Eur. J. Biochem. *170*, 93–98 (1987)

Molecular cloning and sequence determination of the *tuf* gene coding for the elongation factor Tu of *Thermus thermophilus* HB8

Akira KUSHIRO, Masato SHIMIZU and Ken-ichi TOMITA

Faculty of Pharmaceutical Sciences, Osaka University

(Received June 2/August 17, 1987) – EJB 87 0631

The *tuf* gene, which encodes the elongation factor Tu (EF-Tu) of *Thermus thermophilus* HB8, and its flanking regions were cloned and sequenced. The gene encoding EF-G was found upstream of the 5′ end of the *tuf* gene. The *tuf* gene of *T. thermophilus* HB8 had a very high G+C content and 84.5% of the third base in codon usage was either G or C. The deduced primary structure of the EF-Tu was composed of 405 amino acid residues with a M_r = 44658. A comparison of the amino acid sequence of EF-Tu from *T. thermophilus* HB8 with those of *Escherichia coli* and *Saccharomyces cerevisiae* mitochondria showed a very high sequence homology (65–70%). Two Cys residues out of the three found in *E. coli* EF-Tu had been replaced with Val in *T. thermophilus* HB8 EF-Tu. An extra amino acid sequence of ten residues, consisting predominantly of basic amino acids (Met-182–Gly-191), which does not occur in EF-Tu of *E. coli*, was found in *T. thermophilus* HB8.

Elongation factor Tu (EF-Tu), which elongates polypeptide chains during protein biosynthesis, is known to be a multifunctional protein. It plays a major role in the recognition, transport and positioning of the codon-specified aminoacyl-tRNA onto the A site of the ribosome [1, 2]. EF-Tu interacts with GDP, GTP, tRNA, ribosome and EF-Ts, and is also known to be a subunit of bacteriophage Qβ replicase [3] and target protein of the antibiotic kirromycin [4]. We have been interested in studying the three-dimensional structure of EF-Tu. Since thermal and chemical stability of EF-Tu from *Escherichia coli* is poor, the molecule seems to be unsuitable for X-ray crystallographic experiments. The three-dimensional structure of the trypsin-digested and self-digested *E. coli* EF-Tu has been determined [5, 6]. However, the digested EF-Tu cannot bind to tRNA and the EF-Tu structure around the tRNA-binding site is still obscure. Proteins extracted from *Thermus thermophilus* HB8 are known to be active at very high temperatures [7, 8]. EF-Tu from *T. thermophilus* HB8 has been shown to be fully active at 60 °C and also to have high chemical stability [9–11]. So a three-dimensional crystallographic study of its structure may help to solve the structure around the tRNA-binding site in order to elucidate the important mechanism of protein-nucleic acid interaction. The EF-Tu of *E. coli* is encoded by the *tufA* gene and *tufB* gene [12], and that of *Saccharomyces cerevisiae* mitochondria by the *tufM* [13]. All these *tuf* genes were cloned [14, 15] and their nucleotide sequences were also determined [16, 17].

In this paper we describe the molecular cloning of the *tuf* gene from *T. thermophilus* HB8, using the *tufA* gene of *E. coli* as a probe. We also describe the determination of its complete nucleotide sequence as well as the deduced amino acid sequence, the G+C content and codon usage. The amino acid sequence of EF-Tu from this study is compared with that reported for *E. coli* [18, 19].

MATERIALS AND METHODS

Bacterial strains and plasmid

T. thermophilus HB8 (ATCC 27634) was grown at 70 °C in 0.3% Bacto Peptone (Difco), 0.5% yeast extract (Difco), 0.1% glucose, and 0.2% NaCl (pH 7.0). The plasmid pTUA1 was a generous gift from Y. Kaziro (University of Tokyo). It had been obtained by the insertion of the *tufA* gene of *E. coli* into the plasmid RSF2124 [14].

Enzymes and other materials

All the restriction endonucleases, the T4 DNA ligase and the terminal deoxyribonucleotidyltransferase were purchased from Takara Shuzo Co. (Kyoto). Proteinase K and lysozyme from egg white were obtained from Sigma Company. Alkaline phosphatase from calf intestine (CIP) was obtained from Boehringer Mannheim. [α-^{32}P]dCTP (400 Ci/mmol) and [α-^{32}P]ddATP (3000 Ci/mmol) were purchased from Amersham (Amersham Japan). The '7-deaza sequencing' kit was obtained from Takara Shuzo Co. (Kyoto).

Southern hybridization

Chromosomal DNA of *T. thermophilus* HB8 was extracted from cells with lysis in a 1% SDS/NaCl solution at 60 °C for 10 min. Proteinase K was added to a final concentration of 1 mg/ml. After extraction twice with phenol, RNase A was added to a final concentration of 50 mg/ml, and chromosomal DNA was again extracted with phenol. The chromosomal DNA was recovered by precipitation with isopropanol and then completely digested with *Sma*I, *Bam*HI and *Eco*RI, and the fragments were separated on a 0.7% agarose gel. The

Correspondence to K. Tomita, Faculty of Pharmaceutical Sciences, Osaka University, 1-6 Yamadaoka, Suita, Osaka, Japan 565

Abbreviations. EF, elongation factor; ddATP, 2′,3′-dideoxyadenosine 5′-triphosphate.

