Proc. Natl. Acad. Sci. USA
Vol. 77, No. 3, pp. 1265–1269, March 1980
Biochemistry

Leader sequences of *Strongylocentrotus purpuratus* histone mRNAs start at a unique heptanucleotide common to all five histone genes

(polysomal RNA·DNA primer hybrids/reverse transcriptase/2′,3′-dideoxy method for sequence determination/DNA sequence organization)

IRMINGARD SURES*, SHOSHANA LEVY†, AND LAURENCE H. KEDES

Howard Hughes Medical Institute Laboratories and Department of Medicine, Stanford University School of Medicine and Veterans Administration Hospital, Palo Alto, California 94304

Communicated by Stanley N. Cohen, October 30, 1979

ABSTRACT **We have determined the sequence of the untranslated leader nucleotides of all five histone mRNAs from *Strongylocentrotus purpuratus* by the dideoxy chain termination method. Total polysomal RNA from sea urchin embryos was used as a substrate for cDNA synthesis primed by specific DNA restriction fragments. Each of the primers was derived from the 5′-terminal part of the coding region for a different histone protein. The five histone mRNA leader sequences are different in length and primary structure. The 5′ termini of all five histone mRNAs coincide with the unique heptanucleotide Py-Py-A-T-T-C-Pu in genomic DNA. This sequence, which defines the start of the individual histone mRNAs, is preceded by the A+T-rich octanucleotide identified in front of all eukaryotic structural genes where sequences have been determined to date.**

In sea urchins, the genes coding for the five major histone proteins are contained in a 5- to 7-kilobase DNA segment (1, 2) that is tandemly repeated several hundredfold (3, 4). The genes are separated by spacer segments of different length and are oriented in the same direction (2, 5–7). Sequence analysis of histone DNA from two different sea urchin species has previously defined the protein coding regions and has shown them not to be interrupted by intervening sequences (8–12).

Histone mRNAs each contain about 100 untranslated nucleotides (13, 14) distributed in an unknown manner between leader and trailer sequences within the mRNA. It is not clear whether the primary transcripts of sea urchin histone genes differ from the mature mRNAs. The location on the DNA of the termini of mature histone mRNAs would facilitate the assignment of possible functional significance to sequences on the DNA and RNA.

Comparison of the available DNA sequences preceding the coding regions of *Strongylocentrotus purpuratus* histone genes revealed a single common heptanucleotide sequence, 5′ Py-C-A-T-T-C-Pu 3′ (12). Similarly, regions of homology can be found in the DNA following the termination codon (12, 15). The presence of these homologies, coupled with the fact that for each gene the distance between them closely approximates the size of a mature mRNA, suggested to us that these sequences might define the 5′ and 3′ termini of the mRNAs. In fact, we have demonstrated that the 5′ end of the H2B mRNA lies very near the common heptanucleotide (16). In this report we show that the 5′ termini of the other four histone mRNAs map within this same sequence.

MATERIALS AND METHODS

Isolation of RNA. Polysomal RNA was extracted from early blastula *S. purpuratus* embryos as described (17). Purified H2A and H3 histone mRNAs were obtained by hybridization of total polysomal RNA to excess pSp117 DNA immobilized on cellulose as reported (18). RNA was stored in 75% ETOH at −20°C.

Isolation of DNA. The two chimeric plasmids pSp102 and pSp117 were the starting material for all the DNA fragments. pSp102 contains the genes for histones H1, H4, and H2B in a 4.6-kilobase *Eco*RI fragment; pSp117 contains a completely sequenced 1.95-kilobase *Eco*RI fragment containing the H3 and H2A genes. These two cloned fragments comprise an entire *S. purpuratus* histone gene repeat unit; their detailed endonuclease maps and DNA sequences have been published (12). Procedures for the isolation of plasmid DNA as well as preparation of DNA fragments obtained by restriction endonuclease digestion have been described (7, 12, 16). The DNA strands of the templates were separated as described (16, 19, 20). A schematic map of the histone genes and illustration of template and primer sources is shown in Fig. 1.

