Nucleotide sequences of H1 histone genes from *Xenopus laevis*. A recently diverged pair of H1 genes and an unusual H1 pseudogene

P.C.Turner, T.C.Aldridge, H.R.Woodland and R.W.Old

Department of Biological Sciences, University of Warwick, Coventry CV4 7AL, UK

Received 25 February 1983; Revised 26 April 1983; Accepted 20 May 1983

ABSTRACT

Four clones containing H1 histone gene sequences were previously isolated from a Xenopus laevis genomic library (1) and we now present the complete nucleotide sequences of these H1 genes and their flanking regions. Two of these genes code for minor H1 proteins, probably H1C, when expressed in the oocyte transcription/translation system and are present on clones with almost identical overall organization. However, at the nucleotide level these genes differ in showing base insertions and deletions, as well as substitutions. A third gene sequence which is more related to the major X. laevis H1A, corresponds to the 3' two thirds of an H1 gene. This gene has in place of a 5' coding region at least 1800 bp of apparently noncoding sequence, some of which is A-T rich. The junction does not correspond to the consensus sequence of an intron/exon boundary and therefore this H1 sequence is more likely to represent a pseudogene. Comparisons of the coding and flanking regions of these X. laevis H1 genes indicate the kind of differences which can occur among H1 subtypes within a species. A region of homology noted in the 3' noncoding portion of vertebrate histone genes is discussed in relation to the mechanism of termination of transcription.

INTRODUCTION

Compared to the 4 highly conserved core histone proteins H1 histone, which is involved in nucleosome packing (2), shows considerable species and cell-specific variation (e.g. 3,4,5,6). These proteins are composed of three structural domains (7), a variable basic N-terminal portion, a central hydrophobic region which shows the greatest degree of conservation (8) and a variable, very basic C-terminal half (3). Many partial and some complete H1 protein sequences have now been determined (3,4,9) and it has been observed that within the conserved hydrophobic region of these molecules, separate subtypes of the H1 class, such as the erythrocyte type H5, have different, but conserved stretches of amino acids (9).

X. laevis embryos contain 3 resolvable subtypes of H1 histone. H1A, which is quantitatively the major type at all stages of development, H1B and H1C (10,11). While these 3 subtypes are present in all adult tissues so far examined, including erythrocyte and testes, the proportions vary, with H1A not being so

© IRL Press Limited, Oxford, England.

predominant. Adult tissues have in addition small amounts of two faster migrating subtypes H1D and H1E (11). The latter type was found by Risley and Eckhardt (11) to be indistinguishable on 3 gel systems and after partial peptide analysis from a X. laevis erythrocyte subtype referred to as H5 by Destreé et al. (12,13). The X. laevis erythrocyte H5 and the H1D and H1E proteins seem to be enriched in cell types with low rates of cell division (11), and this kind of H1 histone, known as $H1^{o}$ and H5, is found in other species (14,15).

Studies of the primary structure of H1 subtypes have suggested that the different lengths of the termini of H1 proteins are responsible for the changes seen in nucleosome packing in different cell types (2,16,17). The mechanisms by which these adult-specific H1 genes become active in slowly dividing cells may only be understood fully when the nucleotide sequences of the H1 gene regions are known and compared.

In a previous publication we described the isolation of 9 histone clones from a X. laevis genomic library which was constructed by cloning Eco RI digested erythrocyte DNA in λgtWES.λB (1). All these clones contained H4 histone gene sequences since the library was screened using the X. laevis ovarian H4 cDNA clone pcXlH4W1 (18). On initial mapping, 4 of these clones appeared to contain H1 gene sequences. Sequencing of these clones was undertaken to confirm the suspected positions of these genes and to establish the polarity of the genes. It was hoped that sequencing might explain why one of the cloned DNAs, XLHW19 did not produce an H1 polypeptide product after microinjection into oocyte nuclei (1). The data would also allow a comparison of the nucleotide and protein sequences of H1 gene subtypes within a species. Indeed, we now present the first complete nucleotide sequence of a vertebrate H1 histone gene.

MATERIALS AND METHODS

Materials

All enzymes were from Bethesda Research Laboratories, Boehringer Mannheim or New England Biolabs and were used as recommended by the manufacturers. Radiochemicals were from Amersham International, Amersham.

Sequencing Procedures

Both the chemical methods of Maxam and Gilbert (19) and the enzymic method of Sanger et al. (20) applied to M13 single stranded phage subclones were used as indicated in Fig. 1. The majority of the sequences presented were determined using both methods and on both strands. Most sequencing runs were performed at least twice.

