Volume 13 Number 2 1985 Nucleic Acids Research

Characterization of a mouse somatic cytochrome c gene and three cytochrome c pseudogenes

Keith J.Limbach and Ray Wu

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, NY 14853, USA

Received 4 September 1984; Revised and Accepted 17 December 1984

ABSTRACT

Mouse contains two functional, but differentially expressed, cytochrome c genes. One of these genes is expressed in all somatic tissues so far examined. The other gene is expressed only in testis and is assumed to be spermatogenesis-specific. The nucleotide sequence of four mouse cytochrome c-like genes has been determined. One of these genes (MC1) contains an intron and encodes a polypeptide sequence identical to the published mouse somatic cytochrome c amino acid sequence. The other three genes can not properly encode a mouse cytochrome c protein and appear to be pseudogenes which have arisen via an insertion into the mouse genome of a cDNA copy of a cytochrome c mRNA molecule.

INTRODUCTION

Cytochrome c is a small, heme-containing protein found in the mitochondria of all eucaryotic cells. It is one of a series of proteins involved in the electron transport chain, and is therefore essential for aerobic respiration. Due to structural and functional constraints, the cytochrome c proteins of widely divergent organisms exhibit considerable amino acid homology[1]. Assuming this homology to be retained at the DNA level, it was considered likely that a cytochrome c gene from one organism could be used as a specific hybridization probe to isolate the cytochrome c genes of other widely divergent organisms. Utilizing this interspecies homology, cytochrome c genes from six species have now been isolated.

Hybridization analysis of total genomic DNA indicates that most non-mammalian species contain only one or two cytochrome c-like sequences per haploid genome. In contrast, mammals contain approximately 20-30 cytochrome c-like sequences[2], most of which appear to be pseudogenes[3].

At least three different classes of pseudogenes have been isolated. One class appears to have arisen via duplication events and subsequent sequence divergence. These pseudogenes contain introns and are frequently linked to the structural gene. For example, the mouse genome contains

seven β-globin-like sequences within a 70 kb region of DNA. Three of these genes are active and four are pseudogenes[4].

The other two classes of pseudogenes appear to have arisen via an RNA intermediate. These pseudogenes do not appear to be linked to the structural gene and do not contain introns. One of these two classes appears to have arisen via a retroviral intermediate. Two examples of this class of pseudogenes are the mouse α-globin pseudogene, αψ3[5], and the human immunoglobulin pseudogene, $C_{\varepsilon 3}$[6], both of which are flanked by retrovirus LTR-like elements.

The third, and possibly most abundant, class of pseudogenes appears to have arisen via a mRNA intermediate[7]. In addition to lacking introns, these pseudogenes contain a relatively long stretch of A-residues in the 3'-noncoding region, reminiscent of the poly(A) tail of a mRNA molecule. This class of pseudogenes has been observed in numerous systems, such as the rat α-tubulin pseudogenes[8], the rat cytochrome c pseudogenes[3] and the rat glyceraldehyde 3-phosphate dehydrogenase pseudogenes (Tso and Wu, unpublished observations).

The number of pseudogenes per genome appears to vary significantly for different genes. For example, the rat genome contains approximately 25 cytochrome c-like sequences and 300 glyceraldehyde 3-phosphate dehydrogenase-like sequences (Tso and Wu, unpublished observations). Since it appears that the vast majority of these sequences are mRNA-derived pseudogenes, this variation may reflect the relative level of each mRNA species present when these pseudogenes were generated. Therefore, the fact that rat glyceraldehyde 3-phosphate dehydrogenase pseudogenes greatly outnumber rat cytochrome c pseudogenes may indicate that when these pseudogenes were generated, the glyceraldehyde 3-phosphate dehydrogenase gene was expressed at a much higher level than the cytochrome c gene.

To be passed on to subsequent generations, these pseudogenes must have been generated either during spermatogenesis or oogenesis, in primordial germ cells, or in those early embryonic cells which later differentiated into primordial germ cells. A reverse transcriptase-like enzyme is also required.

Mouse contains two different cytochrome c proteins[9]. One of these proteins is found in all somatic tissues so far examined. The other protein is found only in the testis and is assumed to be spermatogenesis-specific[9]. The testis-specific cytochrome c protein differs from the somatic cytochrome c protein at 13 of 104 amino acid

residues[9], therefore, the nucleotide sequence of the two cytochrome c genes is expected to be significantly different.

