PRIMARY STRUCTURE OF THE HISTONE H3 AND H4 GENES AND THEIR FLANKING SEQUENCES IN A MINOR HISTONE GENE CLUSTER OF *XENOPUS LAEVIS*

A. F. M. MOORMAN, P. A. J. DE BOER, R. T. M. DE LAAF, W. M. A. M. VAN DONGEN and O. H. J. DESTRÉE

Department of Anatomy and Embryology, University of Amsterdam, Mauritskade 61, 1092 AD Amsterdam, The Netherlands

Received 31 August 1981; revision received 3 November 1981

1. Introduction

The study of histone gene expression has proved to be particularly useful in obtaining insight into eukaryotic gene regulation [1]. The differential expression of the histone multigene family in oogenesis and early development is especially well documented [2,3]. We have focussed our work on histone gene expression in *Xenopus laevis*, since the study of oogenesis and development of this vertebrate species has given a wealth of information both at the molecular level, especially the regulation of specific genes, and the cellular level [4].

We have detected isolated and molecularly cloned a 5.8 kilobasepair *Eco*RI histone DNA fragment (X1-hi-1) from genomic DNA of *Xenopus laevis* [5]. This histone DNA fragment contains the genes for histones H3, H4, H2A and H2B, in this order. The 5.8 kilobasepair *Eco*RI histone DNA fragment is represented only once or twice in the total genomic DNA as part of a total genomic pool of 40–50 histone gene clusters [6].

Here, we report the nucleotide sequences of genes coding for histones H3 and H4 and their surrounding regions containing putative regulatory sequences. The H3 coding sequence shows 18% base substitutions, resulting in 6 amino acid changes, compared with a gene for histone H3 from the sea urchin *P. miliaris* (clone h19). The H4 gene has 21% nucleotide divergence resulting in only one amino acid substitution compared with the H4 gene in h19 [7]. The 3′-flanking regions of both the H3 and H4 genes display a striking homology with those of *P. miliaris* histone genes in h19, especially with regard to the palindromic structure GGCTCTTTTCAGAGCC. The 5′-flanking regions share with other histone genes several regions of homology.

S1 mapping experiments indicate that at least the H3 gene present in the clone X1-hi-1 is not expressed during oogenesis and early embryogenesis.

2. Materials and methods

2.1. *Preparation of ^{32}P end-labeled DNA fragments*

The preparation of *X. laevis* histone DNA (X1-hi-1), the isolation and purification of the 5.8 kilobasepair *Eco*RI fragment and conditions for restriction endonuclease digestions were as in [5]. Restriction fragments were recovered from agarose, purified by DEAE-cellulose chromatography (Whatman DE-52) and ^{32}P-labeled at their 5′-termini, essentially as in [8]. Optimal incorporation of radioactivity (10^6 cpm (Cerenkov)/pmol DNA fragment) was reached with 0.4 μM [γ-^{32}P]ATP (Amersham) or 25 μCi/20 μl kinase reaction mix. Use of higher ATP concentrations resulted in considerable loss of incorporation. Some DNA fragments were purified by lysine–Sepharose chromatography. Lysine was coupled to CNBr-activated Sepharose 4B (Pharmacia) according to the conditions of the supplier. DNA was applied on the column in 10 mM Tris–HCl (pH 8.0), 1 mM EDTA and eluted in 0.3 M NaCl.

2.2. *DNA sequencing*

The sequencing procedure used was that in [9]. To obtain reproducible yields of DNA in the base-specific reactions we used twice the suggested concentration of tRNA in the stop mixes and shook the tubes during the ethanol precipitations. To eliminate residual piperidine the second lyophilization step was changed: the DNA was dissolved in 0.1 ml water, heated for 10 min at 90°C and lyophilized. The electrophoresis

Published by Elsevier/North-Holland Biomedical Press
00145793/81/0000–0000/$02.75

sample mix contained: 7 M urea, 5 mM Tris–borate (pH 8.3), 0.1 mM EDTA, 0.025% bromophenol blue and 0.025% xylene cyanol FF.

2.3. *Hybridization of RNA and S1 mapping*

RNA from oocytes or gastrula stage embryos [10] was isolated, electrophoresed on 2% agarose gels and transferred to diazotized paper as in [3]. The RNAs were hybridized with the 620 basepair *Eco*RI × *Bam*HI fragment containing the prelude and coding sequences of the gene for histone H3 and with the 320 basepair *Eco*RI × *Hae*III fragment containing the prelude sequences of the gene for histone H3. Labeling of these fragments by nick-translation with ^{32}P-labeled precursors and hybridization was as in [3].

S1 mapping of DNA–mRNA hybrids was essentially as in [11]. 10 ng of end-labeled 620 basepair *Eco*RI × *Bam*HI fragment was hybridized over 5 h to 10 μg total oocyte or gastrula RNA using the conditions in [3]. Hybrids were treated with 20 units S1 nuclease (Sigma) for 30 s at 37°C. S1 resistant DNA was electrophoresed on 2% agarose gels. Autoradiography was at −20°C on Kodak XRI films using intensifying screens.

