BIOCHEMISTRY

 Volume 44, Number 37 September 20, 2005

New Concepts

eIF4G and CBP80 Share a Common Origin and Similar Domain Organization: Implications for the Structure and Function of eIF4G†

Assen Marintchev and Gerhard Wagner*

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, Massachusetts 02115

Received July 1, 2005; Revised Manuscript Received August 9, 2005

ABSTRACT: Eukaryotic translation initiation factor 4G (eIF4G) plays a critical role in protein expression, and is at the center of a complex regulatory network. Together with the cap-binding protein eIF4E, it recruits the small ribosomal subunit to the 5′-end of mRNA and promotes the assembly of a functional translation initiation complex, which scans along the mRNA to the translation start codon. Human eIF4G contains three consecutive HEAT domains, as well as long unstructured regions involved in multiple protein−protein interactions. Despite the accumulating data about the structure and function of eIF4G, the mechanisms of coordination and regulation of its interactions with other factors have remained largely unknown. Here, we present evidence that eIF4G and the large subunit of the nuclear cap-binding complex, CBP80, share a common origin and domain structure. We propose that the organization of the individual domains in eIF4G and CBP80 could also be conserved. The structure of CBP80, in complex with the nuclear cap-binding protein CBP20, is used to build a model for the mutual orientation of the domains in eIF4G and their interactions with other factors. The organization of the CBP80−CBP20 complex suggests how the activity of eIF4G in translation initiation could be regulated through a dynamic network of overlapping intra- and intermolecular interactions centered around the eIF4G HEAT domains.

Control of gene expression is vital for cell proliferation, differentiation, or death. Initiation of translation is the most complex and highly regulated step in protein synthesis. It involves recruitment of the small ribosomal subunit to the mRNA with the help of a set of protein factors, recognition of the proper start codon, and subsequent binding of the large ribosomal subunit, to form a translationally active ribosome. Eukaryotic translation initiation factor 4G (eIF4G) is required for the translation of most proteins and is one of the major targets in regulation of gene expression. It is responsible for recruitment of the small, 40S ribosomal subunit to the 5′-end of the mRNA and is also required for scanning along the 5′-untranslated region (5′-UTR)[1] of the mRNA to the translation start codon. On a number of mRNAs with internal ribosome entry sites (IRES), eIF4G is required for direct binding of the 40S ribosomal subunit at or near the start codon, without scanning from the 5′-end. The association of eIF4G with the 5′-end of mRNA is mediated by an interaction with the 5′-cap-binding protein eIF4E. eIF4G also binds to the 3′-poly(A) tail binding protein (PABP). Recruitment of the 40S ribosomal subunit to mRNA is through interaction of eIF4G with ribosome-bound eIF3. The RNA helicase eIF4A is required for scanning on mRNAs contain-

† A.M. was funded by Postdoctoral Fellowship PF-02-117-01-GMC from the American Cancer Society. The work was supported by NIH Grant CA68262 to G.W.

* To whom correspondence should be addressed. Telephone: (617) 432-3213. Fax: (617) 432-4383. E-mail: wagner@hms.harvard.edu.

10.1021/bi051271v CCC: $30.25
Published on Web 08/24/2005

ing secondary structure; eIF4G helps to recruit eIF4A to the initiation complex and stimulates its helicase activity. It is aided by eIF4H and/or eIF4B, which also stimulate the helicase activity of eIF4A. These two proteins contain homologous RNA recognition motif (RRM) domains, but eIF4B is much larger (reviewed in refs *1*−*4*).

The 5′-m^{7}G cap of the nascent mRNAs in the nucleus is bound to the nuclear cap-binding complex (CBC). The CBC is also involved in mRNA processing and maturation in the nucleus, as well as in the pioneer round of translation (reviewed in ref *5*). In the cytoplasm, the 5′-m^{7}G cap is bound to eIF4E, which in turn interacts with eIF4G. The large subunits of the nuclear and cytoplasmic cap-binding complexes both have MIF4G domains with distant sequence homology (see Figure S1A of the Supporting Information), whereas the two cap-binding subunits, CBP20 and eIF4E, are unrelated. Instead, CBP20 is homologous to most of eIF4H (27% identical, 37% homologous, Figure S1E of the Supporting Information) and a segment of eIF4B (reviewed in refs *1*, *3*, and *4*).

The domain structure of eIF4G has been extensively studied. The N-terminal one-third of the ~1600-residue human eIF4G carries the binding sites for several proteins, including eIF4E and PABP (Figure 1A). No folded domains have been identified in the N-terminus of eIF4G, but individual segments appear to fold upon binding to their respective interaction partners (reviewed in refs *1*−*4*). The C-terminal two-thirds of human eIF4G are composed of three α-helical domains, initially predicted from sequence analysis (*6*−*8*) and subsequently confirmed experimentally (refs *9* and *10* and PDB entry 1UG3, S. K. Burley, unpublished results). The first one, called MIF4G (middle portion of eIF4G) or NIC (NMD2, eIF4G, CBP80), binds to RNA and eIF4A. The second domain, MA3 or MI (MA3, eIF4G), carries a second eIF4A-binding site. The third domain, called W2 or AA (acidic/aromatic) box-containing, binds to the MAP kinase-interacting kinases (MNKs), which in turn phosphorylate eIF4E and regulate its activity (Figure 1A; reviewed in refs *1*−*4*). Both the MIF4G and W2 domains are similar to the HEAT [Huntingtin, elongation factor 3, a subunit of protein phosphatase 2A (PP2A), and target of rapamycin] domain family, and are composed of five and four helical hairpins, respectively (*9*, *10*). The recently released structure of a C-terminal two-domain fragment of human eIF4G1 (MA3+W2 domains) showed that the MA3 domain is also a HEAT-like domain with five helical hairpins (PDB entry 1UG3, S. K. Burley, unpublished results). Thus, the C-terminal two-thirds of human eIF4G are composed of three consecutive HEAT domains. The same three-domain arrangement is found in the translation repressor NAT1/DAP5/p97, homologous to the C-terminal two-thirds of eIF4G (*11*). Below, we will refer to the three HEAT domains of eIF4G as HEAT-1/MIF4G, HEAT-2/MA3, and HEAT-3/W2, to provide reference to their historically established names.

