Core Histones of the Amitochondriate Protist, *Giardia lamblia*

Gang Wu, Andrew G. McArthur,† András Fiser,‡ Andrej Šali,‡ Mitchell L. Sogin,† and Miklós Müller**

*Laboratory of Biochemical Parasitology, The Rockefeller University, New York; †The Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, Massachusetts; and ‡Laboratory of Molecular Biophysics, The Rockefeller University, New York

Genes coding for the core histones H2a, H2b, H3, and H4 of *Giardia lamblia* were sequenced. A conserved organism- and gene-specific element, GRGCGCAGATTTVGG, was found upstream of the coding region in all core histone genes. The derived amino acid sequences of all four histones were similar to their homologs in other eukaryotes, although they were among the most divergent members of this protein family. Comparative protein structure modeling combined with energy evaluation of the resulting models indicated that the *G. lamblia* core histones individually and together can assume the same three-dimensional structures that were established by X-ray crystallography for *Xenopus laevis* histones and the nucleosome core particle. Since *G. lamblia* represents one of the earliest-diverging eukaryotes in many different molecular trees, the structure of its histones is potentially of relevance to understanding histone evolution. The *G. lamblia* proteins do not represent an intermediate stage between archaeal and eukaryotic histones.

Introduction

Histones are basic structural proteins that play an important role in DNA organization and gene regulation in eukaryotes. Histone-like proteins play a similar role in archaea of the Euryachaeota lineage (Isenberg 1979; Pereira and Reeve 1998). Histones are classified into five main types, the core histones H2a, H2b, H3, H4 and the linker histone H1. Two molecules each of the four core histone types are arranged in an octameric structure. Around this octamer, a 146-bp segment of DNA is coiled (Hayes, Clark, and Wolffe 1991). The amino-terminal tails of the core histones interact loosely with DNA and histone H1. The octameric histones and DNA make up the nucleosome core particle, the unit of chromatin organization (Luger et al. 1997). The three-dimensional (3D) structure of the vertebrate nucleosome core particle has been elucidated by X-ray crystallography at 2.8 Å resolution (Luger et al. 1997), confirming earlier results obtained at lower resolution (Arents et al. 1991).

The universal role of histones in eukaryotic chromatin is reflected by their remarkable conservation. The core histones are regarded as one of the most conserved protein families. All four core histones contain a region that forms the easily recognized histone fold, consisting of three α-helices connected by short loops (Luger et al. 1997). The histone folds represent a major part of these proteins. Behind this structural conservation lies an extreme sequence conservation (Isenberg 1979; Wells 1986; Thatcher and Gorovsky 1994; Makalowska et al. 1999). Vertebrate H3 and H4 sequences are almost identical, and the identities across most known H4 genes exceed 95%.

Abbreviations: 3D, three-dimensional; ORF, open reading frame; PDB, protein database.

Key words: evolution, *Giardia lamblia*, histone, promoter, three-dimensional structure.

Address for correspondence and reprints: Miklós Müller, The Rockefeller University, 1230 York Avenue, New York, New York 10021. E-mail: mmuller@rockvax.rockefeller.edu.

Mol. Biol. Evol. 17(8):1156–1163. 2000

No histones are found in bacteria or in many archaea, which have a chromosome organization different from that of eukaryotes (Sandman, Pereira, and Reeve 1998). Histone-like proteins have been observed, however, in archaea belonging to the lineage Euryarchaeota, comprising methanogens and related organisms (Sandman, Pereira, and Reeve 1998). These proteins are smaller than eukaryotic histones and correspond essentially to the eukaryotic histone fold region. Even in this homologous domain, the two groups of proteins are markedly divergent. Archaeal histone-like proteins do not show a differentiation into the four types observed in eukaryotes. Analysis of NMR data complemented with structural modeling revealed that while divergent in sequence, histone-like proteins of the archaeon *Methanothermus fervidus* form typical histone folds (Starich et al. 1996), which are essentially the same as the crystallographically determined 3D structures of all four vertebrate histone types (Luger et al. 1997). The limited but clear sequence similarity and essentially identical 3D structures of eukaryotic histones and archaeal histone-like proteins show that these proteins derive from common ancestral molecules (Sandman, Pereira, and Reeve 1998).

Although numerous eukaryotic histone molecules have been studied, those of unicellular eukaryotes (protists) have received limited attention (Makalowska et. al. 1999). Since the dominant part of eukaryotic diversity is represented by protists (Patterson and Sogin 1992), a more detailed exploration of protists promises to shed light on the limits placed on divergence among eukaryotic histones. In this paper, we present the sequences, the predicted 3D structures and their energy evaluations, and the promoter of the core histones in the highly divergent amitochondriate eukaryote *Giardia lamblia* (Sogin et al. 1989; Adam 1991; Upcroft and Upcroft 1998). Molecular phylogenetic analyses have identified *Giardia,* and related diplomonads, as an early diverging eukaryotic lineage that may have retained some primitive characteristics of the first nucleated cells (Leipe et al. 1993; Stiller and Hall 1997; Roger et al.

1999; but see Embly and Hirt 1998; Stiller et al. 1998; Hirt et al. 1999). Here, we examine the predicted structure of *Giardia* DNA-binding proteins and find them to be similar to those of other eukaryotes and not intermediary between eukaryotes and Archaea.

