Origin and evolution of chromosomal sperm proteins

José M. Eirín-López,[1] *and Juan Ausió*[2]*

[1]Departamento de Biología Celular y Molecular, Universidade da Coruña, E15071 A Coruña, Spain
[2]Department of Biochemistry and Microbiology, University of Victoria, Victoria, BC, V8W 3P6, Canada

In the eukaryotic cell, DNA compaction is achieved through its interaction with histones, constituting a nucleoprotein complex called chromatin. During metazoan evolution, the different structural and functional constraints imposed on the somatic and germinal cell lines led to a unique process of specialization of the sperm nuclear basic proteins (SNBPs) associated with chromatin in male germ cells. SNBPs encompass a heterogeneous group of proteins which, since their discovery in the nineteenth century, have been studied extensively in different organisms. However, the origin and controversial mechanisms driving the evolution of this group of proteins has only recently started to be understood. Here, we analyze in detail the histone hypothesis for the vertical parallel evolution of SNBPs, involving a ''vertical'' transition from a histone to a protamine-like and finally protamine types $(H \rightarrow PL \rightarrow P)$, the last one of which is present in the sperm of organisms at the uppermost tips of the phylogenetic tree. In particular, the common ancestry shared by the protamine-like (PL)- and protamine (P)-types with histone H1 is discussed within the context of the diverse structural and functional constraints acting upon these proteins during bilaterian evolution.

Keywords: arginine; chromatin; evolution; histones; metazoans; sperm nuclear basic proteins; spermatozoa

Sperm nuclear basic proteins (SNBPs) and chromatin organization

Classification and fundamental types of sperm nuclear basic proteins

The main chromosomal proteins that are found associated with DNA in sperm chromatin are called SNBPs.[1] In contrast to the proteins of somatic chromatin (histones), SNBPs

exhibit a greater compositional heterogeneity. Although several classification systems have been proposed for SNBPs in the past based on the compositional features of these proteins,[2–4] a more recent classification which also takes into account additional structural considerations has grouped these proteins into three major types: histone (H)-type, protamine (P)-type, and protamine-like (PL)-type.[5]

The H-type encompasses chromosomal proteins that are compositionally and structurally very similar to the histones that are found in somatic tissues. Somatic histones can be classified into core (H2A, H2B, H3, H4) and linker (H1/H5) histones. Core histones contain a histone-fold domain[6] that is flanked by intrinsically disordered N- (DND) and C-terminal domains (DCD). In contrast to core histones, linker histones exhibit a larger degree of micro-heterogeneity[7,8] and their tertiary structure organization consists of a winged-helix motif[9] domain [winged-helix fold domain (WHD)] that is flanked by DND and DCDs (see Fig. 1A). On the other hand, the P-type encompasses a group of compositionally and structurally heterogeneous small arginine-rich $(Arg \geq 30\%$ mol/mol) (Fig. 1B) proteins (relative molecular mass between 4,000 and 10,000). In addition to arginine, these proteins may also contain cysteine,[10] an amino acid not usually found in other chromosomal proteins. Organisms whose sperm contains P-type SNBPs have the most complex spermatogenic pattern, which involves important histone and protamine post-translational modifications (such as acetylation and phosphorylation) designed to facilitate the histone displacement and protamine assembly processes during the final spermiogenic transitions.[10,11]

The term ''protamine-like'' was first described by Subirana.[12] The PL-type of SNBPs represent a structurally and functionally intermediate group between the H- and P-type SNBPs that are related to histone H1[1,13,14] (Fig. 1A, B). There are basically two main subtypes of PLs: those whose structure consists of a WHD and who are referred to as PL-I (Fig. 1A), and those that lack this domain, like *Mytilus* (bivalve mollusc) PL-III[15] or *Branchiostoma* (tunicate) PL.[16] The latter are smaller in size, exhibit a higher electrophoretic mobility[17] and are evolutionarily related to histone H1 N-terminal domain/C-terminal domains[15] (see Fig. 1A and Box 1).
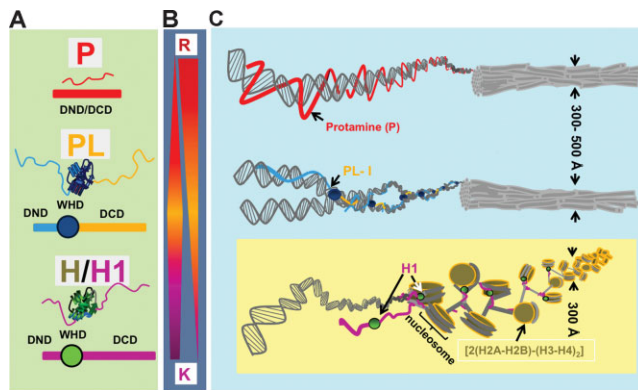
Figure 1. Structural organization of histone H1, PL-I-type, and P-type SNBPs. **A**: Metazoan histone H1 and PL-I proteins exhibit a tripartite organization including a central globular core with a characteristic WHD that is flanked by DND and DCD. Smaller proteins of the SNBP PL-type and protamines (P-type) consist only of either the DND or DCD in the later instance. **B**: The proteins shown in A are structurally and evolutionarily related. During the process of evolution the amount of lysine in the H1 histones increased, which was accompanied by a substantial increase in arginine of PL-type SNBPs. The transition to arginine-rich P-type SNBPs involved a further increase in arginine with a substantial decrease in the lysine content. **C**: The compositional and structural differences between these chromosomal proteins result in different chromatin structures, which all lead to the formation of intermediate fibers with an average diameter of 300–500 Å. In the sperm nuclei (blue section), the three types of 300 Å fibers shown can be present depending on the organism and its SNBP type composition. In somatic cells (yellow section) only the H/H1 type is present. In both sperm and somatic cells, the folding of the 300 Å fiber is mediated and stabilized by the binding of histone H1 to the linker linear DNA connecting adjacent nucleosomes. Figure adapted and modified from.[55]

Box 1

Many species from different bilaterian taxonomic groups retain histones during the entire spermatogenic process[1,18] and in some instances these histones are compositionally and structurally indistinguishable from the somatic counterpart.[19] This represents the most simple type of chromosomal protein transition during spermiogenesis. However, the changes in post-translational modifications of histones during this process are as yet undetermined. Mammals, which all contain protamines in their sperm, provide an example of the highest spermatogenic complexity, requiring the presence of highly specialized testes histones that are replaced by transition proteins[20] before their final replacement by protamines. Like protamines, PL SNBPs partially replace the histones that are present at the onset of spermiogenesis. However, as in the case of the H-type, the detailed molecular mechanisms involved in the process are still unclear.

