Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 3758–3761, May 1991
Genetics

UGA is translated as cysteine in pheromone 3 of *Euplotes octocarinatus*

(stop codon/ciliates/genetic code)

FRANK MEYER*, HELMUT J. SCHMIDT*, EVELYN PLÜMPER*, ANDREJ HASILIK†, GÜNTHER MERSMANN†, HELMUT E. MEYER‡, ÅKE ENGSTRÖM§, AND KLAUS HECKMANN*¶

*Institute of Zoology and †Institute of Physiological Chemistry and Pathobiochemistry, University of Münster, D-4400 Münster, Federal Republic of Germany; ‡Institute of Physiological Chemistry, University of Bochum, D-4630 Bochum 1, Federal Republic of Germany; and §Department of Immunology, University of Uppsala, S-75123 Uppsala, Sweden

Communicated by John R. Preer, Jr., January 31, 1991

ABSTRACT **Pheromone 3 mRNA of the ciliate *Euplotes octocarinatus* contains three in-frame UGA codons that are translated as cysteines. This was revealed from cDNA sequencing and from plasma desorption mass spectrometry of cleaved pheromone 3 in connection with pyridylethylation of the fragments. N-terminal sequence analysis of carboxymethylated protein confirmed this conclusion for the first of the three UGA codons. Besides UGA the common cysteine codons UGU and UGC are also used to encode cysteine. UAA functions as a termination codon. No UAG codon was found. In connection with results reported for other ciliates, this suggests that the role of the classic termination codons had not yet been established when the ciliates started to diverge from other eukaryotes.**

It has generally been assumed that all ciliates deviate from the universal genetic code by translating UAA and UAG as glutamine and using UGA as the sole termination codon. This assumption was based on sequence analyses of genes in four genera: *Paramecium* (1, 2), *Tetrahymena* (3–7), *Stylonychia* (8), and *Oxytricha* (9). Recently this view was challenged by the finding that in *Euplotes crassus* (10) and *Euplotes raikovi* (11), UAA is used as a termination codon, indicating that the Euplotids differ in this respect from other ciliates. Here we report on the cDNA and amino acid sequence of pheromone 3 of *Euplotes octocarinatus*. Pheromone 3 is one of four signal substances secreted by mature cells of *E. octocarinatus* when they are moderately starved (12). It induces competent cells of other mating types to prepare for conjugation. The finding that in *Euplotes* the UGA codon, the sole stop codon in other ciliates, encodes cysteine whereas the UAA triplet, found in other ciliates to encode glutamine, is used as a stop codon suggests that the use of termination codons had not yet been fixed by the time the ciliates diverged from other eukaryotes.‖

MATERIALS AND METHODS

Cells and Culture Conditions. *E. octocarinatus* strain 3(58)-IX was used in this study. The strain is homozygous for the mating type allele mt^3. It was grown in Fernbach flasks in SMB III medium as described (13), using the photosynthetic flagellate *Chlorogonium elongatum* as a food source.

Preparation of RNA. Total RNA was prepared by disruption of 1–3 × 10^7 cells in 8 M urea/4 M LiCl in a Potter–Elvehjem homogenizer, followed by precipitation on ice overnight. RNA was collected by centrifugation, dissolved in 10 mM Mops, pH 7.5/0.5% SDS, and extracted three times with phenol/chloroform/isoamyl alcohol (25:24:1) and once with chloroform/isoamyl alcohol (24:1). Total RNA was then precipitated by addition of 0.1 volume of 4 M LiCl and 2.5 volumes of absolute ethanol.