The mouse cytochrome c pseudogenes should reflect the relative expression of the two cytochrome c genes at the time when these pseudogenes were generated. For example, if the reverse transcription and insertion events leading to the generation of pseudogenes occurred during spermatogenesis, the nucleotide sequence of the resultant pseudogenes should resemble the nucleotide sequence of the putative spermatogenesis-specific cytochrome c gene.

In this communication, we report the existence of a testis-specific cytochrome c transcript. We describe the isolation and characterization of a mouse somatic cytochrome c gene and three mouse cytochrome c pseudogenes. The nucleotide sequence of these pseudogenes indicates that they were derived by an insertion of a cDNA copy of a mouse somatic cytochrome c mRNA molecule.

MATERIALS AND METHODS

The mouse genomic library was provided by the laboratory of Philip Leder (Harvard Medical School). The library was constructed by partially digesting genomic DNA from a Balb/c 12 day old mouse embryo with the restriction endonuclease MboI and ligating fragments approximately 16-20 kb long onto the purified BamHI arms of the lambda phage Charon 28[10]. From the expected size of the mouse DNA inserts (16-20 kb) and the approximate size of the mouse genome (3×10^6 kb), it was estimated that screening approximately 750,000 phage would ensure, with 99% probability, that the entire mouse genome was represented[11]. The hybridization probe was the 0.55 kb BamHI-AccI fragment of pRC4, which contains the entire coding region of the rat (RC4) cytochrome c gene[12]. Hybridization and washing conditions were the same as those previously described[13].

Plasmid DNA, genomic DNA and polyadenylated RNA were isolated and analyzed as previously described. DNA sequence analysis was performed by the dideoxynucleotide chain-termination procedure[14].

RESULTS

Complexity of Cytochrome c-Like Sequences in the Mouse Genome

In order to investigate the complexity of cytochrome c-like sequences in the mouse genome, total DNA was isolated from the liver of an individual Balb/c mouse, cut with various restriction endonucleases, fractionated by

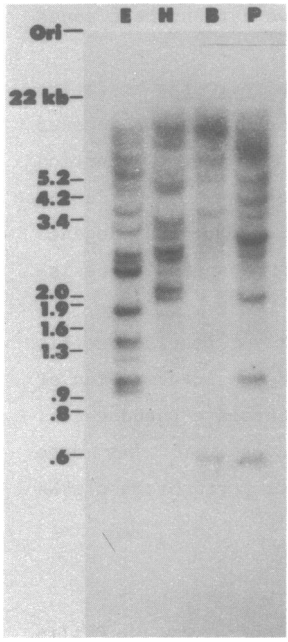


Fig. 1. Hybridization Analysis of Mouse Genomic DNA. Each lane contains 50 µg of genomic DNA from the liver of a Balb/c mouse. Samples were digested with EcoRI (E), HindIII (H), BamHI (B) or PstI (P), resolved by gel electrophoresis, transferred to nitrocellulose and hybridized at 30°C to the ^{32}P-labeled BamHI-AccI fragment of pRC4 (coding region of RC4). Size standards are the EcoRI-HindIII digestion fragments of λ phage DNA.

electrophoresis, transferred to nitrocellulose and hybridized to the coding region of the rat (RC4) cytochrome c gene[12]. Approximately 25 hybridizing fragments are observed in each lane (Fig. 1). This is consistent with previous results which had indicated that mammalian genomes contain approximately 20-30 cytochrome c-like sequences[2].

Analysis of Cytochrome c mRNA from Four Mouse Tissues

Two cytochrome c proteins have been isolated in mouse[9]. It therefore appears that mouse contains at least two functional cytochrome c genes. These two genes appear to be differentially expressed. One gene is expressed in all somatic tissues so far examined. The other gene appears to be expressed only in the testis, and is assumed to be spermatogenesis-specific. The two mouse cytochrome c proteins differ at 13 of 104 amino acid residues (Fig. 2). Although these 13 amino acid substitutions appear to be distributed over the entire cytochrome c molecule, examination of polypeptide tertiary structure indicates that all of the substitutions are restricted to a small region of the molecule behind the heme crevice.

