Association of a truncated cytochrome *c* processed pseudogene with a similarly truncated member from a long interspersed repeat family of rat

Richard C. Scarpulla

Department of Molecular Biology, Northwestern University Medical School, 303 East Chicago Avenue, Chicago, IL 60611, USA

Received 13 November 1984; Revised and Accepted 9 January 1985

ABSTRACT

The cytochrome *c* multigene family of rat contains approximately 30 processed pseudogenes that represent genomic DNA copies of three alternate mRNAs. Here, the DNA sequence of an unusual processed pseudogene reveals that it has a complete 3' noncoding region including a short poly A tail but unlike the others is abruptly truncated at its 5' end, 19 amino acid codons from the translation terminator. At this position the pseudogene is fused through 17 consecutive adenylic acid residues to a 1.3 kb repetitive sequence. This repetitive element is flanked by direct repeats and represents a truncated member from a major long interspersed repeat family. The rat element is a composite of sequences observed in long interspersed repeats from both rodents and primates. Comparison to the equivalent mouse sequences shows that the 5' half of the repeat distal to the pseudogene has an open reading frame and is highly conserved whereas the half adjacent to the pseudogene is evolutionarily unstable. The proportion of cytochrome *c* pseudogene recombinant clones containing this repetitive DNA is 3 fold greater than observed in random isolates and may reflect a general tendency of processed pseudogenes to associate with other repetitive sequences in the genome.

INTRODUCTION

The cytochrome *c* multigene family of rat is unusual because nearly all of the approximately 30 members are processed pseudogenes derived from the mRNA transcripts of a single intron-containing parent gene (1,2). Multiple genomic copies are similarly observed in eight other mammalian genomes (3) and several processed genes are now isolated from human and mouse libraries (unpublished data). These genes may therefore be considered as a repetitive sequence family whose members arise through RNA intermediates. In this respect they resemble interspersed repetitive sequences whose formation and dispersal is also thought to involve genomic integration of DNA copied from RNA transcripts (4). Both processed genes (1,5-11) and repetitive sequences (12,13) end in oligo dA tracts reminiscent of the poly A tails of mRNA and are usually flanked by direct repeats characteristic of a transposition-like integration.

Some interspersed repeat family members from both rodents and primates are estimated to have lengths of over 5 kilobases (for review see 14). However, terminal portions of these large elements can exist as smaller units of approximately a kilobase or less in length and may be independently dispersed throughout the genome. In mouse, the terminal end of some long BamHI repeats can exist as short R elements (13) or as R elements linked to an adjacent Bam5 sequence (15). A truncated and rearranged version of the terminal region from a primate KpnI family sequence designated KpnI-RET is 829bp long and is integrated in satellite DNA (16).

While investigating a family of cytochrome c processed genes in rat, an unusual truncated member was isolated. Here, this pseudogene is shown to be fused through an oligo dA tract to a similarly truncated 1.3kb repetitive sequence. The repeat is homologous to sequences found in both mouse BamHI (17,18) and primate KpnI long interspersed repeats (19). Comparison to equivalent sequences from mouse reveals that approximately one half of the element has a long open reading frame and is highly conserved in evolution whereas the other half adjacent to the pseudogene accumulates numerous insertions, deletions and base substitutions. These results indicate that distinct regions within the same element are subject to very different selective constraints.

MATERIALS AND METHODS

Recombinant clones were isolated from a genomic library constructed using partially digested EcoRI fragments from a Sprague-Dawley rat and cloned into the Charon 4A phage vector(20). Screening was carried out in 0.97 M Na⁺ and 50% formamide at 50° with a 538bp BamHI-AccI coding region fragment from pRC4 as a probe (21). In all procedures phages were propagated using Escherichia coli DP50 sup F (22) and techniques for plating, purification of phage and bulk preparation of phage DNA performed as before (21). For detailed mapping and sequencing, gene-containing restriction fragments from recombinant Ch4A-RC12 were subcloned into plasmid vector pBR322 (23) and plasmid DNA prepared according to the alkaline extraction procedure (24). Total rat DNA for genomic blotting experiments was isolated from the liver of a single Sprague-Dawley rat as described (25) and subjected to restriction enzyme and hybridization analyses.