Template/Primer Hybrids. The appropriate double-stranded primer fragment (10 pmol) was denatured for 3 min at 100°C in 6 μl of ×1.6 H buffer (H buffer, 50 mM NaCl/34 mM Tris, pH 8.3/6 mM $MgCl_2$/5 mM dithiothreitol). The denatured primer was added to 4 μl of ×1.6 H buffer containing either 2 pmol of the appropriate single-stranded DNA template or 2 pmol of an RNA template; hybridization was at 68°C for 45 min. The final concentrations for primer and template were 1 and 0.2 μM, respectively. When polysomal RNA was used as template we assumed that each of the histone mRNAs comprised about 0.25% of the total RNA. To each of the primer DNA fragments, 125 μg of total polysomal RNA containing 2 pmol of each histone mRNA was hybridized. Purified H2A and H3 mRNA were also used at 2 pmol each. DNA·RNA hybrids were prepared under the same conditions as the DNA·DNA hybrids and not in the presence of formamide, as previously reported (16).

Sequencing Reactions. The base-specific chain termination method (21) was used with avian myeloblastosis virus reverse transciptase to determine the sequence of transcripts of mRNA or DNA templates. Reactions were performed in a total volume of 5.25 μl in silanized Eppendorf tubes. Each reaction mixture contained: 1 μl of either 1 mM dideoxy-ATP (ddATP), 0.5 mM ddCTP, 0.5 mM ddGTP, or 0.01 mM ddTTP (P-L Biochemicals); 2 μl of hybrids (in ×1.6 H buffer); 1 μl of dNTPs including dATP, dCTP, and dGTP, each at 500 μM (Sigma) and 10 μM ^{32}P-labeled dTTP (400 Ci/mmol; 1 Ci = 3.7 × 10^{10} becquerels; Amersham); 1 μl of reverse transcriptase (3 units in H buffer); and for RNA templates, 0.25 μl of actinomycin D (1 mg/ml in 50% ethanol). Incubation was for 20 min at 42°C; 1 μl of 0.1 mM dTTP (in ×0.8 H buffer) was added after 10 min. After hydrolysis of RNA, the cDNA was precipitated and analyzed (16).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: dd, dideoxy; bp, base pair(s).

*Present address: Department of Animal Biology, University of Geneva, Geneva, Switzerland.
† Present address: Life Sciences, SRI International, Menlo Park, CA 94025.

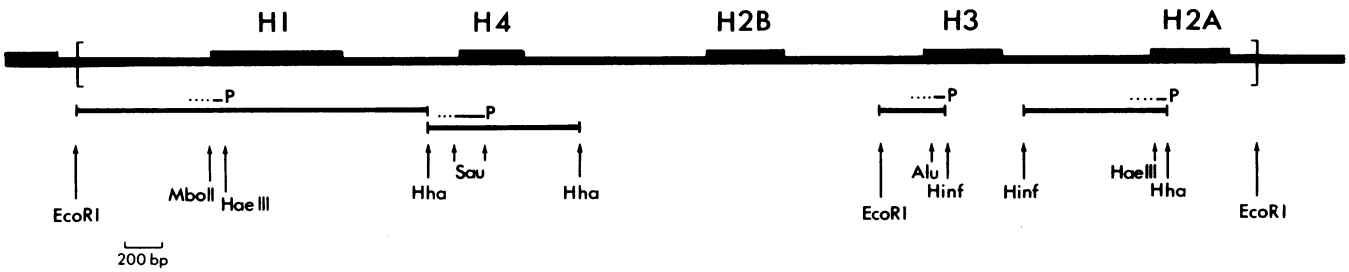

FIG. 1. Schematic map of the histone gene repeat unit of *S. purpuratus* and location of template and primer fragments. The protein coding regions are indicated by thick horizontal lines; the direction of transcription is from left to right. Only those restriction sites that were used to generate the DNA primers and templates (thin horizontal lines) are shown. The primers (P) are extended leftward as indicated by the dotted lines. bp, Base pairs.

RESULTS AND DISCUSSION

Sequencing Strategy. Reverse transcriptase requires a primer fragment to initiate cDNA synthesis on a mRNA template. Once hybridized to the mRNA, the primer can be extended and, if dideoxynucleoside triphosphates are included in the reaction mixture, the resulting cDNA can be subjected to sequence determination (22–27).