Poly(A)$^+$ RNA was prepared by affinity chromatography on oligo(dT)-cellulose (Bethesda Research Laboratories) as recommended by the supplier with the exception that Mops was used as the buffer instead of Tris. Poly(A)$^+$ RNA was precipitated by the addition of LiCl and ethanol as described above and redissolved in water. Quantity and quality were determined spectrophotometrically by measuring absorption at 260 and 280 nm (14).

cDNA Synthesis and Cloning. The cDNA library was constructed (15) in the vector λgt10. The cDNA was treated with S1 nuclease and ligated with *Eco*RI linkers prior to its insertion into the *Eco*RI site of the vector. The pheromone 3 gene was identified by plaque hybridization with the synthetic oligodeoxynucleotide 5′-GTRTANGGYTCYTCCCA-3′, corresponding to the N terminus of the secreted pheromone, and was isolated by standard techniques (14).

DNA Sequencing. Eight positively hybridizing plaques were obtained from 10^5 transformants. Five of them were further subcloned for sequencing by the dideoxy chain-termination method. Their nucleotide sequences were determined from double-stranded and single-stranded templates (pUC12, pT7T3, M13mp18, and M13mp19 as sequencing vectors) according to the sequencing strategy outlined in Fig. 1.

Protein Sequencing. This was carried out by automated Edman degradation (Applied Biosystems model 477A pulsed-liquid sequencer with on-line phenylthiohydantoin amino acid analyzer model 120A). Carboxymethylation was performed by incubating the pheromone sequentially for 30 min each with 50 mM dithiothreitol, 150 mM sodium iodoacetate, 75 mM dithiothreitol, and 200 mM sodium iodoacetate at pH 7.5 at room temperature under N_2 atmosphere in the dark.

Peptide Cleavage. The pheromone 3 fragments shown in Table 1 were produced by digestion with *Staphylococcus aureus* V8 protease (fragments 1–7) or endoproteinase Lys-C (fragment 8). In the first case 11 μg (1 nmol) of the pheromone was dissolved in 20 μl of buffer [25 mM ammonium carbonate, pH 7.8/5% (vol/vol) acetonitrile] to which 1 μl of 20 mM dithiothreitol was added. Digestion was started by adding 3 μl of buffer containing 3 μg of V8 protease. After incubation at 25°C for 3.5 hr, digestion was stopped by evaporation in a vacuum centrifuge. In the second case 5.5 μg (0.5 nmol) of the pheromone was dissolved in 20 μl of buffer [1 mM EDTA/25 mM Tris·HCl, pH 8.5/5% (vol/vol) acetonitrile/1 mM dithiothreitol] to which 5 μl of Lys-C solution (0.2 mg/ml) was added. Digestion was at 25°C for 20 hr.

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.

¶To whom reprint requests should be addressed.

‖The sequence reported in this paper has been deposited in the GenBank data base (accession no. M63389).

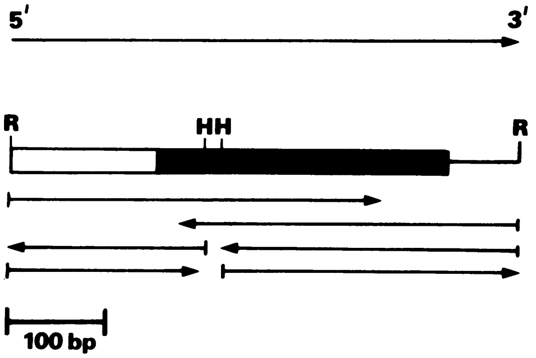

FIG. 1. Restriction map and sequencing strategy for the pheromone 3 gene of *E. octocarinatus*. The arrow above the map shows the direction of transcription; arrows below the map represent the various sequenced portions. The boxed area refers to the coding sequence for the secreted pheromone; the open box indicates the prosequence. H, *Hin*dIII; R, *Eco*RI (generated by added *Eco*RI linkers); bp, base pairs.