Restriction enzymes were purchased from either New England Biolabs or Bethesda Research Laboratories and utilized according to their recommended assay procedures. Restriction fragments from digested phage or genomic DNAs

were resolved by electrophoresis on a 1% agarose gel (electrode buffer: 40mM Tris, 20mM sodium acetate, 2mM EDTA, pH 7.8) and compared to standards of known size. Following transfer of DNA to nitrocellulose, hybridization was performed in 5 x SSC, 50% formamide, 0.2% SDS, 1 x Denhardt's solution (26), 100µg/ml of sonicated, denatured calf thymus DNA and probe, 3-5 x 10⁵cpm per filter for 30 to 48 hrs at 40°. Filters were washed twice for 30 min in 5 x SSC, 50% formamide, 0.2% SDS and 4 times in 2 x SSC, 0.2% SDS at 40° and exposed to X-ray film (Kodak, XAR-5).

DNA sequences were determined by the chemical base modification and cleavage procedure (27). The majority of sequences presented were determined using restriction fragments single end-labeled at their 5' ends with [γ -³²P] ATP (3000 Ci/mmole; New England Nuclear) and T4 polynucleotide kinase, but occasionally fragments 3' end-labeled with [α -³²P] dNTP (3000 Ci/mmole; New England Nuclear) and DNA polymerase I (Boehringer Mannheim) were utilized. After second enzyme digestion or separation of strands, fragments were purified on polyacrylamide gels and subjected to the sequencing chemistry. Nucleotides 3 to 50 from the labeled end were resolved on a thin 20% acrylamide-8M urea gel while nucleotides 30 to 350 were determined from 2 loadings of an 80 cm long 6% acrylamide-8M urea gel.

RESULTS

Structure of a Truncated Processed Gene and its Association with Repetitive DNA

Recent studies demonstrate that the majority of cytochrome c - specific sequences in the rat genome are processed genes representing full-length DNA copies of an 1100 nucleotide mRNA. This mRNA is one of three transcribed from a single intron-containing parent gene (1,2). While investigating the DNA sequence of a collection of recent cytochrome c gene isolates an unusual processed pseudogene was found whose 3' noncoding region including pseudopoly A tail is essentially identical to other genomic copies of the 1100 nucleotide mRNA but whose coding region abruptly ends 19 amino acid codons from the translation terminator. Figure 1 shows a comparison of the sequence of this truncated gene (RC12) to the homologous region of the parent gene (RC4). The two are about 92% homologous over a length of 600bp. Divergence from RC4 at the 3' end coincides with an oligo dA tract found at precisely the same location as in four other processed gene derivatives of the 1100 nucleotide mRNA (2). Direct repeats flanking the pseudogene are not unambiguously defined but either oligo dA or the sequence GAAGAA, which both

Figure 1. Nucleotide sequence comparison of a truncated cytochrome *c* pseudogene to the functional parent gene. The truncated cytochrome *c* pseudogene from Ch4A-RC12 is compared to the functional parent gene from Ch4A-RC4 with the amino acid sequence from the carboxy terminal end of rat cytochrome *c* shown above. Only nucleotide differences with RC4 are shown for the homologous portion of RC12. Repetitive DNA adjacent to the RC12 pseudogene is underlined. Short direct repeats flanking the pseudogene and the repetitive DNA are indicated by horizontal arrows. The sequence GTTTAAAA found at the 3' end of the repetitive element is directly repeated at the 5' end (see Figure 3). The polyadenylation signal sequence is boxed and the S1 nuclease map position for the 3' end of the 1100 nucleotide mRNA marked by a vertical arrow.