Materials and Methods

Organism and Genomic Clones

Giardia lamblia strain WB clone 6 (ATCC 30957) was used throughout this study. The *G. lamblia* genome project (www.mbl.edu/Giardia; Henze et al. 1998) provided the gDNA clones used in this study.

Clone Sequencing and Sequence Evaluation

A number of random gDNA clones obtained by the *G. lamblia* genome project contained complete core histone genes as indicated by BLASTX (Altschul et al. 1997) annotation of single-pass sequencing reads on the project's web page (www.mbl.edu/Giardia). One clone for each of the four core histones was resequenced on both strands and analyzed further. These clones were CI0342 for H2a, NF0363 for H2b, CI0136 for H3, and MD0903 for H4. Motif searches of all *G. lamblia* genome project sequences and NCBI nucleotide databases were performed using the GREF utility of the SEALS package (Walker and Koonin 1997).

Alignment of the Individual Histone Genes

The derived amino acid sequences of the *G. lamblia* open reading frames (ORFs) were first manually aligned with homologous sequences from public databases with the ED program of the MUST package (Philippe 1993). The alignments were subsequently refined with a method that considers the putative 3D structures of the molecules (Šali and Blundell 1993). In view of the high conservation of core histones within most eukaryotic lineages, only one sequence each was selected to represent the animals, plants, and fungi. Protists were sampled more exhaustively. Histone-like proteins from *Methanothermus fervidus* and *Pyrococcus* sp. were used as representatives of archaeal homologs and were aligned only with H4, which is most similar to archaeal histone-like proteins.

Comparative Modeling and Model Evaluation

Comparative 3D models of the histone subunits and the putative nucleosome core particle of *G. lamblia* were constructed by the program MODELLER-5 (Šali and Blundell 1993; Sánchez and Šali 1997) based on the crystallographic structure of the reconstituted *Xenopus laevis* nucleosome core particle (PDB code 1aoi) (Luger et al. 1997). The input to the program is the alignment of the target sequence to be modeled with the known template structures. The output obtained without any user intervention is a 3D model of the target with all nonhydrogen atoms. The multiple-sequence alignments for modeling were calculated by the MALIGN command of MODELLER-5. The accuracy of the models was subsequently evaluated with the program ProsaII (Sippl 1993). The *Z* score given by this program approximates the free energy of a model related to that of a random structure, expressed in units of standard deviation. The more negative the *Z* score, the more accurate is the model likely to be. Larger structures tend to have more negative *Z* scores. The terminal parts of the models that did not overlap with the template structures were omitted from the evaluations. To judge the significance of the *Z* scores of the model, *Z* scores of the crystallographic structures of the template *Xenopus* histones were also calculated.

Sequence Availability

The nucleotide sequence data reported here have been submitted to the GenBank database under accession numbers AF139873–AF139876.

Results

Derived Amino Acid Sequences

The ORFs, uninterrupted by introns, contained in the *G. lamblia* histone H2a, H2b, H3, and H4 genes corresponded to putative translation products of 124, 130, 146, and 99 amino acid residues, respectively (fig. 1). The derived sequences were similar in length to their eukaryotic homologs. As is the case for all histones, the corresponding proteins were rich in positively charged amino acids. Calculated pI values were 10.48 for H2a, 9.38 for H2b, 10.58 for H3, and 10.79 for H4.

The alignment of each gene with its homologs from other organisms was unambiguous. The only exceptions were the amino-termini of H2a and H2b and the carboxyl-terminus of H2a. Remarkably high amino acid conservation was noted in the part of the sequences corresponding to the three α-helices forming the histone fold (table 1). For each core histone, the fungus-plant-animal group showed the highest intragroup amino acid identity. The *G. lamblia* homologs, compared with this group, were consistently less similar. Other protists exhibited various, often intermediate, levels of amino acid identity. Histones H2a, H2b, and H3 of the kinetoplastid protist *Trypanosoma cruzi* showed the lowest identity with their *G. lamblia* homologs, revealing the greatest divergence in the sample studied. Interestingly, the identities of the *G. lamblia* and *T. cruzi* histones with the proteins from other organisms were in the same range. Alignment of the *G. lamblia* core histones with archaeal histone-like proteins was very difficult but revealed the conservation of a number of residues (Pereira and Reeve 1998). Eukaryotic histones clearly represent a separate group from the archaeal histone-like proteins.

Compared with other eukaryotic H2a sequences, *G. lamblia* histone H2a showed unique deletions of three amino acid residues at positions 113–115. The divergent regions were located mostly at two termini of the protein.

