J. evol. Biol. 6: 129–137 (1993)

1010–061X/93/01129–09 $ 1.50 + 0.20/0

An unusual histone H4 gene from *Tilapia nilotica* exhibiting characteristics of both a replication-dependent histone and a basal-expression histone: evolutionary considerations

E. Englander, B. Moav, M. Tabachnik, Y. Shuali[1] and D. Graur[2]

Department of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv 69978, Israel

Key words: Fish; histone sequences; codon usage; evolution.

Abstract

The nucleotide sequence of a histone H4 gene from the fish *Tilapia nilotica* was determined. The deduced amino acid sequence is identical to that of H4 from the trout, *Salmo gairdnerii*. The 3′ untranslated region of the *Tilapia* gene exhibits a unique combination of structural features each of which is usually associated with either a replication-dependent histone or a basal-expression histone, but not with both. The direction of nucleotide substitutions in the *Tilapia* and *Salmo* lineages was inferred. The *Tilapia* gene was found to evolve faster and to exhibit a more biased pattern of substitution and codon usage than its *Salmo* homologue. This combination of features cannot be explained by either mutation or purifying selection. The rapid embryonic development of *Tilapia* prompts us to suggest that the molecular evolution of the histone H4 gene is driven by fixation of advantageous synonymous mutations.

Introduction

The condensation of eukaryotic nuclear DNA in the nucleosome is facilitated by histones, the major proteinaceous constituent of chromatin (Wu et al., 1986). Four different histones comprise the nucleosomal core (H2A, H2B, H3, H4). A fifth "outer histone" (H1) functions in higher folding orders (Finch and Klug, 1976).

---

[1] Deceased

[2] To whom all correspondence should be addressed

Because of the stringent constraints of the packaging function, the primary structures of histones are highly conserved even between distantly related organisms (e.g. Tabata et al., 1983). In contrast, synonymous codon positions change frequently in the course of evolution and give rise to differences in codon usage patterns among gene copies and among closely related organisms. Histone genes, therefore, display average or above average levels of nucleotide sequence variation despite an almost complete lack of variation at the amino acid level.

In most cases, a temporal and functional coupling of the expression of the histone genes with DNA replication was observed. Such histone genes are called replication-dependent (Maxson et al., 1983; Zweidler, 1984; Graves et al., 1987; Pandey and Marzluff, 1987). Replication-dependent histone genes possess a number of unusual features in common. Among these are the absence of introns, the absence of a polyadenylation signal, as indeed the absence of a poly(A) tail, and the presence of a palindromic structure near the 3′ end that may form a hairpin loop-structure (Marzluff and Graves, 1984). In comparison, histone genes which are expressed constitutively throughout the cell cycle and in non-dividing cells, the so-called basal-expression or replication-independent histones, contain both introns and a polyadenylation signal, and lack the palindromic structure (Wells and Kedes, 1985).

Histone variants that represent isoproteins with divergent patterns of expression have been found (Cohen et al., 1975; Wu et al., 1986). The non-coding regions of these variants vary considerably within and between species (Wells et al., 1986; Wells and Herrmann, 1989). The codon usage in the H3 genes was found to exhibit departures from randomness (Taylor et al., 1986). However, whereas in other genes where nonrandomness in the pattern of codon usage correlates well with phylogenetic relationships (Grantham et al., 1980; Ikemura, 1985; Sharp et al., 1988), the codon usage in histone genes was found to correlate with the mode of expression, and not to reflect phylogenetic affinities.

We have isolated and cloned several histone gene clusters from the fish *T. nilotica*, and studied their organization and gene expression (Englander and Moav, 1989). Here, we report the cloning and sequencing of a histone H4 gene, and attempt to infer the evolutionary processes affecting its evolution, in particular the evolution of its codon usage.

