THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 269, No. 13, Issue of April 1, pp. 9436-9444, 1994
Printed in U.S.A.

Isolation and Characterization of cDNA Clones for Chloroplast Translational Initiation Factor-3 from *Euglena gracilis**

(Received for publication, November 3, 1993)

Qiong Lin‡, Lan Ma‡, William Burkhart§, and Linda L. Spremulli‡¶‖

From the ‡Department of Chemistry and the ¶Lineberger Comprehensive Cancer Research Center, University of North Carolina, Chapel Hill, North Carolina 27599 and the §Department of Structural and Biophysical Chemistry, Glaxo Research Institute, Research Triangle Park, North Carolina 27709

A complete cDNA clone encoding *Euglena gracilis* chloroplast translational initiation factor 3 (IF-3_{chl}) has been obtained. Analysis of the sequence indicates that the IF-3_{chl} mRNA contains the spliced leader found at the 5′ end of nuclear encoded mRNAs in *E. gracilis*. The open reading frame for IF-3_{chl} encodes a 537-amino acid protein. IF-3_{chl} appears to be divided into four domains. The first 140 amino acids correspond to a transit peptide required for the import of IF-3_{chl} into the chloroplast. The mature form of IF-3_{chl} encompasses domains 2–4 and is about twice the size of *Escherichia coli* IF-3. The second domain has no homology to other known proteins. It begins with a stretch of 35 residues, of which about 30% are proline. Downstream from this region is a stretch of about 25 amino acids with a repeating $(GX)_n$ motif followed by a very acidic region. The third domain comprises a region of about 175 residues and has between 31 and 37% homology to the IF-3s found in other organisms. The IF-3 homology domain is followed by an acidic region which has no detectable homology to other sequences. Analysis of *E. gracilis* genomic DNA suggests that there are about four copies of the IF-3_{chl} gene, one of which is probably a pseudogene. The activity of IF-3_{chl} is inducible by light. However, the IF-3_{chl} mRNA is present in approximately equal amounts in both dark- and light-grown cells, suggesting that the light-dependent induction of IF-3_{chl} activity is post-transcriptional.

Two chloroplast translational initiation factors (IF-2_{chl} and IF-3_{chl})[1] have been purified and characterized from *Euglena gracilis* (1–3). IF-2_{chl} promotes the binding of fMet-tRNA to chloroplast 30 S ribosomal subunits. It is a large complex factor ranging in size from 200 to 700 kDa. The active forms of IF-2_{chl} occur as dimeric, tetrameric, and probably hexameric complexes of 97–200-kDa subunits. This factor has little activity in binding fMet-tRNA to the *Escherichia coli* ribosomes. IF-3_{chl} enhances the poly(A,U,G)-dependent binding of the initiator tRNA to chloroplast 30 S subunits (3, 4). In *E. coli* IF-3 is thought to play several roles in the initiation process (5, 6). First, it binds to the 30 S subunit and inhibits its association with the 50 S subunit. In this way, IF-3 helps ensure a supply of 30 S subunits for initiation complex formation. Second, IF-3 promotes the conversion of a complex referred to as the preternary complex to the ternary complex (7, 8). The preternary complex contains both the mRNA and the fMet-tRNA on the 30 S subunit, but they are not yet hydrogen-bonded. In the ternary complex, the mRNA and fMet-tRNA interact stably through codon:anticodon hydrogen bonding. IF-3 also plays a proofreading role in initiation by promoting the selection of the initiator tRNA as opposed to elongator tRNAs in the 30 S initiation complex (9–11). No factor directly comparable to IF-3 is found in the eukaryotic cytoplasmic protein biosynthetic system.

IF-3_{chl} can replace *E. coli* IF-3 in promoting 70 S initiation complex formation on *E. coli* ribosomes. This observation suggests that the chloroplast factor has retained many of the features and functions of the corresponding prokaryotic factors. However, IF-3_{chl} is physically unusual compared to the bacterial factor. *E. coli* IF-3 is a monomeric protein with a molecular mass of 22 kDa. IF-3_{chl} has been resolved into three forms designated α, β, and γ. All of these forms of IF-3_{chl} function as monomers but are about twice the size of the corresponding prokaryotic factors. *E. coli* IF-3 is a basic protein, which has a strong RNA binding activity (12). It binds to the 30 S subunit by a specific interaction with the 16 S rRNA. In contrast, IF-3_{chl} is an acidic protein, which does not have a strong RNA binding activity.

Both IF-2_{chl} and IF-3_{chl} are the products of nuclear genes in *E. gracilis*. In addition, the activities of both proteins are induced by exposure of the cells to light (13), as is the development of the chloroplast itself. Chloroplast proteins encoded by nuclear genes are generally synthesized in the cytoplasm as precursors that have extensions at the N terminus. The chloroplasts of *E. gracilis* are surrounded by three membranes as opposed to the two plastid membranes observed in higher plants and in many algae. The transit peptide for chloroplast proteins in this organism must, therefore, contain the information required to direct the precursor polypeptide through all three of these membranes.

EXPERIMENTAL PROCEDURES

Materials—Oligonucleotide primers were made in the Lineberger Comprehensive Cancer Center at the University of North Carolina. [α-^{32}P]dCTP (3000 Ci/mmol) was obtained from DuPont NEN. Pure nitrocellulose blotting membranes were obtained from Schleicher & Schuell, and Zeta-Probe Blotting Membranes were purchased from Bio-Rad. Restriction enzymes were purchased from U. S. Biochemical Corp., Pharmacia, and Life Technologies, Inc. Oligo(dT)-cellulose Type 3 was obtained from Collaborative Research. Sequenase® version 2 DNA sequencing kit was from U. S. Biochemical Corp. Analysis of the sequence was done with the Genetics Computer Group (GCG) sequence analysis programs (version 7.2, October, 1992) running on the University of North Carolina VAX computer.

Light Induction of Chloroplast Development in E. gracilis—The growth of *E. gracilis* Klebs var. *bacillaris Cori* (*E. gracilis* B) and the light induction of chloroplast development were carried out basically as

* This work has been supported in part by National Institutes of Health Grant GM24963. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) L23760.

‖ To whom correspondence should be addressed.

[1] The abbreviations used are: IF, initiation factor; bp, base pair(s); kbp, kilobase pair(s); kb, kilobase(s); PCR, polymerase chain reaction; T_p, annealing temperature; SL, spliced leader; ER, endoplasmic reticulum.

described (13). Cells in 50 ml of media were grown in the dark to a cell density of 2×10^6 cells/ml. The cultures were transferred to 200 ml of resting media and kept in the dark for 3 more days in order to exhaust the carbon source. The cells were then either maintained in the dark or exposed to light for 72 h. Chloroplast development was monitored by determining the chlorophyll content per cell (14). Cells (0.32–0.58 g) were harvested by centrifugation, and RNA was extracted according to the procedure described (15).

Purification of IF-3$_{chl}$ and Peptide Sequence Analysis—IF-3$_{chl}$ β (molecular mass 42 kDa) was purified as described (3), except that the TSKgel heparin-5PW column was omitted. After the TSKgel DEAE-5PW column, about 30 μg of pure IF-3$_{chl}$ β was obtained for partial amino acid sequence analysis. For this procedure, a sample of IF-3$_{chl}$ β was applied to a Hewlett-Packard hydrophobic sequencing column according to the manufacturer's protocol. The protein was reduced and alkylated *in situ* by dripping a buffer containing 20 mM dithiothreitol, 50 mM 4-vinylpyridine in 6 M guanidine hydrochloride containing 0.5 M Tris-HCl, pH 8.6, and 5 mM EDTA through the column at room temperature for 15 min. *In situ* proteolytic digestion with endoproteinase Lys-C (Wako Chemicals) was performed according to Burkhart (16). *In situ* CNBr cleavage was performed on a Hewlett-Packard Protein Chemistry Station according to the manufacturer's protocol using a 6-h cleavage time. After digestion, peptides were separated by reverse phase high performance liquid chromatography using a Hypersil ODS column (2.1 × 100 mm; Hewlett-Packard) employing a linear gradient of 8–48% acetonitrile in 0.1% trifluoroacetic acid over 80 min with a flow rate of 100 μl/min. Peptides were sequenced on the Applied Biosystem model 477A Protein Sequencer with the model 120A PTH Analyzer.

