Revista Chilena de Historia Natural
74: 167-175, 2001

Evolution of the histones: free play with exon shuffling

Evolucion de las histonas: Juego libre con reordenamiento de exones al azar

G. CECILIA TORO

Programa de Biología Molecular y Celular, Instituto de Ciencias Biomédicas, Facultad de Medicina, Universidad de Chile, Santiago, Chile, e-mail: ctoro@machi.med.uchile.cl

ABSTRACT

In higher eukaryotes, the nuclear DNA is organized for transcription, replication and mitosis in competent chromatin and chromosomes. The basic unit of chromatin is the nucleosome. This entity is formed by 168 base pairs of DNA wound around an octamer of histones, this octamer of histones consist of two copies of H2A, H2B, H3 and H4. The DNA is sealed in its input and output point by a histone linker: histone H1. Histones were supposed to be very conserved proteins. However, during the past few years it was found that these proteins present a high degree of divergency in several lower eukaryotes. In *Trypanosoma*, it was found that histones H3 and H4, which are at the center of the nucleosomal organization, showed more than 30 % of divergency, while histone H1 corresponded to only one of the three peptide domains present in higher eukaryotes. These features of *Trypanosoma* histones may explain, at least in part, the unability of chromatin to condense into chromosomes during the cell division in these parasites. Evolution of histones was usually considered as peculiar, with several proposals which are difficult to reconcile with experimental data. In the present work, it is proposed that histones followed the same evolutionary route as many other proteins. Considering that exons code for structural and functional domains in proteins and that, at the origin of eukaryotes, the histones, as other proteins, could be formed by "units" (mecano theory), it was expected that these units or domains eventually would be found in living organisms exhibiting primitive features. Furthermore, those units could work independently. Our results on the structure of *Trypanosoma cruzi* histone genes and proteins as well as the analysis of other histones from different species fit with this proposal.

Key words: histones, evolution.

RESUMEN

En los eucariontes superiores, el DNA nuclear se organiza en cromatina competente y en cromosomas para la transcripción, replicación y mitosis. La unidad básica de la cromatina es el nucleosoma, formado por 168 pares de bases de DNA enrollados en el octámero de histonas, el cual consiste en dos copias de las histonas H2A, H2B, H3 y H4. El DNA es sellado en los puntos de entrada y salida por una histona de unión: la histona H1. Se supone que las histonas son proteínas muy conservadas. Sin embargo, durante los últimos años se ha encontrado que estas proteínas presentan un alto grado de divergencia en varios eucariontes inferiores. En *Trypanosoma* se ha encontrado que las histonas H3 y H4, que están en el centro de la organización nucleosomal, muestran más de un 30 % de divergencia, y que la histona H1 corresponde sólo a uno de los tres dominios presentes en los eucariontes superiores. Estos rasgos en *Trypanosoma* podrían explicar, al menos en parte, la ausencia de condensación de la cromatina en cromosomas durante la división celular en estos parásitos. La evolución de las histonas ha sido considerada como algo peculiar, con varias propuestas difíciles de reconciliar con los datos experimentales. En este trabajo se propone que las histonas han seguido la misma historia evolutiva de muchas otras proteínas. Si consideramos que los exones codifican para dominios estructurales y funcionales en las proteínas y que, en el origen de los eucariontes, las histonas, al igual que otras proteínas, podrían haberse formado por unidades (teoría tipo mecano), se podría esperar que estas unidades o dominios eventualmente fuesen encontrados en organismos actuales que exhiban características primitivas. Mas aún, estas unidades podrían funcionar independientemente. Nuestros resultados relativos a la estructura de los genes de histonas y de sus proteínas en *Trypanosoma cruzi*, y el análisis de otras histonas de diferentes especies confirman la proposición.

Palabras clave: histonas, evolución.

INTRODUCTION

An important number of histone genes and proteins from *Entamoeba* (Binder et al. 1995, Fodinger et al. 1993), *Trypanosoma* (Toro et al. 1992, Toro et al. 1993, Aslund et al. 1994), ciliata (Wu et al. 1986), fungi (May & Morris 1987, Woudt et al. 1983), and intermediate eukaryotes, such as nematoda (Vanfleteren 1988) and annelida (Kmiecik et al. 1985), have been sequenced. All these histones sequences show peculiarities, including a high degree of divergence.

Sequence analysis does not sustain a conclusion as to a common origin of all the histones. For this reason, different evolutive plans involving H2A, H2B, H3 and H4 have been proposed,

(Wuilmart & Wyns 1977, Reeck et al. 1978, Brown 1983). On the other hand, sequence similarity between H1 and nucleosomal histones has not been found.

Van Holde (1989) has proposed the existence of one or more primitive DNA binding proteins in ancient organisms that were the predecessors of all three lines of descent. Such proteins could have evolved independently producing the HU protein in eubacteria, the HTa type proteins of archaebacteria and the histones of eukaryotes. If this were the case, the ancestral molecule should have been shorter than the contemporary ones. Considering, for example, that an HTa segment shows homology with both halves of H2A, each of about 60 residues in length, a gene doubling must be postulated in the H2A evolution from the hypothetical ancestor.

Nevertheless, examining the sequences of the core histones, three groups of investigators have postulated different relationships of lineages between them. Wuilmart & Wyns (1977) proposed that H2A and H2B derived from a proto H4; however these authors did not find an evident relationship between H3 and the rest of the histones. Reeck et al. (1978) found homologies between all four core histones, but the regions described by them as homologous are different than those proposed by Wuilmart & Wyns (1977). On the other hand, Brown (1983) have found that all the histones of the nucleosome core show evidence of a repeating tetrapeptide, which is better seen in H3.

