Crystal structures of ribosome anti-association factor IF6

Caroline M. Groft[1], Roland Beckmann[2,3], Andrej Sali[1] and Stephen K. Burley[1,4]

Ribosome anti-association factor eIF6 (originally named according to translation initiation terminology as eukaryotic initiation factor 6) binds to the large ribosomal subunit, thereby preventing inappropriate interactions with the small subunit during initiation of protein synthesis. We have determined the X-ray structures of two IF6 homologs, *Methanococcus jannaschii* archaeal aIF6 and *Sacchromyces cerevisiae* eIF6, revealing a phylogenetically conserved 25 kDa protein consisting of five quasi identical α/β subdomains arrayed about a five-fold axis of pseudosymmetry. Yeast eIF6 prevents ribosomal subunit association. Comparative protein structure modeling with other known archaeal and eukaryotic homologs demonstrated the presence of two conserved surface regions, one or both of which may bind the large ribosomal subunit.

Protein synthesis in eukaryotes depends critically on the availability of free ribosomal subunits, both large and small, to ensure that they can be separately loaded onto the correct initiation codon[1]. Two translation factors serve as ribosomal chaperones. eIF3 binds to the 40S subunit and prevents premature association with the 60S subunit[2]. eIF6, the subject of this paper, binds to mature 60S subunits and prevents inappropriate interactions with 40S subunits[3–7].

eIF6 is a highly conserved 25 kDa protein that is essential in yeast[8–10]. Pairwise amino acid identities range from 50–98% among 10 eukaryotic homologs of eIF6, some of which are interchangeable *in vivo*[9,10]. Seven archaebacterial genomes contain recognizable aIF6s (pairwise identities 26–90%) that are similar to the eIF6s (for example, *Methanococcus jannaschii* aIF6 and *Saccharomyces cerevisiae* eIF6 are 33% identical), suggesting that all eIF6s and aIF6s share a common three-dimensional structure[11].

Yeast eIF6 conditional mutants arrest in G1 under nonpermissive conditions, as do mutants of various *bona fide* eIFs[12,13]. Instead of causing immediate translation arrest and cell death, however, eIF6 depletion slows translation and leads to a loss of yeast viability long after eIF6 becomes undetectable, indicating that eIF6 is not critical for translation initiation[8]. Yeast death in eIF6-depleted cells can instead be explained by the critically low numbers of free 60S subunits[8,10]. Furthermore, immunofluorescence studies localized eIF6 to the nucleolus and the cytoplasm[10], suggesting that there is a direct link between eIF6–ribosome binding and maintenance of free 60S subunits. Despite recent advances in our understanding of eIF6 function, the precise molecular mechanisms responsible for its interactions with the large ribosomal subunit are unknown.

Human eIF6 has an additional biochemical activity that involves integrins, which are membrane receptors responsible for cell–cell and cell–extracellular matrix adhesion. A yeast two-hybrid screen demonstrated that eIF6 binds to the cytoplasmic domain of β_4 integrin, which is both necessary and sufficient for hemidesmosome formation[14]. eIF6 is expressed in skin cell progenitors at levels greater than those seen in other tissue types[9] and colocalizes with β_4 integrin at the basal surface of these polarized cells[10,15]. Moreover, upregulation of eIF6 expression in human colorectal tumors is correlated with malignant progression[15]. eIF6 must, therefore, be capable of binding to the surface of the 60S ribosomal subunit and, at least in humans, to the two tandem fibronectin type III repeats of β_4 integrin[16]. Knowledge of the structure of eIF6 would enable directed examination of its interactions with these macromolecules, and elucidation of possible biochemical functions in ribosome biogenesis and transport, translation initiation, and cell growth and differentiation.

Our X-ray study of eIF6 also represents a useful example of the potential impact of the Protein Structure Initiative at the National Institute of General Medical Sciences (http://www.nih.gov/nigms/psi.html), which aims to provide experimental structures for more than 10,000 proteins over the next decade. The universe of distinct protein folds is significantly smaller than the universe of all protein sequences (reviewed in ref. 17). Current estimates place the number of distinct, stable three-dimensional arrangements of polypeptide chains at somewhere between 2,000 and 5,000, which is much smaller than the estimated 100,000 or so human genes and reflects an important aspect of evolution. Nature reuses a relatively small number of stable protein domain structures or folds, which have an average size of 153 ± 87 residues. The emerging field of structural genomics seeks to provide one or more experimental structures for each distinct protein fold family[17]. At present, the Protein Data Bank (PDB; http://www.rcsb.org/) contains more than 13,000 structures, representing fewer than 800 distinct protein folds. eIF6 was selected as structure determination target number 111 by the New York Structural Genomics Research Consortium (NYSGRC; http://www.nysgrc.org/), with every expectation that success would provide information about a previously uncharacterized polypeptide chain fold.

Here, we report X-ray crystallographic studies of *M. jannaschii* and *S. cerevisiae* IF6 at 1.3 and 2.5 Å, respectively. These highly similar structures reveal a novel protein structure consisting of five quasi identical α/β subdomains related by a five-fold axis of pseudosymmetry. Comparative protein structure modeling was used to derive high quality model structures for six more known aIF6s and nine more known eIF6s. Analyses of the two resulting ensembles of aIF6 and eIF6 model structures provide insights

[1]Laboratories of Molecular Biophysics and [2]Cell Biology, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA. [3]Current address: Charite, Humboldt Universität, Berlin, Germany. [4]Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA.

