Volume 9 Number 6 1981 Nucleic Acids Research

Hairpin-loop formation by inverted repeats in supercoiled DNA is a local and transmissible property

David M.J.Lilley

Department of Molecular Genetics, Searle Research Laboratories, Lane End Road, High Wycombe, HP12 4HL, UK

Received 21 January 1981

ABSTRACT

Short inverted repeat sequences adopt hairpin stem-loop type structures in supercoiled closed circular DNA molecules, demonstrated by S1 nuclease cleavage. Fine mapping of cleavage frequencies is in good agreement with expected cleavage patterns based upon the interaction between an unpaired loop and a sterically bulky enzyme molecule. Whilst the topological properties of underwound DNA circles depend ultimately upon reduced linkage, necessarily a global molecular property, hairpin loop formation is an essentially local property. Thus molecular size is unimportant for the S1 hypersensitivity of the ColE1 inverted repeat. Furthermore, a 440 bp Sau3AI, EcoRI fragment of ColE1 which contains the inverted repeat has been cloned into pBR322 whereupon it exhibits S1 cleavage similar to ColE1 in the supercoiled recombinant molecule. The effect is therefore both local and transmissible. Direct competition between inverted repeats in the recombinant, coupled with close examination of flanking sequences, enables some simple 'rules' for base pairing in hairpin loops to be formulated. Whilst limited G-T and A-C base pairing appears not to be destabilising, A-G, T-C or loop outs are highly destabilising.

INTRODUCTION

A striking feature of DNA B-form structure is its continuous symmetrical nature (1-3). However from a functional point of view it is equally apparent that discontinuities must be introduced into DNA structure in a sequence-dependent manner in order to delineate functionally significant loci such as promoters, replication origins, phasing signals and other putative sites of DNA-protein interaction. Recently it has become apparent that DNA may be rather more polymorphic and heterologous (4-6) than has previously been considered. Many fundamental genetic processes such as RNA polymerase initiation events, require localised loss of base pairing, and thus any potential mechanism providing specific opening of the double helix is of considerable importance.

I have demonstrated (7) that inverted repeats separated by short non-repetitious sections of DNA are specifically cleaved by the single strand

© IRL Press Limited, 1 Falconberg Court, London W1V 5FG, U.K.

specific nuclease S1 at or very close to the centre of symmetry, provided that the DNA is underwound in a negatively supercoiled structure. The simplest conclusion is that the inverted repeats adopt new secondary structures, formally described by paired hairpin loops, which on theoretical grounds should be stabilised by the free energy of DNA supercoiling. Whilst cruciform and similar secondary structures have been proposed for particular DNA sequences for many years (8), no direct evidence has previously been available.

In this paper I present more detailed evidence for the adoption of hairpin structure by the ColE1 inverted repeat, and demonstrate that the property is local and transmissible. Furthermore, by close examination of sequences flanking the inverted repeats, coupled with competition experiments in a recombinant plasmid, it becomes possible to propose a set of 'rules' for constructing alternative secondary structure in DNA.

EXPERIMENTAL PROCEDURES

Plasmids

ColE1, pDS1118, pVH51 and pBR322 were grown in E. coli K12 for 16 hrs under 150 μg/ml chloramphenicol amplification, followed by lysis and banding by caesium chloride-ethidium bromide isopycnic centrifugation (9). Plasmid bands were removed by side puncture, with minimal exposure to long wavelength UV light in order to minimise nicking of supercoiled circles. Ethidium bromide was removed by two n-butanol extractions and the plasmids ethanol precipitated.

Enzymes

S1 nuclease was a gift from Dr. J.S. Emtage, prepared from *Aspergillus oryzae* according to Vogt (10). Typical S1 reaction conditions are a 20 μl incubation containing 0.1-1 μg plasmid DNA, 1 u S1 (10), 30mM sodium acetate, pH 4.6, 50mM sodium chloride and 1mM zinc chloride. Reactions at 37°C are performed for 1-2 hr whilst low temperature incubations require times between 15 to 40 hr.

Restriction enzymes were obtained from either Bethesda Research Laboratories (BRL) or New England Biolabs.

T4 polynucleotide kinase was obtained from BRL, and DNA 5'-termini labelled according to Maxam and Gilbert (11). T4 DNA ligase and bacterial alkaline phosphatase was obtained from BRL.

Cloning Techniques

A *Sau*3AI, *Eco*RI digest of ColE1 was electrophoresed through a 5% poly-

acrylamide gel and the 440 bp band excised. DNA was eluted from the slice using extraction in 0.5M ammonium acetate, 10mM magnesium acetate, 0.1% SDS and 0.1mM EDTA (11) followed by binding to and elution from DEAE cellulose and ethanol precipitation. pBR322 was restricted with EcoRI and BamHI, and 5'-terminal phosphate residues removed using bacterial alkaline phosphatase. 10 ng of the ColE1,Sau3AI, EcoRI fragment were ligated to 100 ng of the phosphatase treated pBR322, BamHI, EcoRI using 0.1 u T4 ligase in 50mM Tris, pH 7.5, 10mM $MgCl_2$, 20mM DTT and 1mM ATP for 4 hr at 15°C. The ligated molecules were used to transform $CaCl_2$ treated E. coli K12 HB101 cells (12) and plated onto agar containing 100 µg/ml carbenicillin (Pyopen). Colonies were picked onto agar plates containing either 100 µg/ml carbenicillin or 10 µg/ml tetracycline (Sigma), and Ap^rTc^s colonies selected for further analysis. Plasmid was prepared from 0.5 ml overnight incubations using a rapid alkaline extraction procedure (13) coupled with restriction and gel electrophoresis to identify the required recombinants. Pure plasmid was then prepared as above.

Gel Electrophoresis

DNA samples were electrophoresed in 1% or 1.4% agarose (14), 5% polyacrylamide (15) or 8% polyacrylamide sequencing gels (11).

