Characterization of the PL-I-Related SP2 Protein from *Xenopus*†

Lindsay J. Frehlick,‡ Adelina Prado,§ Alison Calestagne-Morelli,‡ and Juan Ausió*,‡

Department of Biochemistry and Microbiology, University of Victoria, Petch Building, 258, Victoria, B.C., Canada, V8W 3P6, and Unidad de Biofisica (CSIC-UPV/EHU) and Departamento de Bioquímica y Biología Molecular, Universidad del País Vasco, PO Box 644, 48080 Bilbao, Spain

Received June 28, 2007; Revised Manuscript Received August 30, 2007

ABSTRACT: The complete cDNA sequence of *Xenopus laevis* sperm specific proteins SP1 and SP2 has been determined. This information when taken together with N-terminal sequencing and mass spectroscopy data indicates that these two proteins share a product precursor relationship in which SP2 results from cleavage of a short N-terminal peptide of SP1. The secondary and tertiary structures of SP2 have been characterized using circular dichroism and three dimension structure prediction. These structural analyses have conclusively shown that SP1/SP2 proteins are related to proteins of the histone H1 family, particularly to vertebrate histone H1x. Hence, they can be considered *bona fide* members of the protamine-like- I (PL-I) group of sperm nuclear basic proteins (SNBPs) that have been described in other vertebrate and invertebrate groups. SP2 binds to nucleosomal DNA in a way that is very similar to that of histone H1. However, its interaction with circular DNA does not exhibit an enhanced preference for the supercoiled conformation, and it appears to be mainly driven by ionic interactions.

Chromatin, the complex of histones, nonhistone proteins, and DNA, is a dynamic structure. Alteration or remodeling of this structure is important for regulation of many cellular events. One of the most dramatic examples of chromatin remodeling occurs within male germ cells, where changes in both the protein composition and compaction of the chromatin take place. During the postmeiotic maturation of sperm (spermiogenesis), chromatin becomes supercondensed and transcriptionally inert. Highly compacted sperm chromatin is needed for a more hydrodynamic sperm head and may also protect the sperm's haploid genome from physical and chemical damage (*1*).

In sperm, DNA is associated with highly basic proteins which often differ from somatic histone proteins. The sperm-specific nuclear basic proteins (SNBPs[1]) are much more diverse than the nucleosomal proteins of somatic cells and can be grouped into three categories: protamine (P-type), protamine-like (PL-type), and histones (H-type) (*2*). P- and PL-type SNBPs are higher in arginine than the somatic histones which they replace, lending to their ability to more densely compact the spermatozoa DNA (*2*). The high levels of compaction are in part due to the fact that arginine has a higher DNA binding affinity than lysine and can bind tighter to DNA.

The PL-type proteins include a group of SNBPs with intermediate composition between the histone and the protamine type (*2*). It has now been shown that this group, and possibly the P-type as well, are structurally related to linker histones (histone H1) (*3*, *4*). Evolutionarily, PL proteins appear to have evolved from a histone H1-related PL-I protein.

While most small molecular weight PL or P protamines do not have any secondary or tertiary structure, PL-I proteins all contain a central folded domain with high similarity to the winged-helix domain, which is characteristic of the members of the histone H1 family. To date, PL-I proteins have been identified in invertebrates (*5*, *6*), chordates (*7*), and vertebrate organisms (*8*). However, low molecular weight PLs lacking the winged-helix domain have only been described in invertebrate organisms (*3*).

Mature *Xenopus laevis* sperm contains six SNBPs, referred to as sperm-specific proteins SPs (SP1 to SP6) (*9*), which are electrophoretically distinct (*10*). In addition, mature sperm chromatin contains somatic-type histones H3 and H4 but has dramatically reduced levels of H2A and H2B (*9*).

X. laevis is a well characterized model organism which has been used extensively to study the process of fertilization. A recent *in vitro* study showed that the histone chaperone nucleoplasmin was able to remove the sperm proteins of *Xenopus* and linker histones from *Xenopus* and chicken erythrocytes with similar efficiency, suggesting the possibility that these proteins may have similar structures (*11*). However, although some information is already available about the primary structure and gene relationship of the low molecular

† This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC), Grant number OGP 0046399 (to J.A.), by a NSERC PGS-D fellowship (to L.J.F), grants from the University of the Basque Country (13505/2001), MEC (BUF2004-0345/BMC), and travel fellowships from the Gobierno Vasco to A.P.

* To whom correspondence should be addressed. Phone: 250-721 8863, Fax: 250-721 8855, e-mail: jausio@uvic.ca.

‡ University of Victoria.

§ Universidad del País Vasco.

[1] Abbreviations: AU, aetic acid−urea; AUT, acetic acid−urea−triton; bp, base pair; CD, circular dichroism; EDTA, ethylenedinitrilotetraacetic acid; EM, electron microscopy; H, histone type; HPLC, high performance liquid chromatography; MALDI-TOF, matrix-assisted laser desorption ionization−time-of-flight; MRW, mean (amino acid) residue weight; P, protamine type; PAGE, polyacrylamide gel electrophoresis; PL, protamine-like; RACE, rapid amplification of cDNA ends; RD, replication dependent; RI, replication independent; SDS, sodium dodecyl sulfate; SNBP, sperm nuclear basic protein.

10.1021/bi701274s CCC: $37.00
Published on Web 10/10/2007

weight SP3−SP6 proteins, the relationship of these proteins to SNBPs of the PL type and the sequence identity of SP1/SP2 has long remained to be determined. Here we provide a detailed structural and biochemical characterization of SP2 and its interactions with circular DNA and with nucleosomes. The data show that SP2 is structurally and functionally a member of the histone H1-related PL-I family. Furthermore, we conclusively show that SP2 is a posttranslational cleavage product of SP1. These proteins share a compositional similarity with SP3−SP6. Taken together these results conclusively show that the SP proteins of *Xenopus* are members of the PL family of SNBPs and share a striking similarity to what has previously been shown in invertebrate groups such as *Mytilus* (blue mussel), which also exhibit a complex heterogeneous SNBP composition.

MATERIALS AND METHODS

Extraction and Purification of Proteins. Samples were obtained from *X. laevis* that were reared at the University of Victoria aquatics facility. SNBPs were extracted from testes with 0.4 N HCl and precipitated with acetone as described (*12*). The protein extract thus obtained was resuspended in HPLC grade distilled water and fractionated by HPLC on a reverse phase 300-Å Vydac C18 column (25 × 0.46 cm) eluted at 0.4 mL/min with a 0.1% TFA−acetonitrile gradient (*13*).

Alternatively, for binding assays and circular dichriosm experiments the sperm proteins were fractionated from sperm nuclear extracts by ion-exchange chromatography using carboxymethyl (CM) C-25-Sephadex as described elsewhere (*14*). The column was equilibrated in 1 M NaCl, 50 mM sodium acetate buffer, pH 6.7, and eluted with a 1−1.5 M NaCl linear gradient in the same buffer.

Gel Electrophoresis. Proteins were analyzed by AU-PAGE [15% acrylamide, 0.1% bis-acrylamide, 5% acetic acid, and 2.5 M urea] as explained previously (*15*), by AUT-PAGE [10% acrylamide, 0.5% bis-acrylamide, 5% acetic acid, 5.25 M urea and 5 mM Triton X-100] as described (*16*) or by SDS-PAGE [15% acrylamide, 0.4% bis-acrylamide] (*17*). Gels were then stained with 0.2% (w/v) Coomassie blue for 1 h in 25% (v/v) 2-propanol/10% (v/v) acetic acid and destained overnight in 10% (v/v) 2-propanol/10% (v/v) acetic acid.

