An intronless gene encoding a poly(A) polymerase is specifically expressed in testis[1]

Young Jae Lee[a], Younghoon Lee[b], Jae Hoon Chung[a,*]

[a] *Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Taejon 305-701, South Korea*
[b] *Department of Chemistry, Korea Advanced Institute of Science and Technology, Taejon 305-701, South Korea*

Received 18 September 2000; accepted 30 November 2000

First published online 13 December 2000

Edited by Takashi Gojobori

Abstract Previous work demonstrated that a single pre-mRNA could generate multiple forms of mammalian poly(A) polymerase mRNAs by alternative splicing or alternative polyadenylation. A cDNA encoding a testis-specific poly(A) polymerase was isolated in this study. The transcription level of *Papt* in testis of a 2 weeks old mouse was much lower than that of the general poly(A) polymerase gene, *Pap*. However, the transcription ratio of *Papt* to *Pap* was reversed in testis of a 4 weeks old mouse. Transient expression analysis showed that GFP-Papt fusion protein is present both in the nucleus and cytoplasm of HeLa cells. These results suggest that Papt is involved in polyadenylation of transcripts expressed during spermatogenesis.

Key words: Intronless gene; Polyadenylation; Poly(A) polymerase; Spermatogenesis; Testis

1. Introduction

Poly(A) polymerase (PAP) is a ubiquitous component of the mRNA 3′-end processing machinery and polyadenylates specifically the cleaved pre-mRNAs. Six mouse *Pap* cDNAs, *PAP I–PAP VI*, have been identified [1], all of which were produced from a common *Pap* pre-mRNA by alternative splicing or alternative polyadenylation. Additionally, an intronless pseudogene was isolated from a mouse genomic DNA library. It was suggested that this pseudogene arose by retrotransposition of a putative *PAP IV* mRNA [1]. PAP II is possibly the major form of the enzyme [1]. Characterization of these cDNAs revealed that PAP contains several functional domains [2,3]: a catalytic core, an RNA binding region, two nuclear localization signals (NLS1 and NLS2), and a long serine/threonine (S/T)-rich domain. The catalytic core is located in the N-terminal half of the protein and is essential for polyadenylation. Conserved aspartate residues in this region are required for PAP's catalytic activity. Two NLSs mediate the nuclear localization. The RNA binding region which overlaps with NLS1 binds to the RNA substrate. The S/T-rich domain is dispensable both for the specific and non-specific polyadenylation in vitro [4,5]. It seems that the S/T-rich domain may play a regulatory role, because the S/T-rich domain interacts with other proteins such as U1 small nuclear ribonucleoprotein particle A protein (U1A) [6,7].

It has been shown that specific sequences at the 3′-untranslated region (3′-UTR) or the poly(A) tail lengths of many genes affect mRNA stability or translational efficiency (reviewed in [8–10]). The mRNA 3′-end processing reaction in testis exemplifies the role of 3′-sequences of mRNA in the post-transcriptional regulation (reviewed in [11]). Some genes have different sequence specificity for polyadenylation site selection in testis compared with other tissue counterparts. Furthermore, The poly(A) tail lengths of some genes vary during spermatogenesis (reviewed in [12,13]). These results suggest that there exists a tissue-specific mRNA 3′-end processing apparatus in testis. Recently, an isoform of 64 kDa subunit of CstF, which is expressed abundantly in mouse male germ cells was identified [14]. It was suggested that the testis-specific isoform of the mRNA 3′-end processing machinery controlled the gene expression in testis.

2. Materials and methods

2.1. RT-PCR analysis

Total RNAs were prepared from mouse tissues using RNeasy mini kit (Qiagen) as recommended by the manufacturer's instruction. Reverse transcription was performed with SuperScript II Reverse Transcriptase (Gibco BRL) using random hexamer. The primers PAP-f (5′-GAAGAGGAACTGCAGCGCAGGAT-3′, nt 305–327) and PAP-r (5′-ACTAGCATAGCCCAGGAAACACC-3′, complementary to nt 878–900) were used for PCR amplification of *Pap* and *Papt*. The primers PAP-P1 (5′-GCTGATCGAGACCCTCCAGCCCT-3′, nt 265–287) and PAP-r were used for specific amplification of *Papt*. PCR amplifications were carried out for 33 cycles in 100 μl of 10 mM Tris–HCl (pH 8.3), 40 mM KCl, 1.5 mM $MgCl_2$, 1 mM DTT, 0.5 μg/ml acetylated BSA, and 200 μM of each dNTP with 50 pmol of each primer and 2.5 units of *Taq* DNA polymerase at 94°C for 1 min, at 55°C for 1 min, and at 72°C for 1 min. The PCR products were analyzed on a 2% agarose gel.