The organization of sperm chromatin

Sperm chromatin is unique in that most, if not all of it, is homogeneously heterochromatinized. Genomic DNA is tightly packed within the highly compacted sperm nuclei irrespective of the SNBP type it is associated with. However, in the early stages of the transition toward this quiescent state, chromatin fibers of ~300–500 Å can be clearly visualized. Figure 1C shows a schematic representation of the different elementary chromatin fibers resulting from the interaction of the three main types of SNBPs with DNA.

Chromatin fibers containing H-type SNBPs exhibit an identical organization to that observed in somatic cell nuclei, in which ~200 bp of DNA are wrapped around a protein core, resulting in a nucleoprotein particle that is called the nucleosome (see Fig. 1C, yellow area). Histones of the H1 family bind to the linker DNA connecting adjacent nucleosomes and further fold and stabilize the DNA into 300 Å chromatin fibers,[21] which can further fold onto themselves and/or produce highly condensed structures in the highly compacted sperm nucleus.[22] In contrast, P-type SNBPs bind to linear DNA at the major groove[23] and extensively neutralize the negative charge of the phosphate backbone allowing for the coalescence of the resulting nucleoprotamine complexes into fibers of 300–500 Å. The presence of cysteine in some of these proteins can provide further structural compaction and stability to the chromatin fibers.

In comparison to the chromatin complexes arising from the interaction of H- and P-type SNBPs with DNA, less information is available on the chromatin assemblies resulting from the interaction of PL-type SNBPs with DNA. Studies carried out in both protostomes[24,25] and deuterostomes[26,27] indicate that, like with the P-type, these proteins interact with linear DNA. Nevertheless, the presence of a WHD in some of these proteins, as in the case of PL-I of the fish *Mullus surmuletus*, appears to confer an enhanced ability for binding to intertwisting DNA resulting in a regularly intertwined organization of DNA[27] (Fig. 1C). As shown in Fig. 1C, the chromatin fibers resulting from the association of the different SNBP types with DNA all exhibit a fairly constant diameter in the range of 300–500 Å independently of the extent of protein folding of the SNBP type involved, which decreases in going from the H- to the P-type.

At the functional level, the tight packing of sperm chromatin results in a significant decrease in the nuclear metabolic activity regardless of the SNBP type involved. However, recent proteomic studies (see[28] for a review) carried out with sperm from different organisms have provided evidence for the presence of functional proteins such as GTPases as well as histone acetyltransferases and deacetylases, among others, which suggests the presence of transcriptional activity. Also, it has been shown that DNA repair can take place late in mammalian spermiogenesis.[29] How these processes can

proceed in the highly condensed chromatin states described above is a very timely and interesting issue that remains unclear.[28] Comparative proteomic analysis carried out with sperm from organisms belonging to different SNBP types may provide important insightful information in this regard as well as in regard to the evolution of the functional sperm proteins.

Sperm chromosomal proteins share a common evolutionary ancestry with histone H1

From a general evolutionary perspective, replication-independent (RI) genes of somatic histones became replication-dependent (RD) during eukaryotic evolution. While RD genes underwent a large increase in their copy number, RI genes became isolated at different chromosomal locations, exhibiting an "orphon origin".[30] The highly expressed RD histones are often referred to as canonical histones (*e.g.*, H2A.1, H2A.2, mammalian H1.1-H1.5), whereas the RI histones include most of the developmentally regulated and replacement histone variants that replace the canonical ones during different stages of the cell cycle[31] (*i.e.*, avian erythrocyte H5, H2A.Z, H2A.Bbd, H3.3, and the sperm-specific histones).

The selective pressure behind the transition from lysine to arginine associated with the process of SNBP evolution (Fig. 1B) appears to be specifically geared toward the preservation of high levels of arginine in P-type SNBPs resulting in (i) increased binding affinity of a protein for DNA[32,33] and tighter chromatin condensation in the sperm nucleus, (ii) additional functional advantage at the time of sperm-egg fertilization (activation of the regulatory pathways in fertilized eggs by way of polyarginine clusters[34]), and (iii) effects on different sperm traits (head shape and sperm motility) with relevant implications for fertility, sperm competition,[18] as well as for the adaptation for internal fertilization,[35] as documented in fish.[36]

SNBPs of the PL-type exhibit structural and functional similarities with specialized variants of the histone H1

The histone H1 family contains one of the largest numbers of variants among histone families. This is especially evident in mammals where 11 different linker histones have been identified (see[13] for review). In contrast to core histones, whose origin can be traced back to archaebacteria,[37] the evolutionary origin of histone H1 appears to have occurred early in eubacteria.[38] However, it seems that, as with core histones, the evolution of the histone H1 family has been driven by a process of "birth-and-death" DNA duplication instead of concerted evolution (see Box 2), allowing for the

appearance of the genetic diversity required for the progressive functional specialization of these proteins.[30,39,40]

Box 2

Concerted evolution represents a form of multigene family evolution in which all the member genes are assumed to evolve as a unit in concert. In this process, a mutation occurring in a repeat spreads through the entire cluster of genes by recurrence of unequal crossover or gene conversion (a form of nonreciprocal recombination in which a DNA segment of a recipient gene is copied from a donor gene). In contrast, the genetic diversity arising from birth-and-death evolution is assumed to result from the creation of new genes by gene duplication. In some instances, the duplicated genes are maintained in the genome for a long period of time, while others are deleted or inactivated through deleterious mutations.[40]

