Nucleic Acids Research, Vol. 20, No. 7 1623–1628

Unusual features of the retroid element PAT from the nematode *Panagrellus redivivus*

Yves de Chastonay+, Heinz Felder, Christopher Link[1], Pierre Aeby[2], Heinz Tobler and Fritz Müller*

Institute of Zoology, University of Fribourg, Pérolles, CH-1700 Fribourg, Switzerland, [1]Department of Biology, University of Denver, Colorado, USA and [2]Baxter Diagnostic, Bonnstraße 9, CH-3186 Düdingen, Switzerland

Received December 30, 1991; Revised and Accepted February 25, 1992 EMBL accession no. X60774

ABSTRACT

The PAT retroid transposable elements differ from other retroids in that they have a 'split direct repeat' structure, i.e., an internal 300bp sequence is found repeated, about one half at each element extremity. A very abundant transcript of about 900nt, the start of which maps to the preferentially deleted portion of PAT elements, is detected on total *Panagrellus redivivus* RNA bearing Northern blots. A potentially corresponding ORF encodes a protein of 265 residues having a carboxy terminal Cystein motif, believed to be exclusively characteristic of the GAG protein in retroid elements. A much fainter, 1800nt long transcript, is also detected on Northern blots and maps slightly downstream of the first ORF. The predicted protein sequence of this region bears motifs typical of reverse transcriptase and RNaseH, as found in the Pol genes of retroid elements. Peptide motif similarities are greatest with the DIRS-1 element derived from *Dictyostelium discoideum*. The possibility of using PAT elements as transposon tagging system for *Caenorhabditis elegans* is discussed.

INTRODUCTION

Mobile genetic elements have been found in a wide variety of organisms, ranging from prokaryotes to higher eukaryotes and humans (for review, see 1). Two main classes of transposable elements can be distiguished, for each of which a different mode of transposition, as well as several structural features, are characteristic. The first class, retroid elements, are mobilized via full length RNA molecules and require reverse transcription prior to possible integration in the genome (2, 3). These can be subdivided into long terminal repeat (LTR) containing and non-LTR containing elements , both of which, however, contain group specific antigen (GAG) and RNA dependent DNA polymerase (RT) (4, 5). First strand cDNA is usually primed by a tRNA molecule and synthesis proceeds outwards through the 5′ LTR (for review, see 6). The LTRs then play a key role by allowing ordered inter- or intrastrand exchanges during reverse transcription (7, 8). In the case of non-LTR containing elements, several priming possibilities that avoid the need for strand exchanges have been reported, including priming by host DNA at the insertion site, a mechanism involving terminal uridine transferase enzyme (for review, see 5) and protein priming (9). For the non-retroid counterparts, all transposition intermediates appear to be composed of DNA. A distinctive structural feature of these elements is the presence of short inverted repeats (IRs) at the element termini and their encoding a so-called transposase function (4). One element, DIRS-1 of Dictyostelium discoideum, shares properties of both classes (i.e., IRs and RT), having led to the proposition of an intramolecular priming mechanism for retroid replication (10).

Representatives of either class have been found in the phylum Nematoda, among which the non-retroid Tc family of *Caenorhabditis elegans* (11) and the retrovirus-like element TAS of *Ascaris lumbricoides* (12). A further nematode transposon, the PAT element of *Panagrellus re*divivus, was discovered upon transposition in a *C. elegans unc-22* gene homologue, the locus of mutation to the twitcher phenotype (13). Like most transposing entities, the PAT element belongs to the repetitive DNA fraction. Distribution is random and elements are represented between 10 and 50 times per haploid genome, depending on strain. In this paper we present the overall structure, relevant open reading frames (ORFs) and transcription of PAT, in an attempt to classify the element among other transposing entities and to evaluate perspectives.

MATERIALS AND METHODS

Nematode cultures, DNA isolation, restriction digestion, TBE electrophoresis and Southern blot hybridization

P. redivivus strain Sc was cultivated as described for *C. elegans* by Sulston and Brenner (14). *E. coli* OP50SR were given as food source and microbial contamination was avoided by the addition

* To whom correspondence should be addressed

+ Present address: ESSEX Chemie AG, Tribschenstr. 11, CH-6005 Luzern, Switzerland

of 10 μg nystatin and 200 μg streptomycin per ml culture medium. Chromosomal DNA was isolated as outlined by Sulston and Hodgkin (15) and restriction digestions were carried out according to the manufacturers specifications. Agarose gels for electrophoresis were cast in TBE buffer and transferred onto nitrocellulose membranes with 20×SSC (16). Hybridization with nick-translated probes was performed in 5×Denhardt (16)/ 0.1% SDS/ 4×SSC/ 10% Na·pyrophosphate/ 10 mM EDTA at 42°C for 16 h. Filters were washed 2×45 min in 0.1×SSC/ 0.1% SDS at 62°C, dried and exposed to X-ray films at −70°C with a tungstate screen. Dehybridization of filters was carried out at 70°C in two changes of 0.1% SDS/ 50% formamide, followed by one change of 0.1%SDS/ 0.1×SSC, for 30 min each. Filters were checked for absence of residual signal by film exposition prior to re-hybridization.

