Volume 15 Number 24 1987 Nucleic Acids Research

Evolution of late H2A, H2B and H4 histone genes of the sea urchin, *Strongylocentrotus purpuratus*

Robert Maxson*, Timothy Mohun[+1], Glen Gormezano[1] and Larry Kedes[1]

Department of Biochemistry and Kenneth R.Norris Cancer Hospital and Research Institute, University of Southern California Medical School, 2025 Zonal Ave., Los Angeles, CA 90033 and [1]Medigen Project, Department of Medicine, Stanford University School of Medicine and Veterans Administration Hospital, 3801 Miranda Ave., Palo Alto, CA 94304, USA

Received April 13, 1987; Revised and Accepted August 18, 1987

ABSTRACT

Sea urchins possess several distinct sets of histone genes, including "early" genes, maximally active in cleavage and blastula stages, and "late" genes, active from the late blastula stage onwards. We determined the nucleotide sequences of six sea urchin (Strongylocentrotus purpuratus) late histone genes located on four genomic segments. Comparative analysis of these sequences identified several conserved elements in 5' flanking regions, including the sequences ATGPyA$^{T}_{A}$N$^{T}_{A}$$^{T}_{A}$ shared by all late genes and GGCGGGAAATTGAAAA shared by two late H4s. Comparisons of protein-coding sequences of late H4 and H2B genes with their early counterparts showed that silent sites have diverged to the theoretical maximum, indicating that early and late histone gene classes diverged at least 200 million years ago. Since extant echinoderms evolved from a common ancestor at about that time, it is likely that early and late histone gene sets are characteristic of all echinoderm groups. Amino acid sequences derived from nucleotide sequences of late H2A and H2B gistone genes differ substantially from amino acid sequences of their late counterparts. Most such differences are in highly mutable positions. A few, however, occur in positions that do not mutate frequently and thus may reflect functional differences between the early and late forms of the H2A and H2B proteins.

INTRODUCTION

At least four different classes of histone genes are expressed over the course of sea urchin development, each during a different interval (1). "Cleavage stage" genes, not yet characterized structurally, are active in oogenesis and early cleavage (2). "Early" genes are transcribed maximally from late oogenesis through the blastula stage and account for the bulk of the histone mRNA synthesized during early development (3). They are arranged in tandem quintets comprising one copy each of H1, H2A, H2B, H3 and H4 histones, and are repeated approximately 400 times per haploid genome. "Late" genes are transcribed at low levels in cleavage and blastula stage embryos and at increased rates in subsequent development (4,5,6). In contrast to the early histone genes, members of the late gene family are arranged in irregular clusters or as isolated single genes, and are repeated only 6-12 times per haploid genome. While the members of the early histone gene family are nearly identical in both gene-flanking and protein-coding nucleotide sequences, those of the late gene family

differ somewhat in protein coding sequence and have diverged substantially in their gene flanking sequences (6,7,8). Finally, sperm histone genes, encoding sperm-specific H2A and H2B proteins, are active in the testis of the adult and are present in the genome in single-copies (9,10).

We wish to understand both how these different classes of histone genes are regulated and the significance of multiple classes of histone genes for development and differentiation. As a step toward these goals we isolated six genes encoding late histone proteins from the sea urchin Strongylocentrotus purpuratus. We showed previously that these genes are located on four distinct cloned genomic segments, L1, L2, L3, L4 (4). Clone L1 contains a late H4 and a late H2B gene, L2 a single late H4 gene, L3 closely linked H2A and H2B genes, and L4 a solitary late H2B gene (4). With the possible exception of the L4 H2B gene (which is still untested), all of these genes are expressed (11).

Here we report the nucleotide sequences of these six late histone genes and flanking regions. A comparative analysis revealed several conserved sequence elements in 5' gene flanking regions. Although these elements are also recognizable in early histone gene flanking sequences, they have diverged considerably, and may therefore be involved in the differential expression of the two gene classes. In addition, the amino acid sequences of late H2A and H2B proteins differ from the corresponding early forms, supporting the view that early and late H2B proteins have different functions in the cell.

METHODS

Plasmids Used for DNA Sequence Analysis and Preparation of Plasmid DNA

Plasmids pSpl-1, pSpl-2, pSpl-3 and pSpl-4 have been described (4), and cloned segments are shown in schematic in Figure 1. Some late gene segments were subcloned in M13 MP 8 or MP 9 vectors (2) prior to sequencing by the dideoxy method.

Plasmids were prepared by the standard cleared lysis procedure (13). Single stranded, M13 replicative form DNA was prepared as described (12).

DNA Sequencing

Restriction enzyme cleavage sites were mapped on cloned DNA segments by digestion with combinations of enzymes (13) and by partial cleavage of terminal-labeled DNA (14). Desired fragments were separated on acrylamide gels, electroeluted, and purified by passage over DEAE sephacel (13). Purified fragments were labeled at their 5' termini with ^{32}P using T4 polynucleotide kinase and ^{32}P-ATP. After strand separation or secondary cleavage to remove one end label, such fragments were sub-

jected to the Maxam-Gilbert base-specific chemical cleavage reactions (15), resolved on urea-acrylamide gels and visualized by autoradiography.