RESULTS

S1 cleavage distribution at the ColE1 inverted repeat

In a previous paper (7) I have demonstrated that restriction of single strand specific nuclease S1 cleaved supercoiled ColE1, pBR322 and ϕX174 RF DNA results in the generation of specific fragments, corresponding to a highly selective cleavage by the enzyme S1. Since S1 is extremely specific for single stranded substrates it was concluded that local unpairing was occurring at inverted repeat sequences. In order to study the cleavage at a specific repeat, a nearby restriction site is required, together with a gel electrophoresis system capable of resolving fragments differing in length by single nucleotides. The ColE1 repeat lies about 98 bp on the origin side of the unique EcoRI site (7), and fragments of this size range are well resolved using standard sequencing gels (11). Thus ColE1 was S1 cleaved at 7°C and the resulting 5' ends labelled using γ-^{32}P ATP and T4 polynucleotide kinase (11). After subsequent EcoRI restriction cleavage and denaturation, the labelled fragments were examined on an 8% sequencing gel, shown in figure 1. It is evident that whilst the cleavage appeared highly specific on gels of lower resolution (7), we can now see that the major cleavage position is

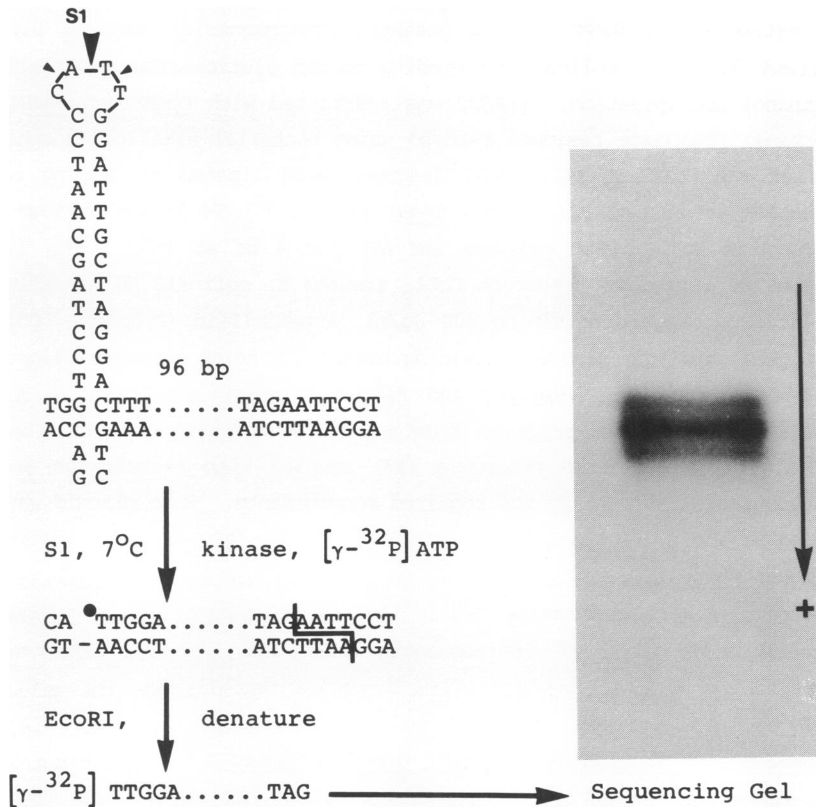


Figure 1

Cleavage frequency distribution by S1 nuclease at the inverted repeat of ColE1. The scheme shows the basic principle of the experiment, the sequence of ColE1 was obtained from Patient (45); ColE1 was digested with S1 nuclease at low temperature for 24 h and the 5'-termini generated were radioactively labelled with ^{32}P. After subsequent cleavage with EcoRI the ~100 nucleotide labelled fragments were examined on a sequencing gel, the autoradiograph of which is shown on the right.

flanked symmetrically by weaker cleavages. These cuts correspond precisely to the central and surrounding P-O bonds of the T-T-A-C loop. Scanning the autoradiograph followed by integration gives relative cutting frequencies of 1:2:1.

Thus the observed cleavages are well in accord with expected interactions between a relatively bulky enzyme and a symmetrical loop/stem structure. The

most accessible bond is at the centre of symmetry, with decreasing accessibility towards the stem. Hairpin formation is stabilised by negative supercoiling, since this corresponds to a considerable twist reduction at an enthalpic net expense of a small number of hydrogen bonds (7). Theoretical considerations are therefore totally in agreement with the experimental data in support of the existence of hairpin-loop structures, or their fourstranded equivalents (16,17).

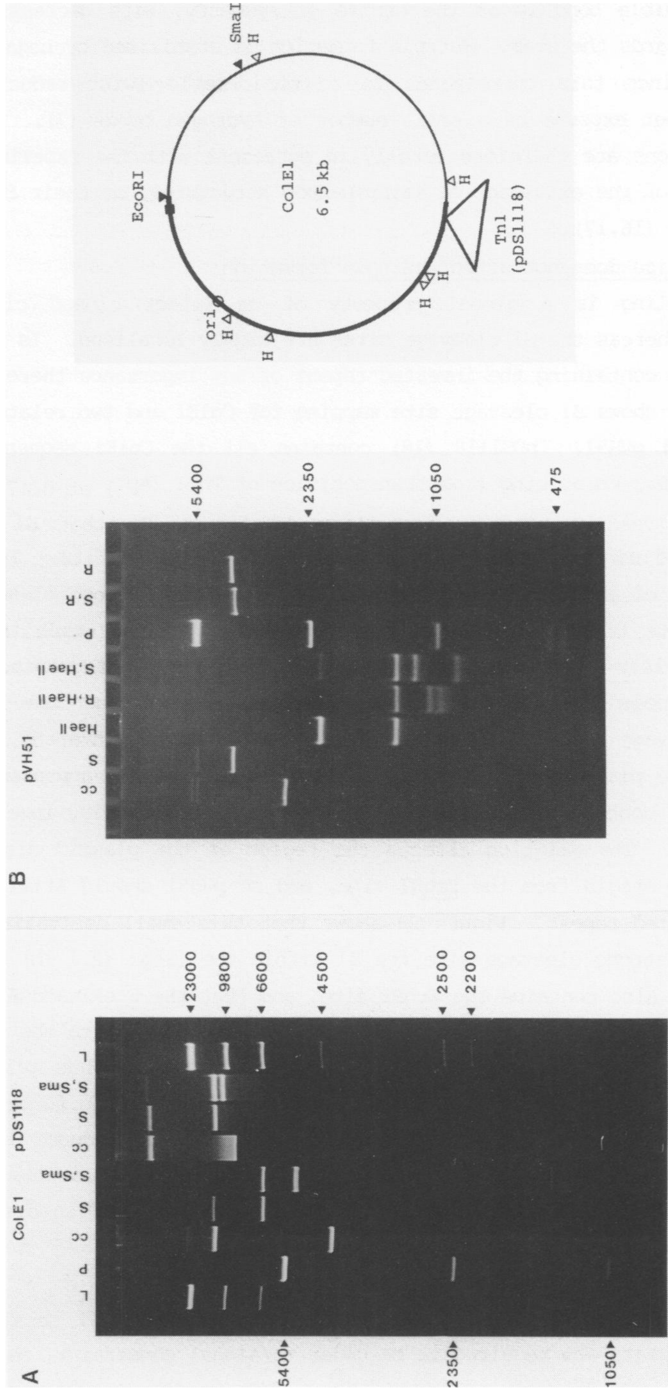
Molecular size does not affect hairpin formation

