Unique features of mRNA translation initiation in trypanosomatids

Anthony Bochler[1,#], Jailson Brito Querido[1,#,‡], Terezie Prilepskaja[2,3#], Heddy Soufari[1], Mayara Lucia Del Cistia[1], Lauriane Kuhn[4], Aline Rimoldi Ribeiro[1], Leoš Shivaya Valášek[2] and Yaser Hashem[1,*]

Affiliations:

[1]INSERM U1212, Institut Européen de Chimie et Biologie, University of Bordeaux, Pessac 33607, France

[2]Laboratory of Regulation of Gene Expression, Institute of Microbiology of the Czech Academy of Sciences, Videnska 1083, Prague, 142 20, the Czech Republic

[3]Charles University, Faculty of Science, Albertov 6, 128 00 Prague 2, Czech Republic

[4]Université de Strasbourg, CNRS, Plateforme Protéomique Strasbourg-Esplanade FRC 1589, F-67000 Strasbourg, France

[#]These authors contributed equally to this work.

[‡]Present address: MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Cambridge, UK

[*]Corresponding author. Email: yaser.hashem@u-bordeaux.fr

Abstract:

The 43S pre-initiation complex (PIC) assembly requires establishment of numerous interactions among eukaryotic initiation factors (eIFs), Met-tRNA$_i^{Met}$ and the small ribosomal subunit (40S). Owing to several differences in the structure and composition of kinetoplastidian 40S compared to their mammalian counterparts, translation initiation in trypanosomatids is suspected to display substantial variability. Here, we determined the structure of the 43S PIC from *Trypanosoma cruzi*, the Chagas disease parasite, showing numerous specific features, such as different eIF3 structure and interactions with the large rRNA expansion segments 9^S, 7^S and 6^S, and the association of a kinetoplastid-specific ~245 kDa DDX60-like helicase. We also revealed a previously undetermined binding site of the eIF5 C-terminal domain, and terminal tails of eIF2β, eIF1, eIF1A and eIF3 c and d subunits, uncovering molecular details of their critical activities.

One Sentence Summary: The 43S pre-initiation complex structure from *Trypanosoma cruzi* reveals kinetoplastid-specific features of translation initiation.

Main Text: The first critical initiation step in eukaryotes is the assembly of the 43S PIC comprising the 40S, the eIF2•GTP•Met-tRNA$_i^{Met}$ ternary complex, and eIFs 1, 1A, 3 and 5 (*1*, *2*). It is followed by the recruitment of the mRNA promoted by the mRNA cap-binding complex comprising eIF4A, 4B and 4F (*3*, *4*), forming the 48S PIC. The 48S PIC then scans the 5' untranslated region (UTR) of mRNA in the 5' to 3' direction till a start codon is encountered, upon which the majority of eIFs sequentially disassemble from the 40S

and the resulting 48S initiation complex (48S IC) joins the large ribosomal subunit (60S) to form an elongation-competent 80S ribosome.

Kinetoplastids is a group of flagellated unicellular eukaryotic parasites that have a complex life cycle. They spend part of their life cycle in the insect guts before being transmitted to the mammalian host upon biting. Common kinetoplastids include human pathogens such as *Trypanosoma cruzi, Trypanosoma brucei* and *Leishmania spp.*, etiologic agents of Chagas disease, African sleeping sickness and leishmaniasis, respectively. However, most of the related public health measures are mainly preventative and therapeutic strategies are extremely limited and often highly toxic. Since kinetoplastids have diverged early from other eukaryotes, their mRNA translational machineries developed unique molecular features unseen in other eukaryotic species. For instance, their 40S contains a kinetoplastid-specific ribosomal protein (KSRP) (*5*) and unusually oversized ribosomal RNA (rRNA) expansion segments (ES^S) (*6*). Since these unique features may play specific roles in kinetoplastidian mRNA translation, they provide potential specific drug targets.

It was proposed that two particularly oversized expansion segments, $ES6^S$ and $ES7^S$ located near the mRNA exit channel on the kinatoplastidian 40S, may contribute to modulating translation initiation in kinetoplastids by interacting with the structural core of the eukaryotic eIF3, specifically via its subunits a and c (*7*). eIF3 is the most complex eIF promoting not only nearly all initiation steps, but also translation termination, stop codon readthrough and ribosomal recycling (*8*). Among its initiation roles, eIF3 critically contributes to the assembly of the 43S PIC through a multitude of contacts that it makes with other eIFs, ensuring their recruitment to the 40S (*8*). Mammalian eIF3 comprises twelve subunits (eIF3a–m; excluding j), eight of which form the PCI/MPN octameric structural core (eIF3a, c, e, f, h, k, l and m) (*9*–*12*). Interestingly, unlike their mammalian hosts, kinetoplastids do not encode the eIF3m subunit (*13*, *14*) co-forming the octameric core in all known "12-subunit" species, strongly suggesting that the structure of their eIF3 core differs from that of mammals.

The 43S PIC assembly is also enhanced by the C-terminal domain (CTD) of eIF5 (*15*). Indeed, biochemical and genetics studies revealed that the eIF5-CTD possesses specific motifs interacting with several eIFs, such as the N-terminal tail (NTT) of the β subunit of eIF2 (*16*, *17*). However, the molecular details underlying the eIF5-CTD critical assembly role remain elusive, and – in contrast to the eIF5-NTD (*18*) – so are the structural details of its binding site within the 43S PIC (*19*). Importantly, structures of terminal tails of several essential eIFs in most of the available cryo-EM reconstructions are also lacking, mainly due to their intrinsic flexibility. Among them stand out the terminal tails of the c and d subunits of eIF3, eIF2β, eIF1 and eIF1A, all critically involved in scanning and AUG recognition.

Here, we solved the structure of the 43S PIC from *Trypanosoma cruzi* at near-atomic resolution and unraveled various new aspects of this complex, some of which are specific to trypanosomatids and others common to eukaryotes. Our structures thus allow us to 1) pin point essential, specific-features of trypanosomatids that could represent potential drug targets, and 2) expand our understanding of the interaction

network between several eIFs within the 43S PIC underlying molecular mechanism of its assembly, as well as of their roles in scanning for start codon recognition.

Results and discussion:

Composition of the 43S PIC in trypanosomatids

We purified endogenous pre-initiation complexes from two different species, *Trypanosoma cruzi* and *Leishmania tarentolae* by stalling the 43S complexes with GMP-PNP, a non-hydrolysable analog of GTP, as previously described (*20*). The proteomic analysis comparison between the stalled *versus* untreated complexes from *T. cruzi* indicated an obvious enrichment in canonical eIFs and ABCE1, as expected (see methods, Fig. 1A-B and S2). Surprisingly, we also identified an orthologue of the human DEAD-box RNA helicase DDX60 (Fig. 1B, S2). A similar repertoire of eIFs can also be found in the 43S PIC from *L. tarentolae* (Fig. S3). Besides initiation factors, several other proteins contaminating the 43S PIC can be found in *T. cruzi* and *L. tarentolae* samples without any apparent link to the translation process. Noteworthy, to date and to the best of our knowledge, DDX60 has never been co-purified with any PICs from any other studied eukaryote. Interestingly, while DDX60 is non-essential in mammals (*21*, *22*), it is required for the cell fitness in kinetoplastids and trypanosomatides (*23*), indicating that it could play a specific role in translation initiation in these parasites. It is not known whether or not it is essential in yeast.

