Nucleic Acids Research, Vol. 20, No. 1 49–53

Retinoblastoma susceptibility genes contain 5′ sequences with a high propensity to form guanine-tetrad structures

Alastair I.H.Murchie and David M.J.Lilley*
Department of Biochemistry, The University, Dundee DD1 4HN, UK

Received October 28, 1991; Revised and Accepted December 3, 1991

ABSTRACT

Retinoblastoma susceptibility genes contain significant runs of oligoguanine at their 5′ ends. Oligonucleotides having these sequences underwent complex formation in the presence of sodium ions, in which there was association of four strands. Formation of this structure was completely prevented if guanine was replaced by 7-deazaguanine, indicating the importance of guanine N7 in the formation of the complex. Complex formation lead to protection of guanine N7 against methylation by dimethyl sulphate, but thymine bases located between oligoguanine blocks were reactive to osmium tetroxide. There was also some sensitivity to S1 nuclease to the 5′ side of the oligoguanine block. The results show that the G-rich regions of the mouse and human retinoblastoma susceptibility genes have a propensity to undergo tetraplex formation of the kind demonstrated in the immunoglobulin switch region.

INTRODUCTION

The retinoblastoma susceptibility gene (Rb) (1–3) encodes a nuclear phosphoprotein that acts as a tumour suppressor by affecting the cell cycle. The gene has been cloned and sequenced from human (2) and mouse (4) cells, revealing long open reading frames encoding proteins with putative DNA binding functions that are mediated by additional cellular proteins (5). The Rb protein may regulate oncogene expression, including that of *c-fos*(6, 7).

Examination of both Rb gene sequences reveals that the 5′ termini of the open reading frames, and the regions upstream of them, are extremely G + C rich. For example, the base composition of the first 100 bp of coding sequence of the mouse Rb gene, and the 100 bp that precede it, are 84% G + C. These regions contain significant runs of oligoguanine. Guanine-rich sequences have recently attracted considerable interest, due to their ability to form a variety of novel DNA structures (8–14) based upon a planar tetrad of guanine bases (15). Sequences with this propensity have been found in the immunoglobulin switch region, and are commonly found at chromosome telomeres. The structure adopted by such G-rich sequences appears to depend on the exact sequence, and on the cations present. In the presence of sodium ions, DNA strands containing runs of oligoguanine may adopt a highly stable structure formed by parallel association of four strands (9). Guanine self-association may also occur between bases in the same strand, and sequences exemplified by $(G_4T_4)_4$ may form a multiple hairpin structures in which the guanine bases are arranged as tetrads (11, 12). Some sequences have the ability to adopt both inter- and intrastranded structures, depending upon the cations present (10).

We demonstrate here that the 5′ sequences of human and mouse Rb genes exhibit a tendency to undergo tetramerisation in the presence of sodium ions, and that this is prevented if the N7 of guanine is replaced by a carbon atom.

MATERIALS AND METHODS

Oligonucleotide synthesis. Oligonucleotides were synthesised by β-cyanoethylphosphoramidite chemistry (16, 17) implemented on an Applied Biosystems 394 DNA synthesiser. After purification by gel electrophoresis, they were radioactively labelled using T4 polynucleotide kinase and [γ-^{32}P]ATP.

Formation of G4 complex and duplex species. Guanine tetraplex structures were formed by incubation of 200 pmol of [5′-^{32}P] radioactively labelled oligonucleotides in 10 mM Tris.HCl, pH 7.5, 0.1 mM EDTA containing 1 M NaCl for 60 h at 60°C. Samples of these oligonucleotides were also hybridised to their complementary sequences to generate perfect duplex species for comparison. These were electrophoresed in 10% polyacrylamide in 90 mM Tris.borate, pH 8.3, 10 mM EDTA, 10 mM NaCl at 10 v/cm. For probing studies, the tetraplex and duplex species were excised from the gel, and the DNA recovered by electroelution and ethanol precipitation.

Chemical and enzyme probing of DNA. [5′-^{32}P] radioactively labelled tetraplex and duplex DNA was studied using chemical and enzyme probes. *Dimethyl sulphate:* Labelled DNA was incubated with 0.5% (v/v) DMS in 50 mM Na cacodylate, pH 8.0, 1 mM EDTA at O°C for 4 min. DNA was recovered by precipitation with ethanol and cleaved with 1 M piperidine at 90°C for 30 min before extensive lyophilisation.