2.2. Cloning of Papt cDNA and genomic DNA clones

A mouse BALB/c testis cDNA library (Clontech) was screened with the random primed *Papt* cDNA (nt 142–684) as the probe. Two positive cDNA clones were isolated and subcloned into the *Eco*RI site of the pGEM-3Zf(+) vector (Promega). These clones were named as pGPapt-1 and pGPapt-2. Two positive genomic DNA clones (λIII-1 and λV-1) were isolated from a mouse 129SVJ genomic DNA library (Stratagene). The DNA inserts were sequenced using a Taq-Track Sequencing System (Promega) and a Bigdye terminator cycle sequencing kit (Perkin-Elmer).

*Corresponding author. Fax: (82)-42-869 2610.
E-mail: jhchung@mail.kaist.ac.kr

[1] The nucleotide sequence reported in this paper has been submitted to GenBank (accession number AF039957).

Abbreviations: PAP, poly(A) polymerase; Papt, testis-specific poly(A) polymerase

0014-5793/00/$20.00
PII: S0014-5793(00)02367-X

2.3. Northern blot and Southern blot analyses

30 μg of total RNAs were electrophoresed on a 1% agarose gel and transferred to a Nytran-Plus membrane (Schleicher and Schuell). 15 μg of mouse genomic DNA were digested with *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I, or *Xba*I electrophoresed on a 0.7% agarose gel, and transferred to a Hybond N^+ membrane (Amersham). The RT-PCR products amplified using the primers PAP-P1 and PAP-r were blotted to a Hybond N^+ membrane. Hybridization was carried out with [α-^{32}P]-dCTP random primed *Papt* cDNA (nt 142–684) whose nucleotide sequence shares 85.6% homology with that of the general *Pap* cDNA in 0.5 M sodium phosphate (pH 7.2), 7% SDS, 1 mM EDTA, and 100 μg/ml denatured salmon sperm DNA at 65°C. Membrane was washed twice in 40 mM sodium phosphate (pH 7.2) and 1% SDS for 15 min at 65°C.

2.4. Transfection and fluorescence microscopic study

To investigate the subcellular localization of Papt, GFP-Papt fusion plasmids were constructed as follows: N-terminal region of *Papt* was

