[Environmental Health and Preventive Medicine 10, 103–110, March 2005]

Original Article

Multiple Variant mRNAs with Different Length Tandem Repeats of (CAYYCC)n Produced from Bovine Selenoprotein P-like Protein Gene

Kiyofumi SAIJOH[1], Tatsuya KOBAYASHI[1], Motoko NAGANO-FUJII[1], Shigeki FUJII[1] and Hironobu KATSUYAMA[2]

[1]*Department of Hygiene, Kanazawa University School of Medicine, Kanazawa, Japan*
[2]*Department of Public Health, Kawasaki Medical School, Okayama, Japan*

Abstract

In contrast to selenoprotein Ps (SelPs) from other animal species, bovine selenoprotein P-like-protein (SelPLP) was found to contain a tandem repeat of $(CAYYCC)_{11}$. During an investigation into whether SelPLP was a bovine substitute for SelP or uniquely bovine, its mRNA was found to consist of multiple variants with different length tandem repeat, namely p(0) with $(CAYYCC)_{11}$, p(-4) lacking $(CAYYCC)_4$, p(-8) lacking $(CAYYCC)_8$, and p(-9) lacking $(CAYYCC)_9$. Although they were encoded on a single gene locus, neither classical GT-AG nor minor class AT-AC donator-acceptor sequences for alternative splicing were identified. A subsequent S1 protection assay using oligonucleotides, whose sequence may occur as variants, performed against bovine poly(A)$^+$RNA identified a total of nine variants. Judging from the sequence of these variants and the branch point mapping, the consensus sequence for recognition of the donator was CACCCCCAC and of the acceptor and the branch point A nucleotide, ACCCCCAT or ACCCCCATCCCCAT. Furthermore, when the p(0) insert mRNA was expressed in COS-7 cells derived from an African green monkey kidney, cDNAs corresponding to p(-8) and p(-9) could be isolated. Therefore, the bovine SelPLP mRNAs consisted of multiple variants probably due to a novel splicing mechanism which was not bovine-specific but common to other mammals.

Key words: selenoprotein P, selenoprotein P-like protein, selenocysteine, alternative splicing

Introduction

Selenoprotein P is unique in that its 10 TGA codons incorporate selenocysteines instead of functioning as termination codons (1–3). We isolated two cDNA clones with a sequence similar to that of rat and human selenoprotein P (1, 3) from bovine brain (3). One clone, designated as p(0), had an insert characterized by having not only 12 instead of 10 TGAs but also a tandem repeat of $(CAYYCC)_{11}$. This tandem repeat and its adjacent region produced a cytosine-rich portion consisting of 57 cytosines out of a total of 84 nucleotides (68%). The amino acid sequence extracted from this portion contained 16 Hiss and 10 Pros out of 28 amino acids (Fig. 1A), making it a His-Pro rich domain. Another clone, p(-8), also had 12 TGAs but lacked $(CAYYCC)_8$ (3). Except for this absence of $(CAYYCC)_8$, the sequence of the insert of p(-8) was identical to that of the p(0) insert. Thus, p(-8) also included a cytosine-rich portion consisting of 20 cytosines out of 36 nucleotides (55.6%) corresponding to 8 Hiss and 3 Pros out of 12 amino acids (Fig. 1A). Although in human and rat selenoprotein P neither the tandem repeat nor the His-Pro rich domain was conserved (1, 3), p(-8) was much like these selenoprotein Ps (Fig. 1A). To confirm whether p(-8) corresponds to a bovine substitute for selenoprotein P, and whether p(0) differs sufficiently from bovine selenoprotein P to qualify as having a unique function, genomic cloning was performed to isolate the genomic loci of these proteins (4). While a genome corresponding to p(0) was cloned and found to consist of five exons with its cytosine-rich portion located at exon 5 (Fig. 1B), neither a gene corresponding to p(-8) nor one corresponding to bovine selenoprotein P without a cytosine-rich portion could be isolated. Therefore, it is possible that these mRNAs were the result of alternative splicing. However, the cytocine-rich portion was found to contain neither classical GT-AG (5) nor minor class AT-AC (6) donator-acceptor sequences for alternative splicing. Moreover,

Received Sep. 24, 2004/Accepted Nov. 15, 2004
Reprint requests to: Kiyofumi SAIJOH
Department of Hygiene, Kanazawa University School of Medicine, 13-1 Takara-machi, Kanazawa 920-8640, Japan
TEL: +81(76)265-2213, FAX: +81(76)243-4232
E-mail: saijohk@med.kanazawa-u.ac.jp

48 bps was very short for a spliced-out intron.

The study presented here showed that bovine selenoprotein P-like protein mRNAs were originated from a single gene locus but consisted of multiple variants with different length tandem repeat of (CAYYCC)n in addition to p(0) and p(-8). The consensus sequence required for recognition not only of the donator-acceptor but also of the branch point could be identified, suggesting that these variants were the result of alternative splicing. Furthermore, using in vivo splicing reaction in COS7 cells derived from an African green monkey kidney, this mechanism was confirmed to be not bovine specific but common to other mammals.