Desorption Mass Spectrometry. For plasma desorption mass spectrometry (16) samples were dissolved in 30 μl of 3 mM Tris/glycine (pH 8.5) containing 0.1% dithioerythritol. From this, samples (8 μl) were spotted on foils coated with nitrocellulose. After addition of a small amount of ethanol, the samples were dried with a stream of nitrogen. Mass spectra were recorded on a Biolon 20 instrument (Biolon Nordic, Uppsala) before and after washing samples with 50 μl of 0.1% trifluoroacetic acid. Molecular masses presented are from positively charged ions. They were determined with an accuracy of ≈0.1%. The remaining sample solution was treated with 2-mercaptoethanol and 4-vinylpyridine at pH 8.5. After dilution of the reaction mixture with 0.1% trifluoroacetic acid the sample was incubated on nitrocellulose-coated foil. Prior to analysis it was washed with 0.1% trifluoroacetic acid. For each alkylation the molecular mass increases by 105 Da.

RESULTS

To elucidate the primary structure of pheromone 3 of *E. octocarinatus*, we constructed a cDNA library from poly(A)$^+$ RNA of moderately starved cells of mating type IX that were homozygous for the pheromone 3 gene. The library was screened with an oligonucleotide corresponding to a portion of the N terminus of pheromone 3. Several positive clones were obtained. Five of them were cloned into sequencing vectors and analyzed. The sequencing strategy and a restriction map of the pheromone 3 gene are shown in Fig. 1. The combined sequence of the five inserts is given in Fig. 2. The sequence consists of 515 nucleotides. It contains an open reading frame encoding 147 amino acids that ends with a TAA codon. The reading frame contains the sequence encoding the 99 amino acids of the secreted pheromone, including three translated TGA triplets. The start of the pheromone sequence was determined by comparing the cDNA sequence with the amino acid sequence of the N terminus of pheromone 3. The pheromone is preceded by a prosequence of 48 amino acids including a portion of a putative signal sequence. The C terminus of the pheromone is followed by a 73-base-pair sequence containing three TAA triplets, a putative polyade-

```
  T TTC ATT ATT TTA GCC ATC CTT ATG GTT ACC CAA GCC TTC    40
     F   I   I   L   A   I   L   M   V   T   Q   A   F

AAG ATG ACA TCC AAG CTC AAC ACT AAG CTC CAG TCT CAG ATC CAG    85
 K   M   T   S   K   V   N   T   K   L   Q   S   Q   I   Q

TCA AAG TTC CAG TCC AAA AAC AAA CTT GCT TCT ACT TTT CAG ACT   130
 S   K   F   Q   S   K   N   K   L   A   S   T   F   Q   T

AGT TCA CAG CTA AAG TAT TAT TGT TGG GAA GAA CCA TAT ACA TCC   175
 S   S   Q   L   K   Y   Y   C   W   E   E   P   Y   T   S
                    Tyr Tyr  -  Trp Glu Glu Pro Tyr Thr Ser
                    (1)                                 (10)
TCA ATT ACT GGT TGT TCT ACA AGC TTA GCT TGT TAT GAA GCT TCC   220
 S   I   T   G   C   S   T   S   L   A   C   Y   E   A   S
Ser Ile Thr Gly  -  Ser Thr Ser Leu Ala  -  Tyr Glu Ala Ser
                                                        (25)
GAT TGA AGC GTA ACT GGA AAT GAT CAA GAC AAA TGA AAT AAT GTT   265
 D   *   S   V   T   G   N   D   Q   D   K   *   N   N   V
Asp  -  Ser Val Thr Gly Asn(Asp)Gln(Gly)Lys             (40)

GGG CAA AAT ATG ATT GAC AAG TTT TTT GAA TTG TGG GGC GTT TGC   310
 G   Q   N   M   I   D   K   F   F   E   L   W   G   V   C
                                                        (55)
ATC AAC GAC TAT GAA ACA TGT CTT CAA TAT GTT GAT AGA GCT TGG   355
 I   N   D   Y   E   T   C   L   Q   Y   V   D   R   A   W
                                                        (70)
ATT CAT TAT AGT GAT TCT GAA TTT TGT GGA TGA ACA AAT CCA GAA   400
 I   H   Y   S   D   S   E   F   C   G   *   T   N   P   E
                                                        (85)
CAA GAA AGT GCA TTC AGG GAT GCA ATG GAT TGC TTG CAA TTT TAA   445
 Q   E   S   A   F   R   D   A   M   D   C   L   Q   F stop
                                                    (99)
AAGTTTTACTTCAGACAGACAAGTAAAATCAGAATAACATAATTTCTGAAGTAAAAATT   504

ACACAAAAAAA                                                   515
```

