Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 3252–3256, May 1989
Genetics

Differential use of termination codons in ciliated protozoa

(actin/tubulin/*Euplote crassus*/genetic code)

DAVID S. HARPER AND CAROLYN L. JAHN

Laboratory for Cell, Molecular, and Developmental Biology, Department of Biological Sciences, University of Illinois at Chicago, P.O. Box 4348, Chicago, IL 60680

Communicated by David M. Prescott, January 12, 1989 (received for review September 29, 1988)

ABSTRACT **Sequence analysis of genes in four species of ciliated protozoa and analysis of tRNAs in *Tetrahymena* has demonstrated that TAG and TAA encode glutamine or glutamic acid in these organisms and TGA is the only stop codon. Thus, it has generally been assumed that all ciliates use a nonuniversal genetic code in which TGA acts as the sole termination codon. We have sequenced the linear DNA molecules that carry an actin gene and a β-tubulin gene from the ciliate *Euplotes crassus*. These genes are shown to use TAA as a termination codon based on homology to known actin and β-tubulin gene sequences. In addition, we have sequenced a portion of the 3′ terminus of the *E. crassus* H4 histone gene and show that it also uses TAA as a termination codon. These data indicate that the timing of genetic code changes in the ciliates must be reconsidered.**

Ciliated protozoa are characterized by the possession of two types of nuclei: a germ-line micronucleus, which is transcriptionally inactive, and a vegetative macronucleus, which is responsible for most, if not all, transcription in the cell. Hypotrichous ciliates such as *Euplotes, Stylonychia*, and *Oxytricha* undergo a macronuclear developmental process that involves the degradation of >90% of all micronuclear DNA sequences and amplification of the remaining macronuclear-destined sequences (for review see ref. 1). These macronuclear molecules range in size from 400 base pairs (bp) to 20 kilobase pairs (kb) and are present in thousands of copies. Each molecule is believed to encode a single gene product and to contain all the information necessary for its replication and transcription. Holotrichous ciliates, such as *Tetrahymena*, undergo a similar, but less drastic, process of macronuclear development, in which 10–20% of all micronuclear sequences are degraded (2), and the resulting macronuclear molecules average 600 kb in size (3). Amplification of macronuclear sequences also occurs in *Tetrahymena*, resulting in ≈45 copies of each macronuclear chromosome (3).

All ciliate macronuclear genes sequenced to date use TGA as a stop codon (4). A number of these genes appear to use TAA and TAG as glutamine or glutamic acid codons at internal positions in the gene. A comparison of the derived amino acid sequence of *Stylonychia lemnae* α-tubulin with other known α-tubulins showed that a single TAA codon in the *Stylonychia* gene corresponded to glutamine (5). In addition, actin and two different H3 histone genes in *Tetrahymena thermophila* appear to use TAA as a glutamine codon (6–8), and two surface antigen genes in *Paramecium* contain both TAA and TAG codons that specify either glutamine or glutamic acid (9, 10). Both a TAA and a TAG codon are internal to an open reading frame that shares significant homology with ADP/ATP carrier proteins from other organisms and is found on an alternatively processed family of macronuclear chromosomes in *Oxytricha fallax* (11). Recent experiments have shown that *T. thermophila* possesses tRNAs with anticodons specific for TAA and TAG codons that are aminoacylated with glutamine (12). Based on this information it has been assumed that TAA and TAG are not stop codons in any of the ciliates.

In this paper, we describe the results of DNA sequence analysis of macronuclear actin and β-tubulin genes from the hypotrich *Euplotes crassus*.* Examination of the derived amino acid sequence indicates that this ciliate uses TAA as a termination codon. In addition, we report that the *E. crassus* actin is among the most atypical actins sequenced to date.

MATERIALS AND METHODS

Macronuclear Gene Isolation and Characterization. Methods for library construction, screening, restriction analysis of clones, subsequent subcloning, and Southern and Northern (RNA) blot hybridization analyses have all been described (13).

DNA Sequence Analysis. Restriction fragments from λ phage clones were subcloned into either M13mp18 and -mp19 (14) phage or into the Bluescribe plasmid vector (Vector Cloning Systems). Both single-stranded (15) and double-stranded (16) sequencing were done using the Sequenase modified T7 DNA polymerase (17).

RESULTS AND DISCUSSION

Three Macronuclear Genes Terminate in TAA. Because most genes in hypotrichous ciliated protozoa exist on individual linear DNA molecules with very short regions of noncoding DNA, localization of a coding region for a known gene is easily accomplished. The actin and β-tubulin genes in *E. crassus* are both examples of linear molecules bearing very short (<100 bp) stretches of 5′ and 3′ noncoding DNA. Initial sequence analysis of these genes indicated that they terminated with TAA codons. To demonstrate that the *E. crassus* code uses termination codons differently from other ciliates, we have sequenced clones of the entire linear molecules for these two genes.

As described previously (13), we used a *T. thermophila* actin gene (6) and a *Chlamydomonas reinhardtii* β-tubulin gene (18), respectively, to probe λ gt10 libraries of *E. crassus* macronuclear DNA. Analysis of the resulting clones indicated that we had isolated a single version of each gene, although Southern blot hybridization analysis showed that three nonallelic actins and two β-tubulins presumed to be allelic existed in the macronucleus (13). These genes were restriction-enzyme digested, and the resulting fragments were subcloned for sequence analysis. Restriction maps and sequencing strategies are shown in Fig. 1. Figs. 2 and 3 present the data obtained from DNA sequence analysis and the deduced amino acid sequence. We have sequenced the entire linear molecule containing the actin and β-tubulin

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: ORF, open reading frame.