Portions of the L1 H4 and L1 H2B genes were sequenced enzymatically (16). M13 clones bearing fragments of interest were annealed with the universal M13 primer and treated with Klenow DNA polymerase in the presence of dideoxy nucleotides and ^{32}P-dATP. The products of this reaction were electrophoresed on buffer-gradient, urea-acrylamide gels and visualized by autoradiography (17). Histone mRNA-coding regions were sequenced at least twice, and, when possible, on both DNA strands.

Data Analysis

A set of computer programs available through the BIONET resource was used to assemble overlapping sequences, to identify histone coding regions, and to compare histone DNA sequences. Sequences used in the comparative analysis were obtained from the latest update of the GENEBANK sequence database.

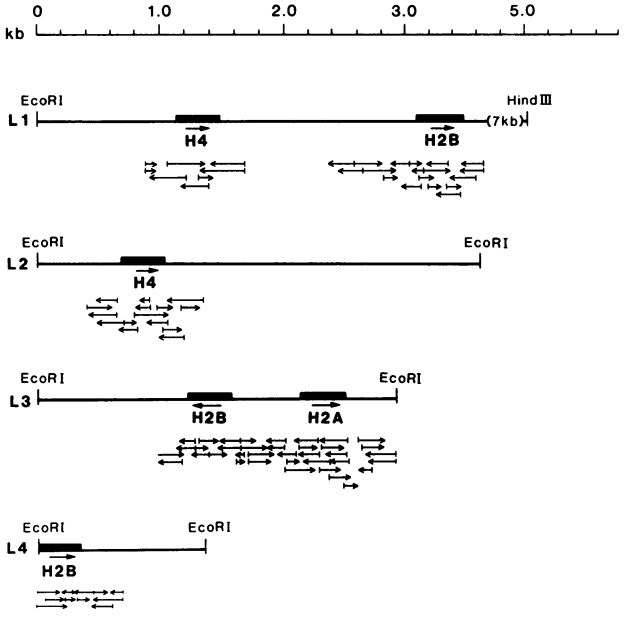
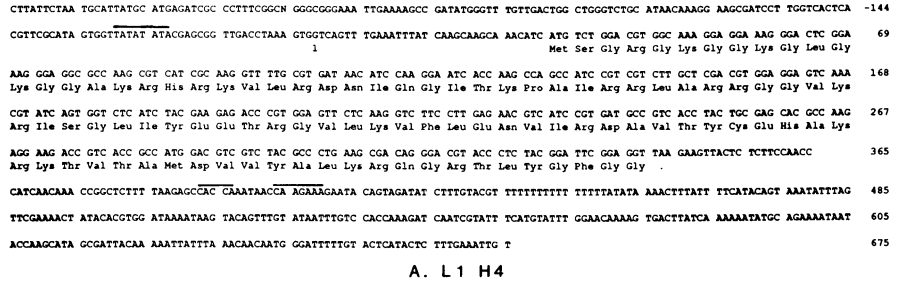


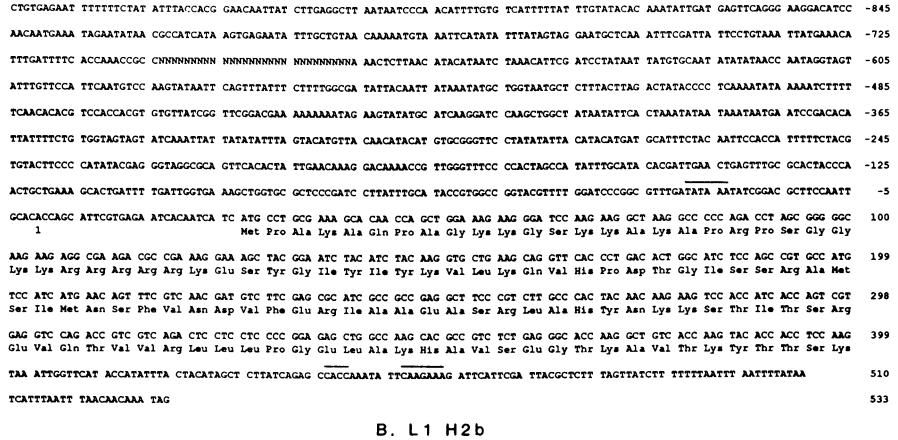
Figure 1. Schematic map of late histone genes showing sequencing strategy. Genomic segments L1-L4 were excised from lambda phage clones and subcloned into pBR322 (4). Nucleotide sequences were determined using a combination of chemical and enzymatic methods as described in Methods. Arrows indicate location and direction of sequences read from single gels. All histone coding regions were sequenced at least twice.