RNA Isolation and Construction of cDNA Libraries—Total RNA was extracted from light-grown *E. gracilis* essentially as described (17). DNA was removed by lithium chloride precipitation. Poly(A)$^+$ RNA was purified by oligo(dT)-cellulose chromatography (18). First strand cDNA was synthesized by reverse transcription of 5 μg of poly(A)$^+$ RNA using either random hexamers or oligo(dT)$_{12-18}$ as primers and Moloney murine leukemia virus reverse transcriptase (Pharmacia). A Sephacryl^R S-400 spun column was used to select cDNAs over 400 bp. Second strand cDNA was synthesized using RNase H nicking, followed by DNA synthesis using *E. coli* DNA polymerase I as suggested by Stratagene. *Eco*RI/*Not*I adaptors were ligated to the double-stranded cDNA by blunt end ligation using T4 DNA ligase. The cDNA was ligated to *Eco*RI-digested λ ZAP II (Stratagene) and packaged with λ phage packaging extract (Stratagene). The random-primed library yielded a total of 3×10^6 independent recombinants, while the oligo(dT)-library contained 5×10^4 independent plaques. The latter library was amplified in *E. coli* XL1-Blue resulting a library containing 1×10^8 plaque-forming units/ml.

Preparation of Specifically Primed cDNA—A first strand cDNA was prepared by priming 5 μg of *E. gracilis* poly(A)$^+$ RNA with 60 pmol of Primer P4 (see Table II) and 1200 units of Moloney murine leukemia virus reverse transcriptase basically as described (19). Aliquots corresponding to 1/10 of the cDNA synthesis reaction were used in the PCR amplification reactions described below.

PCR Amplification—PCR amplifications were performed in a Perkin-Elmer DNA Thermal Cycler. All the primers used carried a convenient restriction site at their 5′ end (see Table II). Reaction mixtures (100 μl) contained 60 pmol of each primer, 2.5 units of *Taq* DNA polymerase (Promega), 0.2 mM dNTPs, the buffer system recommended by Perkin-Elmer, and either 0.1 μg of genomic DNA or an aliquot of the specifically primed cDNA prepared as described above. The template DNA was denatured by incubation for 5 min at 94 °C followed by 5 min at 85 °C prior to the addition of *Taq* DNA polymerase. Amplification reactions were performed using an annealing temperature (T_p) calculated as described (20). During the first three cycles of PCR, the T_p of the primer without the added restriction site was used. The remaining 35 cycles used the T_p of the primer including the added restriction site. For primer combinations that differed in T_p, the lower value was used in the reaction. During each round of PCR, denaturation of the DNA was done at 94 °C for 1 min, primer annealing was at the specified T_p for 1.5 min, and polymerization was carried out at 72 °C for 2 min. In the last cycle, the reaction time at 72 °C was extended to 10 min to allow completion of chains and the incubations were then cooled to 4 °C.

PCR reactions involving a second round of amplification using different primers were prepared as described above, except that they contained 1 μl of the reaction mixture obtained from the first round of PCR. The reaction mixtures were analyzed on 1.4% agarose or 3% NuSieve® GTG-agarose gels. Specific bands were identified by ethidium bromide staining and eluted from the gel using the GeneClean or Mermaid kits from Bio101.

Screening cDNA Libraries—Approximately 1.5×10^6 plaques from the random-primed cDNA library and approximately 250,000 plaques from the oligo(dT)-primed library were plated on *E. coli* XL1-Blue and plaque hybridization was carried out as described (21). Appropriate probes were labeled using the random priming method (22). Positive plaques were selected, replated and rescreened until purified. The pBluescript SK(−) phagemid clones containing positive inserts were excised from the Lambda ZAP II and selected according to the manufacturer's instructions (Stratagene).

DNA Sequencing—All the putative clones were sequenced by the dideoxynucleotide chain termination method (23) and subjected to autosequencing in the University of North Carolina DNA Sequencing Facility. M13 universal and reverse primers, as well as a number of synthetic oligodeoxyribonucleotide primers derived from the IF-3$_{chl}$ sequence were used. Each region was sequenced completely in both directions.

Northern and Southern Blots—RNA was analyzed on 0.9% agarose gels in the presence of 1 M glyoxal and 50% dimethyl sulfoxide as described (24). Total genomic DNA was prepared essentially as described (25), and 20 μg were digested with 200–300 units of the indicated restriction enzyme. Digests were run on a 0.7% agarose gel at 25 V for 17 h. RNA and DNA were transferred from the agarose gels to Zeta-Probe blotting membranes as recommended by the manufacturer. For hybridization reactions the blotted membranes were incubated with 2×10^7 cpm of ^{32}P-labeled probe (2×10^9 cpm/μg) in 1 mM EDTA, 0.5 M NaH_2PO_4, pH 7.2, and 7% SDS at 65 °C for 14–18 h. The membranes were then washed using the standard protocol described by the manufacturer. For reuse of the blot, it was striped by incubation in 0.1 × SSC (0.015 M NaCl, 0.0015 M trisodium citrate) and 0.5% SDS at 95 °C for 30 min. This washing procedure was then repeated prior to reprobing the blot.

RESULTS AND DISCUSSION

Isolation of cDNA Clones for IF-3$_{chl}$—IF-3$_{chl}$ is the product of nuclear gene in *E. gracilis* (13). In order to obtain cDNA clones of this factor, it was necessary to obtain partial peptide sequence information on this protein. Unfortunately, purified IF-3$_{chl}$ is an extremely sticky protein and is easily lost by adsorption onto surfaces when present at the low concentrations resulting from its purification. Techniques commonly used to concentrate and change the buffer of a protein sample lead to significant losses of IF-3$_{chl}$ even when siliconized tubes were used. In order to minimize loses during sequence analysis, a method for the *in situ* digestion of IF-3$_{chl}$ was used (16). With this approach peptides derived from IF-3$_{chl}$ were obtained following digestion with endoproteinase Lys-C or CNBr. Eight peptides ranging from 6 to 34 amino acids in length were obtained from endoproteinase Lys-C cleavage, and five peptides were obtained from the CNBr digestion (Table I). Two sets of peptides overlapped allowing the deduction of two longer stretches of amino acid sequence (peptide A has 54 amino acids, and peptide B has 32 amino acids; Table I). No sequence information could be obtained directly from the N terminus, suggesting that it might be blocked. Several of the peptide sequences obtained showed a significant percentage of amino acid sequence identity to *E. coli* IF-3 (Table I), allowing the prediction of their positions relative to each other. Other peptides, however, showed little or no sequence identity with the prokaryotic factors. This observation is not surprising, since IF-3$_{chl}$ is about twice the size of *E. coli* IF-3 and, thus, would be expected to have significant regions of sequence with little resemblance to the bacterial factor.

Peptides A and B exhibited homology to their prokaryotic counterparts, and their use formed the basis of the initial strategy designed to obtain a portion of the IF-3$_{chl}$ gene. Based on its sequence, peptide A was predicted to be derived from a region of IF-3$_{chl}$ corresponding to a sequence near the N terminus of *E. coli* and *Bacillus stearothermophilus* IF-3 (Fig. 1*A*). Peptide B was predicted to be on the C-terminal side of peptide A. Four degenerate oligonucleotide primers for PCR were designed based on these peptide sequences and on the codon usage of *E.*

TABLE I

*Amino acid sequences of IF-3_{chl} peptides**

Sequences underlined were used for designing degenerate PCR primers. The designation *X* represents an ambiguous residue.

