

## review

## Eukaryotic translation initiation: there are (at least) two sides to every story

Alan B. Sachs<sup>1</sup> and Gabriele Varani<sup>2</sup>

The eukaryotic cap and poly(A) tail binding proteins, eIF4E and Pab1p, play important roles in the initiation of protein synthesis. The recent structures of the complex of eIF4E bound to the methylated guanosine (cap) found at the 5' end of messenger RNA (mRNA), the complex of eIF4E bound to peptide fragments of two related translation factors (eIF4G and 4E-BP1), and the complex of the N-terminal fragment of Pab1p bound to polyadenylate RNA have revealed that eIF4E and Pab1p contain at least two distinct functional surfaces. One surface is used for binding mRNA, and the other for binding proteins involved in translation initiation.

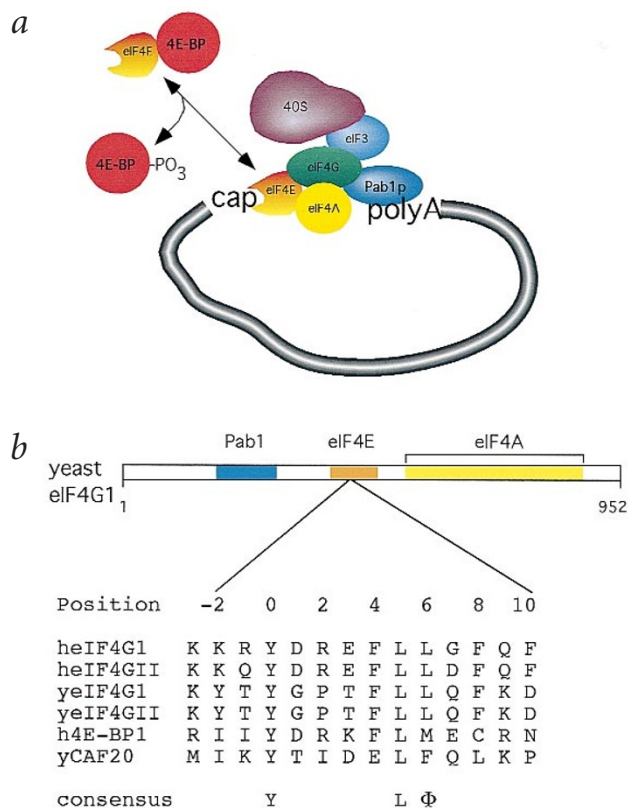
## Translation initiation in eukaryotes

The expression of many eukaryotic genes is regulated at the level of translation initiation. The first step in the initiation process is binding of the small ribosomal subunit (40S) to the mRNA. Since this is often the rate limiting step in initiation, it is also often the target of regulation. This review will discuss recent progress in understanding the proteins involved in mediating ribosome binding to eukaryotic mRNA.

In prokaryotes, base pairing between rRNA and the 'Shine-Delgarno' sequence preceding the initiation codon of each open reading frame plays a dominant role in the ribosome binding step. In eukaryotes, however, the binding step is much more complex. The model that best represents the available data (Fig. 1a) posits that the multisubunit eukaryotic initiation factor 3 (eIF3), which is associated with the 40S ribosomal subunit, also binds to the mRNA-associated eIF4G protein and thereby links the ribosome to the mRNA (reviewed in ref. 1). Understanding how eIF4G becomes associated with mRNA is thus central to understanding the processes underlying ribosome binding and translational control.

In most cases, eIF4G does not only bind directly to mRNA, it also associates with the cap structure at the 5' end and the poly(A) tail at the 3' end (reviewed in ref. 2). Initiation factor eIF4G binds the cap and the poly(A) tail indirectly through its interaction with the cap binding protein eIF4E and the poly(A) binding protein Pab1p, respectively. The simultaneous interaction of eIF4G with eIF4E and Pab1p not only places eIF4G on the mRNA, it also leads to mRNA circularization<sup>3</sup>.

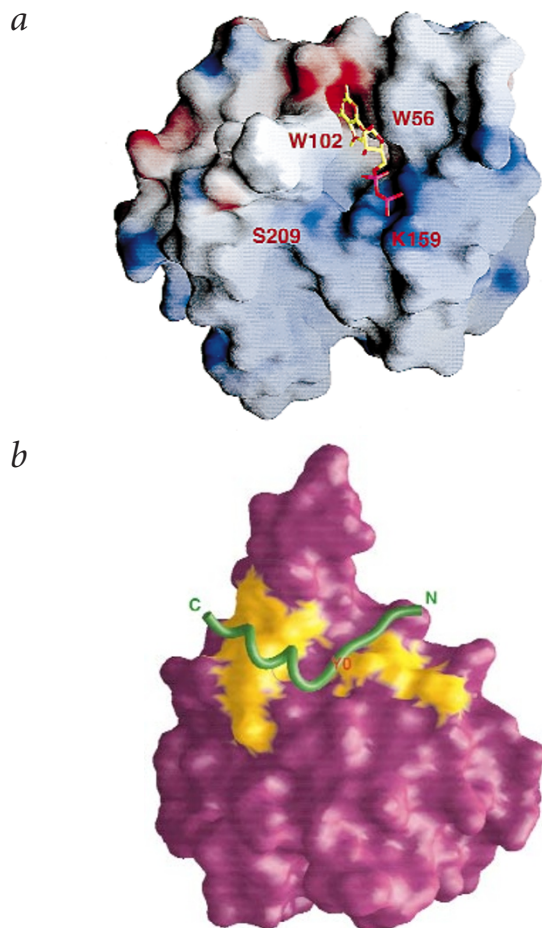
Translation initiation and cell viability depends upon eIF4G binding to eIF4E at the 5' end of mRNA<sup>4,5</sup>. Initiation factor eIF4E binds the methylated guanosine cap of mRNA with submicromolar affinity, and binds nonmethylated precursors at least fivefold less efficiently<sup>6</sup>. Thus, the eIF4G–eIF4E complex binds to the 5' end of mRNA because eIF4E specifically recognizes the cap. In addition, the association of eIF4G with eIF4E increases eIF4E affinity for the cap by at least 10-fold<sup>7,8</sup>. This allosteric regulation provides the initiation competent eIF4E–eIF4G complex a selective advantage in binding to mRNA over the apo eIF4E. The eIF4E–cap interaction is regulated in mammals by the phosphorylation of eIF4E *via* the eIF4G-associated Mnk1 kinase<sup>9</sup>.



