

## The primary structure of elongation factor EF-1 $\alpha$ from the brine shrimp *Artemia*

Formijn J. van Hemert, Reinout Amons,  
Wim J.M. Pluijms, Hans van Ormondt and Wim Möller\*

Department of Medical Biochemistry, Sylvius Laboratories,  
Wassenaarseweg 72, 2333 AL Leiden, Netherlands

\*To whom reprint requests should be sent  
Communicated by W. Möller

**cDNA as well as amino acid sequencing has revealed the complete primary structure of elongation factor EF-1 $\alpha$  from the brine shrimp *Artemia*. A comparison with the published sequences of bacterial EF-Tu, mitochondrial EF-Tu and chloroplastic EF-Tu shows that distinct areas of these polypeptide chains are conserved in evolution. The evolutionary distance between prokaryotic and eukaryotic types of EF-Tu is larger than among bacterial and organellar EF-Tus. A number of regions present in both EF-Tu and EF-G from *Escherichia coli* are also found in EF-1 $\alpha$  from *Artemia*.**

**Key words:** protein biosynthesis/elongation factor I/*Artemia*/protein and DNA sequencing/evolution

### Introduction

In eukaryotic cells, elongation factor EF-1 $\alpha$  catalyses the binding of aminoacyl-tRNA to the ribosome under hydrolysis of GTP (Kaziro, 1978). The abundance and functional importance of the elongation factor in protein synthesis has been a stimulus for its purification and characterization from a large variety of sources (Weissbach, 1980).

The amino acid sequence of the bacterial elongation factor EF-Tu from *Escherichia coli* (EF-TuA and EF-TuB) has been determined (Jones *et al.*, 1980; Laursen *et al.*, 1981; Nakamura *et al.*, 1982), as well as the nucleotide sequence of the corresponding genes, *TufA* and *TufB* (Yokota *et al.*, 1980; An and Friesen, 1980). Recently, nucleotide sequence analysis of an *Euglena gracilis* chloroplast genomic region has revealed a complete sequence for EF-Tu as well as the existence of two introns within the coding region (Montandon and Stutz, 1983). In the yeast *Saccharomyces cerevisiae* an uninterrupted gene codes for the mitochondrial elongation factor EF-Tu and its nucleotide sequence has also been completed (Nagata *et al.*, 1983). Both proteins are very homologous to their bacterial counterpart. We have already reported the partial primary structure of EF-1 $\alpha$  from the brine shrimp *Artemia* (Amons *et al.*, 1983), the nucleotide sequence of a 542-bp cDNA clone (Van Hemert *et al.*, 1983a) and a provisional characterization of the EF-1 $\alpha$  genes (Van Hemert *et al.*, 1983b). Moderate homology between the elongation factors from *E. coli* and *Artemia* has been observed. Here we present the complete primary structure of EF-1 $\alpha$  from *Artemia* as determined by cDNA as well as by protein sequencing. The three EF-Tu proteins mutually show a higher degree of homology compared with EF-1 $\alpha$  from *Artemia*. The results allow an insight into the conservation of the regions, which constitute functional domains, shared by prokaryotic and eukaryotic elongation factors.

### Results

The amino acid sequencing of EF-1 $\alpha$  from *Artemia* as presented previously (Amons *et al.*, 1983) was finished almost completely. Figure 1 summarizes the peptides for the protein sequencing together with the proposed sequence.

