Functional implications related to the gene structure of the elongation factor EF-Tu from *Halobacterium marismortui*

Giuseppe Baldacci, Françoise Guinet^{1*}, Jeanne Tillit, Giuseppe Zaccai² and Anne-Marie de Recondo

GBGM, CNRS UPR 272, 94802 Villejuif Cedex, ¹EMBL, Grenoble outstation, 156X, 38042 Grenoble Cedex and ²ILL, CNRS URA 1333, 156X, 38042 Grenoble Cedex, France

Received October 23, 1989; Accepted December 29, 1989

EMBL accession no. X16677

ABSTRACT

The primary structure of the gene for the elongation factor EF-Tu from the halophilic archaeobacterium *Halobacterium marismortui* (hEF-Tu) is described. It is the first gene of a halophilic elongation factor EF-Tu to be sequenced. When the sequence of hEF-Tu is compared to that of homologous proteins from other organisms, the highest identity (61%) is found with EF-Tu from *Methanococcus vanniellii*, a non-halophilic archaeobacterium. In the search for halophilic characteristics therefore the most appropriate comparison is with the *M. vanniellii* sequence. The excess of acidic amino acid residues in the hEF-Tu sequence (already observed in the composition of other halophilic proteins) results to a large extent from changes of Lys, Asn or Gln to Asp or Glu. A structural analysis algorithm applied to the halophilic sequence places these acidic residues on the surface of the protein. The corresponding residues in the crystal structure of the first domain of EF-Tu from *E. coli* (the only EF-Tu structure available) are grouped in patches on the protein surface, in each of which several residues that may be far apart in the sequence come quite close to each other in the tertiary structure.

INTRODUCTION

Cellular metabolism is strictly dependent on precise physical and chemical conditions. In extreme halophiles the cytoplasmic concentration of KCl is higher than 3.5 M (1). This raises questions concerning the adaptative mechanisms that allow biosynthetic processes in such extreme salt conditions. Because of its role in protein biosynthesis, its interactions with GDP, GTP, tRNAs and ribosomes, the elongation factor EF-Tu from the extremely halophilic archaeobacterium *H. marismortui* (hEF-Tu) is a good model for the study of these adaptative mechanisms. In this paper we present the gene structure of this protein. *H. marismortui* hEF-Tu was purified previously with suitable modifications of methods used for eubacteria and the sequence of its 24 N-terminal amino acids was established (2). This protein remains soluble and functions in a physiological environment with

a salt composition of about 3.5 M KCl and denatures slowly at salt concentrations lower than about 2 M (2). The N-terminal sequence allowed us to synthesize a mixture of oligonucleotides for screening a genomic bank of *H. marismortui* in order to identify and then to sequence the hEF-Tu gene. The comparison of the predicted peptide with homologous proteins indicates that the halophilic nature of hEF-Tu is due to the presence of excess acidic residues forming patches at the surface of the protein. These findings support a recent model for the stabilization of a halophilic protein in which cooperative interactions between acidic amino acid residues and hydrated salt ions play a crucial role (3).

MATERIALS AND METHODS

Using as a starting point the sequence of the N-terminal end of hEF-Tu we have chosen the following mixture of synthetic oligonucleotides:

corresponding to amino acid 3 to 8 as a probe in order to fish hybridizing sequences out of *H. marismortui* DNA (1). Nucleic acids from *H. marismortui* were obtained with repeated phenol extractions of osmotically lysed cells and DNA was purified in CsCl density gradients. Southern blots of restriction enzyme digested DNA and molecular hybridizations were performed as published (4). The 5'-end labeled mixture of oligonucleotides or the labeled 5' portion of the *E. coli* elongation factor gene were

Figure 1. Sequence strategy. All the coding region was sequenced on both strands, all restriction enzyme sites used for M13 cloning were 'passed-through'. The segment corresponding to the dashed line was also sequenced on both strands as directly obtained from genomic DNA.

