Volume 11 Number 22 1983 Nucleic Acids Research

An abundant family of methylated repetitive sequences dominates the genome of *Physarum polycephalum*

Oliver P.Peoples and Norman Hardman

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

Received 8 September 1983; Accepted 19 October 1983

ABSTRACT

A significant portion (20%) of the Physarum genome can be isolated as a HpaII-resistant, methylated fraction. Cloned DNA probes containing highly-repeated sequences derived from this fraction were used to define the pattern of structural organisation of homologous repeats in Physarum genomic DNA. It is shown that the probes detect an abundant, methylated family of sequences with an estimated genomic repetition frequency greater than 2100, derived from a large repeated element whose length exceeds 5.8kb. Sequences comprising the long repetitive element dominate the HpaII-resistant compartment and account for between 4-20% of the Physarum genome. Detailed restriction/hybridisation analysis of cloned DNA segments derived from this compartment shows that HpaII/MspI restriction sites within some copies of the long repeated sequence are probably deleted by mutation. Additionally, segments of the repeat are often found in different organisational patterns that represent scrambled versions of its basic structure, and which are presumed to have arisen as a result of recombinational rearrangement in situ in the Physarum genome. Preliminary experiments indicate that the sequences are transcribed and that the structural properties of the repeat bear some resemblance to those of transposable genetic elements defined in other eukaryotic species.

INTRODUCTION

Repetitive sequences are found in DNA from a wide variety of eukaryotic species. In most cases the repetitive component accounts for a substantial proportion of the total complement of genomic DNA. Both short (1) and long (2) interspersed repeated elements have been identified, in addition to repetitive sequence clusters (3,4), illustrating the diversity of their organisation and structure. Apart from the transposable genetic elements of Drosophila DNA (5,6) surprisingly few highly-abundant 'middle-repetitive' type sequence families have been identified and studied in detail. Perhaps the best-characterised is the 'Alu-1' sequence family; the short interspersed repetitive elements in human DNA which account for 3% of the genome and dominate the middle-repetitive component (7). A family of long (ca. 5kb) interspersed highly-repetitive elements has more recently been identified in human DNA (8).

In neither case is their function known, but it has been argued using indirect lines of evidence that these interspersed sequence families may be transposable genetic elements.

Like human DNA, the genome of the eukaryotic slime mould Physarum polycephalum contains both long and short interspersed repetitive sequences (9) but no individual repetitive elements other than the nucleolar rDNA satellite (10) have so far been characterised. In the present study, cloned repetitive DNA probes were used to identify and characterise an abundant family of sequences derived from a long repetitive element which accounts for between 4-20% of the Physarum genome. These elements thus represent a substantial sequence component which corresponds to an amount of DNA at least two- to four-fold greater than the entire rDNA complement in this organism.

EXPERIMENTAL METHODS

Sources of DNA and cloning methodology

Nuclear DNA from Physarum polycephalum strain LU648 x LU688 (11) was used in this work, and was prepared as described previously (12).

pPH29 and pPH53 plasmid DNA clones containing Physarum nuclear DNA sequences were constructed, selected and characterised by McLachlan & Hardman (13). The PL- series of clones were prepared from Sau3A1 partially-digested Physarum DNA after size-selection and alkaline-phosphatase treatment by cloning into the BamH1 restriction site of the vector Lambda-1059. The clone bank was prepared by Whittaker (14) and contains approximately 2-6 x 10^4 recombinants. Recombinant phage DNA was purified before restriction using the method described by Whittaker (14).

DNA blotting and hybridisation

DNA blots were prepared using DNA fragments separated on agarose using the 'sandwich blot' method (15) based on the technique of Southern (16). Filters were hybridised to ^{32}P nick-translated DNA probes labelled to a specific activity of 0.5-1.0 x 10^8 cpm/microgramme of DNA (17). Pre-hybridisation and hybridisation conditions were as described previously (13) except that the hybridisation mixtures also contained 100 microgrammes of sonicated salmon sperm DNA/ml, and dextran sulphate was omitted. Autoradiograms were developed using Fuji-Rx film after exposure for 4-10 h at -50°C.

