positions b,b′ than at c,c′, i.e., the reactivity increases towards the centre of the loop. In all four cases we have observed double mutations occurring, i.e., simultaneous mutations in bb′ and cc′. In every case the results of restriction enzyme analysis and DNA sequencing agree precisely.

*Modification at different cruciform loops*

The results obtained with pIRekb44 suggest that modification frequencies depend upon the proximity of the cytosine in question to the symmetric centre of the inverted repeat. We therefore investigated two further plasmids containing inverted repeats of similar size to that of pIRekb44, but whose central sequences were AAGCTT (i.e., *Hin*dIII) and GAATTC (i.e., *Eco*RI). pIRbph13 has a potential loop in which the vulnerable cytosine bases are in the a′ position, i.e., maximally central. By contrast, the cytosines of the potential loop in pIRbke8 are in the c′ position. The results obtained with these two plasmids, together with those of pIRekb44, are presented in Table II. They show that the mutation frequency in pIRbph13 (i.e., a′) is even higher than that for pIRekb44 (b′ + c′), whilst that for pIRbke8 (c′ alone) is considerably reduced relative to the others. This reinforces the conclusion that chemical reactivity is higher towards the symmetrical centre of the loop.

*Probing the cruciform four-way junction*

Another region of the cruciform potentially containing unpaired bases is the four-way junction; the interface between the cruciform stems and the main body of the plasmid DNA. The sequencing data discussed above had indicated that the junction sequences were not modified at high frequency, but compression was present on some sequence gels, and we could not conclude with confidence that the junctions were completely refractory to bisulphite attack. Moreover, the number of colonies which could be screened by this method was severely limited. We therefore employed restriction enzyme screening at the junction *Bam*HI sites of pIRbph13. In this way we could probe the terminal C-G base pairs of each cruciform stem. Double digests were performed, i.e., *Eco*RI + *Bam*HI and *Sph*I + *Bam*HI, to examine the right- and left-hand ends, respectively, of the inverted repeat in 40 bisulphite-reacted species. The results are summarised in Figure 6. Modification frequencies are very low, at least an order of magnitude lower than those at loops, but are not zero, however, in that two mutants were detected in which the left-hand *Bam*HI target sequence had become altered. The results suggest that cytosines in the four-way junction are of low reactivity towards bisulphite, a conclusion which has been reinforced by studies on the ColE1 cruciform (K.M.Sullivan, unpublished data).

## Discussion

From the results above it is clear that bisulphite mutagenesis is an excellent probe of perturbed DNA structure. Cytosine deamination was observed at high frequencies in cruciform loops, but background frequencies elsewhere in these molecules were undetectably low. Moreover, the resolution of these experiments is far superior to other chemical probes which we have used previously (Lilley, 1983; Lilley and Paleček, 1984), giving information about the character of single bases. Whilst we have applied this approach successfully to cruciform structures, it could certainly be extended to other structural features, such as the B-Z junction (Singleton *et al.*, 1984).

To derive structural conclusions from these results we must make the assumption that mutation frequency is a direct consequence of the single-strand character of cytosine bases, but is this a correct assumption? No reaction has been detected on fully double-stranded polynucleotides using chemical (Goddard and Schulman, 1972; Shapiro *et al.*, 1973) or genetic (Shortle and Nathans, 1978) analysis. What is the origin of the single-strand selectivity? There are probably at least three components to this. First, initial attack by bisulphite upon cytosine will require an out-of-plane addition, which will be hindered when the base is

stacked into a double-helical region. Second, the adduct formed has a bulky $SO_3^-$ group on C6, which would distort the structure of a DNA helix. Third, this adduct undergoes deamination at C4, mediated by general base catalysis by sulphite, followed by the elimination of $HSO_3^-$ to generate deoxyuracil. Since the exocyclic amino group participates in hydrogen bonding to guanine, this reaction will be facilitated in single-stranded DNA. Could these reactions take place at a C-G base pair located at one end of a double helical segment? It seems feasible that reversible formation of 5,6-dihydrocytosine-6-sulphonate may occur at such a location, but the second-stage conversion of deoxyuracil appears improbable. Comparison with studies on tRNA (Furuichi *et al.*, 1970; Kućan *et al.*, 1971; Goddard and Schulman, 1972; Chambers *et al.*, 1973; Singhal, 1974; Lowdon and Goddard, 1976; Goddard and Lowdon, 1978) is useful, where there is a known secondary and tertiary structure (Kim *et al.*, 1974; Robertus *et al.*, 1974). Goddard and Lowdon (1978) studied the reaction of cytosines in *E. coli* $tRNA^{Phe}$ as a function of temperature. At 25°C they observed high-frequency modification of cytosines in the loop regions, and no detectable reaction with base-paired cytosine, even when present at the ends of helices. However, at 35°C these helix-terminal positions became modified to ~20%. These results suggest that in RNA at least, the reaction is extremely single-strand selective, irrespective of the position along the helix, but that the ends of the helices 'breathe' as the temperature is raised.

