Volume 14 Number 20 1986 Nucleic Acids Research

Comparison of the late H1 histone genes of the sea urchins *Lytechinus pictus* and *Strongelocentrotus purpuratus*

James A.Knowles and Geoffrey J.Childs*

Department of Genetics, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, NY 10461, USA

Received 24 July 1986; Revised and Accepted 11 September 1986

ABSTRACT

We have isolated and sequenced a gene encoding a late H1 histone subtype from the sea urchin species L. pictus. The primary structure of the late H1 subtype encoded by this gene is 209 amino acids in length, and has a net positive charge of 67. This gene is present in a single copy per haploid genome and encodes an mRNA of 752 nucleotides. Late H1 transcripts are detected in the unfertilized egg and are most prevalent in gastrulating embryos. Comparison of 375 bp of 5' flanking sequences of the L. pictus late H1 gene and the H1-gamma gene of a distantly related sea urchin species, S. purpuratus, reveals large blocks of sequences that are identical between the two genes. To determine if these conserved 5' sequences are present in other members of the sea urchin H1 gene family, the analogous region of S. purpuratus H1-alpha, an early H1 gene, was sequenced. The homology between the flanking sequences of the early and late families was limited to consensus sequences which are found upstream of all H1 genes. The possible regulatory implications of these findings are discussed.

INTRODUCTION

Sea urchin histones are encoded by several multigene families which are regulated in both a temporal and tissue specific fashion. The two best characterized families are those encoding the early and late histone subtypes. The early histone gene family contains clustered sequences encoding the alpha subtypes of all five histone proteins which are repeated in several hundred nearly identical tandem arrays (1-4). Lytechinus pictus contains three separately maintained tandem arrays of early genes, the A-D, B-C, and E subfamilies, which differ primarily in their spacer sequences (5-7). Transcripts of these early histone genes are found as stored mRNAs in the unfertilized egg, are synthesized during cleavage, and become most abundant in the early blastula stage of development (3, 4).

The genes encoding the late core histones have been cloned and are repeated only 8-10 times in the sea urchin genome. Unlike their early gene counterparts, the late core histone genes are not arranged in tandem arrays (8-10). The coding regions of the early and late H3 genes of L. pictus differ

by 19% in nucleotide sequence (11). Late gene transcripts for the core histones are also stored in eggs, but are 100-fold less abundant than early gene transcripts (12). The genes encoding the late core histones are activated in cleavage stage embryos 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 the peak of early histone mRNA accumulation. At least part of this accumulation can be accounted for by an increase in the relative rate of transcription of the late gene family at about the blastula stage (7-14 hrs. post fertilization) (12).

Recently, a gene encoding the S. purpuratus late H1 subtype, H1-gamma, has been isolated (13). Unlike its core histone gene counterparts, H1-gamma is not clustered with other late histone genes, and is present in only a single copy per haploid genome. The temporal pattern of expression of H1-gamma is similar to that of the late core histones. H1-gamma mRNA is stored in the unfertilized egg, and is most abundant during gastrulation.

This report details the cloning and sequencing of a late H1 gene from the distantly related species, L. pictus. A comparison of the upstream region of the two late H1 genes revealed striking homologies. To ascertain whether these conserved sequences are common to all members of the sea urchin H1 class of genes, the upstream region of the S. purpuratus early H1 genes was sequenced. The upstream region of the early gene shares little homology with upstream regions of either of the late genes. The significance of this observation to a model for the regulation of these genes is discussed.

MATERIALS AND METHODS

Preparation of RNA

RNA was prepared from L. pictus embryos as previously described (12).

Isolation of Genomic Clones

S. purpuratus late H1 cDNA was subcloned into mp18 (13), radiolabeled by primer extension (12), and used to screen several L. pictus genomic libraries (BamH1 and EcoR1) as described (14). Purified phage were grown in liquid culture and DNA was purified as described (15).

DNA Sequencing

A 3.5 kb XbaI-EcoRI fragment of λLplH1-1 was subcloned into M13mp18 and a series of nested deletions was constructed as described (16). To provide clones to sequence the opposite strand, restriction fragments of the 3.5 kb XbaI-EcoRI fragment were subloned into appropriately cut M13 vectors. For S. purpuratus early H1, an 800 bp SmaI-EcoRI fragment of pSp102 (17) was cloned

into M13mp18. M13 clones were sequenced by the dideoxy sequencing technique (18). DNA and protein sequences were analyzed using the BIONET National Computer Resource.

