MOLECULAR AND CELLULAR BIOLOGY, Jan. 1987, p. 478–485
0270-7306/87/010478-08$02.00/0

Vol. 7, No. 1

Isolation, Characterization, and Expression of the Gene Encoding the Late Histone Subtype H1-Gamma of the Sea Urchin *Strongylocentrotus purpuratus*

JAMES A. KNOWLES, ZHI-CHUN LAI, AND GEOFFREY J. CHILDS*

Department of Genetics, Albert Einstein College of Medicine, Bronx, New York 10461

Received 10 July 1986/Accepted 6 October 1986

We cloned and characterized the gene encoding H1-gamma, a late histone subtype of the sea urchin species *Strongylocentrotus purpuratus*. The predicted primary sequence of H1-gamma is 216 amino acids in length and has a net charge of +70, which is high for a somatic H1 histone. The H1-gamma gene appears to be a unique sequence gene that is not tightly linked to the core histone genes. The 770-base-pair transcribed region of the H1-gamma gene is bordered on the 5′ side by two previously described H1-specific sequence elements and on the 3′ side by a hairpin loop structure and CAGA box sequences. We detected 3,900 stored maternal H1-gamma mRNA transcripts per egg. The number of H1-gamma transcripts per embryo rises by 9.5 h postfertilization, but the maximum rate of accumulation (4,300 molecules per min per embryo) occurs in the late-blastula-stage embryo between 14 and 21 h after fertilization. The number of H1-gamma mRNA molecules peaks 21 h after fertilization when there are 2.0×10^6 molecules per embryo (a 500-fold increase) and then decreases over the next 3.25 h to 1.3 million molecules per embryo. Between 24 and 82 h after fertilization the number of H1-gamma transcripts declines steadily (210 molecules per min per embryo) to reach approximately 5.4×10^5 H1-gamma mRNAs by 82 h postfertilization. Surprisingly, the number of late H1 mRNA molecules per embryo is greater than the number of late H2B mRNA molecules beginning at the early gastrula stage of development.

In the sea urchin, histone proteins are encoded by several multigene families. These families encode the cleavage stage, early (alpha), late (beta and gamma), and testis variants (reviewed in reference 34). Each family is regulated in both a temporal and a tissue-specific fashion. The two best characterized families in the sea urchin are those encoding the early and late subtypes. The early histone gene family is composed of approximately 500 nearly identical tandem repeats each of which encodes one copy of each of the five histone classes (25, 33, 34). The temporal expression of the early family of histone genes has been studied by several methods (8, 18, 24, 27, 32, 35, 36, 39, 53), and a general picture of the pattern of their expression has emerged. About 10^6 mRNA molecules encoding each of the four core histones and 80,000 molecules of H1-alpha mRNA (32) accumulate after meiotic maturation and are stored in the egg pronucleus until fertilization (1). The number of mRNA molecules remains constant until new synthesis begins at the 16-cell stage, resulting in a 10-fold increase in early histone mRNA by the time the embryo becomes a blastula (32, 35, 36, 53).

Members of the core late histone gene family have been cloned and are present in about 5 to 10 copies for each core histone (9, 26, 33). Transcripts encoding the late subtype core histones are also detected in unfertilized eggs but at a much lower level than that observed for the alpha genes (10,000 molecules per egg) (28). The genes encoding the late core histones are activated at the same time as the early histone genes, but late gene transcripts continue to accumulate up to the onset of gastrulation, long past the peak of early histone mRNA accumulation (8, 18, 24, 27, 33). In *Lytechinus pictus*, the amount of mRNA encoding the core histones H3 and H4 has been observed to peak at 1.8×10^6 molecules per embryo. At least part of this accumulation can be accounted for by an increase in the relative rate of transcription between 7 and 14 h after fertilization (27). Although the gene encoding late H1 protein has not previously been cloned, metabolic labeling and in vitro translation of histone mRNA indicate that late H1 mRNA has a pattern of accumulation that is similar to the core histones in both *Strongylocentrotus purpuratus* and *L. pictus* (3, 8, 24, 46).

In *S. purpuratus*, the pattern of histone H1 protein expression has been well characterized (10, 21). During the first three cell divisions H1-cs protein, which is stored in the egg, is the predominant H1 found on the rapidly dividing chromatin (38). As cleavage proceeds H1-alpha is synthesized in large quantities and becomes the major H1 protein on the chromatin. By the blastula stage of development, about 12 h postfertilization, two new H1 proteins, H1-beta and H1-gamma, begin to accumulate and subsequently become the predominant H1 proteins synthesized for the remainder of the life cycle of the organism (29, 41). In the later stage of embryogenesis and in adult tissues a fifth subtype, H1-lambda, can be detected but little is known about this protein (41). In addition to these somatic H1 proteins, there is a well-characterized H1 protein found only in sperm (49).

The H1 proteins were the first class of histones shown to be expressed in a temporal manner in embryos (45, 48); however, owing to large variability in the amino acid sequences of H1 subtypes, the corresponding genes have been reticent to cloning. While examples of sea urchin early, late, and sperm core histone genes have been cloned, only the early H1 gene has been isolated. This report details the

* Corresponding author.

cloning, DNA sequence, and temporal pattern of expression of the H1-gamma gene of *S. purpuratus*.

MATERIALS AND METHODS

Preparation of RNA. For the developmental time series, eggs from a single female were fertilized with sperm from a single male and cultured in Pacific Ocean water at 15°C (27). Samples were withdrawn at various times, and RNA was prepared by phenol extraction (19). 9S RNA for cDNA synthesis was prepared by sucrose gradient centrifugation of RNA extracted from 36-h-old postfertilization embryos.

Synthesis of cDNA. Oligonucleotides were synthesized on an Applied Biosystems automated DNA synthesizer. Oligonucleotide (5'-TTNGC$^{C}_{T}$TT$^{C}_{T}$TNGG-3') (1 μg) was annealed to 25 μg of 9S RNA in 100 μl containing 700 mM KCl, 50 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid) (pH 7), and 1 mM EDTA by heating to 90°C for 2 min, leaving at room temperature for 3.5 h, and then at 4°C overnight. First-strand synthesis was performed as described previously (31) with the addition of 100 μg of actinomycin D per ml. The second strand was constructed as described previously (20). cDNA was then C tailed, annealed to G-tailed *Pst*I-cut pBR322, and transformed into *Escherichia coli* RR1 (40). Colonies were screened by hybridization to a chicken H1 gene (pCHIaBIBKI) (51) which had been recloned into M13 and radiolabeled by primer extension (28). Approximately 2,700 individual colonies were screened, yielding a single positive cDNA clone.

