A nascent micronuclear pseudogene in the ciliate

*Euplotes crassus*

Volker Florian and Albrecht Klein*

Molecular Genetics, Department of Biology, Philipps University, D-35032 Marburg, Germany

Received April 29, 1996; Revised and Accepted July 1, 1996  
EMBL accession nos +

**ABSTRACT**

The macronuclear genome of *Euplotes crassus* contains two different genes, *EFA1* and *EFA2*, encoding EF-1α proteins. They are derived from micronuclear precursors in the course of a sexual process termed conjugation. We have found that two apparent micronuclear precursors exist for *EFA1*. They differ in their potential coding sequences and in the internal sequences interrupting the genes, which are normally removed during the processing from micronuclear into macronuclear genes. One of these genes is not processed into a macronuclear gene and has accumulated C→T transitions in a limited section of the coding region. The gene obviously constitutes a recent duplication which has lost its ability to be processed into a macronuclear gene and has therefore become a micronuclear pseudogene. The true *EFA1* precursor harbours a novel type of internal sequence in addition to a classical AT-rich IES. As usual, only one micronuclear *EFA2* precursor gene was found. Its coding sequence is interrupted by a 79 bp TelIES.

**INTRODUCTION**

Ciliated protozoa are characterized by nuclear dimorphism (1,2). The cells contain generative diploid micronuclei and vegetative macronuclei. This situation can be viewed as a parallel to the germ line–soma differentiation of higher eukaryotes, since only macronuclei are physiologically active. The DNA contents and sequence complexities of the two nuclear types differ. Macronuclei have lower sequence complexity but contain more DNA. This is due to the mode of their generation. During a sexual phase cells of compatible mating types form pairs and exchange meiosis products of their micronuclei. They fuse with the non-transferred haploid nuclei of the respective partner cell to form synkarya. The new diploid nucleus in each of the partner cells divides at least once to form one or more micronuclei, depending on the organism, and the macronuclear precursor, termed the anlage. The anlagen chromosomes undergo rearrangements, which are most pronounced in the hypotrichous ciliates (3–6) to which *Euplotes crassus* belongs. First, the anlagen chromosomes undergo polynization. Different types of sequences which interrupt the macronucleus-destined sequences of the micronuclear genes are then removed in several steps (7,8) and the chromosomes are subsequently fragmented into gene sized molecules. These nascent macronuclear genes are differentially amplified and telomeres are attached (9–11) before a further, general amplification follows to complete the development of the new macronuclear chromosomes. The macronucleus-destined sequences, i.e. the precursor sequences of the macronuclear genes, are clustered in the micronuclear chromosomes (12). Since in addition to the removed intragenic sequences the long intergenic regions are lost during formation of the macronuclear chromosomes, the sequence complexity is reduced by >90%, while the total DNA content increases.

Three classes of sequences that can interrupt the coding regions of macronuclear gene precursors have been described. Internal eliminated sequences (IES) are AT-rich non-coding short sequences (13), transposon-like elements of *E.crassus* (Tec) are very frequent long sequences with open reading frames and inverted repeats (14–16), which probably enable them to transpose (17). TelIES elements are very short relatively GC-rich sequences containing C4A4 motifs reminiscent of macronucle some telomeres (18). All these elements are flanked by TA repeats. A common excision mechanism has been proposed for the Tec and IES sequences involving circular intermediates which are most likely not formed in the course of removal of the very short TelIES (19,20).

We have been interested in the organization and expression of genes encoding EF-1α translation factors in hypotrichous ciliates. *Euplotes crassus* contains two different such genes, *EFA1* and *EFA2*, which are both transcribed (21). They differ in copy number and codon usage, pointing to different expression levels, which have indeed been found. Since we have indications that the macronuclear gene copy number might be correlated with the timing of excision from the polytene anlagen chromosome (Dönhoff and Klein, unpublished results), we were interested to characterize the micronuclear counterparts of the two *EFA* genes. The macronuclear *EFA1* gene is interrupted by an AT-rich IES sequence and in addition contains a long novel interrupting sequence with 12 bp terminal repeats. In addition, we have discovered an *EFA1* pseudoprecursor which is not processed into a macronuclear gene. It contains the same IES element but lacks the novel intervening sequence found in the functional precursor and shows point mutations in a narrow range of its open reading frame. Only one *EFA2* precursor was found, which harbours a short TelIES element.

