The EMBO Journal vol.4 no.13A pp.3557–3562, 1985

Transposon-like properties of the major, long repetitive sequence family in the genome of *Physarum polycephalum*

Douglas H.Pearston, Mairi Gordon and Norman Hardman

Department of Biochemistry, University of Aberdeen, Marischal College, Aberdeen AB9 1AS, UK

Communicated by D.Sherratt

A family of long, highly-repetitive sequences, referred to previously as '*Hpa*II-repeats', dominates the genome of the eukaryotic slime mould *Physarum polycephalum*. These sequences are found exclusively in scrambled clusters. They account for about one-half of the total complement of repetitive DNA in *Physarum*, and represent the major sequence component found in hypermethylated, 20–50 kb segments of *Physarum* genomic DNA that fail to be cleaved using the restriction endonuclease *Hpa*II. The structure of this abundant repetitive element was investigated by analysing cloned segments derived from the hypermethylated genomic DNA compartment. We show that the '*Hpa*II-repeat' forms part of a larger repetitive DNA structure, ~8.6 kb in length, with several structural features in common with recognised eukaryotic transposable genetic elements. Scrambled clusters of the sequence probably arise as a result of transposition-like events, during which the element preferentially recombines in either orientation with target sites located in other copies of the same repeated sequence. The target sites for transposition/recombination are not related in sequence but in all cases studied they are potentially capable of promoting the formation of small 'cruciforms' or 'Z-DNA' structures which might be recognised during the recombination process.

Key words: Physarum/repetitive DNA/transposon

Introduction

Transposable genetic elements are widespread in nature; they are a common feature of both prokaryotic and eukaryotic genomes (Kleckner, 1981; Starlinger, 1984). A number of well-defined eukaryotic transposable elements have been described in widely different phyla, including *Drosophila* (Spradling and Rubin, 1981), yeast (Cameron *et al.*, 1979) and maize (Doring *et al.*, 1984; Shepherd *et al.*, 1984). The structures of different mobile elements are quite similar, and their properties are often compared with those of integrated retrovirus proviruses, leading to speculation that these different classes of genetic elements may share the same evolutionary origin (Temin, 1980; Varmus, 1982). This hypothesis is reinforced by experiments demonstrating amino acid sequence homologies predicted from open reading frames located within the internal domains of both these classes of genetic elements, in regions coding for gene products possibly involved in transposition (Saigo *et al.*, 1984; Toh *et al.*, 1984; Hauber *et al.*, 1985). Additionally, the transposition mechanism of some mobile genetic elements, such as that of Ty in yeast, is known to involve an RNA intermediate (Boeke *et al.*, 1985). Such elements have been referred to as 'retrotransposons'. Mobile genetic elements probably account for a major fraction of the middle-repetitive DNA in the genomes of some organisms, such as *Drosophila melanogaster* and related species (Finnegan *et al.*, 1977; Young, 1979; Dowsett and Young, 1982; D'Eustachio and Ruddle, 1983).

The most abundant middle-repetitive sequence families in mammalian DNA similarly may have evolved by processes involving retrotranscription. *Alu*I-repeats, the most highly-repetitive family of short, interspersed repeats in human DNA (Houck *et al.*, 1979) display a high level of homology with the 7SL RNA sequence and it has been argued that they may have originated from 7SL RNA by reverse transcription, followed by dispersal of pseudogene-like DNA copies by chromosomal integration at some stage in evolution preceding mammalian radiation (Ullu and Tschudi, 1984). The L1 family of repeats in human DNA and its counterparts in other mammalian genomes also possess a pseudogene-like structure (Singer, 1982; Singer and Skowronski, 1985), and other properties suggesting close structural similarity with the transposable F-elements in *Drosophila* (Di Nocera *et al.*, 1983). All these observations point to the general conclusion that mechanisms of replicative transposition have probably played a significant role in the evolution of diverse arrays of repetitive sequences in eukaryotic genomes.

Over one-half of the total complement of repeated DNA in the genome of the eukaryotic slime mould *Physarum polycephalum* consists of a single family of long, methylated, highly-repetitive sequences which we have previously referred to as '*Hpa*II-repeats' (Peoples *et al.*, 1985). They account for a sizeable fraction, possibly up to 20%, of the genome and are found exclusively in clusters, forming methylated 'domains' 20–50 kb long that are a source of about one-half of all the 'foldback' DNA in *Physarum* (Whittaker *et al.*, 1981; Peoples and Hardman, 1983). Using non-homologous segments of the *Hpa*II-repeat as hybridisation probes it was shown that different portions of its sequence are frequently found in various 'scrambled' arrangements in different cloned DNA segments (Peoples and Hardman, 1983). Scrambled clusters of repetitive elements are a general structural feature of eukaryotic genomes (Wensink *et al.*, 1979; Musti *et al.*, 1981; Eden *et al.*, 1981). By studying the structure of segments of *Physarum* DNA containing these sequences we therefore not only gain information on the mechanisms responsible for sequence 'scrambling' and the formation of foldback DNA in this organism, but also hope to provide further insights into the structure and evolution of eukaryotic genomes in general.