Isolation of genomic clones. Late H1 cDNA was subcloned into *Pst*-cut M13mp18, radiolabeled by primer extension (27), and used to screen a partial *Eco*RI *S. purpuratus* genomic library (A-B) as described previously (4). A single clone, λSplH1-1, was isolated from this library. Ten additional overlapping clones corresponding to the same late H1 gene in λSplH1-1 were isolated from a partial *Mbo*I library in EMBL-3 (Bimbo) that was kindly supplied by Eric Davidson. We screened this library with a different H1 cDNA clone that was isolated from testis RNA (Z.-C. Lai and G. Childs, unpublished data). Purified bacteriophage were grown in liquid culture, and DNA was purified as described previously (31).

DNA sequencing and analysis. A 3.2-kilobase (kb) *Hpa*I-*Eco*RI fragment of λSplH1-1 was subcloned into M13mp19, and a series of nested deletions was constructed as described previously (13). To provide clones to sequence the opposite strand, we subcloned restriction endonuclease-generated fragments of the 3.2-kb *Hpa*I-*Eco*RI fragment into appropriately cut M13 vectors. M13 clones were sequenced by the dideoxy sequencing technique (47). DNA and protein sequences were analyzed by using the BIONET National Computer Resource.

Filter hybridization. Chicken H1 probes were hybridized to sea urchin DNA at 37°C in 30% formamide–5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–5× Denhardt solution–50 mM sodium phosphate (pH 6.8)–0.1% sodium dodecyl sulfate–0.1 mg of sonicated salmon sperm DNA per ml. Filters were washed at 65°C in 2× SSC–0.1% sodium dodecyl sulfate–1% sodium PP_i. Genomic DNA hybridizations with homologous probes were carried out at 42°C in the above hybridization buffer with 50% formamide and 10% dextran sulfate added (50% formamide hybridization buffer). The probe used was a nick-translated (43) *Sac*II-*Eco*RI fragment (base pairs [bp] 383 to 1182) of H1-gamma. Filters were washed at 65°C in 0.5× SSC–0.1% sodium dodecyl sulfate–1% sodium PP_i. For RNA slot blots, 1 μg of RNA per slot was loaded onto nitrocellulose as described by the supplier (Schleicher & Schuell, Inc., Keene, N. H.). Sp6 transcripts of equal specific activity were synthesized by mixing all the components necessary for the transcription reaction except DNA (37), dividing this mix into equal aliquots, and adding equal amounts of linerized template to each reaction. Hybridizations were carried out at 58°C in 50% formamide hybridization buffer with a probe concentration of 20 ng/ml. Filters were washed at 65°C in 0.1× SSC–0.1% sodium dodecyl sulfate–1% sodium PP_i and then treated with RNase A in 2× SSC as described previously (27). Using these conditions we could detect no cross-hybridization between early and late genes on Northern blots. Autoradiographic intensity was quantitated with a Cambridge Instruments Quantimat 920 image analysis system.

Sp6 templates. The following Sp6 templates were used. (i) H1-alpha specific. pCO2H1+ was linearized with *Bam*HI, yielding a 1,150-bp transcript. (ii) H1-gamma specific. The 3.2-kb *Hpa*I-*Eco*RI fragment of λSplH1-1 was subcloned into pGEM-2, subcloned DNA linearized with *Ssp*I, yielding a 952-bp transcript. (iii) H2B-L1 specific. pSp65-H2B (30) was linearized with *Bam*HI, yielding an 800-bp template.

RESULTS

Isolation of cDNA clones. In sea urchins, early and late histone mRNAs are present in approximately equal amounts in late-blastula-stage embryos (22 h postfertilization) (27, 33). Under standard hybridization conditions the *S. purpuratus* early H1 gene does not cross-hybridize to late H1 mRNA (unpublished observations), so a chicken histone H1 gene was tested (51). When a chicken histone H1 gene was used to probe a Northern blot of *S. purpuratus* blastula RNA, both early and late H1 mRNAs hybridized with equal intensity (data not shown). Because this probe hybridized to the highly repetitive early genes, the direct screening of genomic libraries for late H1 was precluded. To circumvent this problem, we constructed a cDNA library using 9S RNA from embryos 36 h postfertilization, when the amount of early histone mRNA is negligible compared with late histone mRNA (27, 33). Since late H1, like most histone mRNAs, contains no poly(A) tail (unpublished observations), an oligonucleotide complementary to the amino acid sequence KKAKKP was used to prime first-strand synthesis. A search of all published histone H1 and H5 protein sequences revealed that this amino acid sequence is always present in the C-terminal tail domain. Screening of our cDNA library with the chicken H1 probe yielded a clone, pJK-1, that hybridizes to an mRNA of approximately 750 nucleotides. The RNA complementary to pJK-1 is present in orders of magnitude concentration higher in gastrula-stage embryos than in unfertilized egg RNA (data not shown), in accordance with the expected late H1 pattern of expression. A partial sequence of the cDNA subsequently revealed that the insert encodes a repeating lysine and alanine amino acid sequence typical of H1 proteins, but differing from H1-alpha (data not shown).

Isolation and sequencing of genomic clones. pJK-1 was then used to screen an *S. purpuratus* partial *Eco*RI genomic library (A-B). One clone, λSplH1-1, was found that is highly homologous to pJK-1. This clone contains a single H1 gene but is not homologous to any of the core histone genes. Ten additional overlapping clones were subsequently isolated from a different genomic library (see Materials and Methods). Restriction endonuclease maps of these overlapping