Correspondence should be addressed to S.K.B. *email: burley@rockefeller.edu*

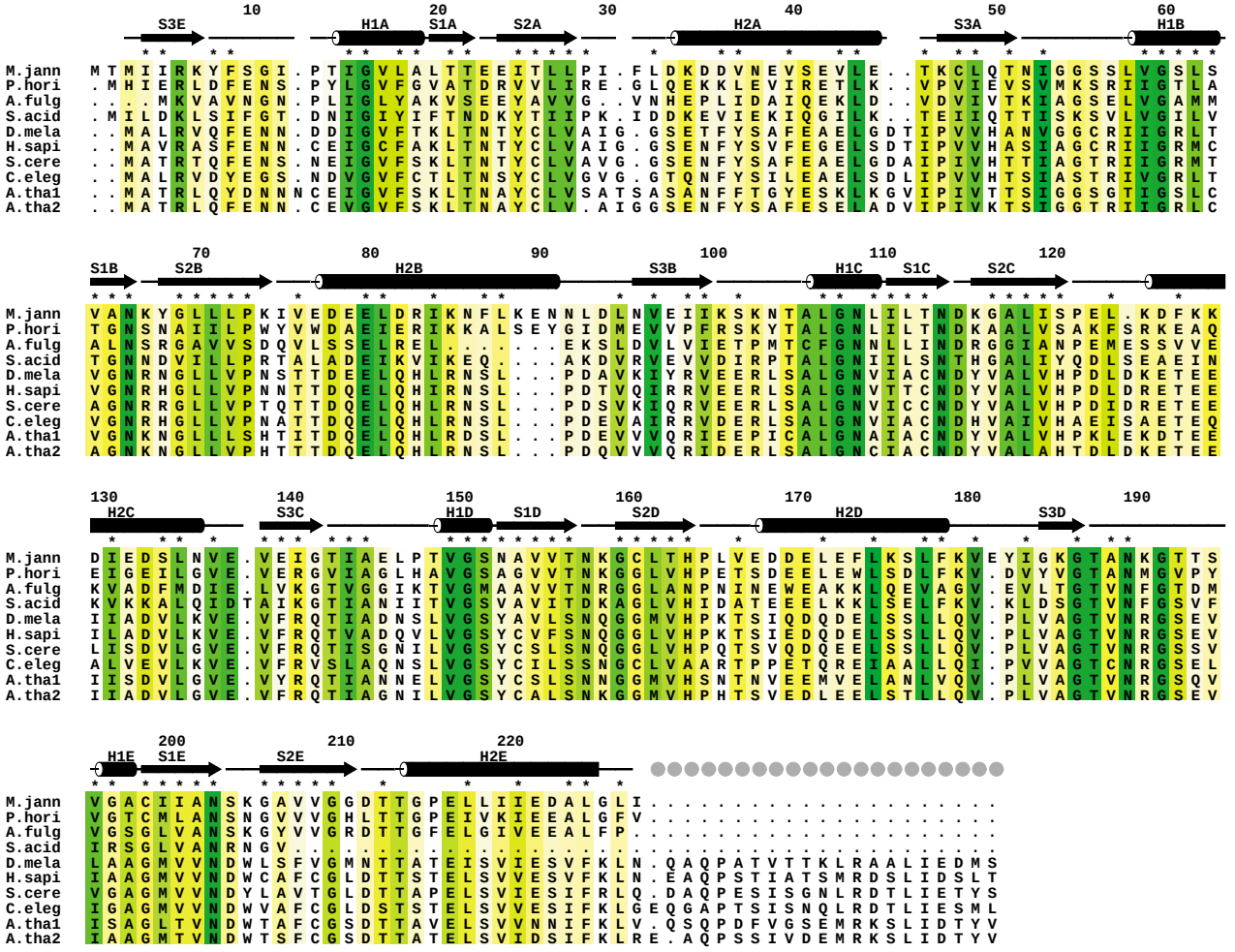

Fig. 1 eIF6 and aIF6 sequence alignments. Secondary structural elements are denoted as follows: α-helices or 3_{10}-helices, cylinders; β-strands, arrows; random coil, lines. Gray circles represent portions of the polypeptide chain that were not present in the crystallization sample. BLOSUM62 (ref. 40) sequence similarity is colored coded using a gradient from white (<40% identity) to dark green (100% identity). Organism names: M.jann, *M. jannaschii*; P.hori, *Pyrococcus horikoshii*; A.fulg, *Archaeoglobus fulgidus*; S.acid, *Sulfolobus acidocaldarius*; D.mela, *Drosophila melanogaster*; H. sapi, *Homo sapiens*; S.cere, *S. cerevisiae*; C.eleg, *Caenorhabditis elegans*; A.thal, *A. thaliana*. There are two eIF6 homologs in *A. thaliana*, labeled '1' and '2'. Asterisks denote buried side chains.

into the mechanisms of IF6 function, including interactions with the large ribosomal subunit and possibly β_4 integrin.

Activity resides in the conserved region of eIF6.

A sequence alignment was performed with all currently available aIF6 and eIF6 homologs (see Fig. 1 for a representative subset). Archaebacterial family members are ~225 residues long, whereas all known eIF6s contain an additional C-terminal extension of 20 amino acids. To determine if the phylogenetically conserved portion of *S. cerevisiae* eIF6 supports ribosome anti-association, a C-terminal deletion eIF6(1–224) was constructed and assayed[4]. The biochemical function of the truncated yeast protein is indistinguishable from its full length counterpart *in vitro* (Fig. 2).

aIF6 and eIF6 structure determination

The structure of full length *M. jannaschii* aIF6 was determined by combining phases from both multiwavelength anomalous dispersion (MAD)[18] and multiple isomorphous replacement (MIR) approaches (Fig. 3). The current refinement model at 1.3 Å resolution consists of residues 3–227, giving a final R-factor of 13.2% and an R_{free} of 17.9% with excellent stereochemistry (Table 1). The structure of *S. cerevisiae* eIF6(1–224) was determined using molecular replacement with a poly-Ala model of *M. jannaschii* aIF6. The current refinement model at 2.5 Å resolution consists of residues 1–224, giving a final R-factor of 19.7% and an R_{free} of 25.0% with excellent stereochemistry. (See Table 1 and Methods for a complete description of structure determinations and refinements.)

eIF6 and aIF6 have novel structures

Both archaeal and eukaryotic IF6 structures contain an internal five-fold axis of pseudosymmetry (Fig. 4) and consist of five copies of a repeating α/β subdomain of about 45 residues, denoted A–E. We have chosen to call this pentameric fold 'pentein'. Each subdomain contains one long (at least two full turns) α-helix, a shorter α-helix or 3_{10}-helix, and three β-strands. Secondary structural elements for both homologs proceed from the N-terminus to the C-terminus as follows (S, strand; H, helix): S3E, H1A, S1A, S2A, H2A, S3A, H1B, S1B, S2B, H2B, S3B, H1C, S1C, S2C, H2C, S3C, H1D, S1D, S2D, H2D, S3D, H1E, S1E, S2E and H2E. (Subdomain boundaries have been chosen to

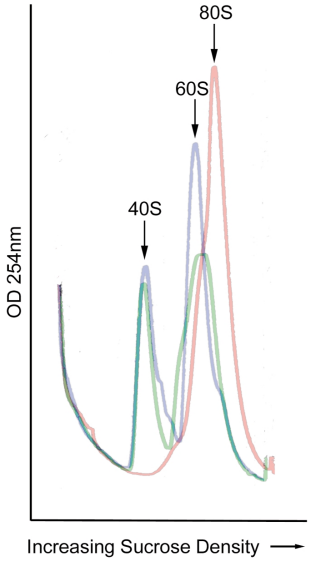

Fig. 2 Yeast eIF6(1–224) prevents association of yeast ribosomal subunit. Sucrose gradient centrifugation results are color coded as follows: blue, full length yeast eIF6; green, yeast eIF6(1–224); and red, full length *M. jannaschii* aIF6. Positions in the sucrose gradient for 40S and 60S ribosomal subunits and 80S ribosomes are denoted with labeled arrows.

position the N- and C-termini within subdomain E.) Structure-based amino acid sequence alignments of various eukaryotic and archaeal homologs (Fig. 1) demonstrate that conserved residues map to the hydrophobic cores of each of the five subdomains comprising the penteins (asterisks in Fig. 1), whereas the locations of most insertions and deletions correspond to surface loops. The sole exception is α-helix H2B, which has different lengths in the archaeal and eukaryotic structures. We conclude, therefore, that the phylogenetically conserved regions of all known aIF6 and eIF6 homologs share the three-dimensional structure illustrated in Fig. 4.

