

Mouse Processed Pseudogenes (Apexp1 and Cbx3p1) for Apex Nuclease and Chromobox Homolog 3

Shuji Seki^{a*}, Yoshie Manabe^a, Takashi Nakamura^b, Ming Yao^b and Yuichi Seki^c

^aDepartment of Human Nutrition, Faculty of Contemporary Life Science, Chugokugakuen University, Okayama 701-0197, Japan,

^bDepartment of Molecular Biology, Institute of Cellular and Molecular Biology, Okayama University Medical School, Okayama 700-8558, Japan and ^cDepartment of Physiology, Okayama University Medical School, Okayama 700-8558, Japan

Mouse processed pseudogenes for Apex nuclease and chromobox homolog 3 were isolated from a mouse genomic library. Inserts of the three isolated clones (clones 2, 6 and 16) were sequenced and analyzed. Clones 2 and 6 were overlapping clones, and the inserts covered a 14.9 kilobase (kb) genomic fragment. The fragment contains a processed Apex pseudogene (designated as mouse Apexp1) and a processed Cbx3 pseudogene (designated as mouse Cbx3p1). These pseudogenes are located in inactive LINE (long interspersed repetitive element) sequences. Calculation of nucleotide substitution rates suggests that the non-functional mouse Apexp1 arose 24 million years ago, and that the non-functional mouse Cbx3p1 arose even earlier, 56 million years ago. The insert sequence of the clone 16 contains a processed Apex pseudogene (designated as mouse Apexp2) differing from the Apexp1.

Key Words: Pseudogenes, Apexp1, Cbx3p1

Introduction

A pseudogene is a DNA segment with high homology with a functional gene but containing nucleotide changes such as frameshift and nonsense mutations that prevent its expression [1]. The various pseudogene sequences reported fall into two general categories [2]. In the first are those that retain the intervening sequences found in their functional counterparts (1, 2). In the second and more abundant category are those lacking the intervening sequences found in their functional counterparts. Such pseudogenes have been termed processed pseudogenes.

APEX nuclease (Apex gene product) is a mammalian multifunctional DNA repair enzyme involved in the repair of apurinic/apyrimidinic sites and single-strand DNA breaks with 3' termini blocked by nucleotide fragments and also in transcriptional regulation via redox activation of the AP-1 transcription factors [3-5]. Chromobox homolog 3 (Cbx3 gene product; Drosophila HP1 gamma homolog) interacts with lamin B receptor (an integral protein of the inner nuclear membrane), implicating in heterochromatin association with inner nuclear membrane [6]. As discussed later, chromosomal localization of Apex gene seems to be different from that of Cbx3 gene.

Processed pseudogenes have been described for numerous housekeeping genes in mammals [2]. However, so far as we know, pseudogenes for mouse Apex and Cbx3 genes have not been reported. In the present paper, DNA sequence and analysis of Apex and Cbx3 pseudogenes in a 14.9 kb mouse genomic fragment are

*Corresponding author.

Shuji Seki, MD, PhD.

Department of Human Nutrition, Faculty of Contemporary Life Science, Chugokugakuen University, 83-Banchi, Niwase, Okayama 701-0197, Japan

Tel. & FAX; + 81 86 293 0247

reported. The time since mouse Apex and Apexp1 diverged and that since Cbx3 and Cbx3p1 diverged are estimated.

Materials and methods

Materials

An adult BALB/c mouse leukocyte genomic library in EMBL-3 was obtained from Clontech Laboratories, Inc., California, U.S.A. The other materials used were obtained as described previously [7].

Cloning of mouse Apex pseudogenes

The genomic library constructed using mouse leukocyte DNA and cloning vector EMBL-3 SP6/T7 (Clontech) was screened using a [³²P]dCMP-labeled mouse Apex cDNA probe as described previously [8].

