ELSEVIER

Biochimica et Biophysica Acta 1245 (1995) 430–438

BBA Biochimica et Biophysica Acta

Structure of genes for sperm-specific nuclear basic protein (SP4) in *Xenopus laevis*

Koichi Mita [a,*], Nobuyuki Ariyoshi [b], Shin-Ichi Abé [b], Kazufumi Takamune [b], Chiaki Katagiri [a]

[a] *Division of Biological Sciences, Graduate School of Science, Hokkaido University, Sapporo 060, Japan*
[b] *Department of Biological Science, Faculty of Science, Kumamoto University, Kumamoto 860, Japan*

Received 17 May 1995; accepted 1 August 1995

Abstract

Nuclear basic proteins in sperm of *Xenopus laevis* consist of 6 sperm-specific proteins (SPs1-6) in addition to somatic core histones. Using a cDNA for SP4 as a probe, we cloned genomic DNA containing SP4 genes from a genomic library constructed from recombinant λ bacteriophage containing 12.0 kbp-*Eco*RI digests of J-strain *X. laevis* liver DNA. Construction of restriction maps based on Southern blot analysis revealed the existence of a total of five SP4 genes which are arranged in a tandemly repeated array forming a cluster of simple multigenes per haploid genome, over a range of 18 kbp. Among these genes, the one located at the most upstream position differed from others in possessing a single base substitution which gave rise to a replacement of one out of 78 amino acid residues. The DNA containing the second to the fourth SP4 genes, arranged at about 3 kbp intervals each, was totally sequenced for 10165 bp. Each gene was found to contain one intron, typical TATA and CCAAT boxes in the 5'-flanking region, and a polyadenylation signal in the 3'-flanking region. Comparative sequence analyses revealed three regions of extensive homology within the upstream non-coding region among three genes, suggesting a possible relevance to their expression at a particular phase of spermatogenesis and/or in testis.

Keywords: Sperm-specific nuclear basic protein; Spermatogenesis; Protamine; (*Xenopus*)

1. Introduction

The nuclear basic proteins of mature sperm in most animals consist of highly basic, arginine-rich proteins termed protamines or sperm histones depending on their biochemical properties (for a review, [1]). Profound biochemical changes in chromosomal proteins occur during spermatogenesis which include partial or complete replacement of basic proteins from somatic histones with sperm-specific basic proteins (SBPs). Although this process includes the appearance of transitional type proteins in some mammals, immunohistochemical [2] and electrophoretic analyses [3,4] using homogeneous spermatogenic cell fractions indicated that protamines are deposited at the final step of spermatogenesis concomitantly with nuclear condensation of elongating spermatids. Studies employing cDNAs for protamines in mammals [5,6], birds [7,8], and an amphibian [9] (possibly also in fish [10]) revealed that the genes for these proteins are transcribed at the round spermatid (haploid) stage. More recent studies using transgenic mice [11,12] or in vitro transcription systems [13] suggest possible cis-regulating elements participating in haploid- and/or testis-specific expression of protamine genes.

The mature sperm of an anuran amphibian *Xenopus laevis* contain six apparent SBPs (SPs1-6) in addition to somatic core histones, but not H1 [14,15]. Using cloned cDNAs for SP4 and SP5, we have previously shown that mRNAs for these proteins are transcribed at the primary spermatocyte (tetraploid) stage [9,16,17], although they are not translated until the final step of spermatogenesis, as is typical for protamines in other vertebrates [15,18,19]. Thus the genes for SBPs in *Xenopus laevis* provide a unique model with which to study the mechanisms of regulating expression of specific genes at tetraploid stage male germ cells.

As an initial effort to study the regulatory mechanisms of expression of these sperm-specific genes, we have isolated genomic DNA containing the SP4 genes and

* Corresponding author. Tel: +81 11 7162111; Fax: +81 11 7575994.

SSDI 0304-4165(95)00124-7

examined its structure by extensive sequence analyses. We report here the unique organization of these genes.

2. Materials and methods

2.1. Genomic DNA

Genomic DNA used in this study was obtained from liver nuclei of adult J-strain *Xenopus laevis* [20], according to the methods of Sambrook et al. [21] with modifications. Freshly excised liver fragments were homogenized with a Teflon-glass homogenizer in 3 volumes of SMT (250 mM sucrose, 5 mM $MgCl_2$, 10 mM Tris-HCl, pH 7.4) and filtered through double Nitex filters. After centrifugation at 1,000 × g for 5 min, the pellet was washed 5 times with SMT, suspended in STE (100 mM NaCl, 200 mM EDTA, pH 8.0, 10 mM Tris-HCl, pH 8.0), and lysed by addition of SDS and proteinase K (SIGMA) to final concentrations of 0.5% and 0.1 mg/ml, respectively. After incubation at 50°C for 4 h, nucleic acids were purified by conventional methods using phenol and dialyzed against a large volume of TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, pH 8.0) for 1 day. The dialysate was treated with 10 μg/ml RNase A (SIGMA) at 37°C for 30 min. Genomic DNA was purified by phenol extraction, followed by dialysis with a large volume of TE for 1 day.

2.2. DNA probe

The probe used for Southern blot analysis and screening of the *Xenopus* genomic library was a 473 bp insert excised with *Eco*RI from the cDNA clone pXSP531, which codes a full length amino acid sequence of SP4 [16]. This fragment was labeled with [α-^{32}P] dCTP (Amersham) using the Megaprime DNA labeling system (Amersham).

2.3. Southern blot analysis

Two μg of genomic DNA was digested with appropriate restriction enzymes, separated by electrophoresis on a 0.5% agarose gel, and transferred to Hybond-N^+ nylon membrane (Amersham) according to the manufacturer's instructions. The blot was prehybridized at 42°C for 4 h with prehybridization buffer (6 × SSC, 5 × Denhardt's solution, 50% formamide, 0.5% SDS, 0.2 mg/ml salmon sperm DNA), and then hybridized at 42°C for 16 h with the ^{32}P-labeled cDNA probe. The blot was washed successively at room temperature for 5 min in 2 × SSC and 0.5% SDS, for 15 min in 2 × SSC and 0.1% SDS, at 37°C for 30 min in 0.1 × SSC and 0.5% SDS, and at 65°C for 30 min in 0.1 × SSC and 0.5% SDS. The blot was exposed to Fuji RX X-ray film (Fuji Photo Film Co., Ltd.) for 5 days at −80°C with intensifying screens or to an imaging plate (Type BAS-III, Fuji Photo Film Co., Ltd.) for 12–24 h at room temperature for estimating the intensity of the hybridized probe using a Bio-Image Analyzer (BA-100, Fuji Photo Film Co., Ltd.).