We have previously used this method to determine the 5′-terminal nucleotides of the H2B histone mRNA (16). cDNA synthesis was primed with a short restriction fragment derived from the coding region of the H2B gene; the template was either purified H2B or total polysomal RNA. The resulting sequences were identical, thus demonstrating that purification of the template RNA is not necessary. The unique primer itself selects by hybridization the appropriate mRNA template. The H2B sequences derived from these RNA templates contained nonspecific breaks at several positions which made a correct nucleotide assignment difficult. During sequence analysis of the other histone mRNAs, we found that these breaks could be eliminated if formamide (which was originally included to favor RNA·DNA hybrid formation) was omitted from the hybridization mixtures. The sequencing procedure is now further simplified by the use of hybrids directly, without precipitation, as substrates for reverse transcriptase.

Using specific primers and total polysomal RNA, we analyzed the untranslated leader nucleotides of the mRNAs for histones H1, H2A, H3, and H4. The primer fragments we chose for H1 [24 base pairs (bp)], H2A (21 bp), and H3 (32 bp) are located *within* the protein coding regions. The distances to the first nucleotide of the initiation codons are 22, 28, and 73 bp,

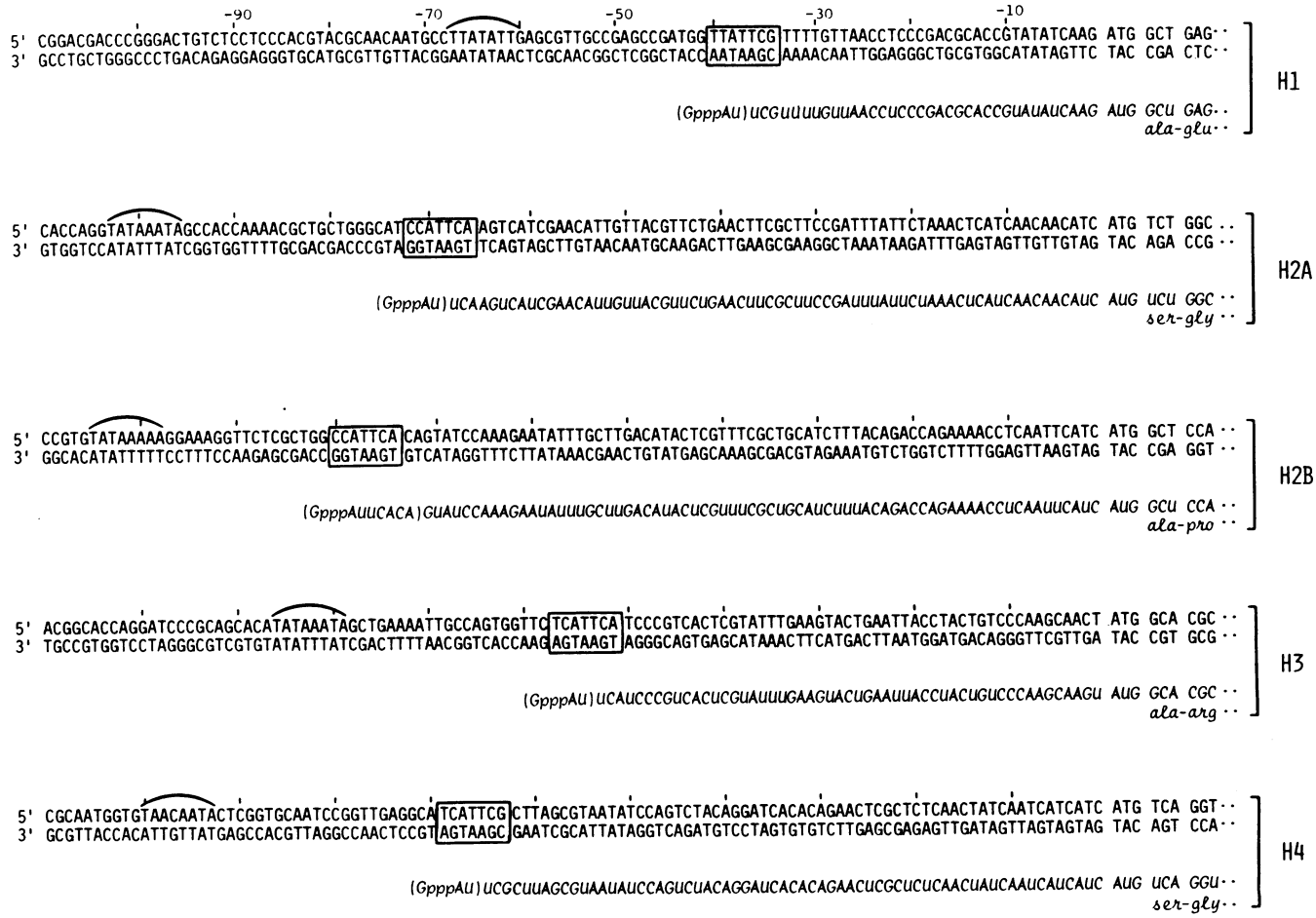

FIG. 2. Leader sequences of the histone mRNAs and their corresponding and bordering DNA regions. Nucleotides are numbered by reference to the first nucleotide of the initiation codon. The common heptanucleotide (boxed) and the A+T-rich nucleotide stretch (bracketed) are indicated. Nucleotides in parentheses at the 5′ ends of the mRNAs were not determined (see *Results and Discussion*).

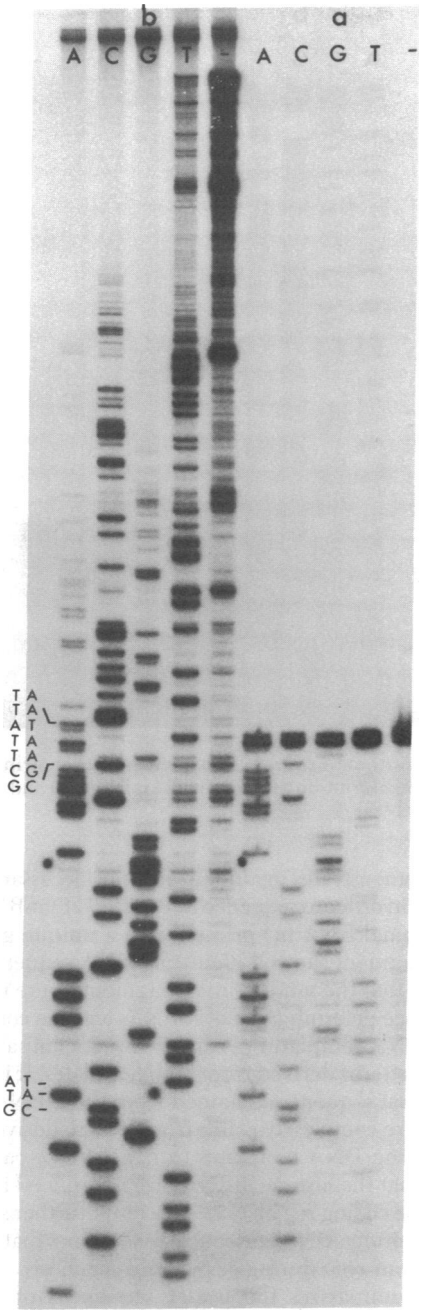

FIG. 3. Autoradiograph of a sequencing gel showing the chain-terminated transcription products of H1 mRNA (*a*) and H1 DNA (*b*) templates. The initiation codon and the common nucleotide 5′ Py-Py-A-T-T-C-Pu 3′ are indicated.

