Biochemistry **2003,** *42,* 8966−8975

Solution Structure and RNA Interactions of the RNA Recognition Motif from Eukaryotic Translation Initiation Factor 4B[†]

Keiran Fleming,[‡] Jamie Ghuman,[§] Xuemei Yuan,[‡] Peter Simpson,[‡] Andrea Szendröi,[§] Stephen Matthews,*[,‡] and Stephen Curry*[,§]

Department of Biological Sciences and Centre for Structural Biology, Imperial College London, Wolfson Laboratory, Exhibition Road, London SW7 2AY, U.K., and Blackett Laboratory, Prince Consort Road, London SW7 2BW, U.K.

Received March 31, 2003; Revised Manuscript Received May 29, 2003

ABSTRACT: Eukaryotic initiation factor 4B (eIF4B) is a multidomain protein with a range of activities that serves primarily to promote association of messenger RNA to the 40S ribosomal subunit during translation initiation. We report here the solution structure of the eIF4B RNA recognition motif (RRM) domain. It adopts a classical RRM fold, with a $\beta\alpha\beta\beta\alpha\beta$ topology. The most striking difference with other RRM structures is in the disposition of loop 3, which connects the β2 and β3 strands and is implicated in RNA recognition. This loop folds down against the body of the RRM and exhibits restricted motion on a milli- to microsecond time scale. Although it contributes to a large basic patch on the RNA binding surface, it does not protrude out from the domain as observed in other RRM structures, possibly implying a different mode of RNA binding. On its own, the core RRM domain provides only a relative weak interaction with RNA targets and appears to require extensions at the N- and C-terminus for high-affinity binding.

Human eukaryotic initiation factor 4B (eIF4B) is a 611 amino acid RNA binding protein of 69 kDa (*1*) which promotes the binding of 40S ribosomal subunits to mRNA (*2*, *3*) and stimulates the RNA helicase activities of eIF4A and eIF4F (*4*, *5*). Deletion and site-directed mutagenesis studies have identified a number of functional domains within eIF4B, two of which are involved in RNA binding and are implicated in linking mRNA to the 40S ribosomal subunit during translation initiation. An N-terminal RNA recognition motif (RRM; residues 97−175) has been shown to bind the 18S rRNA of the 40S ribosomal subunit in UV−cross-linking and competition binding experiments (*6*). A second RNA binding domain is located toward the C-terminus (residues 367−423) and has been termed the basic domain (BD) since it contains two arginine-rich motifs (ARMs) (*7*, *8*). This region, which has not been assigned to a particular structural family, binds RNA nonspecifically but with high affinity and has been proposed to bind mRNA during initiation (*6*, *8*). It is also primarily responsible for the stimulation of the helicase activity of eIF4A although the RRM domain has also been linked to this function (*8*). Between the RNA binding domains there is a region rich in Asp, Arg, Tyr, and Gly amino acids, termed the DRYG domain (residues 214−327) which has been shown, both in vivo and in vitro, to be responsible both for self-association of eIF4B and for binding to the p170 subunit of eIF3 (*9*).

More recently, an interaction between eIF4B and the poly-(A) binding protein (PABP), first demonstrated in wheat germ (*10*), has now been confirmed for mammalian cells (*11*). This interaction appears to stimulate both the eIF4B-mediated activation of the helicase activity of eIF4A (*12*) and binding of poly(A) by PABP (*10*) and is apparently regulated by phosphorylation of both proteins (*13*). One line of evidence indicates that the interaction occurs between the C-terminal homodimerization domain of PABP and a region within the N-terminal 80 amino acids of eIF4B. However, homology between the N-terminal peptide of eIF4B and an internal sequence within eIF4G that binds to the N-terminal pair of RRM domains in PABP suggests that eIF4B may also bind to this region of PABP (*14*). Resolution of these apparently conflicting observations will require further work.

eIF4B has also been shown to interact specifically with the internal ribosome entry sites (IRES) of several picornaviruses which facilitate cap-independent translation initiation (*15*−*18*). The protein appears to bind to stem−loop features just upstream of the initiator AUG codon (*15*−*18*). Although mutations in these sites which abrogate eIF4B binding also severely reduce the initiation activity of the IRES, an absolute requirement for eIF4B in IRES function has not been formally demonstrated; in vitro experiments showed that eIF4B stimulated but was not required for the formation of IRES-mediated 48S preinitiation complexes (*19*).

Despite the functional dissection of eIF4B, the molecular mechanism of its role in translation initiation remains poorly understood. We are undertaking structural studies on the protein in order to assist the interpretation of the accumulating functional data. As a first step we have elucidated the structure of the eIF4B RRM domain.

RRM domains are one of the most ubiquitous structural domains, found in more than 200 different RNA binding

---

[†] J.G. is grateful to the Medical Research Council (U.K.) for a Ph.D. studentship. This work was funded by grant support from the Wellcome Trust, the Biotechnology and Biological Sciences Research Council (U.K.), and the Engineering and Physical Sciences Research Council (U.K.).

* Corresponding authors. S.C.: (44) 20 75947632 (phone); (44) 20 72250960 (fax); s.curry@imperial.ac.uk (e-mail). S.M.: (44) 20 75945315 (phone); (44) 20 72250960 (fax); s.j.matthews@imperial.ac.uk (e-mail).

[‡] Wolfson Laboratory.

[§] Blackett Laboratory.

10.1021/bi034506g CCC: $25.00

proteins. They can appear singly or in multiple repeats, and examples include *Drosophila* sex-lethal (Sxl), PTB, U1A, and PABP (*20*). Several crystal and solution structures exist for these proteins, some of which have been determined in complex with their RNA ligands (*21*−*30*). The RRM domain consists of a conserved sequence of approximately 80−90 residues, with common secondary structural elements that fold into a $\beta\alpha\beta\beta\alpha\beta$ topology. The β-sheet formed by the four strands acts as the RNA binding surface while the α-helices pack against the β-sheet on the opposite face of the molecule. The central pair of β-strands contains two short-sequence motifs (RNP1 and RNP2) that are particularly highly conserved among RRM domains (*20*). These motifs consist of stretches of eight and six residues, respectively, and are principally characterized by the presence of three highly conserved aromatic residues, which have been shown to be responsible for nonspecific ligand interactions via stacking with RNA bases (*23*, *25*, *29*, *30*). Loops between secondary structure elements are the most variable between RRM domains, differing not only in length but also in terms of their amino acid composition and mobility in solution (*20*). Loop 3, which connects β-strands 2 and 3, is known to be particularly variable and has been implicated as a ligand specificity determinant in other RRM domains (*20*, *31*). In the eIF4B RRM domain loop 3 is 12 residues in length and, unusually, contains three evenly spaced proline residues (PxxPxxP) so its structural configuration is of particular interest.

MATERIALS AND METHODS

Cloning and Preparation of the eIF4B RRM Domain. Initial attempts to express fragments containing the eIF4B RRM domain in *Escherichia coli* yielded proteins with low solubility which could not be concentrated for structural analysis without severe precipitation. We therefore mutated Phe99 to Tyr, since this residue is predicted to be exposed on the RNA binding surface, following a strategy that was previously applied to the RRM protein sex-lethal (*29*). The mutated eIF4B RRM domain, comprising residues 88−176 of human eIF4B, was cloned using PCR between the *Bam*HI and *Hin*dIII sites of pQE-9 (Qiagen) and expressed as a recombinant protein within *E. coli*. The expressed fragment contains an N-terminal hexahistidine tag (MRGSHHHHHHG) and was readily concentrated to 1.5 mM. A ^{15}N,^{13}C double-labeled sample of the eIF4B RRM domain construct was produced in minimal media, containing 0.07% $^{15}NH_4Cl$ and 0.2% [^{13}C]glucose, supplemented with 50 μg/mL ampicillin. Protein expression was induced by the addition of 1 mM isopropyl β-D-thiogalactopyroside. Clarified cell lysate was purified essentially to homogeneity in a single step on TALON metal affinity resin (Clontech). Purified eIF4B was dialyzed into 20 mM sodium acetate buffer at pH 5.5 and concentrated to 1.5 mM for NMR.