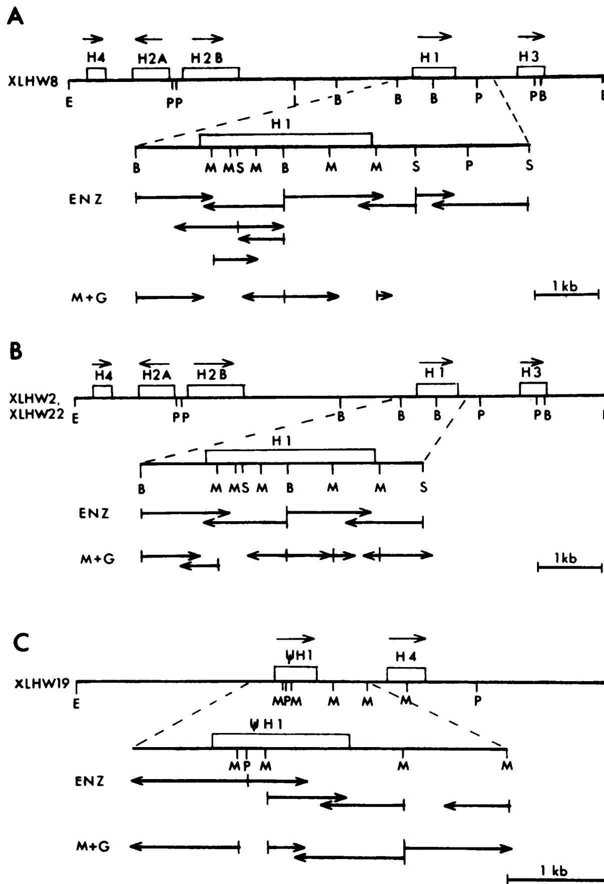

Figure 1. Clone organization and sequencing strategy. The regions sequenced by either the method of Sanger et al. (20) or by the method of Maxam and Gilbert (19) are indicated as ENZ or M+G respectively, below the expanded portions of the maps. Only the restriction sites for Bam HI (B), Pst I (P), Eco RI (E) and Sal I (L) are given for the complete clones. For other enzyme sites see Old et al. (1). In the expanded portions Msp I (M) and Sau 3A (S) sites, used in the M13 subcloning, are also included.

Other Procedures

The nuclear microinjection protocol, hybrid selected translation experiments, Southern blot transfers and gel electrophoresis were all exactly as described by Old et al. (1).

RESULTS

Fig. 1A and B show the first maps of the three clones XLHW2, XLHW8 and

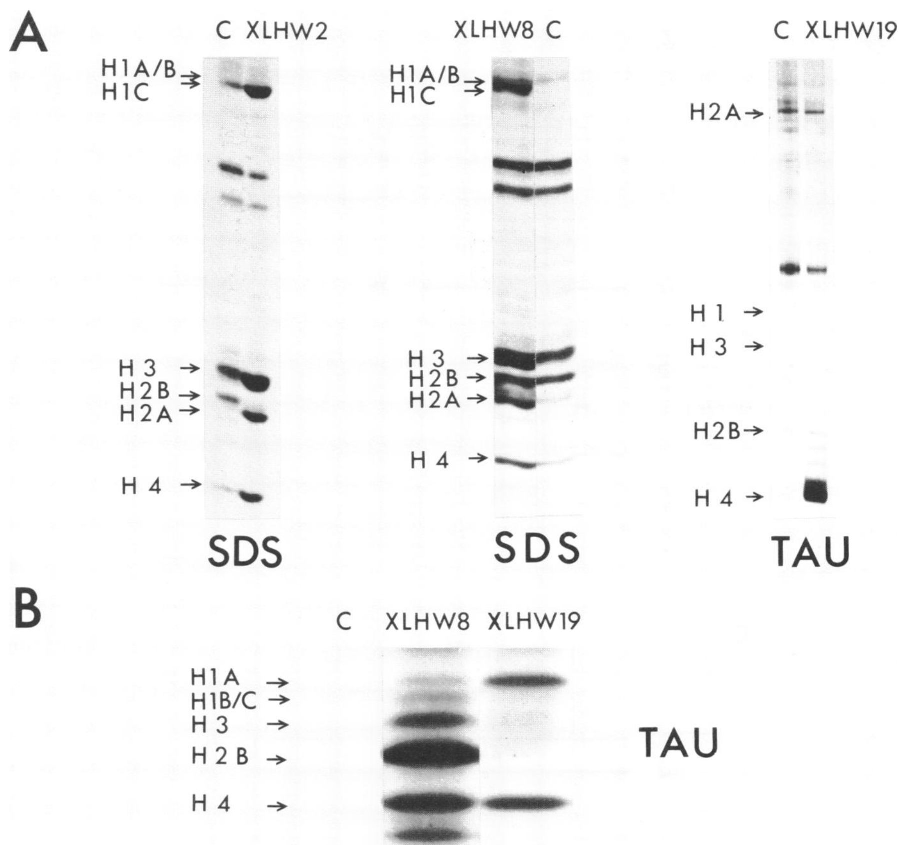

Figure 2. (A) Translation products of XLHW2, XLHW8 and XLHW19 DNAs after micro-injection into *Xenopus* oocyte nuclei. (B) Hybrid selected translation of XLHW8 and XLHW19. The gel systems used were either SDS/acrylamide (18%) or triton/acid urea as indicated. Labelling was with ^{3}H-lysine. Lanes marked C are controls and *X. laevis* erythrocyte histones were used as markers.

XLHW22, which all contain one copy of each of the 5 histone gene types (1). They are very similar in length, organization and restriction site positions. From fine restriction mapping and sequence data (not shown), it now appears that clones XLHW2 and XLHW22 have exactly the same insert which was cloned in opposite orientations in the vector and therefore only XLHW2 is further considered. XLHW2 and XLHW8 have only a few restriction site differences and their similar organization has been underlined by the extensive sequence conservation of the H1 gene regions of these two clones (see below). The clone

XLHW19 (Fig. 1C) has a very different organization and contains only H1 and H4 histone gene sequences.