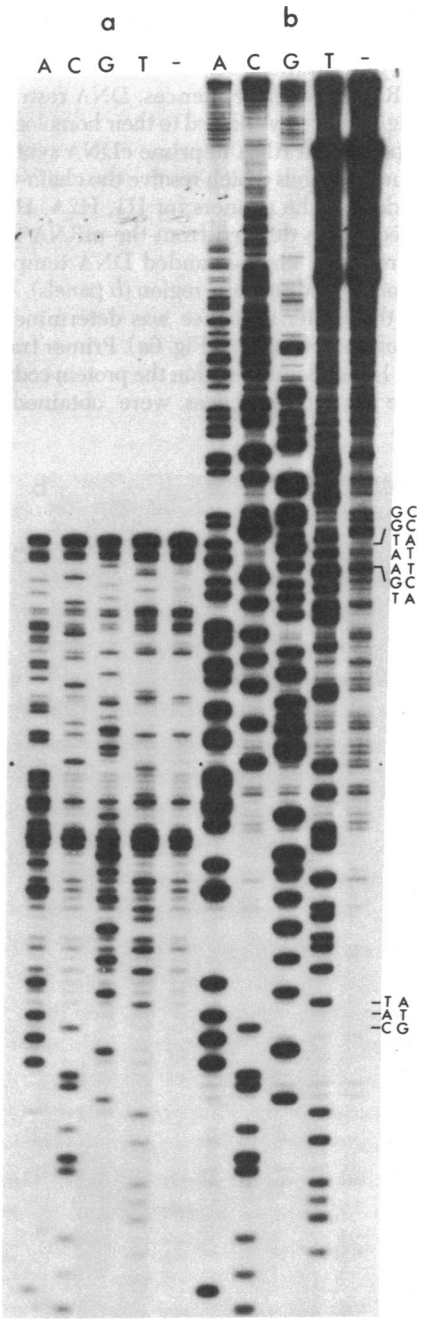

FIG. 4. Autoradiograph of a sequencing gel showing the chain-terminated transcription products of H2A mRNA (*a*) and H2A DNA (*b*) templates.

respectively. The primer fragment for H4 (175 bp) contains 38 bases of the noncoding region. Each of these fragments was used to prime cDNA synthesis on total polysomal RNA. The same fragments were also used to prime cDNA synthesis on single-stranded genomic DNA templates. Each of the DNA templates contains at least 200 nucleotides upstream from the respective initiation codon. The length and the location of the four DNA templates in relation to the primer fragments is shown in Fig. 1.

Thus, the same primer was used for sequence determination, side by side, of mRNA and the homologous genomic DNA region. Originally (16) this was necessary in part because the sequences derived from the RNA templates were ambiguous at several positions and could be assigned only by comparison to the corresponding DNA sequence. Now, after omission of formamide, the sequences are unambiguous and are easily read from the autoradiographs. However, as discussed below, we still considered it advantageous to study genomic DNA in parallel with the mRNA.

DNA Sequences. DNA sequences in front of the H1 and H2A protein coding regions had not been completely determined previously (12). Additional nucleotide sequences preceding these genes have now been obtained and are presented in Fig. 2. As indicated, the DNA region in front of the H1 gene contains the sequence 5′ Py-Py-A-T-T-C-Pu 3′. This heptanucleotide had previously been identified as the only common sequence preceding the H2A, H2B, H3, and H4 genes. Thus, all five histone genes in *S. purpuratus* contain this same sequence.

The sequences determined in this study by the chain termi-

nation method confirm our published data (12) but for one nucleotide: the published H2A leader contained a single error, an additional ($\frac{T}{A}$) at position -33.

Histone mRNA Leader Sequences. DNA restriction fragments (P in Fig. 1) were hybridized to their homologous mRNA templates in polysomal RNA to prime cDNA synthesis. Figs. 3–6 show sequencing gels which resolve the chain-terminated cDNAs extended on the primers for H1, H2A, H3, and H4. Each shows sequences derived from the mRNA template (*a* panels) and from the single-stranded DNA template which contains the corresponding gene region (*b* panels). A relatively short part of the leader sequence was determined at the 5′ terminus end of the H4 mRNA (Fig. 6*a*). Primer fragments for H1, H2A, and H3 are located within the protein coding regions, and complete leader sequences were obtained for these mRNAs.

