Primary structure of the histone H2A and H2B 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 and O.H.J. Destrée

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

Received 19 May 1982

Histone genes Xenopus laevis *Primary sequence* *Expression* *Evolution*

1. INTRODUCTION

The histone protein family comprises the principal structural proteins of eukaryotic chromatin [1]. This family forms an evolutionarily conserved group, reflecting its fundamental role in chromatin structure. Particularly interesting are tissue specific histone subtypes, whose synthesis is developmentally regulated [2], suggesting a specific role of these histone subtypes in the regulation of gene expression. In the cell cycle the bulk of histone synthesis is closely coupled to DNA synthesis [3] and nucleosome structure requires equimolar synthesis of nucleosomal histones and half equimolar synthesis of the H1 class of histones. An essential prerequisite for understanding the regulatory mechanisms operative during development and the cell cycle is insight into the organization of the genome and therefore into the expression of the histone gene family.

In contrast to the uniformity in organization of sea urchin and fruit fly histone genes, a hitherto unexpected variety of different histone gene organizations has been found in different vertebrate species. In sea urchins and fruit flies the majority of the histone genes is repeated and arranged in tandem units, each unit containing the genes for all 5 histones [4,5]. In addition, dispersed polarity histone genes, called orphons have been observed in these organisms [6]. In amphibians the histone genes are clustered but arranged in different ways in *Xenopus* [7] and *Notophthalmus* [8]. The histone genes of *Xenopus* are repeated 45–50 times [7,9]. They are partly (up to 30 copies) arranged in a repeating unit of 14 kilobasepairs, partly located on unique restriction fragments, in varying numbers and different from individual to individual [7].

In the chicken and human genomes the histone genes are clustered but not at all arranged in repeating units [10,11].

We have reported the cloning of a 5.8 kilobasepair genomal histone DNA fragment (Xi-hi-l) from *Xenopus laevis* [12], and have established the nucleotide sequences of the H3 and H4 genes including their flanking sequences [13]. This clone represents a unique histone cluster with a gene order different from that found in the major repeating unit [7]. This paper deals with the nucleotide sequences of the genes coding for histones H2A and H2B of Xl-hi-l including their 5′ and 3′ flanking sequences. It appears that the coding sequences for H3, H4 and H2B are on one strand while those for H2A are on the other. The derived amino acid sequences of histones H2A and H2B of *Xenopus* show more resemblance with the histone protein sequences in mammals than with those in sea urchin. In the 5′ flanking region a ‘TATA box’ can be assigned. The ‘CCAAT box’, present in other eukaryotic polymerase II genes, can be clearly recognized in the prelude sequence of the H2A gene, while the prelude sequence of the H2B gene contains this sequence probably in a different form. The 3′ flanking regions contain a very characteristic GC-rich palindromic structure that can be considered to be typical for histone genes [14].

2. METHODS

The construction of the genomic *Xenopus* histone clone (Xl-hi-l) has been published [12]. Isolation of

Published by Elsevier Biomedical Press
00145793/82/0000–0000/$2.75 © 1982 Federation of European Biochemical Societies

plasmid DNA and DNA-fragments, conditions for restriction endonuclease incubations and the conditions for 5′-terminal labeling of the restriction fragments were as described [12,13]. The DNA sequence analysis was according to [15] with some minor modifications [13]. Of the sequences presented 92% have been determined twice or more.

3. RESULTS AND DISCUSSION

3.1. *Order and polarity of the histone genes in Xl-hi-l DNA*

The arrangement of the 4 genes coding for the nucleosomal histones on the 5.8 kilobasepair cloned DNA of *Xenopus laevis* (Xl-hi-l) is shown in fig.1. The location of the individual histone genes was established by hybridization with individual gene probes derived from cloned *Psammechinus miliaris* histone DNA [16] and by DNA sequence analysis [12]. The presence of an H1 gene could not be established by cross-hybridization with a specific *P. miliaris* H1 probe or by hybridization translation experiments. Sequence analysis ([12,13], this paper) revealed that the genes for histone H3, H4 and H2B have the same polarity but that the H2A gene is of different polarity. Gene order and/or polarity is different in other species [8,16,18]. Surprisingly the order of the histone genes in Xl-hi-l (i.e., H3–H4–H2A–H2B) is different from that in major repeating unit (i.e., H4–H3–H2A–H2B) found in genomic blots of *Xenopus laevis* from our laboratory population, but is identical with that found in another cloned *Xenopus* histone DNA fragment [17]. However, in this case the polarity of the genes has not been reported. This is the first case showing that the order and/or polarity of the histone genes can be different in different histone gene clusters within one species.