Here we report that the 5.8 kb long highly-abundant '*Hpa*II-repeat', identified previously in *Physarum* DNA (Peoples *et al.*, 1985) forms part of an even larger repetitive DNA element with many of the features of a eukaryotic transposon. The disorganised structure typical of DNA segments containing clusters of these elements can thus be explained by transposition-like events that have resulted in the integration of additional copies of the putative mobile element, in either orientation, into preferred target sites located within its own sequence, generating DNA segments with the potential to form 'foldback' structures.

© IRL Press Limited, Oxford, England

Results

Structure of the Physarum genomic DNA clone PL15

Initially, one cloned *Physarum* DNA segment containing the highly-abundant '*Hpa*II-repeat' was chosen for detailed restriction, hybridisation and nucleotide sequence analysis. This sequence (PL15) is one of a number of previously characterised DNA segments derived from the methylated compartment of *Physarum* nuclear DNA (Peoples and Hardman, 1983). The restriction map of PL15 is shown in Figure 1, together with maps of two distinct clones (pPH29 and pPH53a) derived independently from the *Hpa*II-repeat as plasmid clones (McLachlan and Hardman, 1982) and used originally as hybridisation probes to select the PL15 sequence from a lambda genomic DNA library. The restriction map of PL-s5, a 4.3-kb subcloned *Msp*I fragment of a *Hpa*II-repeat obtained from a separate lambda-derived genomic clone PL5 (Peoples *et al.*, 1985) is also presented. pPH29, pPH53a and PL5-s5 were nick-translated and used as probes to various restriction digests of PL15 to determine the location of these sequences. Results are summarised in Figure 1. Those segments previously identified as contiguous, internal portions of the long *Hpa*II-repeat (Peoples *et al.*, 1985) are labelled A – E. The results indicate that the central portion of PL15 contains these segments in the expected register for an intact *Hpa*II-repeat element. Two regions (labelled 1 and 2, Figure 1) indicate the approximate positions of discontinuities in the sequence order in PL15; in region 1 segment A is preceded by B and in region 2 it is preceded by segment E. The nucleotide sequences of these regions (Figure 2) shows the position of the discontinuities. The sequence TGTTGG is located at the precise point at which the two sequences diverge. This suggested that transposition-like events may have given rise to the sequence arrangement found in PL15, since nearly identical small sequence elements are found at the termini of a number of well-characterised mobile genetic elements, including *copia* (Levis *et al.*, 1980), Ty (Gafner and Philippsen, 1980), *Cin*1 (Shepherd *et al.*, 1984) and spleen necrosis virus (Shimotohno *et al.*, 1980).

Figure 3a illustrates a hypothetical scheme for the origin of PL15 based on DNA transposition events involving a repetitive element containing the DNA segments A – F. The structure of PL15 can be accounted for by two insertions of additional copies of the same element, both involving target sites within the repeated element, one to the right of segment B and the other to the right of segment E as depicted. The observed restriction map of PL15 is exactly as predicted on the basis of this scheme, with the exception of two additional restriction sites, one for *Bam*HI and one for *Eco*RI. These differences can be accounted for by minor sequence variations within the *Hpa*II-repeat element since additional copies of this sequence, lacking these particular restriction sites, are found in the same clones. The *Bam*HI-restriction site is located in the *Eco*RI/*Hind*III segment B; four other copies of this segment can be seen in the restriction maps presented in Figure 1 which lack the internal *Bam*HI site. Likewise a copy of the *Bam*HI/*Hpa*II segment D is located in pPH29 (Figure 1) lacking the *Eco*RI restriction site referred to above. Nucleotide sequencing confirms that these differences result from sequence polymorphisms in different copies of the

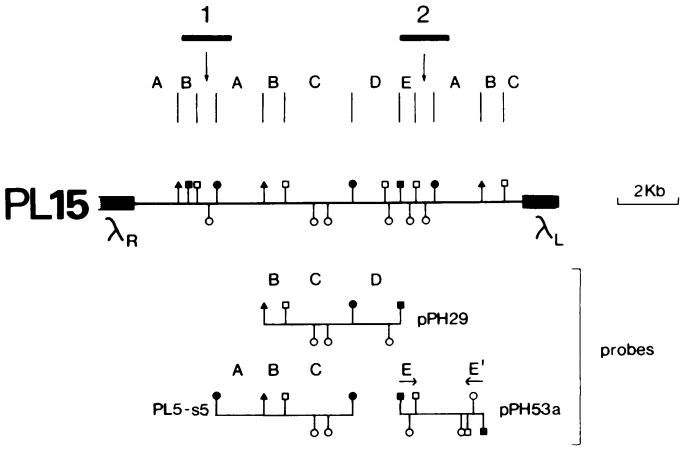

Fig. 1. Restriction/hybridisation analysis of *Physarum* genomic DNA clone PL15. PL15 DNA was restricted with appropriate enzymes. Fragments were separated by gel electrophoresis, and blot-hybridised using the recombinant plasmid DNA probes indicated. Segments with the same restriction/hybridisation properties are labelled A – E. The segment E′ is an inverted copy of E located in pPH53a. Regions 1 and 2 indicate the positions of discontinuities in the sequence order of PL15 DNA segments A – E. ■, *Bam*HI; □, *Eco*RI; ○, *Hind*II; ▲, *Hind*III; ●, *Msp*I.