* To whom correspondence should be addressed

*+X97409–X97412 and X98602–X98605 (incl.)*
Nucleic Acids Research, 1996, Vol. 24, No. 16

3196

MATERIALS AND METHODS

Organisms and growth conditions

Euplotes crassus strains Por-3 and Liv-1 (22), as well as the feeding alga Dunaliella tertiolecta, were kindly supplied by Prof. P. Luporini (University of Camerino, Italy). The ciliates and the algae were grown at 22°C in artificial seawater as described previously.

General methods of DNA purification and analysis

General DNA preparation, electrophoresis, restriction analysis and hybridization techniques have been previously described (23–25).

Preparation of total E.crassus DNA

For the preparation of total cellular DNA E.crassus was grown to a density of 2–3 × 10^9/ml. Feeding was then stopped and the organisms left without food for 3 days. The cells were harvested by two subsequent filtration steps through 30 and 10 µm nylon gauze, rinsed off with seawater and collected by low speed centrifugation. They were resuspended in 3 vol lysis buffer (10 mM Tris–HCl, pH 7.5, 10 mM EDTA, 250 mM NaCl, 0.5% SDS) (8) and lysed at 65°C for 15 min. Proteins were digested at 50°C overnight after addition of 200 µg/ml trypsin K. The solution was extracted with phenol and the DNA dissolved in TE (10 mM Tris–HCl, pH 7.5, 1 mM EDTA) after ethanol precipitation.

Preparation of micronuclear DNA

The cells were collected as described above. They were resuspended in 10 ml TE/ml packed cells. After addition of 0.1% Triton X-100 and 1 µg/ml trypsin the suspension was incubated at room temperature. Lysis of the cells and the macronuclei was followed microscopically. As soon as the macronuclei had been destroyed the trypsin treatment was stopped by transferring the lysate to ice and addition of 2 µg/ml trypsin inhibitor. The nuclei were spun out of the lysate at low speed (500 g) and 4°C, then washed twice with TE to remove most of the macronuclear DNA. After the addition of 1 ml lysis buffer, the micronuclear DNA was prepared by proteinase K digestion, phenol extraction and ethanol precipitation as described above. The DNA was applied to a preparative 1% agarose gel and separated from residual macronuclear DNA. The visible high molecular weight DNA was cut out of the gel and electrophoresed through a similar gel. The DNA was removed from the agarose by electroelution and precipitated by isopropanol in the presence of 1% glycogen.

Polymerase chain reactions and DNA sequencing

DNA amplification by polymerase chain reactions (PCR) and inverse PCR were performed as described (26,27). The products were directly sequenced according to established methods (28,29). The primers used are listed in Table 1. Their positions are shown in Figure 1. Sequencing of cloned DNA was performed according to the dideoxynucleotide termination method (30).

