The Sperm Proteins from Amphioxus Mirror Its Basal Position among Chordates and Redefine the Origin of Vertebrate Protamines

José María Eirín-López,‡ Lindsay J. Frehlick,* Manel Chiva,‡ Nuria Saperas,§ and Juan Ausiò*  

*Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia, Canada; ‡Department of Biología Celular y Molecular, Universidade da Coruña, Campus de A Zapateira s/n, A Coruña, Spain; §Deparment d’Enginyeria Química, ETSEIB, Universitat Politècnica de Catalunya, Diagonal, 647, Barcelona, Spain

The sperm nuclear basic proteins (SNBPs) that participate in chromatin condensation in spermatozoa belong to 3 groups: histone (H), protamine-like (PL), and protamine (P) type. They share a common origin with histone H1 resulting from the segregation of PL components, corresponding to different regions of an H1 precursor molecule (N-terminal, winged-helix, C-terminal domains), becoming independent and following a subsequent process of parallel vertical evolution (H ↔ PL ↔ P). In the present work, we describe the sequence and primary structure of the main SNBP component in the sperm of the cephalochordate Branchiostoma floridae (amphioxus), revealing that it represents the deuterostome counterpart of the PL-III SNBP component from molluscs corresponding to the H1 N-terminal region. Until now, this has been a missing piece needed to complete the evolutionary history of SNBPs in metazoan genomes. The discovery of this PL lineage in deuterostomes definitively validates the parallel vertical evolution of SNBPs across metazoans, giving further support to the “basal” position of amphibious among chordates, with respect to tunicates. Sequence analyses suggest that later on in evolution, the appearance of positively selected arginine-rich protamines, derived from the H1 C-terminal region, led to the extinction of this PL lineage in the genomes of early protostomes and deuterostomes. Given that tunicates are now viewed as a sister group of vertebrates, the lysine to arginine transition responsible for the origin of vertebrate protamines must be set a step back from tunicates.

Introduction

In somatic eukaryotic cells, the DNA is compact within the nucleus by chromatin, a nucleoprotein complex that results from the association of DNA with basic chromosomal proteins known as histones (van Holde 1988). In contrast, sperm cells exhibit much more protein variability, allowing the extreme compaction of chromatin and the subsequent reduction in the volume of the sperm nucleus (Ausiò et al. 2007). The sperm nuclear basic proteins (SNBPs) that participate in the condensation process have been shown to belong to 3 well-defined groups: histone (H), protamine-like (PL), and protamine (P) types (Ausiò 1995, 1999). We have recently shown that PL proteins shared a common evolutionary origin with the histone H1 family early in metazoan evolution, resulting from the segregation of independent PL gene components corresponding to different regions of a precursor molecule (N-terminal, winged-helix, and C-terminal domains) (Eirín-López, Lewis, et al. 2006).

This observation represents an important starting point in assembling the pieces necessary to reconstruct the evolutionary histories of histone H1 and SNBPs. Piece one, we already evidenced that the long-term evolution of the histone H1 family has been subject to a birth-and-death process under strong purifying selection at the protein level, leading to the high levels of diversification and differentiation observed among the family members in higher eukaryotes (Eirín-López et al. 2004; Nei and Rooney 2006). Piece 2, the long-term evolution of SNBPs seems to be best described by a process of parallel vertical evolution (H ↔ PL ↔ P) subject to purifying selection at least in H-type and PL-type lysine-rich lineages in both protostomes and deuterostomes (Eirín-López, Lewis, et al. 2006). Accordingly, only H- or PL-precursor SNBP types would be present in those taxa that arose early in metazoan evolution, whereas the more specialized PL and P types would represent a characteristic feature of those taxa located at the uppermost evolutionary branches of bilaterian evolution. Piece 3, the transition from lysine-rich PLs toward the highly specialized arginine-rich protamines was caused by a frameshift mutation in the PL component corresponding to the C-terminal region of H1 (Lewis et al. 2004) and involved a shift from negative selection to an adaptive selection process (Eirín-López, Lewis, et al. 2006).

