The EMBO Journal Vol.18 No.9 pp.2631–2637, 1999

Structure and interactions of the translation initiation factor eIF1

C.Mark Fletcher, Tatyana V.Pestova[1,2], Christopher U.T.Hellen[1] and Gerhard Wagner[3]

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, [1]Department of Microbiology and Immunology, State University of New York Health Science Center, 450 Clarkson Avenue, Brooklyn, NY 11203, USA and [2]A.N.Belozersky Institute of Physico-Chemical Biology, Moscow State University, 119899 Moscow, Russia

[3]Corresponding author
e-mail: wagner@wagner.med.harvard.edu

eIF1 is a universally conserved translation factor that is necessary for scanning and involved in initiation site selection. We have determined the solution structure of human eIF1 with an N-terminal His tag using NMR spectroscopy. Residues 29–113 of the native sequence form a tightly packed domain with two α-helices on one side of a five-stranded parallel and antiparallel β-sheet. The fold is new but similar to that of several ribosomal proteins and RNA-binding domains. A likely binding site is indicated by yeast mutations and conserved residues located together on the surface. No interaction with recombinant eIF5 or the initiation site RNA GCCACAAUGGCA was detected by NMR, but GST pull-down experiments show that eIF1 binds specifically to the p110 subunit of eIF3. This interaction explains how eIF1 is recruited to the 40S ribosomal subunit.

Keywords: eIF1/structure/translation initiation

Introduction

Eukaryotic translation initiation factor 1 (eIF1) was first purified as a factor stimulating binding of Met-tRNA and mRNA to the ribosome (Schreier *et al*., 1997; Trachsel *et al*., 1977). It is essential for growth in yeast (Yoon and Donahue, 1992). Furthermore, two classes of mutations in yeast eIF1 indicate a role for this protein in ensuring accurate initiation site selection: *sui1* mutations allow initiation to occur at non-AUG codons by mismatch base pairing with the initiator tRNA (Castilho-Valavicius *et al*., 1990; Yoon and Donahue, 1992) and *mof2* mutations increase ribosome frameshifting directed by the yeast killer virus mRNA (Cui *et al*., 1998).

Additional information about the function of eIF1 has been provided by toe printing assays of reconstituted mammalian translation complexes (Pestova *et al*., 1998). These experiments show that in the absence of eIF1 and eIF1A, the ribosomal complex stalls close to the 5′ cap, while in the presence of these two proteins it reaches the initiation codon. eIF1 is therefore essential for scanning. Toe printing experiments also show that eIF1 and eIF1A destabilize the stalled cap-proximal complex and a complex formed by internal entry at an incorrect AUG codon in the encephalomyocarditis virus mRNA. This is consistent with the role in initiation site selection described above.

In yeast, eIF1 is a subunit of the large ribosome-binding factor eIF3 (Naranda *et al*., 1996). Yeast eIF3 is sufficiently similar to its mammalian counterpart to replace it in an *in vitro* assay of translation initiation (Naranda *et al*., 1994). However, eIF1 does not co-purify with mammalian eIF3 (Brown-Luedi *et al*., 1982) and the relationship between these factors remains to be determined. Comparison of the subunit composition of eIF3 in yeast and mammals is currently the subject of much research (Asano *et al*., 1997, 1998; Phan *et al*., 1998).

Human eIF1 is a protein of 113 residues (Figure 1). It is 62% identical in sequence to yeast eIF1, sufficiently similar to restore a normal phenotype to *mof2* mutants (Cui *et al*., 1998). eIF1 homologs are found also in other eukaryotes, archaea and some bacteria (Kyrpides and Woese, 1998).

eIF1 is therefore an essential and universally conserved translation factor with established functions, but the mechanism of its action is unknown. We address this question by describing the high-resolution structure of the protein, a likely interaction site and information about the molecules with which it interacts.

Results and discussion

Determination of the structure of recombinant human eIF1

All of our experiments used recombinant human eIF1 with an N-terminal His tag of the sequence MRGSHHHHHHTDP. This protein construct gives the same results as eIF1 purified from rabbit reticulocytes in toe printing assays of translation initiation (Pestova *et al*., 1998).

The protein was labeled uniformly with ^{15}N or with ^{15}N and ^{13}C for NMR analysis. Spectra generally were recorded at a high salt concentration, 350 mM NaCl, which was necessary to obtain a high concentration of eIF1. Experiments to detect binding to other molecules were carried out at lower salt and protein concentrations, as described below.

We obtained complete backbone assignments, except for the first two residues and overlapping histidines of the N-terminal tag, and complete side chain assignments for most residues. We calculated 30 structures of the entire protein construct, discarding one with a distance restraint violation energy of 73.6 kcal/mol compared with 8.0–20.2 for the remaining 29 structures. Statistics for these structures are given in Table I.

```
                    10        20        30        40        50
Human   MSAIQNLHSFDPFADASKGDDLLPAGTEDYIHIRIQQRNGRKTLTTVQGIADDYDKK
        || |:||:|||||||:  |||   | | :|||||||||||||||||||||: ::||||
Yeast   MS-IENLKSFDPFADT--GDD-ETA-TSNYIHIRIQQRNGRKTLTTVQGVPEEYDLK

                    60        70        80        90       100       110
Human   KLVKAFKKKFACNGTVIEHPEYGEVIQLQGDQRKNICQFLV-EIGLAKDDQLKVHGF
        :::|: || ||||| ::  || ||:|||||||  :|:|:: ::|| |  ::|:|||
Yeast   RILKVLKKDFACNGNIVKDPEMGEIIQLQGDQRAKVCEFMISQLGLQKK-NIKIHGF
                                  **                      *
```

Fig. 1. Sequences of human and yeast eIF1. Asterisks indicate the sites of mutations in yeast eIF1 that are discussed in the text.