DNA sequencing and computer analysis

Nucleotide sequences were established following the chain termination method (17), using Sequenase™ enzyme, according to the manufacturors specifications. Single stranded DNA was obtained as described in the Amersham sequencing handbook for M13 clones, and according to the procedure recommended by Stratagene with VCS-M13 helper phage for pBS-M13 clones. Sequence analysis was performed with the PC-Gene 6.26 program and homologies were searched through the SwissProt13 and EMBL release 21 data banks. Protein alignments were obtained with the Clustal program of PC-Gene, having all computation parameters set to maximum values, except that the window size was set to 10 and transitions were weighted twice as likely as transversions.

RNA isolation, electrophoresis and Northern blot hybridization

Total RNA was extracted from *P. redivivus* cultures as described for *C. elegans* (D. Pilgrim, personal communications), glyoxalated, electrophoresed and transferred onto Biotrans membrane (ICN-Schwarz-Mann) following the manufacturers recommendations. RNA markers, i.e., total RNA and RNA ladder (BRL), were visualized by staining the corresponding lanes with ethidium bromide. Hybridization was carried out in 50% formamide/ 2×SSC/ 10×Denhardt/ 1% SDS/ 50mM sodium phosphate buffer pH 6.5/ 100 mg each, yeast tRNA and denatured salmon sperm DNA, by adjunction of the labelled probe after 3 hrs prehybridization at 50°C. Hybridization was persued 24−65 hrs and followed by 4 changes of 2×SSC/ 0.1% SDS, 2×5 min at room temperature and 2×45 min at 60°C. The filter was then washed for 30 min in 0.1×SSC/ 0.1% SDS at room temperature prior to exposition. Dehybridization technique was that used for Southern blots.

RESULTS

PAT elements have unusual split DR structures

For the analysis of PAT element structure, we have isolated several clones from a *P. redivivus* strain C15 genomic library. Each of the screened entities contains single *Sst*I and *Sph*I restriction sites near their left and right ends, respectively (cf. restriction map Fig. 1B). In order to define the extremities of PAT elements, we have sequenced the borders of two representatives, i.e., the sequences upstream of *Sst*I and downstream of *Sph*I. Homology is nearly complete up to a specific position from where on the sequences diverge. Fig. 1A