*The sequences reported in this paper are being deposited in the EMBL/GenBank data base (accession nos. J04533 and J04534).

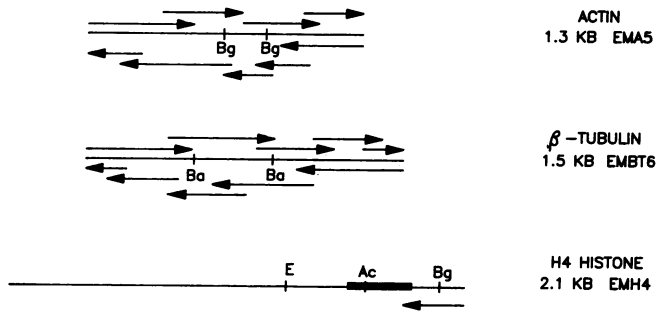

FIG. 1. Partial restriction maps and sequencing strategy for each of the three *E. crassus* macronuclear molecules. Each line with accompanying arrowhead indicates DNA sequence data obtained from either a subclone (if line initiates at a restriction site) or from a synthetic oligonucleotide primer (if line does not initiate at a restriction site). Ac, *Acc* I; Ba, *Bam*HI; Bg, *Bgl* II; and E, *Eco*RI.

genes, and in each case the open reading frame (ORF) identified (Figs. 2 and 3) is the only possible source of an actin or β-tubulin protein. No other ORF of >200 bp exists. Fig. 4 compares the deduced amino acid sequences of the 3′ ends of the actin and β-tubulin genes with the corresponding sequences from *Oxytricha nova* and *C. reinhardtii*, respectively. The reading frame of the actin gene aligns exactly with the *O. nova* actin gene and terminates in TAA. The reading frame of the *E. crassus* β-tubulin gene extends three codons past the *C. reinhardtii* β-tubulin gene ORF in their best alignment and terminates in TAA. This increase in the protein size does not seem particularly unusual. Although β-tubulin genes are highly conserved, divergence at their 3′ ends is not uncommon and could play a part in the detailed organization of microtubules (20). We have looked for other possible termination sites in both genes. Only 53 bp exist 3′ to the actin ORF, with no TGA codons present in frame. Two TAG codons are present in frame in the 30 bp immediately 3′ to the putative termination codon. Thus the only alternative termination sites for actin would be at TAG codons. Only 78 bp are present 3′ to the β-tubulin ORF, with a TGA present in frame that would extend the predicted protein sequence by 23 amino acids if it were used as a termination codon. Two in-frame TAA codons are present in the 20 bp immediately 3′ to the putative termination codon for the β-tubulin gene, with no TAG codons. It seems certain that TGA is not used to terminate the actin gene. Thus, for TGA to terminate the tubulin gene, TAA codons would have to function both as amino acid and termination codons.

We also sequenced the 3′ terminal 300 bp of the *E. crassus* H4 histone macronuclear linear DNA molecule (Fig. 4) and found that 240 bp internal to the telomere is a 60-bp region with identity of 83% on the DNA level and 95% on the amino acid level with the 3′ end of the coding region from the *O. nova* H4 histone gene (13). As can be seen, this gene uses a TAA termination codon. With continuation of the reading frame beyond this TAA, no possible TGA codons are found, and seven TAA and one TAG codons occur in frame. Thus, TGA cannot be the termination codon for this gene.

