The EMBO Journal vol.3 no.13 pp.3311 – 3315, 1984

Identification of two genes coding for the translation elongation factor EF-1α of *S. cerevisiae*

F. Schirmaier and P. Philippsen

Microbiology Department, Biozentrum, University of Basel, Klingelbergstrasse 70, CH-4056 Basel, Switzerland

Communicated by P. Philippsen

The translation elongation factor EF-1α of the yeast *Saccharomyces cerevisiae* is coded for by two genes, called *TEF1* and *TEF2*. Both genes were cloned. *TEF1* maps on chromosome II close to *LYS2*. The location of *TEF2* is unknown. *TEF2* alone is sufficient to promote growth of the cells as shown with a strain deleted for *TEF1*. *TEF1* and *TEF2* were originally identified as two strongly transcribed genes, which most likely code for an identical or nearly identical protein as judged from S1 nuclease protection experiments with mRNA-DNA hybrids. The DNA sequence analysis of *TEF1* allowed the prediction of the protein sequence. This was shown, by a search in the Dayhoff protein data bank, to represent the translation elongation factor EF-1α due to the striking similarity to EF-1α from the shrimp *Artemia*. A search for *TEF1* homologous sequences in several yeast species shows, in most cases, duplicated genes and a much higher sequence conservation than among genes encoding amino acid biosynthetic enzymes.

Key words: gene conservation/gene duplication/*Saccharomyces cerevisiae*/translation elongation

Introduction

During our work with the *LYS2* locus on chromosome II of *Saccharomyces cerevisiae* we noticed that a region upstream of this gene was heavily transcribed (Eibel and Philippsen, 1983, 1984; H. Eibel, unpublished). We became interested in this transcript and the gene encoding it for the following reasons: the gene was found to be heavily transcribed under several growth conditions, all of 21 *Saccharomyces* strains, including different species tested, contained the transcript, the gene seemed to be more conserved among yeast strains than genes encoding amino acid biosynthetic enzymes, and the gene was found to be duplicated.

After cloning of the two genes we sequenced one and conducted a search in the Dayhoff protein data bank for homologies between the predicted amino acid sequence and the amino acid sequences of known proteins. This showed we had characterized the gene for one of the most abundant proteins, the translation elongation factor EF-1α. This protein guides the aminoacylated tRNAs to the acceptor site in ribosomes under GTP consumption (for review, see Kaziro, 1978).

Results

Mapping of the TEF1 gene

Figure 1 shows the restriction map upstream of the *S. cerevisiae LYS2* gene together with the position of a strongly transcribed gene, called *TEF1*. The position of this gene was determined by the S1 mapping experiments presented in Figure 2. The transcription start lies 0.4 kb to the left of the labelled *Eco*RI site (Figures 1 and 2A). The experiment which should locate the end of the transcript to the right of the labelled *Eco*RI site surprisingly showed two signals, one at 1.0 kb the other at 1.1 kb (Figures 1 and 2B). Such double signals could originate from differently processed primary transcripts, from two transcription termination sites or from a second gene copy elsewhere in the genome with slightly different sequences at the 3′ end. The latter interpretation turned out to be the correct one, as shown below.

Analysis of a TEF1 deletion

If the *TEF1* gene codes for an essential protein its deletion should be deleterious to the cell. We obtained from G. Simchen a *lys2* mutant strain (△201) which carries a long deletion including several thousand base pairs of the *LYS2* 5′ region (Simchen *et al.*, 1984). The approximate deletion end-points are indicated in Figure 1. Since the deletion includes the *Eco*RI, *Bgl*II and *Hind*III sites of the *TEF1* gene this strain does not carry a functional *TEF1* gene. Despite this fact it grew well, aerobically and anaerobically. Hybridization of a *TEF1* probe to the total RNA of △201 still showed a strong signal at 1.6 kb (Figure 3A). We therefore had to conclude that △201 carried at least one active gene with strong homology to the *TEF1* gene and the wild-type at least two genes. *TEF1*-specific hybridizations to cleaved DNA from the wild-type and the deletion mutant clearly showed the presence of two gene copies in the wild-type and one copy in the mutant (Figure 3B, lanes a and d). We named the second gene tentatively *TEF2*. A clone carrying the *TEF2* gene was isolated from a cosmid-yeast bank (Meyhack *et al.*, 1982) by screening with a *TEF1* probe. *TEF2* contains the same gene internal restriction sites as *TEF1* (e.g., lanes b and c of Figure 3B) but different flanking sequences (data not shown). The presence of two genes fully explains the double signal observ-