2.4. Construction of genomic DNA library and screening

Genomic DNA was digested to completion with *Eco*RI and fractionated on a linear 5–25% sodium chloride gradient (12 ml) containing 3 mM EDTA, pH 8.0, and 10 mM Tris-HCl, pH 8.0 [22]. After centrifugation for 10 h at 24,000 rpm in a Beckman SW27 rotor, the 12.0 kbp DNA fragments were pooled and ligated with *Eco*RI-digested λDASH II vector (Strategene) as described [21]. Ligated DNA was packaged in vitro with GIGAPACK II Gold (Stratagene). *Escherichia coli* cell line P2392 was used as the host for plaque lifting. Replicate plaque-hybridization filters (Hybond-N^+; Amersham) were screened with ^{32}P-labeled SP4 cDNA probe. Hybridization conditions were the same as used for Southern blot analysis. The filters were then exposed to X-ray film (Fuji Photo Film Co., Ltd.) for 6-16 h at −80°C. Positive clones were purified by repeated screening, and DNA from resulting clones was purified according to standard methods [21].

2.5. DNA sequence analysis

The DNA from recombinant phage λXLSP4 was digested with *Eco*RI, and electrophoresed on a 0.5% agarose gel. Insert DNA (12 kbp) was purified from the gel by electroelution, and digested with *Bgl*II. The resulting four fragments were subcloned into the *Eco*RI-*Bam*HI site or *Bam*HI site of pBluescript II SK^+ (Stratagene). A nested series of deletions was created for each strand of all fragments using Exo-Mung Deletion Kit (Stratagene), and the resulting clones were sequenced by the dideoxynucleotide method [23] using a Sequenase version 2.0 kit (USB).

2.6. Polymerase chain reaction (PCR)

DNA used as a PCR template was prepared as follows. Genomic DNA was digested with *Eco*RI, and separated by electrophoresis on a 0.5% agarose gel. The 2 kbp and 4 kbp DNA fragments were extracted from the gel by electroelution, and approximately 0.1 μg DNA was used for each reaction. The primers for PCR were 5′-AGCAAAGTGAGTGGCGGG-3′ and 5′-AGTGGTTTAACTGCGATA-3′, which correspond to the 71st–89th nucleotide sequence and the sequence complementary to 296th–313th nucleotides of pXSP531, respectively [16]. Amplification reactions were carried out in 50 μl volumes containing 10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% (W/V) gelatin, and 250 μM of each dNTPs. Primers were present at 100 pmol each and Taq DNA polymerase (TAKARA Shuzo Co., Ltd.) at 0.02 units/μl. Twenty cycles of amplification were performed, with the denaturation step at 92°C for 2 min, the annealing

step at 50°C for 2 min, and the extension step at 72°C for 2 min. PCR products were treated with mung bean nuclease, then inserted into the *Sma*I site of pBluescript II SK$^+$ and sequenced as described above.

2.7. Comparative DNA sequence analysis

Sequence homologies were analyzed with *GENETYX-MAC* ver. 7.0.2 (Software Development Co., Ltd.), a program that employs a rapid alignment algorithm.

3. Results

Genomic DNA was digested with restriction endonucleases, *Eco*RI, *Hin*dIII, and *Pvu*II, electrophoresed, and blotted for Southern analysis. The blot was hybridized with labeled SP4 cDNA as a probe, washed under high stringency conditions and autoradiographed (Fig. 1). Two to four bands with variable molecular sizes were present in each digest, depending on the restriction enzymes used.

To isolate genomic DNA containing the SP4 gene(s), a genomic library was constructed from the 12.0 kbp-*Eco*RI fragments of J-strain *Xenopus laevis*. One recombinant clone, λXLSP4, which hybridized with the cDNA probe was isolated, and the positions of SP4 genes were determined. The insert from λXLP4 was cut into four fragments with *Bgl*II and subcloned into pBluescript II SK$^+$. Fig. 2 shows a partial restriction map of this clone. We determined the complete sequence totaling 10 165 nucleotides in the 12.0 kbp-*Eco*RI fragment. This nucleotide sequence has been submitted to the DDBJ (DNA Data Bank of Japan) and will appear in the GSDB, DDBJ, EMBL, and NCBI nucleotide sequence database with the accession number D45253. Sequence analysis revealed that this 12.0 kbp fragment contains three SP4 genes (SP4a, 4b, 4c in Fig. 2) which are tandemly repeated at approximately 3 kbp intervals each. Each gene had an intron containing 494 (SP4a) or 495 bp (SP4b and SP4c), which was located

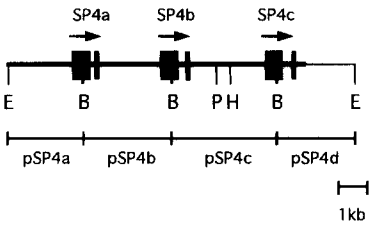

Fig. 2. Restriction map of a 12.0 kbp-*Eco*RI fragment of λXLSP4, based on digestion with *Eco*RI (E), *Bgl*II (B), *Hin*dIII (H) and *Pvu*II (P). Three SP4 genes (SP4a, 4b, 4c) indicated by boxes containing an intron are interspersed with flanking regions. Bold line indicates sequenced region. Arrows show the direction of coding regions. *Bgl*II fragments of λXLSP4 were subcloned into pBluescript II SK$^+$ and these inserts are designated as pSP4a-4d.

