Structural insights into the Kozak sequence interactions within the mammalian late-stage initiation complexes

Angelita Simonetti[1,4], Ewelina Guca[2,4], Anthony Bochler[1,2], Lauriane Kuhn[3] and Yaser Hashem[2,*]

[1] Université de Strasbourg, CNRS, Architecture et Réactivité de l'ARN, UPR9002, F-67000 Strasbourg, France

[2] INSERM U1212 ARNA, Institut Européen de Chimie et Biologie, Université de Bordeaux, Pessac 33607, France

[3] Proteomic Platform Strasbourg - Esplanade, Institut de Biologie Moléculaire et Cellulaire, Université de Strasbourg, Strasbourg 67000, France.

[4] These authors have contributed equally to this work

* Correspondence should be sent to Yaser Hashem, email yaser.hashem@u-bordeaux.fr

ABSTRACT

In eukaryotes, cap-dependent translation initiation represents one of the most complex processes along the mRNA translation regulation pathway. It results in the formation of several transient complexes involving over a dozen eukaryotic initiation factors (eIFs) and culminates in the accommodation of the start codon at the P-site of the small ribosomal subunit (SSU). In higher eukaryotes, the mRNA sequence in direct vicinity of the start codon, called the Kozak sequence (CRCCaugG, where R is a purine), is known to influence the rate of the initiation process. However, the molecular basis underlying its regulatory role remains poorly understood. Here, we present the cryo-electron microscopy structures of late-stage β-globin 48S IC and histone 4 (H4) 48S IC at overall resolution of 3Å and 3.5Å, respectively. We have prepared our complexes in near-native conditions, directly in cell lysates where initiation factors are present at normal cellular levels, and the composition of our complexes is corroborated by mass spectrometry analysis. Our cryo-EM structures shed light on the fine architecture of the mammalian late-stage (LS) initiation complex (IC) in the presence of "strong" (β-globin) and "weak" (H4) Kozak mRNA consensus sequences. From the high-resolution structures we unravel key interactions from the mRNA to eIF1A, eIF2, eIF3, 18S

rRNA, and several ribosomal proteins of the LS48S IC. In addition, we were able to study the structural role of ABCE1 in the formation of native 48S initiation complexes. Our results reveal a comprehensive map of the ribosome/eIFs –mRNA and –tRNA interactions and suggest the impact of mRNA sequence on the structure of the IC after the start-codon recognition. Our high-resolution structures also edify the molecular basis underlying regulation of the Kozak sequence-based translation initiation, including such phenomena as leaky scanning.

KEYWORDS

translation initiation, ribosome, Kozak consensus sequence, 48S initiation complex, eukaryotic initiation factor 1A, eukaryotic initiation factor 3, ATP-binding cassette, cryo-electron microscopy

INTRODUCTION

mRNA translation initiation in mammals is more complex than its bacterial counterpart at all levels. Indeed it includes more steps, more initiation factors and more regulation pathways. One can summarize the overall process in four steps, starting with pre-initiation. During pre-initiation, the ternary complex (TC) is formed by the binding of the heterotrimeric eukaryotic initiation factor 2 (eIF2) to one molecule of guanosine triphosphate (GTP) and the initiator methionylated tRNA ($tRNA_i^{Met}$). The TC then binds to the post-recycled ribosomal small subunit (SSU), also called 40S subunit. TC recruitment is partially mediated by eukaryotic initiation factors attached to the 40S, eIF1, eIF1A, and 13-subunit eIF3 complex. This leads to the formation of the 43S pre-initiation complex (PIC). The architecture of the 43S PIC has been already investigated thanks to several structures from yeast and mammals at intermediate to low resolutions[1–3].

The second step consists on the recruitment of the 5' capped mRNA and leads to the formation of the 48S IC. This step is mediated by the cap-binding complex composed of eIF4F (an assembly of cap-binding protein eIF4E, a scaffold protein eIF4G and an RNA helicase eIF4A), a second copy of eIF4A, and also eIF4B[4–7]. To date and in spite of numerous attempts, this step remains elusive to structural studies.

The third step is the scanning process for the start codon (AUG) in the 5' to 3' direction. This step was first investigated structurally in yeast on cognate and near-cognate start codons[8–9] using *in vitro* reconstituted complexes. Upon start-codon recognition, the codon/anticodon duplex is formed between the mRNA and the $tRNA_i^{Met}$ aided by the eIF1A N-terminal tail (NTT)[10–12]. The GTP is hydrolysed by the γ-subunit of eIF2, eIF1 dissociates from the P-site along with eIF1A C-terminal tail (CTT)[13] and the N-terminal domain (NTD) of eIF5 takes their place on the 40S[10], before eIF5-NTD dissociates in turn at a stage that still remains to be elucidated. This finally results in the formation of the late-stage 48S IC (that we describe later on in this work). The arrest of scanning and a cascade of structural rearrangements lead to the sequential dissociation of most eIFs including eIF2 upon the release of inorganic phosphate (Pi), generated from the GTP hydrolysis. eIF3 stays attached to the remaining complex probably through its peripheral subunits and leaves at a later stage during early elongation cycles[14,15]. During all these steps the post-recycling factor ABCE1 can bind directly to the 40S and act as an anti-ribosomal subunits association factor[16–18].

In the final fourth step, eIF1A and eIF5[19] are released and ABCE1 is replaced by the GTPase eIF5B on helix 14 of 18S rRNA. eIF5B stimulates the joining of the 40S and 60S ribosomal subunits, forming an 80S complex[20].

The efficiency of the translation initiation process has been extensively studied. The sequences flanking the AUG start-codon region have been identified as crucial for start site selection by the initiation complex[21–24]. The optimal sequence for translation initiation in eukaryotes was named after Marylin Kozak, who first discovered this dependence and defined the optimal sequence in vertebrates as CRCCaugG, where R stands for a purine[22,25]. In this motif, the mutations of positions -3 and +4, relative to the start-codon adenine, have the strongest influence on translational efficiency.[24] As a result, a sequence can be dubbed "strong" or "weak" by considering those positions. It was further shown that the substitution of A(-3) for pyrimidine, or mutations of the highly conserved G(+4), lead to a process known as "leaky scanning" with bypass of the first AUG and initiation of translation at the downstream start codon[21,22,26]. More recent studies observed a more extreme case of sequence-dependent translation initiation regulation, dubbed "cap-assisted" for certain cellular mRNAs, such as those encoding histone proteins and in particular histone 4 (H4) mRNA[27,28]. Cap-assisted internal initiation of H4 mRNA implies a very minimalistic scanning mechanism, which is possible thanks to the presence of a tertiary structure on the mRNA at the channel entrance. This element assists in placing the start codon very close to the P-site almost immediately upon its recruitment through the cap-binding complex.

In spite of the tremendous recent advances in understanding this phase of translation, high-resolution structural studies of the initiation process have been conducted by *in vitro* reconstitution of the related complexes. This approach often requires biologically irrelevant molar ratios of the studied eIFs[1–3,8,9], thus limiting insight into more subtle regulatory pathways. Moreover, the structures of the mammalian (pre)initiation complexes are still at intermediate resolutions approximating 6 Å[29–31]. Finally, although the role of ABCE1 was experimentally demonstrated as a ribosomal subunit anti-association factor preventing premature binding of the 60S[16], its impact on the 48S complex formation and conformation is still unclear.

Here, we present cryo-EM structures of LS48S IC formed after the recognition of the start codon in the presence of the "strong" Kozak mRNA (β-globin) and "weak" Kozak mRNA (histone H4), prepared in near-native conditions in rabbit reticulocyte lysate (Fig. 1a). Although the initiation regulation may differ mechanistically between these two archetype mRNAs, our structures provide a high-resolution snapshot on the Kozak sequence-dependent variable interactions in the LS48S IC in mammals.

RESULTS

Overall structure of the mammalian 48S initiation complex

The complexes were prepared using our approach[31] (see Methods for details) that consists on stalling the LS48S IC in rabbit reticulocytes lysate (RRL) using GMP-PNP (a non-hydrolysable analogue of GTP) on the two target cellular mRNAs: human β-globin mRNA (strong Kozak) and mouse histone H4-12 mRNA (weak Kozak) (Fig. 1a). These mRNA were transcribed and capped from BC007075 cDNA (β-globin) and X13235 cDNA (histone H4-12). The advantage of this approach is the ability to prepare ICs bound on different mRNAs of interest directly in nuclease-treated cell extract in the absence of the endogenous mRNAs, allowing for a study of regulatory aspects of the process in natural abundance levels of native eIFs at physiological molar ratios.

The composition of both complexes was investigated by mass spectrometry (Fig. 1b). Our MS/MS analysis reveals the incorporation in both complexes of all eIFs expected to be present after the start-codon recognition (eIF1A, eIF2α, eIF2β, eIF2γ, eIF3 complex, ABCE1). Also expected were extremely poor numbers of peptides and spectra for eIF1 detected in

either complex, corroborating that our complexes are at a late-stage after the start-codon recognition and eIF1 dissociation.

In parallel, we have subjected our prepared complexes to structural analysis by cryo-EM. The structure of the β-globin LS48S IC (3.0Å, 29% of the total number of 40S particles, Fig. 1c-d and Supplementary Fig. 1a-h) shows mRNA, 40S, eIF1A, the ternary complex (TC) and ABCE1. For the H4 mRNA 48S complex (3.5Å, 6.5% of the total number of particles, Fig. 1e-f and Supplementary Fig. 1p-w), the main reconstruction shows mRNA, 40S and the TC. We attribute the lower percentage of H4 LS48S IC formation to contamination with 60S subunits (~30%) (Supplementary Fig. 1w).

Interestingly, both our cryo-EM structures and MS/MS analysis show that the H4 LS48S IC displays a significant reduction in the presence of eIF1A, leaving only residual density for its presence in the cryo-EM reconstruction (Fig. 1e and Supplementary Fig. 2a). Similar observation can be made for ABCE1 in the H4 LS48S IC. Our reconstructions show also another class of IC with eIF3 that we will describe later.

Accommodation of the start codon in the late-stage LS48S IC

In both our reconstructions the codon:anticodon duplex is clearly formed, characterizing the cognate start-codon recognition (Fig. 1c,e and Fig. 2a,b). AUG codons of both mRNAs face the $_{(34)}$CAU$_{(36)}$ of anticodon stem-loop (ASL) tRNA$_i^{Met}$, within hydrogen-bonding distances (~2.7Å). In the case of β-globin mRNA, the codon:anticodon interaction is stabilized further by the N-terminal tail (NTT) of eIF1A (Lys7 interacts with the ribose of G(+3) from mRNA, Fig. 2d). The tail also interacts with the tRNA$_i^{Met}$ A(35) between Gly8 and Gly9. With few exceptions, this eIF1A NTT is highly conserved among eukaryotes (Fig. 2d). Recent fluorescence anisotropy with yeast reconstituted PICs[10] demonstrated that eIF1A binds with lesser affinity to a near-cognate start codon (UUG) compared to a cognate AUG. Along the same lines, only very residual density for eIF1A can be observed in the H4 LS48S IC structure (Supplementary Fig. 2a) (discussed further below), which reflects its weaker binding affinity after the start-codon recognition at this late stage to the 48S complex.

C1696 of 18S rRNA is stacked on the C(34) base at the very tip of the tRNA$_i^{Met}$ ASL that is paired to G(+3) of both β-globin and H4 mRNA (Fig. 2d). This contact between C1696 and C(34) is also found in the yeast partial 48S pre-initiation complex (py48S IC)[8] and it occurs even in the absence of any mRNA[30]. This stacking interaction may partly explain the

difference in recruitment of initiator tRNA between bacteria and eukaryotes. In bacteria, the initiator tRNA is recruited directly at the P site-accommodated start codon, whereas in eukaryotes, the $tRNA_i^{Met}$ is recruited at the pre-initiation stage of the complex before the attachment of mRNA into its channel. The $tRNA_i^{Met}$ ASL also interacts with the C-terminal tails of ribosomal protein uS19, attached to the 40S head (Fig. 2e, Supplementary Fig. 3b) through its Arg140 that contacts A(35).

We then compared the overall conformation of the 40S between both complexes and we observed that in the β-globin IC the head of the SSU is tilted downwards by ~2° and swivelled toward the solvent side by ~3° when compared to its counterpart in the H4 IC (Supplementary Fig. 2c). We attribute these subtle conformational changes to the dissociation of eIF1A in H4 LS48S IC, after the start-codon recognition, due to the loss of contacts between eIF1A and the 40S head.

Interaction network of the Kozak sequence (-4 to +4) with 40S and initiation factors

Thanks to the resolution of our complexes, we are able to clearly visualize the interactions of the mRNA Kozak sequence residues surrounding the start codon with the components of their respective 48S ICs (Fig. 3a). The (+4) position, mainly occupied by a guanine in eukaryotic mRNAs, plays a pivotal regulatory role[21,25]. Our reconstructions demonstrate the structural importance of this position to both mRNAs. In the β-globin LS48S IC, the highly conserved Trp70 from eIF1A is trapped between the G(+4) and the A1819 from h44 18S rRNA of the A-site by stacking interactions (Fig. 2c). Interestingly, the interaction of the position (+4) mRNA position with h44 was shown by cross-linking studies[32]. Our β-globin LS48S IC structure also shows the proximity of uS19 C-terminal tail to (+4) mRNA position, which can also be corroborated by several cross-linking studies[32–34]. In H4 mRNA, G is replaced by a U at position (+4), therefore the stacking interaction with eIF1A appears weaker than when a G is present. Moreover, nucleotides A1818 and A1819 have even more scant densities, indicating their undetermined conformations probably linked to this poor stacking (Fig. 2c). Our reconstructions therefore suggest the structural importance of the (+4) position in the interaction with of eIF1A.

Another crucial position in the Kozak consensus sequence is at (-3), often occupied by an adenine (Fig. 1a). This nucleotide shows several contacts with ribosomal proteins and initiation factors, including Arg55 from domain 1 (D1) of eIF2α (Fig. 3a), which was

reported previously in the py48S IC structure[8,10]. However, in the yeast structure the A(-3) base is in the *syn* conformation and in both our mammalian ICs the adenine is in the *anti* conformation. Noteworthy, the near-cognate yeast mRNA present in the py48S IC structure[10], contains adenines at the positions (-1) and (-2), which in principle could create an ideal stacking context for the A base in (-3), thus explaining this difference in conformation compared to our mRNA where these positions are occupied by two cytosines. The (-3) position further interacts with the G957 nucleotide at the 40S platform (Fig. 3a), highlighted in earlier studies[35]. In addition, cross-linking studies of reconstituted mammalian PIC previously demonstrated that eIF2α and uS7 interact with the (-3) nucleotide, and uS7 with the (-4) nucleotide[32,33]. The interaction of uS7 through its β-hairpin was also suggested in the py48S IC structure, due to their proximity in space[8,10]. However, in our structures, this interaction cannot be confirmed since the electron density at this specific region is very disperse, probably because of the flexibility of this part of uS7 (see Discussion, Supplementary Fig. 6a) compared to its other parts.

Position (-4) of both mammalian mRNAs interacts with ribosomal protein eS26 through its His80. However, we have found that in the case of the β-globin "strong" Kozak mRNA, position (-4) is a cytosine and appears to interact mildly with eS26 His80 (Fig. 3a, left panel), as its weak density suggests. Whereas when this position is an adenine, like in the "weak" Kozak of the H4 mRNA, a stronger stacking interaction occurs, which could further participate in stabilizing the mRNA in its channel (Fig. 3b, right panel). Consequently, the mRNA in this latter case adopts a slightly different conformation. A possible result of this difference is the observed tighter interaction with eIF2α Arg55 residue from domain D1 (Fig. 3a), as its density is better defined in H4 than in β-globin.

Finally, upstream residues near the start codon in our complexes are in contact with 18S rRNA including G1203 from the head rRNA which interacts with the phosphate of A(+1) and stacks with C(-1) of both mRNAs (Fig. 3a).

eIF1A interaction with the Kozak sequence and the 18S rRNA

In addition to the above-mentioned contact with the start codon and G(+4), eIF1A can potentially establish several interactions at more distal positions in the Kozak sequence, closer to the mRNA channel entrance. Indeed, Arg12, Lys67 and Lys68 in eIF1A are in close proximity to C(+7), G(+6) and U(+5) (Fig. 4a). eIF1A NTT also interacts with the 18S rRNA

(Gly9 and Lys10 with C1696; and Lys16 and Asn17 with C1327) (Fig. 2d and Fig. 4a). Other contacts involve the loops of the eIF1A OB domain with the 40S near the A-site (Fig. 4b-d): namely, Asn44 and Arg46 are in contact with A1817-A1819 and C1705 from h44 of 18S rRNA (Fig. 4c); moreover Lys64 and Arg62 contact G604 and C605 of h18 18S rRNA (Fig. 4d). In addition, Arg82, Tyr84 and Gln85 of eIF1A contact Glu58, Leu91 and Gly56 of ribosomal protein uS12 (Fig. 4d); finally, Asp83 is in contact with Arg82 of eS30 (Fig. 4d). Putting together, the above-mentioned interactions might depend on the mRNA sequence and perhaps they can have influence on the stability the cognate start-codon duplex with its anticodon by the NTT of eIF1A (residues Lys7, Gly8, Gly9 and K10).