* To whom correspondence should be addressed

Osmium tetroxide: Labelled DNA was incubated with 1 mM osmium tetroxide, 1% pyridine in 10 mM Tris.HCl, pH 7.5, 0.1 mM EDTA, 10 mM NaCl at O°C for 4 min. DNA was recovered by precipitation with ethanol and cleaved with 1 M piperidine at 90°C for 30 min followed by extensive lyophilisation.

S1 nuclease: Labelled DNA was incubated with 1 U S1 nuclease in 10 μl containing 50 mM Na acetate, pH 4.6, 50 mM NaCl, 1 mM $ZnCl_2$ at O°C for 5 min.

After dissolving in formamide and heating to 100°C for 90 s, cleaved DNA samples were electrophoresed in a 15% polyacrylamide sequencing gel in 90 mM Tris.borate, pH 8.3, 10 mM EDTA containing 7 M urea, followed by autoradiography.

RESULTS

Formation of retarded complexes by DNA sequences from retinoblastoma genes

The relevant 5′ sequences of the human and mouse Rb genes are shown in Figure 1. Both contain significant $G_aX_bG_c$, where X is mainly thymine bases, and b is typically 4. This is very similar to sequences previously demonstrated to undergo tetraplex formation, and thus we examined the electrophoretic mobility of oligonucleotides with and without prolonged incubation with sodium ions. Sen and Gilbert (9) have shown that these conditions promote tetramer formation in the immunoglobulin (l_g) switch region. Figure 2 shows the electrophoretic mobility of human and mouse Rb sequences in polyacrylamide gels. Both are shown as the single G-rich strand after incubation with sodium ions,

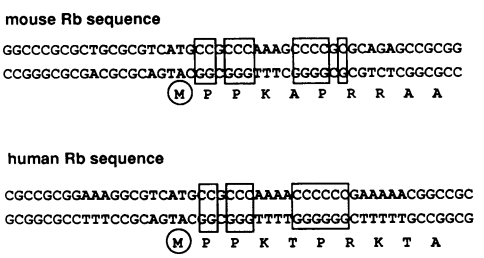

Figure 1. DNA base sequences at the 5′-ends of the coding sequences of mouse and human retinoblastoma genes. The sequences of both strands are given (1, 4), with the blocks of oligoguanine boxed. The oligoguanine stretches are on the lower strands in this representation. The translated protein sequence is given below each DNA sequence, with the initiator methionine residues circled.

compared to a perfect duplex with its complement. Each sequence forms a strongly retarded species in the presence of sodium ions, with similar mobilities to oligonucleotides based on the l_g sequence (data not shown). The Rb sequences appear to undergo tetramerisation efficiently; in the case of the human Rb sequences, no single-stranded species remained after the 60 h incubation. Although the formation of the retarded species was found to be very efficient using long incubations at 60°C in the presence of high sodium ion concentrations, we have also observed its formation at 37°C and 4°C, and using sodium ion concentrations as low as 100 mM.

We carried out experiments similar to those above, in which sodium ions were replaced with potassium ions. The mouse Rb sequence was incubated in the presence of 1M KCl, and electrophoresed in polyacrylamide in the presence of 10 mM KCl. Under these conditions the DNA was slightly retarded relative to duplex, but the migration was considerably faster than in the presence of sodium ions (data not shown). This indicates that the Rb sequence adopted different conformations in sodium and potassium ions, in agreement with earlier work of Sen and Gilbert (10) on the immunoglobulin switch region DNA.

The Rb complexes are tetrameric

To demonstrate the stoichiometry of the complexes formed by the Rb sequences in the presence of sodium ions, we employed the method of Sen and Gilbert (10). Equimolar amounts of two oligonucleotides based on the mouse Rb sequences, that differed in their overall lengths, were incubated and electrophoresed as before. Random incorporation of long (L) and short (S) oligonucleotides into a tetramer should generate five possible species (ie L_4, L_3S, L_2S_2, LS_3 and S_4), and this was just what we observed (Figure 3). We conclude that the retarded species formed by the Rb sequences do indeed contain four strands.