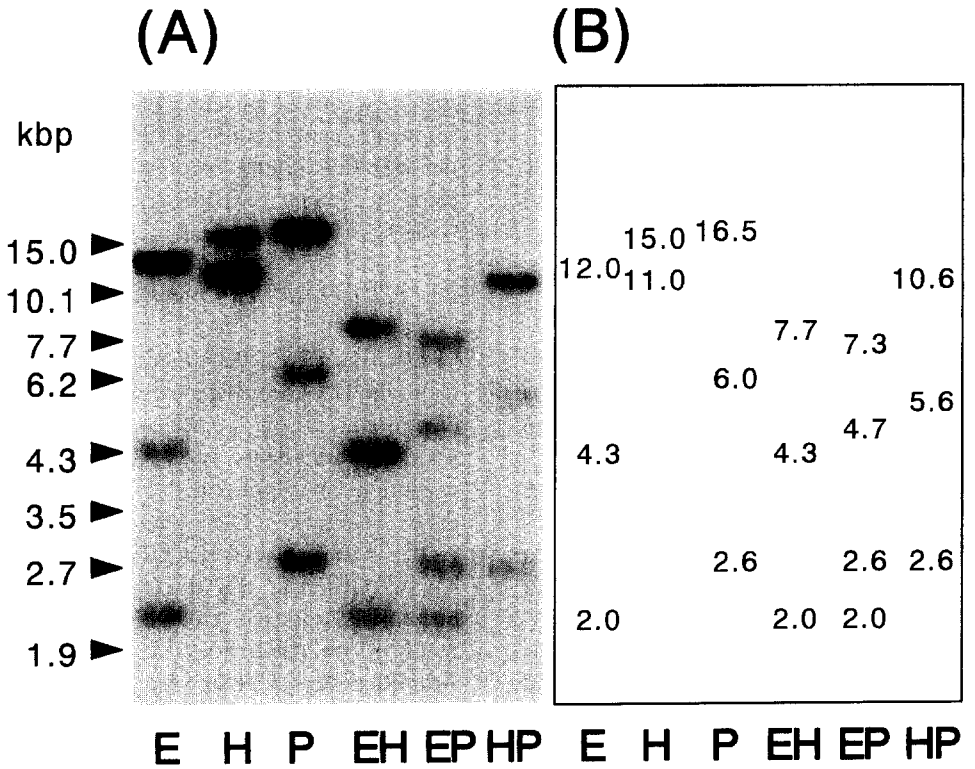

Fig. 1. Southern blot analysis of genomic SP4 genes. (A) Two μg of genomic DNA from J-strain *Xenopus laevis* was digested with restriction enzymes, separated by electrophoresis, transblotted to nylon membrane, and hybridized with radiolabeled SP4 cDNA probes. Restriction enzymes used were *Eco*RI; (E), *Hin*dIII; (H), *Pvu*II; (P), and their combinations (EH, EP, HP). The numbers on the left refer to molecular sizes. (B) The sizes (kbp) of fragments shown in (A) are indicated.

225 bp downstream from the first nucleotide of the coding region, and the 75th deduced amino acid residue from the N-terminal of SP4 was coded by a split codon.

The degree of sequence homology among these three genes including their non-coding region was studied by two-dot matrix homology searches with Harr Plot analysis (Fig. 3). To make this analysis, the insert from λXLSP4 with a total size of 10,165 bp was arbitrarily divided into three fragments at the 3,500th and 6,500th base. Thus, the fragments designated SP4a, 4b, 4c in Fig. 3 contained coding and non-coding regions including approx. 2.0–2.6 kbp upstream from the initiation codon and 0.3–0.4 kbp downstream from the termination codon. Each dot in Fig. 3 represents 16 matches or more in a 20-nucleotide window. It is clear in the figure that the three fragments exhibit extensive sequence homology with one another at three portions in the upstream region. Fig. 4 shows more detailed views of homology within the 5′- and 3′-flanking regions. Within about 300 bp of the 5′-flanking region of each gene (Fig. 4A), there are predicted TATA and CCAAT boxes, the latter being present on another strand as defined by the complementary sequence 5′-ATTGG-3′. Primer extension analysis (data not shown) revealed that the predicted transcriptional initiation sites of SP4a and SP4b are at the 2,152th and 5,282th bases in λXLSP4, respectively (Fig. 4A, arrow head). The positions of the TATA box are −17 bp from these transcriptional initiation sites (Fig. 4A, boxed) and the CCAAT box at −73 (Fig. 4A, shaded). The transcriptional initiation site for SP4c was not determined. TATA box-like motifs are also present at more upstream positions (30 bp and 70 bp upstream from actual TATA box). In the 3′-flanking region, all three genes contain polyadenylation signals AATAAA at positions 135 bp (SP4a) or 137 bp (SP4b and 4c) downstream from the last nucleotide of coding region (Fig. 4B, boxed).

The result of Southern blot analysis after digestion with *Eco*RI (Fig. 1, lane E) suggested the existence of more than three SP4 genes. To determine the total number of genes in the haploid genome, the radioactivity of signals in the Southern blot shown in Fig. 1 was quantitated using a Bio-Image analyzer (Table 1). Since the 12.0 kbp fragment from the *Eco*RI digest contained three SP4 genes as shown in Fig. 2, the relative value of the radioactivity in this fragment was taken as 3. On this basis, the radioactivity in the 4.3 kbp- and 2.0 kbp-*Eco*RI fragments was determined to represent each one gene. Thus we conclude that a total of five genes are present in the haploid genome. Analysis of the radioactivity in the two *Hin*dIII fragments