RESULTS

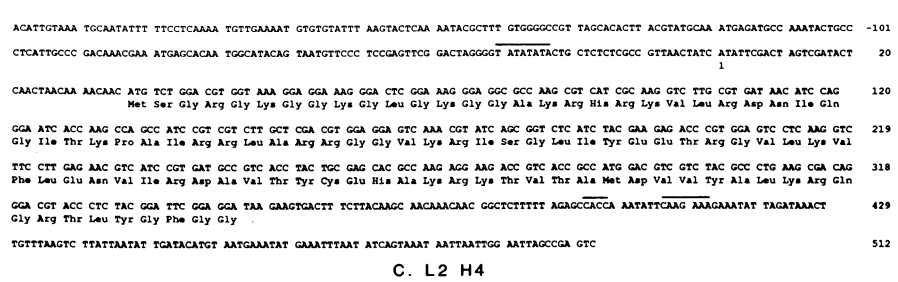
We determined the DNA sequences of six late histone genes following the strategy shown in Figure 1. A set of computer programs (Intelligeneties/BIONET) was used to assemble overlapping sequences and locate histone coding regions. The nucleotide sequences and derived amino acid sequences of the six genes are shown in Figure 2. The 5' portion of the L4 H2B gene was not sequenced because this region was not isolated in the original phage clone (4). DNA sequences of the remaining five genes



A. L1 H4



B. L1 H2b



C. L2 H4

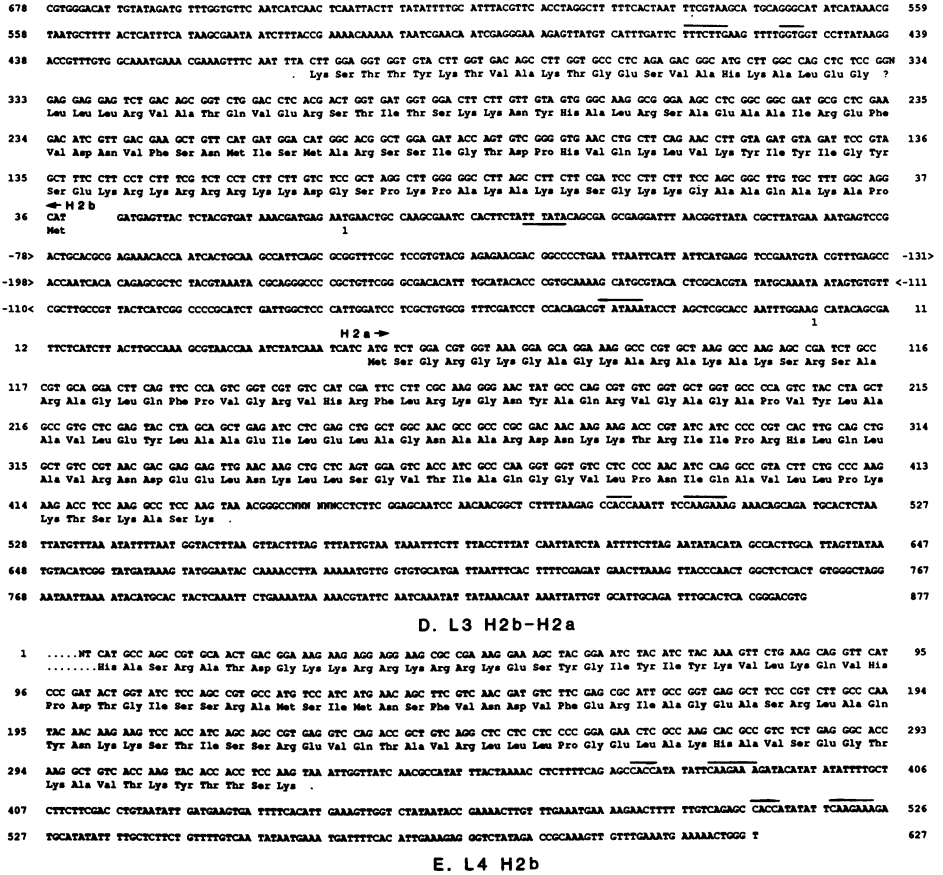


Figure 2. Nucleotide sequences of six S. purpuratus late histone genes. The sequences were derived according to the strategy shown in Figure 1. Only one strand of DNA sequence is shown. Numbering of bases for genes on L1, L2, L3 segments starts at the base representing the 5' end of the mRNA (11). Since L3 H2A and H2B genes are transcribed from opposite strands, we show the transcribed strand for the H2B and the non-transcribed strand for the H2A. Numbering of bases progresses right to left for the H2B sequence and left to right for the H2A sequence (arrows). Numbering of bases of the L4 H2B gene begins in the protein-coding region at the 5' border of the cloned segment. Bases upstream from the transcription initiation sites are indicated by negative numbers. TATA, CACC (mRNA termination site) and CAGA elements are underlined. A: L1 H4 gene; B: L1 H2B gene; C: L2 H4 gene; D L3 H2A and H2B genes; E: L4 H2B.

include at least 200 base pairs each of 5' and 3' flanking sequences, as well as the entire mRNA coding regions.