RESULTS & DISCUSSION

Linkage of pPH29- and pPH53-specific repeated elements in Physarum DNA segments

pPH29 is a BamHl/HindIII restriction fragment of Physarum genomic DNA about 4.3 kb in length, cloned using the plasmid vector pBR322. pPH53 contains a BamHl/HindIII Physarum DNA segment of similar length with an additional internal BamHl cleavage site. The two DNA clones show no cross-homology in Southern blotting experiments (13). It is not known whether the BamHl/BamHl and BamHl/HindIII segments of the pPH53 insert were originally contiguous in the genome or whether they were joined together by ligation during cloning. For some purposes these two segments were separated and purified by sub-cloning and were designated pPH53a and pPH53b, respectively. pPH29 and pPH53a/b all contain highly-repeated sequences which hybridise exclusively to the HpaII-resistant compartment of Physarum genomic DNA (18, 19) which comprises 20% of the total nuclear DNA complement (18). It was therefore of interest to determine whether these repeated elements were associated together in DNA segments derived from this fraction of the genome.

An extensive clone bank of Physarum genomic DNA has been prepared in this laboratory by Whittaker (14). Replicate Benton-Davis (20) filters containing recombinant phage plaques were prepared from this clone bank and hybridised to nick-translated ^{32}P-labelled pPH29 or pPH53 DNA. In one series of experiments 20% (171/850) of all plaques screened generated a positive hybridisation signal using a pPH29 DNA probe, and 22% (189/850) hybridised to a pPH53 DNA probe. About four-fifths of these (151/189) were recombinant phage which showed homology with both DNA probes. No significant hybridisation to these clones was observed using a pBR322 control DNA probe when auto-radiograms were exposed under conditions similar to test filters. Using a DNA probe prepared from total Physarum nuclear DNA, which would be expected to generate the strongest signal with clones containing the most highly-repetitious sequences, the same recombinant phage were detected that displayed homology with pPH29 and pPH53 DNA. Assuming that the DNA clone bank is reasonably representative it can be concluded that the most repetitive sequences are found in about 20% of Physarum genomic DNA fragments 11-16 kb long (the size range of the cloned DNA segments) and that sequences showing homology to the highly-repetitive elements located in pPH29 and pPH53 are linked together in the majority of these DNA segments. A small number of these recombinant phage were selected for more detailed analysis.

Restriction mapping and hybridisation properties of Physarum DNA segments containing highly-repeated sequences

DNA was isolated from a random collection of sixteen lambda-1059 derived Physarum recombinant DNA clones which generated a hybridisation signal with a pPH29 and pPH53 DNA probe. These were designated PL-1 to PL-16. Purified DNA from the clones produced a unique restriction pattern in each case using EcoR1, indicating that distinct cloned DNA segments had been selected. Of these, six DNA clones (PL-3, 5, 12, 13, 15, 16) were analysed in more detail by restriction mapping using BamH1, HindIII and HpaII. Additionally, Southern blots of DNA fragments generated using these restriction nucleases were used to map the position of the regions of sequence homology with the pPH29 and pPH53a/b DNA probes. A summary of the data is presented in Figure 1.

The first conclusion from this analysis is that virtually all DNA fragments in the clones generated by cleavage with these enzymes displayed homology with one or other of the DNA probes. The only regions which failed to hybridise with any of these DNA probes were located at the ends of the cloned segments (Figure 1, PL-13, 15 and 16; L-, R- and R- ends, respectively). In general, all the clones contained segments which displayed an alternating hybridisation pattern with pPH29 or pPH53 sequences. Fragments with identical restriction/hybridisation properties were found in some of the clones. For example, in HindIII/HpaII digests 2.7 kb HindIII/HpaII terminated fragments were detected in PL-5, 13 and 15 which hybridised only to pPH29, and were of similar size to the HindIII/HpaII segment of this plasmid DNA probe as determined from its restriction map (Figure 1). Similarly, 1.6 kb BamH1/HpaII fragments were detected in PL-3, 12 and 15 which annealed only to pPH29 and corresponded in size to the remaining segment comprising the cloned Physarum DNA insert in pPH29. Similar results were obtained for segments of the clones which hybridised only to pPH53; 1.5 kb HindIII/HpaII fragments were found in PL-5 and 15 which annealed only to pPH53b. In these cases, however, no HindIII/HpaII fragment of corresponding size was detected in the pPH53b clone (no HpaII sites were present).