Filter Hybridizations

All hybridizations were done in 50% formamide, 5X SSC, 5X Denhardt's solution, 50mm sodium phosphate pH 6.8, and 0.1 mg/ml sonicated salmon sperm DNA (50% formamide hybridization buffer) (15). S. purpuratus late H1 cDNA (pJK-1) was hybridized to L. pictus genomic libraries at 37°C in this buffer, and the filters were washed at 65°C in 0.5X SSC, 0.1% SDS, 1% sodium pyrophosphate (0.5X wash buffer). Genomic DNA hybridizations were carried out at 42°C with the addition of 10% dextran sulfate. A nick translated (19) SspI-EcoRI fragment was used as a probe (figure 1). For northern blots, an SspI-EcoRI fragment was subcloned into pGem2 and the resulting plasmid was linearized with Hind III yielding a 1.2 kb template. The Sp6 probe was made as described (20) and hybridized to the blot at 58°C in 50% formamide hybridization buffer. The blot was then washed at 65°C in 0.5X wash buffer. Autoradiographic intensity was quantitated using a Cambridge Instruments Quantimat 920 Image Analysis System.

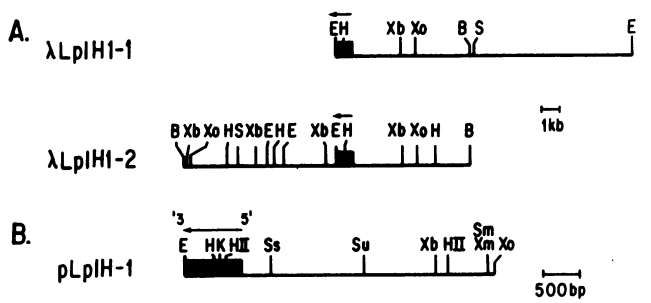

Figure 1. Restriction endonuclease maps of DNA segments containing an L. pictus late H1 gene. Bars denote histone H1 coding sequences and arrows the direction of transcription. Each map is aligned such that the left and right arms of phage DNA are on the left and right of the figure, respectively. Maps were generated from examination of single and double enzyme digestions of phage and plasmid DNA. A.) Two overlapping lambda clones containing a late H1 gene are shown. λLP1H1-1 was isolated from an EcoRI genomic library (generously provided by Dr. Eric Davidson) while λLP1H1-2 was isolated from a BamH1 library (8). The map for λLp1H1-1 contains only those sites detected by hybridization of pJK-1 to appropriately digested λLp1H1-1 DNA. B.) pLp1H1-1 is a subclone containing the 3.5 Xho-EcoRI fragment of λLp1H1-1 cloned into the vector pUC9. The SspI site was determined from the DNA sequence. Abbreviations used for restriction endonucleases are, B-BamHI, E-EcoRI, H-HindIII, HII-HindII, K-KpnI, S-SacI, Sm-SmaI, Ss-SspI, Su-StuI, Xb-XbaI, Xm-XmnI, Xo-XhoI.

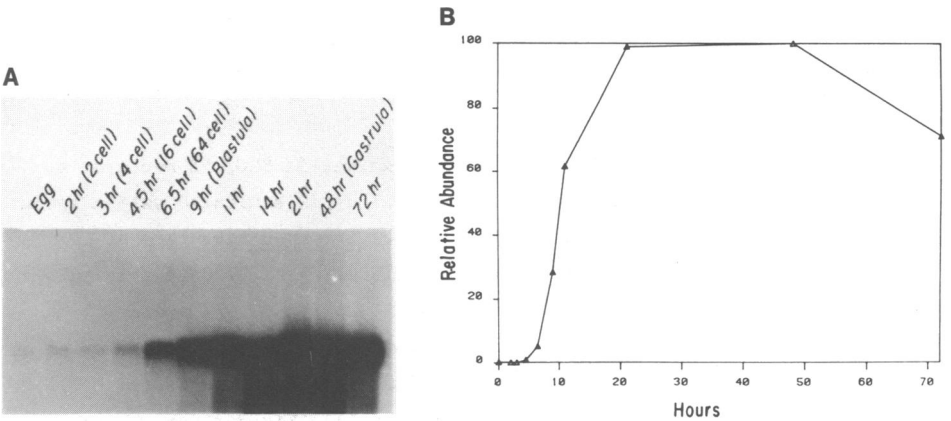

Figure 2. Accumulation of late H1 mRNA during development. A.) RNA was prepared from embryos at various times after fertilization. This RNA was fractionated on a 2% agarose gel containing formaldehyde, blotted to nitrocellulose, and probed with an Sp6 transcript complementary to the late H1 mRNA. B.) Relative autoradiographic densities were determined from this and longer and shorter exposures of this blot and then plotted.