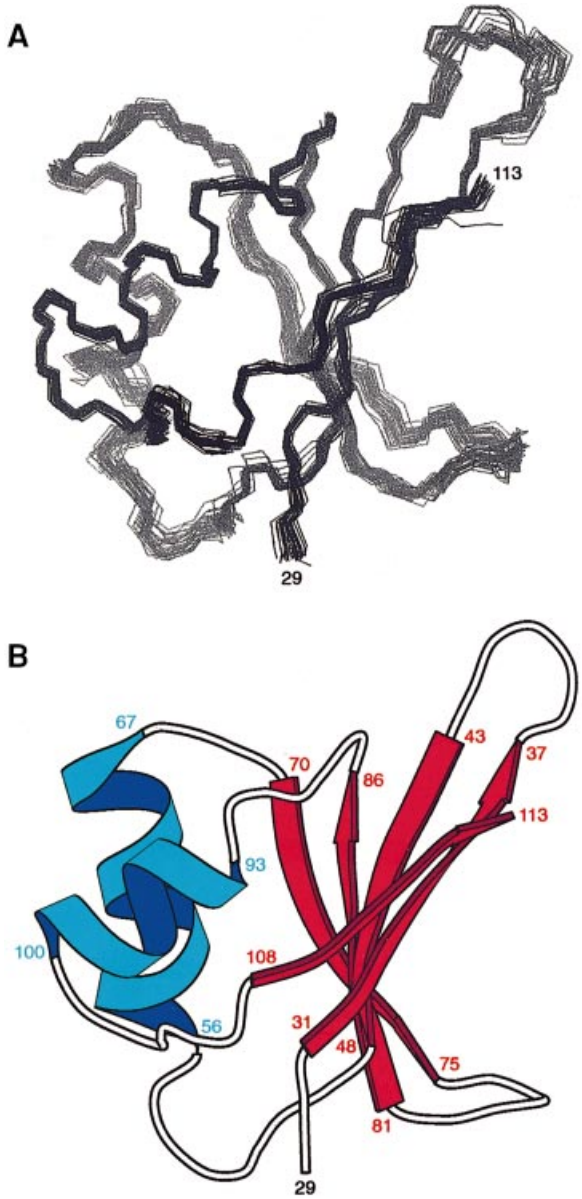

Fig. 2. Structure of the folded region of eIF1. (**A**) Backbone carbon and nitrogen atoms of residues 29–113 of the 29 structures, superimposed on those atoms of the structure with the lowest distance restraint violation energy. (**B**) Smoothed trace through the C_α positions of residues 29–113 of the structure with the lowest distance restraint violation energy.

eIF1 has a new fold but resembles ribosomal and RNA-binding proteins

Residues 29–113 of the native sequence form a tightly folded domain with two α-helices on one side of a five-stranded parallel and antiparallel β-sheet (Figure 2). No other protein domain is known to have an identical fold. However, structures with a general similarity to eIF1, namely a β-sheet with α-helices on one side, are found in three classes of small protein domains: ribosomal protein S6 and RNP RNA-binding domains (Hoffman *et al*., 1991; Lindahl *et al*., 1994; Oubridge *et al*., 1994; Allain *et al*., 1997), ribosomal protein S3 and KH RNA-binding domains (Castiglione Morelli *et al*., 1995; Musco *et al*., 1996), and double-stranded RNA-binding domains, which have a similar fold to ribosomal protein S5 (Ramakrishnan and White, 1992; Bycroft *et al*., 1995; Kharrat *et al*., 1995; Nanduri *et al*., 1998; Ryter and Schultz, 1998).

eIF1 differs from all of these structures in the number and order of secondary structure elements in the sequence and the arrangement of the strands in the β-sheet. Only one feature of the topology is common: the βαββ segment with (+2x, –1) β-strand connections (Richardson, 1981) formed by the second, third and fourth β-strands and the first α-helix of eIF1. The same topology is found in the first three β-strands and first α-helix of the RNP and S6 proteins and in the three β-strands and first α-helix of the KH and S3 proteins. This may reflect an ancient evolutionary relationship between these proteins. Although this part of the RNP domains contains much of the RNA-binding site, which is located on the surface of the β-sheet, there is no evidence for a binding site in the equivalent region of eIF1 or the KH and S3 proteins.

The His tag and first 28 residues of the native sequence are not folded

The N-terminal 41 residues of the protein, comprising the His tag and first 28 residues of the native sequence, have no folded structure. This is deduced from the high levels of backbone flexibility measured by steady-state $\{^1H\}$–^{15}N NOEs (Figure 3), the absence of long-range ^1H–^1H NOE connectivities and ^1H and ^{13}C chemical shifts characteristic of unstructured peptides (Wishart *et al*., 1995).

Although there are no long-range interactions in the N-terminal region, multiple medium-range NOE connectivities are observed at residues 1–4 (together with the last two residues of the His tag) and at residues 10–13. The presence of localized structure in these areas is confirmed by small positive $\{^1H\}$–^{15}N NOEs for residues 2, 13 and 14 (Figure 3). This localized structure seems to involve hydrophobic side chains: Ile4 shows many NOE connectivities to Met1 and residues 10 and 13 are phenylalanines, although an interaction between the two rings cannot be detected because of overlap. Residual structure involving hydrophobic side chains was observed in the urea-dena-

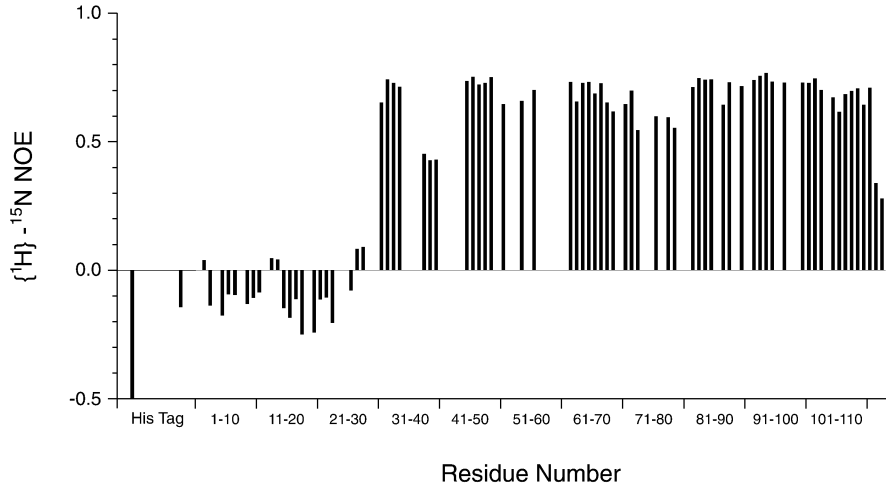

Fig. 3. Steady-state $\{^1H\}$–^{15}N NOEs. The off scale value for the third residue of the His tag is –1.01. Errors range from 1 to 11% with a mean of 3.4%. Gaps in the data are due to proline residues or overlapped signals.