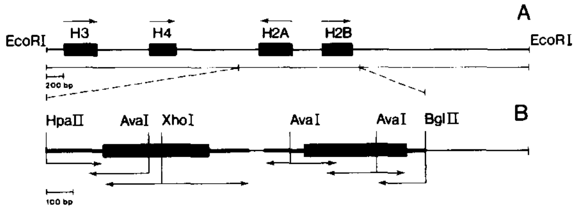

Fig.1. (A) Organization of the *X. laevis* histone genes of clone Xl-hi-l. Arrows indicate the polarity of the genes (5′→3′). (B) Enlargement of the H2A and H2B region indicating the fragments used for sequencing. Arrows indicate stretches only once sequenced. Thick lines indicate the sequences presented in this paper. Bars indicate coding regions.

3.2. *H2A coding sequences*

The nucleotide sequence of the complete coding region of histone H2A has been determined and compared with that of H2A in *P. miliaris* (h19) [14] (fig.2). Compared to *P. miliaris* (h19) the Xl-hi-l H2A coding region is 18 nucleotides longer. This implies that the protein encoded for by this gene has exactly the same length as its mammalian counterpart [19] i.e., 129 amino acids. Besides this increase in length, a 24% basepair difference has accumulated during evolution compared to *P. miliaris*, resulting in 10 (8%) amino acid changes. Compared to the mammalian histone H2A, 8 amino acid changes have occurred as indicated in fig.2. Two amino acid changes in the Xl-hi-l histone H2A appear to be unique for *Xenopus*: a Thr ↔ Ala exchange at position 10 and a Phe ↔ Ala exchange at position 113.

Both in sea urchin and vertebrate species 2 histone H2A variants, containing either methionine or leucine at position 51, occur [2]. Whether this holds also for *Xenopus* has to be tested rigorously. We have indications that H2A (or at least material comigrating with H2A in acid–urea–Triton gels) from *Xenopus* embryos is labeled in vivo with [^{35}S]methionine (Bisschops, C. et al., unpublished).