RESULTS and DISCUSSION

Isolation of a Late H1 gene

A pJK-1 cDNA clone encoding a S. purpuratus late H1 protein (13), was used to screen several L. pictus genomic libraries. Several highly homologous clones were isolated. Restriction endonuclease maps were generated for two of these clones (figure 1), and each appear to be overlapping DNA segments encoding the same H1 gene. Neither clone is homologous to any of the core histone genes (data not shown). Therefore, the L. pictus late H1 gene, like the S. purpuratus H1-gamma gene (13), is not tightly clustered with other histone genes. In order to further localize the region homologous to pJK-1 a 3.5 Xho-EcoRI fragment of λLp1H1-1 was subcloned and mapped in more detail (figure 1).

Temporal Expression of Late H1 mRNA

When the H1 gene encoded by λLp1H1-1 is used to probe a northern blot containing L. pictus RNA from several developmental time points, a single RNA of about 750 nucleotides that accumulates in a temporal pattern typical of the late histone gene family (13,21-23) is observed (figure 2). Like the core late histone mRNAs (12) and S. purpuratus H1-gamma mRNA (13), the L. pictus late H1 mRNA is detected in the egg as well as all subsequent developmental stages. Longer exposures clearly show that the H1 mRNA detected by this probe

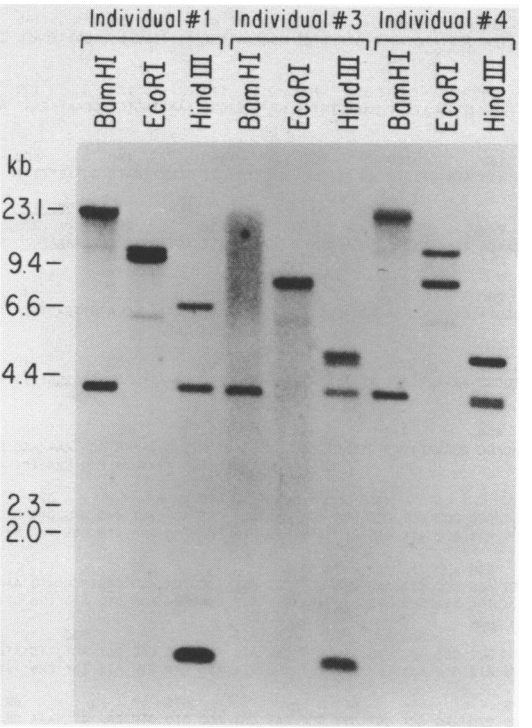

Figure 3. Genomic organization of L. pictus late H1 genes. Sperm DNA from three individuals was digested with the restriction endonucleases indicated; the digestion products were separated on a 0.6% agarose gel, blotted to nitrocellulose, and hybridized to late H1 as described in Materials and Methods.

is the same size as late H1 (data not shown). Therefore, this gene encodes a somatic H1 protein (either H1-gamma or H1-beta) that is also expressed whenever the embryonic early H1 gene is active.