Enzymes (IUB Recommendations, 1984). Restriction endonucleases *Alu*I, *Bam*HI, *Eco*RI, *Hae*III, *Hin*cII, *Kpn*I, *Sma*I (EC 3.1.21.4); T4 DNA ligase (EC 6.5.1.1); terminal deoxyribonucleotidyltransferase (EC 2.7.7.31); calf intestinal alkaline phosphatase (EC 3.1.3.1); proteinase K (EC 3.4.21.14); ribonuclease A (EC 3.1.27.5); lysozyme (EC 3.2.1.17).

separated fragments were then transferred to a nitrocellulose filter (0.45 μm pore, Toyo Roshi Co., Tokyo) as described [20].

The plasmid pTUA1 was digested with *Sma*I. The 0.3-kb *Sma*I fragment, corresponding to the amino acid sequence of Pro-82 – Pro-163 of *E. coli* EF-Tu, was used as a probe by labelling its 3′ end with [α-^{32}P]ddATP using terminal deoxyribonucleotidyltransferase as described [21].

Construction of the genomic library

Chromosomal DNA of *T. thermophilus* HB8 was partially digested with *Sma*I to obtain the main bands of length less than 10 kb involving 1.5-kb and 2.0-kb fragments described later. The plasmid pUC19 [22] was also digested with *Sma*I and dephosphorylated with alkaline phosphatase. The chromosomal DNA digested with *Sma*I was ligated to the *Sma*I-digested pUC19 with T4 ligase and used in transforming *E. coli* HB101.

Screening of the genomic library and sequence determination of the inserted tuf *gene*

The transformants were screened by colony hybridization with the probe described earlier [23]. After the isolation of the plasmid DNA from the positive colony, the inserted chromosomal DNA fragment was recovered. The restriction maps were determined by a single and double digestion of the inserted DNA with various endonucleases and the fragments were cloned to the M13 vectors mp18 and mp19 [22]. The DNA sequencing was carried out by the dideoxy-chain-termination method using the 7-deaza sequencing kit and [α-^{32}P]dCTP [24, 25]. {The 7-deaza sequencing kit contains 2′-deoxy-7-deazaguanosine triphosphate (dc^{7}GTP) instead of 2′-deoxyguanosine triphosphate (dGTP) [26].}

RESULTS AND DISCUSSION

DNA sequence homology between E. coli *and* T. thermophilus *HB8*

From the Southern hybridization analysis, two bands from the chromosomal DNA digested with *Bam*HI were positive. The fragments were of length 4.5 kb and 7.0 kb. Similarly, chromosomal DNA digested with *Sma*I produced two positive bands of length 1.5 kb and 2.0 kb. These results seems to suggest that the chromosomal DNA of *T. thermophilus* HB8 has one or two genes which are homologous to the *tufA* gene of *E. coli*.

Isolation of the DNA fragment from T. thermophilus *HB8 which was homologous to the* tufA *gene*

As a result of screening the genomic library by colony hybridization, we obtained a strong positive signal. This clone, which had a plasmid labelled as pHBTU31, contained the 1.5-kb *Sma*I fragment of chromosomal DNA of *T. thermophilus* HB8. Fig. 1 shows the restriction map of pHBTU31 and Fig. 2 its nucleotide sequence. The open reading frame, which begins with ATG (position 1) and terminates with TGA (position 1219), probably encodes a single polypeptide of 405 amino acid residues of 44658 Da. The five amino acid residues at the amino terminus deduced from this open reading frame were completely identical to those of EF-Tu purified from *T.*

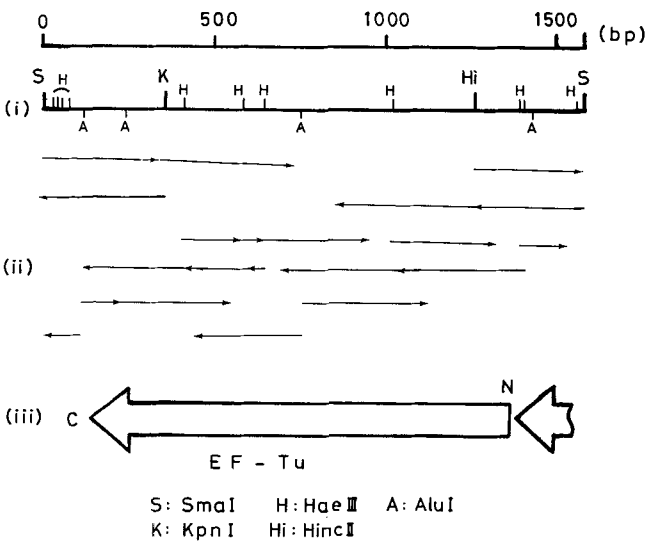

Fig. 1. *Restriction map of the* tuf *gene of* T. thermophilus *HB8*. (i) Restriction sites of the 1.5-kb fragment inserted in the plasmid pHBTU31. (ii) The fragments that were cloned in the M13 mp18 and mp19 vectors are shown. The direction of sequencing is in the direction of the arrowhead. (iii) The open box shows the sequence of the coding region and the direction of transcription

thermophilus HB8 (S. Yokoyama and T. Miyazawa, personal communication).

The amino acid composition predicted from this open reading frame was also in good agreement with that reported for *E. coli* EF-Tu as shown in Table 1 [11]. A comparison of the amino acid sequence deduced from this open reading frame with that of *E. coli* EF-Tu revealed a 70% homology. The amino acid sequences at the active site were highly conserved.

