Structure of the β Subunit of Translation Initiation Factor 2 from the Archaeon *Methanococcus jannaschii*: A Representative of the eIF2β/eIF5 Family of Proteins[†,‡]

Seongeun Cho and David W. Hoffman*

Department of Chemistry and Biochemistry, Institute for Cellular and Molecular Biology, University of Texas at Austin, Austin, Texas 78712

Received October 30, 2001; Revised Manuscript Received February 19, 2002

ABSTRACT: The β subunit of archaeal translation initiation factor 2 (aIF2β) is a representative of a family of proteins whose members include the β subunit of eukaryotic translation initiation factor 2 (eIF2β) and the N-terminal domain within translation initiation factor 5 (eIF5); no members of this family of proteins have been structurally characterized up to this time. In the work presented here, aIF2β from *Methanococcus jannaschii* was expressed in *Escherichia coli*, purified, and analyzed using multidimensional NMR methods. The aIF2β was found to contain two independent structural domains. The N-terminal domain contains a four-stranded antiparallel β sheet and two α helices, and is structurally similar to the DNA-binding domain of a yeast heat shock transcription factor and a domain within ribosomal protein S4. This structural similarity was an unanticipated result, since no significant homology was detected at the level of primary sequence. The C-terminal domain of aIF2β contains a zinc-binding motif of three antiparallel β strands, with four conserved cysteines arranged as two CXXC units separated by 17 residues. Conserved residues on the surface of each domain that are likely candidates for direct interaction with other components of the translational apparatus were identified. The significant primary sequence homology between archaeal aIF2β and the eukaryotic eIF2β and eIF5, when combined with the structural results in the work presented here, permitted structural features to be predicted for these latter two eukaryotic proteins.

The eukaryotic translation initiation factor 2 (eIF2) and its homologue in archaea, aIF2,[1] are heterotrimeric proteins with α, β, and γ subunits. Eukaryotic eIF2, and presumably aIF2 in archaea, play essential roles in the recognition of the correct codon for the start of translation (for an overview, see refs *1*−*3*). eIF2 forms a ternary complex with the methionine initiator tRNA and GTP, which binds to the 40S ribosomal subunit as part of the 43S preinitiation complex. The β and γ subunits of eIF2 are responsible for recruiting the initiator tRNA and GTP (*4*−*6*), while the α subunit is involved in the regulation of the translation initiation process. All three subunits of eIF2 and aIF2 are well conserved among the diverse species of eukaryotes and archaea, but do not occur in prokaryotes. The key role of eIF2 in the initiation of translation has resulted in its being the focus of a large number of biochemical and genetic studies, which have provided insights into function and interactions between the eIF2 subunits and various components of the translation apparatus. However, little is known regarding the structure of either eIF2 or aIF2, since there are no reported high-resolution structural models of either complex or any of the individual subunits. Thus, the understanding of the underlying mechanism of the function of this key translation initiation factor remains relatively poor.

The α subunit of eIF2 (termed eIF2α) is a major site of control of overall protein synthesis in eukaryotes (*7*, *8*). Phosphorylation of eIF2α blocks its ability to recycle by increasing the affinity of eIF2 for GDP and eIF2B, resulting in the inhibition of exchange of GDP for GTP, eventually leading to inhibition of initiator tRNA binding and thus of translation. The most strongly conserved N-terminal region of the α subunit has a primary sequence that predicts an OB-like domain (*9*, *10*) and contains the conserved serine that serves as the site of phosphorylation. The γ subunit of eIF2 (eIF2γ) is believed to be responsible for binding GTP, and is highly conserved among the various eukaryotes, with a level of identity of more than 72% (*6*, *11*−*13*). eIF2γ contains the GTP binding motif found in a superfamily of GTP-binding proteins, which includes prokaryotic translation, elongation, and termination factors. The sequence of eIF2γ is significantly homologous to that of EF-Tu (27% identical, approximately 50% similar), with its sequence being compatible with the three-domain structure of EF-Tu (*14*).

Sequence analysis of the β subunit of eIF2 (eIF2β) from a wide selection of eukaryotes and archaea reveals conserved

[†] This work was supported by a grant from the Robert A. Welch Foundation (F-1353) and a grant from the American Cancer Society (GMC-89306).

[‡] Structures have been deposited with the Protein Data Bank (PDB entries 1K8B and 1K81). Chemical shifts have been deposited with the BioMagResBank (accession number 5428).

* To whom correspondence should be addressed. E-mail: dhoffman@mail.utexas.edu. Telephone: (512) 471-7859. Fax: (512) 471-8696.

[1] Abbreviations: aIF2, archaeal initiation factor 2; IPTG, isopropyl β-thiogalactopyranoside; PCR, polymerase chain reaction; NMR, nuclear magnetic resonance; NOE, nuclear Overhauser effect; HSQC, heteronuclear single-quantum coherence.

10.1021/bi011984n CCC: $22.00
Published on Web 04/12/2002

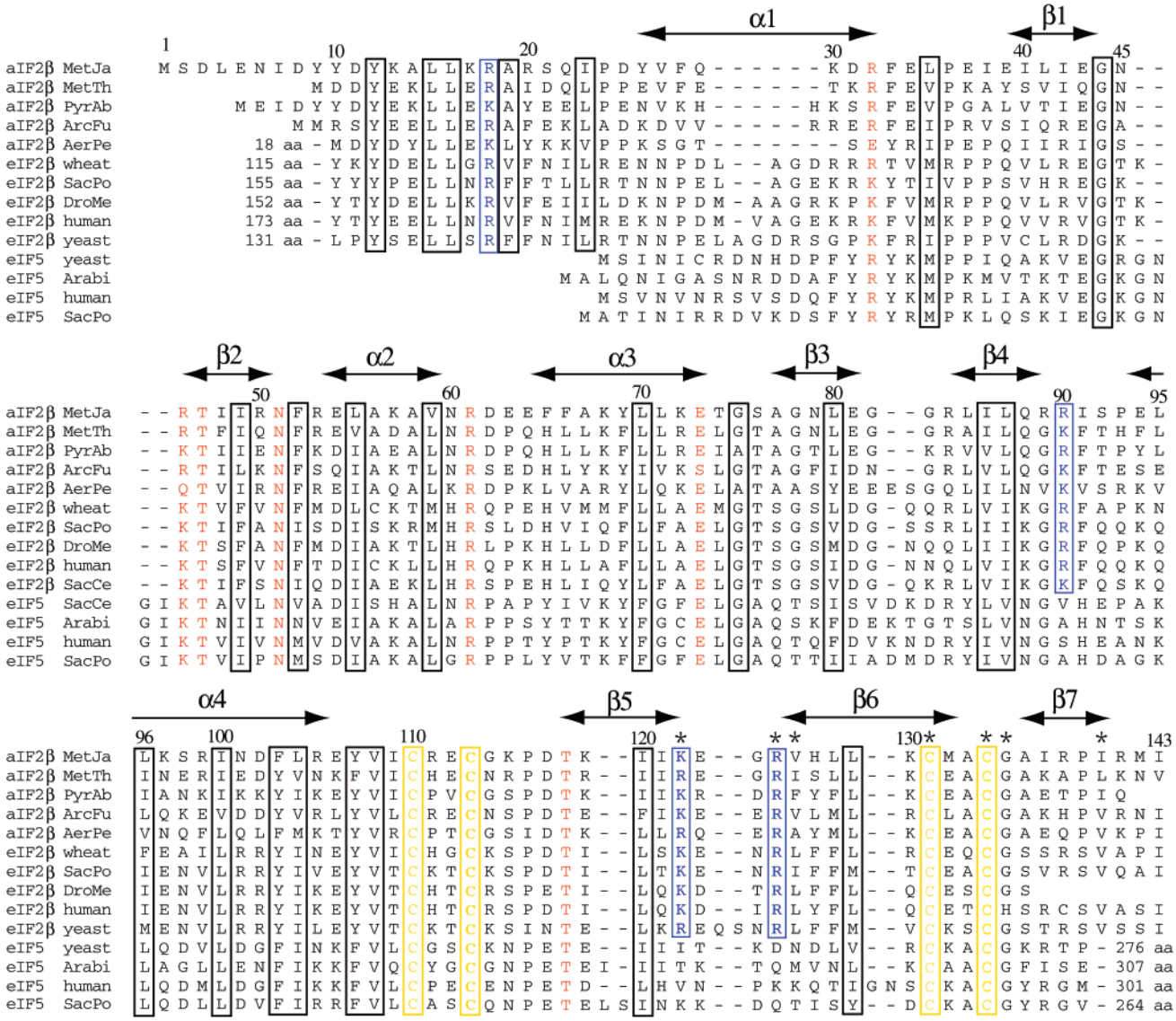

FIGURE 1: Alignment of amino acid sequences of the aIF2β protein from five species of archaea, eIF2β from five species of eukaryotes, and the N-terminal domain of eIF5 from four species of eukaryotes. The host organism and the Swissprot ID number for each represented sequence are given: archaea *M. jannaschii* Q57562 (aIF2β MetJa), *Methanobacterium thermoautotrophicum* O27797 (aIF2β MetTh), *Pyrococcus abyssi* O58312 (aIF2β PyrAb), *Archaeoglobus fulgidus* O27958 (aIF2β ArcFu), and *Aeropyrum pernix* (AerPe); eukaryotes *Triticum aestivum* O24473 (eIF2β wheat), *Saccharomyces pombe* P56329 (eIF2β SacPo), *Drosophila melanogaster* P41373 (eIF2β DroMe), *Homo sapiens* P20042 (eIF2β human), and *Saccharomyces cerevisiae* 320774 (eIF2β yeast); and *S. cerevisiae* P38431 (eIF5 yeast), *Arabidopsis thaliana* (eIF5 Arabi), *H. sapiens* P55010 (eIF5 human), and *S. pombe* Q09689 (eIF5 SacPo). The four conserved zinc-binding cysteines in the zinc ribbon domain are boxed in yellow. Other residues that are most likely to be conserved for structural purposes are boxed in black; these residues provide strong evidence that the structures of aIF2β, eIF2β, and eIF5 proteins are significantly similar. Residues that are accessible on the protein surface and are similar or identical in greater than 90% of the aIF2β, eIF2β, and eIF5 sequences are indicated in red. Lysine and arginine residues indicated in blue are highly conserved among the aIF2β and eIF2β sequences, but have identities that are distinct from the residues at the equivalent position in eIF5; these residues are the most likely to be involved in interactions that distinguish eIF2β from eIF5. Asterisks indicate the positions of residues in eIF2β that have been implicated in initiation site recognition in yeast (*49*). Although only 14 sequences are shown in the figure, a larger set of sequences was compared in deciding which residues are the most well conserved.