Materials and Methods

Genomic Southern analysis

The bovine genome was purified according to the standard procedure (7) from bovine cerebellum obtained from a local slaughterhouse. Ten μg of bovine genome was digested by restriction enzymes, electrophoresed on an agarose gel, and Southern-blotted onto a nitrocellulose filter (7). The cDNA insert of p(0) was radiolabeled using a DNA labeling system (Amersham Pharmacia Biotech) in the presence of 500 μM each of dATP, dGTP, and dTTP, 200 μM of dCTP, and 10 μCi [α-^{32}P]-dCTP (3,000 Ci/mmol, Amersham Pharmacia Biotech). Hybridization of the ^{32}P-radiolabeled cDNA probe was carried out in HB-N (7) at 42°C overnight. The filter was then washed once in 2×SSC/0.1%SDS at room temperature and twice in 0.5×SSC/0.1%SDS at 65°C, dried, and mounted on an imaging plate (BAS-III; Fuji), which was analyzed with an image analyzer BAS-2000 (Fuji).

Genomic PCR

The bovine genome encoding the cytosine-rich portion was PCR amplified by using a set consisting of an upstream primer (U2; TGAAGACGTCTGTAAAAATGT located at 5' terminal of exon 5) and a downstream primer D1 (ATCTGAGACCCTGTCTTTG located at 3' and flanking the cytosine-rich portion) (Fig. 1B).

Cloning of splicing variants

Total RNA was prepared with the guanidium-isothiocyanate/cesium chloride method (7). Poly (A)$^+$RNA was selected from the total RNA and subjected to RT-PCR amplification targeting the cytosine-rich portion. The downstream primer, D1, was utilized but the upstream primer was changed to U1 (ACAAAGATGACTTCCTCATAA) (Fig. 1B), which was hybridized to exon 4. The products were then electrophoretically separated on 2% agarose-ethidium bromide gel and the separated bands were excised out. After electroelution, the products were cloned into a TA-cloning vector, pGemT (Promega). According to the sequence analysis of the inserts of isolated clones, the clones were designated as p(0), p(-4), p(-8) and p(-9).

mRNA protection assay

Radiolabeled antisense cRNAs were synthesized from clones p(0), p(-4), p(-8) and p(-9) using the mCAP mRNA Capping Kit (Stratagene) in the presence of of 500 μM each of rATP, rCTP, and rGTP, 200 μM of rUTP, and 10 μCi [α-^{32}P]-rUTP (3,000 Ci/mmol, Amersham Pharmacia Biotech). Although the insert sizes were 438 bs for p(0), 414 bs for p(-4), 390 bs for p(-8), and 384 bs for p(-9), the corresponding sizes of the transcribed cRNA were 61 bs more, i.e. 499, 475, 451, and 445 bs, because they included the sequence between the multiple cloning site and the T7 RNA polymerase recognition site. Ten μg of poly(A)$^+$RNA was hybridized at 42°C overnight with these ^{32}P-cRNAs as probes in 50 μl of the hybridization buffer consisting of 20 mM Tris/HCl, pH 7.4, 0.5 M NaCl, 2 mM EDTA and 80% formamide. After phenol/chloroform extraction and EtOH precipitation, the reaction was digested at 30°C for 30 min with 60 ng/μl RNase A (Sigma) and 3 ng/μl RNase T1 (Ambion) in 50 μl of the digestion buffer consisting of 20 mM Tris/HCl, pH 7.4, 300 mM NaCl and 1 mM EDTA. After further extraction with phenol/chloroform and EtOH precipitation, the obtained precipitant was separated on denaturing (8.0 M urea, 89 mM Tris/borate, 2.5 mM EDTA) 6% acrylamide gel. The intensities of the undigested ^{32}P-cRNA were ascertained as described above.

S1 protection assay

Oligonucleotides (48 mer) whose sequences could theoretically be for splicing variants were synthesized (Fig. 5). The ^{32}P-labeled cDNA reverse transcribed from 1 μg of poly(A)$^+$ RNA using D1 as a primer was hybridized at 48°C overnight with 50 pmol of the oligonucleotides in 20 μl consisting of 20 mM Tris/HCl, pH 7.4, 0.5 M NaCl, and 2 mM EDTA. After phenol/chloroform extraction and EtOH precipitation, the reaction was digested with 2.5 U/μl S1 nuclease in 100 μl consisting of 0.3 M NaCl, 3 mM $ZnCl_2$ and 30 mM CH_3COONa, pH 4.5, again phenol/chloroform extracted and EtOH precipitated, and finally separated onto a denaturing polyacrylamide gel. The intensities of the undigested ^{32}P-cDNA were ascertained as described above.

Mapping of RNA branchpoints

To obtain RNA splicing intermediates, 10 μg of total RNA was incubated at 30°C for 10 min with 5 mg/ml proteinase K in 50 μl containing 0.2 M Tris/HCl, pH 7.6, 25 mM EDTA, 0.3 M NaCl, and 2% SDS, after which it was phenol/chloroform extracted and EtOH precipitated. The pellet resuspended in 2 μl sterile H_2O was mixed with or without the debranching extract prepared from bovine cerebellum according to the Grabowski method (8), and incubated at 30°C for 30 min. After phenol/chloroform extraction and EtOH precipitation, the reaction was subjected to primer extension performed in 50 μl of the reverse transcription buffer (Toyobo) in the presence of 500 μM dATP, dGTP, and dTTP, 200 μM of dCTP and 10 μCi [α-^{32}P]-dCTP, 0.1 U/μl ReverTra Ace reverse transcriptase (Toyobo), 0.1 U/μl Rnase inhibitor (Toyobo) and 10 pmol D1 as a primer at 50°C for 30 min. The cDNA was then electrophoresed and analysed as described above.