Chemical and enzyme probing of Rb complexes

Further evidence for the nature of the complex formed by the Rb sequences comes from chemical and enzyme probing. It has been suggested that the structure is based on a cyclic G4 tetrad (Figure 4) and it has been shown that for l_g sequences, and for telomeric repeats, hydrogen bonding between N7 and N^2H leads to protection of guanines against reaction with dimethyl sulphate (DMS). Figure 5A shows that there is almost complete protection of the guanines of the $G_aX_bG_c$ sequence of the mouse Rb gene, and equivalent results were obtained with the human Rb sequence, summarised in Figure 5B.

We also probed the character of the bases in between the oligoguanine runs. Since these were mainly thymine bases, we used osmium tetroxide which can react with thymine if the

Table 1. Sequences of oligonucleotides used in these studies. The immunoglobulin switch region (l_g) sequence was identical to that used by Sen and Gilbert to demonstrate tetraplex formation (9). The human and mouse retinoblastoma (Rb) sequences were taken from the published gene sequences (1, 4), with minor modification at the 5′ end to assist in cloning experiments (the modifications are well removed from the guanine blocks). Oligoguanine blocks are highlighted in bold. A shorter version of the mouse Rb sequence (S) was used in some experiments. A version of the mouse Rb (S) sequence was generated in which specified guanine bases were replaced by 7-deazaguanine; these are indicated by lower case (g).

l_g switch region	GGGACCAGACCTAGCAGCTAT**GGGGG**AGCT**GGGG**AAGGT**GGG**AATGTGA
Human Rb	AATTCCGGCCGTTTTTC**GGGGGG**TTTT**GGGCGG**CA
Mouse Rb (L)	AATTCCGGCCGCGGCTCTGCGC**GGGGC**TTT**GGGCGG**CA
Mouse Rb (S)	TCTgCgC**gggg**CTTT**ggcgg**C
Mouse Rb (7-deazaG)	TCTgCgC**gggg**CTTT**gggCgg**C

required out-of-plane attack is unhindered. The results (Figure 5A) show that the thymine bases in the tetraplex were strongly reactive. By contrast, the same sequence incorporated into a duplex was unreactive. Single-stranded character was also probed using S1 nuclease (Figure 5A). Strong cleavage was observed at both ends of the oligonucleotide, consistent with single-stranded overhanging termini, but in addition, some weak cleavage could be seen at the 3′ end of the CTTT sequence, at the junction with one of the G_n blocks. However, the inter-guanine region was not uniformly sensitive to the S1 nuclease. Taken together, the osmium tetroxide and S1 nuclease probing studies suggest that the region between the oligoguanine blocks may be partially structured under the experimental conditions.

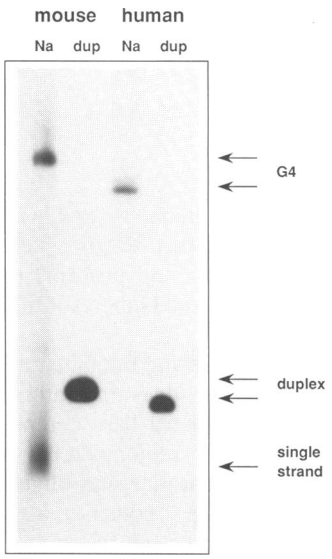

Figure 2. Formation of a retarded species by sequences from the retinoblastoma susceptibility gene. Oligonucleotides based on the human Rb and mouse Rb(L) Rb sequences (Table 1) were radioactively labelled at their 5′-termini with [^{32}P] and incubated in the presence of 1 M NaCl for 60 h at 60°C. Samples of these oligonucleotides were also hybridised to their complementary sequences to generate perfect duplex species for comparison. These were electrophoresed on a 10% polyacrylamide gel in 90 mM Tris.borate, pH 8.3, 10 mM EDTA (TBE) containing 10 mM NaCl, followed by autoradiography. The single-stranded DNA incubated with sodium ions was loaded in the lanes labelled *Na*, and the duplex DNA in those labelled *dup*. Note the generation of the strongly retarded species, indicated by the arrows labelled G4. The mobility of these species is closely similar to that generated by the oligonucleotide having the sequence of the immunoglobulin switch region (Table 1) after an identical incubation in sodium ions (data not shown).