	Amino acid sequence	Homology to *E. coli* IF-3	
		Position	% Identity
Lys C peptides			
1.	LGVMSRQEALELAEAEDIDLVLVSIDTDPPV[a]	34–65	51
2.	RFQNDLADMGK[b]	147–157	30
3.	YPQLXLVXAVPGGRDEK	—[c]	—
4.	VSMEFK[b]	—	—
5.	GSRSSTLSNLGMGLGLGLGXGXGFGXGFGXGRGF	—	—
6.	VIAERRAERDRK[d]		—
7a.	IGQHDYDVRVKQA[d]		—
7b.	FLEGGHRIK[d]	117–125	55
8.	GRENQFVEIGRAVMK[b]	132–146	40
CNBr peptides			
9.	MGLGLGLGKGXGFGXGFGLGXGFPVEEEVEEE[d]	—	
10.	MGKADAVPXKLXTFXILNXAPA	—	
11.	MEFKGRENQFVEIGRAVMKRFQND[d]	128–151	33
12.	MNEDIKYPQLRLVRAVPGGRDEKLGV[a]	15–36	36
13.	MSRQEALELAEAEDIDLVLVSIDTDP[a]	37–64	53

[a] These peptides overlap into peptide A: MNEDIKYPQLRLVRAVPGGRDEKLGVMSRQEALELAEAEDIDLVLVSIDTDPPV.
[b] These peptides overlap into peptide B: VSMEFKGRENQFVEIGRAVMKRFQNDLADMGK.
[c] Indicates the % identity was <20%.
[d] Several ambiguities were noted in the peptide sequence that were resolved following determination of the cDNA sequence.

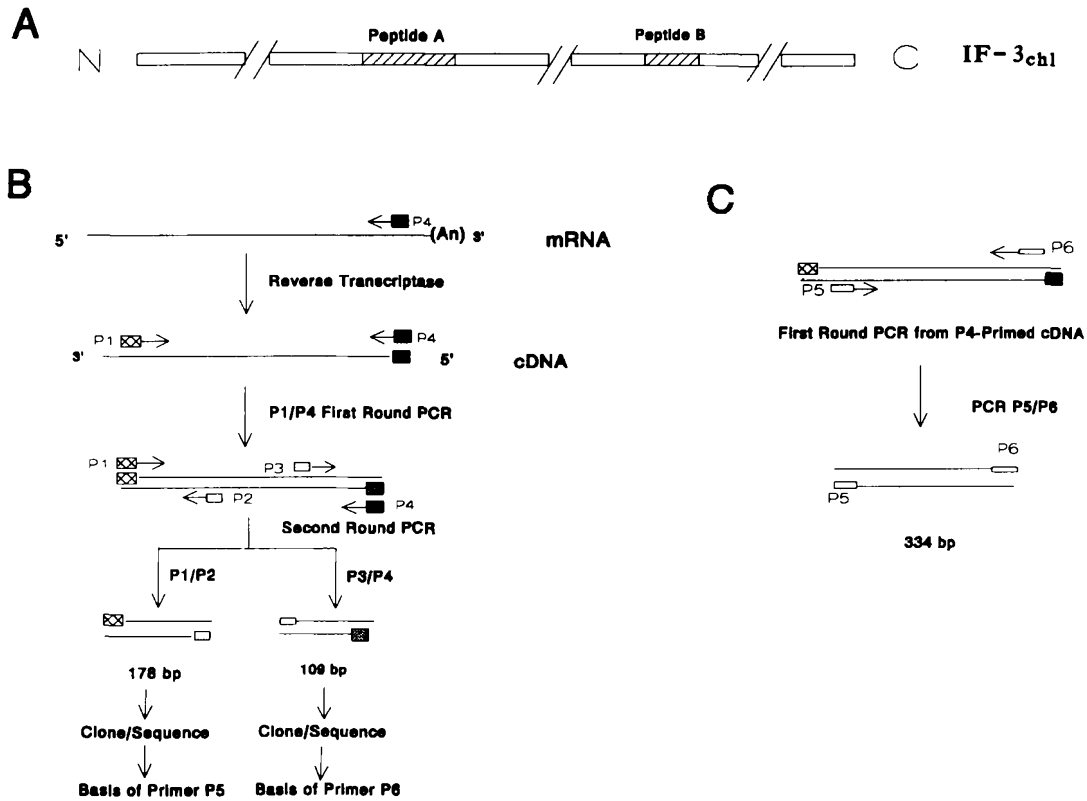

FIG. 1. **PCR strategy used to isolate partial cDNA clones of IF-3_{chl}.** *A*, the relative location of peptides A and B was predicted by alignment with *E. coli* IF-3. The *box* represents the protein sequence of IF-3_{chl} from N to C terminus. The *hatched boxes* represent peptides A and B. *B*, schematic diagram outlining the strategy used in the amplification of a portion of IF-3_{chl} by nested reverse transcriptase-PCR. The primers indicated are shown in Table II. *C*, the amplification of the IF-3_{chl} cDNA probe by PCR using specific primers derived from sequence information determined from the initial clones obtained.

gracilis (Table II). Primers P1 and P2 were derived from the sequences of peptide A. If there are no intervening sequences in this region of the IF-3_{chl} gene, amplification between these two primers should produce a product of 178 base pairs. When genomic DNA was used as a template, however, no band of the appropriate size was observed. A similar approach was used with primers P3 and P4. If this region of the IF-3_{chl} gene does not contain an intron, specific amplification between these two primers should give a product of 109 base pairs. A specific band of about this size was observed following PCR amplification between this pair of primers. This fragment was subcloned into the vector pTZ18R (U. S. Biochemical Corp.) and sequenced. The derived amino acid sequence was consistent with the known protein sequence, indicating that a portion of an IF-3_{chl} gene had been cloned.

In an effort to obtain a portion of the IF-3_{chl} cDNA, primer P4 was selected for use as a primer in the preparation of a specifically-primed cDNA starting with *E. gracilis* poly(A)$^+$ RNA (Fig.

TABLE II
Nucleotide sequences of oligonucleotide primers

Primer	Oligonucleotide (5′ → 3′)
P1	CGGGATCCCATGAATGAGGARATRAAYTARCC
P2	GGAATTCCACIGGIGGATCXGTYTCDAT
P3	CGGGATCCCGTXTCPATGGARTTRAAYGG
P4	GGAATTCCCCCATYTCIGCSAYYTCYTT
P5	CGGGATCCCCGAGAAGCTCGGTGTGATG
P6	GGAATTCCTCTGGAAGCGCTTCATCACC
P7	GGAATTCCGGAGTCCGTGCGCAGCGAT
SL	CGGGATCCTGAGTGTCTATTTTTTTTCG

I is inosine; D is A; G, or T; R is A or G; Y is C or T; X is I or C. The underlined region is the added restriction site.

1*B*). This primer was selected since it appeared to have primed IF-3chl gene sequences specifically in the PCR reaction using genomic DNA. The P4-primed cDNA was used as a template in a first round of PCR using the antisense primer P4 and sense primer P1 (Fig. 1*B*). Gel analysis of this reaction mixture indicated that a large variety of PCR products were present and that the desired product could not be obtained directly by this method. Using this first PCR mixture as a template, additional sets of PCR reactions were carried out utilizing a nested PCR strategy (Fig. 1*B*). Specific DNA fragments of about the expected sizes were observed using primers P1 and P2 together and primers P3 and P4 together. These DNAs were cloned and sequenced. The peptides sequences translated from the DNA sequences matched the known peptide sequences derived from IF-3_{chl}.

In order to obtain a longer probe for use in screening cDNA libraries, two specific primers derived from the determined DNA sequences were prepared (P5 and P6, Table II and Fig. 1*C*). Another PCR reaction was performed using these two primers and the first round PCR product from the P4 primed cDNA as a template (Fig. 1). PCR amplification between these two primers gave a specific product of 343 bp in length. This fragment was used to screen cDNA libraries from *E. gracilis* and as a probe on Northern blots.

