Structural and Mechanistic Studies on Chloroplast Translational Initiation Factor 3 from *Euglena gracilis*†

Nan-Jun Yu‡ and Linda L. Spremulli*,‡

Department of Chemistry CB# 3290 and Lineberger Comprehensive Cancer Research Center, University of North Carolina, Chapel Hill, North Carolina 27599-3290

Received May 20, 1997; Revised Manuscript Received August 29, 1997⊗

ABSTRACT: Chloroplast translational initiation factor 3 ($IF3_{chl}$) from *Euglena gracilis* contains a central region (homology domain) that is homologous to prokaryotic IF3. The homology domain is preceded by a long NH_2-terminal extension (head), and followed by a 64 amino acid COOH-terminal extension (tail). Sequences in these extensions reduce the activity of the homology domain. To gain insight into these effects, a possible three-dimensional structure for the homology region of $IF3_{chl}$ has been modeled using the X-ray coordinates from the N- and C-domains of *Bacillus stearothermophilus* IF3. In *B. stearothermophilus* IF3, these two compact domains are thought to fold independently and are separated by a helical lysine-rich linker. The modeled structure suggests that $IF3_{chl}$ has a similar overall fold although some subtle differences are predicted to occur. Both the head and tail regions of $IF3_{chl}$ are oriented toward the linker region in the homology domain where they may potentially interfere with its function. Circular dichroism spectropolarimetry (CD) indicates that the lysine-rich linker region in $IF3_{chl}$ is not in a helical conformation and is probably a random coil. CD analysis indicates that a portion of the tail region of $IF3_{chl}$ is helical and that the tail has a direct interaction with the linker region in the homology domain. Site-directed mutagenesis of the linker indicates that two conserved lysine residues are important for the function of $IF3_{chl}$ and play a role in the binding of $IF3_{chl}$ to the 30S ribosomal subunit. Mutation of these residues affects the interaction of the homology domain with the tail.

In prokaryotes, the initiation of protein biosynthesis begins on the small subunit of the ribosome. During this process, IF3[1] binds to the 30S subunit preventing its association with the 50S subunit. IF3, thus, ensures a supply of 30S subunits for translational initiation *(1−5)*. IF3 binds the 30S subunit by interacting both with the 16S rRNA and ribosomal proteins *(6−9)*. IF3 also stimulates the binding of IF1, IF2 and fMet-tRNA to 30S subunits *(10, 11)*. Finally, IF3 proofreads the selection of fMet-tRNA and the AUG codon during initiation and, thereby, improves the fidelity of protein synthesis *(12−17)*.

Prokaryotic IF3 folds into two compact protease-resistant domains called the N- and C-domains *(18, 19)*. A model has been proposed suggesting that the N-domain of IF3 is responsible for the proofreading activity, while the C-domain binds the 30S subunit *(18, 19)*. Genetic studies indicate that Tyr75 in the N-domain plays an important role in the proofreading activity *(15)*. Chemical modification of Lys110 or mutation of Lys107 and Lys110 in the C-domain of *Escherichia coli* IF3 decreases the binding of this factor to 30S subunits *(20−22)*. These data support the idea that the N- and C-domains have different functions in IF3. Additional information also indicates that the lysine-rich linker connecting the N- and C-domains plays a role in the binding of IF3 to the 30S subunit *(20, 23)*.

The three-dimensional structures of the N- and C-domains have been determined using NMR with *E. coli* IF3 *(23, 24)* and X-ray crystallography with *Bacillus stearothermophilus* IF3 *(21)*. Both factors show similar topology for the N- and C-domains. Both domains belong to the α/β superfamily. The linker that can be observed in the X-ray structure of *B. stearothermophilus* IF3 exists as a helix *(21)*. Neutron scattering data indicate that the distance between the centers of mass of the domains is ∼45 Å *(18)*. In contrast, in the structure of *E. coli* IF3, derived from the NMR spectra of either the separated domains *(23, 24)* or the whole molecule *(25)*, the lysine-rich linker is highly flexible. The NMR data also indicate that the motions of the two domains are uncorrelated, suggesting that the linker can display almost totally unrestricted motion. The distance between the domains in *E. coli* IF3 fluctuates from 28.2 to 64.5 Å with an average of 46.3 Å *(25)*.

Only one organellar IF3 has been identified and purified to date. This factor ($IF3_{chl}$ from *Euglena gracilis)* is about twice the size of prokaryotic IF3 *(26, 27)*. The mature form of this factor contains a central region that is homologous to prokaryotic IF3 (homology domain). The portion of $IF3_{chl}$ corresponding to *B. stearothermophilus* IF3 ($IF3_{chl}$srH) has been cloned and expressed in *E. coli*. $IF3_{chl}$srH has all the functions of prokaryotic IF3 *(28)*. The homology domain is preceded by a long NH_2-terminal extension (head) and followed by a 64 amino acid COOH-terminal extension (tail). The head and tail negatively affect the function of the homology domain of $IF3_{chl}$. Nine residues at the junction between the head and the homology domain contribute all

† Supported in part by funds from the National Institutes of Health (GM24963).

* Author to whom correspondence should be addressed. Department of Chemistry CB# 3290.

‡ Department of Chemistry.

⊗ Abstract published in *Advance ACS Abstracts,* November 15, 1997.

[1] Abbreviations: CD, circular dichroism spectropolarimetry; $IF3_{chl}$, chloroplast Initiation Factor 3; $IF3_{chl}$srH, region of $IF3_{chl}$ homologous to *Bacillus stearothermophilus* IF3; $IF3_{chl}$sHT, region of $IF3_{chl}$ including the homology domain and the tail.

S0006-2960(97)01185-9 CCC: $14.00

the negative effect from the head. The tail contains mostly acidic amino acid residues and contributes about half of the negative regulatory effect of $IF3_{chl}$ (submitted for publication). Of the prokaryotic IF3s, only the factor from *Myxococcus xanthus* has a COOH-terminal extension *(29)*. *In vivo* studies indicate that this extension is important for certain vegetative and developmental functions but not for viability *(30)*.

In this paper, homology modeling and CD spectroscopy have been used to analyze the three-dimensional structure of the homology domain of $IF3_{chl}$ and to gain insight into the secondary structure of the tail. Mutations in the linker region have been analyzed for their effects on the function of $IF3_{chl}$ and on the interaction between sequences in the tail and the homology domain.