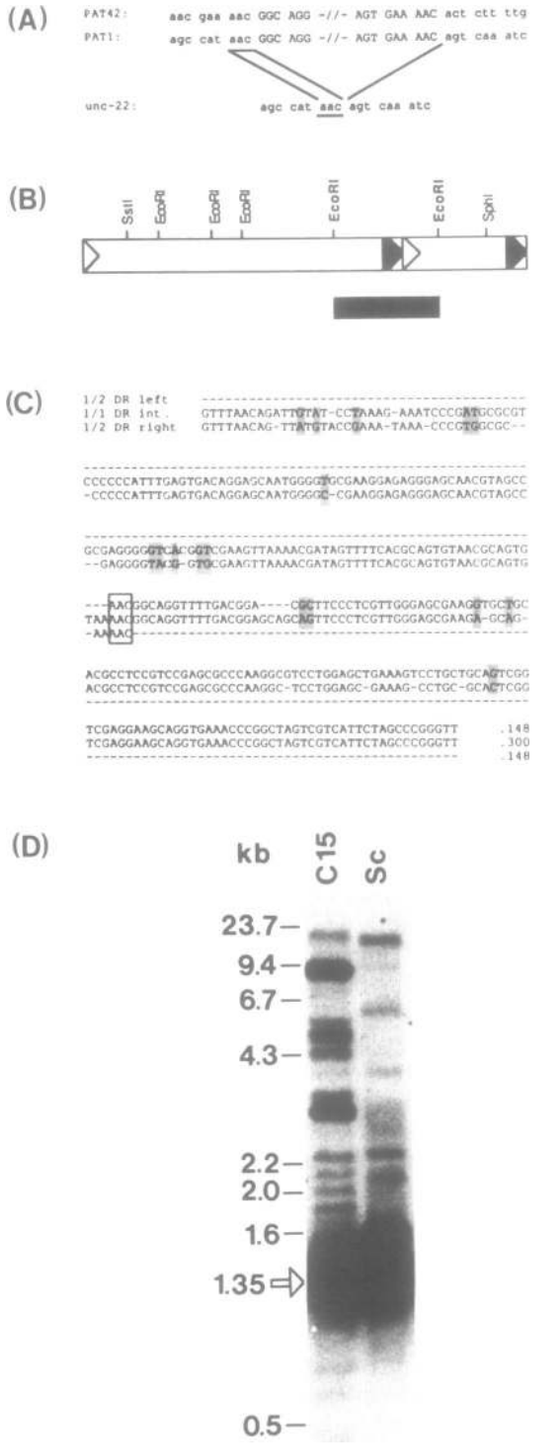

Figure 1. (A) PAT42 and PAT1 border sequences aligned. Nucleotides belonging to the entities are in capitals, insertion site bases are in lower case letters. Further comparison shows the PAT1 insertion site with the corresponding empty site of the *unc-22* homologue of *P. redivivus*. **(B)** DR arrangement in PAT42. The different DR halves are represented as empty and full triangles. *Eco*RI pattern is shown and restriction sites referred to in the text are indicated above the element diagram in bold letters. **(C)** Alignment of the DR halves of PAT42. Differences are indicated by shading and missing bases by a hyphen. **(D)** *P. redivivus* strain C15 and Sc genomic blot of *Eco*RI digested DNA, hybridized with the internal 1.35kb *Eco*RI fragment shown in Figure 1B. Length markers are indicated beside the genomic lanes and the 1.35kb band, implying conservation of DR arrangement, is shown by an arrow.

shows the alignment of the PAT42 border sequences with those of PAT1. Furthermore, the *unc-22* sequence of *P. redivivus* is aligned with the PAT1 insertion site in the mentioned gene, confirming the positions previously defined as element termini.

A feature found for many transposable elements is the presence of direct or inverted repeats at both ends. Single stranded M13 clone hybridizations had shown that PAT elements contain direct repeats (13). Sequencing data of the PAT42 element, however, revealed that these are not LTRs as expected from retroid elements. Instead, an internal 300bp region is found, split, about one half at each end of the element (Fig. 1B). We refer to the PAT element direct repeat (DR) organization as 'split DR' structure. Fig. 1C presents the nucleotide sequence of the internal 300bp DR of PAT42, aligned with the two 148bp border half DR sequences. To which DR half the central AAC trinucleotide belongs, cannot be deduced since AAC is also found at both ends of PAT1 and PAT42 (see also Fig. 1A). The presence of AAC inside the internal DR, however, suggests that one border AAC repeat belongs to the transposable element rather than that the border copies of AAC represent one host trinucleotide and a duplication of the sequence resulting from insertional gap repair.

In order to see whether the split DR structure is representative for PAT elements, we hybridized a Southern blot of *Eco*RI digested genomic *P. redivivus* DNA with a labelled PAT42 *Eco*RI 1.35kb DNA fragment, containing the internal DR (probe represented by a black bar in Fig. 1B). This probe lights up as a very strong 1.35 kb band on the blot (Fig. 1D), indicating that PAT elements do indeed have the split DR structure in common. As expected, many weaker bands are also detected, most if not all of which are due to crosshybridization with border DR sequences. The conclusion that the lengths of these bands are variable was foregone, since they are generated by one *Eco*RI site inside the element and another one in its environment.

PAT elements vary in size

Among the cloned representatives, we have characterized PAT elements of three different lengths. PAT42 is 5.5kb long, while PAT1 and PAT31 have an 0.4kb deletion and PAT32 lacks approximately a 1kb sequence with respect to the 5.5kb form. The deletions were mapped by sequence analysis for PAT1 and by restriction digestions and Southern blot hybridizations for PAT31 and PAT32 (data not shown). Fig. 2A shows the positions of the mentioned deletions, relative to PAT42. As presented in Fig. 2B, we undertook to determine the proportional distribution of 5.5kb and deleted entities in the genome. The 750 bp