Genomic Organization of the Late H1 gene

The late H1 gene was used to probe genomic DNA by Southern blotting (figure 3). One or two bands were detected with restriction endonucleases that do not cleave within the H1 gene, while three or four fragments were detected when genomic DNA is cut with Hind III, (an enzyme that cleaves within the H1 gene). In the case where three bands are detected (individuals #1 and #4), one of the bands is twice as intense as the other two. This result is most consistent with there being only a single H1-gamma gene. HindIII cleaves within H1-gamma (Figure 1) normally yielding two bands of equal intensity.

```
        10         20         30         40         50         60         70
TGTTCTCTCA TGTTATCTAT CTGTTTTTCT GGAAAATATT ATTTTACACA AGATTTCAGC GTTTGTTTTT

        80         90        100        110        120        130        140
GGATTTTAAA GTGCCATTGT CAAAGTGTAA GTTTCGAGGT AAATTACTCT CCAATAGAAT AACCATGGTT

       150        160        170        180        190        200        210
ATTTTTTAAG ATGTCAAGTG AAAGATTGAA AGCTCCTCTT TCGTTTGATG GGTATTTTTTC TCGTTACCGT

                                                            USE II
       220        230        240        250        260        270        280
AGCACGAACA AGAAAGAAAA ACGATATATT CCAACACCTT ACTAACACAT CCAAAAACAC GATTGCCAAA

                          USE I
       290        300        310        320        330        340        350
ATACACGGTA CATGTACAAA ACAAGCGGGC TGTACACACC CTATAGGCGA CTTGACTGCA TTGTTTTCCC

       360        370        380        390        400        410        420
CCACGTCCGC AAAAAGGCGA TATATCTATG GGATAGCGCG GAAATTGAAC ATTACGTTTC TCGCTCAGTG
                                                        ➔
       430        440        450                465
ACAACACAGC CGGAATCTAT CTCACTCACC  ATG TCT GCC GCC AAG CCA AAA ACC GCA
                                  MET Ser Ala Ala Lys Pro Lys Thr Ala

480                495                510                525
AAG AAG GCC CGT GCT GCC CCA GCA CAC CCA CCT ACC TCT CAG ATG GTA GTT GCT
Lys Lys Ala Arg Ala Ala Pro Ala His Pro Pro Thr Ser Gln Met Val Val Ala

       540                555                570                585
GCT ATC ACC GCC CTG AAG GAG CGT GGT GGT TCC TCA AAC CAG GCC ATC AAG AAG
Ala Ile Thr Ala Leu Lys Glu Arg Gly Gly Ser Ser Asn Gln Ala Ile Lys Lys

                600                615                630
TAC ATC GCT GCC AAC TAC AAG GTT GAC ATC AAC AAG CAG GCT ACT TTC ATC AAG
Tyr Ile Ala Ala Asn Tyr Lys Val Asp Ile Asn Lys Gln Ala Thr Phe Ile Lys

    645                660                675                690
CGT GCC CTG AAG GCT GGT GTT GCC AAT GGT ACC CTC GTC CAA GTC AAA GGA AAG
Arg Ala Leu Lys Ala Gly Val Ala Asn Gly Thr Leu Val Gln Val Lys Gly Lys

            705                720                735
GGA GCC AGT GGA TCT TTC AAG CTC GGC AAG GTC AAG GCT GGC AAG ACC GAG GCC
Gly Ala Ser Gly Ser Phe Lys Leu Gly Lys Val Lys Ala Gly Lys Thr Glu Ala

750                765                780                795
CAG AAG GCC CGT GCT GCC GCC AAG AAG GCC AAG CTT GCT GCC AAG AAG AAG GAA
Gln Lys Ala Arg Ala Ala Ala Lys Lys Ala Lys Leu Ala Ala Lys Lys Lys Glu

       810                825                840                855
CAG AAG GAG AAG AAG GCT GCT AAG ACC AAG GCC AGG AAG GAG AAA CTA GCC GCC
Gln Lys Glu Lys Lys Ala Ala Lys Thr Lys Ala Arg Lys Glu Lys Leu Ala Ala

                870                885                900
AAG AAG GCT GCA AAG AAG GCC GCC AAG AAG GTT AAG AAG CCC GCC GCC AAG GCC
Lys Lys Ala Ala Lys Lys Ala Ala Lys Lys Val Lys Lys Pro Ala Ala Lys Ala

    915                930                945                960
AAG AAG CCA GCT AAG AAG GCA GCC AAG AAG CCC GCC GCC AAG AAG GCA GCC AAG
Lys Lys Pro Ala Lys Lys Ala Ala Lys Lys Pro Ala Ala Lys Lys Ala Ala Lys

            975                990               1005
AAG CCC GCC GCT AAG AAG CCA GCT AAG AAG GCC GCC AAG AAG CCT GCT GCT AAG
Lys Pro Ala Ala Lys Lys Pro Ala Lys Lys Ala Ala Lys Lys Pro Ala Ala Lys

1020               1035               1050               1065
AAG GCC GCC AAG CCA GCA AAG AAG GCA GCC AAG AAG CCC GCC GCC AAG AAG GCA
Lys Ala Ala Lys Pro Ala Lys Lys Ala Ala Lys Lys Pro Ala Ala Lys Lys Ala

       1080          1093       1103       1113       1123       1133
GCC AAG AAG TAA ATTCTTAGCG CCACTTGGTG TATTGAGCTT TTTCAGCTCC ACCCCAACGG
Ala Lys Lys  .

      1143       1153       1163
CTCTTATCAG AGCCACCCAA ACTTCAAGAA AGAATTC
```

However, due to the large amount of genomic polymorphism (4% sequence difference) seen between individual sea urchins (24,8), the two bands of lower intensity likely result from polymorphism in the flanking HindIII sites. Consistent with this is the observation that each of the three individuals contain many bands that differ in size. If the H1-gamma gene were present in more than one or two copies per haploid genome, polymorphism would have surely resulted in the generation of a greater number of hybridizing bands as is the case with the late core histone genes (8,9). The faint bands on the autoradiogram are most likely due to cross-hybridization with an H1 gene encoding a different late subtype.

