Primary structure of the *Saccharomyces cerevisiae* gene for methionyl-tRNA synthetase

(DNA sequence analysis/protein sequence comparison)

PHILIPPE WALTER, JEAN GANGLOFF, JACQUES BONNET, YVES BOULANGER, JEAN-PIERRE EBEL, AND FRANCO FASIOLO

Institut de Biologie Moléculaire et Cellulaire du Centre National de la Recherche Scientifique—Laboratoire de Biochimie, 15, rue René Descartes 67084 Strasbourg Cedex, France

Communicated by Boris Magasanik, December 30, 1982

ABSTRACT The sequence of a 5-kilobase DNA insert containing the structural gene for yeast cytoplasmic methionyl-tRNA synthetase has been determined. A unique open reading frame of 2,233 nucleotides encoding a polypeptide chain of 751 amino acids (M₀, 85,500) has been characterized. The data obtained on the purified enzyme (subunit size, amino acid composition, and COOH-terminal sequence) are consistent with the gene structure. The protein sequence deduced from the nucleotide sequence reveals no obvious internal repeats. This protein sequence shows a high degree of homology with that of *Escherichia coli* methionyl-tRNA synthetase within a region that forms the putative methionyl adenylate binding site. This strongly suggests that both proteins derive from a common ancestor.

Aminoacyl-tRNA synthetases play a crucial role in protein synthesis because they catalyze the specific attachment of amino acids to their cognate tRNAs. Knowledge of their primary structure is one of the prerequisites for the complete understanding of the structure–function relationship. So far, the sequence of only one aminoacyl-tRNA synthetase has been completely determined by using classical protein sequence analysis techniques (1). However, the cloning of a number of aminoacyl-tRNA synthetases genes has allowed derivation of the primary structure of the corresponding enzyme from the DNA sequence—i.e., the alanyl- (2), tryptophanyl- (3), glutaminyl- (4), and methionyl- (5) tRNA synthetases from *Escherichia coli*. In our laboratory, a yeast mutant strain lacking functional cytoplasmic methionyl-tRNA synthetase was available. It was complemented with a plasmid containing random fragments of wild-type yeast genomic DNA obtained by a partial ScaI digestion (6). It was thus possible to isolate a 5.1-kilobase piece of DNA containing the methionyl-tRNA synthetase gene (*MES1*). The isolated gene product is a monomer (M₀, 80,000). However, in crude extracts from a wild-type strain the enzyme behaves as a dimer (M₀, 2 × 80,000). Because enzyme purification always led to a monomeric species with no detectable variation of M₀, the existence of a dimeric structure for native methionyl-tRNA synthetase is by no means proven (unpublished data).

In this paper we report the complete nucleotide sequence of the DNA insert containing the structural gene. Indeed, a unique open reading frame whose length corresponds exactly to that expected for the above protein size could be characterized.

MATERIALS AND METHODS

DNA Sequence Analysis. DNA sequence analysis was carried out by the chain termination method of Sanger et al. (7).

Fragments of the cloned DNA were subcloned into the bacteriophage M13mp7 (8) and analyzed by using the primer synthesized by R. Crea and purchased from P.L. Biochemicals. Enzymes and Chemicals. Most restriction enzymes and T4 DNA ligase were obtained from New England BioLabs. The Klenow *E. coli* polymerase and nuclease S1 were from Boehringer Mannheim. All chemicals were from Merck (Darmstadt, Federal Republic of Germany) and were analytical grade.

RESULTS AND DISCUSSION

Sequence of the *MES1* Gene. Fig. 1 shows the complete nucleotide sequence of a 3,215-base-pair-long DNA fragment bearing the structural gene for the yeast methionyl-tRNA synthetase in which a unique open reading frame was identified from nucleotide −33 to nucleotide 2,253 after which the next in-phase stop codon occurs. Nucleotide 1 is A of the first ATG encountered in this frame. As suggested for eukaryotic systems (9), we assume that this is the initiating codon because it is the first ATG after the start of transcription determined by nuclease S1 mapping which showed three major 5′-mRNA ends centered at positions −10, −18, and −17 (unpublished data). Therefore, this DNA section is large enough to encode a polypeptide chain of 751 amino acids (including the initiating methionine). The corresponding peptide sequence is also shown in Fig. 1. It yields a M₀ of 85,500, in good agreement with that measured for the purified protein.

