X-ray crystallography shows that translational initiation factor IF3 consists of two compact α/β domains linked by an α -helix

Valerie Biou¹, Fong Shu and V.Ramakrishnan²

Biology Department, Brookhaven National Laboratory, Upton, NY 11973, USA

¹Present address: European Synchrotron Radiation Facility, Experiments division/Diffraction group BP220, F-38043 Grenoble cedex, France

²Corresponding author

The structures of the two domains of translational initiation factor IF3 from *Bacillus stearothermophilus* have been solved by X-ray crystallography using single wavelength anomalous scattering and multiwavelength anomalous diffraction. Each of the two domains has an α/β topology, with an exposed β -sheet that is reminiscent of several ribosomal and other RNA binding proteins. An α -helix that protrudes out from the body of the N-terminal domain towards the C-terminal domain suggests that IF3 consists of two RNA binding domains connected by an α -helix and that it may bridge two regions of the ribosome. This represents the first high resolution structural information on a translational initiation factor.

Keywords: anomalous scattering/multiwavelength anomalous dispersion/ribosome/selenomethionine/translation

Introduction

Protein synthesis can be divided into roughly three stages: initiation, in which the ribosome is assembled at the start of the coding sequence on mRNA and aminoacylated initiator tRNA is bound to it; elongation, which involves both the synthesis of a peptide bond and movement of the ribosome relative to the mRNA; termination, in which the recognition of a stop codon triggers dissociation of the polypeptide from the ribosome and of the ribosomal subunits from mRNA.

In prokaryotes the first stage, initiation, involves recognition of the initiation site on mRNA by the ribosome and binding of the initiator fmet-tRNA to the P site of the ribosome. This process is facilitated by three initiation factors, IF1, IF2 and IF3, and is accompanied by the hydrolysis of GTP (reviewed by Gualerzi and Pon, 1990).

The role of initiation factor IF1 remains vague. However, it has been shown recently that it is required for the viability of *Escherichia coli* and that cells depleted of IF1 exhibit few polysomes (Cummings and Hershey, 1994). IF2 is a GTPase involved in the binding of fmet-tRNA to the 30S ribosomal subunit (reviewed by Gualerzi and Pon, 1990). The hydrolysis of GTP is linked to ribosome binding and it is thought to be similar to other GTPases, such as elongation factor EF-Tu.

The third initiation factor, IF3, was originally thought

to be a simple dissociation factor, which, by binding to free 30S ribosomes, prevents the formation of 70S ribosomes in the absence of message and thereby kinetically favors the formation of the 30S–mRNA initiation complex (Subramanian and Davis, 1970; Godefroy-Colburn *et al.*, 1975). Subsequent work showed additional roles for IF3. Toeprinting experiments suggested that IF3 is involved, along with IF2, in selection of initiator tRNA and recognizes the anticodon stem-loop of tRNA (Hartz *et al.*, 1989, 1990). Binding studies using fluorescence assays showed that the affinity of IF2 for 30S ribosomal subunits is increased ~3-fold by IF1 and IF3 (Weiel and Hershey, 1982), while the affinity of IF1 for the 30S ribosome is increased ~50-fold in the presence of IF3 (Zucker and Hershey, 1986). However, IF3 binding is not affected by the other two factors (Weiel and Hershey, 1981) and binds with an association constant $>10^7/M$. Previous reports that IF3 promoted the binding of mRNA to the 30S ribosome are contradicted in a recent study using defined oligonucleotides containing the initiation site (Laughrea and Tam, 1989). Thus IF3 binds to free 30S ribosomes, promotes formation of the initiation complex by increasing the affinity of the other factors for the 30S ribosome and is involved in proofreading of the initiation complex.

Where does IF3 bind on the ribosome? IF3 is the only one of the three initiation factors to cross-link efficiently to rRNA. Experiments show that IF3 cross-links primarily to nucleotides (nt) 819–859 in the central domain of 16S rRNA, with a minor cross-link to the region 1506–1529 in the 3'-end (Ehresmann *et al.*, 1986). Recent footprinting experiments showed that IF3 protects residues 700, 703 and 791 from attack by kethoxal (Moazed *et al.*, 1995). It has been UV cross-linked to ribosomal proteins S7, S11, S12, S18 and S21 (MacKeen *et al.*, 1980). It has been shown to bind specifically to a 13 nt sequence that is part of the 3'-terminal cloacin fragment of 16 S rRNA (Wickstrom, 1983). Chemical protection experiments are consistent with IF3 interacting with the central domain (Muralikrishna and Wickstrom, 1989). These biochemical data suggest that IF3 interacts with a region of the 30S ribosome that includes part of the platform, cleft and the head, facing the 50S subunit. These results also agree with the localization observed by immunoelectron microscopy of 30S subunits cross-linked to IF3 (Stöffler *et al.*, 1980).