Data, materials and methods

DNA purification, cloning strategies, and sequencing

Peripheral blood was collected from adult *Tilapia nilotica* fish. Nuclei were pelleted and DNA was extracted by phenol. Genomic DNA was cut with *Mbo*I to yield fragments 5–20 kb in length, and the digested DNA was separated on a 10–40% linear sucrose gradient on an SW40 rotor at 25 000 rpm for 20 hours at 20° C (Maniatis et al., 1982). Since histone gene clusters are believed to be 5–8 kb in length, the longer fragments may contain an entire core histone gene cluster. In

order to construct a genomic library, the DNA fragments were ligated to the modified lambda vector EMBL-3 (Stratagene) under conditions recommended by the manufacturer. Recombinant phages were packed using an *in vitro* packaging mixture (Promega Biotec) and amplified in *E. coli* P2392 (*Stratagene*). Phage DNA was extracted and separated on a CsCl gradient. Using the *in situ* plaque hybridization method (Maniatis et al., 1982), 5×10^5 recombinant phages were screened with histone H4 probes from *Salmo* and mouse (kindly provided by Drs. W. Connor and D. Schumperli, respectively). Fourteen positive clones were identified, and 4 clones (E3H5, E3H6, E3H7, E3H50) containing the core histones were chosen for further analysis (Englander and Moav 1989). A DNA fragment derived from clone E3H50 was found to contain a complete core histone cluster, and was further subcloned. A recombinant clone containing a histone H4 gene was cut with *EcoR*I, ligated into pBR322, and cloned in *E. coli*. A 1.1 kb fragment was cut from the plasmid by *Hind*III and cloned in a M13 vector (Amersham) under conditions recommended by the manufacturer. Nucleotide sequencing was performed using the "Sequenase" sequencing kit (United States Biochemical Corporation).

Data

The sequence of the *Tilapia* H4 gene is reported in Figure 1 (EMBL Data Library, Accession Number X54078). A comparison of the coding region with those from *Salmo* and human (Sierra et al., 1983; Winkfein et al., 1985) is shown in Figure 2.

Nucleotide sequence analysis

The codon usage measure used in this study is derived from information theory and its application to nucleic acid sequences (Gatlin, 1972). The information

```
                                                  M  S  G  R  G  K  G  G  K  G  L  G  K  G
ttaccaatcgaacttggcctggttcaaaaaaccttaccATGTCTGGAAGAGGTAAAGGCGGCAAAGGACTCGGAAAAGGA

 G  A  K  R  H  R  K  V  L  R  D  N  I  Q  G  I  T  K  P  A  I  R  R  L  A  R  R
GGCGCCAAGCGTCACCGTAAGGTTCTCCGTGATAACATTCAGGGCATCACCAAACCAGCCATCCGTCGTCTGGCTCGCCG

  G  G  V  K  R  I  S  G  L  I  Y  E  E  T  R  G  V  L  K  V  F  L  E  N  V  I
TGGTGGCGTCAAGCGTATCTCTGGTCTGATCTACGAGGAGACCCGTGGTGTGTTGAAGGTGTTTCTGGAGAACGTCATCC

R  D  A  V  T  Y  T  E  H  A  K  R  K  T  V  T  A  M  D  V  V  Y  A  L  K  R  Q
GTGACGCCGTCACCTACACTGAGCACGCCAAGAGGAAGACCGTGACCGCCATGGATGTGGTGTACGCTCTGAAGAGGCAG

 G  R  T  L  Y  G  F  G  G  *                                    +25
GGCCGCACTCTGTACGGCTTCGGCGGTTAAactcatgctccttcatccatcaaacggctctttaagagcccacacacttc

actttaagggctttgttctggggccttcttgtgggtaggggttgggttgttgtgtgtggtggtgttgtttttttgttttt

                    +149                                                  +193
ttttttgtcttactacagatttcttgaaatagaaatttatagttaggaaaatgtctgggtaataactttacaatttaag

tcactcagattttattcatctagtagtgatgaccaatgaagctt
```