Unfortunately, NH₂-terminal sequence analyses of the enzyme purified from transformed yeast cells did not give any clear answer. Indeed, the negative results of dansylation, Edman–Chang degradation, and aminopeptidase digestion suggest that the NH₂-terminal residue is blocked. But three lines of evidence establish that the stretch of DNA shown in Fig. 1 does contain the *MES1* gene:

(i) The amino acid composition deduced from the DNA sequence fits well with that calculated from amino acid analysis of the protein purified from the transformed cells (Table 1).

(ii) The COOH terminus predicted from this DNA sequence is Gln-Gln-Val-stop which is in good agreement with that established by both hydrazinolysis and carboxypeptidase A and Y digestions of the protein: (Ser or amide)-(Ser or amide)-Val(OH).

(iii) There is a remarkable homology (see Fig. 2) between our protein sequence and that determined for the *E. coli* crystalized tryptic fragment of methionyl-tRNA synthetase (5) in the region corresponding to the nucleotide binding domain.

Because of the lack of an unambiguous NH₂-terminal sequence as well as small oligopeptides distributed along the DNA sequence, special care was taken to establish the latter. Indeed, 100% of it was read off both strands, thus increasing our con-

-321| AGCTTCCCG AGAACCGAG TCCGATGATA ATGCTGATC CTCATGACG CTGCGCTTGC TAGATTATC TTTGATTGCC GCGAAAGCGA TCAATTATGC

-220| ATGCGCGGAG TAGATTAC TCTTCTCAAT TCTTCTCAAT

-110| AGACATGAC AAATTTTCTT TCTTTCTCAAT

+1| ATG TCT TCC ATT TCC TTT GAT AAA TCA AAG AAA CAT CCT GCC CAT TGG CAG TTA GGC AAT ATT TGG AAG ATT GCC CTA GCA CTT GAA

Met Ser Phe Leu Ile Ser Phe Asp Lys Ser Lys His Pro Ala His Leu Gln Leu Ala Leu Asn Leu Ala Leu Ala Leu Glu

30| TAT GCA AGC AAA AAT TTA AAG CCC GAA GTT GAC AAT GAT AAT GCT GCC ATG GAA TGG CGC AAT CAA AAG GAA CCT TTC TTT GAT Tyr Ala Ser Lys Leu Val Asp Asn Ala Ala Met Glu Leu Arg Thr Lys Glu Pro Leu Phe Asp

60| GCT AAC GCT ATT CTA AGA TAT GTC ATG GAT GAT TTT GAA GGT AAC ACT TCC GAC AAG ATT CAA TTC GGA TAT CCA CAA GCG TTA CAA GAG Ala Asn Leu Arg Tyr Val Met Asp Asp Phe Glu Gly Gln Thr Ser Asp Lys Tyr Gln Phe Ala Leu Ala Ser Leu Glu Asn Leu

90| TTA TAT CAT AAA GAA TTA CCT CAA CAG CAT GTC GAG GTG TTT GCA CCT CAA GAC TAT CCA TGG TCT TCA CAA AAC CTG Tyr Leu His Lys Gln Leu Tyr Val Asp Asn Ala Val Tyr Val Asp Asn Ser Ser Leu Val His Lys Ser Asp Pro Glu Pro Ser Leu Val Ser

120| ACT GCT ACA GAT TGG ATC CTG TTT GCT AAC GTT TAT CTA ATT TCT TCT TTA ATT CCC GAA ATT TGG CCA TGG CCA TCC AAA GTG Thr Ala Asp Leu Ile Leu Phe Ala Asn Val Tyr Ala Leu Asn Ser Ser Leu Val His Lys Ser Asp Pro Glu Pro Ser Leu Val Ser

150| CAT AAC GCT GTA GCA TGG GCT AAA AAG CAT GTT CCA GGT GAT TCT TCT TCT AAA AAC ATC GCC GCA GTG AAA ATC CAA GCT GGA His Asn Ala Val Ala Leu Lys Ala Ala Val Arg Ser Ser Phe Thr Val Asp Ile Leu Gly Asn Ala Val Lys Gly Asp Leu

180| ACA GTT AAC CCA AAG GAT TCA GAA ATT TGG CCT AAG CCA AAC GAA AAG AAC ATC GTG ATC ACT TCG GCA TTA CCT TAT GAC AAT GGT Thr Val Lys Pro Lys Asp Ser Glu Ile Leu Pro Lys Pro Asn Glu Arg Asn Ile Leu Thr Ser Ala Pro Tyr Val Asn Val