Little is known about the structure of IF3 and further progress on understanding its role in initiation will require a greater understanding of its three-dimensional structure. In *Bacillus stearothermophilus* IF3 consists of 172 residues with a molecular weight of 19.678 kDa and other prokaryotic IF3s are similar in size. It has been shown that IF3 can be cleaved proteolytically into two halves, each of which is folded as a domain (Fortier *et al.*, 1994; Kycia *et al.*, 1995). In our previous work (Kycia *et al.*, 1995)

Table I. Data collection

	λ (Å)	f'	f''	Resolution (Å)	Completeness		Multiplicity		R_{sym}^a		$R_{\text{anomalous}}^b$		$I/\sigma(I)$	
					Total	Outer shell	Total	Outer shell	Total	Outer shell	Total	Outer shell	Total	Outer shell
Trimethyl lead acetate IF3-N	0.9500	-12.0	10.0	1.8	95.9	78.0	7.2	3.5	3.9	8.4	7.6	11.9	10.9	8.2
Selenomethionyl IF3-C, λ_1	0.9802	-9.52	3.15	2.0	94.2	88.6	3.8	2.9	4.2	9.0	3.6	7.5	12.8	7.0
Selenomethionyl IF3-C, λ_2	0.9795	-7.35	5.92	2.0	94.4	89.3	3.8	3.0	4.4	10.9	4.9	8.7	12.1	6.1
Selenomethionyl IF3-C, λ_3	0.9300	-2.19	3.46	2.0	94.0	94.0	3.9	3.9	4.8	13.6	3.8	9.1	7.2	4.4
Native IF3-C	1.00	-	-	2.0	99.6	99.6	2.9	2.5	4.2	9.1	-	-	7.8	4.7

$$^a R_{\text{sym}} = \Sigma[(I - \langle I \rangle) / \langle I \rangle].$$

$$^b R_{\text{anomalous}} = \Sigma[(\langle I+ \rangle - \langle I- \rangle) / (\langle I+ \rangle + \langle I- \rangle)].$$

we showed that the *E. coli* outer membrane protease omp_{tin} cleaves IF3 between Lys79 and Gln80. Both the N- and C-terminal fragments resulting from this cleavage, designated IF3-N and IF3-C, could be expressed and crystallized separately. We also showed by small angle scattering that IF3 is an elongated protein, with its two domains separated by ~45 Å. Finally, circular dichroism spectra supported the notion that IF3-N and IF3-C have the same structure when expressed separately as they do *in situ* as part of intact IF3. In this paper we describe the crystal structures of IF3-N and IF3-C. We discuss how these domains could be oriented relative to each other within the constraints of small angle scattering and the biochemical data and the implications of the structures for ribosome binding.

Results

Crystallography

Both structures were solved using the tunability of synchrotron radiation to obtain phase information exclusively from anomalous scattering; no native crystals were used in the initial phase determination of either domain. A summary of the crystallography is described in Tables I–III.

IF3-N. A trimethyl lead acetate derivative of IF3-N was not isomorphous with the native molecule and hence isomorphous differences could not be used. Nevertheless, the structure of the N-terminal domain was solved using just the Bijvoet differences of the derivative, for which data were collected at the 'white line' of the LIII transition edge of lead. At this wavelength the imaginary part f'' of the anomalous scattering component is maximized, leading to the largest possible Bijvoet differences. This resulted in a peak height of 13 standard deviations in the anomalous Patterson map (Figure 1a). The phase ambiguity from this single source of phase information was broken by solvent flattening (Wang, 1985) to give a clearly interpretable map (Figure 1b). As we had much better data from the lead derivative than the native molecule, all refinement was done on the derivative.

IF3-C. The structure of IF3-C was solved using multi-wavelength anomalous dispersion (MAD) on crystals of the selenomethionyl form of the protein (Hendrickson, 1991). Data were collected at three wavelengths around the K edge of selenium, corresponding to the minimum of f' , the maximum of f'' and a remote point, where f' and f'' denote the real and imaginary parts of the anomalous

Table II. Phasing

	R_{cullis}	R_{kraut}	$\langle \text{FOM} \rangle$	Phasing power
IF3-N, trimethyl lead anomalous	–	–	0.50	2.46
Selenomethionyl IF3-C:				
$\lambda_3 - \lambda_1$ isomorphous	0.496	0.035	0.451	2.60
λ_1 anomalous	–	0.019	0.276	1.09
λ_2 anomalous	–	0.036	0.430	2.44
λ_3 anomalous	–	0.042	0.297	1.20

$R_{\text{cullis}} = (\Sigma(|\text{FPH}| \pm |\text{FP}|) - |\text{FH}|) / (\Sigma|\text{FPH}| - |\text{FP}|)$, where FP, FPH and FH are respectively the observed protein, observed derivative and calculated heavy atom structure factors, with the sum over all centric reflections.

$R_{\text{kraut}} = [|\Sigma|\text{FPH}(\text{obs}) - |\text{FPH}(\text{calc})||] / [|\Sigma|\text{FPH}(\text{obs})|]$, with the sum over all acentric reflections.

Phasing power = $\langle \text{FH} \rangle / E_{\text{rms}}$, where E_{rms} is the r.m.s. residual lack of closure.

scattering factor of selenium. A clear peak showing one of the two seleniums was seen in both the anomalous and dispersive Patterson maps (Figure 1c). Phasing was done by treating the multiwavelength data as a special case of multiple isomorphous replacement with anomalous scattering (Ramakrishnan *et al.*, 1993). The λ_1 data (corresponding to the minimum of f') were treated as 'native' and data at the other two wavelengths as derivatives. Table II shows that, as expected, the larger the anomalous signal from f'' , the larger the phasing power of that data set. A similar relationship holds for the dispersive difference. The initial map was solvent flattened to give a map that was easily interpretable (Figure 1d). Subsequent refinement was done using data from a crystal that did not have selenium incorporated.