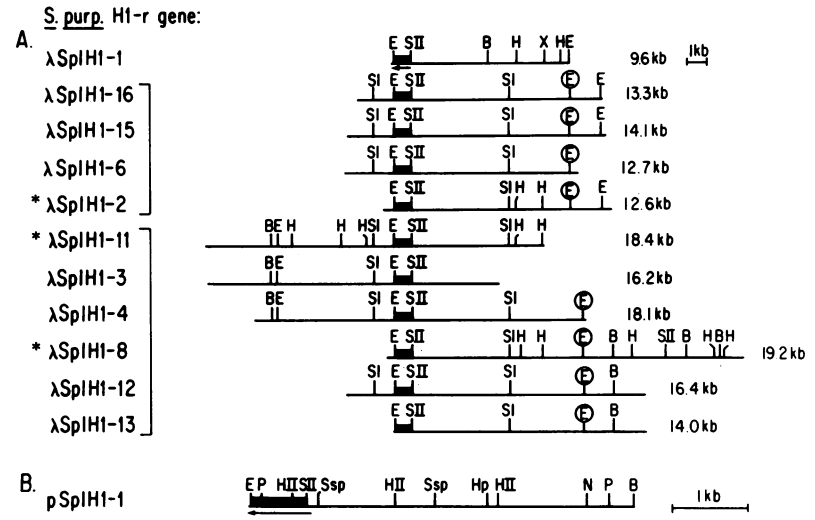

FIG. 1. Restriction endonuclease maps of DNA segments containing the *S. purpuratus* H1-gamma gene. Bars denote histone H1-coding sequences, and arrows show the direction of transcription. Maps were generated from examination of single- and double-enzyme digestions of phage and plasmid DNA. (A) Restriction endonuclease maps generated from eleven H1-gamma phage clones. The size of each entire insert is shown to the right. Brackets enclose the two classes of H1-gamma clones isolated from a genomic library (Bimbo) made from a single individual. Polymorphisms in *Eco*RI (circled) distinguish the two alleles of this individual. * indicates the only three clones in which the location of *Hin*dIII sites was determined. (B) pSplH1-1 is a subclone containing the 5.2-kb *Bam*HI-*Eco*RI fragment of λSplH1-1 cloned into the vector pUC9. Abbreviations used for restriction endonucleases are: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; HII, *Hin*dII; S, *Sac*II; Ssp, *Ssp*I; X, *Xba*I; Hp, *Hpa*I; N, *Nco*I; P, *Pst*I; Sl, *Sal*I.

clones and a subclone, pSplH1-1, containing the 5.2-kb *Bam*HI-*Eco*RI fragment are shown in Fig. 1. The region of λSplH1-1 that is homologous to pJK-1 is located near the *Eco*RI site. We could not detect any other histone genes within the 28 kb of DNA surrounding this H1 gene. The nucleotide sequence of a 1,182-bp region of pSplH1-1 containing the entire coding region as well as the 5′ and 3′ flanking sequences was then determined (Fig. 2).

Analysis of H1-gamma protein. The nucleotide sequence contains an open reading frame of 216 amino acids. Comparison of the amino acid composition of the protein encoded by λSplH1-1 to the known amino acid compositions of *S. purpuratus*, H1-alpha, H1-beta, and H1-gamma (41) predicts that this gene encodes the H1-gamma subtype (Table 1). Based on the overall similarity of its primary structure to that of other H1 proteins, H1-gamma has the tripartite structure (nose, head, and tail) typical of this class of histones (7, 22). Comparison of the amino acid sequences of H1-alpha and H1-gamma reveals no homology in the nose regions; but the head regions (amino acids 17 to 85), which are believed to be globular domains, share a 67% homology. The 3′ tail region of H1-gamma is extremely basic, consisting almost exclusively of lysine, alanine, and proline. H1-gamma contains no serine residues in the C-terminal tail. The last 66 amino acids of H1-gamma are composed almost entirely of two repeated pentapeptides. The peptide PAAKK is repeated four times between amino acids 151 and 171, and the related peptide KPAKK is repeated five times from amino acid 177 to the C terminus (Fig. 3). The large number of lysine residues gives H1-gamma a net positive charge of 70.

Analysis of nucleic acid sequences. Inspection of the DNA sequence reveals two conserved sequences that define the 5′ and 3′ ends of most histone mRNAs (5, 25, 52). S1 nuclease mapping demonstrates the 5′ end of the H1-gamma mRNA to be coincident with a broad band spanning about 10 bp (data not shown). These bands span a degenerate form of a cap box consensus sequence (CATTAC) seen in many sea urchin histone genes. The 3′ end of the mRNA is assumed to be at the adenine residue at bp 1164, following a highly conserved sequence observed at the 3′ end of most histone mRNAs (5). The H1-gamma gene also contains another conserved sequence element (CAAGAAAGA) 6 bp downstream from the putative 3′ end of the mRNA which appears to be necessary for the binding of U7 small nuclear ribonucleoproteins involved in the processing of the 3′ end of the mRNA (5, 50). Interestingly, there is a degenerate form of the 3′ conserved hairpin sequence in the 3′-untranslated region (bp 1111 to 1122).

The late H1 5′ upstream region contains several conserved sequence motifs. The first of these is a TATA box at position −30. Further upstream are two H1-specific sequences. At −100 bp a G+C-rich 11-mer is found which shares 64% homology with a sequence termed upstream sequence element I (USE I) by Perry et al. (42) that has recently been shown to be an essential element of the chicken H1 promoter assayed in *Xenopus* oocytes and HeLa cells (55). Further upstream, another H1-specific sequence element, AAAC ACA (11), which is the core of the longer upstream sequence element II (USE II) (42) is seen at −137 bp. Interestingly, a shorter version of this sequence (AACAC) is repeated again at −147 bp and −158 bp, forming a duplication of this sequence on the same side of the DNA helix. There is no evidence for a sequence homologous to the H1-specific upstream sequence element III (42), but a hexanucleotide, ATTTTC, which is part of the USE II consensus sequence is repeated four times at positions −165, −208, −334, and −374.

Comparison of the nucleotide sequence of the H1-gamma gene coding sequences to the H1-alpha gene showed that they are extensively diverged. The changes in the nucleic acid sequence result in both extensive substitutions and silent third-base changes.