Nucleotide Sequence of the L. pictus Late H1 Gene

The entire late H1 coding region and 400 bp of 5' flanking sequence of λLplH1-1 was then sequenced (figure 4). The nucleotide sequence contains an open reading frame of 209 amino acids. Based on the overall similarity of its primary structure to that of other H1 proteins, this H1 protein probably has the tripartite structure (nose, head, and tail) typical of this class of histones (25-26). The N-terminus of the protein is a variable region that is very basic and rich in alanine and proline. The central domain is the most highly conserved region of H1 proteins and has an amino acid distribution typical of most globulin proteins while the C-terminus is extremely variable in size and sequence but is highly basic. Comparison of the L. pictus late H1 protein to H1-gamma of S. purpuratus indicates that the two proteins are only 75% homologous. Although they differ significantly in primary sequence, a plot of the hydrophobicity of the two proteins is nearly identical (figure 5). The two late H1 proteins therefore contain a large number of conservative amino acid changes. Like H1-gamma (13), the L. pictus late H1 C-terminal tail does not contain serine and is composed of repeating amino acid sequences. The octapeptide, AKKAAKKP, is repeated five times, linked together by pentapeptides similar to those of S. purpuratus H1-gamma. The large number of lysine residues gives this protein a net positive charge of 67.

Examination of the DNA sequences bordering the open reading frame reveals several conserved sequences that define the 5' and 3' ends of most histone mRNAs. The 5' end of the mRNA is probably coincident with bp 401, which lies in the middle of a degenerate form of a conserved cap box sequence (CATTAC) seen in

Figure 4. DNA sequence of the L. pictus late H1 gene. The derived amino acid sequence is shown below the corresponding nucleotide sequence. Conserved DNA sequences are underlined and discussed in the text. The mRNA cap site is denoted by an arrow.

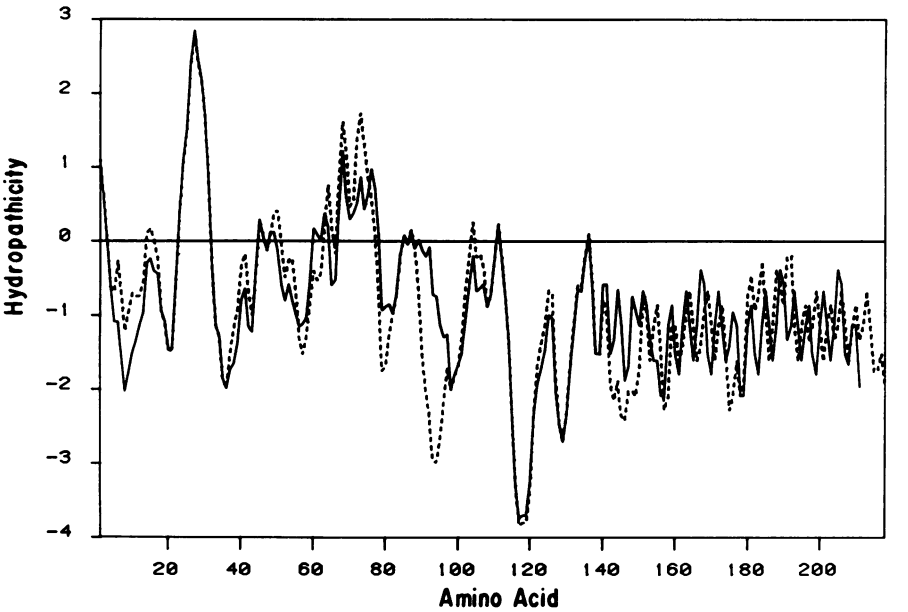

Figure 5. Comparison of the hydrophobicities of the protein encoded by the L. pictus late H1 gene to S. purpuratus H1-gamma. Hydrophobicities were calculated using the parameters of Kyte and Doolittle (38) and plotted using a Textronics graphics plotter.

many sea urchin histone genes (27,4). The 3' end of the non-polyadenylated mRNA is assumed to be at the adenine residue at bp 1152, following a highly conserved region, capable of forming a hairpin loop, observed at the 3' end of many sea urchin histone mRNAs (4). Therefore, this late H1 gene probably encodes a transcript of 752 nucleotides. Downstream of the mRNA is another conserved element, CAAGAAAGA, which appears to be necessary for the binding of U7 snRNPs involved in the processing of the 3' end of the mRNA (28).