Table I. Statistics for the 29 structures

No. of distance restraints	
Intra-residue	809
Sequential	496
Medium range ($2 \leqslant \|i–j\| \leqslant 4$)	338
Long range ($\|i–j\| \geqslant 5$)	531
Hydrogen bond restraints	22
Lower limits	21
Total	2217
Distance restraint violations	
No. of violations >0.1 Å	1–8
Largest violation (Å)	0.376
Mean r.m.s. violation (Å)	0.011 ± 0.0013
R.m.s. deviations from ideal geometry	
Bond lengths (Å)	0.0013
Bond angles (°)	0.315
Improper angles (°)	0.182
Mean coordinate r.m.s. deviations from the mean structure for residues 29–113 (Å)	
Backbone C and N atoms	0.43 ± 0.071
All heavy atoms	0.90 ± 0.065

tured 434 repressor protein (Neri *et al*., 1992). The presence of proline residues at the end of the His tag and at residue 12 may also be significant, perhaps because of their lower backbone flexibility.

It is conceivable that the lack of folded structure in the N-terminal region is an artifact caused by the solution conditions chosen (particularly the high salt concentration) or by the presence of the His tag. This seems unlikely, however, for several reasons. First, the structure is unchanged at 100 mM NaCl and pH 7.0, as shown by an identical ^{15}N HSQC spectrum (this includes signals from residues 1–28 of the native sequence and residues 3–4 and 11–12 of the His tag). Secondly, the His-tagged construct is sufficiently similar to eIF1 purified from rabbit reticulocytes to give identical results in toe printing assays of translation initiation (Pestova *et al*., 1998). Thirdly, the lack of structure seems to be reflected in the evolution of the native sequence: alignment of eIF1 homologs from eukaryotes, archaea and bacteria reveals much lower sequence conservation at the N-terminus than in the folded region (Kyrpides and Woese, 1998).

Yeast mutations and conserved surface residues suggest a binding site

We saw in the Introduction that mutations in yeast eIF1 implicate this protein in maintaining the accuracy of initiation site selection. The *sui1* mutations D88Y, D88G and Q89P allow initiation at non-AUG codons, while the *mof2* mutation G112R causes an increase in programmed ribosomal frameshifting. These phenotypes are closely related, since D88Y also increases frameshifting and G112R allows initiation at a UUG codon (Cui *et al*., 1998). This indicates that the same process or interaction is affected by the two classes of mutations.

All three mutated residues are conserved between yeast and human eIF1 (Figure 1). In the structure of human eIF1, D88, Q89 and G112 are found close together on the surface of the protein. Furthermore, the same region includes the side chains of C69, Q86, G87 and R90, residues that are almost perfectly conserved among eIF1 homologs from eukaryotes, archaea and bacteria (Kyrpides and Woese, 1998). These data suggest that this area of the surface is directly involved in the initiation site selection function of eIF1, most likely as a binding site for another molecule (Figure 4).

Clustered charges and a possible second binding site

Of 26 fully charged residues on the surface of the folded domain, 23 are grouped into clusters of residues with the same charge: three clusters of positively charged residues and two of negatively charged residues. Intermolecular electrostatic interactions between these clusters, perhaps also involving the His tag, may explain why a high level of salt is required to prevent eIF1 from precipitating at high concentration.

One of these clusters is particularly striking, comprising seven lysine residues on the surface of the first α-helix (residues 56–58, 61 and 64–66). Five of these residues are well conserved among eIF1 homologs from eukaryotes, archaea and bacteria (Kyrpides and Woese, 1998). This

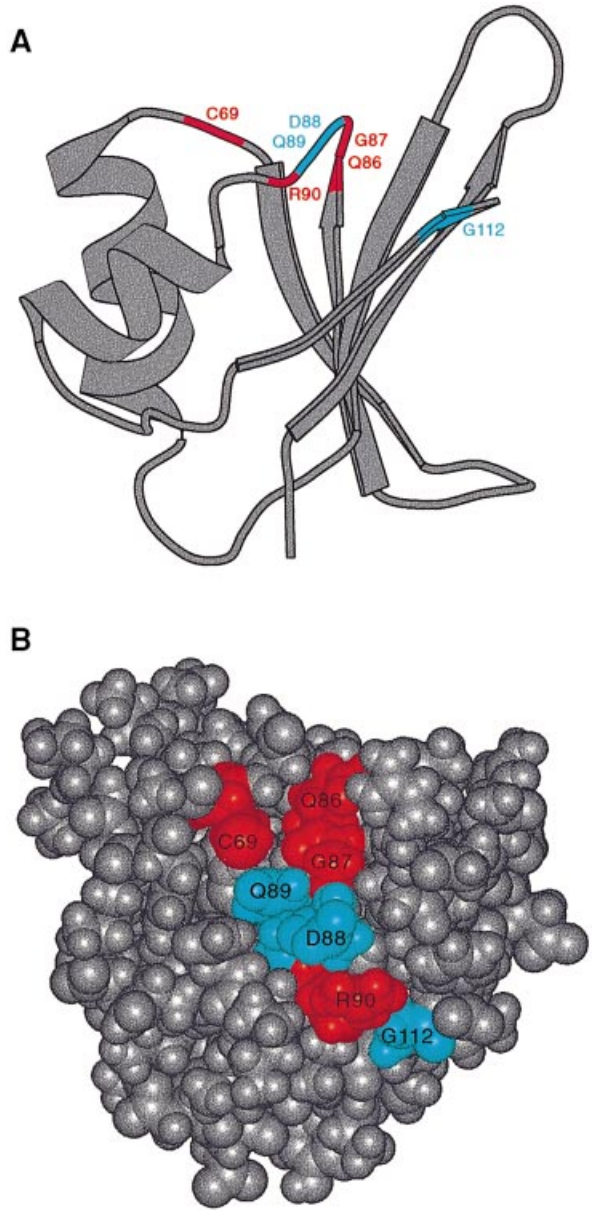

Fig. 4. Yeast mutations and conserved surface residues suggest a binding site. Yeast mutation sites are shown in blue and conserved surface residues in red. (**A**) Structure of the folded region of eIF1 as in Figure 2B. (**B**) Space-filling model of the same structure rotated 90° towards the viewer.