Similar to somatic H1 histones, PLs represent the most structurally heterogeneous group of proteins among SNBPs and exhibit an intermediate amino acid composition between histones and protamines.[2] One such PL, PL-I from the surf clam *Spisula solidissima* (mollusc), was the first PL shown to contain a trypsin-resistant globular core arising from the presence of a WHD with a high sequence similarity (51%) to that of vertebrate histone H5.[41] The intriguing structural similarity of molluscan PL-Is to specialized members of the vertebrate histone H1 family such as H5 is not surprising considering the RI nature of their genes and that they both participate in heterochromatinization in terminally differentiated cell systems (erythrocyte/sperm). Indeed, studies on SNBP evolution reveal that, like H1 histones, PLs are also subject to purifying selection and their presence in both diploblastic and triploblastic animals indicates that their complete differentiation from the somatic H1 took place before the differentiation of these lineages very early in metazoan evolution.[13]

Common evolutionary origin of somatic histone H1 lineages and PL-/P-type SNBPs

The work carried out in molluscs has proven to be very useful for understanding the overall SNBP evolution as this taxonomic group includes organisms whose sperm contains SNBPs representative of the three main types (H, PL, P) in addition to RI and RD somatic histone H1 proteins. Molecular evolutionary studies comparing different SNBP types revealed the presence of regions of homology to different H1/PL-I structural domains as a result of either the

post-translational cleavage of PL-I precursors[42] or the segregation of the PL components into autonomous genes.[15] The presence of these autonomous genes suggests that the appearance of the smaller PLs with a marked increase in their arginine content is the result of a gene segregation process that occurred early on in evolution.[15]

A joint phylogenetic analysis of SNBPs and H1 histones revealed a close relationship between PL proteins and the RI histone H1 lineage, which were shown to constitute a single monophyletic group.[15] These two groups of proteins also have solitary locations in the genome, are present in terminally differentiated cells and have RI expression. The complete functional differentiation of the three lineages (somatic H1 RD, H1 RI, and SNBPs) would have arisen before the split between protostomes and deuterostomes, resulting in a process of parallel evolution. Thus, whereas birth-and-death evolution remained the major mechanism responsible for the functional diversification of RI and RD H1 lineages, PL-I proteins were subject to a purifying selection process that led to the appearance of highly specialized PL proteins. The subsequent transition to protamines (P-type) undergone by PLs in organisms at the uppermost tips of the phylogenetic branches of bilaterian evolution resulted in a shift toward a positive sex-driven selection,[43] which is responsible for the rapid rate of evolution exhibited by these proteins.

Mechanisms of long-term evolution of sperm nuclear basic proteins

Vertical *versus* horizontal hypotheses for the evolution of chromosomal sperm proteins

The notion that protamine (P-type) chromosomal sperm proteins have evolved from a primitive somatic-like histone precursor dates back to 1973 based on comparative amino acid composition analyses.[12] Other alternative hypotheses to this vertical model of SNBP evolution have also been considered. A process of horizontal evolution was put forward based on the apparent random distribution of protamines observed in different taxonomic groups of fish, which suggested that these proteins had a retroviral origin.[11,44] However, detailed systematic analysis on the phylogenetic distribution of SNBP types across different fish taxa conclusively showed that such an apparent sporadic distribution of protamines is not random and could be phylogenetically traced. The study of fish SNBPs[36] has also been very useful in providing potential clues for understanding the putative reversions of the H → PL → P transition[26] and has also provided important information on the constraints imposed on this transition by the different types of fertilization (external *vs*. internal).[36]

Further evidence for the vertical evolution of SNBPs can be obtained from the exclusive presence of H-type SNBPs in sponges[45] and the presence of RI histone H1-related PL proteins (PL-I) in the sperm of different cnidarians.[46,47] Furthermore, SNBPs of the PL- and P-types are more prevalent in the sperm of taxonomic groups located on the upper phylogenetic branches of bilaterian evolution,[14,48] as shown in Fig. 2. These findings are in very good agreement with the predictions made by the vertical model of SNBP evolution. Accordingly, only H-type or PL precursor SNBPs should be present in taxonomic groups that differentiated very early during metazoan evolution with the more specialized protamines being restricted to higher phylogenetic groups.[15]

Different constraints and common trends in the long-term evolution of SNBPs and histone H1

The protein lineage encompassing the different SNBP types represents a highly specialized set of proteins that arose from an ancestral RI H1 precursor[15] (Fig. 2). However, in the course of this evolutionary process, there appears to have been a progressive reduction in the structural protein complexity as illustrated by the segregation of the different domains of the ancestral PL-I proteins.[16,43] In contrast, the somatic RI and RD histone H1 lineages have experienced a genetic diversification without drastic changes in their overall protein structure using birth-and-death under strong purifying selection as a primary mechanism (Fig. 2). The differences between these two strategies are most likely a reflection of the different structural and functional constraints imposed by the packaging of DNA within the sperm and the somatic nucleus (Fig. 1).

However, and despite the striking differences between the long-term evolutionary mechanisms to which somatic H1s and SNBPs are subject, a common trend is evident in both instances as depicted by their progressive differentiation early in metazoans and their parallel segregation across protostomes and deuterostomes. This extends the concept of parallel vertical evolution of SNBPs and H1 histones in metazoans to a higher level that also involves the parallel evolution of male germinal and somatic proteins.