For three of the DNA clones (PL-5, 13 and 15) their restriction maps could be arranged in such a manner that they formed an overlapping pattern (Figure 1). The simplest explanation for the regular DNA restriction and hybridisation properties of these DNA clones is that they contain parts of a large repeated element, one segment corresponding to the BamH1/HindIII pPH29 DNA sequence, and the other consisting of a pPH53b-like element.

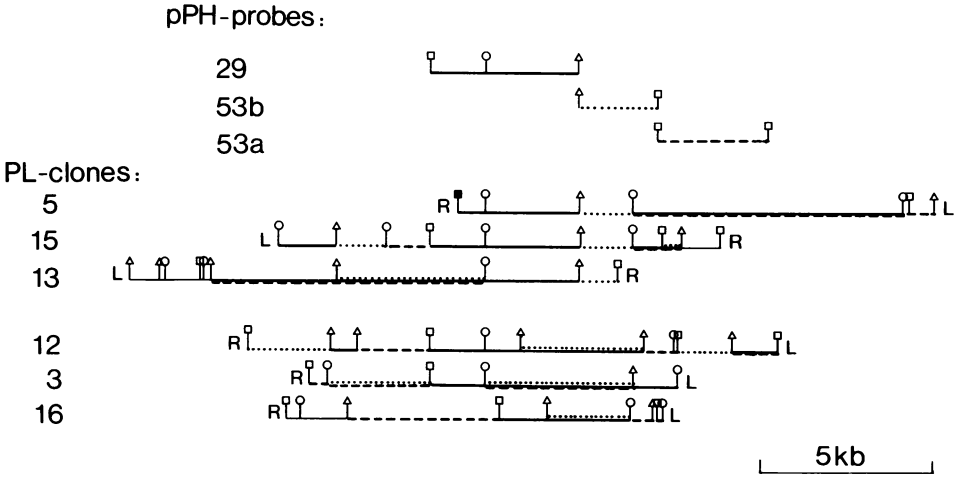

FIGURE 1: Restriction/hybridisation analysis of cloned Physarum DNA segments

The PL- DNA clones analysed, and the pPH- DNA probes used are referred to in the text. Only Physarum DNA inserts are shown. L- and R- refer to the ends of the PL- clones which were attached to the left and right arms of the Lambda DNA cloning vehicle. Restriction maps were constructed from combination digests using BamHl (□); HindIII (△) and HpaII (○). The R- end of PL-5 contained a BamHl cleavage site at its junction with Lambda DNA sequences which was deleted on cloning (■). The restriction maps of the DNA probes used have been published (ref. 22) and are aligned, where possible, with homologous segments in the PL- clones (see text). The probes were used to define the position of fragments showing homology with pPH29 (——); pPH53b (·····); pPH53a (----) or segments with no homology (——).

Additionally, the circularly-permuted arrangement of DNA segments in PL-15 indicated that in this instance two copies of the large repeated element, or segments thereof, may be arranged in tandem.

Close inspection of the data presented in Figure 1 indicates that this simple explanation cannot fully account for the restriction/hybridisation properties of the majority of DNA segments in the clones. Instead the clones appear to consist of scrambled assemblies of DNA fragments showing little resemblance to the regular organisational pattern discussed above, but nevertheless containing some elements with the properties expected for internally-cut regions of the postulated large repeated sequence (for example the 1.6 kb BamHl/HpaII pPH29-specific segments of PH-3 and PL-12).