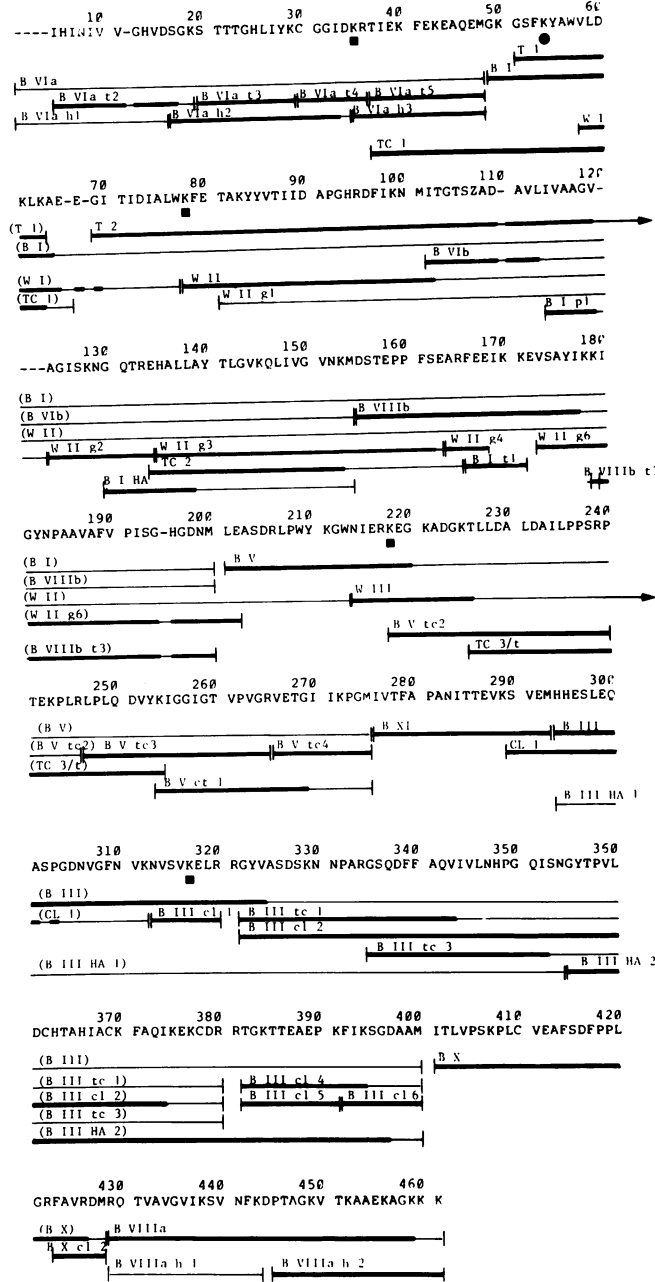
Figure 2a displays the cDNA cloning and sequencing strategy. Our original cDNA clone, renamed pcAE1 (Van Hemert *et al.*, 1983a) was used as a probe to detect more EF-1 $\alpha$  cDNA clones in an oligo(dT)-primed cDNA library. As a result a clone (pcAE2E3) was obtained extending from nucleotide position 630 to the TGA codon followed by an 111-bp part of the 3' non-translated region. This 3' non-coding region of pcAE2E3 does not include a poly(A) tract nor an AATAAA consensus sequence, indicating the absence of the ultimate 3' end of the EF-1 $\alpha$  mRNA.

To clone the 5' end of the intact EF-1 $\alpha$  mRNA, the pcAE2E3 restriction fragment from the *Bst*NI site (891) to the *Bst*NI site (981) was used to prime cDNA synthesis. The resulting library was screened with the 5' *Pst*I-*Bst*NI fragment as a probe. All hybridizing clones contain the expected *Nco*I site (696) and one of them, pcAE13G3, has an insert reaching from position 378 to the *Bst*NI priming fragment. Finally, re-screening the library with the 5' *Pst*I-*Hind*II (525) fragment of pcAE13G3 resulted in the detection of a fourth cDNA clone, pcAE17A5. This clone has the ATG start codon together with 70 upstream non-coding base pairs and shows at its 5' end a 58-bp inverted segment of its 3'-proximal part, containing the *Eco*RI site (435). A mechanism underlying this phenomenon has been proposed, suggesting that the usual small 3' hairpin does not form properly after first strand synthesis (Volkaert *et al.*, 1981). It is unlikely that the coding part of the nucleotide sequence should have been modified by this event, since a primer extension experiment on a poly(A)<sup>+</sup> template confirms the nucleotide sequence of the 5' part and shows that the pcAE17A5 cDNA sequence (Figure 2b) starts within a few nucleotides from the cap site of the EF-1 $\alpha$  mRNA.

