Five TGA "stop" codons occur within the translated sequence of the yeast mitochondrial gene for cytochrome c oxidase subunit II

(frame-shift revertants/DNA sequence/genetic code/protein sequence homology)

THOMAS D. FOX
Department of Biochemistry, Biocenter, University of Basel, CH-4056 Basel, Switzerland

Communicated by Adrian M. Srb, September 21, 1979

ABSTRACT A mitochondrial mutation that genetically maps in the middle of the gene coding cytochrome c oxidase subunit II has been found to be a single-base-pair deletion. Three independently isolated spontaneous revertants of this mutant have different single-base-pair insertions within 15 nucleotides of the mutation. These findings clearly identify the location of the gene and suggest that the mutation causes a frame-shift. The sequence of about 900 base pairs surrounding the mutation has been determined and found to have several chain termination codons in every possible reading frame. The sequence can, however, be translated in one frame by assuming that the codon TGA does not cause chain terminations in yeast mitochondria, as was recently suggested for the human organelle [Barrell, B. G., Bankier, A. T. & Drouin, J. (1979) Nature (London), in press]. If TGA codes for tryptophan residues, as is apparently the case in human mitochondria, a polypeptide can be read from the yeast mtDNA that is identical to bovine cytochrome oxidase subunit II at 37.8% of its residues. Furthermore, the DNA sequences of the frame-shift revertants discussed above predict relative isolectric point differences between the wild-type and various revertant forms of the polypeptide. The detection of these isolectric point differences by two-dimensional electrophoresis of subunit II from the various strains independently confirms the presumed reading frame of the gene. It is concluded that TGA is translated in yeast mitochondria, most probably as tryptophan.

Comparisons between the sequences of genes and their protein products have confirmed the genetic code in all organisms studied so far. Only recently however, have comparisons been made between mitochondrial genes and their highly hydrophobic polypeptide products. In one case, the protein sequence of the yeast (Saccharomyces cerevisiae) mitochondrial ATPase subunit 9 (M13-249) (1) could be compared to the gene sequence (2, 3) to establish 27 of the 64 standard codon assignments for mitochondrion. The determination by Steffens and Buse (4) of the amino acid sequence of cytochrome c oxidase subunit II from beef heart (M, 26,000) has now made it possible to compare longer mtDNA sequences with a gene product. Barrell et al. (5) compared this amino acid sequence with human mtDNA sequences and found that a portion of the DNA would code for a polypeptide highly similar (72.7% sequence identity) to the bovine protein, except that three termination codons (TGA) occur within the sequence. Interestingly, all three TGA codons appeared at positions corresponding to Trp (standard codon TCG) in the protein, suggesting that in mammalian mitochondria TGA codes for Trp.

The region of the yeast mitochondrial chromosome coding for cytochrome oxidase subunit II has recently been identified and isolated (6). This work has made it possible to determine the sequence of the yeast subunit II gene (oxi-1 locus: ref. 7) in mtDNA isolated from wild-type (ρ+) yeast as well as mutant and revertant strains. The present paper demonstrates that the yeast gene for subunit II can also be translated to yield a polypeptide of striking similarity to the bovine protein (37.8% sequence identity), but only if in-frame TGA codons are translated. In parallel to the findings with human mtDNA (5), the positions of four of the five TGA codons in the yeast DNA sequence correspond to Trp residues in the bovine protein. Furthermore, charge changes in the polypeptide, predicted from the DNA sequences of three-frame-shift revertant strains, were detected by isoelectric focusing of subunit II, confirming that this sequence is translated in rito.

MATERIALS AND METHODS

Yeast Strains and Isolation of mtDNA. The wild-type Saccharomyces cerevisiae strain was D273-10B (American Type Culture Collection 25657). The MnCl2-induced oxi-1 mutant M13-249 (7, 8) had been used in a previous study (6). The spontaneous revertants RM215, RM216, and RM220 (7) were obtained from M. Solioz. mtDNA was isolated from stationary-phase cells as described (6).

Restriction Enzyme Digestion and Isolation of DNA Fragments. These procedures were carried out as described (6).

