Retrotransposon-like nature of Tp1 elements: implications for the organisation of highly repetitive, hypermethylated DNA in the genome of *Physarum polycephalum*

Helen M.Rothnie*, Karen J.McCurrach, L.Anne Glover and Norman Hardman[1]
Department of Molecular and Cell Biology, University of Aberdeen, Marischal College, Aberdeen AB9 1AS, UK and [1]Ciba-Geigy Ltd, Biotechnology Section, CH-4002 Basel, Switzerland

Received November 1, 1990; Revised and Accepted December 10, 1990 EMBL accession no. X53558

ABSTRACT

The repetitive fraction of the genome of the eukaryotic slime mould *Physarum polycephalum* is dominated by the Tp1 family of highly repetitive retrotransposon-like sequences. Tp1 elements consist of two terminal direct repeats of 277bp which flank an internal domain of 8.3kb. They are the major sequence component in the hypermethylated (M+) fraction of the genome where they have been found exclusively in scrambled clusters of up to 50kb long. Scrambling is thought to have arisen by insertion of Tp1 into further copies of the same sequence. In the present study, sequence analysis of cloned Tp1 elements has revealed striking homologies of the predicted amino acid sequence to several highly conserved domains characteristic of retrotransposons. The relative order of the predicted coding regions indicates that Tp1 elements are more closely related to *copia* and Ty than to retroviruses. Self-integration and methylation of Tp1 elements may function to limit transposition frequency. Such mechanisms provide a possible explanation for the origin and organisation of M+ DNA in the *Physarum* genome.

INTRODUCTION

Families of mobile genetic elements are now recognised as being widespread in eukaryotic genomes (1). The term 'retrotransposon' (2) describes a class of transposable genetic elements capable of being mobilised by reverse transcription of an RNA intermediate. Retrotransposons are in many ways structurally and functionally analogous to the integrated proviral form of retroviruses and it seems likely that they have a common evolutionary origin (3,4). Retrotransposons are flanked by long terminal direct repeats (LTRs) and encode the functions required for their own transposition, most notably reverse transcriptase (5–8). The most widely documented and best studied examples of retrotransposons are *copia*-like sequence families in *Drosophila melanogaster* (9,10) and Ty elements in *Saccharomyces cerevisiae* (11,12). In addition to yeast and fruit flies, repetitive elements bearing the structural hallmarks of retrotransposons have been identified in a wide range of species including nematodes (13), mammals (14–16) and several plants (17–20). Retrotransposon-like sequences can be present at frequencies ranging from only a few (19) to many thousands of copies per genome (18). In some cases (e.g. in *Drosophila*), families of retrotransposons account for a major fraction of the middle repetitive DNA component (21) and contribute significantly to the frequency of spontaneous mutations. The apparent ubiquity of transposable elements, together with their potential effects on gene function, suggest that these sequences may have played an important role in shaping the eukaryotic genome, thus underlining the importance of reverse transcriptase in evolution (22).

Previous studies in our laboratory have focused on the nature and organisation of repetitive DNA sequences in the genome of the eukaryotic slime mould *Physarum polycephalum*. About one-third of the *Physarum* genome is composed of repetitive DNA sequences (23). Within this fraction, a dominant family of highly repeated sequence elements has previously been identified and partly characterised (24,25). This sequence family (previously referred to as HpaII-repeats) accounts for over half of the repetitive DNA fraction, and possibly for up to 20% of the *Physarum* genome (26). These elements have an unusual organisation in that they appear to be arranged exclusively in 'scrambled' clusters, up to 50kb in length, located within the HpaII resistant, methylated (M+) fraction of the genome (27). Initial sequence analysis (24) indicated that these repetitive sequences (referred to here as Tp1—Transposon *Physarum* 1) were retrotransposon-like in nature and led to the hypothesis that scrambling has resulted from transpositional insertion of copies of Tp1 into target sites present within its own sequence. The characterisation of the LTRs of Tp1 has already been described (24). In the present study evidence is presented that the Tp1 element has the potential to encode peptides with homology to the conserved regions of the nucleic acid binding, protease, endonuclease and reverse transcriptase domains characteristic of

* To whom correspondence should be addressed at Friedrich Miescher Institut, PO Box 2543, CH-4002, Basel, Switzerland

retrotransposons. The relative order of these coding domains allows Tp1 elements to be assigned to the group of eukaryotic retrotransposons which includes *copia* and Ty1.

MATERIALS AND METHODS

Physarum polycephalum genomic DNA was prepared as described (72), from the colonia-derived diploid strain LU648×LU688 (73). Genomic clones in the vector lambda-1059 were constructed and isolated as previously described (28). Subclones were generated in the plasmid vector pUC12 (25). Plasmid clones were propagated in *E.coli* NM522 (*hsd*D5 (rk^-, mk^-), Δ*lac pro*, *thi*, *sup*E, F′*pro*AB^+, *lac*IqzΔM15; 74) or *E.coli* JM83 (*ara*, Δ*lac pro*, *thi*, *str*A. φ80d*lac*ZΔM15, Gibco-BRL).

Standard protocols were used for isolation of plasmids, restriction digests, transformation of *E.coli*, etc. (75). DNA sequencing was performed using the dideoxynucleotide chain termination method (76) with the Sequenase enzyme and the kit supplied by United States Biochemical Corporation. Bluescript™ vectors were obtained from Stratagene (Heidelberg, FRG). In some cases, subclones of appropriate restriction fragments were constructed in the Bluescript™ vector pSKM13$^-$ and deletion clones allowing stepwise sequence analysis of the insert fragments were generated using exonucleaseIII/SI nuclease digestion (77). Alternatively, appropriate restriction fragments were subcloned in M13 vectors. Bluescript clones were propagated in *E.coli* XL1-Blue (*rec*A1, *lac*, *end*A1, *thi*, *hsd*R17, *sup*E44/ F′(*pro*AB, *lac*IqZΔM15), Stratagene). For M13 clones *E.coli* JM101 was used (*lac pro*, *sup*E, *thi*/F′(*tra*D36, *pro*AB, *lac*19, ZΔM15, Gibco-BRL).

DNA and protein homology searches were carried out on the Genbank and PROTEIN data libraries using the FASTN and FASTP programs (78). All other computer manipulations of data were carried out using the University of Wisconsin Genetics Computer Group program package, version 6 (79).