NMR Spectroscopy. The majority of NMR spectra were recorded at 298 K on a 500 MHz four-channel Bruker DRX500 spectrometer equipped with a *z*-shielded gradient triple resonance cryoprobe. Sequence-specific backbone and secondary structure assignments were determined using standard triple resonance methods (*32*−*35*). Side-chain assignments were obtained from HCCH-TOCSY experiments (*36*). 3D ^{1}H−^{15}N/^{13}C NOESY-HSQC (mixing time 100 ms) spectra provided distance restraints used in the final structure calculations. A further 3D ^{1}H−^{13}C NOESY-HSQC spectrum (mixing time 80 ms) was recorded at 800 MHz in order to resolve any overlap. An anisotropic liquid crystal sample of the eIF4B RRM domain was prepared in 3.0% (w/v) C12E5/hexanol ($r = 0.96$) as described (*37*). A total of 64 $^1D_{NH}$ values were measured using interleaved ^{15}N−^{1}H IPAP-HSQC spectra with carbonyl decoupling (*38*); 45 $^1D_{CC\alpha}$ and 44 $^1D_{NC\alpha}$ values were measured using previously described 3D HNCO-based experiments (*39*). All spectra were processed with XWinNmr and analyzed with NMRview (*40*). Amide protons involved in hydrogen bonds were identified by the presence of NH resonances in HSQC spectra recorded 12 h after being dissolved in D_2O. The hydrogen bond acceptors were identified by using distinctive neighboring NOEs. Characteristic $^{13}C_\beta$ and $^{13}C_\gamma$ chemical shifts and sequential NOE information were used to distinguish the cis/trans isomerization of the peptide bond for proline residues.

Structure Calculation. A total of 967 nuclear Overhauser effects (NOE) were assigned, of which 301 intraresidue NOEs were removed from final calculations as they did not contribute significantly to the tertiary structure. Final structures were calculated from random starting coordinates on the basis of 666 distance restraints, comprising 280 long-range NOEs (i to $i > 4$), 126 short-range NOEs ($1 < i \leq 4$), 260 sequential NOEs, and 24 distance restraints for 12 hydrogen bonds. In addition, 76 dihedral angle restraints (37 φ and 39 ϕ angles) from the backbone torsion angle prediction package TALOS (*41*) and 108 residual dipolar couplings were included. In total, this amounts to an average of 13 restraints per residue for the globular eIF4B RMM domain. Structure calculations were performed starting from extended structures and using a hybrid algorithm including torsion angle dynamics (TAD) and Cartesian dynamics executed within the program CNS (*42*). An initial TAD phase was performed at a temperature of 50000 K consisting of 2000 molecular dynamics steps each of 15 fs. A TAD cooling phase to 2000 K followed which included 2000 steps each of 15 fs. Finally, a Cartesian dynamics cooling phase to 300 K was executed using 3000 steps each with a 0.3 fs time step. All parameters were scaled using default CNS values. Residual dipolar couplings where introduced during the Cartesian dynamics cooling phase. One hundred structures were calculated and were ranked in overall energy terms; the best 15% were selected as representative and are presented in Figure 1A. No distance violation greater than 0.5 Å, no dihedral violation greater than 5°, and no residual dipolar coupling violations greater than 1.5 Hz were tolerated in the final family of structures. The structural coordinates have been deposited in the PDB.

RNA Binding Experiments. RNA A2 was identified by in vitro selection as an RNA molecule that binds to the N-terminal domains of eIF4B (residues 1−250) with high affinity ($K_d = 12.5$ nM) (*6*). The original construct, a kind gift from Dr. Nahum Sonenberg, was used to generate RNA for binding assays; it yields a transcript of 101 nucleotides (nt) which contains a core stem−loop structure of 33 nt that is implicated in binding eIF4B (*6*). In vitro transcriptions to generate α-^{32}P-labeled RNA transcripts of the A2 RNA fragment and the EMCV IRES (*43*) and nitrocellulose filter binding assays were performed as described previously (*21*, *44*). For NMR titration experiments, a truncated version of the same RNA was produced; this RNA contains the 33 nt

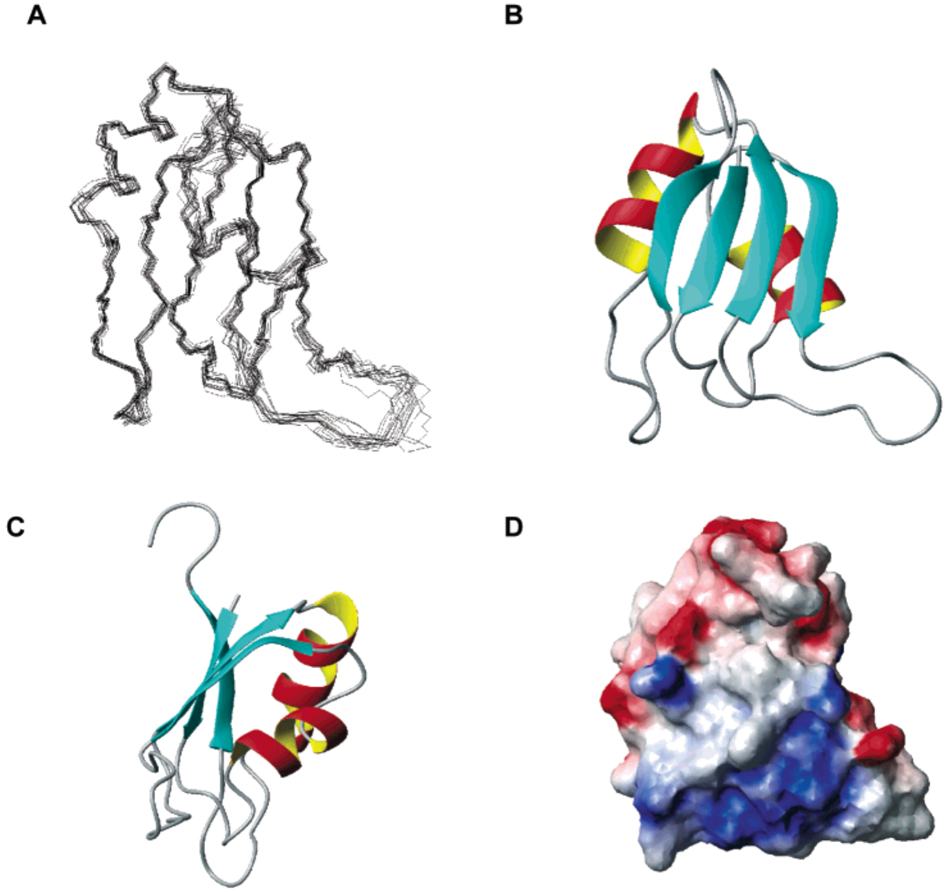

FIGURE 1: Solution structure of the RNA recognition motif (RRM) of eukaryotic translation initiation factor 4B. (A) Backbone trace of the 15 lowest energy structures of the eIF4B RRM domain superimposed for best fit of the N, C′, and Cα atoms of residues 96−129 and 137−172. (B) Ribbon diagram of the lowest energy solution structure in the same orientation as (A). (C) Ribbon diagram showing an approximate 90° rotation of the orientation represented in (B). (D) Electrostatic potential surface in the same orientation as (A) calculated using MolMol (*65*) with areas of positive charge in blue and areas of negative charge in red.