Fig. 1. *Tilapia nilotica* histone H4 gene. The coding region is shown in capital letters, and the amino acid sequence is shown above it. For details on the underlined sequences in the flanking regions, see text.

```
1 TCT GGC CGC GGC AAA GGC GGG AAG GGC CTT GGC AAA GGC GGC GCT AAG CGC CAC CGT
2 TCT GGA AGA GGT AAA GGC GGC AAA GGA CTC GGA AAA GGA GGC GCC AAG CGT CAC CGT
3 TCT GGA AGA GGT AAA GGC GGC AAG GGA CTC GGA AAA GGA GGC GCC AAG CGT CAC CGT
                              *

1 AAA GTA CTG CGC GAC AAT ATC CAT GGC ATC ACC AAA CCA GCC ATT CGA CGC CTT GCC
2 AAG GTT CTC CGT GAT AAC ATT CAG GGC ATC ACC AAA CCA GCC ATC CGT CGT CTG GCT
3 AAG GTT CTC CGC GAT AAC ATC CAG GGA ATC ACC AAG CCC GCC ATT CGC CGT CTG GCT
                *           *       *           *   *       *   *

1 CGC CGC GGC GGC GTG AAG CGC ATC TCC GGC CTC ATC TAC GAG GAG ACT CGC GGG GTG
2 CGC CGT GGT GGC GTC AAG CGT ATC TCT GGT CTG ATC TAC GAG GAG ACC CGT GGT GTG
3 CGC CGT GGC GGC GTG AAG CGT ATC TCC GGG CTG ATC TAC GAG GAG ACC CGC GGT GTC
              *       *               *   *                           *       *

1 CTG AAG GTG TTC TTG GAG AAC GTA ATC CGG GAC GCC GTA ACC TAT ACA GAG CAC GCC
2 TTG AAG GTG TTT CTG GAG AAC GTC ATC CGT GAC GCC GTC ACC TAC ACT GAG CAC GCC
3 CTG AAG GTG TTC CTT GAG AAC GTG ATC CGC GAC GCC GTC ACC TAC ACC GAG CAC GCC
  *             *   *           *       *                           *

1 AAG CGC AAG ACG GTC ACC GCC ATG GAT GTG GTC TAC GCG CTC AAG CGC CAG GGC CGC
2 AAG AGG AAG ACC GTG ACC GCC ATG GAT GTG GTG TAC GCT CTG AAG AGG CAG GGC CGC
3 AAG AGG AAG ACC GTT ACG GCC ATG GAC GTG GTC TAC GCT CTG AAA CGC CAG GGA CGC
                      *   *           *       *                   * * *       *

1 ACC CTC TAC GGT TTC GGT GGT
2 ACT CTG TAC GGC TTC GGC GGT
3 ACC CTG TAC GGT TTC GGC GGT
    *           *
```

Fig. 2. Comparison between H4 coding regions from human, *Tilapia* and *Salmo*. 1 = Human, 2 = *Tilapia*, 3 = *Salmo*. Asterisks denote differences between *Tilapia* and *Salmo*.

content (H) for each of the 8 four-fold codon families was calculated as $H = -\Sigma p_i \ln p_i$, where p_i is the frequency of the i-th codon. Note that in four-fold degenerate codon families, H reaches a maximum value of 1.39 when all codons are equally frequent. The smallest possible value of H is 0, when only one of the four possible synonymous codons is used.

The pattern and number of nucleotide substitutions were deduced from a comparison with an outgroup species (human).

Results

The Tilapia nilotica *histone H4 gene*

The deduced amino acid sequence of H4 histone from *Tilapia nilotica* is identical to its homologue from *Salmo gairdnerii* (Figure 2). Two hundred and fifty nucleotides were determined in the 3′ flanking region. From the comparison, a few consensus sequences were identified: 1. An inverted repeat GGCTCTT-TT-AAGAGCC, 25 bp downstream of the termination codon. 2. An inverted repeat GATTTC-TT-GAAATA, 149 bp downstream the termination codon. 3. A signal for polyadenylation, AATAA, 193 bp downstream from the termination codon. This signal, common to all basal-expression histones but not to H4 replication-dependent histone genes, was also found in the H2B histone gene of *Salmo* (Connor et al., 1984). However, we have no direct evidence for the presence of a poly(A) tail in H4 mRNA in either embryos or adult *Tilapia* tissues (Englander and Moav, 1989).

Relative rates of synonymous substitution and codon usage

We identified 30 synonymous substitutions that have occurred in the pisces lineages. The lineage in which the substitution occurred could not be determined for 6 nucleotide sites. Three of these substitutions were TC transitions, two were TG transversions, and one was a CG transversion. Out of the 24 substitutions for which the lineage in which they occurred could be determined, 15 occurred in the *Tilapia* line and 9 occurred in *Salmo* (Table 1). The numbers of synonymous substitutions per synonymous site between *Tilapia* and human and between *Salmo* and human were 1.4 ± 0.1 and 1.0 ± 0.1, respectively. The difference is statistically significant (t test, df = total number of synonymous sites $- 2 = 88 \times 2 - 2 = 174$, $P < 0.001$).

