Nucleic Acids Research, Vol. 18, No. 1

Structure and expression of the *Euglena gracilis* nuclear gene coding for the translation elongation factor EF-1α

Paul-Etienne Montandon* and Erhard Stutz
Laboratoire de Biochimie Végétale, Université de Neuchâtel, Ch. de Chantemerle 18, CH-2000 Neuchâtel, Switzerland

Received October 17, 1989; Revised and Accepted November 29, 1989 EMBL accession no. X16890

ABSTRACT

A cDNA library from the protist *Euglena gracilis* was used to isolate and sequence an ORF coding for the elongation factor protein EF-1α. The decoded amino acid sequence (MW, 48'515) is to 75 – 80% identical with other eukaryotic EF-1α sequences but only to 24% identical with the *Euglena* chloroplast EF-Tu. Homologous DNA probes interact with multiple fragments of *Euglena* nuclear restricted DNA typical for a multimembered gene family. We present the restriction sites map of four *tef* nuclear gene loci and postulate that the nuclear genome also contains *tef* related sequences (e.g. pseudogenes). Expression of *tef* gene(s) is monitored by Northern hybridization and the 5' end of a stable transcript (1.5 kb) is sequenced and shown to precede the start codon by 29 positions only. The steady state concentration of the 1.5 kb mRNA is not influenced by switching cell growth conditions from dark to light (chloroplast development).

INTRODUCTION

We showed some time ago (1) that the chloroplast specific protein synthesis elongation factor EF-Tu of *Euglena gracilis* is encoded in chloroplast DNA. The tuf gene is part of a gene cluster similar to the str operon of *E. coli* and transcribed into a stable mRNA of 1.95 kb (2). The nucleocytoplasmic counterpart, i.e., the elongation factor protein EF-1α was purified and partially characterized (3) but nothing is known about the corresponding nuclear *tef* gene(s).

More recently several eukaryotic EF-1α genes have been analysed. Studies on cDNAs and genomic clones encoding *tef* were reported, e.g., for *Saccharomyces cerevisiae* (4), *Mucor racemosus* (5), *Lycopersicum esculentum* (6), *Arabidopsis thaliana* (7) and representatives of the animal kingdom (e.g. 8,9). It was shown that two or more gene loci exist and morphology-specific patterns of transcript accumulation were described, e.g. for *Mucor* (5).

Euglena gracilis is an old protist having a complex nuclear genome composed of between 45–50 chromosomes (10). According to renaturation kinetic studies (11) about 12% (40×the *E. coli* genome) are single copy sequences and both middle and highly repetitive sequences are in close vicinity of interspersed single copy DNA. The exact ploidy, however, is unknown, although it was suggested that *Euglena gracilis* most likely is a diploid organism (12).

Very little is known about organisation and expression of specific nuclear genes of *Euglena gracilis*. The nuclear genes coding for chlorophyll a/b binding proteins (CAB) of photosystem I (PSI) and II (PSII) (13,14) are transcribed into unusually large mRNAs coding for polyproteins while the gene(s) for β-tubulin is (are) transcribed into a stable mRNA encoding a single protein (15). The question arises whether genes coding, respectively, for organellar and cytosolic proteins are distinctly organized. To the end of getting more information about gene organisation and expression we established a cDNA library and report in the following about the structure of the coding part and immediate vicinity of the EF-1α genes and their expression into a stable transcript in dark and light grown cells. We determine the 5' end of the mRNA and compare the decoded EF-1α sequence with other elongation factor proteins.

MATERIALS AND METHODS

Enzymes were purchased from Boehringer-Mannheim, or Biolabs and used following instructions of the supplier. (α^{32}P) dATP (400 Ci/mmole) and (γ^{32}P) ATP were from Radiochemical Center Amersham.

Cell culture and preparation of total RNA

Euglena gracilis (Z. strain) was grown heterotrophically in a modified Hutner's medium with vitamin B12 at 50 ng/l (16) and as reported (1). For studying *tef* gene expression during chloroplast development cells were grown first in the dark to stationary phase, then transferred to fresh medium (pH 7.0) with sodium citrate as sole organic carbon source and exposed to light for various periods of time.

Total RNA was extracted essentially as described (17). Approximately 1 g of cells was resuspended in 10 ml of lysis buffer (Tris-HCl pH 9.0, 50 mM; NaCl 100 mM; EDTA pH 8.0 10 mM; sodiumdodecylsulfate 1% and triisopropylnaphtalin sulfonate 1%) and one volume of phenol-cresol–8–hydroxyquinoleine solution (100 ml of solution contains 70 ml of phenol, 10 ml of cresol and 0.1 g of 8-hydroxyquinoleine) was added and extraction was carried out following standard

* Present address: Service de l'Hygiène et de l'Environnement, Avenue Léopold-Robert 36, CH-2300 La Chaux-de-Fonds, Switzerland

procedures. DNA was removed by digestion with DNase I (DNase I, ribonuclease free, Worthington) or by lithium chloride precipitation (18).

cDNA synthesis and preparation of cDNA library

Euglena poly-A$^+$ mRNA was isolated by chromatography on oligo-dT cellulose (19) and cDNA was made following established protocols (20,21). The first strand cDNA was synthesized from *Euglena* poly-A$^+$ mRNA with AMV reverse transcriptase primed with oligo-dT. The second strand was synthesized with DNA polymerase I. After S1 digestion, EcoRI sites were protected by methylation and the double stranded cDNA was ligated to EcoRI linkers using T4 DNA ligase. The cDNA was digested with EcoRI, purified by chromatography on Sepharose 4B column and inserted into the EcoRI site of puc-8 as described (22). Two recombinant clones containing *Euglena* EF-1α sequences were isolated by heterologous hybridization (23) using a Mucor racemosus *tef* DNA probe (5) in a 5×SSPE (1×SSPE is 0.18 M NaCl, 0.01 M sodium phosphate, pH 7.4, 0.001 M EDTA) based solution containing 30% formamide at 37°C. Filter imprints were washed twice in 5×SSPE buffer for 15 min and twice in hybridization buffer at 30°C for 10 min. Filters were autoradiographed using Kodak X-ray film XAR-5.

DNA sequencing

Subfragments of inserts containing EF-1α sequences were inserted into M13mp18/19 and sequenced following standard protocols (24).

Northern blots and RNA:DNA hybridization

RNA was denatured with glyoxal (25). RNA was electrophoresed in 1.2% agarose gels. Filter blotting and hybridization were as published (2).