5' Gene-Flanking Sequences

To identify cis-acting elements that may be involved in the regulation of early and late histone gene expression, we searched gene flanking regions for sequence

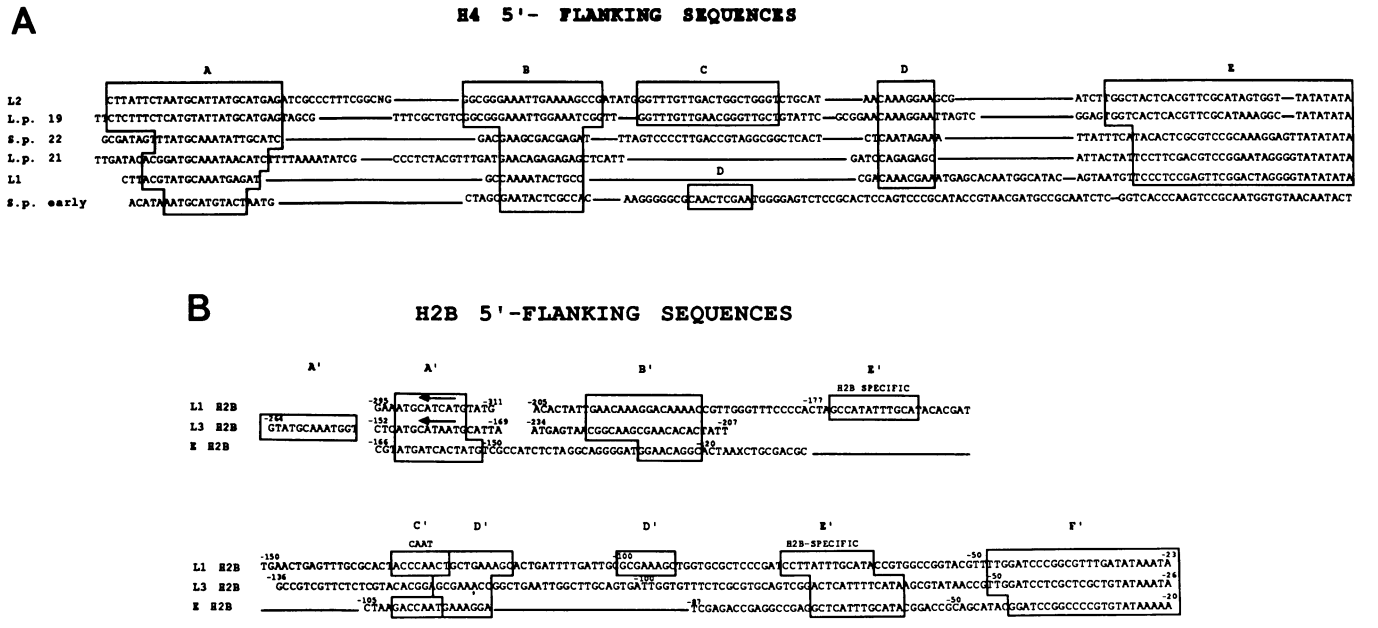


Figure 3. Comparisons of 5' flanking sequences of late H4 and late H2B genes. We searched the 5' flanking segments of several late and early H4 and H2B histone genes for homologous regions using the Needelman-Wunsch algorithm and the BIONET software package. Boxes indicate regions of significant matching. Numbering of bases is relative to the transcription start sites, which are taken as O. Bases upstream from the transcription initiation sites are indicated by negative numbers. Sequences were obtained from the NIH database.

motifs shared among late but not early genes. Comparisons of the region between the transcription start site and position -200 of five late H4 sequences from two sea urchin species revealed five conserved blocks, labeled A, B, C, D and E in Figure 3. Element E comprises the TATA box and approximately 20 base pairs upstream. This sequence is moderately well-conserved among the five late H4 genes, but has diverged significantly in the early H4 gene in comparison to the late H4 gene. Element D is present in all H4 genes examined. Having the consensus CAAAGGAA, this sequence is located 60-80 bp distal to the TATA element. Interestingly, in the early H4 gene, this sequence is located approximately 20 bp farther upstream from the TATA element than is its late counterpart. Element C matched in 15/20 positions when the in L2 H4 gene was compared with one of the two sequenced late H4 genes from L. pictus (L.p. 19 (7)) but was not detected in the other H4 genes that we examined. The B sequence block is also closely conserved in the L2 and L.p. 19 H4s (17/20 positions matching). We found related sequences in the flanking regions of other genes, but they had diverged considerably from those of the L2 H4 gene, sharing only a central GAA and overall purine richness. The A elements of L2 and L.p. 19 H4 genes match in 21/25 positions. Although this region of other H4 genes does not match the L2 sequence closely, 5 of 6 H4 genes share the core sequence ATGA and 3 of 6 share ATGCAAAT. Homologues of the A element are also evident in both L1 and L3 late H2B genes (Figure 3 A' element). Thus, the sequence GTATGCAAATG, located in the region between the divergently transcribed L3 H2A and H2B genes, matches the A element of the L2 H4 gene in 11/11 positions. Also present in this region is the sequence ATGCATAATGCA (in inverted orientation relative to the H2B gene), which matches the L1 H4 A region in 10/12 positions.