The results of refinement of IF3-N and IF3-C are shown in Table III. In each case the final refined structure showed no residues outside allowed regions in the respective Ramachandran plots. The unusually low free R factor for IF3-N was a result of including anomalous differences in the refinement, which probably imposed additional constraints, as well as better accounted for the scattering, resulting in a better model.

Structures of the domains of IF3

Figure 2 shows the alignment of the sequences of IF3 from various prokaryotic species. It shows the assignment of secondary structure to the sequence, based on the crystal structures determined in this work. The designations of the various secondary structure elements in Figure 2

Table III. Refinement

	Resolution (Å)	N_{atoms}	N_{solvent}	R	R_{free}	r.m.s. deviation from ideal geometry		
						Bond length (Å)	Bond angles (°)	Dihedral angles (°)
IF3-N trimethyl lead	6.0–1.8	628	44	0.192	0.233	0.006	1.32	23.8
IF3-C	6.0–2.0	702	40	0.202	0.270	0.012	1.60	23.8

are used in the following description of the structures of IF3-N and IF3-C.

The fold of IF3-N is shown in Figure 3. The structure contains a globular α/β domain consisting of a single α -helix, H1, packed against a mixed four-stranded sheet in which the central strands S1 and S4 are parallel to each other and have antiparallel strands S2 and S3 on the outside. Additionally, the N-terminus of IF3-N also makes some hydrogen bonds with the sheet and extends it partially. The outer strands S2 and S3 of the sheet are highly irregular. While they are characterized as part of the β -sheet based on hydrogen bonding, they were not identified as β -strands by algorithms based on template fitting, such as the YASSPA procedure in the program O (Jones *et al.*, 1991). At the C-terminus of IF3-N a helix H2 protrudes out from the body of the globular α/β domain.

The C-terminal domain also consists of an α/β fold; it consists of two parallel helices, H3 and H4, packed against a mixed four-stranded sheet. In this sheet the first two strands, S5 and S6, are parallel; the third strand, S8, is antiparallel to both S6 and the fourth strand S7. Two views of the structure of IF3-C are shown in Figure 4. As in the case of IF3-N, the outer strands, S5 and S7 in this case, are irregular. Further, although IF3-N and IF3-C each have a mixed four-stranded sheet, the topology of the sheet is not the same in the two domains.

The most striking thing about the structures is the protruding helix H2, comprising residues 62–75 at the C-terminus of IF3-N. For several reasons we believe that the conformation of this helix, especially its protrusion from the rest of the domain, is representative of its conformation in intact IF3. First, the helix is almost entirely hydrophilic and it is unlikely that it could pack against the globular domain of IF3-N. Moreover, the helix makes few crystal contacts except at its C-terminus, thus we do not think the helix is induced by crystallization, although it may be stabilized by the contacts at the end of the helix and the high ammonium sulfate concentration in the crystal. The highly conserved Tyr62 anchors the base of the helix to the core of IF3-N, so we believe the orientation of the helix with respect to the core is relatively fixed. This may have allowed crystallization in the first place. Finally, the end of this helix at residue 75 is very close to the ompT protease cleavage site, which defines the junction between IF3-N and IF3-C (Kycia *et al.*, 1995). It follows that the end of helix H2 must be spatially close to the N-terminus of IF3-C and H2 must connect IF3-C with the globular domain of IF3-N. If one accordingly places the IF3-C structure so that its N-terminus is close to the end of helix H2 on IF3-N, one arrives at a model in which the two domains are separated by ~45 Å. This distance is entirely consistent with small angle scattering data (Kycia *et al.*, 1995), which showed that

IF3 was an elongated protein consisting of two domains separated by this distance.

Thus we conclude that IF3 consists of two relatively compact α/β domains connected by helix H2. A model of this is shown in Figure 5. Residues have been color coded to show the degree of conservation at each position, with blue the most conserved and red the least conserved.

Discussion

Each domain of IF3 has an exposed β -sheet as part of an α/β topology. This motif is reminiscent of the structure of several ribosomal proteins (Wilson *et al.*, 1986; Leijonmarck *et al.*, 1988; Ramakrishnan and White, 1992; Golden *et al.*, 1993) and the RNA binding protein U1A (Nagai *et al.*, 1990; Hoffmann *et al.*, 1991). The topology of the IF3 domains is not the same as in these proteins. Nevertheless, it raises the possibility that each of these domains is involved in binding to the 30S ribosome and, in particular, to 16S rRNA. IF3 is known to bind near the cleft between the 'head' and 'platform' regions. The crystal structures of the two domains suggest that IF3 can span the cleft, with each compact domain binding on either side of the cleft.

There is little information on which residues in IF3 are important for ribosome binding. Reaction with pyridoxal phosphate resulted in modification of lysines at positions 2, 5, 97, 110 and 165 (numbering refers to the *E.coli* sequence). Modification of just Lys110 results in inactivation of IF3 (Ohsawa and Gualerzi, 1981). Of these residues, Lys2 and Lys5 at the N-terminus appear not to have a counterpart in *B.stearothermophilus* IF3 (see Figure 2). The first eight residues of the *E.coli* sequence appear to be highly conserved among mesophilic bacteria, but are deleted in the sequence of *B.stearothermophilus*. It is likely that the thermophilic bacterium has dispensed with the interactions that these residues are involved in.