These represent unusual ciliate macronuclear genes that do not terminate with TGA and that appear to utilize TAA as a termination codon. We have previously shown that both the actin and β-tubulin clones detect transcripts when hybridized to Northern (RNA) blots of *E. crassus* RNA (13). Although three nonallelic actin genes are present in the *E. crassus* macronucleus, they can be distinguished by variation in hybridization stringency. The stringencies used in Northern blot experiments would differentiate among the different members of the actin multigene family; therefore, the transcript must be from this gene. Because the two β-tubulins that we identified cannot be distinguished by differential hybridization, we cannot conclusively state that the gene we sequenced is transcribed. Nevertheless, the fact that we found three different genes terminating in TAA is a strong argument that TAA is a stop codon in *E. crassus*.

```
CCCCAAAACC CCAAAACCCC AAAACCCCTA ATAAAGAAAA TAAAATTGCC CATTCTAATA
TATTTTATTA TATCTATCAA AA  82

                                                    Met Ser Glu Glu      4
                                                    ATG AGC GAA GAA     94

Glu Asn Ala Lys Glu Ala Ile Val Val Asp Asn Gly Ser Gly Val Val      20
GAA AAT GCT AAA GAG GCC ATC GTA GTT GAT AAT GGT TCA GGT GTT GTC     142

Lys Ala Gly Phe Ala Gly Glu Asn Gln Pro Cys Ser Val Phe Pro Ser      36
AAA GCA GGA TTT GCA GGA GAA AAC CAA CCA TGC TCA GTA TTC CCA TCA     190

Val Val Ala Lys Pro Lys Thr Lys Gln Val Ile Val Gly Gly Ala Gly      52
GTT GTT GCT AAG CCA AAG ACT AAG CAA GTT ATT GTC GGT GGA GCT GGA     238

Asn Lys Asp Cys Phe Val Gly Asp Glu Arg Gln Gln Lys Arg Gly Val      68
AAC AAA GAT TGC TTC GTT GGA GAC GAA CGC CAA CAA AAG AGA GGT GTC     286

Cys Thr Leu Ser Tyr Pro Ile Lys Ser Gly Met Ile Lys Asp Trp Asp      84
TGT ACT CTG TCA TAT CCA ATC AAA AGT GGA ATG ATT AAG GAT TGG GAT     334

Gly Met Gln Lys Ile Trp Asp Tyr Thr Phe Tyr Asn Glu Leu Arg Ile     100
GGT ATG CAA AAG ATT TGG GAC TAT ACC TTC TAC AAT GAA CTT AGA ATT     382

Glu Thr Glu Asn His Pro Val Leu Leu Thr Glu Ala Pro Leu Asn Pro     116
GAG ACA GAA AAC CAT CCA GTT CTA TTG ACT GAG GCT CCA TTG AAT CCA     430

Lys Gln Asn Arg Glu Asn Met Cys Arg Ile Met Phe Glu Glu Tyr Asp     132
AAG CAA AAC AGA GAG AAT ATG TGT AGA ATT ATG TTT GAA GAA TAT GAT     478

Phe Pro Ser Met Tyr Ile Gln Ile Gln Ala Val Leu Ser Leu Tyr Ser     148
TTC CCT AGC ATG TAT ATC CAA ATC CAG GCT GTC TTG TCA TTG TAC TCA     526

Ala Gly Arg Thr Thr Gly Ile Val Val Asp Ser Gly Asp Gly Val Thr     164
GCA GGA AGA ACC ACT GGT ATC GTT GTT GAT TCT GGA GAT GGT GTA ACA     574

His Val Val Pro Ile Phe Glu Gly Tyr Gln Ile Pro His Ala Ile Glu     180
CAT GTT GTG CCA ATC TTC GAA GGT TAT CAA ATC CCA CAC GCC ATT GAG     622

Lys Ile Leu Leu Ala Gly Arg Asp Leu Thr Asp Tyr Met Cys Arg Ile     196
AAA ATC CTT CTT GCA GGA AGA GAT CTT ACT GAT TAC ATG TGT AGA ATC     670

Leu Lys Asp Asp Asp Tyr His Phe Glu Thr Thr Ala Glu Lys Glu Thr     212
CTC AAG GAC GAT GAC TAT CAC TTC GAG ACC ACT GCC GAG AAG GAA ACT     718

Val Arg Asp Ile Lys Glu Lys Leu Cys Tyr Val Ala Asp Asp Tyr Glu     228
GTC AGA GAT ATC AAA GAG AAG CTA TGC TAC GTT GCT GAT GAT TAC GAG     766

Ala Glu Leu Lys Lys Ala Gly Glu Gly Gly Glu Leu Glu Glu Ser Tyr     244
GCA GAG CTC AAG AAA GCA GGA GAA GGA GGT GAG CTA GAA GAG TCA TAC     814

Ala Leu Pro Asp Gly Arg Pro Leu Lys Ile Ser Thr Gln Arg Phe Gln     260
GCA TTG CCA GAT GGA CGG CCA CTC AAG ATC TCC ACT CAA AGA TTC CAA     862

Cys Pro Glu Phe Leu Phe Gln Pro Asp Leu Gly Gly Arg Glu Cys Lys     276
TGC CCA GAG TTC TTG TTC CAG CCA GAC TTG GGA GGC AGA GAA TGC AAG     910

Ser Val His Gln Leu Thr Tyr Asp Ser Ile Met Thr Cys Asp Leu Asp     292
AGC GTC CAC CAA CTG ACC TAC GAC TCC ATC ATG ACT TGT GAT TTG GAT     958

Val Arg Lys Asp Leu Tyr Ala Asn Ile Ile Leu Ser Gly Gly Thr Thr     308
GTC AGG AAG GAC TTA TAC GCT AAC ATC ATT CTC TCC GGA GGA ACC ACC    1006

Met Phe Pro Gly Leu Gly Glu Arg Leu Tyr Lys Glu Met Lys Asp Leu     324
ATG TTC CCA GGA CTC GGG GAG AGG CTC TAC AAG GAA ATG AAA GAC TTG    1054

Ala Pro Gln Thr Met Lys Val Lys Val Ile Ala Ser Pro Asp Arg Lys     340
GCT CCA CAG ACA ATG AAG GTC AAG GTC ATT GCG TCC CCA GAC AGA AAA    1102

Tyr Ala Val Trp Arg Gly Gly Ser Thr Leu Ala Lys Leu Ser Thr Phe     356
TAC GCA GTA TGG AGA GGA GGA TCA ACC CTC GCT AAG TTA TCT ACA TTC    1150

Ala Gly Met Trp Val Thr Lys Glu Asp Tyr Ala Glu Phe Gly Glu Ser     372
GCA GGA ATG TGG GTC ACT AAG GAG GAC TAC GCT GAG TTC GGA GAA AGT    1198

Ile Val His Arg Lys Cys Ile ###  379
ATC GTC CAC AGA AAG TGC ATC TAA 1223

TTTCATAGTG ATTATACGGA ATAGAAATAG AATCAATTTA TCAGAATAAC ATAGGGGTTT
TGGGGTTTTG GGGTTTTGGG G  1303
```

FIG. 2. Nucleotide sequence of the entire *E. crassus* macronuclear DNA molecule coding for actin. The deduced amino acid sequence of the actin-coding open reading frame is shown above the nucleotide sequence. Numbering of the nucleotide sequence begins at the telomeric CCCCAAAA repeats at the site where they were ligated to *Eco*RI linkers for cloning. ###, Presumed termination codon.