Comparison of the nucleotide sequence of the L. pictus late H1 gene to the S. purpuratus H1-gamma gene (13), reveals that the two genes are 80% homologous in their coding regions. The nucleotide sequence homology comparable to the amino acid sequence conservation reflects a bias in codon usage in the genes of these two species. When the 5' upstream regions (to -375) of these two genes are compared they are 77% homologous over the entire region, with long sequences of near 100% homology (figure 6). This is quite suprising as these two sea urchin species diverged from a common ancestor more than 65 million years ago (29). The H1-specific upstream

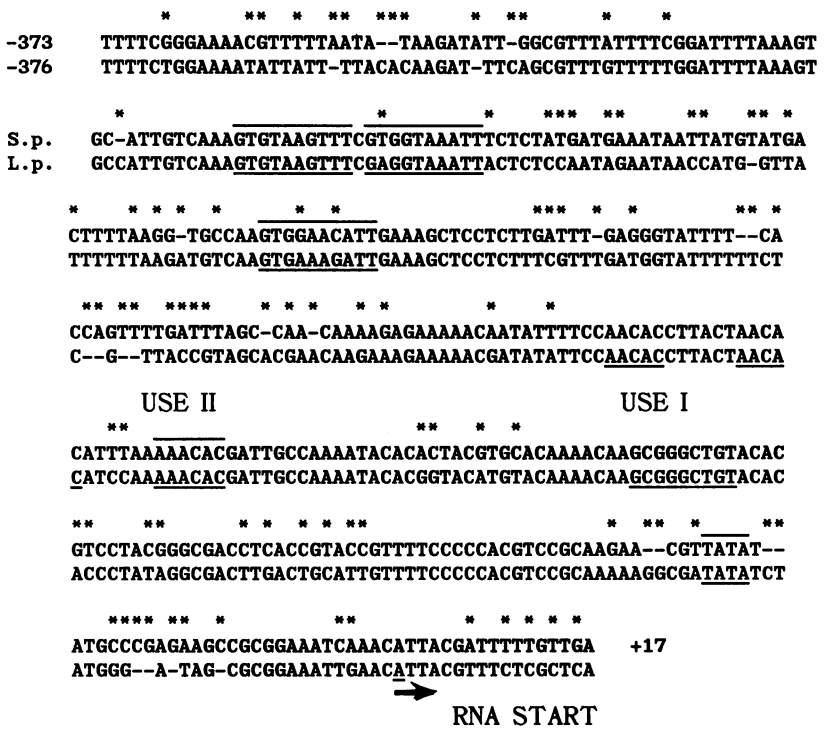

Figure 6. 5' upstream homology of the L. pictus late H1 gene and the S. purpuratus H1-gamma gene. The underlined sequences are discussed in the text. The top sequence is that of the S. purpuratus H1-gamma gene and the bottom the L. pictus late H1. (*) denote mismatched bases between the two sequences and an arrow denotes the mRNA start site for the S. purpuratus gene determined by S1 nuclease analysis (13).

sequence elements (USE elements) I and II (30,31) that are found in the S. purpuratus H1-gamma gene (13) are 100% conserved in the L. pictus late H1 gene. A core sequence of element II (AACAC) is repeated upstream twice more, within a 62 bp conserved region, with the same phasing in both late genes. Further upstream in a region of high homology between the two genes are three copies per gene of a 10 nucleotide consensus sequence, GTGGAAGATT. This sequence is similar to the core consensus sequence (GGTGTGGAAAG) derived from several enhancer elements (32). This is quite different from what is known about 2 non-allelic pairs of late H3 and H4 genes from L. pictus. When the sequences of these genes are compared the coding regions are 96% homologous, but the 5' flanking sequences are less than 40% homologous, with the exception of short USE homologies (11). The large degree of sequence conservation