In order to determine whether the rat cytochrome c gene could be used as a hybridization probe tc detect the mouse testis-specific cytochrome c

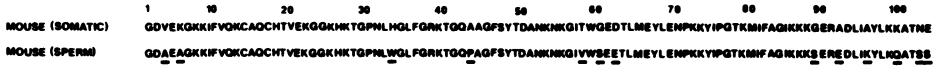


Fig. 2. Amino Acid Sequences of the Mouse Somatic and Testis-Specific Cytochrome c Proteins. The data are taken from Hennig (9). Residues which are not homologous are underlined. The amino acid letter code is as follows: A, Ala; R, Arg; N, Asn; D, Asp; C, Cys; E, Glu; G, Gly; H, His; I, Ile; L, Leu; K, Lys; M, Met; F, Phe; P, Pro; S, Ser; T, Thr; W, Trp; Y, Tyr; V, Val.

transcript, polyadenylated RNA was isolated from mouse liver, testis, kidney and heart, fractionated on an agarose-formaldehyde gel, transferred to a nitrocellulose filter and hybridized to the coding region of the rat (RC4) cytochrome c gene (Fig. 3). In each of the four tissues analyzed, three or four hybridizing mRNA species are observed. In testis, however, an additional cytochrome c mRNA species is observed. This testis-specific transcript is assumed to encode the testis-specific cytochrome c protein.

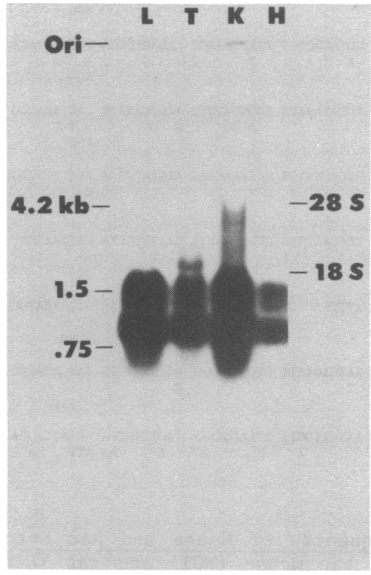


Fig. 3. Hybridization Analysis of Polyadenylated RNA from Four Mouse Tissues. Polyadenylated RNA from liver (L), testis (T), kidney (K) and heart (H) (30 µg, 50 µg, 50 µg and 10 µg, respectively) was resolved by gel electrophoresis, transferred to nitrocellulose and hybridized at 16°C to the ^{32}P-labeled BamHI-AccI fragment of pRC4 (coding region of RC4). Size standards include 28S and 18S mouse ribosomal RNA and the PstI, EcoRI and EcoRI-HindIII fragments of pDC3 (Limbach and Wu, accompanying paper).