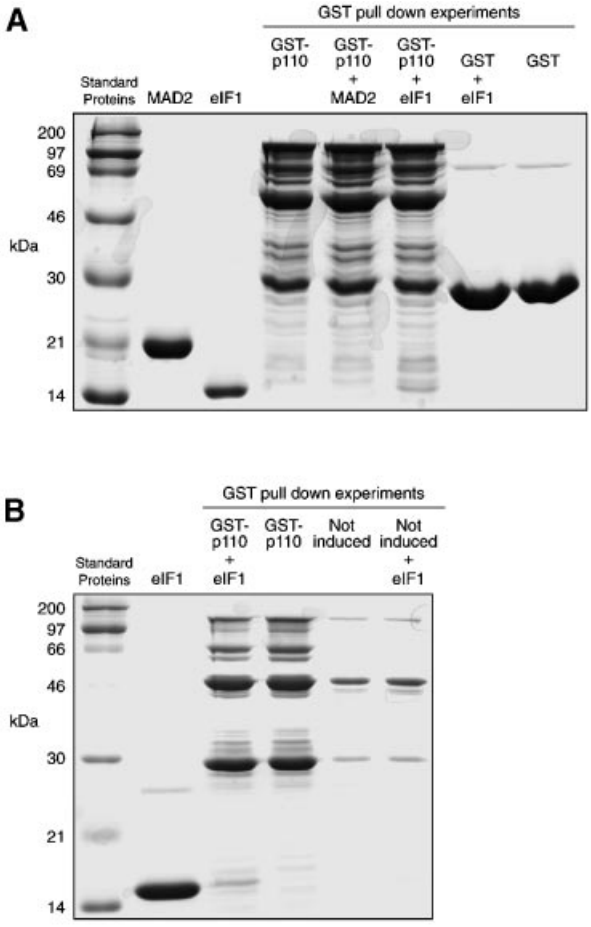

Fig. 5. eIF1 binds to the p110 subunit of eIF3 *in vitro*. SDS–PAGE gels stained with Coomassie Blue, showing the results of GST pull-down experiments along with 25 μg of pure eIF1 and MAD2 (a control protein with a His tag similar to that of eIF1). GST–p110 is unstable when expressed in *E.coli* (Asano *et al*., 1997). (**A**) This experiment was carried out on three occasions with the same result, and the identity of the pulled down eIF1 band was confirmed by N-terminal sequencing. (**B**) Negative control using an *E.coli* culture in which expression of GST–p110 was not induced, to show that the positive result is not due to interaction of eIF1 with an *E.coli* protein. Some expression of GST–p110 appears to occur without induction, and pull-down of eIF1 is still observed (a very faint band not visible in this figure). However, the amount of eIF1 pulled down is proportional to the amount of purified proteins visible and not to the volume of cell culture used (which was the same in all experiments). This shows that eIF1 binds to a protein whose expression is induced, almost certainly p110 or one of its degradation products, and not to an *E.coli* protein.

region is located some distance away from the mutation sites and conserved residues described above and is, therefore, a possible second binding site, with the positive charges making it suitable for interacting with the phosphate backbone of an RNA molecule.

eIF1 does not bind to an initiation site RNA in vitro

The resemblance of eIF1 to RNA-binding domains and the presence of clustered positive charges on its surface suggest that it may bind to RNA. However, we saw no interaction between eIF1 and an oligoribonucleotide with the nearly optimal mRNA initiation site sequence GCCACAAUGGCA (Kozak, 1989), assayed by inspection of the ^{15}N HSQC spectrum of eIF1 for changes upon addition of unlabeled RNA. This negative result is not conclusive because the conditions we chose may have prevented binding, eIF1 may only bind the RNA in the context of the complete translation machinery or it may prefer a different RNA sequence, although if it was a strong RNA-binding protein we might expect to see some indication of non-specific binding in the NMR assay (Harding *et al*., 1997).

eIF1 does not bind to recombinant eIF5 in vitro

An interaction between yeast eIF1 and the 45 kDa factor eIF5 has been detected by yeast two-hybrid assay (H.Yoon and T.F.Donahue, unpublished data). In contrast, we saw no interaction between human eIF1 and a recombinant form of the 49 kDa human eIF5 protein, using NMR to

detect binding as described in the preceding section. There are several possible explanations for this discrepancy. First, the conditions or protein constructs we chose may have prevented binding or the interaction may only occur in the context of the complete translation machinery. Secondly, the interaction between eIF1 and eIF5 may not be conserved between yeast and humans. Thirdly, the two-hybrid result may be explained by an indirect interaction, since both eIF1 and eIF5 proteins of yeast have been shown to bind to the Nip1p subunit of eIF3 by GST pull-down and two-hybrid assays (Asano *et al*., 1998; Phan *et al*., 1998).

eIF1 binds to the p110 subunit of eIF3 in vitro

The interaction between yeast eIF1 and Nip1p that was just mentioned suggests that human eIF1 may interact with the human homolog of Nip1p, the p110 subunit of eIF3. Since p110 is unstable when expressed in *Escherichia coli*, yielding a small amount of degraded protein (Asano *et al*., 1997), we did not try to observe this interaction using NMR. However, GST pull-down assays show that eIF1 binds specifically to p110 *in vitro* (Figure 5). GST–p110 pulls down eIF1 but not MAD2, a control protein with a His tag similar to that of eIF1; GST alone does not pull down eIF1. In addition, the positive result is dependent upon induction of GST–p110 expression and, therefore, is not due to interaction of eIF1 with an *E.coli* protein.

This result provides the first evidence that eIF1 is associated with eIF3 in a mammalian system. It shows that this association, and the specific interaction with the p110/Nip1p subunit, is conserved between humans and yeast. This is consistent with recent work indicating that at least a core complex of eIF3 subunits is conserved between yeast and mammals (Asano *et al*., 1998; Phan *et al*., 1998).

What are the implications of eIF3 association for the activity of eIF1? First, it is consistent with the fact that eIF1 stimulates binding of Met-tRNA and mRNA to the ribosome (Trachsel *et al*., 1977; Schreier *et al*., 1997), since eIF3 is believed to be directly involved in both of these activities, recruiting Met-tRNA via an interaction with eIF2 and mRNA via eIF4F (Merrick and Hershey, 1996). Secondly, it provides a mechanism for recruitment of eIF1 to the 40S ribosomal subunit, in a location close to the mRNA, initiator tRNA and eIF2.