However, an important piece is still needed in order to articulate and validate the whole evolutionary picture of these proteins. In one hand, functional PL components corresponding to the N-terminal region of an H1 precursor have been only identified in mollusc species (protostomes) (Rocchini et al. 1995). This observation is in contrast to the expectations made by the model of parallel vertical evolution of SNBPs, which suggests that the differentiation of the 3 PL components occurred before the split between protostomes and deuterostomes. On the other hand, and more importantly, very recent analyses have traded the position of tunicates with that occupied by cephalochordates (also known as amphioxus or lancelets) in chordate evolution (Bourlat et al. 2006), with the amphioxus now being viewed as the most “basal” chordate and the tunicates as the sister group of the vertebrates (Bourlat et al. 2006; Holland 2007).

In the present work, we found that the SNBP component in the sperm of the cephalochordate Branchiostoma floridae is exclusively composed of the PL fraction corresponding to the N-terminal region of H1. The discovery of this PL lineage in deuterostomes definitively validates the parallel vertical evolution of SNBPs across metazoans, giving further support to the basal position of the amphioxus among chordates, with respect to tunicates. The appearance of arginine-rich protamines later on in evolution led to the extinction of this PL lineage due to its inability to undergo a lysine to arginine transition.

Key words: reproductive proteins, protamines, metazoans, long-term evolution, chromatin.

E-mail: jausio@uvic.ca.

doi:10.1093/molbev/msn121
Advance Access publication May 23, 2008

© The Author 2008. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org
Materials and Methods

SNBP Extraction and Fractionation

SNBPs were extracted from testes of *B. floridae* with 0.4 N HCl and precipitated with acetone as described in Wang and Ausió (2001). The protein extract thus obtained was fractionated by high-performance liquid chromatography (HPLC) on a reverse phase 300-Å Vydac C18 column (25 × 0.46 cm) eluted at 1 ml/min with a 0.1% trifluoroacetic acid–acetonitrile gradient, and the absorbance was monitored at 230 nm (Ausió 1988). Proteins were analyzed by acetic acid–urea polyacrylamide gel electrophoresis (AU-PAGE) (15% acrylamide, 0.1% bisacrylamide, 5% acetic acid, and 2.5 M urea), and gels were then stained with 0.2% (w/v) Coomassie blue for 1 h in 25% (v/v) isopropanol 10% (v/v) acetic acid and destained overnight in 10% (v/v) isopropanol/10% (v/v) acetic acid.

In Silico Isolation of the Amphioxus PL Protein and Genomic DNA and cDNA Sequence Determination

Protein sequences corresponding to the PL component of the sperm of molluscs, tunicates and sea urchins were used in protein–protein basic local alignment and search tool (BLAST) searches against the complete genome sequence of *B. floridae* (unmasked v1.0, DOE Joint Genome Institute, http://www.jgi.doe.gov). The genomic region of the only SNBP candidate found (sequence ID 125082) was subsequently retrieved from the amphioxus database and used in the design of gene-specific primers. Total RNA was extracted from whole *B. floridae* animals using Trizol reagent (GibcoBRL, Burlington, ON), and mRNA was isolated using an mRNA purification kit (Amersham Bioscience, Piscataway, NJ). Genomic DNA was extracted using a DNeasy Tissue Kit (Qiagen, Mississauga, ON) following the manufacturers’ instructions. Gene-specific primers were designed using the genomic sequence retrieved from the amphioxus genome as a reference, and they were subsequently used for polymerase chain reaction with cDNA and genomic DNA templates. In addition, rapid amplification of cDNA ends (RACE) was employed to obtain the complete cDNA sequence using the First Choice RACE Kit (Ambion, Austin, TX). Sequencing was done by the DNA Sequencing Facility, Centre for Biomedical Research at the University of Victoria.

Molecular Evolutionary Analyses

A total of 206 amino acid sequences (see supplementary table, Supplementary Material online) were used in the analyses, including 91 nonredundant histone H1 somatic sequences (68 replication dependent [RD] and 23 replication independent [RI]), 1 oocyte-specific H1, 16 testis-specific H1, 97 SNBP sequences (3 H type, 18 PL type, and 76 P type), and the PL protein from amphioxus. Multiple alignments of the amino acid sequences were conducted using the BIOEDIT (Hall 1999) program, and all molecular evolutionary analyses were conducted using the program MEGA version 3.1 (Kumar et al. 2004). The extent of the amino acid divergence was estimated by means of the uncorrected differences (p distance), and the minimum evolution tree-building method (Saitou and Nei 1987) was used to reconstruct the phylogenetic trees. In order to assess that our results were not dependent on this method, the simplified phylogeny was compared with a maximum parsimony tree reconstructed using the close-neighbor-interchange tree search method. The reliability of the topologies was tested by the bootstrap method (Felsenstein 1985) and the interior branch test (confidence probability) method (Sitnikova 1996), based on 1,000 replicates. Codon usage analyses on PL, P, and H1 genes were performed by using the DnaSP version 3 program (Rozas et al. 2003), using the “effective number of codons” (ENC) to estimate the codon usage bias (Wright 1990).