Genomic organization of late H1 gene. When the H1-

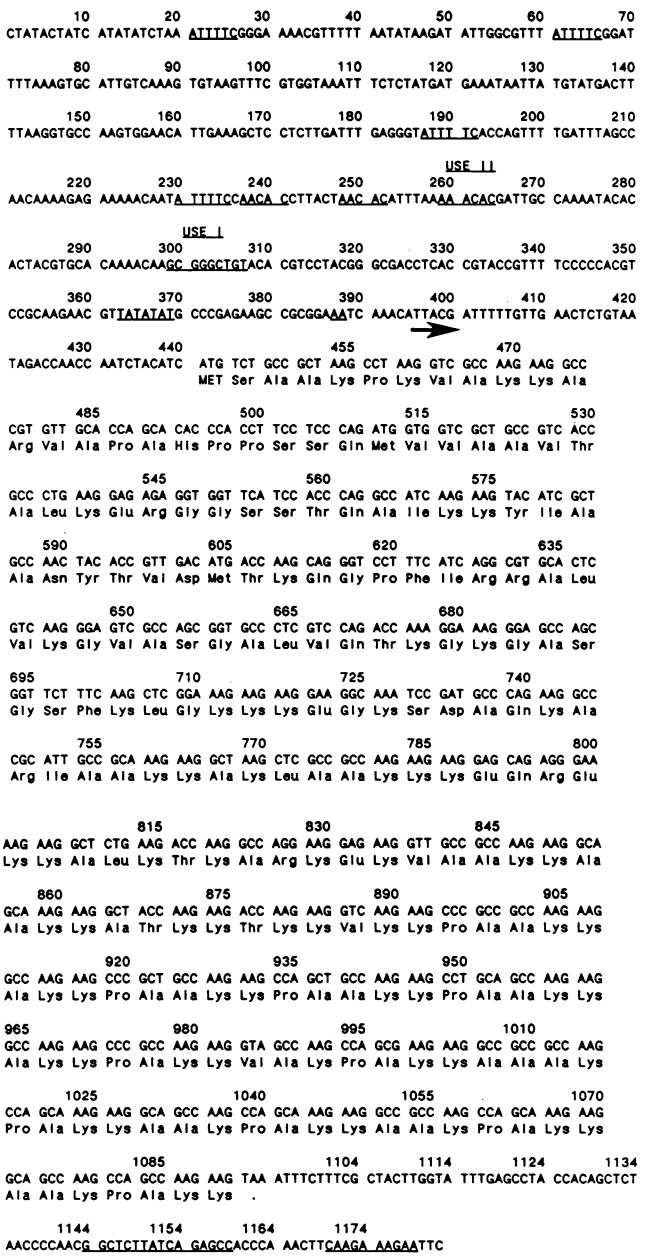

FIG. 2. DNA sequence of the *S. purpuratus* H1-gamma gene. The derived amino acid sequence is shown below the corresponding nucleotide sequence. Conserved DNA sequences are underlined and discussed in the text. The region around the mRNA initiation site is denoted by an arrow.

gamma gene is used to probe sperm DNA from two individual males, either one or two bands are visualized per restriction digest (Fig. 4). In the samples that have two bands we believe that each band represents alternate alleles of the H1-gamma gene and not cross-hybridization to the related H1-beta gene. The large degree of polymorphism seen between the two individuals is common in non-tandemly repeated genes in sea urchins (6, 8a). In fact, among the 10 clones selected from a genomic library made from sperm DNA of a single individual animal (Bimbo) are two classes distinguished by polymorphic *Eco*RI sites (Fig. 1). These clones represent the alternate alleles of this individual animal. Additional polymorphic restriction endonuclease sites are seen in λSplH1-1 which is derived from sperm DNA of a different animal. From the structure of the genomic locus, the intensity of the bands, and the conversion of a single fragment with one enzyme digestion to two fragments of equal intensity with a different enzyme, we believe that the H1-gamma gene is present in a single copy per haploid genome.

Expression of H1-gamma gene. If the H1-gamma gene is used to probe a Northern blot containing RNA from several developmental time points, a single RNA species of about 770 nucleotides that is expressed in a temporal pattern typical of the late gene family is detected (data not shown). Because H1-gamma and H1-beta mRNAs are extremely difficult to resolve owing to their similar size, this signal might represent both mRNA species. It is of some interest to determine the quantities of H1-alpha and H1-gamma mRNA throughout development. This is impossible to do by quantitating hybridization to a single Northern blot probed with both H1-gamma and H1-alpha. The intense H1-alpha mRNA signal in the early stages of development obscures the closely migrating H1-gamma mRNA signal. This problem can be circumvented by using RNA slot blots hybridized under conditions in which there is no cross-hybridization between early and late genes. Slots containing RNA from sequential developmental intervals were hybridized with an excess of Sp6 transcripts of equal specific activity encoding two late genes, H1-gamma and H2B-L1, and an early gene, H1-alpha (Fig. 5). By quantitating the autoradiographic intensity of each slot and comparing it with the H1-alpha mRNA signal in the egg, in which the number of H1-alpha mRNA molecules (80,000) has been accurately determined by RNA titration (32), it is possible to construct a plot of the number of mRNA molecules per embryo throughout early development (Fig. 6).

About 3,900 H1-gamma mRNA molecules and 26,000 H2B-L1 mRNA molecules are estimated to be stored in the unfertilized egg. The accumulation of additional late H1 and H2B transcripts by 9.5 h postfertilization indicates that these genes must have been activated shortly after the 16-cell stage, between 4.25 and 9.5 h after fertilization. H1-gamma mRNA accumulates at a rate of 4,300 molecules per min per embryo between 14 and 21 h postfertilization. H1-gamma mRNA concentration peaks at 21 h postfertilization when there are 2.0×10^6 molecules per embryo. The amount of H1-gamma mRNA then declines at a rate of 3,500 molecules

TABLE 1. *S. purpuratus* H1 protein compositions

Amino acid	H1-alpha[a]	H1-beta[a]	H1-gamma[a]	SplH1-1 gene
Ala	19.5	25.8	25.3	25.9
Asp/Asn	3.6	2.0	1.6	1.4
Glu/Gln	10.4	7.0	5.7	5.1
Phe	1.5	1.0	1.0	0.9
Gly	3.3	4.4	5.0	4.6
His	0.6	0.5	0.6	0.5
Ile	1.9	3.2	1.9	1.9
Lys	29.2	32.4	30.4	31.5
Leu	4.7	3.4	3.0	2.8
Met	1.0	0.5	1.0	0.9
Pro	6.1	5.0	6.9	6.9
Arg	2.2	3.1	3.3	3.2
Ser	4.8	4.1	4.1	4.2

[a] Protein compositions are taken from Pehrson and Cohen (41). Underlined are the values most consistent with SplH1-1 being identical with H1-gamma protein.

per min until 24 h postfertilization, followed by a much slower decrease of 210 molecules per min per embryo until 82 h. Recently, we have isolated an additional late H1 genomic clone whose sequence almost precisely matches the amino acid composition of the H1-beta protein (Lai and Childs, unpublished data). Gene-specific S1 nuclease probes were used to demonstrate that the temporal pattern of accumulation of H1-gamma and H1-beta are virtually identical. Lieber et al. (29) have shown that late H1 mRNAs are found in adult tissues of both *S. purpuratus* and *L. pictus*, and therefore late H1 is a somatic gene.