A random-primed cDNA library was constructed in λ ZAP II and screened with the PCR-amplified probe. Five positive plaques were obtained, and the phagemids were excised *in vivo* and sequenced. The IF-3_{chl}-derived inserts in these recombinants ranged from 0.4 to 0.9 kbp (Fig. 2). Analysis of these clones provided 1.1 kbp of IF-3_{chl} cDNA sequence. Translation of this sequence revealed that it contained most of the known peptide sequences determined from the purified IF-3_{chl} sample. An oligo(dT)-primed library was also prepared and screened. A 1.0-kbp clone was obtained from this library which contained a poly(A) sequence of about 100 bp and had a TGA stop codon 230 bp from the poly(A) tail. This clone overlapped the clones obtained from the random-primed library, and together they provided 1.5 kbp of IF-3_{chl} sequence information (Fig. 2).

The cDNA clones described above contained a continuous open reading frame through the TGA stop codon, suggesting that they did not include the 5′ end and the start codon. A Northern blot was done in order to obtain an indication of the size of the IF-3_{chl} mRNA, and the results obtained suggested that the mRNA was about 2 kb in length (see below). In order to obtain sequences derived from the 5′ end of the mRNA, we took advantage of the observation that the majority of cytoplasmic mRNAs in *E. gracilis* have the same 26-nucleotide spliced leader (SL RNA), which is transferred to premature mRNAs by trans-splicing (26, 27). A special primer (SL primer, Table II) was designed to hybridize to the 5′ end of IF-3_{chl} mRNA. PCR was then carried out using the P4-derived IF-3_{chl} cDNA as a template and P2 and SL as primers followed by a second round of PCR using the SL primer and P7 (Table II, Fig. 2). The sequence of P7 was derived from the sequenced region of the IF-3_{chl} cDNA obtained from the previous clones. PCR followed by gel analysis indicated that a specific 0.6-kb fragment was amplified by this procedure (data not shown). Sequence analysis verified that it overlapped the previous cDNA clones on the 3′ side and that it contained the SL sequence at the 5′ end. When the DNA sequence information from this clone was combined with that obtained previously, approximately 2 kbp of cDNA sequence had been obtained. This length is very close to the size predicted for the IF-3_{chl} mRNA by Northern analysis.

Primary Sequence of IF-3_{chl}—Sequence analysis of the IF-3_{chl} cDNA clones revealed that the mRNA for this protein has a 42-base 5′-untranslated region, an open reading frame of 1611 base pairs, a UGA termination codon, a 3′-untranslated region of 230 bp, and a poly(A) tail (Fig. 3). Translation appears to start from the first AUG codon. A putative polyadenylation signal (AAUAAA) is not found in the 3′-noncoding region. However, such consensus sequences are present in only about 40% of characterized plant genes (28). The open reading frame codes for a protein of 537 amino acids. This sequence contains all of the peptide sequences obtained from the amino acid sequence analysis of the purified IF-3_{chl}, confirming that the cDNA clones obtained do indeed code for IF-3_{chl}. A few differences were observed between the protein sequence predicted from the cDNA clones and the peptide sequence analysis. These differences are seen in peptides 5 and 9. Amino acid sequence analysis indicated the presence of an Arg at position 246, while cDNA sequence analysis indicated that this position was Pro. There may be several copies of the IF-3_{chl} gene that have slightly different sequences (see below) resulting in the apparent difference between the amino acid and nucleotide sequences observed at this position. In addition, several residues indicated as *X* in the peptides in Table I could not be identified as Lys as suggested by the cDNA sequence data. These positions may contain post-translationally modified residues.

An examination of the sequence of IF-3_{chl} suggests that the protein can be divided into four regions (Fig. 4). The N-terminal region encodes a presumptive chloroplast localization signal, the middle region encodes a portion of the mature IF-3_{chl} protein with no homology to other known proteins. This region is followed by a section of about 175 amino acids that has homology to prokaryotic IF-3. Finally there is a short C-terminal tail region. Each of these parts of IF-3_{chl} is discussed in more detail below.

Characterization of the N-terminal Leading Sequence—The molecular mass of IF-3_{chl} deduced from the nucleotide sequence is 58 kDa. The mature IF-3_{chl} β determined by SDS-polyacrylamide gel electrophoresis and gel exclusion chromatography is about 42 kDa. It is apparent that *E. gracilis* IF-3_{chl} is synthesized as a precursor, which is then processed. Since the N terminus of IF-3_{chl} is blocked, no direct sequence information on the junction between the transit peptide and the mature protein is available. However, based on analogy to the other *E. gracilis* proteins that are synthesized in the cytoplasm and then localized to the chloroplast, a leader sequence of about 140 amino acids would be expected. Excluding the putative transit peptide, the predicted mature form of IF-3_{chl} would have a molecular mass of 43 kDa, which is close to the value observed experimentally (3). Overall, the mature form of IF-3_{chl} is an acidic protein with a pI of 4.9. This observation is consistent with previous work, which indicated that IF-3_{chl} behaves as an acidic protein on ion exchange chromatography (4).

The chloroplasts in *E. gracilis*, unlike those in higher plants, are surrounded by three membranes. The inner two membranes are thought to be similar to the double membrane observed in higher plant chloroplasts, while the origin of the outermost membrane is less well understood. One school of

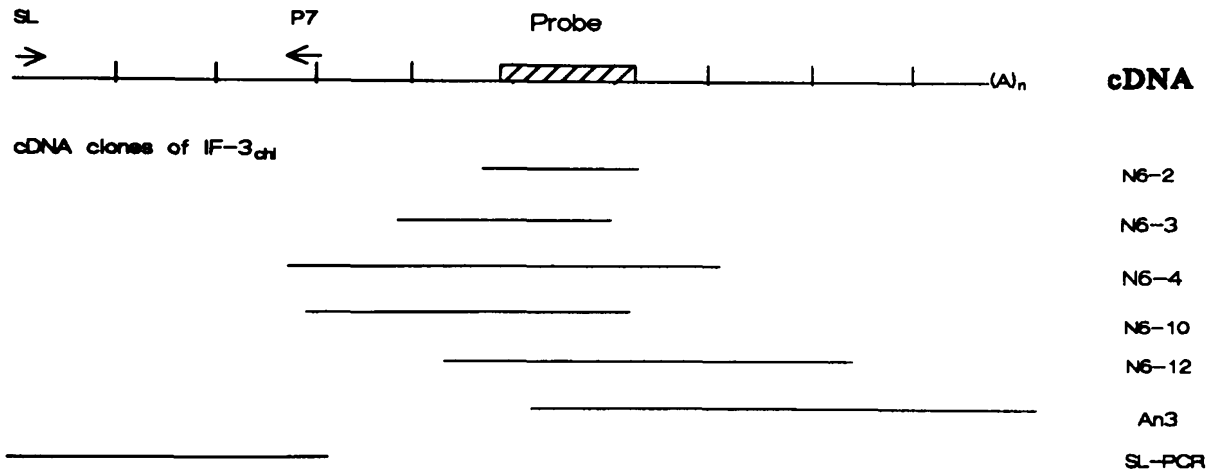

FIG. 2. **IF-3chl cDNA clones isolated from cDNA libraries and the strategy for amplifying the 5′ region of IF-3_{chl} cDNA.** The primers used for the amplification of the 5′ region of the IF-3_{chl} cDNA (SL and P7) are indicated. The relative positions of the inserts derived from IF-3_{chl} sequences present in the *E. gracilis* cDNA clones are represented in the bottom portion of the figure. The clone designated An3 was isolated from the oligo(dT)-primed library. Clones designated N6-n were obtained from the random hexamer-primed library. The clone designated SL-PCR was obtained by amplification of the 5′ end of the IF-3_{chl} mRNA using the SL primer and primer P7 (Table II).

thought suggests that this membrane is derived from the plasma membrane of an original eukaryotic endosymbiotant (29), while a second group believes that it has been derived from the endoplasmic reticulum (30). The transit peptides present in *E. gracilis* proteins targeted to chloroplasts are significantly longer than those observed for higher plant chloroplasts since they must direct localization across this triple membrane system. In general *E. gracilis* chloroplast localization signals appear to have two domains: one that is similar to an ER-targeting sequence and another that has the characteristics of the envelope transit peptides seen in higher plants. It is very likely that the first region is responsible for translocation across the outermost chloroplast membrane, while the second region directs import across the inner two membranes (31). The first 40 amino acids of the IF-3_{chl} leader have a number of features reminiscent of ER localization signals. They contains a stretch of hydrophilic residues at the N terminus followed by a series of non-polar residues (LAAVCATGLLVLVLY) beginning at amino acid 26. Application of the rules of von Heijne *et al.* (32) for the prediction of ER cleavage sites suggests that an ER-type cleavage event could occur after residues 46 or 53. Only one authentic ER transit sequence has been analyzed from *E. gracilis* (33). This sequence resembles those seen in other organisms. It is shorter and has a lower net charge than the "ER-like" leaders observed on chloroplast-localized proteins in *E. gracilis*.