The resulting cDNA sequence is presented in Figure 2b, which compares its deduced amino acid sequence with the determined one. The elongation factor EF-1 $\alpha$  from *Artemia* contains 461 amino acids, has a calculated mol. wt. of 50 274 and is positively charged at pH 7. These data match the properties of the purified enzyme (Slobin and Möller, 1976). A few post-translational modifications are found, namely the presence of four  $\epsilon$ -trimethyllysine residues (positions 35, 78, 218, 317) and another modified amino acid, probably  $\epsilon$ -monomethyllysine (54). The N-terminal amino acid of the isolated protein is blocked, but is not methionine.

### Discussion

The deduced amino acid sequence of EF-1 $\alpha$  from *Artemia* differs at three positions, residues 32, 181, 373, from the amino acid sequence determined in a direct way. An explanation for this difference may be a copying error of one of the



**Fig. 1.** Amino acid sequence analysis of EF-1 $\alpha$  from *Artemia*. The prefixes B and W indicate Br-CN- and BNPS-skatole generated peptides, respectively. The numbering of the BrCN-generated peptides refers to Figure 1 in Amons *et al.* (1983). BrCN I is an aggregate of the peptides 49–101, 49–154 and 49–200. T indicates tryptic peptides (limited digestion conditions), CL clostripain-derived peptides and TC tryptic peptides after citraconylation. Subdigestions are indicated in the figure by h, HA, g, t, tc, /t, ct, p, cl and represent peptides obtained by partial acid hydrolysis (Inglis *et al.*, 1980), hydroxylamine, *Staphylococcus aureus* V8 protease, trypsin, trypsin (after citraconylation), trypsin (limited digestion conditions), chymotrypsin, pepsine and clostripain, respectively. Isolated and identified peptides are indicated by thin lines and, according to the extent of sequencing, by thick lines. Peptide T2 (which has been obtained by limited digestion of the native protein (Amons *et al.*, 1983)) as well as the peptide W III extend to the C-terminal end of the protein, visualized by  $\rightarrow$ . ■ indicates a  $\epsilon$ -trimethyllysine residue (Amons *et al.*, 1983); ● another modified lysine residue, presumably  $\epsilon$ -monomethyllysine. With respect to the partial sequence published, the following additions have been made: Gln-146, Arg-334, Gln-351, Gln-373, Cys-378, as well as the following corrections: Lys-83, an unidentified residue at position 110, Leu-245, Gln-300, Ser-336, Gln-337, Gln-342, His-348.

enzymes used for the cDNA synthesis (Loeb *et al.*, 1979). Another possibility may be that the cloned cDNA originates from minor mRNA species for EF-1 $\alpha$ . *Artemia* cysts may represent a heterogeneous population, although direct amino acid sequence analysis does not indicate microheterogeneity of the EF-1 $\alpha$  polypeptide chain. Since the nucleotide sequence is the result of overlapping sequences of four different cDNA clones, sequence data of different genes may have been combined. Overlapping nucleotide sequences of the four different cDNA clones are identical with one exception (see Figure 2a and Van Hemert *et al.*, 1983a).