Nuclear Microinjection and Hybrid Selected Translation

Fig. 2A shows that when microinjected into the nuclei of *Xenopus* oocytes, XLHW2 and XLHW8 both produce H1 gene products which migrate slightly faster on SDS/acrylamide gels than the H1A/H1B markers. On this basis, Old *et al.* (1) suggested that the H1 gene in XLHW8 was of the H1C subtype. The data of Fig. 2A shows that XLHW2 contains a functional H1 gene which also seems to be of the H1C subtype. Furthermore it appears that the H1 protein band in the XLHW8 track of Fig. 2A migrates fractionally slower relative to the markers than does the equivalent XLHW2 protein and this observation is in keeping with the sequencing data given below.

Fig. 2A also shows the nuclear microinjection analysis for XLHW19 presented by Old *et al.* (1) and is reproduced here to emphasize that no H1 polypeptide could be detected for this clone in this assay system. The only other histone sequence ascribed to this clone, that of H4, was clearly expressed above the low background level of H4 synthesis.

Fig. 2B shows that XLHW19 selects both H1A and H4 mRNA when hybridized with *X. laevis* ovarian total RNA, and must therefore contain H1 sequence. The hybrid selection data for XLHW8 shows that the H1 gene in this clone seems less efficient in hybridizing to ovarian total RNA than the H1 gene in XLHW19. In addition, the selected oocyte H1 mRNA sequences were both the H1A and the H1B/C types. These observations are readily explained if the H1C gene in XLHW8, which corresponds to a minor oocyte mRNA sequence, cross-hybridized weakly, at the stringency employed, with the major oocyte H1A mRNA. The relatively efficient selection of H1A mRNA by XLHW19 would imply that it encodes an H1 gene that is more closely related to the H1A type than to the H1B or H1C types.

Nucleotide Sequences Encompassing the H1 Genes in XLHW2 and XLHW8

Employing the strategies outlined in Figs. 1A and B, the H1 gene regions of XLHW2 and XLHW8 have been sequenced. The complete H1 coding region of XLHW8 and approximately 850 bp of its flanking sequence is shown in Fig. 3. A shorter region of XLHW2 (1071 bp) was sequenced and for clarity, only the base differences which occur in this region are presented. Four small gaps must be introduced to maintain the homology of these sequences.

The H1 gene in XLHW8 is the longer of the two, encoding a protein of 220 amino acid residues, and having the typical features of the H1 proteins whose primary sequence is known. There is a central hydrophobic region of 72 amino

```
-250                                         -200                                                   -150
5'-GGATCCCACACAAACTGGGCACATTTATCATCAGAGACAAGAGAACCCCCCATTTTCCAGACTCTAAAAACACAGACTCCCACCTGCCCCATAGTTTCCACCAGCGAGAATTAATGGGCGGG

      CAAT 2      TATA 2      CAAT 1                    TATA 1       -50                          CAP 2     CAP 1
GTTTGCTTCTCAACCAATGAAGTTTAACTATAAAAGCTGGCCAATCAGGGACCCAGAAAAGATGAATATAAGGAGGTCGATATAAACTGAAAGTTTAGATTTTAGTTCGCGTGAGTGAATTTACTTCAAAG

1   ala         ser                                     50                      lys
met thr glu thr ala ala thr glu thr thr pro ala ala pro pro ala glu pro lys gln lys lys lys     gln gln pro lys lys ala ala gly gly
ATG ACA GAA ACT GCT GCA ACT GAG ACA ACT CCC GCC GCT CCC CCG GCA GAA CCC AAA CAG AAG AAG AAG     CAG CAG CCT AAG AAG GCA GCG GGA GGC

100                                                 150
ala lys ala lys lys pro ser gly pro ser ala ser glu leu ile val lys ser val ser ala ser lys glu arg gly gly val ser leu ala ala leu
GCT AAG GCC AAG AAA CCC TCC GGA CCG AGC GCA TCT GAG CTG ATC GTG AAA TCC GTG TCC GCC TCT AAG GAG CGT GGT GGC GTG TCC CTG GCC GCT CTC

200                             asp                         250
lys lys ala leu ala ala gly gly tyr asn val glu arg asn asn ser arg leu lys leu ala leu lys ala leu val thr lys gly thr leu thr gln
AAG AAG GCC TTG GCT GCC GGA GGT TAC AAT GTG GAG AGG AAC AAC AGT CGC CTC AAG TTG GCT CTC AAG GCT TTG GTC ACT AAA GGG ACT CTC ACC CAA

300                                                         350
val lys gly ser gly ala ser gly ser phe lys leu asn lys lys gln leu glu thr lys val lys ala val ala lys lys lys leu val ala pro lys
GTC AAA GGC AGC GGA GCC TCT GGA TCC TTC AAG CTG AAC AAG AAG CAG CTG GAG ACC AAG GTG AAG GCC GTG GCC AAG AAG AAG CTC GTG GCG CCC AAA

400             thr                                             450
ala lys lys pro val ala ala lys lys lys pro lys ser pro lys lys pro lys lys val ser ala ala ala ala lys ser pro lys lys ala lys lys
GCC AAG AAA CCC GTC GCG GCA AAG AAA AAG CCC AAA TCC CCT AAA AAG CCC AAG AAG GTC TCG GCG GCA GCA GCA AAG AGC CCC AAG AAG GCG AAG AAA

500                                                 ser             550
pro val lys ala ala lys ser pro lys lys pro lys ala val lys pro lys lys val thr lys ser pro ala lys lys ala thr lys pro lys ala ala
CCG GTA AAG GCC GCC AAA AGC CCC AAG AAG CCC AAA GCT GTT AAA CCC AAG AAG GTG ACC AAG AGT CCA GCT AAA AAG GCC ACT AAG CCC AAA GCT GCC

600                                                         650
lys ala lys ile ala lys pro lys ile ala lys ala lys ala ala lys gly lys lys ala ala ala lys lys
AAG GCC AAA ATA GCC AAA CCC AAA ATA GCC AAA GCC AAG GCG GCT AAG GGC AAG AAG GCT GCG GCT AAA AAG TAA CTTGGTTCCGGTGTCCCTGTCCATCCCCCCC

700                                         750                                             800
CAACCTCAAAGCCTCTTTTCAGAGCCACCACCAACTCAGTGAGAAGAGCCGATACTGTGACTGCTCATTGGCTCCATTATGTCAGGGAGGCTTTCTGGTATCTGCTGAATTGTGCGATGCCCAGTCCTGAT

                850                                                 900                                                 950
CTTTCAGCCCCATCTCCCCTTAACTGTGTCTCCCACACAAGTCACTGCGCAGAACTACAGTCTGACATTCCACCTTNTCCAAACGCGTTAAGGTTTATCCTCCGCTAAGAAAGTTCAGTCTAATAGCCGCG

                                                1000                                                1050
TTTCAACTATTCATTTAAGGGCCGATTTTGCCGATGTAATAAAGAGTTTCTATTTAACCACTTCAACTGCAGTCATTCTGCTCAGGTTCTCATTCATATTGATACAAATGTATTTGAGAGGAAGTGACGTG

   1100                                                     1150                                                    1200
TATTAGGACACATTTGTATCAGTCCGTGTCCCTCCCCTTCTTACTGGGCCCCAGAGCCTGGTTTTGGGCAGTTTAAATGATTTAATGAAAGTCTCTCCATAGTAACGAACAGGAGGCGGCGGGTCTCGCGG

                    1250
AAACGAGTTCAGACTCGAGGGAAATTCACGGGGATC-3'
```