* Present address: Rhone-Poulenc Santé, IBV, 13 Quai J.Guesde 94400 Vitry sur Seine, France

Figure 2. Nucleotide sequence and predicted hEF-Tu peptide. The underlined sequence codes for a segment identical to the N-terminal end obtained with the Edman degradation of hEF-Tu (2). Boxes marked with triangles indicate amino acids involved in the GDP-binding site (22, 23). Boxes marked with squares indicate putative components of the halophilic promoter and box marked with a circle indicates a 'Shine-Dalgarno'-like region (24). The arrows indicate sequences able to form a hairpin structure, possibly involved in transcription termination (24).

used as radioactive probes (5). The *H. marismortui* genomic bank was constructed after partial digestion of DNA with the restriction enzyme *Sau*3A I and insertion of suitable fragments at the *Bam*HI sites of the lambda derived phage EMBL3, following published procedures (6). Out of 3.6×10^6 screened recombinant phages we obtained 11 hybridizing phages. The DNA sequence of the portions of inserts responsible for hybridizations was established with the chain termination method (7, 8). Because of the high G+C content of this gene we performed the sequencing reactions

S.c.	1-45	MGKE-----KSHINVVVIGHVDSGKSTTTGHLIYKCGGIDKRTIEKFEKE
M.v.	1-45	MAK-----TKPLNVAFIGHVDAKSTTVGRLLLDGGAIDPQLIVLRKE
H.m.	1-44	MSDE-----QHQLNLAIGHVDHGKSTLVGRLLYETGSVPEHVIEQHKKE
E.c.	1-34	MSKEKFERTKPHVNVGTIGHVDHGKTTLTAATTTV-----
S.c.	46-95	ADELGKSGFYAWLVDLKAERERGITIDIALWKFETPKYQVTVIDAPGH
M.v.	46-95	AEEKGAGFEFAYVMDGLKEERERGVITDVAHKFPKTAKEVTIVDCPGH
H.m.	45-94	AEEKGKGFAYVMDNLAEERERGVITDIAHQEFSTDTYDFITVDCPGH
E.c.	35-84	LAKTYGGAARAFDQIDNAPEEKARGITINTSHVEYDTPTRHVAHVDCPGH
S.c.	96-145	RDFIKNMITGTSQADCAILLIAGGVGEFAGISDKDQGTREHALLAFTLGV
M.v.	96-141	RDFIKNMITGASQADAAVLVNVVDDAKS--GI--QPQTRHVFILRTLGV
H.m.	95-137	RDFVKNMITGASQADNAVLVAADDGV-----QPQTEHVFLARTLGI
E.c.	85-127	ADYVKNMITGAAQMDGAILVVAATDGP-----PQTRHILLGRQVGV
S.c.	146-194	RQLIVAVNKMDSVKWDES-RFQEIIVKTSNFIKKVGYNPKTVFPVPSGW
M.v.	142-191	RQLAVAVNKMMDTVNFSEADYNELKKMIGDQLLKMIGFNPQINFPVVASL
H.m.	138-186	GELIVAVNKMMDLVYDGESEYKQVVE-EVKDILLTVQRFDSNAKFIPIVSAL
E.c.	128-175	PYIIVFLNKCMDVDEE--LLELVEMEVRLLSQYDFPGDDTPIVRGSL
S.c.	195-244	NGDNMIEATTNAPWYKGEKETAAGVVGKTLLEAIDAEIQSPRPTDKPL
M.v.	192-229	HGDNVFKKSERNPWYK-----PTIAEVIDGFPPEKPTNLPL
H.m.	187-224	EGDNIAEESHTGYDG-----EILLEALNELPAPEPTDAPL
E.c.	176-210	KALEGDAAE-----WE-----AKILELAGFLDSYIPEPERAIDKPF
S.c.	245-291	RLPLQDVYKIGGIGTVPVGRVETGVIKPGM--VVFAPAGVTEVKSVE
M.v.	230-276	RLPIQDVYITITGVGTVPVGRVETGIKPGD--KVVFEPAAGIIEIKTVE
H.m.	225-271	RLPIQDVYITISIGITVPVGRVETGILNTGD--NVSFQPSDVSEVKTVE
E.c.	211-259	LLPIEDVFSISGRGTVTVGRVERGIKVGEEVIEVIGKETQ-KSTCTGVF
S.c.	292-341	MHHEQLQGVPGDNVGVFNKNSVKEIRRGVNCDAKNPPKGCASFNAT
M.v.	277-325	MHHEQLPSAEPDGNIGFNVGRVGKDKIRGDLGHTTNPTVA-TDFTAQ
H.m.	272-320	MHHEVPKAEFGDNVGVFNVRGVDKIRRGDVGCPADPPSAET-FQAQ
E.c.	260-307	MFRKLLDEGRAGENVGVLRLGKREIERGQVLAKPGTKPH--TKFESE
S.c.	342-386	VIVLNHPGQISA-----GYSPVLDCHTAHIAACRFDELLEKNRRSGKKLE
M.v.	326-370	IVVLQHPSVLTD-----GYTPVFHTHTAQIACTFAEIQKLNLPATGEVLE
H.m.	321-365	IVVMQHPSVITE-----GYTPVFHAHTAQVACTVESDKKIDPSSGEVAF
E.c.	308-349	VYILSKDEGGRHTPFKRYRPGFYFRITDVTGTIELPEGVEM-----
S.c.	387-436	DHPKFLKSGDAALVKFVPSKPMCEAFSEYPLGRFAVRDMRQTAVAGVI
M.v.	371-420	ENPDFLKAGDAIVKLIPKPMVIESVKEIPQLGRFAIRDMGMTVAAGMA
H.m.	366-415	ENPDFIQNGDAAVTVTRPKPLSIEPSSEIPELGSFAIRDMGTIAAGKV
E.c.	350-388	-----VMPGDNKVMVTLIHPIAMDD-----GLRFAIREGRTVGAGVV
S.c.	437-458	KSVDKTEKAAKVTAKAQAQK
M.v.	421-428	IQVTAQNK
H.m.	416-421	LGVNER
E.c.	389-393	AKVLS