Table 1. Primers used for analysis of the micronuclear EF-1α encoding genes

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
<th>Position a</th>
</tr>
</thead>
<tbody>
<tr>
<td>1F1</td>
<td>GATAAATTTTCGAGTTGAGCG</td>
<td>3</td>
</tr>
<tr>
<td>1R1</td>
<td>CAATAAGCAAGATTTGAGATG</td>
<td>139</td>
</tr>
<tr>
<td>1F2</td>
<td>GAGTCTGGTGAATGGGTAAAGGC</td>
<td>232</td>
</tr>
<tr>
<td>1R2</td>
<td>AGCCTTGGCTATCAGGATTACCC</td>
<td>294</td>
</tr>
<tr>
<td>1F3</td>
<td>GAAGCTGTGCTCTCAGAGAGAAAGG</td>
<td>469</td>
</tr>
<tr>
<td>1F4</td>
<td>CAGGTGCATGTTATCTCAAGAAGG</td>
<td>581</td>
</tr>
<tr>
<td>1F5</td>
<td>GTTGGCTTTTGAGGGCAGGTTACCC</td>
<td>660</td>
</tr>
<tr>
<td>1F6</td>
<td>GCCGTTGATGGTCAATACCTCCGCC</td>
<td>883</td>
</tr>
<tr>
<td>1R4</td>
<td>GCTGCTGGAGCCGCTTTGGACG</td>
<td>981</td>
</tr>
<tr>
<td>1F6</td>
<td>ATGTGACACTGGCCATATGTCGC</td>
<td>1148</td>
</tr>
<tr>
<td>1R5</td>
<td>CTGGTGGAGGATTCTTCCTCAG</td>
<td>1197</td>
</tr>
<tr>
<td>1R6</td>
<td>TGTAGATGTCAGCCATACCTCGGCC</td>
<td>1270</td>
</tr>
<tr>
<td>1F7</td>
<td>GCTGCTGAGATGATGACACG</td>
<td>1336</td>
</tr>
<tr>
<td>1F8</td>
<td>CTGGTGGGACGCTGATTTCATC</td>
<td>1357</td>
</tr>
<tr>
<td>1R7</td>
<td>ATGATGACTATACACAGACAGAG</td>
<td>1488</td>
</tr>
<tr>
<td>1iF</td>
<td>GCCAAGTTCGGTACATTTGAG</td>
<td>1336</td>
</tr>
<tr>
<td>1iR</td>
<td>GTTGGGGCGCCTGTGATACTC</td>
<td>1495</td>
</tr>
<tr>
<td>2F1</td>
<td>CCAATTCATGGTAAAGATTTGAGG</td>
<td>45</td>
</tr>
<tr>
<td>2R1</td>
<td>CCAATTCATGGTAAAGATTTGAGG</td>
<td>685</td>
</tr>
<tr>
<td>2F2</td>
<td>GAGATTGTCATGAGAGGTTAGAG</td>
<td>644</td>
</tr>
<tr>
<td>2R2</td>
<td>CTTGGTGATGAGGCTTCAAG</td>
<td>1004</td>
</tr>
<tr>
<td>2F3</td>
<td>GAAGTTGATGTCACCGTGAG</td>
<td>942</td>
</tr>
<tr>
<td>2R3</td>
<td>CCTCTGTGTAATCAAGAG</td>
<td>1495</td>
</tr>
</tbody>
</table>

aThe numbers are the positions of the 5’-ends of the primers in the published macronuclear sequences of EFA1 (GenBank accession no. U26260) or EFA2 (GenBank accession no. U26267) for primers 2F1 and 2R1. Primers 1iR and 1iF start in the macronuclear sequence and reach into the AT-rich IES element (compare Fig. 3). All primers designated F read in the 5’→3’ direction of the gene. The R primers read in the opposite direction.
Figure 2. Amplification products of macronuclear and micronuclear DNA with primers located in the EF A1 sequence. Macronuclear (lanes 1, 3 and 5) or micronuclear DNA (lanes 2, 4 and 6) was used as templates for PCR reactions employing primer pairs 1F1/1R3 (lanes 1 and 2), 1F3/1R4 (lanes 3 and 4) or 1F5/1R7 (lanes 5 and 6). In lane M a size marker is shown. The fragment sizes are given in kb. Note that a product obtained due to a macronuclear DNA impurity is seen beside the micronucleus-specific product in lane 4. For primer locations compare Figure 1.