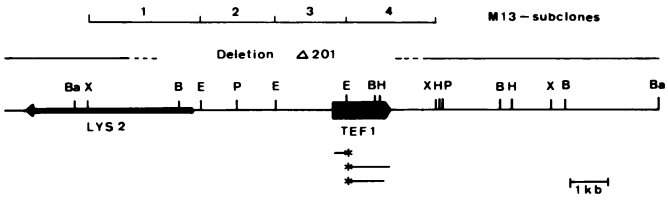

Fig. 1. Map of the *LYS2-TEF1* region in chromosome II. The positions of the *LYS2* gene and the major restriction sites were previously determined (Eibel and Philippsen, 1983). The abundant 1.6-kb *TEF1* transcript was found in 21 *Saccharomyces* species tested when total RNA was probed with labeled λCH2885 DNA. From the four M13-subclones only 3 and 4 hybridized to the 1.6-kb transcript, and subclone 4 reproducibly stronger than subclone 3, indicating that most of the *TEF1* gene resides within the 2.4-kb *Eco*RI-*Xho*I fragment. The lines with the asterisks indicate the positions of the S1 nuclease-protected sequences identified in the experiments of Figure 2. The extent of the deletion in strain △201 is also shown (Simchen *et al.*, 1984 and our own data). The abbrevations used for restriction sites are: X = *Xho*I, B = *Bgl*II, E = *Eco*RI, P = *Pst*I, H = *Hind*III, Ba = *Bam*HI.

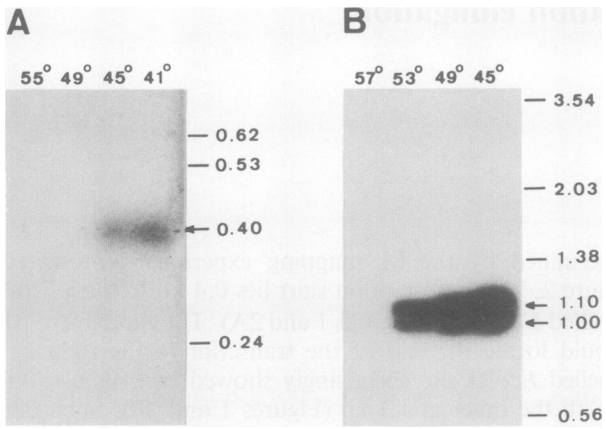

Fig. 2. Autoradiograms of the S1 mapping experiments. **(A)** The 5′-labeled 2.0-kb *Eco*RI fragment of subclone 3 (see Figure 1) was hybridized against an excess of total RNA at 41°, 45°, 49° and 55°C. The samples were digested with S1 nuclease and separated in a 1.5% agarose gel. The size in kb of the protected RNA-DNA hybrids (arrow) was determined by comparison with pBR322 *Hpa*II fragments run in parallel. **(B)** The 2.4-kb *Eco*RI-*Xho*I fragment of subclone 4 (see Figure 1) was labeled at the 3′ end of the *Eco*RI site and hybridized against an excess of total RNA at 45°, 49°, 53° and 57°C. After digestion with S1 nuclease the samples were separated in an 0.8% agarose gel. The sizes in kb of the protected RNA-DNA hybrids (arrows) were determined by comparison with λ*Eco*RI-*Hind*III fragments run in parallel.