From the above results we wish to state that this open reading frame actually encodes the EF-Tu protein of *T. thermophilus* HB8.

A Shine-Dalgarno sequence [27], AGGAGGA, complementary to the 3′ end of 16S ribosomal RNA, was found six bases upstream of the initiation codon (Fig. 2). A promotor region was, however, not detected. To determine what the putative open reading frame upstream of the *tuf* gene encodes, an analysis of the protein data base was made. This analysis revealed a high amino acid sequence homology with that of *E. coli* EF-G [28]. In *E. coli*, *tufA* is cotranscribed sequentially as *rpsL-rpsG-fus-tufA* (*rpsL* and *rpsG* encode the ribosomal proteins S12 and S7 respectively, and *fus* codes EF-G) [29]. It is, therefore, very likely that the *tuf* gene of *T. thermophilus* HB8 may be cotranscribed with the *fus* gene or the other genes.

Codon usage and nucleotide compositions of the tuf *gene of* T. thermophilus *HB8*

The overall G + C content in the *tuf* gene was 62.9%, while the percentage of the third base of the codons being G or C, was 84.5%. This high G + C content has also been shown to be common in the thermophilic bacteria [30, 31]. The codon usage is shown in Table 2. The high percentage G + C content tends to stabilize the DNA structure even at high temperatures.