Preparation of DNA primer and RNA sequencing by primer extension

A ^{32}P labelled DNA primer was synthesized using as template *tef* single strand DNA (position 9 to 156) inserted in phage M13mp19. A 70 bp fragment was prepared by cutting the copied fragment with AccI and BamHI, which cleave respectively, *tef* DNA at position 90 and M13mp19 DNA in the polylinker region. The DNA fragment was electrophoretically purified (acrylamide gel) and eluted by diffusion. Primer extension dideoxysequencing was carried out essentially as described (26) using the 70 bp DNA fragment and 20 μg of total RNA.

Isolation of *Euglena* DNA

Euglena spheroplasts were prepared by incubation of freshly harvested *Euglena* cells with trypsine as described (27). Spheroplasts from 1 g of cells were resuspended into 5 ml of a solution containing Tris-HCl pH 8.0 0.05 M and EDTA 0.1 M, and lysed with Triton X-100 and Sarkosyl at final concentration of 2.5 and 1%, respectively. Proteins were digested with proteinase K (20 μg per ml) for ½ hour at 37°C. 0.78 g of Cesium chloride were added per ml of solution, unlysed cells were removed by centrifugation and DNA was isolated according to (28).

Southern blots and DNA:DNA hybridizations

Southern hybridizations were done using nitrocellulose filters (Schleicher and Schuell, BA83) in 5×SSPE based solutions containing 50% formamide at 42°C (standard conditions) or at 53°C (stringent conditions). Membranes were washed twice in 2×SSC and 0.1% SDS for 15 min at room temperature and twice in 0.5×SSC, 0.1% SDS at 30°C for 15 min (1×SSC is 0.15 M NaCl, 0.015 M sodium citrate).

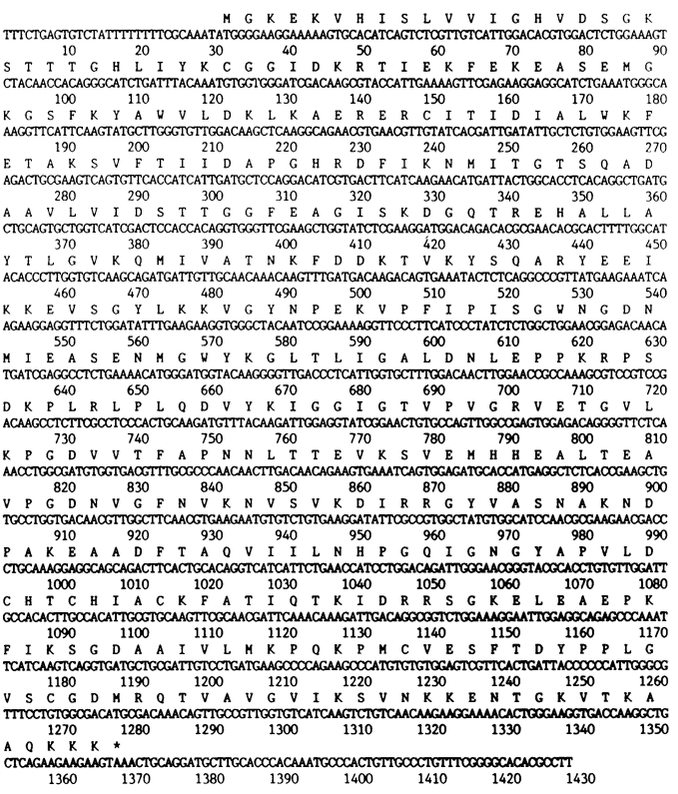

Figure 1. Nucleotide sequence (RNA-like DNA strand) of the EF-1α gene and the decoded amino acid sequence. Numbering of nucleotide sequence starts with the 5′ end of EF-1α mRNA as identified by primer extension dideoxysequencing (consult Fig. 7). The last position (1427) preceds the poly-A tail as established by sequencing cDNA inserts. Amino acid sequence is in one letter code.

RESULTS

cDNA clones and nucleotide sequences

A cDNA library was tested with an EF-1α gene probe from *Mucor racemosus*. Two clones were chosen for sequence analysis having inserts which cover, respectively, positions 9 to 1193 and 412 to 1427 with a poly-A tail as shown in Fig. 1. The overlap of the two inserts amounts to 781 positions, but no sequence divergence was seen, suggesting that the two inserts stem from identical transcripts. As shown in Fig. 1 the start codon is preceded by 29 nucleotides and sequences 1 to 8 were obtained by reverse dideoxy sequencing of transcripts as discussed later (see Fig. 7). A noteworthy feature of the leader part is a run of 8 pyrimidines near the start codon. The non-coding 3′ terminal region consists of 60 nucleotides. The poly-A tail starts at position 1427. The non-translated 3′ region contains no polyadenylation consensus sequence (e.g. AAUAAA). Closest to the consensus comes the pentanucleotide ACAAA (pos. 1388−1392). Nuclear plant genes, however, often lack such a consensus sequence (29).