Characterisation of a major repeated sequence family in Physarum genomic DNA

In order to investigate further the suggestion that pPH29 and pPH53 DNA

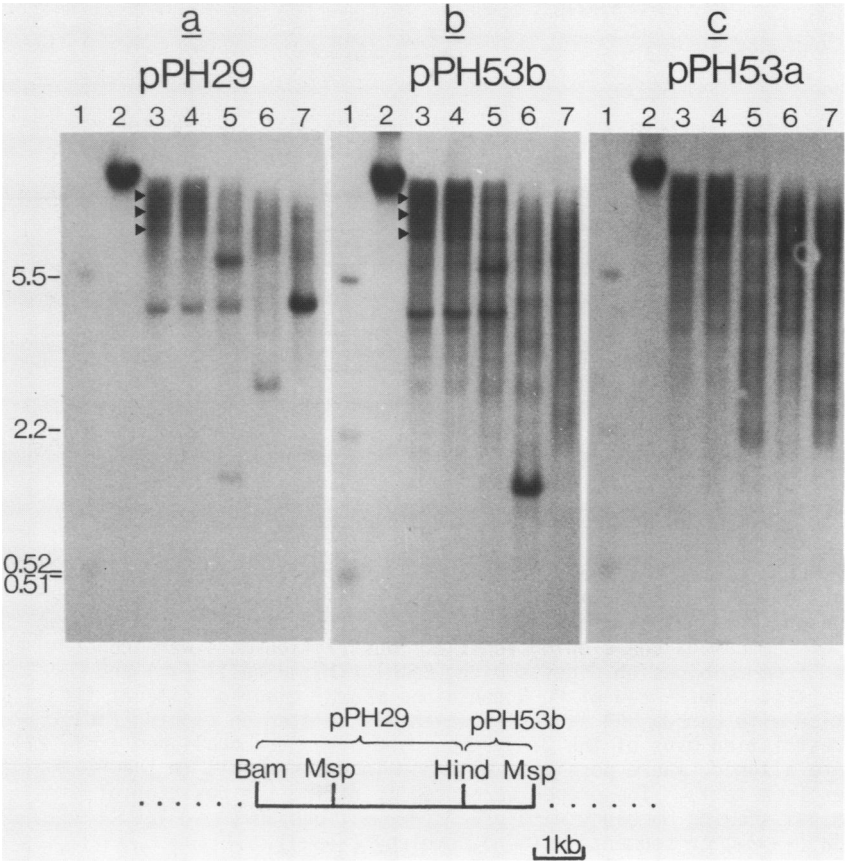

FIGURE 2: Genomic organisation of pPH29- and pPH53-related sequences

Samples of Physarum nuclear DNA (1 microgramme) digested using various restriction nucleases, were electrophoresed on agarose slab gels, transferred to nitrocellulose and probed with either: (a) pPH29, (b) pPH53b or (c) pPH53a, labelled with ^{32}P by nick-translation. Sizes of DNA fragments were determined with reference to Lambda/HindIII + ^{32}P-labelled PM2/HindIII restriction fragments (lane 1). Closed triangles indicate the position of minor, high molecular weight DNA bands in (a) and (b). Physarum DNA samples were digested with: (2) HpaII; (3) MspI; (4) HpaII + MspI; (5) BamHl + MspI; (6) HindIII + MspI; (7) BamHl + HindIII. The most frequent arrangements for the BamHl/MspI/HindIII restriction sites surrounding elements related to pPH29 and pPH53 sequences are shown.

elements form part of a larger repeated sequence, blots of Physarum nuclear DNA digested using various restriction nucleases were probed using pPH29 and pPH53a/b DNA. The results are shown in Figure 2.

The pPH29 DNA probe (Figure 2a) generated a major band of hybridisation

at 4.3 kb in BamHl/HindIII digests of Physarum genomic DNA (lane 7) indicating that this arrangement of restriction sites is the most common one for DNA segments containing pPH29-related repeated sequences. The uniformity in organisation of pPH29-like elements was further underlined by the observation that in HindIII/MspI digests (Figure 2a, lane 6) a band of hybridisation is found at the position expected for the 2.7 kb HindIII/HpaII(MspI) fragment of pPH29 (see Figure 1). Comparison with the MspI tracks (lanes 3 and 4) indicates that these fragments arise by cleavage of a 4.3 kb MspI-generated DNA band. The expected 1.6 kb DNA band corresponding to genomic copies of the other segment of pPH29-like elements can be seen in BamHl/MspI digests (lane 5) together with bands at 4.3 kb and 5.8 kb.

