Sequence, organization and expression of late embryonic H3 and H4 histone genes from the sea urchin, *Strongylocentrotus purpuratus*

John F.Kaumeyer* and Eric S.Weinberg⁺

Department of Biology, University of Pennsylvania, Philadelphia, PA 19104, USA

Received 20 December 1985; Revised and Accepted 7 May 1986

ABSTRACT

A gene pair coding for histones H3 and H4 expressed during late embryonic development has been cloned from the *S. purpuratus* genome. The organization of these genes is similar to the divergently transcribed H3-H4 gene pairs of *Lytechinus pictus*. Whole genome Southern analysis indicates that most late H3 and H4 genes are organized as pairs in both sea urchin genomes. The nucleotide sequences of late and early *S. purpuratus* H3 and H4 genes differ in the coding regions by 17.0% and 15.7%, respectively. Although there is little match between early and late genes outside the transcription unit, there are short stretches of homology in the spacers 5' to the *S. purpuratus* early and late H3 genes and the early and late H4 genes. Sequence comparison of the late H3-H4 *S. purpuratus* pair with two late H3-H4 pairs from *L. pictus* reveals additional striking homologies in the intergenic spacer. At least four late H3 and two H4 RNAs are distinguished by hybridization to Northern electrophoretograms with probes derived from the cloned gene pair. Although most of these RNAs appeared to accumulate coordinately, with the most dramatic increase occurring between 13 and 17 hours of development, some differences in timing of appearance of different H3 RNA species are observed.

INTRODUCTION

The histone genes of the sea urchin are expressed during embryogenesis in a highly regulated pattern (reviewed most recently in 1). At least two distinct gene sets are transcribed in the embryo. The major portion of histone mRNA synthesized in the embryo prior to early blastula is transcribed from a set of highly repeated, tandemly organized, "early" genes. Between the 64-cell stage and early blastula the early histone mRNAs increase approximately ten-fold over the amount stored in the unfertilized egg (2-4). These genes are the templates for 10-15% of the total mRNA synthesis during this period (5), but become transcriptionally inactive as embryogenesis proceeds (4, 7, 8). In the mid-blastula another set of genes, termed "late" histone genes, becomes the source of the bulk of histone mRNA synthesis, and by gastrula the late transcripts account for almost all the histone mRNA in the embryo (8-16). Some late mRNA species

have been detected in small amounts in the egg and increase slowly between the 64-cell stage and early blastula, but the period of rapid accumulation occurs during the mid to late blastula, resulting in a several hundred-fold increase during embryogenesis (8, 15-17). The two gene sets are therefore independently regulated to provide a "switch" in the types of histone mRNA present in the early and late embryo.

Thus far, late H3 and H4 genes have been cloned from the genome of the sea urchin, *L. pictus* (18, 19) and late H2A, H2B, and H4 genes cloned from *S. purpuratus* DNA (14). These clones have been used as probes to show that there are 5-10 copies per genome for these late genes and that their organization is very different from the early genes. Although the late H3 and H4 genes in the two *L. pictus* clones were found to be organized as gene pairs (18, 19), this arrangement was not found in the *S. purpuratus* clones. One H4 sequence was isolated as a single gene and another was found linked to an H2B gene (14). We have isolated a number of H3 and H4 genes from *S. purpuratus* and report here the organization and sequence of an H3-H4 pair arranged similarly to the H3-H4 gene pairs of *L. pictus*. We present evidence that this is the most common arrangement of H3 and H4 genes in *S. purpuratus*. Our sequence data allows us to compare for the first time the sequence divergence of both late and early genes for a single histone from two sea urchin species. The organization of late H3 and H4 genes in the distantly related sea urchins is highly conserved although the sequences are quite diverged, with most of the flanking regions and the spacer separating the H3 and H4 genes showing little homology. Although we do not find in the *S. purpuratus* clone the sequence homologies previously identified in the two late *L. pictus* H3-H4 intragenic spacers (19) we do identify seven other sites of striking homology in the regions 5' to the late genes in the *S. purpuratus* and *L. pictus* clones. When we compare the 5' regions of the *S. purpuratus* early and late genes, we find several stretches of homology in front of the H3 genes and other homologies preceding the H4 genes. We also present information on the timing of accumulation of multiple H4 and H3 RNA species and demonstrate that not all of the multiple late transcripts are regulated coordinately.

METHODS

Lambda Phage Cloning

A genomic library was constructed from the DNA of a single sea urchin individual, denoted S6, (20). To enrich the library for late histone genes

we used the null cloning technique of Childs et al. (18). The DNA was digested with Bgl II, an enzyme which does not cut the early gene repeat unit of *S. purpuratus*. Fragments of 15-20 kb were purified on 15-30% sucrose gradients and the genomic DNA was ligated to purified arms from the vector λ J1 (21) which had been cut with Bam HI (the purified λ J1 arms were a generous gift from Dr. Mary Ellen Digan). The ligated DNA was packaged (22) and the phage were amplified subconfluently on NZCYM plates. Phage were plated at a density of 300 phage/cm² and transferred to nitrocellulose (23). To screen the library for late genes, the filters were hybridized with an Hha I fragment of pC02 which contains the early H4 gene and surrounding spacer (20). After hybridization, the filters were washed at low stringency in 1 x SSPE (0.18 M NaCl, 10 mM NaH₂PO₄, pH 7.4, 1 mM EDTA) at 65° and exposed to X-ray film. The filters were then washed at high stringency in 0.1 x SSPE at 65° and re-exposed. Plaques which hybridized at low stringency but could not be visualized after the high stringency wash were picked and rescreened with the early H4 gene probe and with fragments containing the other four histone genes.

RNA Isolation

Sea urchins were purchased from Pacific Biomarine Supply Co. or Marinus Inc.. Embryos were cultured at 17°C in artificial sea water and were harvested at various developmental stages by centrifugation at 1000 rpm in the Sorvall GSA rotor. The embryos were washed in 10 volumes of TKM buffer (0.35 M KCl, 50 mM Tris pH 7.6, 10 mM MgCl₂) and homogenized in TKM containing 500 ug/ml heparin in a Dounce homogenizer using a tight pestle. Nuclei and cell debris were removed by centrifugation at 1500 rpm for 5 min in the Sorvall SS-34 rotor. The supernatant was precipitated with two volumes of ethanol and resuspended in 20 mM Tris, pH 7.6, 100 mM NaCl, 1 mM EDTA, 0.5% SDS, and 100 ug/ml proteinase K and allowed to digest for 2 hr. The RNA was extracted with phenol, precipitated with ethanol, and redissolved in sterile water for storage at -70°C. RNA to be used for cDNA synthesis was isolated from 48 hr. gastrula embryos. The RNA was enriched for 9S sequences by sedimentation on 5-20% sucrose gradients (24). Fractions containing 9S RNA were pooled and precipitated with ethanol. The 9S was stripped of poly(A)-containing sequences by affinity chromatography on oligo (dT) cellulose (25). RNA not bound to the column after 2 passes was precipitated, dissolved in sterile water and used to prepare 9S poly(A)- cDNA by tailing the RNA with *E. coli* polyA polymerase (26).