Cross-Species Comparison of *E. crassus* Actin and β-Tubulin. We compared the actin and tubulin DNA sequences for similarities to corresponding genes from other organisms. *O. nova* was previously shown to possess one of the most

```
AAAACCCCAA AACCCCAAAA CCCCATTTAT TCAAATTTTA TTTGAAGTTT TATAAAAATA
TTTAAATCTA AAG 73

                                        Met Arg Glu Ile Val His Val      7
                                        ATG AGA GAA ATC GTA CAC GTT      94

Gln Gly Gly Gln Cys Gly Asn Gln Ile Gly Ala Lys Phe Trp Glu Val      23
CAA GGA GGA CAA TGC GGA AAC CAG ATT GGT GCT AAG TTC TGG GAA GTC     142

Ile Ser Asp Glu His Gly Val Asp Pro Thr Gly Thr Tyr His Gly Asp      39
ATC TCT GAC GAA CAT GGT GTT GAC CCA ACT GGT ACC TAC CAC GGA GAC     190

Ser Asp Leu Gln Leu Glu Arg Ile Asn Val Tyr Tyr Asn Glu Ala Thr      55
TCT GAC TTG CAG CTC GAA AGA ATC AAC GTT TAC TAC AAC GAA GCA ACT     238

Gly Gly Arg Tyr Val Pro Arg Ala Val Leu Met Asp Leu Glu Pro Gly      71
GGC GGT AGA TAC GTG CCA AGA GCC GTC TTG ATG GAT CTC GAA CCA GGA     286

Thr Met Asp Ser Val Arg Ala Gly Pro Phe Gly Gln Leu Phe Arg Pro      87
ACC ATG GAC TCC GTC AGA GCC GGA CCA TTC GGA CAG CTC TTC AGA CCA     334

Asp Asn Phe Val Phe Gly Gln Thr Gly Ala Gly Asn Asn Trp Ala Lys     103
GAC AAC TTC GTC TTC GGT CAG ACT GGT GCT GGA AAC AAC TGG GCC AAG     382

Gly His Tyr Thr Glu Gly Ala Glu Leu Ile Asp Ser Val Leu Asp Val     119
GGA CAC TAC ACC GAA GGA GCT GAA CTC ATC GAC TCC GTC CTC GAC GTG     430

Val Arg Lys Glu Ala Glu Gly Cys Asp Cys Leu Gln Gly Phe Gln Ile     135
GTA AGA AAG GAA GCT GAA GGA TGC GAC TGC CTC CAA GGA TTC CAG ATC     478

Thr His Ser Leu Gly Gly Gly Thr Gly Ser Gly Met Gly Thr Leu Leu     151
ACC CAT TCT CTC GGA GGA GGA ACT GGA TCC GGT ATG GGA ACC CTC TTG     526

Ile Ser Lys Ile Arg Glu Glu Tyr Pro Asp Arg Ile Met Glu Thr Phe     167
ATC TCC AAG ATC AGA GAA GAG TAC CCA GAC AGA ATC ATG GAG ACC TTC     574

Ser Val Phe Pro Ser Pro Lys Val Ser Asp Thr Val Val Glu Pro Tyr     183
TCA GTC TTC CCA TCC CCA AAA GTC TCA GAT ACC GTC GTT GAG CCA TAC     622

Asn Ala Thr Leu Ser Val His Gln Leu Val Glu Asn Ala Asp Glu Val     199
AAC GCT ACC CTC TCC GTC CAT CAG CTC GTT GAG AAC GCC GAC GAA GTC     670

Met Val Ile Asp Asn Glu Ala Leu Tyr Asp Ile Cys Phe Arg Thr Leu     215
ATG GTC ATT GAC AAC GAA GCC TTG TAC GAC ATC TGT TTC AGA ACC TTG     718

Lys Leu Thr Thr Pro Thr Tyr Gly Asp Leu Asn His Leu Val Ser Ala     231
AAG TTG ACC ACT CCA ACC TAC GGA GAC TTG AAC CAC TTG GTC TCT GCC     766

Cys Ile Ser Gly Val Thr Ser Cys Leu Arg Phe Pro Gly Gln Leu Asn     247
TGT ATC TCC GGA GTC ACC TCA TGC TTG AGA TTC CCA GGA CAG TTG AAC     814

Ser Asp Leu Arg Lys Leu Ala Val Asn Leu Ile Pro Phe Pro Arg Leu     263
TCT GAC TTA AGA AAG TTG GCT GTC AAC TTG ATC CCA TTC CCA AGA CTC     862

His Phe Phe Met Val Gly Phe Ala Pro Leu Thr Ser Arg Gly Ser Gln     279
CAC TTC TTC ATG GTT GGA TTC GCC CCA TTG ACC TCC AGA GGA TCC CAA     910

Gln Tyr Arg Ala Leu Thr Val Pro Glu Leu Thr Gln Gln Met Phe Asp     295
CAA TAC AGA GCC TTG ACT GTT CCA GAG CTC ACC CAG CAA ATG TTC GAC     958

Ala Lys Asn Met Met Cys Ala Ser Asp Pro Arg His Gly Arg Tyr Leu     311
GCT AAG AAC ATG ATG TGT GCT TCC GAC CCA AGA CAC GGA AGA TAC TTG    1006

Thr Ala Ser Ala Met Phe Arg Gly Arg Met Ser Thr Lys Glu Val Asp     327
ACT GCC TCC GCC ATG TTC AGA GGA AGA ATG TCC ACT AAA GAA GTT GAC    1054

Glu Gln Met Leu Asn Val Gln Asn Lys Asn Ser Ser Tyr Phe Val Glu     343
GAA CAA ATG TTG AAT GTC CAG AAC AAG AAC TCC TCC TAC TTC GTA GAG    1102

Trp Ile Pro Asn Asn Ile Lys Ser Ser Val Cys Asp Ile Pro Pro Lys     359
TGG ATT CCA AAC AAC ATC AAG TCC TCT GTC TGC GAT ATC CCA CCA AAG    1150

Gly Leu Lys Leu Ala Ser Thr Phe Ile Gly Asn Ser Thr Ala Ile Gln     375
GGA CTC AAG CTC GCT TCT ACC TTC ATC GGA AAC TCG ACT GCC ATC CAG    1198

Glu Met Phe Lys Arg Val Ala Glu Gln Phe Thr Ala Met Phe Arg Arg     391
GAA ATG TTC AAG AGA GTC GCC GAA CAA TTC ACT GCC ATG TTC AGA AGA    1246

Lys Ala Phe Leu His Trp Tyr Thr Gly Glu Gly Met Asp Glu Met Glu     407
AAG GCC TTC TTG CAT TGG TAT ACC GGA GAA GGA ATG GAC GAA ATG GAG    1294

Phe Thr Glu Ala Glu Ser Asn Met Asn Asp Leu Val Ser Glu Tyr Gln     423
TTC ACC GAA GCC GAG TCC AAC ATG AAC GAT CTC GTC TCT GAA TAC CAA    1342

Gln Tyr Gln Asp Ala Thr Ala Glu Glu Glu Gly Glu Tyr Val Glu Asp     439
CAA TAC CAG GAT GCC ACT GCT GAA GAA GAA GGA GAG TAT GTC GAA GAC    1390

Glu Asp Glu Met Asp Gly Met ###
GAA GAT GAA ATG GAC GGA ATG TAA

ACTTAATTCA GTTCTTAAAA TCCTCTGACC CTCACTTTTC TCTATTATAA TAGTATATCA
ACTTCATGAA TCTGGTCTGG GGTTTTGGGG TTTTGGGGTT TTGGGG 1521
```