Supercoiling is a global property of an intact closed circular DNA molecule, whereas the S1 cleavage sites are highly localised. Is the size of the species containing the inverted repeat of any importance therefore?

Figure 2 shows S1 cleavage site mapping for ColE1 and two related plasmids pDS1118 and pVH51. pDS1118 (18) contains all the ColE1 sequence plus an additional 5.3 kb arising from transposition of Tn 1 (Ap^r) at 0.47 map units. It can be readily seen by inspection of figure 2A, that S1 introduces cleavages of similar selectivity into both ColE1 and pDS1118. Furthermore, measurement of longer fragment lengths indicate that in both plasmids the S1 cleavage site is 1170 ± 40 bp from the unique SmaI site, which is confirmed by the equality in length of the weakly visible 1.2 kb fragments. Thus the extra transposed DNA appears to affect neither the site nor the selectivity of the cleavage reaction, i.e. hairpin formation by the inverted repeat. By contrast the plasmid pVH51 (19) is a 'mini-ColE1', having approximately half the genetic content of ColE1 plus a small region of ϕ80pt190 inserted at the EcoRI site. The deletion affects the region of the plasmid distal to the replication origin from the EcoRI site, and so pVH51 should still retain the ColE1 inverted repeat. Figure 2B shows that this small derivative of ColE1 contains a strong cleavage site for S1 within the large (2.1 kb) HaeII fragment, which also contains the EcoRI site, and that the EcoRI and S1 sites are separated by 100 ± 5 bp. This is in excellent agreement with the S1 cleavage position at the centre of the ColE1 inverted repeat. Thus plasmid size, covering the range 4.0 to 11.8 kb, in no way affects hairpin formation by the ColE1 inverted repeat, the molecules presumably being reduced to the same negative superhelix density by gyrase. I have observed similarly (unpublished results) that the S1 sites of pBR322 are retained in its partially deleted counterpart pAT153 (20).

Molecular cloning of the ColE1 inverted repeat

The above results begin to suggest that whilst for S1 sensitivity the inverted repeats are required to be under torsional constraint, necessarily a



A
ColE1
pDS1118
L
P
cc
S
S,Sma
23000
9800
6600
4500
2500
2200
5400
2350
1050
B
pVH51
Hae II
R,Hae II
S,Hae II
S,R
R
475
EcoRI
SmaI
ColE1
6.5 kb
ori
H
Tn1
(pDS1118)

global property of the molecule, the essential sequence information is probably local. To test this in greater detail I decided to clone a small fragment of ColE1 into pBR322 to see if the repeat would retain its hairpin properties in its new location. There was a second reason for this experiment. As outlined in discussion it appears that large repeats to some extent 'overpower' smaller ones. pBR322 derives its replication functions from pMB1, which is extensively homologous with ColE1 (21), and the major S1 sequence in pBR322 should also be present in ColE1, although no cleavage is observed. This is slightly complicated, however, since in the sequence of the mini-ColE1 pAO2 by Oka et al. (22), the 11 bp inverted repeat, whilst present, shows one nucleotide difference in the stem from the equivalent structure in pBR322 (23). It is therefore of great interest to transfer the ColE1 repeat to pBR322, and monitor the cleavage of the pBR322 major repeat in the recombinant plasmid.

It was noted from the sequence data for ColE1 (22) that a double digest by Sau3AI and EcoRI should yield a 440 bp fragment which would contain the ColE1 inverted repeat. This is experimentally confirmed in figure 3 which shows Sau3AI and Sau3A1, EcoRI digests of ColE1 DNA. The 660 bp ColE1 Sau3AI band is cleaved by EcoRI to give bands of 440 and 220 ± 10 bp. It may be seen that the 440 bp band is in a region of non-ambiguity and must therefore be the fragment containing the inverted repeat. The cloning scheme I selected is outlined in figure 4. The 440 bp ColE1, Sau3AI, EcoRI band was excised from

Figure 2

S1 cleavage site selection in ColE1, pDS1118 and pVH51. The scheme indicates the relationship between these three plasmids. pDS1118 was derived (18) from ColE1 by transposition of Tn1 at 0.47 ± 0.03 map units, whilst the deletant pVH51 (19) retains approximately half of ColE1, shown as a heavier line. The open triangles, marked H, are the sites for HaeII; the filled square box in this and later figures indicates the inverted repeat.

A. S1 cleavage of ColE1 and pDS1118, shown by 1% agarose gel electrophoresis of the supercoiled plasmids, S1 linearised plasmids, and subsequently SmaI restricted DNA. Note that in these and other experiments, the order in which enzyme reactions are performed is important (7).

B. S1 cleavage of pVH51. 1.4% agarose gel electrophoresis of supercoiled plasmid, plasmid cut by HaeII, EcoRI, and S1, and double digests with S1 and EcoRI, S1 and HaeII, and EcoRI and HaeII. Note that the longest HaeII fragment (2.1 kb) is cut by EcoRI to give 1108 and 1022 bp fragments, whilst prior S1 cleavage results in fragments of 1212 and 925 bp.
Key: cc - covalently closed supercoiled plasmid, S - S1 nuclease treated, R - EcoRI cleaved, P - phage PM2 cleaved by HindIII, L - phage lamda cleaved by HindIII.