MATERIALS AND METHODS

Materials. Chemicals and the ELISA kit were purchased from Sigma Chemical Co. or Fisher Scientific. Syringe filters containing cellulose acetate membranes (3 mm, 0.2 μ pore size) were purchased from Corning Laboratory Sciences. Oligonucleotide primers were made in the Pathology Department at the University of North Carolina. Restriction endonucleases, the Klenow fragment of DNA polymerase I, T4 ligase, Taq DNA polymerase, and T7 RNA polymerase were obtained from Promega. Mini- and Midi-plasmid preparation kits, pQE vectors, *E. coli* M15[pREP4], and Ni^{2+} ion charged nitrilotriacetic acid affinity resin (Ni-NTA) were from Qiagen. The GeneClean kits were from Bio 101. Plasmids encoding a derivative of $IF3_{chl}$ equivalent to *B. stearothermophilus* IF3 ($IF3_{chl}$srH) and a derivative carrying this region followed by the acidic tail ($IF3_{chl}$sHT) were prepared as described elsewhere (submitted for publication). [^{35}S]fMet-tRNA was prepared as described *(31, 32)*. The plasmid pRbcN carrying the 5′ untranslated leader region and the translational start site of the *E. gracilis* chloroplast *rbcL* gene fused in-frame to an internal coding region of the neomycin phosphotransferase gene was prepared as described *(33)*. *In vitro* transcription of this plasmid was carried out as described *(33)* allowing the preparation of the mRNA designated *mRbcN*. *E. coli* ribosomes and initiation factors were prepared as described *(34, 35)*. *E. gracilis* chloroplast 30S subunits, $IF2_{chl}$ and $IF3_{chl}$, were prepared as described previously *(26, 36, 37)*. Antisera was raised against the homology domain of $IF3_{chl}$ as described *(28)*.

Molecular Modeling. Included are sequence alignment, secondary structure prediction, and model building and loop insertion.

Sequence Alignment. The DNA sequences of a number of prokaryotic IF3 genes are available. Only two three-dimensional structures are known. Sequence alignment has been done with 19 IF3s and the homology domain of $IF3_{chl}$ using the Clustalw program *(38)*.

Secondary Structure Prediction. Methods used to predict the secondary structure of $IF3_{chl}$ include the Gibrat method, the Levin homology method, the double prediction method, and the self-optimized prediction method *(39)*. In addition, the PHD program was used to predict the secondary structure of $IF3_{chl}$ and its derivatives based on a multiple sequence alignment *(40)*.

Model Building and Loop Insertion. The starting structures for the N- and C-domains of the homology region of $IF3_{chl}$ were based on the coordinates of *B. stearothermophilus* IF3 obtained from X-ray crystallography and modeled using SYBYL. Sequence replacement was done piece by piece, according to the secondary structural features seen in the X-ray structure and based on the aligned sequence. After each piece of the sequence was replaced, the replaced region was refined by energy minimization while keeping the remainder of the sequence frozen. The Kollman All-Atom force field was chosen for the calculation. Each energy minimization was run using the conjugate-gradient method for 500 steps. After the entire sequence was replaced, the final structure was again energy minimized.

The four-residue insertion in the N-domain of $IF3_{chl}$ was modeled using the Loop Insertion package in SYBYL. Coordinates were assigned to the loop by performing a loop search over the databank of known crystal structures with resolutions better than 2.0 Å. Twenty-five potential loops were generated. The quality of each loop was evaluated and loops with the highest homology values and the lowest rms values were selected. The backbone coordinates of the selected loops were inserted into the coordinates of the N-domain. The side chains were then inserted. The inserted loop was energy minimized first, and the whole molecule was then energy minimized once again.

Site-Directed Mutagenesis. Site-directed mutagenesis was performed using PCR. The mutagenic primers were 5′-TACGAGTCCGAGAAGGCTGCAAAGGACAGCCACAAGAAGGG (2K mutant) 5′-TACGAGTCCGAGGCGGCTGCAGCGGACGACCACAAGAAG (4K mutant). The mutated regions are underlined. Appropriate primers and linear plasmids (cut at the *Pvu*II site) carrying the sequence for $IF3_{chl}$srH or $IF3_{chl}$sHT were used to generate single-stranded mutated DNAs. After purification from a 2% agarose gel, single-stranded DNAs were converted to double strands and amplified by PCR using the original mutated primer and the pQE universal reverse primer. The fragments generated were purified from 2% agarose gels using GeneClean. Full-length copies of the mutated genes were generated by PCR using the purified double-stranded fragments described above and the pQE universal forward primer as primers and the original linearized plasmid DNA (encoding $IF3_{chl}$srH or $IF3_{chl}$sHT) as a template. These DNAs were purified from 1% agarose gels and digested with *Bam*HI and *Hin*dIII. The digested DNAs were then ligated into the pQE16 expression vector providing a His-tag at the COOH-terminus of the expressed proteins. Sequences of the mutated DNAs were confirmed at the automated DNA sequencing facility at the University of North Carolina at Chapel Hill.

Induction and Purification of Various Derivatives of $IF3_{chl}$. Cells were grown, and $IF3_{chl}$ derivatives were induced as described *(28)*. Induction times were 2−3 h for both the wild-type and the mutants. $IF3_{chl}$ mutants were purified using the two-step purification procedure developed for $IF3_{chl}$H *(28)*. The yields of the mutated proteins were the same as their wild-type counterparts.

Assays. The abilities of $IF3_{chl}$ and its derivatives to promote initiation complex formation with *E. coli* 70S ribosomes using poly(A,U,G) as a mRNA were determined as described *(28)*. The activities of $IF3_{chl}$ and its derivatives in promoting initiation complex formation with chloroplast 30S subunits and *mRbcN* were determined as described *(28)*.

Binding of $IF3_{chl}$ to 30S Subunits. Direct measurements of the abilities of various $IF3_{chl}$ mutants to bind to chloroplast

30S subunits were carried out using sucrose density-gradient centrifugation as described (submitted for publication). Briefly, the indicated concentrations of $IF3_{chl}$ and chloroplast 30S subunits were incubated in a total volume of 250 μL in 50 mM Tris-HCl, pH 7.8, 10 mM dithiothreitol (DTT), 50 mM NH_4Cl, and 10 mM $MgCl_2$, at room temperature for 5 min. The mixture was applied to a 5 mL 10 to 30% linear sucrose gradient prepared in the same buffer except that the concentration of Tris-HCl, pH 7.8, was reduced to 10 mM. Samples were subjected to centrifugation at 48 000 rpm for 2 h and fractionated. Aliquots (50 μL) of appropriate fractions were analyzed for the amount of $IF3_{chl}$ present using an ELISA assay *(41)*. A standard curve for each derivative tested was determined in each experiment allowing the amount of $IF3_{chl}$ present to be quantitated using an ELISA.

