MOLECULAR AND CELLULAR BIOLOGY, Mar. 1997, p. 1189–1200 Vol. 17, No. 3
0270-7306/97/$04.00+0

The Five Cleavage-Stage (CS) Histones of the Sea Urchin Are Encoded by a Maternally Expressed Family of Replacement Histone Genes: Functional Equivalence of the CS H1 and Frog H1M (B4) Proteins

BIRGIT MANDL,[1] WOLFGANG F. BRANDT,[2] GIULIO SUPERTI-FURGA,[1]† PAULA G. GRANINGER,[1] MAX L. BIRNSTIEL,[1] AND MEINRAD BUSSLINGER[1]*

Research Institute of Molecular Pathology, A-1030 Vienna, Austria,[1] and Department of Biochemistry, University of Cape Town, Rondebosch 7700, South Africa[2]

Received 13 September 1996/Returned for modification 29 October 1996/Accepted 2 December 1996

The cleavage-stage (CS) histones of the sea urchin are known to be maternally expressed in the egg, have been implicated in chromatin remodeling of the male pronucleus following fertilization, and are the only histone variants present in embryonic chromatin up to the four-cell stage. With the help of partial peptide sequence information, we have isolated and identified CS H1, H2A, H2B, H3, and H4 cDNAs from egg poly(A)$^+$ mRNA of the sea urchin *Psammechinus miliaris*. All five CS proteins correspond to replacement histone variants which are encoded by replication-independent genes containing introns, poly(A) addition signals, and long nontranslated sequences. Transcripts of the CS histone genes could be detected only during oogenesis and in development up to the early blastula stage. The CS proteins, with the exception of H4, are unique histones which are distantly related in sequence to the early, late, and sperm histone subtypes of the sea urchin. In contrast, the CS H1 protein displays highest sequence homology with the H1M (B4) histone of *Xenopus laevis*. Both H1 proteins are replacement histone variants with very similar developmental expression profiles in their respective species, thus indicating that the frog H1M (B4) gene is a vertebrate homolog of the CS H1 gene. These data furthermore suggest that the CS histones are of ancient evolutionary origin and may perform similar conserved functions during oogenesis and early development in different species.

Histones are basic proteins that associate with each other and nuclear DNA to form the nucleosome. This basic unit of chromatin consists of two molecules each of the core histones H2A, H2B, H3, and H4, while the H1 protein interacts both with the histone octamer and with linker DNA and is responsible for packaging nucleosomes into higher-order structures. The core histones H3 and H4 are almost invariant in evolution. The less well conserved H2A and H2B proteins often occur as different subtypes within the same organism, while the H1 protein is the most variable of all histones (reviewed in reference 66).

The expression of the majority of histone genes is tightly coupled to DNA synthesis by specific transcriptional and post-transcriptional mechanisms. These so-called replication histone genes code for mRNAs with short 5′ and 3′ untranslated regions; they lack introns and polyadenylation sequences but instead end in a 3′-terminal palindrome (11, 34) that serves as a recognition sequence for U7 snRNP-mediated 3′ processing of histone pre-mRNA (5). A minor group of histone genes is expressed at a basal level throughout the cell cycle in proliferating cells and at a reduced but significant rate in quiescent cells, where their proteins gradually replace the replication-dependent histones in the chromatin (73, 74). These so-called replacement histone genes contain introns, code for polyadenylated mRNA with long untranslated sequences, and are thus classical RNA polymerase II transcription units (9, 22, 68).

* Corresponding author. Mailing address: Research Institute of Molecular Pathology, Dr. Bohr-Gasse 7, A-1030 Vienna, Austria. Phone: (43/1) 797 30-452. Fax: (43/1) 798 71 53.

† Present address: European Molecular Biology Laboratory, D-69012 Heidelberg, Germany.

The sea urchin genome contains four distinct histone gene families which are sequentially expressed during ontogeny and which code for an extensive repertoire of histone variants (21, 47, 48). The repetitive early histone genes are transcriptionally activated upon meiotic maturation of the egg, are maximally expressed in the rapidly dividing blastula embryo, and are already silenced at the hatching blastula stage, when the transcripts of late histone genes start to accumulate (reviewed in references 13 and 45. The heterogeneous family of late histone genes consists of different single-copy genes that code for distinct H1, H2A, and H2B variants but for identical H3 and H4 proteins. The late histone genes are maximally active during late embryogenesis, and some of their members continue to be expressed in adult tissues (reference 39 and references therein). The family of sperm histone genes is exclusively transcribed during spermatogenesis and codes for specialized H1 and H2B proteins with basic N-terminal extensions which are responsible for the unusually high chromatin condensation in mature sperm (10, 54). Early, late, and sperm histone genes lack introns, contain the conserved 3′ terminal stem-loop structure instead of a poly(A) addition site, and thus show the classical hallmarks of replication-dependent histone genes (references 10 and 41 and references therein).

The so-called cleavage-stage (CS) histones constitute the fourth family of developmental histone variants of the sea urchin. The CS proteins are the first histones to be synthesized after fertilization in the cleaving embryo. Due to this property, the CS H1, H2A, and H2B variants were initially identified as in vivo-labeled proteins which differed in their electrophoretic mobility on high-resolution Triton-acid-urea gels from the early and late histone variants (47). The CS histones appear to be synthesized during oogenesis and in the mature egg, where

they give rise to a large maternal pool of histone proteins (35, 58). The CS proteins are the only histones present in the chromatin of the egg and zygote up to the second cell division (17, 35). Thereafter, the efficient synthesis of the early histones leads to a rapid dilution of the CS histones in the chromatin except in the nondividing small micromeres, where the CS proteins remain the major histones up to the pluteus larva stage (51). The maternally stored CS histones play an important role in remodeling of the sperm chromatin after fertilization (reviewed in reference 54). Upon entry in the egg cytoplasm, the sperm H1 and H2B proteins present in the male pronucleus are rapidly phosphorylated on their N-terminal extensions, and the sperm H1 protein is subsequently replaced in the chromatin by the CS H1 protein (29). This exchange of H1 histones is immediately followed by decondensation of the chromatin (29). At around the time of DNA replication, the CS H2A and H2B proteins start to accumulate in the chromatin of the male pronucleus, which correlates with the transition of the nucleosomal repeat length from 250 bp in the sperm chromatin to 200 bp in the early embryo (55, 59). As a consequence of this massive chromatin restructuring, the paternal genome is transcriptionally activated already at the beginning of S phase in the first cell cycle (46, 53), which further emphasizes the importance of the CS histones for early development. However, the CS histone genes have resisted all previous cloning attempts by screening with heterologous histone gene probes, and hence primary sequence information is still missing for this important class of nucleosomal proteins.

Recently, we have biochemically purified and microsequenced the CS histones of the sea urchin *Parechinus angulosus* (8). This partial peptides sequence information has now been used for reverse transcription-PCR amplification of CS H1, H2A, H2B, H3, and H4 cDNA from egg poly(A)$^+$ RNA of the related sea urchin species *Psammechinus miliaris*. All five genes are shown to be coordinately expressed during oogenesis, in the egg, and in early development up to the 128-cell blastula stage. The presence of introns, poly(A) addition signals, and long noncoding sequences indicates that the CS histone genes are replication-independent transcription units coding for replacement variants. Sequence comparison with known histones identified the CS variants (except for H4) as unique proteins which have evolved at a distant time. An ancient evolutionary origin of the CS histones is further supported by the observation that the H1M (B4) protein of *Xenopus laevis* (62) shares not only a similar developmental expression profile but also high sequence homology with the CS H1 protein. These data therefore identify the frog H1M (B4) gene as a vertebrate homolog of the CS H1 gene.

MATERIALS AND METHODS

Pulse-chase labeling of sea urchin histone proteins. Embryos of the sea urchin *P. miliaris* were raised at Marine Biological Station Millport (Isle of Cumbrae, Scotland). Sea urchin eggs were fertilized with diluted sperm, and embryos (10,000/ml) were cultured at 16°C under constant stirring in Millipore membrane-filtered sea water containing 0.05 mM EDTA and 50 mg each of penicillin and streptomycin per liter. For protein labeling, 50 μM [^{3}H]leucine (120 Ci/mmol; Amersham) and 75 μM [^{3}H]lysine (76 Ci/mmol; Amersham) were added to the embryo cultures. At the end of the labeling period (indicated in Fig. 1), embryos were allowed to settle and subsequently resuspended in seawater containing unlabeled leucine and lysine at a concentration of 10 mM each to start the chase period. After 20 min, the seawater was exchanged once more for seawater containing unlabeled leucine and lysine, each at a concentration of 1 mM. Embryos were collected by centrifugation at the gastrula or pluteus stage for chromatin preparation.

Histone isolation. Embryonic chromatin was prepared as described previously (4), and basic proteins were extracted by stirring the chromatin for 2 h at 4°C in 0.4 M H_2SO_4. Acid-insoluble material was removed by centrifugation, and histones were precipitated at −20°C with 4 volumes of ethanol. After centrifugation, the collected proteins were dialyzed against 0.5% acetic acid and then analyzed on sodium dodecyl sulfate (SDS)–15% polyacrylamide gels.

RNA preparation. Total RNA from ovaries, eggs, and different embryonic stages of the sea urchin *P. miliaris* was prepared by the guanidinium thiocyanate method (18). Egg RNA was fractionated into poly(A)$^+$ and poly(A)$^-$ RNA by three successive runs of oligo(dT)-cellulose chromatography.

Oligonucleotides. The following oligonucleotides were synthesized for PCR cloning of CS histone cDNA fragments: CS H1, 5′-GCGGGATCCATGAAPuAAPyPyTNPyTN(C/A)GNATGGC-3′ and 5′-GCGAAGCTTNGCNACNGCPyTTPyTTPyTTPyTC-3′; CS H2A, 5′-GCGGGATCCAAPyGCNGCN(C/A)GNGAPyAAPyAAPuAA-3′ and 5′-GCGAAGCTTAPuNACNGTPyTG(T/Pu)ATPuTTNGGCAT-3′; CSH2B, 5′-GCGGGATCCACNAAPuGGNGAPuAAPuAAPuCAPuGT-3′ and 5′-GCGAAGCTTC(G/T)PyTCPuAA(T/Pu)ATPuTCPuTTNACPuAA-3′; CS H3, 5′-GCGGGATCCATGGCN(A/C)GNACNAAPuCAPuACNGC-3′ and 5′-GCGCTGCAGPuCANAPuPuTTNGTPuTCPyTCPuAA-3′; and CS H4, 5′-GCGGGATCCGAPyAAPyAT(Py/A)CAPuGGNAT(Py/A)ACNAA-3′ and 5′-GCGAAGCTTGCPuTANACNACPuTCCATNGCNGT-3′. The primers used for PCR amplification from genomic CS H1 DNA were 5′-GCGCTGCAGCCTCGTGAAAACATAGGCGTGGC-3′ and 5′-GCGGAATTCTAGTCGCTGCCTTTGATTTCGGC-3′. Underlining indicates the recognition sequence of the restriction enzyme that was used for cloning.