The cryo-EM structure of the 43S PIC from *T. cruzi*

We next employed cryo-electron microscopy (cryo-EM) to determine the structure of the *T. cruzi* 43S PIC to an overall resolution of 3.3Å (Fig. S4), after image processing and extensive particle sorting. Our reconstruction reveals the so-called "scanning-conducive conformation" of the 43S PIC, in which the head of the 40S is tilted upwards to open up the mRNA channel for the subsequent mRNA loading (*7*, *9*, *24*). Thanks to the conservation of most of the identified initiation factors, we were able to segment the map and assign unambigiously densities corresponding to the 40S, eIF1, eIF1A, eIF2α, eIF2β, eIF2γ, Met-$tRNA_i^{Met}$ and the eIF3 structural core (Fig. 1C-E). Importantly, for the first time we could also identify the entire density corresponding to the N-terminal tail of the eIF3d subunit, implicated in the mRNA-specific translational control (*25*, *26*) (see below).

Furthermore, we observed an unassigned density contacting eIF2γ that has not been seen previously in any equivalent complexes. Since rigid body fitting of the crystal structure of the eIF5-CTD (*27*) showed a close agreement with this unassigned density and previous biochemical and genetics findings suggested a close co-operation between eIF5 and eIF2 on the ribosome (*16*, *28–30*), we assigned this density to the eIF5-CTD (Fig. 1C-E). Because the eIF5-CTD is known to interact with the eIF2β-NTT in both yeasts and mammals (*16*, *17*), we could also for the first time assign part of the eIF2β-NTT to its corresponding density (Fig. 1D) (see below). It is important to highlight that it was possible to assign the above-mentioned densities to eIF5-CTD only due to its general conservation among eukaryotes.

As discussed in detail below, beyond these evolutionary conserved features of the 43S PIC in eukaryotes, our cryo-EM reconstruction also identified several

trypanosomatide and kinetoplastid-specific peculiarities. For instance, the kinetoplastidian eIF2α contains a specific N-terminal domain insertion of unknown function (Fig. S5), and, indeed, an extra density on the eIF2α subunit can be observed (Fig. 1D-E, dashed circle). We also revealed a large density at the 40S interface, in the vicinity of the mRNA channel entrance (Fig. 1C-D), unseen in any of the previous mammalian and yeast 43S PIC reconstructions. Taken into account our proteomic analysis (Fig. 1B and Fig. S1 and S2), the size of this additional density and, above all, its high-resolution features, we were able to assign it unambiguously to the kinetoplastidian DDX60 (k-DDX60) helicase. These same k-DDX60 and eIF2α-NTT densities are also present in the *L. tarentolae* 43S PIC reconstruction (Fig. S6).

The eIF5 C-terminal domain (CTD) in the context of the 43S PIC

Importantly, detailed inspection of our structure allowed us to determine the eIF5-CTD binding site on the 43S PIC. It sits in a pocket formed by the eIF2β-NTT and eIF2γ (Fig. 2A-D). It was proposed that the three conserved poly-lysine stretches (dubbed "K-boxes") within the eIF2β-NTD mediate the eIF2 interaction with the eIF5-CTD (*16*, *17*). Interestingly, the K1 and K2–boxes are conserved in their basic charge character but replaced by R-rich stretches in kinetoplastids (Fig. S7). However, as our structure of eIF2β-NTT is only partial, we cannot validate their involvement in the interaction with eIF5. In contrast, the K3-box is not conserved in sequence among kinestoplastids (Fig. S7), it is replaced by the Q-rich motif, yet its position and orientation towards its binding partner in the eIF5-CTD is conserved (residues Gln 141 and 146 of eIF2β contact Glu333 of eIF5) (Fig. 2A). Additionally, our structure shows numerous other contacts between hydrophobic and charged residues on each side (residues 124 through 137 of eIF2β with residues 260 - 266, 320 - 330, and 360 - 373 of eIF5; Fig. 2A and B; see Table S1 for details). Since residues 360 through 373 correspond to the conserved and essential segment (known as the bipartite motif – AA (acidic/aromatic)-box; Fig. 2B, table S1), which was previously implicated in mediating the eIF5-CTD – eIF2β-NTT (*15*, *16*) contact, our structure provides critical structural evidence supporting earlier biochemical and genetics analysis.

Our structure also provides important molecular insight into the eIF5-CTD interaction with 1) the eIF2γ domain I (G-domain), where Arg229 of eIF5 contacts G-domain Gly223, and with 2) domain III, where Asp204, T205, T237 and N239 of eIF5 interact with domain III Asp432, ArgR469, Trp465 and Phe383 (Fig. 2B, table S1). Noteworthy, the eIF5-CTD shares a common topology with the CTD of the ε subunit of the nucleotide exchange factor eIF2B (*16*); they both fold into a W2-type HEAT domain (*27*) mediating contacts of both factors with the eIF2β-NTT and eIF2γ (*31*). Based on our structure, the arrangement of the eIF5-CTD HEAT domain binding site on eIF2γ in the context of the 43S PIC is similar to that of the eIF2Bε-CTD HEAT domain in the context of the recently solved eIF2-eIF2B complex (*32*, *33*).

Taken together, the eIF5-CTD interaction network revealed here indicates that the interaction between eIF5-CTD and eIF2γ could in principle induce a subtle

conformational change in its G-domain, allowing the eIF5-NTD (a GTPase activating domain of eIF5) to gain access to the GTP-biding pocket to promote reversible GTP hydrolysis on eIF2 during scanning, as demonstrated earlier (*34*).

Extensive interaction network of eIF1 in the context of the 43S PIC

After the GTP hydrolysis by eIF2γ, the release of the inorganic phosphate (P_i) is prevented by eIF1 until an AUG start codon is recognized by the anticodon of Met-$tRNA_i^{Met}$ leading to the full accommodation of TC in the decoding pocket (*2*, *34*) and eIF1 replacement by the eIF5-NTD. Because the access to the GTP-binding pocket on eIF2γ is in part protected by the zinc-binding domain (ZBD) of the eIF2β-CTD (*24*, *35*), it was unclear how eIF1 coordinates the release of free P_i together with the latter factor. Based on biochemical and genetic studies in yeast, eIF2β and eIF3c NTD were implicated in anchoring of eIF1 within the 48S PIC (*36*–*39*), prior to the start codon recognition. However, the molecular basis underlying all these critical interactions remained poorly characterized.