**Fig. 1** Organization of the eukaryotic translation initiation apparatus. **a**, The eIF4G protein serves as a scaffold that connects the mRNA and its associated proteins to the ribosome. Binding of eIF4E to eIF4G is inhibited by the eIF4E binding proteins, the 4E-BPs, while 4E-BP binding to eIF4E is disrupted by phosphorylation. **b**, Domain structure of the yeast eIF4G1 protein. The relative position of binding sites for various translation factors on yeast eIF4G1 is shown. The mammalian eIF4G proteins have extended C-termini that bind to other factors. Shown below are the amino acid sequences surrounding the eIF4E binding site in the human (h) and yeast (y) eIF4G proteins and the eIF4E-binding proteins 4E-BP1 and Caf20p. The positions of the amino acids relative to the conserved tyrosine at position 0, as well as the minimal consensus sequence for this region, including the position of the conserved hydrophobic residue (Φ), are shown.

<sup>1</sup>Department of Molecular and Cell Biology, 401 Barker Hall, University of California, Berkeley, California 94720, USA. <sup>2</sup>MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

Correspondence should be addressed to A.S. email: asachs@uclink4.berkeley.edu or to G.V. email: gv1@mrc-lmb.cam.ac.uk



**Fig. 2** The cap binding protein eIF4E provides two binding surfaces for the methylated cap and peptides derived from eIF4G and the 4E-BPs. **a**, The  $\beta$ -sheet surface of eIF4E provides a binding site for the methylated guanosine cap, which intercalates in a tight binding cleft between two universally conserved Trp residues. The Watson–Crick face of the guanosine interacts with an electronegative (red) patch, whereas the diphosphate moiety points towards a basic patch on the surface of the protein (blue) that defines the likely path for the RNA. Ser 209 and Lys 159 lie on opposite sides of this path. **b**, The  $\alpha$ -helical surface of eIF4E contains conserved hydrophobic and acidic residues (highlighted in yellow) that form two distinctive patches. eIF4G and 4E-BPs derived peptides (the 4E-BP1 peptide is shown) bind to this conserved dorsal surface of eIF4E. Four amino acids (Tyr 0, Phe +4, Leu +5 and  $\Phi$ +6, where  $\Phi$  is any hydrophobic residue, see Fig. 1b) that are nearly universally conserved among all eIF4G and 4E-BP's proteins are highlighted.

ed within the first 200 amino acids of the protein<sup>28–30</sup>. RRM1 and RRM2 are also responsible for its association with eIF4G<sup>19,31</sup>. The structural basis for the ability of the two N-terminal RRM domains of Pab1p to recognize short oligo(A) RNA has been elucidated<sup>32</sup> and is discussed below. This structural work has also defined a putative eIF4G interaction surface on Pab1p.

The machinery responsible for translational initiation utilizes the eIF4E–cap and Pab1p–poly(A) interactions to associate eIF4G with mRNA. These interactions also provide a target for the regulation of protein synthesis. For instance, the amount of available eIF4E in the cell can be determined by titration with a class of proteins that act as competitive inhibitors of the eIF4E–eIF4G interaction, the eIF4E–binding proteins (4E-BPs) (Fig. 1a)<sup>33</sup>. The 4E-BPs are themselves subject to regulation by cellular kinases in response to signal transduction pathways<sup>34</sup>. Under growth promoting conditions, these proteins become hyperphosphorylated and their affinity for eIF4E decreases. The structure of eIF4E bound to a fragment of 4E-BP1 has provided insight into how the 4E-BPs interact with eIF4E and how phosphorylation of the 4E-BPs may control their affinity for eIF4E<sup>13</sup>.

### Structure of the eIF4E–m<sup>7</sup>GDP complex

The three-dimensional structure of mouse eIF4E bound to the cap analog m<sup>7</sup>GDP (ref. 14) and the subsequent structure of the yeast complex<sup>12</sup> provided the first detailed view of a eukaryotic translation initiation factor. (Fig. 2a). Together with several structures of viral capping enzymes<sup>35–38</sup>, these structures revealed how the cap is specifically recognized by eIF4E. They also led to key predictions about how phosphorylation of eIF4E could enhance its affinity for mRNA, and how eIF4G could bind to eIF4E.

The cap binding protein has an  $\alpha\beta$  structure and its shape resembles a baseball glove<sup>12,14</sup>. The  $\alpha\beta$  fold is very common among RNA binding proteins<sup>39</sup>, but the sequence of eIF4E shares no homology to other proteins, including the components of the nuclear cap binding complex that recognizes the cap in the nucleus before mRNA export to the cytoplasm<sup>40</sup>. The eIF4E protein contains a curved eight-stranded antiparallel  $\beta$ -sheet that includes the cap binding site, while its convex surface contains three long  $\alpha$ -helices. Two short helices within the loops bridging the  $\beta$ 1– $\beta$ 2 and the  $\beta$ 3– $\beta$ 4 strands of the  $\beta$ -sheet contain conserved Trp residues involved in cap binding. The loops immediately preceding the Trp containing helices retain some mobility in the complex with the cap analog<sup>12</sup>, suggesting that, in the absence of the cap, these regions of the protein may be disordered.