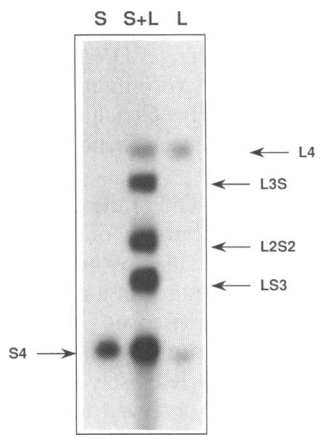

Figure 3. Formation of tetrameric species by the mouse Rb sequence. The short (S) and long (L) versions of the mouse retinoblastoma sequence (Table 1) were incubated with 1 M NaCl as before, both separately and as an equimolar mixture (*S + L*). These were electrophoresed in 12% polyacrylamide, and the autoradiograph presented. Note the formation of five species in the mixed incubation, the slowest and fastest of which correspond to the mobilities of the sodium forms of the pure long and short oligonucleotides respectively. The assignment of the five species are indicated at the sides of the gel.

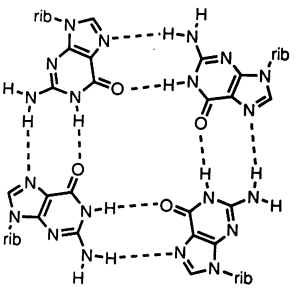

Figure 4. Proposed structure of the tetrad of guanine bases, as the core of the tetrameric structure (10–13, 15).

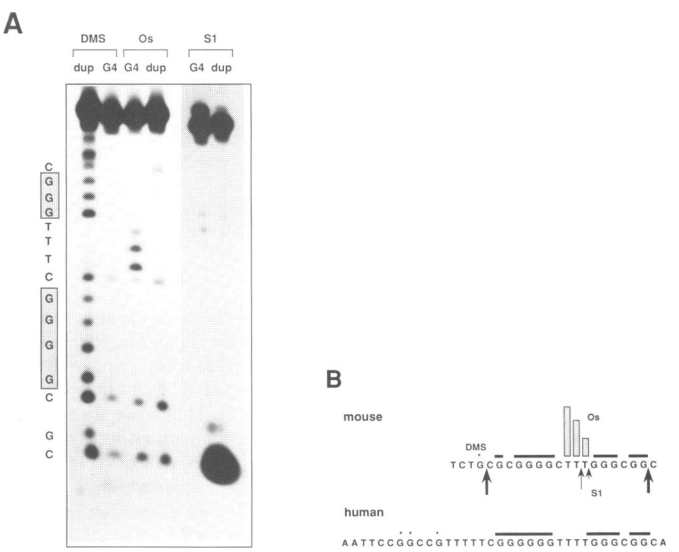

Figure 5. Probing experiments support a structure that is held together through association of guanine bases. **A.** Probing the structure of the tetrameric mouse Rb gene complex with dimethyl sulphate (*DMS*), osmium tetroxide (*Os*) and S1 nuclease (*S1*). For each probe, we compare the reaction of the tetrameric complex (*G4*) with the duplex form (*dup*). The short form of the mouse Rb sequence was [5′-^{32}P] radioactively labelled and hybridised to its complement to form the duplex, or incubated with 1 M NaCl to generate the tetrameric complex. Following chemical modification and piperidine cleavage, or cleavage by S1 nuclease, all six samples were electrophoresed in a 15% sequencing gel in TBE containing 7 M urea, followed by autoradiography. The intense radioactivity at the lower end of the gel after S1 nuclease cleavage of the G4 species is a result of the cleavage of the 5′ single-stranded region. **B.** Summary of the probing results on the Rb sequences. The results of DMS, osmium tetroxide and S1 nuclease are indicated for the mouse Rb sequence, while only DMS experiments have been performed on the human Rb sequence. Regions protected against DMS modification are indicated by the black line; reactive guanine bases are indicated by asterisks (*). Extent of modification of thymine bases by osmium tetroxide are indicated by the stippled bars, and the positions of cleavage by S1 nuclease are shown by arrows.