Noteworthy, eIF1A NTT was shown to interact with eIF5[36,37], but because of its clear involvement in the start codon/anticodon duplex, we suggest that this eIF5 interaction occurs during the pre-initiation phase and very shortly after the recruitment of the mRNA.

mRNA interactions with the 48S beyond the Kozak sequence

The mRNA density at distal positions from the Kozak sequence appears disperse when filtered to high-resolution, suggesting an overall flexibility at both the entrance and the exit of the channel (local resolution of ~6 to ~9 Å). Nevertheless, several contacts can be observed at the entrance and exit sites of the mRNA channel of the β-globin and H4 LS48S ICs. These interactions are common to both complexes and could be more site-specific than they are sequence-specific.

At the entrance of the mRNA channel during this late stage of the initiation process, the mRNA extensively interacts with conserved residues of the 40S ribosomal proteins uS3 and eS30, and with rRNA h16 in positions spaning from +10 to ~ +20 (Fig. 3b). For instance, the conserved Arg117 of the head protein uS3 contacts the mRNA at the channel entrance. This residue was recently indicated as important for stabilizing the P_{IN} closed state of the 48S in yeast IC[10] and for the initiation accuracy in the presence of suboptimal Kozak sequence by *in vivo* assays in yeast[38]. The contribution of this charged residue of uS3 contacting the mRNA is partially corroborated by cross-links in a previous study[33]. More globally, charged amino acid residues from uS3 helix α (residues 117-128) are in close proximity to nucleotides from positions (+14) to (+18) (Fig. 3b). Moreover, residues from a β-hairpin (residues 142-145) can potentially contact nucleotides C(+9) and C(+10) of the mRNA. For ribosomal

protein eS30, Lys126 is in close distance to the bases of G(+12) and A(+13). The proximity of A(+13) of mRNA to Ala133 of uS5 can also be noted (Fig. 3b).

On the other side at the mRNA exit channel, we can observe the exit of both β-globin and H4 mRNAs from their respective 48S ICs (Fig. 3c). The 5' untranslated region (5'UTR) for β-globin mRNA is substantially longer than for H4 (50 nt and 9 nt, respectively) (Fig. 1a, see Methods). We compared the mRNA exit channels of both complexes below the ribosomal head protein RACK1, which unambiguously shows the expected larger 5' UTR for β-globin LS48S IC compared to H4 (Supplementary Fig. 2b). We were able to spot several possible contacts of ribosomal proteins at the exit site with mRNA nucleotides in both LS48S IC, including eS28 (Arg67 with A(-5), and Arg68 with A(-7)) as well as eS26 (Ile41 with C(-8), Arg42 with A(-9) and Arg100 with both these nucleotides) (Fig. 3c), in agreement with previous cross-linking results[32,33].

Interactions of 48S initiation complex with the $tRNA_i^{Met}$

The overall accommodation of the mammalian ASL resembles its yeast counterpart found in the P_{IN} state[8,10]. The 48S IC-$tRNA_i^{Met}$ interaction network is summarized in the Supplementary Fig. 3b. In both IC complexes, we can observe a density attached to A(37), in which we can model the threonylcarbamyol group forming a t^6A modification (Fig. 3a and Fig. 5a). This modification mediates the binding of $t^6A(37)$ to the 2'OH of C(-1) in the mRNA, and therefore can further stabilize the start-codon recognition. It is tempting to suggest that C(-1):$t^6A(37)$ interaction is required for efficient translation in mammals. This mRNA C(-1) position is conserved in higher eukaryotes, as revealed by quantitative sequence analysis[39], and forms part of the Kozak sequence. Interestingly, the electron density for this modification is even stronger in the H4 mRNA complex than in β-globin, even at a lower resolution (3.5Å). We therefore suspect that this interaction could be more important in the case of suboptimal Kozak sequences, where this modification could compensate for the loss of some interactions with the 48S IC, compared to a strong Kozak mRNA. The same interaction in the case of adenine at (-1) position is not excluded, however its nature and conformation will be different.

Despite the universal presence of the t^6A hypermodified base in all organisms, and a crucial role in translation efficiency[40,41], it has only recently been shown that it directly contributes to the AUG recognition accuracy. In the py48S-eIF5 IC structure at a resolution of

3.5 Å[10], t^6A of the initiator $tRNA^{Met}$ was suggested to enhance codon-anticodon base-pairing by interacting with A(-1) and by a stacking on the downstream base-pair involving A(+1). Thanks to our mammalian LS48S IC at 3.0 Å, we can clearly observe the threonylcarbamyol group in a different conformation, placing the carboxyl group within hydrogen-bonding distance (2.7Å) from the 2'OH group of C(-1) (Fig. 5a). However, this modification does not appear to stack over the downstream base-pair as previously suggested. Noteworthy, in yeast there is a preference for an A at position (-1)[21,42], while it is a C(-1) in mammals.

Furthermore, C(34), that is a part of the anticodon, is stabilized by the modified U1248 of helix31 of rRNA ($m^1acp^3\Psi$, 1-methyl-3-(3-amino-3-carboxypropyl) pseudouridine) [43,44] (Fig. 5b), which was previously reported in the py48S-eIF5 (modified U1119)[10]. In addition, the neighbouring nucleotides, C(32) and C(33), are in contact with C-terminal arginine Arg146 of ribosomal head protein uS9 (Fig. 5c).

Aside from its role in the codon-anticodon stabilization, uS19 together with uS13 were found to contact other parts of the $tRNA_i^{Met}$ through their highly conserved C-terminal tails (Fig. 2e and Supplementary Fig. 3b). Thr136 side chain of uS19 interacts with the guanine backbone of three conserved G–C base pairs in ASL that are crucial for stabilization of the initiation complex in eukaryotes[45]. Residues Thr145 and Gly147 of uS13 are in contact with the phosphate groups of U28 and G29 (Fig. 2e). Consistent with previous reports[8,10], several residues from domain D1 of eIF2α (Arg57, Ser58, Asn60 and Lys61) interact with the cysteines backbones of three conserved G–C base pairs of the ASL (phosphate groups of C39-41) (Fig. 5d and Supplementary Fig. 3b).

ABCE1 binding to the initiation complex is NTP-dependent

ABCE1 (named Rli1 in yeast) is a conserved NTP-binding cassette ABC-type multi-domain protein that plays a role in translation initiation as well as translation termination and ribosome recycling[46–51]. It contains two nucleotide-binding domains (NBDs), where the two NTP molecules bind. Its N-terminal NBD contains two iron-sulphur clusters $[4Fe\text{-}4S]^{2+}$ [52–54]. In our β-globin LS48S IC structure, the NTP-binding cassette of ABCE1 displays lower local resolution (between 3.5 and 5 Å, Fig. 6a and Supplementary Fig. 1d) compared to the average resolution, likely due to the flexibility of the NBDs. In H4 LS48S IC structure we observe only a residual density of ABCE1 (Supplementary Fig. 1l), which can be caused by a slightly different conformation h44 18S rRNA in the absence of eIF1A.

The iron-sulfur (Fe-S) binding domain presents a higher local resolution than the NBDs (between 3 and 4 Å, Fig. 6b and Supplementary Fig. 1d). Therefore, we can clearly distinguish the Fe-S clusters in the cryo-EM density map as well as the presence of two bound nucleotides that probably represent GMP-PNP that was used to stall the initiation complexes by blocking eIF2γ (Fig. 6a-c).

NBD1 contacts five nucleotides in the 18S rRNA helix14 (U478, A455, A454, C453 and C452) through residues located in a helix-loop-helix motif (Ser150) and in the hinge-N (Arg306 and Asn310) (Fig. 6f,g, Supplementary Fig. 4a). NBD2 also makes contacts to nucleotides A455, A454, C453 through the hinge-C residues Lys584 and Ile583, conserved in higher eukaryotes (Fig. 6g and Supplementary Fig. 4a). Residues Arg567, Arg576 and Gln588 from NBD2 are in close proximity to the 18S rRNA and can potentially be involved in the interaction with the latter, as suggested by their conservation (Supplementary Fig. 4a). Moreover, NBD1 residues Pro265 and Asp266 interact with residues from the C-terminal helix of ribosomal protein eS24 such as Gly128 (Fig. 6f).

As for the Fe-S binding domain, it interacts with the 18S rRNA (Arg7 from strand β_1 to C471, Lys20 to G1718, Pro43 to A1719 and Asn74 to G470) (Fig. 6d). In addition, ABCE1 interacts with ribosomal protein uS12 residues Ile50, Leu52 and Ile75 via a hydrophobic pocket formed by helices α_2, α_3 and cluster (II) (residues Pro30, Val31, Ile56, Ile60) (Fig. 6d). This is consistent with previous structural and crosslinking studies in yeast and archeal complexes[16,17].

In order to investigate the effect of ABCE1 binding on the structure and composition of the LS48S IC, we purified β-globin complexes using our choice strategy[31] supplemented with 10 mM of ATP (see Methods), thus taking advantage of the ability of ABCE1 to hydrolyse ATP, in contrast to the obligate GTPase eIF2. The β-globin-ATP 48S IC was then analyzed by cryo-EM and yielded two main LS48S IC reconstructions at ~14Å and ~10Å, that differ in the presence and absence of eIF3, respectively (Fig. 6h and Supplementary Fig. 1x,y). The addition of ATP most likely causes the replacement of the GMP-PNP molecules in the NBD pockets by ATP molecules that was then hydrolyzed by ABCE1. Our structures clearly reveal the dissociation of ABCE1 as a result of ATP addition, consistent with recent structural and biophysical studies[16,17,55]. Aside of the absence of ABCE1, the global structure of the β-globin-ATP LS48S IC is identical to its higher-resolution counterpart without ATP (Fig. 6h), thus very likely excluding a direct active role of ABCE1 in the assembly of the initiation complex. It is reasonable to assume that in the cell ABCE1 undergoes on/off cycles

to the IC in an ATP-dependent manner, as we have previously suggested[18]. However, these results do not contradict the demonstrated function for ABCE1 as an anti-ribosomal-subunit association factor.

eIF3 in the late-stage 48S initiation complex

Nearly 15% (~5% of the total particles count) of the particles corresponding to the β-globin LS48S IC structures contain a density for eIF3 at the solvent side. After extensive particle sorting and refinement, a reconstruction of the β-globin LS48S IC showing eIF3 was obtained at a resolution of ~3.6 Å (Supplementary Fig. 1i-o), thus allowing verification of the recognition of the start codon for this specific class (Fig. 7a,b). eIF3 a and c subunits, i.e. those that bind directly to the 40S, are mostly resolved at a resolution ranging from 3.5 to 4.5 Å, enabling for the first time a model of the exact residues in interaction with the 40S subunit (Fig. 7c-g). We find that eIF3a in β-globin LS48S IC shows several contacts with 40S ribosomal body proteins, with eIF3a residues Glu9, Asn16 and Glu17 interacting with eS1 residues Glu78, Ile189 and Pro190, respectively (Fig. 7f,g). In addition, eIF3c residues Lys332, Lys333 and Thr381 contact the small ribosomal subunit protein eS27 via residues Val53, Gln65 and Glu75 (Fig. 7e). Moreover, residues Leu379 and Ala380 interact through their backbone with nucleotides C1113, C1114, A1115 and U1116 part of the apical loop of expansion segment 7.

The eIF3d subunit structure is at lower resolution as compared to the eIF3 octamer core. Nevertheless, secondary structure elements can clearly be depicted when filtered to a lower resolution (6Å), which enabled the fitting of its partial crystal structure in our density[56]. The modelled eIF3d displays contacts with several ribosomal head proteins: helix-α_{12} contacts the N-terminal loop of uS7; the loop between β_9 and β10 contacts RACK1 (loop located between strands β_{6D} and β_{7A}); and β-strand loops and the "RNA gate" insertion contacts eS28 (strand β_3, loops between helices α_8 and α_9).

In spite of the low local-resolution of this particular subunit, our structure provides clues to the demonstrated interactions of eIF3d and eIF3a with the mRNA 5'UTR (Fig. 7d). Thus, shown by the residual electron density traces, the 5'UTR of numerous mRNAs such as β-globin can possibly interact with different parts of eIF3d: N-terminal loop and strand β_2 (residues S166 to E172) and a loop between β_{11} and β_{12} (residues Asn513 and Lys514) (Fig. 7d). These residues of eIF3d are better conserved in higher eukaryotes, indicating a possible

species-specific regulation. An interaction with 5' RNA terminus recognition motif was also previously reported[56]. More insight into the interaction patterns of eIF3d with different mRNA 5'UTRs in the context of translation initiation will be an important goal in future studies.

Finally, the mRNA 5'UTR of β-globin can also interact with eIF3a (residues Tyr4-Pro8, Lys45-His47 and Gln80-Asn82) (Fig. 7d). Similarly to eIF3d, these residues are mainly conserved among higher eukaryotes (Fig. 7g). It was previously shown that the eIF3a–PCI domain (a domain with a common fold for proteasome, COP9, initiation factor 3) is critical for stabilizing mRNA binding at the exit channel[57]. However, because of the low local resolution of the β-globin mRNA 5'UTR in our reconstruction, we do not exclude other patterns of interaction. H4 LS48S IC also shows the residual presence of eIF3 core, however particle sorting reveals a reconstruction containing eIF3 (Supplementary Fig. 1w) at only intermediate resolution because of the low number of particles.

DISCUSSION

Our cryo-EM structures reveal in detail the accommodation of the mRNA containing either "strong" or "weak" Kozak sequences in the context of the late-stage mammalian initiation complex. In our presented IC structures, we did not identify any density corresponding to the eIF1. Combined with codon-anticodon complex formation and several conformational changes characteristic of the stage after the start-codon recognition, we have dubbed our complexes "late-stage 48S IC (LS48 IC)". Initiation on β-globin and H4 mRNA may undergo different regulatory processes as previously reported[27,28], however in our structures we only analyse the mRNA nucleotide interactions of the Kozak sequences without dwelling on the exact regulation mechanism that may be in part influenced by the different interaction patterns that we observe. We therefore believe that our two archetype mRNA Kozak sequences are representative of native cellular mRNAs incorporating weak and strong sequences, as their observed interactions are purely the result of sequence differences, and unrelated to specific regulatory pathways. The interactions described in this work are summarized in the Fig. 7h and Supplementary Table 2.

In the Kozak sequence, the position (+4) appears to play a role in both mRNAs, where it is a G in β-globin and U in H4. At this position, the crucial interaction with Trp70 of eIF1A

appears to be weaker in the case of H4 IC when compared to β-globin IC, as indicated by our mass spectrometry normalized spectral counts and the cryo-EM reconstructions. We suggest that the poor abundance of eIF1A in the H4 LS48S IC cryo-EM reconstruction (Fig. 1, Supplementary Fig. 1l and Supplementary Fig. 2a) might not suggest a negligible role for this initiation factor in suboptimal Kozak consensus mRNAs during scanning. Rather, it simply shows its weaker interaction in the complex after start-codon recognition, supported by our semi-quantitative MS/MS analysis. As expected, the biggest decrease in spectral counts was observed for eIF1A in the H4 LS48S IC compared to its β-globin counterpart (Fig. 1b). The weak above-mentioned stacking interaction with mRNA purine at (+4) position is likely to be the reason of the affinity drop and subsequent weaker interaction between eIF1A and the 48S in the case of H4, after scanning and the recognition of the start codon. This observation is consistent with the suggested special translation initiation mechanism for H4[27,28]. It was suggested that H4 mRNA undergoes an unconventional "tethering mechanism", where the ribosomes are tethered directly on the start codon without scanning[27]. This particular mechanism was proposed to occur thanks to the presence of two secondary structure elements present downstream of the start codon in H4 mRNA (position +19), contacting h16 of rRNA, however, we cannot confirm the presence of these elements by our structure. We believe that eIF1A is present in the H4 complex during pre-initiation and the short scanning process, and only after the recognition of the start codon the affinity can be affected by the mRNA Kozak context.

Regarding the N-terminal tail of eIF1A, it is present in the A-site starting from the scanning process, as shown by our structure and also yeast 43S PIC and closed-48S ICs[9,10]. It is important to emphasize the binding of eIF1A on the H4 IC at a certain stage, because we can observe residual electron density for this initiation factor in the H4 LS48S IC (Supplementary Fig. 1l and Supplementary Fig. 2a). This observation tends to validate the unconventional very short scanning mechanism proposed for H4 mRNA[27], yielding in the faster accommodation of the start codon as compared to β-globin mRNA. Nevertheless, in the early initiation steps eIF1A may interact with eIF5, as demonstrated by biophysical studies of *in vitro* purified proteins[36].

The importance of positions (+4) and (-3) of mRNA has been pointed out in previous studies (reviewed in[22]), however the significant involvement of position (-4) in mammals was not highlighted. Similarly to py48S-eIF5 IC[10], our structures show that this position in the case of "weak" Kozak A(-4), can be stabilized by residue His80 of eS26. This residue is highly conserved in eukaryotes (Fig. 3a), therefore showing a universal mode of interaction.