Nucleosomes were separated by 0.7% agarose in Tris-borate-EDTA (TBE), and plasmids were separated by 1% agarose in Tris-acetate-EDTA (TAE). Agarose gels were stained with ethidium bromide and visualized with UV light.

Trypsin Digestion. SP2 was digested with trypsin (EC 3.4.21.4) (type III) (Sigma-Aldrich, St. Louis, MO). Digestions were carried out in 2 M NaCl, 25 mM Tris/HCl (pH 7.5) buffer (*18*) at an E/S ratio of 1:1000 (w:w) at room temperature. Aliquots of the digestion were collected at different times, mixed with 2× gel electrophoresis sample buffer, and immediately frozen and kept until used for AU-PAGE analysis.

Nucleosome Gel Mobility Retardation Assay. Histone octamers were obtained from hydroxyapaptite (HAP) chromatography of chicken erythrocyte chromatin (*19*). A 208 bp DNA fragment was obtained by RsaI (New England Biolabs, Ipswich, MA) digestion of a 208−12 oligomer consisting of 12 tandemly arranged fragments of from the 5S rRNA gene of the sea urchin *Lytechinus variegatu*.

Nucleosomes were reconstructed by histone octamer assembly onto the 208 bp DNA template by stepwise salt dialysis (*20*). The histone:DNA molar ratio was 1:1.

Aliquots of 300 ng of reconstituted nucleosome particles (in 50 mM NaCl, 10 mM Tris, and 1 mM EDTA) were incubated with increasing amounts of linker histone H1 following a protocol modified from (*21*). After incubation at room temperature for 30 min, 30% sucrose was added to bring the samples to a final 5% sucrose concentration and the samples were then loaded onto a 0.7% agarose gel in 0.5× TBE.

Plasmid Gel Mobility Retardation Assay. Plasmid pBR322 was purified using the QIAprep spin miniprep kit (Qiagen, Misissauga, ON) from *E. coli* grown at 37 °C overnight in LB broth, following the manufacturer's protocol. For the linear plasmid, pBR322 was digested with BamHI (New England Biolab, Ipswich, MA). The reaction was cleaned up using the QIAquick PCR purification kit. DNA concentrations were determined on a Nanodrop spectrometer (Nanodrop Technologies, Wilmington, DE) using an extinction coefficient of 20 mL cm^{-1} mg^{-1} at 260 nm.

Increasing amounts of SP2 or chicken H5 were incubated with 0.5 μg of pBR322, in a total reaction volume of 20 μL and final buffer concentration of 10 mM Tris-HCl (pH 8), 0.2 mM EDTA, 0.02% Triton X-100, and 4 mM NaCl, as described (*22*). The samples were incubated at room temperature for 1 h, 30% sucrose was added to give a final concentration of 5%, and the samples were loaded onto a 1% agarose gel in TAE.

Circular Dichroism and UV Spectroscopy. Circular dichroism (CD) spectroscopy experiments were carried out at 20 °C on a Jasco model J720 spectropolarimeter. Spectra were acquired from 200 to 260 nm using a bandwidth of 1 nm and data pitch of 0.1 nm at a scan speed of 2 nm/min and 1 accumulation per second. Spectra were corrected for solvent contribution, and the CD signal was converted to molar ellipticity [θ], using the formula $[\theta] = \mathrm{CD}\,(\mathrm{MRW}/10 \times l \times c)$, where MRW is the mean residue weight (the molecular mass divided by the number of peptide bonds), l is the path length of the cell, and c is concentration in mg/mL. The extinction coefficients in water at 230 nm were determined by amino acid analysis using norleucine as an internal standard. The values thus obtained were 2.35, 3.03, 4.53, and 8.1 cm^2 mg^{-1} for H5, SP2, SP3−5, and SP6, respectively. The value determined in this way for histone H5 is in very strong agreement with the previously reported value of 2.34 cm^2 mg^{-1} (48 000 M^{-1} cm^{-1}) (*23*) using the same experimental approach. The absorbances were measured on a Cary spectrophotometer (Varian, Palo Alto, CA), and the change in absorbance of the proteins in 20 mM sodium phosphate pH 7.2 and in 10 mM Tris-HCl/0.1 M NaCl pH 7.5 were used to correct the extinction coefficients for buffer effects. On average the extinction coefficients were 15% lower in buffer than in water.

Because histones exhibit a very low β sheet structure, the percentiles of α helix were estimated from the values of the ellipticity [θ] at 220 nm (*24*) using the equation,

$$\% \alpha \text{ helix} = 8.9 - (2.47 \times [\theta]_{220}\,\text{nm}) \times 10^{-3}$$

and at 208 nm (*25*) using the equation,

$$\% \alpha \text{ helix} = [([\theta]_{220}\,\text{nm} - 4000)/(-33000 - 4000)] \times 100$$

N-Terminal Sequencing and Mass Spectroscopy. The N-terminal sequence and partial peptide sequence was obtained using SP2 purified by RP-HPLC. The N-terminal protein sequence was determined by conventional Edman degradation using an ABI Precise protein microsequencer (Applied Biosystems, Foster City, CA), as described previously (*26*). The partial peptide sequence was obtained from Glutamic endopeptidase digestion of the SP2 protein. The resulting peptides were separated by reverse phase HPLC, and the most prominent peptide was sequenced by electrospray quadrupole time-of-flight (Q-TOF) mass spectroscopy.

Molecular masses were determined by mass spectroscopy analysis of *X. laevis* SP1 and SP2 carried out by MALDI-TOF on a Voyager Linear DE (PerSpective Biosystems Inc., Foster City, CA) using a sinapinic acid matrix following a previously described protocol (*27*).

cDNA Sequence Determination. Total RNA from testes was extracted using Trizol reagent (GibcoBRL, Burlington, ON), and mRNA from total RNA was isolated using an mRNA purification kit (Amersham Bioscience, Piscataway, NJ).

The following primers were designed using protein sequence data and cDNA sequences with similarity to SP2 (such as *X. laevis* histone H1x, GeneBank accession number AAH41758):

Forward 1: CAGCCGGGCMRSTACAG
Forward 2: CAGGAACGGCTCGTCCCT
Reverse 1: GTASYKGCCCGGCTGGTT
Reverse 2: GGGACGAGCCGTTCCTCT Using these primers the complete cDNA sequence was amplified using the First Choice Rapid Amplification of *c*DNA Ends (RACE) Kit (Ambion, Austin, TX).

For cDNA sequencing, agarose gel-purified PCR products were ligated into pCR2.1-TOPO vectors (Invitrogen, Burlington, ON) following the instructions of the manufacturer and transformed into TOP10 competent cells (Invitrogen, Burlington, ON). The plasmids were purified with the QIAprep Miniprep kit (Qiagen, Mississauga, ON), and sequencing of the inserts was done by the DNA Sequencing Facility, Centre for Biomedical Research at the University of Victoria. The *X. laevis* SP2 sequence determined was aligned with similar protein sequences using the CLUSTAL_X (*28*) and BIOEDIT programs (*29*) with the default parameters.

Protein Structure. The secondary structures were predicted from the protein sequences using PROFsec (*30*) on the PredictProtein server (*31*). The three-dimensional structure of *X. laevis* SP2 was modeled using the coordinates determined from the crystal structure of the globular core of the chicken erythrocyte histone H5 (*32*) as a reference, using the SWISS-MODEL server (*33*).