core stem−loop structure within a total transcript length of 54 nt. It was generated using the complementary primers A2HindIIIFOR, 5′-AGCTTAATGGTTGGAACGCACAGGCTTGACATCAACCATCAATCCG-3′, and A2EcoRIREV, 5′-AATTCGGATTGATGGTTGATGTCAAGCCTGTGCGTTCCAACCATTA-3′. These primers generate a 5′ *Hin*dIII site and a 3′ *Eco*RI site on hybridization and were used to insert the sequence into pGEM-4Z. The resulting pGEM-4Z-A2 plasmid was linearized with *Eco*RI and ethanol precipitated prior to in vitro transcription using a T7 Ribomax express kit (Promega) according to the manufacturer's protocol. Following transcription, the template was digested with 1 unit of RNase-free DNase/μg of DNA. Unincorporated nucleotides were removed using a NAP-5 column (Amersham Bioscience); the RNA was subjected to phenol−chloroform extraction and precipitated by addition of 0.5 volume of 7.5 M NH_4OAc and 2 volumes of ethanol. The resulting RNA pellet was dried and dissolved in diethyl pyrocarbonate-treated buffer: 100 mM NaCl, 50 mM Na_2HPO_4/NaH_2PO_4, pH 6.5, 1 mM DTT, and 0.02% azide. The purified RNA ran as a single band of the expected size on an 8% polyacrylamide gel containing 8 M urea.

For NMR mapping experiments, the $^{15}N/^{13}C$-labeled eIF4B sample was prepared in 20 mM sodium acetate buffer at pH 5.5 at approximately 30 μM in 0.5 mL. Purified RNA in the same buffer was introduced in four steps, at molar ratios (eIF4B:RNA) of 0, 0.1, 0.5, 1.0, and 2.7:1, and $^{15}N-^{1}H$ HSQC spectra were recorded at each stage under identical experimental conditions. Filter binding assays indicated that the truncated A2 construct bound to eIF4B with affinities that were essentially identical to the longer A2 RNA (data not shown).

RESULTS AND DISCUSSION

Solution Structure of eIF4B. A total of 967 nuclear Overhauser effects (NOE) were assigned, of which 301 intraresidue NOEs were removed from the final calculation as they did contribute to the tertiary structure. A total of 666 structurally significant NOE-derived interproton distance constraints, 24 distance restraints for backbone hydrogen bonds, 76 dihedral angle restraints comprising 37 ϕ and 39 φ angles, 64 backbone one-bond $^{15}N-^{1}H$, 45 one-bond $^{13}C-^{13}C\alpha$, and 44 one-bond $^{15}N-^{13}C\alpha$ residual dipolar couplings, were used for the structure determination, which amounts to an average of 13 restraints per residue for the globular domain. Sequential NOEs and $^{13}C_{\beta}/^{13}C_{\gamma}$ chemical shifts identified the trans isomer for the peptide bonds involving all assigned proline residues. A trans conformation was also given to Pro130, which was unassigned, based on energy terms in preliminary structure calculations. The 15 lowest energy structures were chosen from a total of 100 calculated on the basis of agreement with experimental data and structural quality. In the final family of structures, shown in Figure 1A, no NOE violation greater than 0.5 Å, no torsion

Table 1: Structural Statistics

RMSD from experimental restraints	
distance (Å)	0.058 ± 0.0035
dihedral angle (deg)	1.0244 ± 0.1114
RMSD from idealized covalent geometry	
bonds (Å)	0.0054 ± 0.00029
angles (deg)	0.734 ± 0.0271
impropers (deg)	0.644 ± 0.0427
coordinate RMSD for residues 96−127 and 137−172 (Å)	
all backbone atoms	0.63 ± 0.12
all heavy atoms	1.66 ± 0.14
Ramachandran plot[a]	
residues in most favored regions (%)	67.50 ± 1.89
residues in additionally allowed regions (%)	18.24 ± 2.21
residues in generously allowed regions (%)	12.60 ± 2.02
residues in disallowed regions (%)	1.82 ± 0.67

[a] Structural quality was evaluated using PROCHECK_NMR (*64*).

angle violation greater than 5°, and no residual dipolar coupling violation greater than 1.5 Hz were allowed. The statistics for the structure determination are summarized in Table 1.

As expected, the final structure of the eIF4B RRM domain exhibits a *β*α*ββ*α*β* topology, with the strands arranged in an antiparallel *β*4−*β*1−*β*3−*β*2 fashion. The *β*-sheet is packed against the two α-helices, which themselves lie almost perpendicular to one another (see Figure 1B,C). The three conserved aromatic residues in the RNP motifs on the *β*-sheet surface are facing into solvent, and in common with other RRM domain structures, they are oriented such that their hydrophobic rings are almost parallel to the surface of the *β*-sheet.

All areas of secondary structure are very well defined (Figure 1A); the average pairwise root-mean-squared deviation (RMSD) for the 15 unminimized final structures is 0.63 Å for the backbone atoms and 1.66 Å for the heavy atoms of residues 96−127 and 137−172. This comparison excludes four unstructured residues at the C-terminus of the protein and most of the loop 3 residues. Although loop 3 (residues 127−138) appears to adopt a well-defined conformation, the precision of this part of the structure relative to that of the secondary structure elements is somewhat reduced; the RMSD increases to 0.82 Å for the backbone atoms when loop 3 is included. A relative deficit of long-range NOEs and torsion angle restraints in this region are most likely the primary causes for the decreased accuracy for this loop. Additionally, backbone and side-chain assignments for one of the residues, Pro130, could not be made, which could be due to the attenuation of resonances caused by conformational exchange on an intermediate NMR time scale. The amide resonance of Ser131 was not visible in the ^{1}H−^{15}H HSQC spectrum, which may also be a result of intermediate conformational exchange. Nevertheless, a number of long-range contacts between loop 3 residues and the main body of the structure are observed. In particular, NOEs can be seen from the hydrophobic moiety of Arg135 in loop 3 to the side chains of Leu103, Tyr105, and Val107 in loop 1 which packs underneath and is closely associated with the body of the RRM. Dihedral angle restraints are also in place for several of the loop 3 residues, such as Arg128, Glu129, Arg135, and Leu136. Furthermore, the presence of more than one set of residual dipolar couplings for the majority of the non-proline residues in this area ensures that the loop orientation with respect to the rest of the structure is reasonably well defined.

Backbone Dynamics of the eIF4B RRM Domain. To probe the dynamics of the eIF4B RRM domain and, in particular, to examine the mobility of the atypical loop 3 region, relaxation parameters (^{15}N−T_1, ^{15}N−T_2, and the steady-state heteronuclear [^{1}H]−^{15}N NOE) were measured for resolved ^{1}H−^{15}N backbone cross-peaks (see Figure 2A−C). T_1 and T_2 values are consistent with a molecule the size and shape of eIF4B, with an estimated isotropic overall correlation time (τ_m) of 7.0 (±0.9) ns.

High T_1, T_2 and low NOE values were observed at the N- and C-termini, which reflect the increased flexibility in these regions, especially at the C-terminus where the final *β*-strand ends at residue Ala171. The presence of three proline residues in loop 3 (Pro127, Pro130, and Pro133) and a missing amide assignment on Ser131 mean that this area is not fully sampled. Despite this, a high degree of mobility on the pico- to nanosecond time scale is not apparent, which is consistent with proton−proton NOE evidence. Lipari−Szabo model-free analysis (*45*, *46*) of the experimental backbone ^{15}N relaxation data was carried out assuming isotropic tumbling in solution, and model selection was based on the method described by Mandel and co-workers (*47*). The sequence dependence of the model-free-derived parameters can be seen in Figure 2D−F. Areas of secondary structure have high order parameters, with an average of 0.87 indicating a highly ordered conformation. Aside from the N- and C-termini, which exhibit lower order parameters, there are two additional regions of interest. The first encompasses residues at the end of α1 and in the adjoining loop 2. Slow conformational exchange (R_{ex}) terms are also associated with three residues in this region, Glu114, Phe115, and Arg117, indicating conformation exchange on the micro- to millisecond time scale. The second region occurs at the C-terminal end of loop 3 and extends through to the subsequent *β*-strand, *β*3. Slow conformational exchange for residues Glu134, Leu136, Phe139, and Ala142 in this region is also observed. This indicates that although this part of loop 3 is ordered on the pico- to nanosecond time scale, it exhibits increased mobility relative to other secondary structure elements on a slower micro- to millisecond time scale.