Decoded protein sequence

The decoded aminoacid sequence of EF-1α is aligned with the nucleotide sequence in Fig. 1. Based on this analysis the *Euglena* EF-1α protein is composed of 445 aminoacids (MW = 48′515).

```
EG.cyt  MGKE------KVHISLVVIGHVDSGKSTTTGHLIYKCGGIDKRTIEKFEKEASEMGKGSFK  55
EG.chl  MARQKFERTKPHINIGTIGHVDHGKTTLTAAITMA----------LAATGNSKAK  44
MCV     MAK-------TKPILNVAFIGHVDAGKSTTVGRLLLDGGAIDPQLIVRLRKEAEEKGKAGFE  55
EC      MSKEKFERTKPHVNVGTIGHVDHGKTTLTAAITTV----------IAKTYGGAAR  44
SC.mit  -YAAAFDRSKPHVNIGTIGHVDHGKTTLTAAITKT----------LAAKGGANFL  81
SC.cyt  MGKE------KSHINVVVIGHVDSGKSTTTGHLIYKCGGIDKRTIEKFEKEAAELGKGSFK  55

EG.cyt  YAWVLDKLKAERERGITIDIALWKFETPKYQVTVIDAPGHRDFIKNMITGTSQADAAVLV  115
EG.chl  RYEDIDSAPEEKARGITINTAHVEYETKNRHYAHVDCPGHADYVKNMITGAAQMDGAILV  104
MCV     FAYVMDGLKEERERGVTIDVAHKKFFTAKYEVTIVDCPGHRDFIKNMITGASQADAAVLV  115
EC      AFDQIDNAPEEKARGITINTSHVEYDTPTRHYAHVDCPGHADYVKNMITGAAQMDGAILV  104
SC.mit  DYAAIDKAPEERARGITITSTAHVEYETAKRHYSHVDCPGHADYIKNMITGAAQMDGAIIV  141
SC.cyt  YAWVLDKLKAERERGITIDIALWKFETPKYQVTVIDAPGHRDFIKNMITGTSQADCAILI  115

EG.cyt  IDSTTGGFEAGISKDGQTRTHALLAYTLGVKQMIVAINKFDDKTVKYSQARYEEIKKEVS  175
EG.chl  VSAADGPM---------PQTKEHILLAKQVGVPNIVVFLNKEDQVDDSE----LLELVELEIR  154
MCV     VNVDDAGS---GI---QPQTREHVFLIRTLGVRQLAVAVNKMDTVNGSEAD-YNELKKMIGD  170
EC      VAATDGPM---------PQTREHILLGRQVGVPYIIVFLNKCDMVDDEE----LLELVEMEVR  154
SC.mit  VAATDGQM---------PQTREHLLLARQVGVQHIVVFVNKVDTIDDPE----MLELVEMEMR  191
SC:cyt  IAGGTGEFEAGISKDGQTREHALLAFTLGVRQLIVAVNKMDSVKWDES-RFQEIVK-ETS  173

EG.cyt  GYLKKVGYNPEKVPFIPISGWNGDNMIEASENMGWYKG----------LTLIGALDNL  223
EG.chl  ETLSNYEFPGDDIPVIPGSAL---LSVEALTKNPKITKGEN----KWVDKILNLMDQVDSYI  209
MVC     QLLKMIGFNPEQINFVPVASLHGDNVFKKSERNPWYKG----------PTIAEVIDGF  220
EC      ELLSQYDFPGDDTPIVRGSALKALEGD----------AE-WEAKILELAGFLDSYI  199
SC.mit  ELLNEYGFDGDNAPIIMGSALCALEGR----------QPEIGEQAIMKLLDAVDEYI  238
SC.cyt  NFIKKVGYNPKTVPFVPISGWNGDNMIEATTNAPWYKGWEKETKAGVVKGKTLLEAIDAI  234

EG.cyt  EPPKRPSDKPLRLPLQDVYKIGGIGTVPVGRVETGLLKPGD----VVTFAPNNLTTEVKSV  280
EG.chl  PTPTRDTEKDFLMAIEDVLSITGRGTVATGRVERGTIKVGETVELVGLKDTR-STTITGL  268
MVC     QPPEKPTNLPLRLPIQDVYTITGVGTVPVGRVETGIIKPGD----KVVFEPAGAIGEIKTV  275
EC      PEPERAIDKPFLLPIEDVFSISGRGTVVTGRVERGIIKVGEEVEIVGIKETQ-KSTCTGV  258
SC.mit  PTPERDLNKPFLMPVEDIFSISGRGTVVTGRVERGNLKKGEELEIVGHNSTPLKTTVTGI  298
SC.cyt  EQPSRPTDKPLRLPLQDVYKIGGIGTVPVGRVETGVIKPGM----VVTFAPAGVTTEVKSV  290

EG.cyt  EMHHEALTEAVPGDNVGFNVKNVSCKDIRRGYVASNAKNDPAKEAADLTAQVSILNHPGQ  320
EG.chl  EMFQKSLDEALAGDNVGVLLRGIQKNDVERGMVLAKPRTINPH---TKFDSQVYILTKEEG  326
MVC     EMHHEQLPSAEPGDNIGFNVRGVGKKDIKRGDVLGHTINPPTVA-TDFTAQIVVLQHPSV  334
EC      EMFRKLLDEGRAGENVGVLLRGIKREEIERGQVLAKPGTIKPH---TKFESEVYILSKDEG  316
SC.mit  EMFRKELDSAMAGDNAGVLLRGIRRDQLKRGMVLAKPGTVKAH---TKILASLYILSKEEG  356
SC.cyt  EMHHEQLEQGVPGDNVGFNVKNVSVKEIRRGNVCGDAKNDPPKGCASFNATVIVLNHPGQ  350

EG.cyt  IGN------GYAPVLDCHTCHIACKFATIQTKIDRRSGKELEAEPKFIKSGDAAIVLMKPQ  395
EG.chl  GRHTPFFEGYRPQFYVRTTDVTGKIESF-------RSDN---DNPAQMVMPGDRIKMKVELI  378
MCV     LTD------GYTPVFHTHTAQIACTFAEIQKKLNPATGEVLEENPDFLKAGDAAIVKLIPT  390
EC      GRHTPFFKGYRPQFYFRTTDVTGTI-ELPEGVEM----------VMPGDNIKMVVTLI  363
SC.mit  GRHSGFGENYRPQMFIRTADVTVVMRFPK-EVEDHSMQ----------VMPGDNVEMECDLI  407
SC.cyt  ISA------GYSPVLDCHTAHIACRFDELLEKNDRRSGKKLEDHPKFLKSGDAALVKFVPS  405

EG.cyt  KPMCVESFTDYPPLGVSCG-DMRQTVAVGVIKSVNKKENTGKVTKAAQKKK  445
EG.chl  QPIAIEK-------GMRFAIREGGRTVGAGVVLSIIQ  408
MCV     KPMVIESVKEIPQLGRFAIRDMGMTVAAGMAIQVTAKNK  428
EC      HPIAMDD-------GLRFAIREGGRTVGAGVVAKVLG  393
SC.mit  HPTPLEV-------GQRFNIREGGRTVGTGLITRIIE  437
SC.CYT  KPMCVEAFSEYPPLGRFAVRDMRQTVAVGVIKSVDKTEKAAKVTKAAQKAAKK  458
```

Figure 2. Comparison of EF-1α primary structure with some relevant elongation factor proteins. The overall pattern of sequence alignment follows an analysis of Lercher and Böck (30). Overlines mark invariant positions. Dashed lines within the sequences indicate deletions. At position 213 we note for *Euglena* and *Methanococcus* an identical deletion of amino acids. Numbers at the right margin are accumulated amino acids positions. EG.chl, *Euglena gracilis* chloroplast (1); MCV, *Methanococcus vannielli* (30); EC, *Escherichia coli* (32); SC.mit, *Saccharomyces cerevisiae* mitochondria (33); SC.cyt, *Saccharomyces cerevisiae* cytosol (4).

From a comparative study (not shown) a high degree of sequence identity with other eukaryotic EF-1α proteins becomes apparent, e.g. 74% with yeast (4), 80% with tomato (6).