Results and Discussion

The Main Sperm Chromosomal Protein of Amphioxus Corresponds to the N-terminal Region of a Histone H1 Precursor

Figure 1A shows the SNBP composition of *B. floridae* (lane 5) in comparison to that of representative species from different classes of echinoderms and a tunicate. As seen in this figure (lanes 1–3), echinoderm SNBPs belong to the H type with the exception of the class Holoturoidea, where
a short PL (protein Phi-0) (Subirana 1970) is present in addition to a full histone complement (Azorin et al. 1983) (see fig. 1A, lane 3). In contrast, most of the histones are replaced by highly specialized PL-type proteins in tunicates (Saperas et al. 1992; Chiva et al. 1995), which are highly reminiscent of the P type of SNBPs (Lewis et al. 2004) (fig. 1A, lane 4). The amphioxus stands out in this comparative analysis in that it consists of SNBPs running as a single electrophoretic band with a higher mobility than histones, representing the main sperm chromosomal component. Despite the uniform monodisperse electrophoretic appearance, subsequent HPLC fractionations (fig. 1B) revealed the presence of several distinct peaks, depicting the microheterogeneous nature of SNBPs which is a structural characteristic of many chromosomal sperm proteins (Lewis et al. 2003).

Partial information on the N-terminal residues of the main SNP component of *B. floridae* was obtained (Saperas et al. 1994) and used as a reference in the in silico identification of the amphioxus SNP in the *B. floridae* genome assembly. Direct protein BLAST searches against the amphioxus genome database revealed the highest similarity score with the PL-III component from the mollusc *Mytilus californianus* (mussel). Arginine (R) and lysine (K) residues are indicated in black and gray boxes, respectively, and homology regions are highlighted by black bars below the alignment. Matching residues and gaps are indicated by dots and dashes, respectively. (B) Schematic representation of the exon–intron organization of the *B. floridae* SNP gene. (C) Sequence of the SNP transcript unit (the coding region is highlighted by a gray box) isolated from *B. floridae* cDNA. The locations of the primers F1/R1 used to confirm the presence of this gene in vivo and the additional pair of primers F2/R2 used in RACE amplifications are indicated by boxes in the sequence. Exon boundaries are indicated with open arrowheads. The polyadenylation signal and tail are underlined with solid and discontinuous black lines, respectively. Numbering on the left is referred to the nucleotide sequence and numbering on the right refers to amino acid residues.

Molecular Evolutionary Genomics of SNBPs

The hypothesis for the parallel vertical evolution of SNBPs (Ausió 1999) predicts that only H- or PL-precursor SNP types would be found in those taxa that arose early in metazoan evolution. The phylogenetic relationships among H1 and SNBPs reconstructed in the present work (fig. 3) clearly support this notion, depicting an evolutionary diversification process that mirrors the evolution of triploblastic animals. In fact, histone H1 and protamines appear as
2 clearly differentiated groups, with PLs occupying an intermediate position. The occurrence of RD and RI H1 histones, as well as PL- and P-type SNBPs in protostome and deuterostome representatives, provides support that their evolutionary origin occurred prior to the split between these 2 groups of organisms.

The topology shown in figure 3 is in good agreement with the taxonomic relationships among species, indicating a monophyletic origin for the group of RI H1s and a functional clustering pattern for the different mammalian somatic H1 isoforms (Eirín-López, Lewis, et al. 2006). The major SNBP component of *B. floridæ* locates in the PL region of the topology and shares a monophyletic origin with the PL-III SNBP component from molluscs, supporting the biochemical and molecular analyses presented in the previous section. Given that amphioxus represents the only deuterostome in which the main sperm chromosomal protein corresponds to the N-terminal region of a histone H1 precursor, this result definitively validates the hypothesis of their parallel vertical evolution (Ausió 1999). Accordingly, the products of the segregation of the different SNBP components from a histone H1 molecule should be present in both protostomes and deuterostomes, in agreement with their common evolutionary origin (Eirín-López, Lewis, et al. 2006). However, only components corresponding to the C-terminal region of H1 were found in deuterostomes, which eventually led to the differentiation of highly specialized arginine-rich protamines (Lewis et al. 2004).