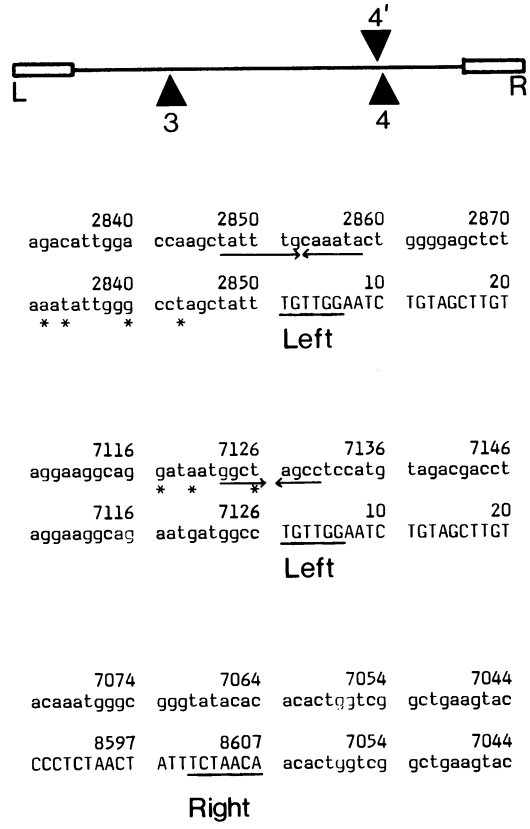

Fig. 2. Nucleotide sequence analysis of sites of discontinuity in DNA segments containing '*Hpa*II-repeats'. Sequence analysis was carried out by subcloning and sequencing appropriate restriction fragments in M13mp8 or M13mp9 vectors. Regions 1 and 2 are sequences of PL15 indicated by arrows (see Figure 1) positioned between *Eco*RI and *Hind*II restriction sites. The first T residue of the underlined sequences TGTTGG have been assigned as nucleotide 1 to indicate their position at the left of the putative transposon-like element. Other regions are numbered to denote their relative position in the element, whose total length is estimated to be 8.6 kb (Figure 4). Positions are approximate and based partly on restriction mapping as well as sequence data. Regions 3 and 4 are sequences of uninterrupted internal regions of the long repeated element with positions as indicated, corresponding to sites to the right of *Eco*RI cleavage sites in PL5-s5 and pPH53a (Figure 1). Asterisks denote base changes in the corresponding homologous segments of regions 1/3 and 2/4. Regions 4′ and 5 are sequences derived from the left and right hand ends, respectively, of the 1.8-kb *Eco*RI segment of pPH53a (labelled F in Figure 3b). The sequence TCTAACA, thought to define the right end of the transposon-like element is underlined. Region 4′ is located on the opposing DNA strand to region 4, in the same region of the repeated element.

*Hpa*II-repeat element.

The scheme predicts that uninterrupted versions of the target sites for the recombination events leading to the formation of PL15 should be located in regions 3 and 4 of the proposed transposon-like element. As shown in Figure 2 this prediction is fulfilled; sequences of the left hand portions of regions 3 and 4 correspond to those immediately to the left of the TGTTGG sequence in regions 1 and 2, respectively.

Structure of the Physarum genomic DNA clone pPH53a

The scheme outlined in Figure 3a suggests that PL15 contains two left hand segments of the putative transposon-like sequence, but is devoid of segment F representing the right hand portion of the element. The *Physarum* genomic DNA clone pPH53a contains a 1.8-kb *Eco*RI-terminated DNA fragment flanked by two E segments arranged in inverted orientation. It was thus supposed that pPH53a might contain the right hand segment of the repeated element (segment F), and that it might have been formed by a transposition-like event involving mutually-inverted copies of the element recombining at a site to the right of segment E. The restriction map of pPH53a is as predicted on the basis of this hypothesis (Figure 3b). The scheme outlined in Figure 3b predicts that sequences at either end of the 1.8-kb *Eco*RI fragment of pPH53a should be closely related or identical up to the