The third base of the codon for Phe or Ile was U instead of G or C, i.e. the codons for Phe and Ile were UUU and

```
                                (-222) 5'-GGGGCAACGCAGGTGATCCGGGCCTTCGTG  -193

          CCCTTGGCGGAGATGTTCGGCTACGCCACGGACCTGCGCTCCAAGACCCAGGGGCGAGGC  -133

          TCCTTCGTCATGTTCTTTGACCACTACCAGGAGGTGCCCAAGCAGGTCCAGGAGAAGCTC   -73

          ATCAAGGGTCAATAGCGCGGCTTCGGGGTGGGCCTTTTAGGGCCCGCCCCTTCCTTGCG   -13

E.coli                        Ser Lys Glu Lys Phe Glu Arg Thr Lys Pro His Val Asn Val Gly   15
HB8                       Met Ala Lys Gly Glu Phe Val Arg Thr Lys Pro His Val Asn Val Gly   15
          AGG AGG ACG GAG ATG GCG AAG GGC GAG TTT GTT CGG ACG AAG CCT CAC GTG AAC GTG GGG   48

E.coli    Thr Ile Gly His Val Asp His Gly Lys Thr Thr Leu Thr Ala Ala Ile Thr Thr Val Leu   35
HB8       Thr Ile Gly His Val Asp His Gly Lys Thr Thr Leu Thr Ala Ala Leu Thr Tyr Val Ala   35
          ACG ATT GGG CAC GTG GAC CAC GGG AAG ACG ACG CTG ACG GCG GCG TTG ACG TAT GTG GCG  108

E.coli    Ala Lys Thr Tyr Gly Gly Ala Ala Arg Ala Phe Asp Gln  -  Ile Asp Asn Ala Pro Glu   54
HB8       Ala Ala Glu Asn Pro Asn Val Glu Val Lys Asp Tyr Gly Asp Ile Asp Lys Ala Pro Glu   55
          GCG GCG GAG AAC CCG AAT GTA GAG GTT AAG GAC TAC GGG GAC ATT GAC AAG GCG CCG GAG  168

E.coli    Glu Lys Ala Arg Gly Ile Thr Ile Asn Thr Ser His Val Glu Tyr Asp Thr Pro Thr Arg   74
HB8       Glu Arg Ala Arg Gly Ile Thr Ile Asn Thr Ala His Val Glu Tyr Glu Thr Ala Lys Arg   75
          GAG CGT GCG CGG GGG ATT ACG ATC AAC ACG GCG CAC GTG GAG TAC GAG ACG GCG AAG CGG  228

E.coli    His Tyr Ala His Val Asp Cys Pro Gly His Ala Asp Tyr Val Lys Asn Met Ile Thr Gly   94
HB8       His Tyr Ser His Val Asp Cys Pro Gly His Ala Asp Tyr Ile Lys Asn Met Ile Thr Gly   95
          CAC TAT TCC CAC GTG GAT TGC CCT GGG CAC GCG GAC TAC ATC AAG AAC ATG ATC ACG GGT  288

E.coli    Ala Ala Gln Met Asp Gly Ala Ile Leu Val Val Ala Ala Thr Asp Gly Pro Met Pro Gln  114
HB8       Ala Ala Gln Met Asp Gly Ala Ile Leu Val Val Ser Ala Ala Asp Gly Pro Met Pro Gln  115
          GCC GCG CAG ATG GAC GGG GCG ATC CTT GTG GTG TCG GCG GCG GAC GGG CCG ATG CCG CAG  348

E.coli    Thr Arg Glu His Ile Leu Leu Gly Arg Gln Val Gly Val Pro Tyr Ile Ile Val Phe Leu  134
HB8       Thr Arg Glu His Ile Leu Leu Ala Arg Gln Val Gly Val Pro Tyr Ile Val Val Phe Met  135
          ACG CGG GAG CAC ATT TTG CTG GCG CGG CAG GTG GGG GTG CCG TAC ATT GTG GTG TTC ATG  408

E.coli    Asn Lys Cys Asp Met Val Asp Asp Glu Glu Leu Leu Glu Leu Val Glu Met Glu Val Arg  154
HB8       Asn Lys Val Asp Met Val Asp Asp Pro Glu Leu Asp Leu Val Glu Met Glu Val Arg  155
          AAC AAG GTG GAC ATG GTG GAC GAC CCC GAG TTG CTG GAC CTG GTG GAG ATG GAG GTG CGG  468

E.coli    Glu Leu Leu Ser Gln Tyr Asp Phe Pro Gly Asp Asp Thr Pro Ile Val Arg Gly Ser Ala  174
HB8       Asp Leu Leu Asn Gln Tyr Glu Phe Pro Gly Asp Glu Val Pro Val Ile Arg Gly Ser Ala  175
          GAC CTT TTG AAC CAG TAC GAG TTT CCT GGG GAC GAG GTT CCG GTG ATT CGG GGG AGT GCT  528

E.coli    Leu Lys Ala Leu Glu Gly  -  -  -  -  -  -  -  -  -  Asp Ala Glu Trp  184
HB8       Leu Leu Ala Leu Glu Gln Met His Arg Asn Pro Lys Thr Arg Arg Gly Glu Asn Glu Trp  195
          CTT TTG GCG CTT GAG CAG ATG CAC AGG AAC CCG AAG ACG AGG CGT GGG GAG AAC GAG TGG  588

E.coli    Glu Ala Lys Ile Leu Glu Leu Ala Gly Phe Leu Asp Ser Tyr Ile Pro Glu Pro Glu Arg  204
HB8       Val Asp Lys Ile Trp Glu Leu Leu Asp Ala Ile Asp Glu Tyr Ile Pro Thr Pro Val Arg  215
          GTG GAC AAG ATT TGG GAG CTG TTG GAC GCG ATT GAC GAG TAC ATT CCC ACG CCG GTG CGG  648

E.coli    Ala Ile Asp Lys Pro Phe Leu Leu Pro Ile Glu Asp Val Phe Ser Ile Ser Gly Arg Gly  224
HB8       Asp Val Asp Lys Pro Phe Leu Met Pro Val Glu Asp Val Phe Thr Ile Thr Gly Arg Gly  235
          GAC GTG GAC AAG CCG TTC TTG ATG CCG GTG GAG GAC GTG TTT ACG ATC ACG GGT CGT GGG  708

E.coli    Thr Val Val Thr Gly Arg Val Glu Arg Gly Ile Ile Lys Val Gly Glu Glu Val Glu Ile  244
HB8       Thr Val Ala Thr Gly Arg Ile Glu Arg Gly Lys Val Lys Val Gly Asp Glu Val Glu Ile  255
          ACG GTG GCC ACG GGT CGG ATT GAG CGG GGC AAG GTG AAG GTT GGG GAC GAG GTG GAG ATT  768

E.coli    Val Gly Ile Lys  -  Glu Thr Gln Lys Ser Thr Cys Thr Gly Val Glu Met Phe Arg Lys  263
HB8       Val Gly Leu Ala Pro Glu Thr Arg Arg Thr Val Val Thr Gly Val Glu Met His Arg Lys  275
          GTG GGC CTT GCT CCG GAG ACG CGG AGG ACG GTG GTG ACG GGT GTG GAG ATG CAC CGG AAG  828

E.coli    Leu Leu Asp Glu Gly Arg Ala Gly Glu Asn Val Gly Val Leu Leu Arg Gly Ile Lys Arg  283
HB8       Thr Leu Gln Glu Gly Ile Ala Gly Asp Asn Val Gly Val Leu Leu Arg Gly Val Ser Arg  295
          ACC TTG CAG GAG GGG ATT GCT GGG GAC AAT GTG GGG GTG CTC CTG CGG GGT GTG AGC CGG  888

E.coli    Glu Glu Ile Glu Arg Gly Gln Val Leu Ala Lys Pro Gly Thr Ile Lys Pro His Thr Lys  303
HB8       Glu Glu Val Glu Arg Gly Gln Val Leu Ala Lys Pro Gly Ser Ile Thr Pro His Thr Lys  315
          GAG GAG GTG GAG CGG GGG CAG GTG CTG GCG AAG CCT GGG AGC ATT ACG CCG CAC ACG AAG  948

E.coli    Phe Glu Ser Glu Val Tyr Ile Leu Ser Lys Asp Glu Gly Gly Arg His Thr Pro Phe Phe  323
HB8       Phe Glu Ala Ser Val Tyr Val Leu Lys Lys Glu Glu Gly Gly Arg His Thr Gly Phe Phe  335
          TTT GAG GCC TCG GTG TAT GTG TTG AAG AAG GAG GAG GGT GGA CGG CAC ACG GGG TTT TTT 1008

E.coli    Lys Gly Tyr Arg Pro Gln Phe Tyr Phe Arg Thr Thr Asp Val Thr Gly Thr Ile Glu Leu  343
HB8       Ser Gly Tyr Arg Pro Gln Phe Tyr Phe Arg Thr Thr Asp Val Thr Gly Val Val Gln Leu  355
          TCG GGG TAC CGT CCG CAG TTT TAC TTT CGG ACG ACG GAC GTG ACG GGG GTG GTG CAG TTG 1068

E.coli    Pro Glu Gly Val Glu Met Val Met Pro Gly Asp Asn Ile Lys Met Val Val Thr Leu Ile  363
HB8       Pro Pro Gly Val Glu Met Val Met Pro Gly Asp Asn Val Thr Phe Thr Val Glu Leu Ile  375
          CCT CCG GGC GTG GAG ATG GTG ATG CCT GGG GAC AAC GTG ACG TTT ACG GTG GAG CTG ATC 1128

E.coli    His Pro Ile Ala Met Asp Asp Gly Leu Arg Phe Ala Ile Arg Glu Gly Gly Arg Thr Val  383
HB8       Lys Pro Val Ala Leu Glu Glu Gly Leu Arg Phe Ala Ile Arg Glu Gly Gly Arg Thr Val  395
          AAG CCG GTG GCG CTG GAG GAG GGT TTG CGG TTT GCC ATC CGT GAG GGT GGG CGG ACC GTG 1188

E.coli    Gly Ala Gly Val Val Ala Lys Val Leu Gly                                        393
HB8       Gly Ala Gly Val Val Thr Lys Ile Leu Glu                                        405
          GGC GCC GGC GTC GTC ACC AAG ATC CTG GAG TGA                                   1221

          GGTGAGGTATGCCCAAGATCCGCATCAAGCTCCGGGGTTTTGACCACAAGACCCTGGACG          1281

          CCTCGGCCCAGAAGATCGTGGAGGCGGCCCGGCGTTCCGGGGCCCAGGTCTCCGGCCCCA         1341

          TCCCCCTACCCACCC-3'   (1356)
```