regions and several notable features (Figure 1). eIF2β ranges from 250 to 333 amino acids in length. The N-terminal regions of the eukaryotic versions of these proteins are divergent in their sequence and length, though they share the notable feature of three polylysine tracts, which have been implicated in binding the GTP exchange factor eIF2B and GTPase-activating protein eIF5 (*15*−*17*). The homologues of eIF2β found in archaea lack the divergent, hydrophilic N-terminal region that contains the polylysine tracts. The C-terminal half of eIF2β is well-conserved across the different species of eukaryotes and is similar to the full length of aIF2β in archaea. This region of eIF2β has been shown to bind to the eIF2γ subunit in a recent study (*18*), contains a C_2−C_2 zinc finger-like motif near its C-terminal end that is conserved among all known archaea and eukaryotic sequences, and is significantly homologous in sequence to the N-terminal region of translation initiation factor eIF5. The relationships between eIF2β and eIF5 in eukaryotes and aIF2β in archaea are represented schematically in Figure 2.

In the work presented here, we have determined the solution structure of the β subunit of archaeal translation initiation factor 2 (aIF2β) from the *Methanococcus jann-*

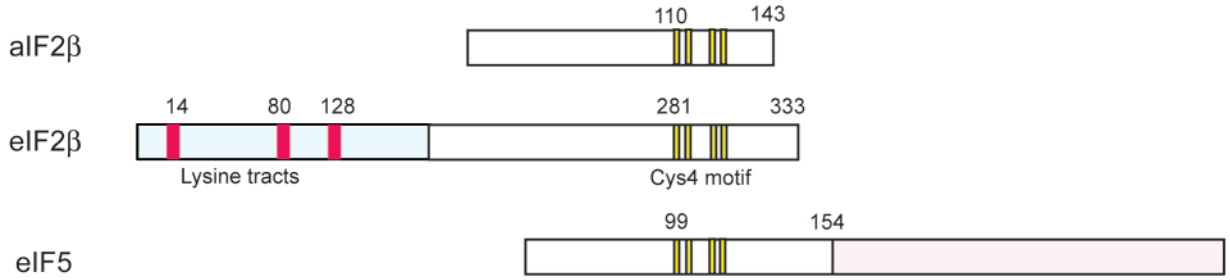

FIGURE 2: Schematic diagram showing the relative locations of conserved structural regions in translation initiation factors aIF2β, eIF2β, and eIF5. Numbers indicate the positions of structural elements within each protein. Residues of aIF2β are numbered as in the *M. jannaschii* sequence. Residues of eIF2β and eIF5 are numbered as in the human sequences. The conserved four-cysteine motif occurs in all known aIF2β, eIF2β, and eIF5 sequences, and is indicated in yellow.

aschii, a 16 kDa protein whose amino acid sequence is 28% similar to that of human eIF2β, 30% similar to that of eIF2β from yeast, and approximately 50% similar to aIF2β sequences of other archaea. A search of the Protein Data Bank showed that aIF2β, eIF2β, and the N-terminal half of eIF5 are not significantly homologous to any protein with known structure. These proteins have been collectively termed members of the eIf2β/eIF5 structural family.

MATERIALS AND METHODS

Protein Cloning, Expression, and Purification. The gene encoding the aIF2β protein was amplified by PCR from a fragment of *M. jannaschii* genomic DNA in an *Escherichia coli* plasmid obtained from American Type Culture Collection (clone AMJPK64). PCR primers were designed so as to incorporate unique *Nde*I and *Bam*HI restriction sites, permitting the insertion of the amplified aIF2β gene into the pET-14b expression vector (Novagen) which encodes a six-residue histidine tag and a thrombin cleavage site at the N-terminus of the protein. DNA sequencing confirmed the correct sequence of the full-length aIF2β gene. The histidine-tagged aIF2β protein was expressed in *E. coli* BL21(DE3) cells. Cells were grown at 37 °C in Luria broth supplemented with ampicillin, and protein expression was induced when the cells reached an OD_{550} of 0.6 with 0.4 mM IPTG. Cells were harvested by centrifugation at 6000 rpm for 10 min. The cell pellet from a 1 L culture was suspended in 10 mL of 0.5 M NaCl, 0.1 M Tris-HCl (pH 7.9) buffer, and the cells were lysed by freezing and thawing. The resulting cell suspension was centrifuged for 20 min at 14 000 rpm. The clear supernatant was loaded onto a nickel affinity column (Pharmacia), and the protein was eluted from the column with 0.5 M imidazole (pH 8). The protein-containing eluent was submitted to dialysis against 10 mM phosphate buffer at pH 7.0, and thrombin was added to remove the N-terminal histidine tag. The dialysate was loaded onto an SP-Sepharose (Pharmacia) column, and pure aIF2β was eluted using a linear salt gradient from 0.0 to 1.0 M NaCl. The fractions containing aIF2β were pooled and dialyzed against 10 mM sodium phosphate buffer at pH 6 before being concentrated in a Centricon-10 microconcentrator (Amicon).

Samples of aIF2β enriched in ^{15}N and/or ^{13}C were prepared as described above, but with M9 minimal medium containing 1 g/L [^{15}N]ammonium chloride and/or 1 g/L [^{13}C]glucose (Cambridge Isotope Laboratories) as the source of nitrogen and/or carbon. A protein sample selectively labeled with [^{15}N]lysine was prepared by growing the cells in M9 minimal medium supplemented with 50 mg/L ^{15}N-labeled lysine and 100 mg of the other 19 unlabeled amino acids per liter. A protein sample selectively labeled with [^{15}N]glycine, serine, and cysteine was prepared using M9 minimal medium supplemented with 100 mg/L ^{15}N-labeled glycine and 100 mg of each of 16 unlabeled amino acid types per liter (all types except glycine, serine, cysteine, and tryptophan). Samples that were simultaneously enriched with [^{15}N]-(uniform) and 1-[^{13}C]leucine (or 1-[^{13}C]isoleucine, 1-[^{13}C]-phenylalanine, or 1-[^{13}C]valine) were prepared by growing the cells in M9 minimal medium containing 1 g/L [^{15}N]-ammonium chloride and 100 mg/L 1-^{13}C-labeled amino acid. A synthetic peptide corresponding to residues 94−143 of *M. jannaschii* aIF2β was obtained from Biosyn, Inc.

NMR Spectroscopy. NMR spectra were recorded at 30 °C using a 500 MHz Varian Inova spectrometer equipped with a triple-resonance probe and *z*-axis pulsed field gradient. The NMR sample typically contained 1.0−1.5 mM aIF2β protein and 10 mM sodium phosphate in a 90% H_2O/10% D_2O solvent at pH 6. Pulse sequences were obtained from L. Kay's group at the Toronto NMR center, and have been optimized before use on local NMR instrumentation. Backbone resonance assignments were obtained using three-dimensional HNCA (*19*), HNCO (*20*), HCACO (*21*), HNCACB (*19*), HACACBCO (*22*), and HN(CO)CACB (*19*) spectra that correlate the backbone protons to the N, C^{α}, C^{o}, and C^{β} signals of the same and adjacent amino acid residues. Side chain resonance assignments were obtained by analyzing three-dimensional ^{15}N-edited HMQC-TOCSY and ^{13}C-edited HCCH-TOCSY (*23*) spectra, and two-dimensional homonuclear 2QF-COSY and TOCSY spectra. NOE cross-peaks were detected using two-dimensional $^{1}H-^{1}H$ NOESY, three-dimensional ^{15}N-edited HSQC-NOESY, and three-dimensional ^{13}C-edited HSQC-NOESY (*24*) spectra. The ^{13}C-edited HSQC-NOESY spectrum was acquired in a 90% H_2O/10% D_2O solvent so that NOE cross-peaks between amide and side chain protons could be resolved by the chemical shift of the ^{13}C nucleus coupled to the side chain proton. In several instances, chemical shift assignment ambiguities were removed with the help of the amino acid type specifically labeled samples. For example, amide resonances of amino acids that follow phenylalanine were identified in an HNCO spectrum of a protein that was uniformly enriched in ^{15}N and specifically enriched with 1-[^{13}C]phenylalanine; an analogous procedure was used to assign resonances of amino acids that follow valine, leucine, and isoleucine. Similarly, assignments of glycine, serine, cysteine, and lysine resonances were confirmed using HSQC spectra of protein samples selectively enriched with the specific amino acid

Table 1: Structural Statistics for aIF2β from *M. jannaschii*

	N-terminal domain (residues 39−92)	C-terminal domain (residues 108−140)
restraints for structure calculation		
total no. of NOE restraints	222	156
intraresidue	70	69
interresidue		
sequential ($\|i-j\|=1$)	77	59
medium-range ($1<\|i-j\|<5$)	27	10
long-range ($\|i-j\|\leq 5$)	48	18
no. of dihedral angle restraints[a] (ϕ, ψ)	72 (36, 36)	34 (16, 18)
no. of hydrogen bonds[a]	28	12
statistics for structure calculation		
rmsd[b] for backbone heavy atoms (Å)	0.87	1.60
rmsd[b] for all heavy atoms (Å)	2.27	3.01
deviation from NOE restraints (Å)	0.065 ± 0.004	0.034 ± 0.002
deviation from dihedral restraints (deg)	0.696 ± 0.157	0.441 ± 0.102
no. of NOE violations of >0.5 Å	none	none
no. of dihedral angle violations of >5 Å	none	none
Ramachandran plot		
residues in most favorable region (%)	74.7	75.5
residues in other favorable regions (%)	22.6	20.2
residues in disallowed region (%)	2.7	4.3
rmsd from ideal bond lengths (Å)	0.0041 ± 0.0004	0.0024 ± 0.0001
rmsd from ideal covalent angles (deg)	0.668 ± 0.027	0.461 ± 0.062
rmsd for improper angles (deg)	0.652 ± 0.035	0.189 ± 0.045

[a] Each hydrogen bond was defined by two distance restraints: one between the donor hydrogen and the acceptor heavy atom (1.9−2.1 Å) and one between the donor heavy atom and the carbonyl carbon atom (3.7−4.3 Å). Hydrogen bond restraints were only included for elements of regular secondary structure that were identified by characteristic NOE cross-peak patterns. Backbone dihedral angles ψ and ϕ were restrained with a range of ±20°. [b] Calculated from residues 39−90 in the N-terminal domain and residues 108−140 in the C-terminal domain.

types labeled with ^{15}N. Data were processed using the program Felix (Hare Research). ^{1}H, ^{15}N, and ^{13}C chemical shifts are referenced as recommended by Wishart et al. (*25*), with proton chemical shifts referenced to internal 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) at 0 ppm. The 0 ppm ^{13}C and ^{15}N reference frequencies were determined by multiplying the 0 ppm ^{1}H reference frequency by 0.251 449 530 and 0.101 329 118, respectively.

