Biochimica et Biophysica Acta, 1153 (1993) 345–347

BBAMEM 70734

Short Sequence-Paper

An *Escherichia coli* gene showing a potential ancestral relationship to the genes for the mitochondrial import site proteins ISP42 and MOM38

Peter Cartwright, Mark Timms, Trevor Lithgow [1], Peter Høj and Nicholas Hoogenraad *

Department of Biochemistry, La Trobe University, Bundoora, Victoria, 3083 (Australia)

(Received 5 August 1993)

Key words: Mitochondrial import site protein; Protein translocation; Sequence comparison; Open reading frame; (*E. coli*)

An ORF (*OrfT*) of 1911 base pairs, upstream of the *hip* operon in *Escherichia coli* at map position 33.82 has been identified. The protein encoded by this sequence is predicted to have a molecular mass of 68 249 Da and the carboxyterminal 276 residues shows 26.8% and 25.4% identity with the import site proteins ISP42 and MOM38 from the mitochondrial outer membrane of *Saccharomyces cerevisiae* and *Neurospora crassa*, respectively. These mitochondrial membrane proteins have been shown to be essential components of the protein translocation apparatus in yeast. These similarities raise the possibility that *OrfT* might represent the bacterial gene from which these eukaryotic genes evolved.

It is now widely accepted that, in the course of evolution of the eukaryotic cell, mitochondria have evolved from endosymbiotic bacteria [1]. The supposed evolution of the mitochondria requires that a system for protein import be established in the outer membrane of the ancestral endosymbiont, since most of the proteins in mitochondria are now nuclear-encoded and synthesized in the cytosol. The mitochondrial protein import machinery is localized in the mitochondrial outer membranes and consists of a number of proteins which include receptor-like components and import site proteins, some of which have now been identified in the fungi, *Saccharomyces cerevisiae* and *Neurospora crassa* [2]. One of the import site proteins (ISP42 in yeast [3]; MOM38 in *Neurospora* [4]) has been shown to be essential for cell viability. A number of genes involved in protein translocation in *E. coli* have also been identified (reviewed in Ref. 5). Surprisingly, to date there is little evidence for homology between components of the bacterial and mitochondrial translocation machinery [2].

The yeast and *Neurospora* outer membrane proteins ISP42 and MOM38 show an overall sequence homology of 53.2%, but in some regions almost 100% identity is observed (Fig. 3). Based on the assumption that

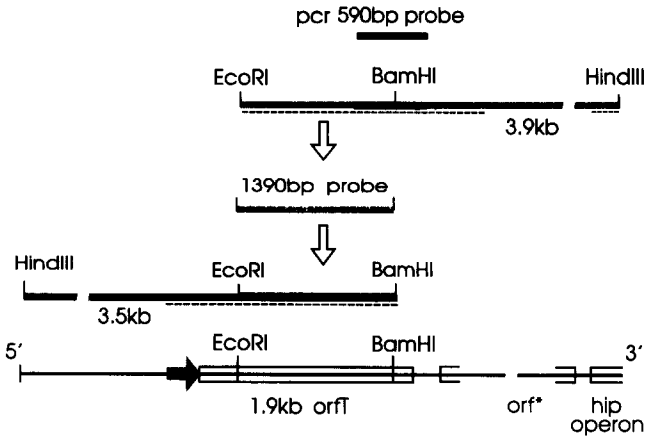

Fig. 1. Relationship of *OrfT* to the *hip* locus of *E. coli*. The overlapping 3.9 kb *Eco*RI-*Hin*dIII and 3.5 kb *Hin*dIII-*Bam*HI fragments contained the 1.9 kb *OrfT*. The *hip* operon is between 3.8 and 3.9 kb downstream of the initiating ATG in *OrfT*, at map position 33.90. A second open reading frame (orf*) is located between *OrfT* and *hip*, but only the 5′ and 3′ ends of this orf has been sequenced. The promoter elements identified for *OrfT* are shown by the solid arrow. The portions of the *E. coli* DNA sequenced is shown by (- - - - -). The 590 bp probe, generated by PCR, used to isolate the 3.9 kb fragment and the 1396 bp *Eco*RI-*Bam*HI fragment used to obtain upstream sequences for *OrfT* are also shown.