All these contradictory proposals made Van Holde (1989) to state that more information from more histone sequences is required to resolve the problem, if a solution is indeed possible.

At present, it was demonstrated that H2A, H2B, H3 and H4 monomers all have a common globular domain, named as the "histone fold", consisting of three amphipathic a-helices separated by short loops and b-strand regions, originally described by Arents et al. (1991), Arents & Moudrianakis (1995) and reviewed by Luger et al. (1997). The central globular domain of each histone is flanked by N-and C terminal regions, that extend beyond the "histone fold", and even beyond the core of the nucleosome, playing roles in gene expression and contributing to the higher order structure of eukaryotic chromatin (Luger & Richmond 1998). Then, the histones present three domains which are structurally and functionally defined.

On the other hand, archaeal histones were first identified in the hyperthermophile *Methanothermus fervidus* (Sandman et al. 1990). The amino acid sequences of HMfA and HMfB, (the histones from *M. fervidus*) are related to those of the four eukaryotic nucleosome core histones, which facilitated an alignment of these histone sequences. This has now been validated by 3D structural studies (Starich et al. 1996). Archaeal histones range from 66 to 69 residues in length and they are essentially only histone folds. At present, histones are found in Eukarya and in only one branch of Archaea, namely the Euryarchaeota (Pereira & Reeve 1998).

Data available in relation to the "histone fold" from archaeal and eukaryotic histones show that H2A, H2B, H3 and H4 are more similar to HMfB than to each other. This fact is consistent with the archael sequence resembling a common ancestor of the four eukaryal core histones. In consequence, they should be considered as homologous, and the evolution of the contemporary "histone fold" probably began with the duplication of a sequence that encoded this motif (Sandman et al. 1998). This duplication is predicted by the 3D structure of contemporary histones, but it is no evident in their amino acid sequences, which must subsequently have undergone extensive divergence. At present, this information leaves without explanation the origin of the three domains in the histone molecules.

On the other hand, Thatcher & Gorovsky (1994) had done a detailed analysis in constructing phylogenetic trees for the H2A, H2B, H3 and H4 histones. They show that H3 and H4 are much more conserved than H2A and H2B, and that the histones forming dimers (H2A and H2B, H3 and H4) have similar reconstructed phylogenies and appear to have coevolved. Moreover, they propose independent evolutionary histories for histone variants (Thatcher et al. 1994).

On the contrary, H1 sequences do not seem related to the sequences of the nucleosomal histones. It also presents a central globular domain and two flanking N- and C-terminal domains, but the central globular domain does not present the "histone fold" (Hill et al. 1989, Maeder & Bohm 1991). This, linked to the fact of the rapid evolution of this protein, darken further its origin.

PROTEIN EVOLUTION

Protein diversity is a consequence of protein evolution. In general, it has been proposed that the evolution of proteins involves changes in residues of single amino acids, insertions and deletions of several residues, and gross changes which result from gene multiplication and gene fusion.

One of the primary mechanisms of protein evolution is gene duplication, in which a cell comes to include two copies (or more) of a single gene. One copy retains its original function, in such a

way that the organism viability is not committed by the lack of an essential protein. The other copy can receive mutations without constraint from natural selection, creating a non functional protein or a protein with a new function (Doolittle & Bork 1993).

On the other hand, Rossman et al. (1971) and Rossman & Liljas (1974) studying several enzymes whose general structures were different, have found that all of them presented a domain with a common folding pattern, having in common the ability to bind certain coenzymes. They postulated that this domain was the ghost of a primitive protein from pre-cellular times (Doolittle & Bork 1993). It seems likely that the first functioning proteins were small and their greatest importance resided in its capacity of binding other molecules. The rudiments of the catalysis could be initiated. Once started, a succession of genic duplications could lead to the establishment of a family of stable proteins. At the beginning, the aggregate of proteins could be inefficient.

Soon after Rossmann suggested that the primitive proteins might have been created by the fusion of useful domains, DNA recombinant studies led to the discovery that the eukaryotic genes are not continuous, but consist of segments that encode part of a protein's structure (exon) separated by segments of noncoding DNA (introns). In some cases the introns were located at or near the boundaries of a protein domain. This correspondence led Gilbert (1978) to propose that the exons are the genomic equivalent of the interchangeable protein parts postulated by Rossmann.

Gilbert suggested that the mosaic structure of eukaryotic genes was the basis of exon shuffling, the recombination of different coding regions through crossing-over. This could bind exons in different transcriptional units and create different products, thus accelerating evolution. Blake (1978) refined this idea by arguing that exons might code for discrete, stable regions of proteins (the domains). Exon might therefore assemble new proteins through the novel reassortment of stable substructures.

According to Gilbert (1985), not only the first proteins were created by assembling stable domains, but the evolution has also maintained the genetic isolation of the domains during the course of time. This organization has the advantage that the continuous assembling of domains in new combinations can give place to new and more complex proteins and create new functions.