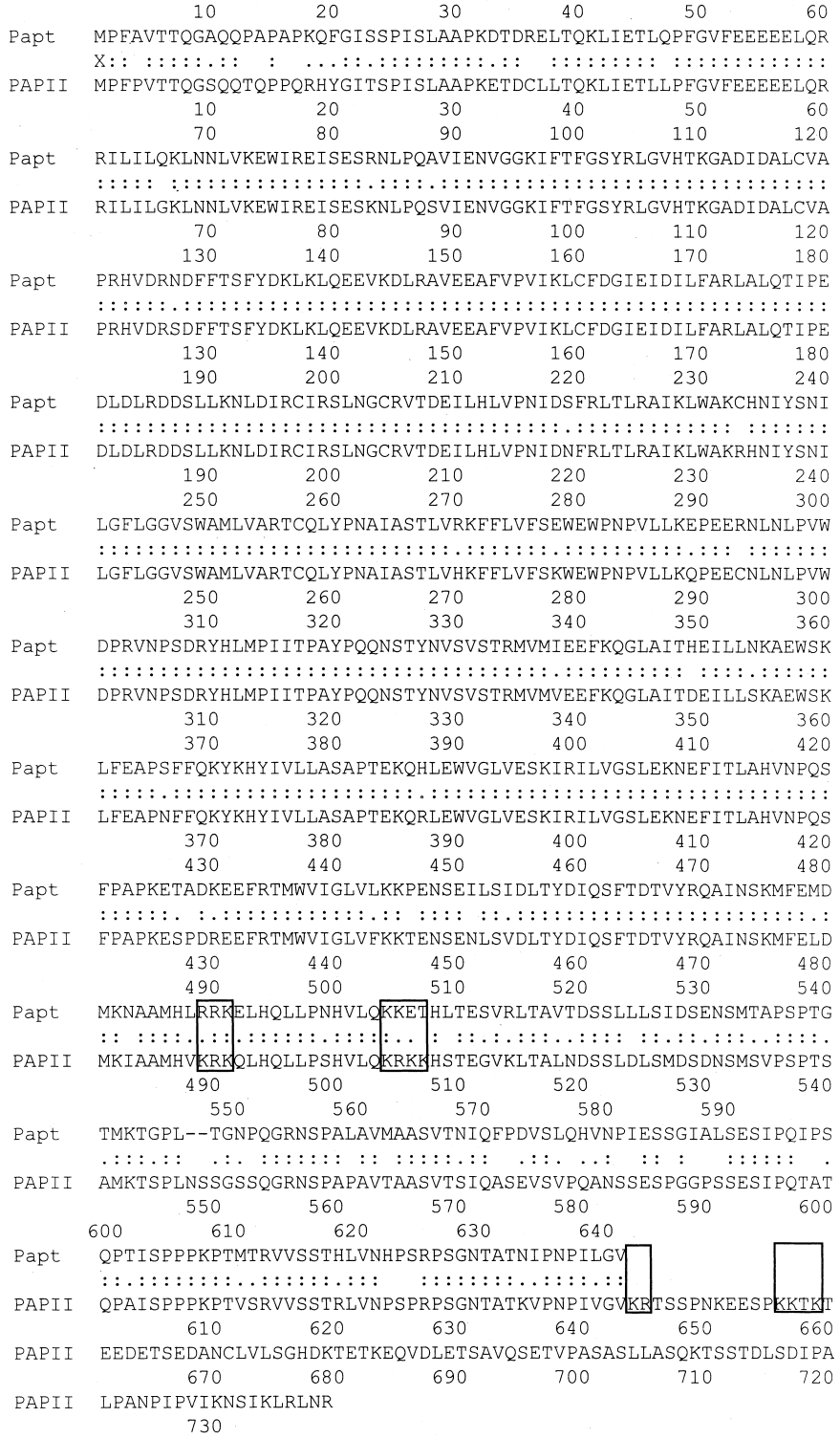

Fig. 1. Comparison of the amino acid sequences of Papt and mouse PAP II. NLSs are boxed. Each NLS contains two clusters of basic residues. Dots indicate the positions of conserved residues shared by two proteins.

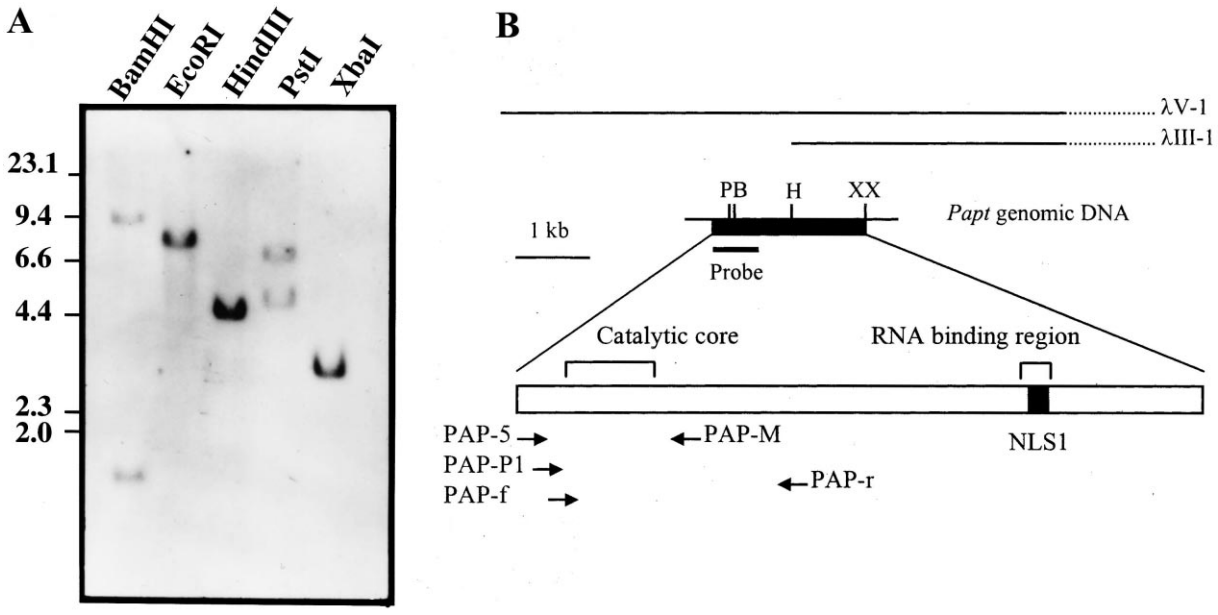

Fig. 2. Genomic structure of *Papt*. A: Southern blot analysis. 15 μg of mouse genomic DNA were digested with *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I, or *Xba*I, electrophoresed, transferred to a nylon membrane, and hybridized with the α-^{32}P-labeled *Papt* cDNA probe. The size markers are indicated in kb. B: Restriction map of *Papt*. The protein coding region is indicated by a solid box. Two positive λ clones (λIII-1 and λV-1) are represented. The PCR primers used in this study are indicated. A catalytic core, an RNA binding region, and a nuclear localization signal (NLS1) are indicated. B, *Bam*HI; H, *Hin*dIII; P, *Pst*I; X, *Xba*I.

obtained by RT-PCR from testis total RNA using the PCR primers PAP-5 (5′-GATGCCATTTGCGGTGAC-3′, nt 142–159) and PAP-M (5′-TCTTCTGGAATAGTCTGC-3′, complementary to nt 667–684). The PCR fragment was cloned into the *Eco*RV site of the pBluescriptIIKS(+) vector (Stratagene), yielding pBSPaptN. The insert fragment of the pGPapt-1 digested with *Eco*RV and *Sal*I was ligated into the pBSPaptN digested with the same restriction enzymes. The *Eco*RI- and *Sal*I-digested whole cDNA fragment was subcloned into the *Eco*RI and *Sal*I sites of the pEGFP-C1 vector (Clontech). This plasmid, pEGFP-Papt, was digested with *Kpn*I and religated to generate pEGFP-PaptΔC, which lacks NLS.