In Fig. 2 we compare the *Euglena* EF-1α protein with organellar, eubacterial and archaebacterial counterparts including the yeast nuclear EF-1α. The following points are noteworthy: 1) The *Euglena* EF-1α has only 108 out of 445 aminoacids positions in common with its chloroplast EF-Tu (24%) but shares 212 positions (48%) with the archaebacterial counterpart; 2) Size and position of a deletion (pos. 213/214) in *Euglena* EF-1α match that in EF-1α of *Methanococcus vannielli* (30). This gap exists also in the *tef* genes of tomato (6), soybean (Aguilar and Stutz, to be published) and *Arabidopsis* (7) but not in fungi (4,5) or animals (8,9). Consensus sequences GXXXXGK (pos. 14), DXXG (pos. 91) and NKXD (pos. 153) corresponding to the three GTP binding domains are preserved.

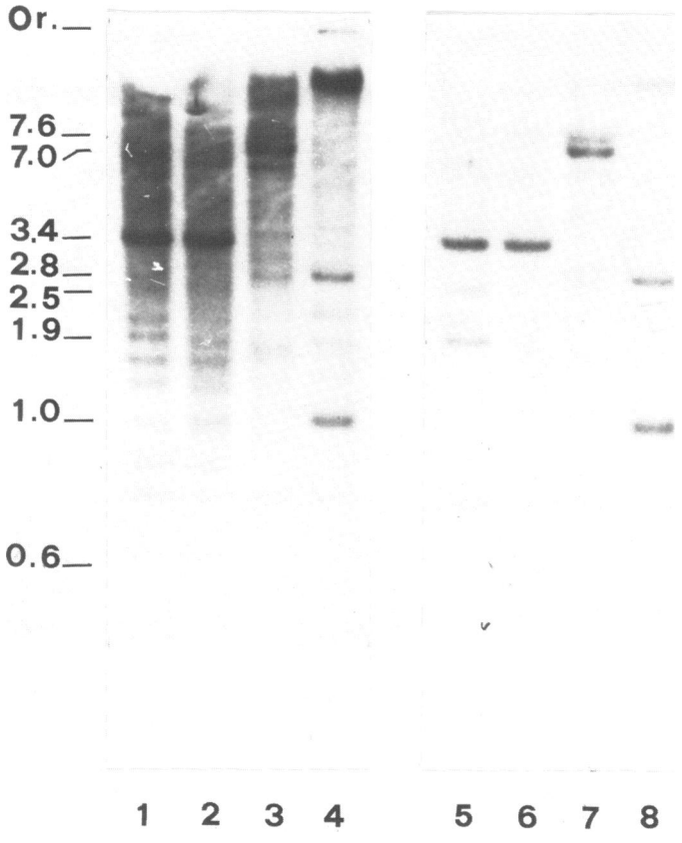

Figure 3. Southern blots obtained under standard (42°C) and stringent (53°C) incubation conditions using restricted Euglena gracilis DNA. DNA (15 μg) was restricted and filter imprints challenged with *tef* cDNA (pos. 366 to 1193). Film exposure with intensifying screen was for 36 h. The filter was first hybridized at 42°C and autoradiographed (lanes 1 to 4), then the label was washed out (34) and the same filter rehybridized at 53°C (lanes 5 to 8). HincII, lane 1,5; AccI + HincII, lane 2,6; AccI, lane 3,7; PstI, lane 4,8; fragment sizes in kb.

Structure and arrangements of EF-1α gene(s)

Single copy DNA sequences comprise about 12% of the *Euglena gracilis* nuclear genome and DNA fragments of 2 kb may contain both single copy and repetitive DNA sequences (11). But so far no detailed information is available concerning the arrangement of specific genes. Using a ^{32}P-labelled cDNA probe (pos. 365 to 1193, Fig. 1) we hybridized filter imprints of nuclear DNA restricted, respectively, with HincII, AccI + HincII, AccI, PstI under standard conditions as specified in Methods. In lanes 1 to 4 of Fig. 3 we see multiple bands, some of them with regular spacing of size intervals between 200 to 300 nucleotides. Interpretation of these banding patterns is difficult, but shifting the hybridization temperature from 42° to 53°C results in much simpler patterns (Fig. 3, lanes 5 to 8) amenable to reasonable predictions of gene anatomy.

At least two features may be the reason for this observation: a) The *Euglena tef* gene family is rather large and members have diverged, interacting more or less strongly with the *tef* cDNA probe; b) and/or *tef* related sequences (e.g. pseudogenes) occur in the genome carrying integrated repetitive elements (see Discussion). In the following hybridisation experiments the stringent conditions (53°C) were used. In order to assess co-linearity between transcript and gene we determined the length of *tef* DNA fragments obtained by restriction of nuclear DNA (Fig. 4,B) and compared these results with the values calculated