The modular assembling vision of extinguished genes is based on the structural common features of superfamilies of genes such as those of the superfamilies of the immunoglobulins (Sakano et al. 1979), which represent the best support for the exon shuffling, both in structural and functional terms. It is also reflected in the history of the lysozyme (Jung et al. 1980), globin (Gò 1981), glyceraldehyde phosphate dehydrogenase (Stone et al. 1985), and so on. In other superfamilies, the old genic structure intron-exon is still visible in some proteins while other members of the same family have lost the introns. Precise excision of an intron has been demonstrated; in this case, a gene is transcribed into RNA, its introns are spliced out, and the processed message returns to the genome via reverse transcription. Other prediction of this theory is the re-use of an exon, and this may be found in the structure of the LDL receptor (lipoprotein of low density) and of the EGF precursor (epithelial growth factor) (Sudhof et al. 1985)

The ancestral character of the introns is based on data which suggests that introns antedate the plant-animal divergence (Marchionni & Gilbert 1986, De Souza et al. 1998). Furthermore, the intron-exon structure can even precede the endosymbiotic incorporation of chloroplasts and mitochondria since these genomes present genes with introns (Obaru et al. 1988). Actually, these organelles must be far more ancient than scientists had thought. They may have originated at the same time or even earlier than the cell nucleus (Vogel 1997, Vellai et al. 1998). Moreover, Dorit et al. (1990) have proposed that the original exons could have had fifteen to twenty amino acids and, due to processes of intron loss, the most complex actual exons would have been originated.

Doolittle (1978) and Gò (1981) speculate that at some early stage in the history of all genes, modules would have been separated by introns and exigencies of space and energetics served to remove introns from prokaryotes, whereas removal in eukaryotes is much less complete. At present several glues from archeobacteria still have introns. On the other hand, Ohno (1981) also postulates oligomeres as the basic coding sequences considering that many polypeptide chains present recurrences implying the repeat of its original coding sequence.

PROPOSAL

The analysis of the data obtained from *Trypanosoma cruzi* histones, points to a group of divergent histones. Considering the early divergence of the flagellates within the eukaryotic lineage (McLaughlin & Dayhoff 1973, Woese et al. 1990)

and that the origin of eukaryotes occurred approximately 2,100 million years ago (Han & Runnegar 1992), their chromosomal proteins may provide clues to the molecular evolution events that must have occurred during the transition from archaeal to eukaryal histones. Evolution is change as a function of time; the dimension of the phenomenon is time. The change is divergence starting from a point of common origin. Therefore, I propose here, a reconstruction of the sequence of events that produced the histones of higher eukaryotes.

It has been demonstrated that histone H1 and all its variants on higher eukaryote, have a tripartite domain structure (Bradbury et al. 1975). One is a central globular domain which binds to DNA at the point in which it enters and leaves the nucleosome, sealing the two turns of DNA (Allan et al. 1980). This globular domain is flanked by amino and carboxy- (N and C) terminal tails domains. These three domains divide the H1 molecule in regions more or less conserved which are chemically, structurally and functionally different (Maeder & Bohm 1991). The available sequences show enormous variations in the N- and C-terminal domains and little variation in the central globular domain. Moreover, alignments of the amino terminal region of the H1 sequences using the recurrent prolines as alignment points has led different authors to postulate an amino-terminal domain with sequence conservation (Bohm & Mitchell 1985). On the other hand, different studies show that the C-terminal domain in the histone H1 is structurally conserved on the basis of the regularities that appear in the distribution of lysine, alanine and proline, which is non-random (Hill et al. 1989). A probable function has been postulated for each domain in relation to their participation in the organization of the nucleosome and of the chromatin. Therefore, in higher eukaryotes we found an histone H1 composed by three different domains, structural and functionally defined as such.

If it is postulated that exons code for domains and functions and that, in the origin, the proteins were formed by these "units" (mecano type theory), it would be expected that these units should be found eventually in organisms preserving primitive features, and, furthermore, that these units could work independently. If Gilbert and Blake's ideas are correct, it would be possible to find, in different histones, common units of folding structure and function derived from an ancestral exon, combined with different units. Therefore, if histones domains represent relics of the process that originally formed these proteins, histones may be analyzed through each domain and not as whole molecules.

Our results on the structure of *Trypanosoma cruzi* histone H1 protein and gene (Toro et al. 1993, Aslund et al. 1994) fit this proposal. We have found that this protein is solely composed by the C-terminal domain of the homologous protein in higher eukaryotes as shown in the deduced aminoacid sequences from genomic and cDNA clones (Fig. 1). These deduced sequences fit with the observed molecular mass of *T. cruzi* histone H1 protein. The same occurs with histone H1 from *Trypanosoma brucei* (Burri et al. 1995), and with histone H1 from *Leishmania* (Fasel et al. 1994, Espinoza et al. 1996).

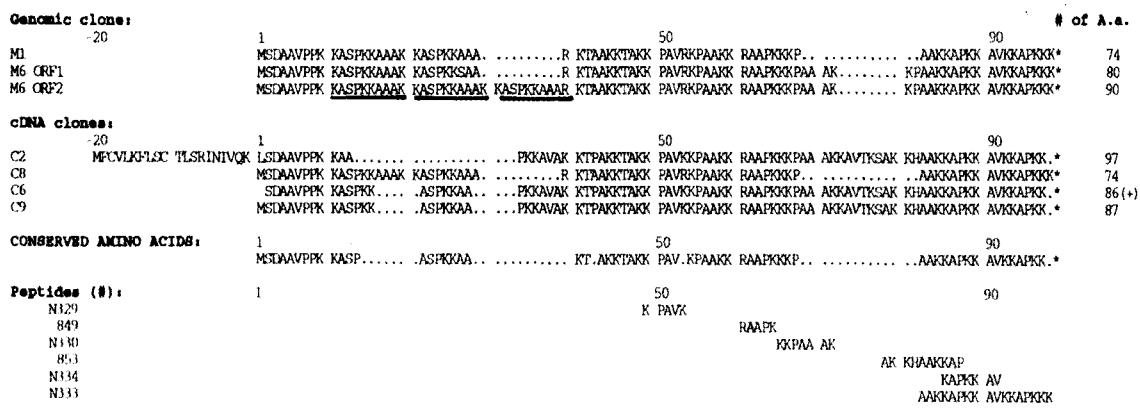

Fig. 1. Alignment of the H1 sequences of *T. cruzi* derived from genomic and cDNA clones according to Aslund et al. (1994). All of them correspond to the C-terminal domain of the homologous protein in higher eukaryotes.