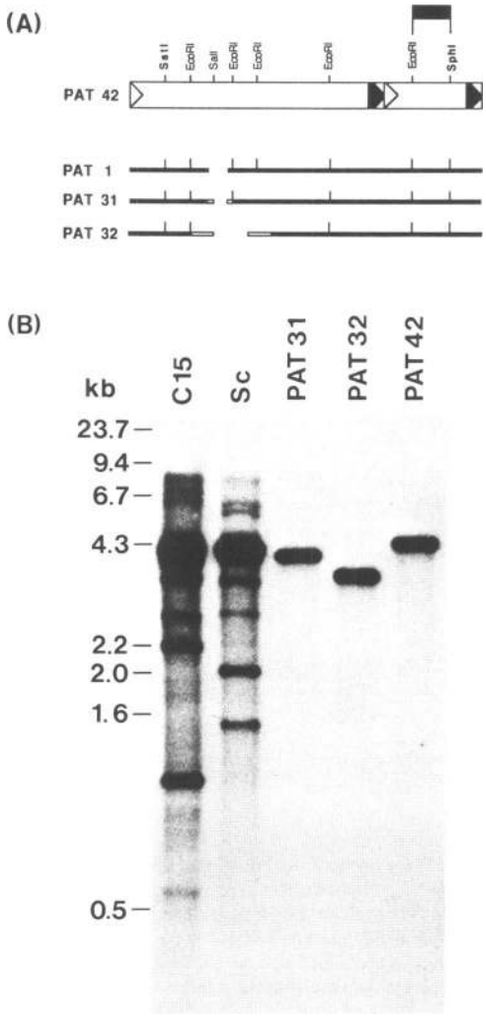

Figure 2. (A) Map of PAT42 and location of internal deletions in different elements shown beneath. The sites referred to for genomic DNA digestions are bold. Conserved restriction sites in deleted elements are indicated individually by vertical bars. The *Eco*RI-*Sph*I fragment denotes the probe used for the Southern blot shown in Figure 2B. Light portions in deleted elements mark the domains in which breakpoints are located. **(B)** Southern blot of *Sph*I / *Sst*I digested DNAs. *P. redivivus* strain, respectively, PAT element clones are indicated above the corresponding lane and length markers are shown to the side.

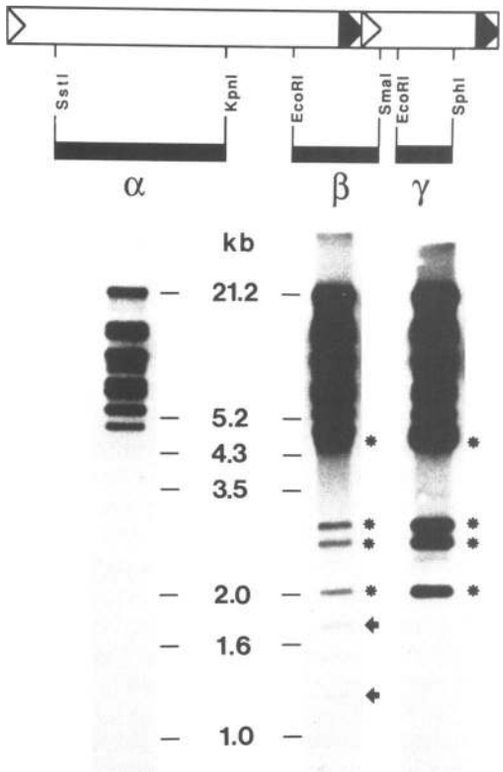

Figure 3. PAT42 map with restriction sites for the probes (dark bars marked a, b and g) used to hybridize genomic *P. redivivus* strain C15 DNA digested with *Cla*I. Length markers are shown along the side. Asterisks and arrows indicate bands referred to in the text.

*Eco*RI/*Sph*I fragment (black bar in Fig. 2A) was used as hybridization probe on a genomic blot of *Sph*I/*Sst*I digested DNA. The bands lighting up are generated by the two restriction sites located inside the different elements. A strong band of about 4.6 kb lights up in both genomic tracks, corresponding in length to the cloned copy PAT42. Several shorter bands are visible in both genomic tracks. In strain C15, at least two of these represent internally deleted elements as deduced from length homologies to the controls PAT31 and PAT32. At this point, we could not discern whether the shorter bands in the genomic tracks are due to restriction site heterogeneities in different elements or to deletion events of further extent. The elucidation of PAT element structure was therefore also attacked by a further series of hybridizations.

Deletions are preferentially confined to one half of the element

Our approach consisted in performing Southern blots of genomic *P. redivivus* DNA, restricted with *Cla*I, an enzyme not cleaving the PAT elements analysed so far. Hybridization was performed with different probes, together spanning the majority of the internal PAT42 domain. For synoptic reasons we present hybridizations to *P. redivivus* strain C15 DNA only, given the fact that conclusions drawn for strain Sc are identical. Fig. 3 shows the PAT42 element and the different probes used for hybridizations (black boxes α, β and γ). Corresponding Southerns are presented below the individual probes.

All bands lighting up with probe α clearly also appear upon hybridization with probes β and γ. This implies that the pattern differences observed are not due to the presence of *Cla*I sites in some of the PAT elements. Rather, this band distribution indicates that PAT element deletions preferentially comprise the region spanned by probe α. Among the bands not detected with α, four are common to probes β and γ (indicated by asterisks) and apparently reflect the presence of severely deleted elements in the genome, perhaps having occurred by homologous recombination of DR sequences at the proviral stage, as found for TOC1 (14). Two further bands (indicated by arrows) are only detected with probe β which spans the internal DR. This indicates the possibility of solo DR sequences being left behind in the genome upon excision of a PAT element, or the existence of elements deleted to yet a further extent.