CD Spectroscopy. After purification, protein samples were concentrated in a Centricon-10 to a volume of <1.0 mL. The samples were desalted, and the buffer was exchanged using a 10 mL G50 gel filtration column equilibrated with 30 mM sodium phosphate (pH 7.0) and 25 mM ammonium sulfate. The fractions containing protein were concentrated using a Centricon-10. Prior to obtaining spectra, the samples were filtered through 3 mm syringe filters containing 0.2 μ cellulose acetate membranes. To determine protein concentrations, $IF3_{chl}$srH or its derivatives were denatured in 6 M guanidinium chloride by incubation at 50 °C for at least 5 h. The samples were then kept at room temperature overnight. Concentrations were calculated from the absorbance at 280 nm using extinction coefficients of 5600 M^{-1} cm^{-1} and 5720 M^{-1} cm^{-1} for $IF3_{chl}$srH and $IF3_{chl}$sHT *(42, 43)*. Spectra were acquired at 25 °C in 1 nm increments from 178 to 260 nm with a single collecting time of 10 s with an AVIV 62DS spectrometer. The wavelength dependence of $[\theta]$ was monitored. Hellma quartz cells with a path length of 0.1 mm were used. The mean residue ellipticities were calculated using the equation $[\theta] = \theta_{obs}/10lcn$, where θ_{obs} is the ellipticity in millidegrees, l is the length of the cell in centimeters, c is the concentration in moles per liter, and n is the total number of residues. Secondary structure contents were estimated with the program Varselec *(44)* using $\Delta\epsilon = [\theta]/3298$.

Thermal denaturation curves were acquired at 222 nm, a wavelength characteristic of α-helix. All the $IF3_{chl}$ constructs were tested in 30 mM sodium phosphate (pH 7.0) in the presence of 25 mM $(NH_4)_2SO_4$. The concentrations used were 0.3−0.6 mg/mL, and data were collected in Hellma quartz cells of 1 mm path length. The temperature range was from 10 to 70 °C with 1 °C steps at a scan rate of 10 °C/h.

RESULTS AND DISCUSSION

Molecular Modeling. Both *E. coli* and *B. stearothermophilus* IF3 fold into two domains (N- and C-domains) separated by a lysine-rich linker. The homology domain of $IF3_{chl}$ is 30−35% identical to various prokaryotic IF3s *(27)*. A derivative of $IF3_{chl}$ corresponding to *B. stearothermophilus* IF3 ($IF3_{chl}$srH) has been expressed in *E. coli*. This derivative is the same length as the *B. stearothermophilus* factor except for a four-residue insertion and residues from the vector providing the His-tag. Sequence alignment was generated using the Clustalw program which places the 4-residue insertion in a loop. The secondary structure predictions suggest that the structures of the N-domain and the C-domains of $IF3_{chl}$ are similar to the corresponding domains in the prokaryotic factors. However, the linker region is predicted to have less helix than expected from examination of the X-ray structure of *B. stearothermophilus* IF3. In addition, two regions in the C-domain that are β-strands in the bacterial factors are predicted to be helical in $IF3_{chl}$.

Homology molecular modeling was performed to provide a structure of the homology domain of $IF3_{chl}$. The coordinates of the N- and C-domains of *B. stearothermophilus* IF3 were selected as the starting coordinates. The structure of the modeled C-domain of $IF3_{chl}$ is similar to that of the *B. stearothermophilus* factor. This domain consists of two helices and a four-stranded β-sheet (Figure 1B). The third strand of the β-sheet in $IF3_{chl}$srH (B3) is, however, comprised of two segments (B3a and B3b) of antiparallel β-sheet.

The N-domain of *B. stearothermophilus* IF3 consists of two helices, including the linker region, with a core comprised of four β-strands. The insertion present in $IF3_{chl}$ occurs in a loop between β-sheet strands B1 and B2 based on the sequence alignment. The distance between the loop and the regions of $IF3_{chl}$ thought to be functionally important, including the linker region, is probably too long to have an effect on the function of the $IF3_{chl}$. Molecular modeling indicates that the N-domain of $IF3_{chl}$ has a structure similar to that of *B. stearothermophilus* although the N-domain of $IF3_{chl}$srH has an additional strand in an antiparallel β-sheet near the NH_2-terminus (open arrow in Figure 1A). The NMR structure of the N-domain of *E. coli* IF3 indicates that this factor also has five strands of β-sheet comparable to the model of $IF3_{chl}$srH produced here *(24)*.

The orientation of the N- and C-domains of IF3 are not known. In intact *B. stearothermophilus* IF3, the centers of mass of the N- and C-domains are separated by about 45 Å *(18)*. Detailed studies by NMR indicate that the linker region in *E. coli* IF3 is highly flexible and that the distance between the N- and C-domains ranges from 28 to 64 Å. However, there is no interaction between these two domains *(25)*. One possible structure for $IF3_{chl}$srH is shown in Figure 1C. Regardless of domain orientation, $IF3_{chl}$srH and prokaryotic IF3 appear to have a dumbbell shape with a basic linker connecting the domains. It is clear from this representation that both the head and tail regions of $IF3_{chl}$ are oriented toward the linker region (Figure 1C). It is, therefore, likely, that the head and tail regions of $IF3_{chl}$ play their negative role in initiation by interacting, directly or indirectly, with the linker region.

Circular Dichroism Spectropolarimetry (CD). The molecular modeling described above indicates that the structure of the homology domain of $IF3_{chl}$ is similar to that of the prokaryotic factors. Homology modeling is based on the assumption that sequence similarity gives rise to structural similarity. No exceptions to this assumption have been observed. Nevertheless, a direct experimental measure of the secondary structural elements present in $IF3_{chl}$srH was desirable to provide confirmation of this assumption and to provide insight into the conformation of the linker. Finally, no information is available on the structure of the tail region of $IF3_{chl}$. Structural measurements can provide insight into the conformation of residues in this region and may suggest how sequences in the tail negatively affect the function of the homology domain.