The other three lysines, corresponding to Arg89, Lys102 and Lys55 respectively in the *B.stearothermophilus* sequence, lie on the same face of the protein, namely that presented by the exposed β -sheet of IF3-C, although Arg89 is actually part of helix H3 (see Figure 3b). This is consistent with the idea that the exposed β -sheet is involved in binding to rRNA. The crystal structure of the complex of U1A with its cognate RNA showed that the RNA was bound to the exposed sheet (Oubridge *et al.*, 1994). Additional evidence for the role of Lys110 (Arg89 in the *B.stearothermophilus* sequence) was found by site-directed mutagenesis (De Bellis *et al.*, 1992). It was found that Lys110 was important for both ribosome binding and proper tRNA selection. This would again be consistent with this face of IF3-C being involved in rRNA binding.

Site-directed mutagenesis also found evidence for the

Fig. 1. Crystallography of the IF3 domains. (a) Harker sections of the anomalous only Patterson map of IF3-N complexed with trimethyl lead acetate. The map shows a strong single peak 13σ above background. (b) Solvent flattened map based on the single anomalous scattering from lead, showing the lead position and neighboring residues. (c) Harker section of Patterson maps generated from the λ_2 anomalous differences (top) and the $\lambda_1 - \lambda_3$ dispersive differences (bottom). Both maps show a strong peak from one of the two selenium atoms in the asymmetric unit. The other peak, barely visible in the anomalous map, was determined by the program HASSP and confirmed in difference Fourier maps. (d) Solvent flattened map showing a salt bridge. The initial map was derived from the MAD experiment by treating MAD as a special case of multiple isomorphous replacement with the inclusion of anomalous scattering.

structure and hence could not possibly make direct contact with a part of the ribosome. However, it makes stacking interactions with the guanidinium plane of Arg139, which itself forms a salt bridge with Glu96 (Figure 4b). It is likely that mutation of the planar aromatic residue to a leucine results in a loss of the stacking interaction resulting in a destabilization of the salt bridge. This in turn may perturb the structure enough to affect binding indirectly. It has been shown that IF3 promotes the binding of IF2

Fig. 4. Stereo pairs showing two views of the structure of IF3-C, the C-terminal half of initiation factor IF3. Residues that have been implicated in ribosome binding are shown. The drawing was made with MOLSCRIPT (Kraulis, 1991).

to the ribosome (Weiel and Hershey, 1982); IF2 is known to be a GTPase that belongs to the same family as Ef-Tu and the *ras* oncogene product. There appears to be a topological similarity between the core of IF3-N and the *ras* binding domain of Raf-1 (Emerson *et al.*, 1994). Both proteins consist of a mixed sheet packed against a helix with identical topology, but in Raf-1 there is an additional strand inserted between the equivalent of S3 and S4 that forms the outer member of the sheet. It is tempting to speculate that some features of protein-protein interactions with the GTPase family of proteins have been preserved. This should become more apparent when the three-dimensional structure of Raf-1 is determined.

It had been known that IF3 could be cross-linked to proteins that are fairly distant in the 30S ribosome. In particular, it can be cross-linked to proteins S7 and S11 (MacKeen *et al.*, 1980). These two ribosomal proteins are on opposite sides of the cleft and have a center-of-mass separation of ~ 70 Å (Ramakrishnan *et al.*, 1981). The structures of IF3-N and IF3-C show how such interactions could be accommodated: each domain of IF3 could interact with the ribosome on either side of the cleft, with the connecting helix H2 spanning the cleft and bridging the two domains. The general architecture of two RNA binding domains connected by a helix has been seen before in the structure of ribosomal protein L9 (Hoffmann *et al.*, 1994),

but the topology of the domains is different in the two cases.

Recently, a fold of the N-terminal domain of IF3 from *E. coli* was determined by NMR (Garcia *et al.*, 1995). At first sight, the NMR structure appears to have a different topology from that of IF3-N described here: it is described as consisting of a five stranded antiparallel sheet instead of the mixed sheet observed in our crystal structure. In fact, a comparison of the NMR coordinates with the crystal structure shows that the topologies of the two structures are very similar. A slightly different orientation of strands S1 and S4 in the NMR structure led to their being characterized as antiparallel by Garcia *et al.* (1995). However, the hydrogen-bonding pattern between S1 and S4 in the crystal structure shows that they are parallel strands and are part of a mixed sheet. Also, we have not characterized residues 6–8 as an additional beta strand, although they make two hydrogen bonds with S3. Apart from these, the main differences between the two structures are a slightly different orientation for helix 1, for which there are not many NOEs; a different conformation of the loop between S3 and S4; and an unravelling of the end of helix H2 in the NMR structure. As the interdomain distance between IF3-N and IF3-C measured by neutron scattering agrees rather well with a model assuming that helix H2 is the connecting helix, we feel that helix H2 in

Fig. 5. A model for the relative disposition of the two halves of IF3. The separation of 45 Å is consistent with neutron scattering data (Kycia *et al.*, 1995). The colors show the variation in sequence at each position based on the alignment of sequences in Figure 2. The drawing was made using the program O (Jones *et al.*, 1991).

the crystal structure of IF3-N is representative of its structure in intact IF3.