Late H2B mRNA accumulates at a relatively steady rate of 2,900 molecules per min per embryo between 4.25 and 25 h after fertilization. Late H2B mRNA accumulation peaks at 1.8×10^6 molecules per embryo at 24 h postfertilization and then declines throughout later development.

DISCUSSION

We cloned and characterized the gene encoding the H1-gamma subtype. Assignment of the H1 gene on λSplH1-1 to the gamma protein is based on three lines of evidence. (i) The open reading frame of the gene encodes a protein that has an amino acid composition almost identical to purified H1-gamma protein and different from H1-beta (Table 1). (ii) Both nucleic acid and amino acid sequences differ substantially from H1-alpha. This rules out the possibility that the gene encoded by λSplH1-1 is an H1-alpha orphon pseudogene (8a). (iii) The RNA encoded by this gene is the same size as late H1 mRNA and accumulates with the temporal pattern of late H1 mRNA (8, 24).

The open reading frame of λSplH1-1 indicates that H1-gamma is 216 amino acids in length and has a charge of +70. By comparison, H1-alpha has a length of 204 amino acids and a charge of +50 (29). Both the charge and length differences of these two proteins may account for the differing physical properties of blastula and gastrula chromatin. The increase in length of H1-gamma compared with H1-alpha may account for the increase in length of the nucleosome repeat that is seen throughout development in *S. purpuratus* (2). The more basic H1-gamma protein could neutralize more of the negative charge in the DNA backbone resulting in tighter binding. This could account for both the decreased rate of digestion of chromatin by DNase I as development proceeds (2) and the observation that H1-gamma is removed from chromatin preparations only by salt conditions that also remove the core histones (L. Cohen, personal communication).

Both H1-alpha and H1-gamma have a serine residue approximately 10 amino acids past the globular domain. Moving toward the carboxy terminus of the proteins, H1-alpha has four serine residues, while H1-gamma has none. Studies of H1 phosphorylation in *S. purpuratus* indicate that

153-Pro-Ala-Ala-Lys-Lys-Ala-Lys-Lys-
Pro-Ala-Ala-Lys-Lys-
Pro-Ala-Ala-Lys-Lys-
Pro-Ala-Ala-Lys-Lys-Ala-Lys-175

176-Lys-Pro-Ala-Lys-Lys-Val-Ala-
Lys-Pro-Ala-Lys-Lys-Ala-Ala-Ala-
Lys-Pro-Ala-Lys-Lys-Ala-Ala-
Lys-Pro-Ala-Lys-Lys-Ala-Ala-
Lys-Pro-Ala-Lys-Lys-Ala-Ala-
Lys-Pro-Ala-Lys-Lys.216

FIG. 3. Repeated pentapeptides in the C-terminal domain of H1-gamma. Amino acid residues 153 through 216 (the C terminus) are aligned to emphasize the repeated pentapeptides PAAKK and KPAKK. The repeated amino acids are underlined.

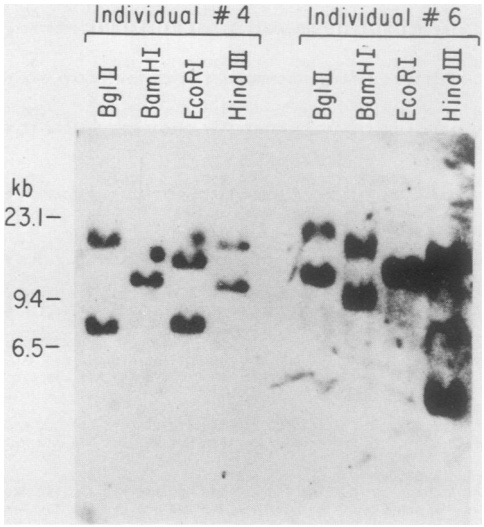

FIG. 4. Homology of restriction endonuclease-digested sperm DNA to the H1-gamma gene. Sperm DNA isolated from two individual animals was digested with the restriction endonucleases indicated; the digestion products were separated on a 0.6% agarose gel, blotted to nitrocellulose, and hybridized to H1-gamma as described in Materials and Methods.

H1-alpha is labeled by ^{32}P, while no label is incorporated into the H1-beta and H1-gamma subtypes (Cohen et al., J. Cell. Biol. **83**:171a, 1979). The C-terminal serine residues present in H1-alpha, but not H1-gamma, are therefore likely candidates for residues that are phosphorylated. Histone H1

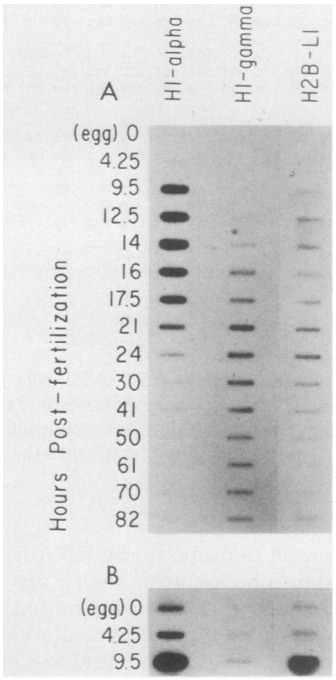

FIG. 5. Accumulation of H1-gamma mRNA during development. (A) RNA was prepared from embryos at various times after fertilization. This RNA was loaded onto nitrocellulose and probed with an Sp6 transcript complementary to either the H1-alpha, H1-gamma or H2B mRNA as described in the Materials and Methods. (B) A longer exposure of the three earliest time points in panel A.