Alineación de las secuencias H1 de *T. cruzi* derivadas de clones genómicos y de ADNc de acuerdo a Aslund et al. (1994). Todas las secuencias corresponden al dominio c-terminal de proteínas homólogas en eucariontes superiores.

At present there is no report on a complete genomic histone H1 sequences homologous to higher eukaryotes, in spite of we can not discard its presence. Moreover, the events described here to explain the formation of a protein, do not correspond with the phylogenetic position of a particular specie, in spite of the proteins present in some species could reflect, how the protein appeared. Since *Trypanosoma* present all core histones with their complete domains, it could be possible that complete histones H1 genes exist and by an RNA processing event in the past, such a mutation (in the acceptor site of Trans-splicing, etc.) could be generate different messages from the gene (the C-terminal domain in this case) and then, this domain, successfully selected, had been incorporated into the genome by reverse transcription. This domain could be selected because it had, by origin, the ability of being an structural and functional domain.

On the other hand, *Tetrahymena thermophila* macronuclei contains a histone H1 composed of two domains, with an intervening sequence of 254 base pairs which interrupts the coding region between the two domains of the single gene for histone H1 of this ciliated protozoan (Wu et al. 1986). Again, numerous lines of evidence suggest that ciliates are an ancient group (Cloud & Glaessner 1982, Horowitz & Gorovsky 1985). In the annelid *Platynereis dumerilu*, which presents a histone H1 with special structural characteristics, we have another step in the history of this chromosomal protein (Kmiecik et al. 1985). Figure 2 shows the proposal for an emerging history of histone H1.

In Fig. 2 the *Trypanosoma cruzi* histone H1 appears as consisting of only one domain which would be coded by the ancestral exon, corresponding only to the C-terminal domain of the homologous histone H1 from higher eukaryotes (Aslund et al. 1994). The *Trypanosoma cruzi* histone H1 protein could reflect the ancestral type. Moreover, it is noticeable that this domain in itself also presents ancestral features, such as the presence of an oligomer of ten amino acids which is repeated three times in the domain (underlined in the figure). This feature is described by Gilbert (1986) and by Ohno (1981), as belonging to proteins that maintain archaic characteristic. At

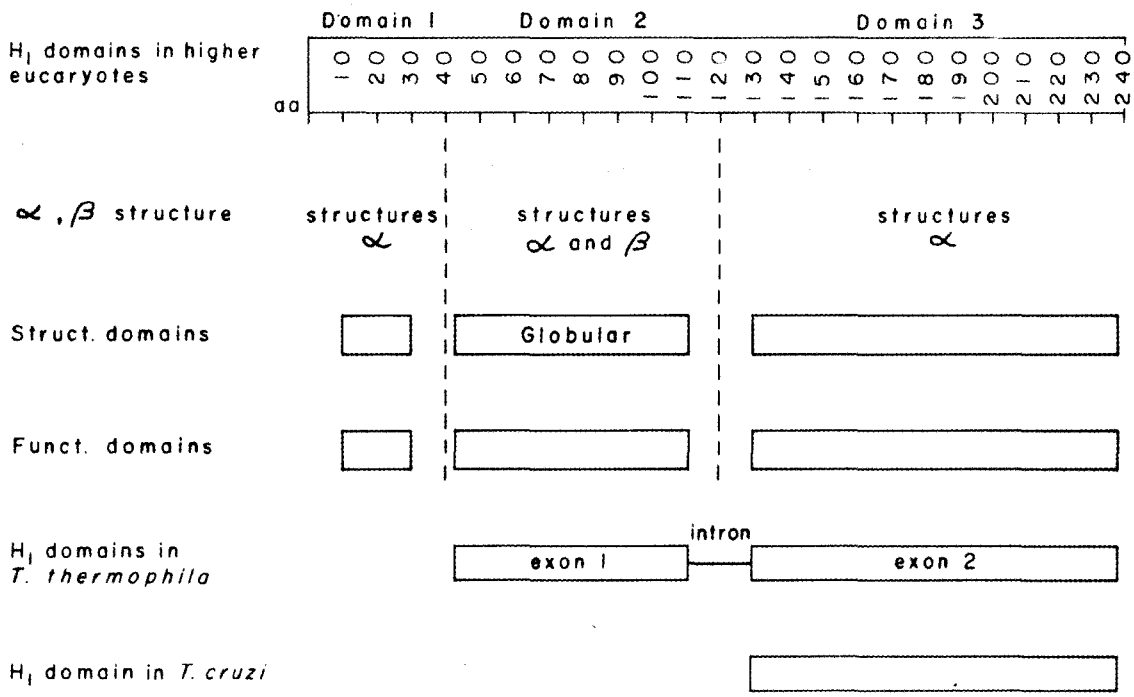

Fig. 2. Correlation between structural and functional parts of histone H1 and the domains of this protein. A linear diagram is shown of the structural elements (a-helical and b-sheet) and the structural and functional domains of the protein that have been described in higher eukaryotes. *Trypanosoma thermophila* presents two domains and *T. cruzi* presents only the C-terminal domain.