Structural Comparison with Other RRM Domains. Structural similarities between the RRM domain of eIF4B and other RRM-containing proteins were assessed using the program DALI (*48*, *49*). The closest structural homologues were found to be the ligand-bound forms of the *Drosophila* sex-lethal RRM1 domain (sequence identity of 20%, an RMSD of 4.1 Å over 74 equivalent Cα atoms, and $Z = 6.5$) (*29*) and the polyadenylate binding protein (PABP) RRM-2 domain (sequence identity of 17%, RMSD of 3.8 Å over 75 residues, and $Z = 6.0$) (*26*). Both of these proteins contain two RRM domains joined by a small linker sequence, and they bind extended stretches of single-stranded RNA. This is in contrast to eIF4B, which contains a single RRM domain and is reported to bind a bulged stem−loop structure (*6*). One of the most significant differences is that helix α1 is approximately one turn shorter in eIF4B in comparison to these other RRM domains, but this feature, rather than being directly involved in RNA binding, is more often implicated in interactions with other proteins (*30*, *50*).

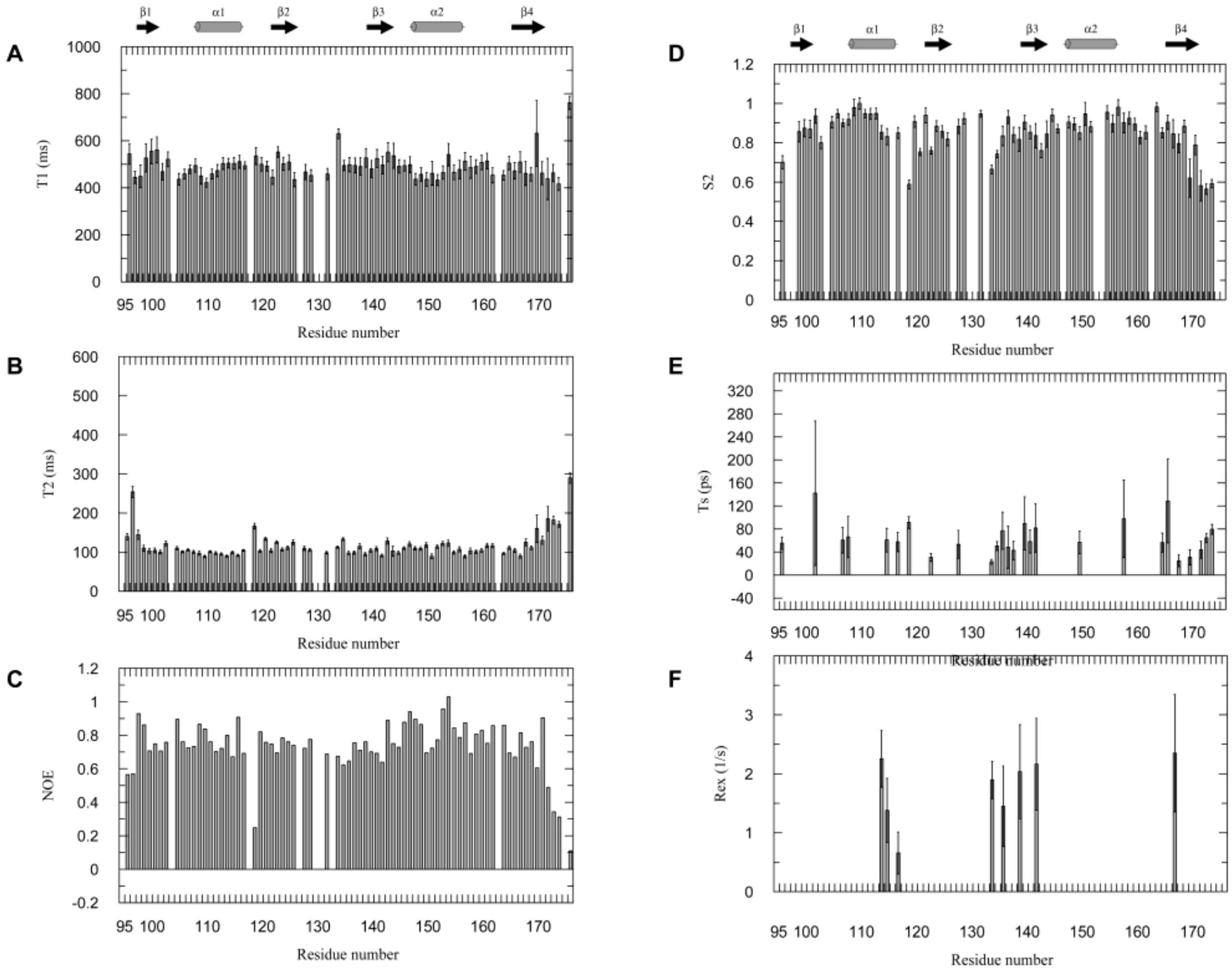

FIGURE 2: Sequence dependence of the relaxation data and model-free analysis for the eIF4B RRM: (A) T_1, (B) T_2, (C) NOE, (D) S^2, (E) τ_s, and (F) R_{ex}.

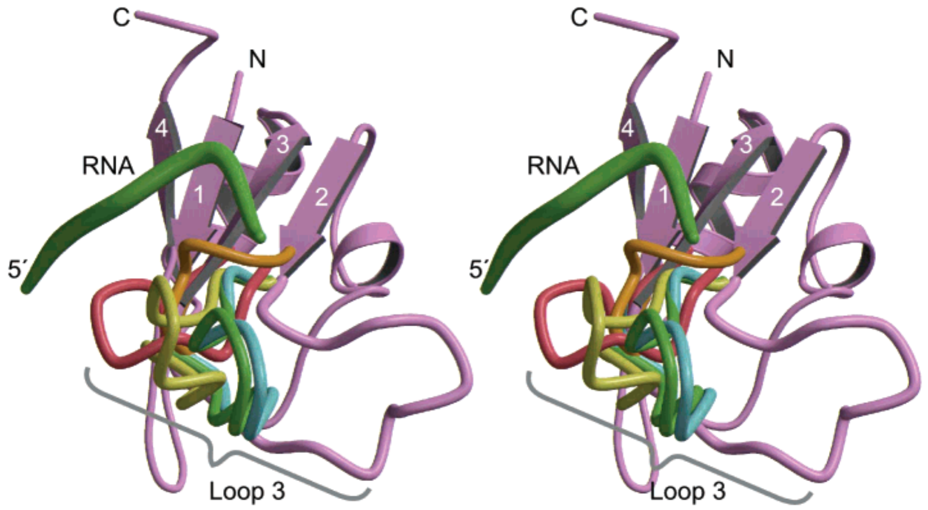

FIGURE 3: Superposition of a selection of loop 3 structures onto the RRM domain of eIF4B (violet). Loop 3 structures were taken from U2B″ [red (*30*)], PTB RRM3 [orange (*21*)], hnRNP-A1 RRM2 [yellow (*27*)], PABP RRM1 [green (*26*)], and sex-lethal RRM2 [blue (*29*)]. To indicate the typical disposition of an RNA ligand, the poly(A) ligand of PABP RRM1 is shown as a green worm. The figure was prepared using Bobscript and Raster3D (*66*, *67*).