3.3. *H2B coding sequences*

The complete nucleotide sequence of the H2B coding region is given in fig.3 in comparison with that of *P. miliaris* (h19) [14]. The derived amino acid sequence has also been compared with calf thymus H2B [20]. Besides an increase in length (the H2B coding region from Xl-hi-l is 9 nucleotides longer) 28% basepair substitutions have occurred, resulting in 28 (23%) amino acid substitutions compared to *P. miliaris* (h19). Compared to calf thymus H2B, 12 amino acid changes have occurred. Two of these amino acid substitutions, Ala ↔ Pro and Ala ↔ Val at positions 10 and 18, respectively, are different from the partial amino acid sequence determined for *Xenopus* erythrocyte histone H2B [21]. However, a number of H2B proteins of other species, among which different urchin species, *Drosophila* and *Patella* [22], also have alanine at one of these positions.

```
H2A CODING SEQUENCES
                                        10                                      20
          *       *   *     * *         * *     * *               *           *   *   *       *
ATG TCT GGA AGA GGC AAA CAA GGC GGC AAG ACT CGC GCT AAG GCA AAG ACT CGC TCA TCT CGG GCC GGG CTG CAG
ATG TCT GGC AGA GGA AAG --- AGT GGA AAG GCC CGC ACC AAG GCA AAG ACG CGC TCA TCC CGT GCA GGG CTC CAG
    Ser Gly Arg Gly Lys Gln Gly Gly Lys Thr Arg Ala Lys Ala Lys Thr Arg Ser Ser Arg Ala Gly Leu Gln
                        --- Ser         Ala     Thr
                                        Ala

                    30                                      40
  *       *   *           *     * * * * * *           *       * *   *     *   *  **           *
TTC CCA GTC GGC CGT GTT CAC CGG CTC TTG AGG AAG GGC AAT TAT GCC GAG CGG GTG GGA GCC GGA GCT CCG GTC
TTT CCA GTG GGA CGT GTT CAT CGG TTT CTC CGA AAG GGC AAC TAT GCA AAG AGG GTC GGC GGT GGA GCT CCT GTC
Phe Pro Val Gly Arg Val His Arg Leu Leu Arg Lys Gly Asn Tyr Ala Glu Arg Val Gly Ala Gly Ala Pro Val
                                Phe                             Lys             Gly

50                                      60                                      70
  * *     *   *       *       *   *   *   *   *           * * *   *               *   *   *       *
TAT CTG GCC GCA GTG CTC GAG TAT CTG ACC GCT GAG ATC TTG GAG TTG GCC GGC AAC GCT GCT CGG GAT AAC AAA
TAC ATG GCT GCC GTC CTA GAG TAC CTC ACT GCC GAA ATC TTG GAA CTC GCA GGC AAC GCT GCC CGC GAC AAC AAG
Tyr Leu Ala Ala Val Leu Glu Tyr Leu Thr Ala Glu Ile Leu Glu Leu Ala Gly Asn Ala Ala Arg Asp Asn Lys
    Met

                    80                                      90
  * * * * *           * * *       *   *               *   *           *           *   * * *   *   *
AAG ACC CGC ATC ATC CCC AGG CAC CTG CAG CTC GCT GTG CGC AAC GAT GAG GAG CTC AAC AAA CTG CTC GGA GGA
AAA TCT AGG ATC ATC CCA CGC CAC CTT CAA CTC GCT GTG CGT AAT GAT GAA GAA CTC AAC AAG CTT TTG GGT GGG
Lys Thr Arg Ile Ile Pro Arg His Leu Gln Leu Ala Val Arg Asn Asp Glu Glu Leu Asn Lys Leu Leu Gly Gly
    Ser
                                                Ile                                             Lys

100                                     110                                     120
  *   *           *   *   *                   *   * **            *               *
GTC ACT ATC GCT CAG GGC GGG GTT CTG CCC AAC ATT CAG TTC GTG CTG CTG CCC AAG AAA ACC GAG AGC TCC AAG
GTG ACG ATC GCT CAA GGT GGT GTT CTG CCC AAC ATC CAA GCC GTG CTG CTT CCC AAG AAA ACT --- --- --- ---
Val Thr Ile Ala Gln Gly Gly Val Leu Pro Asn Ile Gln Phe Val Leu Leu Pro Lys Lys Thr Glu Ser Ser Lys
                                                    Ala                             --- --- --- ---
                                                    Ala                                     His His

      *   * ***  **
TCG GCC AAG AGC AAG TGA X. laevis
--- GCT AAA TCA AGC TAG P. miliaris
Ser Ala Lys Ser Lys     X. laevis
---             Ser     P. miliaris
Lys         Gly         B. taurus
```

Fig.2. Nucleotide sequence of the *X. laevis* histone H2A gene as compared to that of *P. miliaris* [14]. The nucleotide sequence of the sense strand is displayed in the 5'→3' orientation. (*) Nucleotide differences between the *X. laevis* and *P. miliaris* H2A genes. The amino acid sequence derived from the nucleotide sequence is indicated together with amino acid substitutions as compared to the sequences for *P. miliaris* [14] and calf [19].

Thus the importance of these substitutions remains to be evaluated. The lysine at position 31, threonine at position 32 and asparagine at position 85 have not been found in any other species [22]. The other amino acid changes compared to *P. miliaris* and calf thymus histone H2B are also found in H2B from other species [22].

3.4. *Codon usage*

As is evident from table 1 synonymous codon usage in the H2A and H2B genes in Xl-hi-1 is non-random.

A number of codons is not used at all both in the H3 and H4 genes [13] and in the H2A and H2B genes. They comprise AUA (Ile), GUA (Val), ACG (Thr), AGU (Ser) and GGU (Gly). The codons CUU (Leu) and CUA (Leu) are not used in the H2A and H2B genes.