Giardia lamblia H2b had a six-unique-amino-acid-residue insertion in the loop between the α-1 and the α-2 helices. The rather divergent amino-terminal regions were of various lengths in different species but were

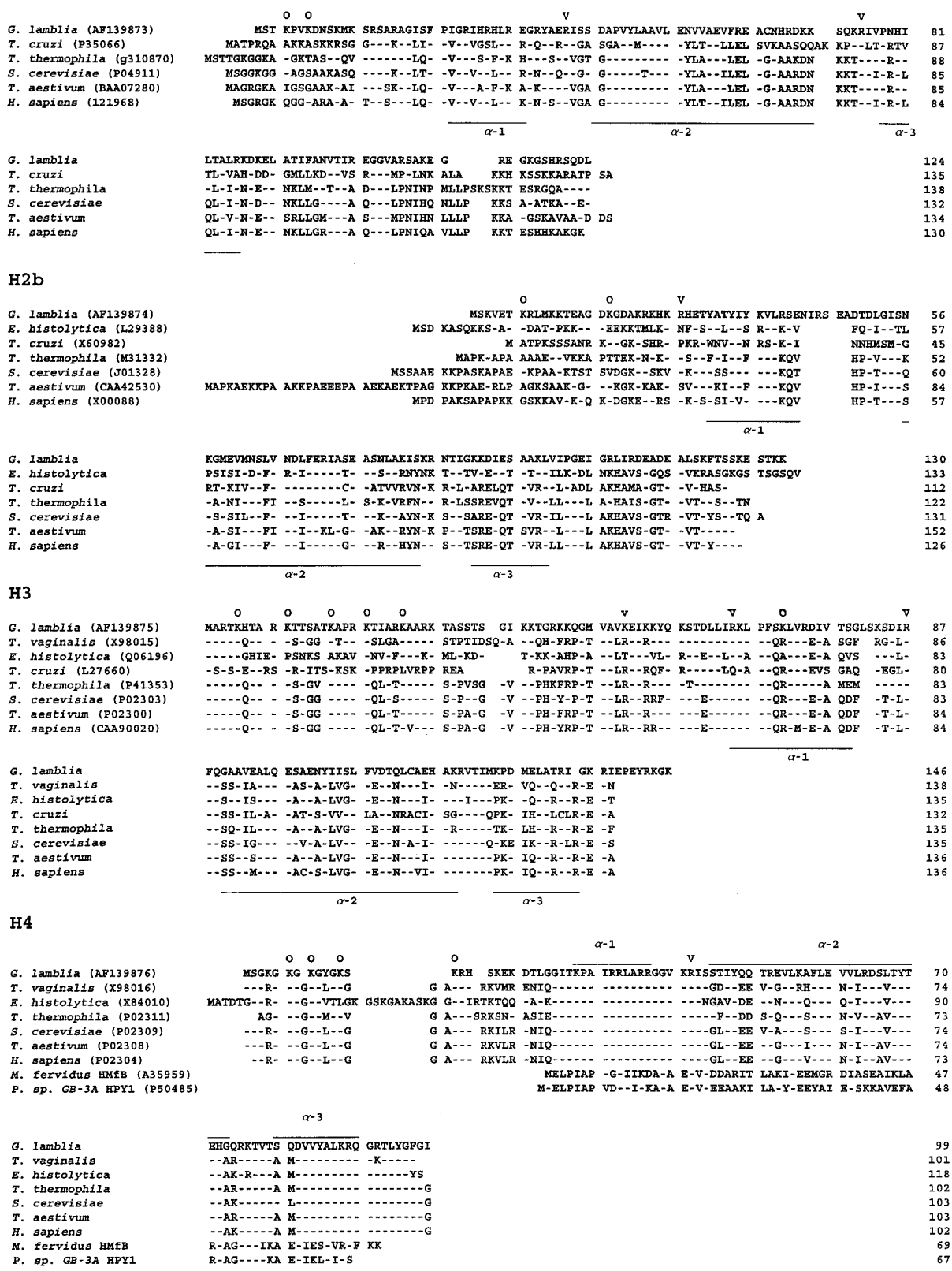

FIG. 1.—Alignment of the derived amino acid sequences of *Giardia lamblia* core histones with a selected set of homologs from other eukaryotes. Residues identical to those in the first sequence are indicated with dashes; deletions are indicated with empty spaces. The three α-helices forming the histone fold are underlined or marked at the top (H4). ''V'' designates an arginine side chain that in the *Xenopus* nucleosome is inserted into the DNA minor groove (Luger et al. 1997). These are shifted by one residue in *G. lamblia* H2b. In H3, one of these is replaced by a lysine. ''O'' designates residues that are involved in acetylation in the *Xenopus* nucleosome (Luger et al. 1997). Database accession numbers are given after the species names.

Table 1
Amino Acid Identities (as percentages) Between Histones H2a, H2b, H3, and H4 of Various Organisms

H2a	*Giardia lamblia*	*Trypanosoma cruzi*	*Tetrahymena thermophila*	*Saccharomyces cervisiae*	*Triticum aestivum*
T. cruzi	35.5				
T. thermophila	54.8	48.4			
S. cerevisiae	50.0	56.5	74.2		
T. aestivum	53.2	53.2	87.1	77.4	
Homo sapiens	48.4	56.5	79.0	88.7	80.7

H2b	*G. lamblia*	*Entamoeba histolytica*	*T. cruzi*	*T. thermophila*	*S. cerevisiae*	*T. aestivum*
E. histolytica	50.0					
T. cruzi	37.5	39.0				
T. thermophila	48.5	46.9	50.0			
S. cerevisiae	53.1	56.2	51.6	64.0		
T. aestivum	45.3	53.1	46.9	68.7	71.9	
H. sapiens	54.7	53.1	46.9	68.7	79.7	78.1