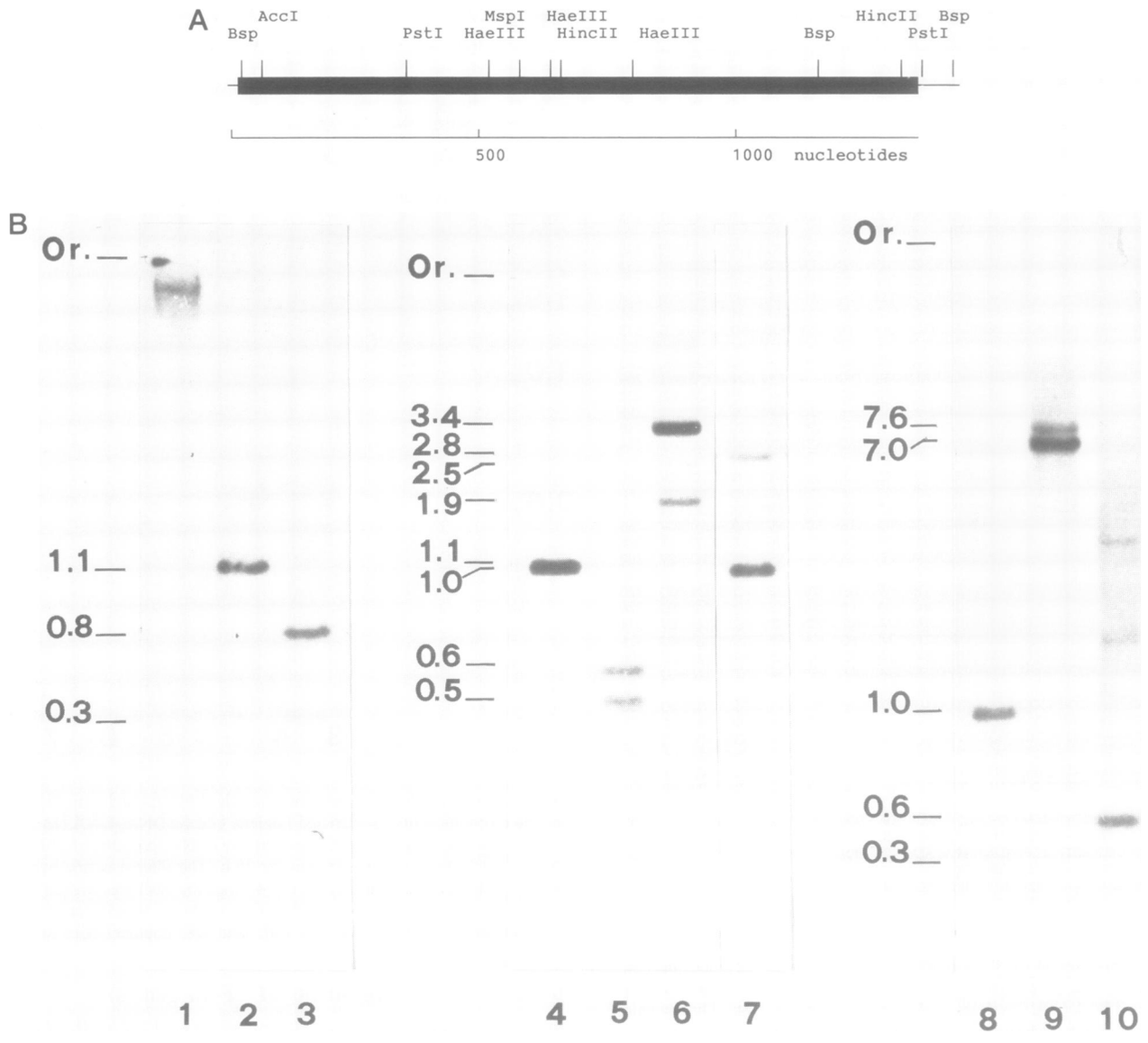

Figure 4. Southern blot analysis for testing transcript co-linearity. A diagram of relevant restriction sites according to sequencing data are shown in (A). Bold line marks coding part of *tef*. Radiograms of filter imprints are given in (B). Restricted DNA samples were 15 μg (lanes 1 to 3, 8 to 10), 6 μg (lanes 4 to 7). Agarose gels were 1.5% (lanes 1 to 7) and 1% (lanes 8 to 10). DNA probes were pos. 9 to 1193 in lane 1 to 3; pos. 9 to 1193 in lane 4 to 7; pos. 9 to 670 in lanes 8 to 10. Film exposure was with intensifying screen for 14 days. 1) Unrestricted DNA; 2) Bsp1286; 3) Bsp1286 + PstI; 4) Bsp1286; 5) Bsp1286 + HincII; 6) HincII; 7) PstI; 8) PstI + AccI; 9) AccI; 10) AccI + HincII. Fragment sizes in kb.

from the sequencing experiment (Table 1). For convenience the relevant restriction sites are diagrammed (Fig. 4,A). We see that good length coincidence exists within the coding part, suggesting that the coding part is not interrupted by an intron of sizable length. However, it was reported e.g., that all three *tef* genes of Mucor contain intron(s), the smallest intron being only 55 nucleotides long. Only genome *tef* sequencing will yield the final answer.

Further hybridization experiments confirming above results were made with PstI:HaeIII and Bsp1286:HaeIII restricted DNA (not shown). We also identified a genomic DNA HaeIII fragment of close to 150 nucleotides long using as DNA probe the HaeIII fragment (pos. 639 to 787). As listed in Table 1 we could not detect a 643 nucleotide HincII fragment. Since we can exclude a technical sequencing error we conclude that either the HincII 1313 site does not exist in all or in most of the *tef* gene copies (see below) or the site is partially protected, e.g., by methylation.

Number and general organization of EF-1α gene(s) were studied in hybridization experiments shown in Fig. 5. The 5′ DNA probe (pos. 9 to 670) recognizes five HincII fragments (1.5, 1.9, 2.5, 3.4 and 5.6 kb) (Fig. 5, lanes 3 and 4) and four PstI fragments (1.0, 2.8, 4.2 and 20 kb) (lane 6). Some fragments interact very strongly, others are very faint. The 3′ DNA probe (pos. 671 to 1193) strongly interacts with a single HincII fragment of 3.4 kb (lane 7) and three PstI fragments (1.0, 2.8 and 4.2) (lane 9). The combined restriction (HincII + PstI) yields four fragments interacting with the 5′ DNA probe which all are shortened by about 300 nucleotides (e.g. 1.9 shifts to 1.6 kb, lane 5). The 3.4 kb band disappears, the conclusion being that the faint 3.4 kb band in lane 4 is due to contamination of the 5′ DNA probe by some 3′ DNA probe, and it is equivalent to the strong 3.4 kb band in lane 7. The four other bands in lane

Table 1. Comparison of calculated (a) and experimentally determined (b) restriction fragment lengths of the EF-1α coding sequences

RE position*	Length (a) nucleotides	Length (b) kb	Reference
Bsp1286			
50:1164	1115	1.1	Fig. 4 lanes 2 and 4
1165:1419	255		Visible on the original
PstI			
366:1373	1008	1.0	Fig. 4 lane 7
HincII			
671:1313	643	–	
Bsp1286:PstI			
1164:366	799	0.8	Fig. 4 lane 3
50:365	316	0.3	Fig. 4 lane 3
Bsp1286:HincII			
1164:671	494	0.5	Fig. 4 lane 5
50:670	621	0.6	Fig. 4 lane 5
AccI:HincII			
91:670	580	0.6	Fig. 4 lane 10
AccI:PstI			
91:365	275	0.3	Fig. 4 lane 8

*Position numbers refer to Fig. 1

4 correspond to four kinds of fragments which are at the 5′ site of the coding region and extend to various degrees into the upstream region. From this we conclude that there are at least four kinds of EF-1α gene loci. The 20 kb PstI fragment (lane b) can only be placed upstream of the Pst 365 site since it does not interact with the 3′ DNA probe. The PstI 2.8 and 4.2 kb fragments (lane 9) interact strongly with the 3′ DNA probe, indicating that these fragment start at the PstI 365 site and extend beyond the PstI 1373 site just outside of the coding part. Again we must assume that this PstI site is partially protected or absent in some gene copies. PstI 1373 is certainly present in the genome as documented by the 1.0 kb fragment (Fig. 4, lane 7).