Our structural and dynamic data for the eIF4B RRM indicate that loop 3, which has frequently been implicated as a determinant of RNA binding specificity by RRM domains (*28*−*30*), shows a better defined conformation than the corresponding feature in other RRM domains (*21*, *51*−*54*). Examination of the seed alignment of 90 RRM sequences in the Pfam database (*55*) suggests that this loop varies from 3 to 20 amino acids in length. In eIF4B loop 3 is 13 amino acids long but, unusually, is punctuated with three Pro residues, a feature observed in only one other sequence in the Pfam seed alignment, the yeast protein NSR1 for which there is no structural information. A superposition

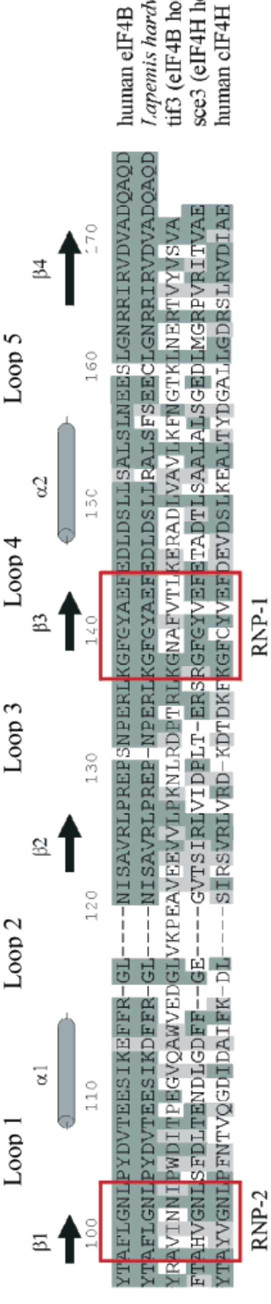

FIGURE 4: Primary sequence alignment of the human eIF4B RRM domain with several homologues. The secondary structure of eIF4B is illustrated above the alignment. Numbering is consistent with full-length human eIF4B. Conserved residues are highlighted, with identical residues in dark gray and similar residues in light gray. Red boxes surround the highly conserved RNP-1 and RNP-2 domains. The alignment was carried out using BLAST (*68*).

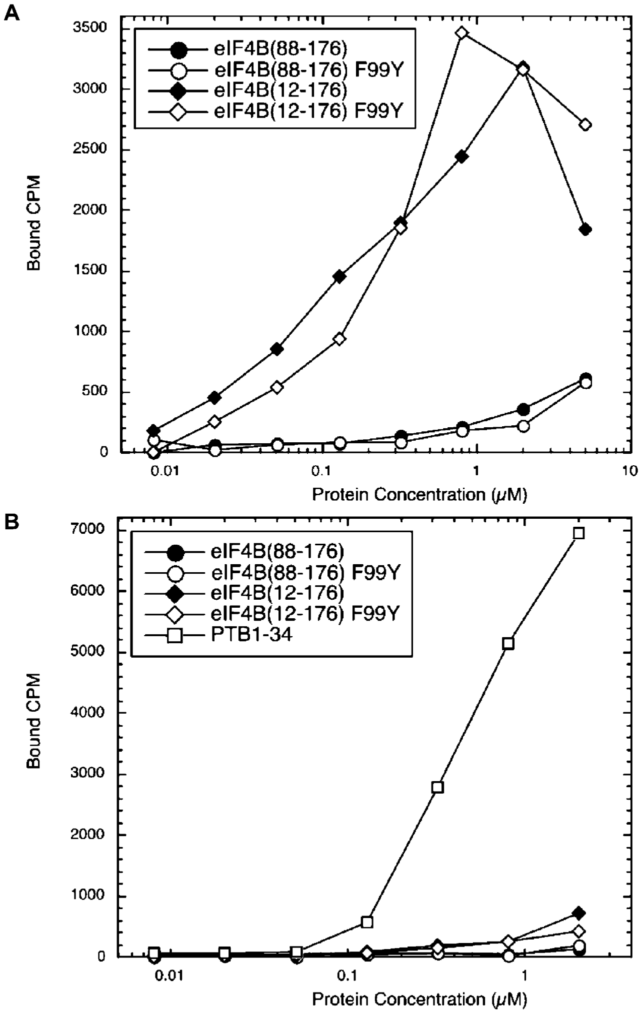

FIGURE 5: RNA binding experiments. Comparative RNA binding of eIF4B(12−176) and eIF4B(88−176) to (A) the A2 RNA fragment obtained by in vitro selection (*6*) and (B) the EMCV IRES (*43*).

of loop 3 structures from a representative variety of RRM domains is shown in relation to the eIF4B structure in Figure 3. The domains in this superposition include U2B″, PTB RRM3, hnRNP-A1 RRM2, PABP RRM1, and sex-lethal RRM2 and have loop 3 lengths of between 7 and 12 amino acids. The disposition of loop 3 in the eIF4B RRM distinguishes it from RRM domains in that it projects along the right-hand flank of the body of the domain rather than protruding out from the bottom end of the β-sheet binding surface where it can contact the underside of the RNA ligand. The novel disposition of loop 3 suggests that the eIF4B RRM may exhibit an altered mode of interaction with its RNA target, though this needs to be confirmed by direct structural analysis.

A sequence alignment, shown in Figure 4, reveals that the eIF4B RRM domain is highly homologous to the RRM domain from eIF4H, a recently discovered eIF4B-related translation initiation factor that can stimulate initiation of protein synthesis in vitro and appears to have an activity similar to that of eIF4B in translation assays (*56*, *57*). Although the RNP1 and RNP2 motifs and core residues are highly conserved between the two proteins (>50% identity), loop 3 is extremely divergent and notably lacks the three proline residues that characterize this feature in eIF4B. The

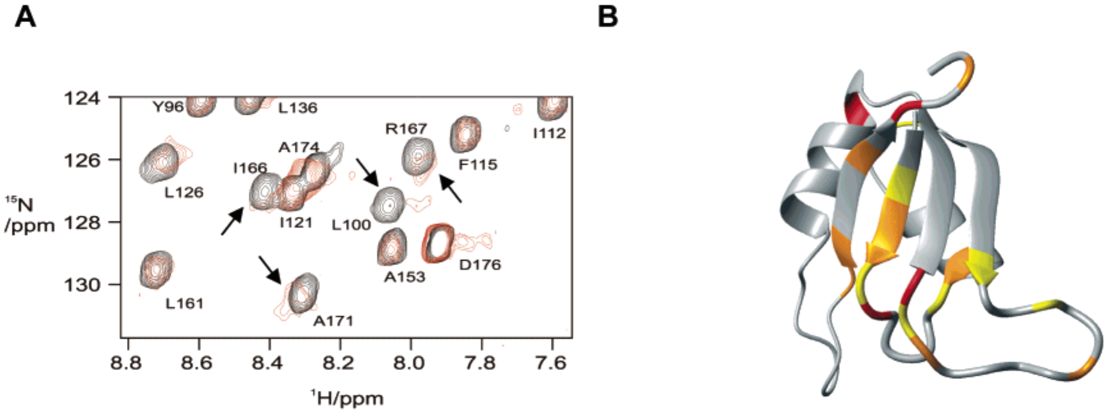

FIGURE 6: Identification of the binding site for the A2 RNA fragment in the RRM of eIF4B. (A) Region of the ^{1}H−^{15}N HSQC NMR spectrum for the ^{13}C,^{15}N-labeled eIF4B RRM domain with the identical region of the spectrum for the eIF4B RRM domain with equimolar amounts of the A2 RNA fragment overlaid (red). Arrows indicate key perturbed amides. (B) Ribbon representation for the structure of the eIF4B RRM domain. Residues shown in red, orange, and yellow experience large, medium, and small ^{1}H/^{15}N chemical shift perturbations, respectively (see Table 2).

marked sequence difference in loop 3 implies a different conformation and mobility for this region and may have important consequences for rRNA binding specificity between eIF4B and eIF4H RRMs.