H3	*G. lamblia*	*Trichomonas vaginalis*	*E. histolytica*	*T. cruzi*	*T. thermophila*	*S. cerevisiae*	*T. aestivum*
T. vaginalis	57.3						
E. histolytica	63.2	76.5					
T. cruzi	41.2	54.4	55.9				
T. thermophila	64.7	76.5	79.4	55.9			
S. cerevisiae	58.8	73.5	73.5	61.8	72.1		
T. aestivum	63.2	82.3	86.8	58.8	80.9	83.8	
H. sapiens	58.8	77.9	79.4	58.8	75.0	79.4	92.6

H4	*G. lamblia*	*T. vaginalis*	*E. histolytica*	*T. thermophila*	*S. cerevisiae*	*T. aestivum*	*H. sapiens*	HMfB
T. vaginalis	75.8							
E. histolytica	74.2	82.3						
T. thermophila	77.4	83.9	79.0					
S. cerevisiae	79.0	87.1	82.3	80.6				
T. aestivum	77.4	91.9	83.9	87.1	87.1			
H. sapiens	77.4	90.3	85.5	85.5	88.7	96.8		
Methanococcus fervidus Hmfb	21.0	25.8	21.0	25.8	22.6	25.8	25.8	
Pyrococcus sp. GB-3A Hpy1	27.4	30.1	27.4	32.3	29.0	32.3	32.3	53.2

NOTE.—Matrices are calculated for the histone fold region from $\alpha - 1$ to $\alpha - 3$. The numbers of positions analyzed were 62 in H2a, 63 in H2b, 67 in H3, and 66 in H4. Histone-like proteins of archaea were compared with H4.

consistently rich in lysine and arginine. The carboxyl-terminal region of the *G. lamblia* H2b, similar to the *Entamoeba histolytica* homolog (Sánchez, Enea, and Eichinger 1994), is somewhat longer than other H2b sequences.

The H3 sequences are similar over their entire length. In *G. lamblia* H3, there is an insertion of two residues, but in a somewhat uncertain position, in the loop between helices α-1 and α-2 and a one-residue deletion close to the carboxyl-terminus. For the best alignment, short gaps had to be introduced into the amino-terminal extensions of protist H3 histones. *Giardia lamblia* H3 is distinguished by an eight-amino-acid extension at its carboxyl-terminus.

Histone H4 is the most conserved core histone, and *G. lamblia H4* was no exception. Single-residue gaps were required for the best alignment of the amino-terminal region. The 10-residue insertion in *E. histolytica* H4 (Binder et al. 1995) was not noted in other H4 sequences.

Of the seven arginine residues in *Xenopus* histones known to be inserted into the DNA minor grove (Luger et al. 1997), six were conserved in *G. lamblia* and one was replaced by lysine (fig. 1). Similarly, almost all lysines known to serve as acetylation sites (Luger et al. 1997) are present in *G. lamblia.*

Protein Structure Modeling

Three-dimensional models of the four *G. lamblia* core histones were built by comparative modeling based on the crystallographic structure of the *X. laevis* nucleosome core particle (Luger et al. 1997) (fig. 2). The four available template folds from the vertebrate nucleosome core particle, corresponding to the four core histones, were used independently as templates to obtain four different models for each of the four histone sequences in *G. lamblia.* Accuracies of the models were quantified by the ProsaII *Z* score (Sippl 1993) (table 2). The actual structures of the *G. lamblia* subunits are expected to be most similar to those vertebrate structures that result in the best comparative models as evaluated by the *Z* score. Furthermore, because the best comparative models have *Z* scores close to those for the template structures, the