3. Results and discussion

3.1. *Sequencing strategy*

The histone gene organization on the 5.8 kilobasepair *Eco*RI fragment (X1-hi-1) is given in fig.1A. The location of the histone genes has been established by hybridization [5] with purified gene probes from *P. miliaris* h22-DNA [12] and by DNA sequencing (this paper and A. F. M. M. et al., unpublished). Fragments used for DNA sequencing are indicated in fig.1B. Most (90%) of the sequences presented have been determined twice or more.

3.2. *H3 and H4 coding sequences*

The complete sequence of the genes coding for histones H3 and H4 has been determined (fig.2,3). In contrast to most eukaryotic genes but analogous to histone genes from other organisms [12] the coding sequences do not contain introns. Compared to the histone genes in a cloned *P. miliaris* DNA fragment (h19) there is 18% nucleotide divergence for H3 and 21% for H4. Part of the sequences of a H4 cDNA clone made against *X. laevis* oocyte H4 mRNA has been published [15]. This partial sequence differs

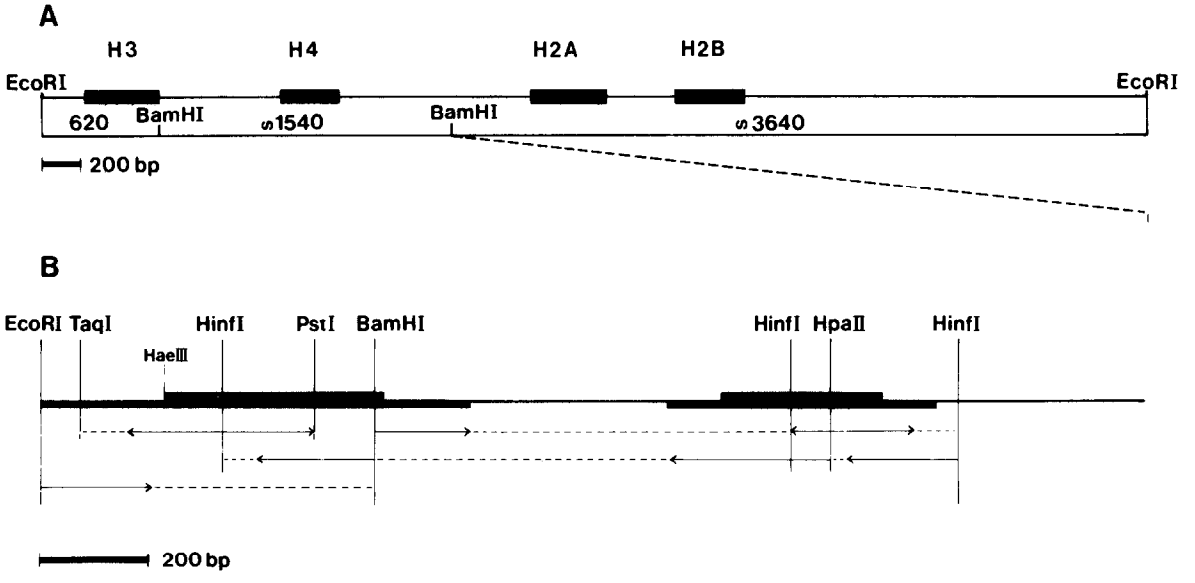

Fig.1. Schematic map of the histone genes of a cloned *X. laevis* histone DNA segment (X1-hi-1) and diagram of restriction fragments used for sequencing of the H3 and H4 regions. (A) Thick lines indicate the location of the coding region of the histone genes on the 5.8 kilobasepair *Eco*RI fragment; numbers indicate the fragment lengths in basepairs of *Bam*HI restriction fragments. (B) Enlargement of the DNA segment comprising the H3 and H4 coding regions. Thick lines indicate the DNA sequences presented in this paper. Arrows indicate singular sequenced stretches; dashed lines, extending from these arrows indicate the rest of the unique and labeled fragments. The *Hae*III site, used in the experiment of fig.6, is indicated also.