PCR cloning of CS histone cDNA probes. Poly(A)$^+$ RNA isolated from *P. miliaris* eggs was transcribed into double-stranded cDNA by priming with $p(dT)_{15}$ as described previously (32). After *Eco*RI linker addition, this cDNA was cloned into bacteriophage λgt10. The *Eco*RI linker consisted of oligonucleotide A (5′-AATTCTCGAGCTCGTCGACA-3′) annealed with oligonucleotide B (5′-TGTCGACGAGCTCGAG-3′). The cloned cDNA inserts were amplified by PCR from DNA of this phage library by using oligonucleotide A as a primer. This amplified cDNA (100 ng) was used as template for amplifying cDNA fragments of the different CS histone genes with the degenerate primer pairs indicated above (50 pmol of each oligonucleotide), using standard PCR conditions (50 mM KCl, 10 mM Tris [pH 8.3], 1.5 mM $MgCl_2$, 0.1% [wt/vol] gelatin, 200 mM each deoxynucleoside triphosphate, 0.05 U of *Taq* DNA polymerase [Perkin-Elmer Cetus] per ml) and 40 cycles. The annealing temperature was adapted for each set of primers (2 min at 60°C for CS H1, 65°C for CS H2A, 53°C for CS H2B, 54°C for CS H3, and 58°C for CS H4), while the same conditions were used for the denaturation (1 min at 94°C) and DNA synthesis (2 min at 72°C) steps. Amplified CS H1 (174-bp), CS H2A (143-bp), CS H2B (200-bp), and CS H4 (197-bp) cDNA fragments were cloned into the *Bam*HI and *Hind*III sites of pSP64, while the CS H3 (333-bp) cDNA fragment was inserted into the *Bam*HI and *Pst*I sites of the same plasmid. The identity of the cloned cDNA inserts was verified by DNA sequencing.

The 5′ part of the CS H1 gene was amplified from *P. miliaris* sperm DNA (100 ng), using the specific primer pair and PCR buffer described above. The conditions for amplification were as follows: initial denaturation at 100°C for 10 min; 5 cycles of 1 min at 94°C, 1 min at 40°C, and 4 min at 72°C; and 35 cycles of 1 min at 94°C, 2 min at 60°C, and 4 min at 72°C. The amplified genomic DNA fragment was cloned into the *Eco*RI and *Pst*I sites of plasmid pSP64.

Cloning of full-length CS histone cDNAs. Recombinant phages (~10^6) of a *P. miliaris* egg cDNA library (constructed in phage λgt10; see above) were separately screened with probes which were radiolabeled by random priming of the cloned CS H1, CS H2A, CS H2B, and CS H4 PCR fragments, respectively. Positive clones were rescreened and plaque purified. The cDNA inserts of the different phages were subcloned as *Sal*I (CS H1 and CS H2A) or *Sac*I (CS H2B and CS H4) DNA fragments into the polylinker of pUC19. Clones pCSH1-2.2, pCSH2A-2.0, pCSH2B-2.5, and pCSH4-1.7, with insert sizes of 2.2, 2.0, 2.5, and 1.7 kb, respectively, were chosen as representative cDNA clones for DNA sequence analysis, which was performed on an automated sequencer (Applied Biosystems) by primer walking.

In vitro translation of CS histone proteins. The open reading frame of clone pCSH1-2.2 was cloned as a 1,266-bp *Ssp*I fragment, a 940-bp *Pst*I-*Eco*RI PCR fragment, and a 1,135-bp *Eco*RV-*Ssp*I fragment into the polylinker of pSP64 to generate constructs R1, R2, and R3, respectively. All fragments were inserted in the sense orientation downstream of the SP6 promoter. The 940-bp *Pst*I-*Eco*RI fragment was amplified by PCR with the two primers 5′-GCGCTGCAGATTCTTCTATATCTTCAAGATG-3′ and 5′-GCGGAATTCTAAAATAAACTGAACATCCATTA-3′. SP6 RNA transcripts of the three templates were in vitro translated in the presence of ^{3}H-labeled lysine (86 mCi/mmol; Amersham) in a nuclease-treated rabbit reticulocyte lysate as previously described (72). The protein products were analyzed by electrophoresis on an SDS–15% polyacrylamide gel, and the ^{3}H-labeled proteins were detected by sodium salicylate fluorography (16).

RNase protection and Northern blot analyses. Total RNA (15 μg) isolated from different embryonic stages was analyzed by RNase protection assay as described (67) except that a hybridization temperature of 60°C was used. Riboprobes were generated by SP6 transcription of the following plasmids which were linearized either by *Pvu*II (all CS probes) or *Afl*III digestion (early H1 probe). The CS H2B, CS H3, and CS H4 probes were transcribed from the respective PCR fragments that were initially cloned in pSP64 (see above). A 203-bp *Pvu*II-*Sty*I DNA fragment of pCSH1-2.2 (from nucleotides 678 to 880; see Fig. 3) was cloned in the antisense orientation into the *Sma*I site of pSP64 and then used as a template to generate the CS H1 riboprobe. The CS H2A probe was transcribed

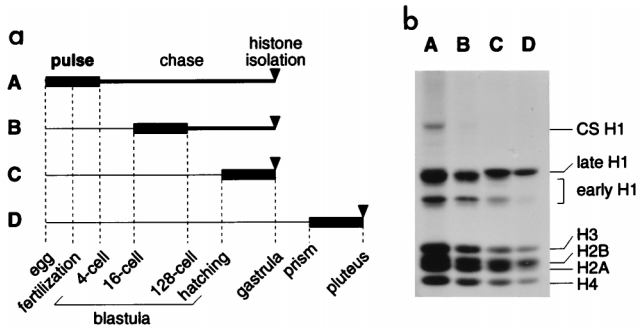

FIG. 1. In vivo labeling of embryonic histones of the sea urchin *P. miliaris*. (a) Schematic diagram of the labeling experiments. Newly synthesized proteins were radioactively labeled by incubating sea urchin embryos with [^{3}H]leucine and [^{3}H]lysine for the indicated time periods (denoted by black boxes). As the preparation of chromatin is more convenient at later times in development, embryos labeled during the early cleavage stages were incubated with nonradioactive leucine and lysine (chase period; indicated by thick lines) until the gastrula stage, when histones were isolated (experiments A and B). The eggs in experiment A were preincubated with the radioactive amino acids for 2 h prior to fertilization (b) SDS-PAGE analysis of the labeled histones. Twenty micrograms of each histone preparation was separated on an SDS–15% polyacrylamide gel followed by detection of the labeled proteins by fluorography. The same amount of histones was present in each lane, as indicated by Coomassie staining (data not shown). The positions of the different histone proteins are indicated to the right. The more slowly migrating protein of the two early H1 histones is encoded by the major h22 repeat unit (60), and the faster-migrating H1 variant is encoded by the minor repeat unit h104 (6) of the sea urchin *P. miliaris.*

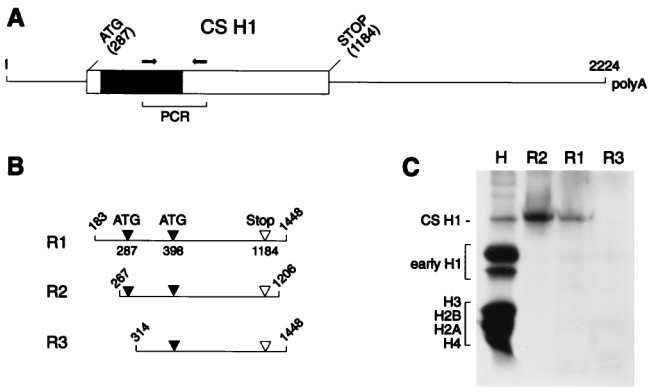

FIG. 2. Identification of the cloned CS H1 cDNA by in vitro translation. (A) Structural organization of the cDNA clone pCSH1-2.2. The coding region (nucleotide positions 287 to 1184) and the conserved core domain (positions 368 to 622) of the CS H1 protein are indicated by open and black boxes, respectively. The positions of the two degenerate PCR primers used for amplification of the CS H1 cDNA probe are indicated by arrows (for sequences, see Materials and Methods). (B) Schematic diagram of the templates used for in vitro translation. Three different templates (R1 to R3) containing the indicated nucleotide sequences of the CS H1 cDNA are schematically depicted. Filled arrowheads indicate the positions of the two start codons which are in frame with the coding sequence. Numbers refer to the nucleotide positions of the CS H1 cDNA sequence shown in Fig. 3. (C) In vitro translation analysis. The in vitro translation products of the templates R1, R2, and R3 were analyzed by SDS-PAGE in parallel with in vivo-labeled histones (H) isolated from four-cell embryos of *P. miliaris* (preparation A in Fig. 1). ^{3}H-labeled proteins were detected by fluorography, and the positions of the different histones are indicated to the left.

from a 315-bp *Sal*I-*Hae*III DNA fragment which was subcloned from pCSH2A-2.0 (from nucleotides 1 to 315) into the *Sal*I and *Sma*I sites of pSP65. A 168-bp *Sac*I-*Bam*HI DNA fragment isolated from the early H1 gene of the h22 repeat unit (60) was inserted into the polylinker of pSP64 to generate the early H1 riboprobe.

For Northern blot analysis, 20 μg of total or poly(A)$^-$ RNA or ~3 μg of poly(A)$^+$ RNA was separated on a 1.5% agarose-formaldehyde gel and then transferred to a nitrocellulose membrane followed by hybridization with randomly primed DNA probes as described previously (20).

Nucleotide sequence accession numbers. The *P. miliaris* CS H1, H2A, H2B, H3, and H4 cDNA sequences have been submitted to GenBank (accession numbers U84113 to U84117, respectively).