* Corresponding author. Fax: +61 3 4792467.
[1] Present address: Biocentre, University of Basel, Switzerland.
The sequence data reported in this paper have been submitted to the EMBL/GenBank Data Libraries under the accession number X73295.

these highly conserved regions are essential for the structure and function of these proteins, oligonucleotide primers were designed corresponding to two such regions (ISP42 residues 126–134 and 300–309) and these were used in a PCR, with *E. coli* chromosomal DNA template, to produce a 590 bp probe. This probe was subsequently used to screen an *E. coli* genomic library and a fragment of approx. 3900 bp was isolated, bounded by unique *Eco*RI and *Hin*dIII sites. A 1390 bp *Eco*RI-*Bam*HI subfragment from the 5′ end of the 3.9 kb clone was used to obtain an overlapping fragment of approx. 3.5 kb with *Hin*dIII-*Bam*HI ends (Fig. 1). Nucleotide sequence was obtained for the overlapping fragments from approx. 600 bp upstream of the *Eco*RI site to approx. 2.2 kb downstream of the *Eco*RI site (Fig. 1). The sequence contained an ORF of 1911 base pairs (*OrfT*) (Fig. 2) and the start of a second ORF approx. 44 bp downstream from *OrfT*. Using DNase to prepare deletion fragments the nucleotide sequence at the 3′ end of the 3.9 kb clone was determined and revealed that the fragment extends into the *hip* locus [6] which has been mapped to position 33.90 on the *E. coli* chromosome [7]. Since this locus is between 3.8 and 3.9 kb downstream of *OrfT*, we propose that the map position of *OrfT* is 33.82. The initiation codon ATG (at position 1, Fig. 2) is preceded by a ribosome binding site (AGG) and promoter elements at positions −35 (TTGGCG) and −10 (TTTAAT) with respect to the transcription start site which show the basic characteristics of *E. coli* promoters [8]. There are no obvious promoter elements between *OrfT* and the second ORF, raising the possibility that these two genes are polycistronic.