Sources of data used for sequence comparisons were as follows: Ta1-3 (17); Tnt1-94 (20); *copia* (5); 1731 (80); Ty3 (8); Ty 912 (7); 17.6 (6); 297 (38); MMoLV (81); RSV (82); HIV-1 (83).

RESULTS AND DISCUSSION

Structure of the *Physarum* genomic clone PL12

PL12 is one of a number of clones isolated from a *Physarum* genomic DNA library constructed in the phage vector lambda-1059 (28). These clones were isolated on the basis of their homology to previously identified, highly repetitive *Physarum* genomic DNA fragments cloned in pBR322: pPH29, pPH53a and pPH53b (29). Characterisation of the insert sequences of the pPH-series and the PL-series of clones has been reported previously (30,27). The PL12 DNA insert consists of two contiguous BamHI fragments of around 5kb and 10kb in length. These two fragments were subcloned in the plasmid vector pUC12 and designated PL12-HP1 and PL12-HP2 respectively (25). In order to further characterise Tp1, PL12-HP1 and PL12-HP2 were examined by restriction mapping, Southern blot hybridisation and nucleotide sequence analysis (25,31). During this analysis, additional non-Tp1 sequences were found, identifying a second retrotransposon-like repetitive sequence family referred to as Tp2. The properties of Tp2 are presented in detail elsewhere (32).

PL12 contains at least five partial copies of Tp1 and two copies of Tp2. Consensus maps of Tp1 and Tp2 and their arrangement in clone PL12 are presented in Figure 1. PL12 does not contain an intact, uninterrupted copy of Tp1. However, it was possible to reconstruct a putative full length sequence for Tp1 by merging the overlapping sequence data from regions I and II in Figure 1. This composite sequence file was used in all of the subsequent analysis.

Nucleotide sequence of Tp1: general features

The composite Tp1 sequence referred to above will appear in the EMBL, Genbank and DDBJ Nucleotide Sequence Databases under the accession number X53558. Two LTRs of 277bp flank an internal region of 8343bp in length. The LTRs are terminated by short inverted repeats and contain possible transcriptional control signals, in addition to primer binding sites analogous to those required for the initation of plus- and minus-strand DNA synthesis in retroviruses and retrotransposons. The putative tRNA binding sites of both Tp1 and the Tp2 element referred to above are identical to that of the *copia* element in *Drosophila melanogaster*, where a mechanism of primer binding involving an internal region of $tRNA_{met}$ has been proposed (33). A poly-purine sequence immediately precedes the 3′-LTR.

The overall base composition of Tp1 is 29% A, 24.4% C, 21.6% G and 25% T, but various internal stretches of the sequence have quite different base composition. For example, several stretches of up to 150bp in length composed almost entirely of homopurine.homopyrimidine (Pur.Pyr) sequence can be identified. It has been suggested that Pur.Pyr sequences can adopt triplex non-B-DNA structures which may play a role in making chromatin accessible to proteins involved in recombination (34). Recently, a Pur.Pyr sequence has been identified within a clustered group of retroposons isolated from the human genome (35). Such structural features may affect targeting of integration of further retro-elements within a cluster. Stretches of Pur.Pyr sequence also occur in the Tp2 element (32).

Another structural feature of the Tp1 sequence is the presence of numerous direct repeats and inverted repeats. The latter could give rise to the foldback structures previously observed in electron micrographs of DNA clones containing portions of the Tp1 element (29).

Retrotransposons are usually flanked by short direct repeats, which arise as a consequence of target site duplication at the site of insertion. Since clone PL12 does not contain a full length copy of Tp1, it was not possible to ascertain whether insertion of a Tp1 element into itself results in a target site duplication, although several sites where integration may have taken place have been identified. In common with most other retrotransposons (exceptions include 297 and 17.6; 36), there appears to be no target site sequence specificity for insertion of Tp1 into its own sequence, although the target sites observed thus far may have structural similarities (i.e. short palindromes or the potential to form Z-DNA structures; 24,25).

Tp1 sequences appear to be present exclusively in the hyper-methylated (M+) fraction of the genome. As expected for a methylated sequence, the frequency of the CpG dinucleotide in the Tp1 sequence is very much reduced compared to the frequency of all other possible dinucleotides. This can be accounted for by the phenomenon known as 'CpG suppression' (37), which results from the loss of CpG by spontaneous deamination of 5-methylcytosine (5-meC) to thymine. The

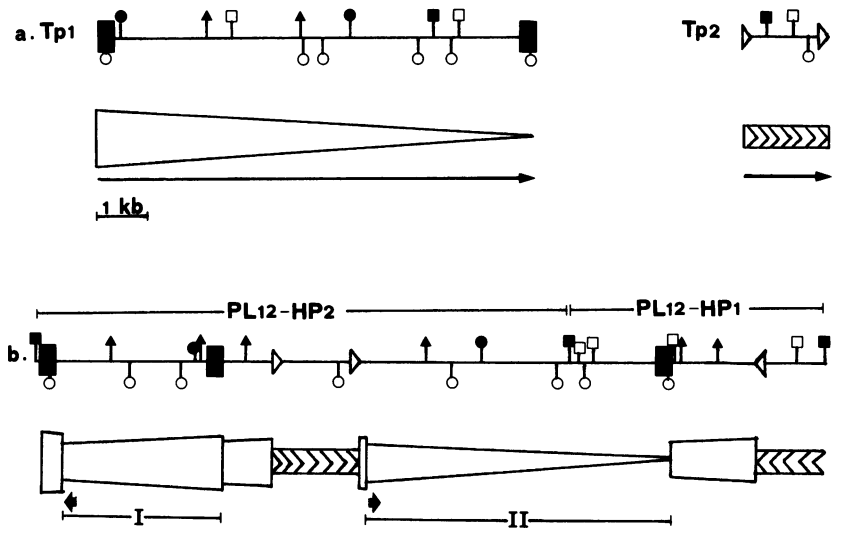

Figure 1a. Consensus maps of Tp1 (Pearston et al., 1985) and Tp2 (McCurrach et al., 1990) together with diagrammatic representations indicating the proposed direction of transcription (arrows). ■ BamHI ▲ HindIII ○ HincII □ EcoRI ● HpaII Tp1 LTR ▮ Tp2 LTR. Fig.1b. Restriction map of Physarum genomic clone PL12 indicating the fragments subcloned as PL12-HP1 and PL12-HP2. The relative positions of copies of Tp1 and Tp2 are represented in the diagram below. The arrowheads indicate the area of overlap used to align sequences from regions I and II to generate the Tp1 sequence analysed in this study.