210| CCA CAC TGG GAT ATT ATC ATC GGT GTT CT TCA GCA GAC ATT TTT GCT GCT TAC TCT GTC ATT ATT GAG GAT TTT AAA TAT TAT AAT GCC TGT ATT ACC TCT GAT CTG AAA TAT TTT ATT GCA CTT TAA TAG CTT GCA GAT TTT GAT GAA CAA TCT GAC AAC ACT CTT GAT GAG

240| TGC GGT ACT GAT GAA TAT GGT ACT GCC ACC GAG AAA ACT AAA GCT TGG GAG GAA GTG ACA CCA AAG AAA CAA CTA TGG GAC AAA AT TAC CAC AAA Cys Gly Thr Asp Glu Tyr Gly Thr Ala Thr Glu Lys Ala Leu Leu Glu Gly Val Thr Pro Arg Glu Leu Cys Leu Asp Lys Tyr His Lys

270| ATC CAC AGT GAC GTT TAC AAG TGG TTC CAA ATT GGA TTT GAT TAT TTC GGT AGA ACT ACG ACG GAT AAG CAA ACG GAG ATT GCT CAA CAT Ile His Ser Asp Val Tyr Lys Trp Phe Glu Ile Gly Phe Asp Tyr Phe Gly Arg Thr Thr Asp Lys Glu Thr Glu Ile Ala Glu His

300| ATT TTT ACA AAG CTT AAT TGC AAT TTA CCT CTA ATT CAA ATT AGG CTA TAG CTG AAA CAA GTT GAC TAA CCA CTA TGG AAC TAC CAT CAT GAT GCT GAT Ile Phe Thr Lys Leu Asn Ser Asp Asp Glu Tyr Leu Ser Lys His Pro Ala His Asn Ser Tyr Leu Ala Asp

330| CGT TAT GGA GTG GAA TGT CCA AAA TGT CAC TAC GAT GCT GTG CTT GAG CAA GTT GAC AAA TGT GGT GCC CTG TTA GAT CCA TTT Arg Tyr Val Glu Gly Gln His Arg Asp Ala Ser Glu Lys Cys His Tyr Asp Ser Ala Arg Asp Glu Cys Asp Lys Gly Ala Leu Leu Asp Pro Phe

360| GAA TTG ATC AAT CCA GTG TGT AAA TTA TAT GCT GCT TCA GCA CAA AAA TAT CAT TGG CAT ATT CAT CAT TGC GCT GAT Ile Phe Thr Lys Asp Leu Asp Ser Ala Ser Pro Glu Pro Tyr Asp Asn Ser Phe Leu Phe Leu Ser Asp Leu Asp Leu Glu

390| AGC CAA ATT TCT GAA TGG GAT GAA AAG GCC TCT GAA GAA GTG ACC TGC TCA AAA AAT TCA AAA ACA ATT ACC CAA TCG TGG AAT GAG Ser Gin Ile Ser Glu Thr Glu Glu Thr Glu Glu Thr Glu Ser Lys Ser Lys Thr Ile Thr Gin Ser Trp Leu Lys Asp

420| GGT TTG AAG CCA CGT GTT ATT ACA AGA GTT TTA GTG TGT GCA GTG CCA TCA AAA TAT TGG AAA AAT TAA AAG AAA CAA TGG TGG AAT ATT TGG AAG GGT GAG TCA TGG GAT GTC TGG GCT TCA CAA GAC ACT GAT TGG Gly Leu Lys Pro Arg Cys Ile Thr Arg Asp Leu Val Trp Gly Thr Pro Leu Lys Tyr Leu Lys Val Leu Tyr Val

450| TTT GAC GCT ACA ATC GCC TAC TTT TCC ATC ACT TCT ATT TAC ACC AAA GAA TGG AAA CAA TGG TGG AAT ATT CAA GAG CAT GTT TCA TTG Phe Asp Ala Thr Ile Gly Tyr Val Ser Ile Ser Asn Tyr Thr Lys Glu Trp Lys Glu Pro Leu Ser Asp

480| TAT CAA TTC ATG GAT GAG AAT GTT CCT TCT CCT GAT CCT GCT GTT CAA CTA TGG TGG AA TGG AAA ATT CAA AAG GAT TCC TGG GCT TAT GAT TGG Gly Leu Lys Pro Arg Cys Ile Thr Arg Asp Leu Val Trp Gly Thr Pro Leu Lys Tyr Leu Lys Val Leu Tyr Val