In vivo splicing reaction in COS7

The inserts of p(0) were bidirectionally subcloned into an prokaryotic/mammalian expression vector, pBK-CMV (Strata-

gene). The clones pBK-p(0) and pBK-p(0)rev could express fusion mRNAs consisting of respectively a galactosidase-p(0) insert in the normal direction or a galactosidase-p(0) insert in the reverse direction. Either pBK-p(0) or pBK-p(0)rev was transiently transfected to subconfluently cultured COS7 (Riken Cell Bank; depositor: K. Todokoro) (9) using TransFast transfection reagent (Promega). The cells were grown in a monolayer culture in DMEM (Sigma) with 10% fetal bovine serum (JRH Bioscience) at 37°C in a 5% CO_2/95% humidified-air atmosphere for 24 hr. Total RNA was isolated from either pBK-p(0) or pBK-p(0)rev transfected cells and subjected to first-strand cDNA synthesis by means of reverse transcription as described in *Mapping of RNA branchpoints* except that [α-^{32}P]-dCTP was omitted and oligodT was used as a primer. PCR amplification was then performed using a set of primers, T3 and M13-20 (Stratagene), in order to avoid amplification from COS7 mRNAs, and separated on 2% ethidium bromide-agarose gel. The products were excised from the gel, extracted, and subjected to direct sequence analysis using U1 as a primer.

Accession numbers

Bovine selenoprotein P-like protein, D25220; rat selenoprotein P, D25001, bovine selsenoprotein P-like protein gene, D88032 and D88033; p(0), AB049004; p(-4), AB049005; p(-8), AB049006; p(-9), AB049007.

Results

Number of gene loci encoding bovine selenoprotein P-like protein

To identify the number of gene loci encoding bovine selenoprotein P-like protein, genomic Southern blot hybridization was perfomed. With EcoRI and HindIII digestion, only a single band of the expected size (Fig. 1B) was hybridized to the probe (Fig. 2A). Even after digestion with enzymes producing fragments which seemed to be sufficiently short to distinguish a