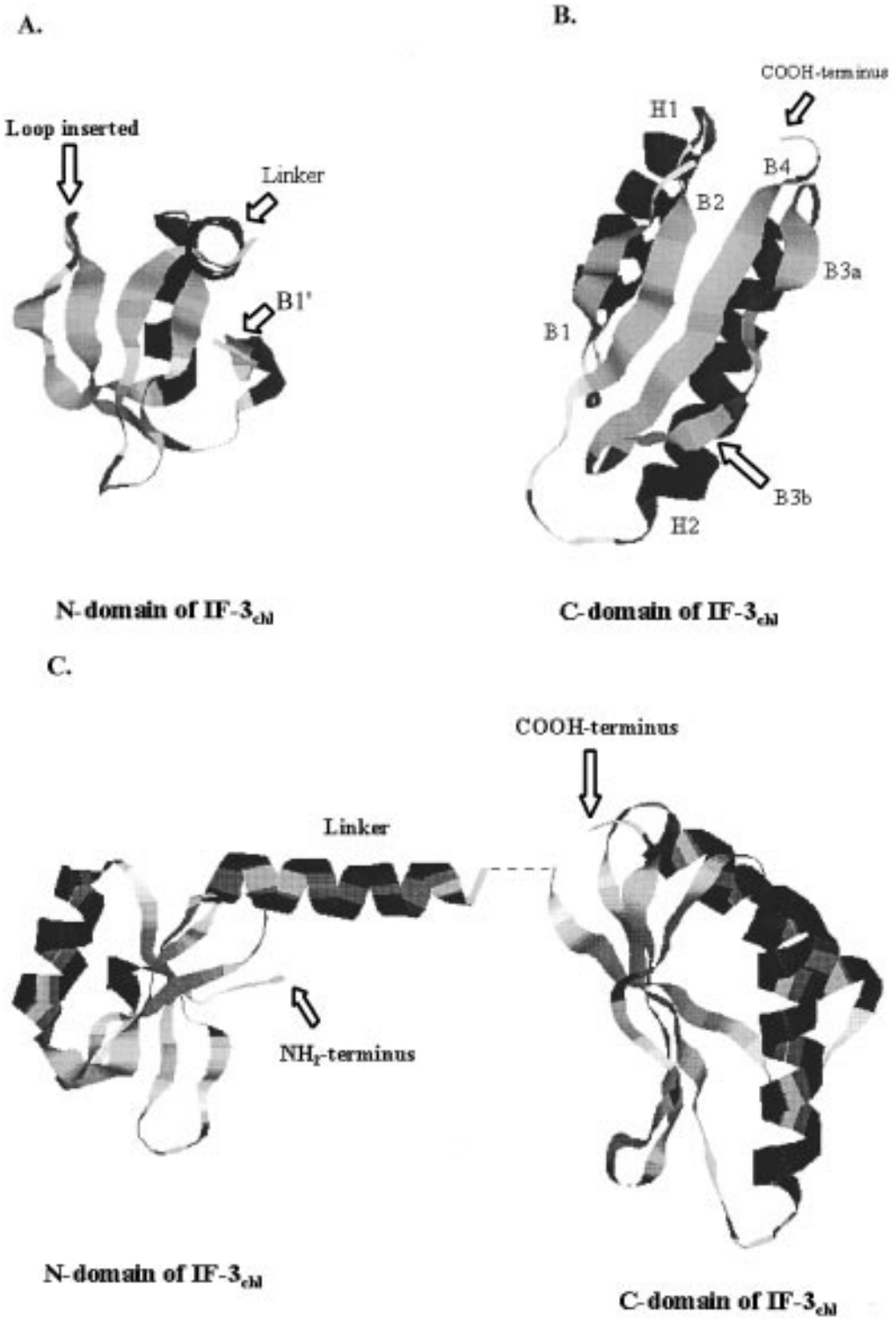

FIGURE 1: Model for $IF3_{chl}$ showing both the N- and C-domains. (A) The model of the N-domain of $IF3_{chl}$ is displayed with RASMOL *(50)*. The arrow indicates the inserted loop. The additional antiparallel β-strand (B1′) predicted at the NH_2-terminus is labeled and shown with an arrow. (B) The model of the C-domain of $IF3_{chl}$ is displayed with RASMOL *(50)*. Helical segments are designated H1 and H2 while β strands are designated B1−B4. The arrow labeled B3b indicates the additional antiparallel β-strand (B3b) predicted to lie between B3a and B4 strands in the C-domain of $IF3_{chl}$. (C) The final models of the N- and C-domains of $IF3_{chl}$ placed adjacent to each other. The relative orientation of these two domains is uncertain. The dashed line indicates four amino acids that are not included in the structures of either *E. coli* or *B. stearothermophilus* IF3.

CD was used to obtain direct information on the structure of $IF3_{chl}$srH (Figure 2A). Secondary structural content was estimated using the Varselec program *(44)*. The percentage of each residue in a particular secondary structure (calculated from mean residue ellipticity) and the number of residues (N) in the structure were calculated (Table 1). It should be noted that CD is more accurate in predicting the content of helical structure than the content of β-sheet structure. $IF3_{chl}$srH, the homologue of *B. stearothermophilus* IF3, has 19% of its residues in helical structure, 28% in antiparallel β-sheet structure, 3% in parallel β-sheet structure, 26% in turns, and 25% in other structures (Table 1). In comparison, *B. stearothermophilus* IF3 has 28% of its residues in helical structure and 19% in antiparallel β-sheet (Table 1) *(18)*. This

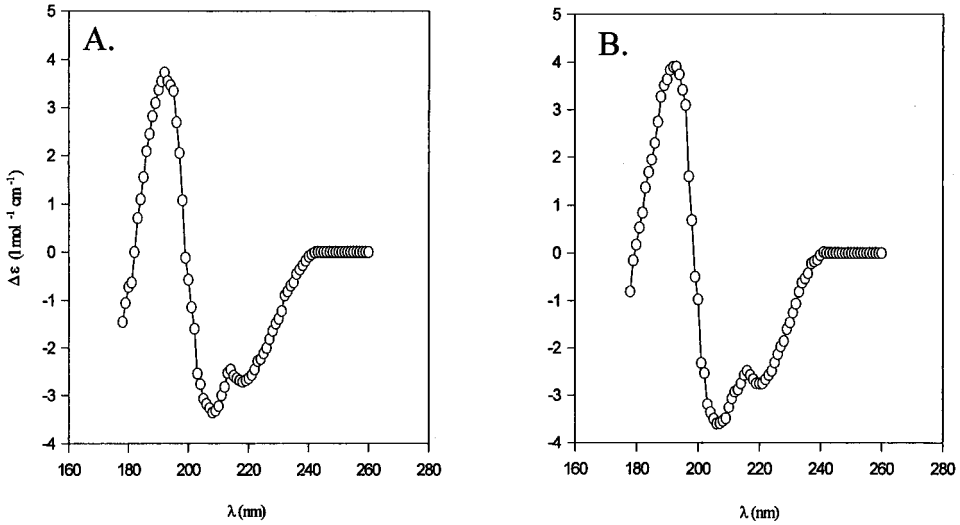

FIGURE 2: CD spectra of $IF3_{chl}$srH and $IF3_{chl}$sHT. The protein concentrations were (A) 1.5 mg/mL for $IF3_{chl}$srH and (B) 0.88 mg/mL for $IF3_{chl}$sHT.

Table 1: Secondary Structural Content Derived from CD Data

content[a]	H	A	P	T	O	TOT[b]
Secondary Structural Content of $IF3_{chl}$srH						
%	19	28	3	26	25	100
N	36	52	6	48	47	189
Secondary Structural Content of $IF3_{chl}$sHT						
%	21	26	1	25	26	99
N	52	65	3	62	65	247
Secondary Structural Content of *B. stearothermophilus* IF3 *(18)*						
%	28	19	7	17	28	99
N	48	32	12	29	48	169

[a] %, content of the secondary structure, N, residues in the secondary structure; H, helix; A, antiparallel beta sheet; P, parallel beta sheet; T, turn; O, others; TOT, total. [b] Includes 14 residues from the vector including the His_6-tag.

CD analysis indicates that $IF3_{chl}$srH has about 12 fewer amino acid residues in a helical structure than does *B. stearothermophilus* IF3. This observation suggests that the linker region in $IF3_{chl}$ is in a highly flexible "random coil" rather than the rigid helical structure seen in *B. stearothermophilus* IF3. It should be noted that the CD data reflects the solution structure of $IF3_{chl}$sH and it is possible that the linker assumes a helical structure when this factor is bound to the ribosome.

