Formerly
PROTISTOLOGICA

European Journal of PROTISTOLOGY

Volume 25 · 1989/90

Gustav Fischer Verlag · Stuttgart · New York

25 / 1

Contents

25 / 2

ARTICLES

25 / 3

ARTICLES

25 / 4

Europ.J.Protistol. 25, 158–167 (1989)
October 27, 1989

European Journal of
PROTISTOLOGY

Regulatory Structures of Gene Expression, DNA-Replication and DNA-Rearrangement in Macronuclear Genes of *Stylonychia lemnae*, a Hypotrichous Ciliate

Elke Helftenbein, Karl Klaus Conzelmann, Karl Friedrich Becker and Hariolf Fritzenschaf
Institut für Biologie III, Abt. Zellbiologie, Universität Tübingen, Tübingen, FRG

SUMMARY

Seven transcriptionally active macronuclear DNA molecules of *Stylonychia lemnae* have been analyzed for potential regulatory sequences of gene expression, DNA replication and DNA rearrangement. Transcription initiation is mediated by a canonical TATA-box or by TATA-box-like sequences, and transcription efficiency may be regulated by gene-specific downstream promoter elements (DPE). Putative signals for RNA 3′-formation are represented by nonconserved palindromes located upstream of the mature RNA 3′-termini. Sequences indicating polyadenylation of the mRNA molecules are not detectable. A single palindrome of A- and T-residues is present within the 80 5′-terminal bases of each macronuclear DNA molecule and very likely functions as the replication origin. The 7 DNA molecules consist of several nonconserved inverted and direct repeats (IR, DR) suggested to play a role in DNA rearrangement during macronuclear development.

Introduction

Recently several hundred eukaryotic genes transcribed by Pol II have been cloned, sequenced and analyzed for consensus sequences involved in regulation of gene expression [4, 5, 7, 29, 39]. Two types of nucleotide sequences involved in regulation of transcription initiation have been identified: promoters and enhancers [31]. Promoters are necessary for accurate and efficient initiation of transcription and are located within the 100 bp immediately upstream of the transcription start sites. A typical promoter consists of several sequence elements. One of these represents the well-conserved TATA-box flanked by GC-stretches that ensures the accurate initiation. It is centred about 30 bases upstream of the transcription start site. The elements located further upstream, designated upstream promoter elements (UPEs), increase the rate of initiation [33]. UPEs differ gene- and tissue-specifically and act regardless of their orientation with respect to the TATA-box. Their regulating function is mediated by binding of nuclear proteins, referred to as transcription factors. Enhancers, the second type of sequence elements regulating gene expression, increase the rate of transcription of promoters and act in an orientation independent manner at great distances upstream and downstream of the transcription unit [46]. They also contain discrete DNA sequence elements specifically interacting with proteins. Transcription is typically initiated between 20 and 80 nucleotides upstream of the coding region preferentially at the motif CA (cap-site) followed by pyrimidines [7].

Formation of precise RNA 3′-termini generally occurs by two different mechanisms signalled by distinct DNA sequences. Accurate termination of transcription is one of these processes and seems to be signalled by T-stretches or AT-rich stem loops located downstream of the transcription termination site [14]. The second mechanism, referred to as RNA 3′-processing, includes 3′-cleavage of pre-mRNA molecules mainly followed by polyadenylation of the resulting mRNA molecules. These processes are preferentially signalled by a cluster of sequences beginning

0932-4739/89/0025-0158$3.50/0

with the highly conserved polyadenylation signal AATAAA, located 10 to 20 bases upstream of the cleavage site CA (functioning as the polyadenylation site) and ending with G/T-clusters within 30 nucleotides downstream of the cleavage site [5, 6, 14, 27, 43]. In contrast to this set of sequences a different, well conserved sequence combination of a palindrome located immediately upstream of the cleavage site and of a downstream positioned nonanucleotide signals RNA 3′-cleavage without polyadenylation in the histone genes of higher eukaryotes [45]. The same mechanism is also found in plastids of higher plants and is regulated by nonconserved inverted sequences located immediately upstream of the mature RNA 3′-termini [49].

The nucleotides preceding the coding regions [CCACCATG] seem to be involved in the initiation of translation in vertebrates [28, 29]. The A-residue at pos. −3 is strongly conserved in all organisms including plants and lower eukaryotes, whereas the Cs may be replaced by As, as it is found in *Drosophila* and yeast [9, 17].

An interesting question is, whether these general eukaryotic signals of gene expression are also existent and functioning in ciliates, a group of lower unicellular eukaryotes, showing very special features of genome organisation. Ciliates contain two morphologically and functionally different nuclei: the generative micronuclei playing a major role in sexual reproduction (conjugation) and showing no or little transcriptional activity in the vegetative cell cycle, and the somatic macronuclei providing all RNA for vegetative cell growth [2, 3]. Macronuclei develop from micronuclei by complex species specific processes, referred to as macronuclear development. In the hypotrichous ciliate *Stylonychia lemnae* this macronuclear development includes chromosome polytenization, degradation of the giant chromosomes, selective DNA elimination, DNA rearrangement and DNA amplification [41]. The resulting transcriptionally active macronuclear genome is organized in about 15 000 different DNA molecules of 0.4 to 20 kbp in length, of which each is present in a distinct copy number of a few to more than 100 000 [3, 10, 18, 19, 30, 48]. Each DNA molecule is thought to represent a transcription and replication unit and consists of a single intronless coding region flanked by short (74 bp to 1000 bp) AT-rich (75%–80%) noncoding sequences terminating in telomeric inverted repeats of C_4A_4. The coding regions show, like those of some other ciliates, an altered genetic code using the "universal" stop codons TAA and TAG to specify glutamine [8, 10, 18, 19, 22, 25, 36, 40].

We examined the noncoding sequences of 7 expressed macronuclear genes from *Stylonychia lemnae* for potential regulatory elements of gene expression, DNA replication and DNA rearrangement. Four of the investigated *Stylonychia* genes represent the two α- and the two β-tubulin genes, each of them present in a high (α_1- and β_1 150 000) and in a low copy number (α_2- and β_2 30 000) [10, 18, 19]. With respect to the biological function of the α- and β-tubulins the coordinated expression of their genes is expected.