Despite the progress in understanding the structure and function of eIF4G, a number of questions about the orga-

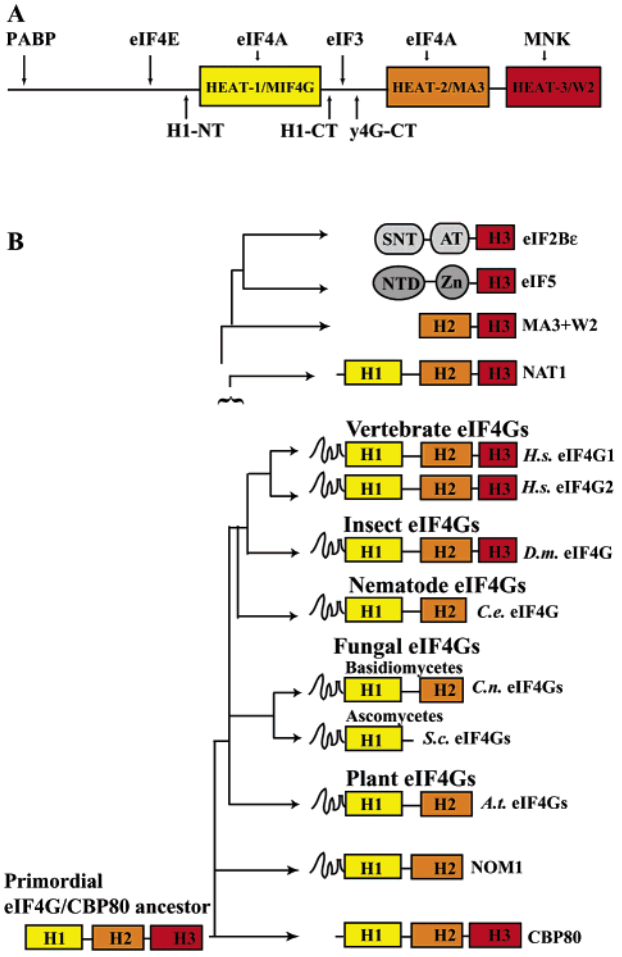

FIGURE 1: Evolution and organization of eIF4G. (A) Domain organization of human eIF4G. Sites of interactions with other proteins are indicated above the diagram. The locations of the three conserved peptides with unknown function discussed in the text are indicated below the diagram. The HEAT-1/MIF4G domain is colored yellow, the HEAT-2/MA3 domain orange, and the HEAT-3/W2 domain red. (B) Simplified evolutionary tree of eIF4G and related families of proteins. The HEAT domains are color-coded as in panel A. Abbreviations: H1, HEAT-1/MIF4G domains; H2, HEAT-2/MA3 domains and the structurally related HEAT-2 domain of CBP80; H3, HEAT-3/W2 domains and the structurally related HEAT-3 domain of CBP80; SNT, sugar-nucleotidyl transferase-like domain; AT, acyl transferase-like domain; NTD, N-terminal domain; Zn, zinc-binding domain; H.s., *Homo sapiens*; D.m., *Drosophila melanogaster*; C.e., *Ca. elegans*; C.n., *Cryptococcus neophormans*; S.c., *S. cerevisiae*; A.t., *Arabidopsis thaliana*. The brace at the NAT1 branch signifies that it is not known with certainty at what point in evolution NAT1 has diverged from eIF4G. The MA3+W2 protein family is often termed BZW (see the Supporting Information). The evolutionary tree is based on the taxonomy adopted by the NCBI web site (http://www.ncbi.nlm.nih.gov/taxonomy). The exact positions of some of the branching points are uncertain (see ref *18*), which would not affect the conclusions of this study.

[1] Abbreviations: eIF, eukaryotic translation initiation factor; 5′-UTR, 5′-untranslated region of mRNA; IRES, internal ribosome entry site; PABP, poly(A) binding protein; CBC, cap-binding complex; CBP, cap-binding protein; MIF4G domain, middle portion of eIF4G; NIC domain, NMD2, eIF4G, CBP80; MI domain, MA3, eIF4G; AA motif, acidic/aromatic motif; MNK, MAP kinase-interacting kinase; HEAT domain, Huntingtin, elongation factor 3, a subunit of protein phosphatase 2A (PP2A), and target of rapamycin; NAT1, novel APOBEC target 1; DAP5, death-associated protein 5; NOM1, nucleolar protein with MIF4G domain 1; PDB, Protein Data Bank; H1-NT peptide (of eIF4G), peptide N-terminal from the HEAT-1/MIF4G domain; H1-CT peptide (of eIF4G), peptide C-terminal from the HEAT-1/MIF4G domain; y4G-CT peptide (of eIF4G), C-terminus of yeast eIF4G; PAM, PABP-interacting motif; PAIP1, PABP-interacting protein 1; NMD2, nonsense-mediated decay protein 2; GEF, guanine nucleotide exchange factor; GAP, GTPase-activating protein.

nization and function of eIF4G remain unanswered. Many of the interactions of eIF4G with other factors are mutually cooperative or anticooperative, even though most of the individual binding sites are not overlapping, and are often far apart along the sequence of eIF4G. For example, although the eIF3 binding site on eIF4G has been reported to be the linker segment after the HEAT-1/MIF4G domain (*12*), mutations or deletions within the HEAT-1/MIF4G domain abolish eIF3 binding (*13*, *14*). Furthermore, binding of eIF4A to the HEAT-1/MIF4G domain has been reported to be cooperative with binding of eIF3 to the adjacent linker region (*12*). Another unanswered question is how the MNK kinases bound near the C-terminus of eIF4G phosphorylate eIF4E, which is bound ~900 residues away along the eIF4G sequence. Free full-length eIF4G is less active in promoting translation initiation on noncapped mRNAs than when it is bound to eIF4E or when its N-terminal one-third (containing the eIF4E-binding site) is deleted. Furthermore, the picornaviral proteases cleave eIF4G in a region located between the eIF4E-binding site and the HEAT-1/MIF4G domain, but only if eIF4E is associated with it (*15*, *16*). eIF3, which binds on the other side of the HEAT-1/MIF4G domain, increases the rate of cleavage, offering even more reasons to suspect that segments from the N-terminal one-third and the C-terminal two-thirds of eIF4G may be in contact with each other (reviewed in ref *2*). This indicates the existence of a complex network of interactions within eIF4G itself. However, even though structures of large segments of eIF4G are now known, information about such a higher-order organization is scarce.