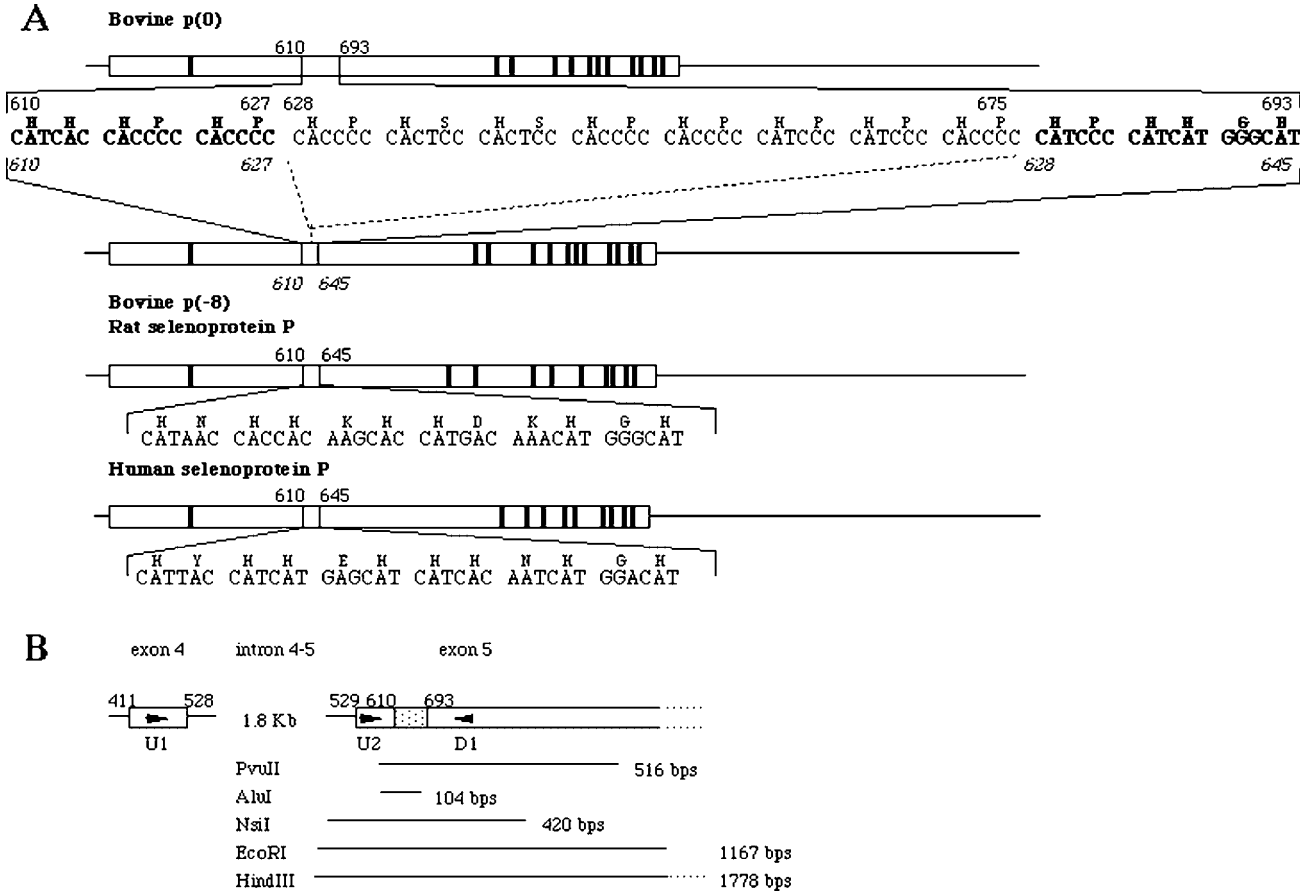

Fig. 1 Schematic diagram of cDNA and gene structure of bovine selenoprotein P-like protein and selenoprotein Ps. A: Differences in cDNA structure between bovine selenoprotein P-like protein, p(0) and p(-8), and rat and human selenoprotein P. Solid lines and the open rectangles represent the 5' and 3' non-coding regions and the protein coding region, respectively. TGAs, corresponding to selenocysteine, are indicated by thick vertical lines. The cytosine-rich portion is indicated by the hatched rectangle. The sequence of the tandem repeat reported as D25220, CACCCCCACCCCCACTCCCACCCCCACCCCCACCCCCATCCCCATCCCCACCCCCACCCCCATCCC (Saijoh et al., 1995), differs from p(0). D25220 and p(0) were obtained from bovines from a local slaughter house in CT, USA and ones in Ishikawa and other prefectures in Japan, respectively. Excluding the tandem repeat their sequences were identical. The present study was done using bovines in Japan. In any case, p(-8) could be isolated from both bovines. B: Structure of genome encoding p(0). The location of primers is indicated by arrow heads. U1 and D1 were utilized to amplify the cDNA of the cytosine rich portion and its flanking region. To avoid genomic amplification, U1 was located on exon 4. U2 and D1 were also used to amplify the genome of the cytosine-rich portion and its flanking region. Underlining indicates the size of the restriction enzyme digested fragments for encoding of the bovine selenoprotein P-like protein on a single gene. PvuII and AluI sites are located within exon 5, and NsiI sites on intron 4–5 and exon 5.

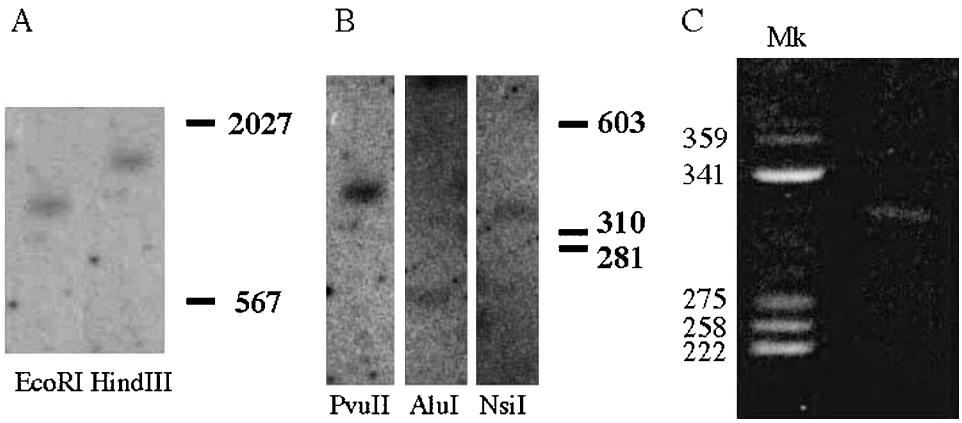

Fig. 2 Genomic Southern blot analysis and genomic PCR. Bovine genome digested with EcoRI and HindIII (A) or with PvuII, AluI and NsiI (B) was electrophoresed on respectively 1% and 2.5% agarose gel, Southern blotted, and hybridized with ^{32}P-radiolabeled cDNA of cytosine rich portion. C: Genomic PCR using the primer set, U2 and D1 (Fig. 1B).

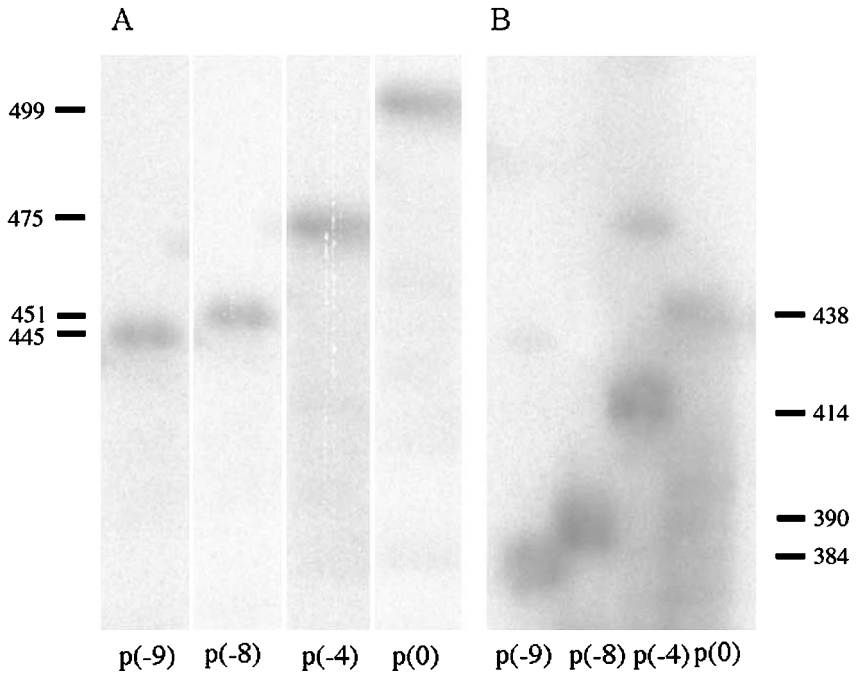

Fig. 3 RNA protection assay for p(0), p(-4), p(-8) and p(-9). A: ^{32}P-radiolabeled antisense cRNAs synthesized from clones p(0), p(-4), p(-8) and p(-9) were used as size markers. The transcribed cRNA were 61 bs more than their respective insert sizes, i.e. 499, 475, 451, and 445bs, because they included the sequence between the multiple cloning site and the T7 RNA polymerase recognition site. B: After hybridizion with poly(A)$^+$RNA from bovine cerebellum, each cRNA was digested with RNase A and T1. The size of the protected cRNA was expected to be 61bs less than that of the corresponding cRNAs after digestion as indicated above.