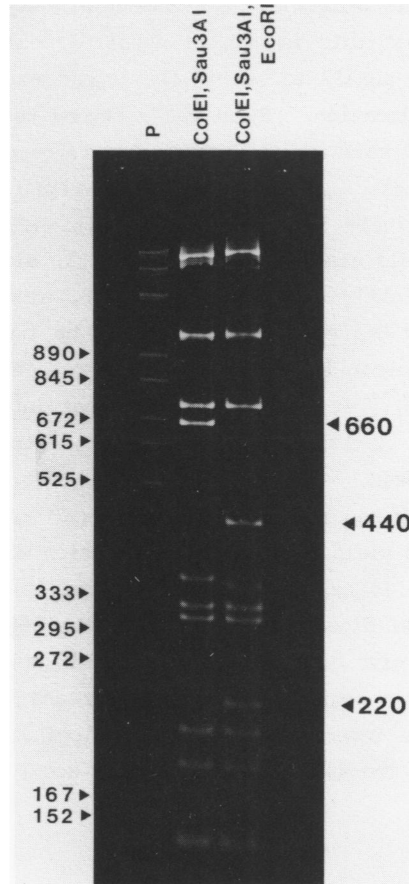


Figure 3

Cleavage of ColE1 by Sau3AI with and without EcoRI cleavage. 5% polyacrylamide gel electrophoresis of resulting DNA fragments.
Key: P - phage PM2 cleaved by HaeIII.

a polyacrylamide gel, purified and ligated into pBR322 which had been cleaved with BamHI and EcoRI. Selection of Ap^r and Tc^s recombinants should yield plasmids of almost identical size to pBR322, but containing the ColE1 inverted repeat flanked by BamHI and EcoRI sites.

Figure 5 shows the restriction characterisation of one such plasmid, pColIR215. Figure 5A shows the mobility of supercoiled pBR322 and pColIR215 on gel electrophoresis in agarose to be very similar. Cleavage with EcoRI to

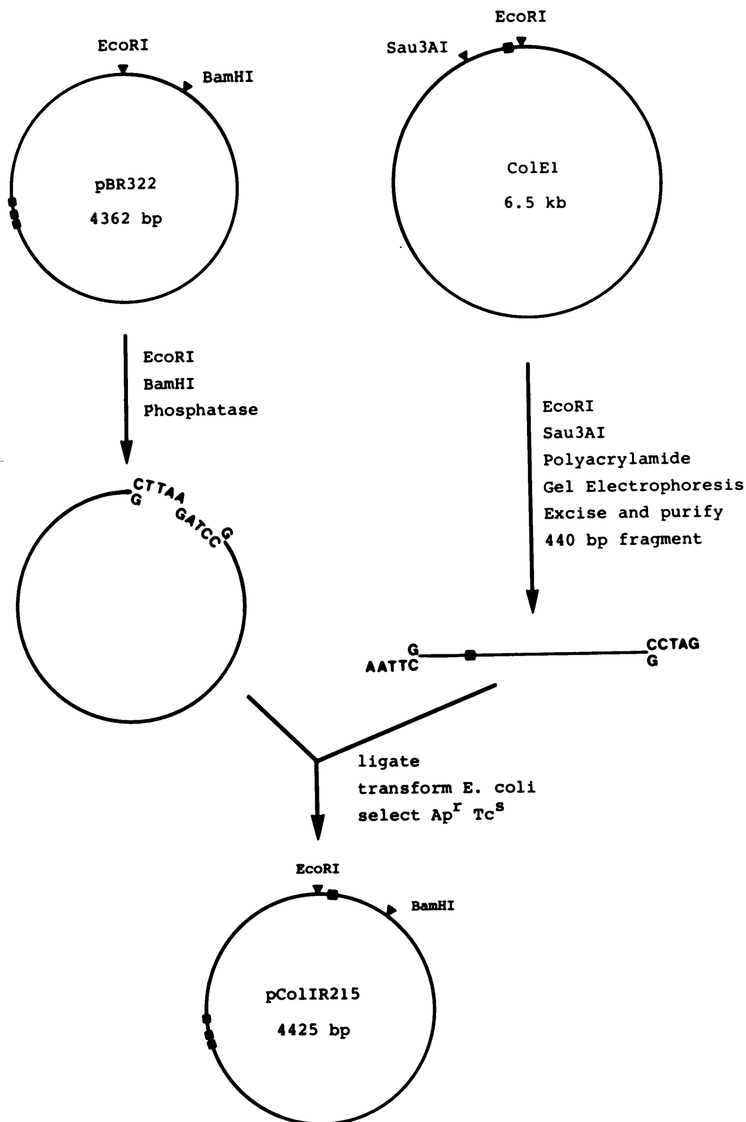
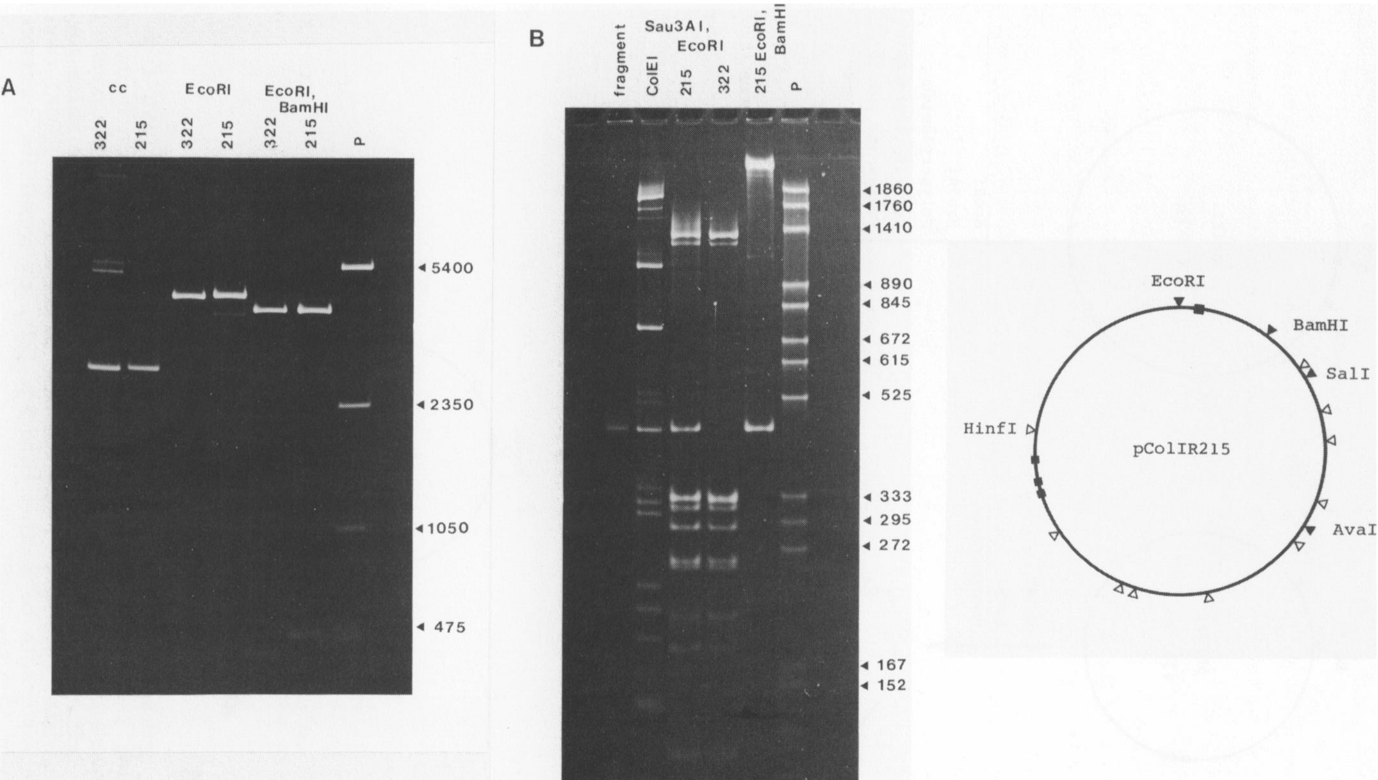


