Protamines, in the Footsteps of Linker Histone Evolution

Jose´ Marı´a Eirı´n-Lo´ pez, Lindsay J. Frehlick, and Juan Ausi´o

From the Department of Biochemistry and Microbiology, University of Victoria, Victoria, British Columbia V8W 3P6, Canada

Histones, Protamines, and Protamine-like Proteins

It was perhaps a lucky coincidence that the early attempts to establish the chemical composition of the cell nucleus were initially carried out on such diverse biological systems as salmon sperm heads (1) and geese and chicken erythrocytes (2). Examination of sperm and erythrocyte systems, respectively, lead to the protamine and histone concepts (3). We know with certainty that, in contrast to the somatic nucleus, sperm lack any chromosomal proteins in their sperm, and some crustaceans (order Decapoda) lack any chromosomal proteins in their sperm (6, 7). Therefore, in contrast to the somatic nucleus, sperm chromatin may have a much more diverse protein composition. It was not until the first attempt of classification of the sperm nuclear basic proteins (SNBPs) by David Bloch (6, 8), an effort later on extended by Harold Kasinsky (7), that a clearer picture started to emerge in this regard.

More recently, an enormous effort has been carried out in several laboratories, including our own (9–15), to extend this analysis to a large number of representative organisms from the different phylogenetic groups. With a broader perspective now available, SNBP heterogeneity can be restricted to three major groups or types: histone (H), protamine (P), and protamine-like (PL) (16).

Protamines are highly compositionally and structurally heterogeneous group of proteins (9). They exhibit a high charge density and a prevalence of arginine in their composition (13), a fact that is most likely related to the highly differentiated histone H5 from bird erythrocytes, another terminally differentiated system (22). A SNBP protein with similar characteristics is extensively represented in many organisms, including our own (9–15), to extend this analysis to a large number of representative organisms from the different phylogenetic groups. One such example can be found in mollusks. An early comparative study of SNBPs within this group (23) revealed the presence, in some instances, of electrophoretically large proteins with lower mobility than histones and a composition rich in both arginine and lysine residues, such as those described in the clam Spisula solidissima. In other instances, such as in the cephalopods Octopus vulgaris and in Eleodes cirrhosa, the SNBPs had higher electrophoretic mobility and a composition that ranged from arginine-rich, such as in the fish and bird protamines, to highly cysteine-rich, such as in mammalian protamines. The proteins of these three organisms have now all been sequenced (24–26).

An initial structural characterization of the Spisula large SNBP component showed that the protein had a tripartite organization with a globular central core, which was found to bear a strong sequence similarity to the winged helix domain of histone H5 from bird erythrocytes, another terminally differentiated system (27). A SNBP protein with similar characteristics is extensively distributed throughout bivalve mollusks and has been called protamine-like protein (1L) (28, 29). The regions flanking this core are unstructured and are rich in both arginine and lysine (26).

PL-1 proteins have now been identified not only in mollusks but also in tunicates and in several fish (9, 30–32) where they represent the major SNBP component. In all instances (see Fig. 1A) they contain an internal folded domain that corresponds to the winged helix motif (33, 34), which is characteristic of the linker histones.

Notably, in the cases of mollusks (35) and tunicates (30) the PL-1 protein can undergo post-translational cleavage giving rise to a series of smaller PL proteins with increasingly higher arginine composition. In some instances, as in Mytilus (mussel), PL-III appear to have become independent genes. These phenomena have been taken as an indication that all PL proteins and possibly protamines are somehow related to a primitive linker histone precursor (16) probably related to the replication-independent (RI) lineage that gave rise to the highly differentiated histone H5 from the nucleated vertebrate erythrocytes (36, 37).

Interestingly, a potential structural relation between histone H5 and protamines has also been described in other invertebrates. Although there is still very little information about the protamines of insects (13), a putative Drosophiila protamine-like protein, which shares some extent of similarity to histone H5 and to the cysteine-rich protamines from mammals, has been identified in screens of transcripts expressed in the male germ line (38).

Protamine-like Proteins from Tunicates and the Lysine to Arginine Transition

One of the major conceptual stumbling blocks in trying to explain the transition from linker histones to protamines has been the difficulty in accounting for the evolutionary transition from the highly lysine-rich (25–30 mol %) composition, which is characteristic of histone H1 molecules (39), to the arginine-rich (30 mol %) composition of protamines. Although all PL proteins exhibit both a lysine- and arginine-rich composition (Arg + Lys = 35–50 mol %), they still have a distinct composition from the predominantly arginine-rich protamines.