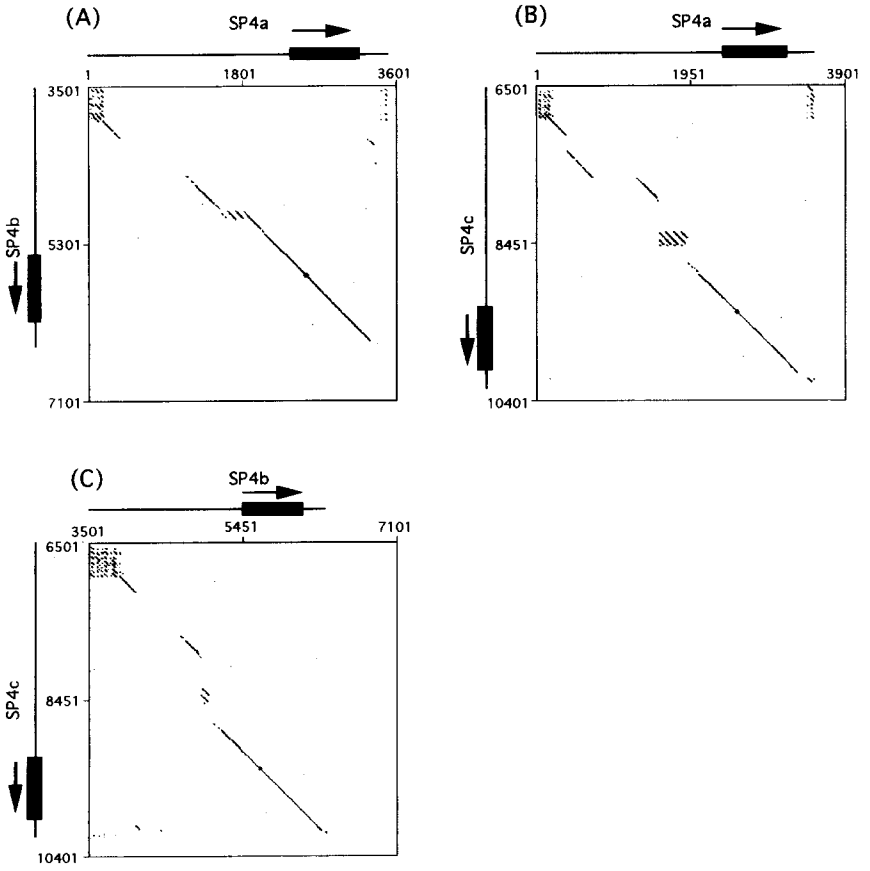

Fig. 3. Dot matrix analyses of three SP4 genes in λXLSP4. The results of dot matrix analyses of SP4a and 4b, 4a and 4c, and 4b and 4c are depicted in (A), (B), and (C), respectively. Each dot represents 16 matches or more in a 20-nucleotide window. The numbers at both the ordinate and the abscissa correspond to those in λXLSP4.

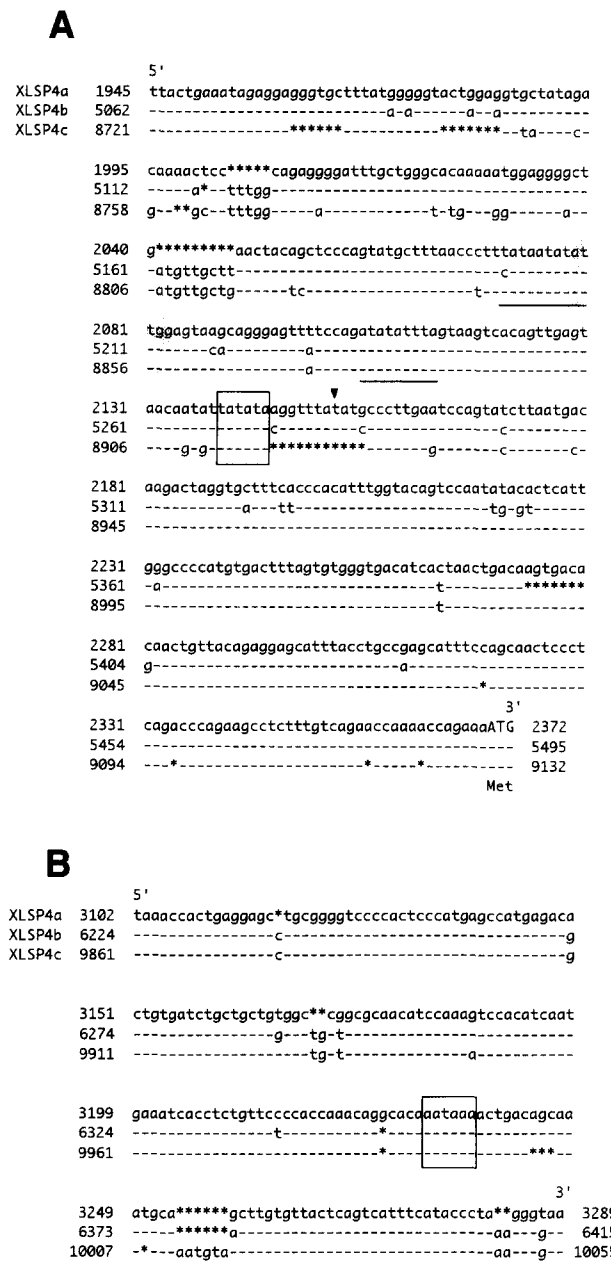

Fig. 4. Nucleotide sequences of (A) 5'- and (B) 3'-flanking regions of three SP4 genes in λXLSP4. Sequence of XLSP4a is shown. For other genes (XLSP4b and 4c), only substituted nucleotides are indicated, with nucleotide matches to XLSP4a shown by dashes. Asterisks indicate deleted nucleotides. Numbers at left and at the end of sequences are the nucleotide number of λXLSP4. In (A), shaded and boxed nucleotides represent CCAAT- and TATA-box, respectively. TATA-like sequences are underlined. An arrow head indicates the transcription initiation site. The last three nucleotides (ATG) are the initiation codon of SP4. In (B), boxed nucleotides represent a polyadenylation signal. The first three nucleotides (taa) are the termination codon of SP4.

gave relative values of 2:3, indicating the existence of five genes. Similarly, the *Pvu*II-digests of 16.5 kbp, 6.0 kbp and 2.6 kbp fragments gave rise to values of three, one and one genes, respectively.

Alignment of these restriction fragments including the

Table 1
Estimation of SP4 gene numbers by quantification of radioactivity on Southern blots after digestion with various restriction enzymes.