DNA sequencing and sequence analysis

Insert DNAs from the isolated bacteriophage clones were subcloned into plasmid vector pBluescript KS (-). The nucleotide sequence was determined on both strands by the dideoxy chain termination method using double-stranded templates and modified AmpliTaq polymerase (ABI PRISM Cycle Sequencing Kit from PE Applied Biosystems) and by an ABI 373S DNA sequencer (Perkin-Elmer Japan Co., Ltd., Chiba, Japan). Sequencing primers used were the M13 universal primers and specific oligonucleotide primers synthesized according to the sequences determined. Sequence analysis was performed using the Genetyx package developed by Software Development Co., Ltd., Tokyo, Japan.

Results and discussion

Several clones hybridized with the [³²P]dCMP-labeled mouse Apex cDNA probe were obtained. Among them, at least 3 clones (clones 2, 6 and 16) contained Apex pseudogenes in the inserts. Clones 2 and 6 were

overlapping clones, and the inserts covered a 14.9 kb genomic fragment (the sequence is deposited in Genome Data Bases with the accession no. AB084238). The insert sequence of the clone 16 covered a 3.2 kb genomic fragment.

Characterization of the 14.9 kb fragment (clones 2 and 6)

The 14.9 kb genomic fragment contained an Apex pseudogene and a Cbx3 pseudogene in inactive LINEs (long interspersed repetitive elements). In order to tentatively assign a base position within the 14.9 kb genomic fragment, the guanine residue at the 5' end of the fragment was designated position + 1. With this as a starting point, positive numbers were given to positions downstream. As shown diagrammatically in Fig. 1, homologous sequences to parts of mouse LINEs (Accession nos. U15647 and AE008686) are located at base positions 2364/3633 and 10413/11863. A processed pseudogene for the Apex gene is located at base positions 5672/6866. The pseudogene is tentatively symbolized as Apexp1 in the present paper. A processed pseudogene (tentatively symbolized as Cbx3p1) for the Cbx3 gene (the gene for chromobox protein homolog 3) is located at base positions 12355/14057.

Characterization of the mouse Apex pseudogene (Apexp1)

Mouse Apexp1 is a processed pseudogene comprising almost the entire sequence of mouse Apex cDNA, from its exon 1 to the polyA signal (AATAAA), with the exception of the polyA tail. Although the mouse Apexp1 sequence is 88.1% identical/1209 base pairs (bp) to that of the mouse Apex cDNA, in-frame termination codons and deletion or insertions that change the reading frame preclude the possibility that this pseudogene encodes a functional APEX nuclease. The Apexp1 is flanked by a direct repeat, TAAATAA, which might be involved in the retrotransposition of the progenitor of the Apexp1

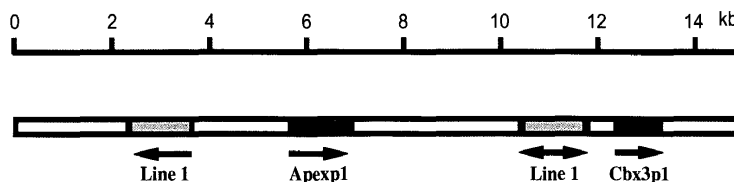


Fig. 1 Organization of the processed mouse Apex pseudogene (Apexp1) and Cbx3 pseudogene (Cbx3p1) in the 14.9 kb genomic fragment (clones 2 and 6). The scale on the top is kilobases (kb). Apexp1 and Cbx3 sequences in the 14.9 kb fragment are indicated by the filled boxes. Parts of inactive LINEs are indicated by the meshed boxes. Arrows under the diagram of the 14.9 kb fragment indicate the coding directions of the progenitor genes (Apex, Cbx3 and Line 1).

gene. The nucleotide sequence encompassing Apexp1 is related to LINEs. The LINEs are incomplete and are thought to be functionally inactive.