The amino acid sequence of EF-1 $\alpha$  from *Artemia* was compared (Figure 3) with those of EF-Tu from *S. cerevisiae* mitochondria (Nagata *et al.*, 1983) from *E. gracilis* chloroplasts (Montandon and Stutz, 1983) and from *E. coli* (Jones *et al.*, 1980; Laursen *et al.*, 1981; Nakamura *et al.*, 1982). Four regions in EF-1 $\alpha$  from *Artemia* are not present in EF-Tu from *E. coli*, reflecting their difference in mol. wt. The EF-1 $\alpha$  polypeptide chain shows an overall homology of ~25% with respect to the three EF-Tu factors, which have a mutual homology of 60–70%. The homology between EF-1 $\alpha$  from *Artemia* and the other elongation factors is unevenly distributed and particularly prominent in the N-terminal third of the molecule. Interestingly, most areas of homology in this N-terminal region nicely match the homologous areas in the amino-terminal parts of the elongation factors EF-Tu and EF-G from *E. coli* (Laursen *et al.*, 1981; see the residues indicated by asterisks in Figure 3). This supports the presence in both eukaryotic and prokaryotic elongation factors of certain conserved regions, which is not unexpected since one factor, EF-Tu (or EF-1 $\alpha$ ), binds tRNA to the ribosome, whereas the other, EF-G (or EF-2) causes release of tRNA from the ribosome (Leder, 1973). The sequences around Met-101 and His-135 (*Artemia* EF-1 $\alpha$  numbering; see Figure 3), however, are confined to only the EF-Tu and EF-1 $\alpha$  proteins and may therefore represent a unique function of these proteins.

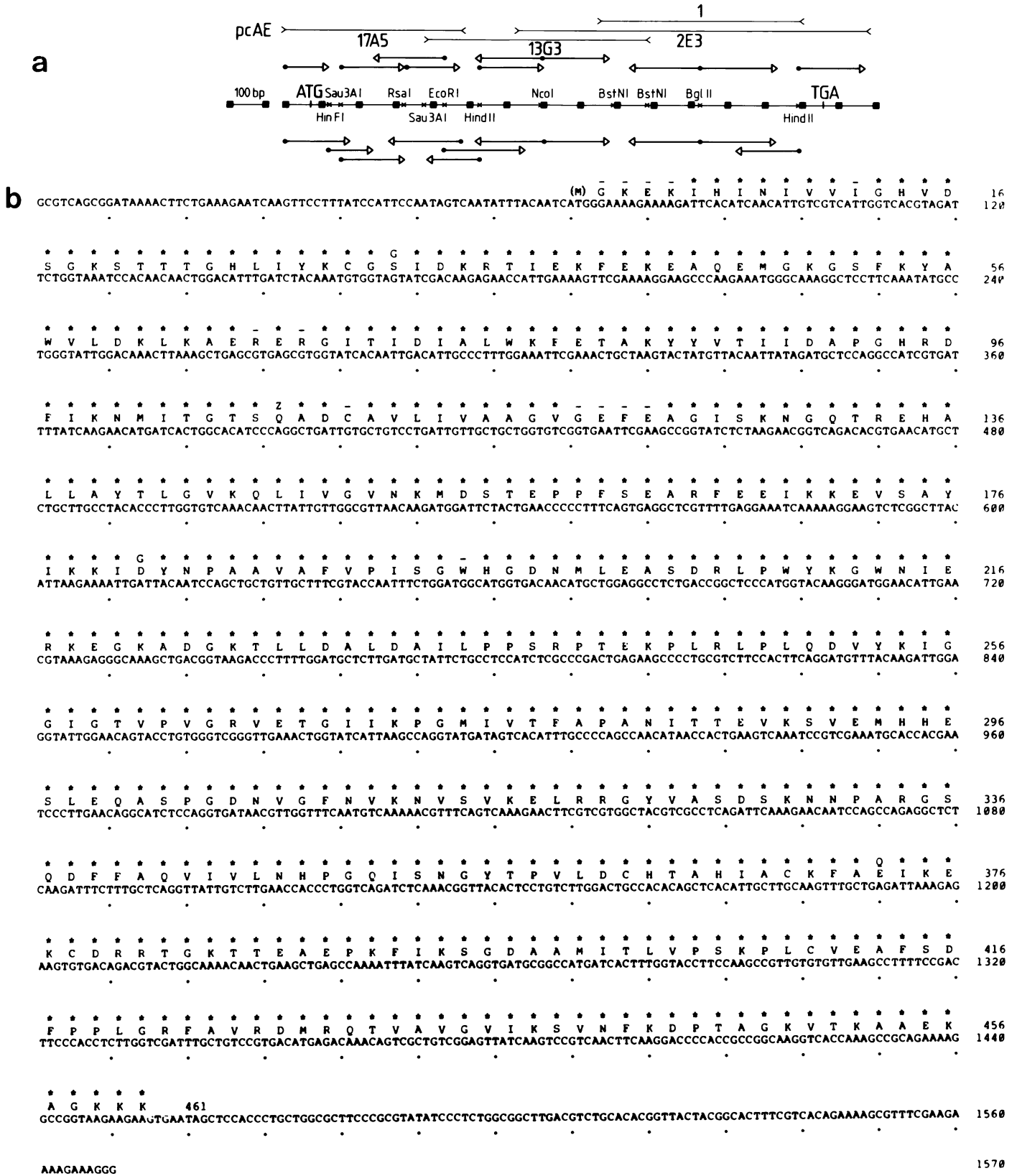
Modification of Cys-81 in *E. coli* EF-Tu abolishes binding of aminoacyl-tRNA as does modification of Cys-137 with respect to the binding of GDP, GTP and EF-Ts. His-66 is the target point when the protein is reacted with bromoacetyl-lysine tRNA (for a review, see Liljas, 1982). These three amino acid residues are part of regions which are similar in the EF-Tu and EF-1 $\alpha$  proteins presented (Figure 3), although these residues themselves are in general not conserved. However, the main site of limited tryptic digestion of *E. coli* EF-Tu (Arg-58/Gly 59) is conserved in all four proteins and is embedded in a region of high homology.