Fragment	Radioactivity (AU) [a]	Relative value [b]
12.0kbp-*Eco*RI	10,822	3.0
4.3kbp-*Eco*RI	3,309	0.9
2.0kbp-*Eco*RI	4,603	1.3
15.0kbp-*Hin*dIII	6,632	1.8
11.0kbp-*Hin*dIII	12,722	3.5
16.5kbp-*Pvu*II	10,842	3.0
6.0kbp-*Pvu*II	3,388	0.9
2.6kbp-*Pvu*II	4,695	1.3

Genomic DNA from J-strain *X. laevis* was digested with the restriction enzymes indicated, electrophoresed and Southern blotted using a SP4 cDNA probe. Radioactivity of the signals shown in Fig. 1(A), lanes E, H, and P was measured using an Image Analyzer

[a] Radioactivity on imaging plate

[b] Relative value taking the radioactivity of the 12.0 kbp-*Eco*RI fragment as 3.0, because it contains 3 SP4 genes.

position of five genes (XLSP41-45) shown in Fig. 5 was constructed in the following way. Sequence analysis of the 12.0 kbp-*Eco*RI fragment (Fig. 2) revealed the presence of both *Pvu*II- and *Hin*dIII-sites between SP4b and SP4c. Thus, two *Hin*dIII fragments containing SP4 genes must be linked together. Existence of three genes in the 16.5 kbp fragment of *Pvu*II digests and the lack of 4.3 kbp signal in the double-digests of *Eco*RI and *Pvu*II suggest that there is 2.0 kbp-*Eco*RI fragment located at a position upstream from the 12.0 kbp-*Eco*RI fragment. Thus the 11.0 kbp-*Hin*dIII fragment must be included in the genes located in the 16.5 kbp-*Pvu*II fragment (Fig. 5). Double digestions with *Eco*RI and *Pvu*II indicate that the gene located in the 4.7 kbp fragment corresponds to that included in the 6.0 kbp-*Pvu*II fragment. Similarly, the gene located in the 4.3 kbp-*Eco*RI fragment must represent that included in the 2.6 kbp-*Pvu*II fragment. Based on this analysis, the alignment of the five genes (XLSP41-45) in relation to the restriction map of the genome is as shown in Fig. 5, where SP4a, SP4b, and SP4c genes in Fig. 2 are

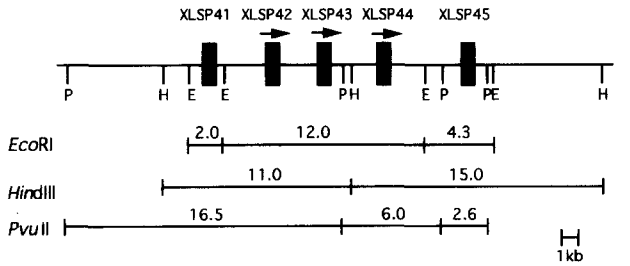

Fig. 5. Alignment of SP4 genes on genomic DNA, with five SP4 genes XLSP41-45 (boxes) containing an intron and flanking regions (solid line). Arrows indicate the direction of the coding regions which were determined by sequencing. XLSP42-44 correspond to SP4a-4c in Fig. 2, respectively. Recognition sites for restriction enzymes are indicated with abbreviated names of enzymes, P (*Pvu*II), H (*Hin*dIII), and E (*Eco*RI). Numbers indicate length of DNA fragments in kbp.

renamed XLSP42, XLSP43, and XLSP44, respectively, according to their order of arrangement in the genome.

To confirm the presence of SP4 genes in the 4.3 kbp- and 2.0 kbp-*Eco*RI fragments, PCR products generated from SP4-specific primers were cloned and sequenced. The results are summarized in Fig. 6, where the pertinent sequences in the 4.3 kbp and 2.0 kbp fragments are referred to as XLSP41 and XLSP45, respectively. As seen in the figure, all five genes are highly homologous, with predicted amino acid sequences which are identical to those of SP4 reported previously [16], with the exception of one possible amino acid replacement in XLSP41.