In general, the 5'-flanking sequences of the L2 H4 gene match those of L.p. 19 more closely than any other gene tested. Figure 3 shows that sequence blocks A-E are virtually identical in these two genes. It is at first puzzling that H4 genes from different species are more similar than those from the same species. Most probably, this simply indicates that the duplication event that gave rise to the two S. purpuratus H4 genes took place long before the branching of the L2 and Lp 19 genes and before the branching of the Strongylocentrotus and Lytechinus lineages.

We found several conserved sequence blocks, labeled A'-F' in Figure 3, in upstream sequences flanking the L1, L3 and early H2B genes. Element "F'" includes the TATA box and 12 bp upstream. The "E'" motif, documented previously in flanking sequences of H2B genes of phylogenetically diverse organisms (18) and recently in the immunoglobulin heavy chain gene enhancer (19), kappa chain promotors (20-22) and several snRNA genes (23-25) is present in flanking regions of each of the H2B

genes that we examined. Interestingly, there are two copies of this element located 82 bp apart in sequences flanking the L1 H2B gene (Figure 3). Element "D", having the consensus GCGAAA was present in L1 and L3 flanking regions immediately distal to the CAAT box regions. A second copy was located 20 bp farther downstream in the L1 flanking region. A similar sequence (4/6 positions matching) was present in the analogous position in the early H2B flanking region. A DNA element with this sequence has been implicated in the periodic transcription of histone genes during the yeast cell cycle (26,27). The B element, located approximately 200 bp upstream from the transcription start site, is a loosely conserved, purine-rich region approximately 18 bp long. Similar sequences were found in H4 genes (element B). Finally, element A, 12 bp in length, was located at different positions in L1, L3 and early H2B genes between -150 and -300. The sequence matches closely that of the "A" element in the H4 gene flanking regions.

In summary, although most of sequence elements identified in our comparative analysis are present in both early and late histone genes, the early and late versions of these sequences have diverged substantially. Thus, these elements may be involved in the differential regulation of the two gene sets.

5' and 3' Untranslated Regions

Both 5' and 3' untranslated regions (UTRs) of sea urchin early histone genes have been implicated in regulation: a positive-acting transcription factor requires sequences in the 5' UTR for its activity (28); the 3' UTR and flanking region are required for correct processing of histone mRNA 3' termini, and may be involved in periodic fluctuations of histone mRNA levels during the cell cycle (29). We therefore examined the 5' and 3' UTRs of the late histone genes for conserved structures. We previously mapped the transcription start sites of the L1 H4, L1 H2B, L2 H4, L3 H2A and L3 H2B genes (11). Using these sites to demarcate the 5' border of the leader sequences, we performed pairwise comparisons of the leaders of the sequenced late genes and the early H2A, H2B and H4 genes. These compairsons showed that the late gene leader sequences have diverged substantially from each other and even more from their early counterparts. We could not detect any sequence elements common to late histone gene leaders. Neither could we identify common secondary structure (data not shown). It appears, therefore, that apart from possessing leader sequences of approximately the same length (30), late histone mRNAs do not shart structural elements in these regions.

The 3' termini of most histone mRNAs are flanked proximally by a palindromic sequence (in the UTR) and distally by the elements ACCA and CAAGA (29). The late histone genes that we have examined share these sequence motifs (Figure 2).

Table I
Sequence Relationships of Sea Urchin Histone Genes

Compared Histone Coding Sequences	Divergence at all sites (%)	Divergence at replacement sites (%)	Divergence at replacement sites corrected for multiple hits (%)	Divergence at silent sites (%)	Divergence at silent sites corrected for multiple hits (%)	Calculated divergence time (millions of years before present)
L1H2B-L3H2B	6	1	2	13	28	56
L1H2B-L4H2B	10	5	6	19	57	114
L3H2B-L4H2B	12	7	9	18	51	102
L1H2B-early H2B	22	7	12	42	100	200
L1H4-L2H4	2	-	-	6	10	20
L1H4-L.p. 19 H4	5.5	-	-	17	41	82
L2H4-L.p. 19 H4	5.5	-	-	15	30	60
L1H4-S.p. 22 H4	7	-	-	19	33	66
L1H4-S.p. 22 H4	7	-	-	19	32	64
L2H4-L.p. 21 H4	6	-	-	18	36	72
L2H4-L.p. 21 H4	6	-	-	16	28	56
L1H4-early H4	16	-	-	47	100	200
L3H2A-early H2A	18	6	6	52	100	200

Evolutionary relationships among histone sequences were analyzed as described in reference (31). Replacement sites are those nucleotides for which a change results in an amino acid replacement; silent sites are those nucleotides for which a change does not results in such a replacement. Differences between genes in silent and replacement sites were totaled and corrected for multiple hits. Divergence times were calculated from the corrected divergence at silent sites by assuming that silent site substitutions occur at a rate of 0.5% per million years (33). L1 H2B, L1 H4, L2 H4, L2 H2A, L3 H2B and L3 H2A are S. purpuratus late histone genes whose sequences are presented in this report. L.p. 19 and L.p. 21 are Lytechinus pictus late H4 genes (7,8). S.p. 22 is an S. purpuratus late H4 gene (6).