The transition toward arginine-rich protamines: biological reasons and evolutionary consequences

The replacement of lysine-rich proteins with arginine-rich protamines occurred by a selective sweep

For many years, one of the main conceptual barriers to validating the histone hypothesis for the vertical evolution of SNBPs has been the lack of a mechanism that explains
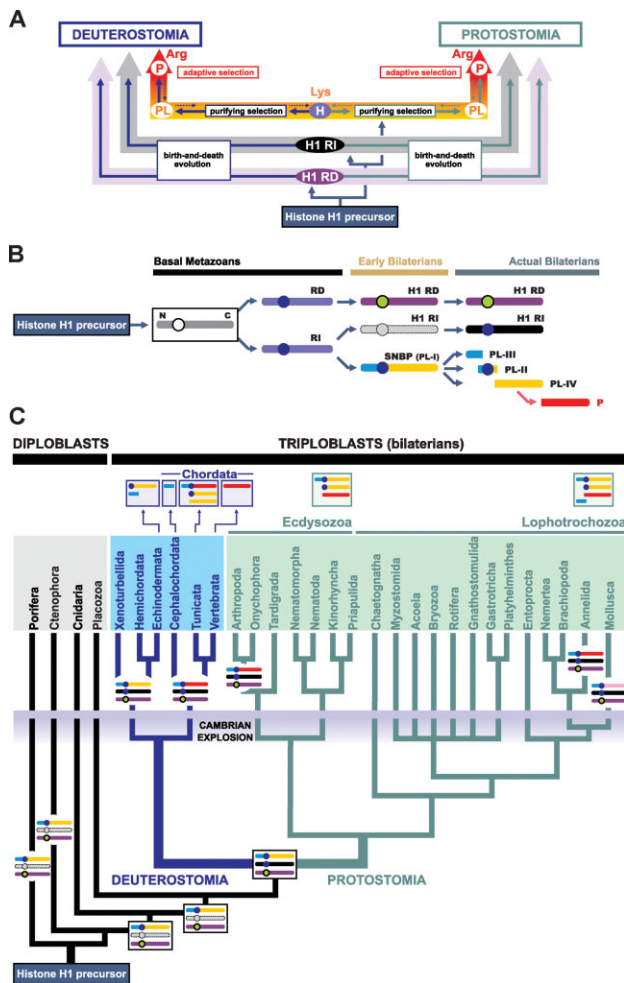
Figure 2. Evolution of histone H1 and SNBPs within the metazoan phylogeny.[52,56] **A**: Common origin and evolution of histone H1 and SNBPs within the context of parallel vertical evolution starting from a histone H1 precursor (H, light blue). The RD histone H1 lineage (purple), the RI histone H1 lineage (gray/black), and the PL-I/P lineages (orange/red) are indicated. The mechanisms acting on somatic RD and RI histone H1 lineages (birth-and-death) and on the PL-type (purifying selection) as well as the positive sex-driven adaptive selection that resulted in the P-type SNBPs are detailed in the diagram. **B**: Schematic representation highlighting how the exclusion of a histone H1 precursor from the main histone gene clusters to a solitary location in the genome resulted in an orphon lineage which initially led to the differentiation of the RD H1 lineage, followed by the RI H1 and SNBP lineages. In this process, PL-I (N-terminal region, light blue; globular domain, dark blue; and C-terminal lysine-rich domain, yellow) experienced different gene segregation events leading to highly specialized PL fractions as well as to the appearance of protamines (red) corresponding to the C-terminal end of an arginine-rich PL-I. **C**: The complete functional differentiation of the three lineages would have had its origin early in the evolution of triploblastic animals before the split between protostomes and deuterostomes. This resulted in a process of parallel evolution across protostomes (indicated in green) and deuterostomes (indicated in blue), in which the arginine-rich SNBPs of the P-type are present in the sperm of organisms at the furthermost tips of the bilaterian phylogenetic branches.[15]

the transition from the lysine-rich composition, which is characteristic of histone H1 molecules to the arginine-rich composition of protamines. In contrast to the progressive evolution from an H1-type into a PL-I precursor (and into specific PL proteins later on), the transition toward arginine-rich protamines involved a further replacement of lysine residues by arginines. In this regard, the analysis of the SNBPs from different related basal metazoans (especially tunicates) revealed the presence of members whose sperm contains lysine-rich PL-Is (such as *Ciona intestinalis*) and arginine-rich PL-Is (such as *Styela plicata* and *Styela montereyensis*).[43] Interestingly, *Styela*'s PL-I has arginines at the very same positions occupied by lysines in the C-terminal tail of *Ciona*'s PL-I, suggesting that a quick PL → P transition had taken place in this group of tunicates (Fig. 3A). This was taken as evidence supporting the origin of the chordate arginine-rich protamines in tunicates[43] that subsequently led to the differentiation of highly specialized protamines in vertebrates (see Box 3 for details).

Box 3

A close comparative examination of *Ciona* and *Styela* PL-Is revealed that while most lysines are encoded by the AAG triplet in *Ciona*'s PL-I, arginines are predominantly encoded by the AGA triplet in the PL-I of *Styela*. The replacement of lysines with arginines was shown to be the result of a single mutational event responsible for a frameshift in one of the nucleotide sequences (Fig. 3B).[43] The evidence obtained from these observations filled a very important gap in the histone hypothesis for the vertical evolution of SNBPs regarding the transition of lysine-rich PL proteins toward more specialized arginine-rich SNBPs of the P-type. Nevertheless, the residues that are found between the short polyarginine and lysine clusters of *Ciona*'s and *Styela*'s PL-Is seem to have not been affected by a frameshift mutation (Fig. 3B). This indicates that while such a unique single mutation event certainly may have played a role in the transition process, additional mechanisms must have also been involved.[13]

The appearance of arginine-rich P-type SNBPs brought about a switch in the evolutionary mechanism operating on them where they were subjected to a rapid process of adaptive selection that led to their establishment in different tunicate and chordate species. Although the presence of adaptive selection among reproductive proteins has been well documented,[49] in the case of protamines this process additionally involves the maintenance of high arginine levels regardless of the position of these residues within the molecule.[50]

The "extinct" PL protein from amphioxus provides further support for the parallel vertical evolution of SNBPs from a histone descent