Figure 3. Primary structures of EF-1 α from *S. cerevisiae* (S.c., 12), of EF-Tus from *E. coli* (E.c., 13, 14) and from archaeobacteria *M. vannielii* (M.v., 15) and *H. marismortui* (H.m.) were aligned to minimize insertions or deletions and to optimize identity. The identity percentages of hEF-Tu are : 61% with EF-Tu from *M. vannielii*, 47% with EF-1 α from *S. cerevisiae* and 34% with EF-Tu from *E. coli*.

TABLE I. PERCENTAGES OF AMINO ACID RESIDUES IN THREE PROKARYOTIC EF-Tus and ONE EUKARYOTIC EF-1 α

	<i>E. coli</i>	<i>H. marismortui</i>	<i>M. vannielii</i>	<i>S. cerevisiae</i>
Residues	393	421	428	458
Ala	6.8	7.3	8.4	8.1
Arg	5.8	3.3	3.7	3.9
Asn	1.8	3.6	4.0	3.5
Asp	6.1	7.6	5.8	5.2
Cys	0.8	0.7	0.5	1.5
Gln	2.0	4.5	4.0	2.6
Glu	9.4	10.4	6.3	6.8
Gly	10.1	8.5	8.6	9.2
His	2.8	3.3	2.3	2.4
Ile	7.3	6.2	7.5	6.5
Leu	7.1	5.2	6.3	5.2
Lys	5.8	3.8	7.9	10.7
Met	2.8	1.7	2.6	1.7
Phe	3.5	3.6	4.2	3.7
Pro	5.1	5.9	6.5	5.0
Ser	2.8	5.2	2.1	4.6
Thr	7.6	5.7	7.2	6.1
Trp	0.3	0.2	0.2	1.3
Tyr	2.5	1.9	1.4	1.7
Val	9.4	11.1	10.3	10.0
G+C %	53	63	39	43
Ref.	(13, 14)		(15)	(12)

Figure 4. A : Differences of corresponding amino acid residues in *E. coli* (E.c.), *M. vannielii* (M.v.) and *H. marismortui* (H.m.) EF-Tus. Numbers indicate positions of amino acid residues in respective EF-Tu protein sequences (see also fig. 3). B : Schematic drawing of predicted antigenicity of hEF-Tu. The profile is a linear combination of calculated hydrophilicity, flexibility and accessibility of hEF-Tu obtained with the procedure of Parker *et al.* (10). Arrows indicate new peaks, dots indicate increased peaks and vertical dashed bars indicate lower peaks in the profile of hEF-Tu compared to that of *M. vannielii*. Letters in the profile indicate acidic amino acid residues of hEF-Tu which are different at the corresponding positions of *M. vannielii* EF-Tu. Stars indicate positions of amino acids involved in the GDP-binding site.

also using dITP instead of dGTP : all the 'G+C compressions' were resolved with this method (9). The computer program of Parker *et al.* was used to predict antigenic determinants (10).