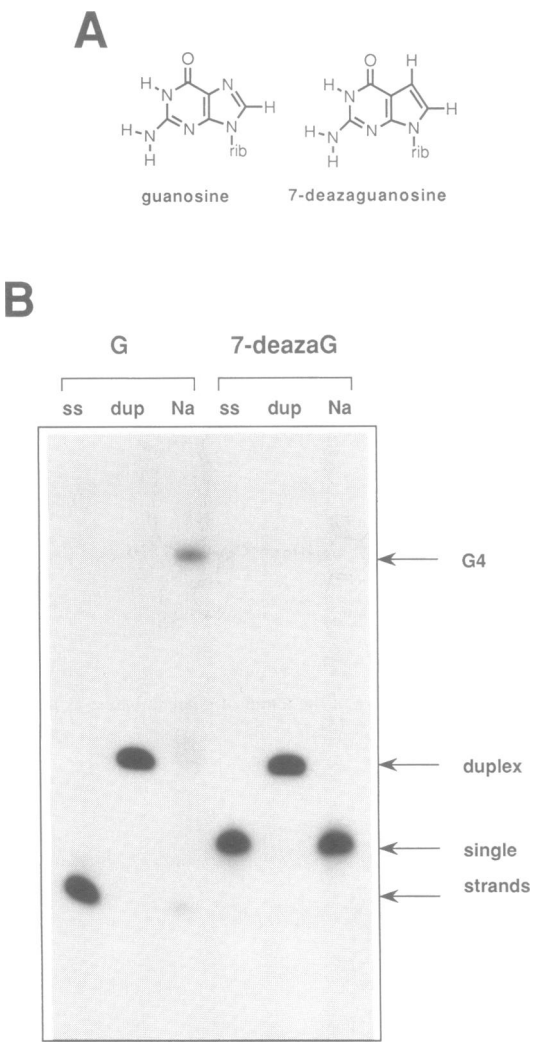

Figure 6. The N7 of guanine is essential for the formation of the tetrameric Rb species. **A.** Structures of guanine and 7-deazaguanine compared. Reference to Figure 4 shows that the modified base will be unable to form a hydrogen bond donated by N^2 of another guanine base. However, the ability to basepair with cytosine via O^6, N1 and N^2 should be unimpaired. **B.** The mouse Rb sequence was synthesised using either normal guanine (*G*) or 7-deazaguanosine (*7-deazaG*), as indicated in Table 1. The oligonucleotides were [5′-^{32}P] radioactively labelled and hybridised to their complementary sequence (*dup*), incubated with 1 M NaCl in 10 mM Tris.HCl, pH 7.5, 0.1 mM EDTA (*Na*), or left as unincubated single strands (*ss*). The six species were electrophoresed on a 12% polyacrylamide gel in TBE, 10 mM NaCl, followed by autoradiography. Note the formation of duplex species by both oligonucleotides, but the complete failure to form a retarded complex in the presence of sodium ions by the 7-deazaguanine-containing oligonucleotide.

Guanine N7 Is essential for the formation of the tetrameric complex

The assumed cyclic tetrad structure (Figure 4) depends upon the ability of the N7 of guanine to accept a hydrogen bond. We therefore employed a new method to test this directly. An analogue of the mouse Rb sequence was synthesised in which the majority of the guanine bases were replaced by 7-deazaguanine (Figure 6A). According to the tetrad model, the ability to adopt the tetraplex structure should be severely impaired. The result is shown in Figure 6B. In contrast to the sequence with normal guanine bases, tetraplex formation by the modified sequence was totally prevented. However, the deazaguanine-containing sequence could form a normal duplex with its complement; as expected because the N7 position does not participate in normal Watson-Crick base pairing. This shows that the guanine N7 is essential for the formation of the tetrameric complex, further supporting the structure of the tetrad shown in Figure 4.

DISCUSSION

We have demonstrated that the mouse and human Rb sequences can adopt a four-stranded structure, in which the N7 position of guanine is critical. This is most probably the same structure as that demonstrated by Sen and Gilbert for the immunoglobulin switch region.

The region for 200 bp flanking the initiator ATG codons are extremely high in G and C residues, in contrast to the remainder of the genes that have a normal base composition. This high G + C content in the coding region can be partially explained by significant runs of oligoproline and oligoalanine in the Rb protein, but the same high G+C content is also found in the 100 bp of noncoding DNA that proceeds the start of translation. This includes sequences such as GGGGAGGCGGGCGGG, that would also be expected to adopt the G4 tetraplex structure.

The significance of these observations is difficult to judge at present. Several factors argue against a biological role for tetraplex formation by these Rb sequences. It is not yet clear how the stability of the tetraplex structure compares with competing duplex formation by hybridisation to the complementary strand, and the kinetics of formation of the structure appear to be slow. Perhaps most significantly, it is not obvious where the four strands required to form the tetraplex would come from, except upon chromosome condensation.