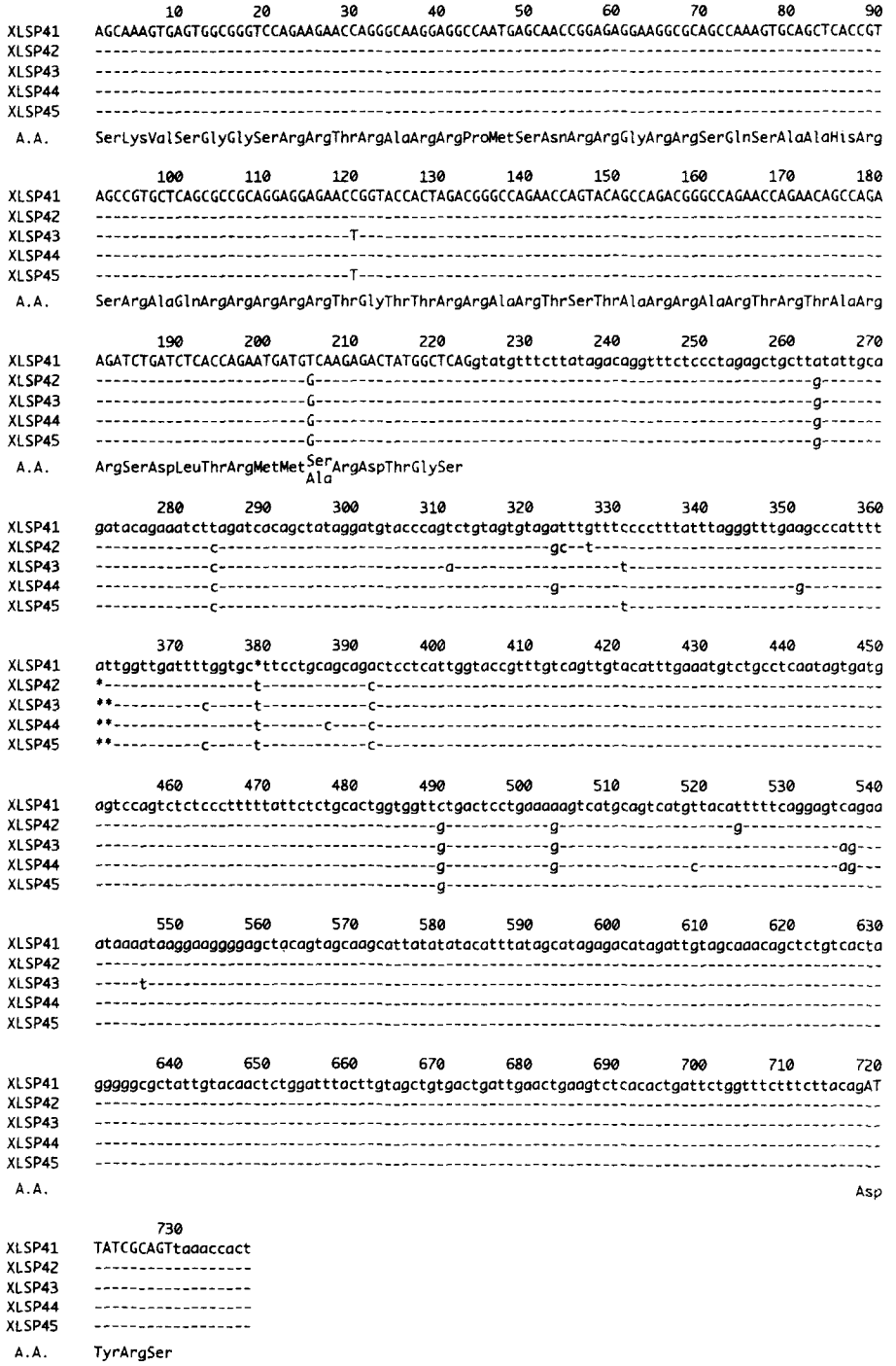

Fig. 6. Nucleotide sequences and predicted amino acid sequences of five SP4 genes (XLSP41-45). Nucleotide sequence of XLSP41 is shown, with exons and introns in capital and small letters, respectively. For other genes (XLSP42-45), only substituted nucleotides are indicated, with nucleotide matches to XLSP41 shown by dashes. Asterisks indicate deleted nucleotides. A.A., predicted amino acid sequence including substitution of Ala by Ser due to a nucleotide substitution of G by T at position 205 in XLSP41.

XLSP41 differs from the other four genes in possessing a base substitution from G to T at position 205 in Fig. 6, giving rise to an amino acid replacement from Ala to Ser.

In comparison with their exons, there are relatively more base substitutions within the introns of the five SP4 genes, including a reduction of one base in XLSP43-45 (Fig. 6). However, in all five genes both intron boundaries consist of a split codon. In addition, introns in all genes follow the GT-AG rule [24], with exon-intron boundaries being CAG/GTATGT and intron-exon boundaries TTTCTTTCTTACAG/A (Fig. 6). These sequences follow the consensus proposed by Mount [25].

4. Discussion

We show in this paper that in *Xenopus laevis*, five SP4 genes each containing one intron are arranged in a tandemly repeated array forming a cluster of simple multigenes per haploid genome. In comparison with the gene structures for sperm proteins in other classes of vertebrates so far studied, the genes for *Xenopus* SP4 share a property with protamine P1 in mammals (bull [26], mouse[27], and human [28]) in containing an intron but differ from intronless genes in fishes [29] and birds (rooster [30] and quail [31]. Regarding the copy numbers, the situation in *Xenopus* is similar to fishes and birds which possess multiple [7,29] and two copies [7,30,31] per haploid genome, respectively, but differs from mammalian counterparts which contain a single copy per haploid genome [26–28]. The testis-specific histone H1 (Hlt) gene in rat is reportedly also a single copy gene [32]. This organization of genes for sperm nuclear basic proteins is in contrast to those for somatic histone genes in most organisms which exist in many copies forming large clusters per haploid genome [33–35]. In *Xenopus*, there are a total of 45–50 genes coding for each of the nucleosomal histones present per haploid genome, forming clusters containing at least one copy of each of the five histone genes [36].

Our finding of tandemly repeated SP4 genes raises a question of whether these genes are all actively transcribed as long heterogeneous RNA or are expressed as separate entities. Available evidence suggests that, at least in the case of the gene XLSP42-44, the latter is more likely. First, all genes contain a polyadenylation signal at about 130 bp downstream from the last codon (cf. Fig. 4B). Second, about 400 bp upstream from each gene is a highly homologous region containing a complementary CCAAT box sequence (5′-ATTGG-3′) and a candidate for a TATA box (cf. Fig. 4A), the latter being known to bind the TFIID factor to determine the location of the transcription initiation site [37,38]. Third, previous Northern blot analyses [9,16] revealed that SP4 mRNAs from primary spermatocytes with sizes of approximately 650–710 nucleotides and lacking an intron give rise upon removal of poly (A) tracts to a size of approximately 560 bases [16] containing only one open reading frame, indicating that each SP4 gene is expressed separately.

It is not known whether all five of the SP4 genes presented here are transcribed or not. There is evidence, however, suggesting that many of them are active. Support for this notion comes from the results of primer extension analysis (unpublished data) showing that there are seven different lengths of single strand DNAs synthesized. These differences in the length of transcripts are likely to be the result of differences in the position and plural number of predicted TATA motifs in the 5′-flanking regions of XLPS42-44 (cf. Fig. 4A). Another line of evidence indicates that the product of the most upstream XLSP41 gene is collected as a separate entity from the typical SP4 gene products in mature sperm. A smaller amount of this protein, which eluted faster than the major SP4 peak (cf. [15]) in HPLC purification, has an amino acid sequence the same as that predicted from the nucleotide sequence of XLSP41 [39]. Thus the amounts of the translational products of XLSP41 and XLSP42-45 genes reflect approximately the copy number of these genes. Finally, Northern blot analysis of transcriptional products of the SP5 gene [17], presumed to be a single copy gene [40], showed that the relative amount of mRNA for SP5 seen in primary spermatocytes is approximately one-fifth of that for SP4. This may again reflect the relative copy numbers of the SP4 and SP5 genes, although a difference in the control and efficiency of their transcription have not been ruled out.