The second domain of the leader peptide is predicted to cover residues from about 50 to 140. This second region corresponds to the chloroplast transit peptide seen for higher plant chloroplasts. These signals generally lack obvious consensus sequences but are enriched in alanine and serine residues and show a lower than average abundance of acidic amino acids. These general features are found in the section of the IF-3_{chl} leader between positions 50 and 140. The amino acid sequence here has about twice the normal abundance of Ala and a somewhat elevated abundance of Ser over the average chloroplast protein. In addition, there are only 2 acidic residues in this 91-amino acid stretch. Von Heijne and Nishikawa (34) have suggested that chloroplast transit peptides are primarily random coils. Secondary structure predictions on residues 50–140 of IF-3_{chl} were carried out using the Chou-Fasman and Garnier-Osguthorpe-Robson algorithms (35, 36). With the exception of the prediction of an α-helical segment between positions 125 and 140, there was no agreement between these two prediction methods on the secondary structure of the putative chloroplast transit sequence.

The Head Region of IF-3_{chl}—The mature form of IF-3_{chl} is thought to include residues from about amino acid 140 through the C terminus at amino acid 537. The mature form is divided into a head, an IF-3 homology domain, and a tail region (Fig. 4). The amino acid composition of each of these regions is quite distinct. The head region is defined as covering residues from position 140 to the beginning of the IF-3 homology domain at amino acid 290. TFASTA analysis of the sequences in the databases did not indicate the presence of any known sequences homologous to the head region of IF-3_{chl}. Analysis using the MOTIFS or BLOCKS programs indicated that this region does not fall into any general family of proteins (37).

The amino acid sequence of the head region shows a number of unusual features. The first part of the head (from residues 140 to 175) is over 30% proline. Overall, the head is about 11% proline compared with about 4–5% for the average protein. The region of the head between residues 254 and 276 has an unusual $(GX)_n$ motif in which Gly alternates with either a basic or large hydrophobic residue. Use of the BLAST program (38) indicated a relationship between the $(GX)_n$ motif and the Gly-rich regions of the keratins and several extracellular structural proteins. The implications of this relationship are unclear. The junction of the head and the IF-3 homology domain is rich in acidic residues with 7 of the last 10 residues in this region being Glu. Secondary structure predictions on the head indicates little α-helix or β-sheet structure, presumably because of the high percentage of Pro and Gly residues. The final 10 residues in the acidic junction between the head and the IF-3 homology domain are predicted to form an α-helix.

The IF-3 Homology Domain—Amino acids 290–474 encompass the IF-3 homology domain. The sequence of IF-3 has been determined from seven different prokaryotic organisms (Refs. 39–41; GenBank). A comparison of their sequences indicates that they are all about 180 amino acids long and are reasonably conserved (Fig. 5 and Table III). In general, these sequences fall into three classes. 1) The IF-3s from a number of Gram-negative organisms are highly conserved (greater than 80%). 2) There is one sequence from a Gram-positive organism (*B. stearothermophilus*), which has about 50% sequence identity to *E. coli* IF-3. 3) Finally, there is the sequence from *Mycoplasma fermentans*, which has between 33 and 39% identity to other IF-3 sequences. The *E. gracilis* IF-3_{chl} appears to fall into a

```
TGAGTGTCTATTTTTTTTCGGGTTTTTTTAAGTGTGAATCACATGGTTAGGAGCAGTTGTCTTCAATGTGACCAACCATCCCAAGGCTCTAATTCAACGT 100
                                M  V  R  S  S  C  L  Q  C  D  Q  P  S  Q  G  S  N  S  T  F

TTGGGTGTGGGGGCCCGCTGGCAGCTGTGTGTGCCACTGGGCTGTTGGTGTTGGTGCTTTACTCCCCCTCATCGCAAACGGCTAATTGGTCGGCACAGGG 200
  G  C  G  G  P  L  A  A  V  C  A  T  G  L  L  V  L  V  L  Y  S  P  S  S  Q  T  A  N  W  S  A  Q  G

TATTTCCACAAAGGCGCTGTATCCGGCAGTGCCTGTGCCTTCCACTCTGCTGCCTGGAAGCGCGCCTGCCAAACACCAGCTGCATGTTTGGAGGGCACAT 300
  I  S  T  K  A  L  Y  P  A  V  P  V  P  S  T  L  L  P  G  S  A  P  A  K  H  Q  L  H  V  W  R  A  H

GCGATGTCGGAGGCCACAACAAACAATTCATTCAAACAGTCATTATTTGGGTACAATGCCATCTCCTCCATTTGGCTTCAACTGGCTGGTGTGGCCGCCA 400
A  M  S  E  A  T  T  N  N  S  F  K  Q  S  L  F  G  Y  N  A  I  S  S  I  W  L  Q  L  A  G  V  A  A  T

CTTTCTTTGCATTTGGAGCTTTGATGGCAGCTGTAACGCAACGCAAGGAGATCGCCGTCTTCTCCGCCTCGGGTCAGGCTGCTGAGCCGGAGGGGGCGGA 500
  F  F  A  F  G  A  L  M  A  A  V  T  Q  R  K  E  I  A  V  F  S  A  S  G  Q  A  A  E  P  E  G  A  E

GCCCCTGAAGCGGCCTTTTCCGTCTCCTGCTGCCAAACCTAAGCCGCTCTTCTCCACCCCGGCAAATTCCTTCAGCAACATCTTCCAGGCGCCTCCATCG 600
  P  L  K  R  P  F  F  P  S  P  A  A  K  P  K  P  L  F  S  T  P  A  N  S  F  S  N  I  F  Q  A  P  P  S

CTGCGCACGGACTCCACCTATGGCCGAGGCCCGCGCTCGACCAGCTTCACCGACATCAGCAACTGGCCCTCCAACAACGCACTCCGCAACCCCCAGTCGG 700
L  R  T  D  S  T  Y  G  R  G  P  R  S  T  S  F  T  D  I  S  N  W  P  S  N  N  A  L  R  N  P  Q  S  V

TGATTGACATCGGGGGAGGAGTCGACTTCCTGGGGGACAGAAGCCCTGGAAACCCGTTCACGCGGCTGCGGGGGTCCCCGAGCTCCACCCTCAGCAACCT 800
  I  D  I  G  G  G  V  D  F  L  G  D  R  S  P  G  N  P  F  T  R  L  R  G  S  P*  S  S  T  L  S  N  L

CGGCATGGGCCTAGGCCTGGGGCTGGGCAAGGGCAAGGGCTTCGGCAAGGGCTTCGGCAAAGGCCGGGGGTTCCCCGTGGAGGAGGAGGTGGAGGAGGAG 900
  G  M  G  L  G  L  G  L  G  K  G  K*  G  F  G  K*  G  F  G  K*  G  R  G  F  P  V  E  E  E  V  E  E  E

CAGGAGGTGCTGTCGTGGGCCGACCGCCGGCGGGCGCTGGCGGACCCCGACGCCCCGCCGATGAACGAGGACATCAAGTACCCGCAGCTGCGGCTGGTGC 1000
Q  E  V  L  S  W  A  D  R  R  R  A  L  A  D  P  D  A  P  P  M  N  E  D  I  K  Y  P  Q  L  R  L  V  R

GGGCCGTGCCGGGCGGCCGGGACGAGAAGCTCGGTGTGATGTCGAGGCAGGAGGCGCTGGAGCTGGCGGAGGCGGAAGACATCGACCTCGTCCTCGTCAG 1100
  A  V  P  G  G  R  D  E  K  L  G  V  M  S  R  Q  E  A  L  E  L  A  E  A  E  D  I  D  L  V  L  V  S

CATCGACACCGACCCCCCGGTGGCCAAGCTAGTAATTACTCAAGTTGAAGTACAGTCCGAGAAGAAGAAGAAGGACAGCCACAAGAAGGGGAAGGTGAAG 1200
  I  D  T  D  P  P  V  A  K  L  V  I  T  Q  V  E  V  Q  S  E  K  K  K  K  D  S  H  K  K  G  K  V  K

GAGGTGAAGGAGCTGAAGGTGTCCCATAAGATCGGCCAGCACGACTACGACGTCCGCGTGAAGCAGGCCCGAAAGTTCCTGGAGGGCGGCCACCGCATCA 1300
E  V  K  E  L  K  V  S  H  K  I  G  Q  H  D  Y  D  V  R  V  K  Q  A  R  K  F  L  E  G  G  H  R  I  K

AGGTGTCGATGGAGTTCAAGGGGCGCGAGAACCAGTTCGTGGAGATCGGCCGCGCGGTGATGAAGCGCTTCCAGAACGACCTGGCGGACATGGGCAAGGC 1400
  V  S  M  E  F  K  G  R  E  N  Q  F  V  E  I  G  R  A  V  M  K  R  F  Q  N  D  L  A  D  M  G  K  A

GGACGCCGTGCCCAAGAAGCTCGGCACCCGGCTGATCCTGAACCTGGCCCCGGCCGGGGAGGCGCTGAAGGTGATTGCGGAGCGGAGGGCAGAGCGCGAC 1500
  D  A  V  P  K  K  L  G  T  R  L  I  L  N  L  A  P  A  G  E  A  L  K  V  I  A  E  R  R  A  E  R  D

AGGAAAGCCGCGGCTGAGGAGGAGGGGGAGGGCGACGACCTCGACTTCGTGGACGAGAACGAGGACGAGGATGTGGAGGGGGAGGGCGAGGAGGAAGAGG 1600
R  K  A  A  A  E  E  E  G  E  G  D  D  L  D  F  V  D  E  N  E  D  E  D  V  E  G  E  G  E  E  E  E  A

CCGAGGAGCTGGAGGAGGAGACAGCGGAGGGGACGGAGGTGCCAACCCGCAGCTGATCGCCGATCCGCGGGGGACAGCCACCTCCCCCCCGGCCTCCCTG 1700
  E  E  L  E  E  E  T  A  E  G  T  E  V  P  T  R  S  *

CCGGGGGCCGGCACCATCCGTCGTTGCGGTGCGGCGCTGCCATCAACGGCCGTCCTTGAGCTTAATGCTCCCGCCCTCCGTTGGCCCGCGGCGGTCGCCA 1800
GGTTGCTGGCCTGGCTGCCCGCAGCTCCTCCCCTCCCCGACTGACACAGTGTGGATGACCGTGATGTGCGCCTTTTCGCCTTC(A)n
```