From the high GC content of the H2A and H2B genes, 60% and 57% GC, respectively, it is expected that codons ending in C and G are preferred. This is indeed the case: 72% and 77% of the H2A and H2B codons, respectively, have G or C at the third codon position. This is not necessarily the consequence of

Table 1

Frequency of each codon in the *X. laevis* histone H2A and H2B genes

		H2A	H2B			H2A	H2B			H2A	H2B			H2A	H2B
Phe	UUU	0	1	Ser	UCU	2	1	Tyr	UAU	3	1	Cys	UGU	0	0
	UUC	2	0		UCC	1	8		UAC	0	4		UGC	0	0
Leu	UUA	0	1		UCA	1	0	Term	UAA	–	–	Term	UGA	–	–
	UUG	3	1		UCG	1	1		UAG	–	–	Trp	UGG	0	0
Leu	CUU	0	0	Pro	CCU	0	2	His	CAU	0	1	Arg	CGU	1	0
	CUC	5	0		CCC	3	1		CAC	2	2		CGC	4	5
	CUA	0	0		CCA	1	2	Gln	CAA	1	0		CGA	0	1
	CUG	8	4		CCG	1	0		CAG	4	3		CGG	4	0
Ile	AUU	1	1	Thr	ACU	3	1	Asn	AAU	1	0	Ser	AGU	0	0
	AUC	4	5		ACC	3	8		AAC	5	4		AGC	2	2
	AUA	0	0		ACA	0	1	Lys	AAA	4	5	Arg	AGA	1	0
Met	AUG	1	3		ACG	0	0		AAG	9	14		AGG	2	2
Val	GUU	2	0	Ala	GCU	7	3	Asp	GAU	2	2	Gly	GGU	0	0
	GUC	3	3		GCC	6	8		GAC	0	1		GGC	7	3
	GUA	0	0		GCA	2	3	Glu	GAA	0	2		GGA	5	0
	GUG	4	5		GCG	0	2		GAG	7	5		GGG	2	3

```
H2B CODING SEQUENCES
                                        10                                      20
                  * *** * *      *   **   *                       *   *  **  ** *** *    *  ***  **  **
ATG CCT GAA CCA GCC AAG TCC GCT CCA GCC GCA AAG AAA GGC TCC AAG AAA GCG GCA ACC AAG ACT CAG AAG AAA GAC
ATG --- --- --- GCT CCA ACA GCT CAA GTT GCT AAG AAA GGC TCC AAG AAG GCA GTC AAG GCC CCT CGG CCC AGC GGT
    Pro Glu Pro Ala Lys Ser Ala Pro Ala Ala Lys Lys Gly Ser Lys Lys Ala Ala Thr Lys Thr Gln Lys Lys Asp
                    Pro Thr     Gln Val                                 Val Lys Ala Pro Arg Pro Ser Gly
                                        Pro                             Val         Ala

                30                                      40                                      50
  *   *     * *  **  *   *                       **   *     * *       *   *   *           *   *   *
GGG AAA AAG CGC AGG AAG ACA AGG AAG GAG AGT TAT GCC ATT TAC GTG TAC AAG GTG CTG AAG CAG GTG CAC CCC GAT
GGC AAG AAG AGG AAC AGG AAA AGG AAG GAG AGT TAT GGA ATC TAC ATC TAC AAA GTC CTC AAG CAG GTT CAT CCA GAT
Gly Lys Lys Arg Arg Lys Thr Arg Lys Glu Ser Tyr Ala Ile Tyr Val Tyr Lys Val Leu Lys Gln Val His Pro Asp
                Asn Arg Lys.                    Gly         Ile
                Lys Arg Ser                     Ser Val

                                60                                      70
              *  ** **           *              **    *           * * *   *       *   *  **   *     * *
ACC GGC ATC TCG TCC AAG GCC ATG AGC ATC ATG AAC TCC TTT GTC AAC GAT GTG TTT GAG CGC ATC GCA GGG GAA GCC
ACC GGC ATC TCC AGT CGG GCC ATG ATC ATC ATG AAC AGC TTC GTC AAC GAC ATC TTC GAG CGA ATT GCC GGC GAA TCT
Thr Gly Ile Ser Ser Lys Ala Met Ser Ile Met Asn Ser Phe Val Asn Asp Val Phe Glu Arg Ile Ala Gly Glu Ala
                    Arg         Ile                                 Ile                             Ser
                                Gly                                 Ile

        80                                      90                                      100
          *       *           * ***   *          *  ***   *       *           *       *   * *   *   *
TCC CGC CTA GCT CAT TAC AAC AAC CGC TCC ACC ATC ACC TCC CGG GAG ATC CAG ACC GCG GTC CGA CTG CTG CTG CCT
TCC CGC CTC GCT CAG TAC AAC AAA AAG TCA ACC ATC AGC AGT CGC GAG ATT CAG ACC GCC GTC CGC CTC ATT CTC CCC
Ser Arg Leu Ala His Tyr Asn Asn Arg Ser Thr Ile Thr Ser Arg Glu Ile Gln Thr Ala Val Arg Leu Leu Leu Pro
                Gln         Lys Lys             Ser                                         Ile
                            Lys

                        110                                     120
  *     *     *   *       *     **        *           *   *   *   *       *  *  *
GGG GAG TTG GCC AAA CAC GCC GTG TCC GAG GGC ACC AAG GCT GTC ACC AAG TAC ACC AGC GCC AAG TAA X.laevis
GGA GAG CTG GCA AAG CAC GCT GTG AGC GAG GGT ACC AAG GCA GTG ACG AAA TAC ACT ACC TCC AAG TAG P.miliaris
Gly Glu Leu Ala Lys His Ala Val Ser Glu Gly Thr Lys Ala Val Thr Lys Tyr Thr Ser Ala Lys     X.laevis
                                                                            Thr Ser         P.miliaris
                                                                                Ser         B.taurus
```