Fig. 2. *Comparison of the amino acid sequence of the EF-Tu of* E. coli *and* T. thermophilus *HB8*. Below the amino acid sequence is shown the nucleotide sequence. The reading frame starts at ATG (position 1) and terminates at TGA (position 1219). On the top and bottom of the main amino acid sequence of EF-Tu are shown the flanking regions. The underlined nucleotide sequence is the Shine-Dalgarno sequence

AUU respectively. These codons are, however, not desirable for thermal stability. *T. thermophilus* HB8 produces the restriction endonuclease, *Tth*HBI, with a recognition sequence TCGA [32]. If UUC or AUC is used as the codon for Phe or Ile, the TCGA sequence will appear; for instance, TTCGAN for Phe-Asp or Glu, and ATCGAN for Ile-Asp or Glu where N indicates any of the four possible bases. In general, bacteria methylate their own DNA by methylase in order to protect themselves from being digested by their own restriction endonuclease. *T. thermophilus* HB8 in particular protects its DNA by not only methylation [33], but also by avoiding the TCGA sequence. As expected, no TCGA sequence was observed in the 1578-bp fragment in this work. The avoidance of the TCGA sequence by *T. caldophilus* GK24 has also been reported [34].

Table 1. *Amino acid composition of EF-Tu from* T. thermophilus *HB8*

Amino acid	Number of residues	Amino acid composition	
		this work	Y. Kaziro et al. [11]
		%	
Ala	28	6.91	6.90
Arg	27	6.67	6.15
Asn + Asp	35	8.65	8.59
Cys	1	0.25	0.23
Gln + Glu	46	11.36	12.57
Gly	39	9.63	9.90
His	12	2.96	2.90
Ile	21	5.19	5.56
Leu	27	6.67	6.75
Lys	20	4.94	4.41
Met	11	2.72	2.77
Phe	12	2.96	2.97
Pro	23	5.68	5.67
Ser	7	1.73	1.81
Thr	31	7.65	7.54
Trp	2	0.49	0.39
Tyr	11	2.72	2.69
Val	52	12.84	12.20
Total	405	100	100

Comparison of the amino acid sequence of EF-Tu of T. thermophilus *HB8 with other EF-Tu proteins and GTP-binding proteins*

A comparison of the amino acid sequence of EF-Tu of *T. thermophilus* HB8 with other EF-Tu proteins and GTP-binding proteins is shown in Table 3. Like those reported for the yeast RAS proteins [35], some homologous regions were highly conserved in *T. thermophilus* HB8 EF-Tu. Comparing our results with those of the X-ray structure of *E. coli* EF-Tu [5, 6], we would like to propose the functions of some homologous regions of *T. thermophilus* HB8 EF-Tu molecule.

a) Part 1 of Table 3 might correspond to the loop which connects the β-1 sheet to the α-A helix in *E. coli* EF-Tu. Gly-Xaa-Xaa-Xaa-Xaa-Gly-Lys is a common sequence found in many purine-nucleotide-binding proteins [5]. This loop might interact with the phosphate group of GDP.

b) Part 2 might correspond to the β-2 sheet [28]. Although this region does not interact with GDP from X-ray analysis, we would like to propose that this region plays an important role in the binding with the phosphate group of tRNA since it is stereochemically near by Cys-81, which is considered to be one of the amino acid residues that interact with tRNA in *E. coli* [36, 37].

c) Part 3 may be the loop which connects the β-3 sheet to the α-B helix. The Cys residue in this region, which is proposed to be a tRNA-binding site [36, 37], is the only one found commonly in the three EF-Tu proteins. The other GTP-binding proteins have been reported to have Thr residue instead of Cys [35]. It is also very likely that the thiol group of Cys is involved in the binding of EF-Tu to tRNA. To solve this problem the site-directed mutation by protein engineering is now being carried out in our laboratory.

d) Part 4 corresponds to the loop connecting the β-5 sheet to the α-D helix. From chemical modification studies it has been reported that Cys-137 may interact with GDP [36, 37]. However, Asn-135 and Asp-138, but not Cys-137, form hydrogen bonds to the guanine base in the crystal structure of *E. coli* EF-Tu · GDP complex. In the EF-Tu protein from *T.*