Figure 3. The nucleotide sequence of the H1 gene regions in XLHW8 and XLHW2. Only the base differences in the H1 gene in XLHW2 are given. Gaps in the sequence, for XLHW8, or in the solid line, for XLHW2, indicate deletions in the molecules. Numbered from ATG=1 and every tenth base of XLHW8 is marked. The sequenced region of XLHW2 ends at 840. Putative promoter elements are indicated.

Figure 4. Comparison of the 5' noncoding region of H1 genes from Xenopus and sea urchin. (A) The 5' noncoding region of the H1 gene in XLHW8 is aligned for maximum homology with the Psammechinus miliaris h22 H1 gene (26) and with the Stronglyocentrotus purpuratus H1 gene (9) using the first possible 'TATA' box in XLHW8. (B) As in A but using the second possible 'TATA' box in XLHW8. Regions of homology are boxed.

acids from codon position 40-112 (121-336 bp, Fig. 3), which shows a high degree of conservation relative to calf thymus, CTL-1 (15) and rabbit thymus, RTL-3 (21) H1 proteins. The H1 protein sequence derived from the gene in XLHW2 is very similar to that in XLHW8, except that it is 4 residues shorter due to two insertion/deletion events in the variable regions of the protein at codon positions 23 and 203 (69 and 610 bp, Fig. 3), and it has 5 amino acid changes relative to XLHW8.

Three possible "TATA" boxes occur in the 254 bp of 5' noncoding sequence of the H1 gene in XLHW8; at positions -51, -65 and -103 bp (the latter 2 being marked in Fig. 3). Two of the 12 substitutions which occur in the corresponding region of XLHW2 would almost certainly make the putative "TATA" box at -51 non-functional. Since no such substitutions are seen in the other two possible "TATA" boxes and since both are preceded by equally good "CCAAT" boxes at -91 and -118 bp (Fig. 3), as are most polymerase II genes (22), then in the absence of other evidence, we are unable to rule out the possibility that both of these pairs of promoter elements function in vivo.

No "cap" site giving a good match to the consensus sequence 5'-CATTC-3' (22) is evident but two possible sites are indicated in Fig. 3, at -28 and -33 bp. In fact, cannonical "cap" sites seem rare in Xenopus histone genes (23,24,25, Fig. 3), the sequence 5'-AGTT-3' occuring more frequently than the "best fit" to the cannonical sequence which is usually 5'-ATT-3'.

Attempts have been made in Fig. 4 to align the 5' noncoding regions of the available H1 genes. By introducing appropriate gaps some regions of homology, in addition to the recognized promoter elements, become apparent. It is noteworthy that in very similar relative positions (approx. -30) the sea urchin genes and the Xenopus genes all have the sequence 5'-TTTT(A)GTT-3'. The first possible "TATA" box in the XLHW8 gene is in about the same relative position as in the sea urchin genes, but produces less extensive homology with

these genes than does the second possible "TATA" box (Fig. 4B). Since both of these alignments in Fig. 4 produce extended homology there is a prima facie case that both "TATA" boxes are functional.

The 3' noncoding region of the H1 genes in XLHW2 and XLHW8 are typical histone postlude sequences in all respects. At position 716 (Fig. 3 and see also Fig. 6) there is the centre of the highly conserved region of dyad symmetry found about 50 bp downstream from the termination codon in all histone 3' noncoding regions except yeast (M. Smith cited in 6) and chicken H5 (27). Between this homology block and the termination codon, homopolymer runs of pyrimidine residues, as at 692 bp (Fig. 3) are a common feature of histone genes (18). There is also no polyadenylation signal, which again is a common feature of histone genes (6). Several short direct, or inverted, repeat sequences in the 3' noncoding region of the XLHW8 gene are indicated in Fig. 3. At present their significance is unknown.