Processed pseudogenes have been described for numerous housekeeping genes in mammals [2]. In terms of Apex-related genes, however, rat Apexp1 is thought to be the only one reported previously at the nucleotide level in vertebrates [7]. The nucleotide sequence analysis indicated that the mouse Apexp1 is not directly related to

the rat Apexp1. So far as we know, the present paper is the first report on mouse Apex pseudogenes.

Pseudogenes are formed during the process of evolution. The extent of nucleotide sequence divergence from a functional mRNA can be regarded as a quantifiable interval in an evolutionary clock [1]. To estimate the time elapsed since the divergence of Apex and Apexp1, we aligned the mouse Apexp1 sequence with those of mouse Apex cDNA (accession no. D90374) [3] and human APEX cDNA (accession no. D90373) [9], and calculated the rate of nucleotide substitutions for each position of the codons, according to the method reported previously [1, 10] (Table 1). According to the computation from Li *et al.* [1], which assumes different substitution rates for inactive pseudogenes and functional genes, it can be estimated that the mouse Apexp1 diverged from the mouse Apex gene and non-functionalized 24 million years ago (Fig. 2A). The result suggests that mouse Apexp1 is a late pseudogene that must be limited to a few species of rodent.

Characterization of the mouse Cbx3 pseudogene 1 (Cbx3p1)

The Cbx3p1 is a processed pseudogene comprising almost the entire sequence of mouse Cbx3 cDNA (accession no. AK002910), from its exon 1 to the polyA signal (AATAAA) with the exception of the polyA tail. Although the mouse Cbx3p1 sequence is 86.5% identical/1734 bp to that of the mouse Cbx3 cDNA, in-frame termination codons and deletion or insertions that change the reading frame preclude the possibility that this

Table 1 Evolutionary rate of mouse Apex and Apexp1 genes^a

	Nucleotide position of codons ^b		
	First	Second	Third
Nucleotide differences ^b (p) between			
mApexp1-mApex	20/311	19/311	27/311
mApexp1-hAPEX	30/311	24/311	102/311
mApex-hAPEX	15/317	7/317	106/317
Nucleotide substitution per site ($-d = (3/4)\ln[1 - (4/3)p]$)			
mApexp1-mApex	0.06723	0.06373	0.09227
mApexp1-hAPEX	0.10326	0.08152	0.43126
mApex-hAPEX	0.04888	0.04888	0.04888
Rate of nucleotide substitution per site per year for functional genes ^c			
	0.24×10^{-9}	0.12×10^{-9}	2.21×10^{-9}

^aCalculated in the manner of Li *et al.* (1981). The sequences are: mApexp1, a pseudogene from the mouse; mApex and hAPEX, coding sequences from the mouse Apex and human APEX genes, respectively.

^bNumber of mismatches per number of aligned codons.

^cTaking into account divergence between the mouse and the human 100 million years (Myr) ago.