The structure of subdomain B of aIF6 is depicted in Fig. 5. A short H1 α-helix leads to two antiparallel β-strands (S1 and S2) followed by a longer α-helix (H2), which packs against the β-strands to form a mini-hydrophobic core. The subdomain finishes with a third β-strand (S3) running parallel to the second β-strand (S2), creating a mixed polarity three-stranded β-sheet. Atomic coordinate submission to DALI[19] returned no statistically significant matches for either experimental structure (the closest match was myoglobin, PDB code 2FAL, Z-score 1.1). DALI submission of each subdomain structure gave several matches with Z-scores of ~2.0. The gelsolin actin modifying proteins possess a single eIF6-like subdomain, suggesting that the IF6 and gelsolin families share a common ancestor.

It is remarkable that the distance between the penultimate amino acid of aIF6 (Leu 227) and Met 3 is 6.5 Å, which would be close enough to make a peptide bond between the N-terminus and C-terminus, cyclizing the pentein. In contrast, all of the eukaryotic eIF6 homologs possess a semi-conserved C-terminal extension of ~20 amino acids, making them less compact than their archaeal counterparts (Fig. 1, gray dots). This extension is not required for large subunit binding *in vitro* (Fig. 2), and could serve as a target for eIF6 regulation *in vivo*.

Subdomain differences break the five-fold symmetry

A structure-based superposition of aIF6 and eIF6 gave a root mean square (r.m.s.) deviation between equivalent Cα atoms of 3.8 Å (224 structurally aligned residues, corresponding to residues 3–227 of *M. jannaschii*). When the H2 α-helices are omitted from the calculation the r.m.s. deviation drops to 1.0 Å, demonstrating that most variation between the two IF6 structures occurs in the long α-helices. A superposition of the Cα backbone traces of archael and yeast eIF6 is shown in Fig. 4 (lower). Deviations >3 Å between the two structures map to subdomain A, where yeast α-helix H2A is kinked due to stacking of aromatic side chains within the hydrophobic core, and to the loops leading out of α-helix H2A and into α-helix H2C. (Note that α-helix H2B is one turn shorter in yeast eIF6, as seen in the structure based sequence alignment illustrated in Fig. 1.) For comparison, the two protomers comprising the asymmetric unit of the aIF6 crystals are related by noncrystallographic symmetry with a r.m.s. deviation of 0.3 Å between equivalent Cα atoms.

Superposition of each subdomain on its neighbors shows that the orientation of the β-strands is as similar among subdomains as it is between the two homologs, with α-helices and loops varying considerably (data not shown). Pairwise Cα atom r.m.s. deviations among the five β-sheets of aIF6 are ~0.65 Å. Inclusion of α-helices in these similarity estimates gives r.m.s. deviations of ~4.5 Å, because each α-helix packs against the β-sheet at a slightly different angle. This finding reflects subtle variations in the structure of the mini-hydrophobic cores within each subdomain. In addition, the longer loops connecting β-strand S3 to α-helix H1 of the next subdomain vary in orientation and length, as do the loops between α-helix H2 and β-strand S3. All long α-helices are oriented with their N-termini pointed toward the face containing the small H1 α-helices (the 'N-terminal' face). Within aIF6, α-helices H2B and H2C differ in length from H2A and H2E, creating a marked asymmetry in the surface of the pentein. Within yeast eIF6, α-helix H2A runs nearly parallel to the five-fold axis and is slightly bent due to interactions within the hydrophobic core of subdomain A (Fig. 4), contributing further to pentein asymmetry.

A structure-based sequence alignment of all five aIF6 subdomains (Fig. 6) shows almost no intramolecular similarity. There is a single invariant residue, a Gly in the middle of each H1

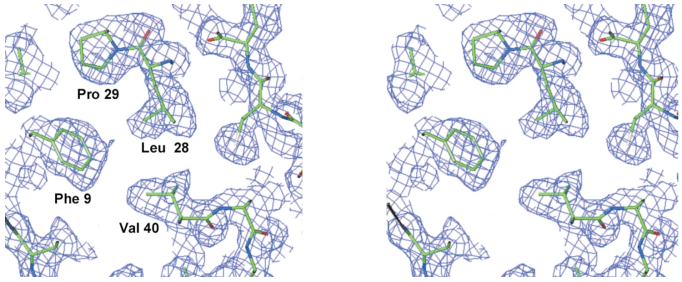

Fig. 3 Stereoview of the experimental electron density for one *M. jannaschii* aIF6 subdomain. The map has been density modified, noncrystallographic symmetry averaged and shows several residues in the hydrophobic core of aIF6 subdomain A (contour level 1.5 σ).

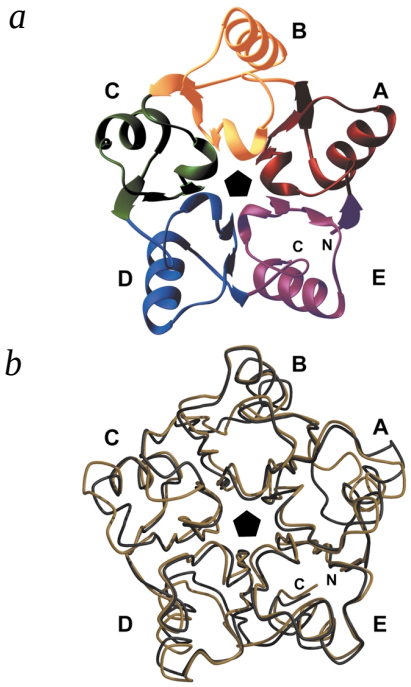

Fig. 4 Pentein structures of aIF6 and eIF6. ***a***, RIBBONS drawing of *M. jannaschii* aIF6 (top), with a black pentagon denoting the five-fold pseudosymmetry axis. Individual aIF6 subdomains are color coded in progression from the N-terminus to the C-terminus as follows: A, red; B, yellow; C, green; D, blue; E, purple. ***b***, Cα backbone traces of the *M. jannaschii* aIF6 (black) and *S. cerevisiae* eIF6 (gold) overlaid to show local differences.

α-helix. Elsewhere, subdomain similarities are minimal: small polar residues are found at the C-termini of the S1 β-strands; the S2 β-strands are hydrophobic; and the H2 α-helices are highly charged. It is unlikely, therefore, that careful study of the sequence would have enabled us to detect the intramolecular five-fold pseudosymmetry.