Correlación entre los elementos estructurales y funcionales de la histona H1 y los dominios de esta proteína. Se muestra un diagrama lineal de los elementos estructurales (alfa-hélice y lámina-beta), y de los dominios estructurales y funcionales de la proteína que han sido descritos en eucariontes superiores. *Thetraymena thermophila* presenta dos dominios, mientras que *T. cruzei* presenta solo el dominio c-terminal.

present, the philogenetic relationship that may exist between histone H1 from *Trypanosoma* and some regulatory proteins in Procaryotes is not known.

Tetrahymena thermophyla histone H1, with two domains, should reflect another step in this history. This outline acquires major relevance due to the fact that the gene that codes for the histone H1 from *Tetrahymena* appears interrupted by one intron located exactly between the two domains of this protein. According to the authors the first exon do not correspond to the central hydrophobic domain (Wu et al. 1986).

Finally, the exon corresponding to the amino-terminal domain should have been recluted to complete the three domain histone H1, somewhere on times.

According to Gilbert, the introns are a primitive feature and eventually they are eliminated as a result of the increasing genomic efficiency (for example; in Eubacteria). In the case of the histones, these proteins are required in large quantities during the synthesis of DNA, which eventually could cause the elimination of the introns of their genes. Interestingly, in some organisms which present histone genes with introns, those genes correspond to histone variants quantitatively less important, whose synthesis is not coupled to DNA replication (Wells & Kedes 1985).

Moreover, in spite of the fact that most of the histone genes do not have introns, a detailed study shows that a significant number of histone genes with introns have been found, and in some cases the positions of the introns are in the boundary of the corresponding structural domain (Engel et al. 1982, Wu et al. 1986). Like histone H1, the histones from the nucleosomal core presents three structurally (Luger et al. 1997) and functionally defined (Wan et al. 1995) domains. Some histone genes possessing introns are in Table 1.

On the other hand, if these proteins were formed by exon shuffling, eventually different products could be found in the domain composition of some of them. This has been described in *Entoamoeba histolytica* and *Trypanosoma cruzi* histones H3 (Bontempi et al. 1994). These histones have extremely divergent amino terminal domains. The homology of these proteins with

TABLE 1

Histone genes which present introns

Genes histona que presentan intrones

- Histone H1 gene of *Tetrahymena thermophila* (Wu et al. 1986)
- Histone H3 gene of chicken (Engel & Sugarman 1982)
- Histone H3 and H4 genes of *Neurospora crassa* (Woudt et al. 1983)
- Histone H3.3 gene of human (Wells & Kedes 1984)
- Histone H2A gene of *Aspergillus nidulans* (May & Morris 1987)
- Histone H4 genes of *Physarum polycephalum* (Wilhelm & Wilhelm 1987)

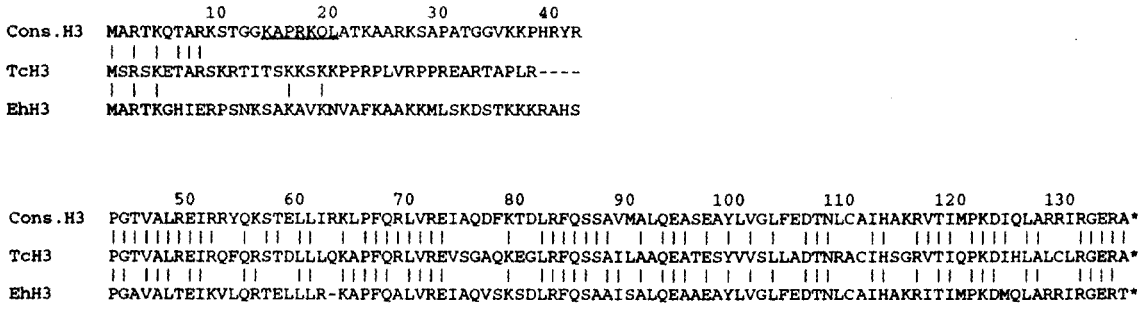

Fig. 3. Protein sequence alignment of the deduced sequence of *T. cruzi* histone H3 with *E. histolytica* and consensus histone H3 sequences (from Bontempi et al. 1994). The globular region is described by Bradbury et al. (1973) betwen aa 42-110; the amino terminal domain in *T. cruzi* and in *E. histolytica* are totally divergent respect to the consensus histone H3 sequence.

Alineación de la secuencia deducida para la histona H3 de *T. cruzi*, de *E. histolytica* y secuencias de consenso para la histona H3 de eucariontes superiores (Bontemi et al. 1994). La región globular entre los aminoácidos 42-110 ha sido descrita por Bradbury et al. (1973); los dominios amino terminales en *T. cruzi* y en *E. histolytica* son totalmente divergentes en relación a la secuencia de la histona de consenso H3.

that of higher eukaryotes is only observed in the globular and C-terminal domains (Fig. 3).

Moreover, Gilbert's theory predicts that, if the proteins were formed by exon shuffling, the same exon could be found in different genes, eventually belonging to other proteins, creating a new transcriptional unit and different products. This is also found in some proteins (Sudhof et al. 1985) and in histones, which are the result of the fusion of sequences of histones and non histone proteins. An example is the macro H2A (from rat liver), described by Pehrson & Fried (1992), which has a segment that resembles a leucine zipper and another segment similar to the COOH-terminal region with 57% identity to sea urchin histone H1 over 30 residues. Another example is CENP-A, an unusual form of histone H3 located in the centromeric heterochromatin, as stated by Palmer et al. (1991).