Figure 2. Restriction enzyme map for the region of Ch4A-RC12 containing the cytochrome c pseudogene and long interspersed repeat. The diagram summarizes the structural features elucidated by restriction mapping, DNA sequencing and genomic blotting experiments. Repetitive DNA is shown by the hatched box, the truncated coding region from the cytochrome c pseudogene by the closed box and the 3' noncoding region extending to the end of the 1100 nucleotide mRNA by the open box. Regions of homology with known mouse repeat sequences designated R(13), Bam5-R composite (15), Bam5(17), L1Md-4 (18) and E24(19) are indicated below the rat repetitive DNA. The 17 adenylic acid residues adjoining the pseudogene and repeat DNA are also indicated. Symbols correspond to restriction enzyme sites as follows: *AccI* (○), *AvaII* (●), *BamHI* (△), *EcoRI* (▽), *HinfI* (■), *PvuII* (▲), *XhoI* (□).

flank the pseudogene could represent a target site duplication.

The rest of the gene is not present on the recombinant phage Ch4A-RC12 which includes about 7 kilobases upstream and 3 kilobases of DNA downstream from the gene boundaries. Immediately upstream from the site of sequence divergence from RC4 at the 5' end an A-rich region comprising a stretch of 17 consecutive adenylic acid residues is observed which resembles the putative pseudopoly A tails found in processed pseudogenes (1,5-11) and certain repetitive sequences (12,13). To investigate whether a repetitive element resides in this upstream region, a 190bp *AvaII* restriction fragment adjacent to the A-rich sequence (see Figure 3) was used as a hybridization probe in genomic blotting experiments. An intense smear of hybridization is observed (not shown) and supports the conclusion that the truncated pseudogene is closely associated with some form of reiterated sequence.

Sequence of the RC12 Repeat and Comparison to Homologous Elements from Mouse

The structure of the repetitive sequence from RC12 is examined by restriction mapping and DNA sequencing and compared to rodent elements of known sequence. A computer search reveals extensive homologies with portions of long interspersed repeats from mouse designated as R (13), Bam5 (17), Bam5-R composite (15), E24 (19) and L1Md-4 (18). Both Bam5 and E24 overlap with L1Md-4, a recently described sequence that contains a long open reading frame. All of these mouse sequences are thought to be derived from

