A regulatory sequence near the 3' end of sea urchin histone genes

Meinrad Busslinger, Rudolf Portmann and Max L. Birnstiel

Institut für Molekularbiologie II der Universität Zürich, Winterthurerstr. 266A, 8057 Zürich, Switzerland

Received 8 May 1979

ABSTRACT

The 3' flanking sequences of all five histone genes have been sequenced in the histone DNA clone h19 of the sea urchin *Psammechinus miliaris*. A large (23 bp) and a small (10 bp) conserved sequence was found by sequence comparison, some 29-40 bp downstream from the termination codon. 12 bases of the larger homology block show a dyad symmetry. The available sequences of clone h22 of the same species and those of the histone clones pSp2 and pSpl7 of *Strongylocentrotus purpuratus*, another sea urchin species, fit well into this comparison. Two types of sequences are involved in the dyad symmetry; one is H1, H3 and H4 specific, the other is H2A and H2B specific. If these conserved sequences are transcribed, a hairpin loop could form in the RNA molecules. This secondary structure might serve as a recognition signal for a regulatory protein.

INTRODUCTION

Recently, interest in the problem of eukaryotic gene organization has focused on the structure of regulatory sequences. The delimitation of such regions by means of classical genetic techniques is greatly impeded in eukaryotes owing to the complexity of the genome. One alternative has been to carry out computer comparisons of DNA sequences in putative regulatory regions, another, to bring cloned genes in their natural or modified form to expression by reintroducing them into the living cell.

The theory underlying the sequence comparison approach has been that regulatory sequences should be conserved, because they are important to the living cell. Such sequences would most easily be detected when comparing a family of genes serving a

similar function within a single species but which are tissue specific or developmentally controlled. The histone genes of the sea urchin constitute such a gene family, showing considerable ontogenetic variety (reviewed in ref.1). A gene cluster coding for histone proteins in Psammechinus miliaris, clone h22, which belongs to the majority type of histone gene repeats has been well characterized and a major part of this clone has been sequenced (2). By screening 70 individual isolates of clones originally obtained using a palette of restriction enzymes, a variant DNA clone, clone h19, was selected. Sequencing data has since revealed that h19 codes for histone variants differing from those encoded by h22 (unpublished results).

Here, we report the results of a comparison of the sequences flanking the 3' ends of the codogenic regions of all five histone genes of clone h19. All regions contain a large (23 bp) and a small (10 bp) conserved sequence block, separated from one another by a non-homologous sequence of 6 bp. The first sequence block shows a dyad symmetry and, if transcribed, would form a hairpin loop. A closely similar sequence arrangement can also be found in the available sequences of clone h22 (2) and in a histone gene cluster of Strongylocentrotus purpuratus (3), a related sea urchin species.

MATERIALS AND METHODS

Material

Psammechinus miliaris h19 histone DNA (4) was prepared and purified as described for h22 histone DNA (5). Restriction endonucleases were purchased from New England Bio Labs, bacterial alkaline phosphatase from Worthington and T₄ polynucleotide kinase from P-L Biochemicals. (γ ³²P)ATP (spec. activity: 1000-3000 Ci/mMol) was obtained from Amersham-Searle. Preswollen DEAE-cellulose powder (DE 52) was from Whatman, hydrazine from Eastman Kodak, piperidine from Merck, acrylamide and bisacrylamide from Serva.

5' End labelling of restriction fragments

The conditions for all endonuclease digestions were those

suggested by the enzyme suppliers. DNA fragments were dephosphorylated and end labelled according to the procedure of Maxam and Gilbert (6) modified as described by Boseley et al. (7).

Isolation of restriction fragments

End labelled DNA fragments were separated on 1% or 2% agarose slab gels by electrophoresis in Loening E buffer. The DNA was eluted from the gel either electrophoretically (8) or by diffusion (6). This DNA was further purified on a DEAE-cellulose column. Each fragment was then cleaved asymmetrically by digestion with a restriction enzyme and the individual labelled fragment recovered by gel electrophoresis.

DNA sequencing

The sequencing procedure was based on that of Maxam and Gilbert (6). For the A-specific reaction 5 μ l of 32 P-labelled DNA was incubated with 20 μ l 1.5M NaOH in a sealed capillary tube at 90°C as described by Maxam and Gilbert. Thereafter, the procedure was modified in that the contents of the capillary were neutralized with 200 μ l 0.3M NaAc/HAc, pH 4.75, and 10 μ l 1M Tris/HAc, pH 7.5 prior to alcohol precipitation. Furthermore, 200 μ l 0.1mM EDTA, but no NaAc was added after hydrazinolysis to the C-specific reaction mixture. 200 μ l 0.3M NaCl, 0.1mM EDTA was added to the T-specific reaction mixture. All subsequent steps were performed again according to Maxam and Gilbert.