Our comparisons did not reveal any additional common homology blocks or secondary structures. However, we did detect a second termination site in the 3' flanking sequence of the L4 H2B gene (Figure 2). Located 134 bp downstream from the "authentic" site, this site closely resembles the authentic one in having the palindromic sequence as well as ACCA and CAAGA elements. We do not know whether this site is used in vivo.

Histone Coding Sequences and the Evolution of Late Genes

Roberts et al (8) estimated that in L. pictus early and late H3 and H4 genes have been evolving separately for at least 240 myr. We asked whether S. purpuratus early and late H2A, H2B and H4 genes are similarly divergent. To answer this question we compared sequences of protein-coding regions of early and late histone genes. Silent site substitutions in sea urchin histone genes occur at an apparently constant rate of 0.5% per million years (33). Thus the divergence of silent sites provides a measure of the time two gene sequences have been evolving separately. The sequence relationships of various early and late histone genes are depicted in Table I. After correction of raw percent divergence for multiple substitutions (31), it is evident that the silent sites of early and late H2A, H2B and H4 genes have diverged 100%.

Comparisons of H4 coding sequences revealed an apparent anomaly in the rates of 5' flanking and coding sequence evolution. Differences in silent sites suggest that the L1 and L2 H4 genes are more closely related to each other than to the L.p. 19 H4 gene. However, when 5' flanking sequences were compared, the L2 and L.p. 19 H4s were most closely related (see above), while flanking regions of L1 and L2 H4s had diverged almost completely. A likely explanation for this apparent paradox is that a gene conversion event between the L1 H4 and the L2 H4 homogenized the H4 coding regions but not the 5' flanking sequences (see Discussion).

Amino Acid Sequences of Late Histone Proteins

There are at least two explanations for why sea urchins possess different, developmentally regulated isotypes of H1, H2A, and H2B histones: (1) Such isotypes may have different functions that have been maintained by natural selection. (2) Such isotypes may be functionally equivalent, and the two gene sets may have arisen as a consequence of selection on another character, such as the mechanism that enables the embryo to produce a large amount of histone protein during early development but reduced amounts in later embryogenesis. According to the first view, early and late forms of H1, H2A and H2B proteins should differ significantly in structure; according to the second, any structural differences should be functionally neutral. As a first step toward distinguishing these possibilities, we compared early and late H2A and H2B histone sequences from several sea urchin species.

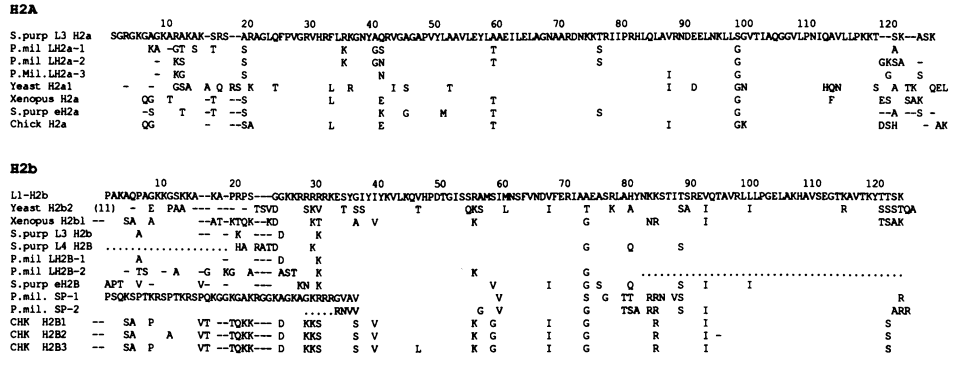


Figure 4. Comparative analysis of H2A and H2B amino acid sequences. Amino acid sequences derived from nucleotide sequences were compared using the Needelman-Wunsch algorithm in the PEP program (BIONET). Differences between the uppermost sequence and those below are indicated by the letter code of the differing amino acid. Dashes symbolize deletions. The N-terminal 11 residues of the yeast H2B2 sequence are not shown. Numbering of amino acids is relative to the amino terminus of the uppermost sequence, which is taken as 1. Sequences were obtained from the NIH database.

Analysis of H2A and H2B amino acid sequences from a variety of organisms has shown that these proteins consist of amino and carboxy terminal regions that evolve at a moderate pace, and a central region that is highly conserved (32). Comparisons of amino acid sequences of the S. purpuratus late H2A and H2B genes with those of other histone proteins are consistent with this general finding (Figure 4).

