Parasitol Res (2000) 86: 259–261

SHORT COMMUNICATION

Julius Lukeš · Dmitri A. Maslov

Unexpectedly high variability of the histone H4 gene in *Leishmania*

Received: 16 August 1999 / Accepted: 10 September 1999

Abstract The sequence of the cDNA of the histone H4 gene of *Leishmania tarentolae* is reported herein. The predicted 100-amino-acid-long protein has the highest degree of identity with the histone H4 gene of *L. infantum* and shares with it a 5′ region that shows a very low degree of identity with the corresponding region of histone H4 genes from other organisms. However, between these two genes is a 7.7% nucleotide difference that results in seven different amino acids, located in the 5′, central, and 3′ regions of the coding sequence. Such a divergence in the H4 gene, which is considered to be one of the most highly conserved genes, between closely related members of the genus *Leishmania* is unexpected and may reflect some unusual features of these important proteins in kinetoplastid flagellates.

Introduction

Kinetoplastid protozoa of the genus *Leishmania* are the causative agents of a wide spectrum of human diseases occurring in most tropical and subtropical regions. Probably due to their ancient character, kinetoplastid flagellates retain many unique features, including unusual gene organization and expression (Donelson 1999). Therefore, the characterization of their histone proteins is of considerable interest. Histones are small basic proteins that are omnipresent in eukaryotic cell nuclei, where they form nucleosomes, which consist of DNA wound around a histone octamer containing two molecules each of the core histones H2A, H2B, H3, and H4. Histone H4 is an extremely slowly evolving histone protein (Thatcher and Gorowski 1994). In the course of a study of nuclear-encoded proteins associated with the respiratory complexes we accidentally obtained the cDNA of histone H4 from the species *L. tarentolae*, and the results are reported herein.

Promastigotes of *L. tarentolae* were cultivated as described elsewhere (Simpson and Braly 1971). On the basis of a partial N-terminal amino-acid sequence (DLPGKIVSV) derived from an as yet unidentified kinetoplast protein, a highly degenerate primer (CTRRANGGNCCNTTYTADCANWSNCA) was designed and, with the total poly(A) mRNA serving as a template, was used for cDNA synthesis. The subsequent polymerase chain reaction was performed with the above-mentioned primer and the mini-exon primer (Campbell et al. 1984). After 30 cycles performed at 95 °C for 2 min, at 50 °C for 1 min, and at 72 °C for 3 min, one major band and several minor bands were amplified. The most abundant 0.6-kb band was cloned into the pT7blue vector (Novagen), and both strands were sequenced (Fig. 1).

Comparison of the sequence obtained with the GenBank data base revealed similarity with several histone H4 genes, particularly with the H4 gene from *L. infantum* (Soto et al. 1997). The cDNA molecule was 585 bp long and contained in the 5′ end the 39-bp spliced leader sequence that is part of all mRNA molecules in trypanosomatids. The region between the spliced leader and the start codon of the histone H4 gene was 85 bp long and had no sequence similarity with the corresponding region of histone H4 of *L. infantum* or with other sequences. The histone H4 gene is 300 bp long, exactly the same length as the *L. infantum* H4 gene. Surprisingly, there are 23 differences between the 2 genes, located mainly in the 5′ and 3′ coding regions. Most of these nucleotide differences are not silent and result in seven differences in predicted amino-acid sequences, localized in three clusters in the 3′, 5′, and

J. Lukeš · D.A. Maslov
Department of Biology, University of California, Riverside, CA 92521, USA

J. Lukeš (✉)
Institute of Parasitology, Czech Academy of Sciences, Branišovská 31, 37005 České Budějovice, Czech Republic
e-mail: jula@paru.cas.cz
Tel.: +420-38-7775416; Fax: +420-38-5300388

The sequence data reported herein have been submitted to GenBank and assigned the accession number AF175386

Fig. 1 Sequence of the cDNA clone. The spliced leader sequence is shown in *boldface*. The protein-coding region is given in *upper-case lettering*. *Single and double asterisks* indicate the start and termination codons, respectively. The sequence identical to the motif located between ribosomal protein genes of *Leishmania infantum* is *underlined*

aactaacgctatataagtatcagtttctgtactttattgc
tgctgcccccgttctacgccttcgttctcagctcatcttc
tcacttttttccctctccacccacattcccgaaacctatcc
aaacATGGCCAAGGGTAAGCGCTCCGCTGATGCCAAGGGC
    *
AGCCAGAAGCGCCAGAAGAAGGTGCTGCGCGACAACATCC
GCGGCATCACGCGCGGCTGCGTCCGCCGTATGGCGCGCCG
CGGTGGCGTGAAGCGCATCTCGAGCGAGATCTACGAAGAG
GTGCGCCGTGTGCTGAAGGCCTACGTGGAGGACATTGTGC
GCTGCAGCACGGCCTACACCGAGTACGCGCGCAAGAAGAC
AGTGACTGCATCCGATGTTGTGAATGCGCTGCGCAAGCGC
GGCCAAATCCTGTACGGCTACGCGTAGggtcattgccaca
                        **
ccgtagtttgtactgagcgcatgcctgcagctcaccagca
cttcatgcagctcacacccaccgatagacctctttcgtt
tcgtctcttttttccctattctcccttccctccgacctac
ctggcaaaatcgtgatcggatcccc(polyA)

central regions of the protein (Fig. 2). The homology search revealed an 18-bp-long region 88 bp downstream of the stop codon, which is identical with a sequence motif located between two ribosomal protein genes of *L. infantum* (Soto et al. 1993).