The central hollow

Solvent accessible molecular surfaces of aIF6 and eIF6 calculated using GRASP[20] are illustrated in Fig. 7. The N-terminal face of aIF6 is flat and largely hydrophobic, with two patches of negative electrostatic potential (Fig. 7*b*, left). Contacts among each of the subdomains do not make a unified hydrophobic center. Instead, there is a hydrophobic torus produced by the association of the subdomains with their neighbors, which in turn creates a cave-like hollow in the center of the pentein (Fig. 8). The yeast eIF6 structure contains an actual closed cavity with a roof. In both structures, the carbonyl groups of the first and third amino acids in the S1 β-strands point into the cavity, while their side chains point toward the S1 β-strand of the adjacent subdomain. There is a continuous hydrophobic ring created by the side chains of the S1 β-strands, which are predominantly small hydrophobic or polar residues. We believe that this structural feature may reflect the evolutionary history of the pentein fold, as models of the other known IF6 sequences derived from the experimental structures preserve the hollow and indicate that no other known sequence can unify the fold into a single hydrophobic core. Therefore, it is possible that the ancestral protein was a symmetric oligomer composed of five small (5 kDa) subunits, each one corresponding to one subdomain. Gene duplication/fusion events could have yielded the observed pentein structure, which subsequently evolved as a single gene product. A similar hypothesis has been proposed for the TIM barrel fold; sequence analysis and structural superpositions were used to propose the existence of two subdomains within bacterial enzymes of related function[21]. A phylogenetic tree derived from structure based alignments of all 17 known eIF6 and aIF6 sequences (not shown) suggests that the archaeal and eukaryotic factors originated from a common ancestor present before divergence of the archaeal and eukaryotic kingdoms.

The central hollow in IF6 does not extend through the protein (Fig. 7, right). In aIF6, there is an opening on the N-terminal face ~6–8 Å in diameter. This 'mouth' is surrounded by the H1 α-helices, which contain the most conserved residues, both among subdomains and among various aIF6 and eIF6 homologs (Figs 1, 6, 7*c*,*d*). The Gly in α-helix H1 is invariant across all species except in subdomain E, where it is either a small hydrophobic (Ala) or polar residue (Ser), or, in two archaea, a positively charged Arg. In all eIF6s except one of the two *Arabidopsis thaliana* homologs, the residue adjacent to the invariant Gly in α-helix H1B is Arg. In our yeast eIF6 structure, the corresponding Arg (residue 61) forms the roof of the closed cavity, and its guanidinium group makes several hydrogen bonds with α-helix H1 Gly residues (Fig. 8*b*). The B-factor for Arg 61 is slightly higher than those of surrounding residues (40 Å^2 *versus* 25 Å^2 on average), suggesting that this basic side chain could flip out to create a cave-like hollow instead of a closed cavity. The C-terminal residues of the S1 β-strands and the N-terminal residues of the loops connecting the S1 and S2 β-strands form the base of the cavity, which is ~15 Å deep and ~9 Å in diameter at its widest point.

Sixteen well-ordered water molecules (some with B-factors as low as 7 Å^2) were detected within the aIF6 cave-like hollow. Ten of these solvent molecules are arranged in two pentagonal layers within the central hollow (Fig. 8*a*,*c*), where they are hydrogen bonded to the carbonyl oxygen atoms of the first and third residues of the S1 β-strands. In addition, three pairs of water molecules were found beneath the two pentagonal layers (Fig. 8*c*). The resolution limit of our yeast eIF6 structure precluded experimental identification of water molecules, but we presume that the closed cavity in the yeast protein is also water filled.

Parallels with the β-propeller fold family.

The β-propeller folds contain repeats of a four-stranded antiparallel β-strand subdomain[22]. Like the IF6 fold, the repeating subdomains are arranged about a central axis, creating a solvent filled tunnel[23]. The central strand for both fold families is composed of small polar and small hydrophobic amino acids, with

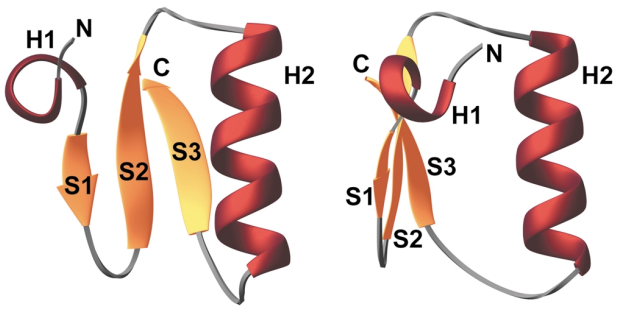

Fig. 5 Two perpendicular views of aIF6 subdomain B. The structure proceeds from the N-terminus to the C-terminus in the order α-helix H1, β-strand S1, β-strand S2, α-helix H2 and β-strand S3.

A	(8-51)	YFSGIPT	IGVLA	LTT	E	EITLL	PIFLD	KDDVNEVSEVLE	T	KCLQT
B	(52-100)	NIGGSSL	VGSLS	VAN	K	YGLLLPK	IVE	DEELDRIKNFLKEN	NLDL	NVEII
C	(101-142)	KSKNTA	LGNL	ILTN	D	KGALIS	PELK	DFKKDIEDSLN	VE	VEIG
D	(143-187)	TIAELPT	VGSN	AVVT	NK	GCLTH	PLVE	DDELEFLKSLF	KVEYI	GKG
E	(188-227;1-7)	TANKGTTS	VGA	CIIAN	SK	GAVVGG	DTT	GPELLIIEDALG	LIMTM	IIRK

Fig. 6 Sequence alignments of the *M. jannaschii* aIF6 subdomains. Subdomain secondary structural elements are depicted above the aligned subdomain sequences, with color coding as in Fig. 5. β-strand S3 of subdomain E is composed of residues 3–8 (Met 1, Thr 2 and Ile 228 were not visible in either half of the asymmetric unit).

backbone carbonyl groups projecting into the tunnel. The medial strands contribute to a stable hydrophobic core, and the outer secondary structure element of each motif is the most divergent among subdomains, as is the H2 α-helix within the IF6 fold. Moreover, the IF6 pentein exploits the same strategy for closing the radial arrangement of subdomains as seen in most known β-propeller structures; the last β-strand at the C-terminus hydrogen bonds with a β-strand from the N-terminal subdomain (called a 'Velcro closure')[22]. These parallels in motif repetition, strand arrangement, and the creation of a solvent filled hollow suggest the existence of a successful evolutionary strategy for creating stable protein structures containing multiple copies of a quasi identical subdomain. If true, there may well be proteins containing six or more (alternatively, four or fewer) IF6 subdomains, as is seen for the repeats in the β-propeller proteins (reviewed in ref. 22). This aspect of modular protein evolution is quite different from the 'beads on a string' property of many eukaryotic proteins (for example, transcription factor IIIA type zinc finger proteins).

Modeling of other known aIF6 and eIF6 homologs

All the other 15 known aIF6 and eIF6 homologs were modeled using MODELLER[24]. The structure of *M. jannaschii* aIF6 served as the modeling template for six archaeal proteins, while the structure of yeast eIF6 provided the template for nine eukaryotic homologs. Model scores in excess of 0.7 (on a scale of 0–1) connote a useful model[25]. All calculated aIF6 and eIF6 models could be modeled using these two experimental templates (see supplementary information at http://www.nysgrc.org/).