FIG. 3. **Primary sequence of the IF-3_{chl} cDNA.** The sequence shown begins at or very near the 5′ end of the mRNA and includes the spliced leader. The amino acid sequence for the IF-3_{chl} precursor is shown beginning with the first ATG codon. The residue designated *P** indicates a difference between the DNA sequence obtained and the peptide sequence information. The residues indicated as *K** may be post-translationally modified.

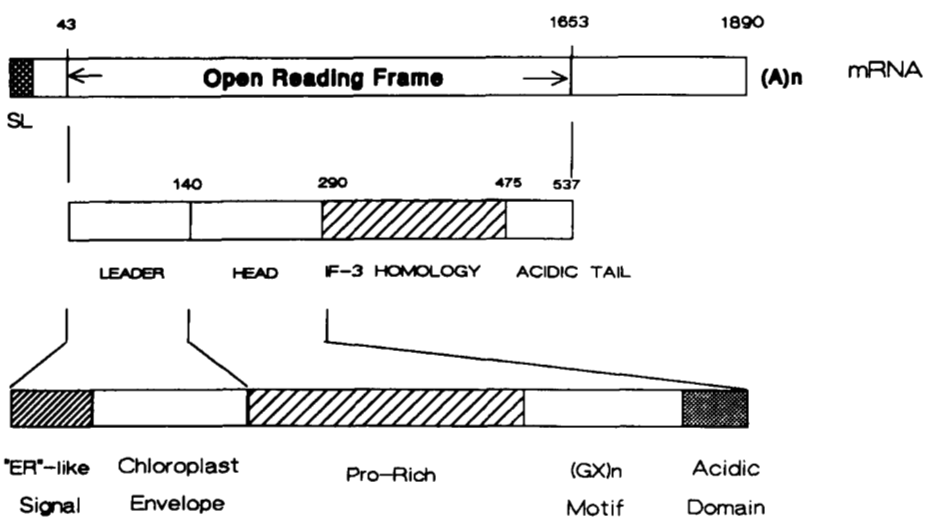

FIG. 4. **Organization of the IF-3chl mRNA and polypeptide.** The general organization of the coding sequence of IF-3_{chl} suggests that it can be divided into a leader specifying import into the chloroplast, a head region, a region with homology to *E. coli* IF-3, and a tail region.