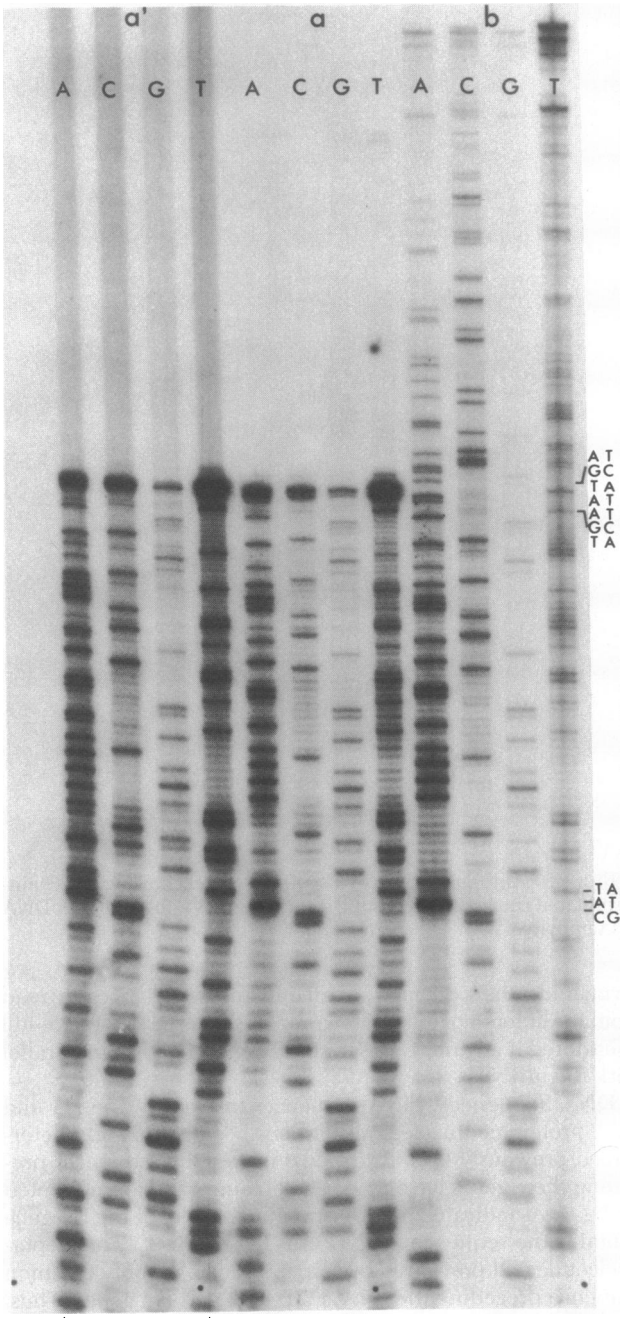

FIG. 5. Autoradiograph of a sequencing gel showing the chain-terminated transcription products of purified H3 mRNA (*a'*), unpurified H3 mRNA (*a*), and H3 DNA (*b*) templates.

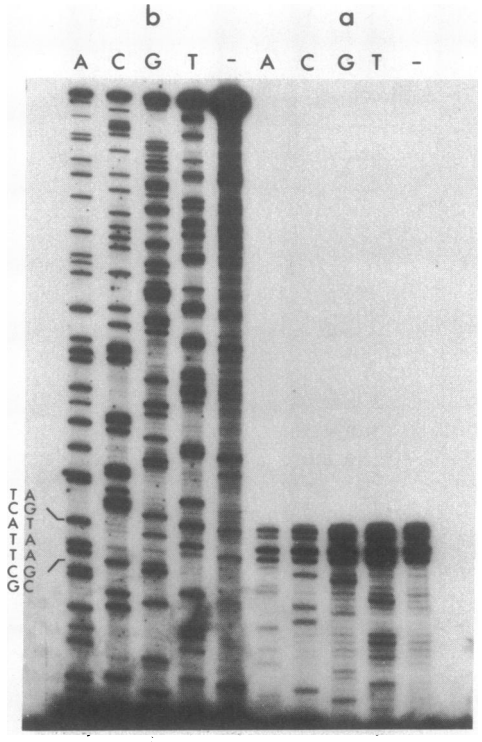

FIG. 6. Autoradiograph of a sequencing gel showing the chain-terminated transcription products of H4 mRNA (*a*) and H4 DNA (*b*) templates.

Fig. 5 demonstrates again that it is not necessary to purify the mRNA in order to sequence it. Purified H3 mRNA (*a'*) and total polysomal RNA (*a*) primed with a unique genomic restriction fragment showed identical cDNA sequences.