The tail region of $IF3_{chl}$ is highly acidic (Figure 3). Since the tail is oriented toward the linker region (Figure 1C), it is likely to have its negative effect on the activity of the homology domain by interacting with basic residues in the linker. Secondary structure predictions on the tail region of $IF3_{chl}$ show two regions with high probabilities of being helical (Figure 3). Insight into the possible structure of residues in the tail region was obtained from the CD spectra of $IF3_{chl}$sHT, the derivative carrying the homology domain and the entire tail (Figure 2B). $IF3_{chl}$sHT is calculated to have 21% of its residues in a helical structure, 26% of the residues in an antiparallel β-sheet structure, 1% of the residues in parallel β-sheet structures, 25% of the residues in turns, and 26% of the residues in other structures (Table 1). Compared to $IF3_{chl}$srH, $IF3_{chl}$sHT has 16 additional residues in a helical conformation. The most likely interpretation of this result is that a portion of the tail region folds into one or more helical segments. Secondary structure predictions based on sequence alignment are consistent with this explanation since they predict two helices in the tail (Figure 3).

Mutation of the Linker Region. Neither the N-domain nor the C-domain of $IF3_{chl}$ alone has any activity in promoting initiation complex formation (data not shown). Neither domain binds to chloroplast 30S ribosomal subunits at a detectable level. These observations imply that the function of $IF3_{chl}$ requires both domains and suggest that the linker region is important for the ability of IF3 to bind to 30S subunits. Since both the head and tail extensions on the homology domain of $IF3_{chl}$ are directed toward the central linker region, these extensions may interact directly with residues in the linker region to affect the activity of the homology domain of $IF3_{chl}$. In particular, the highly acidic tail region may interact with the lysine-rich linker in the homology domain by charge−charge interactions.

Little is known about the roles of specific residues in the linker region on the function of bacterial IF3. Several lysine residues in the linker were targeted for mutagenesis to test the potential importance of the linker region on the function of the homology domain and on the possible interaction of the linker with residues in the tail. Sequence alignment indicated that two lysine residues in the linker (shown in bold in Figure 4) are quite highly conserved. The first of these is present in over 80% of the IF3s that have been sequenced. The other lysine is present in over 60% of the known sequences, and this residue is either lysine or arginine in nearly 90% of them. These two residues will be referred as the conserved lysine residues. Two additional lysine residues are not as highly conserved but are found in $IF3_{chl}$ and a few other species (underlined in Figure 4). They are referred to as the less-conserved lysine residues. Site-directed mutagenesis was used to mutate both the conserved and the less-conserved lysine residues in $IF3_{chl}$srH and $IF3_{chl}$sHT. The lysine residues were mutated to alanine based on the principle of charged to alanine scanning mutagenesis *(45)*. Mutation of the less-conserved lysine residues found in $IF3_{chl}$ lead to the construct designated $IF3_{chl}$srH2K and referred to as the 2K mutant. Mutation of the two additional conserved residues lead to the mutated protein referred to as the 4K mutant.

A.

```
IF3chl     ALKVIAERRAERDRKAAAEEEGEGDDLDFVDENEDEDVEGEGEEEEAEELEEETAEGTEVPTRSRSHHHHHH
Gibrat     HHHHHHHHHHHHHHHHHHHHHHHHCCCHHHHHHHCCCHCHHCCCHHHHHHHHHHHHHHCCECEEEEHHEHEEECC
Levin      HHHHHHHHHHHHHHHHHHHHHHTCTHCHHHCCHCCTTHHHHCCHHHHHHHHHHHHHHCCCCCCCCCTCSCCTC
DPM        HHHHHHHHHHHHHHHHHHHHHHTCTCCCCHHHHTHTHTHCCCCHHHHHHHHHHHHHHCCHCCCCCCCCCCCC
SOPMA      HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHTTHCCCCCHHHHHH
Consensus  HHHHHHHHHHHHHHHHHHHHHHCTCCHHHHHHCCTHHHHCCCHHHHHHHHHHHHHCC-CCCCCC-CCCCCC
```

B.

```
AA      |ALKVIAERRAERDRKAAAEEEGEGDDLDFVDENEDEDVEGEGEEEEAEELEEETAEGTEVPTRS|
PHD     |HHHHHHHHHHHHHH HHHHH                          HHHH  HHHH    HHH     |
Rel     |699997555665322666347776578734123358677506551331565355133324899|

HEL: H=helix, E=extended (sheet), blank=other
PHD: Profile network prediction HeiDelberg
Rel: Reliability index (0-9)
```

FIGURE 3: Secondary structure predictions on the tail region of $IF3_{chl}$. (A) Prediction of the secondary structure of the tail was done using several methods as indicated in Materials and Methods. The amino acid sequence of the tail is shown on the top line. The consensus structure is also shown. H, helical structure; E, β-strand structure; C, coil structure; S, bend. (B) Secondary structure prediction using the PHD program. H: helical structure; E: β strand structure; C: coil structure. The prediction reliability is also indicated. The higher the number, the more reliable the prediction.

Coli	DYGKFLYEKS**K**SS**K**EQK
Bstea	DYGKFRFEQQ**K**KE**K**EAR
$IF3_{chl}$	NYSKLKYESE**K**<u>KK</u>**K**DSH
2K Mutant	NYSKLKYESE**K**AA**K**DSH
4K Mutant	NYSKLKYESEAAAADSH

FIGURE 4: Mutations in the linker region. The sequences of the linker region from $IF3_{chl}$, *E. coli* IF3 (Coli), and *B. stearothermophilus* IF3 (Bstea) were aligned together. The conserved lysine residues are indicated in bold while the less-conserved lysine residues are underlined. Sequence changes in the 2K and 4K mutants are also shown.

The mutated proteins were expressed in *E. coli*, purified, and assayed for activity in initiation complex formation. Figure 5A shows that, in $IF3_{chl}$srH, mutation of the less-conserved lysine residues ($IF3_{chl}$srH2K) does not change the activity of the factor significantly in promoting the binding of fMet-tRNA to *E. coli* 70S ribosomes using poly(A,U,G) as mRNA. In contrast, mutation of all four lysine residues in the homology domain ($IF3_{chl}$srH4K) results in a loss of approximately one-half of the activity of the homology domain. When these mutants are tested with chloroplast 30S subunits using a mRNA which carries a chloroplast translational initiation site *(mRbcN)*, the activity of the 2K mutant ($IF3_{chl}$srH2K) is similar to that of $IF3_{chl}$srH (Figure 5B). Mutation of the four lysine residues in the linker ($IF3_{chl}$srH4K) has a drastic negative effect on the activity of the homology domain resulting in a loss of almost 90% of the activity (Figure 5B). The most likely interpretation of this result is that the two conserved lysine residues are important for the function of $IF3_{chl}$, while the less-conserved lysine residues are not particularly important for its function. Alternatively, it can be argued that mutation of these lysine residues into alanine changes the conformation of the linker from a random coil into a helical structure altering the orientation of the two domains and, thereby, inhibiting the activity of the homology domain.