RESULTS

Figure 1 shows the strategy used for determining the nucleotide sequence of the DNA region recognized by the probes. Figure 2 shows the DNA sequence and the predicted protein which is identical in its N-terminal end to that obtained by the Edman degradation of the peptide (1). The predicted hEF-Tu contains 421 amino acids corresponding to an isoelectric point of 4.64. As the calculated molecular weight is 45702 and that obtained with SDS-PAGE was about 60000 (1), we checked the sequence of the 3'-gene extremity by sequencing it without prior cloning in a bank phage vector. For this purpose we electroeluted regions of hybridizing *H. marismortui* DNA restriction digests corresponding to the dashed fragment of fig. 1 and inserted them in M13 phage derived vectors for DNA sequencing. The DNA sequence so obtained was identical to that present in the EMBL3 recombinant phage. In addition, the size of the RNA band (about 1700 nucleotides) detected by hybridization with an internal gene probe (not shown) fits well with the size of the gene and the

position of putative transcription signals (Fig. 2). The point concerning hEF-Tu size is of interest because several halophilic proteins have apparent *M_r*s in SDS-PAGE greater than the corresponding non-halophilic proteins (11). Though it was speculated that the extra mass might be related to halophilic behaviour, in the case of hEF-Tu it is more likely that its migration in the gel is anomalous.

Figure 3 shows the alignment between hEF-Tu and yeast EF-1 α (12), *E. coli* (13, 14) and *M. vannielii* (15) EF-Tus. The highest value of identity (61%) is with *M. vannielii*, the only other archaeobacterium whose EF-Tu gene is known.

In table 1 the size, G+C content and amino acid composition of three EF-Tus and one EF-1 α are compared: hEF-Tu has the highest level of aspartic and glutamic acid residues and the lowest of lysine residues. Furthermore, the excess acidic residues in hEF-Tu result, in most cases, from mutations of Lys, Asn and Gln in the *M. vannielii* sequence (fig. 2). This result agrees with the previous observations of an increased number of acidic amino acid residues in proteins from halophiles (16).

The major advantage in studying how these primary structure features of hEF-Tu are reflected at higher structural levels is the knowledge of the X-ray diffraction structure of at least a part of EF-Tu from *E. coli* (17, 18). This allowed us to test the quality

of secondary structure prediction obtained with computer programs. An interesting way of analysis and prediction of protein structures was developed by Parker and coworkers: this method attempts to identify peptide segments located on the surface of a protein which are potential antigenic determinants (10). The linear combination of data from HPLC retention times (10), from calculations about peptide flexibility (19) and accessibility (20) applied to the EF-Tu from *E. coli* shows an overall correspondence to its tertiary structure. We analysed the two archaeobacterial EF-Tus with this approach. Not surprisingly, because of the large homology in primary structure, the overall secondary structure elements and the consequent organization of putative antigenic determinants are similar for both EF-Tus. However, figure 4 shows the presence of two additional predicted antigenicity peaks in the profile obtained with hEF-Tu. In addition, the vast majority of basic or neutral amino acid residues changing to aspartic or glutamic acid residues in hEF-Tu are located in the antigenicity peaks, often at the top.

DISCUSSION

The identification of the sequence shown in this paper as the structural gene for the elongation factor EF-Tu of *H. marismortui* essentially derives from the large correspondence with other elongation factors EF-Tus, from the identity of its 24 N-terminal predicted amino acid residues to the Edman degradation of the N-end of the peptide and from the presence of amino acid stretches related to the guanine nucleotide binding function (21, 22).

The positions of deletions allowing the alignment of hEF-Tu with the homologous peptides of fig. 3 and the respective

percentages of identity clearly indicate a closer relationship of the archaeobacterial EF-Tus with the EF-1 α from *S. cerevisiae* than with the *E. coli* EF-Tu. This observation supports the model of an independent evolutive position of archaeobacteria relatively to eubacteria and eukaryotes (23, 24).