The methylated guanosine binds a narrow hydrophobic cleft by intercalating between Trp 56 (in helix  $\alpha$ 1) and Trp 106 (in helix  $\alpha$ 3; Fig. 2a). The favorable binding energy provided by

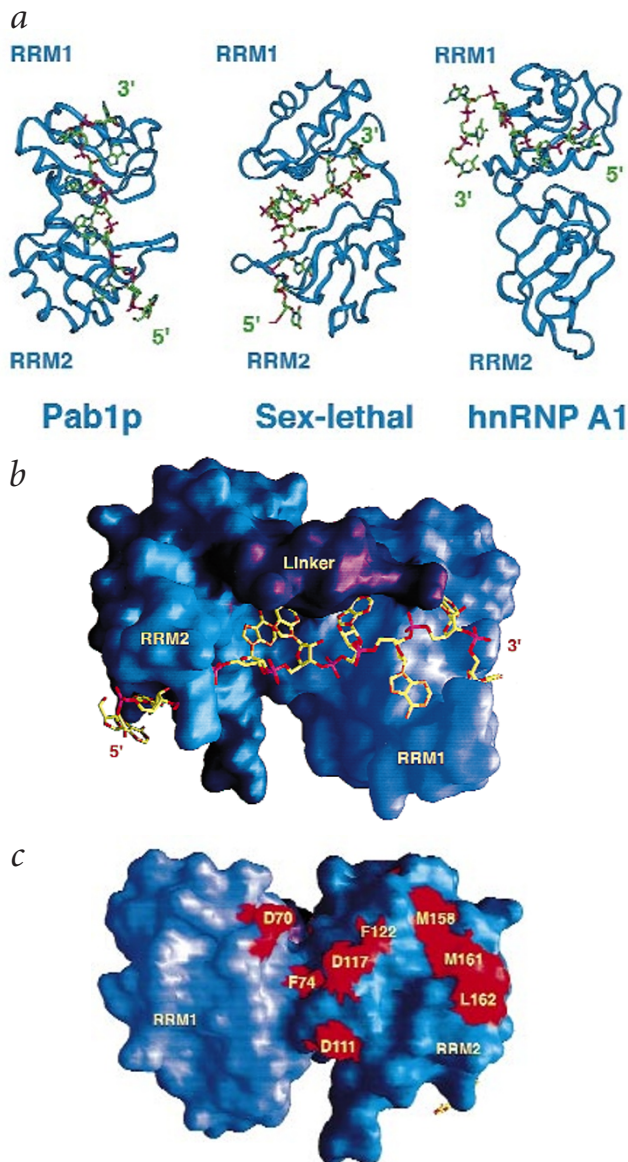
Phosphorylation further enhances eIF4E's affinity for both the cap structure and eIF4G<sup>10,11</sup>. The recent structural work on eIF4E has helped clarify many features of the cap recognition process<sup>12–14</sup>.

Each of the subdomains on eIF4G that bind to eIF4E, Pab1p and other proteins has recently been mapped (Fig. 1b)<sup>5,9,15–20</sup>. The best characterized region of eIF4G is its eIF4E-binding domain. Sequences homologous to this domain have been identified in other eIF4E-binding proteins (Fig. 1b)<sup>15,21</sup>. Peptides as short as 15 amino acids from this domain can also bind to eIF4E with nanomolar affinity<sup>13</sup>. How eIF4E interacts with its binding partners and how related eIF4E binding peptides recognize a common surface on eIF4E are questions that have been addressed by the recent structural work<sup>13</sup>.

The recruitment of eIF4G to mRNA is not only a result of its binding to the cap through eIF4E; eIF4G also associates with the poly(A) tail through Pab1p<sup>16</sup>. The simultaneous association of eIF4E and Pab1p with eIF4G allows for the synergistic activation of translation initiation *in vitro*<sup>22</sup>. *In vivo*, mRNAs that are both capped and polyadenylated are also translated more efficiently than those that are either capped or polyadenylated<sup>23</sup>. The molecular basis for this synergy is unknown.

Pab1p binds with subnanomolar affinity to poly(A)<sup>24</sup>. It is the founding member of the largest class of RNA binding proteins<sup>25,26</sup> and contains the so-called ribonucleoprotein (RNP) or RNA-recognition motif (RRM), a universal RNA recognition module<sup>27</sup>. All Pab1p proteins contain four RRM domains separated by highly conserved linkers<sup>28</sup>. The poly(A) binding activity of Pab1p resides within its two N-terminal RRM domains, locat-

## review



**Fig. 3** The poly(A) binding protein Pab1p provides two binding surfaces for poly(A) and eIF4G. **a**, Many RNA-binding proteins that bind single stranded RNA utilize two contiguous RRM domains. The structure of Pab1p, Sex-lethal and hnRNP A1 show that the arrangement of these domains differs in each of these three proteins, and that the RNA interacts differently with the protein  $\beta$ -sheet surface in each case. Sex-lethal binds U-rich sequences by creating a narrow cleft between RRM1 and RRM2, whereas oligo(A) interacts with an extended  $\beta$ -sheet surface created by the contiguous arrangement of RRM1 and RRM2. In both Pab1p and Sex-lethal, the RNA runs a similar path along the two RRMs: the 5' half of the RNA interacts with RRM2, whereas the 3' half of the sequence interacts with RRM1. **b**, The RNA binding surface of Pab1p, with eight well-ordered adenosine residues interacting with the extended  $\beta$ -sheet surface defined by RRM1 (dark blue) and RRM2 (light blue) and the conserved linker joining the two domains (purple). **c**, The opposite dorsal surface of Pab1p contains conserved hydrophobic and acidic residues (in red). These residues form two distinctive patches on the dorsal surface of the Pab1p structure and originate from both RRM1 (D70 and F74) and RRM2. This surface is likely to provide the binding site for eIF4G, but the structural details of this interaction remain to be elucidated.