Here, we present evidence for a common origin of eIF4G and CBP80 (the large subunits of the cytoplasmic and nuclear cap-binding complexes, respectively) from an ancestor protein, which already contained three consecutive HEAT domains. We then use the compact structure of CBP80 in complex with CBP20 to formulate hypotheses about the structural organization and function of eIF4G and suggest a mechanistic explanation for the observed cooperativity of a number of interactions of eIF4G with other factors.

Evidence for Three Consecutive HEAT Domains in a Primordial eIF4G

The HEAT-1/MIF4G domain is conserved among all eIF4G sequences. Plant eIF4G lacks the HEAT-3/W2 domain, and *Saccharomyces cerevisiae* eIF4G is missing both the HEAT-2/MA3 and HEAT-3/W2 domains. MIF4G domains are also found in other proteins, such as the N-terminal domains of CBP80 and NOM1 (nucleolar protein with MIF4G domain 1) (*7*, *8*). It has been presumed that eIF4G, CBP80, and the other MIF4G-containing proteins have evolved from a single primordial MIF4G domain, and that the additional domains found in eIF4G from higher eukaryotes have been acquired later in evolution (*7*). However, analysis of the sequence data for a number of eukaryotic genomes accumulated in recent years suggests a somewhat different scenario, one in which an eIF4G containing three consecutive HEAT domains existed early in eukaryotic evolution, but the last one or two HEAT domains were subsequently lost in some lineages (Figure 1B; see the Supporting Information for details). Here, we have limited our attention to proteins relevant to the analysis of the origin of eIF4G.

Whereas only eIF4G from higher eukaryotes contains all three HEAT domains, the HEAT-2/MA3 domain is found not only in plant eIF4G but also in a subset of the fungal eIF4Gs (all available eIF4G sequences from basidiomycetes) and in some of the eIF4Gs from lower eukaryotes that have branched out early (e.g., *Dictyostelium*), before the separation of the crown group *Metazoa*, fungi, and plants.

Proteins belonging to other MIF4G domain families also contain consecutive HEAT-1/MIF4G and HEAT-2/MA3 domains (*8*), or HEAT-2/MA3 and HEAT-3/W2 domains (Figure 1B; see the Supporting Information). One such example is NOM1, which was also reported to interact with eIF4A (most likely the nuclear eIF4A-3 isoform involved in splicing) (*17*).

Therefore, at least the HEAT-2/MA3 domain must have been present in early eIF4G, and even in the common ancestor of eIF4G and NOM1, because the probability of the same domain having been acquired independently several times in different lineages during evolution is negligibly small compared to the probability of a domain being lost in several lineages. This appears to be similar to the loss of half of the eIF3 subunits in some yeast species (e.g., *S. cerevisiae*), but not in others (e.g., *Schizosaccharomyces pombe*). An example of a more recent loss of the HEAT-3/W2 domain of eIF4G is its absence in eIF4G from nematodes (e.g., *Caenorhabditis elegans*, Figure 1B). NAT1, which contains all three HEAT domains, is also absent in nematodes; in this case, the entire protein seems to have been lost in *C. elegans*. These observations are consistent with genome-wide analyses indicating that present-day eukaryotes evolved from a fairly complex ancestor, and that evolution of fungi, nematodes, and insects was accompanied by substantial gene loss (see ref *18*).

Common Domain Organization between eIF4G and CBP80

The structure of the CBC, composed of CBP80 and CBP20, reveals that CBP80 consists of three consecutive HEAT domains forming a compact structure in the CBC (*19*−*21*) (Figure 2A). As expected, the N-terminal MIF4G domain is most similar to that of the homologous HEAT-1/MIF4G domain of eIF4G, with a *Z* score for structure similarity from the DALI server (*22*) of 13.2 (where *Z* scores of >2.0 are considered significant). But, what is even more important for the present discussion is the fact that the structure of the C-terminal HEAT domain of CBP80 is most similar to that of the HEAT/W2 domain of eIF2Bϵ (*10*), with a *Z* score of 11.4 (Table S1 of the Supporting Information), indicating that the homologous C-terminal W2 domain of eIF4G is also structurally related to the C-terminal HEAT domain of CBP80. Thus, the first HEAT/MIF4G domains of eIF4G and CBP80 are related in both structure and sequence, and the third HEAT domains of eIF4G (W2) and CBP80 are structurally related, despite the lack of detectable sequence homology.

Therefore, it is logical to extend the above hypothesis about the common ancestor of eIF4G and NOM1 even further, to include their more distant relative CBP80, and propose that the common ancestor of eIF4G, NOM1, and CBP80 already had three consecutive HEAT domains (Figure