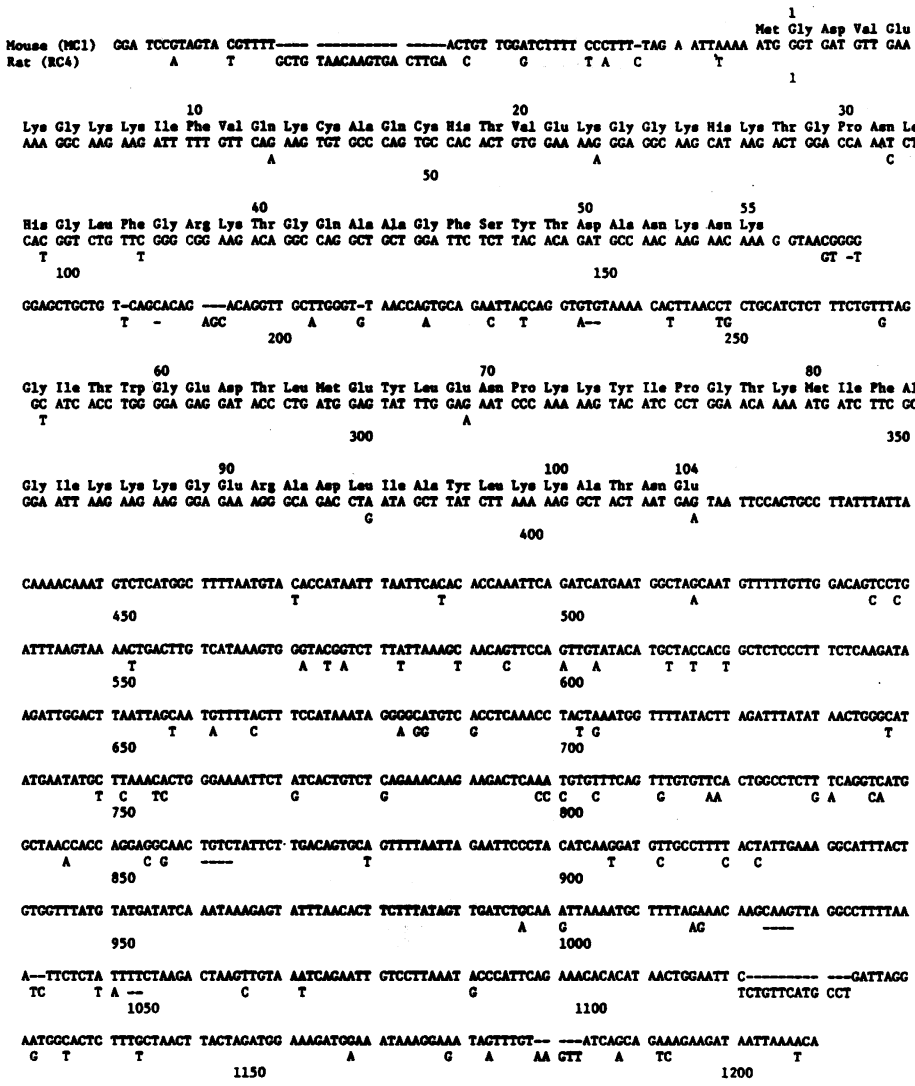


Fig. 4. Nucleotide Sequences of Mouse and Rat Cytochrome c Genes. The nucleotide sequences of the mouse (MC1) and rat (RC4) cytochrome c genes are aligned for maximum homology. The complete sequence of MC1 is presented. Only those nucleotides that differ from MC1 are shown for RC4. The predicted amino acid sequence is displayed on the line above the nucleotide sequences. The numbering system begins from the initiation codon. Nucleotides are numbered on the line below the nucleotide sequences and amino acid residues are numbered on the line above the nucleotide sequences. Putative polyadenylation signal sequences are overlined.

Since the somatic gene probe does not hybridize with equal efficiency to both the somatic and testis-specific transcripts, and because the testis contains both somatic and germ-line cells, a quantitative evaluation of the relative expression of the somatic and spermatogenesis-specific cytochrome c genes in developing spermatozoa can not be made.

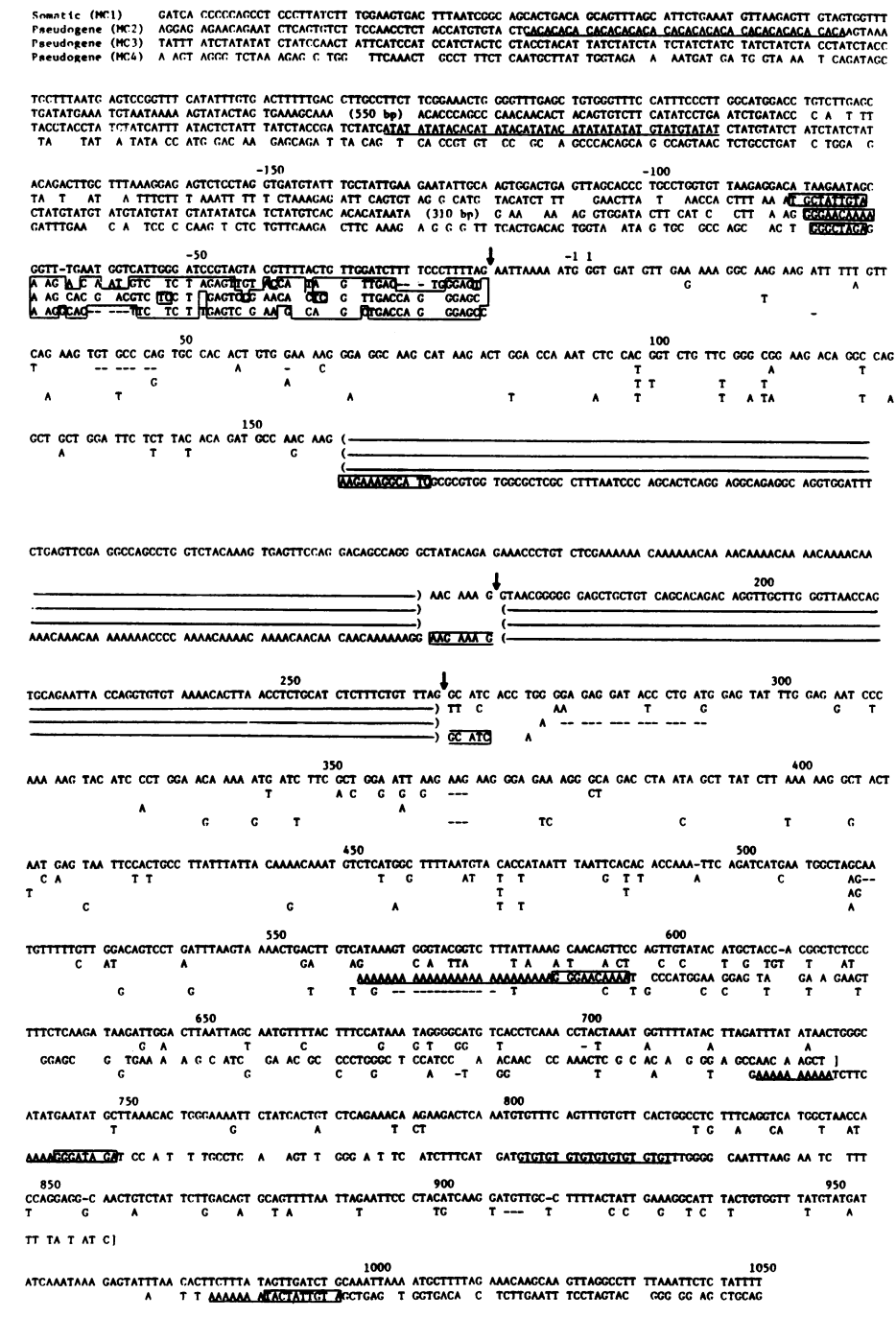
The testis-specific transcript is approximately 1600 nucleotides long. The somatic cytochrome c transcripts are approximately 1300, 1100 and 800 nucleotides long. Densitometry analysis indicates that the 800-nucleotide-long mRNA band is approximately 2.5 times and 3 times more intense than the 1300- and 1100-nucleotide-long mRNA bands, respectively. In addition to being more intense than the other mRNA bands, the 800-nucleotide-long band is broader, and therefore, may represent more than one mRNA species.