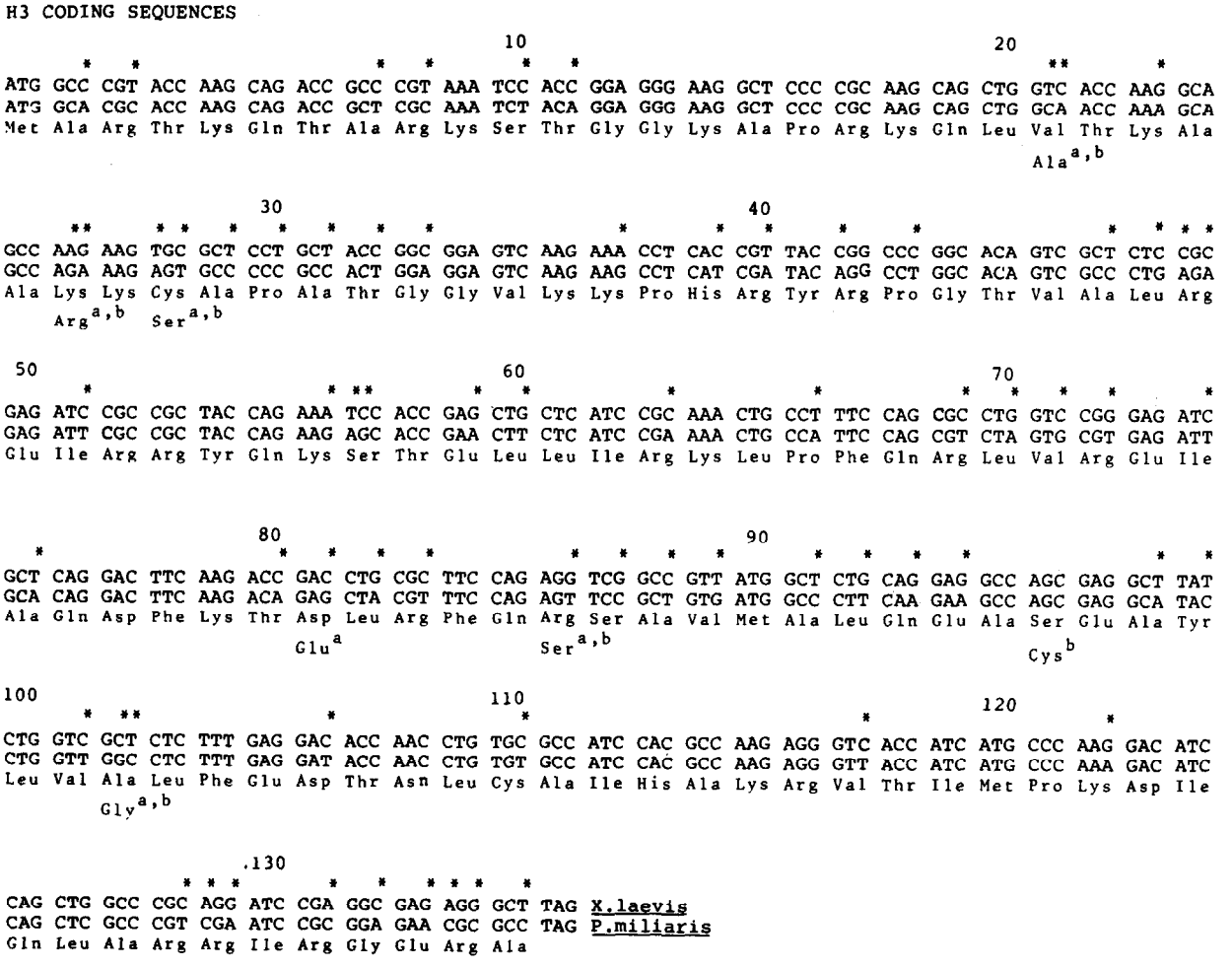

Fig.2. Nucleotide sequence of the *X. laevis* histone H3 gene as compared to that of *P. miliaris* [7]. The nucleotide sequence of the 'sense' strand is displayed in the 5′ → 3′ orientation. Asterisks indicate nucleotide differences between the *X. laevis* and *P. miliaris* H3 genes. The amino acid sequence predicted from the nucleotide sequence is indicated together with amino acid substitutions as compared to (a) *P. miliaris* [7] and (b) calf thymus [13].

from the sequence that we have determined at 10% of the nucleotide bases, without having an effect on the amino acid sequence. Several possibilities can explain this nucleotide divergence:

(i) The animals used, might have originated from different populations in which nucleotide substitutions resulting in synonymous coding changes could have occurred very rapidly if these substitutions were selectively neutral or near neutral in the course of evolution;

(ii) The two H4 sequences might belong to different gene clusters, that are expressed in different developmental stages or in different tissues.

This divergence in degenerate base use within one species is not exceptional: the DNA sequence divergence between the coding regions in h19 and h22, two histone clones from the same species of *P. miliaris*, is 12.4% [7]. It may be due to distinction in expression at the translational level, although this has been demonstrated only in prokaryotes [19]. Histones H3 and H4 display an exceptional evolutionary stability [16] that is believed to reflect the importance of the complete amino acid sequence of these proteins in controlling the expression of genetic information. Thus it may not be surprising that we have found that only a few amino acid substitutions have ocurred in