After the last lyophilisation the DNA was dissolved in 5mM Tris-borate, pH 8.3, 0.1mM EDTA, 7M urea, 0.05% xylene cyanol, 0.05% bromphenol blue (Maxam and Gilbert, personal communication), heated for 1 minute at 90°C and immediately loaded onto 7M urea gels of different acrylamide concentrations (20%, 15%, 10%, 7%) in TBE-buffer (9). These gels of the dimension 48 x 22 x 0.05 cm were pre-electrophoresed at 250 V and were then run at such a voltage that the gels were hot to the touch. After electrophoresis the gel was exposed to a preflashed X-ray film according to Laskey and Mills (10).

RESULTS

Fragments used for sequencing

The 6 kb histone DNA unit of clone λ h19 (4) was transferred to the λ vector Sam 7 for high yield (unpublished results). The phage was grown and purified essentially as described previously (5). A restriction map of the 6 kb unit was constructed and the location of the five histone genes determined first by the Southern technique and then by partial sequencing of the co-genomic regions (data not shown).

The 3' regions adjacent to the codogenic regions of the five histone genes were also sequenced. The locations of the DNA fragments used for end labelling and sequencing are denoted in Figure 1 for each individual gene by arrows above the line.

Figure 1. Sequencing of the 3' regions of all five histone genes in clone h19. For explanation see text.

The arrows below the line denote the asymmetric cleavage points required for sequencing according to the Maxam and Gilbert procedure. The horizontal arrows represent the length of nucleotide sequences analyzed.

The minor modification of the Maxam and Gilbert procedure (6) mentioned in Material and Methods resulted in a more equal yield from the different sequencing reactions and in an improvement in the number of nucleotides which could be determined from a single sequencing experiment. Nevertheless, the initial sequencing gels covering the 3' end of the structural genes showed an area of unequal intensities and spacings of the bands in the C-track. A closer analysis of the surrounding sequence revealed that these were artefacts which arose because the sequences in this region are palindromic (see below). We tried to resolve this problem by running the gels at higher temperatures in order to keep the DNA completely denatured during electrophoresis. This strategy led to unambiguous sequences for the H1, H3 and H4 flanking regions. For corresponding sequences of the H2A and H2B genes, however, where the calculated stability of the palindrome is higher (see below), it was necessary to sequence both DNA strands in opposite directions as is illustrated in the case of the H2A gene in Fig.2a and b. Reading from the labelled end, the first part of the palindromic sequence can be clearly determined, whereas the second part is obscured by renaturation artefacts leading to the reading of a truncated sequence. However, the correct sequence can be read on the other strand running in the opposite direction.

Search for homologies

Since the region downstream of the five structural histone genes presumably contains regulatory signals of some kind, the sequences were analyzed by a computer programme which allowed printing out of homologies at threshold levels up to 9 nucleotides in a row. A compilation of the results is shown in Figure 3, each sequence of the 3' flanking regions downstream of the termination codon being presented in detail. As seen, some 29-40 bases removed from the structural gene, there is an area of

Figure 2. Gels showing the sequence flanking the 3' end of the H2A gene in clone h19: a) sequence determined from the *Alu* I site at the end of the gene towards the spacer region (see Fig.1, line 2), b) the same sequence determined on the opposite strand from the *Hinf* I site in the spacer towards the 3' end of the gene (see Fig.1, line 3). Bracketed nucleotides in gel a) can only be correctly read in gel b) and vice versa.

```

TGATTTCACGCCAACTTTCCCATCTACCAAAACGGCTCTTTTCAGAGCCAC CAAATAATCAAGAAAGAACACTTGTCCAAATG POST H1
TAGAGTTTGCTCCCGCAAATCTTGAACCTCAACGGCCCTATCAGGGCCAC CAATTACTCAGCAAGAAGATTGTTTCATTATG POST H2A
TAGACAGGTCATATCCTGCTCTAATTGGACATACAACGGCCCTTTTCAGGGCCACACAATAATCAAGAAAGAAATGATATCCGTAATA POST H2B
TAGAACCATCGGTACAGCATGTAGCCCATGCACCGCATACACAAACGGCTCTTTTCAGAGCCAC CACAACCCCAAGAAAGAAATCACTGACTTCTAA POST H3
TAAGTGTAACAGACCTCTCTCTCTTGGCTAGAATAACAACGGCTCTTTTCAGAGCCAC CAAATAATCAAGAAAGAAATACTGTTGTATGTT POST H4