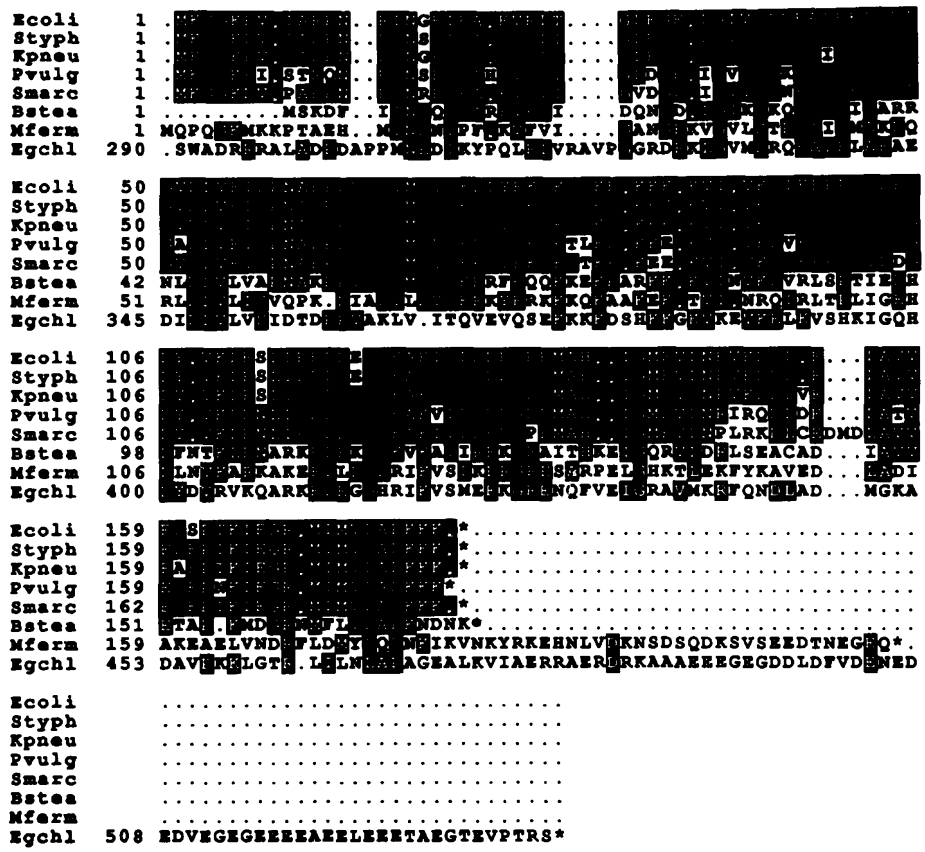

FIG. 5. **Comparison of the amino acid sequences of IF-3 from various organisms.** The alignment of the sequences was carried out using the PILEUP program in the GCG software package and the alignment was displayed using the BOXSHADE program obtained by FTP from netserv@EMBL-Heidelberg.de. The designations used are as follows.: *Ecoli*, *E. coli*; *Styph*, *Salmonella typhimurium*; *Kpneu*, *Klebsiella pneumoniae*); *Pvulg*, *Proteus vulgaris*; *Smarc*, *Serratia marcescens*; *Bstea*, *B. stearothermophilus*; *Mferm*, *Mycoplasma fermentans*; *Egchl*, *E. gracilis* (IF-3chl). Shaded areas indicate residues conserved in most of the IF-3 sequences determined to date (Ref. 39–41 and GenBank).

TABLE III
Identity of the known IF-3 sequences (%)
Designators are defined in legend to Fig. 5.

	Ecoli	*Bstea*	*Mferm*	*Pvulg*	*Smarc*	*Kpneu*	*Styph*	*Egchl*
Ecoli	100	52	39	86	89	97	99	34
Bstea		100	38	51	51	51	52	37
Mferm			100	38	39	39	38	33
Pvulg				100	86	85	86	34
Smarc					100	89	90	31
Kpneu						100	97	34
Styph							100	34
Egchl								100

separate group and is no more closely related to the corresponding factors from Gram-negative bacteria, Gram-positive bacteria, or mycoplasma. The chloroplast factor shows between 30 and 40% sequence identity to all other known IF-3s (Table III).

Complete conservation of 25 of the approximately 180 residues in the homology domain is observed among all of the IF-3 sequences determined to date (Fig. 5). These residues are scattered throughout the protein with the largest stretch of completely conserved residues only 3 amino acids long. There are two clusters that appear to be more highly conserved than the rest of the protein, corresponding to amino acids 33–54 and 117–141 using the *E. coli* numbering system. Numerous other positions show conservative replacements or have been conserved in the majority of IF-3 species. Limited information is currently available on structure-function relationships in prokaryotic IF-3. The crucial residue Tyr-107 that is believed to be important in the binding of *E. coli* IF-3 to 30 S subunits is conserved in the IF-3_{chl} sequence. Another important residue in *E. coli* IF-3 is Lys-110, which is Arg in IF-3_{chl}. Site-directed mutagenesis of *E. coli* IF-3 Lys-110 to Arg results in about a 15-fold lower binding constant to 30 S subunits compared to wild-type IF-3 (42). Many of the basic residues are conserved between *E. coli* IF-3 and the homologous domain in the chloroplast factor. These positively charged residues are thought to be involved in an interaction between the initiation factor and the 16 S rRNA in the small subunit. In contrast to the head region, the IF-3 homology domain is predicted to have extensive α-helical and β-sheet secondary structure based on the Chou-Fasman and Garnier-Osguthorpe-Robson algorithms (35, 36).

The Tail Region of IF-3_{chl}—The last 63 residues of IF-3_{chl} is termed the tail region. The sequence here does not have homology to any protein sequence available in the databases. Analysis using the MOTIFS or BLOCKS programs failed to indicate any relationship between this sequence and known structural patterns. BLAST analysis suggested a relationship to the Glu-rich region of parathymosin, although this relationship may simply reflect the presence of Glu-rich sequences in these proteins. The tail is highly acidic, and this 63-amino acid sequence has 29 acidic residues compared to 7 basic residues. It has a pI value of 3.69. Both the Chou-Fasman and Garnier-Osguthorpe-Robson algorithms predict that this region is largely α-helical. Thus, the IF-3 homology domain, which is basic, is bracketed on both sides by acidic regions that are probably helical.

Structure of the IF-3_{chl} Gene—In *E. gracilis* single-copy DNA sequences comprise about 12% of the genome (43). In order to determine the copy number of the IF-3_{chl} gene, a Southern blot of total DNA digested with *Hin*dIII, *Bam*HI, and *Bgl*II, respectively, was probed using a PCR product amplified between residues 1303 and 1397 (Fig. 6). This region was selected as a probe because PCR amplification of these sequences using genomic DNA as a template suggested that no introns were present in this region (data not shown). Digestion of genomic DNA with *Bam*HI gave four bands of approximately equal intensity on Southern analysis. *Hin*dIII digestion resulted in the presence of three bands, one of which was more intense than the others. Similarly, four major bands were observed with *Bgl*II-digested

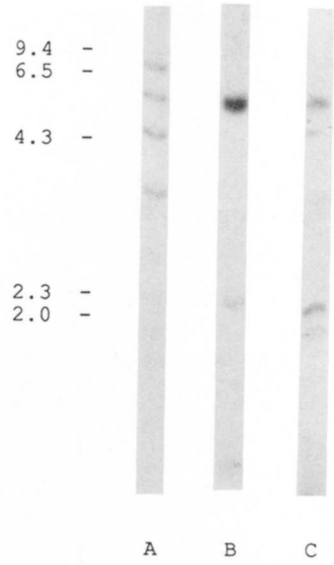

FIG. 6. **Southern analysis of *E. gracilis* genomic DNA for IF-3_{chl} sequences.** Total DNA (20 μg) was digested with *Bam*HI (*lane A*), *Hin*dIII (*lane B*) or *Bgl*II (*lane C*) and the fragments produced were separated by agarose gel electrophoresis as described under "Experimental Procedures." Blots were probed with a PCR product amplified between residues 1303 and 1397. The relative positions of size markers are indicated on the *left*.

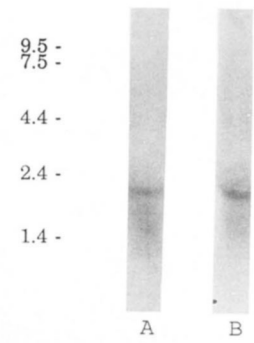

FIG. 7. **Northern blot analysis of chloroplast IF-3_{chl} mRNA.** RNA extracted from 1×10^8 cells was denatured with Me_2SO and glyoxal and subjected to electrophoresis on a 0.9% agarose gel. After transfer to a Zeta-Probe membrane, the blot was hybridized with ^{32}P-labeled probe prepared by PCR amplification between primers P5 and P6. Sizes of RNA molecular weight markers are shown on the *left*. *Lane A*, total RNA from *E. gracilis* cells exposed to light for 72 h; *lane B*, total RNA prepared from cells grown and maintained in the dark.

genomic DNA. Since none of these enzymes cuts within the probe region, these result suggest that the *E. gracilis* genome has at least four copies of IF-3_{chl} gene. It has recently been observed that there are multiple copies of the chloroplast elongation factor Tu gene in higher plants (44).

In order to gain insight into whether one or more of the IF-3_{chl} genes detected could be pseudogenes, genomic DNA sequences between nucleotides 1228 and 1397 were amplified by PCR. The PCR products were separated by gel electrophoresis. Two bands were observed hybridizing to an IF-3_{chl} probe: a weak band of 169 base pairs, which corresponds to the size of the product predicted from the sequence of the cDNA, and a strong band at about 1 kb (data not shown). This observation suggests that at least one of the IF-3_{chl} genes has a different organization from the others. The observation that it lacks an intron in a region that contains an intron in the other copies of the IF-3_{chl} gene suggests that it might be a pseudogene.