An important breakthrough in this direction came from a recent study of the PL proteins from two closely related tunicates: Styela meterreyensis and Ciona intestinalis (30). The former contains an SNBP composition consisting of two PL-1-related proteins P1 and P2 (10, 40). Amino acid sequence analysis showed that these two proteins are indeed related, with the faster electrophoretic component corresponding to the C-terminal domain of the larger
component (Fig. 1B). Furthermore, the faster PL component (P2) had an arginine-rich composition (58 mol %) (10), and more importantly, it consisted of repeated arginine clusters, which are characteristic of many invertebrate and vertebrate canonical protamines (13). However, this SNBP composition appeared to be quite restricted to the genus *Styela*, as other tunicate species consisted only of the larger PL precursor (P1) molecule that apparently had not undergone the post-translational cleavage (10).

*In silico* analysis based on the genome sequence available for the tunicate *Ciona intestinalis* revealed that the single unprocessed PL-I, which is present in this species, had an amino acid sequence that was strikingly similar to that of the *Styela* larger component except for the fact that its C-terminal region was lysine-rich (Fig. 1B). Careful analysis of the genomic nucleotide sequence encoding for *Styela* and *Ciona* PL-I show that the transition from lysine to arginine may have occurred as a result of a single frameshift mutation (30). The mutation occurred at a point (indicated by a blue star) in the C-terminal domain of a lysine-rich precursor of the tunicate *C. intestinalis*. The corresponding amino acid sequence alignments are shown below. Changes from lysine to arginine residues are highlighted by yellow and red shading. The *Ciona* P1 sequence was identified from the draft genome sequence from *C. intestinalis* with the help of a BLAST search using the *Styela* P1 sequence as a template (30). The red arrow points to the site of post-translational cleavage.

This rapid mechanism of evolution is in good agreement with the notion that the reproductive traits (including reproductive proteins) have evolved very quickly (41) and with the experimental evidence that indicates that, despite their rather simple amino acid composition and their high arginine contents, protamines are excellent molecular markers for evolution studies (13).

**Evolution of Histone H1 and Evolution of Protamines, How Related Are They?**

It is interesting to notice that in contrast to core histones, the evolutionary origins of which can now be traced back to archaeabacteria (42, 43), the origin of
Histone H1 appears to have occurred earlier in eubacteria (44) (Fig. 2). Equally interesting is the fact that although archaeal histones contained a histone fold structure characteristic of core histones (45), they lacked the tails flanking this domain found in higher eukaryotes (42). However, linker histones acquired the winged helix folded domain characteristic of higher eukaryotes in a reverse way. In other words, linker histones were initially composed only of a C-terminal region, and the acquisition of the core domain containing the winged helix motif occurred later in their evolution (Fig. 2) (44). Indeed, many protozoans contain a linker histone consisting only of the characteristic APK-rich C-terminal domain, which is critical for the stabilization of the folded chromatin structure (46).

The long term evolution of the histone H1 family was recently shown to be best described by a birth-and-death process (36, 37, 47), a mechanism based on recurrent gene duplication events under a strong purifying selection. This mechanism has favored the great diversification presented by the members of this family and was further enhanced by the presence of strong functional and structural constraints, which ultimately led the different H1 isoforms to the acquisition of specific functions (Fig. 2).

Histone H1 diversification has been maintained throughout its evolutionary process in both plants and animals (48), ranging from the high degree of heterogeneity observed in early protozoans, such as trypanosomes (49), to extreme diversification and specialization in the case of mammals. To date, 11 different mammalian linker histones have been identified: 7 somatic variants, H1.1 to H1.5, H10, H5; 3 sperm-specific variants, H1t, H1t2, and H10o; and the oocyte-specific variant H1foo (56). In addition, H1 evolution has also favored the differentiation of highly specialized isoforms such as histone H5 (57), an H1 replication-independent isoform restricted to terminally differentiated erythrocytes of birds, which also appears to be present in amphibians (58) and reptiles. The PL-I sperm-specific proteins, which appear at the end of spermiogenesis (yet another terminal differentiation process) in some vertebrate and invertebrate organisms (16), would also belong to this classification. PL-I protein evolution and the possible link to protamine evolution is summarized in Fig. 2, where a hypothetical model involving the loss of the winged fold domain upon transition from lysine to arginine in precursor PL-I proteins is shown. Such a loss could be speculatively attributed to a gene duplication process, which has been common in both protamine evolution (59) and H1 evolution (36). Significantly, the PL genes in bivalve mollusks have been shown to occur in hyper-variable restriction fragment length polymorphism (RFLP) regions (60).

In addition, H1 evolution has also favored the differentiation of highly specialized isoforms such as histone H5 (57), an H1 replication-independent isoform restricted to terminally differentiated erythrocytes of birds, which also appears to be present in amphibians (58) and reptiles. The PL-I sperm-specific proteins, which appear at the end of spermiogenesis (yet another terminal differentiation process) in some vertebrate and invertebrate organisms (16), would also belong to this classification. PL-I protein evolution and the possible link to protamine evolution is summarized in Fig. 2, where a hypothetical model involving the loss of the winged fold domain upon transition from lysine to arginine in precursor PL-I proteins is shown. Such a loss could be speculatively attributed to a gene duplication process, which has been common in both protamine evolution (59) and H1 evolution (36). Significantly, the PL genes in bivalve mollusks have been shown to occur in hyper-variable restriction fragment length polymorphism (RFLP) regions (60).