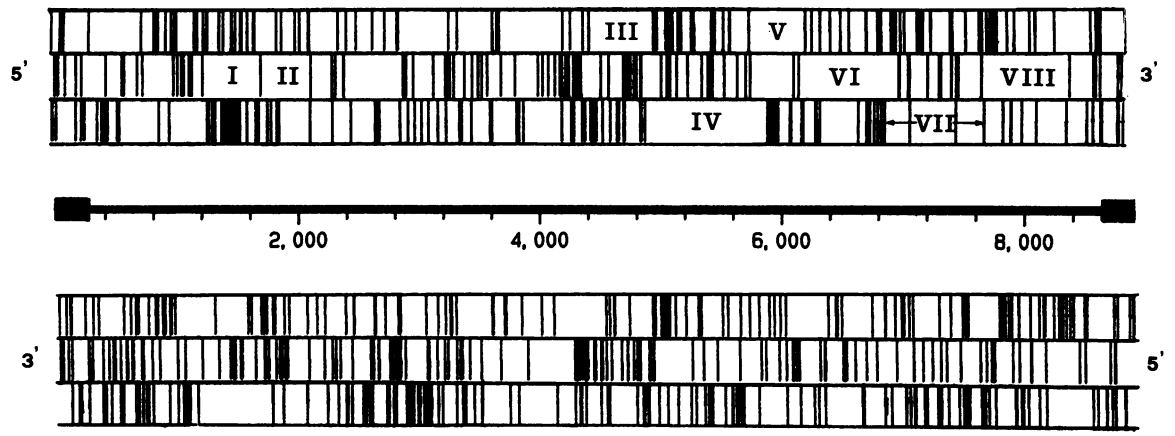

Figure 2. Distribution of termination codons in the amino acid sequence predicted from Tp1 in all six reading frames. Regions labelled I–VIII indicate the position of peptide sequences exhibiting homology to gag and pol-encoded polypeptides as described in the text.

possible significance of methylation with regard to transpositional activity of Tp1 and the structure of M+ DNA will be discussed below.

Coding Potential of Tp1

The distribution of stop codons in the composite Tp1 sequence is shown in Figure 2. Bearing in mind that the sequence comprises parts of two interrupted (and therefore probably non-functional) copies of Tp1, it was not expected that large, continuous open reading frames (ORFs) would be found. Instead, regions which could form part of longer reading frames were identified, peptide sequences were predicted and a search of the available databases made. A number of significant homologies to the nucleic acid binding, protease, endonuclease and reverse transcriptase domains of *copia* and other related elements were found. Significantly, these homologies were found only with peptides encoded by the predicted sense strand of Tp1 (as indicated in Figure 2). In several positions, stop codons were present within regions clearly homologous to retrotransposons. In other cases, e.g. within the region analogous to reverse transcriptase, it was necessary to change reading frames to follow the homology. These features would not be expected of a functional copy of Tp1. Also, in the absence of a functionally complete Tp1 sequence, it is impossible to say whether the functions potentially encoded by the element are contained within a single ORF (as is the case for *copia*, Ta1-3 and Tnt1-94) or if there are two or more overlapping reading frames (as in Ty1, Ty3, 17.6 and 297).

A diagram showing the relative position and degree of homology of conserved domains of retrotransposons with respect to the Tp1 element is presented in Figure 3. Overall, the genetic organisation of Tp1 most closely resembles that of the retrotransposons *copia*, Ta1, Tnt1 and Ty1.

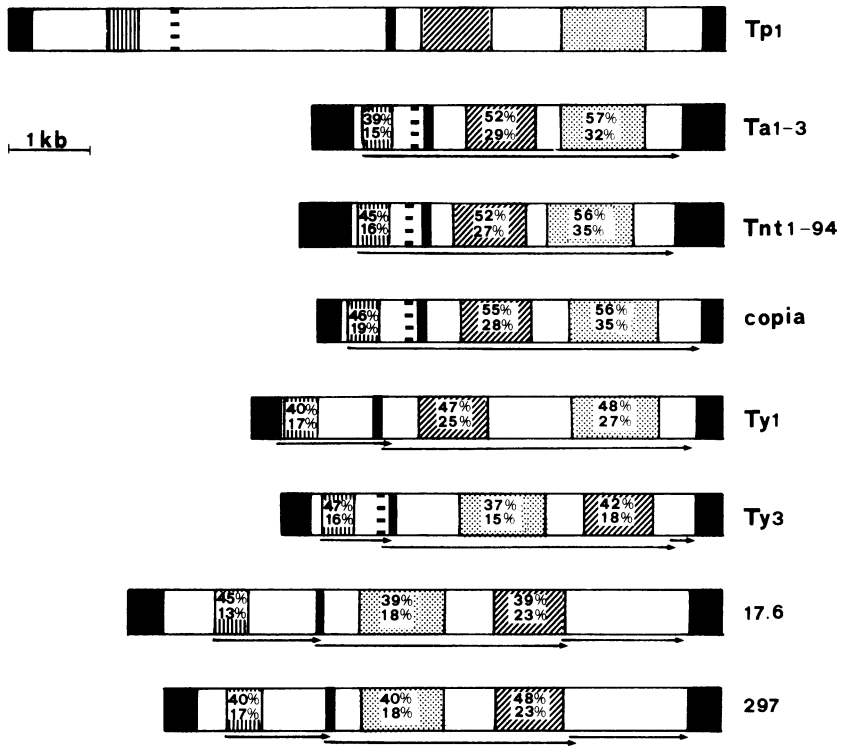

Figure 3. Organisation of conserved amino acid domains among retrotransposons Ta1−3, Tnt1−94, copia, Ty1, Ty3, 17.6 and 297 in comparison with Tp1. Elements are drawn to scale, with arrows representing the position and direction of open reading frames. Numbers in the shaded boxes indicate the% similarity and% identity (upper and lower values respectively), of the amino acid sequence compared with the corresponding region in Tp1. Identity = absolutely conserved amino acids; similarity = amino acids of similar properties as defined in the program BESTFIT/P (79). ▮ LTR, ▥ beginning of ORF1 in Ty1. Ty3, 17.6 and 297 and of the single ORF in Ta1−3, Tnt1−94 and copia, ≣ nucleic acid binding domain, ▮ protease domain, ▨ integrase (endonuclease) domain, ⁘ reverse transcriptase domain.