IF-3_{chl} activity has been resolved into three forms by chromatography on high performance liquid chromatography resins (3). The α form has a molecular mass of about 34 kDa, while the β and γ forms have molecular masses of about 42 kDa. These three forms are functionally indistinguishable in *in vitro* assays. The existence of multiple gene copies for IF-3_{chl} leaves open the possibility that the chromatographically distinct forms of this factor may arise from separate genes. It should be noted, however, that all the IF-3_{chl} cDNA clones obtained from the libraries gave identical sequences throughout their overlapping regions. It is, thus, certainly reasonable to imagine that the three forms of IF-3_{chl} arise through post-translational modification or from minor changes such as deamination events occurring during purification. Further work will be required to examine the structure and organization of the multiple IF-3_{chl} genes.

Effect of Light on the Expression of the E. gracilis IF-3_{chl} Gene—The development of the chloroplast in *E. gracilis* is induced by exposure of dark-grown cells to light. The activities of many nuclear-encoded proteins that function in the chloroplast also appear to be light-inducible. For example, when dark-grown *E. gracilis* cells are exposed to light, a 12-fold induction of IF-3_{chl} activity is observed. However, the levels of certain mRNAs encoding chloroplast proteins do not appear to change significantly in response to light, suggesting that the regulation of expression of these proteins occurs post-transcriptionally (45). The effect of light on the level of the mRNA for IF-3_{chl} has been examined by Northern analysis of RNA obtained from light-induced or dark-grown cells (Fig. 7). Total RNA was extracted from the same number of cells so that the difference in hybridization would reflect the change in the RNA pool per cell. The size of the cell increases during light-induction about 2–3-fold so that different weights of cells were used in this experiment. Northern analysis (Fig. 7) indicates the presence of a single mRNA of about 2 kb, which hybridizes to the IF-3_{chl} probe. No significant differences in the level of IF-3_{chl} mRNA per cell were observed between cells grown and maintained in the dark compared to cells that were grown in the dark and then illuminated for 72 h to allow chloroplast development. The Northern analysis suggests that light has no effect on the level of the IF-3_{chl} mRNA. Thus, the increased activity of IF-3_{chl} observed after light induction appears to be regulated post-transcriptionally.

REFERENCES

1. Ma, L., and Spremulli, L. L, (1990) *J. Biol. Chem.* **265,** 13560–13565
2. Ma, L., and Spremulli, L. L. (1992) *J. Biol. Chem.* **267,** 18356–18360
3. Wang, C. C., and Spremulli, L. L. (1991) *J. Biol. Chem.* **266,** 17079–17083
4. Kraus, B. L., and Spremulli, L. L. (1986) *J. Biol. Chem.* **261,** 4781–4784
5. Gualerzi, C. O., and Pon, C. L. (1990) *Biochemistry* **29,** 5881–5889
6. Van Knippenberg, P. H. (1990) *The Ribosome: Structure, Function and Evolution* (Hill, W. E., Dahlberg, A., Garrett, R. A., Moore, P. B., Schlessinger, D., and Warne, J. R., eds) pp. 265–289, American Society for Microbiology, Washington, DC
7. Pon, C. L., and Gualerzi, C. O. (1974) *Proc. Natl. Acad. Sci. U. S. A.* **71,** 4950–4954
8. Canonaco, M. A., Gualerzi, C. O., and Pon, C. L. (1989) *Eur. J. Biochem.* **182,** 501–506
9. Hartz, D., Binkley, J., Hollingsworth, T., and Gold, L. (1990) *Genes & Dev.* **4,** 1790–1800
10. Hartz, D., McPheeters, D. S., and Gold, L. (1989) *Genes & Dev.* **3,** 1899–1912
11. Berkhout, B., Van der Laken, C. J., and Van Knippenberg, P. H. (1986) *Biochim. Biophys. Acta* **866,** 144–153
12. Wickstorm, E., Tyson, R. W., Newton, G., Obert, R., and Williams, E. E. (1980) *Arch. Biochem. Biophys.* **200,** 296–300
13. Kraus, B. L., and Spremulli, L. L. (1988) *Plant Physiol.* **88,** 993–995
14. Stern, A., Schiff, J., and Epstein, H. (1964) *Plant Physiol.* **39,** 220–226
15. Montandon, P., and Stutz, E. (1990) *Nucleic Acids Res.* **18,** 75–81
16. Burkhart, W. (1992) in *Techniques in Protein Chemistry IV* (Angeletti, R., ed) pp. 399–406, Academic Press, New York
17. Woodcock, E., and Merrett, M. J. (1980) *Arch. Microbiol.* **124,** 33–38
18. Jacobson, A. (1987) *Methods Enzymol.* **152,** 154–261
19. Roth, M. J., Tanese, N., and Goff, S. P. (1985) *J. Biol. Chem.* **260,** 9326–9335
20. Wu, D. Y., Ugozzoli, L., Pal, B. K., Qian, J., and Wallace, B. (1991) *DNA Cell Biol.* **10,** 233–238
21. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
22. Gottesman, M. M. (1987) *Methods Enzymol.* **151,** 363–364

23. Sanger, F., Nicklen, S., and Coulson, A. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5419
24. Carmichael, G. G., and McMaster, G. K. (1980) *Methods Enzymol.* **65,** 380–391
25. Murray, M. G., and Thompson, W. F. (1980) *Nucleic Acids Res.* **8**. 4321–4325
26. Keller, M., Tessier, L. H., Chan, R. L., Weil, J. H., and Imbault, P. (1992) *Nucleic Acids Res.* **20,** 1711–1715
27. Tessier, L. H., Keller, M., Chan, R. L., Fournier, R., Weil, J. H., and Imbault, P. (1991) *EMBO J.* **10,** 2621–2625
28. Hunt, A. G., Chu, N. M., Odell, J. T., Nagy, F., and Chua, N.-H. (1987) *Plant Mol. Biol.* **8,** 23–35
29. Gibbs, S. P. (1981) *Ann. N. Y. Acad. Sci.* **361,** 193–207
30. Lefort-Tran, M. (1981) *Ber. Dtsch. Bot. Ges.* **94**. 463–476
31. Shashidhara, L. S., Lim, S. H., Shackleton, J. B., Robinson, C., and Smith, A. G. (1992) *J. Biol. Chem.* **267,** 12885–12891
32. Von Heijne, G. (1983) *Eur. J. Biochem.* **133,** 17–21
33. Gumpel, N. J., and Smith, A. G. (1992) *Eur. J. Biochem.* **210,** 721–727
34. Von Heijne, G., and Nishikawa, K. (1991) *FEBS Lett.* **278,** 1–3
35. Nishikawa, K. (1983) *Biochim. Biophys. Acta* **748,** 285–299
36. Garnier, J., Osguthorpe, D. J., and Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120
37. Henikoff, S., and Henikoff, J. G. (1991) *Nucleic Acids Res.* **19,** 6565–6572
38. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410
39. Sacerdot, C., Fayat, G., Dessen, P., Springer, M., Plumbridge, J. A., Grunberg-Manago, M., and Blanquet, S. (1982) *EMBO J.* **1,** 311–315
40. Kimura, M., Ernst, H., and Appelt, K. (1983) *FEBS Lett.* **160,** 78–81
41. Hu, W. S., Wang, R. Y., Shih, J. W., and Lo, S. C. (1993) *Gene (Amst.)* **127,** 79–85
42. De Bellis, D., Liveris, D., Goss, D., Ringquist, S., and Schwartz, I. (1992) *Biochemistry* **31,** 11984–11990
43. Rawson, J. R. Y., Eckenrode, V. K., Boerma, C. L., and Curtis, S. (1979) *Biochim. Biophys. Acta* **563,** 1–16
44. Murayama, Y., Matsubayashi, T., Sugita, M., and Sugiura, M. (1993) *Plant Mol. Biol.* **22,** 767–774
45. McCarthy, S. H., and Schwartzbach, S. D. (1984) *Plant Sci. Lett.* **35,** 61–66