RESULTS

Identification of the CS H1 protein in cleaving embryos of the sea urchin *P. miliaris*. CS histones have so far been identified in the sea urchin *Strongylocentrotus purpuratus* as proteins that can be labeled in vivo during the early cleavage stages (47). We therefore investigated whether CS histones are also expressed during early development of *P. miliaris*. For this purpose, we incubated embryos of this sea urchin species with ^{3}H-labeled lysine and leucine at different intervals during embryogenesis, as schematically diagrammed in Fig. 1a. For histone isolation, we took advantage of the fact that the turnover of the CS and early histones, once incorporated in the chromatin, is known to be very low (48). Embryos labeled at early developmental stages were therefore raised in the presence of nonradioactive amino acids up to the gastrula stage, whereupon chromatin was prepared and histones were isolated by acid extraction. As shown by the gel electrophoretic analysis in Fig. 1b, radioactivity was incorporated into all five histone types at the developmental stages analyzed. SDS-polyacrylamide gel electrophoresis (PAGE) analysis did not, however, allow resolution of the different variants of the four core histones due to their minimal size differences. In contrast, distinct histone H1 subtypes could be detected. The CS H1 histone of *S. purpuratus* is known to be the largest of all H1 variants and to be efficiently synthesized only during the first few hours after fertilization (47). A protein with these two characteristic features was also detected in early embryos of *P. miliaris* (Fig. 1, histone preparation A), thus indicating that this sea urchin species also contains CS histone genes. This conclusion was confirmed by identification of the *Psammechinus* CS H2A and CS H2B variants on high-resolution Triton-acid-urea gels (data not shown).

Isolation and characterization of CS H1 cDNA. To clone CS H1 transcripts, we used a strategy based on the following two observations. First, preliminary in vitro translation experiments with poly(A)$^+$ and poly(A)$^-$ RNA isolated from sea urchin eggs suggested that the CS histones mRNAs are polyadenylated, in contrast to the early histone gene transcripts, which partitioned with poly(A)$^-$ RNA (unpublished data). Second, CS histones have recently been identified in four-cell embryos of *Parechinus angulosus*, a species closely related to *P. miliaris* (8). Biochemical purification and microsequencing resulted in a partial amino acid sequence of the *Parechinus angulosus* CS H1 protein (reference 8; see also Fig. 3). This sequence information was used to design degenerate primers, and cDNA transcribed from poly(A)$^+$ RNA of *P. miliaris* eggs was used as the template for PCR amplification of CS H1 DNA. In this manner, a DNA fragment of the expected length of 174 bp was obtained, and its identity was verified by cloning and DNA sequencing (Fig. 2A). This DNA fragment was subsequently used as a probe to screen an egg cDNA λgt10 library for full-length CS H1 cDNA, which resulted in the isolation of 11 positive phages. The majority of these phages carried inserts of 1,000 to 1,400 bp, while one clone contained a longer cDNA fragment. The insert of this latter phage was cloned into pUC19, sequenced, and thus shown to be 2,224 bp long. This cDNA clone contained an open reading frame of 997 bp which was subcloned as a 1,266-bp DNA fragment (construct R1) into pSP64 for in vitro translation analysis (Fig. 2B). As shown by the SDS-PAGE analysis in Fig. 2C, the protein encoded by construct R1 exhibited the same electrophoretic mobility as

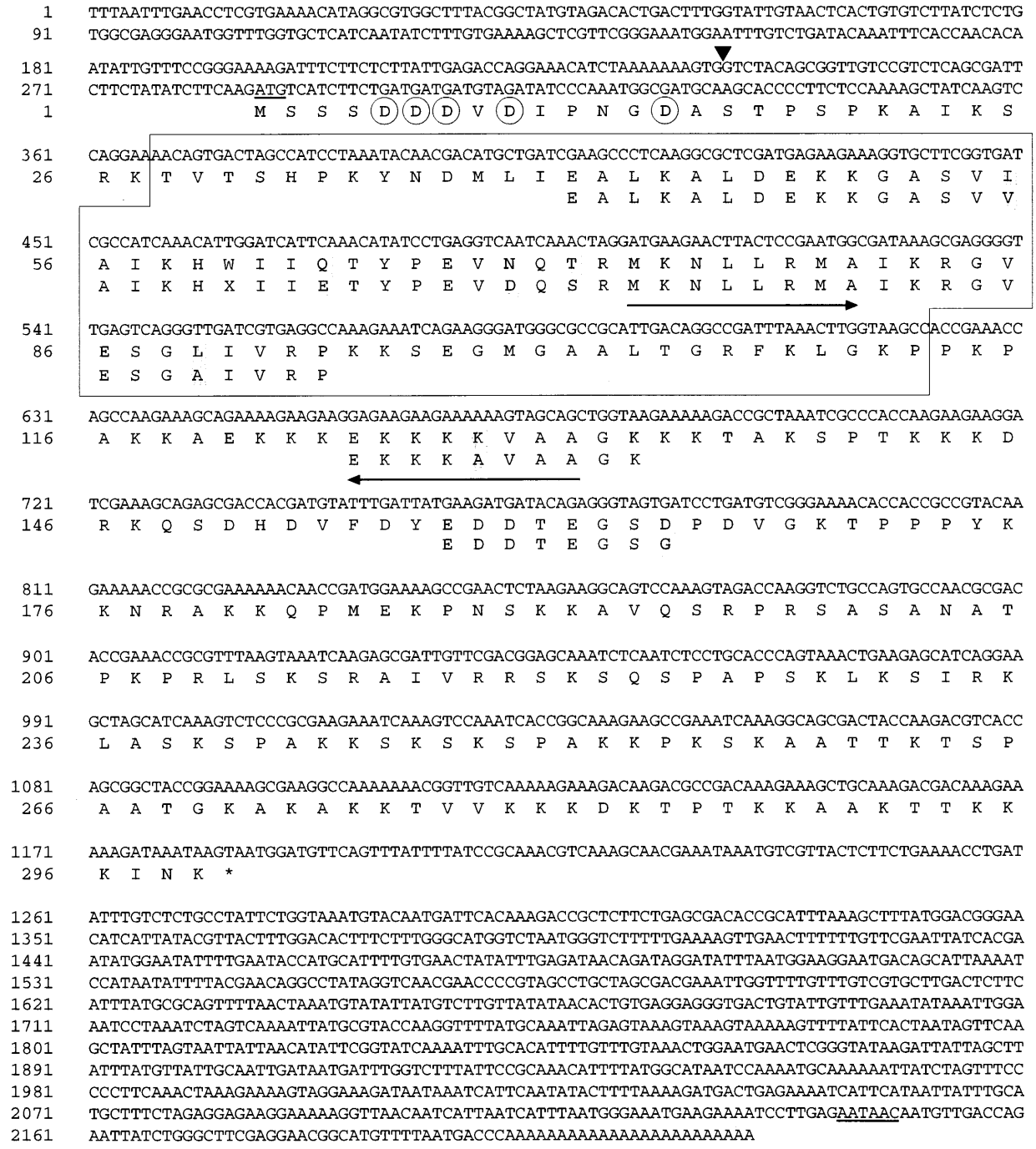

FIG. 3. CS histone H1 sequence. Nucleotide and deduced amino acid sequences of *P. miliaris* CS H1 cDNA (clone pCSH1-2.2). The translation initiation codon and a putative polyadenylation signal are underlined. The position of the intron in the 5′ untranslated sequences (Fig. 4B) is indicated by an arrowhead. The deduced amino acid sequence of the *P. miliaris* CS H1 protein is aligned with the peptide sequences obtained by microsequencing of the *Parechinus angulosus* CS H1 protein (8). Amino acids which differ between the two CS H1 proteins are indicated by grey shading. Aspartic acid residues which are clustered at the N terminus of the CS H1 protein are highlighted by circles. Arrows denote peptide sequences of the *Parechinus angulosus* CS H1 protein which were used for the design of degenerate PCR primers. An unidentified residue in the *Parechinus* peptide sequence is indicated as X. The conserved core domain of the CS H1 protein (boxed sequences) was defined as specified by Wells and McBride (69).

the CS H1 protein of early *P. miliaris* embryos, thus identifying the cloned cDNA as a CS H1 transcript.

The cDNA sequence coding for the CS H1 protein is shown in Fig. 3. A long open reading frame is preceded by two in-frame start codons at positions 287 and 398 in the 5′ region of the cDNA sequence (Fig. 2B and 3). To identify the start site used for translation of the CS H1 protein, we progressively eliminated sequences from the 5′ end of the open reading frame (Fig. 2B) and studied the effects of these deletions in in vitro translation assays (Fig. 2C). These experiments demonstrated that the synthesis of the CS H1 protein was strictly dependent on the presence of the start codon at position 287.

Initiation at this site therefore results in translation of a CS H1 polypeptide of 299 amino acids, which corresponds to the longest H1 protein identified to date. H1 histones are generally composed of three distinct regions consisting of a central globular domain flanked on either side by basic N- and C-terminal arms (69). The CS H1 protein also contains a central globular domain as well as a highly basic C-terminal sequence; in this respect, it resembles conventional H1 histones (Fig. 3B). In contrast to these proteins, the N-terminal region of the CS H1 protein is acidic due to the preponderance of aspartate residues. Reverse transcriptase-PCR analysis demonstrated that the same acidic sequence is also present at the N terminus of the CS H1 protein of the related sea urchin species *Paracentrotus lividus* (data not shown). Hence, this acidic N-terminal sequence appears to constitute an evolutionarily conserved feature of sea urchin CS H1 proteins. Finally, the partial peptide sequences of the *Parechinus angulosus* CS H1 histone (8) display 90% amino acid identity with the protein sequence deduced from our *P. miliaris* cDNA clone (Fig. 3), thus confirming that the cloned cDNA indeed codes for the CS H1 protein.