To examine the accuracy of the modeling procedure, the two experimental structures were used as templates to model each other. The model of yeast eIF6 calculated from the *M. jannaschii* aIF6 structure shows an average Cα r.m.s. deviation of 1.6 Å with the experimental yeast structure. However, several regions of the experimental structure deviated significantly from the model (local modeling errors >3 Å). The obverse calculation, with a *M. jannaschii* model calculated using the yeast experimental structure, yielded similar results. Not surprisingly, the largest differences between the experimental structures and the calculated models correspond to regions that differ the most between the two homologs. Thus, it is a distinct advantage to have experimental structures for more than one member of a particular fold family. As the eukaryotic and archaeal sequences cluster (Fig. 1), our two structures should be sufficient to model all IF6 homologs with reasonable accuracy. An obvious weakness of comparative protein structure modeling becomes apparent whenever the positions of surface accessible side chains are analyzed in detail. For example, the conformation of the side chain of Arg 61 in yeast eIF6 was not correctly predicted from the aIF6 template (Fig. 8*b*). However, homology models are probably accurate enough to identify surface patches of biological interest that can be studied by site-directed mutagenesis of predicted solvent accessible residues.

The resulting combination of two experimental structures and 15 models permitted comparisons of the molecular surfaces of all aIF6 and eIF6 homologs. In every case, the N-terminal face of the pentein is essentially flat and largely hydrophobic with one or two patches of acidic residues (Fig. 7, left). All eukaryotic models (except one of the *A. thaliana* eIF6s) have a basic patch on the N-terminal face that corresponds to Arg 61 of yeast eIF6. Two aIF6s possess a similar Arg residue in H1E. These Arg side chains are probably hydrogen bonded to the Gly carbonyl oxygen atoms in the mouth of the central cavity, as in our yeast experimental structure (Fig. 8*b*). The C-terminal faces of all IF6s (Fig. 7, right) display somewhat rougher surfaces, with a considerable number of solvent accessible polar and aromatic residues. At present, the biological significance of the existence of a second *A. thaliana* eIF6 homolog is not understood, but it is not without precedent. Some plants also possess two TATA box binding protein homologs (reviewed in ref. 26).

Interaction of Yeast eIF6 with the large ribosomal subunit

The molecular surfaces of *S. cerevisiae* eIF6 and *M. jannaschii* aIF6 are shown in Fig. 7 and color coded for electrostatic potential and conservation of solvent accessible residues within their respective kingdoms[27]. The C-terminal surface of the yeast protein displays a single region encompassing the conserved Arg and Tyr residues Arg 67, Tyr 113, Arg 223 and Tyr 202 (Fig. 7*a*,*c*), which represent excellent candidates for amino acids participating in biologically relevant protein–protein interfaces (reviewed in ref. 28). Conserved surface features overlie the center of the pentein and parts of subdomains B–E, mapping to α-helix H2B, the loop between α-helix H2C and β-strand S3C, the loop between β-strand S3D and α-helix H1E, and the C-terminus of α-helix H2E (Fig. 7*c*, right).

The N-terminal face of yeast eIF6 displays fewer conserved, solvent accessible residues, and is considerably smoother and less chemically varied than its C-terminal counterpart (compare Fig. 7*a*, left and right, and the buried residue profile in Fig. 1). Conserved N-terminal surface features include the pentein center and subdomains A and E (Fig. 7*c*, left). It is, therefore, reasonable to suggest that the C-terminal face of the eukaryotic eIF6 serves as the binding site for the 60S ribosomal subunit. Additional, albeit indirect, support for this assertion comes from a similar analysis of the molecular surface properties of *M. jannaschii* aIF6 (Fig. 7*b*,*d*), which cannot bind productively to yeast ribosomes (Fig. 2). The C-terminal faces of the archaeal aIF6s display conserved surface features overlying the pentein center and subdomains C and E (Fig. 7*b*,*d*, right). On the oppo-

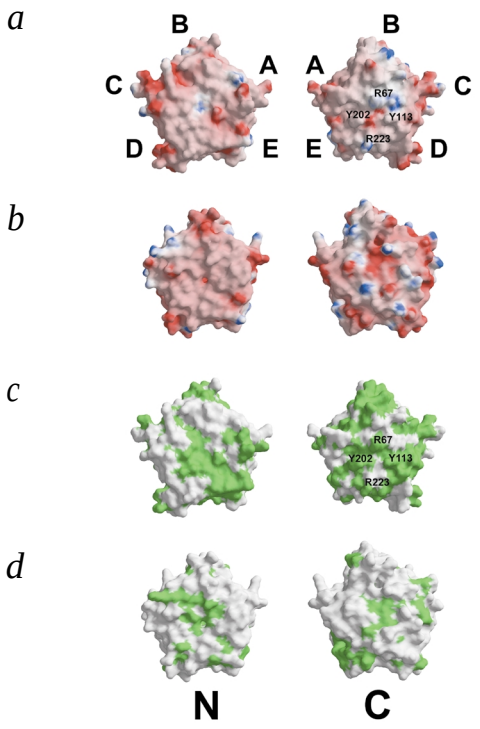

Fig. 7 Molecular surfaces of aIF6 and eIF6. Left, N-terminal face in the same orientation as Fig. 3. Right, C-terminal face showing the structures rotated 180° about the vertical axis. ***a***, Electrostatic surface of *S. cerevisiae* eIF6(1–224)[41]. Regions of negative and positive potential are colored red and blue, denoting <-15 and >+20k_BT, respectively, where k_B is the Boltzman constant and T is temperature. ***b***, Electrostatic surface of *M. jannaschii* aIF6. ***c***, Green color coded surface of *S. cerevisiae* eIF6 denotes underlying residues conserved among all eukaryotic eIF6 homologs. ***d***, Green color coded surface of *M. jannaschii* aIF6 denotes underlying residues conserved among all archaeal aIF6 homologs.

site N-terminal face, conserved solvent accessible residues map to the center and subdomains C, D and E, but not to subdomain A (Fig. 7*b*,*d*, left).

In summary, *S. cerevisiae* eIF6 and *M. jannaschii* aIF6 display common C-terminal pentein surface features, which are conserved among eukaryotes and archaebacteria, respectively. Some of these common surface features are, however, chemically distinct between the archaebacteria and eukaryotes, thereby explaining the observed species specificity of large ribosomal subunit binding. Interestingly, some β-propeller fold family members use aromatic amino acids arranged around the solvent filled tunnel (analogous to the C-terminal IF6 surface, which closes the hollow) to bind target proteins (for example, see ref. 29) in addition to contacts with the edge of the disk.

Potential interactions of eIF6 with other protein factors

Residues 173–183 of *S. cerevisiae* eIF6 (Leu-Ser-Ser-Leu-Leu-Gln-Val-Pro-Leu-Val-Ala) have been proposed to be a nuclear transport signal sequence[9]. This polypeptide chain segment maps to the C-terminus of α-helix H2D, the N-terminus of β-strand S3D, and a short loop connecting these two secondary structural elements (Fig. 1). In our yeast eIF6 structure, Leu 173, Leu 176, Leu 177, and Leu 181 contribute to the hydrophobic core of subdomain D. Only Ser 174 and Ser 175, on the polar surface of α-helix H2D, and Gln 178 and Val 179, comprising the short loop connecting α-helix H2D with β-strand S3D, are solvent accessible. It seems unlikely, therefore, that the proposed Leu-rich nuclear transport signal sequence (Leu 173–Ala 183) supports significant interactions with the nuclear export machinery, provided of course that yeast eIF6 remains folded during nuclear-cytoplasmic transport. A much stronger candidate region for eIF6 transport and regulation is the semiconserved C-terminal tail, which is restricted to eukaryotes and is not required for 60S ribosomal subunit binding *in vitro* (Fig. 2).