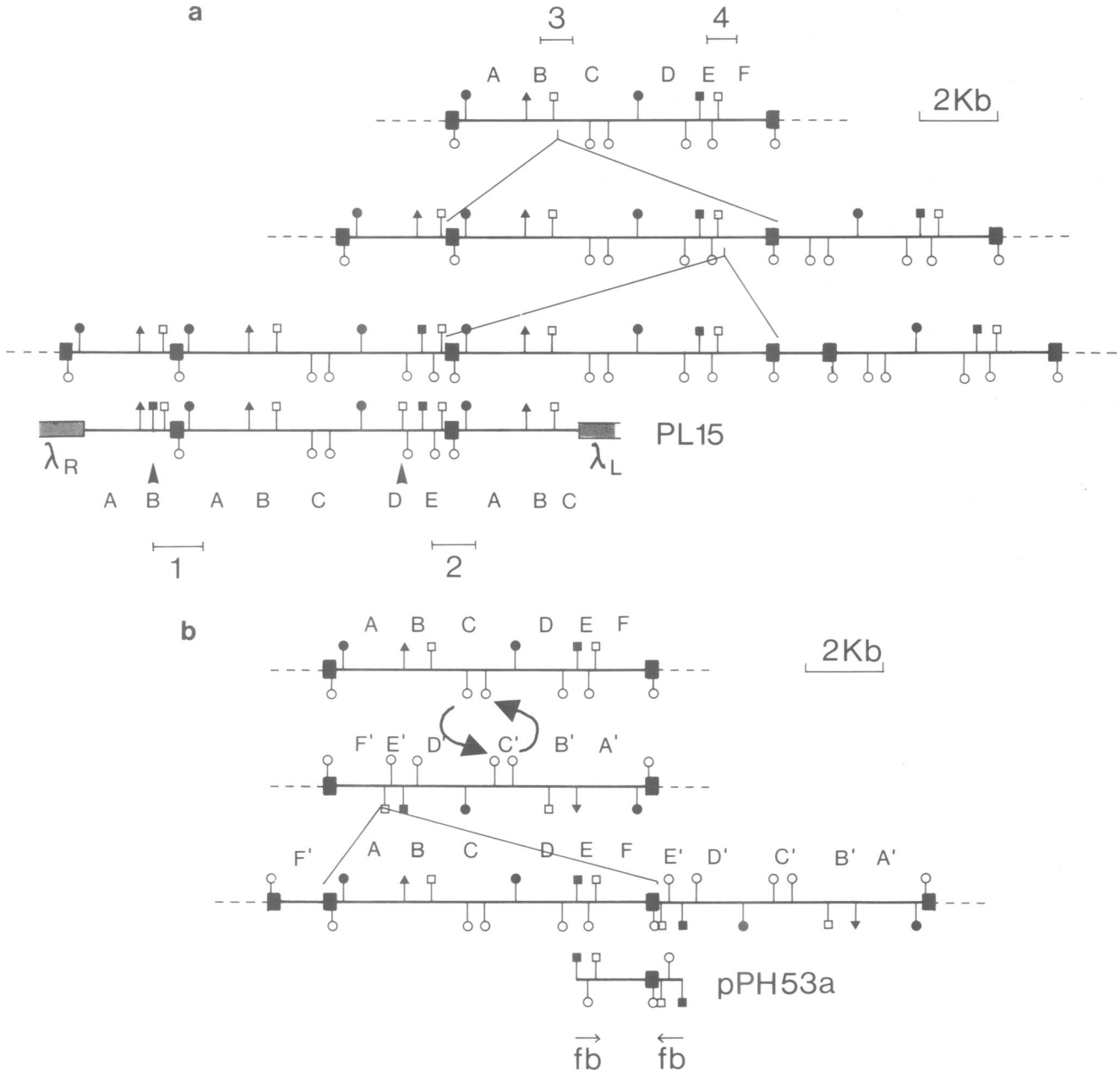

Fig. 3. Hypothetical schemes showing the possible origin of PL15 and pPH53a by DNA transposition. The long transposon-like element containing segments A–F is drawn schematically. The ends of the element are indicated by boxed regions containing *Hind*II restriction sites. Symbols denoting restriction sites are as defined in Figure 1. (**a**) Origin of PL15: two copies of the transposon-like element recombine into regions 3 and 4 in direct register, forming a repeated sequence cluster containing one uninterrupted and two interrupted copies of the element. 'Random' cloning of *Sau*3a1 partials of genomic DNA (Peoples and Hardman, 1983) generates a DNA segment with restriction properties corresponding to those of PL15. (**b**) Origin of pPH53a: a copy of the transposon-like element recombines in reverse orientation into a site located within a second copy of the element around the boundary of E and F. A *Bam*HI segment derived from this structure by plasmid cloning (McLachlan and Hardman, 1982) has the properties of pPH53a. 'fb' indicates the position of foldback elements in pPH53A defined by electron microscopic analysis (McLachlan and Hardman, 1982).