or VP39 leads to a complete loss of function. Changes to other aromatic residues result in partial function<sup>14,38</sup>. In addition, the stacking interactions provide a means for discriminating between methylated and unmethylated bases. Methylation leads to electron deficient  $\pi$ -orbitals, which interact favorably with the electron rich  $\pi$ -orbitals of the Trp indol<sup>41</sup>. Stacking interactions or sandwiching of methylated bases with aromatic side chains has been observed in solution and in the crystal structures of numerous small molecules<sup>36</sup>. Furthermore, the structure of the unmethylated cap analog pGGGp and a viral capping enzyme shows stacking of the unmethylated base with a single aromatic side chain, instead of the characteristic sandwich observed in proteins that recognize the methylated base<sup>35</sup>.

Binding of mRNA to eIF4E is not affected by the identity of the base following the cap. However, binding of a longer cap analog (m<sup>7</sup>GpppA versus m<sup>7</sup>Gpp) results in increased rigidity in helix  $\alpha 6$  and numerous chemical shift changes in the adjacent loop. These regions are likely to be within the mRNA binding site, which is contiguous to the cap binding slot (Fig. 2a)<sup>12</sup>. Treatment of cells with growth factors, hormones and mitogens stimulates translation and phosphorylation of Ser 209 in mammalian eIF4E (reviewed in ref. 33). Ser 209 is located along the edge of the putative mRNA binding cleft (Fig. 2a). Phosphorylated Ser 209 may interact with Lys 159 (located on the opposite edge of the cleft) to produce a bridge to cover the mRNA binding cleft near its entrance<sup>14</sup>. This clamp could stabilize the cap-eIF4E complex and the association of the recruited mRNA with the translation initiation apparatus.

Further insight into the recognition of the remainder of the mRNA by eIF4E was provided by a structure involving VP39 in complex with a capped single stranded RNA hexamer<sup>37</sup>. In this complex, the six nucleotides form two groups of three stacked bases with a roughly helical geometry that are separated by a sharp turn in the phosphate between the third and fourth nucleotide. Only the first three nucleotides interact directly with VP39, primarily through hydrogen bonds and salt bridges with the sugar-phosphate backbone. Thus, the protein appears to recognize in a sequence-independent manner the conformation of a helical trimer of stacked bases.

#### Complexes of eIF4E with eIF4G and 4E-BP1 fragments

The convex dorsal surface of eIF4E lies opposite the cap binding site. The binding site for eIF4G was originally predicted to be on this surface of eIF4E based on two separate observations. First, the addition of mammalian 4E-BP2 to yeast eIF4E resulted in

these stacking interactions is reinforced by van der Waals interactions between these Trp residues and the guanosine sugar. The N7 methyl group is directed towards the interior of the protein, where it contacts a third conserved Trp residue (Trp 166). Glu 103 recognizes the guanosine amino and imino groups through its side chain carboxylate. This residue is conserved in all eIF4E's and its mutation to Ala in human eIF4E abrogates its function<sup>6</sup>. The diphosphate moiety of the cap analog forms electrostatic interactions with conserved Arg and Lys residues.

The structural principles underlying the recognition of methylated bases and discrimination against unmethylated analogs were first illustrated by the complex between the vaccinia virus VP39 capping enzyme and the cap analog m<sup>7</sup>GpppG<sup>36</sup>, and confirmed by the eIF4E structures<sup>12,14</sup>. In both the VP39 and eIF4E structures, the Watson-Crick face of the guanosine is recognized specifically through multiple hydrogen bonding interactions, while the methylated base is sandwiched between aromatic side chains. These stacking interactions play a dominant role in cap binding<sup>38</sup> and therefore in eIF4E function. Changing any of the stacking aromatic residues to Ala in eIF4E



marked changes in ~20 amide crosspeak resonances clustered on the dorsal surface of eIF4E<sup>12</sup>. Second, this same region was shown to contain hydrophobic residues conserved in all known eIF4E proteins<sup>8,14</sup> (Fig. 2*b*). These observations identified this region as the likely binding site for the hydrophobic residues common to eIF4G and the 4E-BPs (Fig. 1*b*). Extensive genetic and biochemical studies lend strong support to this hypothesis<sup>8,42</sup>.

The eIF4E binding region of eIF4G spans ~100 amino acids<sup>43</sup>, but a stretch of only 10 amino acids shared by eIF4G and 4E-BPs provide most of the interaction energy (Fig. 1*b*)<sup>13</sup>. In fact, 15-mer peptides derived from the eIF4E binding region of eIF4G and the 4E-BPs competitively inhibit the eIF4E–eIF4G interaction and translation initiation<sup>13</sup>. The X-ray crystal structures of the ternary complexes between eIF4E, m<sup>7</sup>GDP and eIF4E binding peptides revealed the details of these interactions<sup>13</sup>. Ternary complexes were obtained with peptides derived from both eIF4G2 (one of two isoforms of eIF4G) and 4E-BP1. Both peptides adopt the same L-shaped structure, with an extended stretch preceding a short  $\alpha$ -helix. They each bind to the same dorsal surface of eIF4E and make very similar interactions with invariant or highly conserved eIF4E side chains (Fig. 2*b*). An invariant Tyr in the peptide immediately precedes the beginning of the  $\alpha$ -helix and makes multiple van der Waals contacts with nearly invariant eIF4E residues. An invariant Leu in the peptide and the hydrophobic residue that follows it (Fig. 1*b*) interact with exposed hydrophobic side chains, as well as with each other and the invariant Tyr.