Recently, tRNA cross-linking experiments with EF-Tu from *E. coli* revealed Lys-208 and -237 as being located in the vicinity of the tRNA binding sites (Kraal *et al.*, 1983). Interestingly, these residues are also conserved in the other elongation factors EF-Tu and EF-1 $\alpha$  and are situated in homologous areas in the middle part of their polypeptide chains. In general, the homology observed supports the notion that during evolution certain structural elements have been retained in the elongation factors.

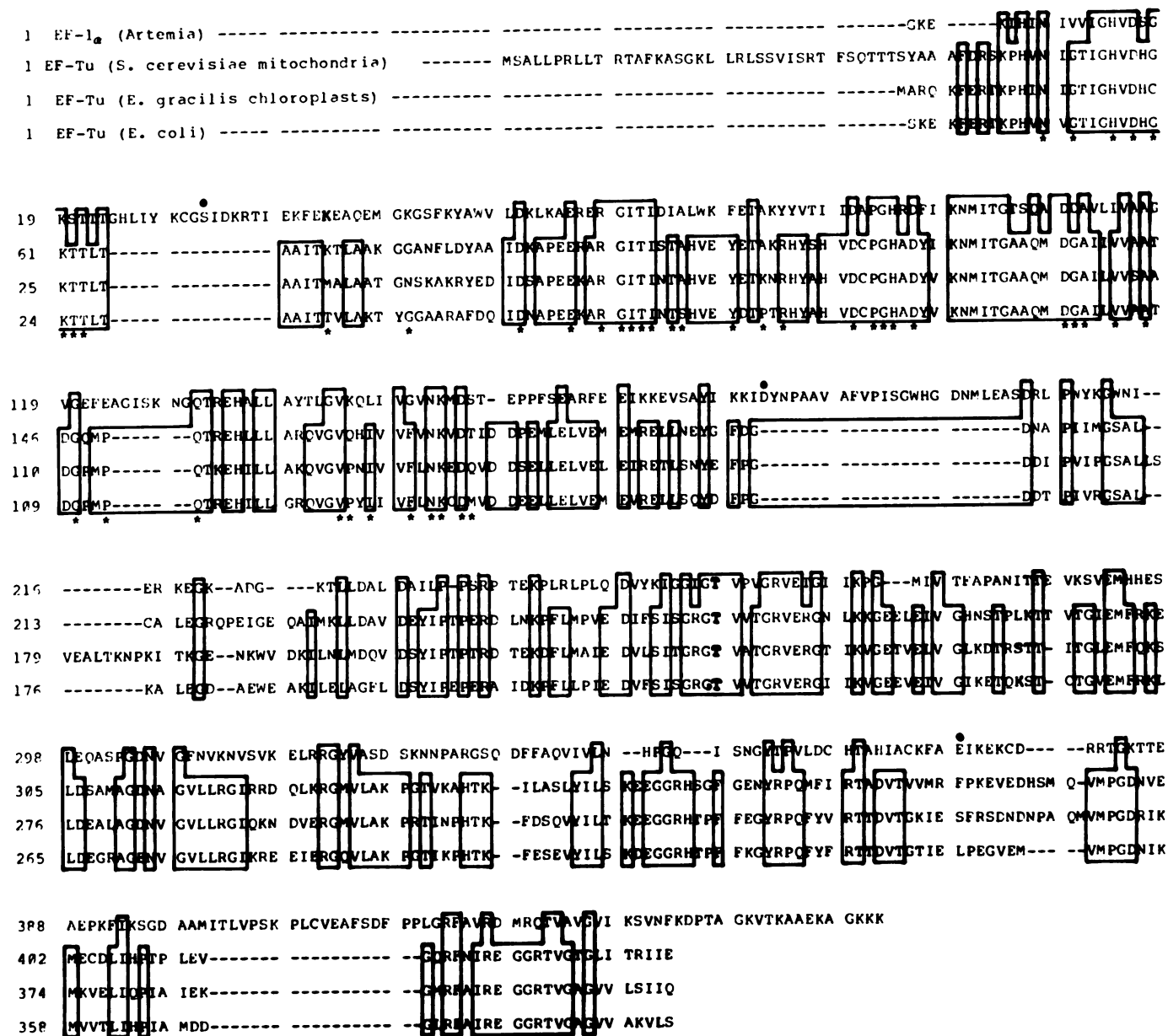
## Materials and methods

### Protein sequence analysis

Most methods have been described in our previous paper (Amons *et al.*, 1983). After fragmentation of the S- $\beta$ -pyridylethylated protein by BNPS-skatole (Hunziker *et al.*, 1980) the products were fractionated on a Sephacryl S200 column (0.9 x 145 cm) with 8 M urea and 0.25% (v/v) HCOOH as an eluant. Purification of the peptides was performed by reverse-phase h.p.l.c.



**Fig. 2. (a)** EF-1 $\alpha$  cDNA clones and sequence strategy. The upper part shows the four overlapping cDNA clones used ( $\blacktriangleleft$ — $\blacktriangleright$ ). The lower part shows the positions of the ATG start codon and the TGA stop codon as well as a map of the relevant restriction sites (—x—). Readings of the nucleotide sequence are given by ( $\bullet$ — $\rightarrow$ ). The sequence of the cDNA clone pcAE1 has been published by Van Hemert *et al.* (1983a). The cDNA clone pcAE2E3, which completely includes pcAE1, displays the amino acid Glu (373) instead of a stop codon in the cDNA clone pcAE1 at this position. **(b)** Sequence of EF-1 $\alpha$  and its cDNA from *Artemia*. Bottom line represents the nucleotide sequence of the coding strand. Middle line is the deduced amino acid sequence. Upper line corresponds to the determined amino acid sequence from protein data. Asterisks (\*) indicate residues, which are identical in the deduced as well as in the directly determined sequence. Dashes (—) indicate residues not determined by protein sequence analysis. In the case of a difference, the residue determined by protein sequencing is given (Z:Glx).