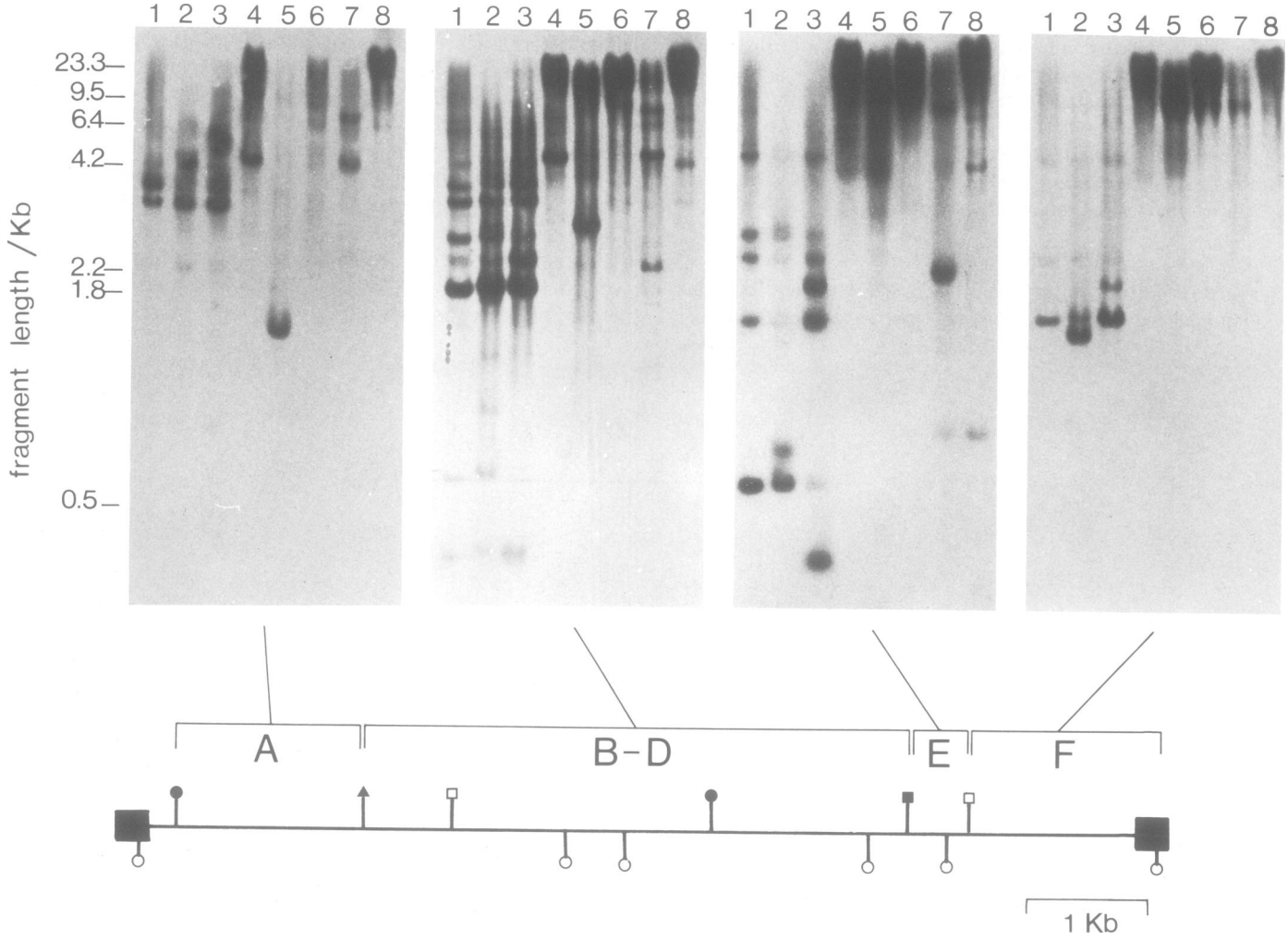

Fig. 4. Co-linear arrangement of *Hpa*II-repeat elements in *Physarum* genomic DNA. The restriction map of a complete element is shown, containing segments A – F. Various non-overlapping segments were obtained by subcloning into pUC8 and pUC9 plasmid vectors and later used as hybridisation probes to various restriction digests containing 1 μg of *Physarum* genomic DNA. These were: **A**, *Hpa*II/*Hind*III segment of PL5-s5; **B – D**, pPH29; **E**, *Bam*HI/*Eco*RI segment of pPH53a; **F**, *Eco*RI segment of pPH53a. Radioactively-labelled lambda *Hind*III restriction fragments (not shown) were used as size markers. *Physarum* genomic DNA was digested using: **lane 1**, *Hind*II; **lane 2**, *Hind*II/*Eco*RI; **lane 3**, *Hind*II/*Bam*HI; **lane 4**, *Msp*I; **lane 5**, *Msp*I/*Hind*III; **lane 6**, *Hind*III; **lane 7**, *Msp*I/*Eco*RI; **lane 8**, *Eco*RI.

point of interruption by the right hand end of the element. Accordingly, this fragment was recovered from pPH53a by gel electrophoresis and sequenced after cloning into M13mp8. The sequences (4′ and 5) are compared in Figure 2. The sequence CTAACA is located at the point of discontinuity. This is an inverted complement of the hexanucleotide sequence found at the putative left hand terminus of the repeated element (Figure 2).

Consensus arrangement of HpaII-repeats in genomic DNA

The restriction map of a proposed uninterrupted copy of the repeated element is shown in Figure 4. It includes the internal portion of the repeating unit previously described as the '*Hpa*II-repeat' extending from the leftmost *Hpa*II restriction site to the *Bam*HI site (segments A and B – D, Figure 4, Peoples and Hardman, 1983; Peoples *et al.*, 1985). Copies of this element so far identified in various cloned *Physarum* DNA segments are interrupted by additional *Hpa*II-repeats. It is nevertheless possible to confirm the co-linear arrangement of restriction segments within the 'consensus' repeating unit by probing Southern blots of appropriate restriction digests of *Physarum* genomic DNA with non-overlapping internal sub-cloned portions of the element (segments A – F, Figure 4). The results are consistent with the restriction map presented in Figure 4. Additional DNA bands seen in some tracks are compatible with loss, or *in vivo* protection, of mapped restriction sites, consistent with sequence polymorphism of genomic copies of the element.