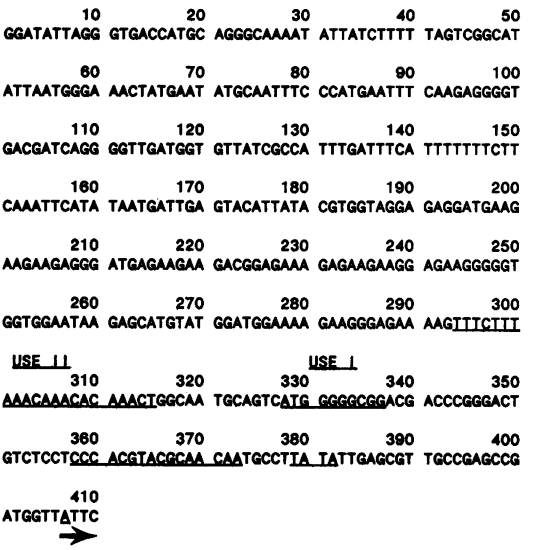

Figure 7. DNA sequence of the 5' flanking region of the S. purpuratus H1-alpha gene. The underlined sequences are discussed in the text. The DNA sequence from bp 345 to bp 410 was determined by Sures et al (27).

upstream of the late H1 genes indicates that some of these sequences must be crucial to the functioning of these genes.

To determine whether these striking homologies were a common feature of sea urchin histone H1 genes or specific to the late genes, the 5' upstream region of an H1-alpha (early) gene was sequenced (figure 7). The upstream sequence element II of the early gene contains only one copy of the AACAC sequence, in the middle of a complete USE II. This USE II is 52% homologous to the consensus sequence of Perry et al (TTTTGAGACTCTAGAAACACAGACTG) (31). The upstream region of this gene also has an upstream sequence element I that is similar to the consensus sequence (61%) (ATGGGCGGGGT) (31). Both the USEs of this gene are 100% homologous to the shorter consensus sequence of Coles and Wells (USE I-GGGCGG; USE II-AAACACA) (30). Within the USE I there is a potential Sp1 transcription factor binding site (GGGCGG) (33), which is not present in the USE I of the late H1 genes. One additional sequence that is shared between the early and late H1 genes is a 16-mer (consensus; CCCACGTNCGCAANAA) centered at -45 (between the TATA box and USE I) that is 88% conserved between the early and late H1 genes. This appears to be a sea urchin specific sequence. The upstream sequence of the early H1 contains no other regions of homology to the two late H1 genes, but does contain two long

purine rich regions. The first of these is just upstream of USE II and consists of 19 consecutive purines (bp 275-293). Further upstream, there is a region that is 63 nucleotides in length that contains 95% purines (bp 187-249).

In one model for the differential temporal expression of multigene families during development, different members of the gene family compete for limiting amounts of shared positive transcription factors (34). As development proceeds, cell division exponentially increases the number of genes in the embryo, but the quantity of transcription factor remains constant or declines. The genes best able to compete for the limiting quantities of transcription factors are able to form stable transcription complexes and therefore continue to be expressed later in development. The Xenopus 5S gene system is the best documental example of this type of regulation (34).

If one aspect of the early/late histone gene switch entails competition for common transcription factors (34), then either a greater number of binding domains (35), or binding domains of higher affinity for regulatory molecules would be predicted to occur within the late gene consensus sequences. The USE I and II regions or the 16-mer centered at position -45 are logical candidates for binding domains of common regulatory proteins. Consistent with this hypothesis, the late H1 genes contain three copies of a sequence, AACAC, in USE II, whereas the early H1 gene contains only one copy of this sequence. Competition for a single shared transcription factor goes a long way to explaining the mechanism of the late gene switch, however, several aspects of the early to late histone gene transition during embryogenesis are inconsistent with a simple model where one positive transcription factor becomes limiting during embryogenesis. The most salient observation that is inconsistent with this model is the a 5-6 fold increase in the relative rate of transcription of the late genes between 7 and 14 hours post-fertilization (12). Two gene families competing for a gene specific transcription factor, as in the case of the 5S genes (34), would be either transcribed or silent but the rate of transcription would not be modulated. It is reasonable to assume that the late genes contain at least one sequence not present in the early genes that serves either as a site for a negative regulator that limits the transcription rate until early blastula or for an additional late gene specific positive transcription factor appearing at early blastula (7-14 hrs. post-fertilization). The three repeated copies of the late gene specific consensus sequence GTGGAAGATT, or any of the other homologies shared between the L. pictus and S. purpuratus late H1 genes represent candidates for such

cis-acting sequences. Site directed mutagenesis of these conserved DNA sequences followed by assays in sea urchin in vitro transcription systems (36) or transient expression in embryos following injection into sea urchin eggs (37) should identify these cis acting regulatory sites.

ACKNOWLEDGEMENTS

We would like to thank Eric Davidson for providing the EcoRI genomic library of L. pictus DNA and to Ken Krauter for helpful comments. This work was supported by grant #GM30333 from the NIH. Jim Knowles is supported by a predoctoral M.D./Ph.D. training grant NIH grant # T32GM7288. Computer resources provided by the BIONET National Computer Resource for Molecular Biology whose funding is provided by NIH grant # 1U41 RR-01685.