RNA Gel Electrophoresis and Blotting

Histone mRNAs were separated on 8M urea, 5% acrylamide gels in 0.45 M Tris, 0.45 M boric acid, pH 8.3, 20mM EDTA. Each sample contained 40 ug of RNA dissolved in buffer containing 8M urea. Samples were heated to 70°C for 10 min. before loading. Gels were run at 200 volts for 18 hr. and then the RNA was transferred directly to a Nylon membrane (Zeta-probe, Bio-Rad) in 40 mM Tris pH 7.4, 20 mM sodium acetate, 1 mM EDTA. The transfers were performed for 3 hr. at 4°C at 1.5 mA, 25 v, in a Hoefer Transphor apparatus. The membranes were then baked at 80°C for 2 hr., prehybridized overnight at 37°C in 40% formamide, 4 x SSPE, 1 mg/ml yeast RNA, 6 x Denhardt's solution, 0.6% SDS and then hybridized overnight at 37°C in 40% formamide, 4 x SSPE, 1 mg/ml yeast RNA, 1 x Denhardt's solution, with 10⁶ cpm/ml of DNA probe labeled with ³²P by nick translation (27). After hybridization, filters were washed with several changes of 1 x SSPE, 0.1% SDS at room temperature and at 65°C. After exposure to X-ray film, the filters were re-washed in 0.1 x SSPE, 0.1% SDS at 65°C.

DNA Sequencing and Computer Analysis

DNA was sequenced by the method of Maxam and Gilbert (28). Labeling of 5' ends was performed with polynucleotide kinase and γ -³²P-ATP; labelling of 3' ends was done with reverse transcriptase and α -³²P deoxynucleotide triphosphates (29). The late H3 and H4 genes were first subcloned from the clone λ Spl22 into pBR322 or PUC13. Homology searches were done with the programs of Pustell and Kafatos (30) on a NorthStar Horizon personal computer.

RESULTS

Isolation of Late Histone Gene Clones

Late histone genes from *S. purpuratus* were isolated by screening a genomic library for variant sequences which were not members of the tandemly repeated array of early histone genes. First, a null library was constructed, as described for *L. pictus* (18), which did not contain the early genes. DNA from a single sea urchin individual was digested with Bgl II, an enzyme which does not cut the early gene repeat. Since the early gene repeats are organized in long tandem arrays, they remain as fragments too long to be packaged into viable phage heads. The library, prepared in the Bam HI site of λ J1, was then screened with a nick-translated probe containing an early H4 gene. The blots were first washed at low stringency (1 x SSPE at 65°C), exposed, and then washed at high

stringency ($0.1 \times$ SSPE at 65°). After repeated testing at high and low stringency, clones were selected which showed reproducible, but only weak, homology to the early probe, thus eliminating early gene fragments such as "orphans" (31).

A preliminary identification of the late histone genes in these clones was made by hybridizing the phage DNA with nick-translated fragments specific for each of the five histone genes (derived from pC02). Only the H3 and H4 probes hybridized to the phage DNA, and as expected, only at low stringency. In the first screen of the library, 13 clones were selected, but they all turned out to be identical copies of one particularly abundant clone in the library. One of these, λ SpL22, was selected for further analysis.

To determine that λ SpL22 contained late histone genes and not some other histone gene set such as sperm or CS variants (1) which may show weak homology to the early genes, the lambda DNA was hybridized with late embryo cDNA (prepared by copying gastrula 9S poly A(-) RNA which had been tailed *in vitro* using E. coli poly A polymerase). The cDNA was hybridized to dots of λ SpL22 DNA and the T_m determined by raising the temperature of the hybridization wash. The T_m of the hybrids was $70-75^\circ\text{C}$, which is the expected range, at this salt concentration, for sequences with the base composition of histone genes. The T_m of the hybrids formed with the cDNA and the early gene clone, pC02, is at least 10° lower, consistent with previous melting experiments with early genes and labeled late histone mRNA (5, 12, 13).

Additional clones were selected from the library by using pLH3-SPE (a subclone of λ SpL22 which contains 150 bp of H3 coding sequence) as a probe. Four clones were isolated which form hybrids with melting characteristics of late genes. All four clones had different restriction maps, none of which were similar to λ SpL22. All contained H4 as well as H3 genes, but not H2A, H2B or H1 genes, as shown by hybridization to Northern blots of late RNA.

Sequence and Organization of a late H3-H4 Gene Pair

To more precisely examine the coding specificity and structure of the late genes in λ SpL22, we mapped the clone with restriction enzymes and sequenced a 2.5 kb region which includes the H3 and H4 structural genes. Fragments of λ SpL22 were subcloned into pBR322 and sequenced by Maxam and Gilbert methods. The restriction map of λ SpL22 and the sequencing strategy are shown in Figure 1. The DNA sequence of the 2.5 kb region is given in

Figure 1. Restriction map of clone λ SpL22 and sequencing strategy. Vector DNA is indicated by the open box and begins just to the right of the Bst HII site. Of the 8 kb insert, only the rightmost 2.5 kb containing the H4 and H3 genes is shown. The solid box denotes the sequence of the mRNAs. The 7 asterisks show the positions of the homologies in the spacers of the three late H3-H4 gene pairs which were compared. Each arrow beneath the map represents a particular sequenced fragment. Filled circles denote 5'-labeled ends and open circles denote 3'-labeled ends.

Figure 2. The overall organization of λ SpL22 consists of single H3 and H4 genes separated by 947 bp of spacer DNA. The genes are oriented in opposite directions with their 5' ends proximal. The spacer has no open reading frames larger than 130 bp. The sea urchin - lambda junctions in the clone are 6 kb downstream from the H4 gene but only 100 bp downstream from the H3 gene.

The sequence and organization of these H3 and H4 genes are dramatically different from the early genes, as reported for other late H3 and H4 genes (14, 18, 19). The late genes of λ SpL22 and the early genes code for identical H3 and H4 proteins; however, the coding regions of the *S. purpuratus* early (32) and λ SpL22 late H3 genes differ in nucleotide sequence by 17.0% and the early H4 (33, 34) and λ SpL22 H4 genes differ by 15.7%. Although the early and late genes have been constrained to produce the same proteins, they have accumulated extensive silent substitutions in their coding sequences.