Transposon-like properties of the repeated element

More extensive nucleotide sequence analysis of the clones containing either end of the repeated element described above reveals several further properties typical of transposable elements already characterised in other eukaryotes. Regions 1 and 2 (containing the left hand portion of the element) and region 5 (containing the right hand portion of the element) show almost exact sequence identity extending for 277 bp, terminating with TGTTGG to the left and CTAACA to the right (Figure 5). This 277-bp region contains several inexact, internal direct and inverted repeats similar to the long terminal repeats (LTRs) of defined transposable elements (Finnegan *et al.*, 1977). Other transposon-like features include a region of 17 bp immediately adjacent to the left LTR which is identical in sequence and position to the retrovirus-like 'tRNA binding site' region of the mobile genetic element *copia* (Flavell and Ish-Horowicz, 1983), a purine-rich region (AGAGGGA) adjacent to the right LTR (Figure 5), putative poly(A) ad-

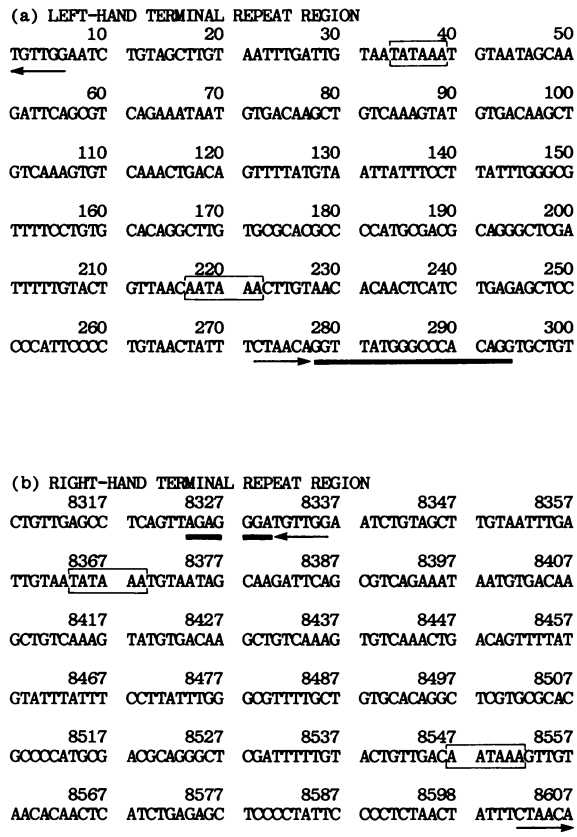

Fig. 5. Nucleotide sequence analysis of left and right LTR-like elements. Subcloning and nucleotide sequence analysis was performed by the Sanger dideoxy-chain termination method using M13mp8 and M13mp9 vectors. (**a**) Sequence of the putative left-hand end of the repeated element. Nucleotides 1 – 20 correspond to region 1 (Figure 2) starting at TGTTGG and extend inwards into the repeat. (**b**) Sequence of the putative right hand end of the repeated element, starting at nucleotide position 8298. Sequences of intervening, internal portions of the repeat are incomplete; nucleotide positions are approximate and based partly on restriction mapping data. Sequences 1 – 277 and 8331 – 8607 are homologous direct repeats terminated by short inverted repeats TGTTGG....CTAACA (underlined with arrows). The 'tRNA binding site region' adjacent to the left LTR, and the oligopurine sequence adjacent to the right LTR, are underlined. Elements in the LTR sequences resembling control elements are boxed and referred to in the text.

dition signals (AATAAA, nucleotide positions 217 and 8546, Figure 5; Montell *et al.*, 1983) and possible 'TATA' box sequences (nucleotide positions 34 and 8362; Figure 5; Goldberg, 1979; Kimmel and Firtel, 1983).

Discussion

The structure of the DNA clone PL15 is typical of segments derived from the *Hpa*II-resistant, hypermethylated compartment of the *Physarum* genome (Peoples *et al.*, 1985). These domains of hypermethylated DNA consist predominantly of multiple copies of partly scrambled DNA segments derived from a single, long repeated element, but occasionally additional less-abundant repetitive elements are also found in this DNA fraction (Peoples *et al.*, 1985). This unusual, apparently disorganised, sequence arrangement is observed for a significant fraction of the repetitive DNA in other eukaryotic genomes (Musti *et al.*, 1981; Eden *et al.*, 1981). Wensink *et al.* (1979) were the first to suggest that scrambled repeated sequence clusters might arise by DNA transposition. Although it has still not been proved directly in this study that such a process has led to the formation of the clusters of methylated repeated sequences in *Physarum* DNA described here, the abundant repetitive element that is a dominant feature of these regions has all the hallmarks of a transposon-like sequence. The observation that LTR-like elements are located at the precise points of discontinuity in all of three instances studied favours the involvement of a directed, *in vivo* recombination mechanism and argues against the assertion that the scrambled arrangements of these sequences may be a DNA cloning artifact. The heat-shock responsive transposable element DIRS in *Dictyostelium* (Capello *et al.*, 1984) is another reported instance of preferential integration of a eukaryotic transposon into sites located within its own sequence.