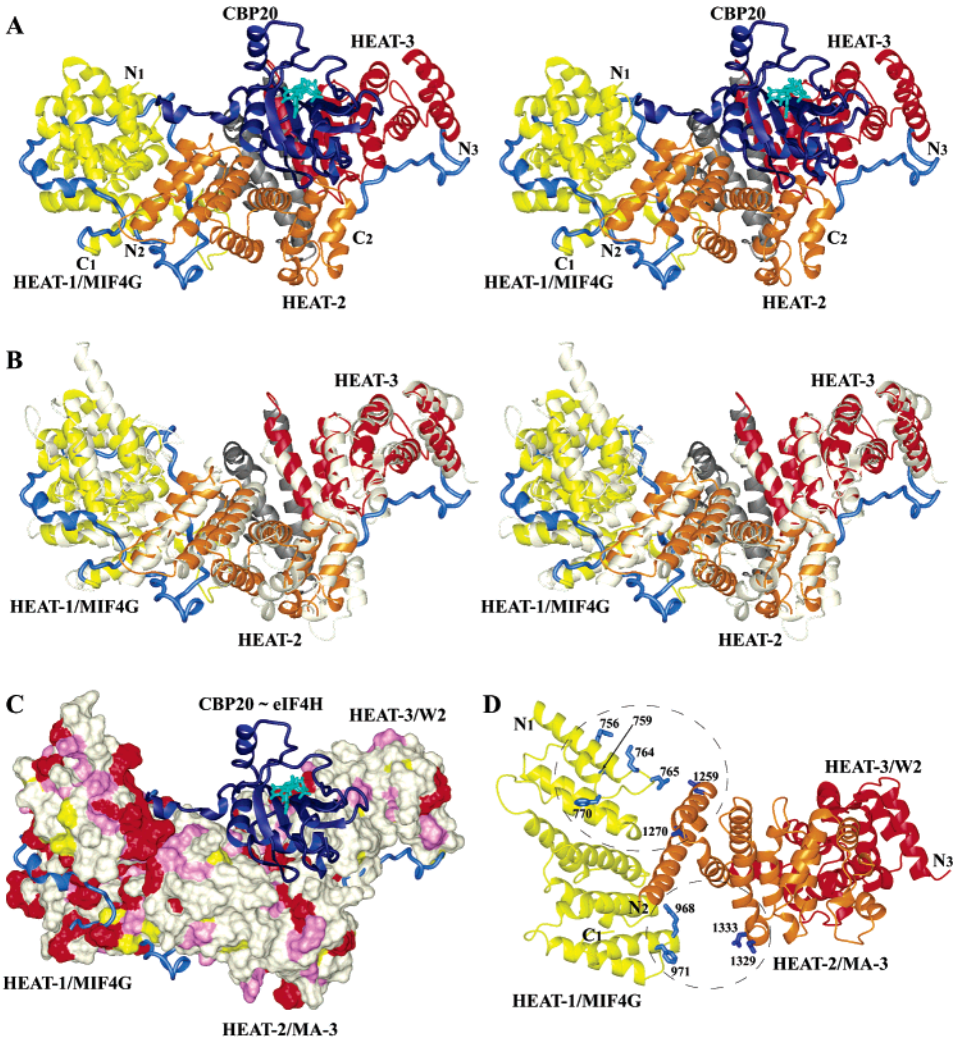

FIGURE 2: Structural parallels between CBP80 and eIF4G. (A) Stereoview of the human nuclear cap-binding complex (CBC) composed of CBP80 and CBP20 [PDB entry 1HT2 (*20*)]. The HEAT-1/MIF4G domain is colored yellow and the HEAT-2 domain orange; the first four helical hairpins of the HEAT-3 domain are colored red, and the last two hairpins (which do not have counterparts in eIF4G) are colored in gray. The interdomain linkers are colored blue. CBP20 is colored purple, and the cap analogue is colored cyan. The N- and C-termini of the domains are labeled, where visible. (B) Structure alignment in stereoview of the three HEAT domains of eIF4G with the corresponding domains of CBP80. The alignment was based on results from the DALI server (*22*), with rmsds of 3.1 Å (over 134 residues), 2.4 Å (over 111 residues), and 1.4 Å (over 99 residues) for the first, second, and third HEAT domains, respectively. The eIF4G domains are colored light gray. CBP80 is as in panel A; CBP20 and the cap analogue are not shown. The HEAT-1/MIF4G domain is from PDB entry 1HU3 (*9*); the HEAT-2/MA3 and HEAT-3/W2 domains are from PDB entry 1UG3 (S. K. Burley, unpublished results). Note that the resulting orientation of the HEAT-2/MA3 and HEAT-3/W2 domains differs from that observed in PDB entry 1UG3. The structure in 1UG3 has a C-terminal deletion of 34 residues, which was modeled here using the corresponding region in the structure of the homologous W2 domain from eIF2Bϵ, PDB entry 1PAQ (*10*). (C) Conservation of the putative eIF4H-contacting surfaces in eIF4G. The three HEAT domains of eIF4G (in surface representation) are oriented as in panel B. The surface is color-coded by sequence conservation in vertebrate eIF4G and NAT1 homologues: residues identical in $\geq$80% of the sequences are colored red, residues conserved in $\geq$80% of the sequences violet, and residues hydrophobic in $\geq$80% of the sequences yellow. The structure of CBP20 was taken directly from the CBC structure in panel A, and its orientation was not modified. Since the sequence of eIF4H is $\sim$30% identical with that of CBP20, its structure is expected to be similar. The interdomain linkers of CBP80 are shown in their original orientation for reference, since the corresponding region in eIF4G could have a similar (but not necessarily identical) location. (D) Mutations in eIF4G that affect eIF4A binding. The three HEAT domains of eIF4G are colored as the corresponding domains of CBP80 in panel A: HEAT-1/MIF4G colored yellow, HEAT-2/MA3 colored orange, and HEAT-3/W2 colored in red. The sites of mutations in the HEAT-1/MIF4G and HEAT-2/MA3 domains of eIF4G that affect eIF4A binding (*9*, *13*, *26*) are colored blue and purple, respectively. The two clusters of mutated residues are circled, and the residue numbers are shown. The residue numbering for the HEAT-1/MIF4G domain corresponds to the structure of the HEAT-1/MIF4G domain of human eIF4G2 [PDB entry 1HU3 (*9*)], whereas numbering of the HEAT-2/MA3 domain corresponds to that in the structure of the HEAT-2/MA3 and HEAT-3/W2 domains of human eIF4G1 (PDB entry 1UG3, S. K. Burley, unpublished results). Residues 756, 859, and 765 are from ref *9*. Residues 764 and 770 and residues 968 and 971 correspond to mutants M1 and M4, respectively, from ref *13*. Please note that the M1 mutant also contains a mutation in the buried residue L762 (not shown), which could affect the domain structure. The mutations in the HEAT-2/MA3 domain are single-residue mutations described in ref *26*. The mutual orientation of the HEAT domains of eIF4G is as described above, but the structure is rotated slightly (up and to the right) compared to the previous panels.

1B). Since members of the eIF4G, NOM1, and CBP80 families can be found in most eukaryotic lineages, including many primitive eukaryotes, the divergence of these families must have happened very early in eukaryotic evolution. The recently released structure of the C-terminal two domains (the MA3 and W2 domains) of human eIF4G (PDB entry

1UG3, S. K. Burley, unpublished results) provides strong support for the above hypothesis: not only does it show that human eIF4G contains three consecutive HEAT domains, but it also allows comparisons of the individual domain structures. As seen in Table S1 of the Supporting Information, the first, second, and third HEAT domains of CBP80 are structurally most similar to the first, second, and third HEAT domains of human eIF4G, respectively (*Z* scores of 13.2, 10.9, and 14.2 for the HEAT-1, HEAT-2, and HEAT-3 domain pairs, respectively), which essentially confirms the hypothesis about a common three-HEAT domain ancestor for eIF4G and CBP80. This hypothesis is also in line with the observation that importin-β, a large horseshoe-shaped HEAT repeat protein (PDB entry 1QGR), is among the closest structural homologues of eIF4G and CBP80 with a *Z* score of 8.5.