```

Figure 3. Comparisons of DNA sequences downstream from the termination codons of all five histone genes in clone h19. The arrows underneath indicate the dyad symmetry.

high sequence homology of 23 bp. After a short gap of 6 bases the near identity extends for another 10 bp. The sequences between the termination codon and the first block of homology reveals nothing unusual in its composition, except for a slight underrepresentation of G. The six nucleotides separating the two homology regions show a preponderance of A's and T's and an absolute lack of G's, while the sequence following the second homology block are typical of AT-rich spacer DNA.

Search for palindromic structures

Since dyad symmetries are frequently found in operator and termination regions directing transcription in prokaryotes we analyzed these homologous sequences for palindromic structures. A major dyad symmetry was encountered in the first homology block and its location is shown in Figure 3 by arrows underneath the common sequences. This dyad symmetry is found at the same location within all the first homology blocks. The non-homologous single bases in the first homology block are perfectly matched by a second base change further along the sequence, thus maintaining the dyad symmetry. It is therefore tempting to speculate that it is the dyad symmetry, rather than the sequence itself, which has been conserved in evolution. In all cases, excepting the H2A sequence, the TTTC group of the unmatched loop is complementary to the sequence AAAG of the second homology block, but the stability of such a structure must be considered to be negligible, unless stabilized by other factors within the living cell.

DISCUSSION

Perhaps the best examples to-date of regulatory sequences subjected to extensive sequence comparisons are the promoters for E.coli polymerase. They also serve to demonstrate the difficulties which may be encountered when trying to identify regulatory sequences on the basis of nucleotide sequence alone. In the case of E.coli polymerase promoters base substitutions are permitted and only by statistical summation of a great many promoters can one deduce a preferred sequence (11). This suggests that if there is a code governing this type of protein-DNA interactions it is likely to be degenerate.

In eukaryotes, similar attempts to identify relatively large areas preceding the 5S (12) or transfer RNA genes (F. Müller, thesis 1978, Zürich) as being important in polymerase III binding have remained unconvincing, especially since it has been shown that no discernable effect on RNA synthesis is produced by removing the sequences near the 5' end of a structural tRNA gene (13). Some of the regulatory signals would be expected to be transcribed into RNA. That secondary structures play a role in RNA protein interaction is well documented, e.g. the interactions of the aminoacyl synthetases with tRNA (14) and the binding of the S₈-ribosomal proteins to ribosomal RNA (15). Such secondary structures may also be important in the processing of RNA precursors of tRNA (16) and rRNA (17). This function is best documented in the case of early mRNA of the bacterial phage T₇ where two such processing signals have been identified and have been sequenced (18,19).

Sequence comparison studies amongst the prelude sequences lying upstream of each of the five structural histone genes of the clone h22 have as yet revealed no extensive homologies (2) with the exception, perhaps, of sequence similarities near the 5' termini of the histone mRNA (Chr. Hentschel, L. Kedes, personal communications) and of the TATAAA sequence motif some 102 and 97 bases removed from the initiator codons of the H2A and the H2B gene. This sequence has also been reported for many other prelude sequences of genes served by polymerase II (cf.

ref.20). The situation is in marked contrast to the extensive homologies observed in the sequences proximal to the 3' termini of the structural genes of clone h19 (Fig.3).

To test whether the homologies found were a general feature of histone genes, we analyzed the available sequences of the h22 clone (2) and those of *S.purpuratus* (3). A compilation of this analysis is shown in Figure 4. The 3' flanking regions of the H1, the H2A and the H2B gene of the h22 clone can be fitted to the general scheme. For this comparison, sequence analysis of the H1 has been extended (Ph.Bucher, unpublished results) and a base denoted as uncertain previously (2) has been identified.