The role of eS26 (and eS28) in the accommodation of the 5' UTR was also highlighted by chemical cross-linking studies performed with 80S ribosome assembled on histone H4 mRNA[28]. Remarkably, in the case of the yeast IC, it was shown that the interaction between the A nucleotide at the position (-4) and His80 of eS26 does occur through stacking[10], in contrast to our study where we can show C and mainly A (-4) stacked below the His80 of eS26. This variation is probably due to the different sequence between our mRNA and those of yeast, where positions -1 to -3 are occupied by A, that favours more stacking between the nucleotides bases and consequently more twist (Fig. 5a). Indeed, the kink in E/P site in py48S-eIF5 IC[10] is sharper than in the case of mammalian 48S IC (Supplementary Fig. 3a). Interestingly, a previous cross-linking study in the context of the human 80S ribosome highlighted the eS26 binding to G and U nucleotides in the region (-4) to (-9) of the mRNA[58]. In a more recent biochemical study in yeast, it was shown that the mRNAs bound to the ribosomes depleted in eS26 (RpS26 in Saccharomyces) translate poorly when compared to those enriched in eS26[59]. It was also reported in the same study that RpS26 is necessary for preferential translation of mRNAs with A at (-4) position and not G, showing that the interaction is very specific and not simply purine/pyrimidine dependent. Therefore, the His80 eS26 recognition is likely optimal for mRNA sequences containing A at this position. In yeast, however, the nucleotide context surrounding the start codon is less critical and but shares with mammals the importance of the (-3) position[21,60]. Further studies on translation of mRNAs containing mutations at these positions will help unveiling the mechanism of scanning in mammals and will shed light on the leaky scanning mechanism.

The tails of uS13 and uS19 have been previously shown to make direct interactions with the ASL of the peptidyl-site-exit-site (P/E) tRNA in presence of elongation factor G (EF-G)-ribosome complex in a pre-translocation state in prokaryotes[61]. To the best of our knowledge, these proteins have not yet been reported to be particularly involved in the initiation process in eukaryotes. Nevertheless, our structures are supported by earlier crosslinking studies of human 80S ribosome showing that the tail of uS19 is located closer to the decoding site than that of prokaryotic S19[58]. In the case of *S. cerevisiae* uS13 and uS19, the C-terminal parts are not conserved, compared to human protein homologues (Fig. 2e) and they have never been observed to interact with the $tRNA_i^{Met}$ [8,10]. Other fungi, such as *N. crassa,* possess very similar sequences to mammalian counterparts and probably would demonstrate similar interactions to $tRNA_i^{Met}$ as shown by our structures. Moreover, and in contrast to yeast, *N. crassa* possesses a mammalian-like eIF3. This is in line with the recent

genome-wide mapping of Kozak impact in the fungal kingdom, showing the particularity of start-codon sequence context in *S. cerevisiae* compared to other fungi [62].

The ribosomal protein uS7 is located close in space to the position (-3) of mRNA, but we cannot confirm its interaction with this nucleotide due to the lack of the density for this part of the protein. Noteworthy, this is the only flexible region in this protein structure (Supplementary Fig. 5a), containing the crucial β-hairpin in the case of bacterial and yeast initiation[63,64]. This region in yeast contains the glycine-stretch GGGG (residues 150-153), whereas in human it is GRAG (residues 129-132). Genetic experiments on single-point mutants of this β–hairpin demonstrated almost unchanged phenotype for human-like G151R and G152A mutations[65], but the G151S mutation was lethal. More recent work from the same research group showed that by using genetic and biochemical approaches, uS7 modulates start-codon recognition by interacting with eIF2α domains in yeast[63]. The residues implicated in the described interactions are highly conserved and are also present in mammals (Supplementary Fig. 5b). Therefore, we speculate that the effect of the studied substitutions of uS7 might result in similar phenotypes in mammals.

When compared to the yeast 40S ribosome-ABCE1 post-splitting complex[16], we do not observe any large structural differences. However, NTP pocket 1 appears to be more open than NTP pocket 2 (Fig. 6a,e, coloured frames) and it is similar to the "open state" found by X-ray crystallography[52]. Recent single-molecule-based fluorescence resonance energy transfer (smFRET) study on archeal ABCE1 showed that two NTP sites are in asymmetric dynamic equilibrium and both NTP sites can exist in different conformations[55]. Therefore, we propose that in ABCE1 in LS48S IC is present in its asymmetric conformer, where NTP pocket 1 is in the open state, whereas pocket 2 is in the closed state. The position of ABCE1 in the IC suggests steric incompatibility with the human re-initiation factor, eIF2D[66]. Indeed, the winged helix (WH) domain of eIF2D was found to interact with the central part of h44 ribosomal RNA in the absence of ABCE1, at the exact position of the ABCE1 Fe-S cluster (I) in our LS48S IC (Supplementary Fig. 4b). This cluster also shows sequence similarity to the C-terminal SUI domain of eIF2D, found to be located in the re-initiation complex at the top of h44 rRNA (Fig. 4a)[66].

Regarding the relatively low abundance of eIF3 in our complexes, we believe that after the LS48S complex formation, eIF3 simply detaches from the 40S, probably during the grid preparation as has been consistently observed in structural studies of analogous complexes. The superposition of the eIF3 octamer to our previous structure of the *in vitro*

reconstituted 43S PIC[30] showed high structural similarity (RMSD 1.3 Å over all atoms). Consistent with our previous study[30], our structure reveals a sizable unassigned density at the mRNA channel exit, interacting mainly with eIF3 a and c (Supplementary Fig. 6b,c). Because of its location, it is tempting to attribute this unassigned density to the 5'UTR of mRNA, however its presence in 43S PIC[30], which does not contain any mRNA, strongly contradicts this assignment. Thus, following our previous suggestion, we believe that this density belongs mainly to flexible segments of eIF3d. Finally, the eIF3 b-i-g module is not visible in our structure, however in the case of py48S-eIF5 IC structure[10], it was demonstrated that these subunits relocate together to the solvent site upon start-codon recognition.

CONCLUSIONS

Our cryo-EM structure at 3.0 Å represents the highest resolution reconstruction of a mammalian 48S initiation complex. It refines our understanding of the architecture of mammalian late-stage 48S initiation complexes and provides structural insights into the Kozak sequence regulatory role in canonical cap-dependent translation initiation. The data presented here suggest a correlation between nucleotides at crucial positions surrounding the start codon and the initiation complex composition, after the recognition of the start codon, but also explain some of the molecular basis underlying the "strength" of a Kozak sequence. Furthermore, our results demonstrate that the binding of ABCE1 during initiation doesn't impact the conformation of the 48S IC. In the presence of ATP, ABCE1 dissociates from the IC, likely after the ATP-dependent hydrolysis. Finally, our structure reveals for the first time the molecular details of the mammalian eIF3 core interactions with the 40S ribosomal subunit, at near-atomic resolution.

ACKNOWLEDGMENTS

We thank Gabor Papai and Julio Ortiz Espinoza (IGBMC, Strasbourg, France) for assistance in data acquisition and Franck Martin for providing the plasmid for β-globin mRNA. We also thank the High-Performance Computing Centre of the University of Strasbourg (funded by the Equipex Equip@Meso project) for IT support and the staff of the proteomic platform of Strasbourg-Esplanade for conducting the nanoLC-MS/MS analysis (funded by LABEX: ANR-10-LABX-0036 NETRNA). The mass spectrometry instrumentation was granted from Université de Strasbourg (IdEx 2015 Equipement mi-lourd). We thank Alan G. Hinnebusch

for critical reading of the manuscript and Cameron Mackereth for numerous useful comments to it. This work was supported by the ANR grant ANR-14-ACHN-0024 @RAction program ‘‘ANR CryoEM80S’’ and the ERC-2017-STG #759120 “TransTryp” (to Y.H.).

AUTHOR CONTRIBUTIONS

A.S. has conducted samples preparation and optimization for cryo-EM study and mass-spectrometry analysis. Y.H. and A.S. performed cryo-electron microscopy experiments. Y.H., E.G. and A.S. have interpreted the data. Y.H. and E.G. carried out data processing, structural analysis and wrote the manuscript with input from all authors. A.B. performed the atomic modelling. L.K. performed the mass spectrometry experiments. All authors read and commented the manuscript. Y.H. directed research.

AUTHOR INFORMATION

Cryo-EM maps and atomic coordinates for the reported structures have been deposited in the Electron Microscopy Data Bank under the accession numbers **EMD-xxxx, EMD-xxxx, EMD-xxxx, EMD-xxxx, EMD-xxxx** and **EMD-xxxx**, and in the Protein Data Bank under the accession numbers **xxxx, xxxx, xxxx** and **xxxx**.

The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to Y.H. (yaser.hashem@u-bordeaux.fr).

METHODS

In vitro transcription and capping

Human β-globin mRNA was prepared as previously described[31]. The template for mouse H4–12 mRNA (375 nt; accession number X13235) was generated by PCR amplification extended on its 3′-end with a 5′-$(CAA)_9$CAC-3′ tail from plasmid containing the gene synthetized by Proteogenix. The PCR product purification and in vitro transcription of mouse H4–12 mRNA were performed as described for the preparation of β-globin mRNA. The pure transcripts

were capped using the ScriptCap™ m^7G Capping System (Epicentre). Radiolabelled transcripts were obtained by substituting the GTP from the kit with [α^{32}P]GTP.

Sample preparation for Cryo-EM

β-globin 48S IC and H4 48S IC were isolated from nuclease-treated rabbit reticulocyte lysate (RRL-promega) as previously described[31] except that lower concentration of Mg (0.5 mM) was used for ribosome complex assembly. Moreover, we have optimized the gradient fraction collection using BioComp Piston Gradient Fractionator™ devise. The formation of translation initiation complexes has been monitored following the ribosome profile via the UV absorbance (OD260) and the radioactivity profile of the ^{32}P labeled β-globin or H4 mRNA. Fractions containing β-globin/48S IC or H4/48S IC were centrifuged at 108,000 rpm (S140AT Sorvall-Hitachi rotor) for 1 h at 4°C and the ribosomal pellet was dissolved in a buffer containing 10 mM Hepes/KOH pH 7.4, 50 mM KOAc, 10 mM NH_4Cl, 5 mM $Mg(OAc)_2$ and 2 mM DTT.

Grids preparation and data collection parameters

The grids were prepared by applying 4 μL of each complex at ~70 nM to 400 mesh holey carbon Quantifoil 2/2 grids (Quantifoil Micro Tools). The grids were blotted for 1.5 sec at 4oC, 100% humidity, using waiting time 30 s, and blot force 4 (Vitrobot Mark IV). The data acquisitions were performed for the β-globin•GMP-PNP and H4•GMP-PNP ICs on a Titan Krios S-FEG instrument (FEI) operated at 300 kV acceleration voltage and at a nominal underfocus of Δz = ~ 0.5 to ~ 3.5 μm using the CMOS Summit K2 direct electron detector 4,096 x 4,096 camera and automated data collection with SerialEM[68] at a nominal magnification of 59,000 x. The K2 camera was used at super-resolution mode and the output movies were binned twice resulting in a pixel size of 1.1Å at the specimen level (The calibrated magnification on the 6.35 μm pixel camera is 115,455 x). The camera was setup to collect 20 frames and frames 3 to 20 were aligned. Total collected dose is ~26 $e^-/Å^2$. In addition, Cryo-EM images of the β-globin•GMP-PNP+ATP were collected on a Polara Tecnai F30 cryo-transmission electron microscope (FEI instruments) operated at 300 keV acceleration voltage and at a nominal underfocus of Δz = ~ 0.5 to ~4.0 μm, using a direct electron detector CMOS (Falcon I) 4,096 x 4,096 camera calibrated at a nominal magnification of 59,000 x, resulting in a pixel size of 1.815 Å.

Image processing

SCIPION[69] package was used for image processing and 3D reconstruction. MotionCor[70] was used for the movie alignment of 8238 movies from the β-globin complex and 8520 movies for the H4 complex. CTFFIND4[71] was used for the estimation of the contrast transfer function of an average image of the whole stack. Particles were selected in SCIPION. Approximately 1,067,000 particles were selected for the β-globin•GMP-PNP IC, 666,000 particles for the H4•GMP-PNP and 200,000 particles for the β-globin•GMP-PNP+ATP IC. RELION[72] was used for particle sorting through 3D classification via SCIPION, (please refer to ED Figure1 for particle sorting details for all three complexes). Selected classes were refined using RELION's 3D autorefine and the final refined classes were then post-processed using the procedure implemented in RELION applied to the final maps for appropriate masking, B factor sharpening, and resolution validation to avoid over-fitting.

Model building, map fitting and refinement

The four initiation complexes were modelled based on the previous initiation Oryctolagus cuniculus complex (PDBID: 5K0Y)[31] resolved at 5.8Å. Adjustments of RNA and proteins were done using the visualization and modelling software UCSF Chimera version 1.12 (build 41623)[73]. Sequences of modelled factors from Oryctolagus cuniculus were retrieved using BLAST[74] tools in the NCBI database[75] using respective template sequence described below. Templates structures were extracted from the PDB[76]. ABCE1 from Saccharomyces cerevisiae 40S complex (PDB ID: 5LL6 chain h)[16] was used as template to thread ABCE1 of Oryctolagus cuniculus in Swiss-model[77] webservice. The core of initiation factor 3 (eIF3) composed of subunits A, C, E, F, H, K, L and M was extracted from corresponding mammalian eIF3 (PDB ID: 5A5T)[30]. Initiation factor 3D (eIF3D) from Nasonia vitripennis (PDB ID: 5K4B)[56] was the template to thread eIF3D of Oryctolagus cuniculus in Swiss-model. Initiation factor 1A (eIF1A) template was extracted from Saccharomyces cerevisiae 48S preinitiation complex (PDB ID: 3JAP chain i)[9] and thread into Oryctolagus cuniculus in Swiss-model. The ternary complex (TC) was affined from Oryctolagus cuniculus initiation complex (PDB ID: 5K0Y)[31]. Both messenger RNA (globin and H4) were modelled using modelling tools of Chimera. Refinements were done on all four complexes in their corresponding maps. The refinement workflow followed four major steps that applied to all initiation complexes. First, a Molecular Dynamic Flexible Fitting (MDFF)[78] ran for 200000

steps with gscale of 1 (potential given to the density map to attract atoms in their density). The trajectories reached a plateau of RMSD curve around frame 160 for the four complexes. A minimization followed the trajectories to relax the system. MDFF ran on VMD[79] 1.9.2 coupled with NAMD2[80] v.1.3. software. Next steps of refinement required the usage of several specialized tools for RNA and proteins geometry included as modules in PHENIX[81] version 1.13-2998-000 software. Phenix.ERRASER[82] is a specialized tool for RNA refinement and Phenix.real_space_refine is specialized for proteins geometry and density fitting refinement. Finally, a last step of minimization using VMD and NAMD2 was applied. Assessment and validation of our models were done by Molprobity[83] webservice. Validation statistics are in tables in supplementary information.

Mass spectrometry analysis and data post-processing

Protein extracts were precipitated overnight with 5 volumes of cold 0.1 M ammonium acetate in 100% methanol. Proteins were then digested with sequencing-grade trypsin (Promega, Fitchburg, MA, USA) as described previously[84]. Each sample was further analyzed by nanoLC-MS/MS on a QExactive+ mass spectrometer coupled to an EASY-nanoLC-1000 (Thermo-Fisher Scientific, USA). Data were searched against the rabbit UniprotKB sub-database with a decoy strategy (UniprotKB release 2016-08-22, taxon 9986 *Oryctolagus cuniculus*, 23086 forward protein sequences). Peptides and proteins were identified with Mascot algorithm (version 2.5.1, Matrix Science, London, UK) and data were further imported into Proline v1.4 software (http://proline.profiproteomics.fr/). Proteins were validated on Mascot pretty rank equal to 1, and 1% FDR on both peptide spectrum matches (PSM score) and protein sets (Protein Set score). The total number of MS/MS fragmentation spectra was used to quantify each protein from three independent biological replicates (Spectral Count relative quantification). Proline was further used to align the Spectral Count values across all samples. The average of three experiments was normalized in respect to the total number of spectral counts (NSC) for all initiation factors (eIFs) (2807 for beta-globin and 2126 for H4). Therefore, the H4 results were multiplied by the normalization factor of 1.32. Then, the multiplicands for different eIFs were added and the results underwent a second normalization according to the number of trypsin sites (>70% of probability) predicted by PeptideCutter provided by the ExPaSy server (https://web.expasy.org/peptide_cutter). The heat maps were generated in respect to the NSC. The coefficients of variations were calculated as standard of deviation between the normalized NCS and are presented in

percentage (see Fig. 1b). The mass spectrometric data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier **PXDxxxx**.

Alignments

The alignments of 8 eukaryotic species (*Homo sapiens*, *Mus musculus*, *Danio rerio*, *Drosophila melanogaster*, *Ceanorhabditis elegans*, *Neurospora crassa*, *Saccharomyces cervisiae*, *Arabidopsis thaliana*) shown in the figures, were done using Constraint-based Multiple Alignment Tool (COBALT)[85] from NCBI and visualized using BoxShade Server (ExPASy).

REFERENCES

1. Hashem, Y. *et al.* Structure of the Mammalian Ribosomal 43S Preinitiation Complex Bound to the Scanning Factor DHX29. *Cell* **153,** 1108–1119 (2013).
2. Erzberger, J. P. *et al.* Molecular architecture of the 40S · eIF1 · eIF3 translation initiation complex. *Cell* **158,** 1123–1135 (2014).
3. Aylett, C. H. S., Boehringer, D., Erzberger, J. P., Schaefer, T. & Ban, N. Structure of a Yeast 40S – eIF1 – eIF1A – eIF3 – eIF3j initiation complex. *Nat. Struct. Mol. Biol.* **22,** 9–12 (2015).
4. Rogers, G. W., Richter, N. J., Lima, W. F. & Merrick, W. C. Modulation of the Helicase Activity of eIF4A by eIF4B, eIF4H, and eIF4F. *J. Biol. Chem.* **276,** 30914–30922 (2001).
5. Gross, J. D. *et al.* Ribosome Loading onto the mRNA Cap Is Driven by Conformational Coupling between eIF4G and eIF4E University of California at Berkeley. *Cell* **115,** 739–750 (2003).
6. Marintchev, A. *et al.* Topology and Regulation of the Human eIF4A/4G/4H Helicase Complex in Translation Initiation. *Cell* **136,** 447–460 (2009).
7. Jackson, R. J., Hellen, C. U. T. & Pestova, T. V. The Mechanism Of Eucaryotic

Translation Initiation And Principles Of Its Regulation. *Nat Rev Mol Cell Biol.* **11,** 113–127 (2010).