Figure 4

Cloning scheme to construct a recombinant plasmid derived from pBR322 containing the 440 bp Sau3AI, EcoRI fragment of ColE1 which includes the ColE1 inverted repeat. Note that this method depends upon the common CTAG 5' termini left by both BamHI and Sau3AI restriction cleavage which may therefore be ligated together. Since the base adjacent to the Sau3AI recognition sequence happens to be another cytosine (22), the resulting junction in the recombinant is a new BamHI site.


A
cc
EcoRI
EcoRI, BamHI
322
215
322
215
322
215
p
5400
2350
1050
475
B
fragment
Sau3AI, EcoRI
ColEI
215
322
215 EcoRI, BamHI
p
1860
1760
1410
890
845
672
615
525
333
295
272
167
152
EcoRI
BamHI
SalI
HinfI
pColIR215
AvaI

give full length linear species reveals the slightly greater length of pColIR215, but a double digestion with EcoRI and BamHI gives identical length fragments of 3.99 kb, with faintly visible short fragments of unequal length. Using the better resolution for smaller fragments of polyacrylamide gel electrophoresis, shown in figure 5B, it can be seen that double digests using EcoRI with either Sau3AI or BamHI result in the production of a fragment identical in size with the original ColE1 Sau3AI, EcoRI 440 bp fragment. All the remaining EcoRI, Sau3AI bands from pColIR215 arise from pBR322 sequences. I have subsequently confirmed the nature of the inserted DNA by Gilbert-Maxam sequencing (11) from the EcoRI site (not shown).

A cloned ColE1 inverted repeat retains its sensitivity to S1 nuclease

Figure 6 shows the results of S1 digestion followed by restriction cleavage of the recombinant plasmid pColIR215. Agarose gel electrophoresis of supercoiled pColIR215 cleaved by S1 at 5°C and subsequently cut with AvaI or SalI localises the S1 cleavage site to the EcoRI end of the inserted ColE1 DNA. Finer mapping using HinfI, shown in figure 5B, demonstrates that the S1 cleavage site lies at 2 ± 3 bp from the centre of the ColE1 inverted repeat. Thus it is evident that pColIR215 is cleaved specifically at the position of the inserted ColE1 inverted repeat in an identical manner to its cleavage in the parent plasmid. Thus the requirement for local sequence information only is proven. This experiment further demonstrates that the ability of inverted repeats to form hairpin structures is transmissible. Whilst this is clearly of considerable importance in its own right it serves further to demonstrate that cloned DNA from any given source can be probed with S1 nuclease for similar hairpin features.

Figure 6A also illustrates the competitive nature of the S1 cleavage.

Figure 5

Restriction enzyme cleavage analysis of the recombinant plasmid pColIR215. The scheme shows the expected distribution of restriction targets, based upon the cloning scheme shown in figure 4. Open triangles denote HinfI target sites.

A. 1% agarose gel electrophoresis of supercoiled, EcoRI cut and EcoRI, BamHI cut pBR322 and pColIR215.
Key: cc - supercoiled plasmid, P - phage PM2 cleaved by HindIII.

B. 5% polyacrylamide gel electrophoresis of fragments of ColE1, pBR322 and pColIR215 derived by double digests with EcoRI and either Sau3AI or BamHI. The first track contains some of the purified 440 bp ColE1, Sau3AI, EcoRI fragment used in the construction of pColIR215.
Key: P - phage PM2 cleaved by HaeIII.