PAT contains two transcribed ORFs

Sequencing of the PAT42 element revealed the presence of two potential coding regions (boxes inside the PAT42 element shown in Fig. 4A). The first ORF has a Met initiation codon at nucleotide position 1084 with respect to the 5′ end, which is 27bp inside of the sequence deleted in PAT1. The ATG translation start signal is bounded by A_{-3} and G_{+1}, a configuration known to favour the initiation of translation by eukaryotic ribosomes (18–20). Lecture from this position on would produce a polypeptide of 265 residues to the first stop codon. Further 72bp downstream, we found an AATAAA motif, which forms an essential part of the recognition sequence for 3′ end processing of primary eukaryotic transcripts (for review, see 21). Another ORF is present, starting at nucleotide position 2122 with, again, a strong ATG translation start signal. ORF2, as we refer to it here, stretches 588 amino acid residues downstream to a stop codon, located 5bp upstream of the internal DR in PAT42. Both, ORFs 1 and 2, are predicted as coding according to Fickett's criteria, with incidences of 92% and 98%, respectively.

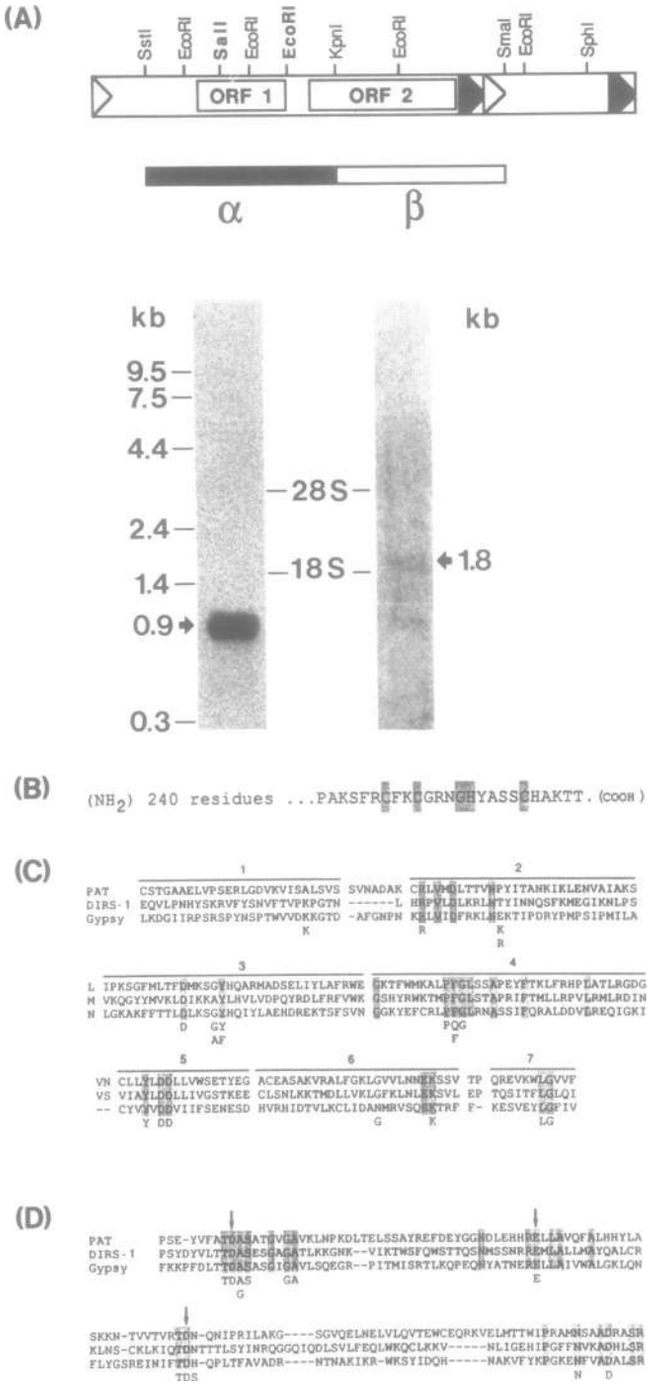

Figure 4.(A) PAT42 map with the probe used for hybridization to a total *P. redivivus* RNA bearing Northern blot. The two major ORFs mentioned in the text are shown as boxes inside the element. The restriction sites used for the generation of different hybridization probes and the *Eco*RI pattern are indicated. The corresponding Northern blot with length standards (i.e., RNA ladder and ribosomal RNA) alongside is presented beneath probe a. **(B)** 25 carboxy terminal residues of ORF1. Amino acids relevant of the GAG Cystein motif are shaded. **(C)** Alignment of PAT, DIRS-1 and Gypsy element RT domains. The conservative subdomains proposed by Xiong and Eickbush (24) are overscored and numbered 1 through 7. Amino acids conserved in all three compared sequences are shaded. The highly conserved residues in the analysis of Xiong and Eickbush (24), comprising 97 different RT sequences, are indicated below, at the corresponding position. **(D)** Alignment of RH sequences derived from PAT, DIRS-1 and Gypsy. Residues identical in all three sequences are shaded and highly conserved amino acids in the analysis of Doolittle *et al.* (23) are noted below. Arrows indicate the three crucial residues referred to in the text.