The Nucleotide Sequence of a Truncated Gene Related to H1A.

Using the strategy outlined in Fig. 1C, approximately 1200 nucleotides encompassing the H1 sequence in XLHW19 have been determined and are presented in Fig. 5. The most significant feature of this sequence is that the 5' one third of the gene is absent, being replaced by an A-T rich sequence containing many short homopolymer runs e.g. at positions 10, 126, 134, 224 and 233. The coding region of this gene appears to start at base pair position 246, with the second of a pair of asparagine residues corresponding to about amino acid position 80 in most vertebrate H1 histones which is in the middle of the central conserved region. No termination codons are present until the end of the presumptive coding region (675 bp, Fig. 5) and all of the 143 triplets code for amino acids which could be present in typical H1 genes at the positions in which they occur, with the possible exception of the glutamic acid residue at 288. It has been noted by Levy et al. (9) that in all H1 and H5 proteins a glycine is present at this position and a single base substitution accounts for this change. The 3' noncoding region of the H1 gene in XLHW19 also seems like a normal histone postlude sequence (see also Fig. 6). Taken with the hybrid selected translation data of Fig. 2B, the sequence in Fig. 5 from 246 onwards most probably resembles very closely that of a genuine X. laevis H1A gene.

The presence of this unusual 5' sequence raises the question of whether the H1 gene in XLHW19 contains an intron? More specifically, does the remainder of the clone contain the missing H1 5' end? We have probed Southern blots of restriction digests of XLHW19 with 5' portions of the H1 gene of XLHW8 not present in the sequenced region of XLHW19 and failed to detect cross

```
5'-CTTTTAATATTTTTTTTTTTTTTTTCCCCGGTACGGGGACCCGTCCTTATAATAGCAACACTAGCATAAGTCAATCTACACTCTCTCAATGGCTGTACATTGTGTCTTCTTTGA

ATATTGTTTTCTTTTTTCTCTTTTTTTTTTTTTCCCTTTATTTTCTCGGGTTGATTATTTGTAAAGTGTCTAATGGTAAAATGTAAAACTTGTAAAAGTAATACAAATTTAAAAAAAACAAAAAAAAAACA

AAC AGC CGC CTC AAG CTG GCT CTC AAG GCT CTG GTC ACG AAG GAG ACC CTG CTC CAA GTC AAA GGC AGC GGA GCC TCC GGT TCC TTC AAG CTC AAC AAG
asn ser arg leu lys leu ala leu lys ala leu val thr lys glu thr leu leu gln val lys gly ser gly ala ser gly ser phe lys leu asn lys

AAG CAG CTG CAG AGC AAG GAC AAG GCC GCC AAG AAG AAG GCG CCG CTA GCA GCC AAA ACC AAG AAA CCG GCG GCA CGA GCC AAG AAG GCG CCA AAG TCT
lys gln leu gln ser lys asp lys ala ala lys lys lys ala pro leu ala ala lys thr lys lys pro ala ala arg ala lys lys ala pro lys ser

CCG AAA AAG CCC AAG AAA GTC TCC GCA GCC GCC AAG AGC CCG AAG AAG GTG AAG AAA CCC GCA AAA GCG GCG CTA GCA GCA AAG AGC CCG AAA AAA AAC
pro lys lys pro lys lys val ser ala ala ala lys ser pro lys lys val lys lys pro ala lys ala ala leu ala ala lys ser pro lys lys asn

AAA GCT GCC AAG CCC AAG AAG GCC ACC AAA AGC CCC GCA AAA AAG ACC GCC GTC AAG CCC AAA ACT GCT GCT GCC AAA AGC CCC GCA AAG GCT AAA GTG
lys ala ala lys pro lys lys ala thr lys ser pro ala lys lys thr ala val lys pro lys thr ala ala ala lys ser pro ala lys ala lys val

GCC AAA GCC AAG AAA GCC GCC CCC AAG AAG AAA TGA GCAGCTCGCTCCCTCGCTCGCTCACTAGTGGCCCATTCAACCAAAGGCTCTTTTAAGAGCCACCACATCCCCCTAAAAAGAGC
ala lys ala lys lys ala ala pro lys lys lys

AGCTTGTTGTCCTATAGCGACTCTCCTCTGCTTCTGCCTACCTACTATTTTTTCTTACTAACACGAGTATCCATCCGG◂Approx. 15 bp▸TCTCGTTGTCGGGAAAAAGCGTCTTTGAGGACAAACCT

GCAATGAATTACACGCACTGCTGGTGCACTTGTTGCCACTTCTATCCTGTACTGCGAGGAATGTTGCGGAAATACCCCAGTGTAACAATGTGGCAAAGTGCCAGTGAAATCTTTCTAGAGAAGTAACAAAC

GGGTGAAAGAGGCAGAATAACAAAGACCTCAGCTGAGAGGCTCATTGGGGTGCCGCTGGGAAGGGGCTGCAAGCGGTTGGCGAGAAAGGAGTAGACGTAGCGGTTTGCTTTTGCCACAAACGTTGANCATA

AATCAATAAAAAAAACCTGGATGAGTCAATGG-3'
```

Figure 5. The nucleotide sequence of the H1 gene region in XLHW19. Numbered from 1 at the 5' end with every tenth base marked.

hybridisation, even at low stringency (1 x SSC, 60°C). We therefore conclude that XLHW19 does not contain any sequence corresponding to the front portion of an H1 gene. Thus if the sequence replacing the 5' portion of the coding region in XLHW19 was an intron, the intron size would exceed 2.1 kb and the next exon would not be contained on this Eco RI genomic clone. However, inspection of the sequence around position 240 (Fig. 5) shows that the consensus 3' intron/exon boundary sequence 5'-YYYNCAG-3' (28) does not occur in this region, except possibly within the coding region itself at position 250. If this site represented the 3' end of an intron it would be remarkable that an asparagine codon occurs at 246 since asparagine is almost exclusively found in this position in vertebrate H1 proteins (9).