FIG. 3. Nucleotide sequence of the entire *E. crassus* macronuclear DNA molecule coding for β-tubulin. The deduced amino acid sequence of the tubulin-coding open reading frame is shown above the nucleotide sequence. Numbering of the nucleotide sequence begins at the telomeric CCCCAAAA repeats at the site where they were ligated to *Eco*RI linkers for cloning. ###, Presumed termination codon.

atypical actins sequenced to date, whereas the *T. thermophila* actin gene bears more resemblance to actins from other organisms. Comparison of *E. crassus* actin to other known actins shows that it possesses 60–65% identity with most known actins on both the DNA and amino acid levels. In contrast, the *E. crassus* and *S. lemnae* (34) β-tubulin gene sequences are 85% similar to each other at the DNA level and 95% at the amino acid level. The *E. crassus* β-tubulin gene shows at least 85% identity to chicken, pig, human, and trypanosome β-tubulin and 72% identity to yeast β-tubulin at the amino acid level. Therefore, while the *E. crassus* β-tubulin appears fairly highly conserved relative to β-tubulins from other organisms, the actin gene seems to be extremely divergent, both from other ciliate actin genes and from actin genes in other organisms.

Codon Usage in *E. crassus*. We also compared the codon usage of the *E. crassus*, *O. nova* (19), and *T. thermophila* (6, 8) actin genes, and the *E. crassus* and *S. lemnae* (34) β-tubulin genes, to determine whether any differences were evident. It has previously been reported that actins from single-celled eukaryotes use between 37–41 codons, whereas those from higher eukaryotes use 51–54 codons (21). Table 1 shows a codon usage table for the *E. crassus* actin and β-tubulin. *E. crassus* uses 54 codons in coding for actin; this corresponds to the number seen in higher eukaryotes. As a result, *E. crassus* exhibits less extreme codon bias than do *Oxytricha* or *Tetrahymena*. In contrast, the *E. crassus* β-tubulin uses only 43 codons as compared with 42 for *Stylonychia*. Similar codon bias is seen for arginine, phenylalanine, asparagine, and proline in both sequences, whereas codon usage for glycine and serine differ. As previously mentioned, the total number of codons used in the two sequences is quite different. It has been shown that codon bias for abundantly expressed genes reflects tRNA abundance (22). In addition, Martindale has recently conducted a survey of codon usage in ciliate macronuclear genes (4). This study indicates that *Tetrahymena* genes expressed at high levels during normal cell growth have a stronger bias towards the use of preferred codons than a gene expressed during the sexual phase of the life cycle. It is not clear that a similar tendency for highly expressed genes to exhibit relatively stronger codon bias than other genes exists in hypotrichs.