The transposon-like repetitive element is estimated to be ~8.6 kb long with direct LTR-like sequences terminated by short inverted repeats (5'TGTTG...TAACA3') typical of certain well-defined eukaryotic transposons (Gafner and Philippsen, 1980; Levis *et al.*, 1980). From the estimated repetition frequency of these elements (2000 – 5000 copies/haploid genome; Peoples and Hardman, 1983) they occupy $\sim 1.7-4.2 \times 10^7$ bp, or 10 – 20% of the *Physarum* genome. Since 20% of the genome is found in the *Hpa*II-resistant compartment in which these sequences are exclusively located (Whittaker *et al.*, 1981) this is in keeping with previous studies suggesting that clusters of these elements account for the major portion of the DNA in this fraction (Peoples and Hardman, 1983). The size of these sequence clusters (20 – 40 kb, Whittaker and Hardman, 1980) suggests that there may be >1000 per genome equivalent, containing up to 4 – 5 repeat copies per cluster. The analysis presented in Figure 3b indicates that transposition/recombination may take place between mutually inverted copies as well as directly oriented copies of the element. This would lead to sequence arrangements with the potential of forming the complex, clustered 'foldback' structures previously recognised as major components of the inverted-repeat fraction of *Physarum* DNA (Hardman *et al.*, 1979a); they probably account for >50% of the foldback structures in *Physarum* DNA (Gerrie *et al.*, 1983). Similar clustered foldback sequences are also a feature of other eukaryotic genomes (Hardman *et al.*, 1979b).

Since sequences within the repeated element reported here are preferred sites for transposition/recombination they were examined for common structural features. They show no obvious sequence homology but it is interesting that in two cases (regions 3 and 4) the axes of symmetry of small perfect palindromes (TATTTG^CAAATA and GGCT^AGCC) lie at the putative site of recombination, and in the third case (region 4') recombination occurs near the centre of a 12-nucleotide alternating pyrimidine/purine sequence (GTATACAC^ACAC) with Z-DNA like properties (Nordheim *et al.*, 1982). It is possible that these special structural features are recognised as favoured sites for transposition/recombination, but it does not in itself offer a ready explanation for preferential targeting to sites within the repeated element since such small sequences might also be expected to occur elsewhere in the genome. Other factors may therefore be involved in modulating the availability of potential target sites for transposition/recombination that have yet to be discovered.

Materials and methods

Strains and DNA clones

Nuclear DNA was isolated from microplasmodia of *P. polycephalum* (Cooke and Dee, 1975) as described previously (Hardman *et al.*, 1979b). The preparation of the DNA library of *Sau*3aI-partially digested *Physarum* genomic DNA fragments using Lambda-1059 was described by Whittaker (1982). PL5 and PL15 were selected by Benton and Davis (1977) hybridisation screening using the *Hpa*II-

repeat-containing *Physarum* genomic DNA clones pPH29 and pPH53a. The construction and characterisation of these DNA clones has been described by McLachlan and Hardman (1982). PL5-s5 is an internal subclone of PL5 containing a 4.3-kb *Msp*I restriction fragment of the *Physarum Hpa*II-repeat constructed by Peoples *et al.* (1985).

DNA hybridisation

Restriction fragments were separated on 1% agarose gels and blotted using the sandwich-blot method of Smith and Summers (1980), based on the original procedure of Southern (1975). Conditions of hybridisation and methods used for the preparation of nick-translated DNA probes have been described previously (Peoples *et al.*, 1985). Typically, probes were labelled to specific activities $>1 \times 10^8$ c.p.m./μg of DNA and autoradiographs were exposed for 16–24 h.

DNA sequence analysis

DNA sequencing was performed using the dideoxynucleotide chain termination method (Sanger *et al.*, 1977). Appropriate subcloned DNA restriction fragments were inserted into M13mp8 or M13mp9 vectors (Messing and Vieira, 1982). Radiochemicals and other materials for DNA sequencing were obtained from Amersham International PLC (UK).

Acknowledgements

We are grateful to the Medical Research Council and Science and Engineering Research Council for supporting various aspects of this work, and to Professor H.M.Keir for provision of research facilities. DHP was in receipt of a Medical Research Council Research Studentship.