**Analysis of single strand conformational polymorphisms (SSCP)**

The method of Hongyo et al. (31) was modified as follows. PCR products were separated on agarose gels to separate them from primer DNA and eluted from the gel. Aliquots of 10 ng each of the DNA samples to be compared were mixed in 15 µl H2O. Then 1 µl sample buffer (15% Ficoll, 0.25% bromophenol blue, 0.25% xylene cyanol FF) and 1 µl 1 N NaOH were added. The DNA was denatured at 90°C for 4 min and immediately chilled on ice. The samples were analysed on 120 × 155 × 1 mm 10% polyacrylamide gels containing 5% glycerol in 0.5 × TBE (4.5 mM Tris–borate, pH 8.3, 0.25 M EDTA). The gels were run at 120 V for 16–18 h at 4 and 20°C in parallel, since the single strand conformation critically depends on temperature. The DNA was visualized by silver staining (32).

**RESULTS**

**A 347 bp IES sequence interrupts an apparent micronuclear EFA1 precursor**

Macronuclear or micronuclear DNA was used as template for the amplification of EF A1 sequences with three primer pairs derived from the coding sequence of the macronuclear gene (Fig. 1). Figure 2 shows that one of the obtained products from the micronuclear DNA was ~350 bp longer than its macronuclear counterpart. Sequence determination yielded a 347 bp AT-rich IES sequence interrupting the open reading frame (Fig. 3). In order to see whether the known coding sequence of the macronuclear and newly obtained micronuclear EF A1 genes were identical, primers (1iF and 1iR) were designed which have their 5′-termini in the coding sequence and their 3′-termini in the IES sequence. When combined with the primers from the non-coding 5′ and 3′ gene regions (1F1 and 1R7) only micronuclear sequences can be obtained by PCR amplification (compare Fig. 1). The products were sequenced directly in order to avoid cloning artifacts. Surprisingly, point mutations were discovered in a limited area of the coding region of the micronuclear copy (Fig. 4). This gene is subsequently called MicEF A1A.

The mutational changes could also be visualized by a subsequent SSCP analysis of amplification products in comparison with their counterparts from macronuclear DNA (Fig. 5). In order to see whether additional mutations had occurred in different parts of the coding region as well, amplification products of micronuclear and macronuclear DNA with primers encompassing the other parts of the genes were also included in the SSCP analysis. Surprisingly, no differences in SSCP behaviour were seen in any other case, confirming that the point mutations were localized to a short region of the micEF A1A gene (Fig. 4). This gene was confirmed by analysis of the parts of the pseudogene corresponding to the potential micronucleus-destined sequences of this pseudoprecursor. These sequences were found to be identical to the respective sequences of the macronuclear EFA1 gene.

When cells carrying this micronuclear gene were mated and the exconjugants analysed, the mutated gene was not found in the new macronuclear DNA, confirming that the micEF A1A gene is

**Figure 3.** Sequence of the IES interrupting the coding sequences in the micEF A1A genes. Identical sections of the micEF A1A and micEF A1B genes are shown. Nucleotides of the coding sequences adjacent to the IES are given in lower case, together with the encoded amino acids. Imperfect inverted repeat sequences are underlined. The terminal repeats are shown in bold.

**Figure 4.** C→T transition mutations found in the coding sequence of the micEF A1A pseudoprecursor gene. Three examples are visible by comparison of the micronuclear micEF A1A (Mi) and the macronuclear EF A1 (Ma) sequences. Two more such transitions were found at positions 1038 and 1073. The numbering follows that of the gene sequences in GenBank.

**Figure 5.** SSCP analysis of amplification products with primers encompassing the other parts of the genes. No differences in SSCP behaviour were seen in any other case, confirming that the point mutations were localized to a short region of the micEF A1A gene (Fig. 4). This gene was confirmed by analysis of the parts of the pseudogene corresponding to the potential micronucleus-destined sequences of this pseudoprecursor. These sequences were found to be identical to the respective sequences of the macronuclear EFA1 gene.
DNA showed that two different micronuclear Southern hybridization of a restriction digest of micronuclear a novel intervening sequence pseudogene. not a precursor for a macronuclear gene but rather a micronuclear sequence as the macronuclear

The DNA was digested with the restriction endonuclease which does not cut within the sequence of the macronuclear gene nor in the IES sequence described above (compare Fig. 3). The DNA was then electrophoretically separated and a Southern blot was hybridized with the macronuclear DNA as a probe. The autoradiogram presented in Figure 6 shows a relatively strong signal caused by a macronuclear impurity and two additional ones. One of these corresponds to the micronuclear sequence as a probe.