Isolation of a Mouse Somatic Cytochrome c Gene and Three Cytochrome c Pseudogenes

Since the rat cytochrome c gene (RC4) was able to hybridize to the mouse somatic and testis-specific cytochrome c transcripts, it was considered likely that the rat cytochrome c gene could be used as a hybridization probe to isolate the two mouse cytochrome c genes from a mouse genomic library.

Approximately 1,000,000 plaques were screened and eight unique clones (MC1-8) isolated. Detailed sequence analysis was performed on four of these clones (MC1-4). MC1 contains a mouse somatic cytochrome c gene which encodes an amino acid sequence identical to the published amino acid sequence (Fig. 4). This clone contains an intron, 104 bp long, which interrupts the coding sequence at the first nucleotide of codon 56. The intron-exon junctions of this intervening sequence follow the obligatory GT/AG rule, and in general, agree with the consensus RNA splice site sequences[15].

The nucleotide sequences of MC2, MC3 and MC4 are very similar to MC1 (Fig. 5). These clones, however, contain all the characteristics commonly associated with processed pseudogenes: they do not contain introns, they contain long tracts of A-residues approximately 20 bp downstream from putative polyadenylation signal sequences[16], and sequence homology with the structural gene is bounded by direct repeats, which are assumed to have been created when the pseudogenes were inserted into the mouse genome. It therefore appears that MC2, MC3 and MC4 originated, not by gene duplication events, but by the insertion of a reverse transcript of a mouse cytochrome c mRNA molecule.



The nucleotide sequences of these three processed pseudogenes resemble the mouse somatic cytochrome c gene (MC1), rather than the putative spermatogenesis-specific cytochrome c gene. It therefore appears that these three pseudogenes were generated from somatic cytochrome c mRNA molecules.

Assuming that MC2, MC3 and MC4 originated from a reverse transcript of an MC1 encoded mRNA molecule, comparative sequence analysis of the structural gene (MC1) and the pseudogenes (MC2-4) should reveal where putative intervening sequences reside, where transcription initiation occurs and where polyadenylation of MC1 transcripts occur. For example, sequence homology with MC1 ends abruptly at position -12. Homology among MC2, MC3 and MC4, however, continues for an additional 60 base pairs (positions -12 to -71). Since the nucleotide sequence of MC1, at position -12, resembles a 3' RNA acceptor splice site, it is likely that MC1 contains an intervening sequence at this position. The 60 base pair region of homology found in MC2, MC3 and MC4 is assumed to be the equivalent of the 5'-noncoding region exon of the expressed gene. It therefore appears that MC1 contains two intervening sequences and that both have been accurately spliced out in MC2, MC3 and MC4.

The 5'-noncoding region exon of MC1 has not been located. Since approximately 380 bp of MC1 5'-noncoding region sequence has been determined, it appears that the 5'-noncoding region intron of MC1 is considerably longer than the coding region intron.