The primary structure of the *S.purpuratus* histone genes is generally similar to that of clone h19 of *P.miliaris*, a finding which will be discussed in detail elsewhere. In *S.purpuratus* the 3' sequences near the H2A and H3 genes are clearly homologous to comparable *P.miliaris* sequences. The H2B region cannot, at first sight, be fitted into this compilation. For this reason, only the second sequence block(3) comprizing 10 bp was recognized as a unifying theme. It should be noted, however, that the H2B sequence in *Strongylocentrotus* was obtained from a "single" Hae III restriction site (GGCC) by sequencing in both

TGATTGACGCGCACTTTCCCATCTACCAAAACGGCTCTTTTCAGAGCCAC	CAATAATCAAGAAAGAACCTTGTCCAAATG	CLONE H19
TAGAGTTTGCTCCCGCAATCTTGAAACCTCAACGGCCCTTATCAGGCGCCAC	CAATTACTCAGAAAGAAATTGTTTCATTATG	POST H1
TAGACAGGTCATATCTGCTCTAATTGGACATACACGGCCCTTTTCAGGCGCCAC	CAATAATCAAGAAAGAAATGATATCCGTAATA	POST H2A
TAGAACCATCGGTACAGCATGTAGCCCATGACCGCATACACAACGGCTCTTTTCAGAGCCAC	CACAACCCCAAGAAAGAAATCACTGACTTCTAA	POST H2B
TAAAGTGAACAGACCTCTCTCTCTTGCGTAGAATAACAACGGCTCTTTTCAGAGCCAC	CAATAATCAAGAAAGAAATCTGTTGTATGTT	POST H3
		POST H4
TGATGTTGCACGCTCTACTCGTGCACACACACAACGGCTCTTTTCAGAGCCAC	CACATTTCACGAAAGAACCAATGCCTATTITT	CLONE H22
TAAATTTGTTTGCTACCTCTTGCACTCAACAACGGCC---		POST H1
	---GGCCAC CAACATCAAGAAAGAAATTGTTGCATAAATC	POST H2A
		POST H2B
TAGATAGAGTTTGCTCCCGCAATCTTGAAACCTCAACGGCCCTTATCAGGCGCCAC	CAATTACTCAGAAAGAAATTGTTTCATTATG	S. PURPURATUS
TAGACAGGTCATATCTGCTCTAATTGGACATAATACACGG	CCXXACAATAATCAAGAAAGAAATGATATCCGTAATA	POST H2A
TAGAACCATCGGTACAGCATGTAGCCCATGACCGCATACACAACGGCTCTTTTCAGAGCCAC	CACAACCCCAAGAAAGAAATCAATTGATTCTAA	POST H2B
		POST H3

Figure 4. Compilation of available 3' flanking sequences of structural histone genes in the sea urchin *P.miliaris* (clone h19, h22) and *S.purpuratus* (clone pSp2, pSp17).

directions. In view of the highly conserved nature of the sequences denoted above for all other histone genes it seems possible to us that there may be two Hae III sites near the H2B gene of the S.purpuratus clone (as is the case for the H2B sequence of clone H19) and that a small intervening sequence was in fact missed in the analysis. Once split into two halves, the H2B sequences of Strongylocentrotus fit the general pattern well enough.

If the first homology block is transcribed into RNA, the RNA in this region would be expected to form a hairpin loop. These loops would fall into two classes as shown in Fig.5a and b for the corresponding DNA sequences. One type of the loop (a) accompanying the H1, H3 and H4 genes in both sea urchins has a stability of -9.8 kcal (calculated for an RNA structure, ref.21). The other type (b) which is found near the H2A and the H2B gene has a stability of -15.2 kcal (21). There is at present some evidence, that in the sea urchins the histone genes are transcribed polycistronically (22,23) and that in in vitro experiments the primary transcripts are at least double the size of mature mRNA (24). There is also some preliminary evidence that the 3' end of the mature H2A mRNA may include part or all of the palindromic sequence (J.Telford, unpublished results). Whatever the exact nature of the primary transcript, it appears likely that the palindromic sequences are transcribed into RNA.

In summary, therefore, two highly conserved nucleotide blocks spaced by six nucleotides have been found to occur in

Figure 5. Hairpin loops that can be formed via the dyad symmetry in the 3' flanking DNA sequences of the histone genes; a) H1, H3, H4 specific loop, b) H2A, H2B specific loop. If transcribed, the RNA would form the same secondary structure.

the 3' non-coding regions of all five types of histone genes from the variant DNA clone hl9. They also occur in histone genes of the majority type clone of Psammechinus miliaris (h22) and are also found in histone genes of another sea urchin species (3). A literature search has revealed that these sequences are not a general feature of all mRNA genes. None of the histone mRNA genes sequenced to-date code the AAUAAA motif found ubiquitously near the 3' end in other eukaryotic mRNA's (25).