It is clear that to arrive at a more detailed understanding of the interaction of IF3 with the ribosome, further work needs to be done. Although the structures of the separate domains show that IF3 consists of two compact α/β domains connected by a helix, the precise orientation of these domains relative to one another is uncertain. We have suggested that the orientation of helix H2 with respect to the core of IF3-N may be relatively fixed, thus allowing crystallization of IF3-N. However, in intact IF3 the junction between helix H2 and IF3-C may be flexible and exposed and this may be why protease cleavage by omptin occurs at this site. This possible flexibility of IF3-C with respect to the rest of the protein may be the reason that intact IF3 has resisted crystallization so far (Gualerzi and Pon, 1990; Kycia *et al.*, 1995). If such crystals are obtained in the future, it should be easy to solve its structure by molecular replacement using the domain structures. Even crystals that diffract to low resolution will suffice for the purpose of determining the relative disposition of the two domains. On the other hand, if flexibility of the C-terminal domain with respect to the central helix is inherent, then it may be difficult to define the structure of the intact protein to any greater precision than is done in the model proposed here.

The structures of the domains should make possible

advances in our understanding of ribosome-IF3 interactions. So far, cross-linking experiments using intact IF3 have resulted in a rather large area on the ribosome being tagged and it is difficult to see exactly how this structure relates to the cross-linking data in its entirety. We have postulated that each of the two domains is involved in ribosome binding. This can be tested by footprinting and cross-linking experiments, using intact IF3 as well as the individual domains. Footprinting and RNA binding studies using individual domains could help determine which domain interacts with which region of the ribosome, thus allowing a more detailed picture to emerge. The structures, by showing the location of various residues, also provide a more rational basis for future site-directed mutagenesis.

Materials and methods

Crystals of the N-terminal and C-terminal domain of IF3, referred to as IF3-N and IF3-C, were grown as described previously (Kycia *et al.*, 1995). Crystals of IF3-N were obtained in 2.7 M ammonium sulfate, 0.1 M Tris-HCl, pH 7.4. These crystals were orthorhombic, with $a = 25.3$, $b = 39.2$ and $c = 63.3$ Å, with one molecule per asymmetric unit and an estimated solvent content of 30%. Crystals of IF3-C were obtained in 2.4 M phosphate, pH 8. The crystals are in space group C2, with cell dimensions of $a = 72.5$, $b = 30.2$ and $c = 43.4$ Å and $\beta = 100.11^\circ$, with one molecule per asymmetric unit and an estimated solvent content of 35%. Selenomethionyl IF3-C was obtained by expression of the gene in the *E. coli* strain B834 (DE3), which is a methionine-requiring auxotroph with an inducible gene for T7 RNA polymerase on the host

chromosome (Studier and Moffatt, 1986). The cells were grown in a synthetic medium containing selenomethionine, along with the other 19 amino acids. Crystals of selenomethionyl IF3-C were obtained by seeding a drop of the selenomethionyl protein with crystals of native IF3-C and had the same cell dimensions as the native crystals.

Data collection

All data were collected at room temperature on beamline X12C at the National Synchrotron Light Source at Brookhaven National Laboratory. The beamline is equipped with a 30 cm MAR image plate detector coupled to a diffractometer equipped with a κ goniostat. For our experiments we used a single horizontal rotation axis, ϕ , perpendicular to the X-ray beam.

For IF3-N data were collected on a single crystal soaked for 24 h in 4 mM trimethyl lead acetate in stabilizing buffer and 200° of data were collected in 2° frames. The wavelength of 0.95 Å was chosen to be at the 'white line' of the LIII transition edge of lead, which corresponds to the maximum of f'' , the imaginary part of the anomalous scattering factor of lead.

For selenomethionyl IF3-C MAD data were collected at three wavelengths, λ_1 , λ_2 and λ_3 . The first two were around the K edge of selenium (0.9798 Å) and corresponded to the minimum of f' and the maximum of f'' respectively, where f' and f'' are the real and imaginary part of the anomalous scattering factor of selenium (Hendrickson, 1991). The third wavelength of 0.93 Å was remote from the K edge. Because the crystal morphology made it difficult to rotate about the b axis and thereby obtain Friedel mates on the same image, Bijvoet pairs were measured by inverse beam geometry. This was done by collecting data in blocks of 10 or 12° and then collecting the same block 180° away in ϕ , before the wavelength was changed and the process repeated. Before collecting each block of data, the wavelength was calibrated by measuring the fluorescence spectrum at right angles to the beam in the horizontal direction. Each image corresponded to a 2 or 3° rotation and two crystals were used to collect the entire three wavelength data set. Data collection was done at 4°C.

Data on native IF3-C were collected in a straightforward manner from a single crystal, with 200° of data collected in a single sweep with 2° rotations per image.

Data reduction and scaling

Data were integrated using the program MOSFLM (Leslie, 1992). Subsequent sorting, scaling and reduction were done using the CCP4 suite of programs (Collaborative Computing Project Number 4, 1994), with scaling done using the programs ROTAVATA and AGROVATA.