In accord with earlier biochemical experiments, our structure reveals that the conserved eIF2β-C terminal tail (eIF2β-CTT) (Fig. 2G), together with the eIF3c-NTD, does anchor eIF1 within the 43S PIC (Fig. 2D-E). In particular, the eIF2β-CTT extends toward the P-site, where it interacts with eIF1 (mainly through Tyr326 with Val77, conserved in character) and with h24 of the 18S rRNA (Arg 333 and 337 with nucleotides U1340 and 1339, respectively) (Fig. 2 E and G). Based on these findings, we examined binding of human eIF2β with eIF1 fused to GST moiety using the GST pull down assay and revealed that the interaction between the CTTs of eIF2β (residues 310 – 333) and eIF1 is also conserved in mammals (Fig. 2H, Fig. S10A). The contact between the eIF3c-NTD and eIF1 involves Arg26 through Thr39 of eIF3c, and Glu95, Asn96, and Asn50 through Arg56 of eIF1 (Fig. 2D-E; table S1). In accord, *T.c.* eIF1 fused to GST moiety interacted specifically with the eIF3c-NTD also *in vitro* (between eIF3c residues 14 and 38) (Fig. 2I). Besides eIF1, the eIF3c-NTD critically promotes scanning for AUG recognition also through its interaction with the eIF5-CTD, which was so far identified only in yeast *S. cerevisiae* (*37*–*40*) but was expected to be conserved among all eukaryotes. Surprisingly then, we did not detect any binding between the *T. cruzi* eIF3c-NTD and eIF5 fused to GST moiety under any experimental conditions that we examined (Fig. 2I). Even though we cannot rule out improper folding as the primary cause of this failure, we speculate that these results may point to a specific evolutionary shift in kinetoplastidian initiation pathway, as will be discussed below.

To further investigate the conservation of the eIF3c-NTD interactions in higher eukaryotes, we fused human eIF1 and eIF5 to GST and tested the resulting fusion proteins against various truncations of the eIF3c-NTD (Fig. 2J). In accord with the earlier yeast data (*37*, *38*), the first ~30 residues of the eIF3c-NTD mediate its binding with eIF5, whereas residues 130 through 325 contact eIF1 (Fig. 2J). These findings contrast with those seen in trypanosomatids, where the tip of the eIF3c-NTD interacted with eIF1 instead of eIF5 (Fig. 2I). Therefore, although the role of the eIF3c-NTD for eIF1

anchoring to the PICs is conserved among eukaryotes, at least in trypanosomatids it seems to be achieved by a species-specific segment.

Besides the eIF1-CTT binding coordinates, our structure also reveals that the N-terminal tail of eIF1 (residues 10 to 22) forms an α-helix that interacts with domains I and III of eIF2γ (Val147 and Gln412, respectively, Fig. 2D and E; table S1), very close to the GTP binding pocket. We propose that these contacts could underlie the role of eIF1 in releasing the P_i by inducing a subtle conformational change in the GTP binding pocket upon sensing the recognition of the start codon through its apical β-hairpin loop at the P-site.

Finally, even though eIF1A appears to interact with eIF1 in a canonical fashion seen in other eukaryotes, it shows that the eIF1A-CTT extends towards the head of the 40S, where it interacts with the rRNA (Arg155 with G1685 and Asn156 with G1714) (Fig. 2F) and ribosomal proteins uS19 (residues Val158 through Asp161 with Lys84, Gln108 and Ala111) and uS13 (residues Asp162 through Leu164 with Val124 and Tyr128; Fig. 2F, table S1), corroborating findings from a previous hydroxyl-radical probing study (*41*).

The specific features and binding site of eIF3 in trypanosomatids

Strikingly, as seen in Figure 3A-D, the unusually large trypanosomatids-specific ES^S are involved in translation initiation by acting as docking platforms for different subunits of eIF3. Similarly to other eukaryotes reported so far, the eIF3 core binds to the 40S through its a and c subunits (Fig. 3C-D). However, unlike in other known eukaryotes, the large $ES7^S$ acts as the main docking point for the eIF3 structural core (Fig. S8A). In particular, the eIF3c is tweezed between $ES7^S$-helix A ($ES7^S$-hA) and $ES7^S$-hB forming a large, kinetoplastid-specific binding site, involving residues Gln204, Lys207, Arg232, Arg243, Gln329 and Arg331 and $ES7^S$ nucleotides A1525, A1523 and U1524, U1476, U1526, G1438 and U1439, respectively (Fig. 3D, table S1). High local resolution of our complex allowed us to assign the identity of the conserved helical domain of the eIF3c-NTD (Fig. 3A, dashed oval) spanning residues 55 through 156. The eIF3c-NTD interacts with the 18S rRNA at the platform region through several evolutionary well-conserved residues on each side of this domain (table S1), suggesting that it has a similar PIC binding mode also in mammals, despite the obvious differences in binding to eIFs 1 and 5 reported above. In addition to these main contacts with the rRNA, a minor interaction of eIF3c can be observed with eS27 (via residues Glu191 and Lys192 with Glu56 and Lys63) (Fig. 3D). In contrast to eIF3c, the eIF3a binding to the ribosomal protein eS1 does not seem to differ from other eukaryotes (residues Arg8, Thr12 and Leu17 contact Thr72, Arg192 and Ile194, respectively) (Fig. 3C).

Another unusually large ES is the kinetoplastidian $ES9^S$ that forms a "horn" on the 40S head, bending towards the mRNA exit channel, where it binds to and stabilizes eIF3d within the 43S PIC (Fig. 3A-B, table S1), representing another important feature that is specific to translation initiation in trypanosomatids. In particular, the eIF3d main globular domain interacts with $ES9^S$ through residues Lys292, Arg294 and Gln296 contacting nucleotides C1867, U1862 and C1868, respectively. Moreover, Arg149,

Lys301 and Asn302 of eIF3d interact with U1863 of ES9^S(Fig. S8A). Noteworthy, structures of ES7^S and the exceptionally large ES6^S undergo drastic conformational changes upon binding of eIF3, as can be observed by comparing this structure with our previous *T. cruzi* 40S lacking eIF3 (*5*) (Fig. S8B). Robustness of these conformational acrobatics indicates their functional importance that, in turn, sets them in the viewfinder for the future drug-targeting studies.

When compared to its mammalian counterpart, the overall conformation of eIF3 structural core differs significantly (Fig. 3E-F, S8C-D), mainly due to the lack of the eIF3m subunit in trypanosomatids, which is in part compensated for by the rearrangements of the other core eIF3 subunits like a, c, e, k, l, but mostly f and h. Indeed, eIF3 f and h shift several α-helices and coils to fill for the absence of the m subunit; this rearrangement is probably required for the maintenance of the eIF3 core central helical bundle (Fig. S8C-D, arrows indicate the direction of the shift). Moreover, a charge surface analysis reveals very different charge distribution patterns between *T. cruzi* eIF3 and its mammalian counterpart (Fig. S9A-B), in part as a consequence of the different 40S binding surface that is mainly represented by rRNA, in contrast to other known eukaryotes.

Importantly, our cryo-EM reconstruction reveals the full structure of eIF3d that appeared separated from the eIF3 structural core in the context of the PIC in all previous studies (*7*, *9*, *42*). We show here that the eIF3d-NTT, unseen in any previous equivalent complexes, extends towards eIF3e, where it interacts with its PCI domain (residues 1-19 of eIF3d and 244-252 of eIF3e) (Fig. 3G-I, table S1). Furthermore, the eIF3d-NTT also comes in a less extensive contact with eIF3a and eIF3c (Fig. 3H and I, table S1). In agreement, the interaction of the eIF3d-NTT (the first 114 residues) with the eIF3 core was previously shown in biochemical and genetics studies (*43*). To support our structural data and investigate the evolutionary conservation of the eIF3d contacts with eIF3 e, a and c subunits within the PIC, we expressed human homologues of all these proteins and subjected them to our GST pull down analysis. As shown in Figures 3J and K and S10B and C, the main contact between eIF3d and eIF3e does involve the first 19 residues (in particular W16, G17, and P18) of the former and residues I246, Q247, and T248 of the latter subunit even in humans. In addition, weak but reproducible binding between eIF3d and eIF3a and eIF3c subunits was also detected, in contrast to other eIF3 subunits (Fig. S10D and E). Since human eIF3d was shown to interact with the mRNA cap (*26*) and, together with several other eIF3 subunits (including eIF3a and eIF3e) proposed to promote recruitment of selected mRNAs to the 43S PIC to control their expression in response to various stresses and cellular signals (*25*, *44*), we speculate that these contacts play pivotal role in coordinating the eIF3d-specific functions with the rest of eIF3 on the ribosome.