48-bp-difference, including NsiI, (Fig. 1B), only a single band proved to be positive for the probe (Fig. 2B). Since one NsiI site was located at intron 4–5 and another at exon 5 (Fig. 1B), the 3' terminal of intron 4–5 was assured to be connected to the 5' terminal of exon 5. Genomic PCR targeting the cytosine-rich portion also amplified only a single band of the expected size (Fig. 2C).

Cloning of other variants

The possible existence of a variant mRNA species was then examined by means of RT-PCR targeting of the cytosine-rich portion and using the primer sets, U1 and D1 (Fig. 1B). As expected, 2% agarose-ethidium bromide electrophoresis showed the products to be separated over a broad band between 400–440 bps, and all of these products were extracted and cloned. Sequence analysis for isolated clones demonstrated that there were at least four different kinds of clones. The clones inserted at 438-bp- and 390-bp-cDNAs were identical to p(0) and p(-8), while the clones with two other inserts, lacking $^{634–658}(CAYYCC)_4$ and $^{622–676}(CAYYCC)_9$ from p(0) (Fig. 1A), were designated as p(-4) and p(-9), respectively. Aside from differences in the length of the tandem repeat, the sequences of these four clones were identical. Thus, even p(-9) had a cytosine-rich portion consisting of 15 cytosines out of 30 nucleotides.

RNA protection assay for cloned variants

To exclude the possibility that the absences seen in (CAYYCC)n were the result of low fidelity in the reverse transcriptase or Taq polymerase, an RNA protection assay was performed using poly(A)$^+$RNA (Fig. 3). All of the cRNAs synthesized from p(0), p(-4), p(-8) and p(-9) were protected, indicating that there were at least four different mRNAs encoding the bovine selenoprotein P-like protein. What we had thought to be exon 5 could be divided into exons 5-1 and 5-2. Since the 3' terminal of p(-4)-exon 5-1 led to ^{658}CATCCCCAT and those of p(-8) and p(-9) to ^{682}CATCCCCAT, such sequences seemed to be required for recognition of acceptor sites. On the other hand, for identification of the donator, CACCCC seemed to be necessary because p(-4)-, p(-8)- and p(-9)-exon 5-1s terminated at CACCCC^{633}, CACCCC^{627}, and CACCCC^{621}, respectively. However, CACCCCs observed at 651 and 657 makes it conceivable that other variants exist.

S1 protection assay and identification of donator-acceptor sequence

Oligonucleotides whose sequences could function as variants were synthesized and an S1 protection assay was performed against poly(A)$^+$RNA (Fig. 4A). Aside from Δ0, Δ42, Δ81 and Δ91, respectively corresponding to p(0), p(-4), p(-8) and p(-9), Δ12, Δ41, Δ52, Δ62, and Δ71 were also protected (Fig. 4B). Except for CACCCC^{657}, CACCCCs$^{621,\ 627,\ 633,\ \text{and}\ 651}$ could be 3' terminals of exon 5-1. Δ12, Δ42, Δ52, and Δ62 led to ^{658}CAT and Δ41, Δ71, Δ81 and Δ91 to ^{676}CATs, but neither ^{664}CAT nor ^{682}CAT could be utilized as an acceptor. It is thus likely that the CACCCCCAC sequence is required for recognition of the donator and CCCCCAT or CATCCCCAT for recognition of the acceptor (Fig. 4A). It is also noted that Δ12 lacked only one (CAYYCC).

RNA branch point and 5' terminal of intron

RNA branch points were mapped by means of a primer extension using reversetranscriptase on cerebellar total RNA (Fig. 5). The primer extension terminated mainly at ^{633}C and $^{653\ \text{and}\ 671}$As. Not ^{631}C but $^{653\ \text{and}\ 671}$As were located 5 bs upstream from $^{658\ \text{and}\ 676}$Cs which supposedly constituted the 5' terminal of exon 5-2 as mentioned earlier. Schematic diagram of lariat formation and following splicing to produce p(-4), p(-8) and p(-9) was illustrated in Fig. 4C.