It is not clear whether a single interaction is sufficient for eIF1 to carry out its function. Pestova *et al*. (1998) suggest that eIF1 or eIF1A enables scanning by clamping the mRNA in place on the 40S ribosomal unit, a role that has been suggested previously for eIF3 in general (Jackson, 1996). Binding to the p110 subunit of eIF3 could position eIF1 to act as a clamp in this way. It is also possible that eIF1 interacts with a second molecule, for which there are many candidates. Binding to the ribosome or to an associated factor could fix the other end of the clamp in place. Alternatively, eIF1 could interact with the mRNA or tRNA to affect scanning and initiation site recognition or it could interact with eIF2 to affect GTP hydrolysis. The question of a second interaction will be addressed in future work.

Materials and methods

Preparation of eIF1

Escherichia coli strain BL21(DE3) were transformed with the plasmid vector pQE(His6-eIF1) (Pestova *et al*., 1998) and grown at 37°C in M9 minimal medium containing 1 g/l NH_4Cl and 2 g/l D-glucose. Expression was induced at $OD_{600} = 0.6$ with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and the cells grown for a further 6 h before harvesting. The cells were lysed by sonication in 50 mM NaH_2PO_4/Na_2HPO_4 pH 8.0, 300 mM NaCl, 10 mM β-mercaptoethanol, 0.5 mM phenylmethylsulfonyl fluoride (PMSF), 0.1% Triton X-100, one Boehringer Mannheim Complete Protease Inhibitor tablet per 200 ml. eIF1 was purified on Ni-NTA agarose resin (Qiagen), washing with 50 mM NaH_2PO_4/Na_2HPO_4 pH 8.0, 300 mM NaCl, 10 mM β-mercaptoethanol, 1 mM EDTA, and eluting with 50 mM NaH_2PO_4/Na_2HPO_4 pH 6.0, 300 mM NaCl, 50 mM imidazole/HCl, 10 mM β-mercaptoethanol. The resulting protein was >95% pure with a single visible impurity of ~28 kDa (Wulfing *et al*., 1994). This was dialyzed against 50 mM NaH_2PO_4/Na_2HPO_4 pH 6.0, 350 mM NaCl, 1 mM dithiothreitol (DTT), 1 mM EDTA, 0.1 mM NaN_3 and concentrated. D_2O was added to 10% or the solvent was replaced by centrifugation through a Sephadex G-25 gel filtration column equilibrated with the same buffer in 99.96% D_2O. NMR samples contained ~3 mM protein in a 300 µl volume using a Shigemi NMR tube.

NMR spectroscopy

NMR spectra were recorded at 27°C on Varian UnityPlus 400, Unity 500 and Inova 750 spectrometers. Backbone ^1H, ^{13}C and ^{15}N assignments were obtained from 400 MHz HNCA, HN(CO)CA, HNCO and HN(CA)CO experiments with residue type information provided by a CBCA(CO)NH spectrum (Bax and Ikura, 1991; Grzesiek and Bax, 1992; Kay *et al*., 1994; Yamazaki *et al*., 1994; Matsuo *et al*., 1996). Side chain ^1H and ^{13}C assignments were obtained from 400 MHz HBHA(CBCACO)NH, 750 MHz ^{15}N TOCSY-HSQC with 60 ms mixing time and 400 MHz HCCH-TOCSY with 23.5 ms mixing time (Bax *et al*., 1990; Grzesiek and Bax, 1993; Zhang *et al*., 1994). Distance restraints, together with aromatic ring proton and side chain NH_2 assignments, were obtained from 750 MHz 2D NOESY and ^{15}N NOESY-HSQC spectra with mixing times of 60 and 120 ms and a 750 MHz 80 ms ^{13}C NOESY-HSQC spectrum (Zuiderweg *et al*., 1986; Ikura *et al*., 1990; Talluri and Wagner, 1996). The latter experiment used WURST-2 pulses to decouple aromatic and aliphatic protons during the indirect proton evolution time (Kupce and Freeman, 1995). Steady-state $\{^1H\}$–^{15}N NOEs were determined from 500 MHz spectra as previously described (Farrow *et al*., 1994). Spectra were processed with Felix (Molecular Simulations Inc.) and analyzed with XEASY (Bartels *et al*., 1995).

Distance restraints

The volumes of non-overlapped cross-peaks in NOESY spectra with 60 and 80 ms mixing times were integrated using the interactive mode of XEASY or the associated program peakint, which gave similar results. We desired to average the volumes of equivalent cross-peaks from either side of the diagonal but add the volumes of multiple cross-peaks from diastereotopic groups of protons, and use the resulting value to generate a single restraint. In practice, these two procedures were replaced by the following simple algorithm: the volumes of all cross-peaks corresponding to a given restraint, from both sides of the diagonal and from all signals of diastereotopic groups, were summed; this value was halved if two or more volumes were included in the summation. This gives restraints that are the same as or looser than those produced by the original separate procedures and is much easier to implement. The resulting volumes were used to derive upper limit distance restraints as follows.

The relationship between cross-peak volume and distance was calibrated for each NOESY spectrum using 5–30 well dispersed cross-peaks corresponding to the following distances: in 2D NOESY and ^{13}C NOESY-HSQC, inter-strand $d_{\alpha\alpha}$ in regular antiparallel β-sheet (2.3 Å) and intra-residue aromatic ring $C^{\delta}H_2$–$C^{\varepsilon}H_2$ (2.46 Å; these volumes were halved because they are generated by two equivalent proton pairs); in ^{15}N NOESY-HSQC, $d_{\alpha N}(i,i+1)$ in regular β-sheet (2.3 Å), $d_{NN}(i,i+1)$ in regular α-helix (2.9 Å) and $d_{\alpha N}(i,i+3)$ in regular α-helix (3.6 Å). These pairs of distances (r) and volumes (V) were fitted to a linear relationship $r = cV^{-1/6}$ using Kaleidagraph (Abelbeck Software).

This relationship was used to calculate the volumes corresponding to the desired upper limits of 2.5, 3.0 and 4.0 Å, the latter used in ^{15}N NOESY-HSQC only because of the longer calibration distances available (Barsukov and Lian, 1993). These values were increased by an arbitrary 10% to allow for inaccurate volume measurement and then used to assign an upper limit distance to each restraint.

Upper limit distances for restraints derived from weaker cross-peaks, overlapped cross-peaks and cross-peaks from longer mixing time NOESY spectra were set to 5 Å (Wuthrich, 1986). All lower bounds were set to 0 Å (Hommel *et al*., 1992), with the exceptions described below.