Structure Calculation. Structure calculations were performed using the restrained simulated annealing protocol in the program CNS (*26*), with the goal of identifying the full range of structures that are consistent with the distance and angle constraints derived from the NMR data while having reasonable molecular geometry, consistent with a minimum value of the CNS energy function. Distance restraints were derived from the intensities of cross-peaks within multidimensional NOE spectra. Whenever possible, cross-peaks were identified in two-dimensional ^{1}H−^{1}H NOE spectra (in either H_2O or D_2O solvent) acquired with a mixing time of 60 ms, to minimize the effects of spin diffusion; additional distance bounds were obtained from the cross-peaks in three-dimensional ^{15}N-edited HSQC-NOESY and ^{13}C-edited HSQC-NOESY spectra. On the basis of the cross-peak intensity in the NOESY spectra, distance restraints were classified as strong (1.8−3.5 Å), medium (1.8−4.0 Å), weak (2.5−5.0 Å), and very weak (2.5−6.0 Å); these distance bounds were calibrated by using interproton distances in regions of regular secondary structure as internal distance standards. For NOEs involving methyl protons, an additional 1.0 Å was applied to the upper limits for the interproton distances, with distances then being measured from the center of the three protons in the methyl group. Backbone dihedral angle restraints were included for residues within the regions of regular α helical or β strand secondary structure, identified by characteristic NOE cross-peaks, chemical shift indices (CSI) (*27*) for C^{α} and C^{o}, and patterns of protection of the amide protons from exchange with the solvent. Each hydrogen bond was defined using two distance restraints: one between the donor hydrogen and the acceptor heavy atom (1.9−2.1 Å) and one between the donor heavy atom and carbonyl carbon atom (3.7−4.3 Å). Hydrogen bond restraints were only included for amide protons that were located within the regions of regular α helical or β strand structure. Covalent restraints for the Cys4 zinc site consisted of a length of 2.3 Å for the cysteine S^{γ}−Zn bond and a value of 109.3° for the cysteine C^{β}−S^{γ}−Zn bond angle. Experimental restraints used for the structure calculations and structural statistics are summarized in Table 1.

For each of the two domains, an initial set of 60 structures was generated from an extended peptide conformation using a simulated annealing protocol with dihedral angle restraints only. From this set, 17 structures with low overall energies were selected for further refinement. Pseudoatom corrections were employed for aromatic rings, methyl groups, and nonselectively assigned methylene protons in the initial simulated annealing process (*28*). After an initial round of structure refinement using NOE-derived distance restraints, a set of structures with low overall energies was inspected to discriminate between possible NOE cross-peak assignments involving the two methyl groups of the leucines and valines and the two δ and ε protons of the phenylalanine and tyrosine rings, thus eliminating the need for some of the pseudoatom corrections. The selected structures were then used as starting points to generate 10 structures each, via restrained simulated annealing using different initial trajectories. A set of refined conformers having the lowest energy were retained for final analysis and evaluation using Procheck-NMR (*29*), with the statistics reported in Table 1. These final structures are a fair representation of the full range of structures that are consistent with the experimental data while

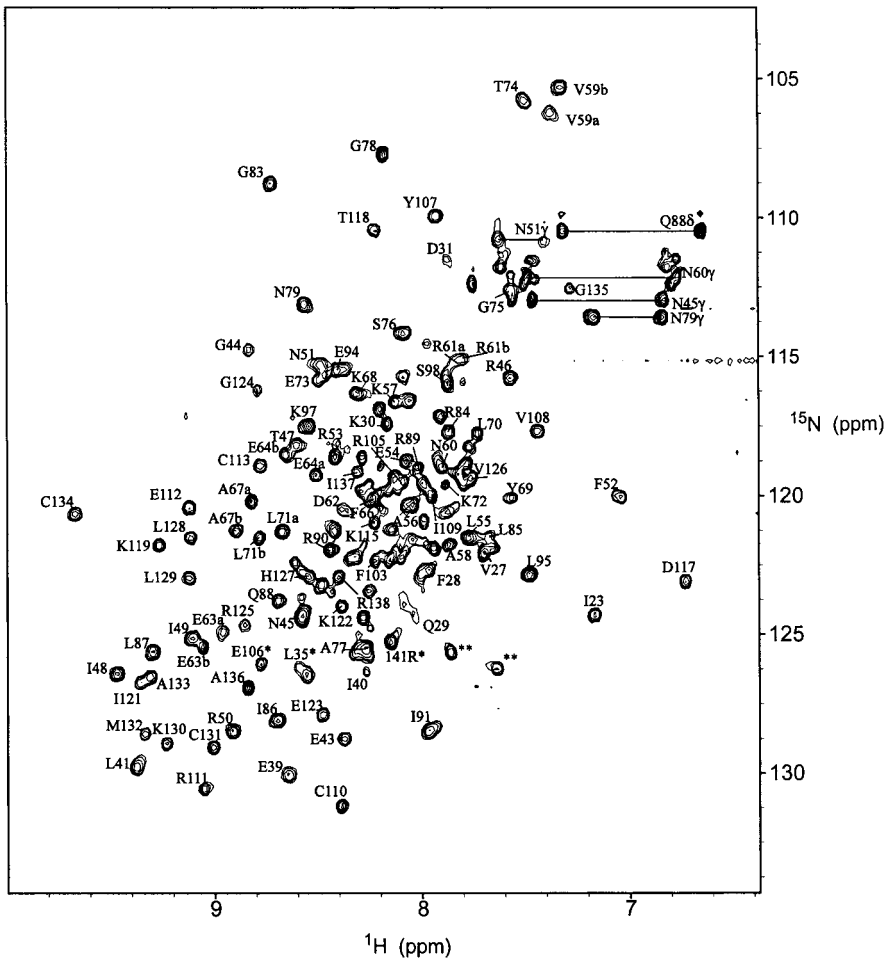

FIGURE 3: ^{15}N−^{1}H correlated HSQC spectrum of the β subunit of translation initiation factor 2 (aIF2β) from *M. jannaschii* obtained at 30 °C and pH 6.8. NMR assignments were obtained for 84% of the backbone resonances of the residues between positions 23 and 141 in the primary sequence. Note that amide resonances from some amino acids appear twice in the spectrum (for example, V59, E63, E64, and A67), a result of conformational or structural heterogeneity in the protein sample, as discussed in the text. Signals for which the assignments are tentative are marked with a single asterisk, and some well-resolved signals that could not be specifically assigned are marked with two asterisks. These resonances remained unassigned due to the lack of observable correlations in the triple-resonance spectra and/or the lack of observable NOEs, and degeneracy in the chemical shifts of C^{α} and C^{β}. Unassigned resonances can be attributed to residues in several (fairly small) gaps in the sequential assignments apparent in Figure 4, or perhaps to resonances in the degraded or partially degraded protein.

having reasonable molecular geometry, and have no NOE-derived distance constraint violations greater than 0.5 Å. Searches for similar structures within the Protein Data Bank were carried out using the Vector Alignment Search Tool (VAST), located at the National Center for Biotechnology Information (NCBI) web site, and the DALI search tool (*30*).