Comprehensive phylogenomic analyses have recently challenged the phylogenetic position of tunicates and cephalo-

chordates in chordate evolution, placing cephalochordates as the most "basal" chordates and the tunicates as the sister group of the vertebrates.[51,52] The implication of this in terms of SNBP evolution is that the transition toward arginine-rich protamines would have had to arise independently in the tunicate/vertebrate sister lineages. A characterization of the
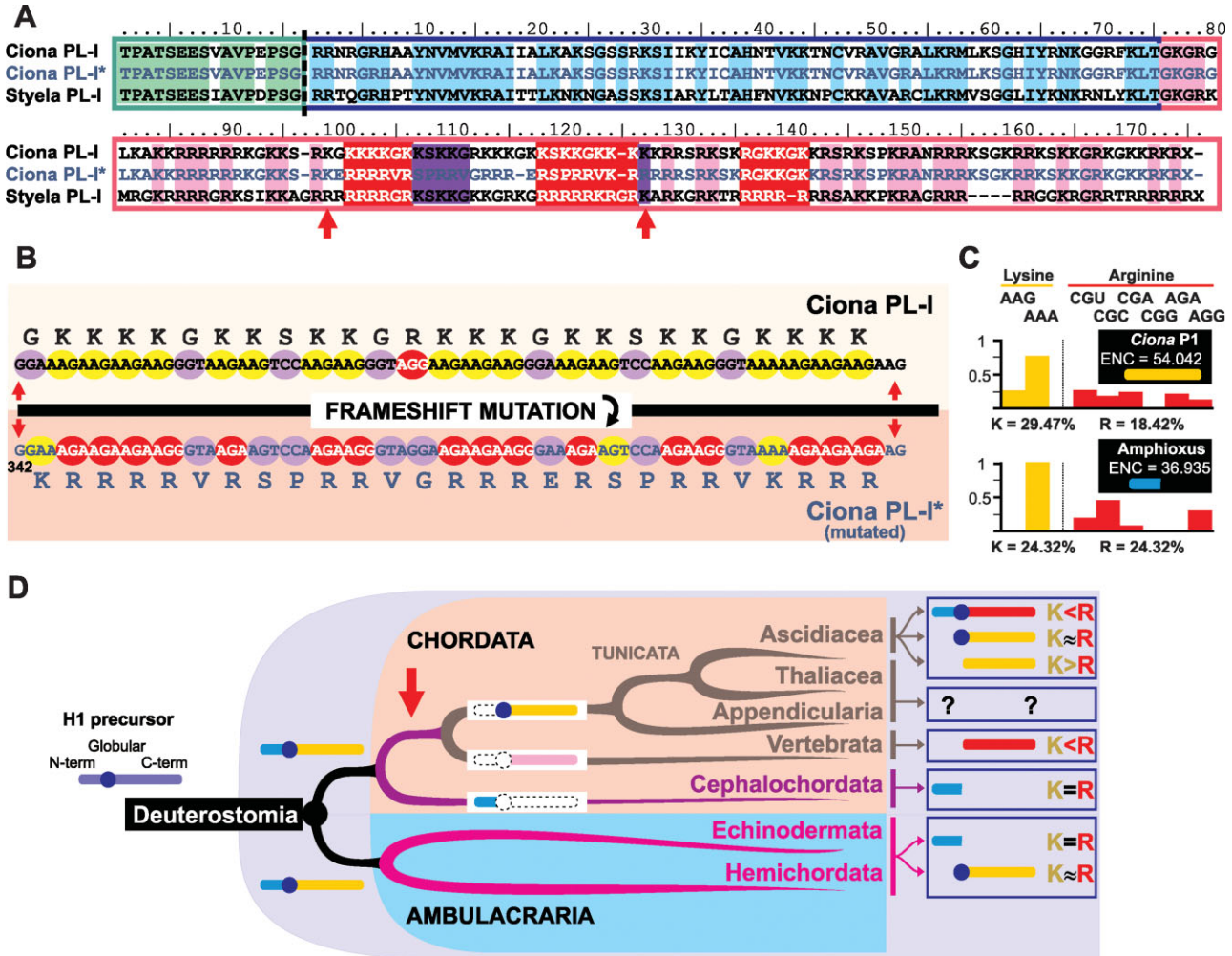