The other 3 cloned macronuclear DNA molecules represent transcriptionally active genes in all tested *Stylonychia* strains, but the functions of their encoded proteins are unknown. The genes p4A3 and pob were chosen with respect to their special features in single *Stylonychia* strains. Parts of the coding region of p4A3 are repeated in several differently sized macronuclear DNA molecules of strain DO (Fritzenschaf and Helftenbein, in prep.) and the DNA molecule pob becomes overamplified during vegetative growth in strain SP (Wegner et al., in prep.). The gene pma1 was investigated as a representative of an extremely short macronuclear DNA molecule (unpublished). A comparison of the detected presumptive regulatory sequences in these 7 macronuclear genes of *Stylonychia* and in the previously reported genes from other ciliates is discussed.

Material and Methods

Cell cultivation and isolation of macronuclear DNA and RNA
Cells of *S. lemnae* strain DO (collected in North Germany; origin of the cloned α_2-, β_1-, β_2-tubulin genes), strain SP (South Germany; origin of the α_1-tubulin gene, pob and p4A3) and 6-Schi (South Germany; origin of pma1) were cultivated in Pringsheim solution as described by Ammermann [1] and DNA was purified from isolated macronuclei [2] as previously described [48]. Total RNA was isolated by the procedure of Glisin [12] and contaminating DNA was eliminated by DNAse I digestion with 3000 units/ml at 37 °C for 40 min.

Cloning and screening of macronuclear DNA molecules

Macronuclear DNA was size-fractionated by preparative agarose gel electrophoresis and cloned by G-C-tailing (α_1-tubulin gene) [18] or after Bal 31 digestion of the protruding 3′-ends as previously described [10]. Screening of recombinant plasmids was done by colony hybridization [16] with radioactive labelled cDNA probes of the α- and β-tubulin genes [10, 18, 19] and of enriched radioactive labelled macronuclear DNA molecules containing the pob and p4A3 genes.

DNA sequencing and S_1-mapping

DNA sequencing was performed by the chemical method of Maxam and Gilbert [32] or by the enzymatic procedure of Sanger et al. [44]. S_1-mapping experiments were carried out as described by Helftenbein [19].

Denaturing agarose gel electrophoresis and hybridization reactions

For denaturing gels RNA and DNA was glyoxylated [34] and gel electrophoresis and hybridization was done as previously described [10].

Results

Transcription initiation sites of the α- and β-tubulin genes

Identification of the transcription start sites located about 40 bases upstream of the coding regions of the α_2-, β_1- and β_2-tubulin genes (Fig. 1) have recently been published [10, 19]. The transcription initiation site of the

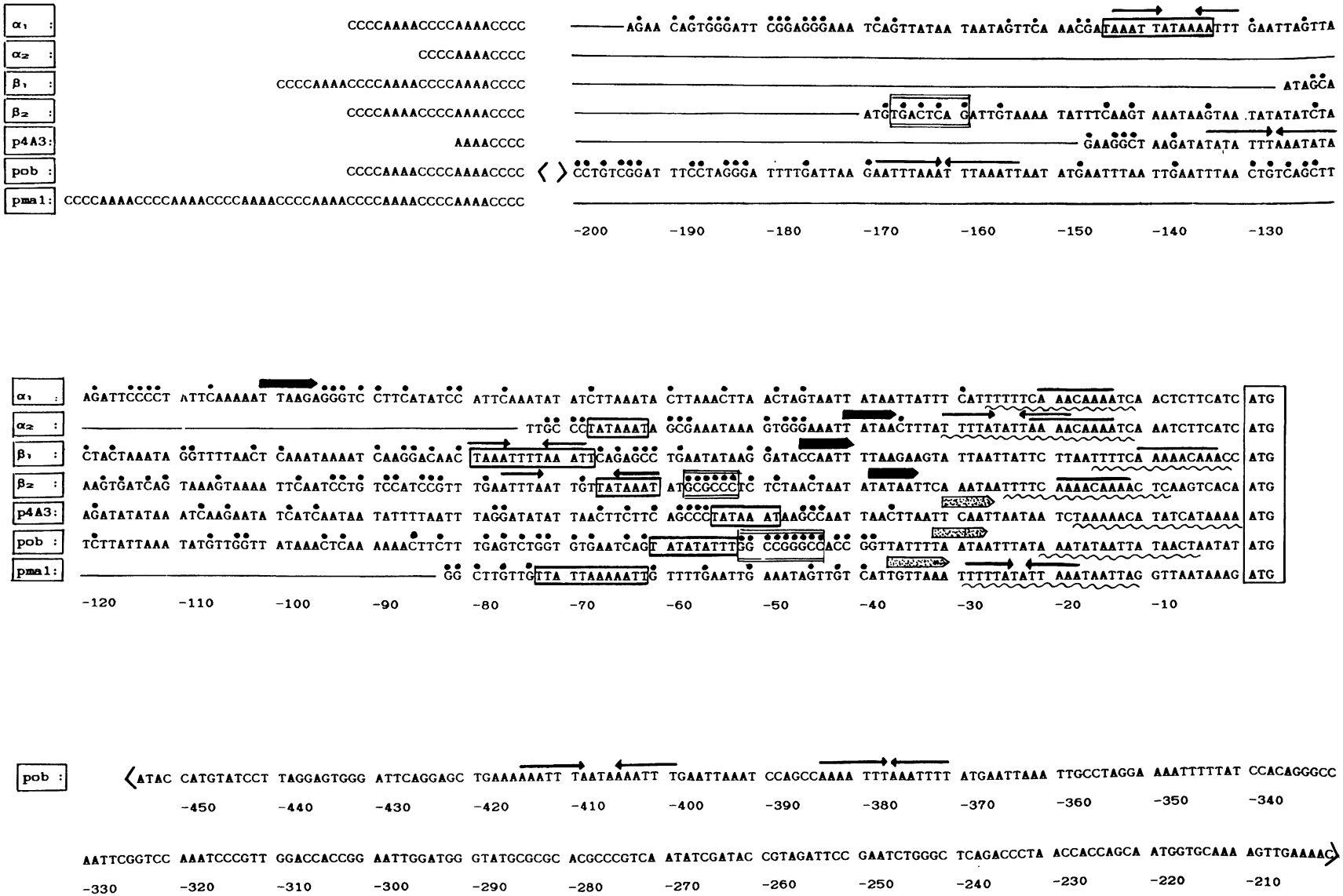

Fig. 1. Nucleotide sequence of the 5′-noncoding regions. All sequences are aligned to their translation start sites (ATG). Gs and Cs are marked by dots, TATA-boxes [31] and TATA-box-like sequences are boxed. Boxes with double lines show eukaryotic recognition sites for transcription factors [33]. Identical sequences are overlined and homologous sequences, representing presumptive gene-specific DPEs, are underlined by ∿. ▬: transcription start sites; ▭: presumptive start sites of transcription; →←: palindromes representing the replication origins; <>: sequence shown in the lower part of the figure.