The mRNA untranslated sequences and the spacers flanking the late genes show limited homologies with the early genes. The 3' untranslated regions do have the typical symmetrical sequence which is common to histone genes (35). The late H3 gene has a TATAAA box starting 59 bp upstream from the ATG start codon. The early and late H3 genes sequenced thus far contain the TATAAA sequence (18, 19, 35). The late H4 gene has the sequence TATATATA starting 66 bp upstream from the ATG codon. This sequence is also found in the two late *L. pictus* H4 clones (18, 19) but is different from the ACAATAC sequence which is the TATA analog in the early H4 genes (35). The 5' ends of the late *S. purpuratus* H3 and H4 mRNAs have

```

10      20      30      40      50      60
aattcttagg ggagggggcc caattttttt ttttttttac cccctacgac gaccacccca
ttaagaatcc cctcccccg gtaaaaaaa aaaaaaatg ggggatgctg ctgggggggt

70      80      90      100     110     120
tgggacactt accccccgga caactaggct aggcctagtg ggtagttaic atttatgaca
accctgtgaa cggggggcct gttgatccga tccggtacac ccatcaatag taatactagt

130     140     150     160     170     180
tgaatttgta catattatat agactataaa taattcataa aaatagtgtg tggcaaaaaa
actttaacat gtataatata tctgatattt attaagtatt ttatcacac accgttttta

190     200     210     220     230     240
agtcaactgt gaccgattgc caaacaaaaa caaatcaaga agaaagaatat cctaggcccta
tcagttgaca ctggcctaacg gtttgttttt gtttagttct tottcttata ggtaccggat

250     260     270     280     290     300
ttttgagaat gagaatacca agggtttaggc ctactgttag tgagcactaa tttttgtttt
aaaaacttta ctcttttaggt tcccaatccg gatgacaatc actcgtgatt aaaaaacaaa

310     320     330     340     350     360
cttcctctaa tcaccccccc cccccctttt tgtttaaaaa aaatggcatg taacagggttc
gaaggaggtt ggtggggggg gggggggaaa acaaattttt ttaccgttat atgttccaa

370     380     390     400     410     420
cctggccctg cctgccttac gagtggagaa attggggaaa gttgggtaat tatcaggccct
ggaccgggac gggacggaaatg ctactcttgg taaccccttt caaccattta atagtccgga

430     440     450     460     470     480
ggaaatcaca gatcatttct tatcactctc cagagtagtc tagctctaga cctagagctag
ccttttagtgt ctagttaaga atagtggagag gtctcatcag atcgagatct ggtactctagc

490     500     510     520     530     540
atctagtttt agactcactt tcgaattaat actttcaaaa gccccgcatt cttcggctcac
tagatcaaaa tctgagtgaa agcttaatta tgaaagtitt cggggcgtaa gaagccagtg

550     560     570     580     590     600
cacacacact tcattttatc tagactctag atcatgtgct gtcaaaactta gttagagatt
gtgtgtgtga ggtaaatatg atctgagatc tagtacacga cagtttgaaat caatctctaa

610     620     630     640     650     660
cttattgatt ctttctttgga tatttggttg cttcaaaaag agccgtttga tatgccgagt
gaataactaa GAAAGAACct ataaACCACC GAGATTTTC TCGGCAAAct atacggctca

670     680     692     704
taaggagtat ttccttc tta acc gcc gaa tcc gta cag ggt acg tcc ttg
attctccata aagggaag AAT TGG CGG CTT AGG CAT GTC CCA TGC AGG AAC
          G   G   F   G   Y   L   T   R   G   Q

716     728     740     752
gcg ctt cag agc gta gac gac gtc cat ggc agt gac ggt ctt tct ctt
CGC GAA GTC TCG CAT CTG CTG CAG GTA CCG TCA CTG CCA GAA AGA GAA
R   K   L   A   Y   V   V   D   M   A   T   V   T   K   R   K

764     776     788     800
ggc gtg ctc gca gta agt gac ggc atc acg gat gac gtt ctc aag gaa
CCG CAC GAG CGT CAT TCA CTG CCG TAG TGC CTA CTG CAA GAG TTC CTT
A   H   E   C   Y   T   V   A   D   R   I   V   N   E   L   F

812     824     836     848
gac ctt gag gac ccc acg ggt ctc ttc gta gat gag acc gct gat acg
CTG GAA CTC CTG GGG TGC CCA GAG AAG CAT CTA CTC TGG CGA CTA TGC
V   K   L   V   G   R   T   E   E   Y   I   L   G   S   I   R

860     872     884     896
ctt gac acc tcc acg tgg ggc aag acg ggc aat ggc ggc ctt ggt gat
GAA CTG TGG AGG TGC AGC CCG TTC TGC CGC TTA CCG CCC GAA GGA CTA
K   V   G   G   R   R   A   L   R   R   I   A   P   K   T   I

908     920     932     944
tcc ttg gat gtt atc acg caa aac ctt ggc atg acg ctt ggc tcc tcc
AGG AAC CTA CAA TAG TGC GTT TTG GAA CGC TAC TGC GAA CCG AGA AGG
G   Q   I   N   D   R   L   V   K   R   H   R   K   A   G   G

956     968     980     ← H4
ctt tcc gag tcc ttt tcc tcc ttt acc acg tcc aga cat attgacttga
GAA AGG CTC AGG AAA AGG AGG AAA TGG TGC AGG TCT GTA taactgaacat
K   G   L   G   K   G   G   K   G   R   G   S   N

1010     1020     1030     1040     1050     1060
gatttgacaa ttgaaaaaat ctgtactgaa aatcgagttt cggcgggtat atataactcc
ctaactggtt aactttttta gacatgactt ttagtctaaa gccgccata TATATTgAGG
          ..
          ..

1070     1080     1090     1100     1110     1120
tttgccgacg cgagtgtact aaataatttc tattgagagt gagccgccta cggccaaggg
AAACGCCTGC GCTCAcatga tttattaaag ataactctca ctgcggcgat gccagttccc