RESULTS

Sample Characterization. The purification strategy of using imidazole affinity chromatography (exploiting the N-terminal histidine tag) followed by cation exchange chromatography was found to be effective for separating the aIF2β from other cellular proteins. However, it was found that purified samples of the 143-residue aIF2β protein quickly degraded to yield two truncated versions of the protein, with migration rates on an SDS−PAGE gel corresponding to deletions of 20−30 residues. N-Terminal amino acid sequence analysis performed on two independently prepared samples confirmed that the two major components had sequences beginning at residues 21 and 31. MALDI mass spectrometry showed that the protein samples contained major components with molecular weights consistent with deletion of the first 20 and 30 residues, as well as several other minor components similar in mass. Interestingly, a sequence alignment of aIF2β and eIF2β sequences from a range of species (Figure 1) indicates a variable-length region just before residue 30, indicating the likely presence of a loop, thus providing an explanation for the high sensitivity of this region to proteolysis. Cleavage of the first 20−30 amino acids of aIF2β occurred within a few hours, even when thrombin was not used to remove the N-terminal histidine tag of the protein. The observation that residues 31−143 of aIF2β form a relatively stable protein fragment provides evidence that the first 30 residues of aIF2β are not essential for the structural integrity of the core regions of the protein. Although residues 31−143 make up a fairly stable protein fragment, it was found that even this fragment has a strong tendency to degrade at its C-terminus, since resonances associated with the C-terminal region of the protein decayed gradually over a period of days. This sample instability made it necessary to prepare a large number of samples to complete the multidimensional NMR data collection. The major compo-

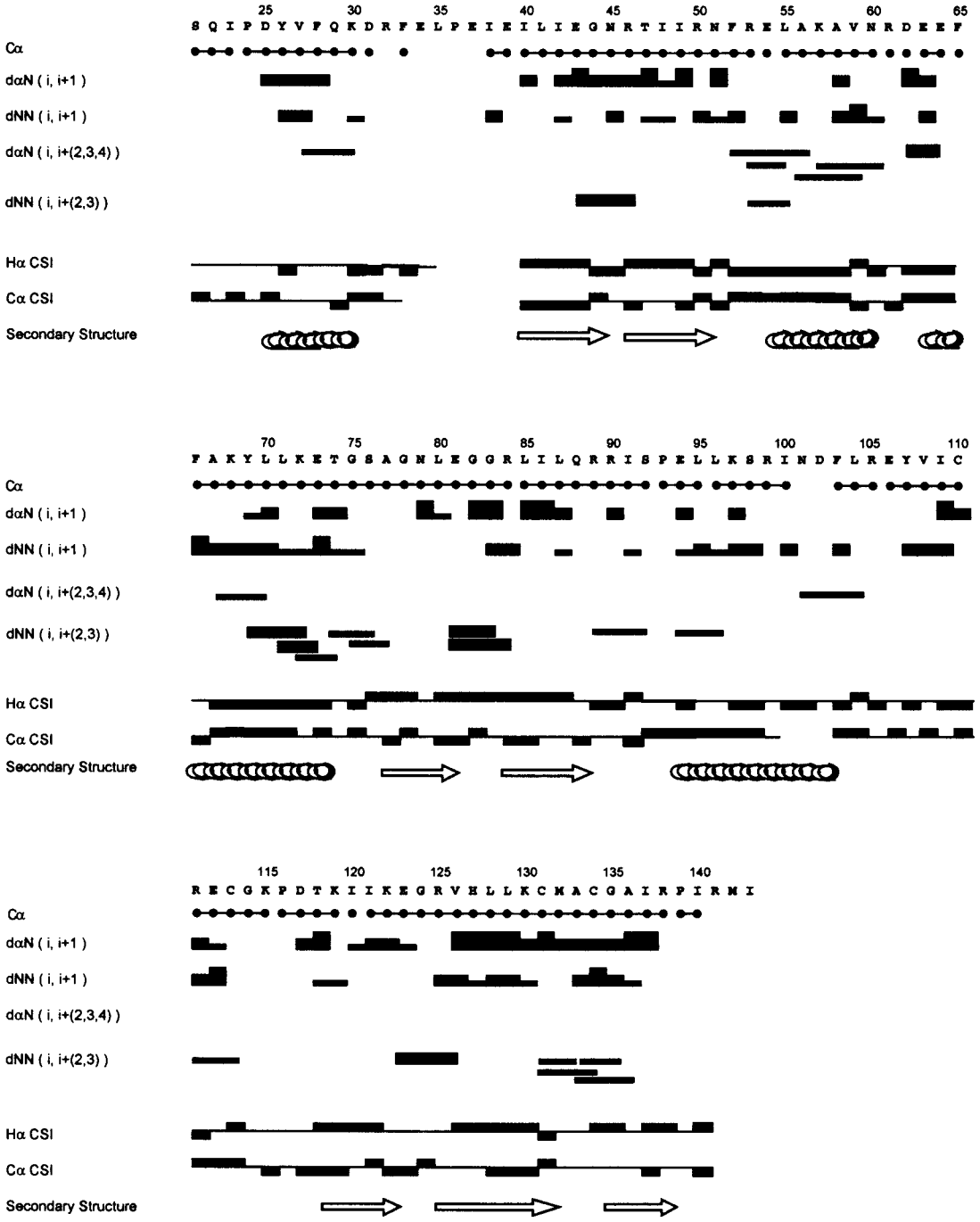

FIGURE 4: Schematic representation of NMR and secondary structure data. Small circles in the row labeled C^{α} indicate residues for which C^{α} chemical shifts have been assigned; the circles are linked where NMR assignments have been made by matching C^{α} and/or C^{β} chemical shifts obtained in triple-resonance spectra. Sequential and medium-range NOE connectivities involving the amide and α protons are shown using horizontal bars, where the thickness of the lines represents the strength of the NOE as strong, medium, or weak. Only unambiguous (nonoverlapping) NOE information is included. H^{α} and C^{α} chemical shift indices (CSI) correlate well with the identified secondary structural elements.

nent within freshly prepared samples used for NMR analysis consisted of residues 31−143 of aIF2β, indicated by mass spectrometry.

NMR Analysis of aIF2β. The structure determination of aIF2β represented a challenging task, for several reasons. The NMR spectrum assignment process was made more complicated by the presence of doubled peaks or shoulder peaks on some of the resonances. This doubling was observed in the chemical shifts of some of the amide protons and nitrogens (Figure 3), as well as the C^{o} and C^{α} shifts observed in the triple-resonance spectra, with the components of the doubled peaks differing in chemical shift by typically 0.2 ppm in each frequency dimension. During the assignment process, it became apparent that the doubled peaks belong exclusively to residues located in the N-terminal domain of the protein, with the doubling being most pronounced in resonances of Lys58, Val59, Asp62, Glu63, Glu64, Ala67, Leu71, and Lys72. The presence of doubled resonances can potentially be attributed to multiple conformations within this region of the protein structure, possibly as a result of interaction with the N-terminal region of the protein that has been shown to be heterogeneous by the mass spectrometry

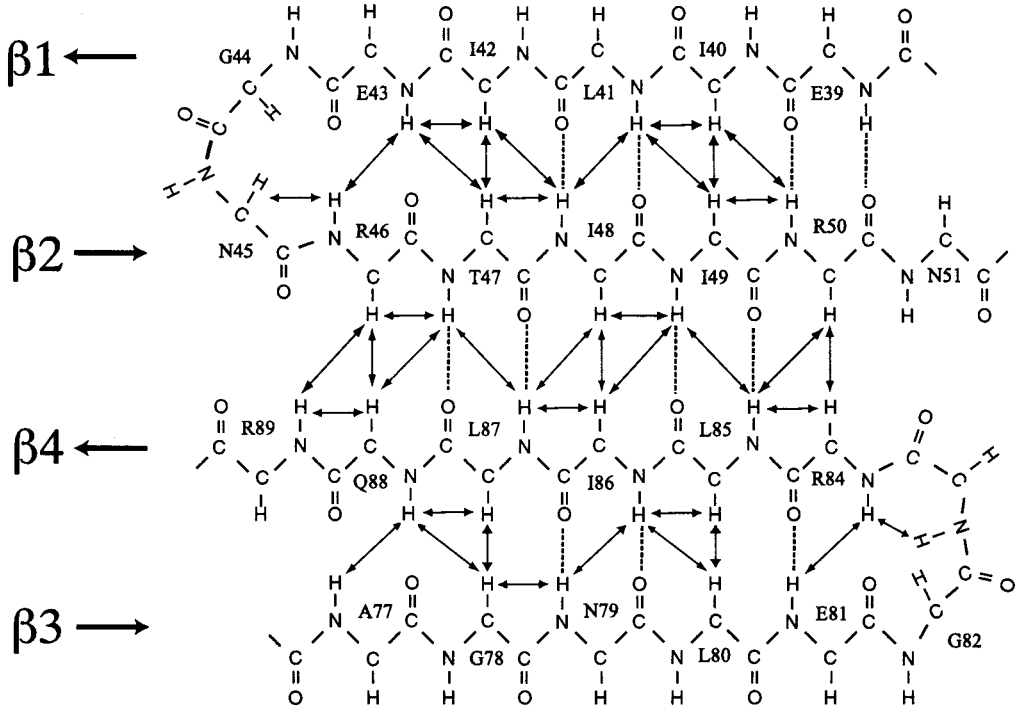

FIGURE 5: Diagram of the four-stranded β sheet structure within the N-terminal domain of aIF2β. Pairs of protons for which unambiguous NOEs are observed are connected by arrows, and interstrand hydrogen bonds are indicated by dotted lines.

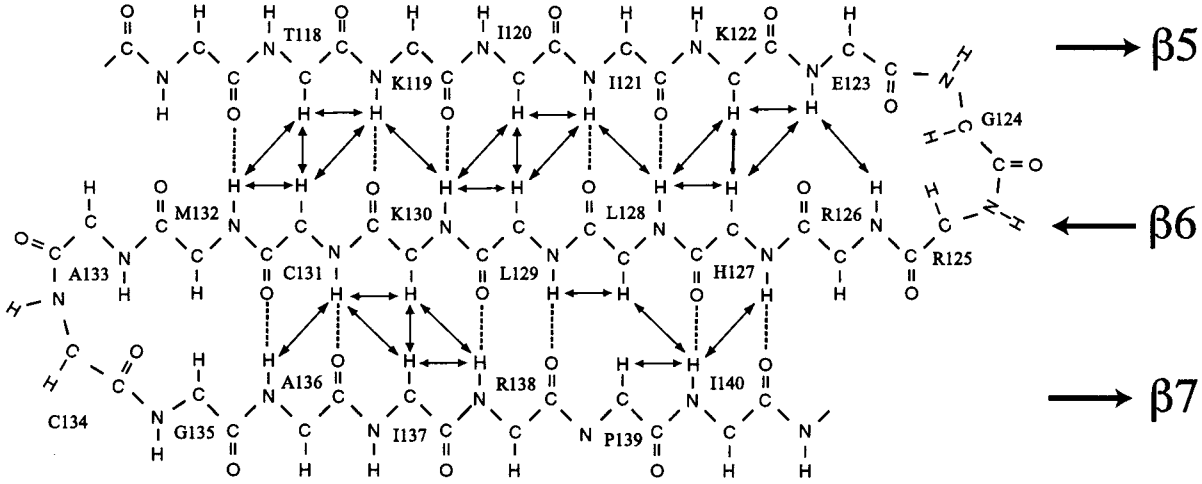

FIGURE 6: Diagram of the three-stranded β sheet structure within the C-terminal domain of aIF2β, as derived from the NMR data. Pairs of protons for which NOE cross-peaks are observed are connected by arrows, and interstrand hydrogen bonds are indicated by dotted lines.

and N-terminal sequencing data. An abundance of isoleucine (16) and leucine (16) residues made the NMR spectrum assignments difficult and ambiguous at several points; these ambiguities were eventually resolved through the use of triple-resonance NMR methods, applied to samples that were enriched with ^{13}C and/or ^{15}N at specific amino acid types. In addition, the core regions of the protein were found to be rich in leucine and isoleucine residues with protons and carbon nuclei resonating within a very narrow chemical shift range, making the assignment of NOE cross-peaks an especially challenging task. Although the large majority (84%) of the backbone resonances were unambiguously assigned by matching of C^{α}, C^{β}, and C^{o} chemical shifts in adjacent residues obtained in triple-resonance experiments (Figure 4), several amide signals that were detected in HSQC spectra remained unassigned (or only tentatively assigned) due to a lack of observable connectivity information. Consequently, the number of NOE-derived distance restraints is not as great as is ideally found in an NMR investigation of protein structure. Chemical shift assignments for aIF2β have been submitted to the BioMagResBank, and assigned accession number BMRB-5294.