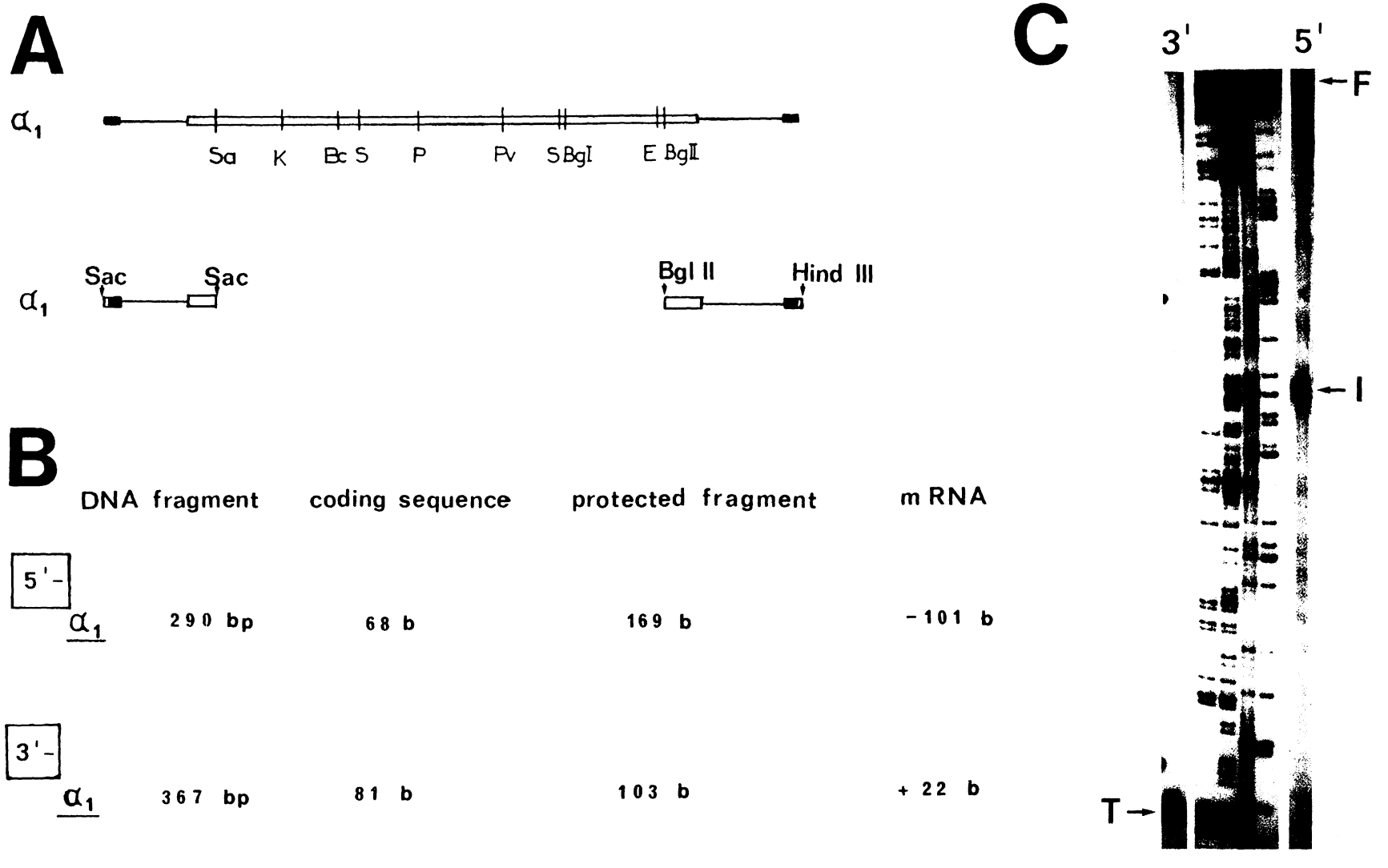

Fig. 2. S_1-mapping of the α_1-tubulin gene. The reactions were carried out as described in Materials and Methods. (A) Restriction map of the α_1-tubulin gene and DNA fragments of the cloned molecule used in the S_1-protection experiments. Open bars represent the coding region; thin lines indicate noncoding sequences and filled boxes represent the terminal repeats. Sa: SacI; K: Kpn; Bc: BclI; S: SalI; P: PstI; Pv: Pvu; BgI: BglI; E: Eco RI: Bg: BglII. (B) summarized data of the S_1-protection experiments. b: bases; bp: base pairs. (C) Autoradiographs of the 5′ and 3′ protected DNA fragments separated in denaturing polyacrylamid gels, which are calibrated by using radioactive products of a sequencing reaction as molecular size markers [19].

α_1-tubulin gene was determined by S_1-mapping experiments as shown in Fig. 2. The single protected DNA fragment of 169 bases demonstrates that transcription is precisely initiated, like in the other tubulin genes, but in a larger distance from the coding region at pos. – 101 (Fig. 1). Thus the length of the four mRNA leaders are in the range of those known from other eukaryotes [7]. However, the first two transcribed nucleotides (cap site) of all tubulin genes are not conserved (α_1: TT; α_2: TA; β_1: CC; β_2: AT) and neither of them represents the eukaryotic cap site motif CA [7].

Common potential promoters of the 7 macronuclear genes

In the α_2- and β_2-tubulin genes and in the gene p4A3 a canonical TATA-box flanked by short GC-rich stretches is present 60 to 70 bases upstream of the coding regions (Fig. 1). The α_1- and β_1-tubulin genes and the genes pob and pma1 coding for still unidentified proteins contain in corresponding distances short AT-stretches similar to a TATA-box (TATA-box-like sequences), which are also flanked by GC-rich regions (Fig. 1). These two types of TATA-boxes are located in a classical distance of 25 to 30 bases upstream of the transcription start sites of the 4 tubulin genes. The GC-rich flanks represent regions with the highest GC-contents of the entire 5′-noncoding regions from all 7 genes, as shown for the tubulin genes in Fig. 3.