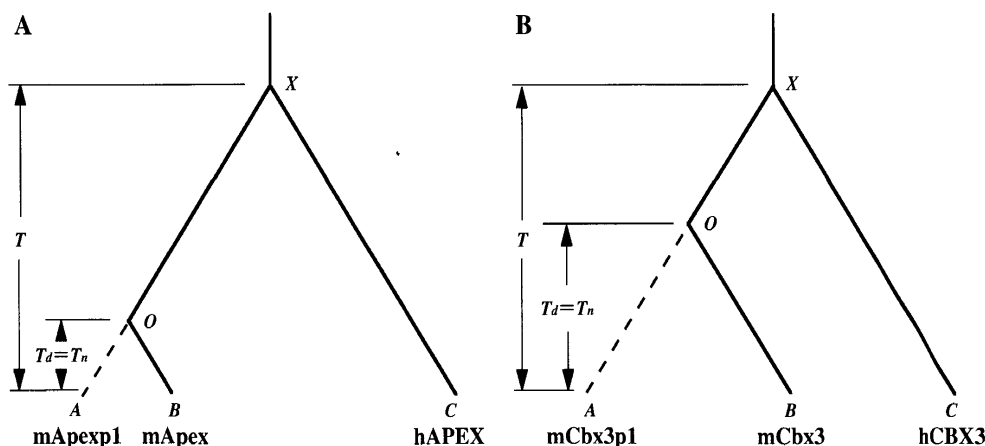


Fig. 2 Plausible phylogenetic trees for mouse Apexp1, mouse Apex and human APEX (A), and for mouse Cbx3p1, mouse Cbx3 and human CBX (B). T (100 million years) denotes the divergence time between mouse and human, and $T_d = T_n$ the time since divergence and non-functionalization of mouse Apexp1 or Cbx3p1. The $T_d = T_n$ s for Apexp1 and Cbx3p1 were estimated to be 24 and 56 million years, respectively.

Table 2 Evolutionary rate of mouse Cbx3 and Cbx3p1 genes^a

	Nucleotide position of codons ^b		
	First	Second	Third
Nucleotide differences ^c (p) between			
mCbx3p1-mCbx3	12/173	9/173	13/173
mCbx3p1-hCBX3	9/173	9/173	38/173
mCbx3-hCBX3	2/184	0/184	52/184
Nucleotide substitution per site ($-d = (3/4)\ln[1 - (4/3)p]$)			
mCbx3p1-mCbx3	0.07279	0.0539	0.07918
mCbx3p1-hCBX3	0.05391	0.0539	0.25991
mCbx3-hCBX3	0.01095	0	0.35468
Rate of nucleotide substitution per site per year for functional genes ^c			
	0.055×10^{-9}	0	1.77×10^{-9}

^aCalculated in the manner of Li *et al.* (1981). The sequences are: mCbx3p1, a pseudogene from the mouse; mCbx3 and hCBX3, coding sequences from the mouse Cbx3 and human CBX3 genes, respectively.

^bNumber of mismatches per number of aligned codons.

^cTaking into account divergence between the mouse and the human 100 million years (Myr) ago.

pseudogene encodes a functional Cbx3 protein. The Cbxp1 is flanked by a direct repeat, AAA(G/T)AAAA, which might be involved in the retrotransposition of the progenitor of the Cbx3p1 gene. The nucleotide sequence encompassing Cbx3p1 is related to LINEs. So far as we know, only a pseudogene, which is different from the Cbxp1, for the Cbx3 gene was reported previously in Genome Data Bases (accession no. AF133300).

Chromosomal localization of the 14.9 kb fragment is not known yet. The locus of human APEX gene is mapped to chromosome 14q11.2-q12 (11), and that of human CBX3 gene to chromosome 7p15.2 (accession no. NM_016587). The mouse gene locus for Apex is mapped to chromosome 14C2-D1 [8], but that for Cbx3 has not been reported yet.

To estimate the time elapsed since the divergence of Cbx3 and Cbx3p1, we aligned the mouse Cbx3p1 sequence with those of mouse Cbx3 cDNA (accession no. AK002910) and human CBX3 cDNA (accession no. NM_016587) [6] and calculated the rate of nucleotide substitutions for each position of the codons, as described above (Table 2). According to the computation from Li *et al.* [1], it can be estimated that the mouse Cbx3p1 diverged from the mouse Cbx3 (or its progenitor) gene and non-functionalized 56 million years ago (Fig. 2B).

Characterization of the 3.2 kb fragment (clone 16)

The 3.2 kb fragment contained another processed Apex pseudogene (tentatively symbolized as Apexp2) comprising a part of mouse Apex cDNA, from its exon 1 to a part of its exon 5. The Apexp2 sequence is 79.6 % identical/683 bp to that of the mouse Apex cDNA, and contains in-frame termination codons and deletion or insertions that change the reading frame. The nucleotide sequence encompassing Apexp2 is related to LINEs.