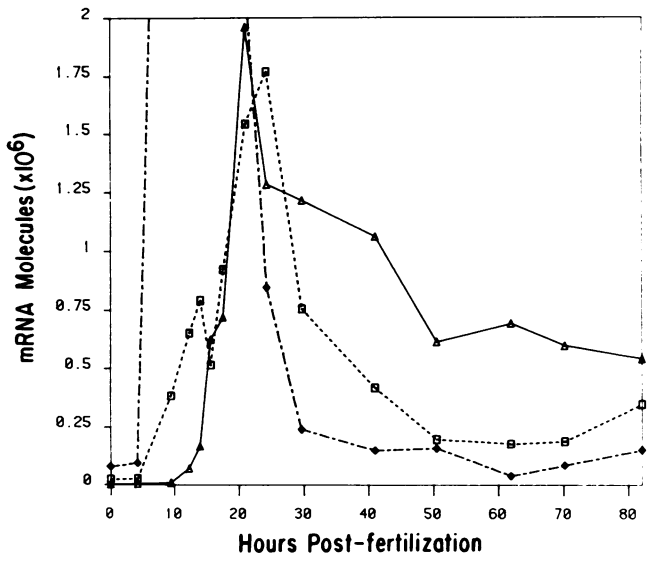

FIG. 6. Quantitative analysis of H1 and H2b mRNA during development. The relative amounts of early and late H1 and late H2b mRNA were determined by analysis of several autoradiographic exposures of the data in Fig. 5. Relative amounts of mRNA were converted to molecules of mRNA per embryo after taking into account the length of each probe as well as the number of [^{32}P]guanine residues in each probe. The data were then plotted as mRNA molecules per embryo versus time after fertilization. Symbols: ◇, early H1; △, late H1; □, late H2B.

proteins have at least two amino acid sequences that can serve as kinase substrates. Cyclic AMP-dependent kinase recognizes the sequence Lys/Arg-X-Ser (23), which is not found in H1-gamma but is present twice in H1-alpha. Histone kinase II recognizes the amino acid sequence Ser-X-Lys (44). This is part of a sequence (Gly-Ala-Ser-Gly-Ser-Phe-Lys) that is conserved in most histone H1 proteins including H1-gamma. H1-alpha contains this site and two additional sites for histone kinase II phosphorylation. This would imply that the Gly-Ala-Ser-Gly-Ser-Phe-Lys sequence found in the central domain of most H1 proteins may be buried in the center of the protein and therefore not available as a substrate in the native protein.

Based on the extensive changes in both third-base positions and amino acid substitutions, the early and late H1 gene families are at least as ancient as their H3 gene counterparts (200 million years) (9). The DNA sequence of the H1-gamma gene, however, contains several conserved sequence elements. Bounding the 5′ and 3′ ends of the transcribed regions are a degenerate form of the 5′ cap box seen in many histone genes (25, 52) and a 3′ conserved palindrome (5). S1 nuclease mapping confirms that the 5′ end of H1-gamma mRNA maps to a region of about 10 bp near the degenerate cap box, and the 3′ palindrome is assumed to represent the 3′ end of the mRNA. Based on these results, the H1-gamma gene encodes an mRNA of about 770 nucleotides. In comparison, the H1-alpha gene encodes an mRNA of about 707 nucleotides (28).

Upstream of the 5′ cap box are four sequences which might be important for the expression of the H1-gamma gene. The first sequence is a TATA box (−30) which is necessary for transcription and the choice of initiation sites on RNA polymerase II transcription units (12). A G+C-rich element (USE I) found upstream of most H1 genes (11, 42) is located at −100 bp. Another H1-specific sequence, AAACACA (USE II), is seen in its characteristic position (−137). What is somewhat novel is that a portion of this sequence, AACAC, is repeated twice more upstream with an 11-nucleotide phasing and could conceivably allow cooperative binding of a protein to the AACAC sequence. Another sequence that might be important for the expression of the H1-gamma gene is ATTTTC, which is seen four times upstream of H1-gamma. Wu and Crothers (54) have shown that double-stranded oligonucleotides containing the sequence CAAAAT have an abnormal gel mobility and thus might be kinked or bent. The ATTTTC sequence might have similar properties that could influence the binding of regulatory factors.

The pattern of expression of late H1 mRNA is similar to that of late H2B, a core histone, but differs in some important respects. Although the number of transcripts from both late genes peaks at about the same point in development and the number of H1-gamma and H2B-L1 transcripts are approximately equal at this time, there is a twofold excess of H1-gamma transcripts later in development.

It is likely that the relative amount of H1-gamma mRNA actually reflects the amount of H1-gamma protein made by the developing sea urchin. In *Strongylocentrotus drobachensis*, a close relative of *S. purpuratus*, and in *Arbacia punctulata*, the ratio of histone H1 to histone H2B rises during development (13a, 17a). In *Urechis caupo*, another marine invertebrate, and in the clam *Spisula solidissima*, the histone H1-to-core histone ratio rises as development proceeds, and this accumulation is due to an increase in the relative abundance of histone H1 mRNA (14–17). H1-gamma might be needed in greater than equimolar amounts to replace H1-alpha in chromatin. H1 proteins have been observed to exchange rapidly in chromatin (30). This would facilitate the replacement of H1-alpha by H1-gamma in chromatin.

Approximately 3,900 H1-gamma transcripts were detected in the unfertilized egg. This is 5% of the amount of H1-alpha mRNA in the egg (32). Previous efforts by Arceci et al. (3) to detect late H1 mRNA in the unfertilized egg by in vitro translation were unsuccessful. They estimated their detection sensitivity to be 1% of the amount of H1-alpha mRNA. Predictions of the secondary structure of the H1-gamma mRNA indicate a potential stem-loop structure ($\Delta G = -50$) near the start of the coding region (Fig. 7). Such a structure might inhibit the in vitro translation of H1-gamma mRNA. H1-alpha does not contain an analogous structure. The finding of late H1 mRNA, as well as mRNA for the late core histones (27), in the unfertilized egg indicates that the term switch is inappropriate in describing the molecular mechanism underlying the expression of these genes. In light of

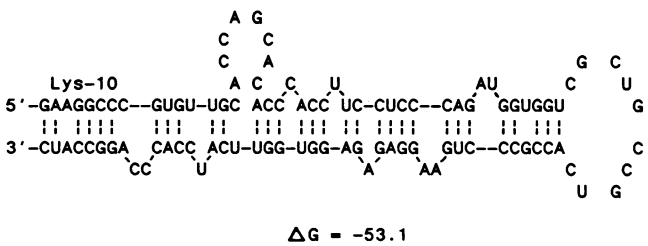

FIG. 7. Potential secondary structure at the 5′ end of H1-gamma mRNA. The region of H1-gamma mRNA with the highest predicted stability is shown. Lys-10 (codon AAG) is the lysine residue at position 10 in the predicted amino acid sequence of H1-gamma.

observation of late histone mRNA in the unfertilized egg, the term modulation might better describe the differential utilization of members of this multigene family.