Table 2. *Codon usage in the* tuf *gene of* T. thermophilus *HB8*
The numbers in parentheses represent the codon usage in the *tufA* gene of *E. coli* [16]

	U			C			A			G		
U	UUU 10	(1)		UCU –	(7)		UAU 3	(1)		UGU –	(1)	
	UUC 2	(13)	Phe	UCC 1	(3)		UAC 8	(9)	Tyr	UGC 1	(2)	Cys
	UUA –	(–)		UCA –	(–)		UAA –	(1)		UGA 1	(–)	Ter
	UUG 11	(–)	Leu	UCG 3	(–)	Ser	UAG –	(–)	Ter	UGG 2	(1)	Trp
C	CUU 5	(1)		CCU 6	(–)		CAU –	(2)		CGU 5	(20)	
	CUC 1	(1)		CCC 2	(–)		CAC 12	(9)	His	CGC –	(3)	
	CUA –	(–)		CCA –	(1)		CAA –	(–)		CGA –	(–)	
	CUG 10	(27)	Leu	CCG 15	(19)	Pro	CAG 9	(8)	Gln	CGG 19	(–)	Arg
A	AUU 13	(3)		ACU –	(12)		AAU 2	(–)		AGU 1	(–)	
	AUC 8	(26)		ACC 3	(15)		AAC 9	(7)	Asn	AGC 2	(1)	Ser
	AUA –	(–)	Ile	ACA –	(2)		AAA –	(17)		AGA –	(–)	
	AUG 11	(10)	Met	ACG 28	(1)	Thr	AAG 20	(6)	Lys	AGG 3	(–)	Arg
G	GUU 4	(22)		GCU 3	(11)		GAU 1	(4)		GGU 8	(19)	
	GUC 2	(1)		GCC 5	(1)		GAC 23	(20)	Asp	GGC 6	(20)	
	GUA 1	(11)		GCA –	(6)		GAA –	(30)		GGA 1	(–)	
	GUG 45	(3)	Val	GCG 20	(9)	Ala	GAG 37	(7)	Glu	GGG 24	(1)	Gly

Table 3. *Comparison of amino acid sequences of EF-Tu proteins and GTP-binding proteins*
The references for the sequence data are *E. coli* EF-Tu [18], *S. cerevisiae* mitochondria EF-Tu [13], *E. coli* EF-G [28], *E. coli* IF-2 [42], *E. coli* LepA [43], *S. cerevisiae* RAS1 [44] and human Ha-ras [45]

Site	Source	Protein	Sequence
1. Phosphate-binding site	*T. thermophilus* HB8	EF-Tu	Gly-His-Val-Asp-His-Gly (18–23)
	E. coli	EF-Tu	Gly-His-Val-Asp-His-Gly (18–23)
	S. cerevisiae	mt EF-Tu	Gly-His-Val-Asp-His-Gly (54–59)
	E. coli	EF-G	Ala-His-Ile-Asp-Ala-Gly (17–22)
	E. coli	IF-2	Gly-His-Val-Asp-His-Gly (398–403)
	E. coli.	LepA	Ala-His-Ile-Asp-His-Gly (11–16)
	human	Ha-ras	Gly-Ala-Gly-Gly-Val-Gly (10–15)
	S. cerevisiae	RAS1	Gly-Gly-Gly-Gly-Val-Gly (17–22)
2. Phosphate-binding site	*T. thermophilus* HB8	EF-Tu	Arg-Gly-Ile-Thr-Ile (59–63)
	E. coli	EF-Tu	Arg-Gly-Ile-Thr-Ile (58–62)
	S. cerevisiae	mt EF-Tu	Arg-Gly-Ile-Thr-Ile (94–98)
	E. coli	EF-G	Arg-Gly-Ile-Thr-Ile (58–62)
	E. coli	IF-2	Gly-Gly-Ile-Thr-Gln (422–426)
	E. coli	LepA	Arg-Gly-Ile-Thr-Ile (50–54)
3. tRNA-binding site	*T. thermophilus* HB8	EF-Tu	Asp-Cys-Pro-Gly-His (81–85)
	E. coli	EF-Tu	Asp-Cys-Pro-Gly-His (80–84)
	S. cerevisiae	mt EF-Tu	Asp-Cys-Pro-Gly-His (116–120)
	E. coli	EF-G	Asp-Thr-Pro-Gly-His (88–92)
	E. coli	IF-2	Asp-Thr-Pro-Gly-His (444–448)
	E. coli	LepA	Asp-Thr-Pro-Gly-His (77–81)
	human	Ha-ras	Asp-Thr-Ala-Gly-Gln (57–61)
	S. cerevisiae	RAS1	Asp-Thr-Ala-Gly-Gln (64–68)
4. Guanine-base-binding site	*T. thermophilus* HB8	EF-Tu	Asn-Lys-Val-Asp (136–139)
	E. coli	EF-Tu	Asn-Lys-Cys-Asp (135–138)
	S. cerevisiae	mt EF-Tu	Asn-Lys-Val-Asp (171–174)
	E. coli	EF-G	Asn-Lys-Met-Asp (142–145)
	E. coli	IF-2	Asn-Lys-Ile-Asp (498–501)
	E. coli	LepA	Asn-Lys-Ile-Asp (131–134)
	human	Ha-ras	Asn-Lys-Cys-Asp (116–119)
	S. cerevisiae	RAS1	Asn-Lys-Leu-Asp (123–126)

thermophilus HB8 or yeast mitochondria, the Cys-137 found in *E. coli* EF-Tu is replaced with Val implying that the Cys residue is not always essential for the GDP-binding activity. It would be interesting to study the effect of replacing Cys with Val on the GDP-binding activity of *E. coli* EF-Tu.