Conclusions

We have characterized a novel pentein fold family using X-ray crystallography to determine two structures, one representative from the archaeal and eukaryotic branches. Both proteins contain a hydrated cavity coinciding with the five-fold axis of pseudosymmetry, which may be important for protein folding, stability and/or function. The phylogenetically conserved N-terminal 224 residues of yeast eIF6 suffice for 60S ribosomal subunit binding *in vitro*, suggesting that the semiconserved C-terminal extension found in eukaryotic eIF6s has an alternative role. The yeast eIF6 structure permitted accurate comparative protein structure modeling of all available eukaryotic eIF6 homologs, which in turn facilitated identification of a putative ribosome interaction surface. Similar analyses may prove useful in functional studies of the aIF6s and their roles in archaeal protein synthesis. Two NYSGRC experimental structures, therefore,

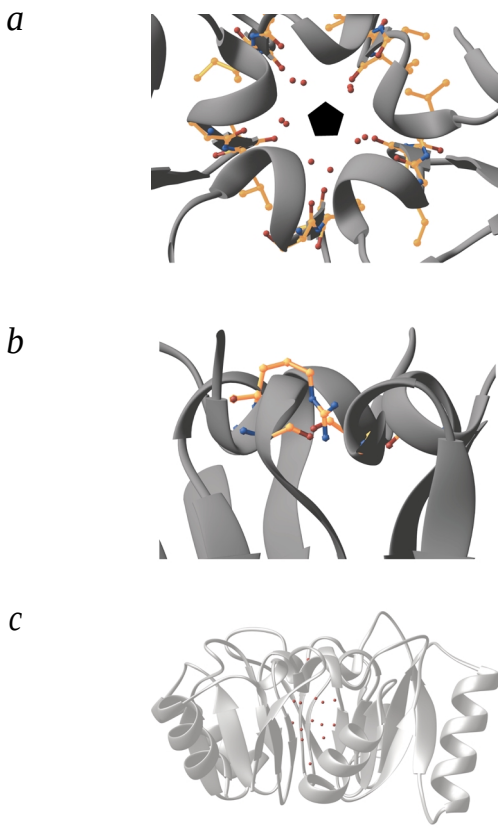

Fig. 8 Perpendicular views of the water-filled hollow of aIF6. ***a***, Top view of the cave-like hollow showing two pentagonal layers of water molecules (red spheres). The black pentagon denotes the five-fold pseudosymmetry axis. ***b***, Side view of *S. cerevisiae* eIF6 showing Arg 61 closing the cavity. ***c***, Side view of aIF6 showing 13 of the total complement of 16 internal water molecules (red spheres).

Table 1 Summary of phase determination and crystallographic refinement

***M. jannaschii* aIF6**

Space group C2

Unit cell a = 118.14 Å, b = 46.75 Å, c = 84.72 Å; β = 98.57°

Data set	Resolution (Å)	Reflections measured / unique	Completeness (%) overall / outer shell	R_{sym}[1] (%) overall / outer shell	Phasing power[2]	R_{cullis}
MAD analysis (two Os sites)					Iso (acen/cen)	Ano/Iso
λ1 = 1.13960 Å	25.0–1.90	254,619 / 69,934	99.1 / 93.0	4.8 / 15.8	–	0.91 / –
λ2 = 1.14023 Å	25.0–1.90	252,671 / 70,183	99.3 / 94.3	4.0 / 11.5	0.32 / 0.26	0.89 / 0.95
λ3 = 1.11903 Å	25.0–1.90	256,979 / 70,679	99.5 / 97.0	3.7 / 13.0	0.26 / 0.21	0.94 / 0.96
MAD figure of merit		0.31				
Derivative 1 (two Hg sites)						
λ = 1.00739 Å	25.0–1.90	244,396 / 68,560	98.8 / 97.0	4.7 / 12.4	–	0.92 / –
MAD + derivative 1 figure of merit		0.34				
Derivative 2 (two Os sites)						
λ = 0.91825 Å	25.5–1.50	365,875 / 139,273	96.8 / 93.0	4.3 / 20.2	0.79 / 0.69	0.94 / 0.79
Derivative 2 figure of merit		0.29				
Overall figure of merit		0.42				
Native (λ = 0.91825 Å)	25.0–1.30	463,351 / 110,239	97.7 / 95.4	3.6 / 23.5	–	–

Refinement statistics

Resolution[3] (Å)	25.0–1.3	R.m.s. deviations	
Completeness[3] (%)	84.6	Bond lengths[5] (Å)	0.019
R-factor[3]	0.132	Bond angles[5] (°)	2.1
R_{free}[3,4]	0.179		

***S. cerevisiae* eIF6**

Space group $P4_32_12$

Unit cell a = b = 56.257 Å, c = 171.703 Å

Native (λ = 1.54 Å)	25.0–2.5	60,623 / 9,926	96.8 / 84.9	8.6 / 34.0	–	–

Refinement statistics

Resolution[6] (Å)	25.0–2.5	R.m.s. deviations	
Completeness[6] (%)	93.5	Bond lengths[5] (Å)	0.006
R-factor[6]	0.197	Bond angles[5] (°)	1.4
R_{free}[4,6]	0.250	Thermal parameters[7] (Å^2)	2.9

[1] $R_{sym} = \Sigma|I - \langle I \rangle| / \Sigma I$, where I is observed intensity, <I> is average intensity obtained from multiple observations of symmetry related reflections.
[2] Phasing power = r.m.s.d. ($|F_H| / E$), where $|F_H|$ is the heavy atom structure factor amplitude and E is residual lack of closure.
[3] Data with $|F| > 4\sigma(|F|)$.
[4] R_{free} was calculated with 10% of data omitted from the structure refinement.
[5] R.m.s. deviation from ideal values.
[6] Data with $|F| > 2\sigma(|F|)$.
[7] R.m.s. deviation between the B-factors of covalently bonded main chain atomic pairs.

yielded three-dimensional structural information for an additional 15 homologous protein sequences derived from the genomes of 14 organisms.