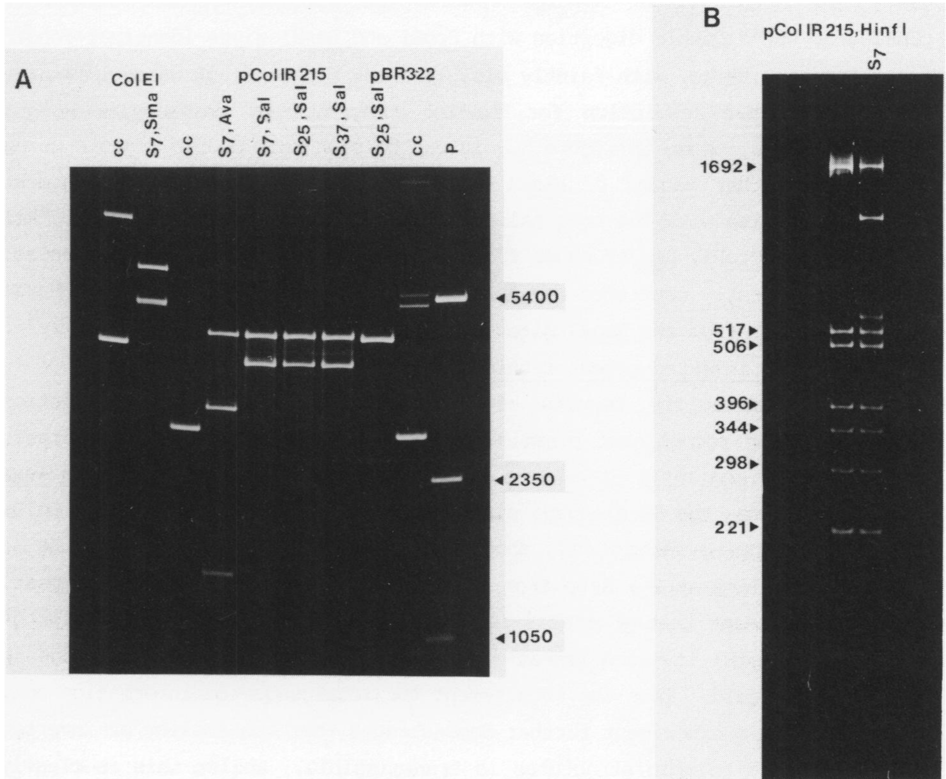


Figure 6

S1 cleavage site selection in pColIR215.

A. Comparison of S1 cleavage sites and selectivity in ColE1, pBR322 and pColIR215 1% agarose gel electrophoresis of supercoiled plasmids, S1 linearised plasmids and restriction enzyme cleaved S1 cut plasmids.

B. 5% polyacrylamide gel electrophoresis of HinfI fragments of pColIR215 with and without prior S1 cleavage.
Key: cc - supercoiled plasmid, S - S1 nuclease digestion at the temperature indicated by the subscript, P - phage PM2 cleaved by HindIII. Refer to the scheme in figure 5 for the position of relevant restriction enzyme targets.

This figure shows the result of SalI restriction of S1 cleaved pBR322 at 20°C. The major inverted repeat is cleaved, albeit rather weakly under these conditions, whereas in pColIR215 under the same conditions this cleavage is virtually abolished, replaced by the far stronger cutting at the inserted

ColE1 repeat. S1 cleavage of pColIR215 is totally dominated by the ColE1 repeat over a range of temperature. It is to be concluded that the ColE1 hairpin structure is of considerably greater stability than even the major repeat of pBR322. As discussed below, this conclusion has informative ramifications concerning the effects of nucleotide mismatch on hairpin stability.

DISCUSSION

Global linkage properties produce local structural perturbation

The evidence presented above argues strongly for the adoption of hairpin loop structures by inverted repeats in supercoiled DNA. This is the first hard evidence for such features in double stranded DNA and is of intrinsic importance in addition to being of considerable significance for several functional aspects of DNA. The requirement for negative superhelicity should be stressed. The majority of natural DNA molecules are either explicitly or intrinsically supercoiled (24) and thus are likely to be under torsional constraint at some stage of their existence. DNA underwinding is an energy requiring process either dependent upon the continuous binding of coiling protein (histone) (25) or linkage reduction by an ATP dependent enzyme (gyrase) (26). It seems likely that the cell should expect some return on its investment of energy in this process, and negative supercoiling has been implicated in replication (27), transcription (28-33) and recombination (34). However, supercoiling is a global property of DNA molecules and so mechanisms which convert global to local constraint are sought. Supercoil driven hairpin formation is just such a mechanism.

The data above show that adoption of hairpin structure by an inverted repeat is independent of the size of the molecule. DNA can be deleted out of, or transposed into, ColE1 without affecting the S1 cleavage of the inverted repeat. More significantly, a relatively small fragment containing this inverted repeat can be cloned into pBR322, whereupon the repeat adopts the hairpin structure in the recombinant molecule. The effect is therefore demonstrably local and transmissible. There is, furthermore, an interesting element of competition involved, since larger repeats tend to 'overpower' smaller ones. S1 cleavage at the major inverted repeat of pBR322 becomes insignificant once the larger ColE1 repeat is recombined into the same molecule.

Thus the inverted repeats exert a pronounced conformational effect upon the molecules in which they exist in a manner which has the following

interesting properties: the effects are (i) local, (ii) transmissible, (iii) hierarchical and (iv) necessarily cis-acting. These are just the properties which might be required in a transposable controlling DNA segment.

Hairpin base-pairing 'rules'

Further analysis of the competitive aspect of these structures, coupled with examination of flanking DNA sequences, allows the formulations of some elementary 'rules' for alternative structure in DNA. Table 1 shows the sequences of five inverted repeats which exhibit hairpin loop formation (7), together with their flanking sequences. Table 2 lists these same repeats in decreasing order of cleavage specificity, including two more repeats identified in the pBR322 sequence by computer analysis, and which are insignificantly S1 cleaved. It is to be expected that hairpin stability will be proportional directly to stem length (or, more precisely, to the length of the entire unit) but inversely to loop length. Thus the repeat of ColE1 will relax approximately 10% of the torsional constraint at the expense of merely four sets of base-pair hydrogen bonds and stacking interactions. It can be seen that the tabulated repeats conform closely to these requirements, with stem length possibly being slightly more important than loop length. Returning to Table 1, examination of the ColE1 repeat indicates that no extension

Table 1: Inverted repeat sequences (underscored) shown (7) to be cleaved by S1 nuclease with high specificity, with some flanking sequences.