The homologous regions of EF-Tu and EF-G were found only at the amino-terminal end of the protein which plays a major role in its interaction with GDP [28].

No major differences were observed in the amino acid compositions of *E. coli* and *T. thermophilus* HB8 (Table 4). The replacement of Ile by Val, Asp by Glu, and Ser by Thr in *T. thermophilus* HB8 was also observed in a comparison of amino acid sequences of several proteins between the mesophilic bacteria and the thermophilic bacteria [38]. The favorable codon for Ile is AUU, but this codon is not suitable for thermal stability. Val, having GUG as its favorable codon, is rather similar to Ile with respect to chemical structure and hydrophobicity. We believe that the replacement of Ile with Val has happened to increase the G+C content to gain higher thermal stability. The only difference in the chemical structures between Ser and Thr, and between Asp and Glu, is a methyl or methylene group respectively. We cannot explain at this stage these minor changes relative to the thermal stability

Table 4. *A comparison of the amino acid residue replacements in EF-Tu of* E. coli *and* T. thermophilus *HB8*
The numbers of replacements occurring more than twice are shown

Amino acid		Number
E. coli	*T. thermophilus* HB8	
Ile	Val	11
Asp	Glu	7
Glu	Asp	4
Val	Ile	3
Glu	Val	3
Ser	Thr	3
Thr	Val	3
Ile	Leu	2
Ser	Ala	2
Lys	Ala	2
Lys	Thr	2
Ala	Ser	2
Leu	Met	2
Cys	Val	2
Glu	Pro	2
Lys	Ser	2
Ala	Asp	2

of *T. thermophilus* HB8 EF-Tu. The reason for these replacements would be much clearer if we could solve the three-dimensional structure of EF-Tu from *T. thermophilus* HB8.

E. coli EF-Tu has three Cys residues, but in *T. thermophilus* HB8 EF-Tu two of these three Cys residues are replaced with Val. The content of Cys residues in the thermophilic bacteria has been reported to be lower than that of the mesophilic bacteria [31, 39]. We think that the thermophilic bacteria have a low preference for Cys residues because of the reactive thiol group.

An additional amino acid sequence, Met-182–Gly-191, was found in the EF-Tu of *T. thermophilus* HB8 (see Fig. 2). This sequence has a high hydrophilicity [40] and basicity. Of the ten residues, five were basic amino acids. This extra sequence probably takes a reverse-turn structure [41]. In *E. coli* EF-Tu the neighboring region was the loop which connected the α-E helix with the α-F helix. Presumably this additional loop may be exposed to solvent and also participate in the binding of *T. thermophilus* HB8 EF-Tu to negatively charged phosphate groups. It would be most interesting to study the effect of the deletion of this loop by genetic mutation on the activity of the EF-Tu of *T. thermophilus* HB8.

Work is presently underway with the expression of the *tuf* gene of *T. thermophilus* HB8 in *E. coli* and the determination of its structure by X-ray crystallography.

We wish to render our sincere gratitude to Professor Y. Kaziro, University of Tokyo, for his generous gift of the *tufA* and *tufB* genes and his encouragement. We wish to thank also Professor T. Oshima, Tokyo Institute of Technology, Professor A. Nakata, Osaka University, Dr Y. Ishino, Takara Shuzo Co., and Dr K. Yagi, Osaka University, for their helpful discussions. We also thank Dr T. Hakoshima, Osaka University, for reading the manuscript and valuable discussions. We are grateful to Dr N. Tanaka, Institute for Protein Research, for his kind help on the use of the computer programs at his institute. Last, but not least, we wish to thank Professor T. Miyazawa and Dr S. Yokoyama, University of Tokyo, for sending us their data on the N-terminal amino acid sequence of EF-Tu from *T. thermophilus* HB8 before its publication. This work was supported by a grant-in-aid for specially promoted research (59065006), from the Ministry of Education, Science and Culture of Japan, and a grant for Japan – U.S.A. Cooperative Research Projects on EF-Tu from the Japan Society for the Promotion of Science.