A second micronuclear EFA1 gene has the same coding sequence as the macronuclear EFA1 gene and contains a novel intervening sequence

Southern hybridization of a restriction digest of micronuclear DNA showed that two different micronuclear EFA1 genes exist. The DNA was digested with the restriction endonuclease StuI, which does not cut within the sequence of the macronuclear gene nor in the IES sequence described above (compare Fig. 3). The DNA was then electrophoretically separated and a Southern blot was hybridized with the macronuclear EFA1 sequence as a probe. The autoradiogram presented in Figure 6 shows a relatively strong signal caused by a macronuclear impurity and two additional ones. One of these corresponds to the StuI fragment containing the micEFA1A gene. The second one is due to a restriction fragment containing most of a second micronuclear EFA1 gene, micEFA1B, as will be shown below. The fact that the signal strength of this smaller micronuclear fragment is lower than that of the larger one indicated that one of the StuI sites was located in an intervening sequence interrupting the coding sequence of micEFA1B. This could indeed be shown by inverse PCR. Micronuclear DNA was digested with AflIII or EaeI restriction endonucleases, which fulfil the same criteria as StuI, and subsequently religated to obtain DNA circles. These were suitable substrates for amplification with the primer pairs 1R2 and 1F3 or 1R6 and 1F7 respectively. The amplification products were re-amplified with primer pairs 1R1 and 1F4 or 1R5 and 1F8 respectively, to make sure that specific products had been obtained. Sequence analyses yielded the nucleotide sequences of the ends of an intervening sequence between nucleotides 1134 and 1135 of the EFA1 sequence. Strikingly, a 12 bp sequence is duplicated at the ends of this intervening sequence, in contrast to the TA duplication found in IES and Tec sequences (Fig. 7). Attempts to amplify the missing part of the intervening sequence with appropriate primer combinations, derived from the sequences obtained after the inverse PCR reactions, failed. This indicates that it may be quite long, because >2 kb amplification products with other primer–template combinations were obtained under the same conditions. The amplification product obtained with the primer pair 1R1 and 1F4 also includes the region containing the 347 bp IES in the micEFA1B gene. Sequence analysis of the resulting product did show the known 347 bp IES in the micEFA1A gene. Sequence analysis of the resulting product did show the known 347 bp IES in the micEFA1B gene in the same location. If this gene is indeed the precursor of the EFA1 gene, identical coding sequences are expected in both genes. Sequence determination of the portion of the micEFA1B gene known to have deviations from the macronuclear sequence in the micEFA1A gene yielded the same sequence as in the macronuclear gene. This supports the view that micEFA1B is the precursor of EFA1.
The precursor of the EFA2 gene contains a TelIES

PCR reactions with micronuclear DNA as template and three sets of primers derived from the sequence of the macronuclear $\text{EF A2}$ gene (compare Fig. 1 and Table 1) yielded overlapping products, one of which (obtained with the primer pair 2F1 and 2R1) was slightly longer than the corresponding PCR product obtained with a macronuclear $\text{EF A2}$ template. Sequence analysis showed a 79 bp TelIES sequence characterized by sequences reminiscent of the telomeric repeats of $\text{E.crassus}$ macronuclear chromosomes (Fig. 8). It ends in a 5 bp repeat.