Site	Sequence	Source of original sequence data
ColE1	ATAAAAGTCCTAGCAATCCAATGGGATTGCTAGGACCAAACA TATTTTCAGGATCGTTAGGTTACCCTAACGATCCTGGTTTGT	(45)
pBR322 major	GCAAACAAACCACCGCTGGTAGCGGTGGTTTTTTTGT CGTTTGTTTGGTGGCGACCATCGCCACCAAAAAAACA	(23)
pBR322 minor	ATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAA TAATAGTTTTTCCTAGAAGTGGATCTAGGAAAATTTAATT	(23)
pBR322 sub-minor	CAGAAAAAAAGGATCTCAAGAAGATCCTTTGATCTTTTCTA GTCTTTTTTTCCTAGAGTTCTTCTAGGAAACTAGAAAAGAT	(23)
φX174RF	CGCCGAAGCGGTAAAAATTTTAATTTTTGCCGCTGAGGGG GCGGCTTCGCCATTTTTAAAATTAAAAACGGCGACTCCCC	(46)

Table 2: Tabulation of inverted repeats found (7) as S1 hypersensitive sites in decreasing order of cleavage frequency.

Site + frequency	ColE1 ≈	φX174 >	pBR322 major >	pBR322 minor >	pBR322 sub-minor >	pBR322* other	
Stem length	13	13	11	10	9	8	9
Loop length	4	2	3	6	5	3	8

*These repeats found by computer analysis of the pBR322 sequence correspond to no observable S1 cleavages.

of the symmetry is possible by manipulation of the flanking sequences. This sets a standard with which to compare the other sequences.

Firstly, the cleavage of the ColE1 and φX174 repeats are of similar specificity, consistent with their similar stem lengths. However, the φX174 stems require the formation of one set of non-Watson-Crick purine-pyrimidine approximations. Therefore it can be concluded that G-T and A-C pairs do not result in a major loss of stability. Secondly, using the flanking region of pBR322 major, the stem size can be increased from 11 to 16 provided a single pyrimidine-pyrimidine or purine-purine approximation is made, see Table 3. Such an increase in stem length should make this structure considerably more

Table 3: Pseudo-symmetric extension of loop-stems for two inverted repeats from pBR322, which are also present in pColIR215.

pBR322 major:

```
G|C A A A|C|A A A C C A C C G C T|G G T|A G C G G T G G T T T|T|T T T G|T
C|G T T T|G|T T T G G T G G C G A|C C A|T C G C C A C C A A A|A|A A A C|A
```

pBR322 sub-minor:

```
                                                          A T
                                                        G     C
C|A G A A A A|A A A G G A T C T|C A A G A|A G A T C C T T T|     |T T T T C T|A
G|T C T T T T|T T T C C T A G A|G T T C T|T C T A G G A A A|     |A A A A G A|T
                                                        C     G
                                                          T A
```

stable than the ColE1 repeat. The results from the recombinant pColIR215 refute this entirely, however, and thus these A-G and T-C base pairs are clearly highly destabilising. Thirdly a similar analysis of the flanking region of pBR322 sub-minor repeat allows a similar exclusion of 'loop-out' structures, shown in Table 3. If four nucleotides are looped-out of the stem of one side of this repeat the new stem size becomes 15. This would once again result in a hairpin of at least comparable stability to that of ColE1. That this is again contrary to the pColIR215 data indicates that loop-outs must be destabilising.

Clearly the requirements for alternative secondary structure are quite rigorous. This is to be expected, for the hairpin loops, unlike comparable structures in single stranded nucleic acids, are always in competition with perfect duplexes. Of course one cannot exclude specific protein binding resulting in the stabilisation of particular structures of lower inherent stability, but in summary the following 'rules' emerge from this analysis:

i) Limited A-C and G-T pairing is permissible;
ii) A-G and C-T pairing are excluded;
iii) Loop-out sections are excluded.

These conclusions are in good agreement with the thermodynamic considerations of Tinoco, Uhlenbeck and Levine (35), and can be considered as an experimental verification in the case of double stranded DNA, subject to reservations concerning the small data base and consequent inability to consider nearest-neighbour interactions.

Some functional implications

Perhaps the most frequently postulated hairpin regions are replication origins. Indeed one can select many such structures from the pages of the 1978 Cold Spring Harbor Symposium. Yet almost all these proposed loops break one or more of the above rules, and certain specific cases may be excluded. The ColE1 replication origin (36,37) is clearly not a strong S1 site, and this is true also for the replication origins of the papova viral DNAs (D.M.J. Lilley, in preparation). Thus these regions do not adopt hairpin structures unaided, although specific replication proteins may bind to and stabilise similar structures.

Hairpin loops are also possible gene promoter contenders, the unpaired loop being attractive as an RNA polymerase initiation region. Again, however, there is presently no single case of direct evidence for this, although superhelix density has been shown (33) to modulate gene expression in bacteria. In the case of eukaryotes perhaps the whole issue of promotion

is still somewhat contentious, but the structure of the Ad2 major late promoter mapped by Ziff and Evans (38) apparently breaks none of the above rules, and would provide a most interesting S1 cleavage substrate in cloned closed circular form. Inverted symmetry has also been noted in the region of the capping site and TATA box of the mouse α-globin gene (39). Class III genes are turning out to provide a fascinating exception to dogma inherited from bacterial molecular genetics (40,41), which could be related to the pronounced secondary structures of their transcripts and, thus, inherently of the genes themselves.

A third cellular process which might take advantage of localised un-base-pairing is genetic recombination. It has been proposed (42) that the formation of Holliday-type intermediates (43) in recombination events may require the interaction of two homologous unbroken strands, and it seems possible that hairpin-type structures might be the sites of primary interaction. Long range interactions of a similar nature may be important in the somatic recombination events thought to occur in immunoglobulin gene rearrangement (44).

In summary DNA under topological constraint can exhibit alternative secondary structure, subject to some rather stringent symmetry requirements. Hairpin loop formation is a local property which is transmissible and cis-competitive. Specific functional significance is unclear, but the general phenomenon is of considerable importance to many dynamic properties of DNA.