Protein sequence analysis based on prediction from nucleic acid studies revealed some striking features for ORFs 1 and 2. Fig. 4B shows the 25 putative carboxy terminal residues of ORF1, bearing a typical Cys motif, which is the correct location for a motif found exclusively in GAG proteins of retroid elements to date (22, 23). Typical for ORF2 is the presence of RT and RNaseH (RH) motifs, separated by a 103 residue tether, again, characteristic features of retroid elements. Fig. 4C shows the alignment of RT sequences derived from PAT42, DIRS-1 and Gypsy. DIRS-1 is the closest match to PAT42 among the sequences analysed by Xiong and Eickbush (24). 21 out of 178 residues are identical between the compared sequences, not taking account of conservative changes. Individually, homologies are 49/178 between PAT and DIRS-1, 41/178 between DIRS-1 and Gypsy and 38/178 between Gypsy and PAT. These figures make of PAT the most distantly related retroid element in the Gypsy family, but this assignment is stronger than to other retroid element classes (analysis not presented). Fig. 4D shows alignment of the RH domains derived from the same elements. Once again, DIRS-1 is the element to which PAT has most similarity when the RHs analysed by Doolittle *et al.* (23) and PAT are compared. Three residues, reported to be absolutely crucial for activity of MMLV and *E. coli* RH (25, 26), are also found in the three compared RHs, namely Asp^{10}, Glu^{48} and Asp^{70}. These residues are indicated by arrows in Fig. 4D.

In order to confirm transcription of the different regions in PAT elements, we performed a series of Northern blot experiments. Hybridization to total *P. redivivus* bearing Northern blot with probe a is shown in Fig. 4A. This probe lights up as a very strong band of 900nt after 2.5 hr filter exposition. The orientation of this transcript was confirmed left to right by hybridization with T3 and T7 probes obtained from a PAT42 *SphI/SstI* pBS-M13 clone. Hybridization with subclones demonstrated that the transcript starts slightly upstream of the first *SalI* site (far left) and ends just downstream of the third *Eco*RI site (data not shown). This indicates that the region spanned by ORF1 is transcribed in the foreseen orientation. Furthermore, the lack of shorter bands implies that elements deleted in this region do not transcribe ORF1 sequences. Another transcript, about 1800nt long, is detected after 15 day film exposition upon hybridization with a probe adjacent to α, namely the *KpnI/SmaI* fragment (see Fig. 4A).

Fig. 5 presents a diagram of ORFs 1 and 2 with putative functional domains boxed. No protease-like domain was identified on account of known motifs, a situation only found in several members of the LINES family and in the DIRS-1 element thus far (23). Also similar to CaMV, CERV, DIRS-1 and some elements of the LINES family, is the lack of an accurate finger motif in what is believed to be an Integrase (Int) domain. On the other hand, a tether separating RT and RH was not found in the Gypsy (including DIRS-1) and copia family elements. A more consistent approach to retroid element classification, though, resides in amino acid homology scans through well defined RT regions of different retroids (23, 24). Using the Clustal program

Figure 5. General repartition diagram of putative functional domains in ORFs 1 and 2. Numbers indicate amino acid position with respect to the methionine codon thought to act as initiator.

of PC gene, the PAT element RT scores best with DIRS-1 (Fig. 4C) and very similar results are obtained comparing RH sequences (Fig. 4D). The distant relationship of PAT and DIRS-1 to the Gypsy family, however, seems to imply the necessity of creating a novel class of retroid elements.

DISCUSSION

A rather unusual feature of the retroid PAT elements is their split DR structure (Fig. 1B and C). The only precedent of internal DR sequences was reported for TOC1 of *Chlamydomonas rein*hardtii (27). The fact that the half DRs are alternate (A...BA..B), implies that PAT elements are not composed of two different transposons having inserted into or next to one another. If DR sequences were to serve for strand exchanges during reverse transcription, the maintenance of overall sequence arrangement would probably require several jumps for first strand cDNA synthesis as proposed for TOC1 (28 < fs >), a peculiarity in terms of retroid elements (8). PAT lacks homology to the 3' end of a tRNA molecule near any one DR sequence, but the possibility of reverse transcription initiating at, and proceeding outwards through, a DR sequence cannot be excluded. A noteworthy example in this context is the exotic nucleic acid priming mode, reported for copia elements (29), where initiation starts at a sequence homologous to an internal stretch of a tRNA molecule. An alternative to outwards cDNA synthesis initiation is terminal priming, either by protein as found for HBV (9), by 3' addition of uridine residues and poly-A priming, or by reverse transcription upon ligating of the element template to the genomic insertion site (for review, see 5).

Several aspects concerning replication of PAT elements remain unanswered, among others, proof is lacking with respect to the existence of virus-like particles (VLPs) as predicted by the presence of a putative GAG protein. If, as shown for Ty (30, 31), VLPs were to be the site of reverse transcription, it should be interesting to analyse the nature and number of encapsidated PAT molecules to gain more knowledge on the replication cycle. It seems clear, however, that PAT encodes functions necessary for replication via RNA intermediates.