HeLa cells grown at 37°C with 5% CO_2 in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal bovine serum (Gibco BRL) and 10 mg/ml gentamicin were harvested and plated on glass coverslips placed in 30 mm diameter plates. After 12 h incubation, cells were transfected with 2 μg of the plasmids using FuGENE 6 Transfection Reagent (Roche Molecular Biochemicals). The cells were further cultured for 28 h and washed with PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, and 1.4 mM KH_2PO_4, pH 7.3). After fixing for 15 min with 4% paraformaldehyde in PBS, the cells were washed with PBS. 7-aminoactinomycin D (Molecular Probes) was used to stain cell nuclei. Fluorescence images were taken using a confocal microscope (Zeiss).

3. Results

3.1. Isolation and characterization of testis-specific poly(A) polymerase gene

The cDNAs encoding poly(A) polymerases were isolated from mouse to investigate the specific role of PAP in the mRNA 3′-end processing reaction in testis. Testis total RNA was reverse transcribed and amplified using the PCR primers PAP-f and PAP-r, which were designed to contain the

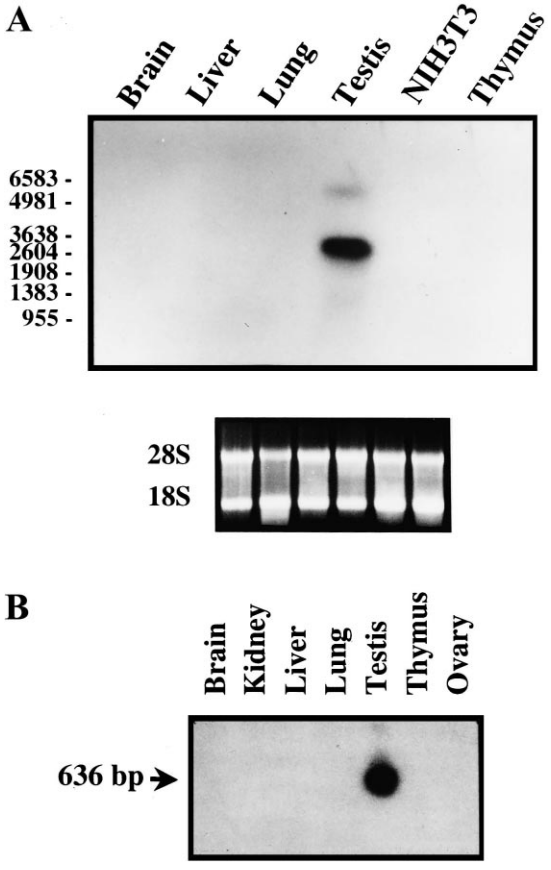

Fig. 3. Expression analyses of *Papt*. A: Northern blot analysis of *Papt*. Northern membrane containing 30 μg of total RNA from each tissue of mouse or from NIH 3T3 cells was hybridized to a *Papt* cDNA probe. The bottom figure shows the ethidium bromide-stained gel before the transfer to a membrane. B: RT-PCR analysis of the tissue-specific expression of *Papt*. The primers PAP-P1 and PAP-r were used for RT-PCR analysis that specifically amplified the *Papt* transcript. A *Papt* cDNA probe was hybridized to the Southern membrane of these RT-PCR products.

conserved sequences between the bovine *Pap* gene and the mouse *Pap* gene. DNA sequence determination of RT-PCR products revealed that two forms of poly(A) polymerase cDNAs were isolated. One group was identical to mouse *PAP III*, *V*, and *VI* cDNAs, which had been previously reported to be derived from the same gene [1]. The other was very similar to the *PAP IV* pseudogene [1]. It was named as *Papt* because of its specific expression in testis (see below). The *Papt* cDNA was isolated from a mouse testis cDNA library using the RT-PCR product of *Papt* as the probe. The coding sequence was predicted, based on the mouse and bovine *Pap* sequences [1,15]. The difference between *Papt* and *PAP IV* pseudogene in open reading frame was a single base pair deletion in *PAP IV* pseudogene at nucleotide position 144. *Papt* potentially encodes a predicted protein of 641 amino acids (Fig. 1). In the open reading frame, *Papt* shares 85.7% identity in the deduced amino acid sequence with mouse *PAP II* (Fig. 1). The predicted protein contains a catalytic core, an RNA binding region, and a nuclear localization signal (NLS1) as PAP II does. The open reading frame, however, is terminated precisely before the second nuclear localization signal (NLS2) resulting in the truncation of a part of the S/T-rich domain as well as NLS2 (Fig. 1). The overall structure of Papt rather resembles the PAP structure of *Saccharomyces cerevisiae* [16] or *Caenorhabditis elegans* [17].