The fact that the same primer fragments were used in parallel sequence determination of mRNAs and the corresponding genomic DNA templates permits a direct comparison of the sequence patterns derived from both templates. It is clear from Figs. 3–6 that sequences obtained from the mRNA templates (*a* panels) are completely colinear with those derived from the corresponding DNA templates (*b* panels). We can therefore conclude that the histone mRNAs are not spliced in either the leader or the coding region (12). This result further suggests that the cloned histone DNA is identical to the genes that are actively transcribed *in vivo* during early embryogenesis.

Fig. 2 summarizes the leader sequences of the histone mRNAs and their bordering DNA regions. In agreement with the "scanning" hypothesis of Kozak (28) the leaders do not contain AUG trinucleotides in front of the AUG codon used for initiation of translation. It is apparent that all five mRNA leaders are very different in size and primary structure. However, projection of the RNA sequences on the homologous DNA regions reveals that all five mRNAs start within the sequence 5′ Py-Py-A-T-T-C-Pu 3′. We believe that this particular sequence is located at or near the end of the RNA templates and is not merely a strong stop signal for reverse transcriptase because the same sequence is copied right through in the corresponding DNA templates (Figs. 3–6). Moreover, the same sequence is contained within another RNA template, the redundant portion of Rous sarcoma virus (29, 30), that is transcribed by reverse transcriptase *in vivo*. It is thus very likely that the 5′ borders of the five histone mRNAs are defined by the sequence 5′ Py-Py-A-T-T-C-Pu 3′ in the histone gene repeat unit of *S. purpuratus*.

What exactly are the first bases at the 5′ ends of the histone mRNAs? Histone mRNAs are capped (31), and reverse tran-

scriptase may not transcribe this unusual structure. It therefore is not clear whether the 5′-terminal nucleotides are present in the cDNA transcript. Sequence analysis of adenovirus hexon mRNA by the chain-termination method (27) revealed that only the 5′-terminal and the penultimate nucleotide of the known cap structure cannot be identified on the sequencing gel. If a similar situation obtained for histone mRNA, then the 5′-terminal nucleotide of the histone mRNAs (with the exception of H2B mRNA) coincides with the A in the common heptanucleotide sequence. This conclusion is supported by the finding that the penultimate nucleotide in 76% of the cap structures of partially purified histone mRNAs from sea urchin embryos is adenine (31).

We note that a similar heptanucleotide sequence, 5′ N-C-A-N-T-Py-Pu 3′, is present at identical locations in front of the four sequenced histone genes of the sea urchin *Psammechinus miliaris* (11). Furthermore, a similar nucleotide sequence, 5′ N-C-Pu-T-T-Py-Pu 3′, is also present in front of three histone genes in *Drosophila melanogaster* (32). The presence of this sequence at analogous positions in histone gene repeats of such evolutionary divergent forms as insects and echinoderms argues strongly for some functional significance.

DNA Regions Preceding the Start of the mRNAs. Comparative analysis of DNA sequences preceding the protein coding regions of several eukaryotic genes has revealed the presence of a very A+T-rich nucleotide stretch (32–36). It has been suggested that this particular sequence—similar to the "Pribnow box" (37)—might be related to transcription initiation because it maps close to the DNA region that codes for the mature mRNA terminus (33, 34). Direct evidence that this A+T-rich sequence has functional significance, however, is still lacking. A similar or identical sequence is present in all five histone genes. As shown in Fig. 2, the octanucleotide 5′ T-A-T-A-A-A-$\overset{T}{\underset{A}{}}$-A 3′ is located 20–25 bases upstream from the 5′ termini of the H2A, H2B, and H3 mRNAs. The H1 and H4 genes do not show this exact sequence but they do contain an A+T-rich nucleotide stretch at analogous positions. The assignment of a possible function to these sequences rests in part on their location relative to the termini of the primary transcripts, which are as yet unknown. Nevertheless, their presence in close proximity to the terminus of the mature mRNA suggests that they play some role in either transcription or the processing of the primary transcript.

We thank Dr. J. W. Beard for reverse transcriptase and Dr. Robert Maxson for help in preparing the manuscript. L.K. is an investigator of the Howard Hughes Medical Institute. The research has been supported by grants from the American Cancer Society, the National Institutes of Health, and the Veterans Administration.