Halophilic proteins are only stable in high concentrations of certain salts (1–3). To address the question of which features are responsible for the halophilic nature of hEF-Tu, the more appropriate comparison is with the primary structure of EF-Tu from *M. vannielii* which is the closest non-halophilic counterpart. The nearly identical size (428 versus 421 residues) and the high level of identity (61%) despite the great difference in G+C content of their genes (table 1) make these two peptides ideal candidates for investigating why one of them is halophilic and not the other.

The increased amount of acidic amino acid residues in halophilic proteins was observed early (16). It is also observed in hEF-Tu so that in the comparison with its counterpart from *M. vannielii* the locations of acidic residues have to be considered in particular. The part A of fig. 4 indicates positions of acidic residues specific to the hEF-Tu sequence and the corresponding residues in the *M. vannielii* peptide (see also fig. 3). With the exception of the aspartic acid residue at position 262 of hEF-Tu, only residues corresponding to 'new' acidic charges have been considered. However, the position in the antigenicity profile of D number 262 of hEF-Tu is on the top of a peak absent in *M. vannielii*. Fig. 4 shows that the vast majority of acidic amino acid residues specific to hEF-Tu are located in regions at the surface of the protein. In addition to the results of secondary structure computer prediction, a strong argument in favour of tertiary structure homology for elongation factors is that they are proteins with quite a complex function that requires them to interact with other elongation factors, GDP, GTP, tRNAs and ribosomes. Presently, the only elongation factor EF-Tu whose tertiary structure is known is the *E. coli* EF-Tu of which the GDP-binding domain is the best defined (25). The amino acid residues in *E. coli* EF-Tu corresponding to the hEF-Tu specific acidic residues shown in figure 4 were identified from the sequence alignment of fig. 3 and from secondary structure predictions of hEF-Tu obtained with computer programs tested with *E. coli* elongation factor EF-Tu. Their positions in the three-dimensional model of domain I of this protein are shown in fig. 5.

When compared to the *E. coli* sequence, the archaeobacterial and eukaryotic sequences show an insertion of 15 amino acids (positions 29 to 44 of hEF-Tu) where hEF-Tu shows two extra acidic charges compared to the *M. vannielii* sequence (fig. 3). The insertions and the adjacent region are predicted to be highly antigenic (fig. 4). The residues corresponding to the other changes between the two archaeobacterial sequences are not spread homogeneously but appear in patches on the protein surface. For example, residues 78 and 82 in hEF-Tu, correspond to residues 68 and 72 in *E. coli* protein (Fig. 3). In the structure, the alpha carbons of these residues are within 10 $\pm$ 2 angstrom of each other, but the distance to residues 187, 193 and 194 (corresponding to 176, 182 and 183 in *E. coli*), which form another patch, is much greater (Fig. 4). The same is true in the case of residues 139, 166 and 174 (corresponding to 129, 155 and 163 in *E. coli*). If our structural interpretation of the sequence comparison is correct, the presence of the acidic amino acid residues in patches at the surface of hEF-Tu could well be correlated with the halophilic nature of the protein. The other acidic amino acid residues specific for hEF-Tu cannot be easily viewed onto the *E. coli* structure because they correspond to

Figure 5. Schematic drawing of the tertiary structure of the GDP-binding domain of *E. coli* EF-Tu (adapted from ref. 17) with locations showing how amino acids mutated to acidic residues in hEF-Tu sequence should appear in patches on the protein surface. The part between amino acids 40 and 59 (asterisks) is missing from the structure. Arrows indicate β -strands and cylinders α -helices. Numbers indicate amino acid locations corresponding to acidic residues specific to hEF-Tu (Fig. 4). Dashed lines indicate acidic patches.

positions in domains II and III where secondary structure elements are less defined (25).

In a recent analysis of biochemical and biophysical experiments on halophilic malate dehydrogenase also from *H. marismortui*, a model was suggested for the stabilization of the protein in highly concentrated salt solutions, that also explains its instability when the solvent salt concentration is lowered (3). In the model, the protein is stabilized by cooperative interactions with hydrated salt ions that result in a larger number of solvent molecules associated with the folded protein than can be accounted for by the hydration of the amino acids when the protein is unfolded. Acidic residues bind about twice as many ordered water molecules as the other polar amino acids (26) and their putative location in patches on the surface of hEF-Tu is consistent with their participation in stabilization through local cooperative hydration networks with salt ions. It is also interesting to note that the patches are not in the vicinity of GDP-binding site (Fig. 5), as if the stabilization 'scaffolding' is located in such a way so as not to interfere with substrate binding. In conclusion, the comparison with other elongation factor EF-Tus shows that the halophilic character of hEF-Tu is due to specific changes in amino acid composition necessary for stabilization without major changes in the rest of the primary structure of the protein.