8. Hussain, T. *et al.* Structural Changes Enable Start Codon Recognition by the Eukaryotic Translation Initiation Complex. *Cell* **159,** 597–607 (2014).

9. Llácer, J. L. *et al.* Conformational Differences between Open and Closed States of the Eukaryotic Translation Initiation Article Conformational Differences between Open and Closed States of the Eukaryotic Translation Initiation Complex. *Mol. Cell* **59,** 399–412 (2015).

10. Llácer, J. L. *et al.* Translational initiation factor eIF5 replaces eIF1 on the 40S ribosomal subunit to promote start-codon recognition. *Elife* **7,** 1–33 (2018).

11. Lomakin, I. B. & Steitz, T. A. The initiation of mammalian protein synthesis and the mechanism of scanning. *Nature* **500,** 307–311 (2013).

12. Hinnebusch, A. G. Molecular Mechanism of Scanning and Start Codon Selection in Eukaryotes. *Microbiol. Mol. Biol. Rev.* **75,** 434–467 (2011).

13. Zhang, F., Saini, A. K., Shin, B. S., Nanda, J. & Hinnebusch, A. G. Conformational changes in the P site and mRNA entry channel evoked by AUG recognition in yeast translation preinitiation complexes. *Nucleic Acids Res.* **43,** 2293–2312 (2015).

14. Beznosková, P. *et al.* Translation Initiation Factors eIF3 and HCR1 Control Translation Termination and Stop Codon Read-Through in Yeast Cells. *PLoS Genet.* **9,** (2013).

15. Beznosková, P., Wagner, S., Jansen, M. E., Von Der Haar, T. & Valášek, L. S. Translation initiation factor eIF3 promotes programmed stop codon readthrough. *Nucleic Acids Res.* **43,** 5099–5111 (2015).

16. Heuer, A. *et al.* Structure of the 40S-ABCE1 post-splitting complex in ribosome recycling and translation initiation. *Nat. Struct. Mol. Biol.* **24,** 453–460 (2017).

17. Kiosze-Becker, K. *et al.* Structure of the ribosome post-recycling complex probed by chemical cross-linking and mass spectrometry. *Nat. Commun.* **7,** 1–9 (2016).

18. Mancera-Martínez, E., Brito Querido, J., Valasek, L. S., Simonetti, A. & Hashem, Y. ABCE1: A special factor that orchestrates translation at the crossroad between recycling and initiation. *RNA Biol.* **14,** 1279–1285 (2017).

19. Fringer, J. M., Acker, M. G., Fekete, C. A., Lorsch, J. R. & Dever, T. E. Coupled Release of Eukaryotic Translation Initiation Factors 5B and 1A from 80S Ribosomes

following Subunit Joining. *Mol. Cell. Biol.* **27,** 2384–2397 (2007).

20. Fernandez, I. S. *et al.* Molecular Architecture of a Eukaryotic Translational Initiation Complex. *Science (80-.).* **342,** 1240585–1240585 (2013).

21. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44,** 283–292 (1986).

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

23. Kozak, M. At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J. Mol. Biol.* **196,** 947–950 (1987).

24. Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger rNAS. *Nucleic Acids Res.* **15,** 8125–8148 (1987).

25. Kozak, M. Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. *Nature* **308,** 241–246 (1984).

26. Lin, F. T., MacDougald, O. A., Diehl, A. M. & Lane, M. D. A 30-kDa alternative translation product of the CCAAT/enhancer binding protein alpha message: transcriptional activator lacking antimitotic activity. *Proc. Natl. Acad. Sci.* **90,** 9606–9610 (1993).

27. Martin, F. *et al.* Cap-Assisted Internal Initiation of Translation of Histone H4. *Mol. Cell* **41,** 197–209 (2011).

28. Martin, F. *et al.* Ribosomal 18S rRNA base pairs with mRNA during eukaryotic translation initiation. *Nat. Commun.* **7,** 1–7 (2016).

29. Eliseev, B. *et al.* Structure of a human cap-dependent 48S translation pre-initiation complex. *Nucleic Acids Res.* **46,** 2678–2689 (2018).

30. des Georges, A. *et al.* Structure of mammalian eIF3 in the context of the 43S preinitiation complex. *Nature* **525,** 491–495 (2015).

31. Simonetti, A. *et al.* eIF3 Peripheral Subunits Rearrangement after mRNA Binding and Start-Codon Recognition. *Mol. Cell* **63,** 206–217 (2016).

32. Pisarev, A. V *et al.* Specific functional interactions of nucleotides at key − 3 and + 4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes Dev.* 624–636 (2006).

33. Pisarev, A. V., Kolupaeva, V. G., Yusupov, M. M., Hellen, C. U. T. & Pestova, T. V. Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO J.* **27,** 1609–1621 (2008).

34. Bulygin, K., Chavatte, L., Frolova, L., Karpova, G. & Favre, A. The first position of a codon placed in the A site of the human 80S ribosome contacts nucleotide C1696 of the 18S rRNA as well as proteins S2, S3, S3a, S30, and S15. *Biochemistry* **44,** 2153–2162 (2005).

35. Demeshkina, N., Repkova, M., Ven'yaminova, A., Graifer, D. & Karpova, G. Nucleotides of 18s rRNA surrounding mRNA codons at the human ribosomal A, P, and E sites: A crosslinking study with mRNA analogs carrying an aryl azide group at either the uracil or the guanine residue. *Rna* **6,** 1727–1736 (2000).

36. Luna, R. E. *et al.* The interaction between eukaryotic initiation factor 1A and eIF5 retains eIF1 within scanning preinitiation complexes. *Biochemistry* **52,** 9510–9518 (2013).

37. Luna, R. E. *et al.* The C-Terminal Domain of Eukaryotic Initiation Factor 5 Promotes Start Codon Recognition by Its Dynamic Interplay with eIF1 and eIF2β. *Cell Rep.* **1,** 689–702 (2012).

38. Dong, J. *et al.* Rps3/uS3 promotes mRNA binding at the 40S ribosome entry channel and stabilizes preinitiation complexes at start codons Jinsheng. *PNAS* **114,** E2126–E2135 (2017).

39. Grzegorski, S. J., Chiari, E. F., Robbins, A., Kish, P. E. & Kahana, A. Natural variability of Kozak sequences correlates with function in a zebrafish model. *PLoS One* **9,** (2014).

40. Thiaville, P. *et al.* Global translational impacts of the loss of the tRNA modification t6A in yeast. *Microb. Cell* **3,** 29–45 (2016).

41. Pollo-Oliveira, L. & De Crécy-Lagard, V. Can Protein Expression Be Regulated by Modulation of tRNA Modification Profiles? *Biochemistry* **58,** 355–362 (2019).

42. Dvir, S. *et al.* Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast. *Proc. Natl. Acad. Sci.* **110,** E2792–E2801 (2013).

43. Taoka, M. *et al.* Landscape of the complete RNA chemical modifications in the human 80S ribosome. *Nucleic Acids Res.* **46,** 9289–9298 (2018).

44. Maden, B. E. H. The Numerous Modified Nucleotides in Eukaryotic Ribosomal RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **39,** 241–303 (1990).

45. Dong, J. *et al.* Genetic identification of yeast 18S rRNA residues required for efficient recruitment of initiator tRNAMet and AUG selection. *Genes Dev.* **22,** 2242–2255 (2008).

46. Khoshnevis, S. *et al.* The iron-sulphur protein RNase L inhibitor functions in translation termination. *EMBO Rep.* **11,** 214–219 (2010).

47. Pisarev, A. V. *et al.* The Role of ABCE1 in Eukaryotic Posttermination Ribosomal Recycling. *Mol. Cell* **37,** 196–210 (2010).

48. Heuer, A. *et al.* Cryoelectron Microscopic Structures of Eukaryotic Translation Termination Complexes Containing eRF1-eRF3 or eRF1-ABCE1. *Cell Rep.* **8,** 59–65 (2014).

49. Young, D. J., Guydosh, N. R., Zhang, F., Hinnebusch, A. G. & Green, R. Rli1/ABCE1 Recycles Terminating Ribosomes and Controls Translation Reinitiation in 3′UTRs In Vivo. *Cell* **162,** 872–884 (2015).

50. Shoemaker, C. J. & Green, R. Kinetic analysis reveals the ordered coupling of translation termination and ribosome recycling in yeast. *Proc. Natl. Acad. Sci.* **108,** E1392–E1398 (2011).

51. Becker, T. *et al.* Structural basis of highly conserved ribosome recycling in eukaryotes and archaea. *Nature* **482,** 501–506 (2012).

52. Karcher, A., Schele, A. & Hopfner, K. X-ray Structure of the Complete ABC Enzyme ABCE1 from. *J. Biol. Chem.* **283,** 7962–7971 (2008).

53. Barthelme, D. *et al.* Ribosome recycling depends on a mechanistic link between the FeS cluster domain and a conformational switch of the twin-ATPase ABCE1. *Proc. Natl. Acad. Sci.* **108,** 3228–3233 (2011).

54. Barthelme, D. *et al.* Structural organization of essential iron-sulfur clusters in the evolutionarily highly conserved ATP-binding cassette protein ABCE1. *J. Biol. Chem.* **282,** 14598–14607 (2007).

55. Gouridis, G. *et al.* ABCE1 Controls Ribosome Recycling by an Asymmetric Dynamic Conformational Equilibrium. *Cell Rep.* **28,** 723–734 (2019).

56. Lee, A. S. ., Kranzusch, P. J., Doudna, J. A. & Cate, J. H. D. eIF3d is an mRNA cap-

binding protein required for specialized translation initiation. *Nature* **536,** 96–99 (2016).

57. Aitken, C. E. *et al.* Eukaryotic translation initiation factor 3 plays distinct roles at the mrna entry and exit channels of the ribosomal preinitiation complex. *Elife* **5,** 1–37 (2016).

58. Graifer, D. *et al.* Variable and conserved elements of human ribosomes surrounding the mRNA at the decoding and upstream sites. *Nucleic Acids Res.* **32,** 3282–3293 (2004).

59. Ferretti, M. B., Ghalei, H., Ward, E. A., Potts, E. L. & Karbstein, K. Rps26 directs mRNA-specific translation by recognition of Kozak sequence elements. *Nat Struct Mol Biol* **24,** 700–707 (2017).

60. Cavener, D. R. & Ray, S. C. Eukaryotic start and stop translation sites. *Nucleic Acids Res.* **19,** 3185–3192 (1991).

61. Zhou, J., Lancaster, L., Donohue, J. P. & Noller, H. F. Crystal structures of EF-G - Ribosome complexes trapped in intermediate states of translocation. *Science (80-.).* **340,** (2013).

62. Wallace, E. *et al.* Start codon context controls translation initiation in the fungal kingdom. Preprint at: https://www.biorxiv.org/content/10.1101/654046v1 (2019).

63. Visweswaraiah, J. & Hinnebusch, A. G. Interface between 40S exit channel protein uS7/Rps5 and eIF2α modulates start codon recognition in vivo. *Elife* **6,** 1–22 (2017).

64. Wimberly, B. T., White, S. W. & Ramakrishnan, V. The structure of ribosomal protein S7 at 1.9 Å resolution reveals a β-hairpin motif that binds double-stranded nucleic acids. *Structure* **5,** 1187–1198 (1997).

65. Visweswaraiah, J., Pittman, Y., Dever, T. E. & Hinnebusch, A. G. The β-hairpin of 40S exit channel protein rps5/uS7 promotes efficient and accurate translation initiation in vivo. *Elife* **4,** 21 (2015).

66. Weisser, M. *et al.* Structural and Functional Insights into Human Re-initiation Complexes. *Mol. Cell* **67,** 447–456.e7 (2017).

67. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 12–21 (2010).

68. Mastronarde, D. N. SerialEM: A program for automated tilt series acquisition on Tecnai microscopes using prediction of specimen position. Microscopy and Microanalysis Vol. 9, Suppl. 2, 1182 (2003).
69. de la Rosa-Trevín, J. M. et al. Scipion: A software framework toward integration, reproducibility and validation in 3D electron microscopy. J. Struct. Biol. 195, 93–99 (2016).
70. Zheng, S. Q., et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo- electron microscopy. Nat Methods. 14, 331-332.
71. Rohou, A. & Grigorieff, N. 2015. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol 192, 216-21 (2017).
72. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. J Struct Biol 180, 519-30 (2012).
73. Pettersen, E. F. et al. UCSF Chimera--a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612 (2004).
74. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. J. Mol. Biol. 215, 403–410 (1990).
75. NCBI Resource Coordinators. Database Resources of the National Center for Biotechnology Information. Nucleic Acids Res. 45, D12–D17 (2017).
76. Berman, H. M. et al. The Protein Data Bank. Nucleic Acids Res. 28, 235–242 (2000).
77. Biasini, M. et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42, W252–W258 (2014).
78. Trabuco, L. G., Villa, E., Mitra, K., Frank, J. & Schulten, K. Flexible Fitting of Atomic Structures into Electron Microscopy Maps Using Molecular Dynamics. Structure 16, 673–683 (2008).
79. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. J. Mol. Graph. 14, 33–38 (1996).
80. Phillips, J. C. et al. Scalable molecular dynamics with NAMD. J. Comput. Chem. 26, 1781–1802 (2005).
81. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221 (2010).

82. Chou, F.-C., Echols, N., Terwilliger, T. C. & Das, R. RNA Structure Refinement Using the ERRASER-Phenix Pipeline. in Nucleic Acid Crystallography 269–282 (Humana Press, New York, NY, (2016).

83. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. 66, 12–21 (2010).

84. Khusainov, I. et al. Structure of the 70S ribosome from human pathogen Staphylococcus aureus. Nucleic Acids Res. 44, 10491-10504 (2016).

85. Papadopoulos, J.S. and Agarwala R. COBALT: constraint-based alignment tool for multiple protein sequences. Bioinformatics, 23(9), 1073-179 (2007).

FIGURE LEGEND

Fig. 1: Overall structure of β-globin and H4 late-stage 48S initiation complexes (ICs). a, mRNA sequences used to form and purify the β-globin and the H4 ICs. Only the sequences near the AUG codon are represented and main differences in the Kozak sequence are indicated in bold. **b,** semi-quantitative mass spectrometry analysis of the eukaryotic initiation factors (eIFs) in both ICs, indicating the abundance of each eIF based on the spectra count normalized. The two rounds of normalization were carried out using the total number of eIFs and estimated number of trypsin cleavage sites (see Methods). The normalized spectra count (NCS) are presented as heat maps with cold colours indicating low abundance and warm colours indicating high abundance. The black points out a high number of NCS for eIF2Bδ which is caused by the detection of three different isoforms of this protein. Small stars indicate the values of the coefficients of variation calculated for each NSC. **c-d,** segmented cryo-EM reconstructions of the main two classes (with and without eIF3) sorted out from the β-globin IC seen from (**c**) solvent, beak and (**d**) platform sides, respectively. The reconstruction shows 40S (in yellow), eIF2γ (in orange), eIF2α (in purple), $tRNA_i^{Met}$ (in magenta), mRNA (in red), eIF1A (in skyblue) and ABCE1 (in green) **e-f,** same as c-d but for the H4 IC. Boxed blowups represent the codon/anti-codon duplexes in all shown reconstructions with their respective atomic models fitted in the corresponding electron densities.

Fig. 2: Key interactions surrounding the start-codon recognition sites in β-globin and H4 late-stage 48S ICs. a, ribbon representation of the atomic model of β-globin LS48S IC viewed from the intersubunit side. **b**, codon:anticodon base-pairing view in both mRNA complexes; left: β-globin, right: H4. **c**, eIF1A (in skyblue) interaction with the mRNA in the β-globin IC (left panel), compared to the corresponding region in the H4 IC, which is mostly free of eIF1A (right panel). **d,** close-up on the eIF1A N-terminal tail (coloured in cyan) showing its intricate interactions with tRNA and mRNA; stacking of C1696 on tip of $tRNA_i^{Met}$. The nucleotides involved in the interactions are coloured in chartreuse. **e**, interaction network of the $tRNA_i$ with ribosomal proteins uS13 and uS19 (coloured in salmon). Residues involved in the interactions are coloured in cyan in uS13 and uS19 and in green in the $tRNA_i$. For eIF1A, uS13 and uS19, sequence alignments of the concerned interacting regions from eight representative eukaryotic species are shown below the panels in black boxes and the described residues are indicated by coloured frames (Hs : *Homo sapiens*, Mm : *Mus musculus*, Dr : *Danio rerio*, Dm : *Drosophila melanogaster*, Ce : *Ceanorhabditis elegans*, Nc : *Neurospora crassa*, Sc: *Saccharomyces cervisiae*, At: *Arabidopsis thaliana*).