Possible reasons for the lack of micEFA1A processing

Analysis of the flanking sequences of the micronuclear $\text{EFA1}$ genes (Fig. 9) did not provide any clue as to why any of them should be processed more or less efficiently. A consensus sequence for chromosome fragmentation has been established for $\text{E.crassus}$ (33). One of the potential fragmentation sequences detected in either the non-coding regions of the potential macronucleus-destined sequences or the neighbouring regions in the micronuclear sequences shows significant sequence similarity with the consensus sequence in each case, while the sequence similarity of the other one is at most moderate. One fragmentation sequence close to each end of the genes would suffice for excision of the macronucleus-destined sequences according to the current model of the excision process (33). Of course, other determinants in the intergenic regions may be necessary, which might be more
distant than the sequences we have analysed adjacent to the two EFA1 genes. Such fragmentation signals would therefore have escaped our attention. Since the micEFA1B gene apparently contains an additional internal sequence, it cannot be ruled out that it could be a prerequisite for the processing of micEFA1B into the macronuclear EFA1 gene. This intervening DNA sequence is most likely >2 kb long. It will therefore be interesting to obtain its entire sequence and look for potential open reading frames.

Such analyses are presently underway.

The mutations encountered in the micEFA1A gene are all C→T transitions. This situation is reminiscent of repeat induced point mutations (RIP) in fungi, in which duplicated genes are mutated in the same way (35–37). In this case the mutations are correlated with methylation of cytosines. There are indications that cytosine methylation can cause elevated rates of deamination, leading to the observed mutation type (38). If this was true, the pseudogene could be expected to be more highly methylated. This might also indicate that duplication has led to its positioning in a special region of the chromosome, most likely outside a gene cluster, since high mutation rates in macronucleus-destined sequences indicate that duplication has led to its positioning in a special sequence. Such fragmentation signals would therefore have been observed in authentic micronuclear precursor sequences and looked for potential open reading frames.

The terminal structure of the apparent novel intervening sequence in the micEFA1B gene suggests a mechanism for its removal, distinct from that of IES, TelIES and Tec elements.

The 12 bp repeats flanking the intervening sequence in the micEFA1B gene apparently contains an additional internal sequence, it cannot be ruled out that it could be a prerequisite for the processing of micEFA1B into the macronuclear EFA1 gene. This intervening DNA sequence is most likely >2 kb long. It will therefore be interesting to obtain its entire sequence and look for potential open reading frames.

Such analyses are presently underway.

The mutations encountered in the micEFA1A gene are all C→T transitions. This situation is reminiscent of repeat induced point mutations (RIP) in fungi, in which duplicated genes are mutated in the same way (35–37). In this case the mutations are correlated with methylation of cytosines. There are indications that cytosine methylation can cause elevated rates of deamination, leading to the observed mutation type (38). If this was true, the pseudogene could be expected to be more highly methylated. This might also indicate that duplication has led to its positioning in a special region of the chromosome, most likely outside a gene cluster, since high mutation rates in macronucleus-destined sequences indicate that duplication has led to its positioning in a special sequence. Such fragmentation signals would therefore have been observed in authentic micronuclear precursor sequences and looked for potential open reading frames.

The terminal structure of the apparent novel intervening sequence in the micEFA1B gene suggests a mechanism for its removal, distinct from that of IES, TelIES and Tec elements.

The 12 bp repeats flanking the intervening sequence in the micEFA1B gene (Fig. 7) suggest that the rearrangement of the coding sequence leading to the coherent macronuclear gene sequence involves a site-specific recombination event. This is reminiscent of the events occurring during formation of several genes in Oxytricha, where macronuclear genes have to be formed from patches of coding sequences in their micronuclear precursors, involving intricate recombination processes between short repeated sequences separating the various macronuclear-destined sequences (reviewed in 4). In the cyanobacterium Anabaena removal of an intervening sequence by site-specific recombination in the course of the activation of genes involved in nitrogen fixation is another example of the transformation of a precursor into a functional gene by site-specific recombination (39). We are presently looking for a potential ring intermediate, which is to be expected if our hypothesis invoking an intrachromosomal element that is removed by a simple recombination step is correct.

ACKNOWLEDGEMENTS

We thank Helga Bestgen for photographic work and Sophie Curtenaz for critical reading of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft.

REFERENCES