RESULTS

The Primary Structure of X. laevis SP1 and SP2. It had been previously revealed that SP2 was processed from a precursor which was probably SP1, as a pulse-chase experiment showed SP2 increased while the SP1 protein band decreased concomitantly (*10*). There was also indirect evidence, based on a partial N-terminal sequence, that suggested that SP1 and SP2 shared a precursor−mature protein relationship (*34*). Here, SP1 and SP2 were purified from other basic proteins of *X. laevis* testes by either reverse phase HPLC (Figure 1A) or cation exchange chromatography (CM Sephedex C25) (Figure 1B). The high absorbance at 230 nm of peak 2 of the cation-exchange column (Figure 1B) was caused by the peptides of the protease inhibitor (shaded area), which eluted in the same position as SP2.

As can be seen in Figure 1A, reverse-phase HPLC allowed us to separate SP1 (lane 4) from SP2 (lane 5). The proteins thus obtained were used in N-terminal and mass spectroscopy sequencing (Figure 2), and their masses were determined by MALDI-TOF (Table 1). The results provided the N-terminal sequence for SP2 and some internal peptide sequence information. Attempts to determine the N-terminal sequence of SP1 failed, probably because of N-terminal blocking, as has been observed with other SNBPs (*8*). Alanine at position 2 of the amino acid sequence had a 83% likelihood of being acetylated as predicted by the Terminator program at http://www.isv.cnrs-gif.fr/terminator2/index.html (*35*), which supports this hypothesis.

In order to obtain the complete primary structure of these proteins and establish the precursor−product relationship between them, we cloned the cDNA (see Figure 2 for the cDNA and translated protein sequence). The N-terminal sequence of SP2 did not start at the initial methionine. The size difference between the N-terminal sequence of SP2 and the starting methionine matched the mass difference between SP1 and SP2 as determined by MALDI-TOF (Table 1). These results, along with the fact that the N-terminus of SP1 was blocked, indicated that the cleavage site on SP1 is at the N-terminus (shown by a triangle in Figure 2B).

The Secondary and Tertiary Structure of SP1/SP2. To further characterize the structure of the protein, SP2 was analyzed by CD to determine the secondary structure composition. SP2 appears to have a structural organization similar to histone H5 (Figure 3A), as can be seen by the shape of the CD spectra, with lows at 208 and 222 nm characteristic of α helices (*25*, *36*). However, the elipticities at 208 and 222 nm of the SP2 curves were not as low as H5, suggesting less α helical content. The calculated α helical content estimated for SP2 was 15.7% and 17.6% and for H5 was 18.7% and 20.3%, for Tris and phosphate buffers, respectively. The α helical content of H5 determined in this way is in good agreement with that of the 19.5% determined from the crystallographic analysis of the winged helix domains of this protein (*32*) and discerns that this is the only structural region in solution. The α helical content estimated from the predicted secondary structure analysis was 31.8% for SP2 and 20% for H5 (Figure 3C). The high predicted α helical content of SP2, compared to that of the calculated content, is most likely due to the fact that the four helices predicted for the tail domain would not likely be folded in solution. The basic charges of the lysine and arginine residues in this region would disrupt the helical fragments and would likely need to be neutralized by binding DNA to form any helical structure (*24*). Also, the prediction of these helices only had a reliability of 70%, relative to 100% for the helices

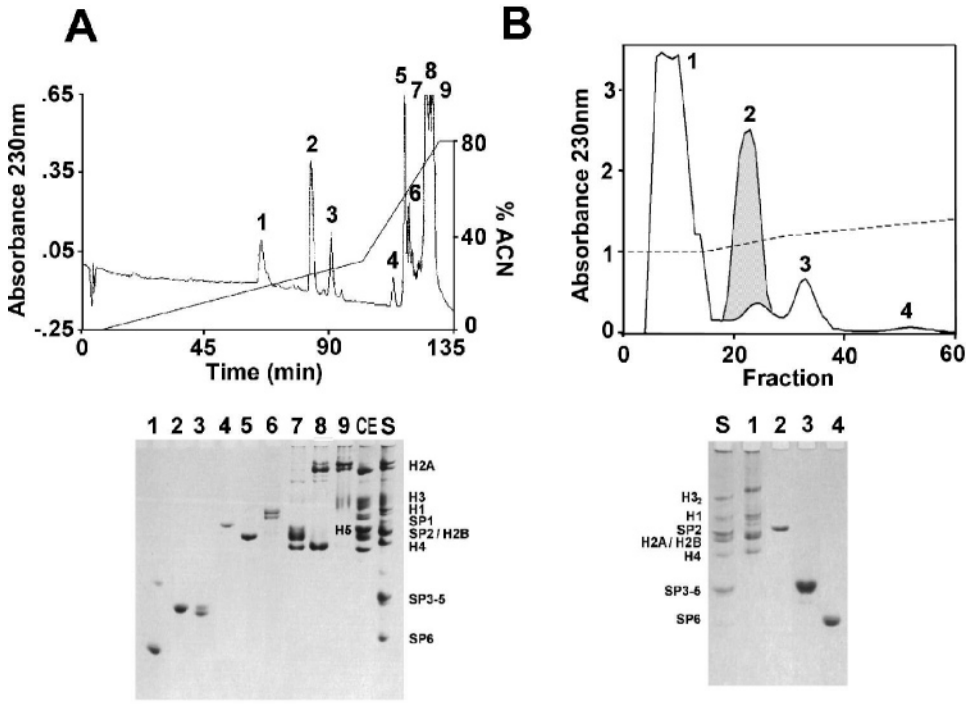

FIGURE 1: Fractionation and purification of *X. laevis* SNBPs. (A) Reversed phase HPLC. Proteins were eluted with an acetonitrile (ACN) gradient. Below is an AUT-PAGE analysis of the indicated peaks from A (1−9). CE, chicken erythrocyte histones; S, starting sample loaded on the HPLC. (B) Ion-exchange [carboxymethyl (CM) C-25 Sephadex] chromatography. The solid line denotes the absorbance at 230 nm, and the dashed line denotes the 1 M to 1.4 M NaCl gradient in 50 mM sodium acetate buffer (pH 6.7). The gel below shows AU-PAGE analysis of the indicated peaks from A (1−4). $H3_2$ indicates an H3 dimer.

in the globular regions. The fact that the SP2 C-terminal tail does not form helices in solution, as well as the bias toward hydrophilic amino acids, indicates that it is likely intrinsically disordered, similar to the C-terminal tail of mouse H1o (*37*).

For comparison, the mixed SP3, SP4, and SP5 fraction and the SP6 protein were also analyzed by CD. The spectra of these proteins is shown in Figure 3B. The amount of α helix determined for these was 15% for the SP3−5 mixture and 26% for SP6, which, despite the shorter amino acid sequence of these proteins, is somewhat similar to that of SP2 and H5. As with SP2, the predicted secondary structure, which consists predominantly of α helix and random coil (Figure 3C), was higher: approximately 28% for SP3−5 and 39% for SP6.

The alignment of the primary structures shown indicates that SP2 shares a substantial amount of sequence similarity with the core domain of linker histones and other PL-Is (Figure 4A). One of the structural signatures of the protein members of the H1 family of linker histones is the presence of a winged-helix domain (*32*), which exhibits trypsin digestion resistance (*38*). To check that this was also the case with SP2, the protein was digested with trypsin in the presence of 2 M NaCl. As can be seen in Figure 4B, and similar to what has been observed with other histone H1-related SNBPs (*6*, *8*, *18*), a trypsin-resistant peptide was observed. Given the sequence similarity to the globular region of H5 (Figure 4A) and the fact that the secondary structure predictions yielded three helices in this region for SP2, which matched the size and spacing of those in H5 (Figure 3C), it is not surprising that the tertiary structure modeling gave a 3D structure that is an almost perfect match to the winged helix of H5 (Figure 4C).