3'-End Labeling of Restriction Fragments. This procedure has been described (6). Restriction fragments were incubated with the large fragment of proteolytically treated DNA polymerase I (Boehringer Mannheim) in the presence of one of the [α-32P]dNTPs [specific activity 350 Ci/mmol (1 Ci = 3.7 × 1010 becquerels), Amersham] depending on the restriction fragment to be labeled. HinII ends were labeled with [α-32P]dATP, Taq I ends with [α-32P]dCTP, and Mbo I ends with [α-32P]dGTP, all at 6°C for 2 hr. Pst I ends were labeled with [α-32P]dGTP at 37°C for 2 hr.

DNA Nucleotide Sequence Analysis. Restriction fragments labeled at a single 3' end, or single-stranded restriction fragments, were subjected to partial chemical degradation as described by Maxam and Gilbert (9). Electrophoretic analysis of the products was carried out on gels (50 cm long, 0.5 mm thick) containing either 20% (9) or 8% (10) polyacrylamide. The gels were exposed to Kodak XR-5 film with Kyokko HS intensifying screens at −70°C.

Labeling of Mitochondrial Proteins and Two-Dimensional Electrophoresis. Yeast cells were labeled with 35SNa2S2O4 in the presence of cycloheximide as described (11). Mitochondria were isolated (12) from the labeled cells and dissolved in sample buffer for isoelectric focusing ("lysis buffer," ref. 13). Isolelectric focusing was carried out as described (14) except that the samples were applied at the basic (cathode) end of the gel. The second dimension sodium dodecyl sulfate (NaDodSO4) gels contained 15% polyacrylamide.

The publication costs of this paper 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.

The abbreviation NaDodSO4, sodium dodecyl sulfate.

6534
RESULTS

DNA Nucleotide Sequence Alterations in an oxi-1 Mutant and Its Revertants. The mitochondrial gene coding cytochrome c oxidase subunit II has recently been localized to a 2400-base-pair Hpa II restriction fragment of mtDNA (6). Although the ends of the gene could not be defined in that study, the position of an oxi-1 mutant, M13-249, that maps genetically in the center of the locus (ref. 6; B. Weiss-Brummer and R. Schweyen, personal communication) was found to lie near the middle of this DNA fragment. Therefore, the region presumed to contain the site of this mutation was examined as a first step towards the DNA sequence analysis of the oxi-1 locus as a whole.

The 2400-base-pair Hpa II fragment was isolated from wild-type and mutant mtDNA and then further digested with the enzyme HinII. HinII makes several cleavages in this fragment, one of which is in the immediate vicinity of the mutation (6). The resulting fragments from each strain were labeled at their 3' ends. Single-stranded molecules labeled at the desired HinII site (indicated \( \bar{H} \) in Fig. 1) were then isolated and subjected to chemical degradation reactions for sequence determination (9). Electrophoretic analysis of the products (Fig. 2) from the wild type and mutant revealed that a G residue present in the wild-type sequence was deleted in the mutant. (This G defines position 0 of the sequence reported here.) This single-base-pair deletion suggested the possibility that the genetic lesion is caused by a frame-shift in the middle of the gene. Sequence analysis of a spontaneous revertant, RM220 (Fig. 2), supported this notion because the revertant still lacked the G deleted in the mutant but had an extra T inserted in the sequence at a position between nucleotides +11 and 15. In addition, two other independently isolated spontaneous revertants were also found to lack the G at position 0 and have single-base-pair insertions. RM215 had an A inserted between nucleotides +9 and 12, and RM216 had a T inserted between nucleotides +4 and 6 (not shown).

Nucleotide Sequence of the Region Surrounding the Mutation. Because the mutation in M13-249 maps genetically in the middle of the oxi-1 locus, the sequence of wild-type mtDNA was determined for roughly 450 base pairs on either side of the mutant site. The sequence determination was carried out as outlined in Fig. 1. The sequence can be checked insofar as it completely agrees with previous restriction mapping (6) as well as more recent mapping of Hph I, Pvu II, and Mbo II sites in this laboratory (unpublished). In addition, different portions of the sequence, covering virtually the entire region determined for wild type, were also established by using DNA from the mutant or one or another of the revertants described above. The sequences obtained from the various strains were identical with the exception of those differences discussed above, and one other change (see below) which appears to be a "silent" mutation in strain M13-249.