The CS H1 protein is encoded by a replacement histone gene. The CS H1 mRNA differs in several important aspects from the transcripts of other sea urchin H1 genes. First, the CS H1 mRNA lacks the 3′-terminal palindrome which is responsible for the U7 snRNP-mediated 3′ processing of replication-dependent histone gene transcripts (reviewed in reference 5). Instead, the CS H1 mRNA ends in a poly(A) tract which is preceded by a putative polyadenylation sequence (AATAAC). Second, the CS H1 cDNA, with a size of 2,224 bp, is unusually long due to the presence of extensive 5′ and 3′ noncoding sequences (286 and 1,040 bp, respectively). To verify both of these features, we used Northern blot hybridization to analyze the CS H1 transcripts in total RNA from sea urchin ovaries and eggs (Fig. 4A). A prevalent mRNA of ~3 kb was detected in both of these tissues, thus indicating that our cloned CS H1 cDNA is incomplete and may lack ~800 bp of leader sequence at the 5′ end. Hence, the CS H1 mRNA appears to contain leader and trailer sequences similar in length (~1,050 bp). Moreover, the CS H1 transcripts could be detected only in the poly(A)$^+$-selected mRNA fraction, not in nonpolyadenylated RNA of the egg. In contrast, the early H1 mRNA was found only in the poly(A)$^-$ RNA, as predicted by the presence of the 3′-terminal palindrome (11). Both polyadenylation and the exceptional length of the CS H1 mRNA indicate that the CS H1 gene belongs to the family of replacement histone genes. As the presence of introns is often an additional hallmark of this class of histone genes, we investigated whether the CS H1 gene is also interrupted by intron sequences. As illustrated in Fig. 4B, PCR amplification from genomic DNA resulted in a DNA fragment that was ~700 bp longer than predicted by the cloned cDNA. Sequence analysis of this PCR fragment revealed the presence of a 707-bp intron which is located 45 bp upstream of the translation start codon in the leader sequence of the CS H1 gene. In summary, we conclude that the CS H1 gene possesses all of the characteristic features (long noncoding sequences, intron, and polyadenylation site) of a replacement histone gene.

The CS H1 protein of the sea urchin shows highest homology with the frog H1M (B4) histone. Of all five histone types, the linker histone H1 is least well conserved during evolution and exhibits the characteristic tripartite structure which is absent in core histone proteins. The central globular domain of H1 is known to interact with the core histones in the nucleosome (56, 66) and has been sufficiently conserved to allow direct sequence comparison (69). In contrast, the unstructured

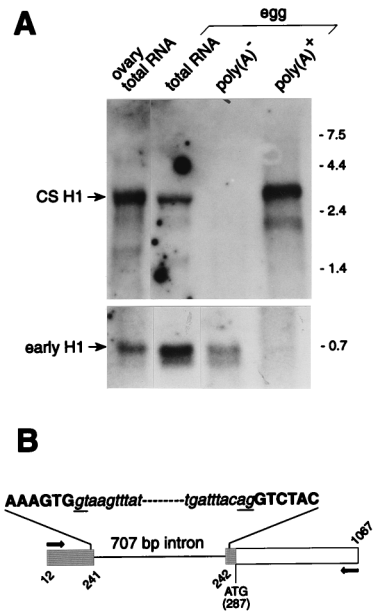

FIG. 4. The CS H1 mRNA is encoded by a replacement histone gene. (A) Northern blot analysis of *P. miliaris* egg and ovary RNA. Total RNA isolated from ovaries and eggs as well as poly(A)$^+$ and poly(A)$^-$ RNA from eggs were subjected to Northern blot hybridization using the *Sal*I insert of clone pCSH1-2.2 as a DNA probe. The filter was subsequently rehybridized with a labeled *Sal*I-*Bam*HI DNA fragment of the early H1 gene (60). (B) Presence of an intron in the 5′ noncoding region of CS H1 gene of *P. miliaris*. Numbers refer to nucleotide positions of the cDNA sequence (Fig. 3). Exon and intron sequences are shown in capital and lowercase letters, respectively. The invariant GT and AG dinucleotides of the 5′ and 3′ splice sites are underlined. Arrows indicate the positions of the two oligonucleotides used for PCR amplification of the CS H1 gene from *P. miliaris* sperm DNA (for primer sequences, see Materials and Methods).

N- and C-terminal arms of the H1 protein interact with linker DNA (66) and are so rich in basic amino acids that any reliable sequence alignment is precluded (69). Sequence comparison with known H1 proteins led to the definition of a central globular domain of 85 amino acids for the CS H1 protein (Fig. 3 and 5). In addition, the CS H1 protein possesses an N-terminal arm of 27 amino acids and a C-terminal tail of 187 residues (35% basic amino acids, predominantly lysines) which are responsible for both the unusually large size and the basic character of this H1 variant.

Six different H1 variants are known to be sequentially expressed during sea urchin development. Expression of the CS H1 protein at the cleavage stages is followed by the synthesis of the early H1 protein (α) in blastula embryos and the subsequent expression of the late H1 subtypes β, γ, and δ during later development (21, 43, 47), while the sperm H1 protein is expressed exclusively during spermatogenesis (63). As shown by the sequence comparison in Fig. 5, all of these sea urchin H1 histones exhibit a relatively low degree of sequence identity (30 to 40%) with the CS H1 protein in the central globular domain, which points to an ancient evolutionary origin of the CS H1 protein. While the H1-α, -β, and -γ subtypes and the sperm H1 variant are encoded by typical replication histone genes, the late H1-δ protein is translated from a polyadenylated mRNA and thus appears to be a replacement variant (43). Of all sea urchin H1 subtypes, the H1-δ protein displays the highest homology (40%) with the CS H1 protein, suggesting that some of the conserved amino acid residues of these two proteins may be characteristic of replacement H1 proteins.

	species	protein	total length	identity
A	P. miliaris	CS H1	299	100 %
B	P. angulosus	CS H1	n.d.	91 %
C	Xenopus laevis	H1M (B4)	273	47 %
D	S. purpuratus	H1-δ	185	40 %
E	S. purpuratus	H1-γ	217	38 %
F	S. purpuratus	H1-β	211	34 %
G	P. miliaris	H1-α	206	30 %
H	P. angulosus	sperm H1	249	29 %

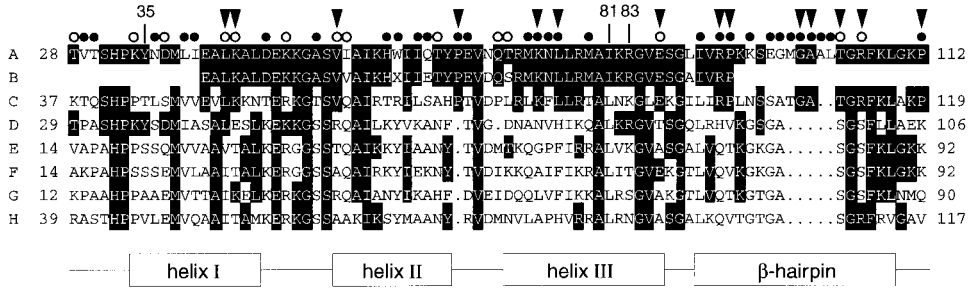

FIG. 5. High sequence conservation of the CS H1 and frog H1M (B4) proteins in the central globular domain. The central globular regions of the CS H1 proteins of the sea urchins *P. miliaris* and *Parechinus angulosus* are compared with those of the early (α) and late (β, γ, and δ) H1 variants of the sea urchins *P. miliaris* and *S. purpuratus* as well as with the H1M (B4) protein of *X. laevis*. The conserved globular domain of H1 proteins, defined as specified by Wells and McBride (69), is shown together with a schematic diagram of its secondary structure as determined by X-ray crystallographic analysis of histone H5 (56) and by nuclear magnetic resonance studies of histone H1 (15). Amino acid residues identical with the *P. miliaris* CS H1 protein are highlighted by black overlay; amino acids unique for the two CS H1 histones in a compilation of 35 H1 and 4 H5 protein sequences from plants and animals (69) are indicated by filled circles; open circles denote amino acid residues which are found in only one or two of the 35 H1 and 4 H5 protein sequences analyzed. The frog H1M (B4) protein was not part of the sequence compilation, and the unique residues of the CS H1 proteins that have been conserved in the frog H1M (B4) sequence are denoted by arrowheads. Numbers to the left and right refer to the first and last amino acid residues, respectively, of the different H1 protein sequences. Three amino acid residues which are characteristic of vertebrate H1° and H5 proteins (69) are also numbered. Gaps which have been introduced for optimal sequence alignment are indicated by dots. The length of each H1 protein is given in numbers of amino acids, and the degree of sequence identity with the *P. miliaris* CS H1 protein was calculated without considering gaps. n.d., not determined. H1 protein sequences (references): *Parechinus angulosus* CS H1 (8); *P. miliaris* H1-α (60); *S. purpuratus* H1-β (42), H1-γ (41), and H1-δ (43); *Parechinus angulosus* sperm H1 (63); *X. laevis* H1M (B4) (62). The South African sea urchin *Parechinus angulosus* and the Pacific species *S. purpuratus* diverged from the North Atlantic species *P. miliaris* about 25 million and 65 million years ago, respectively (12).

Two special features of the CS H1 protein were revealed by sequence comparison with 35 different H1 and 4 H5 proteins which originate from distantly related animals as well as plants (69). The CS H1 protein contains a unique insertion of five amino acids in the C-terminal part of the globular domain (Fig. 5) which has recently been shown to assume a β-hairpin conformation (15, 56). Apart from this insertion, 25 amino acid residues of the CS H1 protein are never found at the corresponding position in any of the H1 and H5 proteins analyzed, while 14 amino acids are present only once or twice at the respective positions in the entire H1 sequence database (Fig. 5). The maternal H1M protein of *X. laevis* (also referred to as B4 [27, 62]) was not part of the histone H1 collection analyzed. However, direct sequence comparison demonstrated that the frog H1M (B4) protein displays the highest homology with the CS H1 protein (47% sequence identity) (Fig. 5). Even more significantly, the globular domain of the H1M (B4) protein contains also a characteristic C-terminal insertion of 3 amino acids and 14 of the CS H1-specific amino acid residues. In addition, the frog H1M (B4) protein, with a size of 273 amino acids, is also an exceptionally long H1 histone similar to the CS H1 protein (Fig. 5). Collectively, these data suggest that the H1M (B4) histone is a vertebrate homolog of the CS H1 protein.