The structures of eIF4E bound to m<sup>7</sup>GDP and the eIF4G2 or 4E-BP1 peptides clarify biochemical data derived from studies on these complexes. They suggest that the bulk of the energy responsible for the high binding affinity between eIF4E and either eIF4G or the 4E-BPs results from nearly identical interactions with eIF4E. Since the two proteins share no sequence homology outside the eIF4E binding site, 4E-BPs function by mimicking eIF4G in an example of convergent evolution. However, additional interactions between either eIF4G or the 4E-BPs and eIF4E must exist to confer differential affinity and specificity. Genetic and biochemical experiments have shown that the binding sites for eIF4G and 4E-BP on eIF4E are not identical; although they share a core set of amino acids, a subset of mutations that decrease eIF4G binding do not affect 4E-BP protein binding<sup>8,42</sup>. Also, regions outside the eIF4E-binding peptide may interact with eIF4E since the entire C-terminal half of 4E-BP1 and a larger fragment of eIF4G were protected from proteolytic degradation when bound to eIF4E<sup>13,43</sup>.

Another important question that can be addressed in the context of these structures concerns how phosphorylation of the 4E-BPs lead to their decreased affinity for eIF4E. A site of phosphorylation on 4E-BP1 (Ser 65) lies just C-terminal to the ordered region observed in the crystal structure. This residue is likely to be located near an acidic patch in eIF4E. Therefore, electrostatic repulsion between the phosphorylated Ser65 in 4E-BP1 and the conserved Glu and Asp residues in eIF4E has been postulated to disrupt the 4E-BP1–4E complex<sup>13</sup>.

Questions that remains unanswered are how binding of eIF4G to eIF4E increases binding of eIF4E to the cap, and how mutations in eIF4E surface residues that reduce the eIF4E–eIF4G interactions also reduce cap–eIF4E binding<sup>7,8,42</sup>. The structures of the binary eIF4E–m<sup>7</sup>GDP and the ternary eIF4E–m<sup>7</sup>GDP–eIF4G complexes are nearly identical<sup>13,14</sup>. This indicates that the cap–eIF4E interaction is not perturbed by the presence of eIF4G or the 4E-BP peptides. Furthermore, the peptides bind ~35 Å away from the cap binding site, and do not

interact directly with the cap or with residues that interact with the cap. One possible explanation for how eIF4G and the 4E-BPs alter eIF4E affinity for the cap is that regions of eIF4G and 4E-BP1 not included in the crystal structure contribute to cap binding by altering the structure of eIF4E. An alternative (or additional) explanation originates from the observation that binding of 4E-BP and eIF4G to eIF4E occurs by induced fit<sup>43–45</sup>, and that the loops within eIF4E that bind the cap retain conformational flexibility in the complex<sup>12</sup>. Perhaps apo eIF4E has a poorly defined m<sup>7</sup>GDP binding site and binding of eIF4G and/or m<sup>7</sup>GDP stabilizes the eIF4E structure, leading to better definition of the cap binding site and increased m<sup>7</sup>GDP and/or eIF4G binding. Phosphorylation of eIF4E could also affect m<sup>7</sup>GDP and eIF4G binding by stabilizing the eIF4E structure in a conformation competent to bind m<sup>7</sup>GDP and eIF4G. The validity of this model remains to be tested by systematic comparisons of the structure and dynamics of eIF4E when free or bound to the cap and other proteins.

### Interaction between Pab1p and oligo(A)

The structure of two RRM of Pab1 bound to oligo(A) is the most recent addition to the growing list of structures of components of the eukaryotic translational initiation apparatus<sup>32</sup>. Other structures of RNA–protein complexes include that of human U1A protein<sup>46,47</sup>, the best studied example. U1A demonstrated how recognition of structured RNA by a single RRM occurs, while two other structures revealed how single stranded RNA can be recognized by proteins containing two RRMs arranged in tandem. These are the cocrystal structures of *Drosophila* Sex-lethal bound to a U-rich sequence derived from a splicing regulatory element<sup>48</sup>, and the structure of heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) bound to a purine-rich DNA telomeric sequence<sup>49</sup>.

Like hnRNP A1 and Sex-lethal, Pab1p requires two RRMs in order to bind poly(A). The first two RRMs of Pab1p are responsible for specific, high affinity binding to oligo(A) and for binding to eIF4G<sup>19,28–31</sup>. In the crystal structure of the Pab1p–oligo(A) complex<sup>32</sup>, the single stranded oligo(A) RNA adopts an extended conformation running through a binding surface lined by the antiparallel  $\beta$ -sheet of each domain and backed on the opposite surface by four  $\alpha$ -helices (Fig. 3*a,b*). The highly conserved linker between RRM1 and RRM2 of Pab1p also contacts the RNA.

With the Pab1p–oligo(A) structure, there are now three distinct examples of nucleic acid recognition by multidomain RRM proteins<sup>48,49</sup>. In each case, two independently folded RNA binding domains act synergistically to increase affinity and specificity for single stranded RNA. The binding surface is created by simultaneous interactions of the nucleic acid with each of the two domains, as well as with the inter RRM linker sequence whose helical structure is stabilized by the nucleic acid. An RNA stabilized  $\alpha$ -helix in the C-terminus of human U1A protein also plays an important role in determining RNA specificity<sup>46</sup>. The interdomain arrangement in the Pab1p–oligo(A) complex, however, is very different from those observed in both the Sex-lethal complex (a V-shaped cleft)<sup>48</sup> and in the hnRNPA1 complex<sup>49</sup> (Fig. 3*a*). The binding cleft between RRM1 and RRM2 is much narrower in the Sex-lethal–U-rich RNA complex than in the Pab1p–oligo(A) complex (Fig. 3*a*)<sup>32,48</sup>, allowing for some discrimination based on the different sizes of purines and pyrimidines.