Upon initial inspection, the most striking feature of this sequence is that it contains termination codons in all reading frames in both directions.

Translation of TGA Codons. Barrell et al. (5) have recently found that the codon TGA, which normally specifies termination, apparently codes for Trp in human mitochondria. If one makes the assumption that TGA codes for Trp in yeast mitochondria, then the oxi-I locus sequence reported here can be translated in one reading frame only (Fig. 3), to yield a protein of molecular weight 28,500 (or 27,000, depending on which initiation codon is used), roughly the size of subunit II (7, 15). Moreover, the predicted yeast amino acid sequence is clearly homologous with the beef heart polypeptide (37.8% identity). In support of the notion that TGA is read as Trp, four of the five TGA codons in the yeast sequence match four of the five Trp residues in the beef heart protein sequence (Fig. 3, boxes). No standard Trp codons (TGG) occur in this frame in the yeast gene. TAA codons, which have been shown to code chain termination in yeast mitochondria (2, 3), occur at the end of the translatable sequence, and seven codons before the first possible initiator triplet. The oxi-1 locus does not appear to contain intervening sequences.

The predicted protein is hydrophobic (polarity of 37.5%, ref. 16) and has an excess of acidic residues over basic. This latter property is consistent with the observed migration of subunit II on two-dimensional gels (18), which indicates an isoelectric point of roughly 5 (17). There are some discrepancies however, between the amino acid composition of the protein predicted here and a published composition for subunit II (15).

Examination of Frame-Shift Revertant Polypeptides Confirms the Reading Frame. Independent confirmation that
The amino acid sequence shown in Fig. 3 was obtained from the nucleotide sequence of the wild-type subunit II. The predicted amino acid sequence is read in the direction and frame indicated. The TGA codon for stop was found in the DNA sequence of the mutant and its revertants. It changes the G at position +333 to a T, producing a TAA stop codon two amino acids before the normal termination. Because the penultimate residue is a Glu, this mutation should add a +1 charge change to all the revertant proteins relative to wild type.

The two-dimensional gel system combining isolectric focusing and NaDodSO4 gel electrophoresis was ideal for application here for two reasons. First, subunit II forms an easily identifiable spot on such gels after electrophoresis of mitochondrial proteins (17). Second, the variant forms of subunit II produced by the revertants all migrate more rapidly than wild-type subunit II during NaDodSO4 gel electrophoresis and differentially from each other (7); the order of increasing electrophoretic mobility is RM216 < RM220 < RM215. Although the \(\alpha\)-\( I\) locus DNA sequence is indeed read in the direction and frame indicated in Fig. 3 would greatly strengthen the conclusion that TGA codons are being translated. The frame-shift revertants examined above provide a convenient test here. The short "garbled" amino acid sequences between the single-base-pair deletion and the three different single-base-pair insertions can be predicted and compared to wild type (Table 1). As indicated in Table 1, all three revertant forms of subunit II should be less acidic than the wild type and differ from each other by a single charge, if the reading frame of Fig. 3 is correct. Completely different predictions of relative charge are obtained if such comparisons are based on any other reading frame. Thus isolectric focusing of subunit II from wild-type and the revertants should provide independent evidence for the correct reading frame. (A second silent mutation was found in the DNA sequence of the mutant and its revertants. It changes the G at position +333 to a T, producing a TAA stop codon two amino acids before the normal termination. Because the penultimate residue is a Glu, this mutation should add a +1 charge change to all the revertant proteins relative to wild type.)

![DNA sequence of the gene for yeast cytochrome c oxidase subunit II](image-url)

**Fig. 3.** DNA sequence of the gene for yeast cytochrome c oxidase subunit II, showing the predicted amino acid sequence of the yeast polypeptide and homology with beef heart subunit II. The nucleotide sequence of the presumed coding strand is written in a reading frame chosen as discussed in the text. The predicted yeast amino acid sequence is written above the nucleotide sequence. The amino acid sequence of beef heart cytochrome c oxidase subunit II (taken from ref. 4) is written above the predicted yeast amino acid sequence. Identities between the predicted yeast protein and the beef heart protein are indicated by horizontal lines above the sequences (three "deletions" were introduced, one in the yeast sequence and two in the beef heart, to reveal the homology). In-frame TAA codons have been boxed and translated as Trp. In-frame TAA codons before and after the coding sequence have also been boxed.