RC12	ACAGCTGGGCGCCAGCGTTTTGTGGGCGCAAGTACCATGGTTGGGCGCTGAAGGTCAGCAATGACATT
100	
RC12	GGCAATGGTGAAGGTTTTCCCTGAGCCAGTCACGCCAAGTAACGCTCTGGTGGCCAGGCCATCTTCCAGC
200	
RC12	CCCTCTTCGAGACGTCGAATCGCCTCTGGCTGATCGCCAGAGGGTTTAAAGCGGGAATTCACAGCTGAGG
L1Md-4	BAATTTCTCACCTGAGG
E24	TTCTCACCCGAGG
300	
RC12	AACAACCCTGAGATTTACCTCACACCAAGTGAGAATGGCTAAGATCAAAAACTCAGGTGACAGCAGATGC
L1Md-4	AACAACCCTGAGATTTACCTCACACCAAGTCAGAATGGCTAAGATCAAAAATTCAGGTGACAGCAGATGC
E24	AACAACCCTGAGATTCATCTCACACCAAGTCAGAATGGGTAAAGATCAAAAATTCAGGTGACAGCAGATGC
400	
RC12	TGGCAAGGATGTGGAGAAAGAGGAACACTCCTCCATTGTTGGTGGGATTGCAGACTGGTAAAACCATTTCT
L1Md-4	TGGCAGGGATGTGGAGAAAGAGGAACACTCCTCCATTGTTGGTGGGATTGCAGGCTGTACAACTCTCT
E24	TGGCAAGGATGTGGAGAAAGGGGAACACTCCTCCATTGTTGGTGGGATTGCAAGCTGTACAACTCTCT
500	
RC12	GGAAATCAGTCTGGAGGTTTCATCAGAAAATTGGACATTGAACTGCCTGAGGATCCAGCTATACCTCTCTT
L1Md-4	GGAAATCAGTCTGGCGGTTCCTCAGAAAATTGGACATAGTACTACCGAGGATCCAGCAATACCTCTCTT
E24	GGAAATCATTTCTGGCAG. TCCTCAGAAAATTGGACATGGTACTACAGAGGA
BAM5-R	GTACTACCGAGATCCAGCAATTCCTCTCTT
BAM5	TCCAGCAATACCCCTCTT
600	
RC12	GGGCATATACCCAAAAGATGCCTCAACATATAAAAGAGACACGTGCTCCACTATGTTTCATGCGAGCCTTA
L1Md-4	GGGCATATATCCGAAGAAGCCCCAACTGGTAAGAAGGACACATGCTCCACTATGTTTCATGCGAGCCTTA
BAM5-R	GGGCATATACCTAGAAGATGTTCCAACT. TGATAAGGACACATG. T. CGCTATGCTCATAGCAGCCTTA
BAM5	GGGCATATATCCAGAAGATGCCCAACTGGTAAGAAGGACACATGCTCCACTATGTTTCATGCGAGCCTCA
700	
RC12	TTTATAATAGCCAGAAAATGGAAAGAACCCAGATGCCCTTCAACAGAGGAATGGATACAGAAAATGTGGT
L1Md-4	TTTATAATAGCCAGAAAATGGAAAGAACCCAGATGCCCTTCAACAGAGGAATGGATACAGAAAATGTGGT
BAM5-R	TTTATAATAGCCAGAAAGCTGGAAAGAACCCAGATGCCCTTCAACAGAGGAATGGATACAGAAAATGTGGT
BAM5	TTTATAATAGCCAGAAAGCTGGAAAGAACCCAGATGCCCTTCAACAGAGGAATGGATACAGAAAATGTGGT
800	
RC12	ACATCTACACAATGGAATATTACTCAGCTATCAAAAACAACGAGTCTATGAAATTCGTAGGCAAAATGGTT
L1Md-4	ACATCTACACAATGGAGTACTACTCAGCTATTAAAAAGAAATGAATTTATGAAATTCCTAGCCAAATGGAT
BAM5-R	ACATTTACACAATGGAGTACTATTTCAGCAATTAAAAACAGTGAATTAATGAAATTCCTAGGCAAACTGGAT
BAM5	ACATCTACACATGGAGTACTACTCAGCTATTAAAAAGAAATGAATTTATGAAATTCCTAGCCAAATGAAT
900	
RC12	GGAACTGGAAAATATCATCTCTGAG. TGAGCTAACC. CAATCACAGA. AAGACATACATGGTATGCACTCA
L1Md-4	GGAACTGGAGAGCATCATCTCTGAG. TGAGGTAA. CAATCACAAA. GGAACCTACACAATATGTAAGTCA
BAM5-R	GCATCTGTAGGATATCATCTCTGAGGTGAGGTAA. ACTAATCACAAA. ATAACAGCATGATATGCACTCA
BAM5	GGAACTGGAGGGCATCATCTCTGAG. TGAGGTAA. CACTCACAAAAGGAAGCTCACACAATATGTAAGT. A
1000	
RC12	TTGATAAGTGGCTATTAGCCCA. AATGCTT. GAATTACCTTAGATCCCTAGAACAAACGAACTC... AA
L1Md-4	CTGATAAGTGGATACTAG
BAM5-R	CTGATACGTGGATATTACCCA. GAAACTTAGAA. TATCCAAGATGC... AACATATGAAACTCAAGAA
BAM5	CTGATAAGTGGATATTAGCCCAAG. CCTAGGA. TA. CCAAGATAT... AAGATAC. AATCTCTTAA
1100	
RC12	GACGGATGATCAAAA... T. GT... GAATGCTTAC... TCCTT. CTTTAAATGAGGA
BAM5-R	GAAGGAAGACCAAG... T. GTG... GAAGGCTTAC... TCCTT. CTTAGAAATGGGA
BAM5	CACATGAAGCTCAGAAAAAT... G. AAGACTGAAGTCT. GG. A. CACTATACCCCTCTTGAAGTGGGA