Methods

eIF6 ribosome anti-association activity assay. *S. cerevisiae* 80S ribosomes were purified as described[30]. One optical density unit of ribosomes was incubated under dissociating conditions (0.5 M KCl, 20 mM Hepes pH 7.4, 1.5 mM $Mg(OAc)_2$, 1 mM DTT) with 100 pmoles of full length *S. cerevisiae* eIF6, *S. cerevisiae* eIF6 (1–224), or full length *M. jannaschii* aIF6. After 15 min at room temperature, the solutions were diluted five-fold to 100 mM KCl, 20 mM Hepes pH 7.4, 20 mM $Mg(OAc)_2$ in 100 μl total volume and incubated for another 15 min. The mixtures were then layered onto 10–30% sucrose gradients (in 20 mM Hepes pH 7.4, 100 mM KOAc, 5 mM $Mg(OAc)_2$), and centrifuged at 48,000 r.p.m. for 75 min in an SW55 rotor. Fractionation was followed by UV absorbance measurements.

Purification, crystallization, and heavy atom derivative preparation of *M. jannaschii* aIF6. A *M. jannaschii* genomic DNA fragment encompassing the aIF6 gene was obtained from TIGR/ATCC[31] and subcloned into a Novagen pET28a *E. coli* expression vector. Conventional protein expression and purification protocols yielded full length *M. jannaschii* aIF6. Matrix assisted laser desorption ionization mass spectrometry (MALDI-MS) confirmed that the protein was purified to homogeneity and was neither proteolyzed nor post-translationally modified (observed mass 24,948 ± 6; predicted mass 24,942). Crystallization trials yielded *M. jannaschii* aIF6 crystals *via* hanging drop vapor diffusion at 4 ºC against 31% (w/v)

PEG-MME 2000, 100 mM MES pH 5.6, 200 mM ammonium sulfate, 3 mM xylitol, 10% (v/v) glycerol, and 1 mM tris (carboxyethyl) phosphine (TCEP) with a protein concentration of 3.5 mg ml^{-1}. Plate-like crystals in the monoclinic space group C2 with two molecules per asymmetric unit (unit cell, a = 118.1 Å, b = 46.7 Å, c = 84.7 Å, β = 98.6°) appeared overnight. The crystals had a Matthews number of 2.36 and a solvent content of 48%. Heavy metal derivatives of aIF6 were prepared by soaking crystals in crystallization solution with 10 mM ethylmercury thiosalicylate for 5 h, or 1 mM osmium tetraoxide for 7 h. Prior to data collection, crystals were transferred into crystallization buffer plus 19% (v/v) ethylene glycol for 30 s, and quick frozen in liquid propane.

aIF6 structure determination. All diffraction data were collected at the X9B beamline at the Brookhaven National Laboratory National Synchrotron Light Source under standard cryogenic conditions. A MAD experiment was conducted by recording diffraction data at three X-ray wavelengths (osmium L_{III} absorption edge inflection point, peak and high energy remote) with a single high quality crystal. Data were processed using DENZO/SCALEPACK[32]. Two osmium sites were located using anomalous Patterson syntheses, and two similar mercury sites were obtained by difference Fourier syntheses. Experimental MAD/MIR phases were estimated at 1.9 Å resolution using MLPHARE[33]. Single anomalous scattering phases were calculated with MLPHARE using optimized osmium anomalous diffraction data and combined with osmium/mercury MAD/MIR phases (Table 1). After solvent flattening and noncrystallographic averaging, 98% of both protomers comprising the asymmetric unit could be built using O[34]. The starting R-factor was 44.8%, and the starting R_{free} was 46.4%, which after the first minimization round dropped to 33% and 38%, respectively. Iterative rounds of model adjustment/refinement were performed using O and CNS[35]. High resolution refinement to 1.3 Å was performed with the program SHELX[36]. The current refinement model at 1.3 Å resolution consists of residues 3–227 in both protomers and a total of 703 waters per asymmetric unit, giving a final R-factor of 13.1% and an R_{free} of 17.9% (Table 1). PROCHECK[37] showed no unfavorable (ϕ,ψ) combinations with excellent stereochemistry (overall G-value -0.1).

Purification, crystallization, and structure determination of yeast eIF6. Initial attempts to express and purify *S. cerevisiae* eIF6 were complicated by proteolytic cleavage of the divergent C-terminus. Truncation beyond residue 224 eliminated this problem, and eIF6(1–224) was purified using the aIF6 purification scheme. MALDI-MS documented no unwanted proteolysis or post-translational modification during expression and purification (observed mass 24,625 ± 5; predicted mass 24,621.8). Tetragonal bipyramidal *S. cerevisiae* eIF6 crystals in the tetragonal space group $P4_32_12$ (unit cell, a = 56.3 Å, c = 171.7 Å) with one molecule per asymmetric unit appeared after two months at room temperature *via* hanging drop vapor diffusion against unbuffered 2 M ammonium sulfate. The crystals had a Matthews number of 2.75 and a solvent content of 55%. A single native data set was collected at room temperature with a CuKα rotating anode X-ray source at 2.5 Å resolution. The structure was determined by molecular replacement with AmoRe[38], using a search model consisting of our aIF6 structure with side chains converted to Ala and α-helix H2A and the loops connecting α-helix H2C to the rest of the structure removed. Iterative rounds of model adjustment/refinement were performed using O and CNS with $2|F_o| - |F_c|$ difference Fourier syntheses generated with the program WARP (autobuilding was not used, but the difference maps were superior to those generated by CNS)[39]. The starting R-factor was 40.8% and the starting R_{free} was 44.4%. The current refinement model consists of residues 1–224, giving a working R-factor of 19.7% and an R_{free} of 25.0% with excellent stereochemistry. Water molecules were not included because of the relatively low resolution limit of the structure. Side chains for residues Glu 32, Asn 33, Asp 44, Gln 79, Gln 82, Arg 85, Asp 90, Lys 93, Arg 100, Glu 124, Gln 178, Arg 188, Tyr 202, Leu 208 and Leu 215 were modeled as Ala, because side chain electron density was not visible beyond the Cβ atom. PROCHECK[37] showed no unfavorable (ϕ,ψ) combinations with main chain and side chain stereochemical parameters better than average (overall G-value 0.3).

Coordinates. Atomic coordinates have been deposited in the Protein Data Bank (*M. jannaschii* aIF6 accession code 1G61; *S. cerevisiae* eIF6 accession code 1G62).

Acknowledgments

At the Brookhaven National Laboratory National Synchrotron Light Source, we thank Z. Dauter and K. R. Rajashankar for their help using beamline X9B. We thank G. Blobel, J. Bonanno, B.T. Chait, M.R. Chance, J. DeAngelis, R.C. Deo, M. Henderson, V. Ilyin, D. Jeruzalmi, J. Kuriyan, C. Kielkopf, C. Lima, J. Marcotrigiano, F. Melo, S. Nair, U. Pieper, G.A. Petsko, S.S. Ray, R. Sanchez, L. Shapiro, F.W. Studier, S. Swaminathan, R.M. Sweet, and W.B. Whitman for many useful discussions. We thank G. He for technical help and T. Niven for editorial assistance. S.K.B. is a member of the Howard Hughes Medical Institute. This work was supported by National Institutes of Health Grants (S.K.B. and A.S.), and by The Rockefeller University (C.M.G.). A.S. is an Alfred P. Sloan Research Fellow and an Irma T. Hirschl Trust Career Scientist, and is also support by The Merck Genome Research Institute, G. Harold and Leila Y. Mathers Charitable Foundation, and the National Science Foundation.