```

```

1130      1140      1150      1160      1170      1180
gactaaaatc tcgtcgcttc gtcgatgcaa tatttgata aactatcgca cggtcggtat
ctgatttttag agcagcggaag cagctacgtt ATAAACGTAT TTgatagcgt gcaagcaata

1190      1200      1210      1220      1230      1240
gaacaaatcg tctttactca gatcggtata aaaaatcgtc aagaacagag ttccggttat
cttgtttagc agaaatgagt ctagccatata ttttttagcag ttcttgcctc aaggccaata

1250      1260      1270      1280      1290      1300
caaatactat ttaaacactt taattcatct caatacagta ttttaagcgt ttagtttaca
gtttatgata aatttgtgaa attaagtaga gttatgtcat aaattcgcat aatcaaatgt

1310      1320      1330      1340      1350      1360
tattgcatag tagacaagag aacattaata tattttactt gacaaaatc gtttacaaca
ATAACGTATC ATctgttctc ttgtaattat ataaaatgaa ctggttttag caaatgttgt

1370      1380      1390      1400      1410      1420
ggtcgccagc accttatgaa taattcatca ggaactcctt aagtcgtttg ctgcgcaaaa
ccagcggtcg tggAACTAT ATTAAGtagt cctgaggaa ctcagcaaac gagcggtttt

1430      1440      1450      1460      1470      1480
tagaaaacaa cgttggaata ttctttcaat ttctactttt gtttggtaga aaaaggtc
atcttttgtt gcacctttat aagaaagtta aagatgaaaa caaaccaatt ttttatcagg

1490      1500      1510      1520      1530      1540
atattatttt ctattcaata tcgatcattt cattttcaat tttgatggat gcttttatgt
tataataaaa gataagttat agctagtata gtaaaagtta aaactacata cgaataaac

1550      1560      1570      1580      1590      1600
agaaagata acatacttct caaaaagcca aaatgtcgac gacaggaaag ccggttttgtt
tattacttat tgtatgaaga gtttttcggt tttagcgtg ctgtccttcc ggccaacaa

1610      1620      1630      1640      1650      1660
aatgaattat tcatctttac agcgatctcg aaaaaccaat ccagtagcat ctttcttcctc
ttacttaata agtaaaaaatg tcgctagagc ttttaggtta ggtcatgcta gaagaagaag

1670      1680      1690      1700      1710      1720
tgaaaccta atgaatatta cgtattagcg tataaatctc tgtaaTACAT TTACAATAc
aattttgat tacttataat gcataatcgc atatttaag acattatgta aatgtttatg

1730      1740      1750      1760      1770      1780
tactttacag cgataacgat gcaatttagg atgtaattaa gtttaattAT TCATatcttt
atgaaatgct gctattgcta cgttaaatcc tacatttaatt caaattataa agtatctaga

1790      1800      1810      1820      1830      1840
ataagcttta ctacaatgag caggttcaaa atcacacgac gaggcccgaa gaaATCATGA
tattgcaaat gatgttactg gtacatgttt tagtgtctg ctccgggctt ctttagtact

1850      1860      1870      1880      1890      1900
atatattaaq aaagcggaca gtacCCAATC gcatactgtg cttgatataq cgaatgCCA
tataaatctc tttgcctgt catgggttag tgtatgacac gaactatata gcttaccggt

1910      1920      1930      1940      1950      1960
ATcactgctt gtcgcagcag taaCCAATca tcttcgtcaa ttttgaTATA AATAcgagtg
tagtgacgaa cagcgtgctg attgggttagt agaagcagtt aaactatat ttagctcac

1970      ***      1990      2000
cgggattttt gaaacatcag ttgatatcac attcagcaaa tcaaa ATG GCC CGT ACC
gccttaaaaa cttttagtc aactatagtg taagtcgttt agttt tac cgg gca tgg
M A R T

2023      2035      2047      2059
AAG CAG ACC GCT CGC AAG TCC ACC GGA GGA AAG GCT CCT CGG AAG CAG
ttc gtc tgg cga gcg ttc agg tgg cct cct ttc cga gga gcg ttc gtc
K Q T A R K S T G G K A P R K Q

2071      2083      2095      2107
CTG GCC ACC AAG GCA GCT CGC AAG TCC GCC CCA GCC ACT GGC GGA GTC
gac cgg tgg ttc cgt cga gcg ttc agg cgg ggt cgg tga ccg cct cag
L A T K A A R K S A P A T G G V

2119      2131      2143      2155
AAG AAG CCC CAT CGT TAC AGG CCC GGT ACC GTC GCT CTC CGT GAG ATC
ttc ttc ggg gta cga atg tcc ggg cca tgg cag cga gag gca ctc tag
K K P H R Y R P G T V A L R E I

2167      2179      2191      2203
CGT CGC TAC CAG AAG AGT ACC GAG CTG CTC ATC CGC AAG CTC CCC TTC
gca cgg atg gtc ttc tca tgg ctc gac gag tag cgc ttc gag ggg aag
R R Y Q K S T E L L I R K L P F

2215      2227      2239      2251
CAG CGT CTG GTC CGT GAG ATC GCT CAG GAC TTC AAG ACC GAG CTC CGT
gtc gca gac ctg gca ctc tag cga gtc ctg aag ttc tgg ctc gag gca
Q R L V R E I A Q D F K T E L R

```



```

2263      2275      2287      2299
TTC CAG AGC TCT GCC GTC ATG GCT CTT CAG GAG GCC AGC GAA GCC TAC
aag gtc tcg aga cgg cag tac cga gaa gtc ctc cgg ctt cgg atg
F   Q   S   S   A   V   M   A   L   Q   E   A   R   E   A   Y

2311      2323      2335      2347
TTG GTC CGT CTT TTC GAG GAC ACC AAC CTT TGC GCC ATC CAC GCC AAG
aac cag gca gaa aag ctc ctg tgg ttg gaa acg cgg tag gtc cgg ttc
L   V   G   L   F   E   D   T   N   L   C   A   I   H   A   K

2359      2371      2383      2395
CGT GTT ACC ATC ATG CCA AAG GAC ATC CAG CTG GCC CGC CGC ATC CGT
gca caa tgg tag tac ggt ttc ctg tag gtc gac cgg cgc ggc tag gca
R   V   T   I   M   P   K   D   I   Q   L   A   R   R   I   R

2407      2426      2436      2446      2456
GGC GAG CGA GCT TAG attgtcagct tgacatctaa ataaaccAAC GGCTCTTTT
cgc ctc gct cga atc taacagtcga actgtagatt tatttgggtt ccgagaaaaa
G   E   R   A

2466      2476      2486      2496
AGAGCCCAAC catttcCAAG AAAGatcaca ttctaaccct tgcgtagatc
tctcgggtgtt gtaagggttc ttctagttt aagatttgag acgcatctag