The trypanosomatid-specific k-DDX60

As mentioned above, our cryo-EM reconstructions of the *T. cruzi* and *L. tarentolae* 43S PICs revealed a large density at the intersubunit side of the 40S (Fig. 1B-D, S6). Known structures of eIFs and ABCE1 (*9*, *18*, *45*) do not fit into this density and

proteomic analysis shows substantial presence of the helicase DDX60 protein in our samples (Fig. 1B, S1 and S2) that we henceforward refer to as kinetoplastidian-DDX60 (k-DDX60). The density was of sufficient resolution to build a near-complete atomic model of k-DDX60, including the helicase recombinase A (RecA) domains (Fig. 4), which fully validates our assignment. Besides the RecA domains, k-DDX60 counts two winged-helices domains, two ratchet domains and one kinetoplastid-specific A-site insert (AI) that protrudes at the end of the RecA2 domain from the C-terminal cassette (Fig. 4C-E).

The presence of k-DDX60 is not due to the use of GMP-PNP, as we did not retrieve any densities resembling GMP-PNP in any of k-DDX60 RecA domains. In addition, its known mammalian DDX60 homologue is an ATP helicase. Next we wanted to inspect structural impact of its ATPase activity by determining the structure of the 43S PIC purified from *T. cruzi* cell lysate supplemented with ATP, in addition to GMP-PNP (Fig. 5A). It is important to stress out that the resolution of the 43S PIC+ATP reconstruction is above 4Å, precluding unambiguous determination of whether ATP hydrolysis took place or not. Nonetheless, the structure reveals a global conformational rearrangement of the 40S head (Fig. 5B-C), which could be driven by the k-DDX60 rearrangement upon ATP hydrolysis (Fig. 5D-F). In addition, we also observe the presence of an extra density at the RecA1 domain of the C-terminal cassette at the position that is unoccupied in the absence of ATP (Fig. 5D).

k-DDX60 binds both to the head and the body of the 40S and the structural dynamics induced by the ATP addition suggest its involvement in remodeling of the 43S PIC mRNA channel due to the head swiveling. Importantly, the AI extended helix of k-DDX60 interacts with the anticodon stem-loop of the Met-tRNA$_i^{Met}$ (Fig. 4C), preventing the codon-anticodon interaction in its presence. The release of k-DDX60, or at least of its AI helix, must therefore precede the rotation of the 40S head and the full accommodation of the Met-tRNA$_i^{Met}$ in the P-site. Moreover, k-DDX60 interacts directly with eIF1A, eIF2β, eIF2γ, eIF3c and eIF5 (Fig. 4E), in addition to the 18S rRNA and ribosomal proteins eS12, uS12, eS30 and eS31 (Fig. 4E), suggesting its direct involvement in structural changes accompanying/driving the AUG recognition process. In fact, we believe that owing to these extensive interactions with numerous components of the 43S PIC, presence of k-DDX60 provided the much needed stabilization support to enable the resolution of flexible tails of most eIFs present in our complexes. In agreement, most of these interactions occur via additional domains and insertions of k-DDX60 that are inexistent in its mammalian homologue (Fig. 4D, S11). It is not clear why translation initiation, perhaps in particular the AUG selection process, in kinetoplastids requires this specific helicase. Interestingly, all mature cytoplasmic mRNAs in kinetoplastids possess a 39-nucleotide spliced leader that confers them an unusual hypermethylated 5′-cap structure (known as cap4)*(46)*. Therefore, the presence of this helicase might be required for an efficient recruitment and handling of these kinetoplastid-specific mRNAs until the start codon has been recognized.

Conclusion

In summary, our structure reveals numerous novel features of the eukaryotic translation initiation machinery, some of which are common to other eukaryotes, such as the placement and proposed roles of terminal tails of eIF1, eIF1A, eIF2β, eIF3c, eIF3d, and, above all, the precise binding site of the eIF5-CTD within the 43S PIC (Fig. 6A-C). Furthermore, our data uncover several striking features of translation initiation specific to kinetoplastids (Fig. 6D-F), such as the role of the oversized kinetoplastidian ES^S in providing a large, unique binding surface for eIF3, as well as the first structural characterization of k-DDX60. These unique molecular features of translation initiation in kinetoplastids represent an unprecedented opportunity to interfere specifically with the initiation process in these "hard-to-combat" parasites, which may stimulate new venues of research and development of new effective drugs against trypanosomiasis and leishmaniasis.

References and Notes:

1. L. S. Valásek, 'Ribozoomin'--translation initiation from the perspective of the ribosome-bound eukaryotic initiation factors (eIFs). *Curr. Protein Pept. Sci.* **13**, 305–30 (2012).
2. A. G. Hinnebusch, Structural Insights into the Mechanism of Scanning and Start Codon Recognition in Eukaryotic Translation Initiation. *Trends Biochem. Sci.* **42**, 589–611 (2017).
3. E. Guca, Y. Hashem, Major structural rearrangements of the canonical eukaryotic translation initiation complex. *Curr. Opin. Struct. Biol.* **53**, 151–158 (2018).
4. Y. Hashem, J. Frank, The Jigsaw Puzzle of mRNA Translation Initiation in Eukaryotes: A Decade of Structures Unraveling the Mechanics of the Process. *Annu. Rev. Biophys.* (2018), doi:10.1146/annurev-biophys-070816-034034.
5. J. Brito Querido, E. Mancera-Martínez, Q. Vicens, A. Bochler, J. Chicher, A. Simonetti, Y. Hashem, The cryo-EM Structure of a Novel 40S Kinetoplastid-Specific Ribosomal Protein. *Structure*. **25**, 1785-1794.e3 (2017).
6. Y. Hashem, A. des Georges, J. Fu, S. N. Buss, F. Jossinet, A. Jobe, Q. Zhang, H. Y. Liao, R. A. Grassucci, C. Bajaj, E. Westhof, S. Madison-Antenucci, J. Frank, High-resolution cryo-electron microscopy structure of the Trypanosoma brucei ribosome. *Nature*. **494**, 385–9 (2013).
7. Y. Hashem, A. des Georges, V. Dhote, R. Langlois, H. Y. Liao, R. A. Grassucci, C. U. T. Hellen, T. V Pestova, J. Frank, Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. *Cell*. **153**, 1108–19 (2013).
8. L. S. Valášek, J. Zeman, S. Wagner, P. Beznosková, Z. Pavlíková, M. P. Mohammad, V. Hronová, A. Herrmannová, Y. Hashem, S. Gunišová, Embraced by eIF3: structural

and functional insights into the roles of eIF3 across the translation cycle. *Nucleic Acids Res.* **45**, 10948–10968 (2017).