Finally, I propose the idea that histone H1 from higher eukaryotes was created by the mixture of exons and therefore its evolutionary history is similar to the evolutionary history of other proteins, and not an exception. Moreover, the other histones of the nucleosomal core could also have been created by a similar mechanism, bringing together certain segments of pre-existing ones by exon shuffling. Then, the history of the core histones could be as shown in Fig. 4, in which the globular domain could had been originated by mutation from the histone fold found in archaebacteria, (reviewed by Sandman et al. 1998), and then, exon – shuffling and mutation could have produced the final proteins composed by three structural and functionally defined domains.

AKNOWLEDGMENTS

This work is dedicated to one of my professors in evolutionary biology, Dr Danko Brncic. I thank Dr. Tito Ureta, Dr. Angel Spotorno and my good friend Dr. Carlos Medina, for their contribution to my evolutionary background. I would also like to thank Dr. Norbel Galanti and Dr. Remigio López for their corrections. This work was supported by grants from Proyecto NRPT/SIDA-SAREC and Fundación Galanti.

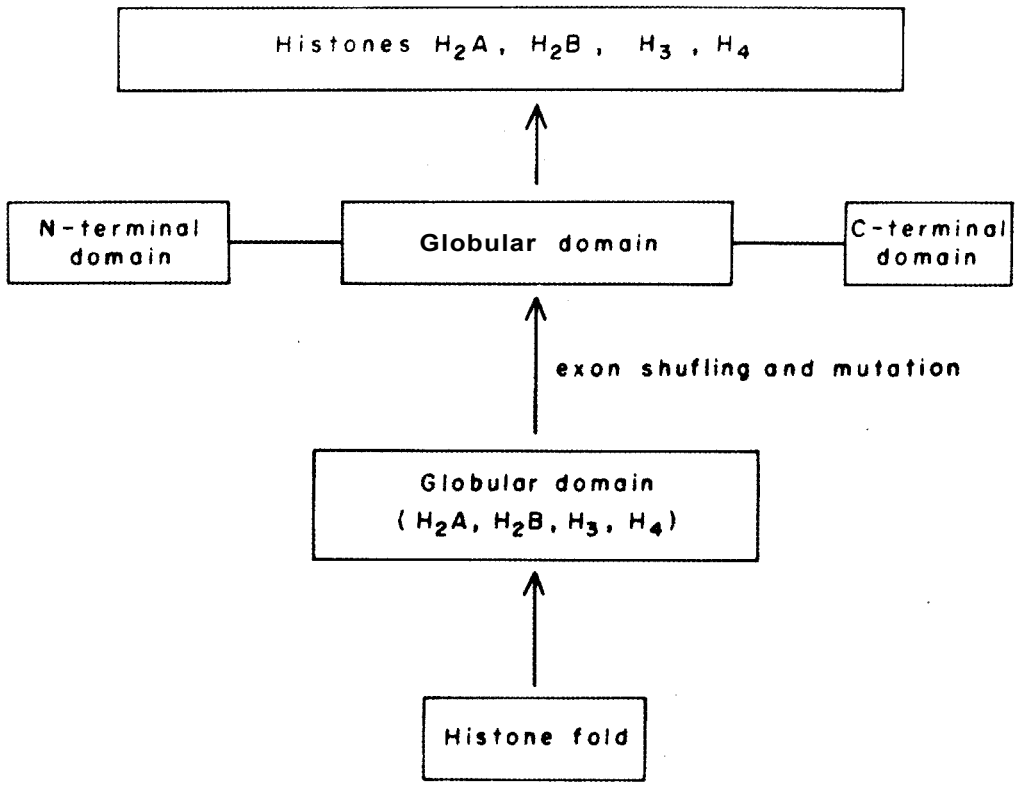

Fig. 4. Proposal for the evolution of histones. Exon shuffling and mutation could produce the final proteins.

Modelo propuesto para la evolución de las histonas. Mutaciones y reubicación de exones podrían producir las proteínas finales.

LITERATURE CITED

ALLAN J, PG HARTMAN, C CRANE-ROBINSON & FX AVILÉS (1980) The structure of histone H1 and its location in chromatin. Nature 288: 675-679.

ARENTS G, RW BURLINGAME, BC WANG, WE LOVE & EN MOUDRIANAKIS (1991) The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix. Proceedings of the National Academy of Sciences USA 88: 10148-10152.

ARENTS G & EN MOUDRIANAKIS (1995) The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. Proceedings of the National Academy of Sciences USA 92: 11170-11174.

ASLUND L, L CARLSSON, J HENRIKSSON, M RYDAKER, GC TORO, N GALANTI & U PETTERSSON (1994) A gene family encoding heterogeneous histone H1 proteins in *Trypanosoma cruzi*. Molecular and Biochemical Parasitology 65: 317-330.

BINDER M, S ORTNER, B PLAIMACUER, M FODINGER, G WIEDERMANN, O SCHEINER & M DUCHENE (1995) Sequence and organization of an unusual histone H4 gene in the human parasite *Entamoeba histolytica*. Molecular and Biochemical Parasitology 71: 243-247.

BLAKE CCF (1978) Do genes in pieces imply proteins in pieces? Nature 273: 267.

BOHM L & TC MITCHELL (1985) Sequence conservation in the N-terminal domain of histone H1. Federation of European Biochemical Societies Letters 193: 1-4.

BONTEMPI E, B PORCEL, J HENRIKSSON, L CARLSSON, M RYDAKER, E SEGURA, A RUÍZ & U PETTERSSON (1994) Genes for histone H3 in *Trypanosoma cruzi*. Molecular and Biochemical Parasitology 66: 147-151.