In most retroid elements, the stoichiometry of GAG and Pol protein expression is regulated by translation from the same messenger molecule and occasional frameshifting to produce slighter levels of GAG-Pol fusion protein than GAG protein alone. This mechanism was well studied for Ty elements (32) and HIV (33). An exception to this retroid feature is found in the case of HBV, which uses separate initiation codons for GAG and Pol, albeit, from the same messenger molecule (34). PAT elements are outsider retroids in this respect since GAG and Pol generate discrete transcripts, as deduced from the lack of precursor and the difference of signal intensity on Northern blots. Thus the stoichiometric proportion of GAG to Pol protein seems to be regulated at the transcriptional level. Several small potential coding regions were detected further 3'wards, none , however, containing a putative transmembrane domain, as one would expect of Env in infectious retroviruses. Hence, PAT presents the features for the classification as nonviral retroelement (35).

The majority of elements are 5.5kb long but several internally deleted forms are also found in both analysed nematode strains. The extent of deletion in different elements is variable and some events could be the consequence of recombination between homologous DR segments. It is not clear when, during the replication cycle or at the proviral stage, these deletions take

place. Retroviral recombination has been shown to occur during reverse transcription, probably by template switching), and for Ty, such deletion events take place during or immediately after reverse transcription, but before genome integration (37). In PAT elements, deletions are confined preferentially to one half of the entities. The lack of shorter bands on the Northern blot of Fig. 4 implies that in PAT1, an element of the 0.4kb deletion form, the start of the major 900nt transcript along with the transcription promotor are lost. We therefore refer to PAT1 as non-autonomous and presume that the missing function was provided in trans. This element did, however, retain the sequences required in cis for transposition (13), and might therefore be used as mutator element for a transposon tagging system.

The nematode *C. elegans* is favored by a growing population of molecular biologists because of its easiness to handle, short generation time, small genome size, detailed knowledge of its genome, development and cell lineage, and because many findings can be extrapolated to other species including man (38). Transposons have been identified (i.e., Tc elements), but the requirement of relatively high copy numbers for efficient transposition (39) make their use as tagging system quite tedious. Moreover, Tc elements are characterized by their target site specificity (11), which could account for a limited number of affectable genes. Not only do P and PAT elements behave similarly with respect to the non-reciprocal effect on fertility in crosses between high and low copy number strains (40, 13), but the notion that PAT elements may provide a useful tagging system for *C. elegans* is indicated by the following observations. First, *P. redivivus* and *C. elegans* are closely related nematodes, and possible transposition of PAT elements in the latter gains support when considering P element mobility in non-Drosophilids (41). Second, the elements actively transpose and cause gene disruption at their insertion site. Third, partially deleted, non-autonomous elements can be mobilized and integrated if the missing function is provided in trans. Fourth, the germline is concerned by transposition of PAT elements, as deduced from the stably inherited *unc-22* mutant (13). Fifth, PAT DNA probes do not crosshybridize with *C. elegans* DNA, thus facilitating screening pro cedures. On the other hand, mobility of PAT elements could be impaired in *C. elegans* by the presence of repressors, inappropriate transcription signals or the lack of positively regulating host factors (for review, see 5). DNA injection experiments have been established (42, 43) and expression of genes in transgenic *C. elegans* was found to be correct (44). Thus, inspirations can henceforth be gathered from transposon tagging and enhancer trapping P element systems in *Drosophila* (45). Moreover, perspectives are opened with recently found marker genes such as roller (46) and genes known to suppress recognizable phenotypes (47), as well as the recently described use of inducible promoters in such systems (48). Libraries of strains with single PAT element insertions would be likely to greatly facilitate gene recognition and characterization in *C. elegans*.

ACKNOWLEDGEMENTS

We thank Mr. H. Gachoud for art work and photography and Mrs. C. Dedual as well as Mrs. C. Stempfel for excellent technical assistance. This work was supported by grants No. 3.113-0.85 and 3.141-0.88 from the Swiss National Science Foundation.