In *Salmo* the ratio of transitions to transversion was 3:6. In *Tilapia*, on the other hand, an inverse ratio of 11:4 was observed. Since about half of the substitutions in *Tilapia* were C→T transitions (8/15), the third position of codons in *Tilapia* shows an increase in the use of *T*, with about 25% of all codons ending in *T*. In comparison, only 14% and 17% of all codons in human and *Salmo*, respectively, end in *T*. In Table 2, we list the *H* values for the 8 four-fold degenerate codon

Table 1. Types of substitutions in the *Salmo* and *Tilapia* lineages.

	Type	*Salmo*	*Tilapia*
Transitions	A→G	1	o
	G→A	1	1
	T→C	1	2
	C→T	0	8
Transversions	C→A	2	1
	A→C	1	0
	G→C	1	1
	C→G	1	2
	G→T	1	0
	Total	9	15

Table 2. *H* values for four-fold degenerate codons in humans (HS), *Salmo* (SG), and *Tilapia* (TN).

Amino acid	HS	SG	TN
Leu	1.08	0.90	0.60
Val	1.06	1.06	0.94
Ser	0.69	0.69	0.00
Pro	0.00	0.00	0.00
Thr	1.15	0.41	0.60
Ala	0.80	0.60	0.60
Arg	0.75	0.68	0.47
Gly	0.80	1.24	1.06
Average	0.79	0.70	0.53

families. The *Tilapia* gene shows the highest bias in codon usage. Thus, the H4 histone gene in *Tilapia* exhibits a unique combination of elevated rates of substitution and a highly biased pattern of substitution and codon usage.

Discussion

Structural features of the Tilapia nilotica *histone H4 gene*

The *Tilapia nilotica* DNA sequence indicates that the H4 gene is a functional gene. First, the coding sequence shows a 100% identity in its deduced amino acid to histone H4 from *Salmo gairdnerii*. Second, this H4 gene is part of a complete core histone gene cluster (E3H50), all the members of which hybridized strongly to mRNA (Englander and Moav, 1989). Therefore, this H4 gene is unlikely to be a pseudogene.

The *Tilapia* H4 3′ untranslated region possesses a combination of features each of which is usually associated with either a replication-dependent gene or a basal-expression gene, but not with both. Specifically, it contains inverted repeats in the 3′ untranslated region, typical of replication-dependent genes, and a signal for polyadenylation, typical of basal-expression genes. The formation of the hairpin loop structure in the mRNA by inverted repeats has been shown to regulate the stability and degradation of the mRNA in the cytoplasm in a replication-dependent manner (Pandey and Marzluff, 1987). In its presence, the mRNA is degraded rapidly when DNA synthesis is inhibited (Graves et al., 1987). On the other hand, addition of a signal for polyadenylation to histone H4 in mouse genes gave rise to poly(A) tail in the 3′ of the mRNA (Seiler-Tuyns and Paterson, 1988). The poly(A) tail keeps the level of the polyadenylated H4 transcripts at a constant level throughout all the stages of the cell cycle, and reduces the level of expression by a factor of 30, independently of the growth demand during the cell cycle. Moreover, the poly(A) tail stabilizes the mRNA even in the presence of a hairpin loop in the 3′ (Alterman et al., 1985; Chodchoy et al., 1987) The fact that the *Tilapia* H4 gene may possess two 3′ hairpin loop structures and a polyadenylation signal constitutes an inconsistency that may generate unexpected patterns of gene expression and mRNA stability.

Mechanisms affecting the evolutionary history of the H4 gene

Basically, two findings require explanation. First, why does the *Tilapia* gene evolve faster than its homologue in *Salmo*, and second, why is the codon usage in *Tilapia* more biased than that in *Salmo*? There can be several reasons for the rate difference: (1) The mutation rate in the *Tilapia* gene may be higher than that in the *Salmo* gene. (2) The *Tilapia* gene may be less constrained than the *Salmo* gene, and thus a lesser proportion of all mutations occuring in the *Tilapia* gene are deleterious than in the *Salmo* lineage. (3) The substitutions observed in *Tilapia* represent

advantageous mutations that have been actively driven towards fixation by positive selection.