The reconstruction of a simplified tree (fig. 4) underscores the close relationship of *B. floridæ* PL with SNBPs of the PL type in molluscs, especially PL-I and PL-III. Furthermore, the PL component from *B. floridæ* appears to be somewhat more closely related to molluscan PL-I and PL-III, and Phi-0 SNBPs, suggesting a more primitive origin and a closer relationship to echinoderms, as supported by recent phylogenetic evidence (Jutglar et al. 1991).

The differentiation of the amphioxus SNBP lineage reported in figure 3 agrees with the evidence provided by
recent reports supporting the status of the amphioxus as the most basal chordate, in exchange of tunicates (Bourlat et al. 2006). In such scenario, the SNBP component from amphioxus (whose protostome counterpart is represented by the PL-III protein from molluscs) would represent the oldest deuterostome PL lineage, prior to the differentiation of PLs corresponding to the C-terminal region of H1.

The Amphioxus PL Lineage Is Subject to Different Evolutionary Constraints than the Protamines or the PL-I Lineage from Tunicates

The progressive specialization of SNBPs has followed a parallel vertical pattern of evolution defined as H \rightarrow PL \rightarrow P (Ausió 1999; Eirin-López, Frehlick, and Ausió 2006), encompassing a progressive replacement of lysines (K) by arginines (R) leading to a more efficient DNA packaging in the sperm as well as in the activation of enzymatic cascades after fertilization in vertebrates (Ausió et al. 1984; Ohtsuki et al. 1996; Rooney and Zhang 1999). The gradual variation in the K/R ratio across H1/SNBP evolution is evident in the phylogeny shown in figure 3, ranging from around 8.7 in the case of H1 to 0.1 in the case of protamines. Furthermore, the increase in arginine is concomitant with the increase in the numbers of nonsynonymous substitutions among P lineages (see fig. 3), suggesting that this trait is positively selected.

It has been recently shown that a frameshift mutation involving only 2 nt would be enough to duplicate the percent of arginine in the PL component of the tunicate Ciona intestinalis, due to the high frequency of AAG and AGA triplets encoding polylysine and polyarginine tracts, respectively (Lewis et al. 2004). By studying the relative codon usage for lysine and arginine residues (fig. 5A), it is evident that, although the SNBP from B. floridae presents similar amounts of lysine and arginine (as does PL-III from molluscs), all lysines are encoded by the same codon (AAA) with a higher proportion of arginine residues encoded by CGU and CGC. This codon preference is in sharp contrast to that shown by histone H1 and the PL components from molluscs, where arginines are mainly encoded by AGA and AGG triplets, and would ultimately represent an insurmountable barrier for a lysine to arginine transition based on a frameshift mutation.

Although molluscan PLs and the amphioxus SNBP are very biased in terms of their codon usage, somatic histone H1s and tunicate PLs exhibit intermediate and very low codon bias, respectively. This may be explained by the fact that although PL proteins rely equally on lysines and arginines for DNA packaging, constraints acting on protamines are only dependent on the overall arginine content of the molecule, rather than on the specific position of the arginines or the nature of their encoding triplets (Rooney et al. 2000; Lewis et al. 2003). This unusual form of purifying selection, together with the lysine to arginine transition, would be consistent with the low codon bias values observed in the arginine-rich PLs from tunicates in comparison with lysine-rich PLs.