REFERENCES

1. Berg, D. and Howe, M. (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. (1988) *Science*, **240**, 1427–1434.
4. Finnegan, D.J. (1989) *TIG*, **5**, 103–107.
5. Boeke, J.D. and Corces, V.G. (1989) *Annu. Rev. Microbiol.*, **43**, 403–434.
6. Varmus, H. (1989) in Berg, D.E. and Howe, M.M. (eds.) *Mobile DNA*, American Society for Microbiology, Washington, DC, 53–108.
7. Arkhipova, I.R., Mazo, A.M., Cherkasova, V.A., Gorelova, T.V., Schuppe, N.G. and Ilyin, Y.V. (1986) *Cell*, **44**, 555–563.
8. Panganiban, A.T. and Fiore, D. (1988) *Science*, **241**, 1064–1069.
9. Bartenschlager, R. and Schaller, H. (1988) *EMBO J.*, **7**, 4185–4192.
10. Cappello, J., Handelsman, K. and Lodish, H.F. (1985) *Cell*, **43**, 105–115.
11. Emmons, S.W. (1988) in Wood, W.B. (ed.) *The Nematode Caenorhabditis elegans*, Cold Spring Harbor Laboratory Press, NY, 47–79.
12. Aeby, P., Spicher, A., de Chastonay, Y., Müller, F. and Tobler, H. (1986) *EMBO J.*, **5**, 3353–3360.
13. Link, C.D., Graf-Whitsel, J. and Wood, W.B. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 5325–5329.
14. Sulston, J.E. and Brenner, S. (1974) *Genetics*, **77**, 95–104.
15. Sulston, J.E. and Hodgkin, J. (1988) in Wood, W.B. (ed.) *The Nematode Caenorhabditis elegans*, Cold Spring Harbor Laboratory Press, NY, 587–606.
16. Maniatis,T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning*, A Laboratory Manual. Cold Spring Harbor Laboratory Press, NY.
17. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci.*, **74**, 5463–5467.
18. Kozak, M. (1984) *Nature*, **308**, 241–246.
19. Kozak, M. (1986) *Cell*, **44**, 283–292.
20. Morlé, F., Lopez, B., Henni, T. and Godet, J. (1985) *EMBO J.*, **4**, 1245–1250.
21. Wickens, M. (1990) *TIBS*, **15**, 277–281.
22. Covey, S.N. (1986) *Nucl. Acids Res.*, **14**, 623–633.
23. Doolittle, R.F., Feng, D.-F., Johnson, M.S. and McClure, M.A. (1989) *Quart. Rev. Biol.*, **64**, 1–30.
24. Xiong, Y. and Eickbush, T.H. (1990) *EMBO J.*, **9**, 3353–3362.
25. Repaske, R., Harley, J.W., Kavilick, M.F., O'Neil, R.R. and Austin, J.B. (1989) J. Virol., **63**, 1460–1464.
26. Kanaya, S. et al. (1990) *J. Biol. Chem.*, **265**, 4615.
27. Day, A., Schirmer-Rahire, M., Kuchka, M.R., Mayfield, S.P. and Rochaix, J.-D. (1988) *EMBO J.*, **7**, 1917–1927.
28. Day, A. and Rochaix, J.-D. (1991) *Nucl. Acids Res.*, **19**, 1259–1266.
29. Kikuchi, Y., Ando, Y. and Shiba,T. (1986) *Nature*, **323**, 824–826.
30. Garfinkel, D.J., Boeke, J.D. and Fink, G.R. (1985) *Cell*, **42**, 507–517.
31. Mellor, J., Malim, M.H., Gull, K., Tuite, M.F. and McReady, Dibbayawan, T., Kingsman, S.M. and Kingsman, A.J.(1985) *Nature*, **318**, 583–586.
32. Kingsman, A.J. and Kingsman, S.M. (1988) *Cell*, **53**, 333–335.
33. Wilson, W., Braddock, M., Adams, S.E., Rathjen, P.D., Kingsman, S.M. and Kingsman, A.J. (1988) *Cell*, **55**, 1159–1169.
34. Schlicht, H.J., Radziwill, G. and Schaller, H. (1989) *Cell*, **56**, 85–92.
35. Hull, R. and Will, H. (1989) *TIG*, **5**, 357–359.
36. Goodrich, D.W. and Duesberg, P.H. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 2052–2056.
37. Xu, H. and Boeke, J.D. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8553–8557.
38. Roberts, L. (1990) *Science*, **248**, 1310–1313.
39. Babity, J.M., Starr, T.V.B. and Rose, A.M. (1990) *Mol. Gen. Genet.*, **222**, 65–70.
40. Rubin, G.M., Kidwell, M.G. and Bingham, P.M. (1982) *Cell*, **29**, 987–994.
41. O'Brochta, D.A. and Handler, A.M. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 6052–6056.
42. Stinchcomb, D.T., Shaw, J.E., Carr, S.H. and Hirsch, D. (1985) *Mol. Cell. Biol.*, **5**, 3484–3496.
43. Fire, A. (1986) *EMBO J.*, **5**, 2673–2680.
44. Fire, A. and Waterston, R.H. (1989) *EMBO J.*, **11**, 3419–3428.
45. Cooley, L., Kelley, R. and Spradling, A. (1988) *Science*, **239**, 1121–1128.
46. Kramer, J.M., French, R.P., Park, E.-C. and Johnson, J.J. (1990) *Mol. Cell. Biol.*, **10**, 2081–2089.
47. Kimble, J., Hodgkin, J., Smith, T. and Smith, J. (1982) *Nature*, **229**, 456–458.
48. Sass, H. (1990) *Gene*, **89**, 179–186.