Timing of Ciliate Genetic Code Fixation. The DNA sequence data presented here show that *E. crassus* uses TAA as a stop codon. In addition, DNA sequence analysis of a cDNA clone encoding a *Euplotes raikovi* pheromone indicates that this *Euplotes* species also uses TAA as a stop codon (23). Furthermore, *in vitro* translation experiments using wheat germ and reticulocyte extracts to translate mRNA from *Euplotes octocarinatus* demonstrate the production of full-length protein products (H. Schmidt, personal communication). This is in contrast to results obtained using *Tetrahymena* (24) or *Paramecium* (25) mRNA in similar *in vitro* translation experiments, where only truncated proteins are produced. Addition of tRNAs from *T. thermophila* to the *in vitro* translation system restores the ability to produce full-length protein products (24). As previously discussed, *T. thermophila* possesses tRNAs that specifically recognize TAA and TAG and are aminoacylated with glutamine (12). Based on this *in vitro* translation data it appears that *E. octocarinatus*, a fresh-water Euplotid, uses the universal genetic code (i.e., TAA, TAG, and TGA used as stop codons); otherwise, it too would produce truncated protein products. Data on *E. crassus* (this paper) and *E. raikovi* (23), both marine Euplotids, only indicate that TAA is used as a stop codon; we do not know how TAG and TGA are used. The fact that both freshwater and marine Euplotids have been shown to differ from other ciliates in their termination codon usage suggests that the Euplotids as a group may share this difference from other ciliates.

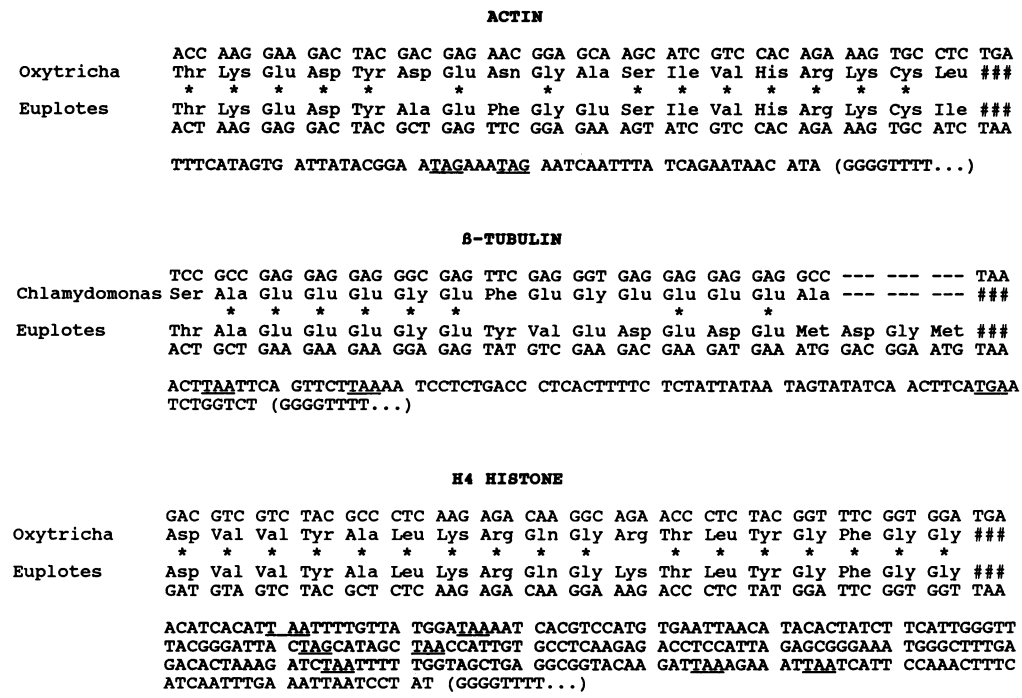

FIG. 4. Comparison of the deduced sequence of the terminal 20 amino acids of *E. crassus* actin, β-tubulin, and H4 histone to sequences from other organisms. Nucleotide and deduced amino acid sequences are shown for the following three comparisons: (*A*) *E. crassus* actin gene vs. *O. nova* actin gene (19). (*B*) *E. crassus* β-tubulin gene vs. *C. reinhardtii* β-tubulin gene (18). (*C*) *E. crassus* H4 histone gene vs. *O. nova* H4 histone gene (13). The nucleotide sequence of the 3′ ends of the actin, β-tubulin, and histone genes are shown with the start of the telomeric GGGGTTTT repeats shown in parentheses. ###, Presumed stop codon in *Euplotes*; *, identity of amino acid sequence. Underlined nucleotides in the 3′ noncoding regions indicate potential in-frame termination codons that are further 3′ to the putative termination codons.

The use of TAA as a stop codon in Euplotids means that the timing of genetic code changes in the ciliates must be reconsidered. Based on previous DNA sequence analysis of ciliate macronuclear genes, it has been postulated that these organisms use TAA and TAG as codons for glutamine or glutamic acid and use TGA as their sole termination codon. Although tRNAs specific for these codons have been isolated and characterized from *T. thermophila* (12), whether *Stylonychia, Oxytricha*, and *Paramecium* possess similar tRNAs is not known. It has been proposed (12, 26) that the duplication of glutamine tRNA UUG and the subsequent mutation of one copy to glutamine tRNA UUA were the crucial events leading to an altered code, with subsequent mutations eliminating the use of TAA and TAG as stop codons. Hanyu *et al.* (12) have proposed that TAA and TAG were rarely used as stop codons in prociliates and that weak suppressor tRNAs may have existed at a very early stage. Therefore, mutations creating TAA and TAG codons within protein-encoding genes would not have been lethal. The findings reported here indicate that these alterations in the genetic code arose much later in the evolution of ciliates than previously thought, suggesting that prociliates used the universal genetic code.