RNA Interactions. RNA targets for RRM domains are diverse in sequence and secondary structure, but one similarity seems to be the presence of unpaired nucleotide bases in the ligand (*23*−*30*, *58*). An iterative in vitro genetic selection study demonstrated that an N-terminal fragment containing the first 250 amino acids of eIF4B (which therefore includes the RRM) selects a similar set of RNA ligands as the full-length protein. These RNAs are predicted to contain a common bulged stem−loop structure (*6*) in which the loop is of three to five nucleotides in length, with a G-C or C-G pair at its base, the 5′ position of the loop occupied by A or U (3:1 likelihood) and a G at the 3′ position. In the leading strand of the stem, a GGAA/C motif was found in a defined structural context, with at least one A nucleotide bulged and at least one G paired at the base of the bulge. However, a search for sequences predicted to adopt this type of secondary structure in 18S rRNA yielded only sequences in promoter or intron regions (*6*).

We used nitrocellulose filter binding assays to investigate the contribution of the RRM to RNA binding. Strikingly, a construct containing just the RRM [eIF4B(88−176)] bound with only very low affinity to an RNA target corresponding to one of the sequences identified using in vitro selection by Methot et al. (*6*) (Figure 5A). High-affinity binding (K_d = 0.2 μM) was observed for a construct with a significantly extended N-terminus [eIF4B(12−176)], though this was still lower than the binding affinity reported for a construct containing N- and C-terminal extensions to the RRM [eIF4B-(1−250); K_d = 0.012 μM] (*6*). It seems clear that these extensions are required for proper recognition of RNA targets by the eIF4B RRM, as has been observed for other RRM proteins (*26*, *29*, *59*−*62*). Previous work has demonstrated specific UV−cross-linking of eIF4B to picornavirus IRES RNA in rabbit reticulocyte lysates (*15*−*18*). To examine the interaction of our eIF4B fragments with such targets, we performed further filter binding assays using the EMCV IRES. In contrast to the results obtained with A2 RNA, neither the eIF4B(88−176) nor the eIF4B(12−176) constructs bound to the EMCV IRES RNA with high affinity (Figure 5B), though a positive control showed that this RNA is tightly bound by a fragment containing the two C-terminal RRM domains of the polypyrimidine tract binding protein (PTB1−34) (*21*). This is consistent with previous UV−cross-linking studies which have indicated that ATP-dependent conformational changes in the IRES may be required to expose the eIF4B binding site (*17*, *18*).

In general, RNA ligands bind to the β-sheet surface of RRM domains. Seven out of nine of the basic Arg or Lys residues in the eIF4B RRM are clustered on or adjacent to the β-sheet RNA binding surface. Arg128, Arg135, and Lys137 are all located within loop 3, and three Arg residues are clustered within the ^{164}RRIR^{167} motif in strand β4 on one side of the binding surface. The remaining Arg on this face of the domain (Arg125) projects from strand β2. Together, these basic residues surround the conserved triad of aromatic residues (Phe99, Phe139, Tyr141), which are likely to be involved in base-stacking interactions with the RNA ligand. Calculation of the electrostatic potential associated with the β-sheet RNA binding surface of the eIF4B RRM domain (Figure 1D) reveals a positively charged area at an angle of approximately 45° to the direction of the β-strands which would be able to make favorable interactions with complementary charges on RNA. Indeed, many RRM domains bind with their ligand at an angle of approximately 45° to the β-strands, and it seems likely that the eIF4B RRM would bind to its target RNA in a similar configuration.

To directly probe the RNA binding site in the eIF4B RRM, we performed an NMR chemical shift mapping experiment with A2 RNA. Analysis of amide line widths and chemical shifts for eIF4B RRM in the presence of the RNA was carried out as described previously (*63*). A comparison of a region from the ^{1}H−^{15}N HSQC spectra without and with RNA is shown in Figure 6. A number of clustered amide resonances move or broaden upon addition of RNA, signifying a likely contact with the ligand, while the majority of the spectrum remains unchanged. Furthermore, a general increase in line widths for unaffected amide resonances is observed, which is consistent with the slower molecular tumbling expected for a 35 kDa eIF4B−RNA complex. Severe resonance broadening that is specific to the affected amides suggests an exchange rate between free and bound states on an NMR intermediate time scale (see Figure 6A).

Table 2: Chemical Shift Perturbation Data

residue	location		combined $^1H/^{15}N$ chemical shift deviation[b]
N102	RNP2	large	2.04
D148	loop 4/α2		0.53
F139	RNP1		0.52
R164	^{164}RRIR^{167}		0.59
D172	C-terminus		1.34
Y99[a]	RNP2	medium	0.46
L100	RNP2		0.42
T108	loop 1/α1		0.35
N132	loop 3		0.33
K137	loop 3/RNP1		0.40
I166	^{164}RRIR^{167}		0.40
R167	^{164}RRIR^{167}		0.37
V170	β4/C-terminus		0.43
A174	C-terminus		0.36
T98	RNP2	small	0.20
G101	RNP2		0.21
V107	loop 1/α1		0.23
L119	loop 2		0.16
R125	β2/loop 3		0.18
L126	β2/loop 3		0.26
E129	loop 3		0.23
G138	RNP1		0.24
G140	β3		0.13
E158	loop 5		0.19

[a] The residue is F99 in the WT sequence. [b] Combined amide chemical shift calculated from the weighted sum $4\Delta\delta^H + \Delta\delta^N$. The ^1H chemical shift weighting and the classification of shift changes as "significant" were determined from a plot of $\Delta\delta^H$ versus $\Delta\delta^N$ as described in ref *63*.

Although line broadening prohibits complete tracking of peak movements throughout the titration, the dissociation constant was estimated to be ~50 μM, which is consistent with the weak binding observed in the filter binding assays ($K_d > 10$ μM). Interestingly, the three largest clusters of shifts occur in RNP2, RNP1, and the ^{164}RRIR^{167} motif in strand β4 (see Figure 6 and Table 2). As expected, most of the affected residues fall primarily on the β-sheet RNA binding surface. However, there are additional residues experiencing chemical shifts in loop 3, supporting the proposed role of loop 3 in RNA recognition. Included among these is Lys137, which is located at the loop 3/RNP1 boundary and was shown by mutagenesis to be involved in RNA binding (*6*). A shift is observed for Y99 (F99 in WT) which is exposed on the RNA binding surface and confirms that the conserved mutation does not prevent RNA binding. Moreover, a number of perturbed residues occur following β4, within the unstructured C-terminus (Figure 6 and Table 2), which suggest that they may contact the RNA and emphasize the contribution of the C-terminal extension to the RRM domain to RNA recognition.

In conclusion, the RRM from eIF4B adopts the classical βαββαβ topology. Though very similar to other RRM domains in overall structure, loop 3 exhibits some restricted motion and is clearly implicated in RNA recognition. Nevertheless, the core RRM domain provides only a relatively weak interaction with RNA targets and requires extensions at both the N- and C-termini for high-affinity binding. Further work is now in progress to examine the structural basis of the interactions of the extended RRM with RNA ligands.

ACKNOWLEDGMENT

We thank Chris Hellen and Tatyana Pestova for the gift of the cDNA of human eIF4B and Nahum Sonenberg for the pKS-A2 plasmid. We also thank Daniel Nietlispach and the BBSRC 800 MHz NMR service at Cambridge University.