References

Benton,W.D. and Davis,R.W. (1977) *Science (Wash.)*, **196**, 180-182.
Boeke,J.D., Garfinkel,D.J., Styles,C.A. and Fink,G.R. (1985) *Cell*, **40**, 491-500.
Cameron,J.R., Loh,E.Y. and Davis,R.W. (1979) *Cell*, **16**, 739-751.
Cappello,J., Cohen,S.M. and Lodish,H.F. (1984) *Mol. Cell. Biol.*, **4**, 2207-2213.
Cooke,D.J. and Dee,J. (1975) *Genet. Res.*, **24**, 175-188.
D'Eustachio,P. and Ruddle,F.H. (1983) *Science (Wash.)*, **220**, 919-924.
Di Nocera,P.P., Digan,M.E. and Dawid,I.B. (1983) *J. Mol. Biol.*, **168**, 715-727.
Doring,H.P., Tillman,E. and Starlinger,P. (1984) *Nature*, **307**, 127-130.
Dowsett,A.P. and Young,M.W. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4570-4574.
Eden,F.C., Musti,A.M. and Sobieski,D.A. (1981) *J. Mol. Biol.*, **148**, 129-151.
Finnegan,D.J., Rubin,G.M., Young,M.W. and Hogness,D.S. (1977) *Cold Spring Harbor Symp. Quant. Biol.*, **42**, 1053-1063.
Flavell,A.J. and Ish-Horowicz,D. (1983) *Cell*, **34**, 415-419.
Gafner,J. and Philippsen,P. (1980) *Nature*, **286**, 414-418.
Gerrie,L.M., Humphreys,J., Peoples,O.P. and Hardman,N. (1983) *Biochim. Biophys Acta*, **741**, 214-223.
Goldberg,M.L. (1979) Ph.D. Thesis, Standford University.
Hardman,N., Bell,A.J. and McLachlan,A. (1979a) *Biochim. Biophys. Acta*, **564**, 372-389.
Hardman,N., Jack,P.L., Brown,A.J.P. and McLachlan,A. (1979b) *Eur. J. Biochem.*, **94**, 179-187.
Hauber,J., Nelbock-Hochstetter,P. and Feldmann,H. (1985) *Nucleic Acids Res.*, **13**, 2745-2759.
Houck,C.M., Rinehart,F.P. and Schmid,J. (1979) *J. Mol. Biol.*, **132**, 289-306.
Kimmel,A.R. and Firtel,R.A. (1983) *Nucleic Acids Res.*, **11**, 541-552.
Kleckner,N. (1981) *Annu. Rev. Genet.*, **15**, 341-404.
Levis,R., Dunsmuir,P. and Rubin,G.M. (1980) *Cell*, **21**, 581-588.
McLachlan,A. and Hardman,N. (1982) *Biochim. Biophys. Acta*, **697**, 89-100.
Messing,J. and Vieira,J. (1982) *Gene*, **19**, 269-276.
Montell,C., Fisher,E.F., Caruthers,M.H. and Berk,A.J. (1983) *Nature*, **305**, 600-605.
Musti,A.M., Sobieski,D.A., Chen,B.B. and Eden,F.C. (1981) *Biochemistry (Wash.)*, **20**, 2989-2999.
Nordheim,A., Lafer,E.M., Peck,L.J., Wand,J.C., Stollar,B.D. and Rich,A. (1982) *Cell*, **31**, 309-318.
Peoples,O.P. and Hardman,N. (1983) *Nucleic Acids Res.*, **11**, 7777-7788.
Peoples,O.P., Whittaker,P.A., Pearston,D.H. and Hardman,N. (1985) *J. Gen. Microbiol.*, **131**, 1157-1165.
Saigo,K., Kugimiya,W., Matsuo,Y., Inouye,S., Yoshioka,K. and Yuki,S. (1984) *Nature*, **312**, 659-661.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Shepherd,N.S., Scwartz-Sommer,Z., Blumberg vel Spalve,J., Gupta,M., Weinand,U. and Saedler,H. (1984) *Nature*, **307**, 185-187.
Singer,M.F. (1982) *Cell*, **28**, 433-434.
Singer,M.F. and Skowronski,J. (1985) *Trends Biochem. Sci.*, **10**, 119-122.
Shimotohno,K., Mizutani,S. and Temin,H.M. (1980) *Nature*, **285**, 550-554.
Smith,G.E. and Summers,M.D. (1980) *Anal. Biochem.*, **109**, 123-129.
Southern,E.M. (1975) *J. Mol. Biol.*, **93**, 51-69.
Spradling,A.C. and Rubin,G.M. (1981) *Annu. Rev. Genet.*, **15**, 219-264.
Starlinger,P. (1984) *Trends Biochem. Sci.*, **9**, 125-127.
Temin,H.M. (1980) *Cell*, **21**, 599-600.
Toh,H., Hayashida,H. and Miyata,T. (1984) *Nature*, **305**, 827-829.
Ullu,E. and Tschudi,C. (1984) *Nature*, **312**, 171-172.
Varmus,H.S. (1982) *Science (Wash.)*, **216**, 812-820.
Wensink,P.C., Tabata,S. and Pachl,C. (1979) *Cell*, **18**, 1231-1246.
Whittaker,P.A. (1982) Ph.D. Thesis, University of Aberdeen.
Whittaker,P.A. and Hardman,N. (1980) *Biochem. J.*, **191**, 859-862.
Whittaker,P.A., McLachlan,A. and Hardman,N. (1981) *Nucleic Acids Res.*, **9**, 801-813.
Young,M.W. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 6274-6278.

Received on 10 September 1985; revised on 11 October 1985