An interesting finding from our comparative sequence analysis concerns the presence of highly conserved upstream sequence in the three SP4 genes (XLSP42, 43, and 44; cf. Fig. 3), although these particular sequences are interspersed by relatively less homologous sequences. These conserved sequence attract attention for their possible relevance to regulation of organ (testis)- and stage (primary spermatocyte)-specific expression of SP4 genes [9,16]. Several promoter sequences have recently been defined for genes specifically expressed in mammalian testis. These include the genes for lactate dehydrogenase C (Ldhc) [41], the E1α subunit of the pyruvate dehydrogenase complex (Pdha-2) [42], phosphoglycerate kinase 2 (Pgk-2) [43,44], and histone variants H1t and TH2B [45,46], expressed in primary spermatocytes, as well as transition protein 1 (TP1) [47], and protamine 1 and 2 [26,27,11], expressed in spermatids. No general consensus regulatory elements which may function for testis-specific expression of these genes have been identified, although candidates for consensus elements limited to both Pgk-2 and Pdha-2 genes [42], and protamine 1 and 2 genes [27] are suggested. Zhou et al. (1994) proposed the importance of the palindromic sequence overlapping the TATA box and transcription initiation site for Ldhc gene expression [41]. It is pertinent to mention in connection with this that in the upstream regions of our SP4 genes, there is no element reminiscent of the sequences and palindromic structures

which were reported previously in genes specifically expressed in testis. We need to determine whether the conserved upstream sequences found in the SP4 genes are crucial as cis-elements for selective expression of these genes at the spermatocyte stage and/or in testis.

Acknowledgements

This study was supported by the Akiyama Foundation, and Grant-in-Aid for Scientific Research (#04454021; #05454654) and priority areas (#05277212) from the Japanese Ministry of Education, Science and Culture.

References

[1] Poccia, D. (1986) Remodelling of nucleoproteins during gametogenesis, fertilization and early development. Int. Rev. Cytol. 105, 1-65.
[2] Roux, Ch., Gusse, M., Chevaillier, Ph. and Dadoune, J.P. (1988) An antiserum against protamines for immunohistochemical studies of histone to protamine transition during human spermatogenesis. J. Reprod. Fert. 82, 35–42.
[3] Mayer, J.F., Jr., Chang, T.S.K. and Zirkin, B.R. (1981) Spermatogenesis in the mouse: 2. Amino acid incorporation into basic nucleoproteins of mouse spermatids and spermatozoa. Biol. Reprod. 25, 1041–1051.
[4] Balhorn, R., Weston, S., Thomas, C. and Wyrobek, A.J. (1984) DNA packaging in mouse spermatids: Synthesis of protamine variants and four transition proteins. Exp. Cell. Res. 150, 298–308.
[5] Hecht, N.B., Bower, P.A., Kleene, K.C. and Distel, R.J. (1985) Size changes of protamine 1 mRNA provide a molecular marker to monitor spermatogenesis in wild-type and mutant mice. Differentiation 29, 189–193.
[6] Mali, P., Sandberg, M., Vuorio, E., Yelick, P.C., Hecht, N.B. and Parvinen, M. (1988) Localization of protamine 1 mRNA in different stages of the cycle of the rat seminiferous epithelium. J. Cell Biol. 107, 407–412.
[7] Oliva, R. and Dixon, G.H. (1991) Vertebrate protamine genes and the histone-to-protamine replacement reaction. Prog. Nucleic Acid Res. Mol. Biol. 40, 25–94.
[8] Oliva, R., Mezquita, J., Mezquita, C. and Dixon, G.H. (1988) Haploid expression of the rooster protamine mRNA in the postmeiotic stage of spermatogenesis. Dev. Biol. 125, 332–340.
[9] Mita, K., Takamune, K. and Katagiri, Ch. (1991) Genes for sperm-specific basic nuclear proteins in *Bufo* and *Xenopus* are expressed at different stages in spermatogenesis. Dev. Growth Differ. 33, 491–498.
[10] Seniuk, N.A., Tatton, W.G., Cannon, P.D., Garber, A.T. and Dixon, G.H. (1991) First expression of protamine message in trout testis. Ann. NY Acad. Sci. 637, 277–288.
[11] Zambrowicz, B.P., Harendza, C.J., Zimmermann, J.W., Brinster, R.L. and Palmiter, R.D. (1993) Analysis of the mouse protamine 1 promoter in transgenic mice. Proc. Natl. Acad. Sci. USA 90, 5071–5075.
[12] Peschon, J.J., Behringer, R.R., Brinster, R.L. and Palmiter, R.D. (1987) Spermatid-specific expression of protamine 1 in transgenic mice. Proc. Natl. Acad. Sci. USA 84, 5316–5319.
[13] Tamura, T., Makino, Y., Mikoshiba, K. and Muramatsu, M. (1992) Demonstration of a testis-specific *trans*-acting factor Tet-1 in vitro that binds to the promoter of the mouse protamine 1 gene. J. Biol. Chem. 267, 4327–4332.
[14] Risley, M.S. and Eckhardt, R.A. (1981) H1 histone variants in *Xenopus laevis*. Dev. Biol. 84, 79–87.
[15] Yokota, T., Takamune, K. and Katagiri, Ch. (1991) Nuclear basic proteins of *Xenopus laevis* sperm: Their characterization and synthesis during spermatogenesis. Dev. Growth Differ. 33, 9–17.
[16] Hiyoshi, H., Uno, S., Yokota, T., Katagiri, Ch., Nishida, H., Takai, M., Agata, K., Eguchi, G. and Abé, S.-I. (1991) Isolation of cDNA for a *Xenopus* sperm-specific basic nuclear protein (SP4) and evidence for expression of SP4 mRNA in primary spermatocytes. Exp. Cell Res. 194, 95–99.
[17] Ariyoshi, N., Hiyoshi, H., Katagiri, Ch. and Abé, S.-I. (1994) cDNA cloning and expression of *Xenopus* sperm-specific basic nuclear protein 5 (SP5) gene. Mol. Reprod. Dev. 37, 363–369.
[18] Abé, S.-I. and Hiyoshi, H. (1991) Synthesis of sperm-specific basic nuclear proteins (SPs) in cultured spermatids from *Xenopus laevis*. Exp. Cell. Res. 194, 90–94.
[19] Moriya, M. and Katagiri, Ch. (1991) Immunoelectron microscopic localization of sperm-specific nuclear basic proteins during spermatogenesis in anuran amphibians. Dev. Growth Differ. 33, 19–27.
[20] Nakamura, T., Maeno, M., Tochinai, S. and Katagiri, Ch. (1987) Tolerance induced by grafting semi-allogeneic adult skin to larval *Xenopus laevis*: possible involvement of specific suppressor cell activity. Differentiation 35, 108–114.
[21] Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd Ed., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[22] Grosveld, F.G., Lund, T., Murray, E.J., Mellor, A.L., Dahl, H.H.M. and Flavell, R.A. (1982) The construction of cosmid libraries which can be used to transform eukaryotic cells. Nucleic Acids Res. 10, 6715–6732.
[23] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[24] Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc. Natl. Acad. Sci. USA 75, 4853–4857.
[25] Mount, S.M. (1982) A catalogue of splice junction sequences. Nucleic Acids Res. 10, 459–472.
[26] Krawetz, S.A., Connor, W. and Dixon, G.H. (1988) Bovine protamine genes contain a single intron. J. Biol. Chem. 263, 321–326.
[27] Johnson, P.A., Peschon, J.J., Yelick, P.C., Palmiter, R.D. and Hecht, N.B. (1988) Sequence homologies in the mouse protamine 1 and 2 genes. Biochem. Biophys. Acta 950, 45–53.
[28] Krawetz, S.A., Herfort, M.H., Hamerton, J.L., Pon, R.T. and Dixon, G.H. (1989) Chromosomal localization and structure of the human P1 protamine gene. Genomics 5, 639–645.
[29] Aiken, J.M., McKenzie, D., Zhao, H.-Z., States, J.C. and Dixon, G.H. (1983) Sequence homologies in the protamine gene family of rainbow trout. Nucl. Acid Res. 11, 4907–4922.
[30] Oliva, R. and Dixon, G.H. (1989) Chicken protamine genes are intronless. J. Biol. Chem. 264, 12472–12481.
[31] Oliva, R., Goren, R. and Dixon, G.H. (1989) Quail (*Coturnix japonica*) protamine, full-length cDNA sequence, and the function of vertebrate protamines. J. Biol. Chem. 264, 17627–17630.
[32] Cole, K.D., Kandala, J.C. and Kistler, W.S. (1986) Isolation of the gene for the testis-specific H1 histone variant H1t. J. Bio. Chem. 261, 7178–7183.
[33] Hentschel, C.C. and Birnstiel, M.L. (1981) The organization and expression of histone gene families. Cell 25, 301–313.
[34] Maxon, R., Cohn, R., Kedes, L. and Mohun, T. (1983) Expression and organization of histone genes. Ann. Rev. Genet. 17, 239–277.
[35] Perry, M., Thomsen, G.H. and Roeder, R.G. (1985) Genomic organization and nucleotide sequence of two distinct histone gene clusters from *Xenopus laevis*: identification of novel conserved upstream sequence elements. J. Mol. Biol. 185, 479-499.
[36] Van Dongen, W., DeLaaf, L., Zaal, R., Moorman, A. and Destree, O. (1981) The organization of the histone genes in the genome of *Xenopus laevis*. Nucl. Acid Res. 9, 2297–2311.