REFERENCES

1. Lucas-Lenard, J. & Lipmann, F. (1971) *Annu. Rev. Biochem. 40*, 409–448.
2. Kaziro, Y. (1978) *Biochim. Biophys. Acta 505*, 95–127.
3. Blumenthal, T. & Carmichael, G. G. (1979) *Annu. Rev. Biochem. 48*, 525–548.
4. Duisterwinkel, F. J., De Graaf, J. M., Schretlen, P. J. M., Kraal, B. & Bosch, L. (1981) *Eur. J. Biochem. 117*, 7–12.
5. la Cour, T. F. M., Nyborg, J., Thirup, S. & Clark, B. F. C. (1985) *EMBO J. 4*, 2385–2388.
6. Jurnak, F. (1985) *Science (Wash. DC) 230*, 32–36.
7. Oshima, T. & Imahori, K. (1974) *Int. J. Syst. Bacteriol. 24*, 102–112.
8. Nojima, H., Ikai, A., Oshima, T. & Noda, H. (1977) *J. Mol. Biol. 116*, 429–442.
9. Arai, K., Ota, Y., Arai, N., Nakamura, S., Henneke, C., Oshima, T. & Kaziro, Y. (1978) *Eur. J. Biochem. 92*, 509–519.
10. Arai, K., Arai, N., Nakamura, S., Oshima, T. & Kaziro, Y. (1978) *Eur. J. Biochem. 92*, 521–531.
11. Nakamura, S., Ohta, S., Arai, K., Arai, N., Oshima, T. & Kaziro, Y. (1978) *Eur J. Biochem. 92*, 533–543.
12. Jaskunas, S. R., Lindahl, L., Nomura, M. & Burgess, R. R. (1975) *Nature (Lond.) 257*, 458–462.
13. Nagata, S., Yokota, Y., Yokota, T., Naito, A. & Kaziro, Y. (1983) *Proc. Natl Acad. Sci. USA 80*, 6192–6196.
14. Shibuya, M., Nashimoto, H. & Kaziro, Y. (1979) *Mol. Gen. Genet. 170*, 231–234.
15. Miyajima, A., Shibuya, M. & Kaziro, Y. (1979) *FEBS Lett. 102*, 207–210.
16. Yokota, T., Sugisaki, H., Takanami, M. & Kaziro, Y. (1980) *Gene 12*, 25–31.
17. An, G. & Friesen, J. D. (1980) *Gene 12*, 33–39.
18. 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. & Wade, M. (1980) *Proc. Natl Acad. Sci. USA 77*, 1326–1330.
19. Jones, M. D., Petersen, T. E., Nielsen, K. M., Magnusson, S., Sottrup-Jensen, L., Gausing, K. & Clark, B. F. C. (1980) *Eur. J. Biochem. 108*, 507–526.
20. Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–517.
21. Yousaf, S. I., Carroll, A. R. & Clarke, B. E. (1984) *Gene 27*, 309–313.
22. Yanisch-Perron, C., Vieira, J. & Messing, J. (1985) *Gene 33*, 103–119.
23. Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl Acad. Sci. USA 72*, 3961–3965.
24. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl Acad. Sci. USA 74*, 5463–5467.
25. Messing, J., Crea, R. & Seeburg, P. H. (1981) *Nucleic Acids Res. 9*, 309–321.
26. Mizusawa, S., Nishimura, S. & Seela, F. (1986) *Nucleic Acids Res. 14*, 1319–1324.
27. Shine, J. & Dalgarno, L. (1974) *Proc. Natl Acad. Sci. USA 71*, 1342–1346.
28. Zengel, J. M., Archer, R. H. & Lindahl, L. (1984) *Nucleic Acids Res. 12*, 2181–2192.
29. Jaskunas, S. R., Fallon, A. M. & Nomura, M. (1977) *J. Biol. Chem. 252*, 7323–7336.
30. Nagahari, K., Koshikawa, T. & Sakaguchi, K. (1980) *Gene 10*, 137–145.
31. Kagawa, Y., Nojima, H., Nukiwa, N., Ishizuka, M., Nakajima, T., Yasuhara, T., Tanaka, T. & Oshima, T. (1984) *J. Biol. Chem. 259*, 2956–2960.
32. Sato, S. & Shinomiya, T. (1978) *J. Biochem. (Tokyo) 84*, 1319–1321.
33. Sato, S., Nakazawa, K. & Shinomiya, T. (1980) *J. Biochem. (Tokyo) 88*, 737–747.
34. Kunai, K., Machida, M., Matsuzawa, H. & Ohta, T. (1986) *Eur. J. Biochem. 160*, 433–440.
35. Ahnn, J., March, P. E., Takiff, H. E. & Inouye, M. (1986) *Proc. Natl Acad. Sci. USA 83*, 8849–8853.
36. Miller, D. L., Hachmann, J. & Weissbach, H. (1971) *Arch. Biochem. Biophys. 144*, 115–121.
37. Arai, K., Kawakita, M., Nakamura, S., Ishikawa, I. & Kaziro, Y. (1974) *J. Biochem. (Tokyo) 76*, 523–534.
38. Argos, P., Rossmann, M. G., Grau, U. M., Zuber, H., Frank, G. & Tratschin, J. D. (1979) *Biochemistry 18*, 5698–5703.
39. Friedman, S. M. (ed.) (1978) *Biochemistry of thermophily*, Academic Press, New York.
40. Hopp, T. P. & Woods, K. R. (1981) *Proc. Natl Acad. Sci. USA 78*, 3824–3828.
41. Chou, P. Y. & Fasman, G. D. (1978) *Annu. Rev. Biochem. 47*, 251–276.
42. Sacerdot, C., Dessen, P., Hershey, J. W. B., Plumbridge, J. A. & Grunberg-Manago, M. (1984) *Proc. Natl Acad. Sci. USA 81*, 7787–7791.
43. March, P. E. & Inouye, M. (1985) *J. Biol. Chem. 260*, 7206–7213.
44. Powers, S., Kataoka, T., Fasano, O., Goldfarb, M., Strathern, J., Broach, J. & Wigler, M. (1984) *Cell 36*, 607–612.
45. Capon, D. J., Chen, E. Y., Levinson, A. D., Seeburg, P. H. & Goeddel, D. V. (1983) *Nature (Lond.) 302*, 33–37.