In the Pab1p complex, the arrangement of RRM1 and RRM2 is stabilized by inter-RRM contacts. However, the interface area defined by these interactions is small (only 550 Å<sup>2</sup>, while the surface area buried by protein–RNA contacts is 2,600 Å<sup>2</sup>). Therefore,

## review

RNA–protein interactions are likely to be important in defining the relative orientation of RRM1 and RRM2. It is also likely that the RNA induced orientation of neighboring RRMs is important in defining the binding specificity of these proteins.

The full length Pab1 protein was shown by titration and nuclease protection experiments to cover 25 nucleotides of poly(A), but its primary binding site is 11–12 nucleotides long<sup>24</sup>. Eight adenines are visible in the electron density map of the Pab1p–oligo(A) complex (Fig. 3*b*). Adenosine recognition by RRM1 and RRM2 of Pab1p is mediated in part by contacts with conserved residues in the two central strands of the  $\beta$ -sheet of each RRM (which contain the RNP-1 and RNP-2 sequences). Some of these residues take part in intermolecular stacking interactions analogous to those observed in all structures of complexes involving other RNP proteins. These interactions are thus likely to be universal features of RRM–RNA recognition<sup>27</sup>. Mutation of the residues involved in these interactions, for example of Phe142 in yeast Pab1p, disrupts high affinity binding of yeast Pab1p to poly(A)<sup>29</sup>. Contacts with conserved residues provide the basal RNA binding activity of the RRM<sup>27</sup>. This would explain why the path of the RNA across the  $\beta$ -sheet surface is similar in the reported structures of RRM–RNA complexes.

Conformational preferences resulting from the strong tendency of adenine residues to form planar stacking interactions are likely to play an important role in the binding specificity of Pab1p for Poly(A). Extensive intramolecular stacking interactions between adenine residues are observed in the Pab1p complex (Fig. 3*b*), whereas only one intramolecular stacking interaction is observed in the Sex-lethal complex. Other interactions observed in the Pab1p complex involve residues that are not conserved between different RRM domains. These presumably provide interactions specific to Pab1p. They include contacts with the phosphodiester backbone and the ribose sugars (seven out of eight sugars in the binding cleft are in contact with protein residues). In addition, direct sequence specific contacts with the purine bases (Fig. 3*b*) are provided by amino acids in loop 3 of each RRM (connecting  $\beta$ 2 and  $\beta$ 3), in the linker between RRM1 and RRM2, and in the C-terminal tail of RRM2.

The pattern of sequence conservation in Pab1p has enabled speculation on the structure of RRM3 and RRM4 of Pab1p and its interaction with poly(A)<sup>32</sup>. The majority of residues participating in RNA recognition are conserved between RRM1 and RRM3 and between RRM2 and RRM4, but not between RRM1 and RRM4 or RRM2 and RRM3. This conservation suggests that RRM3 and RRM4 may form an RNA binding surface similar to that formed by RRM1 and RRM2. The reduced affinity and specificity for poly(A) of RRM3 and RRM4<sup>29</sup> is likely to be a consequence of sequence divergence in RNA binding residues, including the linker region connecting the domains. Future work will address if and how the poly(A) tail interacts simultaneously with all four RRMs of Pab1p.

### Predictions on the Pab1p–eIF4G complex

The  $\alpha$ -helical dorsal surface of Pab1p is a phylogenetically conserved hydrophobic-acidic region (Fig. 3*c*). As with predictions about the conserved dorsal surface of eIF4E and its role in eIF4G/4E-BP binding, it has been predicted that this conserved region would bind to eIF4G<sup>32</sup>. In support of this suggestion, mutagenesis of several residues within this region has already demonstrated its importance for eIF4G binding<sup>50</sup>. One interesting feature about this surface is that it spans two RRMs that are very likely to be juxtaposed as a result of oligo(A) binding (Fig. 3*c*). If the eIF4G binding site really does span the two RRMs, then perhaps the reason that yeast but not mammalian Pab1p requires poly(A) for eIF4G binding<sup>16,19</sup> is that yeast eIF4G by itself cannot stabilize RRM1 and RRM2 packing while mammalian eIF4G can.

The structure of the ternary complex containing Pab1p RRM1 and RRM2, oligo(A) and a suitable fragment of eIF4G is anxiously awaited. Key questions that would be addressed by this structure are how eIF4G binds specifically to Pab1p, and if Pab1p stabilizes a unique eIF4G conformation. Similarly, it will be essential to determine if binding of eIF4G to Pab1p changes the Pab1p–oligo(A) interface. More general questions will include the identification of Pab1p repressors akin to the 4E-BPs based on sequence predictions using the observed interface between Pab1p and eIF4G as a guide. Finally, seeing both Pab1p and eIF4E bound to a large fragment of eIF4G would present unprecedented opportunities for understanding how different components of a translation pre-initiation complex interact with each other.

### Conclusions and perspectives

Translational control is a major regulatory step in normal and abnormal cell growth. The structures of translation initiation factors bound to mRNA and protein fragments have provided critical insight into how mRNAs are delivered to the ribosome. We have now begun to understand how the cap and the poly(A) tail are recognized, and how surfaces of the cap and poly(A) binding proteins function in protein recognition. We still do not know the structure of most translation initiation factors, and thus it is still very difficult to visualize the mechanisms underlying translational regulation. Clearly, we have just begun to scratch the surface of the bewildering beauty and complexity of the apparatus controlling the initiation of protein synthesis.