Figure 3. Nucleotide sequence of the long interspersed repeat in Ch4A-RC12 and comparison to related mouse sequences. The nucleotide sequence of RC12 extending from the PvuII site through the polyA junction with the adjoining cytochrome *c* pseudogene (see Figure 2) is shown. Nucleotide sequences from the various homologous mouse repeat DNAs are shown below. Only that portion of L1Md-4 beginning at the EcoRI site and homologous to the RC12 repeat is shown. For each of the partial mouse elements, nucleotide identities with the rat RC12 repeat are underlined. Gaps are inserted to maximize homologous alignment of the various sequences. Short direct repeats flanking the long interspersed repeat in RC12 are indicated by horizontal arrows.

the BamHI (or MIF-1) family of long repetitive sequences (17, 18).

Figure 2 shows a restriction map of the RC12 repeat and linked cytochrome c pseudogene and also illustrates its relationship to the various published mouse sequences. Figure 3 shows a comparison of their nucleotide sequences. Homology with the overlapping mouse segments extends over 1.3kb from the oligo dA tract at the 3' end adjacent to the pseudogene to the EcoRI site at the 5' end of the repetitive DNA where homology with the equivalent L1Md-4 sequence of mouse precisely ends. Genomic hybridizations also reveal that this EcoRI site defines the upstream boundary of the repetitive DNA in RC12. The 1.3kb segment is flanked by an 8bp direct repeat (GTTTAAAA) that coincides with the 5' and 3' boundaries of homology with the comparable mouse sequences. Similarly, a second random isolate from the same family of rat repeats shares 97% sequence homology with the RC12 element over the same 1.3kb region between the direct repeats but its homology with mouse L1Md-4 continues further upstream. Thus, the rat RC12 element most likely represents a truncated member of a long interspersed repeat family that is derived from the rat equivalent of the mouse BamHI (or MIF-1) family. The direct repeats flanking the element support the notion that it originated from the 3' terminal portion of a longer repeat and was independently transposed to a new genomic location. An alternative explanation, which is difficult to exclude because of the multiplicity of these sequences in the genome, is that the 5' end of the repeat was truncated during construction of the genomic library and fortuitously ligated through the EcoRI site at its 5' end to an unrelated fragment containing the 5' direct repeat sequence.

Optimum alignment of the rat and mouse repetitive sequences (Figure 3) reveals an interesting pattern of sequence conservation and divergence. The first 580bp of the RC12 repeat includes a long open reading frame and displays nearly 90% sequence conservation with the comparable mouse segments. Immediately following this position and for the remainder of the element, numerous gaps must be inserted in both rat and mouse sequences to achieve maximum alignment. Thus, about half of the RC12 repeat closest to the pseudogene has accumulated many small insertions and deletions as well as base substitutions suggesting that selective constraints for this region are much less stringent.