Fig.3. Nucleotide sequence of the *X. laevis* histone H2B gene as compared to that of *P. miliaris* [13]. See for explanation the legend of fig.2. The amino acid substitutions as compared to H2B of *P. miliaris* [14] and calf [20] are indicated.

Table 2

Frequency of doublets CpG and GpC in the *X. laevis* nucleosomal histone genes

Sequence	CpG/GpC				
	H3	H4	H2A	H2B	mean
Translated region					
All codon positions	32/43	21/27	28/43	22/34	0.69
codon position 1–2	14/18	10/7	9/15	6/16	0.68
codon position 2–3	1/14	0/13	2/13	3/10	0.12
codon position 3–1	17/11	11/7	17/15	13/9	1.38
Untranslated region					
Prelude	8/14	3/5	4/8	3/9	0.49
Postlude	1/4	1/4	9/13	3/5	0.50

a high GC content, but seems rather to reflect a general evolutionary tendency [24]. For example, the *P. miliaris* (h19) H2A and H2B genes also display a high GC content (i.e., 55% and 54%, respectively), but only 55% and 69% of the H2A and H2B codons, respectively, have G or C at the third codon position.

Codons ending in A are only slightly underused compared to U in the H2A gene. In the H2B gene A and U are equally used in the third codon position. This is in contrast with the underuse of A as third base in mammalian genes [24].

In agreement with the general bias in eukaryotes against the dinucleotide CpG [23], CpG is underused in the H2A and H2B genes and in the H3 and H4 genes [13]. Only one codon containing CpG is not used. In Table II the use of CpG is shown in more detail. In the translated region the frequency of CpG depends clearly on the codon position. At codon position 3–1 CpG is clearly preferred to GpC, while at codon position 2–3 the use of CpG is rare. Although CpG is underused at codon position 1–2, there seems no strong bias against the use of this doublet at this position, since, e.g., the arginine quartet codons, containing CpG, are overused compared to the arginine duet codons.

Both in the 5′ and in the 3′ untranslated region the CpG doublet has ∼ ½ the frequency of GpC.

The frequency of CpG and GpC in the *P. miliaris* (h19) histone genes (calculated from [14]) is about the same as in *X. laevis* (not shown).

Finally, for human α- and β-globin genes the interesting observation has been made that codons that can mutate by a single step to a termination codon are not used if the genetic code contains other synonymous codons that code for the same amino acid [25]. This is not the case in the *Xenopus* H2A and H2B genes.

3.5. *5′ and 3′ flanking sequences and in vivo expression*

The 5′ and 3′ flanking sequences are of particular importance because of the presence of a number of conserved DNA sequence elements, homology blocks or 'consensus sequences' that have regulatory functions [26].