In vivo deletion reaction of (CAYYCC)n in COS7 cells

To elucidate whether this deletion mechanism was bovine

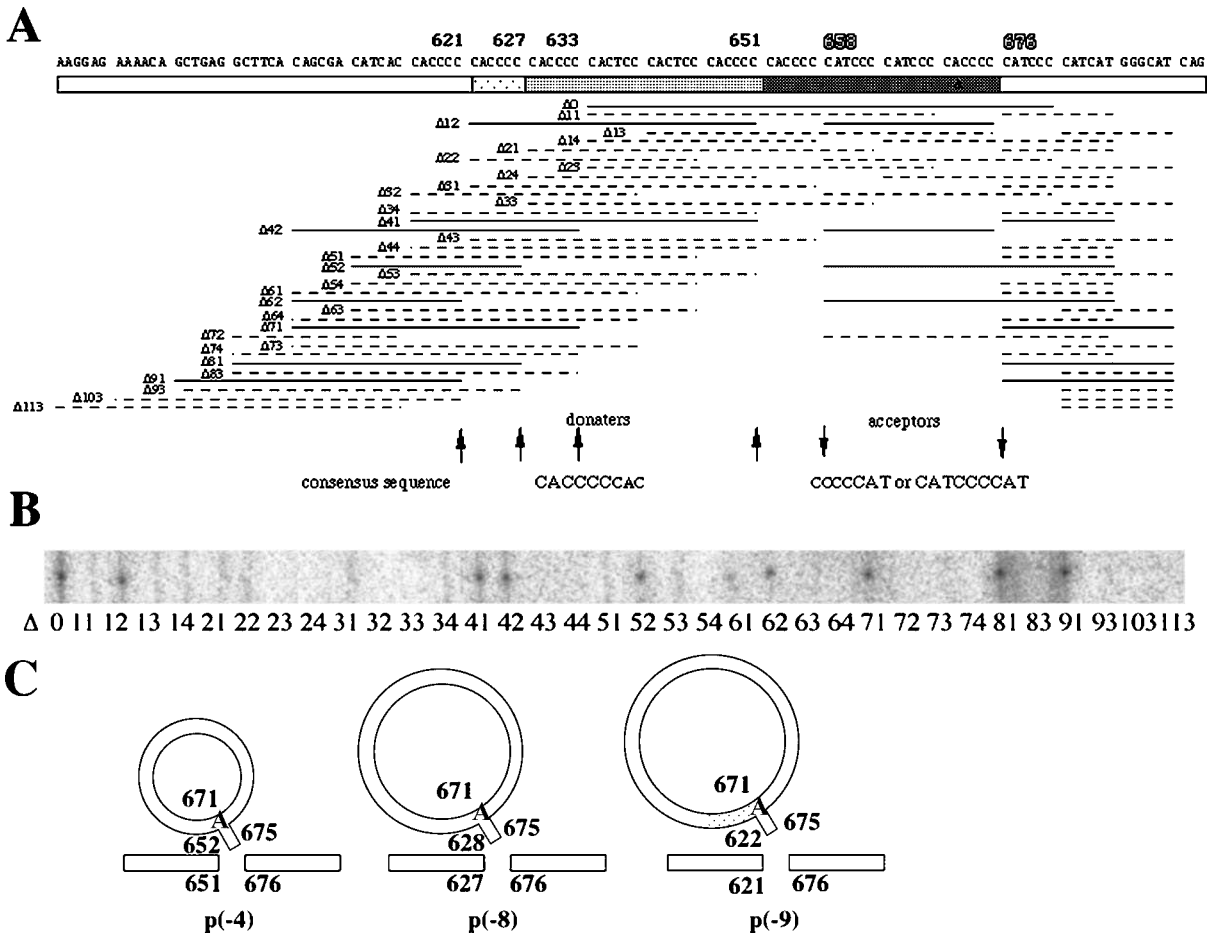

Fig. 4 S1 protection assay using synthetic oligonucleotides whose sequence was assured to be the result of (CAYYCC)n deletion. A: The sequence is indicated as bars: The solid and dotted bars indicate respectively protected and unprotected oligonucleotides according to the autoradiogram (B). The open and hatched boxes under the sequence are used for schematic diagram (C). The location consensus sequence required for recognition of donators and acceptors is indicated by upward and downward arrows, respectively. C: Schematic diagram of alternative splicing producing p(-4), p(-8) and p(-9).

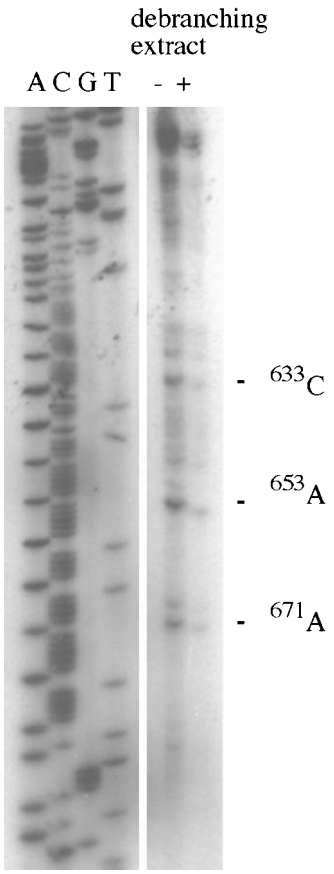

Fig. 5 Mapping of RNA branchpoints at the cytosine-rich portion. The sequence ladder obtained from p(0) using D1 as a primer was used as a size marker. The possible branchpoints (arrows) were obtained by reverse transcription on bovine total RNA using the same primer in the presence and absence of the debranching extract.

specific or common to other mammals, galactosidase-p(0) insert fusion mRNA was expressed in COS7 cells derived from an African green monkey kidney. RNA obtained from COS7 transfected with pBK-p(0), which orthodirectionally contains a p(0) insert, included PCR-amplified bands with sizes of 500, 450, 440, 370, 320 and 270 bps. Since primers whose sequence was located on the vector were used for PCR in order to avoid amplification of COS7 mRNA species, the size of the products was approx. 70 bps longer than the size of the insert. Direct sequence analysis confirmed that the 500-bp-product was identical to the p(0) insert and that the 450- and 440-bp-products corresponded to p(-8) and p(-9), respectively (Fig. 6). Other bands were derived as a result of pseudoamplification. From the RNA of COS7 transfected with pBK-p(0)rev, which contained the p(0) insert in the reverse direction, only a single band corresponding to p(0)rev could be amplified in addition to pseudoamplification.