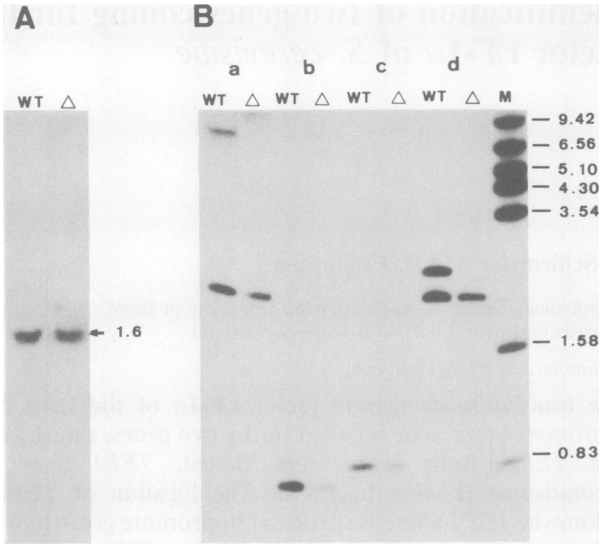

Fig. 3. Hybridization analyses with RNA and DNA from a *TEF1* deletion strain. **(A)** Total RNA from S288C (wild-type) and Δ201 (deletion mutant) was probed with the *TEF1* internal 0.75-kb *Eco*RI-*Bgl*II fragment (see Figure 1). The size of the signal was determined by comparison with the 1.75-kb and 3.35-kb rRNAs which were identified by probing strips run in parallel with rDNA (Philippsen *et al.*, 1978). **(B)** Total DNA from S288C and Δ201 was digested with **(a)** *Eco*RI/*Bam*HI, **(b)** *Eco*RI/*Bgl*II, **(c)** *Eco*RI/*Hind*III and **(d)** *Eco*RI/*Pst*I. The fragments were separated in a 0.7% agarose gel and hybridized with the same probe as in **A**. λ *Eco*RI-*Hind*III fragments were run in parallel as size markers (M). Several of them are visible in the autoradiogram because of a separate hybridization with labeled λCH2885 DNA (Eibel and Philippsen, 1983).

ed in the S1 nuclease protection experiment for mapping the *TEF1* transcription terminus (Figure 2B). *TEF1* and *TEF2* transcripts hybridized to the *TEF1* gene probe and both transcripts have identical or nearly identical sequences up to ~0.1 kb before the transcription terminus of *TEF1*. This predicted point of sequence divergence between the two genes coincides with the end of the open reading frame in *TEF1*, as discussed below.

TEF1 codes for the transcription elongation factor EF-1α

The DNA of the *TEF1* region was subcloned into different M13 vectors and both strands were sequenced by the chain termination method of Sanger *et al.* (1977). The DNA sequence and the deduced amino acid sequence of the only sufficiently long open reading frame (458 amino acids) are presented in Figure 4. The position of the open reading frame agrees with the S1 mapping data of the mRNA shown above. Transcription starts ~30 nucleotides before the ATG and terminates ~0.1 kb after the TAA.

To determine the function of the protein, we searched in the Dayhoff protein data bank for similarities to this sequence using the predicted amino acid sequence of Figure 4. The computer found a small but significant similarity to the *Escherichia coli* translation elongation factor EF-Tu (Jones *et al.*, 1980; Laursen *et al.*, 1981; Nakamura *et al.*, 1982) and an almost 80% similarity to partial sequence information of the translation elongation factor EF-1α from the shrimp *Artemia* (Amons *et al.*, 1983). The complete sequence of EF-1α from this shrimp has recently been published (van Hemert *et al.*, 1984) and shows 80% homology to *TEF1*. The blocks of amino acids identical with the *E. coli* and *Artemia* protein are indicated in Figure 4 by crosses and thin bars, respectively.

TEF1 homologous sequences in other yeasts

DNA from different *Saccharomyces* species and yeast genera was hybridized under stringent conditions to a *TEF1*-specific probe. Examples are shown in Figure 5. The detection of one hybridization signal each with *Kluyveromyces lactis* and *Schizosaccharomyces pombe* DNA is remarkable, because neither a *LYS2, TRP1, HIS3, CEN6* or Ty probe from *S. cerevisiae* showed detectable hybridization to DNA of these two yeast genera (Eibel and Philippsen, 1983; Panzeri and Philippsen, 1982; A. Stotz and P. Philippsen, unpublished).

As far as the copy number of *TEF1*-related sequences in different yeasts is concerned, a final conclusion cannot be drawn from the presently available hybridization data. The majority of the tested strains seem to carry two copies (data not shown) but the copy number may also be one or three as seen from the examples in Figure 5.

Discussion

We wanted to identify the protein coded by a so far uncharacterized yeast gene. The use of the Dayhoff data bank showed convincingly that this gene encodes the yeast translation elongation factor EF-1α which consists of 458 amino acids. Because of the ease with which DNA sequence information can be obtained, computer-assisted comparisons with protein sequences will become common practice. It is an open question how valid the detection of minor homologies will be in the identification of gene products. The case reported here is probably exceptional because of the observed extremely high sequence conservation of the eucaryotic EF-1α proteins. Five blocks of 22–33 amino acids are identical between the *S. cerevisiae* and *Artemia* EF-1α protein and the average homology is 80%.