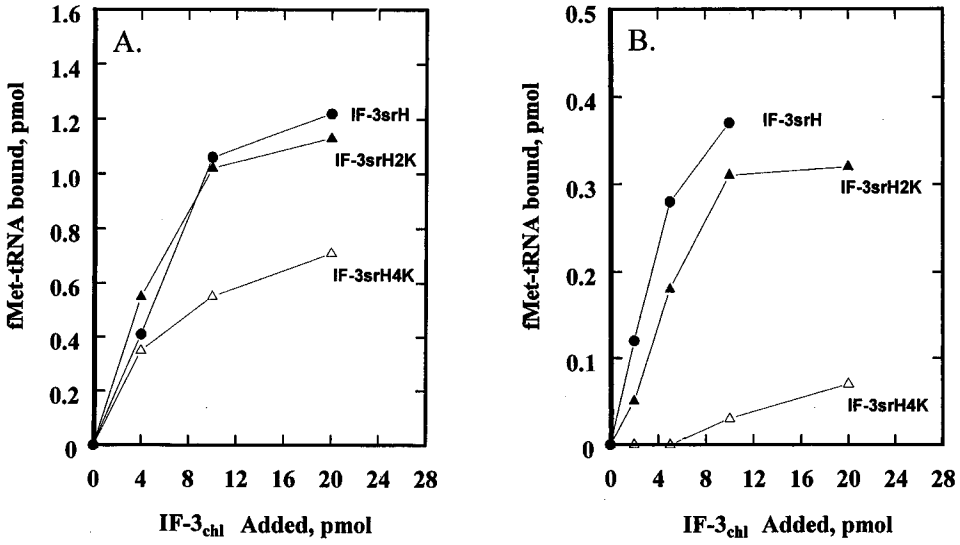

FIGURE 5: Effect of mutation of residues in $IF3_{chl}$srH on initiation complex formation. (A) The activities of $IF3_{chl}$srH2K (▲) and $IF3_{chl}$srH4K (△) were compared to that of $IF3_{chl}$srH (●) in promoting initiation complex formation on *E. coli* ribosomes. A blank (0.05 pmol) representing the amount of fMet-tRNA bound in the absence of $IF3_{chl}$ has been subtracted from each value. (B) Stimulation of initiation complex formation on chloroplast 30S ribosomal subunits in the presence of 10 pmol of *mRbcN (33)*. A blank (0.1 pmol) representing the amount of fMet-tRNA bound in the absence of $IF3_{chl}$ has been subtracted from each value.

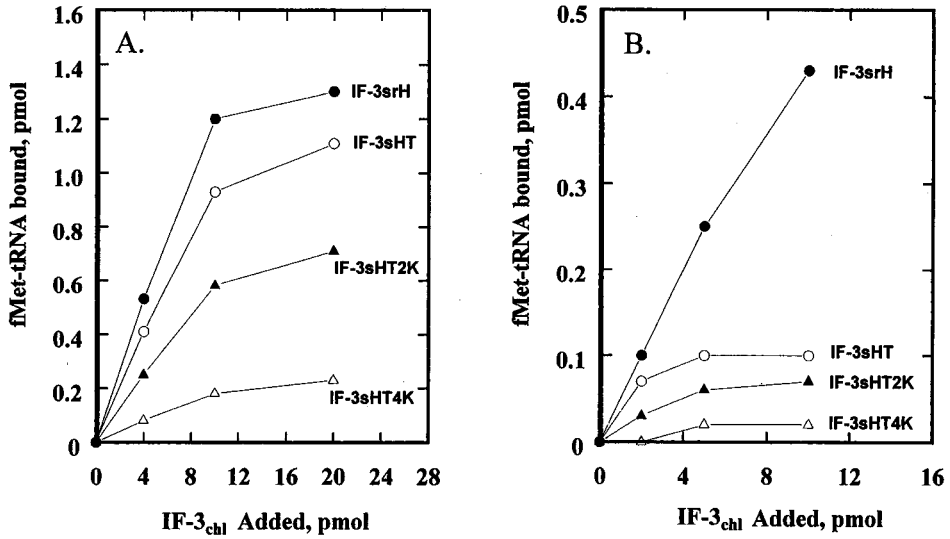

FIGURE 6: Effect of mutation of lysine residues in IF3$_{chl}$sHT on initiation complex formation. A. The activities of IF3$_{chl}$sHT2K (▲) and IF3$_{chl}$sHT4K (△) were compared to those of IF3$_{chl}$sHT (○), and IF3$_{chl}$srH (●) in promoting initiation complex formation on *E. coli* ribosomes. A blank (0.05 pmol) representing the amount of fMet-tRNA bound in the absence of IF3$_{chl}$ has been subtracted from each value. (B) Stimulation of initiation complex formation on chloroplast 30S ribosomal subunits in the presence of 10 pmol of *mRbcN (33)*. A blank (0.1 pmol) representing the amount of fMet-tRNA bound in the absence of IF3$_{chl}$ has been subtracted from each value.

The effects of mutation of the lysine residues in the linker on the activity of IF3$_{chl}$sHT were also examined. When tested on *E. coli* 70S ribosomes using poly(A,U,G) as a mRNA (Figure 6A), IF3$_{chl}$sHT2K has about one-half of the activity observed with IF3$_{chl}$sHT. This observation is surprising since mutation of these residues in IF-3$_{chl}$srH does not have a significant effect on the activity of the homology domain in this assay. When tested for initiation complex formation with chloroplast 30S subunits (Figure 6B), IF3$_{chl}$-sHT2K is also less active than IF3$_{chl}$sHT. Sequences in the tail region partially inhibit the function of the homology domain of IF3$_{chl}$. Mutation of the two less-conserved lysine residues in the linker region appears to affect the interaction between the tail and linker in a way that increases the inhibitory effect of the tail. This inhibitory effect is now seen not only in assays with chloroplast 30S subunits but also in assays on *E. coli* ribosomes.

Mutation of all four lysine residues in the linker region of IF3$_{chl}$sHT has an even more drastic effect. IF3$_{chl}$sHT4K has lost approximately 80% of the activity observed with IF3$_{chl}$-sHT in the *E. coli* assay (Figure 6A). This observation indicates that the conserved lysine residues in the linker region play a significant role in the function of IF3. This observation is in agreement with the negative effect that mutation of these residues has on the activity of the homology domain itself (IF3$_{chl}$srH). Mutation of the conserved and less-conserved lysine residues in IF3$_{chl}$sHT appears to have a cumulative effect. When tested on chloroplast 30S subunits using *mRbcN* as mRNA (Figure 6B), IF3$_{chl}$sHT4K has essentially no activity. Again, mutation of both the less-conserved and conserved lysine residues within the linker region appears to have a cumulative effect on the activity of IF3$_{chl}$sHT.