FIG. 2. Nucleotide and deduced amino acid sequence of cloned cDNA for pheromone 3 of *E. octocarinatus*. The cDNA sequence of the secreted pheromone is shown in boldface type. Inverted repeats are indicated by arrows. A putative polyadenylylation signal is boxed. The amino acids deduced from the cDNA are given in the single-letter code; residues corresponding to TGA triplets are marked by an asterisk. The amino acids determined by direct sequencing of the underivatized pheromone are given by their three-letter abbreviations. Residues that did not provide interpretable signals are marked by a dash, and residues that were interpreted with some ambiguity are presented in parentheses. The sequence represents the compiled information from five subclones.

nylylation signal, and inverted repeats. The cDNA shows no consensus sequence for N-glycosylation.

An unexpected feature of the sequence coding for pheromone 3 is the presence of three in-frame TGA triplets. The first of them is located in the portion of the pheromone that has been sequenced by Edman degradation. No signal for a phenylthiohydantoin amino acid derivative was obtained for the TGA-encoded residue. The determined protein sequence coincides with the deduced one except for one position where the sequence determination was ambiguous due to a high background. A portion of the sequence following Met-44 and all five residues following Met-94 were determined after treatment of the pheromone with cyanogen bromide (data not shown). They are in full agreement with the deduced sequences.

Laser desorption mass spectrometry revealed for pheromone 3 a mass of 11,350–11,400 Da, which is in accordance with the mass calculated from the DNA sequence. Gel filtration chromatography and SDS/polyacrylamide gel electrophoresis had previously indicated a molecular mass of ≈20,000 Da (13). The reasons for this discrepancy are unknown.

Since UGA has been shown to function as a stop codon in other ciliates, one might consider the possibility that the gene sequenced by us is a pseudogene that is transcribed but not translated. However, no other sequence than the one shown in Fig. 2 has been found among five independently isolated clones, and hence this notion appears unlikely.

For several prokaryotes and eukaryotes it has been shown that some of their UGA triplets encode selenocysteine (17–22). We have considered this possibility, but we were unable to detect selenium in pheromone 3 by "ICP" analyses (inductively coupled plasma, JY 70 Plus, Jobin Yvon).

To obtain information on the nature of the UGA-encoded amino acid, we digested pheromone 3 proteolytically and measured the masses of fragments before and after treatment with 4-vinylpyridine (Table 1). Pyridylethylation resulted in enhancement of the masses of fragments containing cysteines and residues encoded by UGA. The increments of the masses indicated that both types of residue were modified. This suggested that the UGA codons specified an amino acid with either a sulfhydryl or a selenohydryl group. The measured mass values strongly argue for cysteine and exclude selenocysteine.

For technical reasons pheromone 3 was initially not available in amounts sufficient for carboxymethylation and direct sequencing. However, we have now been able to do this for the N-terminal part of pheromone 3 and identified residues 3, 15, 21, and 27 (the first of the three UGAs codes for residue 27) as carboxymethyl cysteines.