For MAD data on IF3-C scaling was done in the following manner. Raw intensities for all three wavelengths were included in the same file and were scaled together using ROTAVATA. A scale and B factor for each image were determined using data in the range 6.0–2.0 Å, where 2.0 Å was the limit of usable data. The scale and B factors were then applied to all the data for each wavelength in AGROVATA, so that merged, scaled data were output separately for each wavelength. This put all three wavelengths on a common scale. Data were then reduced to structure factors and placed on an absolute scale using the program TRUNCATE. To place the data on an absolute scale this program applied slightly different scale factors to the data at the various wavelengths, based on a Wilson fit to the data between 3.0 and 2.0 Å. To compensate for this, the structure factors for the λ_2 and λ_3 data were rescaled to the λ_1 data using the program SCALEIT.

Phasing

Phasing for IF3-N was done using the anomalous scattering from lead. An initial estimate of the lead position was obtained from the peak positions in an anomalous Patterson map. The largest 25% of the Bijvoet differences were used to refine the lead position, occupancy and temperature factor, using the program GREF in the PHASES package (Furey and Swaminathan, 1990). This was then used to estimate the phases of the acentric reflections using single anomalous scattering (SAS) and an initial map was calculated using the best (centroid weighted) phases. In this map the phase ambiguity is not resolved, because there is only a single source of phase information. However, such a map consists of true features superimposed on noise (Wang, 1985). The SAS map was solvent flattened using the semi-automated procedure provided in the PHASES suite of programs. Despite the low solvent content of 30%, this was sufficient to both resolve the phase ambiguity of the acentric reflections and estimate the phases of the centric reflections by phase extension. The final solvent flattened map was easily interpretable (see Figure 1b).

MAD phasing for IF3-C was done by treating the data as a special

case of multiple isomorphous replacement: the λ_1 data was considered as the reference, or 'native', and the other two wavelengths as derivatives, as described previously (Ramakrishnan *et al.*, 1993). For the i th wavelength the dispersive term $f'_i - f'_1$ is the source of an isomorphous difference between the wavelengths and the anomalous term f''_i is the source of Bijvoet differences. Therefore, the real and imaginary parts of the scattering factors for the derivative at λ_i were set to $f'_i - f'_1$ and f''_i respectively. The isomorphous difference $f'_2 - f'_1$ was small and did not contribute significant phasing power. This contribution was excluded from the phasing. Thus there were a total of four sources of phase information, the isomorphous difference between λ_1 and λ_3 and the Bijvoet differences within each of the three data sets. The actual values of f' and f'' used for each wavelength were those determined previously for selenomethionyl GH5 from the X-ray fluorescence spectrum around the K edge of selenium (Ramakrishnan *et al.*, 1993). The spectrum of selenomethionyl IF3-C was very similar. However, as the real and anomalous occupancies of the selenium atoms are refined separately, any absolute error in f' and f'' is merely taken up in the apparent occupancies. The first selenium atom was clearly visible in a Harker section of the λ_2 anomalous Patterson map; its position was obtained from the position of this peak in the map. The second position was determined from the λ_2 anomalous Patterson map using the program HASSP (Terwilliger *et al.*, 1987). MIR phasing was done using the PHASES program (Furey and Swaminathan, 1990). Heavy atom parameters were refined using least squares phase refinement against those reflections that had a figure of merit >0.6. The contribution to the protein phase probability distribution from each data set was excluded when parameters for that data set were being refined. The final combined figure of merit estimated by the program was 0.6. A good map was further improved by solvent flattening (Figure 1d) and this was easily interpreted to solve the structure. The use of maximum likelihood, either as incorporated in the PHASES program or in the program MLPHARE (Otwinowski, 1991), did not result in improved maps. In fact, these maps, while also good, had a slightly worse correlation coefficient with the final refined structure than the maps obtained by the procedure described here (Biou and Ramakrishnan, unpublished results).

Model building and refinement

All model building was done using the program O (Jones *et al.*, 1991). In both domains the entire chain was traced and the side chains built in the first round, with the exception of a few residues at the termini. Refinement was done using standard protocols in the program Xplor (Brünger, 1993), which involved positional refinement, simulated annealing and temperature factor refinement. For IF3-N we refined the complex with lead. As the data had large Bijvoet differences, the anomalous scattering term was included in the atomic scattering factor for lead. The occupancies of the lead and water molecules were allowed to vary. Ten percent of the reflections for IF3-C and 5% for IF3-N were set aside for free R factor calculations prior to simulated annealing and temperature factor refinement (Brünger, 1992). The refinement used all measured data between 6 Å and the limit of resolution (1.8 Å for IF3-N and 2.0 Å for IF3-C).

Acknowledgements

We thank R.M.Sweet and V.Stojanoff for help with data collection on beamline X12C at the NSLS at Brookhaven, J.H.Kycia for protein purification and crystallization and J.Flanagan and A.Brünger for useful comments. We thank Dr F.Dardel for communicating the NMR coordinates of *E.coli* IF3-N and discussing the differences with our crystal structure. This work was supported by the Office of Health and Environmental Research of the United States Department of Energy and by grant GM 44973 (to Stephen W.White, Duke University, and V.R.) from the NIH.