BRADBURY EM, PD CARG, GE CHAMPAN, C CRAME-ROBINSON, SE DANBY, HW RATTLE, M BOUBLICK, J PALAU & FJ AVILÉS (1975) Studies on the role and mode of operation of the very-lysine-rich-histone H1 in eukaryote chromatin. European Journal of Biochemistry 52: 605-613.

BROWN AP (1983) Codon-level Analysis of histone primary sequence: evidence of a repeat tetrapeptide origin and later inclusion of transcribed sequence. Journal of Theoretical Biology 104: 401-416.

BURRI M, W SCHLIMME, B BETSCHART, H LINDNER, U KAMPFER, J SCHALLER & H HACKER (1995) Partial amino acid sequence and functional aspects of histone H1 proteins in *Trypanosoma brucei brucei*. Biological Cell 83: 23-31.

CLOUD P & MF GLAESSNER (1982) The Ediacarian period and system: metazoa inherit the Earth. Science 217: 783-792.

DE SOUZA SJ, M LONG, RJ KLEIN, S ROY, S LIN & W GILBERT (1998) Toward a resolution of the introns early/late debate: only phase zero introns are correlated with the structure of ancient proteins. Proceedings of the National Academy of Sciences USA 95: 5094-5099.

DOOLITTLE WF (1978) Genes in pieces: were they ever together? Nature 272: 581-582.

DOOLITTLE RF & P BORK (1993) Módulos móviles en la evolución de las proteínas. Investigación y Ciencia, diciembre: 22-29.

DORIT RL, L SCHOENBACH & W GILBERT (1990) How big is the universe of exons? Science 250: 1377-1382.

ENGEL JD, B SUGARMAN & JB DODGSON (1982) A chicken histone H3 gene contains intervening sequences. Nature 297: 434-436.

ESPINOZA I, GC TORO, U HELLMAN & N GALANTI (1996) Histone H1 and core histones in *Leishmania* and *Crithidia*: comparison with *Trypanosoma*. Experimental Cell Research 224: 1-7.

FODINGER M, S ORTNER, B PLAIMAUER, G WIEDERMANN, O SCHEIMER & M DUCHENE (1993) Pathogenic *Entamoeba histolytica*: cDNA cloning of a histone H3 with a divergent primary structure. Molecular and Biochemical Parasitology 59: 315-322.

FUSEL NJ, DC ROBBER, J MAUL & TA GLASSIER (1994) Identification of a histone H1-like gene expressed in *Leishmania major*. Molecular and Biochemical Parasitology 62: 321-324.

GILBERT W (1978) Why genes in pieces? Nature 271: 501.

GILBERT W (1985) Genes in pieces revisited. Science 228: 823-824.

GILBERT W, M MARCHIONNI & G MCKNIGHT (1986) On the antiquity of introns. Cell 46: 151-154.

GÒ M (1981) Correlation of DNA exonic regions with protein structural units in haemoglobin. Nature 291: 90-92.

HAN TM & B RUNNEGAR (1992) Megascopic eukaryotic algae from the 2.1-billion-year-old Negaunee iron-formation, Michigan. Science 257: 232-235.

HILL CS, S MARTIN & JO THOMAS (1989) A stable a helical element in the carboxy-terminal domain of free and chromatin-bound histone H1 from sea urchin sperm. The European Molecular Biology Organizations Journal 8: 2591-2599.

HOROWITZ S & MA GOROVSKY (1985) An unusual genetic code in nuclear genes of *Tetrahymena*. Proceedings of the National Academy of Sciences USA 82: 2452-2455.

JUNG A, A SIPPEL, M GREZ & G SCHUTZ (1980) Exons encode functional and structural units of chicken lysozyme. Proceedings of the National Academy of Sciences USA 77: 5759-5763.

KMIECIK D, D SELLOS, D BELAICHE & P SANTIERE (1985) Primary structure of the two variants of a sperm-specific histone H1 from the annelid *Platynereis dumerilü*. European Journal of Biochemistry 150: 359-370.

LUGER K, A W MADER, RK RICHMOND, DF SARGENT & TJ RICHMOND (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature 389: 251-260.

LUGER K & TJ RICHMOND (1998). The histone tails of the nucleosome. Current Opininions in Genetic Development 8: 140-146.

MAEDER DL & L BOHM (1991) The C-domain in the H1 histone is structurally conserved. Biochemica et Biophysica Acta 1076: 233-238.

MARCHIONNI M & W GILBERT (1986) The triosephosphate isomerase gene from maize: introns antedate the plant-animal divergence. Cell 46: 133-141.

MAY GS & R MORRIS (1987) The unique histone H2A gene of *Aspergillus nidulans* contains three introns. Gene 58: 59-66.

MCLAUGHLIN PJ & MO DAYHOFF (1973) Eukaryote evolution: a view based on cytochrome C sequence data. Journal of Molecular Evolution 2: 99-116.

OBARU K, T TSUZUKI, C SETOYAMA & K SHIMADA (1988) Structural organization of the mouse aspartate aminotransferase isoenzyme genes: introns antedate the divergence of citosolic and mitochondrial isoenzymes genes. Journal of Molecular Biology 200: 13-22.

OHNO S (1981) Original domain for the serum albumin family arose from repeated sequences. Proceedings of the National Academy of Sciences USA 78: 7657-7661.

PALMER DK, K O'DAY, HL TRONG, H CHARBONNEAU & RL MARGOLIS (1991) Purification of the centromere-specific protein CENP-A and demostration that it is a distinctive histone. Proceedings of the National Academy of Sciences USA 88: 3734-3738.

PEHRSON JR & VA FRIED (1992) Macro H2A, a core histone containing a large monhistone region. Science 257: 1398-1400.