3.2. *Papt is encoded by an intronless gene*

The copy number of the *Papt* gene in the mouse genome was determined by Southern blot analysis (Fig. 2A). A single band was detected in each of the *Eco*RI-, *Hin*dIII-, or *Xba*I-digested DNA lanes. Double bands were detected in each of *Bam*HI- or *Pst*I-digested DNA lanes. Because *Bam*HI and *Pst*I restriction sites are in the probe DNA fragment, the *Papt* gene is present as a single copy in the mouse genome. Genomic DNAs of mouse strains C57BL/6J, BALB/c, and DBA were amplified using *Papt*-specific primer (PAP-P1) and common primer (PAP-r) to verify the possible presence of *Pap IV* pseudogene in mouse genome. Nucleotide sequence determination of the PCR products revealed that *PAP IV* pseudogene did not exist in the tested mouse strains (data not shown). The genomic structure of the *Papt* gene was further investigated by genomic DNA cloning and its sequence determination (Fig. 2B). The most remarkable feature of the

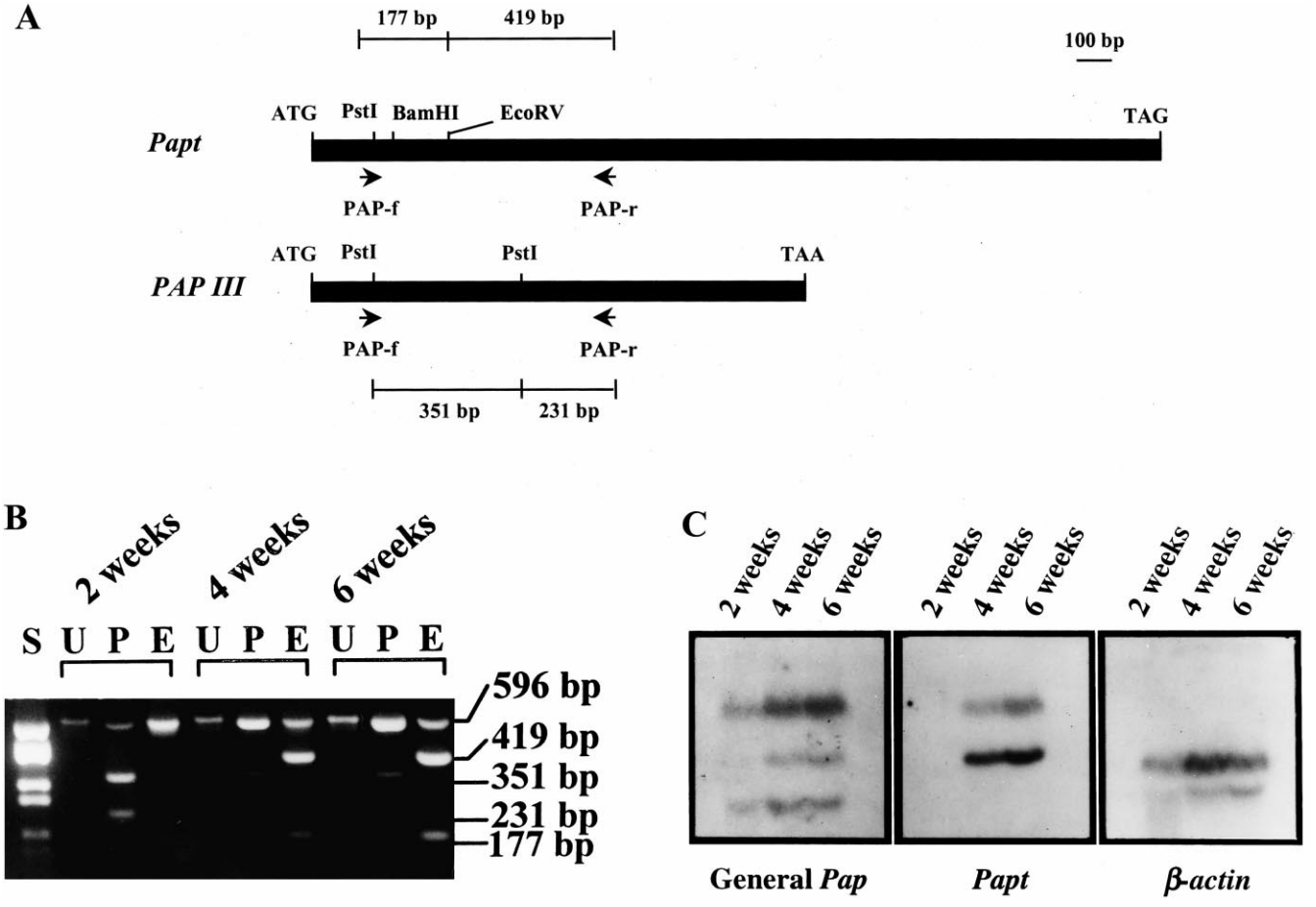

Fig. 4. Developmental analysis of *Papt* in testis. A: Restriction maps of the open reading frame of mouse *Papt* and *PAP III*. Arrows indicate the primers, PAP-f and PAP-r, used for PCR amplification of both *Papt* and general *Paps* (*PAP I–PAP VI*). *Eco*RV and *Pst*I are used for distinction of gene-specific PCR products. B: RT-PCR analysis of the expression of *Papt* in testis. Total RNAs of testis were prepared from 2, 4, and 6 weeks old mice. RT-PCR products amplified using PAP-f and PAP-r primers were digested with restriction enzymes and electrophoresed. S, pGEM-3zf(−)/*Hae*III size marker; U, undigested PCR product; P, *Pst*I-digested PCR product; E, *Eco*RV-digested PCR product. Sizes of the DNA fragments are indicated. C: Northern blot analyses of the expression of general *Pap* and *Papt* in testis. 30 μg of total RNAs prepared from 2, 4, and 6 weeks old mice were electrophoresed and transferred to a Nytran-Plus membrane. The Northern membrane was hybridized to general *Pap*, *Papt*, or β-actin cDNA probes.

Papt gene structure is that it contains no intron at all. In contrast, the general *Pap* gene that codes for *PAP I–PAP IV* has 22 introns [1]. These results suggest that Papt is encoded by a single copy intronless gene plausibly derived from the processed *PAP IV* transcript [1].

3.3. The Papt gene is differentially expressed in tissue-specific and developmental stage-specific manners