These data lead to the postulate that the most common arrangement of restriction sites within and surrounding the pPH29-like repeated elements in Physarum genomic DNA is as depicted in Figure 2. Several additional features of the data are worthy of note. First, comparison of lanes 2-4 (Figure 2a) show that the MspI recognition sites in these regions of genomic DNA are methylated at CpG doublets, thus conferring on the DNA resistance to digestion with HpaII. This confirms earlier work (18). The identical MspI and HpaII+MspI digest patterns (lanes 3 and 4, Figure 2a) additionally shows that external-C methylation at these sites is either absent, or present only as 5'meCpmeCpGpGp3'. The presence of minor, high-molecular weight MspI DNA bands indicates a longer-range organised distribution of restriction sites in Physarum DNA surrounding some pPH29-like repeated elements (Figure 2, lane 3). This could be explained by deletion of MspI restriction sites within a larger repeating unit and/or by the presence of tandem arrangements of the repeat, consonant with the circularly-permuted restriction/hybridisation data for the PL-15 clone discussed above. Control experiments showed that DNA digestion was complete, so that the higher-molecular weight fragments did not arise due to partial cleavage of the DNA with MspI. The idea that MspI restriction sites may be deleted by mutation is not unreasonable, since methylated residues are known to be mutational hotspots (21). That the higher molecular weight MspI DNA bands probably form part of the same repeating structure described above is supported by the results shown in lane 5 (Figure 2a); the bands become less intense in BamHl/MspI double digests and the expected 1.6 kb pPH29-specific DNA band is generated together with an additional 5.8 kb band which would result if a proportion of the BamHl/HindIII pPH29-like segments comprising the proposed repeated sequence (Figure 2) contained a deleted HpaII/MspI site. Restriction site

heterogeneity within different copies of the repeated element similarly might account for the loss of the HpaII cleavage site within the overlapping restriction maps of PL-13 and PL-15 (Figure 1).

The results obtained using pPH53b as a probe (Figure 1 and 2b) suggest that this sequence is often contiguous with those of pPH29 and also forms part of the larger repeating unit described above. The sequence detects a 4.3 kb MspI DNA band (Figure 2b, lanes 3/4), and 5.8 kb DNA band generated by cleavage of genomic DNA using BamHl/MspI (lane 5) and a 1.5 kb fragment in HindIII/MspI digests (lane 6) similar in size to the DNA fragments in the cloned segments showing homology with pPH53b (Figure 1) which are co-linear with pPH29-like sequences in most cases. The strong DNA band in HindIII/HpaII digests at 1.5 kb demonstrates that this is a common arrangement for these restriction sites, and that the plasmid-cloned copy of the pPH53b sequence probably has a sequence in which the HpaII/MspI site has been deleted, possibly by mutation as described above for other HpaII/MspI restriction sites in the vicinity of these repeated elements.

In contrast to pPH29 and pPH53b, a pPH53a DNA probe generated a hybridisation signal (Figure 2c) indicating that it is not derived from within the proposed repeating unit shown in Figure 2. The hybridisation signals for each restriction digest were complex, though structured, indicative of a diverse arrangement of restriction sites in the regions of genomic DNA containing pPH53a-like elements. These sequences are, however, confined to the HpaII-resistant compartment of the genome, as are pPH29 and pPH53b repeated sequences (Figure 2c, lane 2 and ref.18) and they are found in close association with these other elements in all the independently-isolated PL- clones studied here. Moreover, in PL-15, pPH53a-like elements can be seen to flank a pPH29 + pPH53b like sequence whose structure conforms to that of the proposed repeating unit illustrated in Figure 2. It seems reasonable to conclude that pPH53a-related repeats may be adjacent to those of pPH29 + pPH53b repeated units forming part of a larger repeating structure. One explanation for the heterogeneous restriction pattern of DNA segments displaying homology with this sequence may be that the sequences may be positioned at the ends of the large repeated sequence , forming the junctions between copies of the repeated element and the surrounding DNA. More recent work, however, indicates that pPH53a-related sequences may have an inherently more variable structure than those of pPH29 or pPH53b (results not shown).