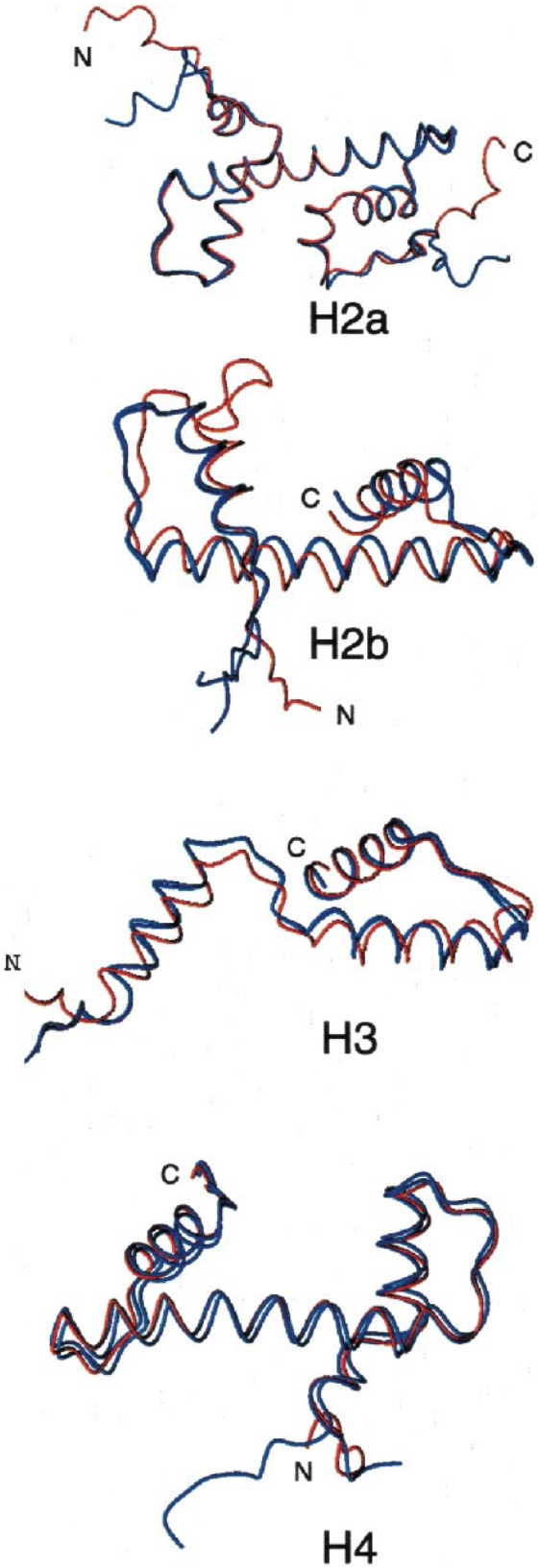

FIG. 2.—Comparative protein structure models of the *Giardia lamblia* core histones (red) based on the known structures of the *Xenopus* histones (blue) (Luger et al. 1997). The models and their evaluations indicate that the sequences of the *G. lamblia* are consistent with the structure of the corresponding *Xenopus* histones, with the exception of their terminal extensions. The noted longer insertion of a six-residue loop in H2b protrudes into the solvent, in the monomer as well as in the octamer complex, without any significant interactions with the rest of the protein.

actual *G. lamblia* structures are likely to be very similar to the corresponding *X. laevis* structures (table 2).

In addition to the models for the individual subunits, a model of the putative *G. lamblia* histone core complex was constructed based on the *X. laevis* eight-chain nucleosome core particle (a pair of four-histone chains). The two halves of the complex (chains A, B, C, D and E, F, G, H) were evaluated as separate units by ProsaII, including in the evaluation all of the interchain nonbonded contacts. The evaluation of the tetramers is consistent with that of monomers (table 2). The models of the tetramer of four *G. lamblia* histones built on chains A–D and E–H of the *X. laevis* core complex have *Z* scores of −11.1 and −10.2, respectively. These values compare well with the scores for the *Xenopus* A–D and E–H tetramers of −10.2 and −10.0, respectively.

Conserved Upstream Sequence Motif

In the upstream untranslated regions of the genes, no canonical TATA and CAAT boxes could be recognized, a circumstance noted for other *Giardia* genes (Gillin et al. 1990; Yee and Dennis 1994; Katiyar, Visvesvara, and Edlind 1995). However, a 15-residue-long motif, GRGCGCAGATTTVGG, was detected in all *G. lamblia* histone genes, located at −41 in H2a, at −34 in H2b, at −35 in H3, and at −34 in H4 (fig. 3). A search of the *G. lamblia* genomic database (about 1.3-fold coverage of the genome, >15 Mb) recognized this motif only in histone genes, indicating its special role in *G. lamblia* histone transcription or translation. A search of the nonredundant GenBank database showed only 35 hits, none of which concerned histone genes. Stretches of adenine found near the start codon in many other upstream sequences of *Giardia* genes were present also in *Giardia* core histone genes (Katiyar, Visvesvara, and Edlind 1995).

Discussion

This study showed that the amitochondriate diplomonad *G. lamblia* contains genes coding for all four core histones. This is in agreement with the results of a preliminary study using SDS-PAGE (Wu, Li, and Lu 1996). The putative translation products correspond to typical eukaryotic histones. They contain the residues critical both in the assembly of histone octamers and in the wrapping of DNA around them, i.e., in the formation of the nucleosome core particle (Luger et al. 1997). The presence of canonical sites for acetylation in the sequences (Luger et al. 1997) and the recognition of genes coding for histone acetylases and deacetylases in the *Giardia* genome database (www.mbl.edu/Giardia) indicate similar regulation processes. Nucleosomes have not been reported yet for this organism, but the data also

Table 2
Evaluation of Comparative Protein Structure Models for *Giardia lamblia* Core Histones

	Xenopus laevis Histone Template Structures			
	H2a	H2b	H3	H4
G. lamblia Histone Models	1aoiC, 1aoiG	1aoiD, 1aoiH	1aoiA, 1aoiE	1aoiB, 1aoiF
H2a	−4.74	−3.42	−0.64	−2.77
H2b	−1.15	−4.34	−0.41	−1.70
H3	−1.35	−0.61	−2.38	−0.41
H4	−2.29	−2.82	−0.26	−4.79
X. laevis Native Structures	−5.41, −5.09	−3.89, −4.05	−2.74, −2.39	−5.23, −4.42

Note.—The models were built on different template structures, corresponding to the four pairs of histone units in the *Xenopus* nucleosome (Protein Data Bank code 1aoi). The values shown are *Z* scores determined with the ProsaII program. The models with the best *Z* scores are underlined. As a reference, *Z* scores are also calculated for the template structures.

suggest that those proteins have similar functions, as in other eukaryotes.

Although the *G. lamblia* histones show some divergent features, these only marginally exceed those observed in the histones of other protists (Bender et al. 1992; Sadler and Brunk 1992; Födinger et al. 1993; Sánchez, Enea, and Eichinger 1994; Binder et al. 1995; Marinets et al. 1996; Galanti et al. 1998). In essence, our results expanded the known sequence space explored by histones in eukaryotic diversification but indicated no unique position for the *G. lamblia* histones. Histone H1, which does not form part of the nucleosome core particle and plays a role in the linking of separate nucleosomes (Garrard 1991), has not been detected by the *Giardia* genome project so far.