Discussion

During an investigation into whether bovine selenoprotein P-like protein (3) was a bovine substitute for selenoprotein P (1–3) or uniquely bovine, multiple variant mRNAs with different length tandem repeats of (CAYYCC)n were found to exist. In addition to the consensus sequence being required for

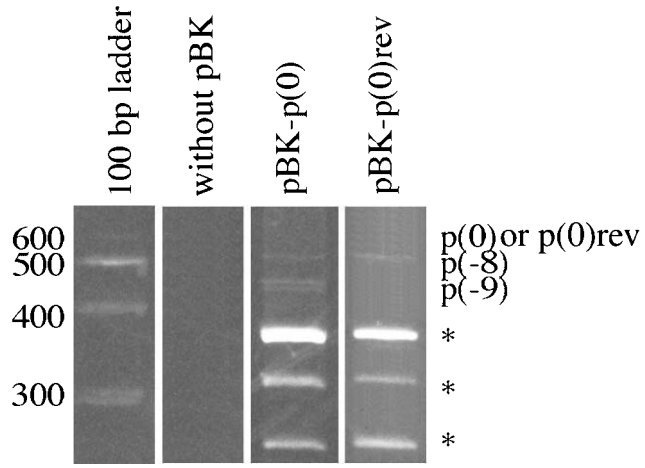

Fig. 6 Splicing variants obtained from COS7 cells transfected with pBK-p(0) or PBK-p(0)rev. The 520-, 470- and 460-bp-products were found to correspond to p(0), p(-8) and p(-9), respectively. Since primers whose sequence was located on the vector were used for PCR in order to avoid amplification of COS7 mRNA species, the size of the products was approx. 90 bps longer than the size of the insert. The 370, 320 and 260-bp-products (asterisks) seem to be obtained by pseudo-amplification, because their sequence could be analyzed neither by U1 nor by D1.

recognition of the donator-acceptor, existence of branch point A suggested that these variants were the result of a novel alternative splicing mechanism. It is noted that this splicing could splice out as short as 24 intron bases.

Two different mRNAs, namely p(0) and p(-8), are known to exist in the bovine encoding selenoprotein P-like protein (3). The insert of clone p(-8) was identical to that of p(0) except for the absence of $(CAYYCC)_8$. Thus, these selenoprotein P-like protein mRNAs differ from selenoprotein Ps from other animal species (1–3) because they have a cytosine-rich portion including a tandem repeat of $(CAYYCC)_{11}$ for p(0) or $(CAYYCC)_8$ for p(-8), as well as having 12, but not 10, TGAs for incorporating selenocysteine (3). The cytosine-rich portion is thought to encode a His-Pro-rich domain, which is reportedly homologous to the PRD repeat found in the segmentation gene *paired* product (10) and the homeobox protein Om(1D) of Drosophila (11). The importance of this His-Pro rich domain remains obscure but it has constructed a region with a slightly negative free energy change (3), and changes in such a region may even affect conformation of the protein itself. The possible existence of a bovine selenoprotein P without the cytosine-rich portion was explored through isolating its gene locus.

In addition to genomic cloning (4), genomic Southern blot analysis and PCR in our study demonstrated that only a single gene locus for the selenoprotein P-like protein exsited in bovines, but not selenoprotein P without the cytosine-rich portion. It is thus likely that both of p(0) and p(-8) are bovine substitutes of selenoprotein P regardless of whether they have unique functions exerted through the His-Pro rich domain or not. Since the sequence of p(-8) was identical to that of p(0) except for the $(CAYYCC)_8$ deletion, it is also possible that these mRNAs were the result of alternative splicing. However, within the cytocine-rich portion, neither classical GT-AG (5, 12) nor minor class AT-AC (13) donator-acceptor sequences for alterna-

tive splicing were found to exist. Instead, CA-CC seemed to function as a donator-acceptor sequence although there were also several CA-CCs in the cytosine-rich portion. If CA-CC can be utilized for alternative splicing, other splicing variants may also exist. As expected, p(-4) lacking $(CAYYCC)_4$ and p(-9) lacking $(CAYYCC)_9$ could be cloned in addition to p(0) and p(-8). An RNA protection assay using cRNA derived from these 4 clones as a probe confirmed that these mRNA species constitutively expressed in bovine and that their absence in (CAYYCC)n were not the result of low fidelity in the reverse transcriptase or Taq polymerase during the cloning procedure.