Despite the difficulties described above, NMR data of sufficient quality for a structural analysis were obtained, and the distance restraints derived from the NOE data were sufficient to provide a useful and reasonably accurate model of the protein structure. The large fraction of antiparallel β sheet structure within aIF2β was fortuitous, since distinctive NOE cross-peak patterns permit regular antiparallel β sheets to be defined using relatively sparse NMR-derived restraints (Figures 5 and 6). aIF2β is the first member of the eIF2β/eIF5 family of proteins to be structurally characterized, by either NMR or X-ray crystallographic methods. The difficulties we experienced in working with aIF2β (specifically,

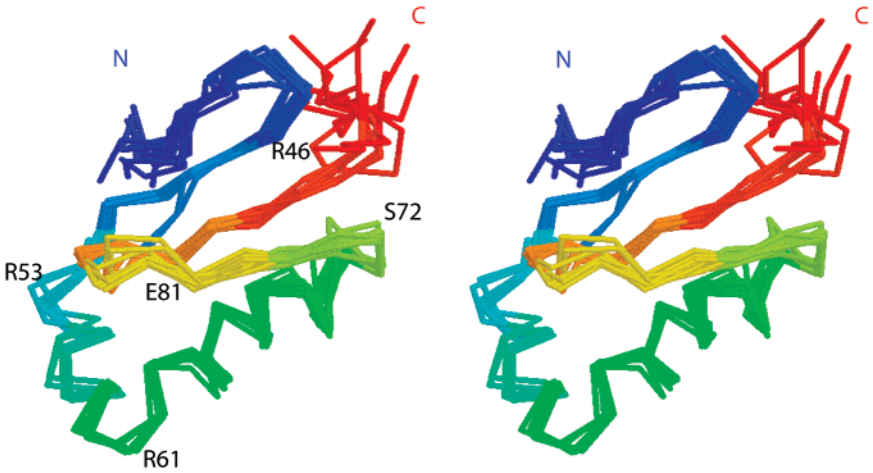

FIGURE 7: Superposition of the backbones of 10 low-energy structures of the N-terminal domain of aIF2β that are equally consistent with the NMR data, color-ramped from blue at residue 39 to red at residue 90. The 10 models are a fair representation of the full range of structural backbones that are consistent with the NMR data.

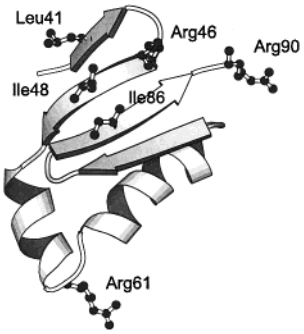

FIGURE 8: Ribbon diagram of the N-terminal domain of protein aIF2β, created using MOLSCRIPT (*57*), with the relative positions of some of the most conserved surface residues indicated. As discussed in the text, possible RNA-binding sites are located at conserved Arg61 at the bend at the junction of two helices, and also at conserved residues Arg46 and Arg90. Surface residues L41, I48, and I86 form a moderately well conserved hydrophobic surface for possible intersubunit contacts within the aIF2 heterotrimer. It should be noted that the rmsd for the side chain coordinates is approximately 2 Å; thus, the NMR data do not establish the precise conformations of the side chains, but do serve to indicate the relative positions of the residues that are on the protein surface.

stability, heterogeneity, and resonance overlap) may provide some explanation as to why no structures have been previously reported for any members of this important protein family.

Description of the N-Terminal Domain. The aIF2β protein was found to contain two distinct structural domains, each of which is internally well defined by interproton distance constraints derived from assigned NOE cross-peaks. Residues 39−90 form a single structural domain (the N-terminal domain), consisting of a β sheet of four strands arranged in an antiparallel manner (Figure 5), and two helices (α2 and α3) packed on the same side of the β sheet (Figures 7 and 8). The β sheet was defined by characteristic long-range and short-range NOE cross-peaks, with the hydrogen-bonded amide protons between the β-strands being protected from exchange with the solvent for a period of several days (Figure 5). Characteristic NOE cross-peak patterns show that strands β1 and β2 are connected by a type I turn (*28*), as are strands β3 and β4. Helices α2 and α3 link strands β2 and β3, and were defined by characteristic NOE cross-peaks between amide protons of sequential amino acids, as well as between α protons and amide and β protons three residues ahead in the sequence. Helix α2 is separated from helix α3 by a loop centered at Arg61, a residue which is conserved among all available sequences of aIF2β and eIF2β proteins. NOE cross-peaks between protons of conserved hydrophobic residues in the interior of the domain (such as Leu49, Ala67, and Leu85) served to orient the helices relative to the β sheet. Superposition of a set of low-energy structures along their backbones demonstrates a well-defined fold in the N-terminal domain (Figure 7); this set is a fair representation of a full range of structures that is consistent with the NMR data and reasonable molecular geometry. A Ramachandran plot for these structures indicates that all except 2.7% of the backbone dihedral angles are within the most favored, allowed, or generously allowed regions according to the program Procheck-NMR (*29*). Structural statistics for the N-terminal domain of aIF2β are summarized in Table 1.

Backbone resonances for residues 21−30 show relatively weak NMR signals, which we attribute to a major component of the protein sample containing only residues 31−143. The negative deviations of the H^{α} chemical shifts and positive deviations of C^{α} chemical shifts from random coil values (*30*) indicate a helical conformation for residues 25−30, as do NOE peaks between sequential amide protons and some (i, $i + 3$) NOE peaks. Chemical shift index data indicate that this helix may extend to residue 33. However, no long-range NOE peaks were observed linking this short helix to other regions of the protein structure. The triple-resonance NMR spectra did not contain signals that could be unambiguously assigned to residues 32 and 34−37; amide proton signals from these residues may be unobserved due to resonance overlap, rapid exchange with the solvent, and/or flexibility.

The coordinates for structures of the core of the N-terminal domain (residues 39−90) were compared against a database of known structures using the VAST and DALI (*31*) search tools. The N-terminal domain of aIF2β is significantly structurally homologous to the DNA binding domain of yeast heat shock transcription factor (HSF) (*32*) and a domain within ribosomal protein S4 (*33*), despite the low level of sequence homology between these proteins (Figure 9). The structures share the same αββαααββ topology, and the rmsds

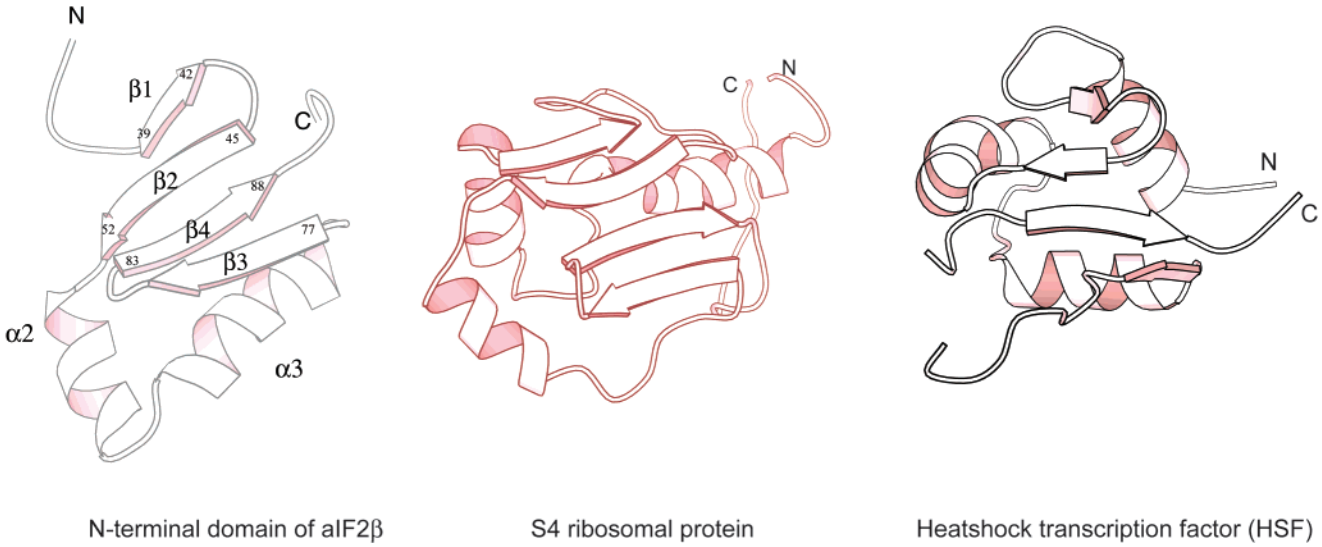

FIGURE 9: Ribbon diagrams comparing the structure of the N-terminal domain of translation initiation factor aIF2β with heat shock transcription factor (HSF) and ribosomal protein S4. The ribbon diagrams were created using coordinates from PDB entry 2HTS for HSF (*32*) and PDB entry 1C05 for ribosomal protein S4 (*33*). The coordinates for aIF2β have been deposited in the Protein Data Bank as entry 1K8B. The rmsd along the C^{α} backbone of the N-terminal domain of aIF2β and S4 is 2.8 Å; the rmsd between aIF2β and the heat shock transcription factor is 3.1 Å.