REFERENCES

1. Kushner, D.J. (1985) in Woese, C. and Wolfe, R. (eds.), *The Bacteria*. Academic Press Inc., New York, Vol III, pp.171–214.
2. Guinet, F., Rainer, F. and Leberman, R. (1988) *Eur. J. Biochem.*, **172**, 687–694.
3. Zaccari, G., Cendrin, F., Haik, Y., Borochoy, N. and Eisenberg, H. (1989) *J. Mol. Biol.*, **208**, 491–500.
4. Maniatis, T., Fritsch, E.F. and Sambrook, J., (1982) *Molecular cloning : A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.
5. Yokota, T., Sugisaki, H., Takanami, M. and Kaziro, Y. (1980) *Gene*, **12**, 25–31.
6. Frischauf, A.M., Lerach, H., Poutska, A. and Murray, N. (1983) *J. Mol. Biol.*, **170**, 827–842.
7. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. natn. Acad. Sci. U.S.A.*, **74**, 5463–5467.
8. Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) *Gene*, **33**, 103–119.
9. Gough, J.A. and Murray, N.E. (1983) *J. Mol. Biol.*, **166**, 1–19.
10. Parker, J.M.R., Guo, D. and Hodges, R.S. (1986) *Biochemistry*, **25**, 5425–5432.
11. Lechner, J. and Sumper, M. (1987) *J. Biol. Chem.*, **262**, 9724–9729.
12. Nagata, S., Nagashima, K., Tsunetsugu-Yokota, Y., Fujimura, K., Miyazaki, M. and Kaziro, Y. (1984) *Embo J.*, **3**, 1825–1830.
13. Arai, K., Clark, B.F.C., Duffy, L., Jones, M.D., Kaziro, Y., Laursen, R.A., L'Italien, J., Miller, D.L., Nagarkatti, S., Nakamura, S., Nielsen, K.M., Petersen, T.E., Takahashi, K. and Wade, M. (1980) *Proc. natn. Acad. Sci. U.S.A.*, **77**, 1326–1330.
14. An, G. and Friesen, J.D. (1980) *Gene*, **12**, 33–39.
15. Lechner, K. and Boeck, A. (1987) *Mol. Gen. Genet.*, **208**, 523–528.
16. Lanyi, J.K. (1974) *Bacteriol. Rev.*, **38**, 272–290.
17. Jurnak, F. (1985) *Science*, **230**, 32–36.
18. Lacour, T.F.M., Nyborg, J., Thirup, S. and Clark, B.F.C. (1985) *Embo J.*, **4**, 2385–2388.
19. Karplus, P.A. and Schulz, G.E. (1985) *Naturwissenschaften*, **72**, 212–213.
20. Janin, J. (1979) *Nature*, **277**, 491–492.
21. Dever, T.E., Glynnias, M.J. and Merrick, W.C. (1987) *Proc. natn. Acad. Sci. U.S.A.*, **84**, 1814–1818.
22. Wooley, P. and Clark, B.F.C. (1989) *Biotechnology*, **7**, 913–920.
23. Woese, C.R. and Olsen, G. J. (1986) *System. Appl. Microbiol.*, **7**, 161–167.
24. Zillig, W., Palm, P., Reiter, W.F., Gropp, F., Puehler, G. and Klenk, H.P. (1987) *Eur. J. Biochem.*, **173**, 473–482.
25. Nyborg, J. and La Cour, T. (1989) in Bosch, L., Kraal, B. and Parmeggiani, A. (eds.) *The guanine nucleotide binding proteins. Common structural and functional properties*. Nato ASI Series A : Life Sciences. Plenum Press New York and London. Vol 65, pp. 3–14.
26. Saenger, W. (1987) *Ann. Rev. Biophys. Chem.*, **16**, 93–114.