The conservation of core histones is a consequence of their role in maintaining chromatin organization and gene regulation. Most of the sequence divergence was found in the amino- and some carboxyl-terminal regions, while the α-helix regions and the loops between α-helices retained highly conserved structures. The six-amino-acid insertion in *G. lamblia* H2b is located just after an α-1 helix without disrupting the main structure. The unique eight-residue extension at the carboxyl-terminus of H3 does not form an α-helix but may affect the interaction of H3 and H4.

Structure modeling shows that *G. lamblia* histones, individually and together, can assume 3D structures indistinguishable from those established for vertebrate histones and nucleosomes. Thus, the divergence noted in their amino acid sequences does not exceed the structural constraints imposed by their role in chromatin organization and gene regulation.

The upstream motif, GRGCGCAGATTTVGG, detected in all *G. lamblia* core histone genes differs from conserved upstream motifs that have been found in various other organisms such as the yeast *Schizosaccharomyces pombe* (Matsumoto and Yanagida 1985), the nematode *Caenorhabditis elegans* (Roberts, Emmons, and Childs 1989), and the green alga *Chlamydomonas reinhardtii* (Fabry et al. 1995). This motif is possibly organism- and gene-specific and may play a role in assuring a correlated transcription of the histone genes.

The protist *G. lamblia* is a typical eukaryote. Several of its features, however, reveal it to be one of the most divergent representatives of its group (Adam 1991; Upcroft and Upcroft 1998). It contains no morphologically or biochemically recognizable mitochondria and has a fermentative core metabolism, several enzymes of which are not found in typical mitochondriate eukaryotes (Brown et al. 1998; Müller 1998; Sánchez 1998; Sánchez et al. 1999). In spite of this great biological divergence, the core histones of *G. lamblia* were found to be typical for a eukaryote. This finding supports the notion that the typical eukaryotic chromatin organization was present in the common ancestor of all eukaryotes and underwent only minor adjustments during their diversification (Starich et al. 1996). The highly unusual chromatin structure of dinoflagellates is an important exception, which probably arose secondarily (Vernet et al. 1990).

Histone-like proteins are found also in archaea of the Euryarchaeota lineage (Pereira and Reeve 1998). These proteins, which form nucleosomes (Starich et al. 1996), are smaller than eukaryotic histones and correspond to their central domain, i.e., the histone fold. The 3D structures of the archaeal histone-like proteins and of the nucleosome are virtually identical to those seen in eukaryotes (Starich et al. 1996). These structural similarities and the clear homology of archaeal and eukaryotic histones demonstrate their shared ancestry (Slesarev et al. 1998). At the same time, the lack of diversification

```
TTAGTTAGAGATGTCCAG GAGCGCAGATTTAGG CATAATTTCAGTTTAAATTGTCCGCGCAGATAAAGAAAGCCATG   H2a
TTTTATAGAAGTGCACAG GGGCGCAGATTTCGG CTGCGGAAGGGGGGAAAAAGCGGGCGGGAAAGAAATG         H2b
GGCGAAGTTAACTTAAGA GGGCGCAGATTTGGG ATCAATCTTTTTGGCCAAAAAGGGCGGGGAACAAAATG        H3
GTTCCTTCCCTGGACGTA GGGCGCAGATTTGGG CTAAAAAAGAAGACGCGGGAAGGGGCAAATAAAAATG        H4
                   ---------------

                   GRGCGCAGATTTVGG
```

Fig. 3.—Conserved motif in the upstream untranslated sequences of *Giardia lamblia* core histone genes. Only about 80 bp from the start codon is shown. The conserved motif is underlined and in boldface type. The consensus sequence is shown in the last line.

into four separate types and the absence of amino- and carboxy-terminal extensions in archaeal histone-like proteins clearly separate this group of proteins from eukaryotic histones. So far, no extant organism has been found that would display characters intermediate between archaeal and eukaryotic histones. The highly divergent protist *G. lamblia* is no exception.

Acknowledgments

We thank Hilary G. Morrison (Woods Hole, Mass.) for providing the clones for sequencing, and Hervé Philippe and Philippe Lopez (Orsay, France) for the MUST package. We also thank Lidya B. Sánchez, Katrin Henze, and Jennifer A. Lee for help and advice. Oligonucleotide synthesis and DNA sequencing at the Rockefeller University were performed by the Nucleic Acid Sequencing Facility. This research was supported by U.S. Public Health Service National Institutes of Health grants AI11942 to M.M., AI43273 and GM32964 to M.L.S., and GM54762 to A.Š., as well as National Science Foundation grant BIR-9601845 to A.Š. A.Š. is a Sinsheimer Scholar and an Alfred P. Sloan Research Fellow. A.F. is a Burroughs Welcome Fellow.

LITERATURE CITED

ADAM, R. D. 1991. The biology of *Giardia* spp. Microbiol. Rev. **55**:706–732.

ALTSCHUL, S. F., T. L. MADDEN, A. A. SCHAFFER, J. ZHANG, Z. ZHANG, W. MILLER, and D. J. LIPMAN. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.