No stereo-specific assignments were obtained. All groups of equivalent or diastereotopic protons were restrained with a single upper limit distance, using r^{-6} summation during the structure calculations to be consistent with the treatment of NOE intensities described above (Fletcher *et al*., 1996). This also allowed ambiguous restraints to be used in cases where a NOESY cross-peak could be assigned to a particular residue but not to a specific proton within that residue, for example restraints to the aromatic ring protons of Phe113 for which only one signal was observed (Nilges, 1995).

Lower limit restraints on $d_{NN}(i,i+1)$ distances were obtained from the 120 ms ^{15}N NOESY-HSQC spectrum, taking the precautions described previously (Neuhaus *et al*., 1992). Slowly exchanging backbone amide protons were identified from a ^{15}N HSQC spectrum recorded within 24 h of replacing the H_2O solvent with D_2O and used to derive hydrogen bond restraints as described previously (Fletcher *et al*., 1994).

Structure calculations

Structures were calculated with X-PLOR 3.851 using simulated annealing in Cartesian coordinate space and starting coordinates with good covalent geometry (Brunger, 1992). Each calculation started from coordinates with different, randomly selected ϕ and ψ angles (M.Nilges, unpublished data). The simulated annealing comprised 24 000 steps of 5 fs at 2000 K and 48 000 cooling steps. Structure calculations included the His tag and all 113 residues of the native sequence.

Initial structures were calculated from distance restraints derived from NOESY cross-peaks that could be assigned using the NMR spectra alone. These structures were refined during several rounds of recalculation by deleting or loosening consistently violated restraints derived from NOESY cross-peaks that were judged likely to be wrongly assigned, overlapped or produced by spin diffusion. This resulted in an ensemble of structures with no restraint violations larger than 0.5 Å and a mean r.m.s. deviation from the mean structure of 0.9 Å for the backbone C and N atoms of residues 29–113. This ensemble was used to assign ambiguous NOESY cross-peaks in further rounds of refinement.

Structure analysis

Coordinate r.m.s. deviations were calculated with Superpose (Diamond, 1992). Searches for similar structures were carried out with DALI and VAST (Holm and Sander, 1993; Madej *et al*., 1995). Figures were created with InsightII (Molecular Simulations Inc.) and Molscript (Kraulis, 1991).

Database depositions

Atomic coordinates and distance restraints are available from the protein databank (http://www.pdb.bnl.gov) under code 2IF1. Chemical shifts are available from the BioMagResBank (http://www.bmrb.wisc.edu) under code 4255.

NMR binding experiments

The oligoribonucleotide GCCACAAUGGCA was synthesized and purified by HPLC and PAGE (Integrated DNA Technologies). This sequence is predicted to form a duplex only below 10°C, and experiments were carried out at 27°C. eIF samples of 0.2 mM in 50 mM NaH_2PO_4/Na_2HPO_4 pH 6.0, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, 0.1 mM NaN_3 or 50 mM NaH_2PO_4/Na_2HPO_4 pH 7.0, 50 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT were used to record ^{15}N HSQC spectra before and after addition of 0.4 mg of RNA (final concentration 0.3 mM). 1D ^{1}H NMR spectra and absorbance at 280 nm confirmed that the RNA had dissolved.

The 49 kDa human eIF5 protein was expressed in *E.coli* with an N-terminal His tag by a method to be described elsewhere (T.V.Pestova, I.Lomakin and C.U.T.Hellen, unpublished data) and purified to 90% in the same way as eIF1. A 0.1 mM sample of ^{15}N-labeled eIF1 in 50 mM NaH_2PO_4/Na_2HPO_4 pH 7.0, 100 mM NaCl, 1 mM DTT, 1 mM EDTA was used to record ^{15}N HSQC spectra before and after addition of unlabeled eIF5 to 0.4 mM in the same buffer, concentrating the mixture to maintain a constant volume. The sample was subsequently filtered and analyzed by SDS–PAGE to confirm that the expected proportion of eIF5 was present.

GST pull-down experiments

The human mitotic checkpoint protein MAD2 (Li and Benezra, 1996) was expressed in *E.coli* with an N-terminal His tag MRGSHHHHHH and purified to 99% by a method to be described elsewhere (X.Luo and G.Wagner, unpublished data). *Escherichia coli* strain BL21(DE3) were transformed with the plasmid vector pGEXp110 (Asano *et al*., 1997) or pGEX-4T-1 (Amersham Pharmacia Biotech) and grown on LB medium. GST–p110 was expressed in 1000 ml cultures grown at 25°C, inducing at OD_{600} = 0.4 with 1 mM IPTG and 0.2 mM PMSF, and harvesting after 30 min. GST was expressed in 10 ml cultures grown at 37°C, inducing at OD_{600} = 0.6 with 1 mM IPTG and harvesting after 2 h.

GST pull-down experiments were carried out at 4°C. Each experiment used cells from 200 ml of GST–p110 culture or 2 ml of GST culture. The cells were lysed by sonication in buffer A (20 mM NaH_2PO_4/Na_2HPO_4 pH 7.3, 150 mM NaCl, 10 mM EDTA, 1 mM DTT, 1 mM PMSF, 2% Triton X-100, one Boehringer Mannheim Complete Protease Inhibitor tablet per 50 ml) and clarified by centrifugation. The lysate was mixed with 150 µl of glutathione–agarose resin (Amersham Pharmacia Biotech), which was washed with 1 ml of buffer A and three times with 1 ml of buffer B (50 mM NaH_2PO_4/Na_2HPO_4 pH 7.0, 100 mM NaCl, 2.5 mM $MgCl_2$, 1 mM EDTA, 1 mM DTT, 1 mM PMSF, 0.1 mM NaN_3, one Boehringer Mannheim Complete Protease Inhibitor tablet per 50 ml). The resin was then mixed for 30 min with 1 ml of buffer B containing 200 µg of eIF1 or MAD2, washed three times with 1 ml of buffer B and resuspended in 50 mM Tris–HCl pH 6.8, 100 mM DTT, 50 mM EDTA, 10% glycerol, 2% SDS, 0.2% bromophenol blue. After boiling for 5 min, 50% of each sample was loaded onto a 14% SDS–polyacrylamide gel.

Acknowledgements

We thank Xuelian Luo for MAD2 and John Hershey for pGEXp110. This work was supported by NIH grant CA68262 (G.W.) and a fellowship from the International Human Frontier Science Program (C.M.F.).

References

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.