Homology to *gag* polypeptides

The degree of similarity of Tp1 to the equivalent *gag* region of retrotransposons was lower than that observed with the endonuclease and reverse transcriptase domains (see Figure 3). Nevertheless, peptide I (Figure 2) could be aligned with the beginning of the single ORF in Ta1-3, Tnt1-94 and *copia*, and with the beginning of ORF1 of Ty1, Ty3, 17.6 and 297. ORF1 in 297 and 17.6 shares some homology with the *gag* region of MMoLV (39). The corresponding region in *copia* has been shown to encode the major capsid protein of *copia*-related virus-like particles (40).

An amino acid motif in peptide II (Figure 2) was identified as being homologous to a highly conserved domain of nucleic acid binding proteins. In all replication competent retroviruses the nucleic acid binding protein is derived from the C-terminal half of the *gag* polypeptide (41). There are conflicting views regarding the function of this protein. Some authors suggest that it is required for the accurate positioning of the tRNA primer on the replication initiation site (42); others have evidence that it is important for ensuring that viral RNA rather than cellular RNA is packaged specifically into virion heads (43). Figure 4a shows the conserved amino acid motif $CX_2CX_4HX_4C$ in Tp1 and the nucleic acid binding domain of several retroviruses and retrotransposons. This motif is also conserved in Tp2 (32).

Protease, endonuclease and reverse transcriptase homologies

Certain amino acid sequences characteristic of the protease, endonuclease and reverse transcriptase domains of retroviruses are also highly conserved among retrotransposons. The active site of retroviral proteases contains the Asp-Thr-Gly sequence characteristic of aspartic proteases (44). Variants of this sequence are conserved in the protease region of retrotransposons. The aspartic acid residue is absolutely invariant and has been shown to be essential for proteolytic function in *copia* (45) and in the plant pararetrovirus CaMV (46). Figure 4b shows a comparison of part of peptide III (Figure 2) with the protease active site region in other retro-elements. The invariant Asp residue is also conserved in this Tp1 peptide.

Peptide IV (Figure 2) encodes an amino acid configuration capable of forming a zinc-binding domain which could interact with DNA. This zinc-binding region, which appears to be absolutely conserved in retroviral endonucleases (47), involves a pair of histidines separated by 20 to 30 residues from a pair of cysteines. The sequence in peptide IV conforms to this structure (Figure 4c) and is present at the appropriate position i.e. at the N-terminal region of the endonuclease domain.

Extensive stretches of homology to the endonuclease and reverse transcriptase domains of group II retrotransposons were identified in peptides IV and V, and VI, VII and VIII, respectively (Figure 2). These regions are shown in detail in Figure 5. One of the most highly conserved domains in most reverse transcriptases is a motif which has been designated the YXDD box (38). The tyrosine and first aspartic acid residue have been shown to be essential for reverse transcriptase activity in the case of the HIV retrovirus (48). A variant of this sequence is located within the putative reverse transcriptase domain of Tp1 at a position (underlined in Figure 5b) corresponding to the YXDD

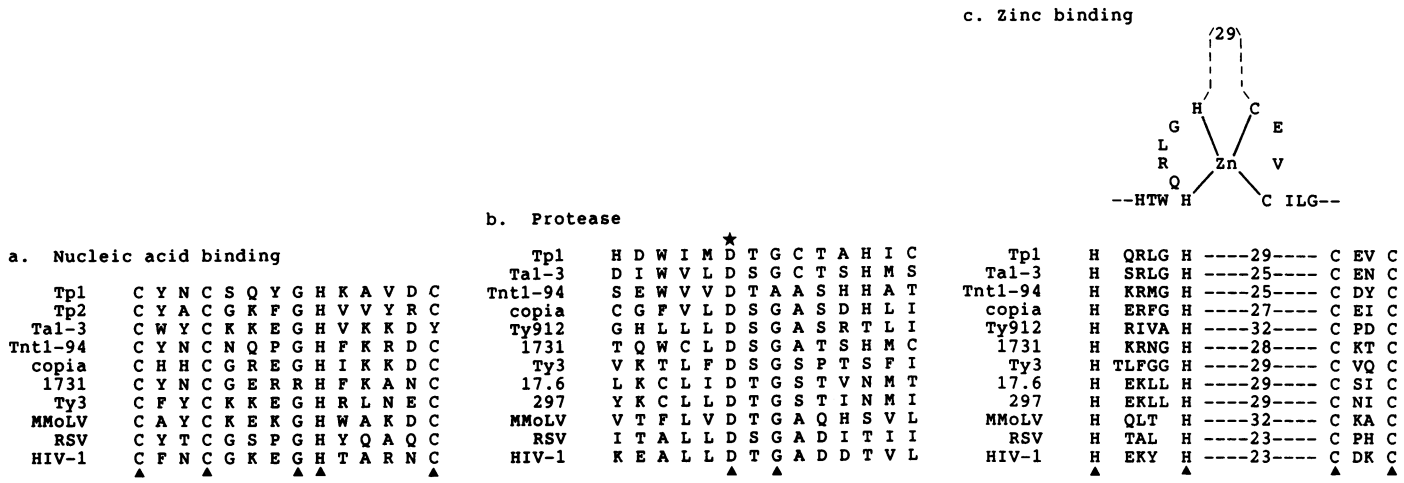

Figure 4. Predicted amino acid sequences of Tp1 aligned with conserved regions of several retrotransposons and retroviruses. ▲ Invariant (or predominantly invariant) residues. References for the sources of sequence data are given under materials and methods. a. Nucleic acid binding domain. Tp1 sequence translated from nucleotides 2071–2113. This sequence motif is not present in Ty912, 17.6 or 297. b. Protease domain. Tp1 sequence translated from nucleotides 4663–4704. ★ Aspartic acid residue corresponding to the protease active site. c. Putative zinc binding domain. Tp1 sequence translated from nucleotides 5076–5210. The diagram depicts the residues potentially involved in zinc metal binding. The number of amino acids separating the histidine and cysteine pairs is as indicated.

box in Ta1-3, Tnt1-94, *copia* and Ty912. In Tp1, the sequence is His-Val-Asp-Glu. The codons for the histidine and glutamic acid residues could have been derived from those of tyrosine and aspartic acid, respectively, by single nucleotide changes. The copy of Tp1 from which this sequence was predicted could be expected to contain many mutations and it is quite possible that functional copies of Tp1 have a sequence at this position which corresponds more closely to the YXDD box.