Fig. 3: Kozak and beyond Kozak interaction networks in β-globin and H4 LS48S ICs. a, close-up of the interactions of upstream start-codon nucleotides top-viewed from the head side in the β-globin (left panel) and H4 (right panel) ICs with ribosomal proteins eS26 and uS11, as well as eIF2α D1 domain, $tRNA_i$ and 18S rRNA. mRNA (-4) position contact with His80 of eS26 is highlighted in dashed line circle. G1203 and G957 of 18S rRNA stacking and interaction with C(-1) and A(-3), respectively, of both mRNAs are shown. **b**, mRNA entry channel seen from the beak side with close-up on the interactions with uS3, eS30 and h16 of 18S rRNA. **c**, mRNA exit channel seen from the solvent side with close-up on the mRNA contacts with ribosomal proteins eS26 and eS28. The nucleotides involved in the interactions are indicated in chartreuse and residues in cyan. Respective sequence alignments are shown in black boxes from eight representative eukaryotic species on the right of the figure panels.

Fig. 4: eIF1A interactions in the A-site of β-globin LS48S IC. a, N-terminal tail interactions with mRNA of downstream start-codon nucleotides and $tRNA_i$. **b**, eIF1A OB-domain interactions with mRNA and 40S. **c**, close-up on interactions of eIF1A with h44 of 18S rRNA (green frame). **d**, zoom in on eIF1A interactions with h18 of 18S rRNA and

ribosomal proteins uS12 and eS30 (blue frame). The nucleotides involved in the interactions are indicated in chartreuse and residues in cyan. Respective sequence alignments are shown in black boxes.

Fig. 5: Initiator tRNA anticodon stem-loop (ASL) interactions with the LS48S IC. a, comparison of the $tRNA_i$ modified t^6A_{37} interaction with mRNA (-1) position in mammalian β-globin (in light pink) and H4 (in red) and yeast[10] (dark purple) initiation complexes. The contact with (-1) mRNA position is labelled by black solid line, black dots and grey solid line for β-globin, H4 and yeast Kozak optimal sequence, respectively. In dashed skyblue circle, comparison of the conformation of highly conserved A(-3) mRNA position: *anti* in mammalian IC and *syn* in yeast. **b**, interaction of the modified $m^1acp^3\Psi$1248 of the 18S rRNA (coloured in chartreuse, overlapping bottom panel) with the C(34) of the $tRNA_i$ and ribosomal protein uS9 (overlapping top panel). **c**, close-up on the interactions of C(32) and C(33) with Arg146 of uS9. **d**, close-up on the interaction of the ASL cysteines from the conserved G-C base-pairs with eIF2α domain D1. Respective sequence alignments are shown in black boxes and interacting residues in coloured frames.

Fig. 6: Atomic model of ABCE1 in the LS48S IC. a and **e**, ribbon representations of ABCE1 (in green) in its electron density with NTP-binding pockets framed in blue (NBD1) and pink (NBD2) seen from different side views. **b**, blowups on the NTP pockets, NBD1 (left blue) and NBD2 (right pink). Although in the purification conditions GMP-PNP was used, ATP molecules were modelled in the electron densities obtained. **c**, mixed ribbon and stick representation of Fe-S binding domain atomic model fitted into its electron density. **d**, close-up of Fe-S binding domain interactions with 40S ribosomal protein uS12 and h44 of 18S rRNA. **f**, close-up on NBD1 interactions with the 40S. **g**, close-up on NBD2 interactions with the 40S. The nucleotides involved in the interactions are indicated in chartreuse and protein residues in cyan. Respective sequence alignments are shown in black boxes. **h**, comparison between the β-globin•GMP-PNP (pink surface) and β-globin•GMP-PNP+ATP (grey surface) LS48S ICs reconstructions without (left superimposition) and with (right superimposition) eIF3. The panel shows that the addition of ATP triggers the dissociation of ABCE1, probably after its hydrolysis. No further conformational changes between both complexes, with and without ATP, can be detected.

Fig. 7: Cryo-EM reconstruction of the eIF3 containing class of the β-globin LS48S IC. a, segmented map showing electron density of the eIF3 core (in rose) attached to the 48S viewed from the platform side. **b**, codon: anticodon base-pairing view in LS48S-eIF3 ICs (identical to both β-globin and H4). **c**, ribbon representation of the atomic model of LS48S-eIF3 IC seen from the platform side. **d**, blowup on the mRNA channel exit, seen from the platform side. mRNA 5'UTR cannot be modelled because of the low local-resolution of the cryo-EM reconstruction in this region, therefore we propose an extrapolation of the mRNA 5'UTR trajectory showing possible interactions with eIF3 d and a subunits (residues coloured in cyan). **e**, close-up on the eIF3c (in navy blue) interaction with 40S: h26 of 18S rRNA and ribosomal protein eS27. **f**, close-up on the eIF3a (in coral) interaction with 40S ribosomal protein eS1. **g**, mixed ribbons and sticks representation of eIF3 core interactions with 40S. The nucleotides involved in the interactions are indicated in chartreuse and protein residues in cyan. Respective sequence alignments are shown in black boxes with interacting residues highlighted by coloured frames. **h**, summary of the mRNA interactions with mammalian LS48S IC. The ribosomal proteins are coloured in orange and 18S rRNA elements in yellow. The mRNA contacts critical for recognition of optimal or poor Kozak context are highlighted by grey frames.

Figure 1

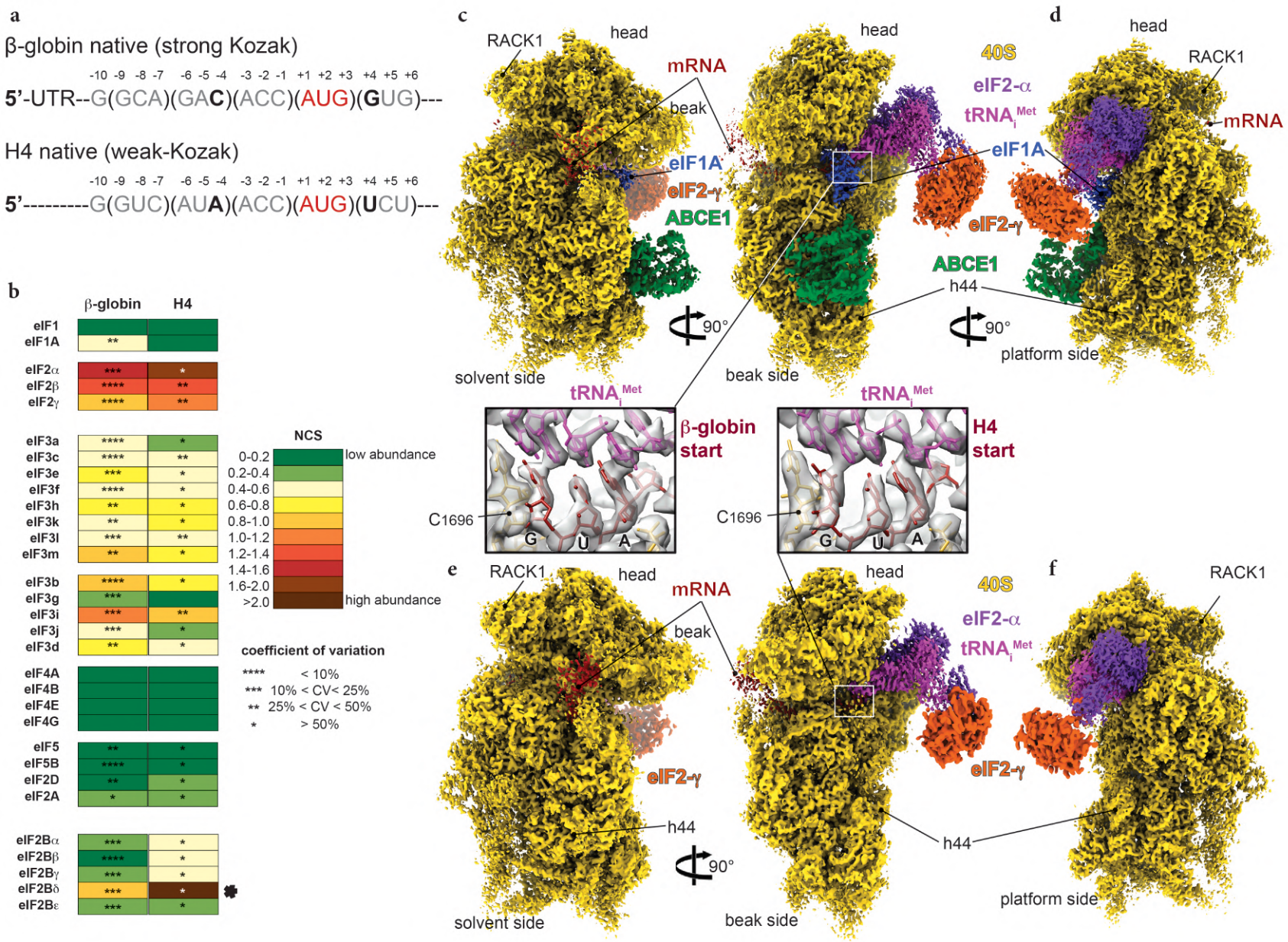
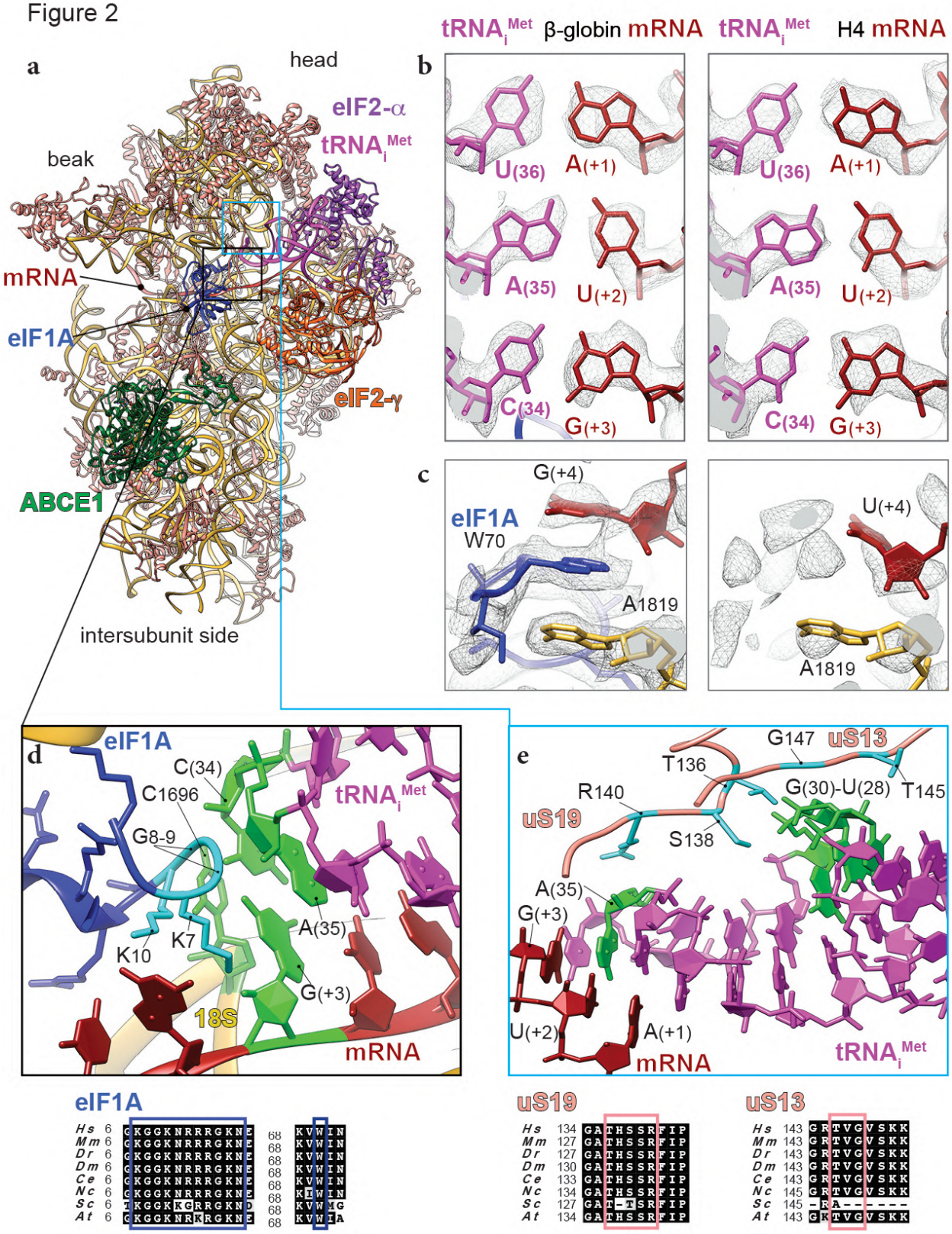

a
β-globin native (strong Kozak)
-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6
5'-UTR--G(GCA)(GAC)(ACC)(AUG)(GUG)---
H4 native (weak-Kozak)
-10 -9 -8 -7 -6 -5 -4 -3 -2 -1 +1 +2 +3 +4 +5 +6
5'---------G(GUC)(AUA)(ACC)(AUG)(UCU)---
b
β-globin
H4
eIF1
eIF1A
eIF2α
eIF2β
eIF2γ
eIF3a
eIF3c
eIF3e
eIF3f
eIF3h
eIF3k
eIF3l
eIF3m
eIF3b
eIF3g
eIF3i
eIF3j
eIF3d
eIF4A
eIF4B
eIF4E
eIF4G
eIF5
eIF5B
eIF2D
eIF2A
eIF2Bα
eIF2Bβ
eIF2Bγ
eIF2Bδ
eIF2Bε
NCS
0-0.2 low abundance
0.2-0.4
0.4-0.6
0.6-0.8
0.8-1.0
1.0-1.2
1.2-1.4
1.4-1.6
1.6-2.0
>2.0 high abundance
coefficient of variation
**** < 10%
*** 10% < CV< 25%
** 25% < CV < 50%
* > 50%
c
RACK1
head
mRNA
beak
eIF1A
eIF2-γ
ABCE1
90°
solvent side
beak side
tRNAi Met
β-globin start
C1696
G U A
d
head
RACK1
40S
eIF2-α
tRNAi Met
mRNA
eIF1A
eIF2-γ
ABCE1
h44
90°
platform side
H4 start
e
f

Figure 2


a
head
eIF2-α
tRNAi Met
beak
mRNA
eIF1A
eIF2-γ
ABCE1
intersubunit side
b
tRNAi Met
β-globin mRNA
tRNAi Met
H4 mRNA
U(36)
A(+1)
A(35)
U(+2)
C(34)
G(+3)
c
G(+4)
eIF1A
W70
A1819
U(+4)
d
eIF1A
C(34)
C1696
tRNAi Met
G8-9
K7
K10
A(35)
G(+3)
18S
mRNA
e
G147
uS13
T136
R140
G(30)-U(28)
T145
uS19
S138
A(35)
G(+3)
U(+2)
A(+1)
mRNA
tRNAi Met

eIF1A
uS19
uS13

Figure 3

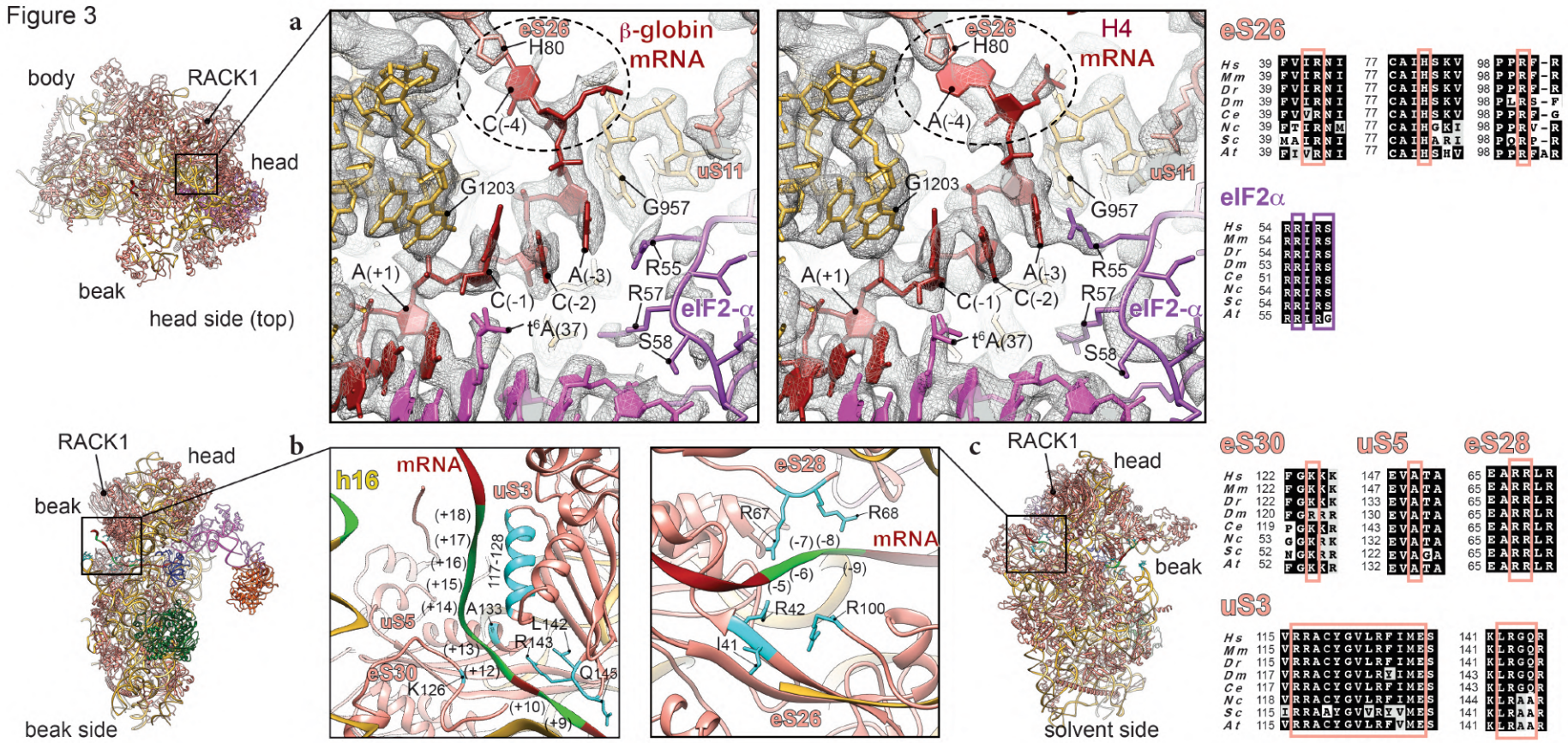

a
body
RACK1
head
beak
head side (top)
eS26
H80
β-globin mRNA
C(-4)
uS11
G1203
G957
A(+1)
A(-3)
R55
C(-1)
C(-2)
R57
eIF2-α
t6A(37)
S58
H4 mRNA
A(-4)
b
RACK1
head
beak
beak side
h16
mRNA
uS3
(+18)
(+17)
(+16)
(+15)
(+14)
117-128
A133
uS5
L142
R143
(+13)
(+12)
eS30
K126
Q145
(+10)
(+9)
eS28
R67
R68
(-7)(-8)
mRNA
(-6)
(-9)
(-5)
R42
R100
I41
eS26
c
RACK1
head
beak
solvent side
eS26
eIF2α
eS30
uS5
eS28
uS3