```

Figure 2. Sequence of the late H3 and H4 genes and surrounding spacer of clone λ SpL22. The sequence shown extends from the Eco RI site 3' of the H4 gene to the border of the cloned fragment in the vector. The encoded amino acids are indicated by single capital letters using the conventional code. Heavily underlined capitalized sequences are the upstream homologies with the late *L. pictus* clones. Sequences underlined with dashes are homologies with upstream regions of the *S. purpuratus* early H3 and H4 genes. The TATA and CAAT boxes and the conserved region of dyad symmetry in the 3' untranslated regions are capitalized but not underlined. The asterisks indicate the 5' ends of the mRNAs as determined by S1 nuclease analysis.

been mapped by S1 nuclease analysis (17). These ends, indicated by asterisks in Figure 2, are located 29 to 31 bases downstream of the beginning of the TATA sequences. The 5' untranslated portions of the H3 and H4 mRNAs are therefore respectively 29–31 and 37–38 bases long. The late H3 gene is preceded by 3 copies of the sequence CCAATCA at positions –52, –78, and –101 with respect to the cap site. This sequence is found upstream from many H3, H2A, and H2B genes but has not been found in front of H4 and H1 genes (35). In addition there are a number of striking homologies between the *S. purpuratus* early and SpL22 late H3 5' flanking regions. There are 6 sequences of at least 7 exactly matching bases at identical sites within the 500 nucleotides upstream from the early and late *S. purpuratus* H3 genes. These are shown underlined with dashes in the sequence of Figure 2. Also underlined with dashes are several sequences just 5' to the SpL22 H4 gene. Since only 130 bp of sequence upstream of the transcription initiation site of the *S. purpuratus* early H4 gene has been reported (18, 19), only a very limited portion of the 5' flanking region of the early and late H4 genes can be compared. Nevertheless, two interesting homologies exist in this region between the early and late genes. The sequence GTCCGCAANGG starts at position –42 upstream of the

late gene and -49 upstream of the early gene. A part of this sequence, CCGCA, occurs at 4 other sites in front of the early gene between -117 and -60, but at no other site in front of the late gene. (This sequence is also homologous to sequences in two late H4 genes from *Lytechinus pictus*, see below). A second sequence, ACTCG, occurs at -11 and -48 in front of the late gene and at four sites between -133 and -25 upstream of the early gene. These sequences common to the 5' regions of early and late H3 and H4 genes might conceivably be recognized by H3-specific or H4-specific factors which would interact with both early and late genes of *S. purpuratus*. It is possible that the number of binding sites would be important in selectively enhancing or repressing transcription of one gene set relative to another.

Late H3 and H4 histone genes have already been isolated from *L. pictus* (18, 19) and late H4 genes from *S. purpuratus* (14), but only the *L. pictus* genes have been sequenced completely. Recently, the sequences of the untranslated 3' and 5' mRNA regions of a number of *S. purpuratus* late genes (including two late H4 genes) have been reported (15). In the case of *S. purpuratus*, these late H4 genes are organized differently from the genes we have cloned. One H4 gene is not closely linked to other histone genes and in the other cloned sequence, an H4 gene is closely linked to an H2B gene (14). Late H3 genes have heretofore not been isolated from this species. The *L. pictus* genes, however, are strikingly similar in organization to our clone λ SpL22 (18, 19). We shall discuss the homologies between these late genes in a section below.

The Late H3 and H4 Gene Family in *S. purpuratus*

Since the late H4 genes which have thus far been cloned from *S. purpuratus* (14) have a quite different arrangement from the two *L. pictus* H3-H4 clones (18, 19), it was of interest to investigate the organization of the late H3 and H4 genes as they exist in the genomes of different *S. purpuratus* individuals. Genomic DNA was cut with Bgl II for Southern analysis and hybridized with SpL22 fragment probes. Three probes were used as shown in Figure 3. Probe A contained most of the H4 coding sequence as well as half of the spacer between the genes. Probe C contained H3 coding sequence exclusively. Probe B contained spacer, the 5' untranslated region of the H3 mRNA, and sequence for the first 30 amino acids of the H3 coding region. Since the tandemly repeated early genes do not contain a Bgl II site, they will remain as high molecular weight DNA in the gel and not interfere with visualization of late gene sequences. In

Figure 3. Southern blot analysis of late H3 and H4 genes in *S. purpuratus*. Sperm DNA from four sea urchin individuals was cut with Bgl II and separated on 0.7% agarose gels. The blots were hybridized with nick-translated fragments subcloned from λ SpL22. The three DNA fragments used are indicated on the map drawn below as A, B, and C. After hybridization the filters were washed at 65°C in 0.1 x SSPE, 0.1% SDS.

addition, at the stringent washing conditions used, these late probes should detect only late sequences.

As shown in Figure 3, the A probe, containing H4 coding sequence, detects between 6 and 11 bands depending on the individual. Each band represents one or two copies per genome as determined by reconstruction experiments. The C probe, containing the H3 coding sequence, detects 8 to 12 bands per individual. Almost every band detected by the A probe is also detected by the C probe. The C probe identifies several additional bands in each lane. This suggests that not all of the late H3 genes are closely linked to H4 genes. In one case where there may be a unique H4 band (4.2 kb in individual 3N), there conceivably could be a Bgl II site in the spacer between the two genes. The B probe, consisting mostly of spacer and untranslated sequence, identifies only 1 or 2 bands in each individual. In each case these bands are among those which are also detected by the A and C probes. The late H3 and H4 gene family therefore consists of homologous

regions within the genes themselves, and a spacer region which is quite different in sequence in at least one H3-H4 unit in each genome. The λ SpL22 clone contains a 2023 bp Bgl II fragment which corresponds exactly to one of the Bgl II bands of individual 6S, from which the lambda library was prepared.

The H3-H4 genes of *S. purpuratus* therefore appear generally to be arranged quite similarly to those of *L. pictus*. The arrangements seen on the *S. purpuratus* clones examined previously (14) are rare and probably exist in the gene of only some individuals. In both *S. purpuratus* (these results) and *L. pictus* (18), the banding pattern is completely different from individual to individual. Even the subset of bands resulting with B probe hybridization shows extensive polymorphism. The high level of polymorphism (as much as 4%) which exists in single copy DNA from individual to individual (36) would result in many changes in restriction enzyme sites. The multiple bands visualized with our spacer specific probe are therefore almost certainly alleles.