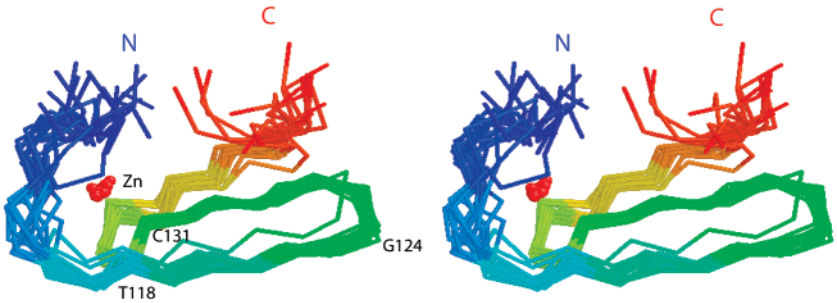

FIGURE 10: Superposition of the backbones of 16 low-energy structures of the C-terminal domain of aIF2β that are equally consistent with the NMR data, color-ramped from blue at residue 108 to red at residue 143. The position of the zinc atom in each structural model is indicated by a red sphere. The 16 models are a fair representation of the full range of structures that are consistent with the NMR-derived constraints.

of the backbone atoms between the N-terminal domain of aIF2β and the two homologues are 3.1 (HSP) and 2.8 Å (S4), respectively. Although there is not any known functional or evolutionary relationship among aIF2β, HSF, and ribosomal protein S4, it is interesting that each of these proteins is associated with nucleic acids. Coordinates for the N-terminal domain of aIF2β have been deposited in the Protein Data Bank as entry 1K8B.

C-Terminal Domain. The C-terminal domain of aIF2β, corresponding to residues 108−143, folds into a "zinc ribbon" structure that is strikingly similar to the structure of ribosomal protein L36 (*34*). The most prominent feature of the domain is a three-stranded antiparallel β sheet (Figure 6). The four conserved cysteine residues (Cys110, Cys113, Cys131, and Cys134) occur in two CXXC clusters located at the turn between strands β6 and β7 and in the loop just prior to strand β5. The four cysteines are clearly positioned close together in the three-dimensional structure (Figures 10 and 11), with the side chains of the four cysteines located on the same side of the sheet in such a way that they can bind the zinc atom (the presence of zinc in the structure was confirmed by atomic absorption spectroscopy, as well as experiments involving a synthetic peptide model of the C-terminal domain described below). The NOE cross-peak pattern characteristic of a metal-coordinating "rubredoxin" turn (*35*−*37*) was observed for residues Cys131−Ala136. NOE data indicate that the loop consisting of residues 108−114 at the beginning of the C-terminal domain (including Cys110 and Cys113) was in neither a regular α helical nor a regular β strand conformation. Conserved hydrophobic residues Ile120 and Leu129 form a small hydrophobic cluster on one side of the sheet and are exposed on the surface of the C-terminal domain. There is a possibility that this hydrophobic cluster may interact with the conserved residues found in the linker region; however, no long-range NOE peaks could be identified.

In addition to its similarity to ribosomal protein L36, the C-terminal domain of aIF2β bears a significant resemblance to the zinc ribbon folds found in the N-terminal domain of *Pyrococcus furiosus* transcription factor TFB (*38*) and the C-terminal domain of eukaryal transcription elongation factor hTFIIS (*39*, *40*) (Figure 11). In each case, the structure contains a three-stranded antiparallel β sheet, with a zinc ion coordination site at one end of the sheet.

The Two Domains Are Structurally Independent and Connected by a Helical Linker. A 15-residue linker, consisting of residues 93−107, connects the N- and C-terminal domains of aIF2β. Although the NMR-derived structural information is fairly sparse for the linker region, due to chemical shift overlap, the sum of the available information

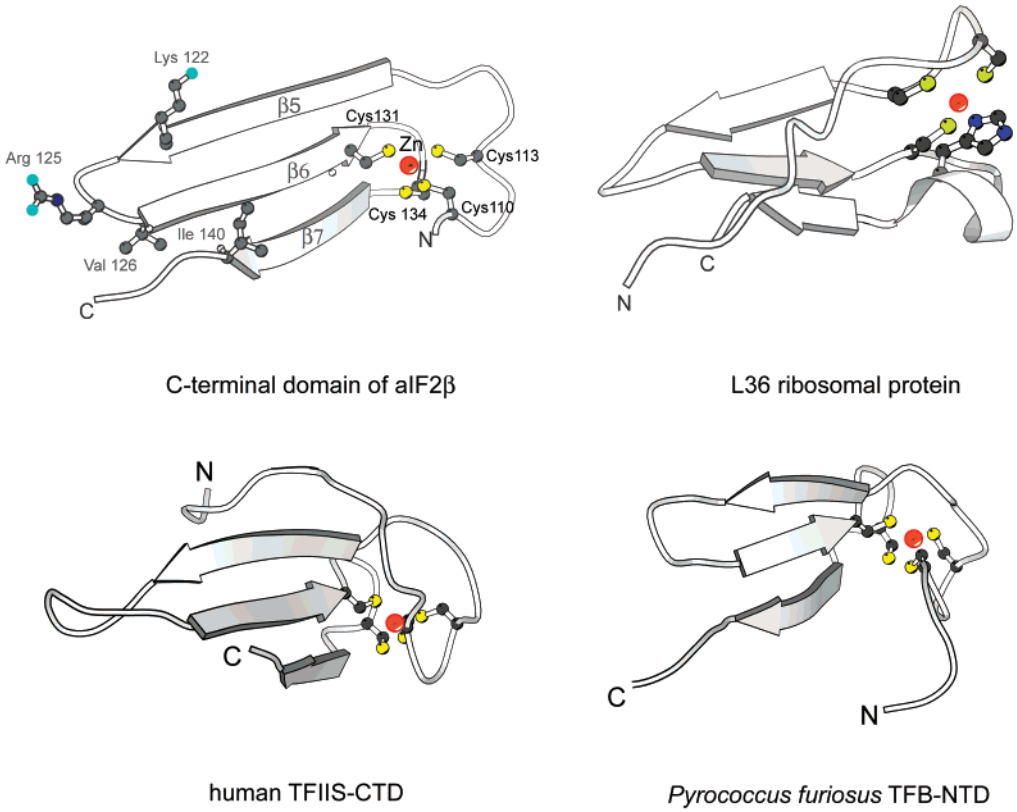

FIGURE 11: Ribbon diagrams comparing the structure of the C-terminal domain of translation initiation factor aIF2β with ribosomal protein L36, human transcription factor TFIIS-CTD, and TFB-NTD from *P. furiosus*, created using coordinates from PDB entries 1DFE, 1TFI, and 1PFT, respectively. The sulfur atoms in the four cysteines of the zinc-binding motif are indicated by yellow spheres; the zinc atom in each structure is indicated by a red sphere. The positions of side chains of Lys122, Arg125, Val126, and Ile140 in aIF2β from *M. jannaschii* are shown; these residues are located at positions that are homologous to residues in eIF2β in yeast that have been implicated in initiation site selection, as discussed in the text. It should be noted that the rmsd for the Lys122, Arg125, Val126, and Ile140 side chain coordinates is approximately 2−3 Å; thus, the NMR data do not establish the precise conformations of the side chains, but do serve to indicate their relative positions on the protein surface, and their positions relative to the Cys4 motif.

indicates that the linker is helical. For example, NOE cross-peaks between sequential amide protons typical of helical structure were observed for most residues in the linker region, although it is noted that we were unable to specifically assign amide protons for residues 101 and 102, and four of the α carbons in the linker, due to resonance overlap. Relatively high C^{α} chemical shifts and low H^{α} chemical shifts relative to random coil values were observed for the majority of the linker residues (Figure 4), consistent with an α helical structure (*30*). C^{o} chemical shifts were assigned for 10 of the 15 residues in the linker, and in all cases were found to have values between 177.2 and 178.6 ppm, also typical of a helical structure (*30*). No long-range NOEs were observed between the protons of the linker and either the N-terminal or C-terminal domain. Furthermore, no long-range NOEs corresponding to interdomain interactions were observed. None of the amide protons of the linker region were observed in the D_2O solvent, indicating that they are not particularly well protected from exchange. The absence of NMR-derived structural constraints linking the N- and C-terminal domains of aIF2β caused us to speculate that the two domains are structurally independent.

The hypothesis that the N- and C-terminal domains are structurally independent was tested using a peptide synthesized (Biosyn) with a sequence corresponding to the last 50 residues of the aIF2β protein (residues 94−143). In the absence of zinc, and in the presence of DTT, most of the proton resonances of the 50-residue peptide have chemical

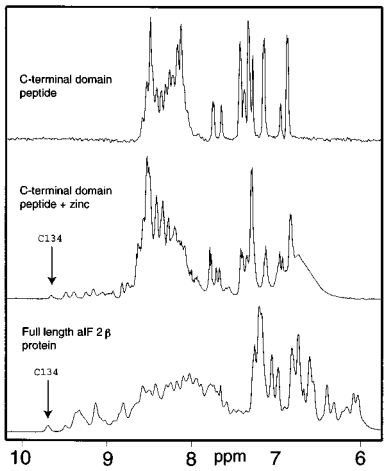

FIGURE 12: Comparison of the one-dimensional proton NMR spectra of a 50-residue synthetic peptide with a sequence corresponding to residues 94−143 of aIF2β from *M. jannaschii* (top), the spectrum of the same peptide after the addition of a millimolar amount of zinc (middle), and the spectrum of the full-length aIF2β protein (bottom). Each spectrum was obtained at 30 °C, with a 10 mM acetate buffer at pH 6. Note that the addition of zinc to the 50-residue peptide results in the appearance of downfield-shifted resonances with chemical shifts that closely resemble those of the full-length protein.

shifts near the values typical of a random coil (Figure 12), indicating an unfolded or perhaps partially folded structure. A two-dimensional NOE spectrum exhibited only weak and

predominantly intraresidue cross-peaks, providing further evidence that the peptide structure was not well ordered.