Characterization of the Interactions of SP2 with DNA and Nucleosomes. Given of the structural similarities between SP2 and histone H5 (Figure 4), and their relation to the members of the histone H1 family, we decided to analyze and compare the binding of these two proteins to DNA and nucleosomes. The presence of a winged helix domain in linker histones imparts them with a preferential binding to DNA cruciform (four-way junction) structures (*39*−*41*), which is reflected by their higher affinity for circular supercoiled DNA and within a more physiological relevant context, to the nucleosome.

Similar to histone H1 (*22*, *42*), histone H5 was found to preferentially bind to supercoiled DNA over linear or relaxed circular DNA (Figure 5A, lanes 9−12). This is shown by the fact that the supercoiled plasmid (lower band) shifts before the linear or relaxed circular DNA. In addition, the protein to DNA ratios (w:w) at which the plasmids shifted were similar to that of H1 (*22*). A complete shift was observed with circular supercoiled DNA at a protein:DNA ratio of 0.25−0.5 (w:w) for both histone H1 (*22*) and histone H5 (Figure 5A, lanes 5−8). In contrast, approximately 5 times the ratio of SP2:DNA compared to that of H5:DNA (1.5 w:w versus 0.3 w:w) was required to shift the plasmids. Another difference between the shifts of H5 and SP2 is that, like in the case of histone H1 (*22*), H5 appears to bind linear DNA nonspecifically, resulting in a sudden aggregation of the DNA (Figure 5A, lanes 1−4). This is in contrast to the supercoiled band, which shifts in a regular manner as the amount of H5 is increased. For our SP2 protein, both the supercoiled and linear DNA exhibits this regular shift, albeit the linear DNA did aggregate at a lower ratio (1−1.5 w:w) than the supercoiled DNA (1.5−2.0 w:w). Although SP2

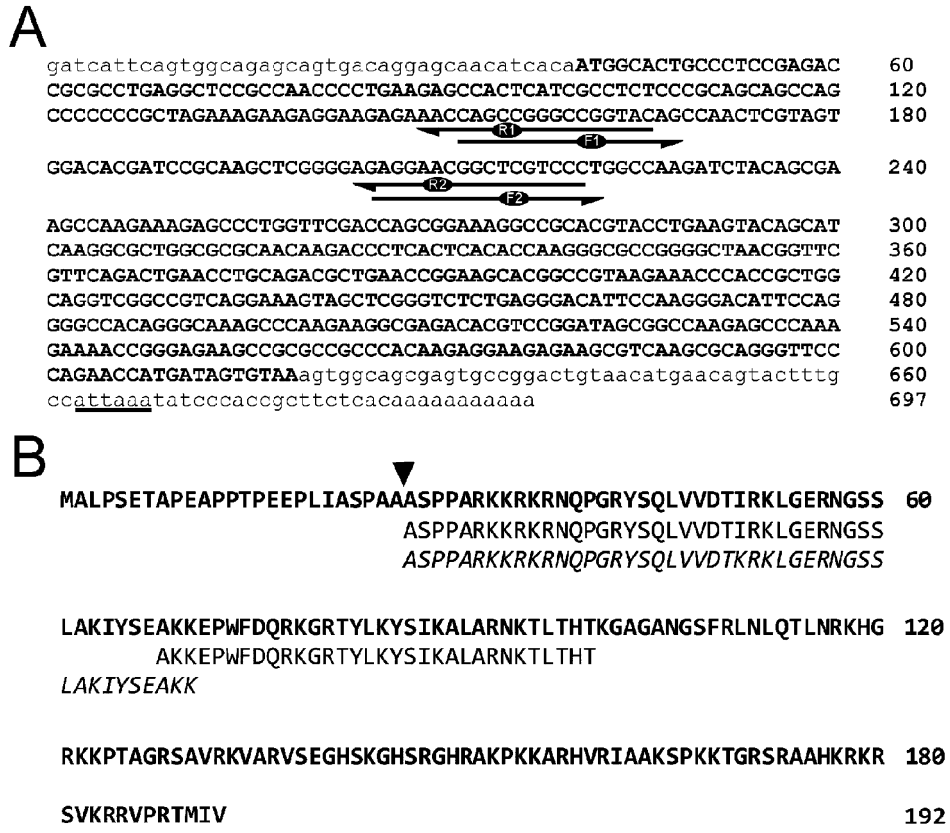

FIGURE 2: *X. laevis* SP1/SP2 sequence. (A) The complete SP1/SP2 cDNA sequence determined from RACE PCR (Genebank accession no. 920865). The 5′ and 3′ untranslated regions are in lower case, and the coding region is in bold capitals. A putative polyadenylation sequence is underlined, and the primers used for RACE are indicated by arrows. (B) The amino acid sequence translated from the SP1/SP2 cDNA is shown in bold. Below the full sequence are the partial peptide sequences confirmed by N-terminal sequencing and mass spectroscopy and in italics the N-terminal sequence previously determined by Ariyoshi et al. (*34*). The predicted cleavage site on the SP1 precursor is shown with an arrowhead.

Table 1: The Predicted Protein Size Compared to the Mass Determined from MALDI-TOF Mass Spectroscopy

protein	no. of amino acids	predicted size (kDa)	size from MALDI-TOF (kDa)
SP1	191	21311[a]	21327
SP2	167	18930	18941

[a] This value is based on the N-terminal methionine being cleaved and the alanine being acetylated (addition of 42kDa).

seems to bind the supercoiled DNA first (Figure 5B, lane 14), at higher ratios both the linear and supercoiled DNA experience a comparable shift (Figure 5B, lanes 15 and 16).

The different ability of SP2 to bind to supercoiled DNA compared to somatic linker histones (H1 and H5) contrasts with the similar ability of these two proteins to bind to nucleosomal DNA (Figure 6). Furthermore, discrete nucleosomal-histone H1 complexes are formed at very similar input ratios (Figure 6).

X. laevis SP Proteins Are Genuine Members of the Sperm Nuclear Basic Proteins of the PL Type. The relationship of SP1/SP2 with members of the histone H1 family clearly identifies these proteins as belonging to the protamine-like (PL-I) type of SNBPs (*2*, *3*). PL-proteins and H1 histones are closely evolutionarily related and are descendants of a common ancient "orphon" group of H1-replication-dependent histones (*43*). PL proteins have been described both in invertebrate (*18*, *26*, *44*, *45*) and chordate groups (*8*, *46*−*48*). Interestingly, in both instances a precursor-product relationship involving protein posttranslational cleavage of the N-terminal (*45*) or C-terminal domain (*49*) of these proteins has been described. For example, N-terminal cleavage of the *X. laevis* SP1 precursor yields the SP2 protein (as shown in this paper), and in mussel, cleavage at a C-terminal cut site in the PL-I precursor yields both PL-II and PL-IV (Figure 7). However, the cleavage site (SPAA*ASP) at the N-terminal region of SP1 has no sequence similarity to the proteolytic processing sites of invertebrates (NKSNN*AK) whether they occur at the N- or C-terminal regions of these proteins (Figure 7B) (*49*, *50*).

DISCUSSION

Xenopus SP2 Is the Cleavage Product from an SP1 Precursor. Nucleotide sequences are available for SP4 and SP5 (*34*, *51*) and partial N-terminal peptide sequences are available for SP1−6 (*34*). Cloning of SP4 and -5 and N-terminal sequence information indicates that the six SPs are derived from three different mRNA species. SP3, -4, and -6 are derived from one mRNA and SP5 from a separate mRNA (*34*). These mRNAs are polyadenylated and likely replication-independent and expressed in primary spermatocytes. From the available N-terminal protein sequence data SP1 and -2 are suspected to be encoded by a common mRNA species (*34*). However, no nucleotide sequence is currently available to support this notion.