Comparison of *S. purpuratus* and *L. pictus* Late H3-H4 Gene Pairs

The two non-allelic H3-H4 pairs cloned from *L. pictus* contain an extremely high degree of sequence homology in the coding regions (95%-96%) but little similarity outside of the transcription unit (19). Except for the expected TATA and CAAT sequences, the only homologies identified in the spacers of the two clones were stretches of 7 bp, at positions -150 and -173 relative to the H3 cap site, and 10 bp, at positions -155 and -173 relative to the H4 cap site (19). When we compare these two clones to the *S. purpuratus* SpL22 clone described here, the coding regions differ by 6.1-7.3% for the H3 and 8.4% for the H4 genes. These values are surprisingly less than the 10.5%-11.2% divergence between the *L. pictus* (19) and *S. purpuratus* (32) early H3 genes.

We have compared the spacer sequences of SpL22 with the spacers of the two *L. pictus* clones (pLpH3H4-21 and pLpH3H4-19, termed here Lp21 and Lp19) to find conserved sequences common to all three clones which may be important for the regulation of these genes. The 7 and 10 bp sequences common to the two *L. pictus* clones did not occur in the *S. purpuratus* clone. However, six regions of short precise homology were found and are noted in Figure 4. Three occur within 300 nucleotides upstream of the SpL22 H3 gene and three occur within 400 nucleotides upstream of the SpL22 H4 gene. The nine sequences in the three clones near the H4 genes are all similar with a consensus sequence of TATGCA whereas the sequences at any

LATE H4 HOMOLOGIESA.

Consensus	CACTCGACGTCCBCAA.AGG.GN TATATATA .CCG
S.p. 22	CACTCG.CGTCCBCAA.AGGaGt TATATATA cCCG
L.p. 19	CACTC.AC GTtCBCAA.AGG.Gc TATATATA .CCG
L.p. 21	CAtTCGACGTCCGgAAtAGG.Gg TATATATA .CCG

B.

	AA * TTT	
Consensus	TTTTATGCAAAA	
S.p. 22	gatagTTTATGCAAAATattg	-132
L.p. 19	catgTATTATGCAATgagtag	-166
L.p. 21	atacAcggATGCAAAATaaca	-138
S.p. 22	tgtcTAcTATGCAATAtgta	-281
L.p. 19	atacTTaTA.GCATAActta	-292
L.p. 21	ataaAAcAATGCAAAActgt	-258
S.p. 22	gatGAATTATtCATAAggt	-354
L.p. 19	tacGAATcATGCATTAttt	-399
L.p. 21	gaaaATTTATGCAAAATatc	-378

LATE H3 HOMOLOGIES

	*	
S.p. 22	aaATCATGAatatat	-140
L.p. 19	accgCATGActaatg	-149
L.p. 21	acATCATGAttggtc	-113
	*	
S.p. 22	aatATTTCATAatc	-206
L.p. 19	ttgCGTTCATAttt	-239
L.p. 21	ttaATTTCATAgta	-218
	*	
S.p. 22	aaTACATTTACAAATActa	-268
L.p. 19	aaatCAgTTACAAAtcgt	-302
L.p. 21	ttTACATTTACAAATAaaa	-303

Figure 4. Homologies between the upstream regions of *S. purpuratus* and *L. pictus* late H3 and H4 genes. The *L. pictus* sequences are taken from Childs et al. (18) and Roberts et al. (19). Numbers to the right of the sequence indicate the distance from the base with an asterisk to the 5' terminus of each gene.

one site upstream of the H3 genes are similar in the three clones but, each site has a different sequence. These sequences homologies in the 5' region of late H4 and H3 genes are not detected in the early *S. purpuratus* H4 and H3 genes.

None of these sequences were identified in the previous match between the two *L. pictus* clones (19). The matches between these two clones are no more than 7 bp and often only 5 bp at the six sites described here. This illustrates the limitation of comparing only two sequences for homology. The more clones that can be compared in a particular region, the clearer

(and more significant) the homologies become. The 7 and 10 bp sequences common to the two *L. pictus* clones (19) are therefore only species specific and not late gene specific. Analysis of other *S. purpuratus* clones will reveal whether a species specific sequence exists in that organism as well. As mentioned above, the SpL22 H3 gene has 3 copies of a CAAT-box sequence 5' to the gene. The two *L. pictus* late H3 genes have CAAT sequences as well. There is, however, no detectable homology among the three clones in the region between the CAAT boxes and the TATA box. The situation is quite different for the late H4 genes. Although there are no CAAT boxes, there is a 21 bp stretch in the SpL22 H4 sequence just 5' to the TATA box which strikingly matches sequences in the two late *L. pictus* H4 genes. As shown in Figure 4, the 21 bp stretch can be written as a 20 bp consensus sequence from which each H4 gene differs at 3 or 4 positions. This sequence also is very similar to the 21 bp just 5' to the TATA box of the early *S. purpuratus* H4 gene. This stretch in the early gene differs from the consensus sequence at 6 positions and contains the sequence GTCGCAANGG mentioned previously in the discussion of homologies between late and early genes. There is little sequence homology from the TATA box up to the AUG start codon in any of the late H4 genes, including the leader sequences of two other *S. purpuratus* clones recently published (15), or between the late H4 gene and the early H4 gene. The spacers of the three late clones, therefore, show the following similarities: three short sequences with consensus of TATGCA in front of the H4 genes, three sequences with a consensus of TTACAA, TTCATA, CATGA in front of the H3 genes, two or three CAAT sequences near the TATA boxes of the H3 genes, and a 21 bp conserved stretch just in front of the H4 TATA boxes. Only the 21 bp stretch is homologous to the early genes. The 3' untranslated regions of all the late genes have the expected region of dyad symmetry involved in termination processing (37-39).

Levels of Late H3 and H4 mRNAs

Subclones of λ SpL22 containing the H3 or H4 gene were used to hybridize to blots of total embryonic RNA separated on 8 M urea acrylamide gels. RNAs from egg and a series of embryonic stages were transferred to a nylon membrane by electroblotting. As can be seen in Figure 5 the acrylamide gel electroblot provides much better resolution than the usual agarose gel blot. The early and late mRNAs, as well as members of the late H3 and H4 class (which may differ by only a few nucleotides), can be easily differentiated. The probes containing coding sequence regions of the SpL22

Figure 5. RNA developmental profiles. RNAs were isolated from a single batch of eggs and embryos at the indicated times. 40 ug of total cytoplasmic RNA was loaded on each lane. The blotted RNAs were hybridized with nick translated pUC13 subclones of a *Hinf* I fragment containing 212 bp of late H4 coding sequence (panels a, b), of a *Pvu* II fragment containing 318 bp of late H3 coding sequence (panels c, d) or of a *Sal* I - *Pvu* II fragment containing 488 bp composed of 90 bp of the 5' end of the late H3 gene and 398 bp of upstream spacer (panels e, f). The filters were washed first at low stringency (1 x SSPE, 0.1% SDS, 65°C; panels a, c, e) and then at high stringency (0.1 x SSPE, 0.1% SDS, 65°C; panels b, d, f). Although RNA samples for each time point were identical for all panels, panels a and b are washes of a single transfer from one gel, panels c and d from a second gel, panel e is from a third gel, and panel f is from a fourth gel. Not all time points were run on each gel.