Evolutionary relationship of Tp1 to retrotransposons

The similarity of retrotransposons and retroviruses has led to speculation about their shared evolutionary origins (reviewed in 84). Since the reverse transcriptase is the most highly conserved domain of retro-elements it has been used to determine the phylogenetic relationship among retroviruses (49,50,84) as well as between these viruses and retrotransposons (51,38,4,84). It has been demonstrated that the diversity of reverse transcriptase-related sequences in retrotransposons is much greater than in retroviruses and appears to consist of two major groups, one represented by 17.6, 297, gypsy and 412 (group I), and the other by *copia* and Ty (group II: 38,4). Group I and II retrotransposons also differ in the relative order of the functional domains encoded by *pol*. Group I elements more closely resemble retroviruses, where the arrangement is protease-reverse transcriptase-integrase. In group II elements, the integrase lies between the protease and reverse transcriptase regions. In view of the relative order of functional domains and the high degree of sequence homology existing between *copia* and Ty, and Tp1, Ta1 and Tnt1, it would be reasonable to propose that the latter three elements are also members of the *copia* and Ty group.

It is interesting that Tp1, Ta1, Tnt1, *copia* and Ty are all very closely related to each other and distinct from the group I retrotransposons, yet they occur in very different species, i.e. acellular slime moulds, plants, insects and yeast. This might indicate that the progenitor of this group of elements is very ancient and was present before the divergence of the above species. The phylogenetic tree constructed from a comparison of reverse transcriptase sequences (84) supports this idea as it shows that *copia* and Ty diverged at a very early stage in evolution from the progenitors of other known retrotransposons and retroviruses. The elements Ta1, Tnt1 and Tp1 might now be added to this branch as other descendants of this ancient event. Alternatively, the occurence of these elements in widely different species may be due to an extensive horizontal transfer of these genetic elements between species. The latter hypothesis is supported by examination of the species distribution of *Drosophila* retrotransposons and comparison of sequence homologies and codon usage in elements found in different phyla (52,8,18).

Possible mechanisms for dispersal of Tp1 sequences

Scrambled clusters of repetitive DNA sequences have long been described as being a general feature of eukaryotic genomes (53,54) and it has been suggested that these structures might have arisen by DNA transposition (55). As with the Tp1 element in *Physarum*, similar structures in other genomes can account for a significant fraction of the repetitive DNA. Although it has not been proved directly that repeated transposition events have led to the formation of the large, scrambled clusters of Tp1 sequences, the observation that Tp1 LTRs are present at most points of discontinuity in the clones examined to date argues in favour of a specific, directed mechanism of integration. The structural homologies described above provide convincing evidence of the relationship of the Tp1 element to retrotransposons, making dispersal via transposition an attractive hypothesis. Equally valid is the possibility that homologous recombination promotes exchange of sequences within, and between, the scrambled arrangements of Tp1 elements, thus contributing to the expansion of the M+ component of the *Physarum* genome. The length (20–50kb) and the repetitive nature of the sequences within M+ tracts would facilitate such a mechanism. Of course, transposition and homologous recombination are not mutually exclusive. Both are known to be involved in the movement of Ty elements in the yeast genome (56), where homologous recombination events are 10^3 times more frequent than transpositional movements (2).