[37] Breathnach, R. and Chambon, P. (1981) Organization and expression of eucaryotic split genes coding for proteins. Ann. Rev. Biochem. 50, 349–383.

[38] Lee, D.K., Horikoshi, M. and Roeder, R.G. (1991) Interaction of TFIID in the minor groove of the TATA element. Cell 67, 1241–1250.

[39] Takamune, K., Teshima, K., Abé, S.-I. and Katagiri, Ch. (1995) The occurrence of a gene-encoded variant of nuclear basic protein (SP4) in sperm of *Xenopus laevis*. FEBS Lett., submitted.

[40] Teshima, K., Abé, S.-I., Katagiri, Ch. and Takamune, K. (1995) Relative amount of basic nuclear proteins SP4 and SP5 in *Xenopus laevis* sperm correlates with number of genes in the genome. Eur. J. Biochem., submitted.

[41] Zhou, W., Xu, J. and Golberg, E. (1994) A 60-bp core promoter sequence of murine lactate dehydrogenase C is sufficient to direct testis-specific transcription in vitro. Biol. Reprod. 51, 425–432.

[42] Iannello, R.C., Kola, I. and Dahl, H.-H.M. (1993) Temporal and tissue-specific interactions involving novel transcription factors and the proximal promoter of the mouse Pdha-2 gene. J. Biol. Chem. 268, 22581–22590.

[43] Robinson, M.O., McCarrey, J.R. and Simon, M.I. (1989) Transcriptional regulatory regions of testis-specific PGK2 defined in transgenic mice. Proc. Natl. Acad. Sci. USA 86, 8437–8441.

[44] Gebara, M.M. and McCarrey, J.R. (1992) Protein-DNA interactions associated with the onset of testis-specific expression of the mammalian Pgk-2 gene. Mol. Cell. Biol. 12, 1422–1431.

[45] Grimes, S.R., Wolfe, S.A. and Koppel, D.A. (1992) Tissue-specific binding of testis nuclear proteins to a sequence element within the promoter of the testis–specific histone H1t gene. Arch. Biochem. Biophys. 296, 402–409.

[46] Kim, Y.-J., Hwang, I., Tres, L.L., Kierszenbaum, A.L. and Chae, C.-B. (1987) Molecular cloning and differential expression of somatic and testis-specific H2B histone genes during rat spermatogenesis. Dev. Biol. 124, 23–34.

[47] Heidaran, M.A., Kozak, C.A. and Kistler, W.S. (1989) Nucleotide sequence of the *Stp-1* gene coding for rat spermatid nuclear transition protein 1 (TP1): homology with protamine P1 and assignment of the mouse *Stp-1* gene to chromosome 1. Gene 75, 39–46.