ACKNOWLEDGMENTS

We thank Doug Engel for providing the chicken H1 clone, Bob Angerer for the Sp6 early H1 subclone, and Eric Davidson for providing the genomic libraries (A-B and Bimbo) of *S. purpuratus* DNA.

This work was supported by Public Health Service grant GM30333 from the National Institutes of Health. J.K. is supported by training grant T32GM7288 from the National Institutes of Health. Computer resources were provided by the BIONET National Computer Resource for Molecular Biology whose funding is provided by grant 1U41 RR-01685 from the National Institutes of Health.

LITERATURE CITED

1. **Angerer, L. M., D. V. DeLeon, R. C. Angerer, R. M. Showman, D. E. Wells, and R. A. Raff.** 1984. Delayed accumulation of maternal histone mRNA during sea urchin oogenesis. Dev. Biol. **101**:477–484.
2. **Arceci, R. J., and P. R. Gross.** 1980. Histone variants and chromatin structure during sea urchin development. Dev. Biol. **80**:186–209.
3. **Arceci, R. J., D. R. Senger, and P. R. Gross.** 1976. The programmed switch in lysine-rich histone synthesis at gastrulation. Cell **9**:171–178.
4. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. Science **196**:180–183.
5. **Birnstiel, M., M. Busslinger, and K. Strab.** 1985. Transcription termination and 3′ processing: the end is in site! Cell **41**: 349–359.
6. **Britten, R. J., A. Cetta, and E. Davidson.** 1978. The single-copy DNA sequence polymorphism of the sea urchin Strongylocentrotus purpuratus. Cell **15**:1175–1186.
7. **Bustin, M., and R. D. Cole.** 1968. Species and organ specificity in very lysine-rich histones. J. Biol. Chem. **243**:4500–4505.
8. **Childs, G., R. Maxson, and L. H. Kedes.** 1979. Histone gene expression during sea urchin embryogenesis: isolation and characterization of early and late messenger RNAs of Strongylocentrotus purpuratus. Dev. Biol. **73**:153–173.

8a. **Childs, G., R. E. Maxson, R. Cohn, and L. H. Kedes.** 1981. Orphons: dispersed genetic elements derived from tandem repetitive genes of eukaryotes. Cell **23**:651–663.

9. **Childs, G., C. Nocente-McGrath, T. Leiber, C. Holt, and J. Knowles.** 1982. Sea urchin (L. pictus) late stage H3 and H4 genes: characterization and mapping of a clustered but nontandemly linked family. Cell **31**:383–393.
10. **Cohen, L., K. Newrock, and A. Zweidler.** 1975. Stage specific switches in histone synthesis during embryogenesis of the sea urchin. Science **190**:994–997.
11. **Coles, L. S., and J. R. E. Wells.** 1985. An H1 histone gene-specific 5′ element and evolution of H1 and H5 genes. Nucleic Acids Res. **13**:585–594.
12. **Corden, J., B. Wasylyk, A. Buchwalder, P. Corsi, C. Kedinger, and P. Chambon.** 1980. Promoter sequences of eukaryotic protein coding genes. Science **209**:1406–1414.
13. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18 S rDNA. Plasmid **13**:31–40.

13a. **Easton, D., and R. Chalkley.** 1972. High-resolution electrophoretic analysis of the histones from embryos and sperm of *Arbatia punctulata*. Exp. Cell Res. **72**:502–508.

14. **Franks, R. R., and F. C. Davis.** 1983. Regulation of histone synthesis during early *Urechis caupo* (Echiura) development. Dev. Biol. **98**:101–109.
15. **Franks, R. R., and F. C. Davis.** 1985. Histone messenger RNA synthesis and accumulation during early development in the echiuroid worm, *Urechis caupo*. Dev. Biol. **109**:118–126.
16. **Gabrielli, F., and C. Baglioni.** 1975. Maternal messenger RNA and histone synthesis in embryos of the surf clam *Spisula solidissima*. Dev. Biol. **43**:254–263.
17. **Gabrielli, F., and C. Baglioni.** 1977. Regulation of maternal mRNA translation in developing embryos on the surf clam *Spisula solidissima*. Nature (London) **269**:529–531.

17a. **Gineitis, A. A., J. V. Stankeviciute, and V. I. Vorob'ev.** 1976. Chromatin proteins from normal, vegetalized, and anamalized sea urchin embryos. Dev. Biol. **52**:181–192.