Received 17 August, 2000; accepted 23 October, 2000.

1. Pestova, T.V. & Hellen, C.U. Ribosome recruitment and scanning: what's new? *Trends Biochem. Sci.* **24**, 85–87 (1999).
2. Trachsel, H., Erni, B., Schreier, M.H. & Staehelin, T. Initiation of mammalian protein synthesis. II. The assembly of the initiation complex with purified initiation factors. *J. Mol. Biol.* **116**, 755–767 (1977).
3. Russell, D.W. & Spremulli, L.L. Mechanism of action of the wheat germ ribosome dissociation factor: interaction with the 60 S subunit. *Arch. Biochem. Biophys.* **201**, 518–526 (1980).
4. Russell, D.W. & Spremulli, L.L. Purification and characterization of a ribosome dissociation factor (eukaryotic initiation factor 6) from wheat germ. *J. Biol. Chem.* **254**, 8796–8800 (1979).
5. Valenzuela, D.M., Chaudhuri, A. & Maitra, U. Eukaryotic ribosomal subunit anti-association activity of calf liver is contained in a single polypeptide chain protein of M_r = 25,500 (eukaryotic initiation factor 6). *J. Biol. Chem.* **257**, 7712–7719 (1982).
6. Raychaudhuri, P., Stringer, E.A., Valenzuela, D.M. & Maitra, U. Ribosomal subunit antiassociation activity in rabbit reticulocyte lysates. Evidence for a low molecular weight ribosomal subunit antiassociation protein factor (M_r = 25,000). *J. Biol. Chem.* **259**, 11930–11935 (1984).
7. Si, K., Chaudhuri, J., Chevesich, J. & Maitra, U. Molecular cloning and functional expression of a human cDNA encoding translation initiation factor 6. *Proc. Natl. Acad. Sci. USA* **94**, 14285–14290 (1997).
8. Si, K. & Maitra, U. The *Saccharomyces cerevisiae* homologue of mammalian translation initiation factor 6 does not function as a translation initiation factor. *Mol. Cell. Biol.* **19**, 1416–1426 (1999).
9. Wood, L.C., Ashby, M.N., Grunfeld, C. & Feingold, K.R. Cloning of murine translation initiation factor 6 and functional analysis of the homologous sequence YPR016c in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **274**, 11653–11659 (1999).
10. Sanvito, F. *et al.* The beta4 integrin interactor p27(BBP/eIF6) is an essential nuclear matrix protein involved in 60S ribosomal subunit assembly. *J. Cell. Biol.* **144**, 823–837 (1999).
11. Sander, C. & Schneider, R. Database of homology-derived protein structures and structural meaning of sequence alignment. *Proteins* **9**, 56–68 (1991).
12. Altmann, M., Sonenberg, N. & Trachsel, H. Translation in *Saccharomyces cerevisiae*: initiation factor 4E-dependent cell-free system. *Mol. Cell. Biol.* **9**, 4467–4472 (1989).
13. Blum, S., Mueller, M., Schmid, S.R., Linder, P. & Trachsel, H. Translation in *Saccharomyces cerevisiae*: initiation factor 4A-dependent cell-free system. *Proc. Natl. Acad. Sci. USA* **86**, 6043–6046 (1989).
14. Biffo, S. *et al.* Isolation of a novel beta4 integrin-binding protein (p27(BBP)) highly expressed in epithelial cells. *J. Biol. Chem.* **272**, 30314–30321 (1997).
15. Sanvito, F. *et al.* Expression of a highly conserved protein, p27BBP, during the progression of human colorectal cancer. *Cancer Res.* **60**, 510–516 (2000).
16. de Pereda, J.M., Wiche, G. & Liddington, R.C. Crystal structure of a tandem pair of fibronectin type III domains from the cytoplasmic tail of integrin alpha6beta4. *EMBO J.* **18**, 4087–4095 (1999).
17. Burley, S.K. *et al.* Structural genomics: beyond the human genome project. *Nature Genet.* **23**, 151–157 (1999).
18. Hendrickson, W. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51–58 (1991).
19. Holm, L. & Sander, C. A review of the use of protein structure comparison in protein classification and function identification. *Science* **273**, 595–602 (1996).
20. Nicholls, A., Sharp, K. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).
21. Lang, D., Thoma, R., Henn-Sax, M., Sterner, R. & Wilmanns, M. Structural evidence for evolution of the beta/alpha barrel scaffold by gene duplication and fusion. *Science* **289**, 1546–1550 (2000).
22. Fulop, V. & Jones, D.T. Beta propellers: structural rigidity and functional diversity. *Curr. Opin. Struct. Biol.* **9**, 715–721 (1999).
23. Sprague, E.R., Redd, M.J., Johnson, A.D. & Wolberger, C. Structure of the C-terminal domain of Tup1, a corepressor of transcription in yeast. *EMBO J.* **19**, 3016–3027 (2000).
24. Sanchez, R. *et al.* MODBASE, a database of annotated comparative protein structure models. *Nucleic Acids Res.* **28**, 250–253 (2000).
25. Sanchez, R. & Sali, A. Comparative protein structure modeling in genomics. *J. Comp. Phys.* **151**, 388–401 (1999).
26. Nikolov, D.B. *et al.* Crystal structure of TFIID TATA-box binding protein. *Nature* **360**, 40–46 (1992).
27. Nicholls, A., Sharp, K. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281–296 (1991).
28. Conte, L.L., Chothia, C. & Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198 (1999).
29. Wall, M.A. *et al.* The structure of the G protein heterotrimer Gi alpha 1 beta 1 gamma 2. *Cell* **83**, 1047–1058 (1995).
30. Beckmann, R. *et al.* Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex. *Science* **278**, 2123–2126 (1997).
31. Bult, C.J. *et al.* Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* **273**, 1058–1073 (1996).
32. Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
33. Otwinowski, Z. Proceedings of the Daresbury CCP4 study weekend (1991).
34. Jones, T.A., Zou, J.Y., Cowan, S.W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
35. Brünger, A. *et al.* Crystallography and NMR system: a new software system for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
36. Sheldrick, G.M. Isomorphous replacement and anomalous scattering. In *Proc. CCP4 Study Weekend* (eds, Wolf, W., Evans, P.R. & Leslie, A.G.W.) 23–28 (Daresbury Laboratory, Warrington, UK; 1991).
37. Laskowski, R.J., MacArthur, M.W., Moss, D.S. & Thornton, J.M. PROCHECK: a program to check stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–290 (1993).
38. Navaza, J. AMoRe: an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
39. Lamzin, V.S. & Wilson, K.S. Automated refinement of protein models. *Acta Crystallogr. D* **49**, 129–149 (1993).
40. Henikoff, S. & Henikoff, J.G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. USA* **89**, 10915–10919 (1992).
41. Gilson, M., Sharp, K. & Honig, B. Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comput. Chem.* **9**, 327–335 (1988).