A search of the EMBL/GenBank Data Libraries revealed that part of the *OrfT* sequence (from residue 1740 to 2130) shows 98.0% identity with a fragment of DNA from landmark loops of lampbrush chromosomes from the amphibian newt, *Pleurodeles waltlii* [9]. Since the *OrfT* sequence is contiguous with that of the *E. coli hip* operon, it appears likely that the newt sequence is in fact an *E. coli* sequence and therefore a cloning artefact. The 1911 bp *OrfT* sequence encodes

```
-281                                                                                         GGGCAGAACTCTCTTCTTGAAACATTCAATATGTACGTAGG
-240  GACATCTGGTACGGGTACGTTAACCCTGACGAATAACGGTACGCTGAATGTCGAAGGTGGAGAAGTTTACTTAGGTGTTTTTGAGCCTGCTGTAGGAACGCTAAACATTGGTGCTGCTCA
                                                     (c)                       (b)                       (a)
-120  CGGTGAGGCGGCGCAGATGCCGGATTTATCACCAATGCGACGAAAGTGGAGTTTGGTCTTGGCGAAGGCGTTTTTGTCTTTAATCATACCAATAACAGCGATGCCGGCTACCAGGTCGAT

   1  ATGCTGATTACAGGTGACGATAAAGACGGAAAAGTGATCCATGATGCAGGCCATACGGTGTTCAATGCAGGGAATACTTATAGCGGTAAAACGCTGGTCAATGACGGCCTCCTGACCATT
       M  L  I  T  G  D  D  K  D  G  K  V  I  H  D  A  G  H  T  V  F  N  A  G  N  T  Y  S  G  K  T  L  V  N  D  G  L  L  T  I   40

 121  GCGTCTCATACGGCAGATGGGGTAACGGGCATGGGGTCGAGTGAAGTAACCATTGCAAACCCCGGTACGCTCGACATTCTCGCATCAACGAACAGTGCAGGAGATTACACGCTGACCAAT
       A  S  H  T  A  D  G  V  T  G  M  G  S  S  E  V  T  I  A  N  P  G  T  L  D  I  L  A  S  T  N  S  A  G  D  Y  T  L  T  N   80

 241  GCGCTCAAAGGCGATGGCTTGATGCGAGTGCAGCTGTCATCCTCCGACAAGATGTTTGGCTTTACCCATGCAACAGGGACTGAATTCGCCGGTGTTGCCCAACTGAAAGACAGTACCTTC
       A  L  K  G  D  G  L  M  R  V  Q  L  S  S  S  D  K  M  F  G  F  T  H  A  T  G  T  E  F  A  G  V  A  Q  L  K  D  S  T  F  120

 361  ACTCTGGAACGCGACAACACCGCTGCGCTTACTCACGCGATGTTGCAGTCTGACAGTGAAAATACCACATCGGTAAAAGTTGGAGAGCAATCCATTGGTGGACTGGCCATGAATGGAGGT
       T  L  E  R  D  N  T  A  A  L  T  H  A  M  L  Q  S  D  S  E  N  T  T  S  V  K  V  G  E  Q  S  I  G  G  L  A  M  N  G  G  160

 481  ACCATCATTTTCGATACGGATATTCCTGCTGCGACGCTCGCGGAGGGATATATCAGCGTCGATACGCTGGTTGTCGGCGCGGGTGACTACACCTGGAAAGGCCGTAAGTATCAGGTAAAC
       T  I  I  F  D  T  D  I  P  A  A  T  L  A  E  G  Y  I  S  V  D  T  L  V  V  G  A  G  D  Y  T  W  K  G  R  K  Y  Q  V  N  200

 601  GGGACGGGCGACGTGCTTATCGACGTGCCTAAACCGTGGAATGATCCCATGGCGAATAACCCTCTGACGACGCTCAATTTGCTGGAACACGACGATAGCCATGTCGGCGTTCAACTGGTG
       G  T  G  D  V  L  I  D  V  P  K  P  W  N  D  P  M  A  N  N  P  L  T  T  L  N  L  L  E  H  D  D  S  H  V  G  V  Q  L  V  240

 721  AAGGCGCAAACGGTTATTGGGTCGGGTGGCTCATTAACGTTACGTGATTTACAGGGCGACGAGGTGGAAGCGGACAAAACGTTACACATTGCGCAAAACGGAACGGTGGTCGCCGAGGGT