The propensity for tetramerisation by these G-rich sequences suggests a role in recombination events, and it is proposed that this could be the function of the immunoglobulin switch regions sequences (9). Similar sequences are also present in the regions responsible for the isomerisation of the HSV1 genome, and these have been shown to adopt a novel DNA structure *in vitro* (18). It is noteworthy that the recombinational hotspot chi in bacteriophage λ is similarly G-rich (19, 20), as are the hypervariable satellite sequences found in the human genome (21).

It is interesting that the 5′ end of the gene encoding a protein that has many of the characteristics of a DNA-binding protein important in gene regulation, itself exhibits a predisposition for adopting a non-B DNA structure. In a recent study, Chittenden *et al.* (5) have identified a consensus DNA binding site for an Rb-cellular protein complex having the sequence TTTTGGCGGG; this is clearly closely similar to the sequences that are responsible for the observed tetramerisation.

ACKNOWLEDGEMENTS

We are very grateful to Peter Yesudian for exploratory experiments, to David Lane and Carol Midgeley for valuable discussion, and to the Cancer Research Campaign for financial support.

REFERENCES

1. Friend,S.H., Bernards,R., Rogeli,S., Weinberg,R.A., Rapaport,J.M., Alberts,D.M. and Dryja,T.P. (1986) *Nature* **323**, 643–646.
2. Lee,W.-H., Shew,J.Y., Hong,F.D., Sery,T.W., Donoso,L.A., Young,L.-J., Bookstein,R. and Lee,E.Y.-H.P. (1987) *Nature* **329**, 642–645.
3. Lee,E.Y.-H.P., To,S., Shew,J.Y., Bookstein,R., Scully,P. and Lee,W.H. (1988) *Science* **241**, 218–221.
4. Bernards,R., Schackleford,G.M., Gerber,M.R., Horowitz,J.M., Friend,S.H., Schartl,M., Bogenmann,E., Rapaport,J.M., McGee,T., Dryja,T.P. and Weinberg,R.A. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6474–6478.
5. Chittenden,T., Livingston,D.M. and Kaelin Jr,W.G. (1991) *Cell* **65**, 1073–1082.
6. Robbins,P.D., Horowitz,J.M. and Mulligan,R.C. (1990) *Nature* **346**, 668–671.
7. Kim,S.-J., Lee,H.-D., Robbins,P.D., Busam,K., Sporn,M.B. and Roberts,A.B. (1991) *Proc. Natl. Acad Sci. USA* **88**, 3052–3056.
8. Henderson,E., Hardin,C.C., Walk,S.K., Tinoco Jr.I. and Blackburn,E. (1987) *Cell* **51**, 899–908.
9. Sen,D. and Gilbert,W. (1988) *Nature* **334**, 364–366.
10. Sen,D. and Gilbert,W. (1990) *Nature* **344**, 410–414.
11. Sundquist,W.I. and Klug,A. (1989) *Nature* **342**, 825–829.
12. Williamson,J.R., Raghuraman,M.K. and Cech,T.R. (1989) *Cell* **59**, 871–880.
13. Panyutin,I.G., Kovalsky,O.I., Budowsky,E.I., Dickerson,R.E., Rikhirev,M.E. and Lipanov,A.A. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 867–870.
14. Jin,R., Breslauer,K.J., Jones,R.A. and Gaffney,B.L. (1990) *Science* **250**, 543–546.
15. Gellert,M., Lipsett,M.N. and Davies,D.R. (1962) *Proc. Natl. Acad. Sci. USA* **48**, 2013–2018.
16. Beaucage,S.L. and Caruthers,M.H. (1981) *Tetrahedron Letters* **22**, 1859–1862.
17. Sinha,N.D., Biernat,J., McManus,J. and Koster,H. (1984) *Nucl. Acids Res.* **12**, 4539–4557.
18. Wohlrab,F., McClean,M.J. and Wells,R.D. (1987) *J. Biol. Chem.* **262**, 6407–6416.
19. Lam,S.T., Stahl,M.M., McMilin,K.D. and Stahl,F.W. (1974) *Genetics* **77**, 425–423.
20. Smith,G.R., Kunes,S.M., Shultz,D.W., Taylor,A. and Triman,K.L. (1981) *Cell* **24**, 429–436.
21. Jeffreys,A.J., Wilson,V. and Thein,S.L. (1985) *Nature* **314**, 67–73.