REFERENCES

1. Milburn, S. C., Hershey, J. W. B., Davies, M. V., Kelleher, K., and Kaufman, R. J. (1990) Cloning and Expression of Eukaryotic Initiation Factor-4b Cdna−Sequence Determination Identifies a Common Rna Recognition Motif, *EMBO J. 9*, 2783−2790.
2. Benne, R., and Hershey, J. W. B. (1978) The mechanism of action of protein synthesis initiation factors from rabbit reticulocytes, *J. Biol. Chem. 253*, 3078−3087.
3. Trachsel, H., Erni, B., Schreier, M. H., and Staehelin, T. (1977) Initiation on mammalian protein synthesis: The assembly of the initiation complex with purified initiation factors, *J. Mol. Biol. 116*, 755−767.
4. Rozen, F., Edery, I., Meerovitch, K., Dever, T. E., Merrick, W. C., and Sonenberg, N. (1990) Bidirectional RNA helicase activity of eucaryotic translation initiation factors 4A and 4F, *Mol. Cell. Biol. 10*, 1134−1344.
5. Pause, A., Méthot, N., Svitkin, Y., Merrick, W. C., and Sonenberg, N. (1994) Dominant negative mutants of mammalian translation initiation factor eIF-4A define a critical role for eIF-4F in cap-dependent and cap-independent initiation of translation, *EMBO J. 13*, 1205−1215.
6. Methot, N., Pickett, G., Keene, J. D., and Sonenberg, N. (1996) In vitro RNA selection identifies RNA ligands that specifically bind to eukaryotic translation initiation factor 4B: The role of the RNA recognition motif, *RNA 2*, 38−50.
7. Naranda, T., MacMillan, S. E., and Hershey, J. W. (1994) Purified yeast translational initiation factor eIF-3 is an RNA-binding protein complex that contains the PRT1 protein, *J. Biol. Chem. 269*, 32286−32292.
8. Methot, N., Pause, A., Hershey, J. W. B., and Sonenberg, N. (1994) The Translation Initiation Factor-Eif-4b Contains an Rna- Binding Region That Is Distinct and Independent from Its Ribonucleoprotein Consensus Sequence, *Mol. Cell. Biol. 14*, 2307−2316.
9. Methot, N., Song, M. S., and Sonenberg, N. (1996) A region rich in aspartic acid, arginine, tyrosine, and glycine (DRYG) mediates eukaryotic initiation factor 4B (eIF4B) self-association and interaction with eIF3, *Mol. Cell. Biol. 16*, 5328−5334.
10. Le, H., Tanguay, R. L., Balasta, M. L., Wei, C. C., Browning, K. S., Metz, A. M., Goss, D. J., and Gallie, D. R. (1997) Translation initiation factors eIF-iso4G and eIF-4B interact with the poly-(A)-binding protein and increase its RNA binding activity, *J. Biol. Chem. 272*, 16247−16255.
11. Bushell, M., Wood, W., Carpenter, G., Pain, V. M., Morley, S. J., and Clemens, M. J. (2001) Disruption of the interaction of mammalian protein synthesis eukaryotic initiation factor 4B with the poly(A)-binding protein by caspase- and viral protease-mediated cleavages, *J. Biol. Chem. 276*, 23922−23928.
12. Bi, X., and Goss, D. J. (2000) Wheat germ poly(A)-binding protein increases the ATPase and the RNA helicase activity of translation initiation factors eIF4A, eIF4B, and eIF-iso4F, *J. Biol. Chem. 275*, 17740−17746.
13. Le, H., Browning, K. S., and Gallie, D. R. (2000) The phosphorylation state of poly(A)-binding protein specifies its binding to poly(A) RNA and its interaction with eukaryotic initiation factor (eIF) 4F, eIFiso4F, and eIF4B, *J. Biol. Chem. 275*, 17452−17462.
14. Groft, C. M., and Burley, S. K. (2002) Recognition of eIF4G by rotavirus NSP3 reveals a basis for mRNA circularization, *Mol. Cell 9*, 1273−1283.
15. Rust, R. C., Ochs, K., Meyer, K., Beck, E., and Niepmann, M. (1999) Interaction of eukaryotic initiation factor eIF4B with the internal ribosome entry site of foot-and-mouth disease virus is independent of the polypyrimidine tract-binding protein, *J. Virol. 73*, 6111−6113.
16. Ochs, K., Saleh, L., Bassili, G., Sonntag, V. H., Zeller, A., and Niepmann, M. (2002) Interaction of translation initiation factor eIF4B with the poliovirus internal ribosome entry site, *J. Virol. 76*, 2113−2122.
17. Ochs, K., Rust, R. C., and Niepmann, M. (1999) Translation initiation factor eIF4B interacts with a picornavirus internal

ribosome entry site in both 48S and 80S initiation complexes independently of initiator AUG location, *J. Virol. 73*, 7505−7514.

