Complete sequence of the extrachromosomal rDNA molecule from the ciliate *Tetrahymena thermophila* strain B1868VII

Jan Engberg and Henrik Nielsen
Department of Biochemistry B, The Panum Institute, University of Copenhagen, Blegdamsvej 3, DK-2200 Copenhagen N, Denmark

Received September 4, 1990; Revised and Accepted October 29, 1990 EMBL accession no. X54512

ABSTRACT

The recent development of rDNA vectors for transformation of *Tetrahymena* combined with improved microinjection technology should lead to a renewed interest in this organism. In particular, the rDNA itself constitutes an attractive system for biochemical studies. The rDNA is amplified to a level of 2% of the total DNA and exists as extrachromosomal molecules. Furthermore, the rDNA is homogeneous in sequence because it is derived from a single gene during sexual reorganization. In order to facilitate studies of this molecule, we report here a compilation of previously published sequence information together with new sequence data that completes the entire sequence of the 21 kb rDNA molecule.

INTRODUCTION

Tetrahymena thermophila is a dinucleate ciliate. The genetic nucleus, the micronucleus, is diploid, and the somatic nucleus, the macronucleus, is polyploid (n= ca. 45). *Tetrahymena* can be grown on a complex or chemically defined medium and is easily subjected to labelling conditions of cellular components and to subcellular fractionation procedures. Cell cultures can be synchronized through the vegetative cell cycle and through sexual reorganization (conjugation). Many aspects of cellular metabolism, cell physiology and genetics of *Tetrahymena* have been studied thoroughly as presented in ref.1. Most topics concerning the molecular biology of *Tetrahymena* and other ciliates are described in a more recent book (2).

The ribosomal RNA genes of *Tetrahymena* are among the best characterized eukaryotic genes. During conjugation, the single locus micronuclear gene is excised from its chromosomal origin and rearranged into extrachromosomal, palindromic, 21 kb molecules each containing two divergently arranged transcription units. These events have been diagrammed in Fig. 1. As a result of the amplification and subsequent polyploidization, each macronucleus contains approximately 20,000 rRNA genes which are essentially identical due to their derivation from the single micronuclear copy (the B strain is inbred and thus homozygous for most loci) (cf. reviews, 3 and 4). This situation is in contrast to the chromosomal organization of tandem arrays of rRNA genes found in most other eukaryotes and makes *Tetrahymena* very well suited for genetic analysis of the rRNA gene.

The only apparent transcript from the extrachromosomal rDNA molecules is a 35S rRNA precursor containing the 17S, 5.8S and 26S rRNAs. Transcription and processing of the 35S pre-rRNA has been studied in detail (cf.5) including the selfsplicing reaction performed by the intron of the 26S rRNA (6). A simplified account of these processes has been diagrammed in Fig.1. The origin of replication of the rDNA molecule has been mapped to a position close to the center of the molecule (7) and the replication of the telomers has been studied extensively (8,19). Also known is the nucleosome pattern (cf. review, 9) and the specific cleavage sites for topoisomerases (10). Specific proteins