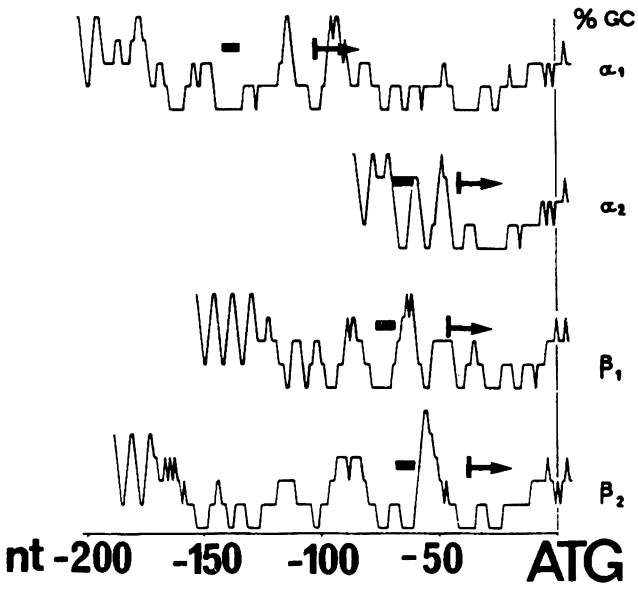

Fig. 3. Base composition of the 5′-noncoding regions in the tubulin genes. The relative GC-contents are plotted at each base position and the base compositions were determined every base in a 5 bp range, using Pustell Sequence Analysis Programs, International Biotechnologies, Inc. Bars represent TATA-boxes in the α_2- and β_2-tubulin genes and TATA-box like sequences in the α_1- and β_1-tubulin genes; arrows indicate the transcription start sites; nt: nucleotides.

Upstream of the two types of TATA-boxes neither the sequences of the UPEs found in higher eukaryotes nor conserved sequences functioning as potential ciliate specific UPEs could be detected. One exception is given by the β_2-tubulin gene, which shows the recognition site of the Sp1 transcription factor in an unusually large distance from the TATA-box at pos. – 169 (Fig. 1).

Almost identical T- and A-stretches interrupted by single C-residues are present in the transcribed regions of the tubulin genes, located very close to the translational start sites (12, 12, 2, 11 bp upstream of ATG in α_1, α_2, β_1, β_2, respectively) (Fig. 1). In corresponding regions of the other 3 macronuclear genes similar but not identical sequences of more alternating As and Ts without single Cs are present. Neither these sequences nor those of the tubulin genes do resemble known eukaryotic signals involved in transcription or translation processes.

Putative regulatory sequences for initiation of translation

The pentanucleotide upstream of the ATG in the α-tubulin genes resembles the CCACC (ATG) of vertebrates [28, 29], whereas a transition to the pentanucleotides of Drosophila and yeast, prefering As [9, 17], is visible within the β-tubulin genes. Pentanucleotides consisting almost completely of As are found in the other 3 macronuclear genes. The assumed indispensable A-residue in pos. – 3 is present in all genes except for the gene pob, where it is replaced by a T-residue.

Potential signals for the formation of mRNA 3′-termini

The recently reported transcribed parts of the 3′-noncoding regions from the α_2-, β_1- and β_2-tubulin genes [10, 19] are shown in Fig. 4. The identification procedure of the α_1-tubulin mRNA trailer by S_1-mapping experiments is explained in Fig. 2, showing a single precisely defined 3′-terminus located 25 bases downstream of the coding region (pos. TGA + 25 in Fig. 4). Such single well defined RNA 3′-termini were also found in the β-tubulin genes (β_1: TGA + 73; β_2: TGA + 57). However, two RNA 3′-termini located at slightly separated positions TGA + 48 and TGA + 44 were found in the α_2-tubulin gene. Thus the tubulin gene trailers are unusually short in comparison to those of other eukaryotes, which show lengths of several hundred to more than 1000 nucleotides [5].

The last two transcribed bases of all the tubulin genes are given by the dinucleotides TA or CA representing the eukaryotic polyadenylation site. However, the polyadenylation signal AATAAA generally centred 15 bases upstream of the polyadenylation site is not present in either of the transcribed regions of the tubulin genes. Surprisingly such a sequence was detected in the α_2-tubulin gene in the nontranscribed region starting at pos. TGA + 68 and in that of the β_1-tubulin gene at pos. TGA + 175, +271 and +282. Shortened versions of the eukaryotic GT-clusters normally located 15 bases downstream of the RNA 3′ cleavage site are also present in the tubulin genes but in a