### Acknowledgments

We thank L. Varani for help with the figures, NATO and HFSP for support.

Received 7 February, 2000; accepted 14 March, 2000.

1. Merrick, W.C. & Hershey, J.W.B. The pathway and mechanism of eukaryotic protein synthesis. In *Translational Control* (eds. Hershey, J.W.B., Matthews, M. & Sonenberg, N.) 31–69 (Cold Spring Harbor Press, Cold Spring Harbor; 1996).
2. Sachs, A.B., Sarnow, P. & Hentze, M.W. Starting at the beginning, middle, and end: translation initiation in eucaryotes. *Cell* **89**, 831–838 (1997).
3. Wells, S.E., Hillner, P., Vale, R. & Sachs, A.B. Circularization of mRNA by eukaryotic translation initiation factors. *Mol. Cell* **2**, 135–140 (1998).
4. Altmann, M. & Trachsel, H. Altered mRNA cap recognition activity of initiation factor 4E in the yeast cell cycle division mutant *cdc33*. *Nuc. Acids Res.* **17**, 5923–5931 (1989).
5. Tarun, S.Z. & Sachs, A.B. Binding of eIF4E to eIF4G represses translation of uncapped mRNA. *Mol. Cell. Biol.* **17**, 6876–6886 (1997).
6. Morino, S., et al. Analysis of the mRNA cap-binding ability of human eukaryotic initiation factor-4E by use of recombinant wild-type and mutant forms. *Eur. J. Biochem.* **239**, 597–601 (1996).
7. Haghhighat, A. & Sonenberg, N. eIF4G dramatically enhances the binding of eIF4E to the mRNA 5'-cap structure. *J. Biol. Chem.* **272**, 21677–21680 (1997).
8. Ptushkina, M., et al. Cooperative modulation by eIF4G of eIF4E-binding to the mRNA 5' cap in yeast involves a site partially shared by p20. *EMBO J.* **17**, 4798–808 (1998).
9. Pyronnet, S., et al. Human eukaryotic translation initiation factor 4G (eIF4G) recruits Mnk1 to phosphorylate eIF4E. *EMBO J.* **18**, 270–279 (1999).
10. Minich, W.B., Balasta, M.L., Goss, D.J. & Rhoads, R.E. Chromatographic resolution of *in vivo* phosphorylated and nonphosphorylated eukaryotic translation initiation factor eIF-4E: increased cap affinity of the phosphorylated form. *Proc. Natl. Acad. Sci. USA* **91**, 7668–7672 (1994).
11. Feigenblum, D. & Schneider, R.J. Cap-binding protein (eukaryotic initiation factor 4E) and 4E-inactivating protein BP-1 independently regulate cap-dependent translation. *Mol. Cell. Biol.* **16**, 5450–5457 (1996).
12. Matsuo, H., et al. Structure of translation factor eIF4E bound to m7GDP and interaction with 4E-binding protein. *Nature Struct. Biol.* **4**, 717–724 (1997).
13. Marcotrigiano, J., Gingras, A.C., Sonenberg, N. & Burley, S.K. Cap-dependent translation initiation in eucaryotes is regulated by a molecular mimic of eIF4G. *Mol. Cell* **3**, 707–716 (1999).
14. Marcotrigiano, J., Gingras, A.C., Sonenberg, N. & Burley, S.K. Cocrystal structure of the messenger RNA 5' cap-binding protein (eIF4E) bound to 7-methyl-GDP. *Cell* **89**, 951–961 (1997).
15. Mader, S., Lee, H., Pause, A. & Sonenberg, N. The translation initiation factor eIF-4E binds to a common motif shared by the translation factor eIF-4 $\gamma$  and the translational repressors 4E-binding proteins. *Mol. Cell. Biol.* **15**, 4990–4997 (1995).
16. Tarun, S.Z. & Sachs, A.B. Association of the yeast poly(A) tail binding protein with translation initiation factor eIF-4G. *EMBO J.* **15**, 7168–7177 (1996).
17. Morino, S., Imataka, H., Svitkin, Y.V., Pestova, T.V. & Sonenberg, N. Eukaryotic translation initiation factor 4E (eIF4E) binding site and the middle one-third of eIF4G1 constitute the core domain for cap-dependent translation, and the C-terminal one-third functions as a modulatory region. *Mol. Cell. Biol.* **20**, 468–77 (2000).
18. Dominguez, D., Altmann, M., Benz, J., Baumann, U. & Trachsel, H. Interaction of translation initiation factor eIF4G with eIF4A in the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* **274**, 26720–6 (1999).
19. Imataka, H., Gradi, A. & Sonenberg, N. A newly identified N-terminal amino acid sequence of human eIF4G binds poly(A)-binding protein and functions in poly(A)-dependent translation. *EMBO J.* **17**, 7480–7489 (1998).
20. Neff, C. & Sachs, A.B. Eukaryotic translation initiation factors 4G and 4A from *Saccharomyces cerevisiae* interact physically and functionally. *Mol. Cell. Biol.* **19**, 5557–5564 (1999).
21. Stebbins-Boaz, B., Cao, Q., de Moor, C.H., Mendez, R. & Richter, J.D. Maskin is a CPEB-associated factor that transiently interacts with eIF-4E. *Mol. Cell* **4**, 1017–27 (1999).
22. Tarun, S.Z., Wells, S.E., Deardorff, J.A. & Sachs, A.B. Translation initiation factor eIF-4G mediates *in vitro* poly(A) tail-dependent translation. *Proc. Natl. Acad. Sci. USA* **94**, 9046–9051 (1997).
23. Gallie, D.R. The cap and the poly(A) tail function synergistically to regulate mRNA translational efficiency. *Genes Dev.* **5**, 2108–2116 (1991).