Most of the *S. cerevisiae* ribosomal proteins are encoded by duplicated genes (Fried *et al.*, 1981; Molenaar *et al.*, 1984). The gene for the translation elongation factor EF-1α also exists in two copies, called *TEF1* and *TEF2*, per haploid

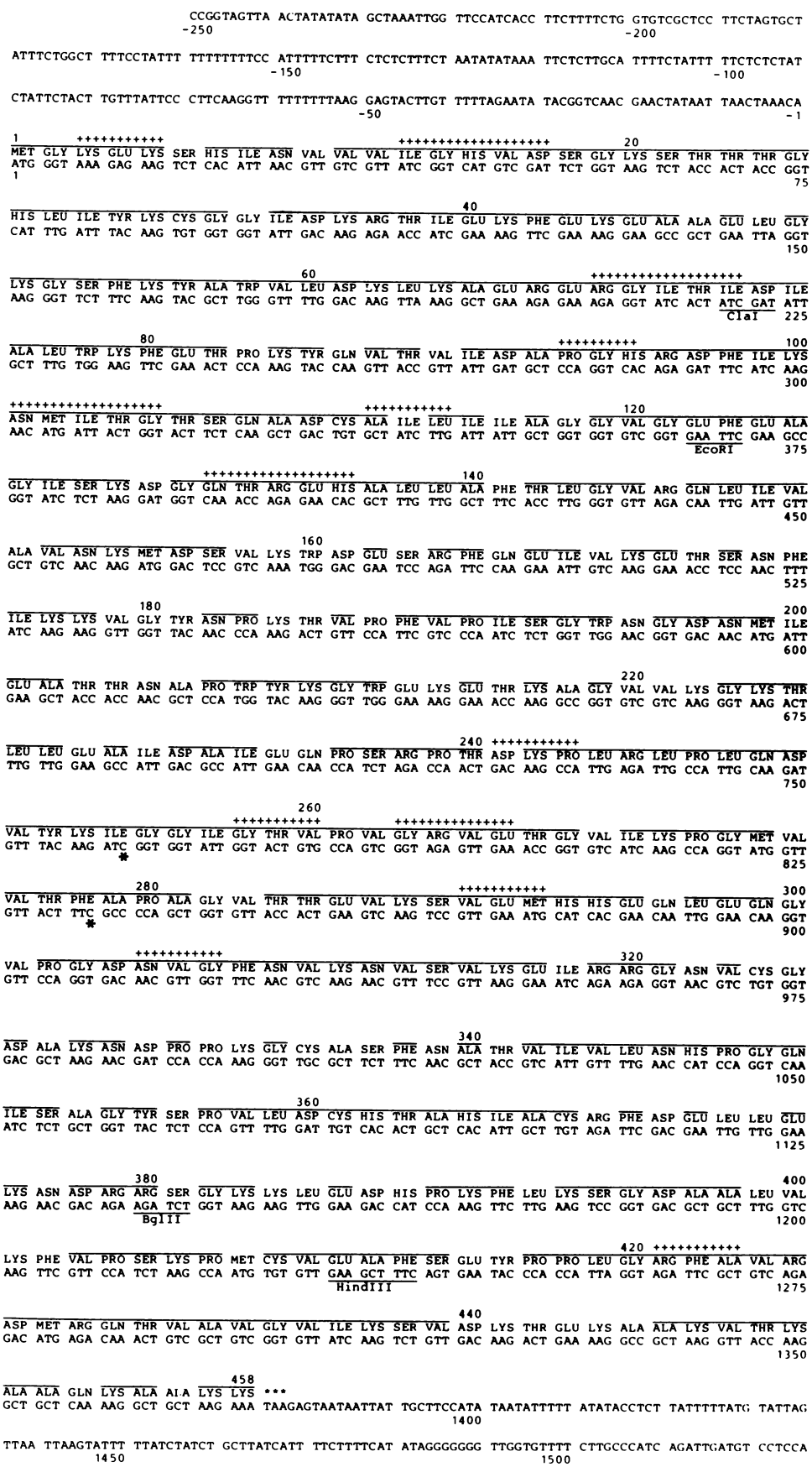

Fig. 4. Nucleotide sequence of the *S. cerevisiae TEF1* gene and the predicted amino acid sequence of EF-1α. Homology of three and more amino acids with EF-Tu from *E. coli* is indicated by crosses. Homologies with EF-1α from the shrimp *Artemia* are shown by thin lines. The two single base pair differences between the open reading frames of *TEF1* and *TEF2* (sequence taken from Nagata *et al.*, 1984) are marked by asterisks.