Upon addition of $ZnCl_2$ to the peptide sample, a new set of dispersed NMR peaks appeared, with α protons with chemical shifts of >4.8 ppm and amide proton shifts of >8.8 ppm (Figure 12), indicating a transition to a folded structure. Proton resonance assignments for the peptide were obtained from two-dimensional homonuclear NOESY and TOCSY spectra. Chemical shifts in the peptide showed a strong similarity to those of the corresponding residues in the full-length aIF2β protein, in most cases differing by a few hundredths of a part per million or less, although a few differences of ~0.2 ppm were observed. The most downfield proton resonance in the peptide is the amide proton of Cys134 with a chemical shift near 9.6 ppm (Figure 12); this corresponds to the most downfield resonance in the full-length protein, which is also Cys134, with a similar chemical shift near 9.7 ppm (Figure 12). NOE cross-peak patterns observed in the two-dimensional spectra of the peptide were characteristic of a rubredoxin turn for residues Cys131−Ala136, strikingly similar to that observed within the full-length protein, and confirmed the presence of the three-strand antiparallel β sheet. NOE data showed that the local geometry around the four cysteine residues in the 50-residue peptide was not measurably different from that found in the full-length protein, despite the slight differences in some of the chemical shift values. In summary, the NMR spectra of the peptide provide strong evidence that the C-terminal domain of aIF2β (1) is a structurally independent unit and (2) contains the zinc binding site. The data also provide evidence that the binding of the zinc atom is important for the folding of the C-terminal domain.

An additional (and unintentional) experiment provided evidence that the N-terminal domain of aIF2β also forms an independent structural unit. A sample of what was intended to be full-length ^{15}N-labeled aIF2β protein was found to have degraded so that no resonances of the C-terminal domain were observed in the NMR spectra, while the familiar resonances and NOE cross-peaks of the N-terminal domain and the part of the helical linker nearest the N-terminal domain were observed at their usual chemical shift values. These observations provided further evidence in support of a model where the N- and C-terminal domains of aIF2β do not strongly interact with each other, and behave as two independently folded structural units connected by a linker. Structural models of the C-terminal domain of aIF2β have been deposited in the Protein Data Bank as entry 1K81.

NMR spectroscopy was used to address the question of whether the linker connecting the N- and C-terminal domains is ordered or relatively flexible. NMR-observable parameters related to molecular motion in proteins include the ^{15}N−^{1}H heteronuclear NOE and the imino ^{15}N relaxation rates R_1 and R_2 (reciprocals of the longitudinal and transverse relaxation times T_1 and T_2, respectively); measurements of these three relaxation parameters were obtained using a sample of ^{15}N-enriched aIF2β. Although the quantitative analysis of the ^{15}N relaxation data was hampered by the tendency of the protein samples to gradually degrade during the data collection, data were obtained that were sufficient for drawing at least qualitative conclusions regarding the degree of ordering in the linker between the two protein domains. Residues E94, L95, K97, S98, F103, and V108 within the linker region exhibit strongly positive ^{15}N−^{1}H heteronuclear NOEs, with values near 0.7, and R_1 and R_2 relaxation rates near 1.9 and 9 s^{-1}, respectively; these values are typical of ordered structure and are similar to relaxation data obtained for residues within β sheet regions of the protein. We note that ^{15}N relaxation data could not be obtained for other linker residues, due to either the resonances being in overlapped regions of the two-dimensional ^{15}N−^{1}H correlated spectra or, in a few cases, the assignments being ambiguous. In summary, we find that the NMR data provide evidence that (at least) a large fraction of the linker region is well ordered, with the degree of ordering being uncertain in two short segments containing residues 99−102 and 104−107, for which relaxation data were not obtained.

Although our results provide strong evidence that the N- and C-terminal domains of aIF2β fold independently, it is likely that the linker region does play an important role in maintaining the relative orientations of the two domains. Evidence that supports this hypothesis can be found in the alignment of primary sequences (Figure 1), which shows that the length of the linker is conserved. In addition, there are several conserved hydrophobic residues within the linker region, such as L96, I100, F103, and L104. These conserved residues may make essential contacts within the context of the complete heterotrimeric aIF2 protein; these contacts may be with either of the two domains of the β subunit, or perhaps with the α or γ subunits.

DISCUSSION

The sequences of archaeal aIF2β and eukaryotic eIF2β are significantly homologous, to the extent that their structures are certain to contain similar features (*41*). Specifically, the sequence alignment in Figure 1 shows that the C_2−C_2 zinc ribbon domain within the C-terminus of the protein is universally conserved among the eukaryotes and archaea, and eukaryotic eIF2β contains a central domain that is homologous to the N-terminal domain of archaeal aIF2β, although the archaeal versions of the protein lack the variable-length N-terminal region that is present in the eukaryotes and contains the lysine boxes (Figures 1 and 2). The homology between eIF2β and aIF2β implies that structural information gained in the work presented here can be used to improve our understanding of eukaryotic eIF2β, and provides a framework for interpreting some of the earlier biochemical data.

The eIF2β subunit of the heterotrimeric eIF2 complex has been shown to be involved in interactions with several other proteins in the assembly of the 43S preinitiation complex. Interactions of eIF2β with the eIF2γ subunit (*18*), subunits of eIF2B (*17*, *42*), and the C-terminal domain of eIF5 (*16*, *17*) have been reported, and it is plausible that interactions with other initiation factors, rRNA, and the 40S ribosome may also play roles in the initiation process.

While we were working on the structure of aIF2β, Thompson et al. (*18*) reported the identification of three separate functional segments in yeast eIF2β: an N-terminal fragment responsible for binding to the C-terminal domain of eIF5 and subunits of eIF2B, a central fragment for binding to the eIF2γ subunit, and a C-terminal fragment for initiation site recognition. The observation of a three-domain functionality in the eukaryotic protein is consistent with our two-

domain structure of archaeal aIF2β derived from NMR data. The central fragment and the C-terminal fragment of yeast eIF2β correspond to the N-terminal and C-terminal domains in aIF2β of *M. jannaschii*, respectively. The lack of the eukaryotic N-terminal fragment bearing the lysine tracts in aIF2β (Figure 2) is consistent with the observation that archaea do not appear to contain either eIF5 or eIF2B within their genome (*43*−*45*).

Within the N-terminal domain of aIF2β (amino acids 39−90), conserved residues Gly44, Arg46, Thr47, Asn51, and Arg90 map to a relatively small surface among the secondary structural elements (Figure 8), suggesting a functional role. The positions of positively charged and conserved residues Arg46 and Arg90, which are adjacent in the structure, suggest that their side chains may be important for electrostatic interactions with a ligand, possibly the negatively charged phosphate backbone of a nucleic acid. Well-conserved hydrophobic residues form hydrophobic patches on the two sides of the β sheet in the N-terminal domain. Leu49, Leu85, and Leu87 contribute to the hydrophobic core of the domain on the side of the β sheet closest to helices α2 and α3, while another group of conserved residues (Leu41, Ile48, and Ile86) forms a small hydrophobic patch on the surface of the β sheet opposite helices α2 and α3. A recent study shows that yeast eIF2β containing only the conserved regions that correspond to the full length of aIF2β interacts with the γ subunit of eIF2 to the same extent as the full-length wild-type protein (*18*), and the minimum fragment of eIF2β required for binding eIF2γ in yeast shown in their study corresponds to residues 2−81 in the *M. jannaschii* sequence (*18*). The solvent accessible hydrophobic residues Ile41, Thr48, and Leu86 within the N-terminal domain of aIF2β, and their homologues in eukaryotes, are thus likely candidates for involvement in the interaction with the γ subunit.

Initiation site recognition by the translational apparatus in eukaryotes involves primarily the base pairing interaction between the anticodon of $tRNA_i$-Met and the AUG codon (*46*). In addition to this key interaction, genetic mutational analyses have identified several additional components in yeast that are important determinants for correct start site recognition. Mutations in eIF1 (*47*, *48*), all three subunits of eIF2 (*49*−*53*), and eIF5 (*53*) allowed translation initiation at non-AUG codons. In eIF2β from yeast, 10 independent mutations involving six residues in or adjacent to the conserved C_2−C_2 motif allowed initiation at non-AUG codons. Among these six residues, four residues are identical or conserved in all available eukaryotic eIF2β and archaeal aIF2β sequences. On the other hand, none of the six residues are conserved within the eIF5 family. When the corresponding residues are mapped onto the structure of aIF2β, the four conserved residues Lys122, Arg125, Val126, and Ile140 are found close together at the surface of one end of the β sheet in the C-terminal domain, as shown in Figure 11. The detailed mechanism through which eIF2 participates in selecting the correct AUG codon is unknown, and most archaea genes have a prokaryote-like Shine-Dalgarno sequence for selecting the starting codon. However, the mutational and structural data that are now available suggest that Lys122, Arg125, Val126, and Ile140 may serve as a binding site for other conserved translational components involved in ribosomal initiation site recognition common in eukaryotes and archaea.

On the basis of primary sequence similarity (Figure 1), the N-terminal domain of eIF5 is expected to share a common fold with the aIF2β and eIF2β proteins; these proteins have been recognized as being members of a single structural family commonly termed eIF2β/eIF5. The structural features shared by eIF2β and eIF5 do not necessarily imply functional similarity, since it has been shown in several cases that one protein fold can contain a wide range of functional diversity (*54*). For example, the zinc ribbon fold has been found in many proteins with diverse functions (*55*, *56*). The identities of conserved surface residues and residues identified as being most likely to be critical for initiation site recognition in eIF5 (*53*) differ from those shared with eIF2β and aIF2β (Figure 1). Differences between eIF5 and the eIF2β/aIF2β proteins, such as the presence of a C-terminal domain in eIF5 that is not present in eIF2β/aIF2β, and differences in conserved residues within the domains that are homologous are likely to contribute to the ability of each protein to recognize its own specific target, and the control of GTPase activating activity attributed to eIF5.