*To whom correspondence should be addressed

REFERENCES

1. Cohn, R.H., Lowry, J.C., and Kedes, L.H. (1976). Cell 9, 147-161.
2. Schaffner, W., Gross, K., Telford, J., and Birnstiel, M. (1976). Cell 8, 471-478.
3. Kedes, L.H. (1979). Ann. Rev. Biochem. 48, 837-870.
4. Hentschel, C.C. and Birnstiel, M.L. (1981). Cell 25, 301-313.
5. Cohn, R.H. and Kedes, L.H. (1979). Cell 18, 843-853.
6. Cohn, R.H. and Kedes, L.H. (1979). Cell 18, 855-864.
7. Holt, C.A. and Childs, G. (1984). Nucleic Acids Res. 12, 6455-6471.
8. Childs, G., Nocente-McGrath, C., Leiber, T., Holt C., and Knowles, J. (1982). Cell 31, 383-393.
9. Maxson, R., Mohun, T., Gromezano, G., Childs, G., Kedes, L. (1983). Nature 301, 120-125.
10. Kaumeyer, J.F. and Weinberg, E.S. (1986). Nucleic Acids Res. 14, 4557-4576.
11. Roberts, S.B., Weisser, K.E., and Childs, G. (1984). J. Mol. Biol. 174, 647-662.
12. Knowles, J.A. and Childs, G. (1984). Proc. Nat. Acad. Sci. 81, 2411-2415
13. Knowles, J.A., Lai, Z.C. and Childs, G. (1986). Mol. & Cell Biol., submitted.
14. Benton, W.D. and Davis, R.W. (1977). Science 196, 180-183.
15. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). "Molecular Cloning: A Laboratory Manual." Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
16. Dale, R.M.K., McClure, B.A., and Houchins, J.P. (1985). Plasmid 5, 31-40.
17. Sures, I., Lowery, J., and Kedes, L.H. (1978). Cell 15, 1033-1044.
18. Sanger, F.S., Nicklen, S., and Coulson, A.R. (1977). Proc. Nat. Acad. Sci. 74, 5463-5467.
19. Rigby, C.W., Dieckmann, M., Rhodes, C., and Berg, P. (1977). J. Mol. Biol. 113, 237-251.
20. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K., and Green, M.R. (1984). Nucleic Acids Res. 12, 7035-7056.
21. Ruderman, J.V. and Gross, P.R. (1974). Dev. Biol. 36, 286-298.
22. Cohen, L.H., Newrock, K.M. and Zweidler, A. (1975). Science 190, 994-997.
23. Childs, G., Maxson, R. and Kedes, L.H. (1979). Dev. Biol. 73, 153-173.
24. Britten, R., Etton, A., and Davidson, E. (1978). Cell 15, 1175-1186.

25. Bustin, M. and Cole, R.D. (1968). J. Biol. Chem. 243, 4500-4505.
26. Hartman, P.G., Chapman, T.M., and Bradbury, E.M. (1977). Eur. J. Biochem. 77, 45-51.
27. Sures, I., Levy, S., and Kedes, L.H. (1980). Proc. Nat. Acad. Sci. 77, 1215-1269.
28. Strub, K., Galli, G., Busslinger, M., and Birnstiel, M.L. (1984). EMBO Journal 3, 2801-2807.
29. Smith, A.B. (1981). Paleontology 24, 779-801.
30. Coles, L.S. and Wells, J.R.E. (1985). Nucleic Acids Res. 13, 585-594.
31. Perry, M., Thomsen, G.H. and Roeder, R.G. (1985). J. Mol. Biol. 185, 479-499.
32. Khoury, G., and Gruss, P. (1983). Cell 33, 313-314.
33. Dynan, W.S. and Tjian, R. (1985). Nature 316, 774-778.
34. Brown, D.D. (1984). Cell 37, 359-368.
35. Reeder, R.H. and Roan, J.A. (1984). 38, 39-44.
36. Morris, G.F., Price, D.H. and Marzluff, W.F. (1986). Proc. Nat. Acad. Sci. 83, 3674-3678.
37. McMahon, A.P., Novak, R.T., Britten, R.J. and Davidson, E.S. (1984). Proc. Nat. Acad. Sci. 81, 7490-7494.
38. Kyte, J. and Doolittle, R.F. (1982). J. Mol. Biol. 157, 105-132.