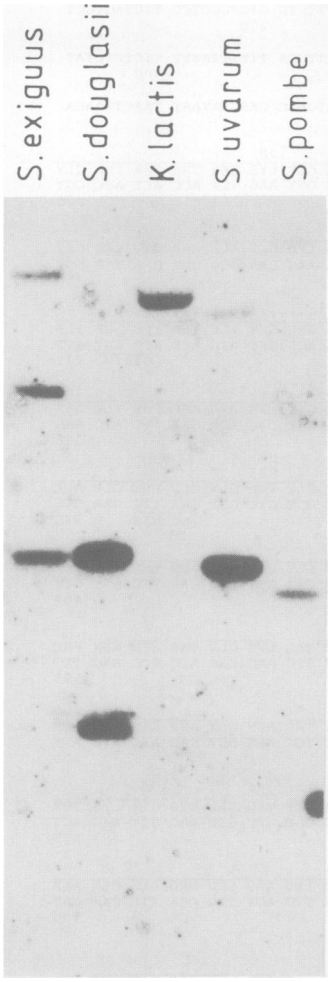

Fig. 5. *TEF1*-specific hybridisations to various yeast DNAs. DNA from *S. exiguus, S. douglasii, K. lactis, S. uvarum* and *S. pombe* was cleaved to completion with *Eco*RI (assayed with a rDNA probe), separated on a 0.7% agrose gel, transferred to nitrocellulose and probed with the *TEF1* internal *Eco*RI-*Hind*III fragment (see Figure 1). Approximately 0.8 μg of DNA were loaded in each lane. The signal intensity is therefore a rough measure for the degree of sequence similarity to the probe.

genome. Our S1 protection analysis of mRNA-DNA hybrids indicates that *TEF1* and *TEF2* have identical or nearly identical sequences. We have sequenced so far *TEF1*. The codon usage is biased as expected for a heavily expressed *S. cerevisiae* gene (Bennetzen and Hall, 1982). Recently, the sequence of *TEF2* was published by Nagata *et al.* (1984) who isolated the two genes for the elongation factor by a direct approach using antibodies against purified EF-1α and DNA probes partially homologous to the EF-1α gene. The open reading frames of *TEF1* and *TEF2* differ in two codons but in both cases in the third position without alteration of the predicted amino acid sequence (see asterisks in Figure 4). The interactions of the elongation factor EF-1α with the aminoacylated tRNAs and the ribosome components around the tRNA acceptor site probably do not, or very rarely, allow changes in the primary structure of EF-1α without loss of function. This selection pressure and the use of biased codons explains the highest degree of DNA homology observed so far for open reading frames of gene pairs in *S. cerevisiae*. Outside of the open reading frames the homology between *TEF1* and *TEF2* drops to 40%.

The cloning of the two genes for the translation elongation factor EF-1α and the demonstration that the cells are viable with one gene copy opens the way for a genetic analysis of this key protein in the translation process. One aim will be the *in vitro* construction of conditional lethal mutants. Selection and characterization of extragenic suppressors for these mutants should allow the identification of those macromolecules most likely interacting with the EF-1α protein.

Materials and methods

Recombinant DNA

λCH2885 a recombinant λ-yeast phage carries the 15-kb *Bam*HI fragment with 3 kb of the *S. cerevisiae LYS2* gene and 12 kb of its 5′-flanking sequences (Eibel and Philippsen, 1983). DNA from this phage was the source for cloning and characterization of the *TEF1* gene. Subclones in M13 vectors were processed in *E. coli* strain JM101 (Messing, 1983). Subclones in pBR322 were processed in the *E. coli* K12 strain 294 (Eibel and Philippsen, 1983). λgt30-Sc310 a recombinant λ-yeast hybrid phage carries the 9.1-kb *S. cerevisiae* ribosomal DNA unit (Philippsen *et al.*, 1978). Radioactive labelled DNA from this phage was used for size calibrations in Northern-type hybridizations and as probe for assaying complete digestions of yeast DNAs.