Asano,K., Kinzy,T.G., Merrick,W.C. and Hershey,J.W.B. (1997) Conservation and diversity of eukaryotic translation initiation factor eIF3. *J. Biol. Chem.*, **272**, 1101–1109.

Asano,K., Phan,L., Anderson,J. and Hinnebusch,A.G. (1998) Complex formation by all five homologues of mammalian translation initiation factor 3 subunits from yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.*, **273**, 18573–18585.

Barsukov,I.L. and Lian,L.-Y. (1993) Structure determination from NMR data I: analysis of NMR data. In Roberts,G.C.K. (ed.), *NMR of Macromolecules: A Practical Approach*. IRL Press, Oxford, UK, pp. 315–357.

Bartels,C., Xia,T., Billeter,M., Guntert,P. and Wuthrich,K. (1995) The program XEASY for computer-supported NMR spectral analysis of biological molecules. *J. Biomol. NMR*, **6**, 1–10.

Bax,A. and Ikura,M. (1991) An efficient 3D NMR technique for correlating the proton and ^{15}N backbone amide resonances with the α carbon of the preceding residue in uniformly ^{15}N/^{13}C enriched proteins. *J. Biomol. NMR*, **1**, 99–104.

Bax,A., Clore,G.M. and Gronenborn,A.M. (1990) ^{1}H–^{1}H correlation via isotropic mixing of ^{13}C magnetization, a new three dimensional approach for assigning ^{1}H and ^{13}C spectra of ^{13}C enriched proteins. *J. Magn. Reson.*, **88**, 425–431.

Brown-Luedi,M.L., Meyer,L.J., Milburn,S.C., Yau,P.M.P., Corbett,S. and Hershey,J.W.B. (1982) Protein synthesis initiation factors from human HeLa cells and rabbit reticulocytes are similar: comparison of protein structure, activities and immunological properties. *Biochemistry*, **21**, 4202–4206.

Brunger,A.T. (1992) *X-PLOR Version 3.1: A System for Crystallography and NMR*. Yale University Press, New Haven, CT.

Bycroft,M., Grunert,S., Murzin,A.G., Proctor,M. and St. Johnston,D. (1995) NMR solution structure of a dsRNA binding domain from *Drosophila* staufen protein reveals homology to the N-terminal domain of ribosomal protein S5. *EMBO J.*, **14**, 3563–3571.

Castiglione Morelli,M.A., Stier,G., Gibson,T.J., Joseph,C., Musco,G., Pastore,A. and Trave,G. (1995) The KH module has an $\alpha\beta$ fold. *FEBS Lett.*, **358**, 193–198.

Castilho-Valavicius,B., Yoon,H. and Donahue,T.F. (1990) Genetic characterization of the *Saccharomyces cerevisiae* translation initiation suppressors *sui1*, *sui2* and *SUI3* and their effects on *HIS4* expression. *Genetics*, **124**, 483–495.

Cui,Y., Dinman,J.D., Kinzy,T.G. and Peltz,S.W. (1998) The mof2/sui1 protein is a general monitor of translational accuracy. *Mol. Cell. Biol.*, **18**, 1506–1516.

Diamond,R. (1992) On the multiple simultaneous superposition of molecular structures by rigid body transformation. *Protein Sci.*, **1**, 1279–1287.

Farrow,N.A. *et al*. (1994) Backbone dynamics of a free and a phosphopeptide complexed src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry*, **33**, 5984–6003.

Fletcher,C.M., Harrison,R.A., Lachmann,P.J. and Neuhaus,D. (1994) Structure of a soluble, glycosylated form of the human complement regulatory protein CD59. *Structure*, **2**, 185–199.

Fletcher,C.M., Jones,D.M.N., Diamond,R. and Neuhaus,D. (1996) Treatment of NOE constraints involving equivalent or non-stereoassigned protons in calculations of biomacromolecular structures. *J. Biomol. NMR*, **8**, 292–310.

Grzesiek,S. and Bax,A. (1992) Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J. Am. Chem. Soc.*, **114**, 6291–6293.

Grzesiek,S. and Bax,A. (1993) Amino acid type determination in the sequential assignment procedure of uniformly ^{13}C/^{15}N enriched proteins. *J. Biomol. NMR*, **3**, 185–204.

Harding,M.M., Krippner,G.Y., Shelton,C.J., Rodger,A., Sanders,K.J., Mackay,J.P. and Prakash,A.S. (1997) DNA binding studies of XSPTSPSZ, derivatives of the intercalating heptad repeat of RNA polymerase II. *Biopolymers*, **42**, 387–398.

Hoffman,D.W., Query,C.C., Golden,B.L., White,S.W. and Keene,J.D. (1991) RNA binding domain of the A protein component of the U1 small nuclear ribonucleoprotein analyzed by NMR spectroscopy is structurally similar to ribosomal proteins. *Proc. Natl Acad. Sci. USA*, **88**, 2495–2499.

Holm,L. and Sander,C. (1993) Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.*, **233**, 123–138.

Hommel,U., Harvey,T.S., Driscoll,P.C. and Campbell,I.D. (1992) Human epidermal growth factor: high resolution solution structure and comparison with transforming growth factor α. *J. Mol. Biol.*, **227**, 271–282.

Ikura,M., Kay,L.E., Tschudin,R. and Bax,A. (1990) Three dimensional NOESY-HMQC spectroscopy of a ^{13}C labeled protein. *J. Magn. Reson.*, **86**, 204–209.

Jackson,R.J. (1996) A comparative view of initiation site selection mechanisms. In Hershey,J.W.B., Mathews,M.B. and Sonenberg,N. (eds), *Translational Control*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 71–112.

Kay,L.E., Xu,G.Y. and Yamazaki,T. (1994) Enhanced sensitivity triple resonance spectroscopy with minimal H_2O saturation. *J. Magn. Reson.*, **A109**, 129–133.

Kharrat,A., Macias,M.J., Gibson,T.J., Nilges,M. and Pastore,A. (1995) Structure of the dsRNA binding domain of *E.coli* RNase II. *EMBO J.*, **14**, 3572–3584.

Kozak,M. (1989) The scanning model for translation: an update. *J. Cell Biol.*, **108**, 229–241.

Kraulis,P.J. (1991) Molscript—a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.

Kupce,E. and Freeman,R. (1995) Adiabatic pulses for wideband inversion and broadband decoupling. *J. Magn. Reson.*, **A115**, 273–276.