9. A. des Georges, V. Dhote, L. Kuhn, C. U. T. Hellen, T. V Pestova, J. Frank, Y. Hashem, Structure of mammalian eIF3 in the context of the 43S preinitiation complex. *Nature*. **525**, 491–5 (2015).
10. C. Sun, A. Todorovic, J. Querol-Audí, Y. Bai, N. Villa, M. Snyder, J. Ashchyan, C. S. Lewis, A. Hartland, S. Gradia, C. S. Fraser, J. A. Doudna, E. Nogales, J. H. D. Cate, Functional reconstitution of human eukaryotic translation initiation factor 3 (eIF3). *Proc. Natl. Acad. Sci. U. S. A.* **108**, 20473–8 (2011).
11. S. Wagner, A. Herrmannová, R. Malík, L. Peclinovská, L. S. Valášek, Functional and biochemical characterization of human eukaryotic translation initiation factor 3 in living cells. *Mol. Cell. Biol.* **34**, 3041–52 (2014).
12. S. Wagner, A. Herrmannová, D. Šikrová, L. S. Valášek, Human eIF3b and eIF3a serve as the nucleation core for the assembly of eIF3 into two interconnected modules: the yeast-like core and the octamer. *Nucleic Acids Res.* (2016), doi:10.1093/nar/gkw972.
13. K. Li, S. Zhou, Q. Guo, X. Chen, D. Lai, Z. Lun, X. Guo, The eIF3 complex of Trypanosoma brucei: composition conservation does not imply the conservation of structural assembly and subunits function. *RNA*. **23**, 333–345 (2017).
14. A. M. Rezende, L. A. Assis, E. C. Nunes, T. D. da Costa Lima, F. K. Marchini, E. R. Freire, C. R. S. Reis, O. P. de Melo Neto, The translation initiation complex eIF3 in trypanosomatids and other pathogenic excavates--identification of conserved and divergent features based on orthologue analysis. *BMC Genomics*. **15**, 1175 (2014).
15. K. Asano, A. Shalev, L. Phan, K. Nielsen, J. Clayton, L. Valásek, T. F. Donahue, A. G. Hinnebusch, Multiple roles for the C-terminal domain of eIF5 in translation initiation complex assembly and GTPase activation. *EMBO J.* **20**, 2326–37 (2001).
16. K. Asano, T. Krishnamoorthy, L. Phan, G. D. Pavitt, A. G. Hinnebusch, Conserved bipartite motifs in yeast eIF5 and eIF2Bepsilon, GTPase-activating and GDP-GTP exchange factors in translation initiation, mediate binding to their common substrate eIF2. *EMBO J.* **18**, 1673–88 (1999).
17. S. Das, T. Maiti, K. Das, U. Maitra, Specific interaction of eukaryotic translation initiation factor 5 (eIF5) with the beta-subunit of eIF2. *J. Biol. Chem.* **272**, 31712–8 (1997).
18. J. L. Llácer, T. Hussain, A. K. Saini, J. S. Nanda, S. Kaur, Y. Gordiyenko, R. Kumar, A. G. Hinnebusch, J. R. Lorsch, V. Ramakrishnan, Translational initiation factor eIF5 replaces eIF1 on the 40S ribosomal subunit to promote start-codon recognition. *Elife*. **7** (2018), doi:10.7554/eLife.39273.
19. J. Zeman, Y. Itoh, Z. Kukačka, M. Rosůlek, D. Kavan, T. Kouba, M. E. Jansen, M. P. Mohammad, P. Novák, L. S. Valášek, Binding of eIF3 in complex with eIF5 and eIF1 to the 40S ribosomal subunit is accompanied by dramatic structural changes. *Nucleic Acids Res.* **47**, 8282–8300 (2019).
20. A. Simonetti, J. Brito Querido, A. G. Myasnikov, E. Mancera-Martinez, A. Renaud, L. Kuhn, Y. Hashem, eIF3 Peripheral Subunits Rearrangement after mRNA Binding

and Start-Codon Recognition. *Mol. Cell*. **63**, 206–217 (2016).

21. M. Miyashita, H. Oshiumi, M. Matsumoto, T. Seya, DDX60, a DEXD/H box helicase, is a novel antiviral factor promoting RIG-I-like receptor-mediated signaling. *Mol. Cell. Biol.* **31**, 3802–19 (2011).
22. H. Oshiumi, M. Miyashita, M. Okamoto, Y. Morioka, M. Okabe, M. Matsumoto, T. Seya, DDX60 Is Involved in RIG-I-Dependent and Independent Antiviral Responses, and Its Function Is Attenuated by Virus-Induced EGFR Activation. *Cell Rep.* **11**, 1193–207 (2015).
23. S. Alsford, D. J. Turner, S. O. Obado, A. Sanchez-Flores, L. Glover, M. Berriman, C. Hertz-Fowler, D. Horn, High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Res.* **21**, 915–24 (2011).
24. J. L. Llácer, T. Hussain, L. Marler, C. E. Aitken, A. Thakur, J. R. Lorsch, A. G. Hinnebusch, V. Ramakrishnan, Conformational Differences between Open and Closed States of the Eukaryotic Translation Initiation Complex. *Mol. Cell*. **59**, 399–412 (2015).
25. A. S. Y. Lee, P. J. Kranzusch, J. H. D. Cate, eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature* (2015), doi:10.1038/nature14267.
26. A. S. Lee, P. J. Kranzusch, J. A. Doudna, J. H. D. Cate, eIF3d is an mRNA cap-binding protein that is required for specialized translation initiation. *Nature*. **536**, 96–9 (2016).
27. Z. Wei, Y. Xue, H. Xu, W. Gong, Crystal structure of the C-terminal domain of S.cerevisiae eIF5. *J. Mol. Biol.* **359**, 1–9 (2006).
28. R. E. Luna, H. Arthanari, H. Hiraishi, J. Nanda, P. Martin-Marcos, M. A. Markus, B. Akabayov, A. G. Milbradt, L. E. Luna, H.-C. Seo, S. G. Hyberts, A. Fahmy, M. Reibarkh, D. Miles, P. R. Hagner, E. M. O'Day, T. Yi, A. Marintchev, A. G. Hinnebusch, J. R. Lorsch, K. Asano, G. Wagner, 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).
29. C. R. Singh, Y. Yamamoto, K. Asano, Physical association of eukaryotic initiation factor (eIF) 5 carboxyl-terminal domain with the lysine-rich eIF2beta segment strongly enhances its binding to eIF3. *J. Biol. Chem.* **279**, 49644–55 (2004).
30. C. R. Singh, R. Watanabe, W. Chowdhury, H. Hiraishi, M. J. Murai, Y. Yamamoto, D. Miles, Y. Ikeda, M. Asano, K. Asano, Sequential eukaryotic translation initiation factor 5 (eIF5) binding to the charged disordered segments of eIF4G and eIF2β stabilizes the 48S preinitiation complex and promotes its shift to the initiation mode. *Mol. Cell. Biol.* **32**, 3978–89 (2012).
31. P. V Alone, T. E. Dever, Direct binding of translation initiation factor eIF2gamma-G domain to its GTPase-activating and GDP-GTP exchange factors eIF5 and eIF2B epsilon. *J. Biol. Chem.* **281**, 12636–44 (2006).
32. L. R. Kenner, A. A. Anand, H. C. Nguyen, A. G. Myasnikov, C. J. Klose, L. A. McGeever, J. C. Tsai, L. E. Miller-Vedam, P. Walter, A. Frost, eIF2B-catalyzed nucleotide

exchange and phosphoregulation by the integrated stress response. *Science*. **364**, 491–495 (2019).