ARENTS, G., R. W. BURLINGAME, B. C. WANG, W. E. LOVE, and E. N. MOUDRIANAKIS. 1991. The nucleosomal core histone octamer at 3.1 Å resolution: a tripartite protein assembly and a left-handed superhelix. Proc. Natl. Acad. Sci. USA **88**:10148–10152.

BENDER, K., B. BETSCHART, J. SCHALLER, U. KAMPFER, and H. HECKER. 1992. Sequence differences between histones of procyclic *Trypanosoma brucei brucei* and higher eukaryotes. Parasitology **105**:97–104.

BINDER, M., S. ORTNER, B. PLAIMAUER, M. FÖDINGER, G. WIEDERMANN, O. SCHEINER, and M. DUCHÊNE. 1995. Sequence and organization of an unusual histone H4 gene in the human parasite *Entamoeba histolytica*. Mol. Biochem. Parasitol. **71**:243–247.

BROWN, D. M., J. A. UPCROFT, M. R. EDWARDS, and P. UPCROFT. 1998. Anaerobic bacterial metabolism in the ancient eukaryote *Giardia duodenalis*. Int. J. Parasitol. **28**:149–164.

EMBLEY, T. M., and R. P. HIRT. 1998. Early branching eukaryotes? Curr. Opin. Genet. Dev. **8**:624–629.

FABRY, S., K. MÜLLER, A. LINDAUER, P. B. PARK, T. CORNELIUS, and R. SCHMITT. 1995. The organization structure and regulatory elements of *Chlamydomonas* histone genes reveal features linking plant and animal genes. Curr. Genet. **28**:333–343.

FÖDINGER, M., S. ORTNER, B. PLAIMAUER, G. WIEDERMANN, O. SCHEINER, and M. DUCHÉNE. 1993. Pathogenic *Entamoeba histolytica*: cDNA cloning of a histone H3 with a divergent primary structure. Mol. Biochem. Parasitol. **59**: 315–322.

GALANTI, N., M. GALINDO, V. SABAJ, I. ESPINOZA, and G. C. TORO. 1998. Histone genes in trypanosomatids. Parasitol. Today **14**:64–70.

GARRARD, W. T. 1991. Histone H1 and the conformation of transcriptionally active chromatin. BioEssays **13**:87–88.

GILLIN, F. D., P. HAGBLOM, J. HARWOOD, S. B. ALEY, D. S. REINER, M. MCCAFFERY, M. SO, and D. G. GUINEY. 1990. Isolation and expression of the gene for a major surface protein of *Giardia lamblia*. Proc. Natl. Acad. Sci. USA **87**: 4463–4467.

HAYES, J. J., D. J. CLARK, and A. P. WOLFFE. 1991. Histone contributions to the structure of DNA in the nucleosome. Proc. Natl. Acad. Sci. USA **88**:6829–6833.

HENZE, K., H. G. MORRISON, M. L. SOGIN, and M. MÜLLER. 1998. Sequence and phylogenetic position of a class II aldolase gene in the amitochondriate protist, *Giardia lamblia*. Gene **222**:163–168.

HIRT, R. P., J. M. LOGSDON, B. HEALY, M. W. DOREY, W. F. DOOLITTLE, and T. M. EMBLY. 1999. Microsporidia are related to Fungi: evidence from the largest subunit of RNA polymerase II and other proteins. Proc. Natl. Acad. Sci. USA **96**:580–585.

ISENBERG, I. 1979. Histones. Annu. Rev. Biochem. **48**:159–161.

KATIYAR, S. K., G. S. VISVESVARA, and T. D. EDLIND. 1995. Comparisons of ribosomal RNA sequences from amitochondrial protozoa: implications for processing, mRNA binding and paromomycin suceptibility. Gene **152**:27–33.

LEIPE, D. D., J. H. GUNDERSON, T. A. NERAD, and M. L. SOGIN. 1993. Small subunit ribosomal RNA+ of *Hexamita inflata* and the quest for the first branch in the eukaryotic tree. Mol. Biochem. Parasitol. **59**:41–48.

LUGER, K., A. W. MÄDER, R. K. RICHMOND, D. F. SARGENT, and T. J. RICHMOND. 1997. Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature **389**:251–260.

MAKALOWSKA, I., E. S. FERLANTI, A. D. BAXEVANIS, and D. LANDSMAN. 1999. Histone sequence database: sequences, structures, post-translational modifications and genetic loci. Nucleic Acids Res. **27**:323–324.

MARINETS, A., M. MÜLLER, P. J. JOHNSON, J. KULDA, O. SCHEINER, G. WIEDERMANN, and M. DUCHÊNE. 1996. The sequence and organization of the core histone H3 and H4 genes in the early branching amitochondriate protist *Trichomonas vaginalis*. J. Mol. Evol. **43**:563–571.

MATSUMOTO, S., and M. YANAGIDA. 1985. Histone gene organization of fission yeast: a common upstream sequence. EMBO J. **4**:3531–3538.

MÜLLER, M. 1998. Enzymes and compartmentation of core energy metabolism of anaerobic protists—a special case in eukaryotic evolution? Pp. 109–131 *in* G. H. COOMBS, K. VICKERMAN, M. A. SLEIGH, and A. WARREN, eds. Evolutionary relationships among protozoa. Kluwer, Dordrecht, the Netherlands.