Homology with MC1, in the 3'-noncoding region, does not end at the identical position in the three pseudogenes. In each case, however, sequence homology ends at a putative polyadenylation addition site. For example, in MC2, sequence divergence occurs 20 bp downstream from the sequence AATAAA (position +980). In MC4, sequence divergence occurs 22 bp downstream from the sequence ACTAAA (position +723). In MC3, sequence

Fig. 5. Nucleotide Sequences of a Mouse Cytochrome c Gene and Three Mouse Cytochrome c Pseudogenes. The nucleotide sequences of the mouse cytochrome c gene (MC1) and three cytochrome c pseudogenes are aligned for maximum homology. The complete sequence of MC1 is presented. Only those nucleotides that differ from MC1 are shown for MC2, MC3 and MC4. The numbering system is the same as Fig. 4. Putative polyadenylation signal sequences are overlined. Alternating $(purine\text{-}pyrimidine)_n$ sequences and poly(A)-like sequences are underlined. Direct repeats and 5'-noncoding region homology among MC2, MC3 and MC4 are boxed. Horizontal lines represent insertions or deletions. Arrows indicate the positions of intron boundaries. Nucleotides -740 to -241 are not shown for MC2 and nucleotides -430 to -121 are not shown for MC3.

divergence occurs 18 bp downstream from the sequence AAGTAAA (position +562). In addition, a relatively long stretch of A-residues, reminiscent of a poly(A) tail, is found at each of these sites. For example, MC3 contains a stretch of 26 consecutive A-residues, MC4 contains a stretch of 10(14) A-residues and MC2 contains a stretch of seven A-residues.

These observations suggest that MC1 transcripts are polyadenylated at at least three different places. Since an additional polyadenylation signal sequence is also observed at position +1164, it appears that four cytochrome c mRNA species may be encoded by the MC1 gene. By subtracting the length of the intron and allowing 200 nucleotides for the poly(A) tail, these putative transcripts should be approximately 1300, 1100, 950 and 800 nucleotides long.

Hybridization analysis of mouse polyadenylated RNA detects three or four cytochrome c mRNA species, 1300, 1100 and 800 nucleotides long (Fig. 2). The broadness and increased intensity of the 800-nucleotide-long band, however, had indicated that an additional mRNA species, approximately 900 nucleotides long, might exist. Utilization of the putative polyadenylation signal sequence, at position +723, would result in a mRNA molecule approximately 950 nucleotides long. Since MC4 appears to have originated from a mRNA transcript which had utilized this polyadenylation signal sequence, it would appear that four cytochrome c mRNA species exist.

Due to numerous mutational events, MC2, MC3 and MC4 can not encode functional cytochrome c proteins. For example, a C→T substitution at codon 12 of MC2 results in the termination codon, TAG. Similarly, a G→A substitution at codon 59 of MC3 results in the termination codon, TGA. Transcription and translation of these clones would therefore result in truncated, nonfunctional proteins being produced. In addition, due to a one bp deletion at codon 9, transcription and translation of MC4 would result in a nonsense polypeptide being produced. Obviously, these three pseudogenes can not produce functional cytochrome c proteins.

By examining the number of mutational events which have occurred in each of the three pseudogenes, it is possible to calculate when each of these sequences was inserted into the mouse genome. Using the value 7×10^{-9} substitutions/nucleotide/year as an estimate of the rate of silent site mutations[17], and by counting the number of mutations which have accumulated in a defined, nonselective sequence (position +420 to +560), it was calculated that MC2, MC3 and MC4 were inserted into the mouse genome approximately 22 MYs, 4 MYs and 9 MYs ago, respectively. Performing

similar calculations on a putative selective sequence (coding region), values of 16 MYs, 6 MYs and 12 MYs were obtained. If these pseudogenes had been under selective pressure at one time, the values obtained for the coding region should be smaller than the values obtained for the noncoding region. Since the two values are approximately the same, it appears that MC2, MC3 and MC4 were never under selective pressure and were never functionally active. This is not surprising since these pseudogenes do not contain promoter sequences or other transcription control regions.

MC4 contains a 232 bp insertion at position +160. This insertion is flanked by two 12 bp direct repeats, which are assumed to have been created when insertion occurred. Since the 3' direct repeat created by this insertion spans the site of the cytochrome c coding region intron (MC1), it appears that MC4 did not contain this intron when the insertion event occurred. This is consistent with the putative origin of this processed pseudogene.

The 232 bp insertion is extremely homologous (90%) with a highly repetitive, AluI-like, mouse sequence called B1[18,19]. Approximately 40,000-80,000 copies of B1 are scattered throughout the mouse genome. A block of short repeats, A_mC (m= 2-8), is always found flanking the 3' end of B1 sequences. A block of short A_mC repeats is also found at the 3' end of the MC4 insertion.