As far as possibility (1) is concerned, it has been proposed that generation time, i.e., the number of germline DNA replications per year, may be one of the causes for the differences in the rate of mutation (Li et al., 1987). In the present case, it is unlikely that the differences in the rates of substitution are due to different generation times, since the two species have about the same generation time, despite markedly different life histories (Dr. L. Fishelson, personal communication).

According to population genetics theory, if strong purifying selection is involved, the rate of substitution should be reduced and the bias in codon usage increased (Li et al., 1985; Sharp et al., 1988). In the present case, however, we observe an opposite relationship, i.e., the rate of substitution in the *Tilapia* H4 gene is higher than that in *Salmo*, and at the same time its codon usage is more biased. Therefore, purifying selection cannot explain the differences in the rates and patterns of substitution.

We must, thus, consider advantageous selection for synonymous mutations as the reason behind the synonymous rate differences between *Tilapia* and *Salmo*. In replication-dependent histones there is a strong bias towards utilization of nonrandom codons. This bias has been claimed to correlate well with the rates of production of the proteins. In other words, a gene with a biased codon usage is likely to be translated by more abundant tRNA species than a gene using synonymous codons at random (Sharp et al., 1986). If a need exists for the rapid translation of a protein, synonymous mutations changing a codon into one that is recognized by a more abundant tRNA species is likely to confer an advantage. Interestingly, such a need exists in *Tilapia* but not in *Salmo*. *Tilapia*, a tropical fish, reaches a blastula stage within 16 hours at 24° C, while *Salmo*, which develops in cold water, reaches the same stage in 30 days or more (Hochenberg-Grumet and Moav, 1983). The large demand for the rapid production of histones in *Tilapia* means that the translation and expression of histone genes must proceed 20–30 times more rapidly in *Tilapia* than in *Salmo*. The strongly biased codon usage observed in the *Tilapia* H4 gene may, thus, represent the end result of a process of selection for advantageous mutants. The fact that most substitutions in the *Tilapia* are of one type, C → T, supports our conclusion.

Acknowledgements

We thank Drs. W. Connor and D. Schumperli for their kind donation of probes, Dr. L. Fishelson for information regarding the natural history of *Tilapia*, and I. Judkiewicz for salvaging virus-infected disks. D.G. has been supported in this study by a grant from the U.S.-Israel Binational Science Foundation.

References

Alterman, R. B. M., C. Sprecher, R. A. Graves, W. F. Marzluff, and A. I. Skóultchi. 1985. Regulated expression of a chimeric mouse histone gene introduced into mouse fibroblasts. Mol. Cell. Biol 5: 2316–2324.

Chodchoy, N., B. J. Levine, C. Sprecher, A. I. Skóultchi, and W. F. Marzluff. 1987. Expession of mouse histone genes: Transcription into 3′ intergenic DNA and cryptic processing downstream from the 3′ end of H3 gene. Mol. Cell. Biol 7: 1039–1047.

Cohen, L. H., K. M. Newrock, and A. Zweidler. 1975. Stage-specific switches in histone synthesis during embryogenesis of the sea urchin. Science 190: 994–997.

Connor, W., J. C. States, J. Mezquita, and G. H. Dixon. 1984. Organization and nucleotide sequence of rainbow trout histone H2A and H3 genes. J. Mol. Evol. 20: 236–250.

Englander, E., and B. Moav. 1989. Cloning and characterization of a histone gene family in *Tilapia* fish. Biochim. Biophys. Acta 1007: 277–282.

Finch, J. T., and A. Klug. 1976. Solenoidal mode for superstructure of chromatin. Proc. Natl. Acad. Sci. USA 79: 1897–1901.

Gatlin, L. L. 1972. *Information Theory and the Living System.* Columbia University Press, New York.

Grantham, R., C. Gautier, M. Gouy, M. Mercier and a. Pavè. 1980. Codon catalog and the genome hypothesis. Nucleic Acids Res. 8: r49–r80.