Association of Repetitive Sequences with Other Pseudogene Recombinant Clones

Genomic hybridizations suggest that the rat repeat of RC12 is widely distributed throughout the genome and it thus seemed probable that related

Figure 4. Detection of repetitive DNA in rat cytochrome *c* recombinant clones Ch4A-RC4, 5, 6, 8, 9, 10, 12 and 13. Each lane depicts the electrophoretic banding pattern from 0.5 μ g of EcoRI digested recombinant phage DNA visualized either by ethidium bromide staining (left panel) or by Southern transfer hybridization (right panel) using the 32 P-labeled EcoRI-XhoI restriction fragment from the RC12 repeat (see Figure 2) as a hybridization probe. Common restriction fragments from the left and right Charon 4A vector arms are indicated by the arrows. The restriction fragment containing the cytochrome *c* gene or pseudogene is marked by a closed triangle for each recombinant clone. Size standards from HindIII digested λ DNA are indicated at the left.

copies would reside on other cytochrome *c* recombinant phages. The eight nonallelic genomic clones whose cytochrome *c* sequences are already established (2) were therefore tested for the presence of DNA related to the RC12 repeat. Figure 4 shows that six of the seven pseudogene clones contain sequences homologous to the RC12 repeat. Only the parent gene RC4 and pseudogene RC5 are devoid of repeat DNA. Clones RC8 and 13 display a weaker hybridization signal than the others suggesting the presence of a more distantly related member of the repeat family. Both RC9 and 12 contain the repetitive sequence on the same Eco RI fragment as the pseudogene. In RC9 the repetitive DNA is located 2kb upstream from the 5' boundary of the

```

RC5      TGTGTATCTTGGAGAGCTCATAAGACCTGTAACCTGAGCTATGATAGGCATAAGACACA
           ───────────┐──────────┐──────────┐──────────┐──────────┐
RC6      TAAGGTAAGTCCACAGACCATAAAATTTGAATTGAAGAAATTTACCTGTACAAAAATATGGT
           ───────────┐──────────┐──────────┐──────────┐──────────┐
RC8      TGTTTAAAAACAGGATAAGAGATAAGATAGAAAGGAACTGGATGTGGTAATGTGAGGCTG
           ───────────┐──────────┐──────────┐──────────┐──────────┐
RC9      TCTATATTCAAATATTGCTTCAAATTCCTCTTCAATGCTCAATTAAGTTCCTTATAAGTC
           ───────────┐──────────┐──────────┐──────────┐──────────┐
RC10     ATTTTACAGAGAAATGCAGGTATAAAGAGTGAACATTTTATGAAAGTGACAGGATGGG
           ───────────┐──────────┐──────────┐──────────┐──────────┐
RC12     AAAAAATGTTAAAAAAAAAAAAAAAAAAGAGAAATACAGAAATAAAGTCCTTCTAGAATG
           ───────────┐──────────┐──────────┐──────────┐──────────┐
RC13     TGCAGTGAAATGTGCCCAAGTAAGAAAATCTTAAATCAACTGATTCACGTCTTTAATATT
           ───────────┐──────────┐──────────┐──────────┐──────────┐
12-REPEAT AATCGCCTCTGGCTGATCGCCAGAAGGTTTAAAAAAAAAAAAAAAAAAGAGAGGGAGAA
           ───────────┐──────────┐──────────┐──────────┐──────────┐

```

Figure 5. Comparison of DNA flanking seven cytochrome *c* pseudogenes and the RC12 repeat DNA. Putative target sites for integration of cytochrome *c* pseudogenes and the long interspersed repeat from RC12 are reconstructed by overlapping the short direct repeats (underlined) flanking the 5' and 3' boundaries of each element and deleting the sequence in between. Direct repeats flanking each target sequence are marked by horizontal arrows.

cytochrome *c* sequence. Thus, nearly all of the cytochrome *c* recombinant clones have both a pseudogene and repetitive sequence within the same 12 to 15kb insert of genomic DNA. To examine the distribution of the RC12 repeat within the original rat Charon 4A library from which all of the cytochrome *c* recombinants were isolated (2,21) the proportion of phages homologous to the RC12 repeat was determined for 5000 plaques plated at random. Approximately 25% of the recombinant clones are detected with the RC12 repeat as a hybridization probe. Although the hybridization signal was somewhat weaker, essentially the same result is achieved with an independently constructed Balb-C mouse library in λ L47. These results indicate that approximately 1.7% of the genome represents sequences related to the rat repeat. Because six out of seven or 86% of the cytochrome *c* pseudogene clones also contain regions homologous to the RC12 repeat we conclude that the association between these two in recombinant isolates occurs approximately 3 times more frequently than expected on a random basis.