DISCUSSION

Pseudogene Production

Mouse appears to contain two functional, but differentially expressed, cytochrome c genes. One of these cytochrome c genes appears to be expressed in all tissues. The other gene, however, appears to be expressed only in testis and is assumed to be spermatogenesis-specific[9]. A mouse somatic cytochrome c gene and three cytochrome c pseudogenes have been isolated. These three pseudogenes have all the characteristics commonly associated with processed pseudogenes and are assumed to have arisen by an insertion into the genome of a cDNA copy of a cytochrome c mRNA molecule. To be successfully transmitted to subsequent generations, these pseudogenes must have been generated either during spermatogenesis or oogenesis, in primordial germ cells, or in those early embryonic cells which later differentiated into primordial germ cells. The mouse cytochrome c pseudogenes should reflect the relative expression of the two cytochrome c genes at the time when these pseudogenes were generated. Therefore, if the

mouse cytochrome c pseudogenes were generated during spermatogenesis, the nucleotide sequence of these pseudogenes should resemble the sequence of the putative spermatogenesis-specific cytochrome c gene. However, the nucleotide sequence of the three mouse pseudogenes which we analyzed resemble the somatic cytochrome c gene, rather than the spermatogenesis-specific cytochrome c gene, indicating that the somatic cytochrome c transcripts were the major cytochrome c mRNA molecules present when these pseudogenes were generated.

The pseudogene, MC4, is interrupted by a 232 bp insertion. MC4 was calculated to have inserted into the mouse genome approximately 10 MYs ago. Therefore, this region of DNA has been the site of two major insertion events within the last 10 MYs. These insertion events may not be random. They may reflect a stretch of DNA which contains a favorable region for these insertions to occur.

MC4 also contains the sequence $(GT)_{10}$, at position +802, approximately 50 bp downstream from the site of MC4 insertion. MC2 contains the complementary sequence $(CA)_{21}$, at position -830, approximately 760 bp upstream from the site of MC2 insertion. A similar sequence also appears to be present in Charon 28-MC3, Charon 28-MC5 and Charon 28-MC6 because each recombinant phage contains at least one restriction fragment which hybridizes to a probe containing the $(CA)_{21}$ sequence (results not shown). In fact, in each of the three recombinant phage clones, the restriction fragment that hybridizes to the cytochrome c probe also hybridizes to the $(CA)_{21}$-containing probe.

$(GT)_n$ sequences are quite numerous in mammalian genomes. It has been estimated that approximately 10^5 copies of $(GT)_n$ sequences exist in mouse[20]. Since the mouse genome consists of approximately 3×10^6 kb, this sequence should occur approximately once every 30 kb. In this study, 2.1 kb of pseudogene flanking nucleotide sequence was determined and two $(GT)_n$ sequences were observed. The probability of two $(GT)_n$ sequences being observed in 2.1 kb is very low.

MC3 contains another novel sequence approximately 350 bp upstream from the site of MC3 insertion. This region of DNA is approximately 250 bp long (position -680 to -430) and consists of a series of two bp and four bp direct repeats. One particular segment of this sequence consists of 44 consecutive base pairs of alternating purine-pyrimidine residues (position -544 to -501). It has been demonstrated that alternating purine-pyrimidine sequences, such as poly(dG-dT), poly(dC-dA) and poly(dA-dT), have the

ability to adopt a Z-DNA conformation[21]. The above data suggest the possibility that alternating purine-pyrimidine sequences can form a DNA configuration which is favorable for insertion events.

Comparison of Yeast, Drosophila, Chicken, Rat and Mouse Somatic Cytochrome c Genes

Cytochrome c genes from six species have now been isolated and sequenced; the CYC1[22] and CYC7[23] genes of the yeast Saccharomyces cerevisiae, the PoCYC gene of the fission yeast Schizosaccharomyces pombe[24], the DC3 and DC4 genes of Drosophila melanogaster (Limbach and Wu, accompanying paper), the CC9(CC10) alleles of chicken[13], the RC4 gene of rat[12] and the MC1 gene of mouse.

The mouse cytochrome c gene (MC1) appears to contain two intervening sequences; one in the 5'-noncoding region and one in the coding region. The rat cytochrome c gene (RC4) also contains two intervening sequences (R. Scarpulla, personal communication). These introns are located at the identical positions as the mouse introns and appear to be of comparable lengths. The chicken cytochrome c gene contains the coding region intron, but does not appear to contain the 5'-noncoding region intron. The yeast and Drosophila cytochrome c genes do not contain either intervening sequence. It therefore appears that intervening sequences are not a constant feature of the cytochrome c gene throughout evolution.