       K  A  Q  T  V  I  G  S  G  G  S  L  T  L  R  D  L  Q  G  D  E  V  E  A  D  K  T  L  H  I  A  Q  N  G  T  V  V  A  E  G  280

 841  GATTATGGATTCCGCCTCACGACCGCACCAGGTAATGGTTTGTACGTTAACTATGGGCTGAAAGCGCTGAACATCCATGGTGGGCAAAAGCTGACGTTAGCCGAACATGGCGGAGCCTAT
       D  Y  G  F  R  L  T  T  A  P  G  N  G  L  Y  V  N  Y  G  L  K  A  L  N  I  H  G  G  Q  K  L  T  L  A  E  H  G  G  A  Y  320

 961  GGCGCAACGGCCGATATGTCGGCAAAAATCGGTGGTGAAGGGGATCTGGCAATCAATACGGTGCGACAGGTTTCGCTTTCCAACGGTCAGAACGACTATCAGGGGGCAACCTACGTTCAG
       G  A  T  A  D  M  S  A  K  I  G  G  E  G  D  L  A  I  N  T  V  R  Q  V  S  L  S  N  G  Q  N  D  Y  Q  G  A  T  Y  V  Q  360

1081  ATGGGGACATTACGTACCGATGCGGATGGTGCACTTGGCAACACCCGGGAACTGAACATCAGCAACGCGGCCATCGTCGATCTTAATGGATCGACGCAGACGGTAGAGACATTCACCGGG
       M  G  T  L  R  T  D  A  D  G  A  L  G  N  T  R  E  L  N  I  S  N  A  A  I  V  D  L  N  G  S  T  Q  T  V  E  T  F  T  G  400

1201  CAGATGGGTTCGACTGTTTTGTTCAAAGAGGGGGCGCTGACGGTAAATAAAGGTGGGATCAGTCAGGGTGAACTGACAGGTGGCGGAAACCTGAATGTTACAGGGGGAACGCTGGCTATC
       Q  M  G  S  T  V  L  F  K  E  G  A  L  T  V  N  K  G  G  I  S  Q  G  E  L  T  G  G  G  N  L  N  V  T  G  G  T  L  A  I  440

1321  GAGGGGCTTAATGCACGCTACAATGCGTTAACCAGCATTAGCCCAAATGCGGAAGTCAGCCTCGATAATACTCAGGGGTTAGGCAGAGGAAATATTGCCAATGACGGTCTGTTAACGCTA
       E  G  L  N  A  R  Y  N  A  L  T  S  I  S  P  N  A  E  V  S  L  D  N  T  Q  G  L  G  R  G  N  I  A  N  D  G  L  L  T  L  480

1441  AAAAACGTGACTGGCGAACTGCGTAATAGCATAAGCGGGAAGGGTATCGTGAGCGCAACCGCCAGGACAGATGTAGAGTTGGATGGCGATAATAGCCGCTTTGTGGGGCAATTCAACATT
       K  N  V  T  G  E  L  R  N  S  I  S  G  K  G  I  V  S  A  T  A  R  T  D  V  E  L  D  G  D  N  S  R  F  V  G  Q  F  N  I  520

1561  GATACAGGCAGCGCGCTCAGCGTCAACGAGCAGAAAAACCTGGGTGATGCTTCCGTTATCAATAATGGCCTGCTCACCATCTCCACTGAGCGTAGCTGGGCGATGACGCACAGTATCAGC
       D  T  G  S  A  L  S  V  N  E  Q  K  N  L  G  D  A  S  V  I  N  N  G  L  L  T  I  S  T  E  R  S  W  A  M  T  H  S  I  S  560

1681  GGTAGCGGTGATGTGACAAAACTGGGTACCGGGATCCTGACTCTTAACAACGATTCCGCGGCGTATCAGGGTACGACGGATATCGTGGGGGGGGAAATTGCTTTCGGTTCCGACTCTGCC
       G  S  G  D  V  T  K  L  G  T  G  I  L  T  L  N  N  D  S  A  A  Y  Q  G  T  T  D  I  V  G  G  E  I  A  F  G  S  D  S  A  600

1801  ATTAATATGGCAAGTCAACACATTAATATTCATAACAGCGGTGTGATGTCGGGAAATGTCACCACTGCAGGTGATATGAACGTTAGTCCTGGGGGGGGCACTGCGTGTCGCTAAAACCAC
       I  N  M  A  S  Q  H  I  N  I  H  N  S  G  V  M  S  G  N  V  T  T  A  G  D  M  N  V  S  P  G  G  G  T  A  C  R  -     637

1921  TATCGGCGGCAACCTGGAGAATGGCGGCACGGTTCAAATGAATAGCGAAGGGGGGAAACCGGGGAATGTACTGACCGTTAACGCAACTATCCGGAAACAATGGCCTGATGACGTTCAACG
2041  CGACGTTGGGCGGCATAATTCGCCCACCGATAAGATGAACGTGAAAGGCGATACCCAAGGGAACACTCGCGTTCGGGTTGATAACATTGGC
```