One potential explanation for the association of processed pseudogenes and long interspersed repeats is that they integrate at similar target sequences. A comparison of the DNA flanking the seven processed cytochrome *c* pseudogenes and the RC12 repeat (Figure 5) shows no strong homologies common to all of the putative target sites. In all cases however, the DNA adjacent to the target site duplication is A and T rich (60 to 75% A + T) and the majority also have short direct repeats flanking the original

duplicated sequence. These characteristics are also observed in the DNA flanking R-elements which are dispersed throughout the mouse genome (13). No conservation of sequence or length is apparent in the short direct repeats at the pseudogene boundaries.

DISCUSSION

In this work the structure of an unusual cytochrome c processed pseudogene and its association with a member from a major long interspersed repeat family is described. The linkage of these two within the rat genome is intriguing because they each represent truncated members of their respective repeat families. In addition, they each have flanking direct repeats and short poly A tails characteristic of a growing list of sequences thought to arise from RNA transcripts and to integrate through a transposition-like mechanism. This repetitive sequence of rat differs from Bam5-R elements of mouse (15) by having an additional 274bp of the rat equivalent of E24, an independently described mouse repeat (19). It also differs from KpnI-RET (16) a truncated element from the primate KpnI family which has portions of its sequence deleted and rearranged relative to longer members from the same family. Comparison of the RC12 truncated repeat to an independently isolated longer version reveals 97% sequence homology and no such deletions or rearrangements.

While this work was in progress, a region from the BamHI family of mouse designated LIMd-4 was reported to contain a long open reading frame (984bp) and to evolve as a protein encoding sequence (18). This sequence contains the earlier described E24 (19) and Bam5 (17) sequences but continues further upstream in longer elements. Here it is found to be homologous to the region of the RC12 repeat that is highly conserved between rat and mouse. As already documented for various mouse species (18) the equivalence of replacement and silent changes observed here between rat and mouse, despite the predominance of replacement sites, is also consistent with a possible protein encoding function for this region. The existence of this putative protein remains to be established. Nevertheless, the 3' half of the RC12 repeat adjacent to the pseudogene contrasts sharply with the 5' half by having many short insertions and deletions relative to the mouse sequences suggesting that very different selective constraints operate on these regions.

Both Bam5-R elements of mouse (15) and the present RC12 rat element begin at different positions within the region comprising the open reading

frame. If RNA intermediates are involved in the propagation and dispersal of these truncated elements, they may be integrated DNA copies of RNA transcripts initiated from different positions within a longer repeat. They may also be the products of incomplete reverse transcripts of longer sequences. The existence of the truncated cytochrome c pseudogene of RC12 suggests that incomplete reverse transcripts may be integrated since it seems unlikely that the pseudogene is the product of aberrant transcription initiation. Failure to detect the remainder of the pseudogene in at least 6kb upstream from the 5' end of the repetitive sequence argues against a simple integration of the truncated repeat into a complete processed gene. It is possible that the observed structure is the product of multiple integration or recombination events. The absence of direct repeats flanking the entire region containing both the repeat and processed pseudogene indicates that the entire fused structure probably did not integrate as a unit.

Because so little is known about the interactions and functions of repetitive sequences, it is not clear whether the association between the pseudogene and the reiterated sequence in RC12 has functional or evolutionary significance. The compositional and structural similarities between the DNA flanking the processed cytochrome c pseudogenes and that surrounding other dispersed elements (13) are subtle but nevertheless intriguing. One possibility is that repetitive sequences may facilitate the dispersion of all or part of the genetic information in processed pseudogenes thereby increasing the chances of desirable interactions between otherwise nonfunctional units. We find that the distribution of the RC12 repeat homologies among the cytochrome c pseudogene recombinant clones is several fold greater than found in a random collection of phages from the same genomic library. Thus it appears that these two types of sequence element have a tendency to associate with one another in the rat genome.