The results provided in the preceding section show that a precursor-product relationship exist between SP1 and SP2 where SP2 is the result of posttranslational cleavage of SP1

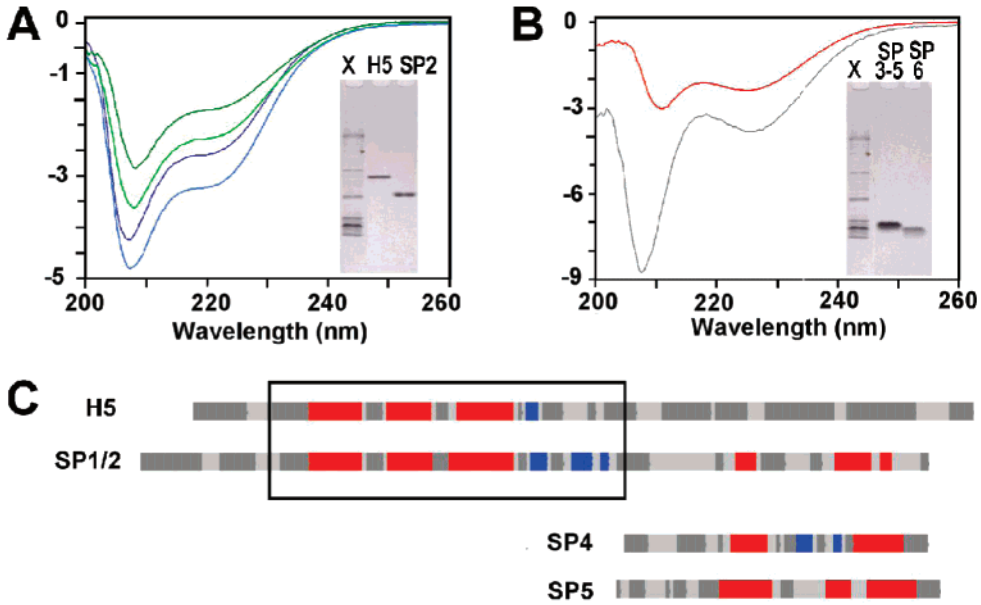

FIGURE 3: (A) CD spectra of *X. laevis* SP2 (green) and chicken H5 (blue) in 10 mM Tris/0.1 M NaCl buffer (pH 7.5) (dark green and blue) and 20 mM phosphate buffer (pH 7.2) (light green and blue). (B) CD spectra of *X. laevis* the SP3, SP4, and SP5 fraction (red) and SP6 (gray), both in 20 mM phosphate buffer (pH 7.2). The inserts show the SDS-PAGE of the proteins used in the analysis: Xl is *X. laevis* SNBPs. (C) The predicted secondary structures of H5, SP1/SP2, SP4, and SP5. Red, helix; blue, extended (sheet); dark gray, other (loop); light gray, no prediction is made for these residues, as the reliability is to low (<5). The globular domains of H5 and SP1/SP2 are within the black box.

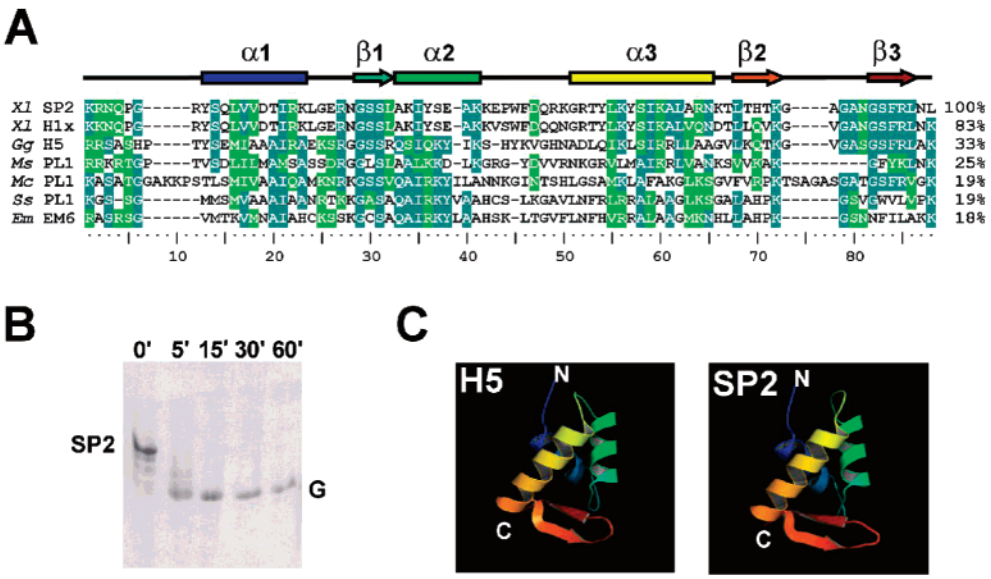

FIGURE 4: SP2 contains a trypsin-resistant winged-helix motif. (A) The schematic secondary structure of the globular winged-helix domain is above the corresponding amino acid sequences. β-Turns and strands are indicated by arrows, and α-helices are indicated by boxes. Sequence alignment of the amino acid region corresponding to the globular winged-helix domain in SP2 in comparison to other protein members of the histone H1 family. The light green shading indicates similar amino acids and the teal shading indicates identical amino acids, in at least 50% of the sequences compared. The sequences and Genebank accession numbers are: *Xl* SP2, *X. laevis* SP2; *Xl* H1x, *X. laevis* histone H1x (AAH41758); *Gg* H5, *Gallus gallus* histone H5 (NP001038138); *Ms* PLI, *Mullus surmuletus* PL-I (Q08GK9); *Mc* PLI, *Mytilus californianus* PL-I; *Ss* PLI, *Spisula solidissima* PL-I (AAT45384); Em EM6, *Ensis minor* EM6 (AAA98076). The numbers at the right-hand side designate percent identity. (B) AU-PAGE analysis of the time course trypsin digestion of SP2, carried out in the presence of 2 M NaCl at room temperature. G, the resistant globular core of the protein. The digestion times (0, 5, 15, 30 and 60 min) are indicated on top of the lanes. (C) Tertiary structure of the trypsin-resistant core of chicken erythrocyte H5 obtained from the crystallographic data determined in ref *32*. The H5 structure was used as a template to model the tertiary structure of the trypsin-resistant core of SP2.

with removal of the first 25 N-terminal amino acids. A precursor product relationship between *X. laevis* SP1 and SP2 had been earlier demonstrated by Katagiri and co-workers using [^{14}C]arginine -[^{14}C]lysine incorporation during spermiogenesis (*10*). However, on the basis of N-terminal sequencing of the two proteins, it was concluded that the protein processing must have taken place at the C-terminal end of SP1 (*34*). The disagreement regarding the site of cleavage is most likely due to the method (AUT gel purification and extraction of the bands with 0.4 N SO_4H_2) as reported previously (*34*). As we have shown, the N-terminal end of SP1 appears to be blocked and contamination and/or partial cleavage by the method of extraction used could have led to the misidentification of the SP1 N-terminus