The amino acid sequences of the L3 H2A, L1 H2B and L3 H2B proteins differ from those of their early counterparts by 10%, 14% and 14% respectively (Figure 4). In the case of H2A, 5 of these differences are in the N-terminal 19 amino acids, 5 in the central 90 amino acids, and 3 in the C-terminal 14 amino acids. All changes but one, a met-leu at position 51, occur in positions known to be mutable from comparisons of histone sequences from a variety of species (Figure 4).

A comparison of the early H2B with the L1 and L3 late H2Bs reveals a similar distribution of amino acid changes. In the case of the L1 there are 10 changes in the N-terminal 29 amino acids, and 9 in the remainder of the protein. Of these nine, eight occur in highly mutable positions, but one ser—ala (position 74) is unique to the S. purpuratus early H2B-L1 late H2B pair, suggesting that it may be functionally significant. Arguing against this possibility is the fact that early and late H2Bs of P. miliaris do not differ at this position (both contain ala (9)).

DISCUSSION

Gene Flanking Regions

Comparisons of 5 late H4 and 2 late H2B 5' gene flanking sequences from two different sea urchin species have identified several conserved elements in addition to the well-documented TATA, CAAT and "H2B specific" (or octamer) elements (18). These include the consensus sequence ATGCATA, noted previously in surveys of H3 and H4 genes (6,8) and here identified in sequences flanking H4 and H2B genes and between a divergently transcribed H2A-H2B gene pair. Interestingly, the two genes showing the greatest similarity in 5' flanking sequences were the L2 H4 gene (from S. purpuratus) and the L.p. 19 H4 gene (from L. pictus). The conservation of specific sequence elements in the 5' flanking regions of these genes over the 65 million years since the L. pictus and S. purpuratus lineages diverged (34) suggests that such elements serve important functions, likely in the expression or regulation of the genes.

Homologues of most of the sequence elements found to be conserved among late histone genes are also present in the flanking regions of early genes. However, with the exception of the H2B-specific motif, they have diverged significantly from their late counterparts. Thus, the early and late versions of these elements may be recognized by different DNA binding factors, or may interact differently with the same set of factors to cause the differential expression of the early and late histone genes. Furthermore, the relative positions of some conserved sequence elements vary between early and late genes, and such differences could also play a role in determining the different temporal patterns of expression of early and late genes.

The sequence, spacing and even number of putative regulatory elements vary between individual late histone genes, although less dramatically so than between late and early genes. For example, the A sequences of the L1 and L3 late H2B genes differ in 5 of 10 positions and the B sequences in 8 of 15 positions, and these elements are located in different relative positions in the two genes. Such differences between late genes could account for their slightly different temporal profiles of expression during development (11).

Histone Gene Evolution

Our sequence comparisons enable us to estimate when early and late forms of H2A, H2B and H4 genes diverged. The rate of change of silent sites in sea urchin early histone genes is a nearly constant 0.5% per million years (33). Applying this rate to the sequence differences among early and late histone genes, and correcting for multiple substitutions (31), we found that silent sites had diverged to 100% of the theoretical maximum in all cases. Thus, early and late forms of S. purpuratus H2A, H2B and H4 histone genes diverged a minimum of 200 million years ago. This

estimate is similar to that of Roberts and coworkers (8) for the divergence time of early and late H4 genes of Lytechnius pictus. Since all exant Echinoderms are thought to have evolved from a common ancestor in the Triassic, approximately 200 million years ago (34), we predict that early and late histone gene families exist in all Echinoderm groups.

Comparisons of late H4 gene sequences revealed an anomaly in the rate of sequence evolution: the protein-coding sequences of L2 H4 gene are most closely related to those of the L1 H4 gene, while the 5' gene-flanking sequence of the L2 H4 is most closely related to the Lp 19 gene of L. pictus. There are at least two ways of explaining this apparent paradox: either the rate of silent site substitution of one or more genes may deviate from the value of 0.5% per million years, or there may have been a gene conversion event between the L1 and L2 H4 genes that homogenized the coding sequences but not the 5' flanking regions. Such a conversion event would make the two genes seem more closely related than they actually are. Since gene conversions involving only coding sequences have been documented in the sea urchin late histone gene family (8), and since the apparent rate of evolution of coding sequences is not known to vary significantly from 0.5% per million years, a gene conversion event seems the more likely explanation for this anomaly.

Late H2A and H2B Histone Protein Sequences

Our data show that early and late H2A and H2B proteins differ substantially in amino acid sequence, but the majority of these differences occur in positions known to be highly mutable on the basis of comparisons of a wide variety of histone sequences, and thus may not be functionally significant. We did observe two amino acid substitutions, one in the early H2A and one in the early H2B, that were not in highly mutable positions. The early H2A sequence contains a methionine at position 74, the late H2A a leucine. The early H2B has a serine residue at position 74 and its late H2B counterpart on alanine. Early and late H2A sequences of P. miliaris also show the met-ser substitution at position 51, suggesting that it may have functional significance. However, the ser-ala substitution observed in the S. purpuratus early and late H2Bs is not present in early and late H2B sequences of P. miliaris; both P. miliaris H2Bs contain an alanine at this position. The ser-ala substitution has therefore appeared in the S. purpuratus early H2B sequence in the recent evolutionary past and is not likely to be functionally important.