One possible explanation for use of different termination codons by the Euplotids as compared with other ciliates is that two or more distinct lines of ciliates have evolved that use either the universal code or a universal code altered by the fixation of different termination codons and suppressor tRNAs. If so, then further comparisons of DNA or RNA sequences, along with characterization of termination codon use (i.e., by *in vitro* translation) should demonstrate a distinct divergence of ciliate lines that use different genetic codes. Although 5S RNA (27) and small subunit ribosomal RNA (28) sequences demonstrate significant divergence of the Euplotids from *Tetrahymena* and *Paramecium* (27) or from *Oxytricha* and *Stylonychia* (28), the sequence data base for ciliates is much too small to argue for or against the divergence of distinct lines of ciliates.

An alternative view is that the genetic code allows for a far greater degree of flexibility than previously realized and provides for rapid evolution towards differential use of termination codons. This view is suggested by the presence

Table 1. Codon usage in *E. crassus* actin and β-tubulin

	Number*			Number			Number			Number	
Codon	A	T	Codon	A	T	Codon	A	T	Codon	A	T
TTT-Phe	2	0	TCT-Ser	2	8	TAT-Tyr	6	2	TGT-Cys	4	3
TTC-Phe	12	23	TCC-Ser	4	16	TAC-Tyr	11	14	TGC-Cys	6	5
TTA-Leu	2	1	TCA-Ser	8	3	TAA-***	1	1	TGA-***	0	0
TTG-Leu	9	17	TCG-Ser	0	1	TAG-***	0	0	TGG-Trp	4	4
CTT-Leu	4	0	CCT-Pro	1	0	CAT-His	2	4	CGT-Arg	0	0
CTC-Leu	7	15	CCC-Pro	0	0	CAC-His	4	6	CGC-Arg	1	0
CTA-Leu	3	0	CCA-Pro	16	19	CAA-Gln	11	10	CGA-Arg	0	0
CTG-Leu	2	0	CCG-Pro	0	0	CAG-Gln	3	11	CGG-Arg	1	0
ATT-Ile	8	3	ACT-Thr	10	10	AAT-Asn	5	1	AGT-Ser	2	0
ATC-Ile	15	15	ACC-Thr	7	17	AAC-Asn	5	20	AGC-Ser	3	0
ATA-Ile	0	0	ACA-Thr	4	1	AAA-Lys	9	2	AGA-Arg	13	22
ATG-Met	12	20	ACG-Thr	0	0	AAG-Lys	20	13	AGG-Arg	2	0
GTT-Val	11	8	GCT-Ala	10	10	GAT-Asp	16	6	GGT-Gly	9	7
GTC-Val	11	19	GCC-Ala	3	16	GAC-Asp	10	21	GGC-Gly	1	1
GTA-Val	4	3	GCA-Ala	9	1	GAA-Glu	16	27	GGA-Gly	20	30
GTG-Val	1	2	GCG-Ala	1	0	GAG-Glu	16	9	GGG-Gly	1	0

*Numbers of occurrences of each codon are given for actin (A) and β-tubulin (T).

of natural suppressor tRNAs in a number of other organisms (for review, see ref. 29). These tRNAs are thought to recognize and translate termination codons in order to insert modified amino acids or to regulate the synthesis of a specific protein. Recognition of termination codons by these natural suppressor tRNAs appears to depend on subtleties of tRNA structure and/or codon context. Therefore, it is possible that the ciliates represent an extreme form of this context recognition process and that different ciliates use suppressors to different degrees. Ciliates, in general, may use termination codons to specify both amino acids and termination with quantitative variation in suppressor tRNA amounts and the degree of fixation of internal stop codons. In any case, further study of the evolution of ciliates should provide interesting data concerning the evolution of the genetic code.

Evolution of Macronuclear Development. The DNA sequence data presented here lead to questions concerning the evolution of macronuclear structure in the ciliates. Traditionally, the classification of ciliates on the basis of morphology has grouped together the Euplotids (e.g., *E. crassus*) and Oxytrichids (*Oxytricha* and *Stylonychia*) as hypotrichs (1). This classification scheme fits with data concerning macronuclear development within these organisms. In the Oxytrichids and Euplotids, the time course of mating, meiosis, and changes in the micronucleus leading to formation of a macronucleus is very similar and substantially different from *Tetrahymena* (1, 30). The DNA processing events during macronuclear development, including polytenization of micronuclear chromosomes, elimination of 90–95% of the DNA during a so-called "vesicle" stage, and the formation of "gene-sized" linear molecules terminating in repeats of the sequence CCCCAAAA, are also related (1). In addition, all of these organisms show an unusual macronuclear DNA replication process involving "replication bands." The hypotrich mode of macronuclear development has generally been viewed as the most highly evolved and specialized, while organisms that produce larger macronuclear molecules and eliminate less DNA (e.g., *Tetrahymena, Paramecium*) are considered as precursors to the hypotrich mode (31).

Several new lines of evidence suggested a change in the classification of hypotrichs. A more detailed consideration of morphology has led to the separation of the Euplotids and the Oxytrichids into two separate subphyla (32). Unfortunately, this revision of the taxonomy does not agree with the results presented here, as the Euplotids are included in the same subphylum as *Paramecium* and *Tetrahymena*, with the Oxytrichids in a separate subphylum. DNA sequence analysis of the small subunit rRNA genes from *O. nova, Stylonychia pustulata*, and *Euplotes aediculatus* demonstrated a large divergence between the Oxytrichid (i.e., *O. nova* and *S. pustulata*) and Euplotid sequences—with the two Oxytrichid sequences being closely related (28). Comparison of these rRNAs to the *Saccharomyces cerevisiae* small subunit rRNA sequence indicated that the *E. aediculatus* sequence has evolved at a faster rate than the other two, although the observed difference in evolutionary rate was not considered substantial enough to account for the divergence of the Euplotid and Oxytrichid sequences. The divergence of Oxytrichids and Euplotids as two separate lines is also supported by the lack of homology of actin genes as documented above.