Figure 3. Transition toward arginine-rich protamines based on a frameshift mutation in tunicates. **A**: Amino acid sequence alignment of tunicate SNBPs including the lysine-rich PL-Is from *C. intestinalis*, the arginine-rich PL-I from *S. montereyensis* and the hypothetical arginine-rich mutated PL-I protein from *C. intestinalis* (PL-I*). The leading sequence (green box), the WHD (blue box) and the C-terminal region (red box) are highlighted. Solid red boxes indicate areas containing polyarginine and polylysine tracts, and solid purple boxes highlight unmatching residues between the arginine-rich hypothetical PL-I* protein from *C. intestinalis* derived from the frameshift mutation event and the arginine-rich PL-I protein found in *S. montereyensis*. The thick red arrows point to the boundaries of the lysine-rich region of *Ciona* PL-I and the corresponding arginine-rich region of *Styela* PL-I. **B**: Detailed representation of the putative gene frameshift mutation that led to the lysine (PL-I) to arginine (PL-I*) transition at the C-terminal region of *Ciona* PL-I. Lysine and arginine codons are indicated in yellow and red, respectively, while codons for other residues are indicated in purple. A deletion of a nucleotide at position 342 results in a frameshift mutation that converts 16 lysine codons to arginine codons. **C**: Relative codon usage for lysine (yellow) and arginine (red) residues in the PL-I protein of *C. intestinalis* and in the major PL protein of *B. floridae* (amphioxus). The codon usage bias is expressed as the effective number of codons (ENCs). The main SNBP component of cephalochordates corresponding to a PL-I N-terminal region had no lysine residues encoded by AAG and hence could not have undergone a similar transition. **D**: Schematic representation of the evolutionary process leading to the segregation and specialization of SNBPs within the deuterostome phylogeny.[16] A lysine-rich SNBP PL-I at the onset of the diversification between the tunicate and vertebrate lineages was subject to a process of adaptive selection resulting from the lysine (yellow) to arginine (red) transition (indicated by a red arrow) and giving rise to highly specialized protamines. The inability of the SNBP component of amphioxus to undergo this transition, together with the rapid evolution of protamines, led to the extinction of this lineage (blue), which got restricted to cephalochordates and perhaps some echinoderms.
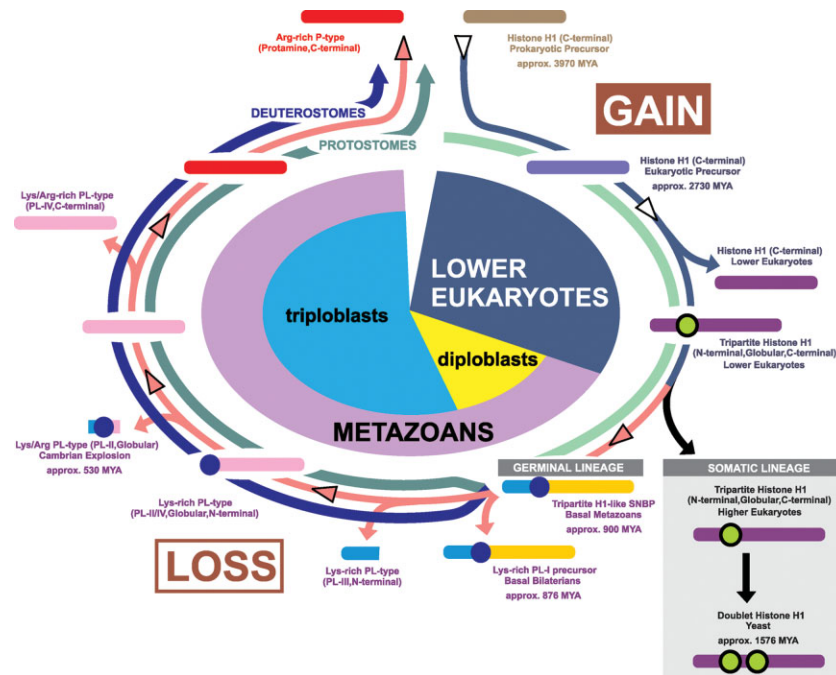
Figure 4. Representation of the "structural evolutionary cycle" undergone by histone H1, starting from its prokaryotic origin through to its eukaryotic evolution. The eubacterial lysine-rich ancestor and the early eukaryotic versions of this histone consisted initially of what later on became the C-terminal domain of metazoan H1. In metazoa, histone H1 evolved to have a more complex tripartite organization resulting from the acquisition of the winged-helix motif domain that was maintained both in the somatic and germinal variants. The evolutionary preservation of such structure was possibly due to the critical roles of the individual winged-helix domains and N- and C-terminal domains in the modulation of chromatin compaction and metabolism. In contrast, the less functionally stringent but greater and more efficient compaction of male germ cell chromatin resulted in an arginine enrichment and finally in the segregation of the C-terminal domain. This eventually led to the prevalent presence of arginine-rich protamines in the sperm chromatin of the species at the uppermost phylogenetic branches of bilaterian evolution.

cephalochordate *Branchiostoma floridae* (amphioxus) SNBPs revealed that they consist of a single lysine-rich PL fraction with homology to an N-terminal region of histone H1,[16] in which the codon used for lysine residues was exclusively the AAA triplet (Fig. 3C). This would systematically preclude a lysine-to-arginine transition based on a frameshift mutation such as that observed in tunicates[43] from being responsible for a transition toward an arginine-rich protein. Therefore, the differentiation of arginine-rich protamines must have been set up a step back from tunicates, immediately before the differentiation between tunicates and vertebrates, in order to account for the presence of protamines in both chordate lineages with a common evolutionary origin (Fig. 3D).

H1 histones and protamines – back and forth from an intrinsically disordered C-terminal domain.

In considering the processes involved in the evolution of H1 histones and their further specialization into PL-I proteins that eventually led to SNBPs of the P-type, it is important to note the curious evolutionary cycle experienced by their ancestral precursors (Fig. 4). Early in eukaryotic histone H1 evolution, the protein consisted of what is now the C-terminal region of

metazoan H1 (*i.e.*, in trypanosomes). It was initially recruited to assist with the chromatin compaction processes in these organisms.[53] The subsequent acquisition of the WHD and its full establishment in metazoans provided H1 with an enhanced specificity in terms of both chromatin packing and metabolism.[31] The mechanism of recurrent gene duplication involving this structural gain took place twice in yeast where two WHDs are present in Hho1p.[54] The concomitant differentiation of the somatic histone H1 and some of the germinal SNBP lineages led to highly specialized PL-I proteins whose genes underwent a segregation from the increasingly arginine-rich C-terminal domains. This resulted in the interesting closing of an evolutionary circle initiated with a lysine-rich intrinsically disordered C-terminal tail acquired from eubacteria.[38]

Conclusions

For almost a century since their discovery, SNBPs have been considered a group of small basic and conserved proteins with little or no interest besides their structural role in DNA packing in the male germ cells. There is now plenty of

evidence to suggest that in addition to this role, these proteins also support important functions that may impinge on critical biological processes such as sperm fertility.[18] The structural and compositional diversity of SNBPs has also been long recognized.[3,35] In recent years, the study of the massive flow of molecular and structural data obtained from genomic databases has provided strong support for the "histone hypothesis for the parallel vertical evolution of SNBPs". Accordingly, somatic "histone" H1 and SNBPs shared a common evolutionary origin very early in metazoan evolution. This was followed by a "parallel" differentiation birth-and-death DNA duplication process (rather than concerted evolution) of the somatic and germinal RI H1 histones in both protostomes and deuterostomes. In this way, the SNBP lineage underwent a "vertical" transition from a histone to a PL and finally protamine types ($H \rightarrow PL \rightarrow P$), the last one of which is present in the sperm of organisms at the uppermost tips of the phylogenetic tree. This transition involved a purifying (negative selection) of PL-type SNBP and an adaptive (positive) selection of protamines. In contrast to somatic H1s that experienced a genetic diversification without drastic changes in their overall protein structure during the evolution process, SNBPs have undergone a progressive reduction in their structural complexity. The structural commonalities between these proteins are striking, while the dissimilarity of their lysine/arginine contents may simply reflect the greater ability of arginine[32–36] to interact with DNA and achieve the efficient chromatin packing favored by the sperm of highly evolved organisms with internal fertilization.