ACKNOWLEDGEMENTS

We thank Susan Halsell and Peter Evans for expert technical assistance and Drs. Amy Lee and R.E.K. Fournier for critically reviewing this manuscript. This work was supported by NIH grants to RM and LK.

*To whom correspondence should be addressed

+Present address: CRC Molecular Embryology Group, Department of Zoology, Downing Street, Cambridge CB2 3EJ, UK

REFERENCES

1. Maxson, R., Cohn, R., and Kedes, L. (1983) Annu. Rev. Genet. **17**, 239-277.
2. Newrock, K., Cohen, L., hendricks, M., Donelly, R. and Weinberg, E.S. (1978) Cell **14**, 247-257.
3. Kedes, L.H. (1979) Annu. Rev. Biochem. **48**, 847-870.
4. Maxson, R., Mohun, T., Gormezano, G., Childs, G. and Kedes, L. (1983) Nature **301**, 120-125.
5. Knowles, J., and Childs, G. (1984) Proc. Natl. Acad. Sci. USA **81**, 2411-2415.
6. Kaumeyer, J.F., and Weinberg, E.S. (1986) Nuc. Acids Res. **14**, 4557-4576.
7. Childs, G., Nocente-McGrath, C., Lieber, T., Holt, C. and Knowles, J. (1982) Cell **31**, 383-393.
8. Roberts, S.B., Weisser, K.E., and Childs, G., (1984) J. Mol. Biol. **174**, 647-662.
9. Busslinger, M., and Barberis, A., (1985) Proc. Natl. Acad. Sci. USA **82**, 5676-5680.
10. Lieber, T., Weisser, K., and Childs, G., (1986) Molec. Cell. Biol. **6**, 2602-2612.
11. Mohun, T.J., Maxson, R., Gormezano, G., and Kedes, L. (1985) Dev. Biol. **108**, 491-502.
12. Messing, J. (1983) in "Meth. Enzymol" **101**, Wu, R., Crossman, L. and Moldave, K. eds. pp. 20-78.
13. Maniatis, T., Fritsch, E. and Sambrook, J. (1982) "Molecular Cloning: A Laboratory Manual" Cold Spring Harbor Laboratory.
14. Smith, H.O., and Birnstiel, J. (1976) Nucleic Acids Res. **3**, 2387-2397.
15. Maxam, A. and Gilbert, W. (1979) Proc. Natl. Acad. Sci. USA **74**, 560-564.
16. Sanger, F., Nicklen, S., and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA **74**, 5463-5467.
17. Biggen, M.D., Gibson, T.J., and Hong, G.F. (1983) Proc. Natl. Acad. Sci. USA **80**, 3963-3965.
18. Harvey, R.P., Robins, A.J., and Wells, J.R.E. (1982) Nucleic Acids Res. **10**, 7851-7863.
19. Ephrussi, A., Church, G.M., Tonegawa, S. and Gilbert, W. (1985) Science **227**, 134-140.
20. Falkner, F.G., and Zachau, H.G. (1984) NaturE **310**, 71-74.
21. Bergman, Y., Rice, D., Grosschedl, R., and Baltimore, D. (1984) Proc. Natl. Acad. Sci. USA **81**, 7041-7045.
22. Parslow, T., Blair, D.L., Murphy, W.F. and Granner, D.K. (1984) Proc. Natl. Acad. Sci. USA **81**, 2650-2654.
23. Ares, M., Jr., Mangin, M., and Weiner, A.M. (1985) Mol. Cell. Biol. **5**, 1560-1570.
24. Mataj, I., Lienhard S., Jiricny, J., and deRobertis, E.M. (1985) Nature **316**, 163-167.
25. Krol, A., Lund, E., and Dahlberg, J.E. (1985) EMBO J. **4**, 1529-1535.
26. Osley, M.A., Gould, J., Kim, S., Kane, M., and Hereford, L. (1986) Cell **45**, 537-544.
27. Smith, M.M., and Andresson, O.S., (1983) J. Mol. Biol. **169**, 663-690.
28. Mous, J., Stunnenberg, H., Georgiev, O., and Birnstiel, M. (1985) Molec. Cell. Biol. **5**, 2764-2769.
29. Birnstiel, M., Busslinger, M., and Strub, K. (1985) Cell **41**, 349-359.
30. Childs, G., Maxson, R., and Kedes, L.H. (1979) Dev. Biol. **73**, 153-173.
31. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., and Dodgson, J. (1980) Cell **20**, 555-565.
32. Isenberg, I. (1979) Ann. Rev. Biochem. **48**, 159-191.
33. Busslinger, M., Rusconi, S., and Birnstiel, M. (1982) EMBO J. **1**, 27-33.
34. Smith, A.B. (1981) Paleontology **24**, 779-801.
35. Sures, I., Lowry, J. and Kedes, L.H. (1978) Cell **15**, 1033-1044.