Fig.4 presents the 5′ flanking sequences (prelude sequences) of the H2A and H2B genes. They do not display much homology either with each other or with the prelude sequences of, e.g., the *P. miliaris* (h19) histone genes [26]. DNA sequence elements, possibly homologous with the 'consensus sequences' have been indicated tentatively. The most clearly recognizable sequence motif is the 'TATA box' involved in the initiation of transcription by RNA polymerase II [27]. Further upstream from the 'TATA box' a 'CCAAT box' can be assigned. The sequence motif: GATCC, characteristic for histone genes and usually present ∼ 10 basepairs upstream of the TATA box is not clear. A possible

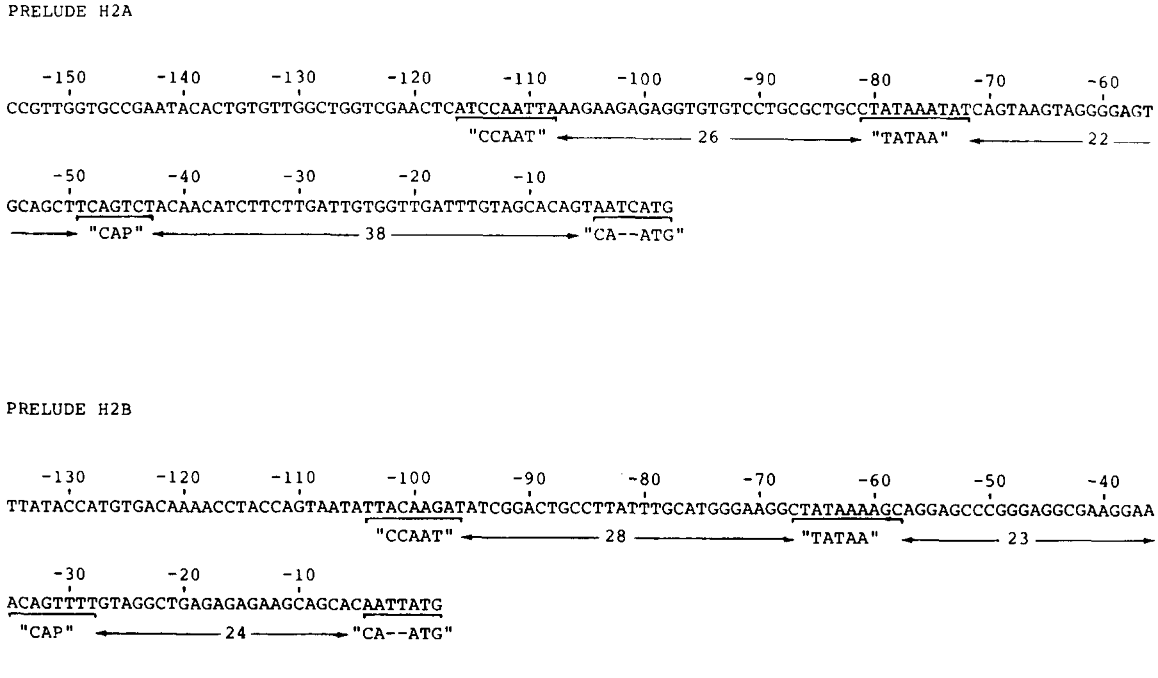
PRELUDE H2A
-150 -140 -130 -120 -110 -100 -90 -80 -70 -60
CCGTTGGTGCCGAATACACTGTGTTGGCTGGTCGAACTCATCCAATTAAAGAAGAGAGGTGTGTCCTGCGCTGCCTATAAATATCAGTAAGTAGGGGAGT
"CCAAT"
26
"TATAA"
22
-50 -40 -30 -20 -10
GCAGCTTCAGTCTACAACATCTTCTTGATTGTGGTTGATTTGTAGCACAGTAATCATG
"CAP"
38
"CA--ATG"
PRELUDE H2B
-130 -120 -110 -100 -90 -80 -70 -60 -50 -40
TTATACCATGTGACAAAACCTACCAGTAATATTACAAGATATCGGACTGCCTTATTTGCATGGGAAGGCTATAAAAGCAGGAGCCCGGGAGGCGAAGGAA
"CCAAT"
28
"TATAA"
23
-30 -20 -10
ACAGTTTTGTAGGCTGAGAGAGAAGCAGCACAATTATG
"CAP"
24
"CA--ATG"

Fig.4. 5′-Prelude nucleotide sequences of the *X. laevis* H2A and H2B genes in Xl-hi-1. Putative blocks of homology [26] are indicated.