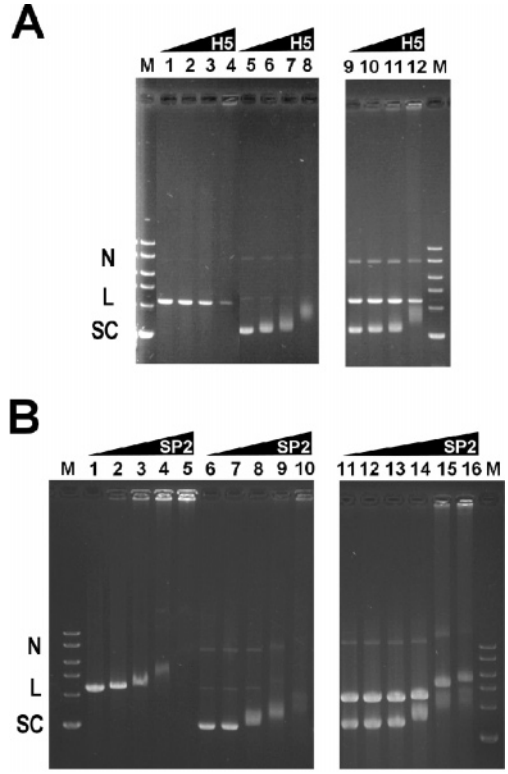

FIGURE 5: Gel mobility retardation assay comparing the binding of H5 and SP2 to supercoiled and linear pBR322 plasmid DNA. (A) Increasing amounts of H5 were incubated with 0.5 μg of BamH1 cut pBR322 (lanes 1−4), uncut pBR322 (lanes 5−8), or both uncut and cut plasmid (lanes 9−12). After incubation, the resulting complexes were analyzed on a 1% agarose gel. The H5 to plasmid ratios (w:w) were 0, 0.075, 0.15, and 0.3 for lanes 1−4, 5−8, and 9−12, respectively. (B) Increasing amounts of SP2 were incubated with 0.5 μg of BamH1 cut pBR322 (lanes 1−5), uncut pBR322 (lanes 6−10), or both uncut and cut plasmid (lanes 11−15). The SP2 to plasmid ratios (w:w) were 0, 0.5, 1, 1.5, and 2 from lanes 1−5, 6−10, and 11−15, respectively; lane 16 had a ratio of 2.5. A 1 kilobase marker (M) was loaded with bands of 10, 8, 6, 5, and 3 kb from the top.

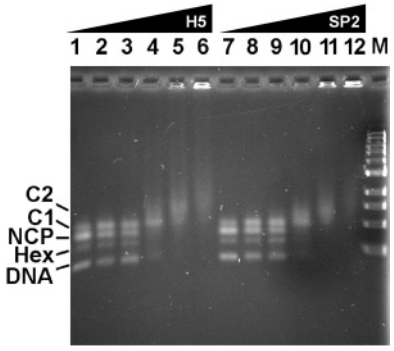

FIGURE 6: Gel mobility retardation assay comparing the binding of H5 and SP2 to nucleosomes. Increasing amounts of H5 were incubated with 300 μg of nucleosomes and analyzed on a 0.8% agarose gel. The H5 to nucleosome ratios (mol:mol) were 0, 1, 2, 4, 6, and 8 for lanes 1−6. Increasing amounts of SP2 were incubated with 300 μg of nucleosomes. The SP2 to plasmid ratios (mol:mol) were 0, 1, 2, 4, 6, and 8 for lanes 7−12. A 1 kilobase marker (M) was loaded with bands of 10, 8, 6, 5, and 3 kb from the top. C1, chromatosome with one linker protein; C2 chromatosome with two linker proteins; NCP, nucleosome core particle; Hex, hexamer; DNA, free DNA.

(*34*). The process that leads to the N-terminal cleavage observed by us is currently unknown but is highly reminiscent of the N-terminal processing undergone by some

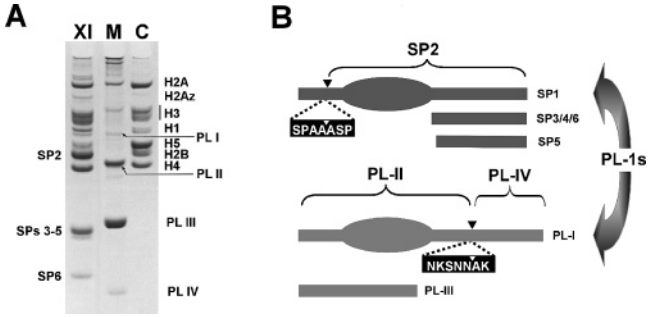

FIGURE 7: (A) The SP proteins from *X. laevis* are structurally related to the SNBPs of the PL-type. Acid-urea-Triton (AUT)-PAGE characterization of SNBPs from *X. laevis* (Xl) and *Mytilus californianus* (M) and a histone marker from chicken erythrocytes (CE). PL, protamine-like. (B) Schematic representation of sperm protein structure in *X. laevis* (SP1 to SP6) and *Mytilus* (PL-I, -II, and -III), showing the globular (trypsin-resistant) portions (oval) and tails (rectangular). The predicted cleavage sites are shown with arrowheads and the corresponding sequences are shown below the sites. See text for details.

vertebrate and invertebrate protamines where a terminal portion of the protein of similar length is removed during spermiogenesis (see ref *52* for a review).

Sperm-Specific SP1/SP2 Proteins Are Closely Related to Vertebrate Histone H1x. The protein sequence alignment shown in Figure 4 indicates that, of all the histone H1 proteins, SP1/SP2 exhibits the highest similarity with H1x. Histone H1x has been recently described to be present in mammalian cells and to be closely related to the replacement subtype H1o. The genes for these two linker histones are orphons and like SP1/SP2 are transcribed into polyadenylated mRNA (*53*). Interestingly, human H1x is present in chromatin regions that are resilient to micrococcal nuclease digestion, suggesting that they enhance chromatin compaction (*53*) and accumulate in the nucleolus during G1 phase at inactive ribosomal genes (*54*).

Although SP1/SP2 shows a high similarity with H1x proteins within the core domain, there is greater variability in the N- and C-tail regions. Following cleavage of the SP1 precursor, the resulting SP2 protein has only a short highly basic C-terminal tail, and together the C- and N-terminal tail domains of SP2 are higher in arginine than a typical H1. This high arginine content is clearly seen when comparing the amino acid composition of *X. laevis* SP2 to *X. laevis* H1x. SP2 has 15.6% arginine whereas H1x only has 4.1% arginine. This increase in arginine is at the expense of lysine, as SP2 has only 15.6% lysine and H1x has a 21.2% lysine content. The high arginine content is characteristic of PL and P proteins (*2*).

The chromatin compacting role of the histone SP1/SP2-related histone H1s is in functional agreement with the role of SP proteins in the condensation of *X. laevis* chromatin during spermiogenesis. The molecular mechanisms by which SP2 contributes to the highly condensed chromatin organization of the mature sperm nucleus are not clear. Our results indicate that SP2 can bind to nucleosomes in a similar way to histone H5 and other related histone H1 proteins. Such binding does not appear to be impaired by the partial depletion of H2A-H2B dimers (see Figure 6 Hex) which is present to a higher extent in native *X. laevis* sperm chromatin (*9*). However, in contrast to histone H1 (*22*) or to histone H5, for SP2 no preferential binding for supercoiled DNA is

observed (Figure 5). This suggests that, as in the case of other vertebrate PL-I proteins which bind where nucleosomes are nonexistent (*8*) and in the case of protamines (*52*), the interaction of SP2 with DNA is electrostatically driven. These results are in good agreement with the previous *in vitro* findings that *Xenopus* nucleoplasmin removed SP2 with similar efficiency to the removal of H1/H5 (*11*), suggesting that SP2 could be arranged like an H1 on chromatin. The data in this current paper go further to show that indeed SP2 has a similar structure to H1/H5.