33. K. Kashiwagi, T. Yokoyama, M. Nishimoto, M. Takahashi, A. Sakamoto, M. Yonemochi, M. Shirouzu, T. Ito, Structural basis for eIF2B inhibition in integrated stress response. *Science*. **364**, 495–499 (2019).
34. M. A. Algire, D. Maag, J. R. Lorsch, Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Mol. Cell*. **20**, 251–62 (2005).
35. E. Stolboushkina, S. Nikonov, A. Nikulin, U. Bläsi, D. J. Manstein, R. Fedorov, M. Garber, O. Nikonov, Crystal structure of the intact archaeal translation initiation factor 2 demonstrates very high conformational flexibility in the alpha- and beta-subunits. *J. Mol. Biol.* **382**, 680–91 (2008).
36. A. Thakur, L. Marler, A. G. Hinnebusch, A network of eIF2β interactions with eIF1 and Met-tRNAi promotes accurate start codon selection by the translation preinitiation complex. *Nucleic Acids Res.* **47**, 2574–2593 (2019).
37. E. Obayashi, R. E. Luna, T. Nagata, P. Martin-Marcos, H. Hiraishi, C. R. Singh, J. P. Erzberger, F. Zhang, H. Arthanari, J. Morris, R. Pellarin, C. Moore, I. Harmon, E. Papadopoulos, H. Yoshida, M. L. Nasr, S. Unzai, B. Thompson, E. Aube, S. Hustak, F. Stengel, E. Dagraca, A. Ananbandam, P. Gao, T. Urano, A. G. Hinnebusch, G. Wagner, K. Asano, Molecular Landscape of the Ribosome Pre-initiation Complex during mRNA Scanning: Structural Role for eIF3c and Its Control by eIF5. *Cell Rep.* **18**, 2651–2663 (2017).
38. M. Karásková, S. Gunišová, A. Herrmannová, S. Wagner, V. Munzarová, L. S. L. S. Valášek, S. Gunisová, L. S. Valásek, Functional Characterization of the Role of the N-terminal Domain of the c/Nip1 Subunit of Eukaryotic Initiation Factor 3 (eIF3) in AUG Recognition. *J. Biol. Chem.* **287**, 28420–34 (2012).
39. L. Valásek, K. H. Nielsen, F. Zhang, C. A. Fekete, A. G. Hinnebusch, Interactions of eukaryotic translation initiation factor 3 (eIF3) subunit NIP1/c with eIF1 and eIF5 promote preinitiation complex assembly and regulate start codon selection. *Mol. Cell. Biol.* **24**, 9437–55 (2004).
40. L. Phan, X. Zhang, K. Asano, J. Anderson, H. P. Vornlocher, J. R. Greenberg, J. Qin, A. G. Hinnebusch, Identification of a translation initiation factor 3 (eIF3) core complex, conserved in yeast and mammals, that interacts with eIF5. *Mol. Cell. Biol.* **18**, 4935–46 (1998).
41. Y. Yu, A. Marintchev, V. G. Kolupaeva, A. Unbehaun, T. Veryasova, S.-C. Lai, P. Hong, G. Wagner, C. U. T. Hellen, T. V Pestova, Position of eukaryotic translation initiation factor eIF1A on the 40S ribosomal subunit mapped by directed hydroxyl radical probing. *Nucleic Acids Res.* **37**, 5167–82 (2009).
42. B. Eliseev, L. Yeramala, A. Leitner, M. Karuppasamy, E. Raimondeau, K. Huard, E. Alkalaeva, R. Aebersold, C. Schaffitzel, Structure of a human cap-dependent 48S translation pre-initiation complex. *Nucleic Acids Res.* **46**, 2678–2689 (2018).
43. M. D. Smith, L. Arake-Tacca, A. Nitido, E. Montabana, A. Park, J. H. Cate, Assembly of eIF3 Mediated by Mutually Dependent Subunit Insertion. *Structure*. **24**, 886–96

(2016).

44. M. Shah, D. Su, J. S. Scheliga, T. Pluskal, S. Boronat, K. Motamedchaboki, A. R. Campos, F. Qi, E. Hidalgo, M. Yanagida, D. A. Wolf, A Transcript-Specific eIF3 Complex Mediates Global Translational Control of Energy Metabolism. *Cell Rep.* **16**, 1891–902 (2016).
45. J. P. Erzberger, F. Stengel, R. Pellarin, S. Zhang, T. Schaefer, C. H. S. Aylett, P. Cimermančič, D. Boehringer, A. Sali, R. Aebersold, N. Ban, Molecular Architecture of the 40S·eIF1·eIF3 Translation Initiation Complex. *Cell*. **158**, 1123–1135 (2014).
46. S. Michaeli. Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future Microbiol*. **6**, 459-474 (2011).
47. S. Q. Zheng, E. Palovcak, J. P. Armache, K. A. Verba, Y. Cheng and D. A. Agard (2017). "MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy." *Nat Methods* **14**, 331-332.
48. K. Zhang. "Gctf: Real-time CTF determination and correction." *J Struct Biol* **193**, 1-12 (2016).
49. J. Zivanov, T. Nakane, B. O. Forsberg, D. Kimanius, W. J. Hagen, E. Lindahl and S. H. Scheres (2018). "New tools for automated high-resolution cryo-EM structure determination in RELION-3." *Elife* **7**.
50. A. Kucukelbir, F. J. Sigworth and H. D. Tagare (2014). "Quantifying the local resolution of cryo-EM density maps." *Nat Methods* **11**, 63-6551. E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng, T. E. Ferrin, UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605-12 (2004).
52. P. Emsley and K. Cowtan, Coot: model-building tools for molecular graphics. *Acta Crystallogr.* **D60**, 2126-2132 (2004).
53. P. D. Adams, P. V. Afonine, G. Bunkóczi, V. B. Chen, I. W. Davis, N. Echols, et al., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst.* **D66**, 213-221 (2010).
54. W. Humphrey, A. Dalke and K. Schulten, VMD - Visual Molecular Dynamics. *J. Molec. Graphics* **14**, 33-38 (1996).
55. K. F. Santos, S. M. Jovin, G. Weber, V. Pena, R. Luhrmann, M. C. Wahl, Structural basis for functional cooperation between tandem helicase cassettes in Brr2-mediated remodeling of the spliceosome. *Proc.Natl.Acad.Sci*.USA **109**: 17418-17423 (2012).
56. V. A. Simossis and J. Heringa, Optimally segmented consensus secondary structure prediction. Bioinformatics (2004).
57. A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296-W303 (2018).

58. L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols* **10**, 845-858 (2015)
59. C. Bieniossek, P. Schutz, M. Bumann, A. Limacher, I. Uson, U. Baumann, The Crystal Structure of the Carboxy-Terminal Domain of Human Translation Initiation Factor Eif5. *J.Mol.Biol.* **360**, 457 (2006).
60. L. G. Trabuco, E. Villa, K. Mitra, J. Frank, and K. Schulten, Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673-683 (2008).
61. F. C. Chou, P. Sripakdeevong, S. M. Dibrov, T. Hermann, and R. Das, Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat Methods* **10**, 74-76 (2012).