Additional evidence for several *tef* gene loci is obtained from results obtained with HincII plus HindIII digests. HindIII alone yields two fragments of 8.6 and 9.6 kb interacting with the 5′ probe (Fig. 5, lane 1). HindIII shortens the 1.9 kb HincII fragment (lane 3) by about 300 nucleotides to 1.6 kb (lane 2). The faint 5.6 kb HincII (lane 3) disappears and the signal of the 2.5 kb HincII + HindIII band (lane 2) is reinforced, suggesting that HindIII cuts the 5.6 kb HincII fragment to about the same length as the 2.5 kb HincII fragment. The faint 3.4 kb band (contamination of probe) remains uncut. We propose that one kind of *tef* genes, represented by the 1.9 kb HincII fragment, carries a HindIII site about 900 nucleotides upstream of the 5′ end of the coding part, while the other type (5.6 kb HincII) has a HindIII site about 1.8 kb upstream of the 5′ end of the coding part. This HindIII site would be very close to the upstream HincII site of the 2.5 kb HincII fragment, but we cannot decide whether the gene represented by the 2.5 kb HincII band also carries a HindIII site. The existence of these two HindIII sites was confirmed by results obtained with HindIII + PstI restricted DNA (not shown). In Fig. 5,A we diagram the relevant restriction sites in four kinds of EF-1α gene loci.

According to Fig. 5, lane 1 two types of HindIII fragments carry EF-1α genes. One is about 9.6 kb, the other about 8.6 kb long. This is in line with the relative map positions of the two HindIII sites in the upstream region of the coding part. This means that the HindIII sites in the downstream region are at (about) identical positions relative to the coding part. Also the AccI banding pattern (Fig. 4, lane 9) clearly demonstrates the existence of two bands with 7.0 kb and 7.6 kb. In Fig. 5, lane 6 we see PstI fragment(s) (20 kb) which must contain the 5′ end region of the coding part and a large upstream segment. Considering these results, we conclude that EF-1α gene loci are dispersed in the genome and, e.g., better than 20 kb apart. In the MspI pattern (Fig. 5, lane 8) we notice a 0.4 kb band interacting with the 3′ DNA probe. We sequenced a single MspI site (pos. 582) but none was identified towards the 3′ end, i.e., there are *tef* sequences in the genome carrying this second MspI site.

Expression of EF-1α gene(s)

The Southern analysis gives good evidence for the existence of several (e.g. four) *tef* gene loci, varying in abundance in the genome. Furthermore we find *tef* related sequences, whose organisation and function, if any, remains obscure.

To test gene(s) expression we probed RNA filters with the *tef* DNA probe (pos. 366 to 1193). RNA was isolated from cells grown in the dark and then exposed for various time to light. We show in Fig. 6 that, independent of light exposure, a single major RNA band of 1.5 kb lights up. No sizable length difference from one sample to the next was detectable. We conclude that the various gene loci (if all are transcribed) yield a uniform kind of stable mRNA and *tef* related sequences seem not to be stably transcribed.

The 5′ end of the stable transcript was determined by primer extension for three RNA samples harvested from cells exposed for 0, 12 and 24 hours light. Again we identified a single 5′ end (not shown) and the dideoxy sequencing result given in Fig. 7 corroborates this result. A single 5′ start sequence reading 5′ TTTCTGAG.. (see also Fig. 1) is detected. Both, Northern and sequencing data indicate that neither expression level nor processing of the EF-1α gene(s) transcripts vary in function of growth conditions (dark-light shift).

DISCUSSION

EF-1α protein

Beck and Spremulli (3) have purified and partially characterized an abundant *Euglena* cytosolic protein of apparent MW of 56′000 which had strong translation elongation activity with wheat germ ribosomes. According to our results the *tef* gene (cDNA) codes for a protein of MW 48′515 which has a high degree of sequence identity with other eukaryotic EF-1α proteins. In spite of the discrepancy in MW we postulate that the purified protein is the translation product of the sequenced *tef* region. Positive arguments are a) A single major class of mRNA of about 1.5 kb was found to interact with *tef* DNA. b) A single 5′ end was identified by primer extension and reverse transcriptase sequencing. c) There exists excellent coincidence between mRNA length (Northern) and sequenced cDNA. So we conclude that the apparent MW 56′000 is an overestimate due to analytical circumstances.

Tef gene expression

Sofar, knowledge about the organisation of specific nuclear genes in *Euglena* is scarce. Schantz and coll. (13,14) investigated the CAB genes of PSI and PSII on a nucleotide level and found that gene families exist in both cases without, however, determining their number or general organisation. Quite interestingly, they found that in both cases the cDNA probes interact with large transcripts (e.g. 4.7 and 7.5 kb) which contain information for several consecutive proteins. On the other hand they found that