Mouse and rat diverged approximately 30 MYs ago[8]. The recency of this event is reflected in the relatively conserved nature of the two coding region intervening sequences (Fig. 4). Only 17 individual substitution events have occurred in these two introns. Using the value 7×10^{-9} substitutions/nucleotide/year as an estimate of the rate of silent site mutations, mouse and rat were calculated to have diverged 23 MYs ago, which is relatively close to the referenced value.

Since mouse and rat have diverged relatively recently, it is possible that rat also contains a functional testis-specific cytochrome c gene. Hybridization analysis of polyadenylated RNA from rat testis, however, did not detect a testis-specific cytochrome c transcript (results not shown).

Hybridization analysis of mouse polyadenylated RNA reveals the existence of four cytochrome c mRNA species. Hybridization analysis of rat polyadenylated RNA detects only three cytochrome c mRNA species[3]. Like mouse, the three rat mRNA species appear to be encoded by a single cytochrome c gene (RC4). S1 mapping of the 3' end of rat cytochrome c mRNA reveals that polyadenylation occurs at three sites. These sites are approximately 20 bp downstream from putative polyadenylation signal

sequences and are at the same position as those polyadenylation sites used in mouse. Mouse, however, appears to contain an additional site of polyadenylation at position +723. Analysis of the nucleotide sequence of the 3'-noncoding region of the rat (RC4) and mouse (MC1) cytochrome c genes reveals that the putative polyadenylation signal sequence in mouse, at position +702, ACTAAA, has been mutated in rat to ATTGAA. Since the 3'-noncoding region of mouse and rat are highly conserved, it would appear that the mutations within this putative control sequence are enough to alter its ability to function as a polyadenylation signal sequence.

ACKNOWLEDGEMENTS

We would like to thank J. Tso and Joseph Calvo for valuable criticisms. This work was supported by research grant GM 29179 from the National Institute of Health, U.S. Public Health Service.

REFERENCES

1. Margoliash, E. and Schejter, A. (1966) Adv. Protein Chem. 21, 114-286.
2. Scarpulla, R.C., Agne, K.M. and Wu, R. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 739-743.
3. Scarpulla, R.C. and Wu, R. (1983) Cell 32, 473-482.
4. Jahn, C.L., Hutchison, C.A., Phillips, S.J., Weaver, S., Haigwood, N.L., Voliva, C.F. and Edgell, M.H. (1980) Cell 21, 813-844.
5. Leuders, K., Leder, A., Leder, P. and Kuff, E. (1982) Nature 295, 426-428.
6. Ueda, S., Hakai, S., Nishida, Y., Hisajima, H. and Honjo, T. (1982) EMBO J. 1, 1539-1544.
7. Hollis, G.F., Hieter, P.A., McBride, O.W., Swan, D. and Leder, P. (1982) Nature 296, 321-325.
8. Lemischka, I. and Sharp, P.A. (1982) Nature 300, 330-335.
9. Hennig, B. (1975) Eur. J. Biochem. 55, 167-183.
10. Rimm, D.L., Horness, D., Kucera, J. and Blattner, F.R. (1980) Gene 12, 301-309.
11. Clarke, L. and Carbon, J. (1976) Cell 9, 91-99.
12. Scarpulla, R.C., Agne, K.M. and Wu, R. (1981) J. Biol. Chem. 256, 6480-6486.
13. Limbach, K.J. and Wu, R. (1983) Nucl. Acids Res. 11, 8931-8950.
14. Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J. and Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
15. Lewin, B. (1980) Cell 22, 324-326.
16. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211-214.
17. Perler, F., Efstratiades, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1980) Cell 20, 555-566.
18. Krayev, A.S., Kramerov, D.A., Skryabin, K.G., Ryskov, A.P., Bayev, A.A. and Georgiev, G.P. (1980) Nucl. Acids Res. 8, 1201-1215.
19. Jelinek, W.R. and Schmid, C.A. (1982) Ann. Rev. Biochem. 51, 813-844.
20. Hamada, H., Petrino, M.G. and Kakunaga, T. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 6465-6469.
21. Arnott, S., Chandrasekaran, R., Birdsall, D.L., Leslie, A.G.W. and Ratliff, R.L. (1980) Nature 283, 743-745.
22. Smith, M., Leung, D.W., Gillam, S. and Astell, C.R. (1979) Cell 16, 753-761.
23. Montgomery, D.L., Leung, D.W., Smith, M., Shalt, P., Faye, G. and Hall, B.D. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 541-545.
24. Russell, P.R. and Hall, B.D. (1982) Mol. Cell Biol. 2, 106-116.