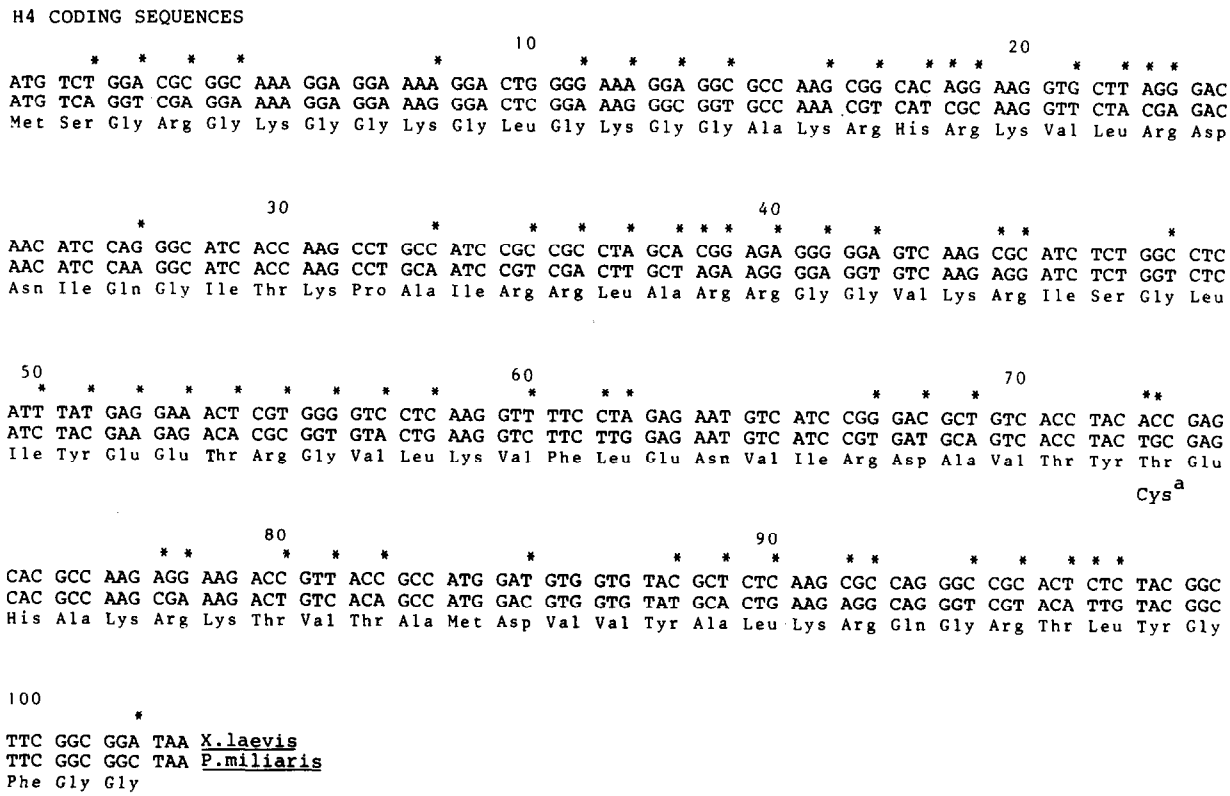

Fig.3. Nucleotide sequence of *X. laevis* histone H4 gene as compared to that of *P. miliaris* [7] (see fig.2). One amino acid substitution as compared to calf thymus [14] is indicated (a).

the 500×10^6 years, since the evolutionary lines resulting in the present-day sea urchin and *Xenopus* diverged. The derived amino acid sequence of *X. laevis* H3 differs from that of sea urchin [7] calf, carp, shark and chicken at position 21 (Val ↔ Ala), 26 (lys ↔ arg), 28 (Cys ↔ Ser), 86 (Arg ↔ Ser) and 102 (Ala ↔ Gly) [16]. At position 81 there is a (Asp ↔ Glu) exchange compared to sea urchin and at position 96 a (Ser ↔ Cys) exchange has occurred compared to calf. Some of these amino acid substitutions may be considered 'non-conservative' (Arg ↔ Ser; Ser ↔ Cys). However, the effect of these amino acid substitutions on the protein structure has to be established.

Cysteine at position 110 has been preserved throughout evolution. This cysteine might participate in the formation of a disulfide bridge with a second histone H3 molecule in a nucleosomal particle [17]. The H3 proteins of higher eukaryotes display heterogeneity (Cys ↔ Ser) at position 96 even within one species [18]. In contrast to calf thymus H3, the *X. laevis* H3 amino acid sequence has a serine at position 96 but a cysteine at position 28.

The derived amino acid sequence of *X. laevis* H4 (fig.3) is completely identical to that determined for calf [14] and other vertebrates [16] and has only one amino acid exchange compared to sea urchin at position 73 (Thr ↔ Cys).

3.3. *Codon usage*

Both in prokaryotes and in eukaryotes triplet analysis has shown that synonymous codons are used nonrandomly [19]. This is also the case in *Xenopus*. The codon usage of *Xenopus* genes for H3, H4 and globin and the *P. miliaris* genes for H3 and H4 is given to permit inter- and intra-species comparisons (table 1). Several points regarding the codon usage are worth emphasizing.