Consideration of evolution of the processed pseudogenes, Apexp1 and Cbx3p1

Considering that the mouse Cbx3p1 had arisen from quite different generation from Apexp1 and that the mouse active Cbx3 gene is thought to locate on a different chromosome from the active Apex gene, as described above, these pseudogenes (Cbx3p1 and Apexp1) are thought to join without an inevitable consequence (or accidentally) on the genomic region.

The generally accepted explanation for the formation of processed pseudogenes is that cDNA copies of the corresponding mRNA were integrated into the genome [1]. The cDNA could have been made *in situ* from the corresponding mRNA either by reverse transcriptase or by DNA polymerase γ [12]. Vanin [2] indicated that all the processed pseudogenes so far analyzed arose after mammalian radiation (approximately 100 million years ago). The present results are consistent with these ideas.

Acknowledgement. This work was supported in part by a Grant-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Science, Sports and Culture of Japan (Repair, Recombination and Mutagenesis, 08280101) (to S.S.) and in part by a Grant of the Promotion and Mutual Aid Corporation for Private School of Japan (to S.S.).

References

- Li WH, Gojobori T and Nei M: Pseudogenes as a paradigm of neutral evolution. *Nature* (1981) **292**, 237-239.
- Vanin EF: Processed pseudogenes: characteristics and evolution. *Annu. Rev. Genet.* (1985) **19**, 253-272.
- Seki S, Akiyama K, Watanabe S, Hatsushika M, Ikeda S and Tsutsui K: cDNA and deduced amino acid sequence of a mouse DNA repair enzyme (APEX nuclease) with significant homology to *Escherichia coli* exonuclease III. *J. Biol. Chem.* (1991) **266**, 20797-20802.
- Xanthoudakis S, Miao G, Wang F, Pan Y-CE and Curran T: Redox activation of Fos-Jun DNA binding activity is mediated by a DNA repair enzyme. *EMBO J.* (1992) **11**, 3323-3335.
- Friedberg EC, Walker GC and Siede W: DNA repair and mutagenesis. ASM Press, Washington, D.C. (1995).
- Ye Q and Worman HJ: Interaction between an integral protein of the nuclear envelope inner membrane and human chromodomain proteins

- homologous to *Drosophila* HP1. *J. Biol. Chem.* (1996) **271**, 14653-14656.
7. Yao M, Akiyama K, Tan Y, Sarker AH, Ikeda S, Alam SS, Tsutsui K, Yoshida MC and Seki S: Genomic structure of the rat major AP endonuclease gene (Apex) with an adjacent putative O-sialoglycoprotease gene (Prsmg1/Gcpl1) and a processed Apex pseudogene (Apexp1). *Acta Med. Okayama* (1999) **53**, 245-252.
 8. Akiyama K, Nagao K, Oshida T, Tsutsui K, Yoshida MC and Seki S: Cloning, sequence analysis, and chromosomal assignment of the mouse Apex gene. *Genomics* (1995) **26**, 63-69.
 9. Seki S, Hatsushika M, Watanabe S, Akiyama K, Nagao K and Tsutsui K: cDNA cloning, sequencing, expression and possible domain structure of human APEX nuclease homologous to *Escherichia coli* exonuclease III. *Biochim. Biophys. Acta* (1992) **1131**, 287-299.
 10. Lazard M, Agou F, Cavarelli J, Latreille MT, Moras D and Mirande M: Genomic organization of the rat aspartyl-tRNA synthetase gene family: A single active gene and several retropseudogenes. *Gene* (1996) **180**, 197-205.
 11. Akiyama K, Seki S, Oshida T and Yoshida MC: Structure, promoter analysis and chromosomal assignment of the human APEX gene. *Biochim. Biophys. Acta* (1994) **1219**, 15-25.
 12. Weissbach A: Eukaryotic DNA polymerases. *Ann. Rev. Biochem.* (1977) **46**, 25-47.
-

Accepted March 29, 2002.