PATTERSON, D. J., and M. L. SOGIN. 1992. Eukaryote origins and protistan diversity. Pp. 13–46 *in* H. HARTMAN and K. MATSUNO, eds. The origins and evolution of the cell. World Scientific, Singapore.

PEREIRA, S. L., and J. N. REEVE. 1998. Histones and nucleosomes in Archaea and Eukarya: a comparative analysis. Extremophiles **2**:141–148.

PHILIPPE, H. 1993. MUST, a computer package of management utilities for sequences and trees. Nucleic Acids Res. **21**: 5264–5272.

ROBERTS, S. B., S. W. EMMONS, and G. CHILDS. 1989. Nucleotide sequences of *Caenorhabditis elegans* core histone genes. Genes for different histone classes share common flanking elements. J. Mol. Biol. **206**:567–577.

ROGER, A. J., O. SANDBLOM, W. F. DOOLITTLE, and H. PHILIPPE. 1999. An evaluation of elongation factor 1α as a

phylogenetic marker for eukaryotes. Mol. Biol. Evol. **16**: 218–233.

SADLER, L. A., and C. F. BRUNK. 1992. Phylogenetic relationships and unusual diversity in histone H4 proteins within the *Tetrahymena pyriformis* complex. Mol. Biol. Evol. **9**: 70–84.

ŠALI, A., and T. L. BLUNDELL. 1993. Comparative protein modeling by satisfaction of spatial restraints. J. Mol. Biol. **234**: 779–815.

SÁNCHEZ, L. B. 1998. Aldehyde dehydrogenase(CoA-acetylating) and the mechanism of ethanol formation in the amitochondriate protist, *Giardia lamblia.* Arch. Biochem. Biophys. **354**:57–64.

SÁNCHEZ, L. B., V. ENEA, and D. EICHINGER. 1994. Increased levels of polyadenylated histone H2B mRNA accumulate during *Entamoeba invadens* cyst formation. Mol. Biochem. Parasitol. **67**:137–146.

SÁNCHEZ, L. B., H. G. MORRISON, M. L. SOGIN, and M. MÜLLER. 1999. Cloning and sequencing of an acetyl CoA-synthase (ADP forming) gene from the amitochondriate protist, *Giardia lamblia.* Gene **233**:225–231.

SÁNCHEZ, R., and A. ŠALI. 1997. Evaluation of comparative protein structure modeling by Modeller-3. Proteins Suppl. 1:50–58.

SANDMAN, K., S. L. PEREIRA, and J. N. REEVE. 1998. Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome. Cell. Mol. Life Sci. **54**:1350–1364.

SIPPL, M. J. 1993. Recognition of errors in three-dimensional structures of proteins. Proteins **17**:355–362.

SLESAREV, A. I., G. I. BELOVA, S. A. KOZYAVKIN, and J. A. LAKE. 1998. Evidence for an early prokaryotic origin of histones H2A and H4 prior to the emergence of eukaryotes. Nucleic Acids Res. **2**:427–430.

SOGIN, M. L., J. H. GUNDERSON, H. J. ELWOOD, R. A. ALONSO, and D. A. PEATTIE. 1989. Phylogenetic meaning of the kingdom concept: an unusual ribosomal RNA from *Giardia lamblia.* Science **243**:75–77.

STARICH, M. R., K. SANDMAN, J. N. REEVE, and M. F. SUMMERS. 1996. NMR structure of HMfB from the hyperthermophile, *Methanothermus fervidus,* confirms that this archaeal protein is a histone. J. Mol. Biol. **255**:187–203.

STILLER, J. W., E. C. S. DUFFIELD, and B. D. HALL. 1998. Amitochondriate amoebae and the evolution of DNA-dependent RNA polymerase II. Proc. Natl. Acad. Sci. USA **95**:11769–11774.

STILLER, J .W., and B. D. HALL. 1997. The origin of red algae: implications for plastid evolution. Proc. Natl. Acad. Sci. USA **94**:4520–4525.

THATCHER, T. H., and M. A. GOROVSKY. 1994. Phylogenetic analysis of the core histones H2A, H2B, H3, and H4. Nucleic Acids Res. **22**:174–179.

UPCROFT, J. A., and P. UPCROFT. 1998. My favorite cell: *Giardia.* BioEssays **20**:256–263.

VERNET, G., M. SALA-ROVIRA, M. MAEDER, F. JAQUES, and M. HERZOG. 1990. Basic nuclear proteins of the histone-less eukaryote *Crypthecodinium cohnii* (Pyrrophyta): two-dimensional electrophoresis and DNA-binding properties. Biochim. Biophys. Acta **1048**:281–289.

WALKER, D. R., and E. V. KOONIN. 1997. SEALS: a system for easy analysis of lots of sequences. Intell. Syst. Mol. Biol. **5**:333–339.

WELLS, D. J. 1986. Compilation analysis of histones and histone genes. Nucleic Acids Res. **14**(Suppl.):119–149.

WU, G., J. LI, and S. LU. 1996. Preliminary study on the histones of *Giardia lamblia.* Zool. Res. (Kunming) **17**:301–305 [in Chinese with English abstract].

YEE, J., and P. P. DENNIS. 1994. The NADP-dependent glutamate dehydrogenase of *Giardia lamblia*: a study of function, gene structure and expression. Syst. Appl. Microbiol. **16**:759–767.

GEOFFREY MCFADDEN, reviewing editor

Accepted April 11, 2000