The CS H2A, H2B, and H3 genes code for unique histone variants of the sea urchin. We next cloned the remaining four members of the CS histone gene family. Partial peptide sequence information was available for the CS H2A, H2B, and H3 proteins of *Parechinus angulosus*, while there was no biochemical evidence for a CS-specific H4 variant (8). Using degenerate PCR primers (Fig. 6), we amplified cDNA fragments of 143, 200, 333, and 197 bp from H2A, H2B, H3, and H4 mRNAs present in poly(A)$^+$ mRNA of *P. miliaris* eggs. Screening of an egg poly(A)$^+$ cDNA library with these probes resulted in the isolation of two H2A cDNA clones of 2 and 3 kb, two H2B cDNA clones of 2.3 and 2.5 kb, and two H4 cDNA clones of 1.7 and 1.8 kb. As shown by DNA sequence analysis, the 3′ region of each of these cDNA clones contained a polyadenylation signal followed by a poly(A) stretch (data not shown), thus indicating that the corresponding mRNAs are also encoded by replacement histone genes analogous to the CS H1 transcript. The H2A and H2B protein sequences deduced from these *P. miliaris* cDNAs deviated only at two and four amino acid positions from the *Parechinus angulosus* CS H2A and H2B polypeptides, respectively (Fig. 6A and B). Moreover, the H3 sequence encoded by the cloned PCR fragment was identical with that of the known CS H3 peptides (Fig. 6C). Together, these data identified the cloned cDNAs as transcripts of the CS H2A, CS H2B, and CS H3 genes. In the absence of protein data, it was impossible to positively identify the two H4 cDNA clones as CS H4 transcripts. In particular, both H4 cDNA clones coded for the same H4 protein (data not shown) as the early and late sea urchin histone genes (17, 57, 60). However, the expression analysis shown below strongly suggested that the two cDNA clones correspond to transcripts of the CS H4 gene.

Histone H3, similar to H4, is known to be highly conserved throughout evolution, as evidenced by the fact that the early and late H3 genes of the sea urchin encode the same protein (17, 57, 60). Interestingly, the CS H3 histone differs at three

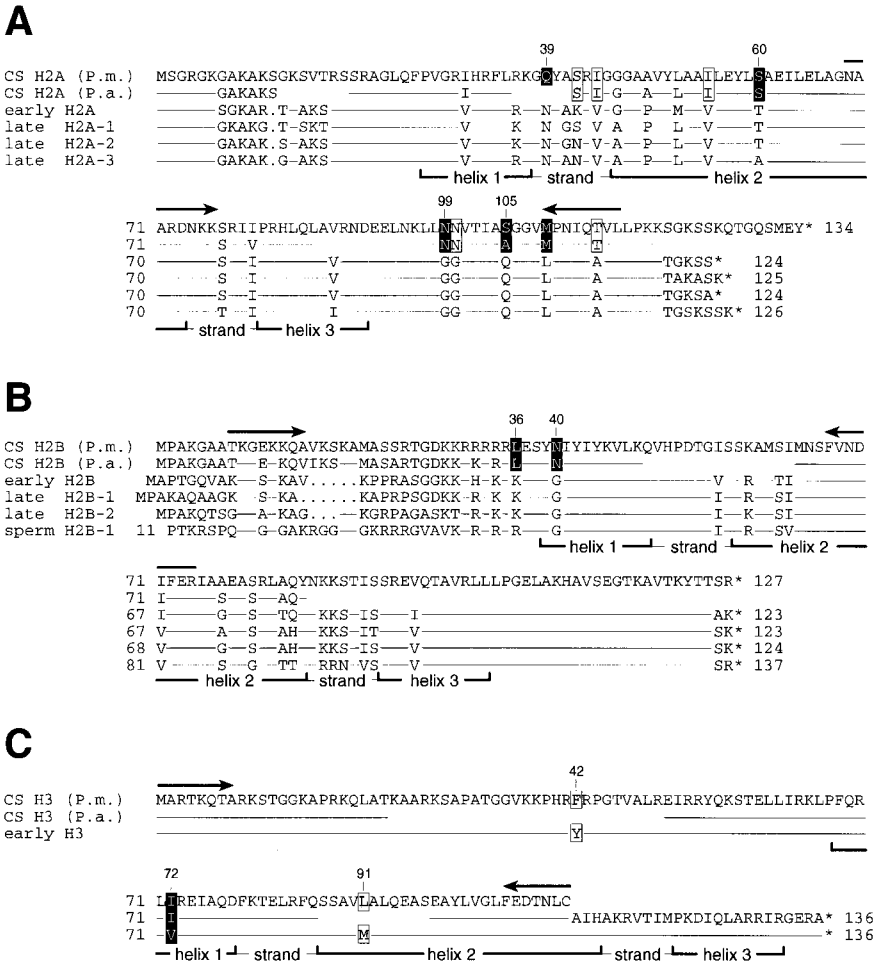

FIG. 6. Sequence comparison of the CS H2A, H2B, and H3 proteins with the corresponding histone subtypes of *P. miliaris*. The CS H2A (A), CS H2B (B), and CS H3 (C) protein sequences of *P. miliaris* (P.m.) were deduced from the cDNA sequences of clones pCSH2A-2.0 and pCSH2B-2.5 as well as from the cloned CS H3 PCR fragment, respectively. The partial sequences of the CS H2A, H2B, and H3 proteins of *Parechinus angulosus* (P.a.) were obtained by microsequencing (8). The early histone sequences are encoded by the genes of the *P. miliaris* repeat unit h22 (60), while the late and sperm histone sequences have been published by Busslinger and Barberis (10). The partial H2B-2 sequence was extended at the C terminus by the H2B-2.1 sequence (39). Amino acids which are identical with the corresponding CS histone sequence of *P. miliaris* are indicated by lines. Numbers refer to amino acid positions within the different CS histone proteins. Residues which are unique to the CS histones in a compilation of 51 H2A, 46 H2B, and 57 H3 protein sequences from plants and animals (69) are highlighted by a black overlay. Grey shading indicates those amino acids within the CS histones which are rarely found at the corresponding positions in other proteins of the same histone type. Amino acids constituting the α-helices and β-strand segments of the histone fold motif (3) are indicated by brackets. Arrows denote the peptide sequences of the *Parechinus angulosus* CS histones which were used to design degenerate primers for PCR amplification of the corresponding *P. miliaris* cDNA sequences.

amino acid positions from the early H3 counterpart (Fig. 6C). Isoleucine 72 is unique to the CS H3 histone, as it is not found at this position in any of 57 H3 proteins from plants to humans (69). Phenylalanine 42 and leucine 91 are not observed in H3 protein of the animal kingdom but are present in H3 histones of plants and ciliated protozoa (69). Hence, the CS H3 protein exhibits very ancient evolutionary traits.

The histones H2A and H2B are known to evolve at a more rapid pace than the H3 and H4 proteins. In particular, the N-terminal sequences of H2B proteins are often divergent enough to render sequence alignment difficult. As shown in Fig. 6B, the remainder of the CS H2B protein harbors two amino acid residues (36 and 40) which are not encountered at these positions in any other H2B protein (69). Likewise, the CS H2A protein contains five unique amino acids (39, 60, 99, 105, and 109), while the residues at five additional positions (42, 44, 55, 100, and 114) rarely occur in other H2A histones (Fig. 6A). In summary, we conclude that the CS proteins constitute a distinct family of ancient histones.

Coordinate expression of all five CS histone genes during oogenesis and early development of the sea urchin. We next studied the developmental expression profile of the different CS histone genes by using RNase protection analysis to measure steady-state mRNA levels. For this purpose, we generated gene-specific riboprobes to detect the distinct CS histone transcripts in total RNA isolated from ovary, egg, and different embryonic stages. As shown in Fig. 7A, all five CS histone mRNAs are most abundantly present in the ovary, indicating that the CS histone genes are maternally expressed during oogenesis. The five CS mRNAs are present at a slightly lower but constant level in the egg and during the cleavage stages up to the 128-cell blastula embryo. Thereafter, the CS mRNAs rapidly decline with the exception of the CS H2B and H4 transcripts, which are still detectable at a relatively high level in

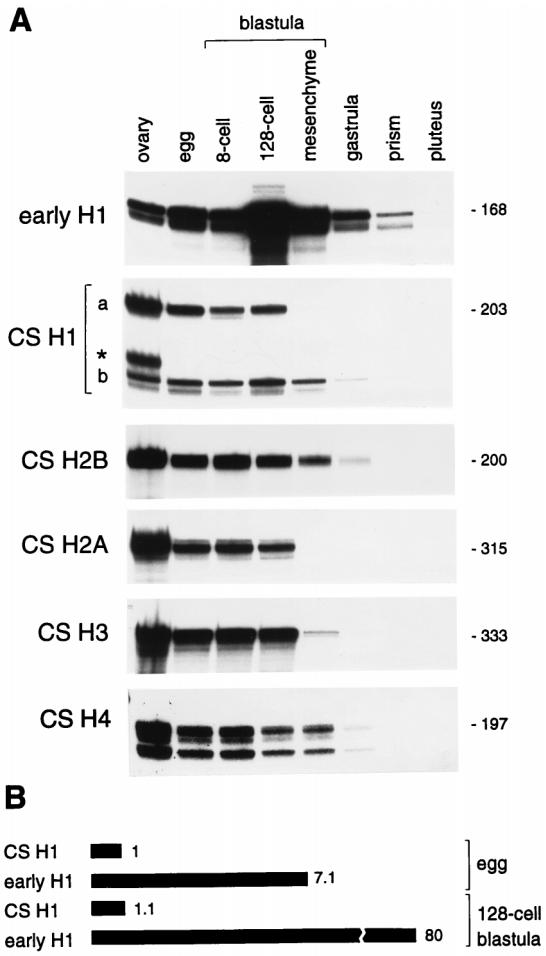

FIG. 7. Coordinate expression of the different CS histone genes during oogenesis and early development. (A) RNase protection analysis. Total RNA (15 μg) isolated from different developmental stages of *P. miliaris* was analyzed by RNase protection assay with riboprobes that were specific for the early H1 mRNA and the indicated CS histone gene transcripts. Only the relevant part of the autoradiograph containing the RNase-protected fragments of the indicated sizes (given in nucleotides to the right) is shown. The autoradiograph containing the early H1 mRNA signals was exposed for 6 h, while all other autoradiographs were exposed for 1 day. The CS H1 probe consistently resulted in two distinct RNase-protected signals (a and b), the origin of which is explained in the text. A third RNase-protected fragment (indicated by an asterisk) was obtained only with the ovary RNA preparation and may reflect the presence of a polymorphic sequence in the CS H1 gene. (B) Quantitation of H1 mRNAs. Decreasing amounts (5 to 0.2 μg) of total RNA from the egg and blastula embryo were mapped with the early H1 riboprobe, while 15-μg aliquots of the same RNAs were analyzed in parallel with the CS H1 probe. The RNase-protected signals of duplicate experiments were quantitated on a PhosphorImager and are shown as average values relative to the CS H1 mRNA level in the egg.

the mesenchyme blastula embryo. Analysis of various tissues indicated that the CS H1 transcript could be detected in the ovary but not in any other tissue of the adult sea urchin (data not shown). Hence, expression of the CS histone genes appears to be restricted to oogenesis and early embryonic development.