Yeast strains and media

S. cerevisiae S288C (α *mal gal2*) is a widely used laboratory strain which is called wild-type in this paper. *S. cerevisiae* SR36/201 (*a his4-912 ura3-52 lys2-201*) carries a deletion (△201) of *TEF1* and part of *LYS2* (Simchen *et al.*, 1984) and was kindly donated by G. Simchen, Jerusalem. The RNA hybridizations to a *TEF1* probe includes S288C, two diploid *S. cerevisiae* laboratory strains, four *S. cerevisiae* wine yeasts, ten *S. cerevisiae* natural isolates, two *S. uvarum*, one *S. oviformis* and one *S. douglasii* strains which were obtained from the Australian wine yeast collection, from the Centraal Bureau voor Schimmel Cultures in Delft, Holland, from Ron Davis, Stanford and from Don Hawthorne, Seattle. The DNA hybridizations to a *TEF1* probe included DNA from *S. cerevisiae, S. servazii, S. kluyvery, S. telluris, S. dairensis, S. unisporus, S. exiguus, S. uvarum, S. douglasii, K. lactis* and *S. pombe* which were obtained from the Centraal Bureau voor Schimmel Cultures in Delft. All yeasts were grown at 30°C in YPD (1% yeast extract, 2% peptone, 2% dextrose).

Enzymes

Restriction endonucleases were purchased either from Biolabs, Bethesda Research Laboratories or Boehringer, Mannheim, and were used essentially as recommended by the supplier. DNA ligase and Klenow polymerase were purchased from Biolabs, bacterial alkaline phosphatase, T4 kinase and *E. coli* DNA polymerase I from Bethesda Research Laboratories and S1 nuclease from Sigma. [γ-^{32}P]ATP and [d-^{32}P]dNTPs, were obtained from Amersham.

DNA-RNA hybridization

Northern analysis was carried out essentially as described by Thomas *et al.* (1980) except that the applied gel system also contained 6% formaldehyde and the samples were resuspended in 50% formamide, 2.2 M formaldehyde, 10 mM $NaPO_4$ instead of the Glyoxal solution. Total cell RNA was run on a gel, transferred to a nitrocellulose filter and hybridized with nick-translated λCH2885 DNA.

Nick-translation was performed with [α-^{32}P]dTNPs (410 Ci/mmol, Amersham) to yield ~1 x 10^6 c.p.m. (μg DNA). Pre-hybridization was performed at 42°C overnight in a mixture of 50% formamide, 5 x SSC, 50 mM $NaPO_4$, pH 6.5–7.0, 1 x Denhardts. Hybridization was done at 42°C overnight in a mixture of four parts pre-hybridization buffer, one part 50% Dextran sulfate, containing the nick-translated probe. The filter was washed at 55°C with 0.2 x SSC, 0.2% SDS, dried and exposed to X-ray film.

S1 mapping

S1 mapping was carried out essentially as described by Sharp *et al.* (1980) and Maniatis *et al.* (1982). Either 5′ or 3′ end-labelled fragments were mixed with ~20 μg of total cell RNA, precipitated and the pellet resuspended in hybridization buffer (40 mM Pipes, 1 mM EDTA, 0.4 M NaCl, 80% formamide). After denaturing at 78°C for 15 min, the probes were renatured, trying a set of different temperatures (41–57°C), allowing DNA-RNA hybrids to form. After 3 h of hybridization, the samples were diluted 10-fold with S1 digestion buffer containing 500 U/ml S1 nuclease and incubated at 37°C for 30 min. The reactions were stopped with termination mix (4 M ammonium acetate, 0.1 M EDTA). The samples were phenol extracted and then ethanol precipitated.

Southern hybridization

Southern hybridization was carried out essentially as described by Southern (1975). Yeast DNA was digested, separated on a 0.7% agarose gel treated for

2 x 15 min each with 0.25 M HCl, 0.5 M NaOH and with 1 M Tris-HCl, pH 7.5 containing 2 M NaCl, transferred to a nitrocellulose filter and hybridized overnight at 65°C in hybridization mix (6 x SSC, 0.2% SDS, 1 x Denhardts), containing a nick-translated probe. The filter was washed at 55°C with a mixture of 2 x SSC, 0.2% SDS, dried and exposed to X-ray film.

DNA sequence determination

DNA sequence analysis was carried out by the dideoxy method (Sanger *et al.*, 1977). Single-stranded templates were obtained by transforming *E. coli* JM 101 cells with recombinant bacteriophage M13-RF DNA. The M13 vectors used were mp8 or mp9. Each sequence was determined at least twice.