18. **Grunstein, M.** 1978. Hatching in the sea urchin Lytechinus pictus is accompanied by a shift in histone H4 gene activity. Proc. Natl. Acad. Sci. USA **75**:4135–4139.
19. **Grunstein, M., P. Schedl, and L. H. Kedes.** 1976. Sequence analysis and evolution of sea urchin (Lytechinus pictus and Strongylocentrotus purpuratus) histone H4 messenger RNAs. J. Mol. Biol. **104**:351–369.
20. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25**:263–269.
21. **Harrison, M., and F. Wilt.** 1982. The program of H1 histone synthesis in *S. purpuratus* embryos and the control of its timing. J. Exp. Zool. **223**:255–256.
22. **Hartman, P. G., T. M. Chapman, and E. M. Bradbury.** 1977. Studies on the role and mode of operation of the very-lysine-rich histone H1 in eukaryote chromatin. Eur. J. Biochem. **77**:45–51.
23. **Hashimoto, E., M. Takeda, Y. Nishizuka, K. Hamana, and K. Jwai.** 1976. Studies on the sites in histones phosphorylated by adenosine 3′:5′-monophosphate-dependent and guanosine 3′:5′-monophosphate-dependent protein kinases. J. Biol. Chem. **251**: 6287–6293.
24. **Heiter, P. A., M. B. Hendricks, K. Hemminiki, and E. S. Weinberg.** 1979. Histone gene switch in the sea urchin embryo: identification of late embryonic histone mRNAs and the controls of their synthesis. Biochemistry **18**:2707–2716.
25. **Hentschel, C. C., and M. L. Birnstiel.** 1981. The organization and expression of histone gene families. Cell **25**:301–313.
26. **Kaumeyer, J. F., and E. S. Weinberg.** 1986. Sequence, organization and expression of late embryonic H3 and H4 histone genes from the sea urchin, Strongylocentrotus purpuratus. Nucleic Acids Res. **14**:4576–4577.
27. **Knowles, J. A., and G. C. Childs.** 1984. Temporal expression of late histone messenger RNA in the sea urchin Lytechinus pictus. Proc. Natl. Acad. Sci. USA **81**:2411–2415.
28. **Levy, S., I. Sures, and L. Kedes.** 1982. The nucleotide and amino acid coding sequence of a gene for H1 histone that interacts with euchromatin. The early embryonic H1 gene of the sea urchin *Strongylocentrotus purpuratus*. J. Biol. Chem. **257**:9438–9443.
29. **Lieber, T., K. Weisser, and G. Childs.** 1986. Analysis of histone gene expression in adult tissues of the sea urchins *Strongylocentrotus purpuratus* and *Lytechinus pictus*: tissue-specific expression of sperm-specific histone genes. Mol. Cell Biol. **6**:2602–2612.
30. **Leuters, O. L., and R. Chalkley.** 1985. Exchange of histones H1, H2A, and H2B *in vivo*. Biochemistry **24**:3080–3085.
31. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
32. **Mauron, A., L. Kedes, B. R. Hough-Evans, and E. H. Davidson.** 1982. Accumulation of individual histone mRNAs during embryogenesis of sea urchin *Strongylocentrotus purpuratus*. Dev. Biol. **94**:425–434.
33. **Maxson, R., T. Mohun, G. Gormezano, G. Childs, and L. Kedes.** 1983. Distinct organizations and patterns of expression of early and late gene sets in the sea urchin. Nature (London) **301**: 120–125.
34. **Maxson, R. E., T. J. Mohon, R. Cohn, and L. Kedes.** 1983. Expression and organization of histone genes. Annu. Rev. Genet. **17**:239–277.
35. **Maxson, R. E., and F. H. Wilt.** 1981. The rate of synthesis of histone mRNA during the development of sea urchin embryos (Strongylocentrotus purpuratus). Dev. Biol. **83**:380–386.
36. **Maxson, R. E., and F. H. Wilt.** 1982. Accumulation of the early

histone messenger RNAs during the development of *Strongylocentrotus purpuratus*. Dev. Biol. **94**:435–440.
37. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA probes from plasmids containing a bacteriophage Sp6 promoter. Nucleic Acids Res. **12**:7035–7056.
38. **Newrock, K. M., C. R. Alfageme, R. V. Nardi, and L. H. Cohen.** 1977. Histone changes during chromatin remodeling in embryogenesis. Cold Spring Harbor Symp. Quant. Biol. **92**: 421–431.
39. **Newrock, K. M., L. H. Cohen, M. B. Hendricks, R. J. Donnelley, 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.
40. **Peacock, S. L., C. M. McIver, and J. J. Monahan.** 1981. Transformation of *E. coli* using homopolymer-linked plasmid chimeras. Biochim. Biophys. Acta **655**:243–250.
41. **Pehrson, J. R., and L. H. Cohen.** 1984. Embryonal histone H1 subtypes of the sea urchin *Strongylocentrotus purpuratus*: purification, characterization, and immunological comparison with H1 subtypes of the adult. Biochemistry **23**:6761–6764.
42. **Perry, M., G. H. Thomsen, and R. G. Roeder.** 1985. Genomic organization and nucleotide sequence of two distinct histone gene clusters from Xenopus laevis: identification of novel conserved upstream sequence elements. J. Mol. Biol. **185**:479–499.
43. **Rigby, C. W., M. Dieckmann, C. Rhodes, and P. Berg.** 1977. Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. J. Mol. Biol. **113**:237–251.
44. **Romhanyi, T., J. Seprodi, F. Antoni, G. Meszaros, and A. Farago.** 1985. Specific substrate for histone kinase II: a synthetic nonpeptide. Biochim. Biophys. Acta **827**:144–149.
45. **Ruderman, J., and P. Gross.** 1974. Histones and histone synthesis in sea urchin development. Dev. Biol. **36**:286–298.
46. **Ruderman, J. V., C. Baglioni, and P. R. Gross.** 1974. Histone mRNA and histone synthesis during embryogenesis. Nature (London) **247**:36–38.
47. **Sanger, F. S., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
48. **Seale, R. L., and A. I. Aronson.** 1973. Chromatin-associated proteins of the developing sea urchin embryo. J. Mol. Biol. **75**:647–658.
49. **Strickland, W. N., M. Strickland, W. F. Brandt, C. Von Holt, A. Lehman, and B. Wittman-Liebild.** 1980. The primary structure of histone H1 from sperm of the sea urchin Parenchinus angulosus. Sequence of the C-terminal CNBr peptide and the entire primary structure. Eur. J. Biochem. **104**:567–578.
50. **Strub, K., G. Galli, M. Busslinger, and M. L. Birnstiel.** 1984. The cDNA sequences of the sea urchin U7 small nuclear RNA suggest specific contacts between histone mRNA precursor and U7 RNA during RNA processing. EMBO J. **3**:2801–2807.
51. **Sugarman, B. J., J. B. Dodgson, and J. D. Engel.** 1983. Genomic organization, DNA sequence, and expression of chicken embryonic histone genes. J. Biol. Chem. **258**:9005–9016.
52. **Sures, I., S. Levy, and L. H. Kedes.** 1980. Leader sequences of Strongylcentrotus purpuratus histone mRNAs start at a unique heptanucleotide common to all five histone genes. Proc. Natl. Acad. Sci. USA **77**:1215–1269.
53. **Weinberg, E. S., M. B. Hendricks, K. Hemminiki, P. E. Kuwabara, and L. E. Farrelly.** 1983. Timing and rates of synthesis of early histone mRNA in the embryo of Strongylocentrotus purpuratus. Dev. Biol. **98**:117–129.
54. **Wu, H., and D. M. Crothers.** 1984. The locus of sequence-directed and protein-induced DNA bending. Nature (London) **308**:509–513.
55. **Younghusband, H. B., R. Strum, and J. R. E. Wells.** 1986. Mutagenesis of conserved 5′ elements and transcription of a chicken H1 histone gene. Nucleic Acids Res. **14**:635–644.