<table>
<thead>
<tr>
<th>Table 1. Predicted amino acid sequence alterations in the frame-shift revertant forms of subunit II, assuming the reading frame of Fig. 3</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Strain</strong></td>
</tr>
<tr>
<td>Wild type</td>
</tr>
<tr>
<td>RM216</td>
</tr>
<tr>
<td>RM220</td>
</tr>
<tr>
<td>RM215</td>
</tr>
</tbody>
</table>
the cause of this anomalous NaDodSO₄ electrophoretic mobility is not known, it provides a convenient means of identifying the different revertant forms of subunit II on the basis of their mobilities in the second dimension.

To examine the isoelectric point of subunit II, mitochondria were isolated from cells of wild type and each of the revertant strains that had been radioactively labeled in the presence of cycloheximide, and mitochondrial proteins were subjected to two-dimensional electrophoresis. Fig. 4 displays autoradiograms of the region of the gels that contained subunit II (the acidic pole is to the right). When electrophoresed separately, subunits II from each of the strains produced single spots (Fig. 4 A–D). When mitochondria from wild type and RM216 were mixed and electrophoresed on the same gel, two spots corresponding to subunit II were observed (Fig. 4E). The revertant protein, identifiable by its more rapid migration in the second dimension, focused at a less acidic position than the wild-type protein. Similarly, pairwise combination of wild type plus RM220 (Fig. 4F) and wild type plus RM215 (Fig. 4G) showed that the revertant species were less acidic than wild-type subunit II. Finally, mitochondria from all four strains were mixed and electrophoresed (Fig. 4H). The four forms of subunit II, identifiable by their relative mobilities in the second dimension, were observed to focus in the following order of decreasing acidity: wild type > RM216 > RM220 > RM215.

These results are in excellent agreement with the predicted charge differences for the revertant polypeptides (Table I) and thus provide independent evidence that the oxi-I locus is translated as shown in Fig. 3.

**DISCUSSION**

The codon TGA normally codes for polypeptide chain termination in prokaryotes (18, 19) and the nuclear genes of eukaryotes (20, 21), including yeast (22, 23). This paper demonstrates that the yeast mitochondrial gene for subunit II of cytochrome c oxidase contains five TGA codons in the reading frame that would code for a polypeptide whose sequence clearly shows homology with the corresponding bovine protein. Furthermore, an examination of the polypeptide products of frame-shift revertants in this gene independently supports the idea that the reading frame containing these TGA codons is indeed translated in two. Four of the five TGA codons correspond in position to four of the five Trp residues in the bovine protein, and there are no standard Trp codons (TGG) in this reading frame, suggesting that TGA is translated as Trp in yeast mitochondria. These findings support the previous observation of Barrell et al. (5) that human mtDNA contains TGA codons at positions corresponding to several Trp residues in the bovine protein. Proof of this novel codon assignment will, however, require the comparison of a protein and its corresponding DNA (or RNA) from the same organism.

A direct comparison of yeast mitochondrial protein and DNA sequence has already been made for the ATPase subunit 9 (1–3), a short polypeptide (M, 8000) that does not contain Trp. These studies rigorously established 27 standard codon assignments for yeast mitochondria. In the gene sequence reported here, 13 additional codons appear whose assignments have not yet been rigorously established for mitochondria (marked with asterisks in Fig. 5). (To this list must be added the Leu codons CTT and CTG, which do not occur in the wild-type gene but are read in the frame-shift revertants.)