Histone H4 is generally considered to be among the most highly conserved proteins, since at the amino-acid level there is no difference among vertebrate H4 genes and there are very few differences between those of vertebrates, invertebrates, and plants (Thatcher and Gorowski 1994). With the exception of *L. infantum* and *L. tarentolae*, the complete H4 sequences of other trypanosomatid flagellates are not available; however, a comparison of only partial H4 amino-acid sequences between *Trypanosoma brucei* and *T. cruzi* has revealed three substitutions (Hecker 1993). Although both are members of the monophyletic genus *Trypanosoma*, these species constitute two different lineages with relatively long branches in the ribosomal RNA-based trees (Lukeš et al. 1997).

On the other hand, *Leishmania* species generally have very similar ribosomal RNA gene sequences (Noyes et al. 1997), which indicates that in comparison with trypanosomes, radiation within the genus has occurred more recently. Therefore, the 7.7% nucleotide (7% amino-acid) sequence difference between the histone H4

Fig. 2 The inferred *L. tarentolae* amino-acid sequence of histone H4 aligned with that of *L. infantum*. Identical residues are indicated by a *dot*

L.tarentolae MAKGKRSADAKGSQKRQKKVLRDNIR
L.infantum .......T......R..........

GITRGCVRRMARRGGVKRISSEIYEEVRRVLKAYVEDIVR
.....................T.V.................

CSTAYTEYARKKTVTASDVVNALRKRGQILYGYA*
.....................T....Q.H......*

genes of *L. tarentolae* and *L. infantum* is much greater than would be expected on the basis of their close phylogenetic relatedness. Kinetoplastid histones H3 and H4 are the most divergent histones reported thus far, with an extremely low degree of conservation being noted for the 5′ region (Soto et al. 1994, 1997). We assume that such an unusual divergence may reflect observed variability in the chromatin structure (Schlimme et al. 1995) or the existence of unknown features unique for the trypanosomatid genome that are responsible for accelerated evolution of this gene. Since the Southern hybridization of total DNA digested with the hexanucleotide-target restriction enzymes detected the presence of several copies of the histone H4 gene in the *L. tarentolae* genome (data not shown), whereas seven copies are known from the *L. infantum* genome (Soto et al. 1997), the differences may also reflect the sequence heterogeneity of this protein within the genome. Sequencing of the entire genome of *L. major*, which is under way (Ivens and Blackwell 1999), and the availability of more trypanosomatid histone H4 sequences may shed more light on this interesting discrepancy.

Acknowledgements This work was supported in part by grant A6022903 from the Grant Agency of the Czech Academy of Sciences, awarded to J.L., and by grant AI40634 from the National Institutes of Health, awarded to D.A.M., who is also a Burroughs Wellcome New Investigator in Molecular Parasitology.

References

Campbell D, Thornton D, Boothroyd J (1984) Apparent discontinuous transcription of *Trypanosoma brucei* variant surface antigen genes. Nature 311: 350–355

Donelson JE, Gardner MJ, El-Sayed MN (1999) More surprises from Kinetoplastida. Proc Natl Acad Sci USA 96: 2579–2581

Hecker H (1993) Man and sea urchin – more closely related than African and American trypanosomes. Parasitol Today 9: 57

Ivens AC, Blackwell JM (1999) The *Leishmania* genome comes of age. Parasitol Today 15: 225–231

Lukeš J, Jirků M, Doležel D, Král'ová I, Hollar L, Maslov DA (1997) Analysis of ribosomal RNA genes suggests that trypanosomes are monophyletic. J Mol Evol 44: 521–527

Noyes HA, Arana BA, Chance ML, Maingon R (1997) The *Leishmania hertigi* (Kinetoplastida; Trypanosomatidae) complex and the lizard *Leishmania*: their classification and evidence for a neotropical origin of the *Leishmania-Endotrypanum* clade. J Eukaryotic Microbiol 44: 511–517

Schlimme W, Burri M, Betschart B, Hecker H (1995) Properties of the histones and functional aspects of the soluble chromatin of epimastigote *Trypanosoma cruzi*. Acta Trop (Basel) 60: 141–154

Simpson L, Braly P (1971) Synchronization of *Leishmania tarentolae* by hydroxyurea. J Protozool 17: 511–517

Soto M, Requena JM, Garcia M, Gomez LC, Navarrete I, Alonso C (1993) Genomic organization and expression of two independent gene arrays coding for two antigenic acidic ribosomal proteins of *Leishmania*. J Biol Chem 29: 21835–21843

Soto M, Requena JM, Morales G, Alonso C (1994) The *Leishmania infantum* histone H3 possesses an extremely divergent N-terminal domain. Biochim Biophys Acta 1219: 533–535

Soto M, Quijada L, Alonso C, Requena JM (1997) Molecular cloning and analysis of expression of the *Leishmania infantum* histone H4 genes. Mol Biochem Parasitol 90: 439–447

Thatcher TH, Gorowski MA (1994) Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. Nucleic Acids Res 22: 174–179