H3 and H4 genes hybridize to multiple bands. As shown in Figure 5, the late H4 probe hybridizes to both early and late RNAs (panel a) but after a high stringency wash, only two late H4 RNAs are visualized (panel b). The late H3 probe also hybridizes to early and late RNAs (panel c) but after a high stringency wash the three major late RNAs remain as stable hybrids (panel d). In the case of the H3 probe hybridization, two bands at the early RNA position (528 bases) can be seen even after the high stringency wash. Although this is due in part to cross-hybridization (note that a

band at the early RNA position is still seen in panel d in the samples from egg and 5 hr, stages at which there is very little, if any, late RNA detected on the blots), the remaining hybrid signal may also result from additional late H3 RNAs which co-migrate with the early H3 RNA. We think the latter explanation is probable in the case of the larger of the two species which run at approximately 528 bases, since the signal does not diminish after the high stringency wash as much as does the lower molecular weight band (clearly seen by comparing the ratio of the two bands in the 528 base region in the 13, 17, and 21 hr lanes of panels c and d). The SpL22 H3 gene product is contained in the band showing the strongest hybridization running at 491 bases. As seen in Figure 5 (panel f) only the middle of the 3 major late RNA bands is visualized after the high stringency wash when a probe containing only the 5' end of the SpL22 H3 mRNA sequence is used.

During the first 21 hr of development, the accumulation of late H4 and two of the three major H3 RNAs is fairly coordinate. The lowest molecular weight late H3 species may be regulated somewhat differently than the two other major late H3 RNAs since it is not detected until 17 hr and the ratio between this band and the slowest of the three bands clearly changes from 17 to 21 hr. Furthermore, the two bands seen at the 528 base position in the late samples after high stringency wash also appear to have an accumulation pattern which differs from the major late H3 band. In this case, the increase occurs earlier and the RNA levels remain constant from 13 to 21 hr. The Northern blots do not detect the late RNAs in the egg or 5 hr samples although S1 nuclease protection clearly shows that small amounts of late H3 and H4 RNAs can be found at these stages in some cases (8, 15, 17). The late H3 and H4 RNAs are first clearly visible in our blots at 9 hr and increase most strikingly from 13 to 17 hr. The expression of the late H3 and H4 genes may thus be more complex than a simple coordinate regulation of all of the members of the gene family.

The multiple bands are not allelic variants. The RNAs analyzed in Figure 5 were obtained from embryos of a mating between a single male and a single female. Identical results were obtained with RNAs obtained from another single mating. We do not know, however, if all of the mRNA species are transcribed from different genes. It is possible that some species (eg. the two late H4s) are derived from closely spaced promoters on the same gene.

DISCUSSION

Gene Organization

We have shown that the late H3 and H4 genes are generally found as pairs in the *S. purpuratus* genome. This result was somewhat surprising since the two previously isolated H4 genes from *S. purpuratus* were not linked to H3 genes (14). Yet our results, illustrated in Figure 3, show that virtually every restriction fragment which hybridized with an H4 probe also hybridized with an H3 probe. Some of these fragments, however, do not hybridize with equal intensity. One reason for this may be that only a portion of one or another of the genes is contained on that particular restriction fragment. The more intense bands may be due to gene pairs which may happen to occur on similarly sized fragments or to duplications of gene pairs. Our conclusion is that almost all H4 genes are linked to H3 genes but there are additional H3 genes which are contained on separate Bgl II fragments.

The number of late H3 and H4 genes may vary from individual to individual. Individual 6S appears to have many more genes than 5N or 11N. Some of the late genes may be pseudogenes or selection for the number of functional genes may not be stringent. There may be two allelic copies of a particular gene pair in each individual since the "B" spacer probe visualizes two bands, each representing one copy, in individuals 3N and 11N. The single band in individual 5N is consistent with this idea since the band is considerably darker indicating that the two copies of the pair in this individual are on restriction fragments of the same size. Individual 6S, however, does contain a high molecular weight fragment in addition to the two fragments of about 2 kb. In some individuals, therefore, intergenic spacers may be common to more than the two copies expected of allelic genes.

Sequence Divergence of Late and Early Gene Sets

As described above, there is only limited homology between regions flanking the early and late genes, exceptions being the TATA boxes, CAAT sequences in front of the H3 genes, six positions at which there are homologies of at least 7 bp within the 500 nucleotides upstream of the *S. purpuratus* H3 genes and two homologous sequences within the 130 bp upstream of the H4 genes. The early and late protein sequences, however, are identical. The nucleotide sequence divergence between the early and late *S. purpuratus* coding sequence is 17% for H3 and 15.7% for H4. The comparison of *L. pictus* H3 coding sequence shows a divergence of 18.4-19.7%

for early and late genes, depending on which of the two early repeat types present in this species is compared to the two published late sequences (19). The early *L. pictus* H4 sequences are not available so a similar comparison cannot be done for the early and late *L. pictus* H4 genes. Although the divergence is somewhat less for the *S. purpuratus* genes, an analysis of specific base changes clearly indicates a common ancestry for all late genes sequenced thus far. The divergence of the late gene lineage occurred well before the divergence of the *Lytechinus* and *Strongylocentrotus* species progenitors.

As noted above, there are six homologies of at least 7 bp in the 500 nucleotides 5' to the early and late *S. purpuratus* H3 genes and two homologies of 8 and 5 bp in the 130 bp upstream of the early and late H4 genes. These may serve as recognition signals for *S. purpuratus* H3- and H4-specific factors. The 8 bp sequence is actually part of a 20 bp consensus sequence, just 5' of the TATA box, which is also found in two *L. pictus* late H4 genes (shown in Figure 4). In addition to this H4 sequence, the three late gene pairs have 6 regions of homology in the region between the H3 and H4 genes. These homologies (see Figure 4) are not found in the early gene flanking regions and may be recognition sites for late gene specific factors.