There are additional aspects of the results presented here which cast

doubt on an entirely simple, straightforward explanation for the nature and organisation of the repeated sequences described here. The complex, structured background hybridisation observed in genomic DNA digests probed using pPH29 or pPH53b demonstrates that a portion of the segments containing these repeated elements also have a diverse arrangement of restriction sites similar to pPH53a sequences, consistent with the partially scrambled nature of the repeated sequences within the cloned DNA segments shown in Figure 2. Several mechanisms can be envisaged to lead to apparent 'scrambling' of the restriction patterns of segments containing the large repeated sequence; addition or deletion of restriction sites by mutation as described above, or insertion/deletion/inversion of sequences brought about by some recombinational mechanism. We have insufficient data at present to distinguish between these possibilities, but the complex arrangements of repeats shown in Figure 1 already excludes the possibility that scrambling simply results from restriction site heterogeneity alone.

Genomic organisation of the HpaII-repeated sequence family and comparison with other eukaryotic species

Although we have been unable to completely define some of the complex organisational arrangements involving the repeated sequence family described here, which we term the HpaII-repeat family, some useful comparisons with data obtained for other eukaryotic DNAs can be made. The basic unit of the HpaII-repeat family appears to be a sequence in excess of 5.8 kb long, which is confined to DNA segments comprising about 20% of the Physarum genome. DNA clones have been isolated which contain comparatively long, contiguous intact segments of the repeated sequence, but a large number of the clones also contain scrambled arrangements of fragments of the repeated element. This 'scrambling' is in some respects similar to that observed for the methylated, highly-abundant sequence component of chicken DNA (3) and the clustered repetitive elements in Drosophila DNA (4). It may therefore be a general phenomenon in eukaryotic genomes. Scrambling precludes an accurate estimation of the repetition frequency of the HpaII sequence family from the data presented, but an approximate minimum estimate can be made from the frequency of occurence of recognisable fragments of the repeat in the cloned segments shown in Figure 1. Each of the three major restriction fragments forming the 5.8 kb of the repeat defined here are each represented three times in the six DNA clones analysed in detail. Together, the cloned segments contain 78.6 kb of Physarum DNA. Since the genome of Physarum contains 2.7×10^5 kb, and the repeats are confined

to 20% of the genome (Figure 1 and ref. 18) it can be calculated that approximately 2100 copies of these elements are present in this DNA compartment in a recognisably unaltered form. Collectively this accounts for 11.9 x 10^3 kb of DNA, or 4% of the genome. This is clearly a considerable underestimate of the true repetition frequency of elements related to pPH29 or pPH53b, since at least as many additional homologous DNA fragments are found in the scrambled arrangements of these repeats in the DNA clones shown in Figure 1. These sequences thus form a major component of the Physarum genome; they represent between 4-20% of the genomic DNA complement and dominate the highly-repetitive DNA compartment (18, 19).

The Physarum HpaII-repeat family in some respects resembles the clustered repeats in chicken DNA (3) in that they appear to be confined to a minority fraction (20%) of the genome, they are highly-abundant, and they are methylated and organised in scrambled arrangements. In Physarum DNA methylated HpaII and HhaI restriction sites are found in regions containing the HpaII-repeats and the HpaII sites occur infrequently such that its DNA can be isolated as a HpaII-resistant fraction. A contributing factor to the low density of methylated HpaII sites found in this fraction may be that they are subject to a high degree of mutation (21). As described above, this may account for the observed loss HpaII/MspI restriction sites in some copies of the HpaII-repeat family which dominates this compartment. The structural properties of the DNA compartment containing these highly-repeated sequences leads to some interesting questions as to how it may have arisen and evolved. Wensink and co-workers (4) were the first to suggest that clustered arrangements of repetitive elements might be a consequence of multiple DNA transposition events involving target sites which already contained copies of homologous sequences, acting as hotspots for integrative recombination. DNA rearrangements, often associated with the presence of transposable elements, might accompany or follow the formation of such sequence clusters leading to 'scrambling' of their structure. If such an idea were correct the nature of the sequences which display homology with pPH53a are of special interest since they may include flanking regions and segments derived from the ends of the HpaII-repeats, as described above. Hence they may provide information about the way in which the sequences have undergone transposition. It may be of significance in this connection that similarities exist in the structure of pPH53a DNA sequences (ref. 22 and unpublished work) and the FB-transposable elements of Drosophila (23).