ACKNOWLEDGEMENT

I thank Randall P. French and Mary C. Kilar for their excellent technical assistance.

This work was supported by grant GM32525 from the National Institutes of Health, by the Earl M. Bane Biomedical Research Fund and in part by grant 83-34 from the Illinois Division of the American Cancer Society.

REFERENCES

1. Scarpulla, R.C. and Wu, R. (1983) *Cell* **32**, 473-482.
2. Scarpulla, R.C. (1984) *Mol. Cell. Biol.* **4**, 2279-2288.
3. Scarpulla, R.C., Agne, K.M. and Wu, R. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 739-743.
4. Jagadeeswaran, P., Forget, B.G. and Weissman, S.M. (1981) *Cell* **26**, 141-142.
5. Nishioka, Y., Leder, A. and Leder, P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2806-2809.
6. Vanin, E.G., Goldberg, G.I., Tucker, P.W., and Smithies, O. (1980) *Nature* **286**, 222-226.
7. Hollis, G.F., Hieter, P.A., McBride, O.W., Swan, D. and Leder, P. (1982) *Nature* **296**, 321-325.
8. Karin, M. and Richards, R.I. (1982) *Nature* **299**, 797-802.
9. Lemischka, I. and Sharp, P.A. (1982) *Nature* **300**, 330-335.
10. Masters, J.N., Yang, J.K., Cellini, A. and Attardi, G. (1983) *J. Mol. Biol.* **167**, 23-36.
11. Lee, M.G.-S., Lewis, S.A., Wilde, C.D. and Cowan, N.J. (1983) *Cell* **33**, 477-487.
12. Rubin, C.M., Houck, C.M., Deininger, P.L., Friedmann, T. and Schmid, C.W. (1980) *Nature* **284**, 372-374.
13. Gebhard, W., Meitinger, T., Hocht, J. and Zachau, H.G. (1982) *J. Mol. Biol.* **157**, 453-471.
14. Singer, M.F. (1982) *Cell* **28**, 433-434.
15. Wilson, R. and Storb, U. (1983) *Nucleic Acids Res.* **11**, 1803-1817.
16. Thayer, R.E. and Singer, M.F. (1983) *Mol. Cell. Biol.* **3**, 967-973.
17. Fanning, T.G. (1983) *Nucleic Acids Res.* **11**, 5073-5091.
18. Martin, S.L., Voliva, C.F., Burton, F.H., Edgell, M.H. and Hutchison III, C.A. (1984) *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2308-2312.
19. Singer, M.F., Thayer, R.E., Grimaldi, G., Lerman, M.I. and Fanning, T.G. (1983) *Nucleic Acids Res.* **11**, 5739-5745.
20. Sargent, T.D., Wu, J.R., Sala-Trepat, J.M., Wallace, R.B., Reyes, A.A. and Bonner, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3256-3260.
21. Scarpulla, R.C., Agne, K.M. and Wu, R. (1981) *J. Biol. Chem.* **256**, 6480-6486.
22. Leder, P., Tiemeier, D. and Enquist, L. (1977) *Science* **196**, 175-177.
23. Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heyneker, H.L. and Boyer, H.W. (1977) *Gene* **2**, 95-113.
24. Birnboim, H.C. and Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513-1523.
25. Blin, N. and Stafford, D.W. (1976) *Nucleic Acids Res.* **3**, 2303-2308.
26. Denhardt, D.T. (1966) *Biochem. Biophys. Res. Commun.* **23**, 641-646.
27. Maxam, A.M. and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499-560.