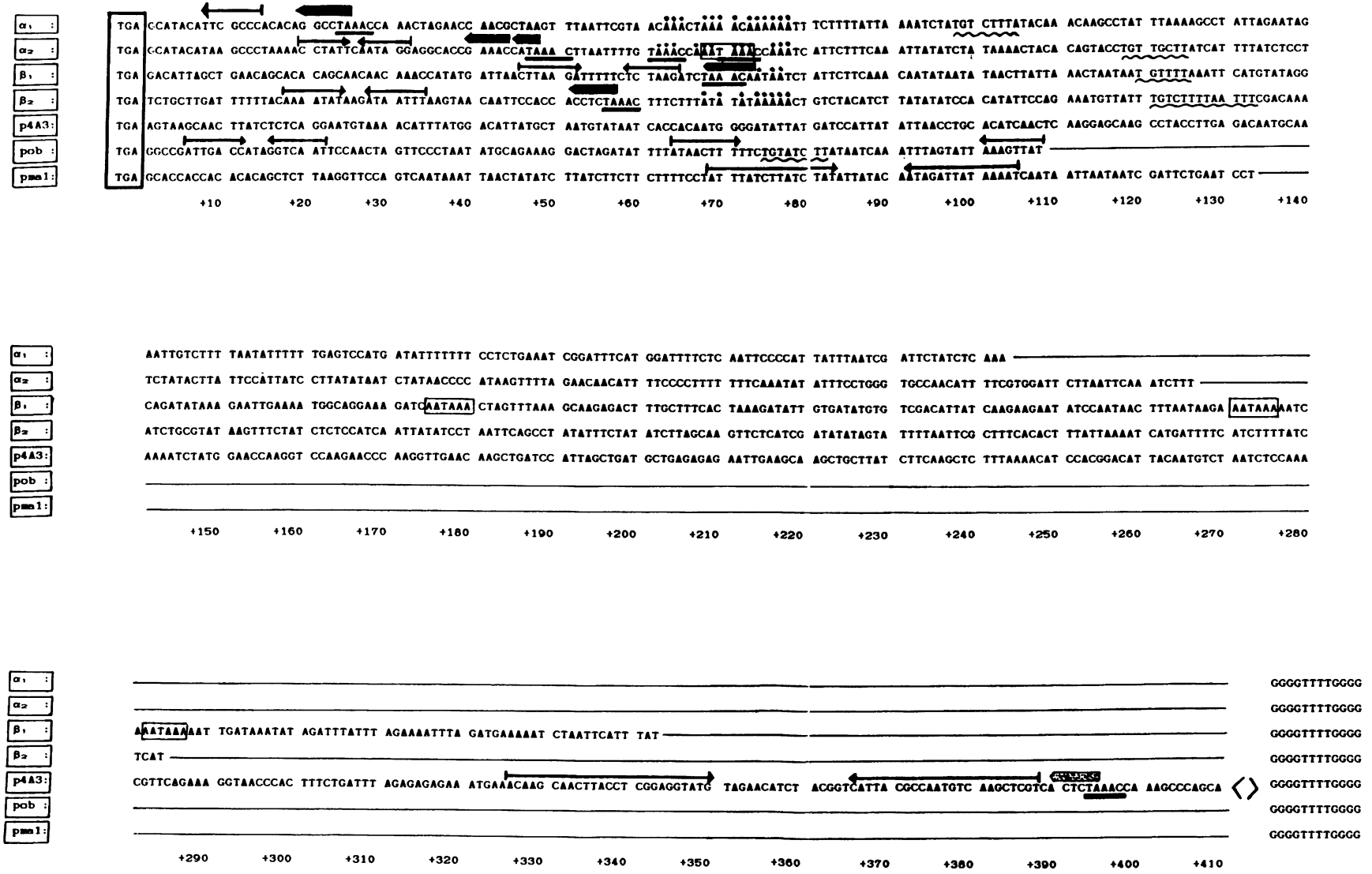

Fig. 4. Nucleotide sequences of the 3′-noncoding regions. All sequences are aligned to their coding regions (TGA); sequences identical to the eukaryotic polyadenylation signal [5] are boxed and sequences similar to the GT-clusters present in the sequence set mediating RNA 3′-processing in eukaryotes [5] are marked by ~~. ▬ : represent identified RNA 3′-termini and ⟨⋯⟩ shows putative RNA 3′-termini; the suggested gene-specific signal TAAAC is underlined. The RNA 3′-terminal inverted repeats are indicated by arrows. <>: additional nucleotide sequence, which is not shown and which does not contain potential regulatory signals.

larger distance of about 60 bases around pos. TGA + 120 (Fig. 4).

The mRNA 3′-termini of the 4 tubulin genes are located within an identical sequence TAAAC and each of the 4 mRNA trailers contains a palindrome of a nonconserved sequence very close to its 3′-end (Fig. 4). Such a combination of the sequence TAAAC and an immediately upstream located nonconserved palindrome is also present in the gene p4A3 between pos. TGA + 330 and TGA + 400 (Fig. 4), where the mRNA 3′ terminus has been localized by preliminary experiments (data not shown). The other two cloned genes, pob and pma 1, do not contain the motif TAAAC but also show nonconserved palindromic sequences in their 3′-noncoding regions. The positions of the palindromes (pob: TGA + 6 to TGA + 22; pma1: TGA + 69 to TGA + 115) (Fig. 4) are in good agreement with the mRNA lengths identified in Northern blots and preliminary S_1-mapping experiments (data not shown).

Potential hairpin structures or larger T-stretches downstream of the RNA 3′-end could not be detected in any of the tubulin genes.

Inverted and direct repeats within the 7 macronuclear genes

Each of the 7 cloned genes contains an inverted repeat consisting only of A- and T-residues within the first 80 nucleotides of their 5′-noncoding regions (Fig. 1). Such potential hairpin structures are unique in each DNA molecule, except for the gene pob, which contains two additional inverted repeats of A- and T-residues in the 5′-noncoding region, beginning at pos. – 383 and – 169 (Fig. 1). These AT-structures show high homology to autonomously replicating sequences identified in mouse rDNA (50).

Short inverted repeats (IR) of 5 to 12 bases are localized at both ends of each DNA molecule and with different frequency within each molecule (Fig. 5). The lengths of the nucleotide stretches inside each IR varies extremely from about 20 bases in the 3′-terminal IR of gene p4A3 to about 580 bases in the 5′-terminal IR of the same gene. Adjacent IRs are located side by side or separated by 1 to 3 bases, or by direct repeats (DR) (Fig. 5). These short IRs and DRs might play a role in DNA rearrangement during macronuclear development, since such short nonconserved sequences are found to surround the several short parts of the macronuclear DNA molecules within the highmolecular weight DNA of the micronucleus (Helftenbein and Richter, in prep).

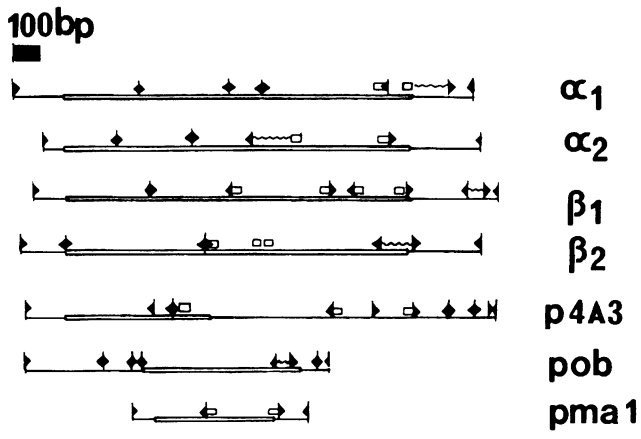

Fig. 5. Schematic representation of inverted (IR) and direct repeats (DR) within the 7 macronuclear DNA molecules. Open bars represent coding regions and thin lines show noncoding sequences. IRs are shown by: ▶◀ and DRs by: □. Sequences of IRs or DRs are not identical and each comprises 5 to 12 nucleotides. ∿: represents sequences, which are not flanked by IRs or DRs.