Fig. 2. Nucleotide sequence and deduced amino acid composition of *OrfT*. The translated sequence is given in the single letter code. The numbering of nucleotides is shown on the left hand side and amino acid residues on the right hand side. The initiation codon at position 1 is preceded by a putative ribosome binding site (a) and promoter elements at −10 (b) and −35 (c). The potential transcription start site is approx. 10 bases down-stream of (b).

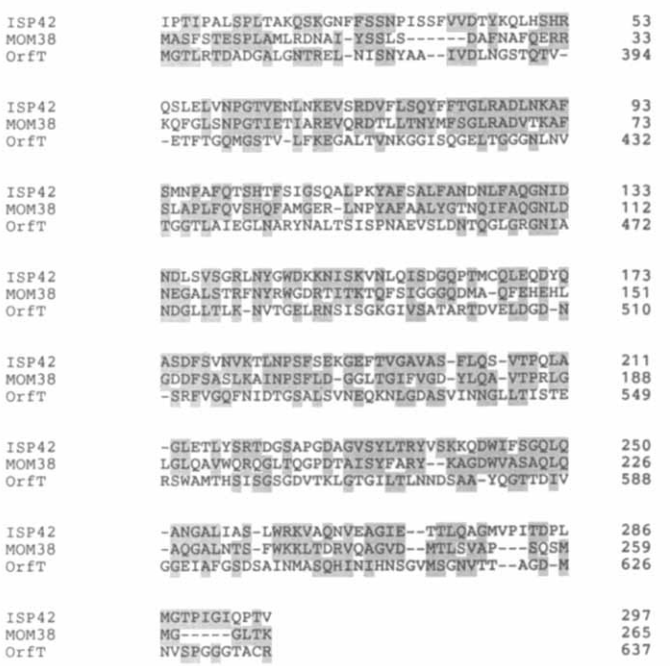

Fig. 3. Amino acid similarities between *E. coli* ORFT, yeast ISP42 and *Neurospora* MOM38. Sequences starting with residue 361 of OrfT, 13 of ISP42 and 1 of MOM38 were compared. Identical residues and conservative replacements defined as T/S, D/E, K/R/H, I/L/V, G/A and Q/N are shaded grey.

a protein of 637 amino acids (Fig. 2) with an estimated molecular mass of 68249 Da. The protein is very acidic, having a predicted pI of 4.26. Comparison of the protein sequence encoded by *OrfT* (ORFT) with ISP42 from *S. cerevisiae* and MOM38 from *N. crassa* shows that the carboxyterminal half of the molecule (residues 361 to 637) is 26.8% and 25.4% homologous, respectively (Fig. 3). This compares with a similarity of 53.2%, between ISP42 and MOM38. As is the case with the integral membrane protein ISP42, ORFT also has no obvious α-helical transmembrane domains.

ORFT represents the first example of a bacterial protein with a similarity to the membrane components of the mitochondrial translocation machinery. In the absence of functional data, we cannot assume that the protein encoded by *OrfT* functions in bacterial protein translocation, but the similarity between ORFT and ISP42 and MOM38 suggests a potential relationship between these genes. We believe that the identification of *OrfT* provides further support for the hypothesis that genes from an endosymbiont ancestor were transmuted by the host cell to create a system for protein import into mitochondria.

This work was supported by a grant from the Australian Research Council. We gratefully acknowledge the assistance of Ms A. Wright in the preparation of this paper.

References

1 Gray, M.W. (1989) Trends Genet. 5, 294–299.
2 Hannavy, K., Rospert, S. and Schatz, G. (1993) Current Opinion Cell Biol. 5, 694–700.
3 Baker, K. and Schatz, G. (1991) Nature 349, 205–208.
4 Kiebler, M., Pfaller, R., Söllner, T., Griffith, G., Horstmann, R., Pfanner, N. and Neupert, W. (1990) Nature 348, 610–616.
5 Wickner, W., Driessen, A.J.M. and Hartl, F.-U. (1991) Annu. Rev. Biochem. 60, 101–124.
6 Black, D.S., Kelly, A.J., Mardis, M.J. and Moyed, H.S. (1991) J. Bacteriol. 173, 5733–5739.
7 Kröger, M., Wahl, R., Schachtel, G. and Rice, P. (1992) Nucleic Acid Res. 20, 2119–2144.
8 Harley, C.B. and Reynolds, R.P. (1987) Nucleic Acid Res. 15, 2343–2361.
9 Penrad-Mobayed, M., Sourrouille, P., Bonnafant-Jäis, M.L., N'Da, E., Edström, J.E. and Angelier, N. (1991) Chromosoma 101, 180–188.