Expression pattern of the *Papt* gene in various mouse tissues was investigated by Northern blot analysis (Fig. 3A). The *Papt* cDNA probe detected *Papt* transcripts only in testis (Fig. 3A). RT-PCR analysis using *Papt*-specific primer (PAP-P1) and common primer (PAP-r) unambiguously confirmed that *Papt* was not expressed in brain, kidney, liver, lung, thymus, ovary, heat, colon, spleen, and muscle (Fig. 3B and unpublished data). These results demonstrated that the *Papt* gene is exclusively expressed in testis. The size of predominant *Papt* message was approximately 3.0 kb. In addition to the 3.0 kb transcript, approximately 5.5 kb of barely detectable message was observed in testis (Fig. 3A).

The developmental expression pattern of the *Papt* gene in testis was investigated by RT-PCR and Northern blot analyses (Fig. 4). Both general *Pap* mRNA and *Papt* mRNA were simultaneously amplified using the primers PAP-f and PAP-r. The differential endonucleolytic cleavage of the amplified products using *Pst*I or *Eco*RV could discriminate the template RNAs from which the products were reverse transcribed. The RT-PCR product of general *Pap* would be cleaved into 351 bp and 231 bp bands by *Pst*I digestion, while the RT-PCR product of *Papt* would be cleaved into 419 bp and 177 bp bands by *Eco*RV digestion (Fig. 4A). General *Pap* was predominantly expressed in testis of 2 weeks old mouse while *Papt* was detected at a low level (Fig. 4B and C). However, *Papt* was most abundantly expressed in testes of 4 and 6 weeks old mice (Fig. 4B and C). The transcription ratio of *Pap* and *Papt* was reversed in testis of 4 weeks old mouse. The increase of expression level of the *Papt* gene chronologically correlates with the progression of the first meiotic division at day 17–19 after birth [12,18]. Papt seems to be the main PAP as well as the functional PAP in spermatids which develop at day 22–24 after birth [18].

3.4. Papt was detected in both the nucleus and the cytoplasm

Transient expression experiments were carried out to investigate the subcellular localization of Papt. The expression plasmids encoding Papt derivatives were constructed. These expression plasmids encoded N-terminal GFP fusion proteins. The intact Papt (pEGFP-Papt) expression plasmid and truncated Papt (pEGFP-PaptΔC) expression plasmid were transfected into HeLa cells. Fluorescence imaging of GFP in the transfected cells with the expression plasmids was used to determine possible subcellular localization of the fusion proteins. The GFP-Papt fusion protein (pEGFP-Papt) was dispersively localized throughout the cell (Fig. 5). However, the truncated GFP-Papt fusion protein (pEGFP-PaptΔC) which lacks NLS showed its exclusive cytoplasmic localization (Fig. 5). The result indicates that NLS of Papt is required for the transport of Papt into nucleus.