Appearance of Late H3 and H4 RNAs

Although most of the late H3 and H4 RNAs detected on our electroblots accumulate from 9 to 21 hr of development, the regulation of the late gene family is not completely coordinate. The lowest molecular weight form of the late H3 RNAs is first seen somewhat later than the other two major bands and has a different accumulation pattern. At least one of the two RNA species in the late samples which run at the 528 base position also has a different accumulation pattern since there is no increase from 13 to 17 hr. There have been two recent reports indicating that not all late gene transcripts are coordinately regulated. Mohun et al. (15) found that in *S. purpuratus*, two late H4 RNAs and a late H2b RNA accumulated from 5 to 12 hr and then remained fairly constant until 72 hr. They also showed that RNAs encoded by closely linked H2a and H2b genes accumulated slowly during this period and increased greatly only after 16 hr. Busslinger and Barberis (16) showed that in *Psammachinus miliaris* different late H2a and H2b mRNAs accumulated at different times and, moreover, one particular late H2a species could be found in the testes whereas another could not. The major late H3 and H4 RNAs which we observe accumulate at times very similar

to the late H3 and H4 RNAs identified in *L. pictus* (8). We have not identified a late H4 species which increases most dramatically from 5-12 hr (15) but, as described above, we have demonstrated H3 species which appear earlier and somewhat later than the major late H3 RNAs which accumulate most dramatically from 13-17 hr during the blastula stage.

ACKNOWLEDGEMENTS

We would like to thank Dr. Peter Curtis for his gift of poly (A) polymerase and Dr. Mary Ellen Digan for λ J1 arms. This work was supported by NIH grant GM22155 to E.S.W.

* Current address: Miles Laboratories, Elkhart, IN 46515

+ To whom reprint requests should be directed.

REFERENCES

1. Maxson, R.E., Mohun, T., Cohn, R., and Kedes, L.H. (1983). *Ann. Rev. Genetics* **17**, 239-277.
2. Mauron, A., Kedes, L.H., Hough-Evans, B.R., and Davidson, E.H. (1982). *Dev. Biol.* **94**, 425-434.
3. Maxson, R.E. and Wilt, F.H. (1982). *Dev. Biol.* **94**, 435-440.
4. Weinberg, E.S., Hendricks, M.B., Hemminki, K., Kuwabara, P.E. and Farrelly, L.A. (1983). *Dev. Biol.* **98**, 117-129.
5. Kunkel, N.S. and Weinberg, E.S. (1978). *Cell* **14**, 313-326.
6. Maxson, R.E. and Wilt, F.H. (1981). *Dev. Biol.* **83**, 380-386.
7. Knowles, J.A. and Childs, G.J. (1984). *Proc. Natl. Acad. Sci. USA* **81**, 2411-2415.
8. Weinberg, E.S., Hendricks, M.B. and Hemminki, K. (1981). In: *The Role of RNA in Development and Reproduction*, M.C. Niu and H.H. Chuang, Eds. pp. 842-863. (Science Press, Beijing; distributed by Van Nostrand Reinhold, New York)
9. Newrock, K., Cohen, L., Hendricks, M.B., Donnelly, R. and Weinberg, E.S. (1978). *Cell* **14**, 327-336.
10. Grunstein, M. (1978). *Proc. Natl. Acad. Sci. USA* **75**, 4135-4139.
11. Childs, G., Maxson, R., and Kedes, L.H. (1979). *Dev. Biol.* **73**, 153-173.
12. Hieter, P.A., Hendricks, M.B., Hemminki, K., and Weinberg, E.S. (1979). *Biochemistry* **18**, 2707-2715.
13. Maxson, R.E., Mohun, T.J., and Kedes, L.H. (1983). *Nature (London)* **301**, 120-125.
14. Mohun, T., Maxson, R., Gormezano, G., and Kedes, L., (1985). *Dev. Biol.* **108**, 491-502.
15. Busslinger, M. and Barberis, A. (1985). *Proc. Natl. Acad. Sci. USA* **82**, 5676-5680.
16. Kaumeyer, J.F., Yager, L.N., Shupe, K. and Weinberg, E.S. In preparation.
17. Childs, G., Nocente-McGrath, C., Lieber, T., Holt, C. and Knowles, J. (1982). *Cell* **31**, 383-393.
18. Roberts, S.B., Weissner, K.E., and Childs, G. (1984). *J. Mol. Biol.* **174**, 647-662.

19. Overton, C. and Weinberg, E.S. (1978). *Cell* **14**, 247-257.
20. Mullins, J.I., Brody, D.S., Binari Jr., R.C., and Cotter, S.M. (1984). *Nature (London)* **308**, 856-858.
21. Hohn, B. and Murray, K. (1977). *Proc. Natl. Acad. Sci. USA* **74**, 3259-3263.
22. Benton, W.D. and Davis, R.W. (1977). *Science* **196**, 180-182.
23. Weinberg, E.S. and Overton, G.C. (1977). *Methods Cell Biol.* **19**, 273-286.
24. Aviv, H. and Leder, P. (1972). *Proc. Natl. Acad. Sci. USA* **69**, 1408-1412.
25. Sippel, A.E. (1973). *Eur. J. Biochem.* **37** 31-40.
26. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* **113**, 237-251.
27. Maxam, A.M. and Gilbert, W. (1980). In *Methods in Enzymology*, L. Grossman and K. Moldave, Eds., Vol 65, pp. 499-513 (Academic Press, New York).
28. Smith, D.R. and Calvo, J.M. (1980). *Nuc. Acids Res.* **8**, 2255-2274.
29. Pustell, J. and Kafatos, F.C. (1982). *Nuc. Acids Res.* **10**, 4765-4782.
30. Childs, G., Maxson, R., Cohn, R., Kedes, L.H. (1981). *Cell* **23**, 651-663.
31. Sures, I., Lowry, J. and Kedes, L.H. (1978). *Cell* **15**, 1033-1044.
32. Grunstein, M., Diamond, K.E., Knoppel, E. and Grunstein, J.E. (1981). *Biochemistry* **20**, 1216-1223.
33. Yager, L.N., Kaumeyer, J.F., and Weinberg, E.S. (1984). *J. Mol. Evol.* **20**, 215-226.
34. Hentschel, C.C. and Birnstiel, M.L. (1981). *Cell* **25**, 301-313.
35. Britten, R.J., Cetta, A., and Davidson, E.H. (1978). *Cell* **15**, 1175-1186.
36. Birnmeier, C., Grosschedl, R., and Birnstiel, M.L. (1982). *Cell* **28**, 433-440.
37. Birnmeier, C., Folk, W., and Birnstiel, M.L. (1983). *Cell* **35**, 433-440.
38. Birnmeier, C., Schumperli, D., Sconzo, G., and Birnstiel, M.L. (1984). *Proc. Natl. Acad. Sci. USA* **81**, 1057-1061.
39. Smith, M.M., and Murray, K. (1983). *J. Mol. Biol.* **169**, 641-661.
40. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., and Dodgson, J. (1980). *Cell* **20**, 555-565.