Discussion

The detailed sequence analyses of the noncoding regions of 7 macronuclear genes from *Stylonychia* identified common features, some of which with similarities to known eukaryotic regulatory signals for gene expression, DNA replication and DNA rearrangement. Out of the conserved sequence set of typical eukaryotic promoters consisting of UPEs and a TATA-box flanked by GC-rich stretches [31] only the TATA-box element is present in 3 of the 7 genes. The other 4 genes contain at corresponding positions TATA-box-like sequences with GC-flanks on either side. Thus these sequences may also signal, like TATA-boxes in higher eukaryotes, accurate initiation of transcription in *Stylonychia* as found in the α_1- and β_1-tubulin genes (Fig. 1). The lacking of eukaryotic UPEs as well as the missing of conserved *Stylonychia* specific sequences suggest that the efficiency of transcription initiation may be regulated in *Stylonychia* differently compared with other eukaryotes.

Three different mechanisms of gene specific regulation of the transcription initiation in *Stylonychia* are conceivable. The first one might be represented by the sequence of the TATA-box itself, leading to a high transcription initiation rate of genes containing a canonical TATA-box, whereas the genes with only a TATA-box-like sequence may be transcribed with a lower efficiency. Such a mechanism could be realized in the α- and β-tubulin genes, of which the highly amplified α_1- and β_1-tubulin genes (150 000 copies) contain only TATA-box-like sequences and the less abundant α_2- and β_2-tubulin genes show canonical TATA-box sequences. Quantitative measurements of the transcription products of the 4 tubulin genes support this hypothesis (unpublished). A second regulation mechanism may be enabled by the conserved A- and T-rich sequences found immediately upstream of the coding regions within the mRNA leaders (Fig. 1). These sequences, which contain identical nucleotide stretches in the tubulin genes and slightly different nucleotides in the other 3 genes, could function as downstream promoter elements (DPEs) comparable to the UPEs found in higher eukaryotes [31]. Thus the assumed regulation mechanism should be mediated by the binding of these sequences to specific transcription factors. The third regulation mechanism of the transcription initiation efficiency is represented by the specific amplification of the genes during the macronuclear development occuring in *Stylo-*

nychia [10, 18, 19, 36, 48] and presumably in all hypotrichous ciliates.

The general eukaryotic sequence sets regulating transcription termination or RNA 3′-processing followed by polyadenylation are almost completely absent in the macronuclear genes of *Stylonychia*. We therefore assume that the formation of the distinct single RNA 3′ termini of the tubulin genes is mediated by other motifs than the known ones. The detected 3′-terminal nonconserved palindromic sequences and the downstream adjacent conserved pentanucleotide TAAAC of the mRNA molecules of the tubulin and the p4A3 genes seem to represent such *Stylonychia* specific signals (Fig. 4). The existence of only palindromes upstream of the roughly determined RNA 3′-ends of the genes pob and pma1 led us to assume that in *Stylonychia* a potential hairpin structure of a nonconserved sequence in combination with the dinucleotides CA or TA located at the RNA 3′-termini is generally sufficient for RNA 3′-formation. Similar signals were found for RNA 3′-processing without polyadenylation in plastids of higher plants [49] and in histone genes of higher eukaryotes [45]. Thus the investigated RNA molecules of *Stylonychia* may also lack polyadenylated regions.

However, the motif TAAAC present only in some genes (tubulin and p4A3) could represent an additional sequence necessary for accurate RNA 3′-processing as found in the tubulin genes, or it might act as a signal for a *Stylonychia* specific RNA 3′ modification. Whether this assumed modification is generally different from the eukaryotic polyadenylation process remains to be investigated. Evidences for such a hypothesis are the poor enrichment of tubulin mRNA molecules by oligo dT-cellulose chromatography of total RNA (unpublished) and the identification of unusually short A stretches, which are only attached to about 0.2% of all RNA molecules (unpublished).

The existence of the species specific poorly conserved pentanucleotides preceding the coding regions [28, 29] in the 7 investigated genes, suggests that translation initiation is similarly regulated in *Stylonychia* and in higher eukaryotes.

The discussed results lead to the conclusion that the 7 investigated macronuclear genes might have contained earlier in ciliate evolution the general eukaryotic promoter sequences, of which they maintained only the TATA-box or a TATA-box-like sequence inevitable for accurate transcription initiation. However, all eukaryotic upstream promoter elements acting as modulators of the transcription initiation efficiency were lost. The efficiency might instead be regulated by ciliate specific mechanisms, namely by the specific amplification of genes during macronuclear development and/or by gene specific DPEs, which might be created as a result of the very short noncoding regions (e.g., α_2: 74 bp) (Fig. 1) of the macronuclear DNA molecules. From the sequence set acting in eukaryotic RNA 3′-formation only a few remnants could be detected in the 7 macronuclear genes. Such sequences are the polyadenylation site-motif CA and TA, conserved at all 7 mRNA 3′-termini, the polyadenylation signal AATAAA, present only in the α_2- and β_1-tubulin genes but at strange distances from the mRNA 3′-termini, and the shortened GT-clusters found only in some genes but also at unusual positions (Fig. 4). Whether CA/TA and AATAAA represent indeed significant signals within the very AT-rich (75%–80%) noncoding regions of the macronuclear DNA molecules is questionable. Thus it could not be decided, whether the ciliates possessed in the past the eukaryotic signals for RNA 3′-processing and polyadenylation, or whether the ancestors of ciliates had already evolved the assumed present mechanism of RNA 3′-formation without polyadenylation like that in the eukaryotic histone genes [45] or plastids of plants [49].

These general assumptions of gene expression regulation in *Stylonychia* are supported by sequence analysis of other recently investigated ciliate genes. Although transcription is initiated in the actin gene of the hypotrichous ciliate *Oxytricha nova* divergently from *Stylonychia* at 3 distinct sites, two of the transcribed RNA molecules show leaders comparable in length to those found in *Stylonychia* and the gene contains also a TATA-box-like sequence with GC-rich flanks at a corresponding distance from the coding region [15]. In all of the other so far reported genes of *Oxytricha* without identified transcription start sites, such a TATA-box-like sequence arrangement is present between 70 and 90 nucleotides upstream of the coding regions [20, 24, 26]. Even the published sequences for histone and actin genes of *Tetrahymena*, a holotrichous ciliate, do share these characteristics [11, 21, 23, 37], whereas in the surface antigen genes of *Paramecium*, another holotrichous ciliate, the accurate initiation of transcription at a single site seems to be signalled differently [13, 35]. In accordance with the 7 investigated genes of *Stylonychia* none of these other ciliate genes contains the eukaryotic sequence set involved in RNA 3′-formation followed by polyadenylation. Although the RNA trailers of the *Tetrahymena* histone genes are approximately 4 times longer than the trailers of genes from *Stylonychia* and *Paramecium*, in all of these genes only one single RNA 3′-terminus could be detected, preceded also by a short nonconserved IR, except for the histone H4I gene of *Tetrahymena* [23]. However, the motif TAAAC found at the RNA 3′-termini in some genes of *Stylonychia* is not present in either of the few other investigated ciliate mRNA molecules. Thus TAAAC seems to be a potential gene- or species-specific signal.