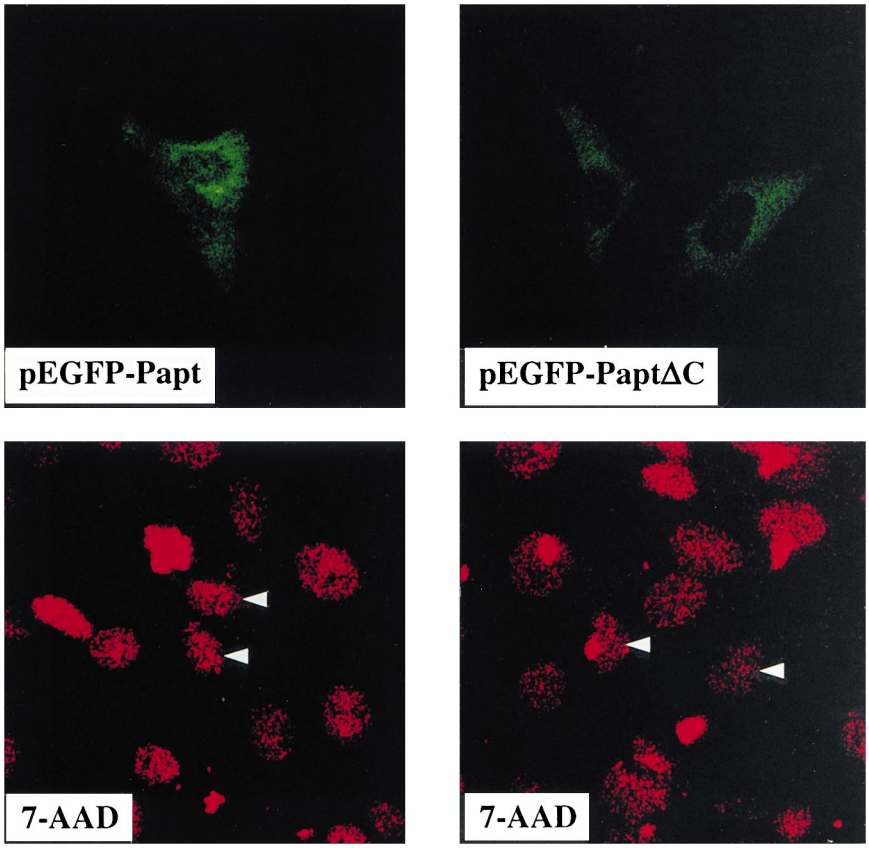

Fig. 5. Subcellular localization of Papt in HeLa cells. Fluorescence images of the HeLa cells transfected with the full-length Papt expression plasmid (pEGFP-Papt) and truncated Papt expression plasmid (pEGFP-PaptΔC). Nuclei were stained by 7-aminoactinomycin D (7-AAD) (Molecular Probes). Arrows indicated nuclei of transfected cells.

4. Discussion

Papt, an intronless gene encoding a poly(A) polymerase was isolated and characterized in this study. Our results suggest that the *Papt* gene is specifically expressed in mouse testis. Although several forms of cDNAs for poly(A) polymerases have been isolated in mammals, they were produced by alternative splicing or alternative polyadenylation of a pre-mRNA from a single gene [1]. Therefore, the *Papt* gene is the first example of an additional *Pap* gene in mammals. It was thought that *Papt* is derived from retrotransposition of a putative *PAP IV* mRNA [1]. Although most retrotransposons are non-functional, some encode functional proteins. For example, *Pgk-2*, *Pdha-2*, *Zfa*, and *G6pd-2* which are expressed in spermatogenic cells replacing their X-chromosomal isotype genes were reported to be functional retrotransposons [19–22]. Also, several autosomal genes are replaced or complemented by isotypic genes during spermatogenesis [23–25].

The transcription level of the *Papt* gene in testis of 2 weeks old mouse is very low. However, the transcription level in testis of 4 or 6 weeks old mouse increases significantly. Since mouse spermatids emerge at day 20 after birth [12], the increase of the *Papt* transcription in testis chronologically coincides with the emergence of spermatids. A previous study revealed that several components of the RNA polymerase II transcription machinery, such as RNA polymerase II, TATA binding protein, and TFIIB, are highly accumulated in round spermatids [26]. It was suggested that these components synthesize all of the mRNAs necessary for spermiogenesis in a relatively short period of burst synthesis. The developmental expression pattern of the components in testis is similar to that of the *Papt* gene. Therefore, Papt may polyadenylate the mRNAs synthesized and accumulated in round spermatids.