**Fig. 3.** Sequence homology of EF-1 $\alpha$  from *Artemia* and EF-Tu from *S. cerevisiae* mitochondria (Nagata *et al.*, 1983), from *Euglena gracilis* chloroplasts (Montandon and Stutz, 1983) and from *E. coli* (Jones *et al.*, 1980; Laursen *et al.*, 1981; Nakamura *et al.*, 1982). Boxes indicate residues which are conserved in all four proteins (boxes of four lines) or only in the EF-Tu elongation factors (boxes of three lines). The number of homologies is considerably larger if any three out of four different sequences are boxed at positions of identity. Dots above the upper sequence mark the three positions where the amino acid and the nucleotide sequence determinations have gained different residues. Asterisks under the lower sequence represent homologous residues in EF-Tu and EF-G from *E. coli* for the first 140 amino acids (Laursen *et al.*, 1981). Following residue 140 the homology is substantially less and therefore not shown (Ovchinnikov *et al.*, 1982).

essentially as described by Amons *et al.* (1983), but using a gradient of 0.2% (v/v) trifluoroacetic acid to 0.1% (v/v) trifluoroacetic acid and 70% (v/v) CH<sub>3</sub>CN. This system has been utilized for most other peptide separations including purification of the products of limited tryptic digestion of the native proteins (Slobin *et al.*, 1983; Amons *et al.*, 1983).

**Molecular cloning and nucleotide sequence analysis**

The isolation of RNA, the size fractionation of poly(A)<sup>+</sup> RNA and the procedures for synthesis and cloning of cDNA have been described before (Van Hemert *et al.*, 1981; Maniatis *et al.*, 1982; Van Hemert *et al.*, 1983a). *E. coli* C600 instead of *E. coli* RR1 was used as a host for transformation. cDNA synthesis by priming with a restriction fragment was essentially done according to Ullrich *et al.* (1983). After the S1 nuclease treatment the cDNA was fractionated by means of chromatography on Sephacryl S400 in order to remove the primer.

Screening of the ordered cDNA libraries was performed according to Gergen *et al.* (1979). cDNA clones were sequenced by the chemical modification method (Maxam and Gilbert, 1977), either directly in the pBR322 vector or after recloning the insert in the *Pst*I site of pUC8 (Vieira and Messing, 1982). For RNA sequencing the chain-terminating method of Zimmern and Kaesberg (1978) was adapted. The 32-bp *Hin*I (130)-*Sau*3A1 (152) restriction fragment was dissolved in hybridization buffer together with poly(A)<sup>+</sup> RNA from a gradient fraction enriched for EF-1 $\alpha$  encoding sequences (Van Hemert *et al.*, 1981). After heating at 70°C for 15 min the mixture was hybridized for 6 h under slow cooling from 40 to 30°C.

Enzymes were obtained from NE-Biolabs (UK), BRL (USA), Worthington (USA), Boehringer (FRG), CBL (UK) and The Radiochemical Centre, Amersham (UK). <sup>32</sup>P-Labelled nucleoside triphosphates were purchased from The Radiochemical Centre, Amersham.

## Acknowledgements

The authors express their gratitude to Mrs. R. de Ru for reading the manuscript and to Mr. G.M.C. Janssen and Mrs. J.H.G.M. Brands for expert technical assistance. This research was supported by The Netherlands Organization for the Advancement of Pure Research (ZWO) and by The Netherlands Foundation for Chemical Research (SON).