<table>
<thead>
<tr>
<th></th>
<th>U</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>Phe</td>
<td>8</td>
<td>4</td>
<td>Tyr</td>
</tr>
<tr>
<td>U</td>
<td>Leu</td>
<td>27</td>
<td>2*</td>
<td>Cys</td>
</tr>
<tr>
<td>C</td>
<td>Leu</td>
<td>0</td>
<td>0</td>
<td>His</td>
</tr>
<tr>
<td>C</td>
<td>Leu</td>
<td>0</td>
<td>0</td>
<td>Arg</td>
</tr>
<tr>
<td>A</td>
<td>Ile</td>
<td>24</td>
<td>2*</td>
<td>Ser</td>
</tr>
<tr>
<td>A</td>
<td>Thr</td>
<td>6*</td>
<td>0</td>
<td>Gly</td>
</tr>
<tr>
<td>G</td>
<td>Val</td>
<td>11</td>
<td>7</td>
<td>Asp</td>
</tr>
<tr>
<td>G</td>
<td>Ala</td>
<td>0</td>
<td>9</td>
<td>Glu</td>
</tr>
</tbody>
</table>

**FIG. 5.** Apparent codon usage in the mitochondrial gene for yeast cytochrome oxidase subunit II. The codon UGA has been assigned as Trp. This and other codon assignments that have not been rigorously established for mitochondria (see Discussion) are marked with an asterisk.
Although the codon TGA (UGA) causes chain termination in both Escherichia coli and Salmonella typhimurium, wild-type strains of both these bacteria carry a low-level background UGA-suppressor activity. This background suppressor causes limited read-through of some bacteriophage genes, both in etco (24) and in vitro (25), and results in the "leaky" phenotype of some bacterial UGA mutations (26). However, this activity would be too weak to yield an appreciable amount of product from an mRNA with five UGA codons in series. Indeed, translation of yeast mitochondrial RNA in an E. coli system in vitro yields short-fragment polypeptides but no detectable full-length products (27). It is interesting to note in this connection that at least some chloroplast mRNAs can be translated in heterologous systems in vitro (28–30), indicating that nonsense codons are probably not present in the genes of these organelles. An efficient E. coli UGA suppressor has been shown to be the result of an alteration in the tRNA<sub>trp</sub>, not affecting the anticodon, that facilitates the reading of both UGG and UGA codons by this tRNA<sub>trp</sub> (31). Such nonstandard wobble may play a role in the normal translation of UGA codons in the mitochondrial system.

Several nucleotides 15 to 19 bases upstream from the first possible initiator codon (positions -429 to -434) could form a four-base-pair hybrid in either of two ways with the Shine and Dalgarno sequence (32) at the 3′ terminus of E. coli 16S ribosomal RNA. However, the significance of this observation is clearly open to question. No such sequence was found preceding the mitochondrial gene for ATPase subunit 9 (3). As in the case of the ATPase gene (2, 3), the gene for cytochrome oxidase subunit II is terminated by a TAA codon and immediately followed by a long stretch (more than 100 base pairs) of DNA containing few G and C residues. The existence of such (A+T)-rich spacers had been previously predicted (33).

The observation that TGA is translated, probably as Trp, in the mitochondria of both humans (5) and yeast suggests that the modern organelles could have inherited a slightly altered genetic code from a common ancestor. Following this line of speculation, it seems possible that this difference in the code may have played a role in preserving the mitochondrial genetic system, because it would tend to prevent transfer of mitochondrial genes to the nucleus (see ref. 34) by blocking their expression at the level of translation in the cytoplasm. The finding that the gene coding for the ATPase subunit 9 is located in the mitochondria in yeast, whereas the homologous protein in Neurospora crassa is coded by a nuclear gene (1), strongly suggests that movement of genes between the mitochondria and the nucleus has occurred. In this case however, the protein does not contain Trp.

On the basis of similarities between the bovine subunit II and known copper-binding proteins, Steffen and Buse (4) have suggested that cytochrome oxidase subunit II is a copper-binding protein. The sequence reported here supports this conclusion.

Note Added in Proof. The sequence of otx-1 locus DNA from a <sup>−</sup> strain has been independently determined (35, 36).

I thank B. G. Barrell and colleagues for communicating their results prior to publication. I also thank J. E. Walker, K. Ineichen, P. Philippeps, and T. Bickle for helpful discussions, T. Catlin for excellent technical assistance, and T. Mason for a critical reading of the manuscript. This work was supported by a grant from the Swiss National Science Foundation (3.172-1.77).