Fig. 4.—The tree was rooted with the H1-like sequence from the protist Entamoeba histolytica as it represents one of the most primitive eukaryotes for which a H1-related protein has been characterized (Kasinsky et al. 2001). The different lineages of H1 (RD/RI) as well as PL-type and P-type SNBPs are indicated in the right margin of the topology together with the mechanisms of 3’ mRNA processing specific for each group. Transition proteins and testis-specific H1 histones (H1t) are marked as (*) to indicate that their expression pattern is intermediate between RD and RI H1 histones. Numbers for interior branches are representative of bootstrap and interior branch test, followed by the bootstrap values obtained in the reconstruction of the maximum parsimony tree (in boldface). Numbers in parentheses by the species names indicate the number of sequences used.
The PL component from Amphioxus Provides a Footprint in the Process of Parallel Evolution Observed in SNBPs across Metazoan Genomes

There is now plenty of evidence indicating that the exclusion of specialized H1 genes from the main repetitive histone units to a solitary location in the genome resulted in the generation of an “orphon” group of H1 genes early in the evolution of metazoans (Eirín-López et al. 2002, 2004). The independent evolution of this group eventually led to the differentiation of RD and RI H1 variants, followed by the subsequent differentiation of SNBPs within the second lineage (Eirín-López, Lewis, et al. 2006). During the differentiation and diversification of these chromosomal proteins, the transition toward arginine-rich protamines appears to have involved in the progressive specialization of a PL protein corresponding to the C-terminal region of H1 (Lewis et al. 2004), which eventually suffered a shift from a purifying selection process (present in PLs) to adaptive selection favoring high arginine contents in protamines (fig. 5B) (Balhorn 2007). However, the presence and fate of the SNBP component corresponding to the N-terminal part of H1 have remained elusive until now.

Conclusions

The identification of a deuterostome counterpart (represented by the amphioxus PL reported here) to the protostome PL component corresponding to the N-terminal region of a precursor histone H1 (represented by the PL-III protein from molluscs) provides a footprint that definitively validates the parallel vertical evolution of SNBPs (summarized in fig. 6). This result is even more critically relevant in light of recent studies placing the amphioxus in the basal position in chordate evolution (Bourlat et al. 2006), in place...
of tunicates. In this new scenario, the lysine to arginine transition observed in tunicates could no longer be considered as an early event in bilaterians or stand for the origin of protamines in vertebrates. The existence of tunicate PLs which are already highly specialized and reminiscent of P-type SNBPs, together with the sharp contrast they present with respect to the primitive amphioxus PL-type SNBP lineage, make tempting to hypothesize that such observations support the basal condition of amphioxus among chordates (fig. 5B). Furthermore, the Phi-0 protein from the sea cucumber Holothuria tubulosa (Prats and Comudella 1995) displays certain sequence similarities, including the presence of introns, with the B. floridae SNBP as well as with PL-III (figs. 3 and 4). This further supports the origin of this
SNBP lineage as early as before the differentiation between protostomes and deuterostomes. However, and even by taking into account that protamines have been successfully used as molecular markers in the past (Eirín-López, Frehlick, and Ausió 2006; Frehlick et al. 2006), these results must be interpreted with caution given that tunicates have very derived genomes and body plans.

The SNBP component corresponding to the H1 N-terminal region appears to be the initial choice for sperm DNA packaging in early metazoans. The extinction of this PL lineage in the genome was probably due to its inability to undergo a lysine to arginine transition based on the observed patterns of codon usage bias for lysine residues. In this regard, the PL-III protein from molluscs and the SNBP component identified in amphioxus would represent end points of the SNBP evolutionary process in deuterostome and protostome genomes that were promptly discarded by the fast specialization experienced by arginine-rich PL- and P-type SNBPs which arose through a process of positive selection.

Supplementary Material

Supplementary table and figure are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/). The sequence described in the present work has been deposited in the GenBank database with accession number EU271675.

Acknowledgments

We are very thankful to Linda Holland of the Marine Biology Research Division at the Scripps Institution of Oceanography (UCSD) for providing us with specimens of B. floridae during the early biochemical characterization of SNBPs and to Nancy Sherwood and her laboratory at the Department of Biology at UVic for providing us with material for the molecular biology analyses. This work was supported by Natural Science and Engineering Research Council of Canada (NSERC) Grant OGP 0046399-02 to J.A. and by Grants BFU 2005-00123 and BFU2006-04035/BMC from the Spanish MCYT PGC-FEDER. J.M.E.-L. is a recipient of a Postdoctoral Marie Curie International Fellowship within the sixth European Community Framework Program. L.J.F. holds an NSERC PGSD fellowship.

Literature Cited


Billie Swalla, Associate Editor

Accepted May 20, 2008