## References

- Amons,R., Pluijms,W., Roobol,K. and Möller,W. (1983) *FEBS Lett.*, **153**, 37-42.
- An,G. and Friesen,J.B. (1980) *Gene*, **12**, 33-39.
- Gergen,J.P., Stern,R.H. and Wensink,P.C. (1979) *Nucleic Acids Res.*, **7**, 2115-2136.
- Hunziker,P.E., Hughes,G.J. and Wilson,K.J. (1980) *Biochem. J.*, **187**, 515-519.
- Inglis,A., McKern,N., Roxburgh,C. and Strike,P. (1980) in Birr,C. (ed.), *Methods in Peptide and Protein Sequence Analysis*, Elsevier Biomedical, Amsterdam/NY, pp. 329-343.
- Jones,M.D., Petersen,T.E., Nielsen,K.M., Magnusson,S., Sottrup-Jensen, L., Gausing,K. and Clark,B.F.C. (1980) *Eur. J. Biochem.*, **108**, 507-526.
- Kaziro,Y. (1978) *Biochim. Biophys. Acta*, **505**, 95-127.
- Kraal,B., Van Noort,J.M. and Bosch,L. (1983) in Abraham,A.K. (ed.), *Translation and Post-Translational Events*, Humana Press Inc., Clifton, NY, in press.
- Laursen,R.A., L'Italien,J.J., Nagarkatti,S. and Miller,D.L. (1981) *J. Biol. Chem.*, **256**, 8102-8109.
- Leder,P. (1973) *Adv. Protein Chem.*, **27**, 213-239.
- Liljas,A. (1982) *Prog. Biophys. Mol. Biol.*, **40**, 161-228.
- Loeb,L.A., Weijmouth,L.A., Kunkel,T.A., Gopinathan,K.P., Beckman, R.A. and Duke,D.K. (1979) *Cold Spring Harbor Symp. Quant. Biol.*, **43**, 921-927.
- Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, a Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
- Maxam,A.M. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560-564.
- Montandon,P.-E. and Stutz,E. (1983) *Nucleic Acids Res.*, **11**, 5878-5892.
- Nagata,S., Tsunetsugu-Yokota,Y., Naito,A. and Kaziro,Y. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 6192-6196.
- Nakamura,S., Nakayama,N., Takahashi,K. and Kaziro,Y. (1982) *J. Biochem.*, **91**, 1047-1063.
- Ovchinnikov,Y.A., Alakhov,Y.B., Bundulis,Y.P., Bundule,M.A., Dovgas, N.V., Kozlov,V.P., Motuz,L.P. and Vinokurov,L.M. (1982) *FEBS Lett.*, **139**, 130-135.
- Slobin,L.I. and Möller,W. (1976) *Eur. J. Biochem.*, **69**, 351-375.
- Slobin,L.I., Clark,R.V. and Olson,M.O.J. (1983) *Biochemistry (Wash.)*, **22**, 1911-1917.
- Ullrich,A., Gray,A., Berman,C. and Dull,T.J. (1983) *Nature*, **303**, 821-825.
- Van Hemert,F.J., Roobol,K. and Möller,W. (1981) *Eur. J. Biochem.*, **120**, 137-142.
- Van Hemert,F.J., Van Ormondt,H. and Möller,W. (1983a) *FEBS Lett.*, **157**, 289-293.
- Van Hemert,F.J., Lenstra,J.A. and Möller,W. (1983b) *FEBS Lett.*, **157**, 295-299.
- Vieira,J. and Messing,J. (1982) *Gene*, **19**, 259-268.
- Volkaert,G., Tavernier,J., Derynck,R., Devos,R. and Fiers,W. (1981) *Gene*, **15**, 215-223.
- Weissbach,H. (1980) in Chambliss,G. et al. (eds.), *Ribosomes, Structure, Function and Genetics*, University Park Press, Baltimore, MD, pp. 377-411.
- Yokota,T., Sugisaki,H., Takanimi,M. and Kaziro,Y. (1980) *Gene*, **12**, 25-31.
- Zimmern,D. and Kaesberg,P. (1976) *Proc. Natl. Acad. Sci. USA*, **75**, 4257-4261.

Received 10 February 1984