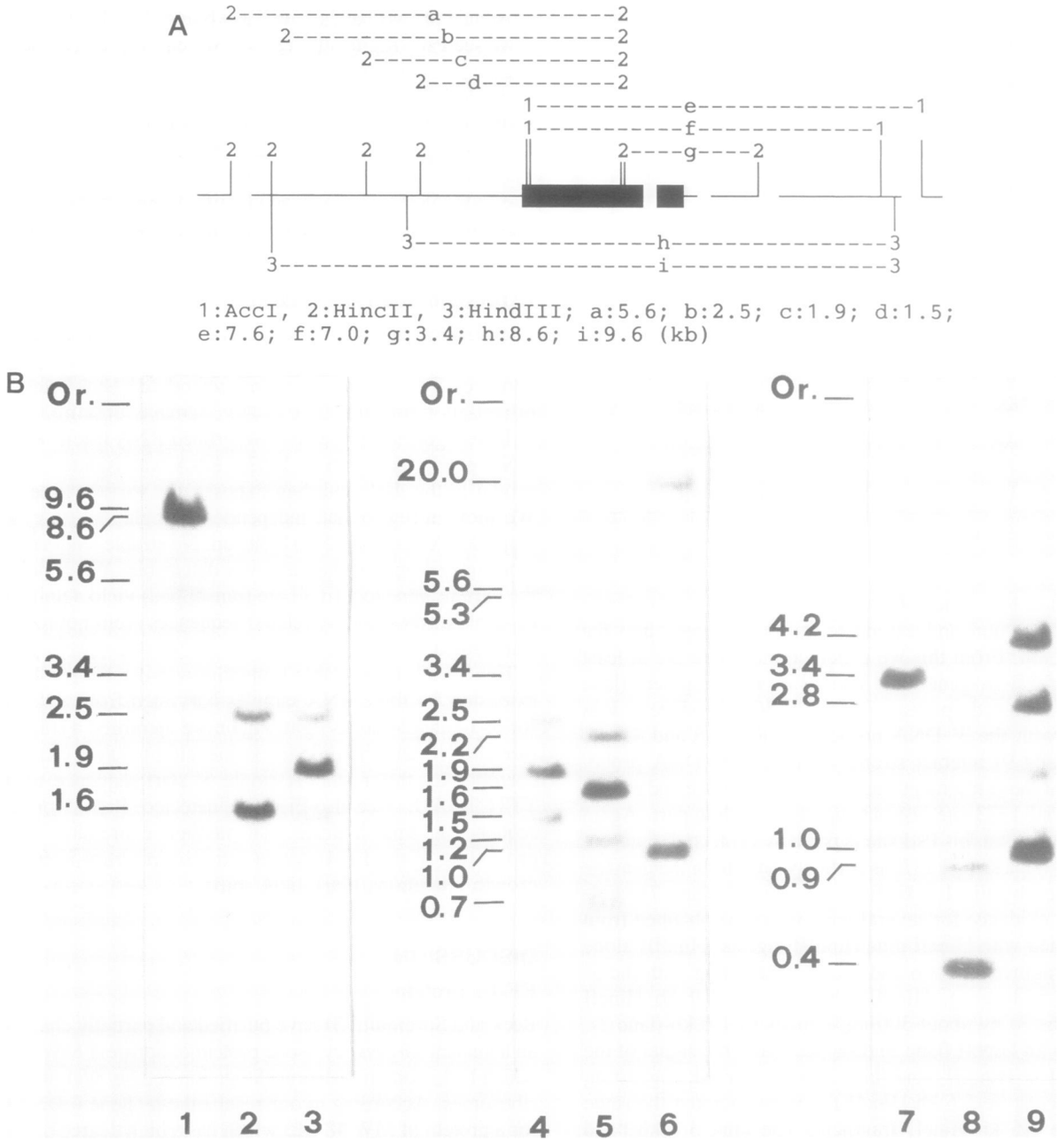

Figure 5. Southern blot analysis for identifying different tef gene loci. A: A diagram of restriction sites in the 5′- and 3′-vicinity of different *tef* loci according to Southern analysis. Gapped bold line marks coding part of *tef*. Gapped horizontal solid line marks DNA segments. Distances between restriction sites are marked by dashed lines with letters which correlate with the distances given in kb (see footnote). Gaps indicate that the drawing is not in scale. Single and double vertical bars mark restriction sites identified by Southern analysis and sequencing, respectively. B: Radiograms of filter imprints. Restricted DNA samples were 15 μg, separated in 1% agarose gels, and hybridized at 53°C. Film exposure, 14 days with intensifying screen. DNA probes were pos. 9 to 670 in lanes 1 to 6; pos. 671 to 1193 in lanes 7 to 9. 1) HindIII; 2) HindIII + HincII; 3) HincII; 4) HincII; 5) HincII + PstI; 6) PstI; 7) HincII; 8) MspI; 9) PstI. Fragments length in kb.

a *Euglena* cDNA sequence encoding β-tubulin (15) strongly hybridizes with a single 1.9 kb mRNA, what corresponds to a 'normal' length. A tentative conclusion is that large mRNAs coding for proteins destined for chloroplast import are translated into polyproteins which undergo post-translational cleavage, while mRNA encoding cytosolic proteins are of the usual length. This assumption is corroborated by our results, i.e., the major stable transcript encoding EF-1α is of the expected length (1.5 kb), on the other hand we have seen that a *Euglena* DNA probe coding for the RUBISCO small subunit protein interacts also with a huge mRNA (Montandon, unpublished).

Looking at the 5′ untranslated sequence we notice close to the start codon the sequence -GTGTCTA$(T)_8$CG-. Exactly the same nucleotide sequence (17 mer) is found upstream and close to the start codon of the β-tubulin gene. We suggest this sequence to be involved in controlling the expression of *Euglena* cytolosic proteins.

The unicellular protist *Euglena gracilis* can differentiate from an etiolated cell to a green cell upon exposure to light. We wondered whether this differentiation step requires a differential expression of *tef* gene(s). This seems not to be the case since no significant qualitative nor quantitative change in *tef* mRNA

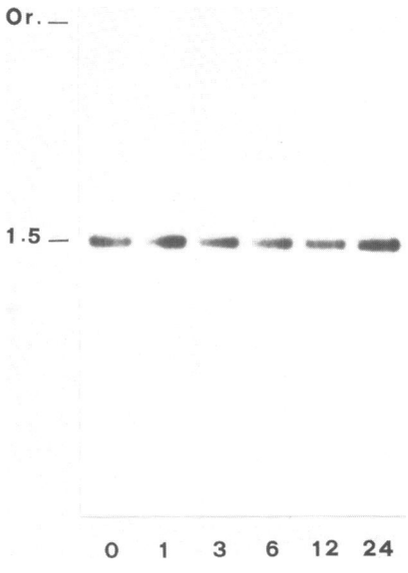

Figure 6. Northern blot analysis to identify transcripts of tef gene(s) in dark and light grown Euglena gracilis. 1 μg of total RNA obtained from *Euglena* cells grown in the dark (0) or exposed for 1,3,6,12,24 hours to light was hybridized to *tef* DNA (pos. 366 to 1193). Film exposure with intensifying screen was for 4 days. Size in kb.

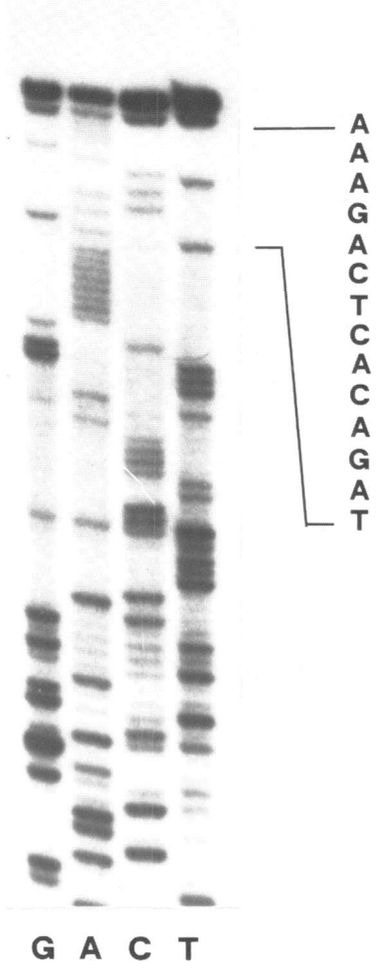

Figure 7. Dideoxysequencing by primer extension of the 5′ end region of the tef mRNA. The sequence complementary to the 5′ end of the RNA is given (consult also Fig. 1).

could be detected in function of growth conditions. Stage specific differences in *tef* gene expression were reported, e.g., for the fungus *Mucor racemosus* (5) and *Drosophila melanogaster* (9).