Binding of the Mutated Derivatives to 30S Subunits. To test if lysine residues in the linker region are important for the binding of IF3$_{chl}$ to the chloroplast 30S subunit, direct measurements of the binding of the normal and mutated proteins to 30S subunits were carried out. In these experiments, IF3$_{chl}$ or its derivatives were incubated with chloroplast 30S subunits. The IF3$_{chl}$ bound to the subunits was separated from the free factor by sucrose density-gradient centrifugation. The amount of IF3$_{chl}$ bound to the subunits was calculated based on the standard curve of an ELISA using antibodies raised against the homology domain. Each derivative shows a similar response to the antibody (data not shown). The amounts of each derivative bound to 30S subunits and the K_{obs} values were calculated (Table 2). Mutation of the less-conserved lysine residues present in IF3$_{chl}$ has no effect on the ability of the homology domain of IF3$_{chl}$ to bind 30S subunits. Similarly, mutation of these residues in IF3$_{chl}$sHT has no effect on the binding of this derivative to ribosomal subunits. These observations suggest that these two lysine residues are not directly involved in the interaction of IF3$_{chl}$ with 30S subunits. However, when the conserved lysine residues are also mutated to give the 4K derivatives, IF3$_{chl}$srH4K and IF3$_{chl}$sHT4K, over a 100-fold decrease in affinity is observed. This result indicates that these two conserved lysine residues in the linker region play a direct role in the binding of IF3$_{chl}$ to 30S subunits. By analogy, it is likely that these conserved lysine residues are also important in the binding of the corresponding bacterial factors to the small subunit of the ribosome.

Table 2: Binding of IF3$_{chl}$ to 30S Subunits[a]

IF3 type	$[IF3]_0$, M	$[30S]_0$, M	IF3 bound, pmol	K_{obs}, M^{-1}
IF3srH	1×10^{-7}	0.8×10^{-7}	8.5	$(1.1-1.5) \times 10^7$
IF3srH2K	1×10^{-7}	0.8×10^{-7}	8.8	$(1.1-1.5) \times 10^7$
IF3srH4K	1×10^{-7}	0.8×10^{-7}	<0.2	$<1 \times 10^5$
IF3sHT	1×10^{-7}	0.8×10^{-7}	4.3	$(3.0-3.3) \times 10^6$
IF3sHT2K	1×10^{-7}	0.8×10^{-7}	4.6	$(3.0-3.7) \times 10^6$
IF3sHT4K	1×10^{-7}	0.8×10^{-7}	<0.2	$<1 \times 10^5$

[a] Approximate binding constants (K_{obs}) for the interaction of IF3$_{chl}$ and its derivatives with chloroplast 30S subunits were calculated according to the equation, 30S + IF3$_{chl}$ ⇌ [30S:IF3$_{chl}$]. The equilibrium association constant was calculated as K_{obs} = [30S:IF3$_{chl}$]/[30 S][IF3$_{chl}$]. Bound IF3$_{chl}$ was quantified by an ELISA. The initial concentrations used are designated $[IF3]_0$ and $[30S]_0$.

Thermal Stability of IF3$_{chl}$ Derivatives. To address the question of the interaction between the tail region and the linker region, CD-detected thermal denaturation studies were

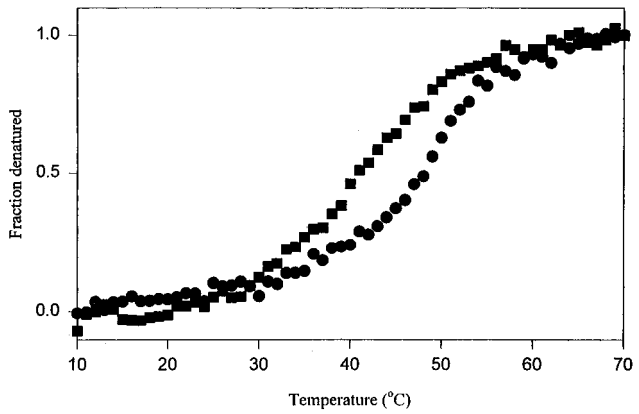

FIGURE 7: CD-monitored thermal denaturation of $IF3_{chl}$srH and $IF3_{chl}$sHT. Denaturation of $IF3_{chl}$srH (■) and $IF3_{chl}$sHT (●) was monitored from 10 to 70 °C as described in Materials and Methods.

Table 3: Melting Temperatures for $IF3_{chl}$ and Its Derivatives

IF3 type	T_m (°C)
IF3srH	40
IF3srH2K	40
IF3srH4K	40
IF3sHT	48
IF3sHT2K	44
IF3sHT4K	40

performed on the normal and mutated proteins. Figure 7 shows the temperature-dependent denaturation curves for $IF3_{chl}$srH and $IF3_{chl}$sHT. The CD signal collected at 222 nm (characteristic for a helical structure) has been normalized using the signal at 10 °C as 100% native and the signal at 70 °C as 100% denatured. The melting temperature (T_m) is the temperature at which 50% of the signal disappears. The results show that $IF3_{chl}$srH has a melting temperature of ∼40 °C, while $IF3_{chl}$sHT has a melting temperature of about 48 °C. These data indicate that $IF3_{chl}$sHT melts as a single cooperative unit, leading to the conclusion that the tail and homology domain of $IF3_{chl}$ may interact to stabilize the protein as a whole. Thermal denaturation is reversible for $IF3_{chl}$srH, but not for $IF3_{chl}$sHT (data not shown). Quantitative thermodynamic data, thus, cannot be obtained.

Temperature-dependent denaturation profiles were also determined with the 2K and 4K mutants of $IF3_{chl}$srH and $IF3_{chl}$sHT. Both the $IF3_{chl}$srH2K and the $IF3_{chl}$srH4K mutants have the same melting temperature as $IF3_{chl}$srH (Table 3). This observation suggests that mutating the lysine residues in the linker region does not affect the thermal stability of the homology domain. In contrast, changing the two less-conserved lysine residues in $IF3_{chl}$sHT changes the melting temperature from 48 to ∼44 °C. Further mutation of the conserved lysine residues in the linker region reduces the melting temperature even more, to about 40 °C (Table 3). These data are consistent with the idea that the tail region stabilizes the homology domain by interacting with the lysine residues in the linker region of the homology domain. These observations provide direct evidence that the tail region and the linker region interact with each other. Mutation of these lysine residues alters this interaction resulting in a decrease in the melting temperature. The data presented here support the idea that the tail region of $IF3_{chl}$ plays a negative role in the initiation of protein synthesis by interacting with the linker region in the homology domain.

The negative effects of sequences in the head and tail regions on the activity of the homology domain of $IF3_{chl}$ raises the interesting question of how these effects are modulated *in vivo*. It is clear from numerous studies (*46*−*49*) that chloroplast protein synthesis is regulated through a complex series of trans-acting mRNA binding proteins that couple the synthesis of certain proteins to the presence of light. The head and tail regions of $IF3_{chl}$ may play a role in coupling the activity of $IF3_{chl}$ to translational regulatory proteins in the chloroplast. It is likely that sequences in the head and tail region are involved in protein−protein interactions that modulate the effect of the head and tail on the activity of the homology domain during the initiation of translation. These protein−protein interactions may play a crucial role in the control of chloroplast translation by light.