Hypotheses about the Structure and Function of eIF4G

In addition to offering insights into the evolution of proteins involved in various aspects of mRNA metabolism, the common origins and domain organization of eIF4G and CBP80 have direct practical implications, because they allow us to use information available about one of the proteins to formulate testable hypotheses about the structure and function of the other. In the structure of the complex of CBP80 with CBP20, the three HEAT domains of CBP80 pack together. The interdomain linker regions in CBP80 interact with the numerous grooves formed by the helical hairpins, and the linker between HEAT-1 and HEAT-2 wraps around the HEAT-1 domain (*19*) (Figure 2A). CBP20 is responsible for binding to the 5′-m^{7}G cap. Cap recognition is mediated by the RRM domain and flanking regions (*20*, *21*). Free CBP20 is mostly unstructured (*20*) and needs to be bound to CBP80, for folding and stable interaction with the cap. It binds CBP80 through an extensive contact surface, mostly to the second and third HEAT domains. Cap binding causes folding of additional segments of CBP20, and in the cap-bound complex, CBP20 also interacts with the first HEAT domain (*20*, *21*). CBP20 is homologous to eIF4H (see Figure S1E of the Supporting Information), an initiation factor that, together with eIF4G, stimulates the helicase activity of eIF4A (*23*). Interestingly, free eIF4H also appears to be mostly disordered (H. Matsuo and G. Wagner, unpublished observations). The structure of free CBP80 in the absence of CBP20 is not known, and it is not clear if binding to CBP20 influences the orientation of the HEAT domains in CBP80 in the complex.

By analogy with the CBC, we can formulate the following hypotheses about eIF4G, and try to offer explanations for some of the unanswered questions about the structure and function of eIF4G:

(1) The orientation of the three HEAT domains of eIF4G may be similar to that found in CBP80 in its complex with CBP20 (Figure 2B). The linker between the HEAT-1/MIF4G and HEAT-2/MA3 domains of eIF4G may wrap around the HEAT-1/MIF4G domain and also thread between the two domains, as observed in CBP80. For example, this would bring the linker segment reported to bind eIF3 (*12*) close to the HEAT-1/MIF4G domain and explain why mutations and deletions in the HEAT-1 domain and the linker affect eIF3 binding (*13*, *14*).

(2) The mutual orientation of the eIF4G HEAT domains need not be fixed: it may be dynamic and dependent on the binding of other factors.

(3) Peptides from other segments of eIF4G may also interact with the HEAT domains. The interaction sites for individual peptides may overlap with each other and with the binding surfaces for other factors. Such an intricate interaction network offers a plausible mechanism for the observed mutual cooperativity of most interactions of eIF4G with other factors. In particular, in free eIF4G, parts of the eIF4E-binding region (located ~100 residues N-terminal from the HEAT-1/MIF4G domain) could interact with surfaces of the HEAT domain, and folding upon eIF4E binding (*24*) may change these interactions. Such a scenario can explain the effects of eIF4E binding or cleavage of the eIF4G N-terminus on the activity of eIF4G on uncapped mRNAs, as well as of eIF4E and eIF3 binding on cleavage by picornaviral proteases between the eIF4E-binding site and the HEAT-1/MIF4G domain (reviewed in ref *2*).

(4) If the HEAT domains in eIF4G are oriented as in CBP80, the MNK kinase-binding HEAT-3/W2 domain would be brought into the proximity of the HEAT-1/MIF4G domain (Figure 2B). Then it would be easier for eIF4E, bound ~100 residues N-terminal from the HEAT-1 domain, to contact MNK. This hypothesis can be taken one step further, by suggesting that a peptide in or around the eIF4E-binding region of eIF4G could be associated with the HEAT-1/MIF4G domain even after eIF4E is bound. This in turn would bring eIF4E even closer to its kinase.

The evidence listed above leaves little doubt that eIF4G and CBP80 have evolved from a common ancestor with three consecutive HEAT domains (see Figure 1B). However, the further hypotheses that structural organization and interactions of the HEAT domains are also conserved between eIF4G and CBP80 necessitate additional analysis. Three issues in particular need to be addressed. (1) The interdomain orientation between the HEAT-2/MA3 and HEAT-3/W2 domains of eIF4G observed in the crystal structure of the two-domain construct (PDB entry 1UG3, S. K. Burley, unpublished results) is different from that predicted by analogy to CBP80. (2) The linker length between the HEAT-1/MIF4G and HEAT-2 domains in CBP80 is ~60 residues, whereas the corresponding linker in eIF4G is ~250 residues. (3) The HEAT-3 domain of CBP80 contains six helical hairpins, whereas the HEAT-3/W2 domain is missing the last two, which reduces the number of expected interdomain contacts (see Figure 2A,B).

If an interaction is conserved among homologous proteins, the corresponding surfaces would be expected to have similar (and complementary) overall shapes and a certain degree of sequence conservation. Sequence analysis and alignment of the structures of the three HEAT domains of eIF4G demonstrate that the surfaces corresponding to interdomain contacts in CBP80 show significant conservation in both sequence and shape (see Figures S1 and S2 of the Supporting Information), indicating that the three HEAT domains of eIF4G could be oriented like those of CBP80. Furthermore, there is no obvious sequence conservation at the interdomain interface between the HEAT-2/MA3 and HEAT-3/W2 domains of eIF4G found in the crystal structure of the two-domain construct (PDB entry 1UG3, S. K. Burley, unpublished results), suggesting that the observed contacts could

be due to crystal packing and not represent the actual interface. It should also be noted that the two-domain eIF4G construct (HEAT-2 and HEAT-3) used for structure determination is C-terminally truncated by 34 residues, corresponding to most of the fourth helical hairpin of the W2 domain, and therefore is missing one of the predicted interdomain contact points (see Figure 2A and Figures S1B and S2 of the Supporting Information). The missing fragment shares a high degree of sequence homology (~35% identical and ~50% similar) with the C-terminus of the HEAT/W2 domain of eIF2Bϵ, and it is safe to assume that its structure is very similar. Alternatively, the interdomain orientation in eIF4G could change depending on the presence or absence of interacting partners.