Computer analysis

Protein homologies were searched by comparing blocks of four amino acids with the Dayhoff protein sequence bank, which contains 2372 sequences with a total of 430 262 amino acids. The computer revealed homology with fragments of EF-1α of brine shrimp (Amons *et al.*, 1983) and with prokaryotic translation elongation factor EF-Tu (Jones *et al.*, 1980; Laursen *et al.*, 1981). The protein data bank used is the following: Dayhoff Library: Protein Identification Resource (PIR), National Biomedical Research Foundation, Georgetown University Medical Center, 3900 Reservoir Road, N.W., Washington, DC 20007, USA, (202) 625-2121.

Acknowledgements

We would like to thank Christoph Nager for providing and handling the computer programs, David Pridmore for his advice in DNA sequencing and Agathe Stotz for help in the DNA-DNA hybridization experiment to various yeast DNAs. We are also grateful to Ursula Fleig and Giora Simchen for the generous gift of recombinant DNAs and strains, and to A. Hinnen for his yeast cosmid gene bank. The stimulating discussion with Hermann Eibel and Ursula Fleig are gratefully acknowledged. We thank Tom Bickle and Hans Trachsel for critically reading the manuscript and Elvira Amstutz for typing it. This work was supported by the Deutsche Forschungsgemeinschaft, Bonn, FRG.

References

Amons,R., Pluijms,W., Roobol,K. and Möller,W. (1983) *FEBS Lett.*, **153**, 37-42.

Bennetzen,J.L. and Hall,B.D. (1982) *J. Biol. Chem.*, **257**, 3026-3031.

Eibel,H. and Philippsen,P. (1983) *Mol. Gen. Genet.*, **191**, 66-73.

Eibel,H. and Philippsen,P. (1984) *Nature*, **307**, 386-388.

Fried,H.M., Pearson,N.J., Kim,C.H. and Warner,J.R. (1981) *J. Biol. Chem.*, **256**, 10176-10183.

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.

Laursen,R.A., L'Italien,J.J., Nagarkatti,S. and Miller,D.L. (1981) *J. Biol. Chem.*, **256**, 8102-8109.

Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.

Messing,J. (1983) *Methods Enzymol.*, **101**, 20-78.

Meyhack,B., Bajwa,W., Rudolph,H. and Hinnen,A. (1982) *EMBO J.*, **1**, 675-680.

Molenaar,C.M.T., Woudt,L.P., Jansen,A.E.M., Mager,W.H., Planta,R.J., Donovan,D.M. and Pearson,N.J. (1984) *Nucleic Acids Res.*, in press.

Nagata,S., Nagashima,K., Tsunetsugu-Yokota,Y., Fujimura,K., Miyazaki, M. and Kaziro,Y. (1984) *EMBO J.*, **3**, 1825-1830.

Nakamura,S., Nakayama,N., Takahashi,K. and Kaziro,Y. (1982) *J. Biochem.*, **91**, 1047-1063.

Panzeri,L. and Philippsen,P. (1982) *EMBO J.*, **1**, 1605-1611.

Philippsen,P., Thomas,M., Kramer,R.A. and Davis,R.W. (1978) *J. Mol. Biol.*, **123**, 387-404.

Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.

Sharp,P.A., Berk,A.J. and Berget,S.M. (1980) *Methods Enzymol.*, **65**, 750-768.

Simchen,G., Winston,F., Styles,C.A. and Fink,G.R. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2431-2434.

Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.

Thomas,P.S. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5201-5205.

van Hemert,F.J., Amons,R., Pluijms,W.J.M., van Ormondt,H. and Möller, W. (1984) *EMBO J.*, **3**, 1109-1113.

Received on 25 September 1984; revised on 15 October 1984

Note added in proof

Both *TEF* genes of *S. cerevisiae* were also cloned and characterized by Cottrelle, Thiele, Price, Memet, Micouin, Marck, Buhler, Sentenac and Fromaget (*J. Biol. Chem.*, in press). These authors also use the gene symbols *TEF1* and *TEF2*. Three letter codes plus a number are the accepted rules for nomenclature in yeast genetics. Unfortunately, our *TEF1* is their *TEF2* and our *TEF2* is their *TEF1*. In order to avoid confusion in the scientific community we will in future use the designation proposed by Cottrelle *et al.*