The single AT-rich IR detected within the last 80 5′-terminal bases of each macronuclear DNA molecule represents very likely the origin of DNA replication, which is expected from electron microscopical investigations close to one end of each DNA molecule. This assumption was confirmed by testing the 3 reiterated potential hairpin structures of the gene pob (Fig. 1) for autonomous replication in a heterologous mouse L-cell system (Helftenbein et al., submitted). These experiments demonstrate that each of the 3 palindromes initiates DNA replication and indicate that the exceptional reiteration of such structures in the gene pob mediates the observed overamplification of this macronuclear gene during vegetative cell-growth (Helftenbein et al., submitted). In all reported macronuclear DNA molecules of the closely related ciliates *Stylo-*

nychia pustulata [38] and *Oxytricha fallax* and *O. nova* [15, 20, 24, 26] such a unique AT-rich IR is detected at corresponding positions in the 5′- or 3′-noncoding regions, presumably also functioning as the replication origin.

The short nonconserved IR sequences localized at the 5′- and 3′-termini of each investigated macronuclear DNA molecule from *Stylonychia,* as well as the adjacent irregularly distributed ones (Fig. 5) may represent signals comparable to the IRs found at the termini of transposable elements in higher eukaryotes [47]. This assumption is strongly supported by the finding of such structures in micronuclear DNA, where they mark the macronuclear DNA sequences (Helftenbein and Richter, in prep.). This result and the identified short macronuclear stretches within the micronuclear DNA are in strong contrast to the results reported for *Oxytricha,* showing macronuclear sequences of larger extents in the micronuclear genome. Instead of terminal IRs of macronuclear DNA stretches, short DRs have been found at one end of the maintained macronuclear sequence and at one end of the eliminated internal sequence in *Oxytricha* [26, 42]. Thus in *Oxytricha* no signals of DNA rearrangement were detectable in macronuclear DNA molecules.

Acknowledgements

We would like to thank Prof. W. Seyffert for his generous help with computer analyses and Prof. D. Ammermann and Dr. Helmut Schmidt for their critical comments on the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft.

References

1 Ammermann D. (1965): Cytologische und genetische Untersuchungen an dem Ciliaten *Stylonychia mytilus* Ehrenberg. Arch. Protistenk., *108,* 109–152.
2 Ammermann D. (1971): Morphology and development of the ciliates *Stylonychia mytilus* and *Euplotes aediculatus.* Chromosoma, *33,* 209–238.
3 Ammermann D., Steinbrück G., Von Berger L. and Hennig W. (1974): The development of the macronucleus in the ciliated protozoan *Stylonychia mytilus.* Chromosoma, *45,* 401–429.
4 Birchmeier C., Schümperli D., Sconzo G. and Birnstiel M. L. (1984): 3'editing of mRNA's: sequence requirements and involvement of a 60 nucleotide RNA in maturation of histone mRNA precursors. Proc. Natl. Acad. Sci. USA, *81,* 1057–1061.
5 Birnstiel M. L., Busslinger M. and Strub K. (1985): Transcription termination and 3'processing: the end is in site! Cell, *41,* 349–359.
6 Boardman M., Basi G. S. and Storti R. V. (1985): Multiple polyadenylation sites in a *Drosophila* tropomyosin gene are used to generate functional mRNA's. Nucl. Acids Res., *13,* 1763–1776.
7 Bucher P. and Trifonov E. (1986): Compilation and analysis of eukaryotic Pol II promoter sequences. Nucl. Acids Res., *14,* 10009–10026.
8 Caron F. and Meyer E. (1985): Does *Paramecium primaurelia* use a different genetic code in its macronucleus? Nature, *314,* 185–188.
9 Cavener D. R. (1987): Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. Nucl. Acids Res., *15,* 1353–1361.
10 Conzelmann K. K. and Helftenbein E. (1987): Nucleotide sequence and expression of two β-tubulin genes in *Stylonychia lemnae.* J. Mol. Biol., *198,* 643–653.
11 Cupples C. G. and Pearlman R. E. (1986): Isolation and characterization of the actin gene from *Tetrahymena.* Proc. Natl. Acad. Sci. USA, *83,* 5160–5164.
12 Glisin V., Crkvanjakov R. and Byus C. (1974): Ribonucleic acid isolated by cesium chloride centrifugation. Biochemistry, *13,* 2633–2637.
13 Godiska R. (1987): Structure and sequence of the H surface protein gene of *Paramecium* and comparison with related genes. Mol. Gen. Genet., *208,* 529–536.
14 Grass D. S., Jove R. and Manley J. L. (1987): RNA polymerase II terminates transcription in vitro in the SV 40 origin region. Nucl. Acids Res., *15,* 4417–4436.
15 Greslin A. F., Loukin S. H., Oka Y. and Prescott D. M. (1988): An analysis of the macronuclear actin genes of *Oxytricha.* Genes and Develop. (in press).
16 Grunstein M. and Wallis J. (1979): Colony hybridization. Methods Enzymol., *68,* 379–389.
17 Hamilton R., Watanabe C. K. and de Boer H. A. (1987): Compilation and comparison of the sequence context around the AUG start-codons in *Saccharomyces cerevisae* mRNAs. Nucl. Acids Res., *15,* 3581–3593.
18 Helftenbein E. (1985): Nucleotide sequence of a macronuclear DNA molecule coding for α-tubulin from the ciliate *Stylonychia lemnae.* Special codon usage: TAA is not a translation termination codon. Nucl. Acids Res., *13,* 415–433.
19 Helftenbein E. and Müller E. (1988): Both α-tubulin genes are transcriptionally active in *Stylonychia lemnae.* Curr. Genetics, *13,* 425–432.
20 Herrick G., Hunter D., Williams K. and Kotter K. (1987): Alternative processing during the development of a macronuclear chromosome family in *Oxytricha fallax.* Genes and Develop., *1,* 1047–1058.
21 Hirono M., Endoh H., Okada N., Numata O. and Watanabe Y. (1987): *Tetrahymena* Actin. Cloning and sequencing of the *Tetrahymena* actin gene and identification of its gene product. J. Mol. Biol., *194,* 181–192.
22 Horowitz S. and Gorovsky M. A. (1985): An unusual genetic code in nuclear genes of *Tetrahymena.* Proc. Natl. Acad. Sci. USA, *82,* 2452–2455.
23 Horowitz S., Bowen J. K., Bannon G. A. and Gorovsky M. A. (1987): Unusual features of transcribed and translated regions of the histone H4 gene family of *Tetrahymena thermophila.* Nucl. Acids Res., *15,* 141–160.
24 Kaine B. P. and Spear B. B. (1982): Nucleotide sequence of a macronuclear gene for actin in *Oxytricha fallax.* Nature, *295,* 430–432.
25 Klobutcher L. A., Swanton M. T., Donini P. and Prescott D. M. (1981): All gene-sized DNA molecules in four species of hypotrichs have the same terminal sequence and an unusual 3'terminus. Proc. Natl. Acad. Sci. USA, *78,* 3015–3019.
26 Klobutcher L. A., Jahn C. L. and Prescott D. M. (1984): Internal sequences are eliminated from genes during macronuclear development in the ciliated protozoan *Oxytricha nova.* Cell, *36,* 1045–1055.