ACKNOWLEDGEMENTS

I should like to thank many of my colleagues at Searle for discussion, especially Drs. J.S. Emtage, W. Tacon and A. Porter. I also wish to thank Prof. D. Sherratt and Drs. R.K. Patient and R.D. Wells for provision of bacterial strains.

REFERENCES

1. Watson, J.D. and Crick, F.H.C. (1953) Cold Spring Harbor Symp. Quant. Biol. 18, 123-131
2. Langridge, R., Wilson, H.R., Cooper, C.W., Wilkins, M.H.F. and Hamilton, L.D. (1960) J. Mol. Biol. 2, 19-37
3. Wing, R., Drew, H., Takano, T., Broka, C., Tanaka, S., Itakura, K. and Dickerson, R.E. (1980) Nature 287, 755-758
4. Wang, H-K., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G. and Rich, A. (1979) Nature 282, 680-686

5. Drew, H., Takano, T., Tanaka, S., Itakura, K. and Dickerson, R.E. (1980) Nature 286, 567-573
6. Arnott, S., Chandrasekaran, R., Birdsall, D.L., Leslie, A.G.W. and Ratliff, R.L. (1980) Nature 283, 743-745
7. Lilley, D.M.J. (1980) Proc. Natl. Acad. Sci. USA 77, 6468-6472
8. Gierer, A. (1966) Nature 212, 1480-1481
9. Katz, L., Kingsbury, D.K. and Helinski, D.R. (1973) J. Bacteriol. 114, 557-591
10. Vogt, V.M. (1973) Eur. J. Biochem. 33, 192-200
11. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560-564
12. Cohen, S.N., Chang, A.C.Y. and Hsu, L. (1972) Proc. Natl. Acad. Sci. USA 69, 2110-2114
13. Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. 7, 1513-1523
14. Sharp, P.A., Sugden, B. and Sambrook, J. (1973) Biochemistry 12, 3055-3063
15. Maniatis, T., Jeffrey, A. and Van de Sande, H. (1975) Biochemistry 14, 3787-3794
16. McGavin, S. (1971) J. Mol. Biol. 55, 293-298
17. Lim, V.I. and Mazanov, A.L. (1978) FEBS Lett. 88, 118-123
18. Dougan, G. and Sherratt, D. (1977) Molec. gen Genet. 151, 151-160
19. Hershfield, V., Boyer, H.W., Chow, L. and Helinski, D.R. (1976) J. Bacteriol. 126, 447-453
20. Twigg, A.J. and Sherratt, D. (1980) Nature 283, 216-218
21. Betlach, M.C., Hershfield, V., Chow, L., Brown, W., Goodman, H.M. and Boyer, H.W. (1976) Fed. Proc. 35, 2037
22. Oka, A., Nomura, N., Morita, M., Sugisaki, H., Sugimoto, K. and Takanami, M. (1979) Mol. gen. Genet. 172, 151-159
23. Sutcliffe, J.G. (1979) Cold Spring Harbor Symp. Quant. Biol. 43, 77-90
24. Bauer, W.R. (1978) Ann. Rev. Biophys. Bioeng. 7, 287-313
25. Germond, J.E., Hirt, B., Gross-Bellard, M. and Chambon, P. (1975) Proc. Natl. Acad. Sci. USA 72, 1843-1847
26. Gellert, M., Mizuuchi, K., O'Dea, M.H. and Nash, H.A. (1976) Proc. Natl. Acad. Sci. USA 73, 3872-3876
27. Gellert, M., Mizuuchi, K., O'Dea, H., Itoh, T. and Tomizawa, J. (1977) Proc. Natl. Acad. Sci. USA 74, 4772-4776
28. Botchan, P., Wang, J.C. and Echols, H. (1973) Proc. Natl. Acad. Sci. USA 70, 3077-3081
29. Hosenlopp, P., Oudet, P. and Chambon, P. (1974) Eur. J. Biochem. 41, 397-411
30. Lescure, B., Chestier, A. and Yaniv, M. (1978) J. Mol. Biol. 124, 73-85
31. Levine, A.D. and Rupp, W.D. (1978) in Microbiology 1978, ed. Schlessinger, D. (American Society of Microbiology, Washington) pp. 163-166
32. Lilley, D.M.J. and Houghton, M. (1979) Nucleic Acids Res. 6, 507-523
33. Yang, H-Y., Heller, K., Gellert, M. and Zubay, G. (1979) Proc. Natl. Acad. Sci. USA 76, 3304-3308
34. Mizuuchi, K. and Nash, H.A. (1976) Proc. Natl. Acad. Sci. USA 73, 3524-3528
35. Tinoco, Jr., I., Uhlenbeck, O.C. and Levine, M.D. (1971) Nature 230, 362-367
36. Bolivar, F., Betlach, M.C., Heyneker, H.L., Shine, J., Rodriguez, R.L. and Boyer, H.W. (1977) Proc. Natl. Acad. Sci. USA 74, 5265-5269
37. Tomizawa, J-I., Ohmori, H. and Bird, R.E. (1977) Proc. Natl. Acad. Sci. USA 74, 1865-1869
38. Ziff, E.B. and Evans, R.M. (1978) Cell 15, 1463-1475
39. Nishioka, Y. and Leder, P. (1979) Cell 18, 875-882

40. Engelke, D.R., Ng, S-Y., Shastry, B.S. and Roeder, R.G. (1980) Cell 9, 717-728
41. Sakonju, S., Bogenhagen, D.F. and Brown, D.D. (1980) Cell 19, 13-25
42. Wilson, J.H. (1979) Proc. Natl. Acad. Sci. USA 76, 3641-3645
43. Holliday, R. (1964) Genet. Res. 5, 282-304
44. Max, E.E., Seidman, J.G. and Leder, P. (1979) Proc. Natl. Acad. Sci. USA, 76, 3450-3454
45. Patient, R. (1979) Nucleic Acids Res. 6, 2647-2665
46. Sanger, F., Coulson, A.R., Friedmann, T., Air, G.M., Barrell, B.G., Brown, N.L., Fiddes, J.C., Hutchison, C.A., Slocombe, P.M. and Smith, M. (1978) J. Mol. Biol. 125, 225-246