Figure 4

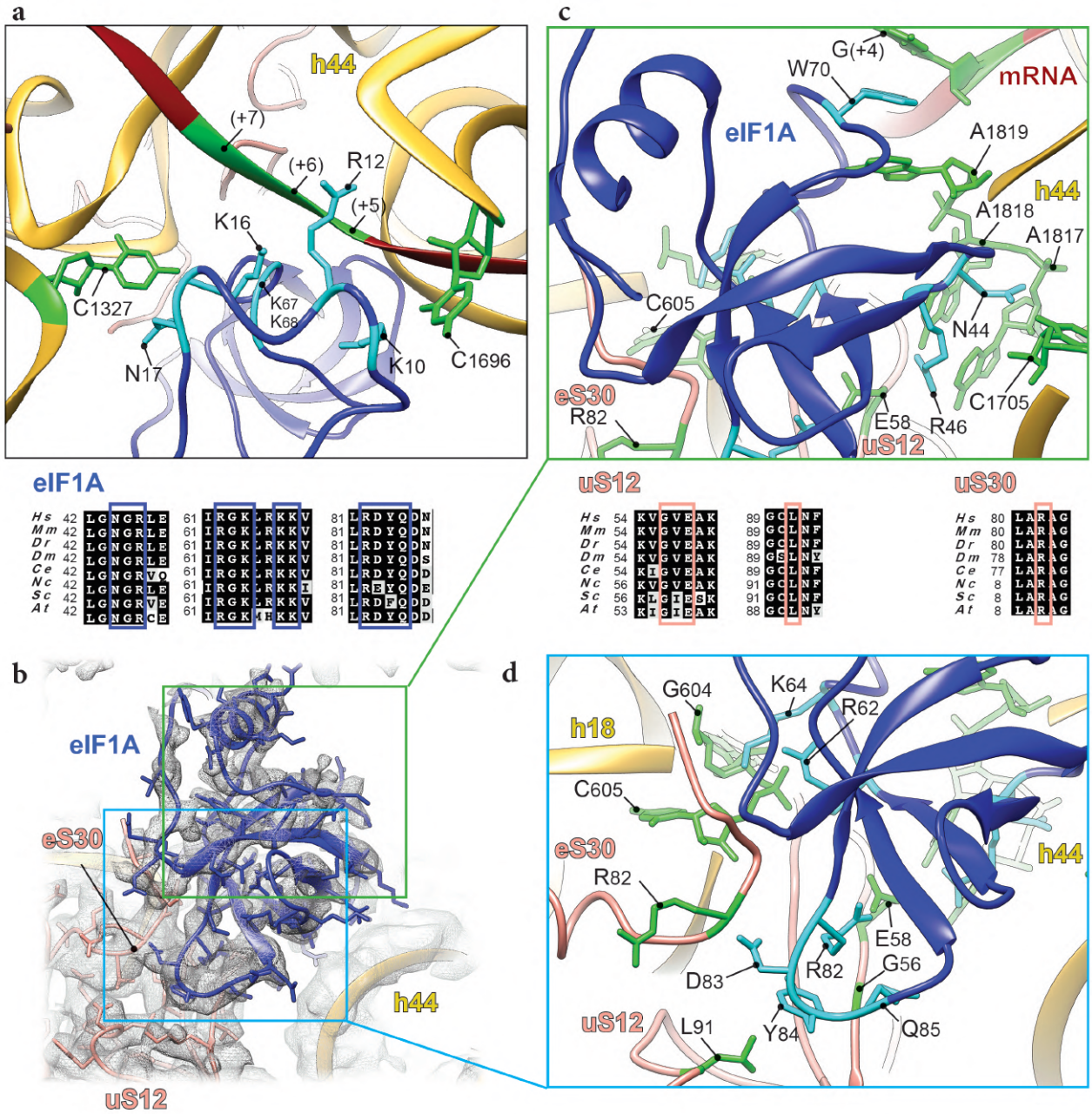

a
h44
(+7)
(+6)
R12
(+5)
K16
C1327
K67
K68
N17
K10
C1696
eIF1A
Hs 42 LGNGRLE 61 IRGKLRKKV 81 LRDYQDN
Mm 42 LGNGRLE 61 IRGKLRKKV 81 LRDYQDN
Dr 42 LGNGRLE 61 IRGKLRKKV 81 LRDYQDN
Dm 42 LGNGRLE 61 IRGKLRKKV 81 LRDYQDS
Ce 42 LGNGRVQ 61 IRGKLRKKV 81 LRDYQDD
Nc 42 LGNGRLE 61 IRGKLRKKI 81 LREYQDE
Sc 42 LGNGRVE 61 IRGKLRKKI 81 LRDFQDD
At 42 LGNGRCE 61 IRGKMFKKV 81 LRDYQDD
c
G(+4)
W70
mRNA
eIF1A
A1819
h44
A1818
A1817
C605
N44
C1705
eS30
R82
E58
R46
uS12
uS12
Hs 54 KVGVEAK 89 GGLNF
Mm 54 KVGVEAK 89 GGLNF
Dr 54 KVGVEAK 89 GGLNF
Dm 54 KVGVEAK 89 GSLNY
Ce 54 KIGVEAK 89 GGLNF
Nc 56 KVGVEAK 91 GGLNF
Sc 56 KLGIESK 91 GGLNF
At 53 KIGIEAK 88 GGLNY
uS30
Hs 80 LARAG
Mm 80 LARAG
Dr 80 LARAG
Dm 78 LARAG
Ce 77 LARAG
Nc 8 LARAG
Sc 8 LARAG
At 8 LARAG
b
eIF1A
eS30
h44
uS12
d
G604
K64
R62
h18
C605
eS30
h44
R82
E58
D83
R82
G56
uS12
L91
Y84
Q85

Figure 5

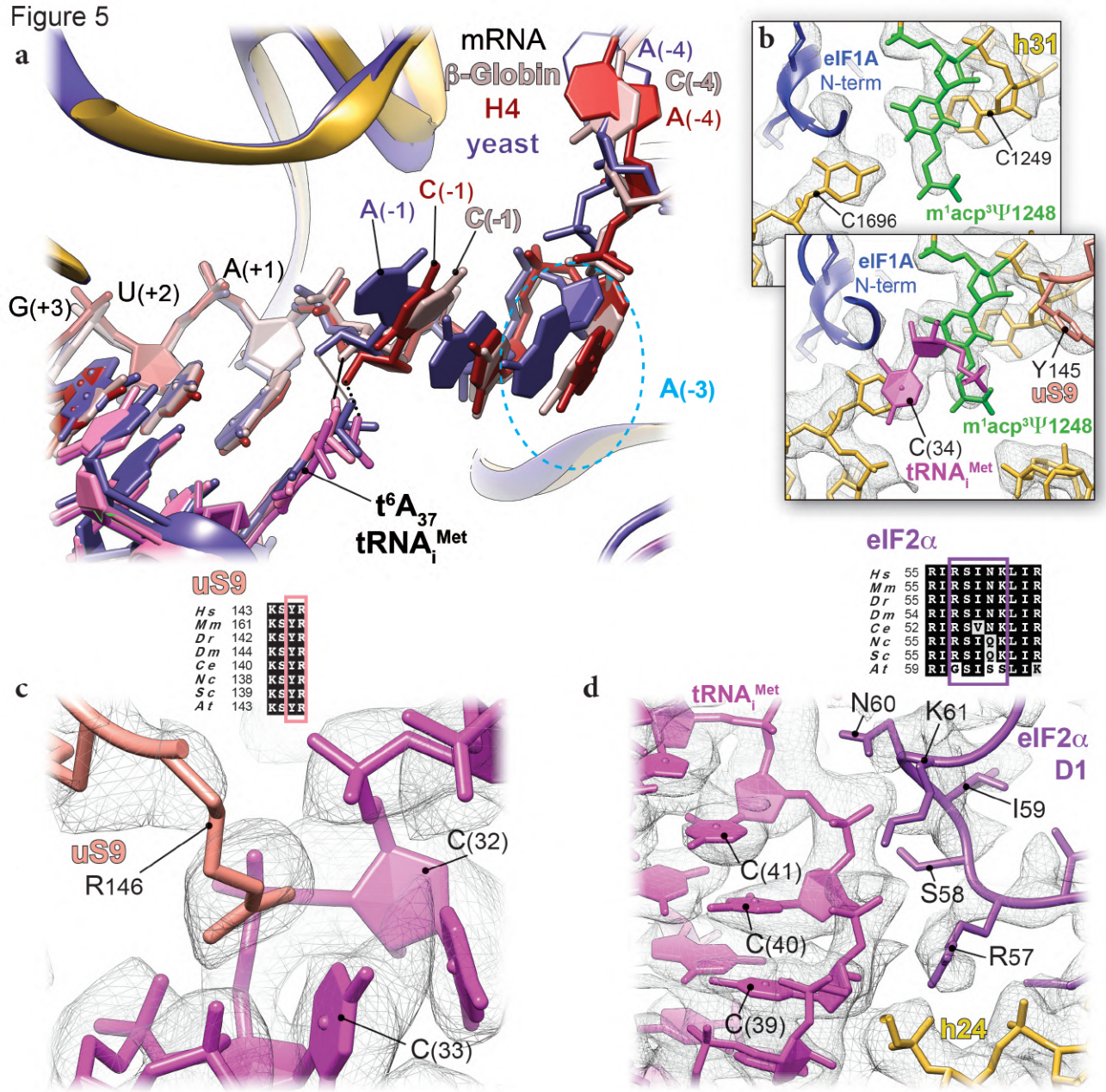

a
mRNA
β-Globin
H4
yeast
A(-4)
C(-4)
A(-4)
C(-1)
A(-1)
C(-1)
A(+1)
U(+2)
G(+3)
A(-3)
t6A37
tRNAiMet
b
eIF1A
N-term
h31
C1249
C1696
m1acp3Ψ1248
eIF1A
N-term
Y145
uS9
m1acp3Ψ1248
C(34)
tRNAiMet
uS9
Hs 143 KSYR
Mm 161 KSYR
Dr 142 KSYR
Dm 144 KSYR
Ce 140 KSYR
Nc 138 KSYR
Sc 139 KSYR
At 143 KSYR
eIF2α
Hs 55 RIRSINKLIR
Mm 55 RIRSINKLIR
Dr 55 RIRSINKLIR
Dm 54 RIRSINKLIR
Ce 52 RIRSVNKLIR
Nc 55 RIRSIQKLIR
Sc 55 RIRSIQKLIR
At 59 RIGSISSLIK
c
uS9
R146
C(32)
C(33)
d
tRNAiMet
N60
K61
eIF2α
D1
I59
C(41)
S58
C(40)
R57
C(39)
h24

Figure 6

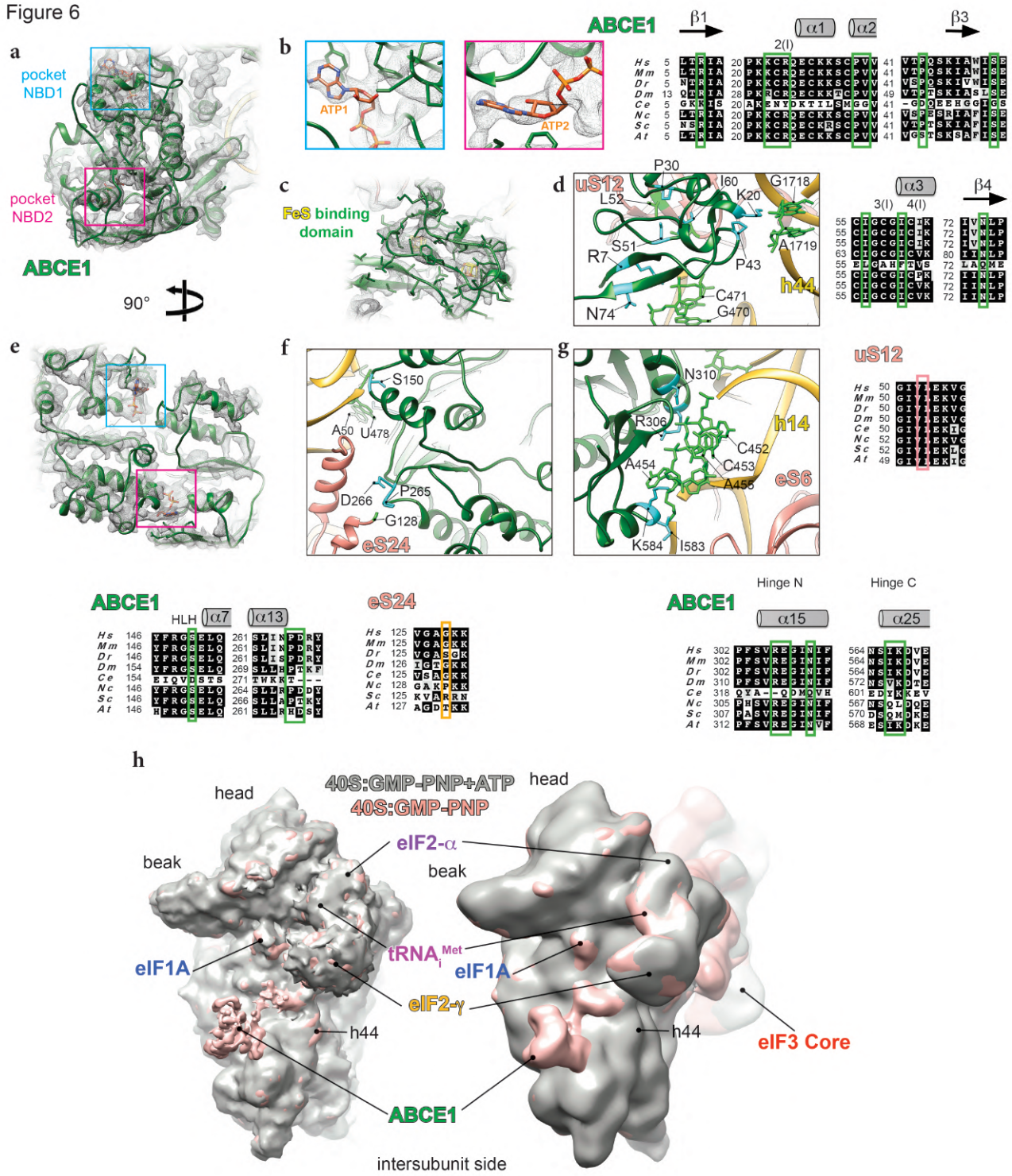
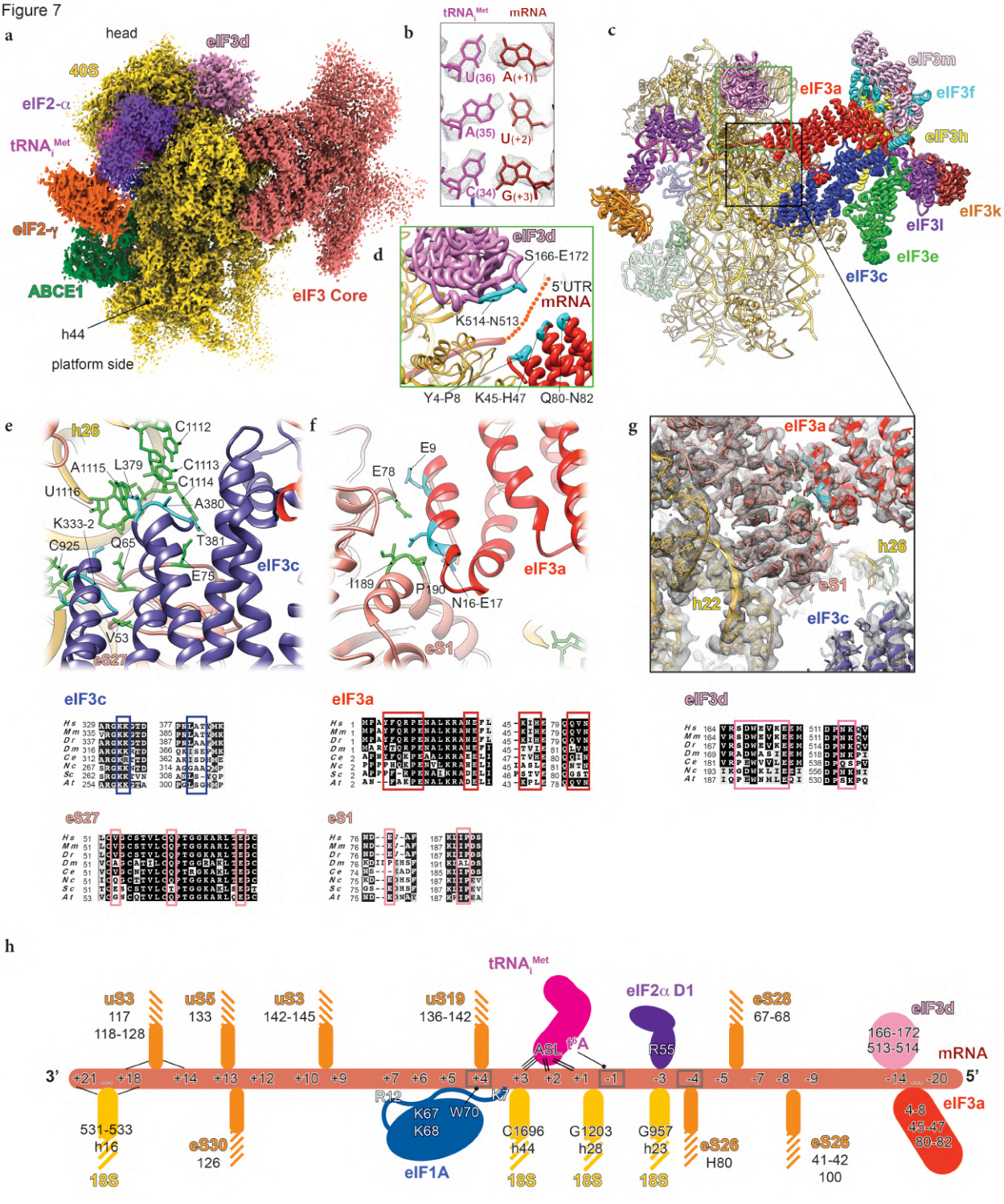