In view of the absence of introns in all other histone genes except a variant chicken H3 gene (29), we feel that the most probable explanation of the structure of XLHW19 is that there has been a recent mutational event, and the existing coding portion of the H1 gene has changed little in sequence since it was functional. In any case, the structure observed for this H1 gene readily explains why we observed no H1 gene product when this clone was microinjected into Xenopus oocyte nuclei (Fig. 2A).

3' Homologies of X. laevis Histone Genes.

In Fig. 6 all the available 3' noncoding regions of H1 genes have been compiled in the upper part of the figure, and in the lower part are the same regions of all the additional available X. laevis histone genes and a mouse histone H4 gene. By introducing appropriate gaps in these sequences we observe considerable homology of the H1 genes, although the sea urchin H1 genes are somewhat shorter. The extent of homology goes beyond the region of dyad symmetry in the 5' direction and the presence of variations of the sequence 5'-CAACACAAAC-3' noted in histone H3 and H4 genes (18) is readily apparent in both of these H1 genes and the other X. laevis histone genes. 3' to this conserved region of dyad symmetry there is another region of homology. This region was noted by Harvey et al. (30) in a comparison of chicken histone genes. Its divergence from the sea urchin sequence prompted these workers to suggest that the consensus sequence they observed might be specific to the chicken. However, the consensus sequence derived for X. laevis histone genes is so similar to the chicken consensus and to the only mouse histone sequence available (31) that one must conclude that this sequence (boxed in Fig. 6) is common at least to vertebrates as a whole and it may be that some species, e.g. chicken, tend to show one particular variation more than others. No regions other than those in Fig. 6 are conserved between the two X. laevis H1 3' noncoding regions presented here.

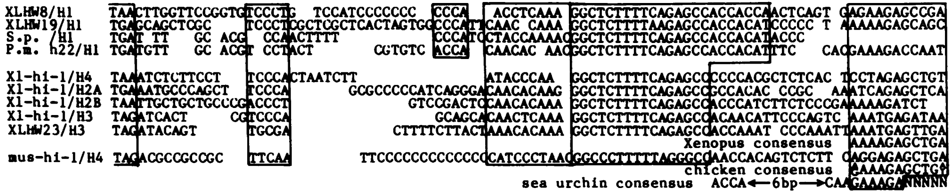

Figure 6. Comparison of the 3' noncoding region of H1 genes from sea urchin and Xenopus (Top). The 3' noncoding regions of other Xenopus histone genes and a mouse histone H4 (Bottom). Sequences are written 5' to 3' and gaps have been introduced to maximize homology. Regions of homology have been boxed and consensus sequences for the most 3' region of homology are given for Xenopus, chicken (30) and sea urchin (6). Sequence data for mus-hi-1 ref. 31, Xl-hi-1 refs. 23 and 24, XLHW23 ref. 25 or as in Fig. 4.

DISCUSSION

Noncoding Sequence Comparisons

Comparisons of the 5' noncoding region of H1 nucleotide sequences (Fig. 4) indicated the conservation of the sequence 5'-TTTT(A)GTT-3' about 30 bases upstream of the ATG. In view of the presence of a different H2B-specific sequence about 100 bp 5' to the translation initiation codon in all H2B sequences so far analysed (30) it will be interesting to see if the conserved H1 sequence is present in other H1 genes.

The consensus sequence 5'-AAAAGAGCTGA-3' (Fig. 6), just downstream from the region of dyad symmetry in the 3' noncoding region, is well conserved among vertebrates, being present in Xenopus, chicken and mouse and therefore cannot be considered specific to the chicken as proposed by Harvey et al. (30). Its presence in an altered form in sea urchins (6) suggests that it may have a fundamental function. In the experiments of Birchmeier et al. (32), faithful termination of transcription of a sea urchin H2A gene was shown to require sequences downstream from the conserved region of dyad symmetry. The presence of a conserved sequence in this region is exactly what would be predicted by these experiments.

Thus the present model for termination of transcription of histone genes is that termination probably occurs at an "ACCA" motif just 3' to the highly conserved region of dyad symmetry, by a mechanism which involves the recognition of the region of dyad symmetry and a region downstream of the "ACCA" sequence (32). This proposed second region of recognition in the sea urchin is closer to the "ACCA" motif and slightly shorter than the consensus sequence seen in the vertebrates (Fig. 6) and this implies that the mechanism of transcription termination may differ somewhat between sea urchins and these

vertebrates. On this point, it is noteworthy that the H3 gene in the Psammechinus miliaris clone h22, fails to terminate correctly when injected into the frog oocyte (33). The sequence of this sea urchin H3 gene, immediately after the 'CAAGAAAGA' does not match any of the next few bases in the vertebrate consensus sequence, whereas all the other histone genes in the P. miliaris h22 clone have at least one matching base in the adjacent 4 nucleotides. Recent experiments demonstrating that faithful termination of transcription of this sea urchin H3 gene in the Xenopus oocyte can be achieved by co-injection of a crude protein fraction from sea urchin (33) strengthens the belief that the sea urchin and vertebrate mechanisms may differ slightly.