The obvious changes in the structural properties of the proteins, derived from the lysine to arginine transition, determined the new mechanisms underlying protamine evolution. It is their high arginine content that allows protamines to tightly condense chromatin in the sperm nucleus. This feature represented a new (and the most important) constraint driving their evolution, which differed to that presented by somatic H1 proteins. In fact, both positive Darwinian (adaptive) selection (61) and an unusual form of purifying selection (62) are the major mechanisms to which protamines are subject in their evolutionary process.

The lysine to arginine transition and gene segregation may have taken place several times in the course of evolution. Whether this has involved a mecha-
MINIREVIEW: Histone H1 and the Evolution of Protamines

nism similar to that described in tunicates remains to be established. In this regard, it is interesting to note that the process of post-translational cleavage of PL-1 precursors has occurred repeatedly in completely unrelated groups of organisms such as bivalve mollusks and ascidian tunicates. Remarkably, in both instances the next step in the evolution of these two groups, cephalopods (25, 63) and cephalochordates (10), has involved the acquisition of an independent protamine gene encoding for a protein with characteristics almost identical to those of the PL-1 arginine-rich fragments.

In conclusion, if it can finally be proven that protamines with independent genes are related to linker histones through the process described above, this relationship would have resulted in the closure of an interesting evolutionary cycle in which the C-terminal domain of linker histones would have returned to the initial independent existence of its eubacterial/protozoan ancestry (see Fig. 2).

REFERENCES

1. Miescher, F. (1874) Hoppe-Seyler’s J. Physiol. 8, 511–515
2. Kossel, A. (1984) Z. Physiol. Chem. 8, 511–515
3. Kossel, A. (1928) The Protamines and Histones, Longmans Green and Co., London
4. Felix, K. (1960) Adv. Protein Chem. 15, 1–56
5. Ando, T., Yamashiki, M., and Suzuki, K. (1973) Protamines: Isolation, Characterization, and Function, Springer-Verlag, New York
6. Bloch, D. P. (1976) Histones of Sperm, Plenum Press, New York
7. Kasinsky, H. E. (1989) in Histones and Other Basic Nuclear Proteins (Hnilica, L. S., Stein, G. S., and Stein J., L., eds) pp. 73–163, CRC Press Inc., Boca Raton, FL
8. Bloch, D. P. (1969) Genetics 61, (suppl.) 93–111
9. Ausiò, J. (1999) Mol. Cell. Biochem. 115, 163–172
10. Ausiò, J. (1991) Comp. Biochem. Physiol. B 98, 439–449
11. Lewis, J. D., Saperas, N., Song, Y., Zamora, M. J., Chiva, M., and Ausiò, J. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 4148–4152
12. Waterson, C. E., and Davies, P. L. (1998) J. Biol. Chem. 273, 6157–6162
13. Watson, C. E., Gauthier, S. Y., and Davies, P. L. (1999) Eur. J. Biochem. 262, 258–267
14. Graziano, V., Gerchman, S. E., Winacott, A. J., Sweet, R. W., Wells, J. R., White, S. W., and Ramakrishnan, V. (1990) J. Mol. Biol. 212, 253–257
15. Ramakrishnan, V., Finch, J. T., Graziano, V., Lee, P. L., and Sweet, R. M. (1993) Nature 362, 219–223
16. Carlos, S., Hunt, D. F., Rocchini, C., Aronst, D. P., and Ausiò, J. (1993) J. Biol. Chem. 268, 195–199
17. Eirin-Lopez, J. M., Gonzalez-Tizon, A. M., Martinez, A., and Mendez, J. (2004) Mol. Biol. Evol. 21, 1992–2003
18. Eirin-Lopez, J. M., Ruiz, M. F., Gonzalez-Tizon, A. M., Martinez, A., Ausiò, J., Sanchez, L., and Mendez, J. (2005) J. Mol. Biol. 31, 398–407
19. Russell, S. R., and Kaiser, K. (1993) Genetics 134, 293–308
20. Cole, R. D. (1984) Anal. Biochem. 136, 24–30
21. Fischle, W., Wang, Y., and Allis, C. D. (2003) EMBO Rep. 4, 311–315
22. Prado, A., Ramos, I., Frehlick, L. J., Muga, A., and Ausiò, J. (2004) J. Biol. Chem. 279, 36410–36417
23. Ausiò, J., Tournadre, A., McParland, R., Becker, R. R., Johnson, W. C., and van Holde, K. E. (1987) Bioc hemistry 26, 975–982