a
pocket NBD1
pocket NBD2
ABCE1
90°
b
ATP1
ATP2
ABCE1
β1
α1
α2
β3
2(I)
c
FeS binding domain
d
uS12
P30
I60
G1718
L52
K20
S51
R7
P43
A1719
h44
C471
G470
N74
α3
β4
3(I)
4(I)
e
f
S150
A50
U478
D266
P265
G128
eS24
g
N310
R306
h14
C452
A454
C453
A455
eS6
K584
I583
uS12
ABCE1
HLH
α7
α13
eS24
ABCE1
Hinge N
Hinge C
α15
α25
h
40S:GMP-PNP+ATP
40S:GMP-PNP
head
head
beak
beak
eIF2-α
tRNA$_i^{Met}$
eIF1A
eIF1A
eIF2-γ
h44
h44
eIF3 Core
ABCE1
intersubunit side

Figure 7


a
head
eIF3d
40S
eIF2-α
tRNAi Met
eIF2-γ
ABCE1
eIF3 Core
h44
platform side
b
tRNAi Met
mRNA
U(36)
A(+1)
A(35)
U(+2)
C(34)
G(+3)
c
eIF3m
eIF3a
eIF3f
eIF3h
eIF3k
eIF3l
eIF3e
eIF3c
d
eIF3d
S166-E172
5'UTR
mRNA
K514-N513
Y4-P8
K45-H47
Q80-N82
e
h26
C1112
A1115
L379
C1113
C1114
U1116
A380
K333-2
T381
C925
Q65
eIF3c
E75
V53
eS27
f
E9
E78
eIF3a
I189
P190
N16-E17
eS1
g
eIF3a
h26
eS1
h22
eIF3c
eIF3c
eIF3a
eIF3d
eS27
eS1
h
tRNAi Met
uS3
117
118-128
uS5
133
uS3
142-145
uS19
136-142
eIF2α D1
eS28
67-68
eIF3d
166-172
513-514
mRNA
ASL
i6A
R55
3'
+21
+18
+14
+13
+12
+10
+9
+7
+6
+5
+4
+3
+2
+1
-1
-3
-4
-5
-7
-8
-9
-14
-20
5'
R12
K7
K67
W70
K68
eIF3a
4-8
45-47
80-82
531-533
h16
eS30
126
eIF1A
C1696
h44
G1203
h28
G957
h23
eS26
H80
eS26
41-42
100
18S
18S
18S
18S

SUPPLEMENTARY FIGURE LEGEND

Supplementary Fig. 1: Particle classification outputs, average and local resolutions of the different β-globin and H4 LS48S ICs cryo-EM reconstructions. a to **f**, local resolution of the β-globin late-stage IC reconstruction representing a class lacking eIF3 seen from the intersubunit side (**a**), beak side (**b**), solvent side (**c**), platform side (**f**). **d** to **e**, insets showing local resolution of eIF1A and ABCE1 (yellow frame) and of mRNA and $tRNA_i^{Met}$ anticodon stem-loop (ASL) (green frame). **g**, average resolution of the reconstruction of the class without eIF3. **h**, particle classification output of the β-globin LS48S IC. **i** to **n**, local resolution of the β-globin LS48S IC reconstruction representing a class showing eIF3 octamer seen from intersubunit side (**i**), beak side (**j**), solvent side (**k**), platform side (**n**). **l** to **m**, insets showing local resolution for eIF1A and ABCE1 (yellow frame) and for mRNA and $tRNA_i^{Met}$ ASL (green frame). **o**, average resolution of the reconstruction of the class of β-globin LS48S IC with eIF3. **p** to **u**, local resolution of the H4 LS48S IC reconstruction representing a class lacking eIF3 seen from the intersubunit side (**p**), beak side (**q**), solvent side (**r**), platform side (**u**). The local resolution of eIF3 is not shown, as its average resolution is relatively low. **s** to **t**, insets showing local resolution of eIF1A and ABCE1 (yellow frame) and of mRNA and $tRNA_i^{Met}$ ASL (green frame). **v**, average resolution of the reconstruction of the class. **w**, particle classification output of the H4 LS48S IC. **x** to **y**, average resolutions of the reconstruction of β-globin+ATP LS48S IC reconstructions after their particle sorting (**x**). Similarly to the counterpart complex with ATP, two major classes stand out, with and without eIF3 (**y**).

Supplementary Fig. 2: Structural differences between β-globin and H4 LS48S ICs. a, residual density for eIF1A in H4 LS48S IC in the A-site, highlighted by dashed blue oval. **b**, comparison of mRNA channel exit between β-globin (orange surface) and H4 (grey surface) 40S reconstructions. **c**, comparison between β-globin (gold ribbons) and H4 (grey ribbons) ICs 40S atomic models. Blowups highlight the conformational changes between both types of complexes at the mRNA channel entrance, the beak and the A, P and E –sites. Dashed coloured ovals indicate the conformational changes in several ribosomal proteins (uS3, eS30 and uS7).

Supplementary Fig. 3. Differences between mammalian and yeast mRNA trajectories in the mRNA channel and the $tRNA_i^{Met}$ interactions. a, superimposition of yeast optimal mRNA Kozak consensus sequence in py48S-eIF5 IC structure[10] and mammalian mRNAs showing a smoother P/E kink in the case of β-globin and H4 ICs, indicated by arrows. **b**, the cloverleaf representation of the $tRNA_i^{Met}$ summarizing the interactions within the LS48S IC described in the Results section. The interaction of $t^6A(37)$ modification with (-1) mRNA position is highlighted by dashed-line oval.

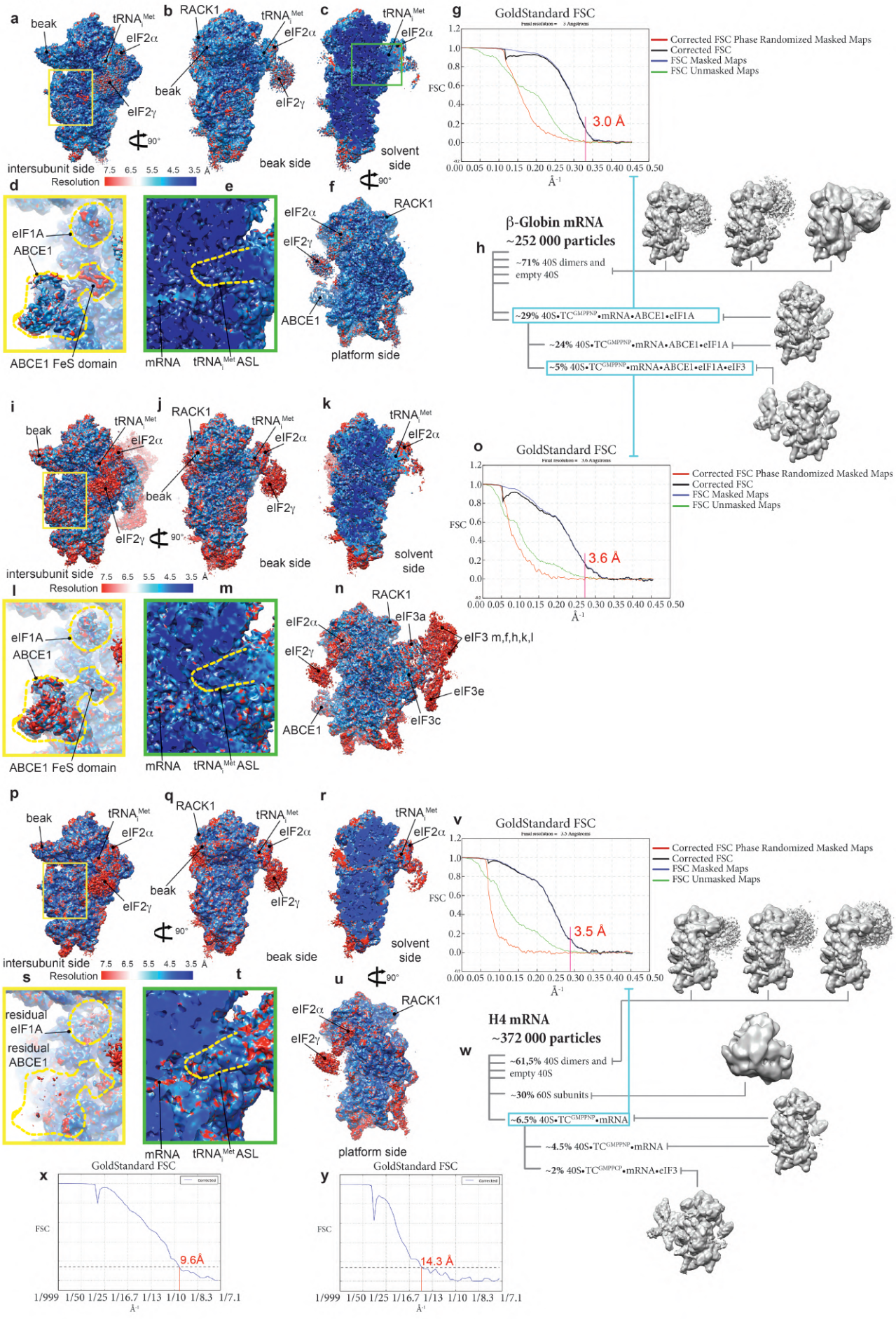
Supplementary Fig. 4: ABCE1 sequence comparison between eukaryotes and a sterical incompatibility with eIF2D. **a**, alignment of ABCE1 among eight representative eukaryotic species with secondary structure elements labelled, helices α and β-sheets (according to reference 52), and functional domains, indicated by bottom coloured lines. The residues involved in the interactions described in Results section are framed in green. In brown frame, the sequence of Fe-S cluster that showed similarity to the N-terminal part of eIF2D SUI domain (~40% sequence identity). The coordinated cysteines of cluster (I) (*) and (II) (**) are labelled in yellow. In orange frames, the residues responsible for proper ATP/GTP binding in nucleotide binding domains (NBD1 and NBD2) pockets. The accession numbers used for the alignment: *Hs*: NP_002931.2, *Mm*: NP_056566.2, *Dr*: NP_998216.2, *Dm*: NP_648272.1 (pixie), *Ce*: NP_506192.1 ABC transporter class F, *Nc*: XP_963869.3, *Sc*: AJV19484.1, *At*: OAP04903.1). **b**, superimposition of ABCE1 in β-globin LS48S IC (in green) with eIF2D (PDB ID: 5OA3)[55], showing the sterical clashing of Fe-S cluster (I) with winged-helix (WH) domain of eIF2D (in grey).

Supplementary Fig. 5: Atomic model of uS7 in mammalian LS48S IC. a, ribbon representation of uS7 in its electron density, showing the lack of density only in the mRNA-contacting β-hairpin, indicating the flexibility of this part of the protein. **b**, sequence conservation of the interacting residues in eIF2α and uS7, showed by reference 63, among eukaryotic species showed in coloured frames.

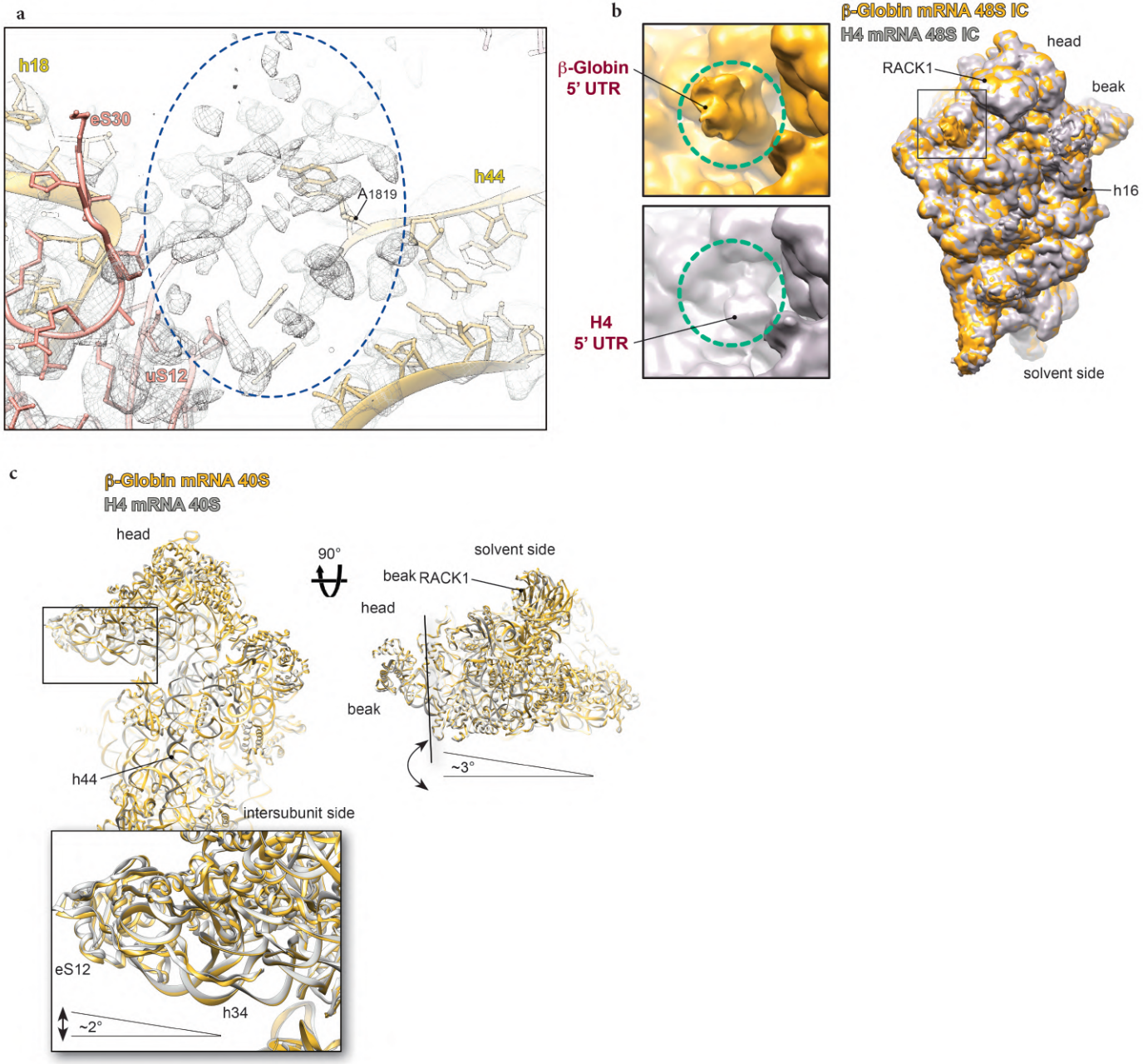
Supplementary Fig. 6: Comparison between eIF3 of ICs from near-native conditions and *in vitro* assembly. a, atomic model of eIF3 octamer core along with eIF3d subunit (in coloured ribbons). **b**, segmented cryo-EM densities of eIF3 octamer core along with eIF3d

(filtered to 8 Å), showing unidentified density (green surface) in contact with eIF3 a and c subunits. **c**, fitting of eIF3d partial crystal structure into its cryo-EM density.