Examination of the linkers between the first two HEAT domains of CBP80 and eIF4G indicates that the sequence homology between the corresponding HEAT-1/MIF4G domains of the two proteins extends at least 15 residues beyond the end of the domains. Remarkably, the corresponding segment in eIF4G was part of the MIF4G-containing fragment resistant to limited proteolysis (*9*). Here we call this peptide H1-CT (C-terminal from the HEAT-1/MIF4G domain). The next ~15 residues of the linker are highly conserved in eIF4G homologues from vertebrates (see Figure S1A,D of the Supporting Information). Therefore, the interdomain linker in eIF4G seems to wrap at least halfway around the HEAT-1/MIF4G domain. Although large portions of the ~250-residue interdomain linker in eIF4G are not conserved, individual segments in it are nearly identical among eIF4Gs from vertebrates (see Figure S1D of the Supporting Information) and could contact either the first (MIF4G) or the second (MA3) HEAT domain, or be buried between the two, as seen in the CBP80 structure (Figure 2A). These conserved sequences are part of the reported eIF3-binding site (*12*), and a C-terminal truncation into this region of human eIF4G abolished eIF3 binding by a larger eIF4G fragment that also contained the MIF4G domain (*14*). Surprisingly, there is no obvious conservation of this linker region between eIF4G and NAT1, which also binds eIF3. Weak sequence homology is observed in this region between vertebrate eIF4Gs and eIF4Gs from fungi that have retained the HEAT-2/MA3 domain (*Basidiomycetes*) but not those that have lost it (e.g., *S. cerevisiae*). As expected, the domain surfaces, predicted to be in contact with peptides from the linker, are also conserved (see Figures S1 and S2 of the Supporting Information). Surprisingly, surfaces in the MIF4G domain predicted to form interdomain and/or domain−peptide contacts appear to be conserved even in yeast eIF4G (not shown), which has only the first (MIF4G) HEAT domain and an ~100-residue C-terminal peptide, corresponding in sequence to the first half of the interdomain linker of human eIF4G. This observation suggests that the C-terminal peptide in yeast eIF4G wraps around the MIF4G HEAT domain even in the absence of the second and third HEAT domains or that the corresponding surface interacts with other proteins, other segments of eIF4G, or both.

The surfaces in eIF4G, corresponding to the CBP20-binding surfaces in CBP80, are remarkably similar in overall shape to those in CBP80 (Figure 2C). Therefore, if the domain arrangement in eIF4G is indeed similar to that in CBP80, eIF4H could bind to eIF4G in a manner similar to that of CBP20 binding to CBP80. This is particularly interesting because such an interaction could explain how eIF4G and eIF4H can stimulate the helicase activity of eIF4A simultaneously. No direct binding between eIF4G and eIF4H has been reported, to our knowledge. Therefore, this putative eIF4G−eIF4H interaction would likely be much weaker than that between CBP80 and CBP20, and could also depend on other binding partners, RNA and eIF4A, for example.

Indications for a Contiguous eIF4A-Binding Site Formed by the HEAT-1 and HEAT-2 Domains

Several of the binding partners of eIF4G have been reported to bind directly to one or more of the HEAT domains. While the exact binding surfaces are not known, there are mutation data for some of them: RNA binding to the HEAT-1/MIF4G (*9*, *25*) and eIF4A binding to the HEAT-1 (*9*, *13*, *25*) and HEAT-2 (*26*) domains. The vast majority of the mutations affecting RNA and eIF4A binding are solvent-exposed in the model presented here. Remarkably, mutations in the HEAT-1 and HEAT-2 domains that affected the interaction with eIF4A come close to each other in the model, forming two clusters, each composed of residues from both HEAT domains (Figure 2D). This suggests that two sets of adjacent or partially overlapping determinants on eIF4A could be recognized by the two HEAT domains of eIF4G. A greater degree of overlap and competition is somewhat unlikely, because the translation repressor Pdcd4 (which contains two consecutive HEAT/MA3 domains) was reported to compete for eIF4A binding with the HEAT-2/MA3 but not with the HEAT-1/MIF4G domain of eIF4G (*27*). The two domains of eIF4A could contact both regions simultaneously (*28*), but there is not enough information to predict from which side it would approach eIF4G, because the eIF4A-binding clusters map to nearly opposite surfaces on the HEAT domains (Figure 2D). It is also possible that there is a single contiguous eIF4A-binding surface on the HEAT-1/MIF4G and/or HEAT-2/MA-3 domains of eIF4G, but that the regions between the two clusters have not been sampled extensively by site-directed mutagenesis.

Sequence Conservation in Linker Peptides Suggests Interactions with the HEAT Domains and/or Other Proteins

The two best-studied regions from the largely unstructured N-terminus of eIF4G are the eIF4E- and PABP-binding sites. Both human and yeast (*S. cerevisiae*) eIF4G bind eIF4E and PABP; the respective binding sites are in similar locations, and the eIF4E-binding site is highly conserved. However, there is hardly any conservation between the PABP-binding sequences. As discussed above, there are indications that the eIF4E-binding region may interact with the HEAT-1 domain and/or a nearby segment of eIF4G. In the absence of direct functional data, it is difficult to speculate whether the PABP-binding site would also contact any of the HEAT domains.

Analysis of eIF4G sequences from various organisms reveals three short peptides conserved among most eIF4G homologues (see Figure 1A and Figure S1C,D of the Supporting Information), although no function has been ascribed to the corresponding regions of eIF4G. One of them is located just N-terminal to the HEAT-1/MIF4G domain, and we will designate it as H1-NT. Its location appears to

be similar to that of the PABP-interacting motif 2 (PAM-2) found at the N-terminus of the HEAT/MIF4G domain of PAIP1 (PABP-interacting protein 1) (*29*). The sequence of the eIF4G peptide resembles that of the PAM-2 motif, indicating that the two may have a common origin but have diverged. The second universally conserved eIF4G peptide, H1-CT, is found at the C-terminus of the HEAT-1/MIF4G domain. As discussed above, H1-CT is conserved not only among eIF4G sequences but also between eIF4G and CBP80 (see Figure S1A,D of the Supporting Information) and probably contacts the HEAT-1 domain. The third conserved peptide is at the C-terminus of *S. cerevisiae* eIF4G and in the corresponding location in the linker between the HEAT-1/MIF4G and HEAT-2/MA3 domains of human eIF4G. Here, we will call this peptide y4G-CT. Again, no specific function has been ascribed to this region, but a C-terminal deletion including this segment of yeast eIF4G was reported to affect eIF1 and (to a lesser degree) eIF5 binding, even though the primary eIF1/eIF5-binding site was found to be N-terminal from the MIF4G domain (*30*).