For recognition of both classical GT-AG and minor class AT-AC donator-acceptor sequences, a consensus sequence is required which flanks these dinucleotides (5, 6, 12). In addition to the investigation of the possible existence of other variants, an S1 protection assay was performed to explore the consensus sequence required for recognition of CA-CC. This assay used as probes, synthetic oligonucleotides whose sequences were able to function as variants. The presence of the protected oligonucleotides suggests that CACCCCCAC is required for recognition of the donator, and CCCCCAT or CATCCCCAT for recognition of the acceptor. For cleavage of 5' splice site, both GT-AG and AT-AC splicing required a lariat formation at the A nucleotide which is located close to the 3' splice site in the intron sequence (5, 6, 12). That is, this A nucleotide cleaves 5' splice site and the cut 5' end of the intron sequence covalently links with this A nucleotide to form a lariat, so that it is known as branch point A. The OH base at the 3' end of the 5' splice site adds to the beginning of the exon sequence in the 3' splice site and releases the intron sequence as a lariat. As a result, the two exon sequences are joined (12). The location of branch point A can be detected as a termination when primer extention is performed on total RNA priming 30–60 nucleotide downstream of the branch point. Since this branch point is an intermediate product during a splicing reaction performed in the nucleus as mentioned above, such a termination disappears after total RNA is treated with a debranching nuclear extract (8). In our study, terminations were observed not only at $^{653 \text{ and } 671}$As but also at ^{631}C. For recognition of the branch point A nucleotide, YNYTRAY for GT-AG splicing, and CTRAC or TCCTTRAC for AT-AC splicing are reportedly required (6). $^{653 \text{ and } 671}$As were characterized as located 5 bs upstream to $^{658 \text{ and } 676}$CATCCCCATs, respectively. ^{658}CATCCCCAT was 5' terminal of exon 5-1 of p(-4) and ^{676}CATCCCCAT was that of p(-8) and p(-9). Thus, these A nucleotides seem to be branch points and ACCCCCAT and/or ACCCCCATCCCCAT were confirmed to be important for recognition of the branch point as well as of the acceptor. In addition to A nucleotides, G nucleotides also function as catalysts and are included in the self-splicing intron observed in *Tetrahymena*, bacteriophage T4, etc. (12). In contrast to these purines, C nucleotides are less likely to have catalytic activity, so that the significance of its presence in the termination at ^{633}C is not clear but it is unlikely that this particular C nucleotide plays a branch point role. It is still possible that the variants were produced during transcription. Neither snRNAs nor the spliceosome responsible for recognition of these splicing sites could be yet identified. Nonetheless, existence of branch point A as well as of consensus sequence for recognition of donator-acceptor suggested that these variants were the result of a novel alternative splicing which does not use classical GT-AG (5, 12) or minor class AT-AC (6) as donator-acceptor sequences.

From mRNA of COS7 cells transfected p(0) insert orthodirectionally, only three variants corresponding to p(0), p(-8) and p(-9) could be isolated, while multiple variants including the Δ12 lacking $^{652}(CAYYCC)_1$ were identified in bovines. It is possible that the length required for intron recognition and deletion might be different for primates and bovines. Even in bovines, other variants with shorter depletions excluding p(-4) lacking $(CAYYCC)_4$ seemed not to be the major variants, because they could not be cloned and could be identified by S1 protection assay alone. Especially, the variant without $(CAYYCC)_1$ was not considered to be enriched because only one C nucleotide, i.e., ^{652}C, was found to exist between the 5' splice site and branch point A, indicating that no lariat could be constructed even when the A nucleotide could attack the 5' splice site. It is likely that this mechanism, which could delete as short as 24 intron bases, was not bovine-specific but common to other mammals.

In the gene database, no gene containing both CACCCCCAC and ACCCCCAT or -ACCCCCATCCCCAT has been registered yet so that it is at present impossible to determine exactly how many genes were the targets for this deletion. It is thus necessary to explore not only how stringent a consensus of sequences is required for recognition of the donator-acceptor and the branch point A nucleotide, but also whether this mechanism is utilized for other proteins containing the His-Pro rich domain to regulate the function of this domain.

Acknowledgement

This work was in part supported by the Grants-in-Aid of Ministry of Science, Education and Culture, Japan (C2-16590474).

References

(1) Hill KE, Lloyd RS, Yang JG, Read R, Burk RF. The cDNA for rat selenoprotein P contains 10 TGA codons in the open reading frame. J Biol Chem. 1991; 266: 10050–10053.

(2) Hill KE, Lloyd RS, Burk RF. Conserved nucleotide sequence in the open reading frame and 3' untraslated region of selenoprotein P mRNA. Proc Natl Acad Sci U S A. 1993; 90: 537–541.

(3) Saijoh K, Saito N, Lee MJ, Fujii M, Kobayashi T, Sumino K. Molecular cloning of cDNA encoding a bovine selenoprotein P-like protein containing 12 selenocysteines and a (His-Pro) rich domain insertion, and its regional expression. Mol Brain Res. 1995; 30: 301–311.

(4) Fujii M, Saijoh K, Kobayashi T, Fujii S, Lee MJ, Sumino K. Analysis of bovine selenoprotein P-like protein gene and

availability of metal responsive element (MRE) located in its promoter. Gene. 1997; 199: 211–217.
(5) Kramer A. The structure and function of proteins involved in mammalian pre-mRNA splicing. Ann Rev Biochem. 1996; 65: 367–409.
(6) Wu Q, Krainer AR. U1-mediated exon definition interactions between AT-AC and GT-AG introns. Science. 1996; 276: 1005–1008.
(7) Davis LG, Kuehl M, Battey J. Basic Methods in Molecular Biology. 2nd ed., New York: Elsevir Science Publishing; 1994.
(8) Grabowski PJ. Characterization of RNA. In: Higgins SJ, Hames BD, editors. RNA Processing Vol. 1. New York: IRL Press; 1994. p. 49–55.
(9) Gluzman Y. SV40-transformed simian cells support the replication of early SV40 mutants. Cell. 1981; 23: 175–182.
(10) Frigerio G, Burri M, Bopp D, Baumgartner S, Noll M. Structure of the segmentation gene paired and the DrosophillaPRD gene set as part of a gene network. Cell. 1986; 47: 735–746.
(11) Tanda S, Corces VG. Retrotransposon-induced overexpression of a homeobos gene causes defects in eye morphogenesis in Drosphila. EMBO J. 1991; 10: 407–417.
(12) Alberts B, Bray D, Lewis J, Raff M, Roberts K, Watson JD. Molecular Biology of the Cell, 3rd ed. Garland New York; 1994. p. 365–378.