Graves, R. A., N. B. Pandey, N. Chodchoy, and W. F. Marzluff. 1987. Translation is required for regulation of histone mRNA degradation. Cell 48: 615–626.

Hochenberg-Grumet, R., and B. Moav. 1983. Biosynthetic patterns of developmental histones in *Sorotherodon* species and their hybrids. In: Fishelson, L., and Z. Yaron, (eds.) *Tilapia* in Aquaculture. pp. 219–228. Tel Aviv University Press, Tel Aviv.

Ikemura, T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. 2: 13–34.

Li, W.-H., C.-C. Luo, and C.-I Wu. 1985. Evolution of DNA sequences. In: MacIntyre, R. J., (ed.) *Molecular Evolutionary Genetics.* pp. 1–94. Plenum, New York.

Li, W.-H., M. Tanimura, and P. M. Sharp. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. J. Mol. Evol. 25: 330–342.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbur, New York.

Marzluff, W. F., and R. A. Graves. 1984. Organization and expression of mouse histone genes. In: Stein G. S., J. L. Stein, and W. F. Marzluff, (eds.) *Histone Genes.* pp. 339–371. Wiley Interscience, New York.

Maxson, R., R. Cohn, and L. Kedes. 1983. Expression and organization of histone genes. Ann. Rev. Genet. 17: 239–277.

Pandey, N. B. and W. F. Marzluff. 1987. The stem-loop structure at the 3′ end of histone mRNA is necessary and sufficient for regulation of histone mRNA stability. Mol. Cell. Biol. 7: 4557–4559.

Seiler-Tuyns, A. and B. M. Paterson. 1988. A chimeric mouse histone H4 gene containing either introne or poly(A) addition signal behaves like a basal histone. Nucleic Acids Res. 14: 8845–8862.

Sharp, P. M., E. Cowe, G. Higgins, D. C. Shields, K. H. Wolfe, and F. Wright. 1988. Codon usage patterns: a review of the considerable within-species diversity. Nucleic Acids Res. 16: 8207–8211.

Sharp, P. M., T. M. F. Tuohy, and K. R. Mosurski. 1986. Codon usage in yeast: cluster analysis clearly differentiates highly expressed genes. Nucleic Acids Res. 14: 5125–5143.

Sierra, F., G. Stein, and J. Stein. 1983. Structure and in vitro transcription of a human H4 histone gene. Nucleic Acids Res. 11: 7069–7086.

Tabata, T., A. Sasaki, and M. Iwabuchi. 1983. The structural organization and DNA sequence of a wheat histone H4 gene. Nucleic Acids Res. 11: 5865–5875.

Taylor, J. D., S. E. Wellman, and W. F. Marzluff. 1986. Sequences of four mouse histone H3 genes: Implication for evolution of mouse histone genes. J. Mol. Evol. 23: 242–249.

Wells, D., and L. Kedes. 1985. Structure of a human histone cDNA: Evidence that basally expressed histone genes have intervening sequences and encode polyadenylated mRNAs. Proc. Natl. Acad. Sci. USA 82: 2834–2838.

Wells, D., W. Bains, and L. Kedes. 1986. Codon usage in histone gene families of higher eukaryotes reflects functional rather than phylogenetic relationships. J. Mol. Evol. 23: 224–241.

Wells, D., and J. Herrman. 1989. Functionally constrained codon usage in histone genes. Int. J. Biochem 21: 1–6.

Winkfein, R. J., W. Connor, J. Mezquita, and G. H. Dixon. 1985. Histone H4 and H2b genes in the rainbow trout (*Salomo gairdnerii*). J. Mol. Evol. 22: 1–19.

Wu, R. S., H. T. Panusz, C. L. Hatch, and W. M. Bonner. 1986. Histones and their modifications. CRC Crit. Rev. Biochem. 20: 201–248.

Zweidler, A. 1984. Core histone variants of the mouse: primary structure and differential expression. In: Stein, G. S., J. L. Stein, and W. F. Marzuff, (eds.) *Histone Genes*. pp. 339–371. Wiley Interscience, New York.

Received 25 August 1991;
accepted 25 April 1992.
Corresponding Editor: B. Levin