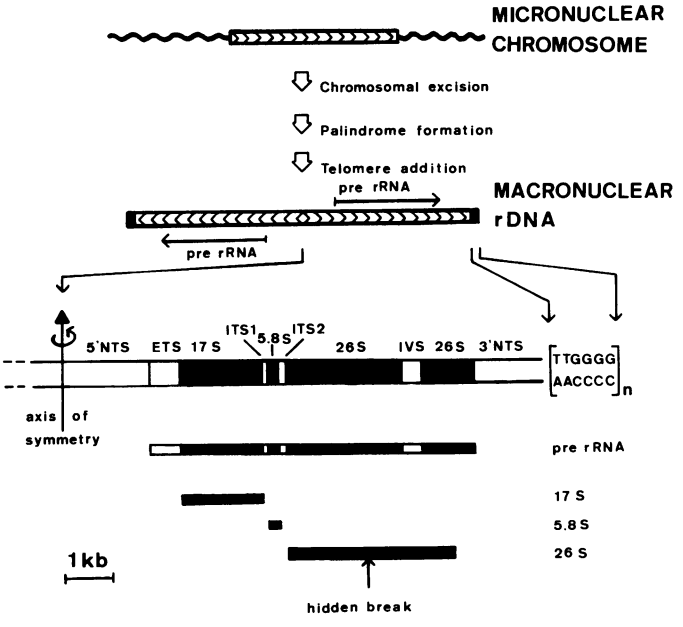

Fig. 1. Diagram of the rDNA in the micronucleus and the macronucleus of *Tetrahymena* combined with a simplified transcription map of the macronuclear rDNA (cf. text). NTS = non transcribed spacer; ETS = externally transcribed spacer; ITS 1 and 2 = internally transcribed spacer regions 1 and 2; IVS = intervening sequence. The number (n) of tandem telomeric repeats is about 50 (34). The 'hidden break' is introduced into the mature 26 S rRNA (35) and has been mapped at the nucleotide level (25).

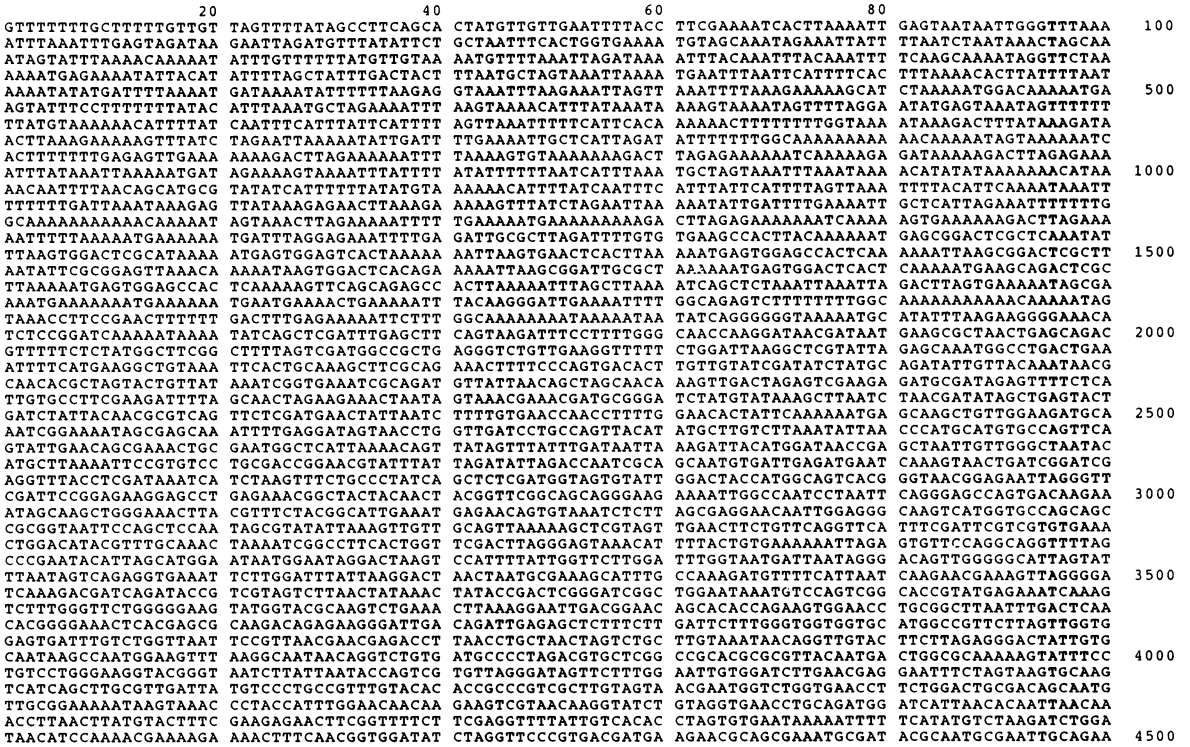

Fig. 2. Display of the sequence of the right hand half of the extrachromosomal, palindromic rDNA molecule from *T.thermophila* (cf. text). The sequence has been compiled from the following sources: pos. 1–61 (ref. 21); pos. 62–1907 (ref. 31); pos. 1908–2464 (ref. 26); pos. 2465–2534 (ref. 27); pos. 2535–4287 (ref. 28); pos. 4288–6957 (this work); pos. 6958–7474 (ref. 32); pos. 7475–8364 (this work); pos. 8365–8781 (ref. 30); pos. 8782–9958 (ref. 24); pos. 9959–10264 (R. Pearlman; pers. comm.); pos. 10265–10315 (ref. 33).

binding to the telomeric (11) and promotor regions (12) have been described.