References

1. **Ausió, J.,** Histone H1 and evolution of sperm nuclear basic proteins. *J Biol Chem* 1999. **274**: 31115–31118.
2. **Bloch, D. P.,** A catalog of sperm histones. *Genetics* 1969. **61**: 93–111.
3. **Bloch, D. P.,** Histones of sperm. In: King, R. C. editor. Handbook of genetics. New York, Plenum Press, 1976. pp. 139–167.
4. **Subirana, J. A.,** Nuclear proteins in spermatozoa and their interactions with DNA. In: André, J. editor. Proceedings of the Fourth International Symposium on Spermatology. The Hague, Maltinus Nijhoff, 1983. pp. 197–213.
5. **Kossel, A.,** Uber einen peptonartigen bestandteil des zellkerns. *Z Physiol Chem* 1884. **8**: 511–515.
6. **Arents, G. and Moudrianakis, E.,** The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. *Proc Natl Acad Sci U S A* 1995. **92**: 11170–11174.

7. **Cole, R. D.,** A minireview of microheterogeneity in H1 histone and its possible significance. *Anal Biochem* 1984. **136**: 24–30.
8. **Cole, R. D.,** Microheterogeneity in H1 histones and its consequences. *Int J Pept Protein Res* 1987. **30**: 433–449.
9. **Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L. and Sweet, R. M.,** Crystal structure of globular domain of histone H5 and its implications for nucleosome binding. *Nature* 1993. **362**: 219–223.
10. **Lewis, J. D., Song, Y., de Jong, M., Bagha, S. and Ausió, J.,** A walk though vertebrate and invertebrate protamines. *Chromosoma* 2003. **111**: 473–482.
11. **Oliva, R. and Dixon, G. H.,** Vertebrate protamine genes and the histone-to-protamine replacement reaction. *Prog Nucleic Acid Res Mol Biol* 1991. **40**: 25–94.
12. **Subirana, J. A., Cozcolluela, C., Palau, J. and Unzueta, M.,** Protamines and other basic proteins from spermatozoa of molluscs. *Biochim Biophys Acta* 1973. **317**: 364–379.
13. **Eirín-López, J. M., Frehlick, L. J. and Ausió, J.,** Protamines, in the footsteps of linker histone evolution. *J Biol Chem* 2006. **281**: 1–4.
14. **Ausió, J.,** Histone H1 and the evolution of the nuclear sperm-specific proteins. In: Jamieson, B. G. M., Ausió, J. and Justine, J. L. editors. Advances in spermatozoal phylogeny and taxonomy. Paris, Memoires du Museum National d'Histoire, Naturelle, 1995. pp. 447–462.
15. **Eirín-López, J. M., Lewis, J. D., Howe, L. and Ausió, J.,** Common phylogenetic origin of protamine-like (PL) proteins and histone H1: evidence from bivalve PL genes. *Mol Biol Evol* 2006. **23**: 1304–1317.
16. **Eirín-López, J. M., Frehlick, L. J., Chiva, M., Saperas, N. and Ausió, J.,** The sperm proteins from amphioxus mirror its basal position among chordates and redefine the origin of vertebrate protamines. *Mol Biol Evol* 2008. **25**: 1705–1713.
17. **Ausió, J.,** Structural variability and compositional homology of the protamine-like components of the sperm from the bivalve mollusks. *Comp Biochem Physiol, Part B: Biochem Mol Biol* 1986. **85**: 439–449.
18. **Ausió, J., Eirín-López, J. M. and Frehlick, L. J.,** Evolution of vertebrate chromosomal sperm proteins: implications for fertility and sperm competition. In: Roldan, E. R. S. and Gomendio, M. editors. Spermatology. Nottingham, Nottingham University Press, 2007. pp. 63–79.
19. **Munoz-Guerra, S., Azorin, F., Casas, M. T., Marcet, X., Maristany, M. A.,** et al. Structural organization of sperm chromatin from the fish *Carassius auratus. Exp Cell Res* 1982. **137**: 47–53.
20. **Meistrich, M. L., Mohapatra, B., Shirley, C. R. and Zhao, M.,** Roles of transition nuclear proteins in spermiogenesis. *Chromosoma* 2003. **111**: 483–488.
21. **van Holde, K. E.,** Chromatin. New York, NY, Springer-Verlag, 1988.
22. **Daban, J. R.,** High concentration of DNA in condensed chromatin. *Biochem Cell Biol* 2003. **81**: 91–99.
23. **Subirana, J. A.,** Protein-DNA interactions in spermatozoa. In: Baccetti, B. editor. Comparative spermatology 20 years later. New York, Serono Symposia Publications, Raven Press, 1991. pp. 89–92.
24. **Ausió, J. and van Holde, K. E.,** A dual chromatin organization in the sperm of the bivalve mollusc *Spisula solidissima. Eur J Biochem* 1987. **165**: 363–371.
25. **Casas, M. T., Ausio, J. and Subirana, J. A.,** Chromatin fibers with different protamine and histone compositions. *Exp Cell Res* 1993. **204**: 192–197.
26. **Frehlick, L. J., Prado, A., Calestagne-Morelli, A. and Ausio, J.,** Characterization of the PL-I-related SP2 protein from *Xenopus. Biochemistry* 2007. **46**: 12700–12708.
27. **Saperas, N., Chiva, M., Casas, M. T., Campos, J. L., Eirín-López, J. M.,** et al. A unique vertebrate histone H1-related protamine-like protein results in an unusual sperm chromatin organization. *FEBS J* 2006. **273**: 4548–4561.
28. **Oliva, R., de Mateo, S. and Estanyol, J. M.,** Sperm cell proteomics. *Proteomics* 2009. **9**: 1004–1017.
29. **McMurray, C. T. and Kortun, I. V.,** Repair in haploid male germ cells occurs late in differentiation as chromatin is condensing. *Chromosoma* 2003. **111**: 505–508.
30. **Eirín-López, J. M., González-Tizón, A. M., Martínez, A. and Méndez, J.,** Birth-and-death evolution with strong purifying selection in the histone H1 multigene family and the origin of *orphon* H1 genes. *Mol Biol Evol* 2004. **21**: 1992–2003.