SP2 Interaction with DNA Is Electrostatically Driven but Is Able To Form Chromatosomes. The molecular weights of H5 and SP2 are 20.6 kDa (189 amino acids) and 18.0 kDa (170 amino acids), respectively, and their basic amino acid compositions are 31.5% and 35%. From this it is possible to calculate that in both instances charge saturation (mol − ve/mol + ve) in terms of their binding to DNA is achieved at 1:1 w:w ratios. However, while H1 and H5 can shift circular supercoiled DNA at a ratio of 0.3−0.5 w:w below the charge neutralization, in SP2 a shift is only observed above charge saturation, approximately 5-fold higher w:w. This indicates that SP2 binds nonspecifically to circular supercoiled DNA, and this binding appears to be electrostatically driven. However, the ability of SP2 to produce chromatosome structures (Figure 6) shows that this electrostatic component does not impede the protein from assembling into nucleosome substrates to form chromatosomes containing 1−2 molecules of SP2 (Figure 6, C1 and C2).

These results lead us to hypothesis that SP2 can bind to the H2A/H2B-deficient nucleosomes (*9*) in the *Xenopus* sperm chromatin through its winged-helix motif, while using its N- and C-terminal tails to neutralize the charge in the linker domains. *In vivo* this would be achieved in conjunction with SP3−SP6 to create a highly compacted sperm chromatin organization (*55*).

Evolutionary Implications. In addition to the biochemical information on amphibian PL-I-related proteins, this work has some evolutionary relevance. Despite the clear relationship between SP1/SP2 and PL-I proteins, the origin of *X. laevis* SP3−SP6 and their relation to other proteins of the PL type is not as clear. In some *Xenopus* species, such as *X. tropicalis*, these smaller sperm-specific proteins SP3−6 are absent. This absence also occurs in some molluscs. For instance, in *Mytilus*, SNBPs contain a small PL-III in addition to PL-I (PL-II/PL-IV) (Figure 7); however, the former is absent in other species such as the surf clam *Spisula solidissima* (*7*), whereas Mytilus PL-III can be distinctively related to the N-terminal region of some PL-I proteins (*4*), a blast search for proteins similar to the full length precursor SP4 or SP6 produced only matches to themselves or each other. Interestingly, in the case of *X. laevis*, there is a trend toward an increase in arginine at the expense of lysine in going from the PL-I like SP2 to SP4 (precursor to SP3, SP4, and SP6) and SP5, which are more similar to the protamine group of SNBPs (*34*, *51*). This trend is reminiscent of the K to R transition that has been observed in tunicates (*46*). There is also a substantial decrease in size between SP1/SP2 and SP3−SP6, which is another characteristic feature of the protamine group of SNBPs (*2*). These trends lend support to the hypothesis that the smaller forms of PLs or SPs represent an early attempt in the gene transition from H1-related PL-I proteins to the protamines.

The coexistence of PL proteins containing (PL-I, SP2) and lacking (PL-III, SP3−SP6) the winged-helix motif appears to have occurred repeatedly in different phylogenetic groups (*2*). The lack of a structural relationship between the shorter PLs in these different groups provides support to the notion that this is most likely the result of a process of convergent evolution which is constrained by the structural requirement needed to achieve sperm chromatin compaction. This would also apply to the seemingly unrelated posttranslational cleavage of these proteins in different unrelated organisms (Figure 7).

REFERENCES

1. Braun, R. E. (2001) Packaging paternal chromosomes with protamine, *Nat. Genet. 28*, 10−12.
2. Ausió, J. (1999) Histone H1 and evolution of sperm nuclear basic proteins, *J. Biol. Chem. 274*, 31115−31118.
3. Eirin-Lopez, J. M., Frehlick, L. J., and Ausio, J. (2006) Protamines, in the footsteps of linker histone evolution, *J. Biol. Chem. 281*, 1−4.
4. Eirin-Lopez, J. M., Lewis, J. D., Howe, L. A., and Ausio, J. (2006) Common phylogenetic origin of protamine-like (PL) proteins and histone H1: Evidence from bivalve PL genes, *Mol. Biol. Evol. 23*, 1304−1317.
5. Zhang, F., Lewis, J. D., and Ausió, J. (1999) Cysteine-containing histone H1-like (PL-I) proteins of sperm, *Mol. Reprod. Dev. 54*, 402−409.
6. Jutglar, L., Borrell, J. I., and Ausió, J. (1991) Primary, secondary, and tertiary structure of the core of a histone H1-like protein from the sperm of *Mytilus*, *J. Biol. Chem. 266*, 8184−8191.
7. Lewis, J. D., McParland, R., and Ausio, J. (2004) PL-I of *Spisula solidissima*, a highly elongated sperm-specific histone H1, *Biochemistry 43*, 7766−7775.
8. Saperas, N., Chiva, M., Casas, M. T., Campos, J. L., Eirin-Lopez, J. M., Frehlick, L. J., Prieto, C., Subirana, J. A., and Ausio, J. (2006) A unique vertebrate histone H1-related protamine-like protein results in an unusual sperm chromatin organization, *FEBS J 273*, 4548−4561.
9. Mann, M., Risley, M. S., Eckhardt, R. A., and Kasinsky, H. E. (1982) Characterization of spermatid/sperm basic chromosomal proteins in the genus *Xenopus* (Anura, Pipidae), *J. Exp. Zool. 222*, 173−186.
10. Abe, S., and Hiyoshi, H. (1991) Synthesis of sperm-specific basic nuclear proteins (SPs) in cultured spermatids from *Xenopus laevis*, *Exp. Cell Res. 194*, 90−94.
11. Ramos, I., Prado, A., Finn, R. M., Muga, A., and Ausio, J. (2005) Nucleoplasmin-mediated unfolding of chromatin involves the displacement of linker-associated chromatin proteins, *Biochemistry 44*, 8274−8281.
12. Wang, X., and Ausió, J. (2001) Histones are the major chromosomal protein components of the sperm of the nemerteans *Cerebratulus californiensis* and *Cerebratulus lacteus*, *J. Exp. Zool. 290*, 431−436.
13. Ausió, J. (1988) An unusual cysteine-containing histone H1-like protein and two protamine-like proteins are the major nuclear proteins of the sperm of the bivalve mollusc *Macoma nasuta*, *J. Biol. Chem. 263*, 10141−10150.
14. Ausió, J., and Subirana, J. A. (1982) Conformational study and determination of the molecular weight of highly charged basic proteins by sedimentation equilibrium and gel electrophoresis, *Biochemistry 21*, 5910−8.
15. Ausio, J. (1992) Presence of a highly specific histone H1-like protein in the chromatin of the sperm of the bivalve mollusks, *Mol. Cell. Biochem. 115*, 163−172.
16. Frehlick, L. J., Eirin-Lopez, J. M., Field, E. D., Hunt, D. F., and Ausio, J. (2006) The characterization of amphibian nucleoplasmins yields new insight into their role in sperm chromatin remodeling, *BMC Genomics 7*, 99.
17. Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4, *Nature 227*, 680−685.
18. Ausió, J., Toumadje, A., McParland, R., Becker, R. R., Johnson, W. C., and van Holde, K. E. (1987) Structural characterization of