PEREIRA SL & JN REEVE (1998) Histones and nucleosomes in Archaea and Eukarya: a comparative analysis. Extremophiles 2: 141-148.

REECK G, E SWANSON & DC TELLER (1978) The evolution of histones. Journal of Molecular Evolution 10: 309-317.

ROSSMAN MG, MJ ADAMS, M BUEHNER, GC FORD, ML HACKERT, PJ LENTZ JR, A MCPHERSON JR, RW SCHEVITZ & IE SMILEY (1971) Structural constraints of possible mechanisms of lactate dehydrogenase as shown by high resolution studies of the apoenzyme and a variety of enzyme complexes. Cold Spring Harbor Symposium in Quantitative Biology 36: 179-191.

ROSSMANN MG & A LILJAS (1974) Recognition of structural domains in globular proteins. Journal of Molecular Biology 85: 177-181.

SAKANO H, JH ROGERS, K HUPPI, C BRACK, A TRAUNECKEER, R MAKI, R WALL & S TONEGAWA (1979) Domains and the hinge region of an immunoglobulin heavy chain are encoded in separate DNA segments. Nature 277: 627-633.

SANDMAN K, JA KRZYCKI, B DOBRINSKI, R LURZ & JN REEVE (1990) HMf, a DNA binding protein isolated from the hyperthermophilic archaeon *Methanothermus fervidus* is most closely related to histones. Proceedings of the National Academy of Sciences USA 87: 5788-5791.

SANDMAN K, SL PEREIRA & JN REEVE (1998) Diversity of prokaryotic chromosomal proteins and the origin of nucleosome. Cellular and Molecular Life Sciences 54: 1350-1364.

STARICH MR, K SANDMAN, JN REEVE & MF SUMMERS (1996) NMR structure of HMfB from the hyperthermophile, *Methanothermus fervidus*, confirms that this archaeal protein is a histone. Journal of Molecular Biology 255: 187-203.

STONE EM, KN ROTHBLUM & RJ SCHWARTZ (1985) Intron-dependent evolution of chicken glyceraldehyde phosphate dehydrogenase gene. Nature 313: 498-500.

SUDHOF TC, JL GOLDSTEIN, MS BROWN & DW RUSSELL DW (1985) The LDL receptor gene: a mosaic of exons shared with different proteins. Science 228: 815-822.

THATCHER TH & M GOROVSKY (1994) Phylogenetic analysis of the core histones H2A, H2B, H3 and H4. Nucleic Acids Research 22: 174-179.

THATCHER TH, J MACGAFFEY, J BOWEN, S HOROWITZ, DL SHAPIRO & M GOROVSKY (1994) Independent evolutionary origin of histone H3.3 like variants of animals and *Tetrahymena*. Nucleic Acids Research 22: 180-186.

TORO GC, C WERNSTEDT, C MEDINA, N JARAMILLO, U HELLMAN & N GALANTI (1992) Extremely divergent histone H4 sequence from *Trypanosoma cruzi*: evolutionary implications. Journal of Cellular Biochemistry 49: 266-271.

TORO GC, N GALANTI, U HELLMAN & C WERNSTEDT (1993) Unambiguous identification of histone H1 in *Trypanosoma cruzi*. Journal of Cellular Biochemistry 52: 431-439.

VANFLETEREN J, SM VAN BUN & JJ VAN BEEUMEN (1988) The primary structure of the major isoform (H1.1) of histone H1 from the nematode *Caenorhabditis elegans*. Journal of Biochemistry 255: 647-652.

VAN HOLDE KE (1989) The proteins of chromatin. I. Histones. Springer-Verlag, New York, New York. 500 pp.

VELLAI T, K TAKACS & G VIDA (1998) A new aspect to the origin and evolution of eukaryotes. Journal of Molecular Evolution 46: 499-507.

VOGEL G (1997) Searching for living relics of the cell's early days. Science 277: 1604.

WAN JS, RK MANN, M GRUNSTEIN (1995) Yeast histone H3 and H4 N termini function through different GAL1 regulatory elements to repress and activate transcription. Proceedings of the National Academy of Sciences USA 92: 5664-5668.

WELLS D & L KEDES (1985) Structure of a human histone cDNA: evidence that basally expressed histone genes have intervening sequences and encode polyadenylylated mRNAs. Proceedings of the National Academy of Sciences USA 82: 2834-2838.

WILHELM M & X WILHELM (1987). Both histone H4 genes of *Physarum polycephalum* are interrupted by an intervening sequence. Nucleic Acids Research 15: 5478-5479.

WOESE CR, O KANDLER & ML WHEELIS (1990) Towards a natural system of organism: proposal for the domains Archaea, Bacteria and Eucarya. Proceedings of the National Academy of Sciences USA 87: 4576-4579.

WOUDT L, A PASTINK, A KEMPERS-VEENSTRA, A JANSEN, W MAGER & R PLANTA (1983) The genes coding for histone H3 and H4 in *Neurospora crassa* are unique and contain intervening sequences. Nucleic Acids Research 11: 5347-5360.

WU M, CD ALLIS, R RICHMAN, RG COOK & MA GOROVSKY (1986) An intervening sequence in an unusual histone H1 gene of *Tetrahymena thermophila*. Proceedings of the National Academy of Sciences USA 83: 8674-8678.

WUILMART C & L WYNS (1977) An evolutionary scheme for the histones as derived from a study of internal repetitions and homologies among the different classes. Journal of Theoretical Biology 65: 231-252.

Invited Editors: R. Godoy-Herrera and G. Gajardo
Received April 15, 2000; accepted September 18, 2000