Much of the original work on the ribosomal RNA genes in *Tetrahymena* relied on the ability to isolate the native gene in a pure form. A protocol was devised based on a modified Hirt extraction followed by banding of the rDNA according to buoyant density in a CsCl-gradient (13). This isolation procedure gave sufficient material for a variety of biochemical studies to be performed, including sequencing of native rDNA (14). Protocols for isolation of native r-chromatin (15) and for preparation of extracts for *in vitro* transcription and processing (16) of rRNA have also been developed.

The lack of an efficient system for genetic transformation of *Tetrahymena* was an obstacle for the molecular biologist in the field for many years. Then transformation by microinjection (17) and electroporation (18) using native rDNA was described. This was followed by the development of an rDNA based vector system, where a circular plasmid containing a single rRNA gene is propagated in *E.coli* and manipulated as a standard vector. Upon introduction in *Tetrahymena*, this plasmid is converted into a linear, palindromic molecule by recombination. Drug resistant markers in the introduced gene allows for selection of transformants, and the rearranged vector completely replaces the endogenous rDNA after a number of generations because an allele is used which contains an origin of replication with a replication advantage over that of the host rDNA. This transformation system has been used successfully in several studies, *e.g.* in characterization of the template function of telomerase RNA (19). A series of very promising rDNA genetics experiments in which conjugating cells were transformed with altered rDNA molecules should also be mentioned (20).

In view of the above mentioned advances, we would like to suggest that *Tetrahymena* be used as an experimental organism in studies of various aspects of DNA metabolism, in particular when these studies require microgram amounts of native material of a specific gene (the rDNA). *Tetrahymena* is also a good experimental organism for studies involving DNA damage because the high ploidity level of its genome results in high viability of the cells thus permitting repair of the damage to be investigated.

RESULTS

Despite the fact that the extrachromosomal rDNA molecule is usually described as a gigant palindromic sequence, it contains a small region (28 bp) at the very centre of the molecule which does not belong to the overall symmetry (21). This feature has been noted in several *Tetrahymena* species (14) and has important implications for the molecular processes leading to the formation of the palindromic rDNA molecule from the chromosomally integrated micronuclear copy (cf. ref. 4).

The compiled rDNA sequence is listed in Fig.2. The sequence displays the RNA-like strand of the right hand half of the palindromic molecule reading in the 5′ to 3′ direction beginning at the non-palindromic center (pos 1−28) of the molecule. Only one of the approximately 45 telomeric TTTTGG-repeats has been included in the Fig. (pos. 10310−10315). A variety of sequencing protocols, including chemical and enzymatic sequencing, were used by the different contributing laboratories. In order to complete the entire rDNA molecule, we have sequenced about 3.5 kilobases of DNA from different regions of the molecule. Furthermore, the previously published sequences of the ends of restriction fragments were reinvestigated by sequencing overlapping fragments.

Table I: Map co-ordinates of the transcribed regions of the rDNA molecule of *T.thermophila*.

Sequence position	Feature
1887-8522	pre-rRNA
2535-4287	17S rRNA
4418-4571	5.8S rRNA
4749-8508	26S rRNA
6342-6345	hidden break deletion
7010-7422	group I intron

In no cases did we detect discrepancies in the sequence obtained from individual clones derived from different rounds of cloning experiments. The displayed sequence is 10315 bp in length with an overall A+T composition of 64%. The A+T composition of the transcribed region is 57% which is lower that that of the overall rDNA and also of the bulk DNA which is 75% (22). An analysis of the occurrence of repeats in the rDNA sequence has been published (23,24). Different features concerning transcription of the rDNA are given in table I. The listed map coordinates were determined experimentally using Sl nuclease protection, primer extension and reverse sequencing procedures. The 5′ end of pre-rRNA and the entire 5.8S rRNA were determined by RNA sequencing (26,29).