Transposition of Tp1 via a reverse transcriptase-mediated

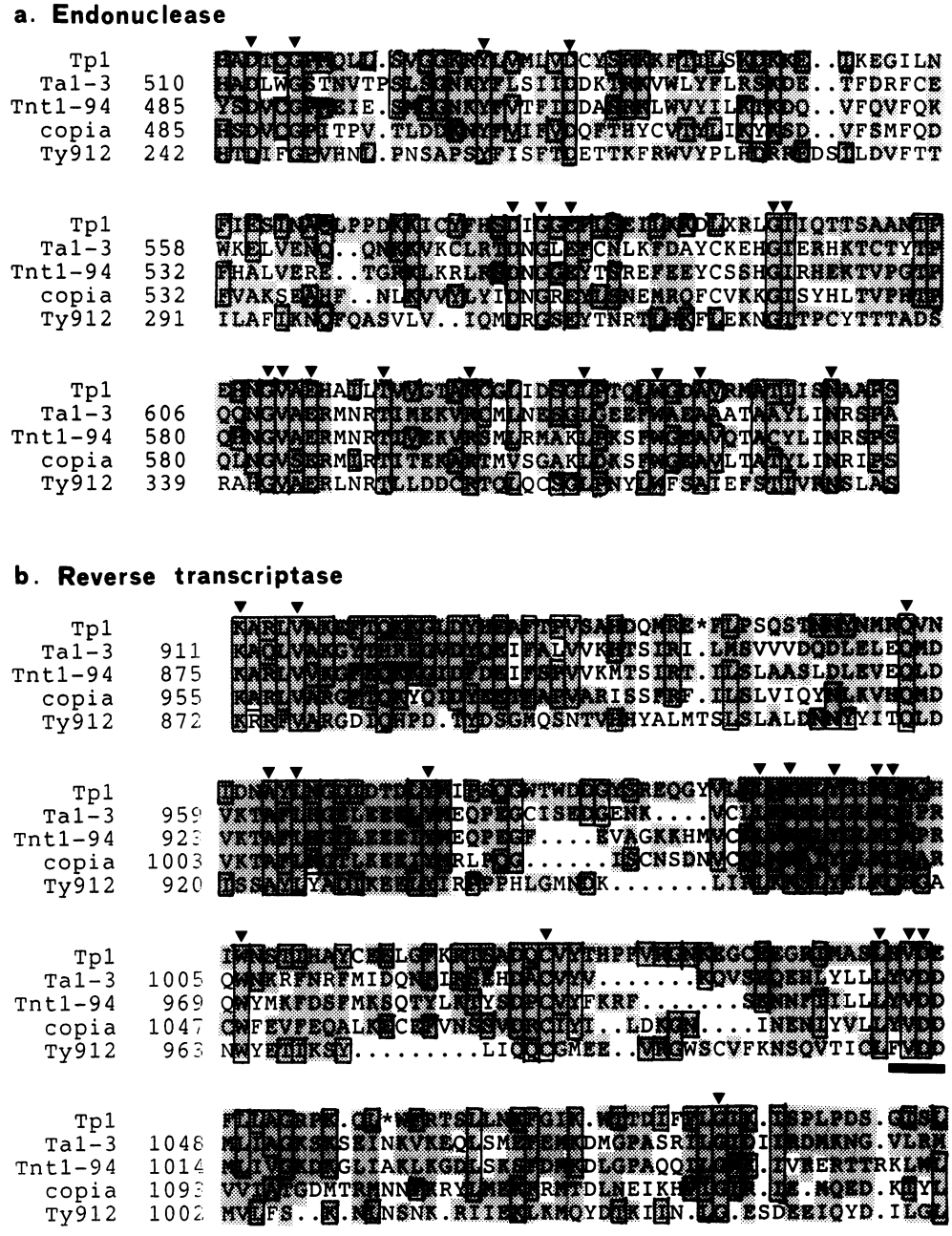

Figure 5. Alignment of predicted amino acid sequences from Tpl with (a) the endonuclease and (b) the reverse transcriptase domains of Ta1-3, Tnt1-94, copia and Ty912. Numbers refer to amino acid positions (unassigned in Tpl). Boxed residues indicate amino acids which are identical with respect to the Tpl sequence. Residues which are similar (in size and/or chemical properties) to the corresponding amino acid in Tpl are shaded. The YXDD box is underlined. ▼ Absolutely conserved amino acids.

mechanism would require the production of a full length RNA intermediate. To date, no full length Tpl-specific transcripts have been detected, but truncated transcripts detected in preliminary Northern blot experiments suggest that Tpl elements are indeed transcribed and that expression is developmentally regulated (Rothnie, Pallotta and Lemieux, unpublished data). Full length mRNA might only be produced during certain developmental stages during the *Physarum* life cycle, as has been observed for some classes of *Drosophila* retrotransposons (57).

Preliminary studies have revealed the presence of particulate structures (morphologically indistinguishable from Ty virus-like particles—VLPs) in lysates of *Physarum* myxamoebae (31). These VLP-like structures have not yet been characterised further, but if they are indeed associated with Tpl, the analogy with other retrotransposons would be further strengthened and they would be invaluable as a potential source of full length, active forms of Tpl.

Preferential integration of Tpl into self: possible implications

Self-insertion is not unique to Tpl. The heat-shock responsive element DIRS in *Dictyostelium* (58) is another example of a eukaryotic transposable element which preferentially integrates into its own sequence. Several retro-elements of different types have been found associated with each other in the human genome

(35). Clusters of Ty elements in which Ty1, Ty3 and Ty4 elements interrupt each other have also been described (59,60). The latter structures are commonly associated with tRNA genes and are probably distinct from the recently reported multimeric arrays of Ty, which appear to have arisen by recombination of intermediates during transposition (61).

To date, the Tp1 element has not been identified outwith the scrambled clusters of its own sequence and has therefore not been found in association with other gene loci. This is in contrast with many other retrotransposons, which are often found to be the cause of spontaneous mutations in structural genes (62,63). In some cases, such mutations may confer a selective advantage on the host organism and may thus contribute to rapid evolutionary adaptation. What might the advantage be of preferential integration of a retrotransposon into copies of its own, or a related, sequence? Since only a very small proportion of insertions into structural genes would be expected to be advantageous to the host, a mechanism to limit deleterious mutations may be necessary. It could be postulated that scrambled clusters of repetitive elements could act as reservoirs of sites where further integration of transposons would be tolerated. An extension of this idea is that each scrambled sequence cluster should contain an intact, potentially functional, copy of the most recently inserted element. In this way, the transposon, acting as purely selfish DNA (64), ensures its own propagation, and the host organism maintains a limited supply of active transposons which could potentially contribute to favourable mutations. Also, limiting the number of active elements in this way would limit the number of functional transcripts which in turn could control the frequency of transposition. An alternative, or additional, mechanism by which transposition could be limited is discussed below.

Tp1 elements and DNA methylation

Digestion of *Physarum* genomic DNA with HpaII generates M+ (HpaII-resistant) and M− (HpaII-sensitive) fractions which account for 20% and 80% of the nuclear DNA, respectively (65). M+ and M− DNA have similar G + C contents but the density of 5-methylcytosine (5-meC) residues in M+ DNA is approximately fivefold greater than in the M− fraction, demonstrating that M+ DNA is hypermethylated (26). The M+ fraction is comprised almost exclusively of sequences belonging to the Tp1 family (27) and to date, no Tp1 sequences have been found outwith M+ DNA. These observations raise the interesting question of whether selective methylation of sequences in M+ DNA serves to modulate expression and mobilisation of Tp1.

Several recent reviews have summarised the current state of knowledge regarding DNA methylation and its effects on gene expression and chromatin structure (66−69). In general, there is a negative correlation between methylation and gene activity, both at the level of specific 5-meC residues associated with the regulation of tissue- and development-specific genes, and as part of a more 'global' mechanism of gene inactivation i.e. inactive chromatin is heavily methylated. Clusters of Tp1 elements in M+ DNA might therefore be expected to be inactive. Further copies of Tp1 becoming integrated into a resident cluster by reverse transcriptase-mediated transposition would be unmethylated, making it necessary to postulate that *de novo* methylation of these insertions maintains the structure of M+ DNA. *De novo* methylation is thought to occur upon integration of adenovirus DNA into the genome of adenovirus-transformed cells (70). Maintenance of methylation may be part of a mechanism by which Tp1 sequences are rendered transcriptionally silent, thus limiting transposition frequency. Such a mechanism has been described in the mutator system in maize (71). If all copies of Tp1 in M+ DNA were inactivated by methylation, transcriptionally competent copies might be expected to occur elsewhere in the genome. Tp1 elements occuring at very low frequency in M− DNA may have escaped detection thus far. Unmethylated copies of Tp1 could be responsible for the generation of Tp1 transcripts. Preferential insertion of the products of these transcripts into M+ DNA would result in their subsequent inactivation by methylation. Again, like self-insertion, this is a possible mechanism for limiting the number of transposition-competent copies of Tp1 in the genome. Once inactivated, the lack of selective pressure to maintain functionally important coding sequences would result in the accumulation of mutations in these copies of Tp1, which in turn would ensure that inactivation is effectively irreversible.