H1 Coding Region Comparisons and Protein Evolution

The derived molecular weights of the H1 protein in XLHW2, which comigrates with H1C on two gel systems, and that in XLHW8, which migrates fractionally slower than H1C on SDS gels, are 22,187 and 22,589 daltons respectively. These values suggest a consistent, small underestimation in the estimated molecular weights of X. laevis H1A/B, H1D and H1E made by Risley and Eckhardt (11).

There are 24 base substitutions in the 1065 bp of sequence which can be compared between XLHW2 and XLHW8 (Fig. 3), an average of 2.3%, and this suggests that these pair of genes have derived from a common ancestral gene quite recently in evolutionary terms. 63% of these base changes are transitions and there is a tendency for the substitutions to be grouped, particularly in the noncoding region e.g. at -40 in Fig. 3 there are 6 changes in 13 bp. The 15 substitutions which occur in the noncoding region show that, in common with many other protein-coding genes, the rate of base substitution in the noncoding regions (3.5% base changes) is 2.6 times the total rate in the coding region (1.4%). The rate of silent site substitution is 2.6%.

There are 3 major insertions or deletions observed when comparing the H1 genes in XLHW2 and XLHW8 (Fig. 3), and their locations coincide in all 3 cases with the presence of short tandemly duplicated sequences e.g. at 611, there is an exact repeat of 14 bp. Mutational events of this kind could be caused by "slipped mispairing" (34) or unequal cross-over.

Figure 7 presents the coding region of the H1 gene in XLHW8 compared with the coding region of the partial H1 gene in XLHW19. 9 gaps have been introduced to maintain the homology of the derived proteins and all of these occur in the 3' variable region, there being 6 gaps in the XLHW8 gene. In view of the variability of this region of H1 proteins, the degree of homology observed in this pair of genes is quite high and may underline the equality of function of

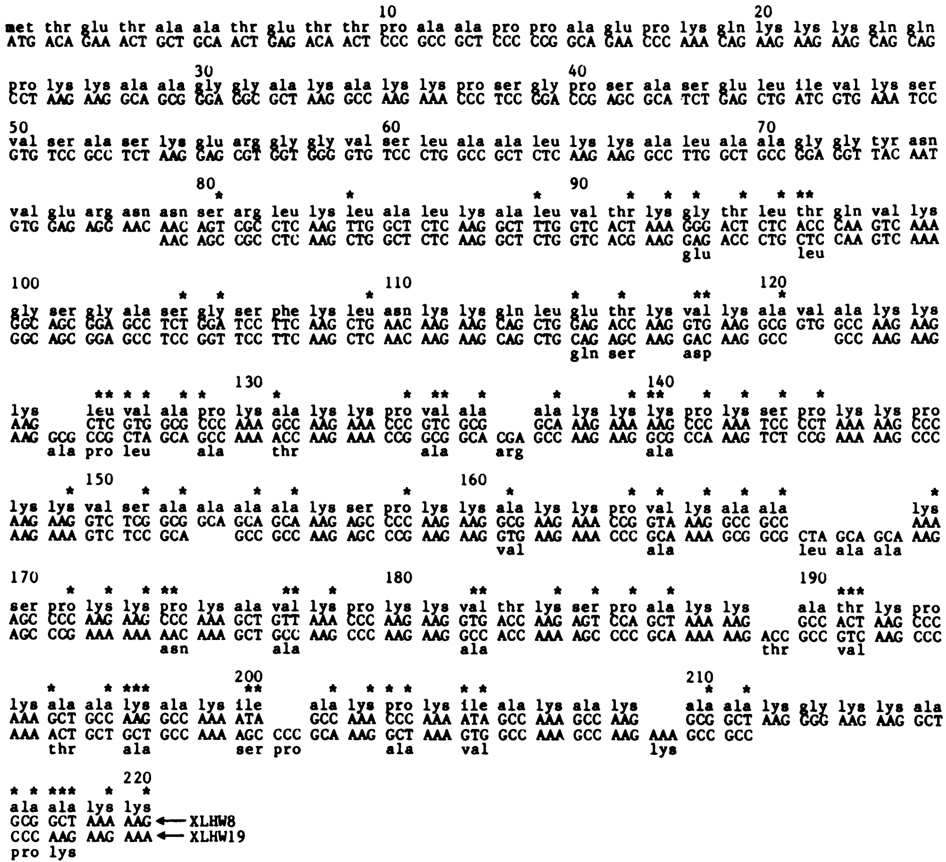

Figure 7. A comparison of the coding regions of the H1 genes in XLHW8 and XLHW19. Gaps, introduced for homology, base changes (*) and amino acid substitutions are indicated. Amino acid numbering is for XLHW8 excluding the first codon.

the X. laevis H1A and H1C subtypes.

Ignoring the gaps, 135 amino acids can be compared and 89 base changes (22%) have resulted in 24 amino acid replacements, the amount of replacement site substitution being 9.3%. Thus this pair of H1 genes shows 15 fold more base changes in the coding region than the XLHW2, XLHW8 pair.

One point that is clear from this study is that there seems to be an unsuspected degree of variation of H1 proteins even within a given subtype, i.e. the two H1C genes reported here have 5 amino acid replacements and one gene has 4 extra amino acids. There are about 90 H4 histone genes in X. laevis and these are present on at least 2 different types of genomic histone cluster (35).