18. Lopez de Quinto, S., and Martinez-Salas, E. (2000) Interaction of the eIF4G initiation factor with the aphthovirus IRES is essential for internal translation initiation in vivo, *RNA 6*, 1380−1392.
19. Pestova, T. V., Hellen, C. U. T., and Shatsky, I. N. (1996) Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry, *Mol. Cell. Biol. 16*, 6859−6869.
20. Kenan, D. J., Query, C. C., and Keene, J. D. (1991) RNA recognition: towards identifying determinants of specificity, *Trends Biochem. Sci. 16*, 214−220.
21. Conte, M. R., Grune, T., Ghuman, J., Kelly, G., Ladas, A., Matthews, S., and Curry, S. (2000) Structure of tandem RNA recognition motifs from polypyrimidine tract binding protein reveals novel features of the RRM fold, *EMBO J. 19*, 3132−3141.
22. Nagai, K., Oubridge, C., Jessen, T. H., Li, J., and Evans, P. R. (1990) Crystal structure of the RNA-binding domains of the U1 small nuclear ribonucleoprotein A, *Nature 348*, 515−520.
23. Oubridge, C., Ito, N., Evans, P. R., Teo, C.-H., and Nagai, K. (1995) Crystal structure of the RNA-binding domains of the U1A spliceosomal protein complexed with an RNA hairpin, *Nature 372*, 432−438.
24. Varani, L., Gunderson, S. I., Mattaj, I. W., Kay, L. E., Neuhaus, D., and Varani, G. (2000) The NMR structure of the 38 kDa U1A protein−PIE RNA complex reveals the basis of cooperativity in regulation of polyadenylation by human U1A protein, *Nat. Struct. Biol. 7*, 329−335.
25. Wang, X., and Tanaka Hall, T. M. (2001) Structural basis for recognition of AU-rich element RNA by the HuD protein, *Nat. Struct. Biol. 8*, 141−145.
26. Deo, R. C., Bonanno, J. B., Sonenberg, N., and Burley, S. K. (1999) Recognition of polyadenylate RNA by the poly(A)-binding protein, *Cell 98*, 835−845.
27. Ding, J. Z., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R., and Xu, R. M. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA, *Genes Dev. 13*, 1102−1115.
28. Allain, F. H., Bouvet, P., Dieckmann, T., and Feigon, J. (2000) Molecular basis of sequence-specific recognition of pre-ribosomal RNA by nucleolin, *EMBO J. 19*, 6870−6881.
29. Handa, N., Nureki, O., Kurimoto, K., Kim, I., Sakamoto, H., Shimura, Y., Muto, Y., and Yokoyama, S. (1999) Structural basis for recognition of the tra mRNA precursor by the sex-lethal protein, *Nature 398*, 579−585.
30. Price, S. R., Evens, P. R., and Nagai, K. (1998) Crystal structure of the spliceosomal U2B″-U2A′ protein complex bound to a fragment of U2 small nuclear RNA, *Nature 394*, 645−650.
31. Burd, C. G., and Dreyfuss, G. (1994) RNA binding specificity of hnRNP A1: significance of hnRNP A1 high- affinity binding sites in pre-mRNA splicing, *EMBO J. 13*, 1197−204.
32. Grzesiek, S., and Bax, A. (1992) An Efficient Experiment for Sequential Backbone Assignment of Medium-Sized Isotopically Enriched Proteins, *J. Magn. Reson. 99*, 201−207.
33. Grzesiek, S., and Bax, A. (1992) Improved 3d Triple-Resonance NMR Techniques Applied to a 31-Kda Protein, *J. Magn. Reson. 96*, 432−440.
34. Muhandiram, D. R., and Kay, L. E. (1994) Gradient-Enhanced Triple-Resonance 3-Dimensional NMR Experiments with Improved Sensitivity, *J. Magn. Reson., Ser. B 103*, 203−216.
35. Kay, L. E., Xu, G. Y., and Yamazaki, T. (1994) Enhanced-Sensitivity Triple-Resonance Spectroscopy with Minimal H2o Saturation, *J. Magn. Reson., Ser. A 109*, 129−133.
36. Bax, A., Clore, G. M., and Gronenborn, A. M. (1990) H-1-H-1 Correlation Via Isotropic Mixing of C-13 Magnetization, a New 3-Dimensional Approach for Assigning H-1 and C-13 Spectra of C-13-Enriched Proteins, *J. Magn. Reson. 88*, 425−431.
37. Ruckert, M., and Otting, G. (2000) Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments, *J. Am. Chem. Soc. 122*, 7793−7797.
38. Ottiger, M., Delaglio, F., and Bax, A. (1998) Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra, *J. Magn. Reson. 131*, 373−378.
39. Permi, P., Rosevear, P. R., and Annila, A. (2000) A set of HNCO-based experiments for measurement of residual dipolar couplings in N-15, C-13, (H-2)-labeled proteins, *J. Biomol. NMR 17*, 43−54.
40. Johnson, B. A., and Blevins, R. A. (1994) NMR View−a Computer-Program for the Visualization and Analysis of NMR Data, *J. Biomol. NMR 4*, 603−614.
41. Cornilescu, G., Delaglio, F., and Bax, A. (1999) Protein backbone angle restraints from searching a database for chemical shift and sequence homology, *J. Biomol. NMR 13*, 289−302.
42. Brunger, 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) Crystallography & NMR system: A new software suite for macromolecular structure determination, *Acta Crystallogr. 54*, 905−921.
43. Kaminski, A., Hunt, S. L., Patton, J. G., and Jackson, R. J. (1995) Direct evidence that polypyrimidine tract binding protein (PTB) is essential for internal initiation of translation of encephalomyocarditis virus RNA, *RNA 1*, 928−938.
44. Yuan, X., Davydova, N., Conte, M. R., Curry, S., and Matthews, S. (2002) Chemical shift mapping of RNA interactions with the polypyrimidine tract binding protein, *Nucleic Acids Res. 30*, 456−462.
45. Lipari, G., and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules 1. Theory and Range of Validity, *J. Am. Chem. Soc. 104*, 4546−4559.
46. Lipari, G., and Szabo, A. (1982) Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules 2. Analysis of Experimental Results, *J. Am. Chem. Soc. 104*, 4559−4570.
47. Mandel, A. M., Akke, M., and Palmer, A. G. (1995) Backbone Dynamics of Escherichia-Coli Ribonuclease Hi−Correlations with Structure and Function in an Active Enzyme, *J. Mol. Biol. 246*, 144−163.
48. Holm, L., and Sander, C. (1993) Protein-Structure Comparison by Alignment of Distance Matrices, *J. Mol. Biol. 233*, 123−138.
49. Holm, L., and Sander, C. (1995) Dali−a Network Tool for Protein-Structure Comparison, *Trends Biochem. Sci. 20*, 478−480.
50. Sachs, A. B., and Varani, G. (2000) Eukaryotic translation initiation: there are (at least) two sides to every story, *Nat. Struct. Biol. 7*, 356−361.
51. Avis, J. M., Allain, F. H. T., Howe, P. W. A., Varani, G., Nagai, K., and Neuhaus, D. (1996) Solution structure of the N-terminal RNP domain of U1A protein: The role of C-terminal residues in structure stability and RNA binding, *J. Mol. Biol. 257*, 398−411.
52. Inoue, M., Muto, Y., Sakamoto, H., Kigawa, T., Takio, K., Shimura, Y., and Yokoyama, S. (1997) A characteristic arrangement of aromatic amino acid residues in the solution structure of the amino-terminal RNA-binding domain of Drosophila sex-lethal, *J. Mol. Biol. 272*, 82−94.
53. Ito, T., Muto, Y., Green, M. R., and Yokoyama, S. (1999) Solution structures of the first and second RNA-binding domains of human U2 small nuclear ribonucleoprotein particle auxiliary factor (U2AF(65)), *EMBO J. 18*, 4523−4534.
54. Nagata, T., Kanno, R., Kurihara, Y., Uesugi, S., Imai, T., Sakakibara, S., Okano, H., and Katahira, M. (1999) Structure, backbone dynamics and interactions with RNA of the C-terminal RNA-binding domain of a mouse neural RNA-binding protein, Musashi1, *J. Mol. Biol. 287*, 315−330.
55. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M., and Sonnhammer, E. L. (2002) The Pfam protein families database, *Nucleic Acids Res. 30*, 276−280.
56. Richter, N. J., Rogers, G. W., Jr., Hensold, J. O., and Merrick, W. C. (1999) Further biochemical and kinetic characterization of human eukaryotic initiation factor 4H, *J. Biol. Chem. 274*, 35415−35424.
57. Richter-Cook, N. J., Dever, T. E., Hensold, J. O., and Merrick, W. C. (1998) Purification and characterization of a new eukaryotic protein translation factor. Eukaryotic initiation factor 4H, *J. Biol. Chem. 273*, 7579−7587.
58. Allain, F. H., Howe, P. W., Neuhaus, D., and Varani, G. (1997) Structural basis of the RNA-binding specificity of human U1A protein, *EMBO J. 16*, 5764−5772.
59. Lutz-Freyermuth, C., Query, C. C., and Keene, J. D. (1990) Quantitative determination that one of two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA, *Proc. Natl. Acad. Sci. U.S.A. 87*, 6393−6397.

60. Gorlach, M., Burd, C. G., and Dreyfuss, G. (1994) The Determinants of Rna-Binding Specificity of the Heterogeneous Nuclear Ribonucleoprotein-C Proteins, *J. Biol. Chem. 269*, 23074−23078.
61. Query, C. C., Bentley, R. C., and Keene, J. D. (1989) A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein, *Cell 57*, 89−101.
62. Scherly, D., Boelens, W., van Venrooij, W. J., Dathan, N. A., Hamm, J., and Mattaj, I. W. (1989) Identification of the RNA binding segment of human U1 A protein and definition of its binding site on U1 snRNA, *EMBO J. 8*, 4163−4170.
63. Simpson, P. J., Bolam, D. N., Cooper, A., Ciruela, A., Hazlewood, G. P., Gilbert, H. J., and Williamson, M. P. (1999) A family IIb xylan-binding domain has a similar secondary structure to a homologous family IIa cellulose-binding domain but different ligand specificity, *Structure 7*, 853−864.
64. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR, *J. Biomol. NMR 8*, 477−486.
65. Koradi, R., Billeter, M., and Wuthrich, K. (1996) MOLMOL: A program for display and analysis of macromolecular structures, *J. Mol. Graphics 14*, 51.
66. Esnouf, R. (1997) An extensively modified version of Molscript that includes greatly enhanced colouring capabilities, *J. Mol. Graphics 15*, 133−138.
67. Merritt, E. A., and Bacon, D. J. (1997) Raster3D: Photorealistic molecular graphics, *Methods Enzymol. 277*, 505−524.
68. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res. 25*, 3389−3402.

BI034506G