the trypsin-resistant core in the nuclear sperm-specific protein from *Spisula solidissima*, *Biochemistry 26*, 975−982.
19. Ausió, J., and Moore, S. C. (1998) Reconstitution of chromatin complexes from high-performance liquid chromatography-purified histones, *Methods (San Diego, Calif.) 15*, 333−342.
20. Tatchell, K., and Van Holde, K. E. (1977) Reconstitution of chromatin core particles, *Biochemistry 16*, 5295−5303.
21. Sera, T., and Wolffe, A. P. (1998) Role of histone H1 as an architectural determinant of chromatin structure and as a specific repressor of transcription on *Xenopus* oocyte 5S rRNA genes, *Mol. Cell Biol. 18*, 3668−3680.
22. Ellen, T. P., and van Holde, K. E. (2004) Linker histone interaction shows divalent character with both supercoiled and linear DNA, *Biochemistry 43*, 7867−7872.
23. Carter, G. J., and van Holde, K. (1998) Self-association of linker histone H5 and of its globular domain: evidence for specific self-contacts, *Biochemistry 37*, 12477−12488.
24. Verdaguer, N., Perello, M., Palau, J., and Subirana, J. A. (1993) Helical structure of basic proteins from spermatozoa. Comparison with model peptides, *Eur. J. Biochem. 214*, 879−887.
25. Greenfield, N., and Fasman, G. D. (1969) Computed circular dichroism spectra for the evaluation of protein conformation, *Biochemistry 8*, 4108−4116.
26. Carlos, S., Jutglar, L., Borrell, I., Hunt, D. F., and Ausió, J. (1993) Sequence and characterization of a sperm-specific histone H1-like protein of *Mytilus californianus*, *J. Biol. Chem. 268*, 185−194.
27. Hunt, J. G., Kasinsky, H. E., Elsey, R. M., Wright, C. L., Rice, P., Bell, J. E., Sharp, D. J., Kiss, A. J., Hunt, D. F., Arnott, D. P., Russ, M. M., Shabanowitz, J., and Ausió, J. (1996) Protamines of reptiles, *J. Biol. Chem. 271*, 23547−23557.
28. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., and Higgins, D. G. (1997) The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools, *Nucleic Acids Res. 25*, 4876−4882.
29. Hall, T. A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, *Nucl. Acids Symp. Ser. 41*, 95−98.
30. Rost, B., and Sander, C. (1993) Prediction of protein secondary structure at better than 70% accuracy, *J. Mol. Biol. 232*, 584−599.
31. Rost, B., Yachdav, G., and Liu, J. (2004) The PredictProtein server. *Nucleic Acids Res. 32*, W321−326.
32. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., and Sweet, R. M. (1993) Crystal structure of globular domain of histone H5 and its implications for nucleosome binding, *Nature 362*, 219−223.
33. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003) SWISS-MODEL: An automated protein homology-modeling server, *Nucleic Acids Res. 31*, 3381−3385.
34. Ariyoshi, N., Hiyoshi, H., Katagiri, C., and Abe, S. I. (1994) cDNA cloning and expression of *Xenopus* sperm-specific basic nuclear protein 5 (SP5) gene, *Mol. Reprod. Dev. 37*, 363−369.
35. Frottin, F., Martinez, A., Peynot, P., Mitra, S., Holz, R. C., Giglione, C., and Meinnel, T. (2006) The proteomics of N-terminal methionine cleavage, *Mol. Cell Proteomics 5*, 2336−2349.
36. Townend, R., Kumosinski, T. F., Timasheff, S. N., Fasman, G. D., and Davidson, B. (1966) The circular dichroism of the beta structure of poly-L-lysine, *Biochem. Biophys. Res. Commun. 23*, 163−169.
37. Hansen, J. C., Lu, X., Ross, E. D., and Woody, R. W. (2006) Intrinsic protein disorder, amino acid composition, and histone terminal domains, *J. Biol. Chem. 281*, 1853−1856.
38. Hartman, P. G., Chapman, G. E., Moss, T., and Bradbury, E. M. (1977) Studies on the role and mode of operation of the very-lysine-rich histone H1 in eukaryote chromatin. The three structural regions of the histone H1 molecule, *Eur. J. Biochem. 77*, 45−51.
39. Thomas, J. O., Rees, C., and Finch, J. T. (1992) Cooperative binding of the globular domains of histones H1 and H5 to DNA, *Nucleic Acids Res. 20*, 187−194.
40. Varga-Weisz, P., van Holde, K., and Zlatanova, J. (1993) Preferential binding of histone H1 to four-way helical junction DNA, *J. Biol. Chem. 268*, 20699−20700.
41. Varga-Weisz, P., Zlatanova, J., Leuba, S. H., Schroth, G. P., and van Holde, K. (1994) Binding of histones H1 and H5 and their globular domains to four-way junction DNA. *Proc. Natl. Acad. Sci. U.S.A. 91*, 3525−3529.
42. Ivanchenko, M., Zlatanova, J., and van Holde, K. (1997) Histone H1 preferentially binds to superhelical DNA molecules of higher compaction, *Biophys. J. 72*, 1388−1395.
43. Eirin-Lopez, J. M., Gonzalez-Tizon, A. M., Martinez, A., and Mendez, J. (2004) Birth-and-death evolution with strong purifying selection in the histone H1 multigene family and the origin of orphon H1 genes, *Mol. Biol. Evol 21*, 1992−2003.
44. Ausió, J. (1986) Structural variability and compositional homology of the protamine-like components of the sperm from bivalve molluscs, *Comp. Biochem. Physiol. 85B*, 439−449.
45. Bandiera, A., Patel, U. A., Manfioletti, G., Rustighi, A., Giancotti, V., and Crane-Robinson, C. (1995) A precursor-product relationship in molluscan sperm proteins from *Ensis minor*, *Eur. J. Biochem. 233*, 744−749.
46. Lewis, J. D., Saperas, N., Song, Y., Zamora, M. J., Chiva, M., and Ausio, J. (2004) Histone H1 and the origin of protamines, *Proc. Natl. Acad. Sci. U.S.A. 101*, 4148−4152.
47. Watson, C. E., and Davies, P. L. (1998) The high molecular weight chromatin proteins of winter flounder sperm are related to an extreme histone H1 variant, *J. Biol. Chem. 273*, 6157−6162.
48. Watson, C. E., Gauthier, S. Y., and Davies, P. L. (1999) Structure and expression of the highly repetitive histone H1-related sperm chromatin proteins from winter flounder, *Eur. J. Biochem. 262*, 258−267.
49. Carlos, S., Hunt, D. F., Rocchini, C., Arnott, D. P., and Ausió, J. (1993) Post-translational cleavage of a histone H1-like protein in the sperm of *Mytilus*, *J. Biol. Chem. 268*, 195−199.
50. Agelopoulou, B., Cary, P. D., Pataryas, T., Aleporou-Marinou, V., and Crane-Robinson, C. (2004) The sperm-specific proteins of the edible oyster (European flat oyster (Ostrea edulis)) are products of proteolytic processing, *Biochim. Biophys. Acta 1676*, 12−22.
51. Hiyoshi, H., Uno, S., Yokota, T., Katagiri, C., Nishida, H., Takai, M., Agata, K., Eguchi, G., and Abe, S. (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.
52. Lewis, J. D., Song, Y., de Jong, M. E., Bagha, S. M., and Ausio, J. (2003) A walk though vertebrate and invertebrate protamines, *Chromosoma 111*, 473−482.
53. Happel, N., Schulze, E., and Doenecke, D. (2005) Characterisation of human histone H1x, *Biol. Chem. 386*, 541−551.
54. Stoldt, S., Wenzel, D., Schulze, E., Doenecke, D., and Happel, N. (2007) G1 phase-dependent nucleolar accumulation of human histone H1x, *Biol. Cell. 99*, 541−552.
55. Lewis, J. D., and Ausió, J. (2002) Protamine-like proteins: evidence for a novel chromatin structure, *Biochem. Cell Biol. = Biochimie Et Biologie Cellulaire 80*, 353−361.

BI701274S