If the H1 genes in the X. laevis genome are also divided among different types of genomic histone cluster, then perhaps the 3 major H1 protein types, H1A, H1B and H1C reflect 3 more or less independently evolving histone cluster types. Each cluster type might contain a set of non-identical H1 genes encoding a group of related proteins such as the H1A type. X. borealis, which has a single major H1 protein type (10) has a homogeneous major histone gene cluster arrangement within its genome (35). The observation of Zernik et al. (36) that variant H1 genes in X. laevis were present on genomic clones which had different gene orders, strengthens the possibility that genes for a given X. laevis H1 protein subtype will be present on a particular genomic cluster type. However, considering the degree of protein variation seen within a subtype there may be no absolute correlation between protein subtype and genomic cluster type.

ACKNOWLEDGEMENTS

We thank Elizabeth Ballantine for providing unpublished data, Neil Coveney for help in computer processing the DNA sequences and Dianne Simpson for typing the manuscript.

REFERENCES

1. Old, R.W., Woodland, H.R., Ballantine, J.E.M., Aldridge, T.C., Newton, C.A., Bains, W.A. and Turner, P.C. (1982) Nucleic Acids Res. 10, 7561-7580.
2. Morris, N.R. (1976) Cell 9, 627.
3. Isenberg, I. (1979) Annu. Rev. Biochem. 48, 159-191.
4. Von Holt, C., Strickland, W.N., Brandt, W.F. and Strickland, M.S. (1979) FEBS Lett. 100, 201-218.
5. Newrock, K.M., Cohen, L.H., Hendrick, M.B., Donnelly, R.J. and Weinberg, E.S. (1978) Cell 14, 327-336.
6. Hentchel, C.C. and Birnstiel, M.L. (1981) Cell 25, 301-313.
7. Bustin, M. and Cole, R.O. (1970) J. Biol. Chem. 245, 1458-1460.
8. Macleod, A.R., Wong, N.C.W. and Dixon, G.H. (1977) Eur. J. Biochem. 78, 281-291.
9. Levy, S., Sures, I. and Kedes, L. (1982) J. Biol. Chem. 257, 9438-9443.
10. Flynn, J.M. and Woodland, H.R. (1980). Develop. Biol. 75, 222-230.
11. Risley, M.S. and Eckhardt, R.A. (1981). Develop. Biol. 84, 79-87.
12. Destree, O.H.J., D'Aelhart-Toorop, H.A. and Charles, R. (1973) Cell-Differ. 2, 229-242.
13. Destree, O.H.J., Hoenders, H.J., Moorman, A.F.M. and Charles, R. (1979) Biochem. Biophys. Acta 577, 61-70.
14. Yaguchi, M., Roy, C. and Seligy, V.L. (1979) Biochem. Biophys. Res. Commun. 90, 1400-1406.
15. Pehrson, J.R. and Cole, R.D. (1981) Biochemistry 20, 2298-2301.
16. Noll, M. (1976) Cell 8, 349-355.

17. Weintraub, H. (1978) Nucleic Acids Res. 8, 1179-1188.
18. Turner, P.C. and Woodland, H.R. (1982) Nucleic Acids Res. 10, 3769-3780.
19. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
20. Sanger, F., Nicklen, S. and Coulson, A.R. (1977). Proc. Nat. Acad. Sci. U.S.A. 74, 5463-5467.
21. Cole, R.D. (1977) In: The Molecular Biology of the Mammalian Genetic Apparatus (Ts'o, P.O.P. ed). Vol. 1, pp 93-104, Elsevier/North Holland Biomedical Press, Amsterdam, N.Y.
22. Corden, J., Wasylyk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) Science 209, 1406-1414.
23. Moorman, A.F. M., De Boer, P.A.J., De Laaf, R.T.M., Van Dongen, W.M.A.M. and Destree, O.H.J. (1981). FEBS Lett. 136, 45-52.
24. Moorman, A.F.M., De Boer, P.A.J., De Laaf, R.T.M. and Destree, O.H.J. (1982) FEBS Lett. 144, 235-241.
25. Mohammed, A., Aldridge, T. C. and Old, R. W. In preparation.
26. Schaffner, W., Kunz, G., Daetwyler, M., Telford, J., Smith, H.O. and Birnstiel, M.L. (1978) Cell 14, 655-671.
27. Krieg, P.A., Robins, A.J., Colman, A. and Wells, J.R.E. (1982) Nucleic Acids Res. 10, 6777-6785.
28. Sharp, P. (1981) Cell 23, 643-646.
29. Engel, J.D., Sugarman, B.J. and Dodgson, J.B. (1982) Nature 297, 434-436.
30. Harvey, R.P., Robins, A.J. and Wells, J.R.E. (1982) Nucleic Acids Res. 10, 7851-7863.
31. Seiler-Tuyns, A. and Birnstiel, M.L. (1981) J. Mol. Biol. 151, 607-625.
32. Birchmeier, C., Grosschedle, R. and Birnstiel, M.L. (1982) Cell 28, 739-745.
33. Stunnenberg, H.G. and Birnstiel, M.L. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 6201-6204.
34. Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., De Riel, J.K., Forget, B.G., Weissman, S.M., Slightom, J.L., Blechl, A.E., Smithies, O., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) Cell 21, 653-668.
35. Turner, P.C. and Woodland, H.R. (1983) Nucleic acids Res. 11, 971-986.
36. Zernik, M., Heintz, N., Boime, I. and Roeder, R.G. (1980) Cell 22, 807-822.