(1) G and C are preferred to a high extent at codon position 3. The G + C content of the *Xenopus* H3

Table 1

Codon usage of *Xenopus laevis* histone H3 and H4 in comparison with that of *Xenopus* globin and *P. miliaris* histone H3 and H4

	Xenopus			h19			Xenopus			h19			Xenopus			h19			Xenopus			h19	
	H3	H4	Hb	H3	H4		H3	H4	Hb	H3	H4		H3	H4	Hb	H3	H4		H3	H4	Hb	H3	H4
Phe UUU	1	0	5	1	0	Ser UCU	0	2	3	1	1	Tyr UAU	1	1	6	0	1	Cys UGU	0	0	0	1	0
UUC	3	2	5	3	2	UCC	2	0	1	1	0	UAC	2	3	3	3	3	UGC	2	0	1	0	1
Leu UUA	0	0	0	0	0	UCA	0	0	1	0	1	term UAA						term UGA					
UUG	0	0	4	0	2	UCG	1	0	0	0	0	term UAG						Trp UGG	0	0	0	0	0
Leu CUU	0	1	4	2	1	Pro CCU	3	1	3	2	1	His CAU	0	0	6	1	1	Arg CGU	3	1	2	4	4
CUC	3	4	2	3	2	CCC	3	0	2	3	0	CAC	2	2	7	1	1	CGC	8	6	1	7	2
CUA	0	2	2	2	1	CCA	0	0	1	1	0	Gln CAA	0	0	1	1	1	CGA	1	0	0	3	4
CUG	9	1	8	5	2	CCG	0	0	0	0	0	CAG	8	2	3	7	1	CGG	2	3	3	0	0
Ile AUU	0	1	2	2	0	Thr ACU	0	2	2	1	1	Asn AAU	0	1	3	0	1	Ser AGU	0	0	3	2	0
AUC	7	5	2	5	6	ACC	9	5	3	6	2	AAC	1	1	6	1	1	AGC	1	0	5	2	0
AUA	0	0	2	0	0	ACA	1	0	0	3	3	Lys AAA	4	3	2	4	2	Arg AGA	0	1	2	2	1
Met AUG	3	2	3	3	2	ACG	0	0	0	0	0	AAG	10	8	9	9	9	AGG	4	3	0	2	3
Val GUU	1	2	4	2	1	Ala GCU	9	2	8	3	1	Asp GAU	0	1	6	1	1	Gly GGU	0	0	4	0	6
GUC	6	4	1	2	5	GCC	8	4	10	10	3	GAC	4	2	7	2	2	GGC	3	7	3	2	5
GUA	0	0	1	0	1	GCA	1	1	2	5	3	Glu GAA	0	1	4	3	1	GGA	2	7	4	4	6
GUG	0	3	4	2	2	GCG	0	0	2	0	0	GAG	7	3	3	5	3	GGG	1	3	1	1	0

Frequencies of use for each codon are given for the genes for *X. laevis* (X1-hi-1) histones H3 and H4, as presented here, for *X. laevis* globin gene [20] and for the genes for *P. miliaris*, histone H3 and H4 (h19). Frequencies that are boxed represent frequencies of dispensable pre-termination codons, double-boxed frequencies represent those of obligatory pre-termination codons [21]. Pre-termination codons are those codons that can mutate to a termination codon by a single nucleotide change. If they specify amino acids that are also encoded for by triplets that cannot mutate to a termination codon by a single base substitution, they are called dispensable pre-termination codons. Those pre-termination codons that encode amino acids only specified by pre-termination codons are called obligatory pre-termination codons

coding sequence is 68% (58% in sea urchin) and is 60% (54% in sea urchin) for the H4 coding sequence. The frequency of usage of G and C as third base is 80% for H3 (62% in sea urchin) and 71% for H4 (57% in sea urchin). This 'overuse' of G and C at the third base position could reflect a general evolutionary tendency, since mammalian genes also show this tendency [19]. In addition, in mammalian genes the use of A in third base position is reduced to half that of U. In contrast, sea urchin H3 and H4 genes use A slightly more frequently than U as third base.

Grantham has postulated that each taxonomic category has a particular coding strategy, i.e., a particular choice of degenerate bases [19]. The *Xenopus* H3 and H4 genes over use G and C and under use A at the third position and, therefore could agree with the genome hypothesis, but a correspondence analysis will be necessary. However, in contrast, *Xenopus* globin (α- and β-partial sequences, see [20]) does not display an overuse of G and C at the third base position. So, clearly more sequences of higher taxonomic categories are necessary to establish the general validity of Grantham's genome hypothesis.

(2) In *Xenopus* H3 and H4 genes, codons with U or A at the third base position are rare or absent. This is the main difference in codon use compared with that in the *Xenopus* globin gene and in the *P. miliaris* H3 and H4 genes. There are two exceptions to this rule: codon GCU (Ala) and GGA (Gly) are relatively frequently used.