Tef gene organisation

The Southern hybridization experiments done under standard conditions yielded a very complex result. A striking observation was that regularly spaced bands appeared in DNA restricted with frequently cutting enzymes. *Euglena gracilis* nuclear DNA probably contains *tef* pseudogenes with integrated repetitive elements as, e.g., described for a mouse gene family (31).

The results obtained under more stringent conditions revealed that 1) Within the coding part there essentially exists colinearity between transcript and gene(s). At least we can exclude the existence of an intron larger than 50 nucleotides. 2) the *tef* gene family has at least four members which show different restriction patterns near the 5′ end of the coding part and are probably not equally abundant in the genome as suggested by the different signal intensities under comparable conditions. Differences in signal intensities might also be due to sequence divergence. We cannot exclude, of course, that partial modification of restriction sites (e.g. HincII) contributes to the multiplicity of hybridizing fragments. The cDNA probes used are expected to be copies of the most abundant mRNAs and therefore we may argue that, e.g., the *tef* gene represented, respectively, by the AccI 7.0 kb, HindIII 8.6 kb and HincII 1.9 kb, which all strongly interact with the same DNA probe, is the prominent member of the *tef* gene family. Experiments are under way to elucidate this question on the genome level.

ACKNOWLEDGEMENTS

We thank Dr. P.S. Sypherd U.C., Irvine for the *Mucor* DNA probe. We are grateful to C. Bettinelli for secretarial and S. Marc-Martin and M.C. Grand-Guillaume-Perrenoud for technical help. These studies receive support from Fonds National Suisse de la Recherche Scientifique (to E.S.).

REFERENCES

1. Montandon,P.E. and Stutz,E. (1983) Nucleic Acids Res., 11,5877–5892.
2. Montandon,P.E. and Stutz,E. (1987) Nucleic Acids Res., 15,7809–7822.
3. Beck,C.M. and Spremulli,L.L. (1982) Arch. Biochem. Biophys. 215,414–424.
4. Nagashima,K., Kasai,M., Nagata,S. and Kaziro,Y. (1986) Gene, 45,265–273.
5. Linz,J.E. and Sypherd,P.S. (1987) Mol. Cell. Biol. 7,1925–1932.
6. Pokalsky,A.R., Hiatt,W.R., Ridge,N., Rasmussen,R., Houck,C.M. and Shewmaker,C.K. (1989) Nucleic Acids Res. 17,4661–4673.
7. Axelos,M., Barolet,C. Liboz,T., Le Van Thai,A. Curie,C. and Lescure,B. (1989) Mol. Gen. Genet., in press.
8. Lenstra,J.A., Van Vliet,A., Arnberg,A.C., Van Hemert,F.J. and Möhler,W. (1986) Eur. J. Biochem., 155,475–483.
9. Hovemann,B., Richter,S., Walldorf,U. and Cziepluch,C. (1988) Nucleic Acids Res., 16,3175–3194.
10. Leedale,G.F. (1958) Nature, 181,502–503.
11. Rawson,J.R.Y., Eckenrode,V.K., Boerma,C.L. and Curtis,S. (1979) Biochim. Biophys. Acta, 563,1–16.
12. Rawson,J.R.Y. (1975) Biochim. Biophys. Acta, 402,171–178.
13. Houlne,G. and Schantz,R. (1987) Curr. Gen., 12,611–616.
14. Houlne,G. and Schantz,R. (1988) Mol. Gen. Genet., 213,479–487.
15. Schantz,M.L. and Schantz,R. (1989) Nucleic Acids Res., 17,6727.
16. Vasconcelos,A.C., Pollak,M., Mendiola,L.R., Hoffmann,H.P., Brown,D.H. and Price,C.A. (1971) Plant Physiol., 47,217–221.
17. Woodcock,E. and Merrett,M.J. (1980) Arch. Microbiol., 124,33–38.
21. Huynh,T.V., Young,R.A. and Davis,R.W. (1985) In Glover,D.M. (ed.), DNA Cloning, IRL-Press, Oxford, Vol. I, pp. 49–78.

22. Hannahan,O. (1985) In Glover,D.M. (ed.), DNA Cloning, IRL Press, Oxford, vol. I, pp. 109–135.
23. Grunstein,M. and Hogness,D. (1975) Proc. Natl. Acad. Sci. USA, 72,3961–3965.
24. Sanger,F., Coulson,A.R., Barrel,B.G., Smith,A.J.H. and Roe,B.A. (1980) J. Mol. Biol., 143,161–178.
25. Müller,R., Slamon,D.J., Tremblay,J.M., Cline,M.J. and Verma,I. (1982) Nature, 299,640–644.
26. Goldenberg,C.J. and Hauser,S.D. (1983) Nucleic Acids Res., 11,1337–1348.
27. Ortiz,W., Reardon,E.M. and Price,C.A. (1980) Plant Physiol., 66,291–294.
28. Curtis,S.E. and Rawson,J.R.Y. (1981) Gene, 15,237–247.
29. Hunt,A.G., Chu,N.M., Odel,J.T., Nagy,F. and Chua,N.H. (1987) Plant Mol. Biol., 8,23–35.
30. Lechner,K. and Böck,A. (1987) Mol. Gen. Genet., 208,523–528.
31. Man,Y.M., Delius,H. and Leader,D.P. (1987) Nucleic Acids Res., 15,3291–3304.
32. Au,G. and Friesen,J.D. (1980) Gene, 12,33–39.
33. Nagata,S., Nagashima,K., Tsunetsugu-Yoka,Y., Fujimura,K., Miyazaki,M. and Kaziro,Y. (1984) EMBO J., 3,1825–1830.
34. Mason,P.J. and Williams,J.G. (1985) In Hames,B.D. and Higgins,S.J. (eds) Nucleic Acid Hybridization, A practical approach, IRL Press, Oxford, pp. 113–160.