The present results indicate that there is considerable single-strand character at the centres of several inverted repeat sequences in supercoiled DNA. These are the regions which correspond to the loops of extruded cruciform structures, and so these data may be taken as further proof of the existence of cruciform structures, if this is needed. More detailed analysis of the results shows that the single-stranded character is greatest for the central four nucleotides (a, a', b and b'), is reduced, but nevertheless significant, for the next nucleotide pair (c and c'), but is undetectable beyond this point. This would be consistent with a loop size of four nucleotides, where the first base pair of the stem is loose, possibly in a dynamic equilibrium between open and closed states. Recently Hilbers and colleagues (1985) have studied a series of hairpin-forming DNA fragments of sequence $ATCCTAT_nTAGGAT$, where n was systematically varied between 0 and 7. Thermodynamic measurements indicated that the most stable loop contained four to five thymidine residues. We therefore find ourselves in excellent agreement with the physical data.

The low-frequency mutation events observed at the four-way junction are not consistent with models in which there is permanent unpairing of bases in this region. The results would be explained most reasonably in terms of a fully base-paired junction, with a low degree of opening due to thermal fluctuations.

How general are our conclusions? It is likely that the structure of loops will show some sequence-dependent variation. For example, S1 nuclease cleaves at >10 phosphodiester bonds at the centre of an $(AT)_{34}$ cruciform (Greaves *et al.*, 1985), whereas in the ColE1 cruciform the distribution is more tightly limited to three bonds (Lilley, 1981). It is possible, therefore, that A + T-rich loops may be looser than more G + C-rich ones. Similar considerations may apply to the geometry of the four-way junction. We have observed a very low frequency of bisulphite attack at the junction of pIRbhp13, but it is possible that this value could be higher or lower for different sequences. In broad terms, however, we feel that these studies have given some indications of which bases have greatest single strand character in three cruciforms, and should provide a basis for thinking about the three-dimensional structure of the cruciform.

## Materials and methods

### *Plasmids*

The construction of the plasmids used in these studies has been described previously (Lilley and Markham, 1983; Gough and Lilley, 1985). pIRekb44 and pIRbke8 were made by synthesising two 13-base oligonucleotides which hybridise together leaving cohesive *Eco*RI and *Bam*HI termini. After ligation the DNA was cut with either *Eco*RI or *Bam*HI and cloned into the corresponding site of pAT153, to generate pIRekb44 or pIRbke8, respectively. pIRbph13 contains a 46-bp inverted repeat created by self-ligation of the *Bam*HI to *Hin*dIII region of the polylinker of phage M13mp8, cloned into pAT153.

Supercoiled DNA was prepared from chloramphenicol-amplified *E. coli* by lysozyme, EDTA, SDS lysis and isopycnic centrifugation in caesium chloride, ethidium bromide gradients.

### *Sodium bisulphite modification reactions*

2–6 μg of supercoiled plasmid DNA were incubated for 24 h at 37°C in 1.4 M sodium metabisulphite, 0.84 M sodium sulphite, 2 mM hydroquinone, 3.5 mM NaCl and 0.35 mM Na citrate, pH 7.0. DNA was then recovered free of salts by passage down a Sephadex G-50 column and twice precipitated with ethanol. The DNA was redissolved in 10 mM Tris-HCl pH 7.5, 1 mM EDTA and incubated at 37°C for at least 9 h before precipitation with ethanol. The modified DNA was redissolved in sterile water and transformed into *E. coli* BD1528 (Duncan *et al.*, 1978) with selection for $Ap^r$.

### *Restriction analysis of DNA*

DNA from individual colonies of transformed BD1528 was prepared by the alkaline or heat treatment mini-preparation procedures (Birnboim and Doly, 1979; Holmes and Quigley, 1981). DNA was cleaved with restriction enzymes (Bethesda Research Laboratories), and analysed by electrophoresis on agarose or polyacrylamide gels.

### *Sequencing*

DNA required for sequencing was prepared from chloramphenicol-amplified cells, lysozyme, EDTA, SDS lysis and two rounds of caesium chloride, ethidium bromide centrifugation. DNA was sequenced by the dideoxy procedure (Sanger *et al.*, 1977), modified for application to plasmid DNA (Wallace *et al.*, 1981). 1 μg of each DNA to be sequenced was linearised by restriction cleavage, followed by addition of primer. The mixture was fully denatured at 100°C for 3 min and placed immediately on ice, followed by reannealing at 37°C for 30 min. After addition of deoxynucleoside triphosphates, the appropriate dideoxynucleotide and [$\alpha$-$^{35}$S]dATP, 1 unit of the Klenow fragment of DNA polymerase I was added to start the reaction. Terminated DNA chains were examined by electrophoresis on 8% polyacrylamide gels containing 7 M urea, and autoradiography. The primer employed in these studies was 5'GACGTCTAAGAAACC3', which anneals at a position 60 bp to the left of the *Eco*RI site on the conventional map of pBR322 (Sutcliffe, 1979). It was designed for minimal illicit hybridisation, and synthesised by phosphoramidite chemistry (Beaucage and Caruthers, 1981).