(3) A well known tendency in viral and eukaryotic sequences is the GpC over CpG excess [22]. This is

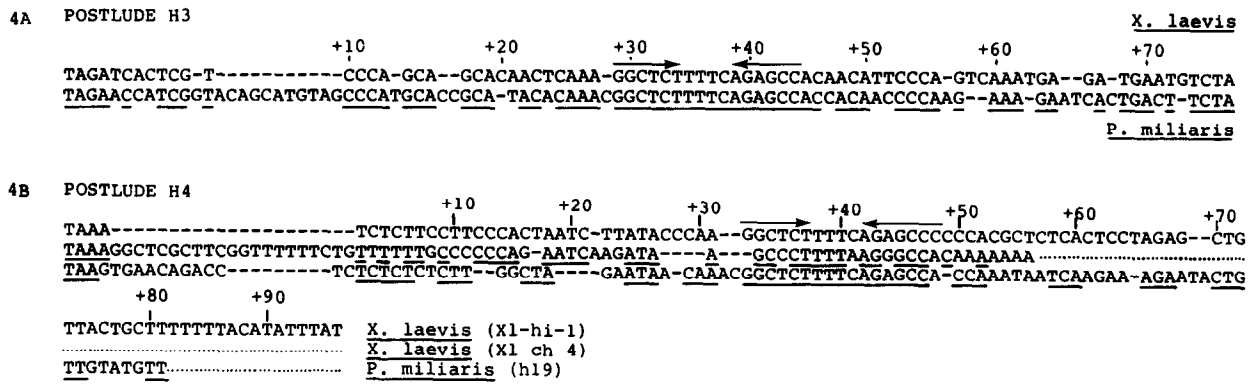

Fig.4. 3′-Flanking nucleotide sequence of *X. laevis* X1-hi-1 H3 (A) and H4 (B) compared with those of *P. miliaris* [7] and *X. laevis* (X1-ch-4) [15]. Sequences are aligned for maximal homology. Homologous stretches are underlined. The palindromic sequence is indicated by →.

also the case for the *Xenopus* H3 and H4 genes: CpG is used only once at position 2–3, while GpC is used 27 times at this position. However, there is no significant difference in use of CpG and GpC at position 1–2 or 3–1. This is particularly clear for the arginine codon sextet, consisting of one codon quartet (CGU, CGC, CGA and CGG) and a codon duet (AGA and ACG). The duet codons are not preferred compared to the quartet codons although the quartet codons have CpG at codon position 1–2. Why CpG at codon position 2 and 3 is nearly completely forbidden, is as yet unclear.

(4) The so-called dispensable pre-termination codons (see table 1) are not used at all in human α- and β-globin genes. This is obviously not the case for the *Xenopus* H3, H4 and globin genes, although they are used rarely, except for the codon GGA (Gly).

3.4. *3′-Flanking sequence (postlude)*

Fig.4 shows the 3′-flanking sequences of *X. laevis* histone genes. They are compared to *P. miliaris* and for H4 also compared to *X. laevis* H4 (X1-Ch-4). If aligned for maximum homology by introducing some deletions an impressive level of homology can be found (up to 80% for the H3 postlude). As for the sea urchin histone genes [7] about 30–40 nucleotides downstream from the terminator codon, a conserved homology block can be found. This block consists of the palindromic sequence GGCTCTTTTCAGAGCC, resembling putative polymerase III terminators [23]. The 3′-terminus of the mRNA has been located just downstream from this palindromic structure [15,24]. However, *S. purpuratus* H4 early and late gene transcripts do not contain this palindromic sequence [25]. Similar to the cloned sea urchin histone genes [7] and the *Xenopus* H4 cDNA [15] clone, the AAUAAA sequence, that is present on other eukaryotic mRNAs [26] and that possibly plays a role in polyadenylation, is not encoded for by the *Xenopus* histone H3 and H4 genes.

The biological significance of the other homologous sequences within the postlude sequence remains to be established.

3.5. *5′-Flanking sequences (prelude)*

The 5′-sequences preceding the genes for H3 and H4 are indicated in fig.5. They display much more divergence compared to those of the cloned *P. miliaris* histone genes (h19) than the 3′-postlude sequences. Conserved regions or homology blocks that appear to be characteristic for histone genes have been tabulated and called 'consensus sequences' [1]. Such sequence motifs have been indicated tentatively in fig.5, based on homology with the 'consensus sequence'. A 'TATA box', that might play a role in promoting specific initiation of transcription by RNA polymerase II [27,28], can be clearly assigned. 'CCAAT', 'GATC', 'Cap' and 'CAPyNATG' motifs, partially homologous to the 'consensus sequence', can be assigned (function discussed in [1]). Experiments, testing the role of these sequence motifs and/or others in the function and expression of this cloned *Xenopus* histone DNA cluster are currently performed.