24. Sachs, A.B., Davis, R.W. & Kornberg, R.D. A single domain of yeast poly(A)-binding protein is necessary and sufficient for RNA binding and cell viability. *Mol. Cell. Biol.* **7**, 3268–3276 (1987).
25. Sachs, A.B., Bond, M.W. & Kornberg, R.D. A single gene from yeast for both nuclear and cytoplasmic polyadenylate-binding proteins: domain structure and expression. *Cell* **45**, 827–35 (1986).
26. Adam, S.A., Nakagawa, T., Swanson, M.S., Woodruff, T.K. & Dreyfuss, G. mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. *Mol. Cell. Biol.* **6**, 2932–2943 (1986).
27. Varani, G. & Nagai, K. RNA recognition by RNP proteins during RNA processing. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 407–445 (1998).
28. Burd, C.G., Matunis, E.L. & Dreyfuss, G. The multiple RNA-binding domains of the mRNA poly(A)-binding protein have different RNA-binding activities. *Mol. Cell. Biol.* **11**, 3419–3424 (1991).
29. Deardorff, J.A. & Sachs, A.B. Differential effects of aromatic and charged residue substitutions in the RNA binding domains of the yeast poly(A)-binding protein. *J. Mol. Biol.* **269**, 67–81 (1997).
30. Kuhn, U. & Pieler, T. Xenopus poly(A) binding protein: functional domains in RNA binding and protein-protein interactions. *J. Mol. Biol.* **256**, 20–30 (1996).
31. Kessler, S.H. & Sachs, A.B. RNA Recognition motif 2 of Yeast Pab1p is required for its functional interaction with eukaryotic translation initiation factor 4G. *Mol. Cell. Biol.* **18**, 51–57 (1998).
32. Deo, R.C., Bonanno, J.B., Sonenberg, N. & Burley, S.K. Recognition of polyadenylate RNA by the poly(A)-binding protein. *Cell* **98**, 835–45 (1999).
33. Gingras, A.C., Raught, B. & Sonenberg, N. eIF4 initiation factors: effectors of mRNA recruitment to ribosomes and regulators of translation. *Ann. Rev. Biochem.* **68**, 913–963 (1999).
34. Gingras, A.C., et al. Regulation of 4E-BP1 phosphorylation: a novel two-step mechanism. *Genes Dev.* **13**, 1422–37 (1999).
35. Hakansson, K. & Wigley, D.B. Structure of a complex between a cap analogue and mRNA guanylyl transferase demonstrates the structural chemistry of RNA capping. *Proc. Natl. Acad. Sci. USA* **95**, 1505–1510 (1998).
36. Hodel, A.E., Gershon, P.D., Shi, X., Wang, S.M. & Quijcho, F.A. Specific protein recognition of an mRNA cap through its alkylated base. *Nature Struct. Biol.* **4**, 350–4 (1997).
37. Hodel, A.E., Gershon, P.D. & Quijcho, F.A. Structural basis for sequence-nonspecific recognition of 5'-capped mRNA by a cap-modifying enzyme. *Mol Cell* **1**, 443–7 (1998).
38. Hu, G., Gershon, P.D., Hodel, A.E. & Quijcho, F.A. mRNA cap recognition: dominant role of enhanced stacking interactions between methylated bases and protein aromatic side chains. *Proc. Natl. Acad. Sci. USA* **96**, 7149–54 (1999).
39. Nagai, K. RNA-protein complexes. *Curr Opin Struct Biol* **6**, 53–61 (1996).
40. Izaurralde, E., et al. A cap-binding protein complex mediating U snRNA export. *Nature* **376**, 709–712 (1995).
41. Ishida, T., Doi, M., Ueda, H., Inoue, M. & Sheldrick, G.M. Specific ring stacking interaction on the tryptophan-7-methylguanine system: comparative crystallographic studies of indole derivatives-7-methylguanine base, nucleoside, nucleotide complexes. *J. Amer. Chem. Soc.* **110**, 2286–2294 (1988).
42. Ptushkina, M., von der Haar, T., Karim, M.M., Hughes, J.M. & McCarthy, J.E. Repressor binding to a dorsal regulatory site traps human eIF4E in a high cap-affinity state. *EMBO J.* **18**, 4068–75 (1999).
43. Hershey, P.E., et al. The Cap-binding protein eIF4E promotes folding of a functional domain of yeast translation initiation factor eIF4G1. *J. Biol. Chem.* **274**, 21297–304 (1999).
44. Fletcher, C.M., et al. 4E binding proteins inhibit the translation factor eIF4E without folded structure. *Biochemistry* **37**, 9–15 (1998).
45. Fletcher, C.M. & Wagner, G. The interaction of eIF4E with 4E-BP1 is an induced fit to a completely disordered protein. *Protein Sci.* **7**, 1639–42 (1998).
46. Allain, F.H., et al. Specificity of ribonucleoprotein interaction determined by RNA folding during complex formulation. *Nature* **380**, 646–50 (1996).
47. Oubridge, C., Ito, N., Evans, P.R., Teo, C.H. & Nagai, K. Crystal structure at 1.92 Å resolution of the RNA-binding domain of the U1A spliceosomal protein complexed with an RNA hairpin. *Nature* **372**, 432–438 (1994).
48. Handa, N., et al. Structural basis for recognition of the tra mRNA precursor by the Sex-lethal protein. *Nature* **398**, 579–585 (1999).
49. Ding, J., et al. Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* **13**, 1102–1115 (1999).
50. Otero, L., Ashe, M. & Sachs, A. The yeast poly(A)-binding protein Pab1p stimulates *in vitro* poly(A)-dependent and cap-dependent translation by distinct mechanisms. *EMBO J.* **18**, 3153–3163 (1999).