Nucleotide sequence of the gene encoding the largest subunit of the DNA-dependent RNA polymerase III of *Giardia lamblia*

Martin Lanzendörfer, Peter Palm*, Bernd Grampp, Debra A.Peattie[1] and Wolfram Zillig
Max-Planck-Institut für Biochemie, Martinsried, Am Klopferspitz 18, FRG and [1]Department of Tropical Public Health, 665 Huntington Avenue, Boston, MA 02115, USA

Submitted November 28, 1991 EMBL accession no. X60325

On the basis of 16S rRNA analysis (1) it was concluded that the intestinal parasite *Giardia lamblia*, which does neither contain mitochondria nor a Golgi apparatus, represents the lowest known lineage in the eucaryal domain. In order to gain a better insight into the phylogenetic relationship between the three evolutionary domains of life, we included this primitive eukaryote in the comparison of sequences of the largest (A) subunits of eucaryal RNA polymerases and corresponding components (β',A and A'plus A" respectively) from *Bacteria* and *Archaea* (2).

The amino acid sequence **NADFDGD(E/Q)M(N/A)** is conserved in all β',A (respectively A') subunits of bacterial, eucaryal and archaeal RNA polymerases known so far (3, 4, 5). Frequently, an oligonucleotide primer derived from this sequence, specifically hybridized to three *G.lamblia* chromosomal DNA fragments, whether digested with SacI, AvaI, BamHI, XbaI, HindIII, SalI, PstI or PvuII (Figure 1). Thus it is probable that *G.lamblia* like other eukaryotes contains three different nuclear RNA polymerases.

A 7.6 kbp PstI-fragment identified in this way was cloned into pBluescript™ II SK+ and sequenced in both directions. It was found to contain a complete open reading frame (ORF) encoding a putative protein of 529 amino acids (from 1439 bp to 3022 bp) and the larger part of a long ORF encoding a protein sequence showing homology to the A components of RNA polymerases. The highest similarity was obtained when the putative protein sequence was compared to the A subunits of RNA polymerases III of *Trypanosoma brucei* (6) and *Saccharomyces cerevisiae* (4).

In order to clone the missing part of the gene, encoding the C terminus of the component, chromosomal DNA was digested with BclI, circularized by ligation and subjected to an inverse polymerase chain reaction (7) using two oligonucleotides complementary to the 3'-end of the 7.6 kbp PstI fragment. The product, 800 bp in length, contained the sequence encoding the missing C-terminus and the stop codon.

The pol III A gene is 5205 bp long. Compared to other known pol III A genes it contains a 450 bp insert (from about position 1540 to position 1990). Nuclease S1 mapping of enriched mRNA showed that this region was fully protected, indicating that the insert is transcribed and that it is not an intron.

From an alignment with the sequences of genes encoding the β' and A components of RNA polymerases of organisms from all domains of life, phylogenetic trees were calculated using various methods (8). Parsimony and distance matrix analysis indicate that the *G.lamblia* pol III A gene sequence clusters with other eucaryal RNA polymerases III and that it is in close phylogenetic neighbourhood to that of *T.brucei* (9). This places *Giardia* higher than previously proposed (1), but additional pol III A genes from other protists would help confirm this finding. The genes encoding the largest subunits of RNA polymerases I and II of *G.lamblia* are now under investigation and may provide further insight into the evolution of *Eukarya*.

REFERENCES

1. Sogin,M.L., Gunderson,J.H., Elwood,H.J., Alonso,R.A. and Peattie,D.A. (1988) *Science* **243**, 75–77.
2. Zillig,W., Palm,P., Klenk,H.P., Pühler,G., Gropp,F. and Schleper,C., (1991) in General and Applied Aspects of Halophilic Microorganisms edited By Francisco Rodriguez-Valera, Plenum Press, New York and London, 321–332.
3. Cornelissen,A.W.C.A., Evers,R. and Köck,J., (1988). *Oxford Surveys Euk. Gen.* **5**, 91–131.
4. Memet,S., Gouy,M., Marck,C., Sentenac,A. and Buhler,J.-M. (1988) *J. Biol. Chem.* **263**, 2830–2839.
5. Pühler,G., Leffers,H., Gropp,F., Palm,P., Klenk,H.-P., Lottspeich,F., Garrett,R.A. and Zillig,W. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4569–4573.
6. Köck,J., Evers,R. and Cornelissen,A.W.C.A. (1988) *Nucl. Acids Res.* **16**, 8753–8772.
7. Erlich,H.A. (Editor), PCR Technology, Stockton Press, 1989
8. Felsenstein,J. (1981) *J. Mol. Evol.* **17**, 368–376.
9. Zillig,W., Palm,P., Langer,D., Klenk,H.P., Lanzendörfer,M., Hüdepohl,U. and Hain,J. (1991) Biochemical Society Transactions (in press).

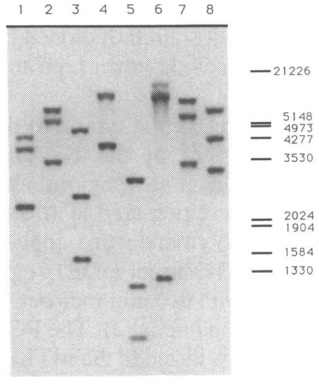

Figure 1. Southern blot of chromosomal DNA digested with different enzymes. The hybridisation with the oligonucleotide 5'-CATTTCGTCKCCGTCGAAG-TCKGCGTT-3' was carried out under stringent conditions (60°C, 6×SSC) Lanes: 1, SacI; 2, AvaI; 3, BamHI; Xbal; 5, HindIII; 6, SalI; 7, PstI; 8, PvuII.

* To whom correspondence should be addressed