510| TAT CAA TTC ATG GAT GAG AAT GTT CCT TCT CCT GAT CCT GCT GTT CAA CTA TGG TGG AA TGG AAA ATT CAA AAG GAT TCC TGG GCT TAT GAT TGG Gly Leu Lys Pro Arg Cys Ile Thr Arg Asp Leu Val Trp Gly Thr Pro Leu Lys Tyr Leu Lys Val Leu Tyr Val

540| (Fig. 1 continues on next page.)
AAG ACT ATT CTC GTA AGA TAC TAC CTG CTA GCA TCT GTT AGA CCA CCA

TAT ATG CCA GAA ATA GGT GAG AAA ATA AAC AAG ATG TTA AAT GCA CCA GCT TTA AAA ATT GAT AGA TTC CAT TTG GCA ATC GTA

GAA CAT GTA GTA AAG GTA GAA AAG AAA AGT GAA TGG AGA CCC AAA TAT GGT GTG CAA GGA ACT TAA ATG

GTG TAA GGGTATCGG AAGCATTTTT TTCTCTCTTC TTTCTCTT CAAATTCACA TGAAAGAGG CTTGCAATAA TATAGGATTG TTGGTCTGA ATGGAGAA TATAGGGATG TATTCTCTGA TATAAACTG

AGTTAAAA ACAATTCTCC GCCTAACAGT TTATACCT TTCTTTTAT TGATATTTT CTGAGAGGAT ATGAGA ATGTCCTGA ATGAGAAGACCTTGAGAGG

CTGCTTTATA ATACACAAATG TTAACTGTG CTCGCCATA ATATATAG TATTAGGGAA ATAAAAACCTG AAAACAAAA ATACCTCTGA TCGGCTGAA TGGAGACAGC

**Fig. 1.** Sequence of the *MES* gene and adjacent regions. This is the complete sequence of a region encompassing the MES structural gene and 320 nucleotides of flanking DNA in each direction. Only the mRNA identical strand (i.e., plus strand) of the gene is shown. The second base of stop codons is marked by an asterisk, thus indicating the reading frame. Numbers above the sequence are those of nucleotides (position +1 is assigned to A of the first ATG); those below the sequence correspond to amino acid positions (starting from the first putative translated methionine). Dyad symmetries are indicated by lines above the sequence.

**Table 1.** Amino acid compositions of methionyl-tRNA synthetase as calculated from amino acid analyses and deduced from DNA sequence.

<table>
<thead>
<tr>
<th>Amino acid</th>
<th>Composition*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>From analyses</td>
</tr>
<tr>
<td>Aax</td>
<td>89</td>
</tr>
<tr>
<td>Thr</td>
<td>29</td>
</tr>
<tr>
<td>Ser</td>
<td>46</td>
</tr>
<tr>
<td>G1x</td>
<td>87</td>
</tr>
<tr>
<td>Pro</td>
<td>34</td>
</tr>
<tr>
<td>Gly</td>
<td>37</td>
</tr>
<tr>
<td>Ala</td>
<td>45</td>
</tr>
<tr>
<td>Cys+</td>
<td>9</td>
</tr>
<tr>
<td>Val</td>
<td>37</td>
</tr>
<tr>
<td>Met†</td>
<td>9</td>
</tr>
<tr>
<td>Ile</td>
<td>38</td>
</tr>
<tr>
<td>Leu</td>
<td>72</td>
</tr>
<tr>
<td>Tyr</td>
<td>28</td>
</tr>
<tr>
<td>Phe</td>
<td>28</td>
</tr>
<tr>
<td>His</td>
<td>20</td>
</tr>
<tr>
<td>Lys</td>
<td>56</td>
</tr>
<tr>
<td>Trp+</td>
<td>13</td>
</tr>
<tr>
<td>Arg</td>
<td>24</td>
</tr>
<tr>
<td>Total</td>
<td>701</td>
</tr>
</tbody>
</table>