Recently, two organisms with different morphological characteristics have been shown to possess small linear DNA molecules in their macronuclei, and at least one of these polytenizes its chromosomes during macronuclear development (33). Thus, the hypotrich mode of macronuclear development may be more widespread than previously thought and may not represent the most-evolved form. Gene-sized macronuclear molecules could represent a less-evolved form of organization; larger macronuclear chromosomes would evolve by a decrease in the frequency of telomere addition during macronuclear development. This possibility can be tested by analyzing the macronuclear DNA of a wide range of ciliates and determining how many possess hypotrich-like macronuclear chromosomes.

The authors thank Mark Krikau for technical assistance and Dr. Helmut Schmidt for providing important information prior to publication. This research was supported by a grant from the National Institutes of Health (GM37661) to C.L.J.

1. Klobutcher, L. A. & Prescott, D. M. (1987) in *The Molecular Biology of Ciliated Protozoa*, ed. Gall, J. (Academic, New York), pp. 111–154.
2. Yao, M.-C. & Gorovsky, M. A. (1974) *Chromosoma* **48**, 1–18.
3. Altschuler, M. I. & Yao, M.-C. (1985) *Nucleic Acids Res.* **13**, 5817–5831.
4. Martindale, D. W. (1989) *J. Protozool.*, in press.
5. Helftenbein, E. (1985) *Nucleic Acids Res.* **13**, 415–433.
6. Cupples, C. G. & Pearlman, R. E. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5160–5164.
7. Horowitz, S. & Gorovsky, M. A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 2452–2455.
8. Hirono, M., Endoh, H., Okada, N., Numata, O. & Watanabe, Y. (1987) *J. Mol. Biol.* **194**, 181–192.
9. Caron, F. & Meyer, E. (1985) *Nature (London)* **314**, 185–188.
10. Preer, J. R., Preer, L. B., Rudman, B. M. & Barnett, A. J. (1985) *Nature (London)* **314**, 188–190.
11. Herrick, G., Hunter, D., Williams, K. & Kotter, K. (1987) *Genes Dev.* **1**, 1047–1058.
12. Hanyu, N., Kuchino, Y. & Nishimura, S. (1986) *EMBO J.* **5**, 1307–1311.
13. Harper, D. S. & Jahn, C. L. (1989) *Gene* **75**, 93–107.
14. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78.
15. Messing, J., Crea, R. & Seeburg, P. M. (1981) *Nucleic Acids Res.* **9**, 309–321.
16. Hattori, M. & Sakaki, Y. (1986) *Anal. Biochem.* **152**, 232–238.
17. Tabor, S. & Richardson, C. C. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 4767–4771.
18. Youngbloom, J., Schloss, J. A. & Silflow, C. D. (1984) *Mol. Cell. Biol.* **4**, 2686–2696.
19. Greslin, A. F., Loukin, S. H., Oka, Y. & Prescott, D. M. (1988) *DNA* **7**, 529–536.
20. Cleveland, D. W. & Sullivan, K. F. (1985) *Annu. Rev. Biochem.* **54**, 331–365.
21. Nudel, U., Zakut, R., Shani, M., Neuman, S., Levy, Z. & Yaffe, D. (1983) *Nucleic Acids Res.* **11**, 1759–1771.
22. Ikemura, T. & Ozeki, H. (1983) *Cold Spring Harbor Symp. Quant. Biol.* **47**, 1087–1097.
23. Miceli, C., La Terza, A. & Melli, M. L. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 3016–3020.
24. Andreasen, P. H., Dreisig, H. & Kristiansen, K. (1987) *Biochem. J.* **244**, 331–335.
25. Meyer, E., Caron, F. & Guiard, B. (1984) *Biochimie* **66**, 403–412.
26. Osawa, S. & Jukes, T. H. (1988) *Trends Genet.* **4**, 191–198.
27. Kumazaki, T., Hori, H. & Osawa, S. (1983) *J. Mol. Evol.* **19**, 411–419.
28. Sogin, M. L., Swanton, M. T., Gunderson, J. H. & Elwood, H. J. (1986) *J. Protozool.* **33**, 26–29.
29. Valle, R. P. C. & Morch, M.-D. (1988) *FEBS Lett.* **235**, 1–15.
30. Gorovsky, M. A. (1980) *Annu. Rev. Genet.* **14**, 203–239.
31. Steinbruck, G. (1986) in *Germ Line—Soma Differentiation*, Results and Problems in Cell Differentiation, ed. Hennig, W. (Springer, Berlin), Vol. 13, pp. 105–174.
32. Small, E. B. & Lynn, D. H. (1985) in *An Illustrated Guide to the Protozoa*, eds. Lee, J. J., Hutner, S. H. & Bovee, E. C. (Allen, Lawrence, KS), pp. 393–575.
33. Metenier, G. & Hufschmid, J. D. (1988) *J. Protozool.* **35**, 71–73.
34. Conzelmann, K. K. & Helftenbein, E. (1987) *J. Mol. Biol.* **198**, 643–653.