ACKNOWLEDGMENT

The authors thank Drs. Iosif Vaisman and Alex Tropsha for help with the molecular modeling.

REFERENCES

1. Paci, M., Pon, C., Lammi, M., and Gualerzi, C. (1984) *J. Biol. Chem. 259*, 9628−9634.
2. Dottavio-Martin, D., Suttle, D. P., and Ravel, J. M. (1979) *FEBS Lett. 97*, 105−110.
3. Van Knippenberg, P. (1990) in *The ribosome: Structure, function and evolution* (Hill, W., Dahlberg, A., Garrett, R., Moore, P., Schlessinger, D., and Warner, J., Eds.) pp 265−274, American Society for Microbiology, Washington, DC.
4. Gualerzi, C., and Pon, C. (1990) *Biochemistry 29*, 5881−5889.
5. Gualerzi, C., La Teana, A., Spurio, R., Canonaco, M., Severini, M., and Pon, C. (1990) in *The Ribosome: Structure, function and evolution* (Hill, W., Dahlberg, A., Garrett, R., Moore, P., Schlessinger, D., and Warner, J., Eds.) pp 281−291, American Society for Microbiology, Washington, DC.
6. Firpo, M., Connelly, M., Goss, D., and Dahlberg, A. (1996) *J. Biol. Chem. 271*, 4693−4698.
7. Muralikrishna, P., and Wickstrom, E. (1989) *Biochem. 28*, 7505−7510.
8. Laughrea, M., and Tam, J. (1991) *Biochemistry 30*, 11412−11420.
9. Moazed, D., Samaha, R. R., Gualerzi, C., and Noller, H. F. (1995) *J. Mol. Biol. 248*, 207−210.
10. Canonaco, M., Gualerzi, C., and Pon, C. (1989) *Eur. J. Biochem. 182*, 501−506.
11. Teana, A., Pon, C., and Gualerzi, C. (1996) *J. Mol. Biol. 256*, 667−675.
12. Berkhout, B., van der Laken, C. J., and van Knippenberg, P. H. (1986) *Biochim. Biophys. Acta 866*, 144−153.
13. Hartz, D., McPheeters, D., and Gold, L. (1989) *Genes Dev. 3*, 1899−1912.
14. Hartz, D., Binkley, J., Hollingsworth, T., and Gold, L. (1990) *Genes Dev. 4*, 1790−1800.
15. Sussman, J., Simons, E., and Simons, R. (1996) *Mol. Microbiol. 21*, 347−360.
16. Mangroo, D., Wu, X., and RajBhandary, U. (1995) *Biochem. Cell Biol. 73*, 1023−1031.
17. Butler, J. S., Springer, M., and Grunberg-Manago, M. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 4022−4025.
18. Kycia, J., Biou, V., Shu, F., Gerchman, S., Graziano, V., and Ramakrishnan, V. (1995) *Biochemistry 34*, 6183−6187.
19. Fortier, P., Schmitter, J., Garcia, C., and Dardel, F. (1994) *Biochimie 76*, 376−383.
20. Box, R., Wooley, P., and Pon, C. (1981) *Eur. J. Biochem.* 116, 93−99.
21. Biou, V., Shu, F., and Ramakrishnan, V. (1995) *EMBO J. 14*, 4056−4064.
22. Bellis, D., Liveris, D., Goss, D., Ringquist, S., and Schwartz, I. (1992) *Biochemistry 31*, 11984−11990.

23. Garcia, C., Fortier, P., Blanquet, S., Lallemand, J., and Dardel, F. (1995) *J. Mol. Biol. 254,* 247−259.
24. Garcia, C., Fortier, P., Blanquet, S., Lallemand, J., and Dardel, F. (1995) *Eur. J. Biochem. 228,* 395−402.
25. Moreau, M., de Cock, E., Fortier, P., Garcia, C., Albaret, C., Blanquet, S., Lallemand, J., and Dardel, F. (1997) *J. Mol. Biol. 266,* 15−22.
26. Wang, C., and Spremulli, L. (1991) *J. Biol. Chem. 266,* 17079−17083.
27. Lin, Q., Ma, L., Burkhart, W., and Spremulli, L. (1994) *J. Biol. Chem. 269,* 9436−9444.
28. Lin, Q., Yu, N., and Spremulli, L. (1996) *Plant Mol. Biol. 32,* 937−945.
29. Cheng, Y., Kalman, L., and Kaiser, D. (1994) *J. Bacteriol. 176,* 1427−1433.
30. Kalman, L., Cheng, Y., and Kaiser, D. (1994) *J. Bacteriol. 176,* 1434−1442.
31. Graves, M., and Spremulli, L. (1983) *Arch. Biochem. Biophys. 222,* 192−199.
32. Pon, C. L. and Gualerzi, C. (1979) *Meth. Enzymol. 60* (*230-9*), 230−239.
33. Wang, C., Roney, W., Alston, R., and Spremulli, L. (1989) *Nucleic Acids Res. 17,* 9735−9747.
34. Graves, M., Breitenberger, C., and Spremulli, L. (1980) *Arch. Biochem. Biophys. 204,* 444−454.
35. de Smit, M., and van Duin, J. (1994) *J. Mol. Biol. 235,* 173−184.
36. Gold, J., and Spremulli, L. (1985) *J. Biol. Chem. 260,* 14897−14900.
37. Ma, L., and Spremulli, L. (1992) *J. Biol. Chem. 267,* 18356−18360.
38. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) *Nucleic Acids Res. 22,* 4673−4680.
39. Geourjon, C., and Deleage, G. (1995) *CABIOS 11,* 681−684.
40. Rost, B., Sander, C., and Schneider, R. (1994) *CABIOS 10,* 53−60.
41. Browning, K., Humphreys, J., Hobbs, W., Smith, G., and Ravel, J. (1990) *J. Biol. Chem. 265,* 17967−17973.
42. Gill, S., and von Hippel, P. (1989) *Anal. Biochem. 182,* 319−326.
43. Elwell, M. L., and Schellman, J. A. (1977) *Biochim. Biophys. Acta 494,* 367−383.
44. Johnson, W. C. (1990) *Proteins: Struct., Funct., and Genet. 7,* 205−214.
45. Gibbs, C., and Zoller, M. (1991) *Methods 3,* 165−173.
46. Danon, A., and Mayfield, S. (1991) *EMBO J. 10,* 3993−4001.
47. Hirose, T., and Sugiura, M. (1996) *EMBO J. 15,* 1687−1695.
48. Danon, A., and Mayfield, S. (1994) *EMBO J.* 13, 2227−2235.
49. Zerges, W., and Rochaix, J. (1994) *Mol. Cell. Biol. 14,* 5268−5277.
50. Sayle, R. A., and Milner, E. J. (1995) *Trends Biochem. Sci. 20,* 374−376.

BI971185Y