DISCUSSION

Deviations from the standard genetic code have been reported primarily for mitochondria, but also for nuclear genes of several organisms (23). In the prokaryotes *Mycoplasma capricolum, M. pneumoniae, M. genitalium,* and *M. gallisepticum,* UGA is not used for termination but codes for tryptophan instead (24, 25). That UGA is not used for termination in these prokaryotes is attributed to the presence of a tRNA_{UCA} that can translate both UGA and UGG codons. *M. pneumoniae* and *M. genitalium* appear to have only the tRNA_{UCA} gene, while *M. capricolum* and *M. gallisepticum* contain in addition a tRNA_{CCA} recognizing UGG. Similar deviations are known for mitochondria and are explained by assuming a simplification of codon–anticodon pairing in which one tRNA recognizes all four members of a codon family (26). Deviations among eukaryotes are known for the yeast *Candida cylindracea,* where CUG is translated as serine instead of leucine (27), and for the ciliates *Paramecium* (1, 2), *Tetrahymena* (3–7), *Stylonychia* (8), and *Oxytricha* (9) and the algae *Acetabularia cliftonii* and *A. mediterranea* (28), where UAA and UAG encode glutamine. In the case of mammals and some prokaryotes—which otherwise use the standard genetic code—UGA directs the incorporation of selenocysteine in selenoproteins (17–22). Our results show that in the pheromone 3 gene of *E. octocarinatus,* UGA is translated, like UGU and UGC, as cysteine and that UAA functions as a termination codon. UAG codons have not been found. At the moment it is unclear whether the UGA codon is used in *Euplotes* as a third cysteine codon throughout the genome or whether it is used only in some genes, perhaps genes expressed under specialized conditions. In the first case, we would be confronted with a situation comparable to that known for *Paramecium, Tetrahymena,* and other ciliates, where stop codons are used to encode an amino acid apparently throughout the genome, with the difference, however, that UGA in *Euplotes*—the sole stop codon in these ciliates—would now code for an amino acid. In the latter case we would be dealing with a site-specific variation in coding, perhaps comparable to the situation in the genes coding for selenoproteins. Whether the translation of the UGA codon into cysteine in *Euplotes* is due to an extra tRNA^{Cys} or a reduced codon specificity on the side of a common tRNA^{Cys}, or to some other mechanism, must be left for further investigations. However, independent of the possibility that the

Table 1. Measured and calculated masses of pheromone 3 fragments

Fragment		Mass, Da			
		Experimental		Calculated	
No.	Sequence	Pyridyl-ethylated	Untreated	For cysteine	For se cyste
1	YYCWEEPYTSSITGCSTSLACYE	2969 (3)	2656	2658	
2	ASD*SVTGNDQDK*NNVGQNMIDKFFE	3189 (2)	2978	2981	306
3	LWGVCINDYE	1317 (1)	1213	1211	
4	TCLQYVDRAWIHYSDSE	2191 (1)	2086	2087	
5	FCG*TNPE	1080 (2)	n.o.	1081†	112
6	QE	n.o.	n.o.	276	
7	SAFRDAMDCLQF	1509 (1)	1404	1404	
8	FFELWGVCINDYETCLQYVDRAWIHYSD-SEFCG*TNPEQESAFRDAMCLQF	ND	6192	6195	624

The fragments were produced by digestion with *S. aureus* V8 protease (fragments 1–7) or endoproteinase Lys-C (fragment 8) and their m were measured before and after treatment with 4-vinylpyridine. For each pyridylethylation (numbers in parentheses) the molecular increases by 105 Da. Both the measured and the calculated masses refer to the protonated species. n.o., Not observed; ND, not dor UGA-encoded amino acid.

†Mass calculated for pyridylethylated fragment to facilitate a comparison with the determined value.

UGA-encoded cysteines might be restricted to particular genes—perhaps the translation of their UGA codons being used for regulatory purposes—the data are also of interest with respect to the evolution of the genetic code. They support the view that the Euplotids separated early from other ciliate branches (29) and suggest that the rules used by eukaryotes to translate mRNA into polypeptides had not yet been fixed by the time that the ciliates diverged from other eukaryotes, about a billion years ago.

We thank Drs. M. Karas for mass determination of pheromone 3 by laser desorption mass spectrometry, G. Huber for the selenium analyses, and G. H. Beale, K. Müller, and H.-W. Kuhlmann for critical reading of the manuscript. This work was supported by the Deutsche Forschungsgemeinschaft, the Fonds der Chemischen Industrie, the Minister für Wissenschaft und Forschung des Landes Nordrhein-Westfalen, and the Swedish National Board for Technological Development.