The expression pattern of the H4 mRNA closely resembled that of the other CS histone transcripts, indicating that the cloned H4 cDNAs originate from the CS H4 gene(s). The two isolated CS H4 cDNAs differ in the coding sequences by 5% of silent nucleotide substitutions (data not shown), which suggests the existence of two nonallelic CS H4 genes in the sea urchin genome. Consistent with this conclusion, the CS H4 probe gave rise to two distinct RNase-protected signals (Fig. 7A). A similar observation was also made with the CS H1 probe, which consisted of a 203-nucleotide-long cDNA sequence (positions 678 to 880 in Fig. 3) encoding part of the variable C-terminal tail of the CS H1 protein. This probe not only was fully protected by the homologous CS H1 mRNA (band a) but also resulted in a shorter RNase-protected signal (band b) (Fig. 7A). Since PCR amplification from genomic DNA did not reveal an intron at a corresponding position in the CS H1 gene (Fig. 4), we interpret this result to mean that a second, closely related CS H1 gene with a similar expression pattern may exist in the sea urchin genome.

To quantitate the levels of the CS H1 mRNA relative to its early embryonic counterpart, we have also mapped the early H1 transcript at the same developmental stages. Early histone gene expression is known to be initiated in late oogenesis, thus resulting in a maternal store of histone mRNAs that are entirely sequestered in the nucleus of the egg (23). After fertilization, these early histone mRNAs accumulate to maximal levels in the 128-cell blastula embryo (44, 45), as is also demonstrated in Fig. 7A. Throughout early embryogenesis, the CS H1 mRNA appeared to be less prevalent than the early H1 mRNA. Direct quantitation of the two H1 transcripts demonstrated that the CS H1 mRNA was indeed ~7- and ~80-fold less abundant than the early H1 mRNA in the egg and blastula embryo, respectively (Fig. 7B).

DISCUSSION

The CS histones of the sea urchin were discovered 18 years ago as a distinct class of variants that constitute the only histones in the chromatin of the egg and the early embryo up to the four-cell stage (47). These histones have since been shown to participate in chromatin remodeling of the male pronucleus in the zygote (reviewed in reference 54). Here we have described the cloning and characterization of cDNA transcripts for all five members of this histone family and have identified the encoded proteins as CS histones based on partial peptide sequence information, in vitro translation experiments, and expression analyses. These experiments have indicated that the CS proteins are unique histones of ancient origin which may perform conserved functions in oogenesis and early development.

The CS histones are encoded by a family of replacement histone genes. While the majority of histone genes code for nonpolyadenylated mRNAs ending in a stem-loop structure, we have isolated all five CS histone gene transcripts (CS H1, H2A, H2B, H3, and H4) from poly(A)$^+$ mRNA of the egg. As shown by RNase protection experiments, these CS histone genes are transcribed only during a short period of the sea urchin life cycle. They are coordinately expressed during oogenesis, in the egg, and in the embryo up to the blastula stage, while their transcripts could be detected neither during later development nor in somatic tissues of the adult sea urchin. This mRNA expression pattern is consistent with available data on CS protein synthesis (35, 47) and is compatible with a role of these histones at the onset of sea urchin development (55).

All five CS histone mRNAs not only are polyadenylated but also contain long 5′ and 3′ noncoding sequences, which distinguishes them from transcripts of the replication-dependent early, late, and sperm histone gene families. In addition, we have shown that the CS H1 gene contains an intron in its 5′ noncoding region. The CS histone genes exhibit therefore all characteristic features of replacement histone genes, which are known to be expressed at a basal level throughout the cell cycle

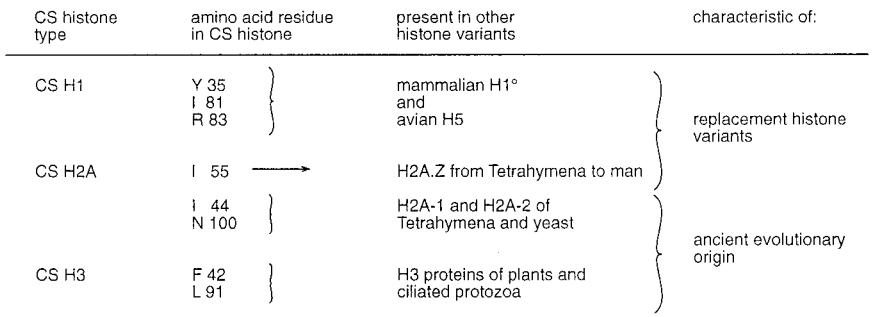

FIG. 8. Evolutionarily conserved amino acid residues of CS histones. The histone sequence compilation of Wells and McBride (69) was used as a database for comparison with the CS histones. The H2A.Z variant of the chicken has been referred to as H2A.F (22), and the corresponding variant of *Tetrahymena* is designated histone hv1 (71). For references of the different histone sequences, see reference 69. Additional H2A.Z sequences have been published for the sea urchin (28) and *Drosophila melanogaster* (65).

as well as in quiescent cells (9, 22, 68). The expression profile of the CS histone genes further supports the hypothesis that these genes are expressed in a replication-independent manner. Before fertilization, the CS histone genes are transcribed in the oocyte and egg at a time when histone synthesis is uncoupled from DNA replication. Hence, a replication-independent mode of CS histone gene expression can explain the accumulation of relatively large stores of CS proteins in the unfertilized egg (35, 58). After fertilization, the CS histone transcripts are still present at the same basal level up to the blastula stage, although this early period of sea urchin development is characterized by rapid cell divisions. In contrast, expression of the replication-dependent early histone genes increases at least 10-fold during the same time period (10, 44) (Fig. 7), leading to a rapid increase in early histone proteins and to a dilution of the CS histones in the chromatin of proliferating cells (47) (Fig. 1). The small micromeres, which originate at the vegetal pole of the 16-cell embryos, are the first cells to withdraw from the embryonic cell cycle and later become part of the coelomic sacs of the pluteus larva (51). Interestingly, the CS histones are still the major histones in the chromatin of these nondividing cells at the pluteus larva stage (51).

The CS proteins are specialized histones involved in chromatin remodeling of the paternal genome. The CS proteins, with the exception of H4, are unique histones, as they contain amino acid residues which have so far not been found at the same positions in any of the corresponding histones of plants and animals (Fig. 5 and 6). For instance, one-third of the central globular domain of the CS H1 protein consists of unique amino acids which are absent in all other H1 proteins of eukaryotes (Fig. 5). The most illustrative example is, however, the CS H3 protein. While different early and late H3 genes of the sea urchin encode the same protein (17, 57, 60), the CS H3 histone deviates from this protein at three amino acid positions. Isoleucine 72 is entirely unique to the CS H3 protein (Fig. 6). Two other amino acid residues, phenylalanine 42 and leucine 91, are present in H3 proteins of plants and ciliated protozoa but not in any H3 histone of the animal kingdom (Fig. 8). Likewise, the CS H2A protein contains, in addition to its unique residues, two amino acids which have been detected so far only in the H2A proteins of the lower eukaryotes *Saccharomyces cerevisiae* and *Tetrahymena thermophila* (Fig. 8). Hence, the CS proteins of the sea urchin share features with histones of lower eukaryotes and plants which may reflect either convergent evolution or, more likely, a common ancestry of these proteins. The CS H1 and H2A proteins furthermore contain amino acid residues which are characteristic of replacement histone variants. First, the CS H1 histone shows highest sequence similarity among all sea urchin H1 proteins with the replacement variant H1-δ (43) (Fig. 5). Second, the CS H1 protein comprises three amino acid residues, tyrosine 35, isoleucine 81, and arginine 83, which are present at the corresponding positions only in the replication-independent mammalian H1° and avian H5 proteins (Fig. 8). Moreover, isoleucine 55 of the CS H2A histone is exclusively found in H2A.Z variants from *Tetrahymena* to human (Fig. 8), which are known to be expressed at a basal level throughout the cell cycle (22, 33, 70).

All four core histones possess a three-dimensional structure, the so-called histone fold, which consists of three α-helices separated by short loop and β-strand segments (1, 3) (Fig. 6). This structural motif appears to be a tandem duplication of a fundamental element, the helix-strand-helix (HSH) motif, with the junction point near the middle of the long second helix of the histone fold (2). Sequence comparisons indicated that the amino acid residues of the N-terminal HSH1 motif are less stringently conserved than those of the C-terminal HSH2 segment (3). Consistent with this notion, the unique amino acids of the CS H2A, H2B, and H3 histones are located only within the N-terminal HSH1 motif or outside the histone fold altogether (Fig. 6).

The function of the CS histones appears to be intimately linked with that of the sperm histones, since both histone gene families are transcribed during gametogenesis. While the CS histones are expressed in oogenesis and early development (references 35 and 47 and this study), the sperm H2B and H1 proteins are synthesized exclusively during spermatogenesis (13, 54). These sperm histones are first expressed in spermatogonia and subsequently function as normal histones in transcription and replication during meiosis due to the fact that their basic N-terminal extensions are phosphorylated and thus neutralized (52). At the late spermatid stage, these N-terminal sequences become dephosphorylated (52) and subsequently stabilize the highly condensed and transcriptionally inert chromatin of the mature sperm by strong ionic interaction with linker DNA (30, 36). In contrast, the CS histones constitute the chromatin of the transcriptionally active oocyte and egg (14, 35) and participate as maternally stored proteins in chromatin remodeling of the male pronucleus in the zygote (55). Within 5 min after fertilization, the sperm H1 and H2B proteins become rephosphorylated followed by replacement of the sperm H1 histone with the CS H1 protein and by chromatin decondensation (29). Later, the CS H2A and CS H2B proteins start to accumulate in the male pronucleus, which coincides with the transition of the nucleosomal repeat length from 250 bp in

sperm to 200 bp in the embryo (55). In analogy to the involvement of the basic N-terminal repeats of the sperm histones in chromatin condensation, we speculate that the unique sequences of the CS histones may have evolved to facilitate the assembly of active chromatin in the oocyte and egg as well as to promote chromatin remodeling of the paternal genome in the zygote.