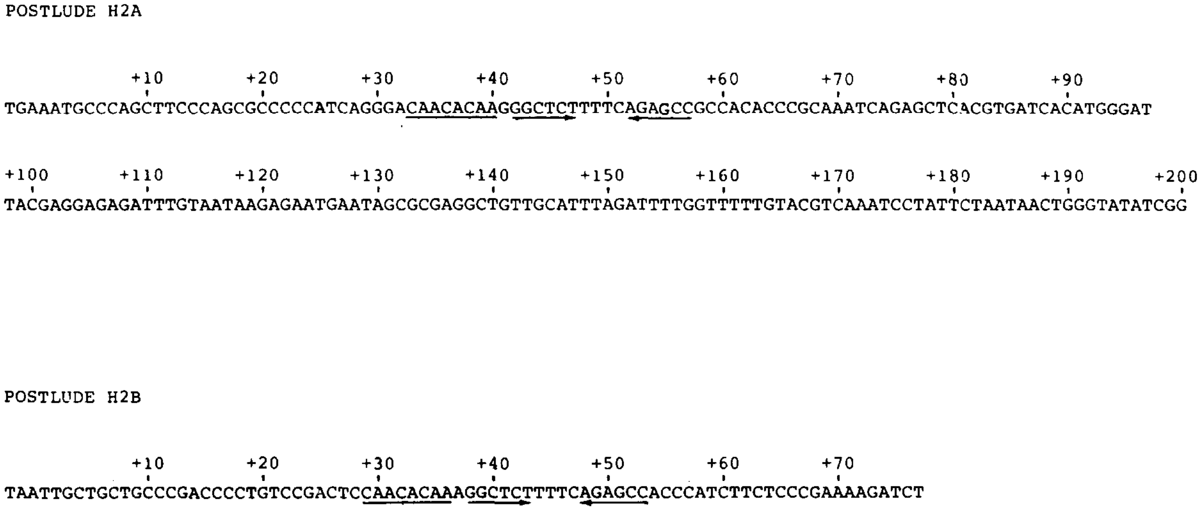
POSTLUDE H2A
+10 +20 +30 +40 +50 +60 +70 +80 +90
TGAAATGCCCAGCTTCCCAGCGCCCCCATCAGGGACAACACAAGGGCTCTTTTCAGAGCCGCCACACCCGCAAATCAGAGCTCACGTGATCACATGGGAT
+100 +110 +120 +130 +140 +150 +160 +170 +180 +190 +200
TACGAGGAGAGATTTGTAATAAGAGAATGAATAGCGCGAGGCTGTTGCATTTAGATTTTGGTTTTTGTACGTCAAATCCTATTCTAATAACTGGGTATATCGG
POSTLUDE H2B
+10 +20 +30 +40 +50 +60 +70
TAATTGCTGCTGCCCGACCCCTGTCCGACTCCAACACAAAGGCTCTTTTCAGAGCCACCCATCTTCTCCCGAAAAGATCT

Fig.5. 3′-Postlude nucleotide sequences of the *X. laevis* H2A and H2B genes in Xl-hi-1. The palindromic sequence is indicated by arrows. An AC-rich block of homology is underlined.

'Cap box' is indicated tentatively; however, it is not present in the form of: 5'-pyrimidine–CATTC–purine–3'. Unambiguous identification of the possible regulatory elements requires further experiments using specific deletion mutants.

The 3' flanking sequences (postlude sequences) of the histone H2A and H2B genes are given in fig.5. They are more divergent from each other or/from other 3' flanking histone sequences than the postlude sequences of *Xenopus* H3 or H4. However, a block of impressive homology is present at 30–50 nucleotides downstream of the terminator codon. This block is also present in the postlude sequences of *Xenopus* H3 and H4 genes and in those of the sea urchin histone genes. This block consists of the palindromic sequence GGCTCTTTTCA-GAGCC preceded by an AC-rich conserved motif. The palindromic sequence is probably also present in the mRNAs encoded [17]. The sequence AAUAAA, that may be involved in polyadenylation of eukaryotic mRNAs [28] is not present in any of the histone genes in Xl-hi-l. This is similar to the situation in the histone genes of sea urchins.