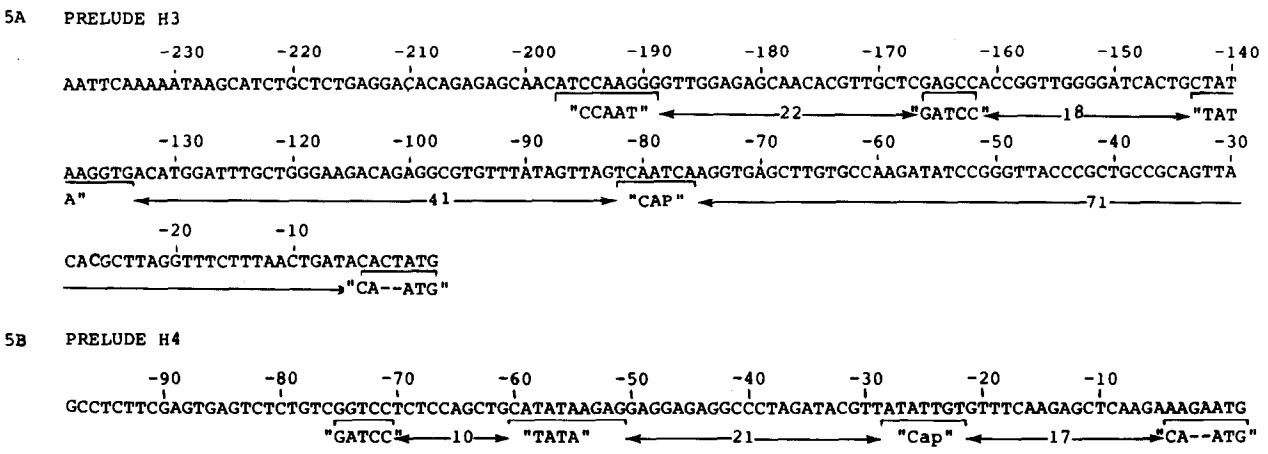

Fig.5. 5′-Flanking nucleotide sequences of *X. laevis* X1-hi-1 H3 (A) and H4 (B). Putative blocks of homology [1] are indicated.

3.6. *In vivo expression*

Expression of the genes present in the cloned histone DNA fragment during oogenesis and early embryogenesis was tested by homology mapping of hybrids between oocytes or embryo RNA and the end-labeled 620 basepair *Eco*RI × *Bam*HI fragment. If oocyte or embryo H3 mRNA form perfectly matched hybrids with this DNA-fragment a piece of DNA of 475 basepair, between the *Bam*HI site and the presumed 'Cap' site is expected to become protected against nuclease S1 action. However, it appears that not more than 390 ± 20 basepair are protected against the nuclease (fig.6A(c,d)) both by oocyte and gastrula RNA. This indicates that the prelude sequences are not protected by the mRNA. This point is reinforced by the experiment of fig.6B. Electrophoretically separated oocyte and gastrula RNA are hybridized with ^{32}P-labeled 620 basepair *Eco*RI × *Bam*HI fragment or with ^{32}P-labeled 230 basepair *Eco*RI × *Hae*III fragment (see fig.1B). This latter fragment contains besides non-coding sequences, the complete prelude of the H3 gene up to the codon for the first amino acid. As shown in fig.6B(c,d) hybridization of H3 mRNA cannot be detected with this fragment, in contrast to the fragment containing the H3 coding sequences (fig.6B(a,b)). We conclude from this experiment that the H3 present in the cloned DNA is not, or only at a very low level, expressed during oogenesis and early embryogenesis in *Xenopus*. This might imply that these genes are under developmental control and that other histone gene clusters must be expressed that contain different prelude sequences.