The linker histone H1M (B4) of *X. laevis* is a vertebrate homolog of the CS H1 protein implicated in the organization of early embryonic chromatin. The cloning and characterization of the CS H1 gene have also led to the surprising discovery of an orthologous relationship between the H1M (B4) gene of *X. laevis* and the CS H1 gene of the sea urchin. The B4 mRNA was originally identified as a maternal transcript in frog oocytes by differential cDNA cloning (62). Both the H1M (B4) and CS H1 proteins are the longest H1 histones known to date and have the highest degree of sequence identity (47%) in their central globular domains among all H1 proteins analyzed (Fig. 5). The two H1 histones are also classical replacement variants, as both are encoded by polyadenylated transcripts of intron-containing genes (references 19 and 62 and this study). Even the expression patterns of the two H1 genes are very similar in their respective species. The H1M (B4) mRNA is initially transcribed during oogenesis, remains at a constant level throughout the blastula stages, starts to decline in the gastrula embryo, and is absent at later developmental stages and in adult tissues (62). Consistent with this expression pattern, the H1M (B4) protein is present in the chromatin of the oocyte, egg, and embryo up to early gastrula stage (24, 27, 37, 62), comparable to the situation observed with the CS H1 histone in sea urchin development. Moreover, the H1M (B4) protein is also maternally stored in the *Xenopus* egg and has been shown to participate in rapid remodeling of the sperm chromatin in frog egg extracts (25, 50). In summary, all of these common features indicate that the *Xenopus* H1M (B4) gene is a vertebrate homolog of the CS H1 gene.

The structural role of H1M (B4) in chromatin organization has recently been studied by in vitro reconstitution experiments. The H1M (B4) protein was shown to associate with the nucleosome by binding to linker DNA in the same manner as histone H1 (49). Hence, the specific incorporation into nucleosomal DNA identifies H1M (B4) as a classical linker histone. However, the H1M (B4) protein binds with sixfold-lower affinity to reconstituted nucleosomes and can only inefficiently repress 5S rRNA gene transcription in vitro compared to histone H1 (64). The presence of the H1M (B4) protein in early frog embryos is therefore thought to contribute to the formation of more extended and less stable chromatin structures which are more easily replicated and transcribed (64). The H1M (B4) protein is present in the embryonic chromatin up to the midblastula transition, when it starts to be diluted out by the de novo synthesis of somatic histone H1 (24). The progressive replacement of embryonic H1M (B4) by adult H1 in the chromatin is responsible for the specific repression of the oocyte 5S rRNA genes in the gastrula embryo (7, 38), in agreement with recent genetic evidence that histone H1 variants of *Tetrahymena* are also involved in gene-specific rather than global transcriptional regulation (61). A role for H1M (B4) in gene regulation is further supported by the correlation observed between the incorporation of H1M (B4) and the acquisition of transcriptional competence in somatic nuclei following chromatin remodeling in frog egg extracts (26). Together, these data indicate a role for H1M (B4) in the establishment of accessible and transcriptionally competent chromatin structures. This biological function may have been conserved since the divergence of echinoderms and chordates, as the sea urchin CS H1 and frog H1M (B4) proteins share many conserved features (see above).

Given the evolutionary conservation of the CS H1 and H1M (B4) histones, it is likely that the frog genome also contains counterparts for the other CS histone genes of the sea urchin. Indeed, an H2A variant, H2A.X, which differs in electrophoretic mobility from somatic H2A histones, has been identified as a maternally stored protein in frog egg extracts (40, 50). Histone H2A.X, like the H1M (B4) protein, is involved in remodeling of the sperm chromatin and furthermore becomes phosphorylated upon mobilization from the storage pool and incorporation into the male pronuclear chromatin (25, 40). The same phosphorylation behavior has been observed for the CS H2A histone of the sea urchin (31), thus strengthening the hypothesis that the H2A.X variant is the frog equivalent of the sea urchin protein. The availability of the CS histone sequences should now greatly facilitate the cloning and identification of the frog H2A.X and other CS histone homologs of vertebrates.

ACKNOWLEDGMENTS

We are grateful to G. Schaffner for oligonucleotide synthesis, to R. Kurzbauer and I. Botto for DNA sequencing, and to M. Cotten and P. Pfeffer for critical reading of the manuscript.

This work was supported in part by a grant from the Austrian Industrial Research Promotion Fund.

REFERENCES

1. **Arents, G., R. W. Burlingame, B.-C. Wang, W. E. Love, and E. N. Moudrianakis.** 1991. The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix. Proc. Natl. Acad. Sci. USA **88:**10148–10152.
2. **Arents, G., and E. N. Moudrianakis.** 1993. Topography of the histone octamer surface: repeating structural motifs utilized in the docking of nucleosomal DNA. Proc. Natl. Acad. Sci. USA **90:**10489–10493.
3. **Arents, G., and E. N. Moudrianakis.** 1995. The histone fold: a ubiquitous architectural motif utilized in DNA compaction and protein dimerization. Proc. Natl. Acad. Sci. USA **92:**11170–11174.
4. **Bentinnen, L. C., and D. G. Comb.** 1971. Early and late histones during sea urchin development. J. Mol. Biol. **57:**355–358.
5. **Birnstiel, M. L., M. Busslinger, and K. Strub.** 1985. Transcription termination and 3′ processing: the end is in site! Cell **41:**349–359.
6. **Birnstiel, M. L., R. Portmann, M. Busslinger, W. Schaffner, W. Probst, and A. Kressmann.** 1979. Functional organization of the histone genes in the sea urchin *Psammechinus*: a progress report. Alfred Benzon Symp. **13:**117–129.
7. **Bouvet, P., S. Dimitrov, and A. P. Wolffe.** 1994. Specific regulation of *Xenopus* chromosomal 5S rRNA gene transcription in vivo by histone H1. Genes Dev. **8:**1147–1159.
8. **Brandt, W. F., S. Schwager, G. Rodriques, and M. Busslinger.** Isolation and partial amino acid sequences of the cleavage stage (CS) histones of the sea urchin. Eur. J. Biochem., in press.
9. **Brush, D., J. B. Dodgson, O.-R. Choi, P. W. Stevens, and J. D. Engel.** 1985. Replacement variant histone genes contain intervening sequences. Mol. Cell. Biol. **5:**1307–1317.
10. **Busslinger, M., and A. Barberis.** 1985. Synthesis of sperm and late histone cDNAs of the sea urchin with a primer complementary to the conserved 3′ terminal palindrome: evidence for tissue-specific and more general histone gene variants. Proc. Natl. Acad. Sci. USA **82:**5676–5680.
11. **Busslinger, M., R. Portmann, and M. L. Birnstiel.** 1979. A regulatory sequence near the 3′ end of sea urchin histone genes. Nucleic Acids Res. **6:**2997–3008.
12. **Busslinger, M., S. Rusconi, and M. L. Birnstiel.** 1982. An unusual evolutionary behavior of a sea urchin histone gene cluster. EMBO J. **1:**27–33.
13. **Busslinger, M., D. Schümperli, and M. L. Birnstiel.** 1985. Regulation of histone gene expression. Cold Spring Harbor Symp. Quant. Biol. **50:**665–670.
14. **Carroll, A. G., and H. Ozaki.** 1979. Changes in the histones of the sea urchin *Strongylocentrotus purpuratus* at fertilization. Exp. Cell Res. **119:**307–315.
15. **Cerf, C., G. Lippens, V. Ramakrishnan, S. Muyldermans, A. Segers, L. Wyns, S. J. Wodak, and K. Hallenga.** 1994. Homo- and heteronuclear two-dimensional NMR studies of the globular domain of histone H1: full assignment, tertiary structure, and comparison with the globular domain of H5. Biochemistry **33:**11079–11086.