Regarding their possible function, preliminary work has shown that pPH29 and pPH53 DNA sequences are transcribed in Physarum to give RNA molecules of heterogeneous chain length (Peoples, O.P. & Hardman, N., unpublished work). Similar results have been obtained for a long repeated sequence family in human DNA (24). The role that such transcripts might serve is not known, though one suggestion has been that they may represent intermediates for transposition of DNA sequences. If this is the case for the HpaII-repeat family then it is also necessary to postulate the involvement of de novo sequence-specific methylation to account for the conserved methylation pattern of these sequences, since new, transposed copies of DNA sequences generated by such a mechanism would presumably lack hemimethylated template sequences.

ACKNOWLEDGEMENTS

The authors are grateful to the Medical Research Council for financial support, Mairi Gordon for excellent technical assistance, and Professor H M Keir for the provision of facilities in the Department of Biochemistry.

REFERENCES

1 Davidson, E.H., Hough, B.R., Amenson, C.S. & Britten, R.J.(1973) J . Mol. Biol. 77, 1-24
2 Singer, M.F. (1982) Cell 28, 433-434
3 Musti, A.M., Sobieski, D.A., Chen, B.B. & Eden, F.C. (1981) Biochemistry 20, 2899-2999
4 Wensink, P.C., Tabata, S. & Pachl, C. (1979) Cell 18, 1231-1246
5 Potter, S.S., Brorein, W.J. Jr., Dunsmuir, P. & Rubin, G.M. (1979) Cell 17, 415-427
6 Levis, R., Dunsmuir, P. & Rubin, G.M. (1980) Cell 21, 581-588
7 Houck, C.M., Rinehart, F.P. & Schmid, C.W. (1979) J. Mol. Biol. 132, 289-306
8 Manuelidis, L. & Biro, P.A. (1980) Nucleic Acids Res. 10, 3221-3239
9 Hardman, N., Jack, PL., Fergie, R.C. & Gerrie, L.M. (1980) Eur. J. Biochem. 103, 247-257
10 Vogt, V.M. & Braun, R. (1976) J. Mol. Biol. 106, 567-587
11 Cooke, D.J. & Dee, J. (1975) Genet. Res. 15 35-41
12 Hardman, N., Jack, P.L., Brown, A.J.P. & McLachlan, A. (1979) Eur. J. Biochem. 94, 179-187
13 McLachlan, A. & Hardman, N. (1982) Biochim. Biophys. Acta 697, 89-100
14 Whittaker, P.A. (1982) Ph.D. Thesis, University of Aberdeen
15 Smith, G.E. & Summers, M.D. (1980) Anal. Biochem. 109 123-129
16 Southern, E.M. (1975) J. Mol. Biol. 94, 51-69
17 Rigby, P.W.J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) J. Mol. Biol. 113, 237-251

18 Whittaker, P.A., McLachlan, A. & Hardman, N. (1981) Nucleic Acids Res. 9, 801-804
19 Gerrie, L.M., Humphreys, J., Peoples, O.P. & Hardman, N. (1983) Biochim. Biophys. Acta, in the press
20 Benton, W.D. & Davis, R.W. (1977) Science 196, 180-182
21 Bird, A.P. (1980) Nucleic Acids Res. 8, 1499-1504
22 Peoples, O.P., Robinson, A.C., Whittaker, P.A. & Hardman, N. (1983) Biochim. Biophys. Acta, in the press
23 Potter, S.S. (1982) Nature 297, 201-204
24 Kole, L.B., Haynes, S.R. & Jelinek, W.R. (1983) J. Mol. Biol. 165, 257