References

- Brünger, A.T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–475.
- Brünger, A.T. (1993) X-PLOR, Version 3.1. Yale University.
- Collaborative Computational Project Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr.*, **D50**, 760–763.
- Cummings, H.S. and Hershey, J.W. (1994) Translation initiation factor IF1 is essential for cell viability in *Escherichia coli*. *J. Bacteriol.*, **176**, 198–205.
- De Bellis, D., Liveris, D., Goss, D., Ringquist, S. and Schwartz, I. (1992)

- Structure-function analysis of *Escherichia coli* translation initiation factor IF3: tyrosine 107 and lysine 110 are required for ribosome binding. *Biochemistry*, **31**, 11984–11990.
- Ehresmann, C., Moine, H., Mougél, M., Dondon, J., Grunberg-Manago, M., Ebel, J. P. and Ehresmann, B. (1986) Cross-linking of initiation factor IF3 to *Escherichia coli* 30S ribosomal subunit by trans-diamminedichloroplatinum(II): characterization of two cross-linking sites in 16S rRNA; a possible way of functioning for IF3. *Nucleic Acids Res.*, **14**, 4803–4821.
- Emerson, S.D., Waugh, D.S., Scheffler, J.E., Tsao, K.L., Prinzo, K.M. and Fry, D.C. (1994) Chemical shift assignments and folding topology of the Ras-binding domain of human Raf-1 as determined by heteronuclear three-dimensional NMR spectroscopy. *Biochemistry*, **33**, 7745–7752.
- Fortier, P.L., Schmitter, J.M., Garcia, C. and Dardel, F. (1994) The N-terminal half of initiation factor IF3 is folded as a stable independent domain. *Biochimie*, **76**, 376–383.
- Furey, W. and Swaminathan, S. (1990) PHASES: a package of programs for protein crystallography. *Am. Crystallogr. Assoc. Meeting Abstr.*, **PA33**, 18, 73.
- Garcia, C., Fortier, P.-L., Blanquet, S., Lallemand, J.-Y. and Dardel, F. (1995) ¹H and ¹⁵N resonance assignments and structure of the N-terminal domain of *Escherichia coli* initiation factor 3. *Eur. J. Biochem.*, **228**, 395–402.
- Godefroy-Colburn, T., Wolfe, A.D., Dondon, J., Grunberg-Manago, M., Dessen, P. and Pantaloni, D. (1975) Light-scattering studies showing the effect of initiation factors on the reversible dissociation of *Escherichia coli* ribosomes. *J. Mol. Biol.*, **94**, 461–487.
- Golden, B.L., Ramakrishnan, V. and White, S.W. (1993) The structure of ribosomal protein L6: structural evidence of gene duplication from a primitive RNA binding protein. *EMBO J.*, **12**, 4901–4908.
- Gualerzi, C.O. and Pon, C.L. (1990) Initiation of mRNA translation in prokaryotes. *Biochemistry*, **29**, 5881–5889.
- Hartz, D., McPheeters, D.S. and Gold, L. (1989) Selection of the initiator tRNA by *Escherichia coli* initiation factors. *Genes Dev.*, **3**, 1899–1912.
- Hartz, D., Binkley, J., Hollingsworth, T. and Gold, L. (1990) Domains of initiator tRNA and initiation codon crucial for initiator tRNA selection by *Escherichia coli* IF3. *Genes Dev.*, **4**, 1790–1800.
- Hendrickson, W.A. (1991) Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science*, **254**, 51–58.
- Hoffmann, 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.
- Hoffmann, D.W., Davies, C., Gerchman, S.E., Kycia, J.H., Porter, S.J., White, S.W. and Ramakrishnan, V. (1994) Crystal structure of prokaryotic ribosomal protein L9: a bi-lobed RNA-binding protein. *EMBO J.*, **13**, 205–212.
- Hu, W.S., Wang, R.Y.-H., Shih, J.W.-K. and Lo, S.-C. (1993) Identification of a putative *infC-rpmI-rplT* operon flanked by long inverted repeats in *Mycoplasma fermentans* (incognitus strain). *Gene*, **127**, 79–85.
- Jones, T.A., Zou, J.-Y., Cowan, S.W. and Kjeldgaard, M. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr.*, **A47**, 110–119.
- Kraulis, P. (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.*, **24**, 946–950.
- Kycia, J.H., Biou, V., Shu, F., Gerchman, S.E., Graziano, V. and Ramakrishnan, V. (1995) Prokaryotic translation initiation factor IF3 is an elongated protein consisting of two crystallizable domains. *Biochemistry*, **34**, 6183–6187.
- Laughrea, M. and Tam, J. (1989) Ribosomal protein S1 and initiation factor IF3 do not promote the ribosomal binding of approximately 19-nucleotide-long mRNA and mRNA models. *Biochem. Cell. Biol.*, **67**, 812–817.
- Leijonmarck, M., Appelt, K., Badger, J., Liljas, A., Wilson, K.S. and White, S.W. (1988) Structural comparison of the prokaryotic ribosomal protein L7/L12 and L30. *Proteins*, **3**, 243–248.
- Leslie, A.G.W. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *CCP4 ESF-EACMB Newsl. Protein Crystallogr.* No. 26, Daresbury Laboratory, Warrington, UK.