Table II shows the sequence differences observed between strain B and strain C3 rDNA. These data have been kindly provided by the laboratories of Ed Orias and Meng-Chao Yao. Strain C3 is a strain with a dominant replication origin and is often used in transformation experiments. The complete sequence of the rDNA from a different species of *Tetrahymena*, *T.pyriformis*, has been determined recently and will be published elsewhere (T.Higashinakagawa, pers.communication). A comparison of the 26S rRNA sequences of *T.thermophila* and *T.pyriformis* has been published recently (25).

Plasmid clones covering the entire rDNA molecule from *T.thermophila* is available from the authors. Strains of *Tetrahymena* are available from the American Type collection (*T.thermophila* B1868VII has ATCC no. 30377) and from some of the authors in the cited references.

ACKNOWLEDGEMENTS

We would like to thank all the colleagues in the *Tetrahymena* field who contributed to the compilation, especially Cliff Brunk, Ron Pearlman, Ed Orias, Meng-Chao Yao, Drena Larson and Peter Yeager who also provided unpublished data and critically proofread the compiled sequence. This work was supported by the Danish Natural Science and Medical Science Research Councils and by a Nato Research Council travelling grant (to JE) no. 86/0650.

REFERENCES

1. Biology of Tetrahymena (1973) Elliott, A.M. (ed). Dowden Hutchinson and Ross, Stroudsbury, Pennsylvania.
2. The Molecular Biology of Ciliated Protozoa (1986) Gall, J.G. (ed). Academic Press, Inc.
3. Engberg,J. (1985) Eur. J. Cell Biol. **36**, 133−51.
4. Yao,M.-C. (1986) In Gall,J.G. (ed) The Molecular Biology of Ciliated Protozoa. Academic Press, Inc. pp. 179−201.

Table II: Sequence differences between the rDNA molecules of *T.thermophila* strains B and C3 (Data kindly provided by E. Orias and M.-C. Yao).

Nucleotide position	B rDNA	C3 rDNA	Comments	Reference
70	C	A	possible, unconfirmed	a
584	T	A	possible, unconfirmed	a
1014	G	A	B-spec. SphI site	b
1226	A	42bp (*)	Confers repl. advantage	c
1277	AAAAA-	AAAAAA	Confirmed	c
1308	A	T	Confirmed	c
1366	C	T	Confirmed	d
1417	AAAAA	AAAA-	possible, unconfirmed	a
1442	A	G	possible, unconfirmed	a
1507	C-	CA	possible, unconfirmed	a
8050	G	-		e
8091	C	T		e
8560	T--	TTT	possible, unconfirmed	f
8567	T	-	possible, unconfirmed	f
8573	G	A	possible, unconfirmed	f
8623	C	CC	possible, unconfirmed	f
8629	G	GG	possible, unconfirmed	f
8653	AC	TT	possible, unconfirmed	f
8665	A	-	possible, unconfirmed	f
8707	AAAA	----	confirmed	f
8766	G	A	possible, unconfirmed	f
8963	A	T	possible, unconfirmed	f
9120	T	G	possible, unconfirmed	f
9148	C	A	C3 spec. BamH1 site	g
9155	G	C	C3 spec. TaqI site	b
9158	A	C	possible, unconfirmed	f
9272	C	T	possible, unconfirmed	f
9462	GT	TG	possible, unconfirmed	f
9628	A	-	possible, unconfirmed	f
9680	C	T	C3 spec. NcoI site	h
9685	TG	AA	possible, unconfirmed	f
9756	G	A	possible, unconfirmed	f
9882	G	T	B spec. Sau3A site	f
9938	T	-	possible, unconfirmed	f
10072	T	G	possible, unconfirmed	f
10238	G	C	possible, unconfirmed	f
10248	C	T	possible, unconfirmed	f
10266	A	T	possible, unconfirmed	f
10302	A	ATTAAA	possible, unconfirmed	f

(*) 42 bp insertion following B rDNA position A1226:
CCTTCCGAACTTTTGCAACTTTTGAGACTTCGTGAAAAAAGA

References: (a) Yu, G.-L. and E. H. Blackburn, pers. comm. to E. Orias. (b) Yu, G.-L. et al. Proc. Natl. Acad. Sci. U.S.A. **85,** 5151 5155 (1988). (c) Larson, D. D. et al. Cell **47,** 229-240 (1986). (d) Luehrsen, K. R. et al. Gene **55,** 169-178 (1987). (e) M.-C. Yac pers. comm. (f) E. Orias, pers. comm. (g) Pan, W.-C. and E. H. Blackburn Cell **23,** 459-466 (1981). (h) P. Yaeger, unpublished.