CONCLUSIONS

Although a direct demonstration of mobility would be the ultimate proof that the Tp1 element can indeed function as a transposable genetic element, it is not always necessary to demonstrate transposition in order to place a new element in a class along with *copia*, Ty and other retro-elements. The presence of long terminal direct repeats with their associated controlling sequences, and the capacity of the internal sequence to encode proteins with homology to reverse transcriptase and other proteins involved in transposition suggest very strongly that the sequence element is, or was, a retrotransposon. The evidence presented here clearly shows the relationship of the Tp1 elements to other eukaryotic retrotransposons, although the Tp1 clusters described here are novel both in their genetic organisation and in the fact that they are heavily methylated. A combination of the mechanisms discussed above for limiting transposition of Tp1 provides a plausable explanation for the origin and propagation of the scrambled clusters of Tp1 which dominate M+ DNA in *Physarum*. As yet, there is no evidence that the Tp1 element actively transposes, but whatever the mechanism, amplification and dispersal of these sequences has clearly contributed greatly to the structural organisation of the *Physarum* genome.

ACKNOWLEDGEMENTS

We would like to thank Domenico Ammaturo, Tim Bullock, Felix Businger and Mairi Thomson for helping to generate some of the sequence data. Thanks also to Keith Gull and Margaret Bryans for helpful discussions and suggestions. HR and KMcC were in receipt of research studentships from the Science and Engineering Research Council. The authors are grateful to Professor H.M. Keir for provision of research facilities at Aberdeen University.

REFERENCES

1. Berg,D.E. and Howe,M.M. (eds) (1989) *Mobile DNA*, American Society for Microbiology, Washington DC.
2. Boeke,J.D., Garfinkel,D.J., Styles,C.A. and Fink,G.R. (1985) *Cell* **40**, 491−500.
3. Varmus,H. (1982) *Science* **216**, 812−820.
4. Xiong,Y. and Eickbush,T.H. (1988) *Mol. Biol. Evol.* **5**, 675−690.
5. Mount,S.M. and Rubin,G.M. (1985) *Mol. and Cell Biol.* **5**, 1630−1638.
6. Saigo,K., Kugimiya,W., Matsuo,Y., Inouye,S., Yoshioka,K. and Yuki,S. (1984) *Nature* **312**, 659−661.