Papt and *PAPs* of *S. cerevisiae* [16] and *C. elegans* [17] contain only a single NLS, designated as NLS1. NLS1 is known to be necessary for nuclear localization and for RNA substrate binding. The C-terminal deletion analysis of bovine PAP showed that NLS2 is dispensable for the catalytic activity and for RNA substrate binding activity of PAP [2,3]. However, both NLS1 and NLS2 are required for efficient nuclear localization of bovine PAP [3]. The N-terminal region including NLS1 (amino acids 1–507) of Papt shares 91.3% amino acid identity with that of mouse PAP II. On the contrary, the C-terminal region (amino acids 508–641) of Papt is less conserved with amino acid identity of 65.4%. It seems that the sequence at the C-terminal region of Papt diverged during or after the retrotransposition. Two molecules of U1A bind specifically to two loops within the 3′-UTR of the *U1A* pre-mRNA. The bound U1As interact with the C-terminal 20 amino acids of PAP, resulting in the inhibition of the polyadenylation and the coupling of splicing with polyadenylation [6,7]. Consequently, Papt could have additionally acquired special features such as alternation of interactions with regulatory proteins.

Acknowledgements: We thank Dr. John P. Richardson (Indiana University, Bloomington) and Dr. Gary B. Braedt (Louisiana State University, New Orleans) for reading the manuscript, and Jin Woo Kim (KAIST, Taejon) for animal cell culture and transfection experiment. This work was supported by a grant of Molecular Medicine Research Group Program (99-J03-01-01-A-05) from the Ministry of Science and Technology.

References

[1] Zhao, W. and Manley, J.L. (1996) Mol. Cell. Biol. 16, 2378–2386.
[2] Martin, G. and Keller, W. (1996) EMBO J. 15, 2593–2603.
[3] Raabe, T., Murthy, K.G.K. and Manley, J.L. (1994) Mol. Cell. Biol. 14, 2946–2957.
[4] Colgan, D.F., Murthy, K.G.K., Prives, C. and Manley, J.L. (1996) Nature 384, 282–285.
[5] Colgan, D.F., Murthy, K.G.K., Zhao, W., Prives, C. and Manley, J.L. (1998) EMBO J. 17, 1053–1062.
[6] Gunderson, S.I., Beyer, K., Martin, G., Keller, W., Boelens, W.C. and Mattaj, I.W. (1994) Cell 76, 531–541.
[7] Gunderson, S.I., Vagner, S., Polycarpou-Schwarz, M. and Mattaj, I.W. (1997) Genes Dev. 11, 761–773.
[8] Ross, J. (1996) Trends Genet. 12, 171–175.
[9] Hake, L.E. and Richter, J.D. (1997) Biochim. Biophys. Acta 1332, M31–M38.
[10] Wickens, M., Anderson, P. and Jackson, R.J. (1997) Curr. Opin. Genet. Dev. 7, 220–232.
[11] Kleene, K.C. (1993) Dev. Biol. 159, 720–731.
[12] Erickson, R.P. (1990) Trends Genet. 6, 264–269.
[13] Wolgemuth, D.J. and Watrin, F. (1991) Mamm. Genome 1, 283–288.
[14] Wallace, A.M., Dass, B., Ravnik, S.E., Tonk, V., Jenkins, N.A., Gilbert, D.J., Copeland, N.G. and MacDonald, C.C. (1999) Proc. Natl. Acad. Sci. USA 96, 6763–6768.
[15] Raabe, T., Bollum, F.J. and Manley, J.L. (1991) Nature 353, 229–234.
[16] Lingner, J., Kellermann, J. and Keller, W. (1991) Nature 354, 496–498.
[17] Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T. and Cooper, J. et al. (1994) Nature 368, 32–38.
[18] Nebel, B.R., Amarose, A.P. and Hackett, E.M. (1961) Science 134, 832–833.
[19] McCarrey, J.R., Berg, W.M., Paragioudkis, S.J., Zhang, P.L., Dilworth, D.D., Arnold, B.L. and Rossi, J.J. (1992) Dev. Biol. 154, 160–168.
[20] Takakubo, F. and Dahl, H-H.M. (1992) Exp. Cell Res. 199, 39–49.
[21] Ashworth, A., Skene, B., Swift, S. and Lovell-Badge, R. (1990) EMBO J. 9, 1529–1534.
[22] Hendriksen, P.J.M., Hoogerbrugge, J.W., Baarends, W.M., de Boer, P., Vreeburg, J.T., Vos, E.A., van der Lende, T. and Grootegoed, J.A. (1997) Genomics 41, 350–359.
[23] Welch, J.E., Schatte, E.C., O'Brien, D.A. and Eddy, E.M. (1992) Biol. Reprod. 46, 869–878.
[24] Kalab, P., Visconti, P., Leclerc, P. and Kopf, G.S. (1994) J. Biol. Chem. 269, 3810–3817.
[25] Fulcher, K.D., Welch, J.E., Klapper, D.G., O'Brien, D.A. and Eddy, E.M. (1995) Mol. Reprod. Dev. 42, 415–424.
[26] Schmidt, E.E. and Schibler, U. (1995) Development 121, 2373–2383.