To our knowledge, potential binding of the eIF4G peptides described above to the HEAT domains or other proteins, or interaction of the HEAT domains with each other or with other regions of eIF4G, has not been directly tested. The MA3/Pdcd4 translation repressor, consisting of two consecutive HEAT/MA3 domains, was reported to bind to the HEAT-1/MIF4G domain of eIF4G (*27*), but it is not clear if the same is true for the isolated HEAT-1/MIF4G and HEAT-2/MA3 domains of eIF4G, which would be consistent with the hypotheses presented above.

Evolution of eIF4G and Translation Initiation

It appears that the evolution of eIF4G and the other related families of proteins has involved numerous events of duplication and loss of individual domains and entire genes. Among the closer eIF4G homologues, it appears that all domains can be found alone or in combination in individual proteins, whereas all known CBP80 sequences contain all three HEAT domains. Why are the evolutionary fates of eIF4G and CBP80 so different? One factor that may have favored greater divergence of the eIF4G lineage in both domain organization and function is that eIF4G is often present in more than one copy per genome. Another possible explanation is that the HEAT-3 domain of CBP80 contains two more helical hairpins at its C-terminus (Figure 2A,B), which form a number of contacts with the rest of the protein and likely contribute to the overall compact structure. In the absence of such contacts in eIF4G, it may have been easier to lose one or two domains from the C-terminus, without disrupting the structure and stability of the entire protein, provided the lost domains did not perform essential function or had been first duplicated in another protein. For example, the W2 domains of eIF2Bϵ and eIF5 are close homologues of the HEAT-3/W2 domain of eIF4G (which is lost most frequently in evolution) and likely resulted from gene duplication of the corresponding domain of eIF4G (see Figure 1B and the Supporting Information). A certain degree of functional redundancy may have existed shortly after the duplication of the W2 domain into eIF2Bϵ and eIF5, which could explain the relatively frequent loss of the W2 domain from eIF4G in a number of species. eIF2B and eIF5 are the guanine nucleotide exchange factor (GEF) and GTPase-activating protein (GAP) of eIF2, respectively, and both their HEAT/W2 domains bind to the N-terminus of the eIF2β subunit (reviewed in refs *1*, *4*, *31*, and *32*). It is tempting to speculate that the W2 domain in the primordial eIF4G may have interacted with eIF2β, before the emergence of eIF5 and eIF2B. No binding of eIF4G to eIF2 has been reported, even though the two AA (acidic/aromatic) motifs implicated in eIF2β binding are well conserved between the HEAT-3/W2 domain of eIF4G and the W2 domains of eIF2Bϵ and eIF5. Instead, the HEAT-3/W2 domain of eIF4G was reported to bind to the MNK kinases. Since the AA motifs in the eIF4G HEAT-3/W2 domain are partially occluded in our model, the domain might still be able to bind to eIF2β, but only upon structural rearrangement. We are not aware if the isolated eIF4G W2 domain has been tested for eIF2β binding.

In summary, our findings about the remarkable similarity between CBP80 and eIF4G allow experimental information available for CBP80 to be used to predict the interdomain orientation in eIF4G and offer insights into the function of eIF4G and the mechanisms by which interactions within eIF4G and with other factors are coordinated.

ACKNOWLEDGMENT

We thank Dr. Stephen Burley for releasing the coordinates of the C-terminal fragment of human eIF4G (PDB entry 1UG3) before publication.

SUPPORTING INFORMATION AVAILABLE

Details about the sequence and structure analyses employed in this work. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Hershey, J. W. B., and Merrick, W. C. (2000) The pathway and mechanism of initiation of protein synthesis, in *Translational Control* (Sonenberg, N., Hershey, J. W. B., and Mathews, M. B., Eds.) pp 33−88, Cold Spring Harbor Laboratory Press, Plainview, NY.
2. Jackson, R. J. (2000) Comparative view of initiation site selection mechanisms, in *Translational Control* (Sonenberg, N., Hershey, J. W. B., and Mathews, M. B., Eds.) pp 127−184, Cold Spring Harbor Laboratory Press, Plainview, NY.
3. Prevot, D., Darlix, J. L., and Ohlmann, T. (2003) Conducting the initiation of protein synthesis: The role of eIF4G, *Biol. Cell 95*, 141−156.
4. Marintchev, A., and Wagner, G. (2005) Translation initiation: Structures, mechanisms and evolution, *Q. Rev. Biophys.*, 1−88.
5. Maquat, L. E. (2004) Nonsense-mediated mRNA decay: Splicing, translation and mRNP dynamics, *Nat. Rev. Mol. Cell Biol. 5*, 89−99.
6. Koonin, E. V. (1995) Multidomain organization of eukaryotic guanine nucleotide exchange translation initiation factor eIF-2B subunits revealed by analysis of conserved sequence motifs, *Protein Sci. 4*, 1608−1617.
7. Aravind, L., and Koonin, E. V. (2000) Eukaryote-specific domains in translation initiation factors: Implications for translation regulation and evolution of the translation system, *Genome Res. 10*, 1172−1184.
8. Ponting, C. P. (2000) Novel eIF4G domain homologues linking mRNA translation with nonsense-mediated mRNA decay, *Trends Biochem. Sci. 25*, 423−426.
9. Marcotrigiano, J., Lomakin, I. B., Sonenberg, N., Pestova, T. V., Hellen, C. U., and Burley, S. K. (2001) A conserved HEAT domain within eIF4G directs assembly of the translation initiation machinery, *Mol. Cell 7*, 193−203.
10. Boesen, T., Mohammad, S. S., Pavitt, G. D., and Andersen, G. R. (2004) Structure of the catalytic fragment of translation initiation