The in vivo expression of the H2A gene in Xl-hi-l is under investigation by testing the S1 nuclease resistance of hybrids between H2A gene and mRNAs. Preliminary experiments [29] show that none, or only a low amount, of the histone mRNAs from oocytes and gastrula stage embryos is completely homologous to the H2A sequence in Xl-hi-l. The differences between the nucleotide sequences of the gene and the mRNA for H2A are localized in the non-coding region as was found for the H3 gene [13].

REFERENCES

[1] McGhee, J.D. and Felsenfeld, G. (1980) Annu. Rev. Biochem. 49, 1115–1156.

[2] Zweidler, A. (1980) in: Gene Families of Collagen and Other Proteins (Prockop, D.J. and Champe, P.C. eds) pp. 47–56, Elsevier Biomedical, Amsterdam, New York.

[3] Elgin, S.C.R. and Weintraub, H. (1975) Annu. Rev. Biochem. 44, 725–774.

[4] Kedes, L.H. and Birnstiel, M.L. (1971) Nature New Biol. 230, 165–169.

[5] Kedes, L.H. (1979) Annu. Rev. Biochem. 48, 837–870.

[6] Childs, G., Maxson, R., Cohen, R. and Kedes, L. (1981) Cell 23, 651–663.

[7] Van Dongen, W., De Laaf, L., Zaal, R., Moorman, A. and Destrée, O. (1981) Nucleic Acids Res. 9, 2297–2311.

[8] Stephenson, E.C., Erba, H.P. and Gall, J.G. (1981) Cell 24, 639–647.

[9] Jacob, E., Malacinsky, G. and Birnstiel, M.L. (1976) Eur. J. Biochem. 69, 45–54.

[10] Engel, J.D. and Dodgson, J.B. (1981) Proc. Natl. Acad. Sci. USA 78, 2856–2860.

[11] Heintz, N., Zernik, M. and Roeder, R.G. (1981) Cell 24, 661–668.

[12] Moorman, A.F.M., De Laaf, R.T.M., Destrée, O.H.J., Telford, J. and Birnstiel, M.L. (1980) Gene 10, 185–193.

[13] Moorman, A.F.M., De Boer, P.A.J., De Laaf, R.T.M., Van Dongen, W.H.A.M. and Destrée, O.H.J. (1981) FEBS Lett. 136, 45–52.

[14] Busslinger, M., Portmann, R., Irminger, J.C. and Birnstiel, M.L. (1980) Nucleic Acids Res. 8, 957–976.

[15] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 449–560.

[16] Schaffner, W., Kunz, G., Daetwyler, H., Telford, J., Smith, H.O. and Birnstiel, M.L. (1978) Cell 14, 655–671.

[17] Zernik, M., Heinz, N., Boime, I. and Roeder, R.G. (1980) Cell 22, 807–815.

[18] Lifton, R.P., Goldberg, M.L., Karp, R.W. and Hogness, D.S. (1977) Cold Spring Harbor Symp. Quant. Biol. 42, 1047–1051.

[19] Yeoman, L.C., Olson, W.O., Sugano, N., Jordan, J.J., Taylor, C.W., Starbuck, W.C. and Busch, H. (1972) J. Biol. Chem. 247, 6018–6023.

[20] Iwai, K., Hayashi, H. and Ishikawa, K. (1972) J. Biochem. 72, 357–367.

[21] Van Helden, P., Strickland, W.N., Brandt, W.F. and Van Holt, C. (1978) Biochim. Biophys. Acta 533, 278–281.

[22] Von Holt, C., Strickland, W.N., Brandt, W.F. and Strickland, M.S. (1979) FEBS Lett. 100, 201–218.

[23] Nussinov, R. (1980) Nucl. Acid Res. 8, 4545–4562.

[24] Grantham, R. (1980) Trends Biochem. Sci. 5, 327–331.

[25] Modiano, G., Battistuzzi, G. and Motulsky, A.G. (1981) Proc. Natl. Acad. Sci. USA 78, 1110–1114.

[26] Hentschel, C.G. and Birnstiel, M.L. (1981) Cell 25, 301–313.

[27] Bogenhagen, D.F. and Brown, D.D. (1981) Cell 24, 266–270.

[28] Proudfoot, N.J. and Brownlee, G.G. (1976) Nature 263, 211–214.

[29] Van Dongen, W.M.A.M. (1982) PhD Thesis, Rodopi Editions NV, Amsterdam.