31. **Eirín-López, J. M.**, **González-Romero, R.**, **Dryhurst, D.**, **Méndez, J.** and **Ausió, J.**, Long-term evolution of histone families: old notions and new insights into their diversification mechanisms across eukaryotes. In: Pontarotti, P. editor. Evolutionary Biology: Concept, Modeling, and Application. Berlin Heidelberg, Springer-Verlag, 2009. in press.

32. **Ausió, J., Greulich, K. O., Haas, E. and Wachtel, E.,** Characterization of the fluorescence of the protamine thynnine and studies of binding to double-stranded DNA. Biopolymers. 1984. **23**: 2559–2571.

33. **Puigdomenech, P., Martínez, P., Palau, J., Bradbury, E. M. and Crane-Robinson, C.,** Studies on the role and mode of operation of the very-lysine-rich histones in eukaryote chromatin. Nuclear-magnetic-resonance studies on nucleoprotein and histone phi 1-DNA complexes from marine invertebrate sperm. Eur J Biochem 1976. **65**: 357–363.

34. **Ohtsuki, K., Nishikawa, Y., Saito, H., Munakata, H. and Kato, T.,** DNA-binding sperm proteins with oligo-arginine clusters function as potent activators for egg CK-II. FEBS Lett 1996. **378**: 115–120.

35. **Kasinsky, H. E.,** Evolution and origins of sperm basic proteins. In: Jamieson, B. G. M., Ausió, J. and Justine, J. L. editors. Advances in spermatozoal phylogeny and taxonomy. Paris, France, Memoires de Museum National d'Histoire, Naturelle, 1995. pp. 447–462.

36. **Frehlick, L. J., Eirín-López, J. M., Prado, A., Su, H. W., Kasinsky, H. E.,** et al. Sperm nuclear basic proteins of two closely related species of Scorpaeniform fish (Sebastes maliger, Sebastolobus sp.) with different sexual reproduction and the evolution of fish protamines. J Exp Zoolog A Comp Exp Biol 2006. **305**: 277–287.

37. **Sandman, K. and Reeve, J. N.,** Archaeal histones and the origin of the histone fold. Curr Opin Microbiol 2006. **9**: 520–525.

38. **Kasinsky, H. E., Lewis, J. D., Dacks, J. B. and Ausió, J.,** Origin of H1 histones. FASEB J 2001. **15**: 34–42.

39. **Malik, H. S. and Henikoff, S.,** Phylogenomics of the nucleosome. Nat Struct Biol 2003. **10**: 882–891.

40. **Nei, M. and Rooney, A. P.,** Concerted and birth-and-death evolution in multigene families. Annu Rev Genet 2006. **39**: 121–152.

41. **Ausió, J. and Subirana, J. A.,** A high molecular weight nuclear basic protein from the bivalve mollusc Spisula solidissima. J Biol Chem 1982. **257**: 2802–2805.

42. **Carlos, S., Hunt, D. F., Rocchini, C., Arnott, D. P. and Ausió, J.,** Post-translational cleavage of a histone H1-like protein in the sperm of Mytilus. J Biol Chem 1993. **268**: 195–199.

43. **Lewis, J. D., Saperas, N., Song, Y., Zamora, M. J., Chiva, M.,** et al. Histone H1 and the origin of protamines. Proc Natl Acad Sci U S A 2004. **101**: 4148–4152.

44. **Jankowski, J. M., States, J. C. and Dixon, G. H.,** Evidence of sequences resembling avian retrovirus long terminal repeats flanking the trout protamine gene. J Mol Evol 1986. **23**: 1–10.

45. **Ausió, J., Van Veghel, M. L., Gómez, R. and Barreda, D.,** The sperm nuclear basic proteins (SNBPs) of the sponge Neofibularia nolitangere: implications for the molecular evolution of SNBPs. J Mol Evol 1997. **45**: 91–96.

46. **Rocchini, C., Marx, R. M., Carosfeld, J. S., Kasinsky, H. E., Rosenberg, H.,** et al. Replacement of nucleosomal histones by histone H1-like proteins during spermiogenesis in Cnidaria: evolutionary implications. J Mol Evol 1996. **42**: 240–246.

47. **Rocchini, C., Zhang, F. and Ausió, J.,** Two highly specialized histone H1 proteins are the major chromosomal proteins of the sperm of the sea anemone Urticina (Tealia) crassicornis. Biochemistry 1995. **34**: 15704–15712.

48. **Rocchini, C., Rice, P. and Ausió, J.,** Complete sequence and characterization of the major sperm nuclear basic protein from Mytilus trossulus. FEBS Lett 1995. **363**: 37–40.

49. **Swanson, W. J. and Vacquier, V. D.,** The rapid evolution of reproductive proteins. Nat Rev Genet 2002. **3**: 137–144.

50. **Rooney, A. P., Zhang, J. and Nei, M.,** An unusual form of purifying selection in a sperm protein. Mol Biol Evol 2000. **17**: 278–283.

51. **Bourlat, S. J., Juliusdottir, T., Lowe, C. J., Freeman, R., Aronowicz, J.,** et al. Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. Nature 2006. **444**: 85–88.

52. **Dunn, C. W., Hejnol, A., Matus, D. Q., Pang, K., Browne, W. E.,** et al. Broad phylogenomics sampling improves resolution of the animal tree of life. Nature 2008. **452**: 745–749.

53. **Burri, M., Schlimme, W., Betschart, B. and Hecker, H.,** Characterization of the histones of Trypanosoma brucei brucei bloodstream forms. Acta Trop 1994. **58**: 291–305.

54. **Ali, T. and Thomas, J. O.,** Distinct properties of the two putative "globular domains" of the yeast linker histone, Hho1p. J Mol Biol 2004. **337**: 1123–1135.

55. **Ausió, J., Saperas, N. and Chiva, M.,** Sperm nuclear basic proteins and sperm chromatin organization in fish. In: Tiersch, T. and Green, C. editors. Cryopreservation in aquatic species, 2nd edition. Baton Rouge, Louisiana, World Aquaculture Society, 2009.

56. **Telford, M. J.,** Resolving animal phylogeny: a sledgehammer for a tough nut? Dev Cell 2008. **14**: 457–459.