Although the general roles of eIF2β and eIF5 within the translation process have become increasingly well established by recent biochemical studies, the detailed mechanisms by which these proteins participate in the recognition and recruitment of initiator tRNA, how phosphorylation of eIF2α can affect the binding of eIF2B to eIF2β, how the GTPase activity of eIF5 is triggered by the initiation complex, and what interactions bring the 40S ribosome to the eIF2−GTP−$tRNA_i$ ternary complex are unknown. The structural insights provided by the results of this work serve as a starting point for further elucidation of the mechanisms by which translation is initiated and regulated.

REFERENCES

1. Pestova, T. V., and Hellen, C. U. (2000) *Cell. Mol. Life Sci. 57*, 651−674.
2. Hershey, J. W. B., and Merrick, W. C. (2000) in *Translational Control of Gene Expression*, pp 33−88, Cold Spring Harbor Laboratory Press, Plainview, NY.
3. Pestova, T. V., Kolupaeva, V. G., Lomakin, I. B., Pilipenkom, E. V., Shatskym, I. N., Agol, V. I., and Hellen, C. U. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 7029−7036.
4. Anthony, D. D., Kinzy, T. G., and Merrick, W. C. (1990) *Arch. Biochem. Biophys. 281*, 157−162.
5. Bommer, U. A., Kraft, R., Kurzchalia, T. V., Price, N. T., and Proud, C. G. (1991) *Biochim. Biophys. Acta 1079*, 308−315.
6. Gaspar, N. J., Kinzy, T. G., Scherer, B. J., Humbelin, M., Hershey, J. W., and Merrick, W. C. (1994) *J. Biol. Chem. 269*, 3415−3422.
7. Clemens, M. J. (1996) in *Translational Control*, pp 139−172, Cold Spring Harbor Laboratory Press, Plainview, NY.
8. Hinnebusch, A. G. (2000) in *Translational Control of Gene Expression*, pp 185−244, Cold Spring Harbor Laboratory Press, Plainview, NY.
9. Gribskov, M. (1992) *Gene 119*, 107−111.
10. Bycroft, M., Hubbard, T. J., Procor, M., Freund, S. M. V., and Murzin, A. G. (1997) *Cell 88*, 235−242.
11. Erickson, F. L., Harding, L. D., Dorris, D. R., and Hannig, E. M. (1997) *Mol. Gen. Genet. 253*, 711−719.
12. Hannig, E. M., Cigan, A. M., Freeman, B. A., and Kinzy, T. G. (1994) *Mol. Cell. Biol. 13*, 506−520.
13. Ehrmann, I. E., Ellis, P. S., Mazeyrat, S., Duthie, S., Brockdorff, N., Mattei, M. G., Gavin, M. A., Affara, N. A., Brown, G. M., Simpson, E., Mitchell, M. J., and Scott, D. M. (1998) *Hum. Mol. Genet. 7*, 1725−1737.

14. Kjeldgaard, M., and Nyborg, J. (1992) *J. Mol. Biol. 223*, 721−742.
15. Koonin, E. V. (1995) *Protein Sci. 8*, 1608−1617.
16. Das, S., Maiti, T., Das, K., and Maitra, U. (1997) *J. Biol. Chem. 272*, 31712−31718.
17. Asano, K., Krishnamoorthy, T., Phan, L., Pavitt, G. D., and Hinnebusch, A. G. (1999) *EMBO J. 18*, 1673−1688.
18. Thompson, G. M., Pacheco, E., Melo, E. O., and Castilho, B. A. (2000) *Biochem. J. 347*, 703−709.
19. Muhandiram, D. R., and Kay, L. E. (1994) *J. Magn. Reson., Ser. B 103*, 203−216.
20. Grzesiek, S., and Bax, A. (1992) *J. Magn. Reson. 96*, 432−440.
21. Grzesiek, S., and Bax, A. (1993) *J. Magn. Reson., Ser. B 102*, 103−106.
22. Kay, L. E. (1993) *J. Am. Chem. Soc. 115*, 2055−2057.
23. Kay, L. E., Xu, G. Y., Singer, A. U., Muhandiram, D. R., and Forman-Kay, J. D. (1993) *J. Magn. Reson., Ser. B 101*, 333−337.
24. Pascal, S. M., Muhandiram, D. R., Yamazaki, T., Forman-Kay, J. D., and Kay, L. E. (1994) *J. Magn. Reson., Ser. B 103*, 197−201.
25. Wishart, D. S., Bigam, C. G., Yao, J., Abildgaard, F., Dyson, H. J., Oldfield, E., Markley, J. L., and Sykes, B. D. (1995) *J. Biomol. NMR 6*, 135−140.
26. Brünger, A. T., Adams, P. D., Clore, G. M., Delano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., and Warren, G. L. (1998) *Acta Crystallogr. D54*, 905−921.
27. Wishart, D. S., and Sykes, B. D. (1994) *Methods Enzymol. 239*, 363−392.
28. Wüthrich, K. (1986) in *NMR of Proteins and Nucleic Acids*, John Wiley, New York.
29. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thorton, J. M. (1996) *J. Biomol. NMR 8*, 477−486.
30. Wishart, D. S., Bigam, C. G., Holm, A., Hodges, R. S., and Sykes, B. D. (1995) *J. Biomol. NMR 5*, 67−81.
31. Holm, L., and Sander, C. (1993) *J. Mol. Biol. 233*, 123−138.
32. Harrison, C. J., Bohm, A. A., and Nelson, H. C. (1994) *Science 263*, 224−227.
33. Markus, M. A., Gerstner, R. B., Draper, D. E., and Torchia, D. A. (1998) *EMBO J. 17*, 4559−4571.
34. Härd, T., Rak, A., Allard, P., Kloo, L., and Garber, M. (2000) *J. Mol. Biol. 296*, 169−180.
35. Summers, M. F., South, T. L., Kim, B., and Hare, D. R. (1990) *Biochemistry 29*, 329−340.
36. Fourmy, D., Dardel, F., and Balnquet, S. (1993) *J. Mol. Biol. 231*, 1078−1089.
37. Perez-Alvarado, G. C., Miles, C., Michelsen, J. W., Louis, H. A., Winge, D. R., Beckerle, M. C., and Summers, M. F. (1994) *Nat. Struct. Biol. 1*, 388−398.
38. Zhu, W., Zeng, Q., Colangelo, C. M., Lewis, M., Summers, M. F., and Scott, R. A. (1996) *Nat. Struct. Biol. 3*, 122−124.
39. Qian, X., Gozani, S. N., Yoon, H., Jeon, C. J., Agarwal, K., and Weiss, M. A. (1993) *Biochemistry 32*, 9944−9959.
40. Qian, X., Jeon, C., Yoon, H., Agarwal, K., and Weiss, M. A. (1993) *Nature 365*, 277−279.
41. Sander, C., and Schneider, R. (1991) *Proteins 9*, 56−68.
42. Kimball, S. R., Heinzinger, N. K., Horetsky, R. L., and Jefferson, L. S. (1998) *J. Biol. Chem. 273*, 3039−3044.
43. Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., Blake, J. A., FitzGerald, L. M., Clayton, R. A., Gocayne, J. D., Kerlavage, A. R., Dougherty, B. A., Tomb, J. F., Adams, M. D., Reich, C. I., Overbeek, R., Kirkness, E. F., Weinstock, K. G., Merrick, J. M., Glodek, A., Scott, J. L., Geoghagen, N. S., and Venter, J. C. (1996) *Science 273*, 1058−1073.
44. Klenk, H.-P., Clayton, R. A., Tomb, J. F., White, O., Nelson, K. E., Ketchum, K. A., et al. (1997) The complete genome sequence of the hyperthermophilic sulphate-reducing archaeon *Archaeoglobus fulgidus*, *Nature 390*, 364−370.
45. Smith, D. R., Doucette-Stamm, L. A., Deloughery, C., Lee, H., Dubois, J., Aldredge, T., Bashirzadeh, R., Blakely, D., Cook, R., et al. (1997) Complete genome sequence of the *Methanobacterium thermoautotrophicum* delataH: functional analysis and comparative genomics. *J. Bacteriol. 179*, 7135−7155.
46. Cigan, A. M., Feng, L., and Donahue, T. F. (1988) *Science 242*, 93−97.
47. Cui, Y., Dinman, J. D., Kinzy, T. G., and Peltz, S. W. (1998) *Mol. Cell. Biol. 18*, 1506−1516.
48. Yoon, H. J., and Donahue, T. F. (1992) *Mol. Cell. Biol. 12*, 248−260.
49. Castilho-Valavicius, B., Thompson, G. M., and Donahue, T. F. (1992) *Gene Expression 2*, 297−309.
50. Cigan, A. M., Pabich, E. K., Feng, L., and Donahue, T. F. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2784−2788.
51. Dorris, D. R., Erickson, F. L., and Hannig, E. M. (1995) *EMBO J. 14*, 2239−2249.
52. Erickson, F. L., and Hannig, E. M. (1996) *EMBO J. 15*, 6311−6320.
53. Huang, H. K., Yoon, H., Hannig, E. M., and Donahue, T. F. (1997) *Genes Dev. 11*, 2396−2413.
54. Orengo, C. A., Jones, D. T., and Thornton, J. M. (1994) *Nature 372*, 631−634.
55. Wang, B., Jones, D. N., Kaine, B. P., and Weiss, M. A. (1998) *Structure 6*, 555−569.
56. Laity, J. H., Lee, B. M., and Wright, P. E. (2001) *Curr. Opin. Struct. Biol. 11*, 39−46.
57. Kraulis, P. J. (1991) *J. Mol. Graphics 14*, 51−55.

BI011984N