16. **Chamberlain, J. P.** 1979. Fluorographic detection of radioactivity in polyacrylamide gels with the water-soluble fluor sodium salicylate. Anal. Biochem. **98:**132–135.
17. **Childs, G., C. Nocente-McGrath, T. Lieber, C. Holt, and J. A. Knowles.** 1982. Sea urchin (Lytechinus pictus) late-stage histone *H3* and *H4* genes: characterization and mapping of a clustered but nontandemly linked multigene family. Cell **31:**383–393.
18. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18:**5294–5299.
19. **Cho, H., and A. P. Wolffe.** 1994. *Xenopus laevis B4*, an intron-containing oocyte-specific linker histone-encoding gene. Gene **143:**233–238.
20. **Church, G. M., and W. Gilbert.** 1984. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81:**1991–1995.
21. **Cohen, L. H., K. M. Newrock, and A. Zweidler.** 1975. Stage-specific switches in histone synthesis during embryogenesis of the sea urchin. Science **190:** 994–997.
22. **Dalton, S., A. J. Robins, R. P. Harvey, and J. R. E. Wells.** 1989. Transcription from the intron-containing chicken H2A.F gene is not S-phase regulated. Nucleic Acids Res. **17:**1745–1756.
23. **DeLeon, D. V., K. H. Cox, L. M. Angerer, and R. C. Angerer.** 1983. Most early-variant histone mRNA is contained in the pronucleus of sea urchin eggs. Dev. Biol. **100:**197–206.
24. **Dimitrov, S., G. Almouzni, M. Dasso, and A. P. Wolffe.** 1993. Chromatin transitions during early *Xenopus* embryogenesis: changes in histone H4 acetylation and in linker histone type. Dev. Biol. **160:**214–227.
25. **Dimitrov, S., M. C. Dasso, and A. P. Wolffe.** 1994. Remodeling sperm chromatin in Xenopus laevis egg extracts: the role of core histone phosphorylation and linker histone B4 in chromatin assembly. J. Cell Biol. **126:**591–601.
26. **Dimitrov, S., and A. P. Wolffe.** 1996. Remodeling somatic nuclei in Xenopus laevis egg extracts: molecular mechanisms for the selective release of histones H1 and H1^{o} from chromatin and the acquisition of transcriptional competence. EMBO J. **15:**5897–5906.
27. **Dworkin-Rastl, E., H. Kandolf, and R. C. Smith.** 1994. The maternal histone H1 variant, H1M (B4 protein), is the predominant H1 histone in *Xenopus* pregastrula embryos. Dev. Biol. **161:**425–439.
28. **Ernst, S. G., H. Miller, C. A. Brenner, C. Nocente-McGrath, S. Francis, and R. McIsaac.** 1987. Characterization of a cDNA clone coding for a sea urchin histone H2A variant related to the H2A.F/Z histone protein in vertebrates. Nucleic Acids Res. **15:**4629–4644.
29. **Green, G. R., and D. L. Poccia.** 1985. Phosphorylation of sea urchin sperm H1 and H2B histones precedes chromatin decondensation and H1 exchange during pronuclear formation. Dev. Biol. **108:**235–245.
30. **Green, G. R., and D. L. Poccia.** 1988. Interaction of sperm histone variants and linker DNA during spermiogenesis in the sea urchin. Biochemistry **27:**619–625.
31. **Green, G. R., and D. L. Poccia.** 1989. Phosphorylation of sea urchin histone CS H2A. Dev. Biol. **134:**413–419.
32. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25:**263–269.
33. **Hatch, C. L., and W. M. Bonner.** 1988. Sequence of cDNAs for mammalian H2A.Z, an evolutionarily diverged but highly conserved basal histone H2A isoprotein species. Nucleic Acids Res. **16:**1113–1124.
34. **Hentschel, C. C., and M. L. Birnstiel.** 1981. The organization and expression of histone gene families. Cell **25:**301–313.
35. **Herlands, L., V. G. Allfrey, and D. Poccia.** 1982. Translational regulation of histone synthesis in the sea urchin *Stronglyocentrotus purpuratus*. J. Cell Biol. **94:**219–223.
36. **Hill, C. S., J. M. Rimmer, B. N. Breen, J. T. Finch, and J. O. Thomas.** 1991. Histone-DNA interactions and their modulation by phosphorylation of -Ser-Pro-X-Lys/Arg- motifs. EMBO J. **10:**1939–1948.
37. **Hock, R., A. Moorman, D. Fischer, and U. Scheer.** 1993. Absence of somatic histone H1 in oocytes and preblastula embryos of *Xenopus laevis*. Dev. Biol. **158:**510–522.
38. **Kandolf, H.** 1994. The H1A histone variant is an *in vivo* repressor of oocyte-type 5S gene transcription in *Xenopus laevis* embryos. Proc. Natl. Acad. Sci. USA **91:**7257–7261.
39. **Kemler, I., and M. Busslinger.** 1986. Characterization of two nonallelic pairs of late histone H2A and H2B genes of the sea urchin: differential regulation in the embryo and tissue-specific expression in the adult. Mol. Cell. Biol. **6:**3746–3754.
40. **Kleinschmidt, J. A., and H. Steinbeisser.** 1991. DNA-dependent phosphorylation of histone H2A.X during nucleosome assembly of *Xenopus laevis* oocytes: involvement of protein phosphorylation in nucleosome spacing. EMBO J. **10:**3043–3050.
41. **Knowles, J. A., Z.-C. Lai, and G. J. Childs.** 1987. Isolation, characterization, and expression of the gene encoding the late histone subtype H1-gamma of the sea urchin *Strongylocentrotus purpuratus*. Mol. Cell. Biol. **7:**478–485.
42. **Lai, Z.-C., and G. Childs.** 1988. Characterization of the structure and transcriptional patterns of the gene encoding the late histone subtype H1-β of the sea urchin *Stronglocentrotus purpuratus*. Mol. Cell. Biol. **8:**1842–1844.
43. **Lieber, T., L. M. Angerer, R. C. Angerer, and G. Childs.** 1988. A histone H1 protein in sea urchins is encoded by a poly(A)$^{+}$ mRNA. Proc. Natl. Acad. Sci. USA **85:**4123–4127.
44. **Mauron, A., L. Kedes, B. R. Hough-Evans, and E. H. Davidson.** 1982. Accumulation of individual histone mRNAs during embryogenesis of the sea urchin *Strongylocentrotus purpuratus*. Dev. Biol. **94:**425–434.
45. **Maxson, R., R. Cohn, and L. Kedes.** 1983. Expression and organization of histone genes. Annu. Rev. Genet. **17:**239–277.
46. **Maxson, R. E. J., and J. C. Egrie.** 1980. Expression of maternal and paternal histone genes during early cleavage stages of the echinoderm hybrid *Strongylocentrotus purpuratus* × *Lytechinus pictus*. Dev. Biol. **74:** 335–342.
47. **Newrock, K. M., C. R. Alfageme, R. V. Nardi, and L. H. Cohen.** 1978. Histone changes during chromatin remodeling in embryogenesis. Cold Spring Harbor Symp. Quant. Biol. **42:**421–431.
48. **Newrock, K. M., L. H. Cohen, M. B. Hendricks, R. J. Donnelly, and E. S. Weinberg.** 1978. Stage-specific mRNAs coding for subtypes of H2A and H2B histones in the sea urchin embryo. Cell **14:**327–336.
49. **Nightingale, K., S. Dimitrov, R. Reeves, and A. P. Wolffe.** 1996. Evidence for a shared structural role for HMG1 and linker histones B4 and H1 in organizing chromatin. EMBO J. **15:**548–561.
50. **Ohsumi, K., and C. Katagiri.** 1991. Occurrence of H1 subtypes specific to pronuclei and cleavage-stage nuclei of anuran amphibians. Dev. Biol. **147:** 110–120.
51. **Pehrson, J. R., and L. H. Cohen.** 1986. The fate of small micromeres in sea urchin development. Dev. Biol. **113:**522–526.
52. **Poccia, D., M. V. Simpson, and G. R. Green.** 1987. Transitions in histone variants during sea urchin spermatogenesis. Dev. Biol. **121:**445–453.
53. **Poccia, D., R. Wolff, S. Kragh, and P. Williamson.** 1985. RNA synthesis in male pronuclei of the sea urchin. Biochim. Biophys. Acta **824:**349–356.
54. **Poccia, D. L., and G. R. Green.** 1992. Packaging and unpackaging the sea urchin sperm genome. Tends Biochem. Sci. **17:**223–227.
55. **Poccia, D. L., T. Greenough, G. R. Green, E. Nash, J. Erickson, and M. Gibbs.** 1984. Remodeling of sperm chromatin following fertilization: nucleosome repeat length and histone variant transitions in the absence of DNA synthesis. Dev. Biol. **104:**274–286.
56. **Ramakrishnan, V., J. T. Finch, V. Graziano, P. L. Lee, and R. M. Sweet.** 1993. Crystal structure of globular domain of histone H5 and its implication for nucleosome binding. Nature **362:**219–223.
57. **Roberts, S. B., K. E. Weisser, and G. Childs.** 1984. Sequence comparison of non-allelic late histone genes and their early stage counterparts. J. Mol. Biol. **174:**647–662.
58. **Salik, J., L. Herlands, H. P. Hoffmann, and D. Poccia.** 1981. Electrophoretic analysis of the stored histone pool in unfertilized sea urchin eggs: quantification and identification by antibody binding. J. Cell Biol. **90:**385–395.
59. **Savic, A., P. Richman, P. Williamson, and D. Poccia.** 1981. Alterations in chromatin structure during early sea urchin embryogenesis. Proc. Natl. Acad. Sci. USA **78:**3706–3710.
60. **Schaffner, W., G. Kunz, H. Daetwyler, J. Telford, H. O. Smith, and M. L. Birnstiel.** 1978. Genes and spacers of cloned sea urchin histone DNA analyzed by sequencing. Cell **14:**655–671.
61. **Shen, X., and M. A. Gorovsky.** 1996. Linker histone H1 regulates specific gene expression but not global transcription in vivo. Cell **86:**475–483.
62. **Smith, R. C., E. Dworkin-Rastl, and M. B. Dworkin.** 1988. Expression of a histone H1-like protein is restricted to early *Xenopus* development. Genes Dev. **2:**1284–1295.
63. **Strickland, W. N., M. Strickland, W. F. Brandt, C. von Holt, A. Lehmann, and B. Wittmann-Liebold.** 1980. The primary structure of histone H1 from sperm of the sea urchin Parechinus angulosus. Eur. J. Biochem. **104:**567–578.
64. **Ura, K., K. Nightingale, and A. P. Wolffe.** 1996. Differential association of HMG1 and linker histones B4 and H1 with dinucleosomal DNA: structural transitions and transcriptional repression. EMBO J. **15:**4959–4969.
65. **van Daal, A., E. M. White, M. A. Gorovsky, and S. C. R. Elgin.** 1988. *Drosophila* has a single copy of the gene encoding a highly conserved histone H2A variant of the H2A.F/Z type. Nucleic Acids Res. **16:**7487–7497.
66. **van Holde, K. E.** 1989. Chromatin. Springer Verlag, New York, N.Y.
67. **Vitelli, L., I. Kemler, B. Lauber, M. L. Birnstiel, and M. Busslinger.** 1988. Developmental regulation of micro-injected histone genes in sea urchin embryos. Dev. Biol. **127:**54–63.
68. **Wells, D., and L. Kedes.** 1985. Structure of a human histone cDNA: evidence that basally expressed histone genes have intervening sequences and encode polyadenylated mRNAs. Proc. Natl. Acad. Sci. USA **82:**2834–2838.
69. **Wells, D., and C. McBride.** 1989. A comprehensive compilation and alignment of histones and histone genes. Nucleic Acids Res. **17:**r311–r346.
70. **White, E. M., and M. A. Gorovsky.** 1988. Localization and expression of

mRNA for a macronuclear-specific histone H2A variant (hv1) during the cell cycle and conjugation of *Tetrahymena thermophila*. Mol. Cell. Biol. **8:**4780–4786.

71. **White, E. M., D. L. Shapiro, C. D. Allis, and M. A. Gorovsky.** 1988. Sequence and properties of the message encoding *Tetrahymena* hv1, a highly evolutionarily conserved histone H2A variant that is associated with active genes. Nucleic Acids Res. **16:**179–198.
72. **Wrighton, C., and M. Busslinger.** 1993. Direct transcriptional stimulation of the ornithine decarboxylase gene by Fos in PC12 cells but not in fibroblasts. Mol. Cell. Biol. **13:**4657–4669.
73. **Wu, R. S., and W. M. Bonner.** 1981. Separation of basal histone synthesis from S-phase histone synthesis in dividing cells. Cell **27:**321–330.
74. **Wu, R. S., S. Tsai, and W. M. Bonner.** 1982. Patterns of histone variant synthesis can distinguish G0 from G1 cells. Cell **31:**367–374.