factor 2B and identification of a critically important catalytic residue, *J. Biol. Chem. 279*, 10584−10592.
11. Imataka, H., Olsen, H. S., and Sonenberg, N. (1997) A new translational regulator with homology to eukaryotic translation initiation factor 4G, *EMBO J. 16*, 817−825.
12. Korneeva, N. L., Lamphear, B. J., Hennigan, F. L., and Rhoads, R. E. (2000) Mutually cooperative binding of eukaryotic translation initiation factor (eIF) 3 and eIF4A to human eIF4G-1, *J. Biol. Chem. 275*, 41369−41376.
13. Imataka, H., and Sonenberg, N. (1997) Human eukaryotic translation initiation factor 4G (eIF4G) possesses two separate and independent binding sites for eIF4A, *Mol. Cell. Biol. 17*, 6940−6947.
14. Morino, S., Imataka, H., Svitkin, Y. V., Pestova, T. V., and Sonenberg, N. (2000) Eukaryotic translation initiation factor 4E (eIF4E) binding site and the middle one-third of eIF4GI constitute the core domain for cap-dependent translation, and the C-terminal one-third functions as a modulatory region, *Mol. Cell. Biol. 20*, 468−477.
15. Lamphear, B. J., Kirchweger, R., Skern, T., and Rhoads, R. E. (1995) Mapping of functional domains in eukaryotic protein synthesis initiation factor 4G (eIF4G) with picornaviral proteases. Implications for cap-dependent and cap-independent translational initiation, *J. Biol. Chem. 270*, 21975−21983.
16. Haghighat, A., Svitkin, Y., Novoa, I., Kuechler, E., Skern, T., and Sonenberg, N. (1996) The eIF4G-eIF4E complex is the target for direct cleavage by the rhinovirus 2A proteinase, *J. Virol. 70*, 8444−8450.
17. Simmons, H. M., Ruis, B. L., Kapoor, M., Hudacek, A. W., and Conklin, K. F. (2005) Identification of NOM1, a nucleolar, eIF4A binding protein encoded within the chromosome 7q36 breakpoint region targeted in cases of pediatric acute myeloid leukemia, *Gene 347*, 137−145.
18. Koonin, E. V., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Krylov, D. M., Makarova, K. S., Mazumder, R., Mekhedov, S. L., Nikolskaya, A. N., Rao, B. S., Rogozin, I. B., Smirnov, S., Sorokin, A. V., Sverdlov, A. V., Vasudevan, S., Wolf, Y. I., Yin, J. J., and Natale, D. A. (2004) A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes, *Genome Biol. 5*, R7.
19. Mazza, C., Ohno, M., Segref, A., Mattaj, I. W., and Cusack, S. (2001) Crystal structure of the human nuclear cap binding complex, *Mol. Cell 8*, 383−396.
20. Mazza, C., Segref, A., Mattaj, I. W., and Cusack, S. (2002) Large-scale induced fit recognition of an m(7)GpppG cap analogue by the human nuclear cap-binding complex, *EMBO J. 21*, 5548−5557.
21. Calero, G., Wilson, K. F., Ly, T., Rios-Steiner, J. L., Clardy, J. C., and Cerione, R. A. (2002) Structural basis of m7GpppG binding to the nuclear cap-binding protein complex, *Nat. Struct. Biol. 9*, 912−917.
22. Holm, L., and Sander, C. (1993) Protein structure comparison by alignment of distance matrices, *J. Mol. Biol. 233*, 123−138.
23. Richter-Cook, N. J., Dever, T. E., Hensold, J. O., and Merrick, W. C. (1998) Purification and characterization of a new eukaryotic protein translation factor. Eukaryotic initiation factor 4H, *J. Biol. Chem. 273*, 7579−7587.
24. Gross, J. D., Moerke, N. J., von der Haar, T., Lugovskoy, A. A., Sachs, A. B., McCarthy, J. E., and Wagner, G. (2003) Ribosome loading onto the mRNA cap is driven by conformational coupling between eIF4G and eIF4E, *Cell 115*, 739−750.
25. Lomakin, I. B., Hellen, C. U., and Pestova, T. V. (2000) Physical association of eukaryotic initiation factor 4G (eIF4G) with eIF4A strongly enhances binding of eIF4G to the internal ribosomal entry site of encephalomyocarditis virus and is required for internal initiation of translation, *Mol. Cell. Biol. 20*, 6019−6029.
26. Yang, H. S., Cho, M. H., Zakowicz, H., Hegamyer, G., Sonenberg, N., and Colburn, N. H. (2004) A novel function of the MA-3 domains in transformation and translation suppressor Pdcd4 is essential for its binding to eukaryotic translation initiation factor 4A, *Mol. Cell. Biol. 24*, 3894−3906.
27. Yang, H. S., Jansen, A. P., Komar, A. A., Zheng, X., Merrick, W. C., Costes, S., Lockett, S. J., Sonenberg, N., and Colburn, N. H. (2003) The transformation suppressor Pdcd4 is a novel eukaryotic translation initiation factor 4A binding protein that inhibits translation, *Mol. Cell. Biol. 23*, 26−37.
28. Oberer, M., Marintchev, A., and Wagner, G. (2005) Structural basis for the enhancement of eIF4A helicase activity by eIF4G, *Genes Dev.* (in press).
29. Roy, G., De Crescenzo, G., Khaleghpour, K., Kahvejian, A., O'Connor-McCourt, M., and Sonenberg, N. (2002) Paip1 interacts with poly(A) binding protein through two independent binding motifs, *Mol. Cell. Biol. 22*, 3769−3782.
30. He, H., von der Haar, T., Singh, C. R., Ii, M., Li, B., Hinnebusch, A. G., McCarthy, J. E., and Asano, K. (2003) The yeast eukaryotic initiation factor 4G (eIF4G) HEAT domain interacts with eIF1 and eIF5 and is involved in stringent AUG selection, *Mol. Cell. Biol. 23*, 5431−5445.
31. Hinnebusch, A. G. (2000) Mechanism and regulation of initiator methionyl-tRNA binding to ribosomes, in *Translational Control* (Sonenberg, N., Hershey, J. W. B., and Mathews, M. B., Eds.) pp 185−244, Cold Spring Harbor Laboratory Press, Plainview, NY.
32. Dever, T. E. (2002) Gene-specific regulation by general translation factors, *Cell 108*, 545−556.

BI051271V