27 Kobrin B. J., Milcarek C. and Morrison S. L. (1986): Sequence near the 3'secretion specific polyadenylation site influence levels of secretion-specific and membrane-specific IgG 2b mRNA myeloma cells. Mol. Cell. Biol., *6*, 1687–1697.
28 Kozak M. (1984): Compilation and analysis of sequences up-stream from the translational start site in eukaryotic mRNAs. Nucl. Acids Res., *12*, 857–872.
29 Kozak M. (1986): Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell, *44*, 283–292.
30 Lipps H. J. and Steinbrück G. (1978): Free genes for rRNAs in the macronuclear genome of the ciliate *Stylonychia mytilus*. Chromosoma, *69*, 21–26.
31 Maniatis T., Goodbourn S. and Fischer J. A. (1987): Regulation of inducible and tissue-specific gene expression. Science, *236*, 1237–1245.
32 Maxam A. and Gilbert W. (1977): A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA, *74*, 560–564.
33 McKnight S. and Tjian R. (1986): Transcriptional selectivity of viral genes in mammalian cells. Cell, *46*, 795–805.
34 Mc Master G. K. and Carmichael G. G. (1977): Analysis of single and double-stranded nucleic acids on polyacrylamid and agarose gels by using glyoxal and acridine orange. Proc. Natl. Acad. Sci. USA, *74*, 4835–4838.
35 Meyer E., Caron F. and Baroin A. (1985): Macronuclear structure of the G surface antigen gene of *Paramecium primaurelia* and direct expression of its repeated epitopes in *Escherichia coli*. Mol. Cell. Biol., *5*, 2414–2422.
36 Meyers G. and Helftenbein E. (1988): Transfection of the hypotrichous ciliate *Stylonychia lemnae* with linear DNA vectors. Gene, *63*, 31–40.
37 Nomoto M., Imai N., Saiga H., Matsui T. and Mita T. (1987): Characterization of two types of histone H2B genes from macronuclei of *Tetrahymena thermophila*. Nucl. Acids Res., *15*, 5681–5697.
38 Oka Y. and Honjo T. (1983): Common terminal repeats of the macronuclear DNA are absent from the micronuclear DNA in the hypotrichous ciliate, *Stylonychia pustulata*. Nucl. Acids Res., *11*, 4325–4333.
39 Platt T. (1986): Transcription termination and the regulation of gene expression. Ann. Rev. Biochem., *55*, 339–372.
40 Preer J. R., Jr., Preer L. B., Rudman B. M. and Barnett A. J. (1985): A deviation from the universal code: the gene for surface protein 51A in *Paramecium*. Nature, *314*, 188–190.
41 Raikov J. B. (1982): The protozoan nucleus. In: Alfert M., Beermann W.,, Goldstein L., Porter K. R. and Sitte P. (eds.): Cell biology monographs, vol. 9, pp. 266–364. Springer Verlag, Wien–New York.
42 Ribas-Aparicio R. M., Sparkowsky J. J., Proulx A. E., Mitchell J. D. and Klobutcher L. A. (1987): Nucleic acid splicing events occur frequently during macronuclear development in the protozoan *Oxytricha nova* and involve the elimination of unique DNA. Genes and Develop., *1*, 323–336.
43 Ryner L. C. and Manley J. L. (1987): Requirements for accurate and efficient mRNA 3' end cleavage and polyadenylation of a Simian virus 40 early pre-RNA in vitro. Mol. Cell. Biol., *7*, 495–503.
44 Sanger F., Nicklen S. and Coulson A. R. (1977): DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA, *74*, 5463–5467.
45 Schaufele F., Gilmartin G. M., Bannwarth W. and Birnstiel M. L. (1986): Compensatory mutations suggest that base-pairing with a small nuclear RNA is required to form the 3' end of H3 messenger RNA. Nature, *323*, 777–781.
46 Serfling E., Jasin M. and Schaffner W. (1985): Enhancers and eukaryotic gene expression. Trends Genet., *1*, 224–230.
47 Shapiro J. A. (1983): Mobile genetic elements. Academic Press, New York.
48 Steinbrück G., Haas I., Hellmer K. H. and Ammermann D. (1981): Characterization of macronuclear DNA in five species of ciliates. Chromosoma, *83*, 199–208.
49 Stern D. B. and Gruissem W. (1987): Control of plastid gene expression: 3'inverted repeats act as mRNA processing and stabilizing elements, but do not terminate transcription. Cell, *51*, 1145–1157.
50 Wegner M., Klavinius A., Schwender S., Müller F., Zastrow G., Luksza H., Hoppe J., Koehler U. and Grummt F. (1989): Cis-acting sequences from mouse rDNA enable plasmid amplification and persistence in mouse cells: implication of a HMG I-like protein in their function (in press).

Key words: Hypotrichous ciliates – Macronuclear genes – Gene expression – DNA-replication – DNA-rearrangement

Elke Helftenbein, Institut für medizinische Virologie und Infektionsepidemiologie, Hölderlinplatz 10, D-7000 Stuttgart 1, FRG