1. Kedes, L. H., Cohn, R. H., Lowry, J. C., Chang, A. C. Y. & Cohen, S. N. (1975) *Cell* **6**, 359–370.
2. Schaffner, W., Gross, K., Telford, J. & Birnstiel, M. L. (1976) *Cell* **8**, 471–478.
3. Kedes, L. H. & Birnstiel, M. L. (1971) *Nature (London) New Biol.* **230**, 165–169.
4. Weinberg, E. S., Birnstiel, M. L., Purdom, I. F. & Williamson, R. (1972) *Nature (London)* **240**, 225–228.
5. Gross, K., Schaffner, W., Telford, J. & Birnstiel, M. L. (1976) *Cell* **8**, 479–484.
6. Wu, M., Holmes, D. S., Davidson, N., Cohn, R. H. & Kedes, L. H. (1976) *Cell* **9**, 163–169.
7. Cohn, R. H., Lowry, J. C. & Kedes, L. H. (1976) *Cell* **9**, 147–161.
8. Sures, I., Maxam, A., Cohn, R. H. & Kedes, L. H. (1976) *Cell* **9**, 495–502.
9. Birnstiel, M. L., Schaffner, W. & Smith, O. H. (1977) *Nature (London)* **266**, 603–607.
10. Grunstein, M. & Grunstein, J. E. (1977) *Cold Spring Harbor Symp. Quant. Biol.* **42**, 1083–1092.
11. Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, O. H. & Birnstiel, M. L. (1978) *Cell* **14**, 655–671.
12. Sures, I., Lowry, J. & Kedes, L. H. (1978) *Cell* **15**, 1033–1044.
13. Holmes, D. S., Cohn, R. H., Kedes, L. H. & Davidson, N. (1977) *Biochemistry* **16**, 1504–1512.
14. Grunstein, M. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4135–4139.
15. Busslinger, M., Portmann, R. & Birnstiel, M. L. (1979) *Nucleic Acids Res.* **6**, 2997–3008.
16. Levy, S., Sures, I. & Kedes, L. H. (1979) *Nature (London)* **279**, 737–739.
17. Kedes, L. H. & Gross, P. R. (1969) *Nature (London)* **223**, 1335–1339.
18. Childs, G., Levy, S. & Kedes, L. H. (1979) *Biochemistry* **18**, 208–213.
19. Szalay, A. A., Grohmann, K. & Sinsheimer, R. L. (1977) *Nucleic Acids Res.* **4**, 1569–1578.
20. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.*, in press.
21. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
22. Zimmern, D. & Kaesberg, P. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 4257–4261.
23. McGeoch, D. J. & Turnbull, N. T. (1978) *Nucleic Acids Res.* **5**, 4007–4024.
24. Akusjarvi, G. & Pettersson, U. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5822–5826.
25. Bina-Stein, M., Thorn, M., Salzman, N. & Thompson, A. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 731–735.
26. Kan, Y. W. & Chang, J. C. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 2886–2889.
27. Akusjarvi, G. & Pettersson, U. (1979) *Cell* **16**, 841–850.
28. Kozak, M. (1978) *Cell* **15**, 1109–1123.
29. Haseltine, W. A., Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 989–993.
30. Schwartz, D. E., Zamecnik, P. C. & Weith, H. L. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 994–998.
31. Surrey, S. & Nemer, M. (1976) *Cell* **9**, 589–595.
32. Goldberg, M. (1979) Dissertation (Stanford Univ., Stanford, CA).
33. Ziff, E. B. & Evans, R. M. (1978) *Cell* **15**, 1463–1475.
34. Tsujimoto, Y. & Suzuki, Y. (1979) *Cell* **16**, 425–436.
35. Smith, M., Leung, D. W., Gillam, S., Astell, C. R., Montgomery, D. L. & Hall, B. D. (1979) *Cell* **16**, 753–761.
36. Gannon, F., O'Hare, K., Perrin, F., LePennec, J. P., Benoist, C., Cochet, N., Breathnach, R., Royal, A., Garapin, A., Cami, B. & Chambon, P. (1979) *Nature (London)* **278**, 428–434.
37. Pribnow, D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 784–788.