ACKNOWLEDGMENTS

We thank Christoph Diebolder, and Ludovic Renault (NeCEN, Leiden, Holland) as well as Julio Ortiz Espinoza, Corinne Crucifixand and Christine Ruhlmann (IGBMC, Strasbourg, France) for assistance with data acquisition. We would like also to thank Angelita Simonetti for her help in the samples purification and the High-Performance Computing Center of the University of Strasbourg for IT support. The mass spectrometry instrumentation was funded by the University of Strasbourg, IdEx "Equipement mi-lourd" 2015. **Funding:** This work was supported by ERC-2017-STG #759120 "TransTryp" (to Y.H.), Labex: ANR-10-LABX-0036_NETRNA (to Y.H.), ANR-14-ACHN-0024 - CryoEM80S (to Y.H.), the Grant of Excellence in Basic Research (EXPRO 2019) provided by the Czech Science Foundation (19-25821X to L.S.V.), and Charles University Grant Agency (project GA UK No. 244119 to T.P.).

Author contributions: J.B.Q. purified and characterized the complexes from *T. cruzi*, M.L.D.C. and A.R.R. purified the complex from *L. tarentolae*. T.P. preformed the GST pulldown assays and analyzed the data together with L.S.V. Y.H. and H.S. performed the cryo-EM data processing. L.K. performed MS/MS analysis. A.B., J.B.Q. and Y.H. interpreted the cryo-EM data. A.B. and Y.H. performed the molecular modeling. J.B.Q., T.P., A.B., L.S.V. and Y.H. wrote the manuscript. Y.H. supervised the research.

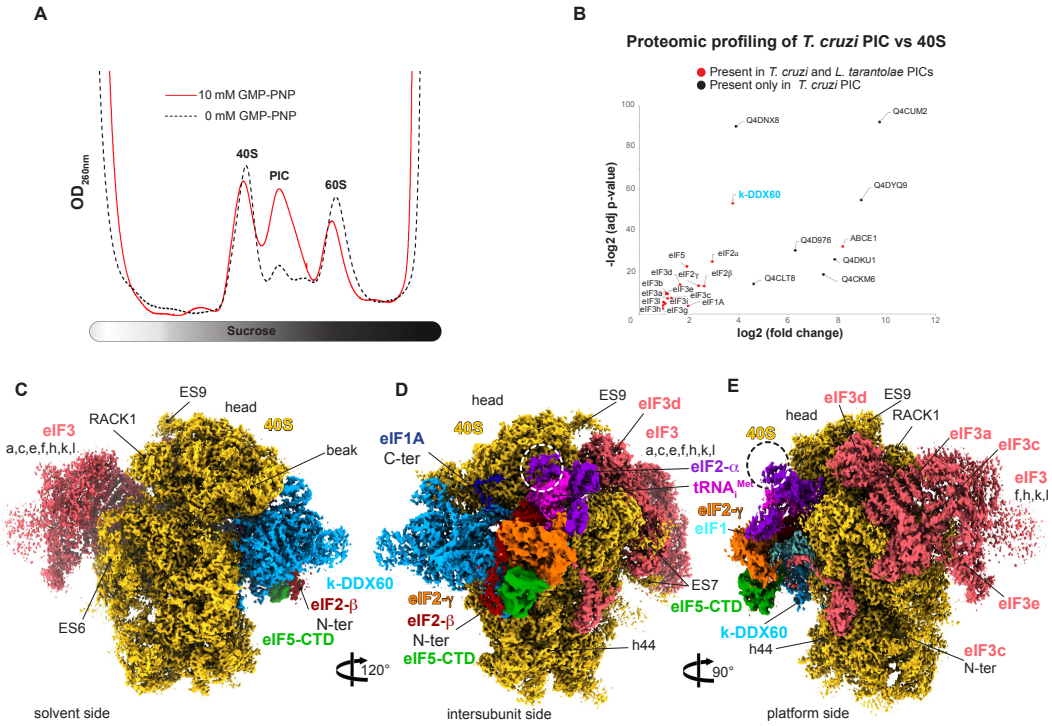

Fig. 1. Composition and cryo-EM structure of the T. cruzi 43S PIC. (**A**) The effect of the GMP-PNP treatment on the 43S PIC stabilization in the *T.cruzi* lysate assessed by UV absorbance profile analyses (**B**) Proteomic profiling of the endogenous pre-initiation complex in comparison with native 40Ss purified from the *T. cruzi* cell lysate (see methods for the validation). (**C**) The overall structure of the *T. cruzi* 43S PIC shown from the intersubunit side. The initiation factors are colored variably. (**D**) The 43S PIC reconstruction focused on the solvent side. Extra density of eIF2α corresponding to the kinetoplastidian specific N-terminal insertion is encircled by a dashed line. (**E**) The 43S PIC reconstruction focused on eIF3 and the 40S platform.