5. Kister,K.-P., Kaffenberger,W. and Eckert,W.A. (1988) Eur. J. Cell Biol. **46**, 233–243.
6. Cech,T.R. and Bass,B.L. (1986) Ann. Rev. Biochem. **55**, 599–629.
7. Cech,T.R. and Brehm,S.L. (1981) Nucl. Acids Res. **9**, 3531–3543.
8. Greider,C.W. and Blackburn,E.H. (1987) Cell **51**,887–898.
9. Gorovsky,M.A. (1986) In Gall,J.G. (ed) The Molecular Biology of Ciliated Protozoa. Academic Press, Inc. pp. 227–261.
10. Bonvén,B.J., Gocke,E. and Westergaard,O. (1985) Cell **41**, 541–551.
11. Henderson,E. (pers. communication).
12. Niles,E.G. (1985) J. Biol. Chem. **260**, 672–678.
13. Din,N. and Engberg,J. (1979) J. Mol. Biol. **134**, 555–574.
14. Engberg,J. (1983) Nucl. Acids Res.**11**, 4939–4945.
15. Colavito-Shepanski,M. and Gorovsky,M.A. (1983) J. Biol. Chem. **258**, 5944–5954.
16. Sutiphong,J., Matzura,C. and Niles,E.G. (1984) Biochemistry **23**, 6319–6323.
17. Tondravi,M.M. and Yao,M.-C. (1986) Proc. Natl. Acad. Sci. USA.**83**, 4369–4373.
18. Brunk,C.F. and Navas,P. (1988) Exp. Cell Res. **174**, 525–532.
19. Greider,C.W. and Blackburn,E.H. (1989) Nature **337**, 331–337.
20. Sweeney,R. and Yao,M.-C. (1989) EMBO J. 8,**3**, 933–938.
21. Kiss,G.B. and Pearlman,R.E. (1981) Gene **13**, 281–287.
22. Karrer,K.M. (1986) In Gall,J.G. (ed) The Molecular Biology of Ciliated Protozoa. Academic Press, Inc. pp. 85–110.
23. Niles,E.G., Sutiphong,J. and Haque,S. (1981) J. Biol. Chem. **256**, 12849–12856.
24. Challoner,P.B., Amin,A.A., Pearlman,R.E. and Blackburn, E.H. (1985) Nucl. Acids Res. **13**, 2661–2680.
25. Engberg,J., Nielsen,H., Lenaers,G., Murajama,O., Fujitani,H. and Higashinakagawa,T. (1990) J. Mol. Evol. **30**, 514–521.
26. Engberg,J., Din,N., Saiga,H. and Higashinakagawa, T. (1984) Nucl. Acids Res. **12**, 959–972.
27. Spangler,E.A. and Blackburn,E.H. (1985) J. Biol. Chem. **260**, 6334–6340.
28. Sogin,M., Ingold,A. Karlok,M., Nielsen,H. and Engberg,J. (1986) EMBO J. **5**, 3625–3630.
29. van Bell, C.T. (1985) J. Protozool. **32**, 640–644.
30. Din,N., Engberg,J. and Gall,J.G. (1982) Nucl. Acids Res. **10**, 1505–1513.
31. Luehrsen,K.R., Baum,M.P. and Orias,E. (1987) Gene **55**, 169–178.
32. Kan,N.C. and Gall,J.G. (1982) Nucl. Acids Res. **10**, 2809–2822.
33. Kiss,G.B.; Amin,A.A. and Pearlman,R.E. (1981) Mol. Cell. Biol. **1**, 535–543.
34. Blackburn,E. (1986) In Gall,J.G. (ed.) The Molecular Biology of Ciliated Protozoa. Academic Press, Inc. pp. 155–178.
35. Eckert,W.A., Kaffenberger,W., Krohne,G. and Franke,W.W. (1978) Eur. J. Biochem **87**, 607–616.