- Liveris, D., Schwartz, J.J., Geertman, R. and Schwartz, I. (1993) Molecular cloning and sequencing of *infC*, the gene encoding translation initiation factor IF3, from four enterobacterial species. *FEMS Microbiol. Lett.*, **112**, 211–216.
- MacKee, L.A., Kahan, L., Wahba, A.J. and Schwartz, I. (1980) Photochemical cross-linking of initiation factor-3 to *Escherichia coli* 30S ribosomal subunits. *J. Biol. Chem.*, **255**, 10526–10531.
- Moazed, D., Samaha, R.R., Gualerzi, C. and Noller, H.F. (1995) Specific protection of 16S rRNA by translational initiation factors. *J. Mol. Biol.*, **248**, 207–210.
- Muralikrishna, P. and Wickstrom, E. (1989) *Escherichia coli* initiation factor 3 protein binding to 30S ribosomal subunits alters the accessibility of nucleotides within the conserved central region of 16S rRNA. *Biochemistry*, **28**, 7505–7510.
- Nagai, K., Oubridge, C., Jessen, T.H., Li, J. and Evans, P.R. (1990) Crystal structure of the RNA-binding domain of the U1 small nuclear ribonucleoprotein A. *Nature*, **348**, 515–520.
- Ohsawa, H. and Gualerzi, C. (1981) Structure-function relationship in *Escherichia coli* initiation factors. *J. Biol. Chem.*, **256**, 4905–4912.
- Otwinowski, Z. (1991) In Wolf, W., Evans, P.R. and Leslie, A.G.W. (eds), *Isomorphous Replacement and Anomalous Scattering*. Science and Engineering Research Council, Daresbury Laboratory, Daresbury, UK, pp. 80–86.
- 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.
- Pon, C.L., Brombach, M., Thamm, S. and Gualerzi, C.O. (1989) Cloning and characterization of a gene cluster from *Bacillus stearothermophilus* comprising *infC*, *rpmI* and *rplT*. *Mol. Gen. Genet.*, **218**, 355–357.
- Ramakrishnan, V. and White, S.W. (1992) Structure of ribosomal protein S5 reveals sites of interaction with 16S RNA. *Nature*, **358**, 768–771.
- Ramakrishnan, V., Yabuki, S., Sillers, I.-Y., Schindler, D.G., Engelman, D.M. and Moore, P.B. (1981) Position of proteins S6, S11 and S15 in the 30S ribosomal subunit of *Escherichia coli*. *J. Mol. Biol.*, **153**, 739–760.
- Ramakrishnan, V., Finch, J.T., Graziano, V., Lee, P.L. and Sweet, R.M. (1993) Crystal structure of the globular domain of histone H5 and its implications for nucleosome binding. *Nature*, **362**, 219–223.
- Sacerdot, C., Fayat, G., Dessen, P., Springer, M., Plumbridge, J.A., Grunberg-Manago, M. and Blanquet, S. (1982) Sequence of a 1.26-kb DNA fragment containing the structural gene for *E. coli* initiation factor IF3: presence of an AUU initiator codon. *EMBO J.*, **1**, 311–315.
- Stöffler, G., Bald, R., Kastner, B., Lüthmann, R., Stöffler-Meilicke, M. and Tishendorf, G. (1980). In Chambliss, G., Craven, G.R., Davies, J., Davis, K., Kahan, L. and Nomura, M. (eds), *Ribosomes: Structure, Function and Genetics*. University Park Press, Baltimore, MD, pp. 171–205.
- Studier, F.W. and Moffatt, B.A. (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.*, **189**, 113–130.
- Subramanian, A.R. and Davis, B.D. (1970) Activity of initiation factor 3 in dissociating *Escherichia coli* ribosomes. *Nature*, **228**, 1273–1275.
- Terwilliger, T.C., Kim, S.-H. and Eisenberg, D. (1987) Generalized method of determining heavy-atom positions using the difference Patterson function. *Acta Crystallogr.*, **A43**, 1–5.
- Vingron, M. and Argos, P. (1991) Motif recognition and alignment for many sequences by comparison of dot-matrices. *J. Mol. Biol.*, **218**, 33–43.
- Wang, B.C. (1985) Resolution of phase ambiguity in macromolecular crystallography. *Methods Enzymol.*, **115**, 90–112.
- Weiel, J. and Hershey, J.W.B. (1981) Fluorescence polarization studies of the interaction of *Escherichia coli* protein synthesis initiation factor 3 with 30S ribosomal subunits. *Biochemistry*, **20**, 5859–5865.
- Weiel, J. and Hershey, J.W.B. (1982) The binding of fluorescein-labeled protein synthesis initiation factor 2 to *Escherichia coli* 30S ribosomal subunits determined by fluorescence polarization. *J. Biol. Chem.*, **257**, 1215–1220.
- Wickstrom, E. (1983) Nuclease mapping of the secondary structure of the 49-nucleotide 3' terminal cloacin fragment of *Escherichia coli* 16S RNA and its interactions with initiation factor 3. *Nucleic Acids Res.*, **11**, 2035–2052.
- Wilson, K.S., Appelt, K., Badger, J., Tanaka, I. and White, S.W. (1986) Crystal structure of a prokaryotic ribosomal protein. *Proc. Natl Acad. Sci. USA*, **83**, 7251–7255.
- Zucker, F.H. and Hershey, J.W.B. (1986) Binding of *Escherichia coli* protein synthesis initiation factor IF1 to 30S ribosomal subunits measured by fluorescence polarization. *Biochemistry*, **25**, 3682–3690.

Received on April 5, 1995; revised on May 12, 1995