7. Clare,J. and Farabaugh,P. (1985) *Proc. Natl. Acad. Sci. USA* 82, 2829–2833.
8. Hansen,L.J., Chalker,D.L., and Sandmeyer,S.B. (1988) *Mol. and Cell Biol.* **8**, 5245–5256.
9. Rubin,G.M. (1983) in Shapiro,J.A. (ed.), *Mobile Genetic Elements*, Academic Press, New York, pp 329–361.
10. Finnegan,D.J. (1985) *Int. Rev. Cyt.* **93**, 281–326.
11. Mellor,J., Kingsman,A.J. and Kingsman,S.M. (1986) *Yeast* **2**, 145–152.
12. Boeke,J.D. and Garfinkel,D.J. (1988) in Koltin,Y. and Leibowitz,M.J. (eds.), *Viruses of Fungi and Simple Eukaryotes*, Marcel Dekker Inc., New York, Basel, pp 15–39.
13. Aeby,P., Spicher,A., de Chastonay, Y., Müller,F. and Tobler,H. (1986) *EMBO J.* **5**, 3353–3360.
14. Paulson,K.E., Deka,N., Schmid,C.W., Misra,R., Schindler, C.W., Rush,M.G., Kadyk,L. and Leinwand,L. (1985) *Nature* **316**, 359–361.
15. Rotman,G., Itin,A. and Keshet,E. (1986) *Nucleic Acids Res.* **14**, 645–658.
16. Kuff, E.L., Feenstra, A., Lueders,K., Smith,L., Hawley,R., Hozumi,N. and Shulman,M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1992–1996.
17. Voytas,D.F. and Ausubel,F.M. (1988) *Nature* **336**, 242–244.
18. Smyth,D.R., Kalitsis,P., Joseph,J.L. and Sentry,J.W. (1989) *Proc. Natl. Acad. Sci.* **USA** 86, 5015–5019.
19. Jin,Y-K. and Bennetzen,J.L. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6235–6239.
20. Grandbastien,M-A., Spielmann,A. and Caboche,M. (1989) *Nature* **337**, 376–380.
21. Spradling,A.C. and Rubin,G.M. (1981) *Ann. Rev. Genet.* **15**, 219–264.
22. Baltimore,D. (1985) *Cell* **40**, 481–482.
23. Hardman,N., Jack,P.L., Fergie,R.C. and Gerrie,L.M. (1980) *Eur. J. Biochem.* **103**, 247–257.
24. Pearston,D.H., Gordon,M. and Hardman,N. (1985) *EMBO J.* **4**, 3357–3562.
25. Parkinson,H.M. (1987) PhD Thesis, Aberdeen University.
26. Peoples,O.P., Whittaker,P.A., Pearston,D.H. and Hardman,N. (1985) *J.Gen. Microbiology* **131**, 1157–1165.
27. Peoples,O.P. and Hardman,N. (1983) *Nucleic Acids Res.* **11**, 7777–7788.
28. Whittaker,P.A. (1982) PhD Thesis, Aberdeen University.
29. McLachlan,A. and Hardman,N. (1982) *Biochim. Biophys. Acta* **697**, 89–100.
30. Peoples,O.P., Robinson,A.C., Whittaker,P.A. and Hardman,N. (1983) *Biochim. Biophys. Acta* **741**, 204–213.
31. McCurrach,K.J. (1989) PhD Thesis, Aberdeen University.
32. McCurrach,K.J., Rothnie,H.M., Hardman,N. and Glover,L.A. (1990) *Current Genetics* **17**, 403–408.
33. Kikuchi,Y., Ando,Y. and Shiba,T. (1986) *Nature* **323**, 824–826.
34. Collier,D.A., Griffin,J.A. and Wells,R.D. (1988) *J. Biol. Chem.* **263**, 7397–7405.
35. Liu,Q-R. and Chan,P.K. (1990) *J. Mol. Biol.* **212**, 453–459.
36. Inouye,S., Yuki,S. and Saigo,K. (1984) *Nature* **310**, 332–333.
37. Bird,A.P. (1980) *Nucleic Acids Res.* **8**, 1499–1504.
38. Yuki,S., Ishimaru,S., Inouye,S. and Saigo,K. (1986) *Nucleic Acids Res.* **14**, 3017–3030.
39. Inouye,S., Yuki,S. and Saigo,K. (1986) *Eur. J. Biochem.* **154**, 417–425.
40. Emori,Y., Shiba,T., Kanaya,S., Inouye,S., Yuki,S. and Saigo, K. (1985) *Nature* **315**, 773–776.
41. Covey,S.N. (1986) *Nucleic Acids Res.* **14**, 623–633.
42. Prats,A.C., Sarih,L., Gabus,C., Litvak,S., Keith,G. and Darlix,J.L. (1988) *EMBO J.* **7**, 1777–1783.
43. Gorelick,R.J., Henderson,L.E., Hanser,J.P. and Rein,A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8420–8424.
44. Pearl,L.H. and Taylor,W.R. (1987) *Nature* **329**, 351–354.
45. Yoshioka,K., Honma,H., Zushi,M., Kondo,S., Togashi,S. Miyake,T. and Shiba,T. (1990) *EMBO J.* **9**, 535–541.
46. Torruella,M., Gordon,K. and Hohn.T. (1989) *EMBO J.* **8**, 2819–2825.
47. Johnson,M.S., McClure,M.A., Feng,D-F., Gray,J. and Doolittle,R.F. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 7648–7652.
48. Larder,B.A., Purifoy,D.J.M., Powell,K.L. and Darby,G. (1987) *Nature* **327**, 716–717.
49. Chiu,I-M., Yaniv,A., Dahlberg,J.E., Gazit,A., Skuntz, S.F., Tronick,S.R. and Aaronson,S.A. (1985) *Nature* **317**, 366–368.
50. McClure,M.A., Johnson,M.S., Feng,D-F. and Doolittle,R.F. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2469–2473.
51. Toh,H., Kikuno,R., Hayashida,H., Miyata,T., Kugimiya,W., Inouye,S., Yuki,S. and Saigo,K. (1985) *EMBO J.* **4**, 1267–1272.
52. Martin,G., Wiernasz,D and Schedl,P. (1983) *J. Mol. Evol.* **19**, 203–213.
53. Musti,A.M., Sobieski,D.A., Chen,B.B. and Eden,F.C. (1981) *Biochemistry* **20**, 2989–2999.
54. Eden,F.C., Musti,A.M. and Sobieski,D.A. (1981) *J.Mol. Biol.* **148**, 129–151.
55. Wensink,P.C., Tabata,S. and Pachl,C. (1979) *Cell* **18**, 1231–1246.
56. Roeder,G.S. and Fink,G.R. (1980) *Cell* **21**, 239–249.
57. Parkhurst,S.M. and Corces,V.G. (1987) *EMBO J.* **6**, 419–424.
58. Cappello,J., Cohen,S.M. and Lodish,H.F. (1984) *Mol. and Cell Biol.* **4**, 2207–2213.
59. Clark,D.J., Bilanchone,V.W., Haywood,L.J., Dildine,S.L. and Sandmeyer,S.B. (1988) *J. Biol. Chem.* **263**, 1413–1423.
60. Stucka,R., Lochmüller,H. and Feldmann,H. (1989) *Nucleic Acids Res.* **17**, 4993–5001.
61. Weinstock,K.G., Mastrangelo,M.F., Burkett,T.J., Garfinkel, D.J. and Strathern,J.N. (1990) *Mol. and Cell Biol.* **10**, 2882–2892.
62. Roeder,G.S. and Fink,G.R. (1983) Shapiro,J.A. (ed.), *Mobile Genetic Elements*, Academic Press, New York, pp. 299–328.
63. Green,M.M. (1988) in Lambert,M.E., McDonald,J.F. and Weinstein,I.B. (eds.), *Eukaryotic Transposable Elements as Mutagenic Agents* 30th Banbury Report, Cold Spring Harbour University Press, pp 41–50.
64. Orgel,L.E. and Crick,F.H.C. (1980) *Nature* **284**, 604–607.
65. Whittaker,P.A. and Hardman,N. (1980) *Biochem. J.* **191**, 859–862.
66. Adams,R.L.P. (1990) *Biochem. J.* **265**, 309–320.
67. Cedar,H. and Razin,A. (1990) *Biochim. Biophys. Acta* **1049**, 1–8.
68. Magill,J.M. and Magill,C.W. (1989) *Developmental Genetics* **10**, 63–69.
69. Selker,E.U. (1990) *TIBS* **15**, 103–107.
70. Sutter,D. and Doerfler,W. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 253–256.
71. Chandler,V.L. and Walbot,V. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 1767–1771.
72. Hardman,N., Jack,P.L., Brown,A.J.P. and McLachlan,A. (1979) *Eur. J. Biochem.* **94**, 179–187.
73. Cooke,D.J. and Dee,J. (1975) *Genet. Res. Camb.* **24**, 175–187.
74. Gough,J.A. and Murray,N.E. (1983) *J. Mol. Biol.* **166**, 1–19.
75. Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor University Press, Cols Spring Harbor.
76. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
77. Henikoff,S. (1984) *Gene* **28**, 351–359.
78. Lipman,D.J. and Pearson,W.R. (1985) *Science* **227**, 1435–1441.
79. Devereux,J. and Haeberli,P. (1986) Introduction to the Sequence Analysis Software Package of the University of Wisconsin Genetics Computer Group, version 6. University of Wisconsin, Madison, WI USA.
80. Fourcade-Peronnet,F., d'Auriol,L., Becker,J., Galibert,F. and Best-Belpomme,M. (1988) *Nucleic Acids Res.* **16**, 6113–6125.
81. Shinnick,T.M., Lerner,T.A. and Sutcliffe,J.G. (1981) *Nature* **293**, 543–548.
82. Schwartz,D.E., Tizard,R. and Gilbert,W. (1983) *Cell* **32**, 853–869.
83. Ratner,L., Haseltine,W., Patarca,R., Livak,K.J., Starcich,B., Josephs,S.F., Doran,E.R., Rafalski,J.A., Whitehorn,E.A., Baumeister,K., Ivanoff,L., Petteway, S.R.Jr., Pearson,M.L., Lautenberger,J.A., Papas,T.S., Ghrayeb,J., Chang,N.T., Gallo,R.C. and Wong-Staal,F. (1985) *Nature* **313**, 277–284.
84. Doolittle,R.F., Feng,D-F., Johnson,M.S. and McClure,M.A. (1989) *The Quarterly Review of Biology* **64**, 1–30.