1. Caron, F. & Meyer, E. (1985) *Nature (London)* **314,** 185–188.
2. Preer, J. R., Jr., Preer, L. B., Rudman, B. M. & Barnett, A. J. (1985) *Nature (London)* **314,** 188–190.
3. Cupples, C. G. & Pearlman, R. E. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 5160–5164.
4. Hanyu, N., Kuchino, Y. & Nishimura, S. (1986) *EMBO J.* **5,** 1307–1311.
5. Hirono, M., Endoh, H., Okada, N., Numata, O. & Watanabe, Y. (1987) *J. Mol. Biol.* **194,** 181–192.
6. Horowitz, S. & Gorovsky, M. A. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2452–2455.
7. Kuchino, Y., Hanyu, N., Tashiro, F. & Nishimura, S. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 4758–4762.
8. Helftenbein, E. (1985) *Nucleic Acids Res.* **13,** 415–433.
9. Herrick, G., Hunter, D., Williams, K. & Kotter, K. (1987) *Genes Dev.* **1,** 1047–1058.
10. Harper, D. S. & Jahn, C. L. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 3252–3256.
11. Miceli, C., La Terza, A. & Melli, M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 3016–3020.
12. Heckmann, K. & Kuhlmann, H.-W. (1986) *J. Exp. Zool.* **237,** 87–96.
13. Schulze Dieckhoff, H., Freiburg, M. & Heckmann, K. (1987) *Eur. J. Biochem.* **168,** 89–94.
14. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
15. Gubler, U. & Hoffmann, B. J. (1983) *Gene* **25,** 263–269.
16. Sundqvist, B. U. R. & Macfarlane, R. D. (1985) *Mass Spectrom. Rev.* **4,** 421–460.
17. Chambers, I., Frampton, J., Goldfarb, P., Affara, N., McBain, W. & Harrison, P. R. (1986) *EMBO J.* **5,** 1221–1227.
18. Forchhammer, K., Leinfelder, W. & Böck, A. (1989) *Nature (London)* **342,** 453–456.
19. Leinfelder, W., Zehelein, E., Mandrand-Berthelot, M.-A. & Böck, A. (1988) *Nature (London)* **331,** 723–725.
20. Mullenbach, G. T., Tabrizi, A., Irvine, B. D., Bell, G. I. & Hallewell, R. A. (1987) *Nucleic Acids Res.* **15,** 5484.
21. Yoshimura, S., Takekoshi, S., Watanabe, K. & Fujii-Kuriyama, Y. (1988) *Biochem. Biophys. Res. Commun.* **154,** 1024–1028.
22. Zinoni, F., Birkmann, A., Leinfelder, W. & Böck, A. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 3156–3160.
23. Fox, T. D. (1987) *Annu. Rev. Genet.* **21,** 67–91.
24. Inamine, J. M., Ho, K.-C., Loechel, S. & Hu, P.-C. (1990) *J. Bacteriol.* **172,** 504–506.
25. Yamao, F., Muto, A., Kawauchi, Y., Iwami, M., Iwagami, S., Azumi, Y. & Osawa, S. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 2306–2309.
26. Barrell, B. G., Anderson, S., Bankier, A. T., de Bruijn, M. H. L., Chen, E., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. J. H., Staden, R. & Young, I. G. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 3164–3166.
27. Kawaguchi, Y., Honda, H., Taniguchi-Morimura, J. & Iwasaki, S. (1989) *Nature (London)* **341,** 164–166.
28. Schneider, S. U., Leible, M. B. & Yang, X.-P. (1989) *Mol. Gen. Genet.* **218,** 445–452.
29. Lynn, D. H. & Sogin, M. L. (1988) *BioSystems* **21,** 249–254.