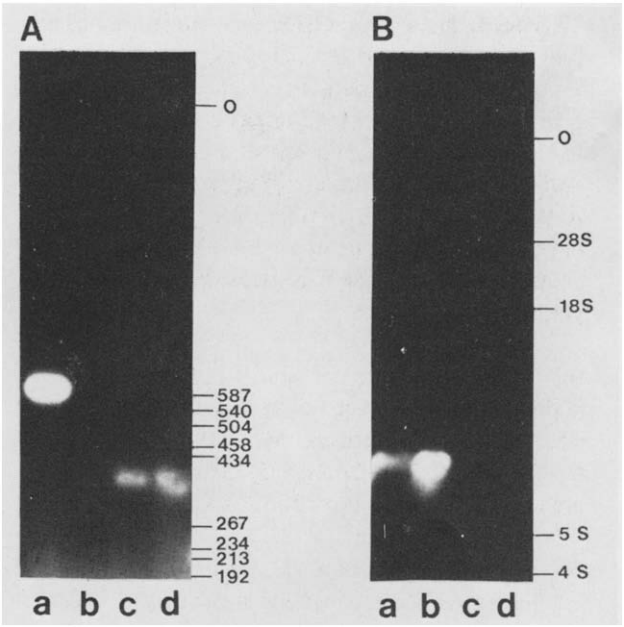

Fig.6. (A) S1 mapping of hybrids between the end-labeled 620 basepair *Eco*RI × *Bam*HI fragment and H3 mRNA. End-labeled 620 basepair fragment (a); fragments protected against nuclease S1 attack after hybridization of the 620 basepair fragment with wheat germ tRNA (b) oocyte RNA (c) and gastrula RNA (d). As marker a digest of PBR 322 DNA with *Hae*III is used. Their positions and length in basepair is indicated. O is origin. (B) Hybridization of electrophoretically separated oocyte RNA (b,d) and gastrula RNA (a,c) with ^{32}P-labeled 620 basepair *Eco*RI × *Bam*HI DNA fragment (a,b) or 230 basepair *Eco*RI × *Hae*III DNA fragment (c,d). The position of 4 S, 5 S, 18 S and 28 S RNA are indicated.

Indeed, we have found that the prelude sequence of the H3 gene from the clone X1-hi-1 does not hybridize to at least one of the H3 genes in a number of different recently cloned *Xenopus* histone genes.

Acknowledgements

We are grateful to the staff of the workshop of our institute for their diligent manufacturing of the DNA sequencing equipment. This work was supported in part by a grant to O.D. from the Netherlands Foundation for Fundamental Medical Research (FUNGO) with financial aid from the Netherlands Organization for the Advancement of Pure Research (ZWO) (grant 13-50-09).

References

[1] Hentschel, C. G. and Birnstiel, M. L. (1981) Cell 25, 301–313.
[2] Woodland, H. R. (1980) FEBS Lett. 121, 1–7.
[3] Van Dongen, W., Zaal, R., Moorman, A. and Destrée, O. (1981) Dev. Biol. 86, 303–314.
[4] Davidson, E. H. (1976) Gene activity in early development, 2nd edn, Academic Press, New York.
[5] Moorman, A. F. M., De Laaf, R. T. M., Destrée, O. H. J., Telford, J. and Birnstiel, M. L. (1980) Gene 10, 185–193.
[6] Van Dongen, W., De Laaf, L., Zaal, R., Moorman, A. and Destrée, O. (1981) Nucleic Acids Res. 9, 2297–2311.
[7] Busslinger, M., Portman, R., Irminger, J. C. and Birnstiel, M. L. (1980) Nucleic Acids Res. 8, 957–977.
[8] Clarkson, S. G., Kurer, V. and Smith, H. O. (1978) Cell 14, 713–724.
[9] Maxam, A. M. and Gilbert, W. (1980) Methods Enzymol. 65, 499–560.
[10] Nieuwkoop, P. D. and Faber, J. (1956) Normal table of *Xenopus laevis* (Daudin ed) North-Holland, Amsterdam.
[11] Berk, A. J. and Sharp, P. A. (1977) Cell 12, 721–732.
[12] Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H. O. and Birnstiel, M. L. (1978) Cell 14, 655–671.
[13] De Lange, R. J., Hooper, J. A. and Smith, E. L. (1973) J. Biol. Chem. 248, 3261–3274.
[14] De Lange, R. J., Fambrough, D. M., Smith, E. L. and Bonner, J. (1969) J. Biol. Chem. 244, 319–334.
[15] Zernik, M., Heinz, N., Boime, I. and Roeder, R. G. (1980) Cell 22, 807–815.
[16] Von Holt, C., Strickland, W. N., Brandt, W. F. and Strickland, M. S. (1979) FEBS Lett. 100, 201–218.
[17] Camerini-Otero, R. D. and Felsenfeld, G. (1977) Proc. Natl. Acad. Sci. USA 74, 5519–5523.
[18] Franklin, S. G. and Zweidler, A. (1977) Nature 266, 273–275.
[19] Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. and Mercier, R. (1981) Nucleic Acids Res. 9, r43–r74.
[20] Richardson, C., Cappello, J., Cochran, M. D., Armentrout, R. W. and Brown, R. D. (1980) Dev. Biol. 78, 161–172.
[21] Modiano, G., Battistuzzi, G. and Motulsky, A. G. (1981) Proc. Natl. Acad. Sci. USA 78, 1110–1114.
[22] Grantham, R. (1980) FEBS Lett. 121, 193–199.
[23] Bogenhagen, D. F. and Brown, D. D. (1981) Cell 24, 266–270.
[24] Hentschel, C., Irminger, J. C., Bucher, P. and Birnstiel, M. L. (1980) Nature 285, 147–151.
[25] Grunstein, M., Diamond, K. E., Knoppel, E. and Grunstein, J. (1981) Biochemistry 20, 1216–1223.
[26] Proudfoot, N. J. and Brownlee, G. G. (1976) Nature 263, 211–214.
[27] Grosschedl, R. and Birnstiel, M. L. (1980) Proc. Natl. Acad. Sci. USA 77, 1432–1436.
[28] Mathis, D. J. and Chambon, P. (1981) Nature 290, 310–315.