Kyrpides,N.C. and Woese,C.R. (1998) Universally conserved translation initiation factors. *Proc. Natl Acad. Sci. USA*, **95**, 224–228.

Li,Y. and Benezra,R. (1996) Identification of a human mitotic checkpoint gene: *hsMAD2*. *Science*, **274**, 246–248.

Lindahl,M. *et al*. (1994) Crystal structure of the ribosomal protein S6 from *Thermus thermophilus*. *EMBO J.*, **13**, 1249–1254.

Madej,T., Gibrat,J.-F. and Bryant,S.H. (1995) Threading a database of protein cores. *Proteins*, **23**, 356–369.

Matsuo,H., Li,H. and Wagner,G. (1996) A sensitive HN(CA)CO experiment for deuterated proteins. *J. Magn. Reson.*, **B110**, 112–115.

Merrick,W.C. and Hershey,J.W.B. (1996) The pathway and mechanism of eukaryotic protein synthesis. In Hershey,J.W.B., Mathews,M.B. and Sonenberg,N. (eds), *Translational Control*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 31–69.

Musco,G., Stier,G., Joseph,C., Castiglione Morelli,M.A., Nilges,M., Gibson,T.J. and Pastore,A. (1996) Three dimensional structure and stability of the KH domain: molecular insights into the fragile X syndrome. *Cell*, **85**, 237–245.

Nanduri,S., Carpick,B.W., Yang,Y., Williams,B.R. and Qin,J. (1998) Structure of the double stranded RNA binding domain of the protein kinase PKR reveals the molecular basis of its dsRNA mediated activation. *EMBO J.*, **17**, 5458–5465.

Naranda,T., MacMillan,S.E. and Hershey,J.W.B. (1994) Purified yeast translational initiation factor eIF3 is an RNA binding protein complex that contains the PRT1 protein. *J. Biol. Chem.*, **269**, 32286–32292.

Naranda,T., MacMillan,S.E., Donahue,T.F. and Hershey,J.W.B. (1996) SUI1/p16 is required for the activity of eukaryotic translation initiation factor 3 in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **16**, 2307–2313.

Neri,D., Billeter,M., Wider,G. and Wuthrich,K. (1992) NMR determination of residual structure in a urea denatured protein, the 434 repressor. *Science*, **257**, 1559–1563.

Neuhaus,D., Nakaseko,Y., Schwabe,J.W.R. and Klug,A. (1992) Solution structure of two zinc-finger domains from SWI5 obtained using two-dimensional ^{1}H nuclear magnetic resonance spectroscopy: a zinc-finger with a third strand of β-sheet. *J. Mol. Biol.*, **228**, 637–651.

Nilges,M. (1995) Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. *J. Mol. Biol.*, **245**, 645–660.

Oubridge,C., Ito,N., Evans,P.R., Teo,C.H. and Nagai,K. (1994) Crystal structure at 1.92 Å resolution of the RNA binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature*, **372**, 432–438.

Pestova,T.V., Borukhov,S.I. and Hellen,C.U.T. (1998) Eukaryotic ribosomes require eIFs 1 and 1A to locate initiation codons. *Nature*, **394**, 854–859.

Phan,L., Zhang,X., Asano,K., Anderson,J., Vornlocher,H.-P., Greenberg,J.R., Qin,J. and Hinnebusch,A.G. (1998) Identification of a translation initiation factor 3 (eIF3) core complex, conserved in yeast and mammals, that interacts with eIF5. *Mol. Cell. Biol.*, **18**, 4935–4946.

Ramakrishnan,V. and White,S.W. (1992) The structure of ribosomal protein S5 reveals sites of interaction with 16S rRNA. *Nature*, **358**, 768–771.

Richardson,J.S. (1981) The anatomy and taxonomy of protein structure. *Adv. Protein Chem.*, **34**, 167–339.

Ryter,J.M. and Schultz,S.C. (1998) Molecular basis of double stranded RNA–protein interactions: structure of a dsRNA binding domain complexed with dsRNA. *EMBO J.*, **17**, 7505–7513.

Schreier,M.H., Erni,B. and Staehelin,T. (1997) Initiation of mammalian protein synthesis. I. Purification and characterization of seven initiation factors. *J. Mol. Biol.*, **116**, 727–753.

Talluri,S. and Wagner,G. (1996) An optimized 3D NOESY-HSQC. *J. Magn. Reson.*, **A112**, 200–205.

Trachsel,H., Erni,B., Schreier,M.H. and Staehelin,T. (1977) Initiation of mammalian protein synthesis. II. The assembly of the initiation complex with purified initiation factors. *J. Mol. Biol.*, **116**, 755–768.

Wishart,D.S., Bigam,C., Holm,A., Hodges,R.S. and Sykes,B.D. (1995) ^{1}H, ^{13}C and ^{15}N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest neighbour effects. *J. Biomol. NMR*, **5**, 67–81.

Wulfing,C., Lombardero,J. and Pluckthun,A. (1994) An *Escherichia coli* protein consisting of a domain homologous to FK506-binding proteins (FKBP) and a new metal binding motif. *J. Biol. Chem.*, **269**, 2895–2901.

Wuthrich,K. (1986) *NMR of Proteins and Nucleic Acids*. John Wiley, New York, NY.

Yamazaki,T., Lee,W., Revington,M., Mattiello,D.L., Dahlquist,F.W., Arrowsmith,C.H. and Kay,L.E. (1994) An HNCA pulse scheme for the backbone assignment of ^{15}N, ^{13}C, ^{2}H labeled proteins: application to a 37 kDa Trp repressor DNA complex. *J. Am. Chem. Soc.*, **116**, 6464–6465.

Yoon,H. and Donahue,T.F. (1992) The *sui1* supressor locus in *Saccharomyces cerevisiae* encodes a translation factor that functions during $tRNA_i^{Met}$ recognition of the start codon. *Mol. Cell. Biol.*, **12**, 248–260.

Zhang,O., Kay,L.E., Olivier,J.P. and Forman-Kay,J.D. (1994) Backbone ^{1}H and ^{15}N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced sensitivity pulsed field gradient NMR techniques. *J. Biomol. NMR*, **4**, 845–858.

Zuiderweg,E.R.P., Hallenga,K. and Olejniczak,E.T. (1986) Improvement of 2D NOE spectra of biomolecules in H_2O solution by coherent suppression of the solvent resonance. *J. Magn. Reson.*, **70**, 336–343.

Received February 4, 1999; revised and accepted March 2, 1999