Cloning and analysis of the macronuclear gene for histone H1 from *Euplotes eurystomus*

Loren J.Hauser*, Mary L.Treat+ and Donald E.Olins
University of Tennessee — Oak Ridge Graduate School of Biomedical Sciences, PO Box 2009, Biology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-8077, USA

Received May 1993; Accepted June 4, 1993 GenBank accession no. L15293

Euplotes is a hypotrichous ciliate that contains both a macronucleus (MAC) and a micronucleus in the same cell. The DNA in the MAC is composed of small, linear highly amplified fragments, each of which contains one gene, two telomeres and all the necessary regulatory elements. Although there is a large amount of condensed chromatin in the MAC, it appears not to be organized in 30 nm fibers as seen in higher eukaryotes (9). Since histone H1 is responsible for the formation of the 30 nm fiber, analysis of this protein in *Euplotes* may provide some insight into the unique chromatin structure in the MAC.

Histone H1 from the *Euplotes* MAC has been previously identified, purified and partially characterized (7). The N-terminus was sequenced and was used to construct an oligonucleotide probe. This probe identified a single gene on Southern blots of MAC DNA and was successfully used to screen a MAC DNA library (5). The double-stranded portion of the gene is 1310 bp and contains a 405 bp open reading frame, coding for a 134 aa protein (the N-terminal Met is removed in the mature protein). There are two putative TATA boxes within 55 bp upstream of the coding region, but no CCAAT boxes, which are seen upstream in other ciliate histone genes (1, 6). Quantitative Southern blots estimate 9×10^4 copies of this gene per MAC (data not shown).

A search of the SwissPro database (using the GCG programs, (4)) reveals that 20 of the top 25 matches were to histone H1 proteins. Despite an identity of about 30%, little or none of the homology was to the 80 aa globular domain, the most conserved domain among H1. However, the search matched *Euplotes* H1 with *Tetrahymena* H1, which also does not contain a globular domain. The *Euplotes* H1 does not contain the SPKK (2) or KTPKKAKKP (3) repeat motifs seen in some higher eukaryotic H1, that are believed to form beta turns and interact with the minor groove. On the other hand, it does contain a number of repeats of a 9−11 aa motif (see Figure 1) some of which could form beta turns. The beta turn would center on the sequence KKA(G/A)ARK. A computer search (4) revealed that all of the repeats contain at least one phosphorylation motif for either cAMP dependent kinase [(R,K)(R,K)X(S,T)] or protein kinase C [(S,T)X(R,K)]. In addition, many contain sites that could be phosphorylated by other kinases (8). These would provide an appropriate opportunity for a phosphorylation event to modify the binding of the H1 to DNA. Seven of the repeats are on the N-terminal portion of the protein, separated from the last two repeats by a 19 aa segment rich on A and P (the segment which has the greatest potential for forming α-helices and is double underlined in Figure 1).

Although the evolutionary relationship of *Euplotes* H1 to other H1 is in question, this protein displays many of the properties and performs many of the functions associated with H1 (7).

ACKNOWLEDGEMENTS

We thank C.Smith (University of Tennessee Health Science Center) for performing the protein sequencing; A.Herrmann for technical assistance; and A.Olins and R.Mural for critical reading of the manuscript. This research was sponsored by the U.S. Department of Energy, under contract DE-AC05-84OR21400 with Martin Marietta Energy Systems, Inc.; NSF grant MCB9117123 to D.E.O. The GenBank accession number for the MAC histone H1 DNA sequence is L15293.

REFERENCES

1. Brunk,C.F. and Sadler, L.A. (1990) *Nucleic Acids Res.* **18**, 323−329.
2. Churchill,M.E.A. and Suzuki,M. (1989) *EMBO J.* **8**, 4189−4195.
3. Erard,M., Lakhdar-Ghazal,F. and Amalric,F. (1990) *Eur. J. Biochem.* **191**, 19−26.
4. Genetics Computer Group (1991), Program Manual for the GCG Package, Version 7, April, 1991, 575 Science Drive, Madison, WI 53711, USA
5. Hauser,L.J. and Olins,D.E. (1993) *Gene*, submitted.
6. Hauser,L.J., Ready,T.R. and Olins,D.E. Manuscript in preparation.
7. Herrmann,A.L., Cadilla,C.L., Cachiero,L.H., Carne,A.F., and Olins,D.E. (1987) *Eur.J. Cell Biol.* **43**, 155−162.
8. Kemp,B.E. andPearson,R.B. (1990) *Trends Biochem. Sci.* **15**, 342−346.
9. Olins,A.L., Olins,D.E., Derenzini,M., Hernandez-Verdun,D., Grounon,P., Robert-Nicoud,M. and Jovin,T.M. (1988) *Biol. Cell* **62**, 83−93.

```
                CONSENSUS:  KKAGA--RKXS
                                  A   T
                     1   PAKTATA-VKRTT      4/8
                    13   TTKKSAA-KRKTS      7/8
                    25   KAVKKAG--KRTQS     6/8
                37  KAKGAQKVKKAAT-RRTPS     6/8
                    55     KSAGA--RKAT      7/8
                    64     KKAGA--RKAS      8/8
                    73     TKRSAT--KKTT     4/8
83  AAPAAAAAPATDAPAAAATPSKATGSAKKASA--RKSS  7/8
                   119   AKKPAKGGKKKSAKKN   5/8
```

Figure 1. Peptide sequence of the *Euplotes* histone H1 with the 9−11 aa repeat motifs aligned. The consensus of the repeat motif is aligned above the sequence. The fraction to the right of each repeat designates the number of matches (underlined) with the 8 aa that make up the consensus. The number to the left of each repeat is the aa number at the beginning of the that line.

* To whom correspondence should be addressed
+Present address: BioWhittaker, Inc., 8830 Biggs Ford Road, Walkersville, MD 21793-0127, USA