The finding that closely similar sequences occur adjacent to five different histone genes in two sea urchin species is a strong indication that these sequences play an important role in the activity of these genes. Moreover, since the sequences have been conserved with remarkable accuracy, it is clear that they must function in some system which does not permit sequence degeneracy of the type exemplified by the E.coli polymerase-promoter interactions. We therefore feel that we are most probably dealing with a signal which acts via an RNA structure as would be the case, if, for instance, these signals were active in the maturation or termination of histone DNA transcripts.

ACKNOWLEDGEMENT

We wish to thank Heidi Schernitzki for excellent technical assistance and Fritz Ochsenbein for preparing the graphs. We are grateful to Dr. M. Chipchase for critical reading of the manuscript. This work was supported by a grant from the State of Zürich and by the Swiss National Research Council, grant No. 3.O66.076 and No.3.257.077.

REFERENCES

1. Brandt, W.F., Strickland, N.W., Strickland, M., Carlisle, L., Woods, D. and Von Holt, C. (1979) Eur.J.Biochem. 94, 1-10
2. Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H.O. and Birnstiel, M.L. (1978) Cell 14, 655-672
3. Sures, I., Lowry, J. and Kedes, L.H. (1978) Cell 15, 1033-1044
4. Clarkson, S.G., Smith, H.O., Schaffner, W., Gross, K.W. and Birnstiel, M.L. (1976) Nucl.Acids Res. 3, 2617-2632
5. Birnstiel, M.L., Schaffner, W. and Smith, H.O. (1977) Nature 266, 603-607

6. Maxam, A.M. and Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* 74, 560-564
7. Boseley, P., Moss, T., Mächler, M., Portmann, R. and Birnstiel, M. (1979) *Cell* 17, 19-31
8. Wienand, U., Schwarz, Z. and Feix, G. (1979) *FEBS Letters* 98, 319-323
9. Peacock, A.C. and Dingman, C.W. (1967) *Biochemistry* 6, 1818-1827
10. Laskey, R.A. and Mills, A.D. (1977) *FEBS Letters* 82, 314-316
11. Scherer, G.E.F., Walkinshaw, N.D. and Arnott, S. (1978) *Nucl. Acids Res.* 5, 3759-3773
12. Korn, L.J. and Brown, D.D. (1978) *Cell* 15, 1145-1156
13. Telford, J.L., Kressman, A., Koski, R.A., Grosschedl, R., Müller, F., Clarkson, S.G. and Birnstiel, M.L. (1979) accepted in *Proc. Natl. Acad. Sci. USA*
14. Rich, A. and Schimmel, P.R. (1977) *Nucl. Acids Res.* 4, 1649-1665
15. Zimmermann, R.A., Mackie, G.A., Muto, A., Garrett, R.A., Ungewickell, E., Ehresmann, C., Stiegel, P., Ebel, J.P. and Fellner, P. (1975) *Nucl. Acids Res.* 2, 279-302
16. Smith, J.D. (1976) *Prog. Nucl. Acids Res. Mol. Biol.* 16, 25-74
17. Nikolaev, N., Birge, C.H., Gotoh, S., Glazier, K. and Schlessinger, D., in *Brockhaven Symposia in Biology*, Vol. 26 Dunn, J.J., Ed. (1974) pp. 175-193
18. Robertson, H.D., Dickson, E. and Dunn, J.H. (1977) *Proc. Natl. Acad. Sci. USA* 74, 822-826
19. Robertson, H.D. in *Nucleic Acid and Protein Recognition*, Vogel, H.S., Ed., Academic Press N.Y. (1977) pp. 549-567
20. Gannon, F., O'Hare, K., Perrin, F., LePennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P. (1979) *Nature* 278, 428-434
21. Tinoco, I., Borer, N.P., Dengler, B. and Levine, M.D. (1973) *Nat. New Biol.* 246, 40-41
22. Kunkel, N.S., Hemminki, K. and Weinberg, E.S. (1978) *Biochemistry* 17, 2591-2598
23. Spinelli, G., Melli, L., Arnold, E., Casano, C., Gianguzza, F. and Caccio, M. submitted to *J. Mol. Biol.*
24. Levy, S., Childs, G., and Kedes, L. (1978) *Cell* 15, 151-162
25. Proudfoot, N.J. and Brownlee, G.G. (1976) *Nature* 263, 211-214