* Numbers correspond to residues per mol on the basis of *M* = 80,000.
† Cysteine and methionine were determined as cysteic acid and methionine sulfone, respectively, after performic acid oxidation of the protein.
‡ Tryptophan was determined after hydrolysis with methanesulphonic acid.
Fig. 2. Sequence homologies between an active tryptic fragment of the E. coli methionyl-tRNA synthetase and the entire yeast cytoplasmic enzyme. In both cases the amino acid sequences were deduced from the respective cloned genes with corroborative protein chemical data available for the bacterial methionyl-tRNA synthetase only. Deletions have been introduced at positions 31, 209–212, 244, and 308–309 in the bacterial enzyme sequence and 461–465 in the yeast enzyme sequence in order to maximize the alignment. Amino acids that are identical in both sequences are boxed; those that are similar are underlined. The secondary structural features of the Rossman fold (β-sheets and α-helices) are also indicated on the E. coli sequence and were kindly provided by J. L. Risler and C. Zelwer. For clarity, the yeast sequence is limited to residues 181–594 and the E. coli sequence, to residues 1–401.

Therefore, although the topology of this region of the active site is probably essential for adenylate formation, we can deduce little about the identity of the key catalytic residues. However, recent studies (16) have emphasized the essential role of cysteine-35 in the binding of ATP by tyrosyl-tRNA synthetase.

Finally, we would like to point out that Putney et al. (2) have proposed a structural model for E. coli alanyl-tRNA synthetase in which the NH2-terminal domain (residues 1–400) contains the catalytic site wherein the alanyl-adenylate is formed whereas the COOH-terminal domain is involved in the transfer step and subunit association. In addition, the enzyme contains clusters of arginine residues and mixed-charge residues distributed along the sequence. Such charged areas could interact with the cognate tRNA. This model does not fit entirely with the yeast methionyl-tRNA synthetase in which the putative nucleotide-binding domain corresponds to the middle of the molecule (between residues 192 and 594). But it is worth noting that mixed-charge clusters can be found in the NH2- and COOH-terminal parts (residues 8–12, 183–186, 603–608, 635–637, 709–713, and 733–745).

Fig. 3. Sequence homologies between yeast and E. coli methionyl-tRNA synthetase and Bacillus stearothermophilus tyrosyl-tRNA synthetase. The homologous cysteine and histidine residues previously detected by Barker and Winter (13) in the two bacterial enzymes are boxed. Vertical lines indicate homologies between the three enzymes (solid = identity; dotted = conservative replacement). The structural signature for tyrosyl-tRNA synthetase is taken from Bhat et al. (14).
Internal Sequence Homology. Extensive sequence repeats were found by tryptic mapping in E. coli methionyl-tRNA synthetase, and similar experiments performed on other aminoacyl-tRNA synthetases from various sources led to the widespread view that those enzymes with subunit sizes ranging from Mr 62,000 to 130,000 could contain such long stretches of repeated sequences in polyptide chain (for a review, see ref. 17). No obvious internal redundancy was found in the predicted sequence of yeast methionyl-tRNA synthetase and the same is true in E. coli alanlyl-, glutaminyl-, and methionyl-tRNA synthetase (2, 4, 5). Therefore, all these results cast serious doubts on the validity of the above hypothesis.

To conclude, we emphasize that transformation to prototrophy of our yeast K₄₉ mutant could be due to the cloning of either cytoplasmic or mitochondrial methionyl-tRNA synthetase genes because both belong to the nuclear genome and both enzymes aminoacylate cytoplasmic tRNA (18). However, the two enzymes can be distinguished on the basis of two criteria (18): (i) their chromatographic behavior on hydroxypatite columns, the mitochondrial enzyme being eluted at 0.1 M phosphate and the cytoplasmic one at 0.2 M; and (ii) the specificity of aminoacylation. Indeed, mitochondrial methionyl-tRNA synthetase aminoacylates both yeast cytoplasmic and E. coli tRNA at similar rates whereas its cytoplasmic counterpart aminoacylates E. coli tRNA at a rate 1/7th that of yeast cytoplasmic tRNA (18). We recently have found that antibodies raised against the purified gene product do not inhibit the mitochondrial methionyl-tRNA synthetase. These differences show that our cloned gene is indeed that of the cytoplasmic enzyme.

The authors thank G. de Marcillac for his help with sequence computer analysis, Dr. M. Yaniv who introduced F. P. to Sanger’s sequence analysis method, and Dr. D. Barker for preprints of his articles before publication. The skillful technical assistance of G. Nussbaum is gratefully acknowledged. This work was supported by grants from the Centre National de la Recherche Scientifique. P. W. was the recipient of a fellowship from the Délégation Générale à la Recherche Scientifique et Technique.