Supplementary Figure 1

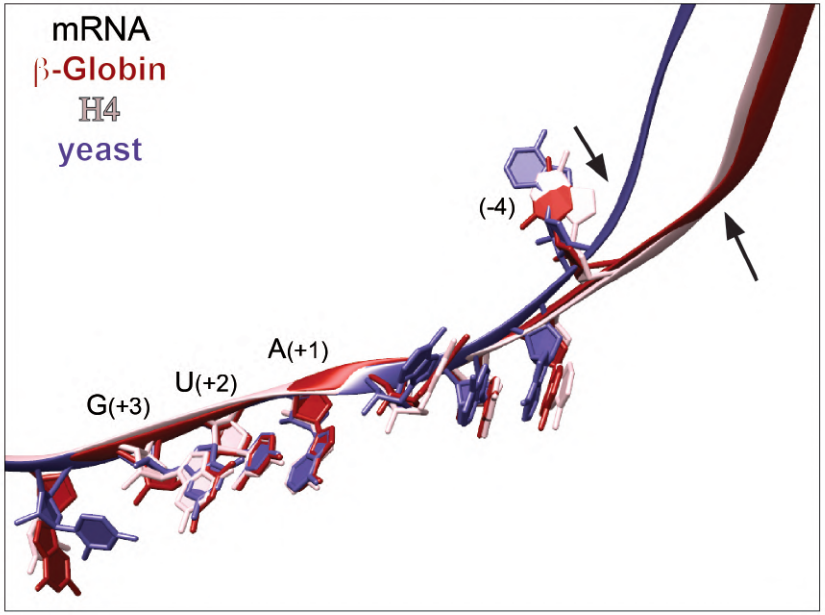


Supplementary Figure 2

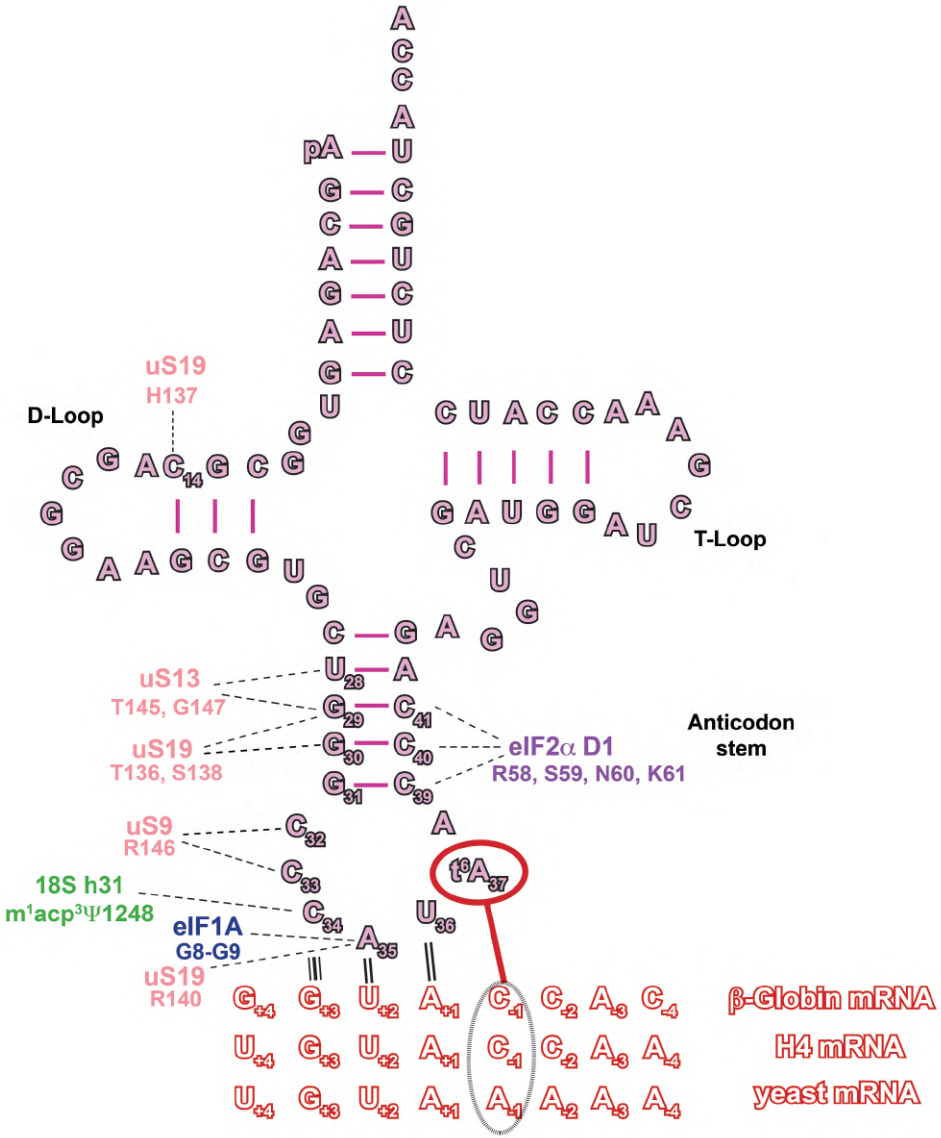


Supplementary Figure 3

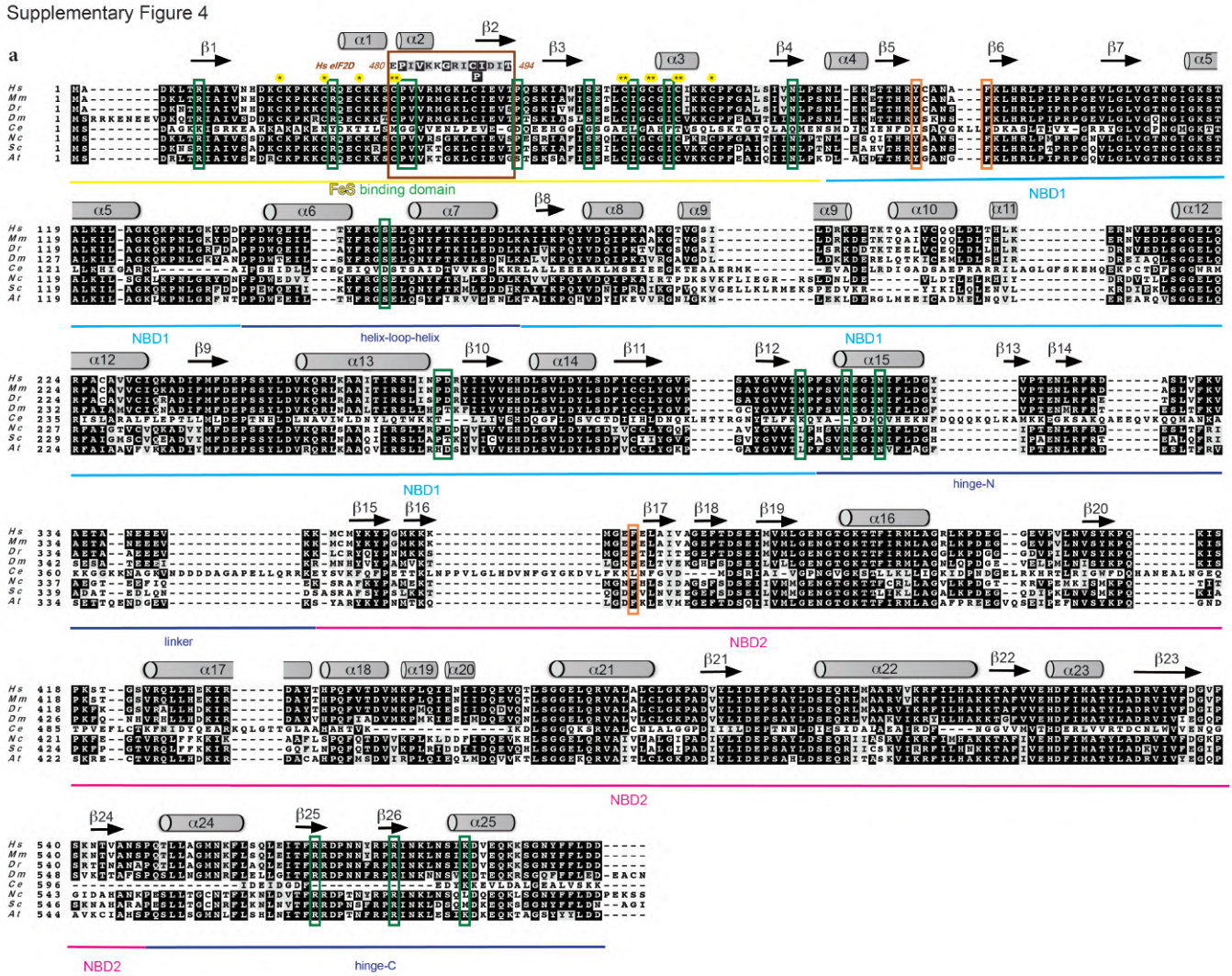
a



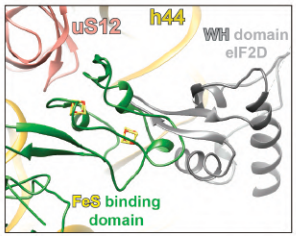
b



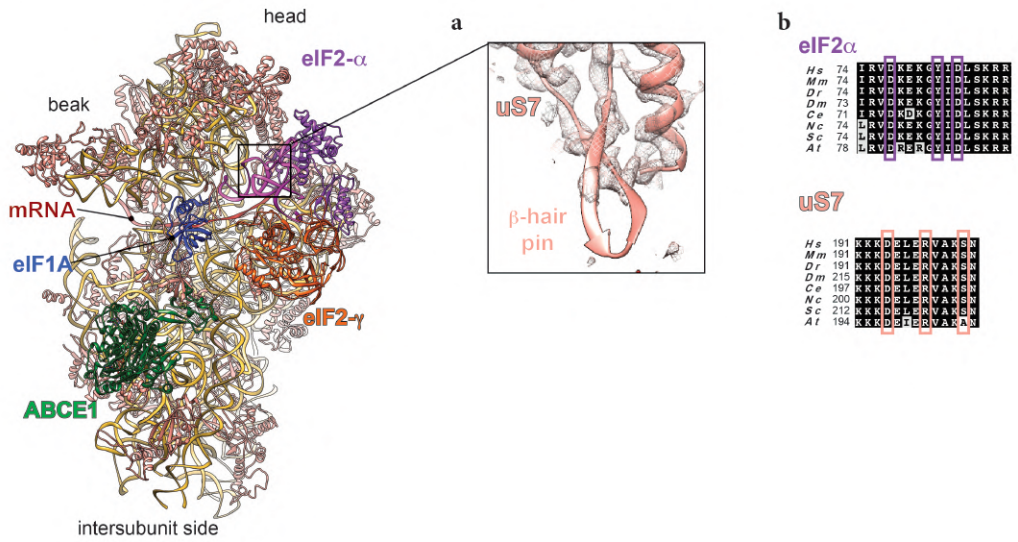
Supplementary Figure 4



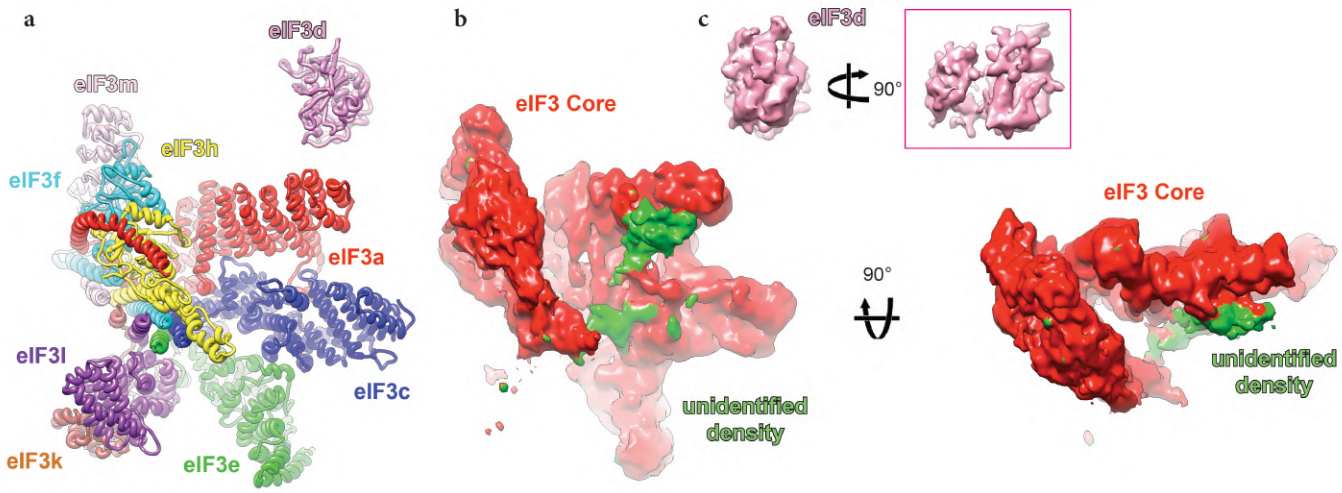
b



Supplementary Figure 5



Supplementary Figure 6



SUPPLEMENTARY TABLES

Table S1. Refinement and validation statistics for the four LS48S IC structures from O.cuniculus.

Validation[b]	48S+eIF3-β-globin	48S-NOeIF3-β-globin	48S+eIF3-H4	48S-NOeIF3-H4
Clashscore[c]	4.96 (94th percentile)	4.81 (94th percentile)	0.66 (99th percentile)	1.15 (99th percentile)
Proteins				
Molprobity score	1.91 (80th percentile)	1.99 (76th percentile)	1.50 (99th percentile)	1.60 (93rd percentile)
Favored rotamers	94.45%	93.38%	92.54%	93.38%
Allowed rotamers	4.40%	5.12%	5.64%	4.96%
Poor rotamers	1.15%	1.50%	1.82%	1.66%
Ramachandran favored	87.74%	87.60%	88.62%	87.83%
Ramachandran allowed	9.66%	9.48%	8.17%	8.66%
Ramachandran outliers	2.60%	2.92%	3.21%	3.51%
RNA				
Correct sugar puckers	99.68%	99.83%	99.35%	99.68%
Correct backbone conformation	87.04%	87.04%	86.99%	85.45%

[b] Compiled using MolProbity[67]

[c] Clashscore is the number of serious steric overlaps (> 0.4 Å) per 1,000 atoms.

Table S2. Summary of the interaction found in the β-globin LS48S and H4 LS48S ICs, described in this work. The contacts found for the first time are highlighted in blue.

protein factors	mRNA position β-globin	mRNA position H4	tRNA position	40S component	function
eIF1A					
Lys7	G(+3)				
Gly8-Gly9			A(35)		start-codon recognition by the eIF1A NTT, Kozak dependent
Trp70	G(+4)			18S rRNA A1819	
Arg12	**C(+7)**	**eIF1A absent**			
Lys67	G(+6)				
Lys68	U(+5)				
Gly9, Lys10				18S rRNA C1696	influence on the stability the cognate codon:anticodon duplex by the eIF1A NTT
Lys16, Asn17				C1327	
Asn44, Arg46				C1705 A1817-A1819,	
Lys64 , Arg 62				G604, C605	might depend on the mRNA sequence
Arg82, Tyr84, Gln58				uS12 Leu91, Gly56	
Asp83				eS30 Arg82	
	A(+1), C(-1)			18S rRNA G1203	stabilizing the ASL in the P-site
			C(34)	$m^1acp^3\Psi$ U1248	
				uS19	
			A(35)	**Arg140**	
			G(29), G(30)	**Thr136**	
				uS13	
			U(28), G(29)	**Thr145, Gly147**	
				uS9	
			C(33), C(32)	**Arg 146**	
				eS26	reinforcement of start-codon recognition in case of suboptimal Kozak sequence
	C(-4)	**A(-4)**		**His80**	
	(-8)			Ile41	stabilizing the mRNA channel exit site
	(-9)			Arg42	
	(-8), (-9)			Arg100	
				eS28	
	(-5)			Arg67	
	(-7)			Arg68	

protein factors	mRNA position		tRNA position	40S component	function
	β-globin	H4			
eIF2α Arg55 Arg57, Ser58, Asn60, Lys61	C(-3)		C(39-41)		stabilizing the ASL in the P-site
	mRNA entry channel (+14) to (+18) (+9), (+10) (+12), (+13) A(+13)			uS3 Arg117 helix α (117-128) β-hairpin (142-145) eS30 Lys126 uS5 Ala133	stabilizing the mRNA
ABCE1 HLH and NBD1: Ser150, Arg306, Asn310 NBD2: Lys584, Ile583 NBD1: **Pro265, Asp266** Fe-S cluster: Arg7 Lys20 Pro43 Asn74 Pro30, Val31, Ile56, Ile60				18S rRNA U478, A455, A454, C453, C452 A455, A454, C453 eS24 **C-terminal helix Gly128** 18S rRNA C471 G1718 A1719 G470 uS12 Ile50, Leu5, Ile75	binding to 40S
eIF3c **Lys332, Lys333, Thr381** **Leu379 and Ala380** eIF3a **Glu9, Asn16, Glu17**				eS27 **Val53, Gln65, Glu75** 18S rRNA **C1113, C1114, A1115, U1116** eS1 **Glu78, Ile189, Pro190**	binding to the 40S platform site
Tyr4-Pro8, Lys45-His47 and Gln80-Asn82 eIF3d **S166 to E172, Asn513, Lys514**	**5'UTR** **5'UTR**				binding to mRNA 5'UTR