## Acknowledgements

We gratefully thank Dr John Goddard for discussion and helpful advice on performing bisulphite reactions, Dr Kees Hilbers for communicating results prior to publication and Dr Bruce Duncan for bacterial strains. We thank the Medical Research Council and the Royal Society for financial support.

## References

Beaucage,S.L. and Caruthers,M.H. (1981) *Tetrahedron Lett.*, **22**, 1859-1862.
Birnboim,H.C. and Doly,J. (1979) *Nucleic Acids Res.*, **7**, 1513-1523.
Chambers,R.W., Aoyagi,S., Furakawa,Y., Zawadska,H. and Bhanot,O.S. (1973) *J. Biol. Chem.*, **248**, 5549-5551.
Duncan,B.K., Rockstroh,P.A. and Warner,H.R. (1978) *J. Bacteriol.*, **134**, 1039-1045.
Furuichi,Y., Wataya,Y., Hayatsu,H. and Ukita,T. (1970) *Biochem. Biophys. Res. Commun.*, **41**, 1185-1191.
Gellert,M., Mizuuchi,K., O'Dea,M.H., Ohmore,H. and Tomizawa,J. (1979) *Cold Spring Harbor Symp. Quant. Biol.*, **43**, 35-40.
Gierer,A. (1966) *Nature*, **212**, 1480-1481.
Goddard,J.P. and Lowdon,M. (1978) *Eur. J. Biochem.*, **89**, 531-541.
Goddard,J.P. and Schulman,L.H. (1972) *J. Biol. Chem.*, **247**, 3864-3867.
Gough,G.W. and Lilley,D.M.J. (1985) *Nature*, **313**, 154-156.
Greaves,D.R., Patient,R.K. and Lilley,D.M.J. (1985) *J. Mol. Biol.*, **185**, 461-478.
Hayatsu,H. (1976) *Prog. Nucleic Acids Res. Mol. Biol.*, **16**, 75-124.
Hayatsu,H., Wataya,Y., Kai,K. and Kida,S. (1970) *Biochemistry*, **9**, 2858-2865.

Hilbers,C.W., Haasnoot,C.A.G., de Bruin,S.H., Joordens,J.J.M., Van der Marel, G.A. and Van Boom,J.H. (1985) *Biochimie*, in press.
Holmes,D.S. and Quigley,M. (1981) *Anal. Biochem.*, **114**, 193-197.
Kim,S-H., Suddath,F.L., Quigley,G.J., McPherson,A., Sussman,J.L., Wang, A.H-J., Seeman,N.C. and Rich,A. (1974) *Science*, **185**, 435-440.
Kućan,Z., Freude,K.A., Kućan,I. and Chambers,R.W. (1971) *Nature New Biol.*, **232**, 177-197.
Lilley,D.M.J. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 6468-6472.
Lilley,D.M.J. (1981) *Nucleic Acids Res.*, **39**, 1271-1289.
Lilley,D.M.J. (1983) *Nucleic Acids Res.*, **11**, 3097-3112.
Lilley,D.M.J. (1985) *IARC Scientific Publication* 70, *Nucleic Acid Cyclic Adducts and Carcinogenesis*, in press.
Lilley,D.M.J. and Markham,A.F. (1983) *EMBO J.*, **2**, 527-533.
Lilley,D.M.J. and Paleček,E. (1984) *EMBO J.*, **3**, 1187-1192.
Lowdon,M. and Goddard,J. (1976) *Nucleic Acids Res.*, **3**, 3383-3396.
Panayotatos,N. and Wells,R.D. (1981) *Nature*, **289**, 466-470.
Peden,K.W.C. and Nathans,D. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7214-7217.
Platt,J.R. (1955) *Proc. Natl. Acad. Sci. USA*, **41**, 181-183.
Robertus,J.D., Ladner,J.E., Finch,J.T., Rhodes,D., Brown,R.S., Clark,B.F.C. and Klug,A. (1974) *Nature*, **250**, 546-551.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Shapiro,R., Braverman,B., Louis,J.B. and Servis,R.E. (1973) *J. Biol. Chem.*, **248**, 4060-4064.
Shapiro,R., Servis,R.E. and Welcher,M. (1970) *J. Am. Chem. Soc.*, **92**, 422-424.
Shortle,D. and Nathans,D. (1978) *Proc. Natl. Acad. Sci. USA*, **75**, 2170-2174.
Singhal,R.P. (1974) *Biochemistry*, **13**, 2924-2932.
Singleton,C.K., Kilpatrick,M.W. and Wells,R.D. (1984) *J. Biol. Chem.*, **259**, 1963-1967.
Sutcliffe,J.G. (1979) *Cold Spring Harbor Symp. Quant. Biol.*, **43**, 77-90.
Wallace,R.B., Johnson,M.J., Suggs,S.V., Miyoshi,K-I., Bhatt,R. and Itakura,K. (1981) *Gene*, **16**, 21-26.

*Received on 18 October 1985*