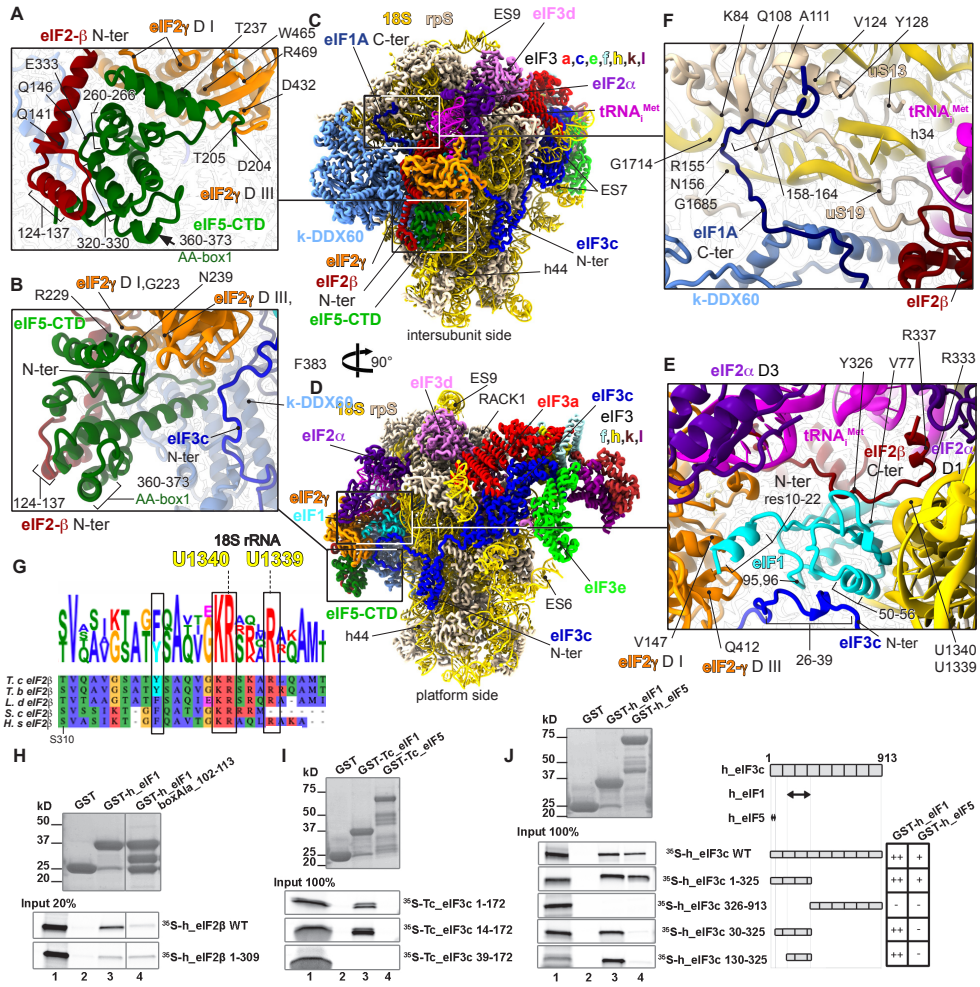

Fig. 2. Atomic model of the 43S PIC showing the interaction network of various eIFs. (**A**) Close-up view of an atomic model of the eIF5-CTD (in green), the eIF2β-NTT (in cherry red) and eIF2γ (in orange) shown from the intersubunit side. (**B**) Close-up view of the eIF5-CTD (in green) and its interaction with eIF2 from the platform side. (**C**) The overall view of atomic model of the 43S PIC from the intersubunit and **(D)** the platform side. (**E**) Close-up view of the P-site, showing eIF1 (in cyan) and its biding partners the eIF2β-CTT (in cherry red) and the eIF3c-NTD (in blue). (**F**) Close-up view of the eIF1A-CTT and its interactions with h34, uS13 and uS19. (**G**) Polypeptide sequence alignment of the eIF2β-CTT, highlighting residues involved in the interaction with 18S rRNA and eIF1; *T. cruzi, T. brucei, L. donovani, S. cerevisiae* and *H. sapiens*. Residue numbreding from *H. sapiens* was used (**H**) *In vitro* protein-protein binding analysis of the interaction between human eIF2β and GST-eIF1. (**I**) Binding analysis between the *T. cruzi* eIF3c-NTD and GST-eIF1 and GST-eIF5. (**J**) Binding analysis between human eIF3c-NTD and GST-eIF1 and GST-eIF5.

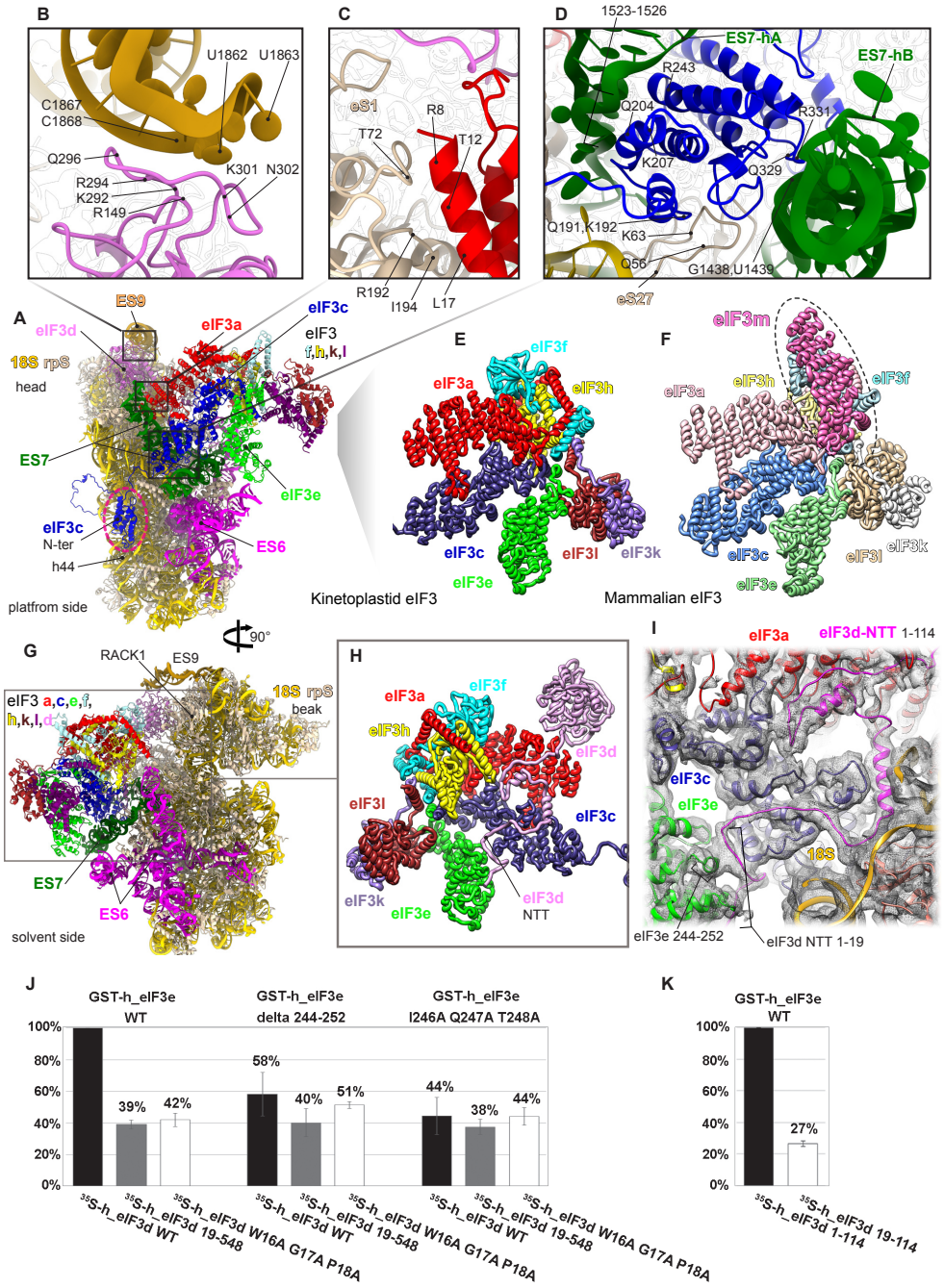

Fig. 3. Kinetoplastidian eIF3 and its unique binding site. (A) The overall view of the atomic model of the 43S PIC from the platform side. The conserved helical domain of the eIF3c-NTD is encircled with dashed line, eIF3 subunits are colored variably and 18S RNA in yellow. (**B**) Close-up view of the interaction between the ES9S (honey yellow) and eIF3d (in pink). (**C**) Close-up view of the interaction between eIF3a (in red) and eS1 (in beige) (**D**) Close-up view of the interaction between the ES7S (in green) and eIF3c (in blue). (**E**) Cartoon representation of the atomic model of the kinetoplastidian eIF3 structural core. (**F**) Cartoon representation of an atomic model of the mammalian eIF3 structural core. Subunit eIF3m, which is not encoded by kinetoplastids, is marked by dashed oval. (**G**) The overall view of an atomic model of the 43S PIC from the solvent side. (**H**) Cartoon representation of the atomic model of the kinetoplastidian eIF3 focused on the eIF3d-NTT (in pink). (**I**) Fitting of the eIF3d-NTT model into its cryo-EM

map. (**J, K**) binding analysis between human eIF3d and GST-eIF3e, expressed in plots showing normalized data from three different dilutions of GST-proteins (see Fig. S10A).

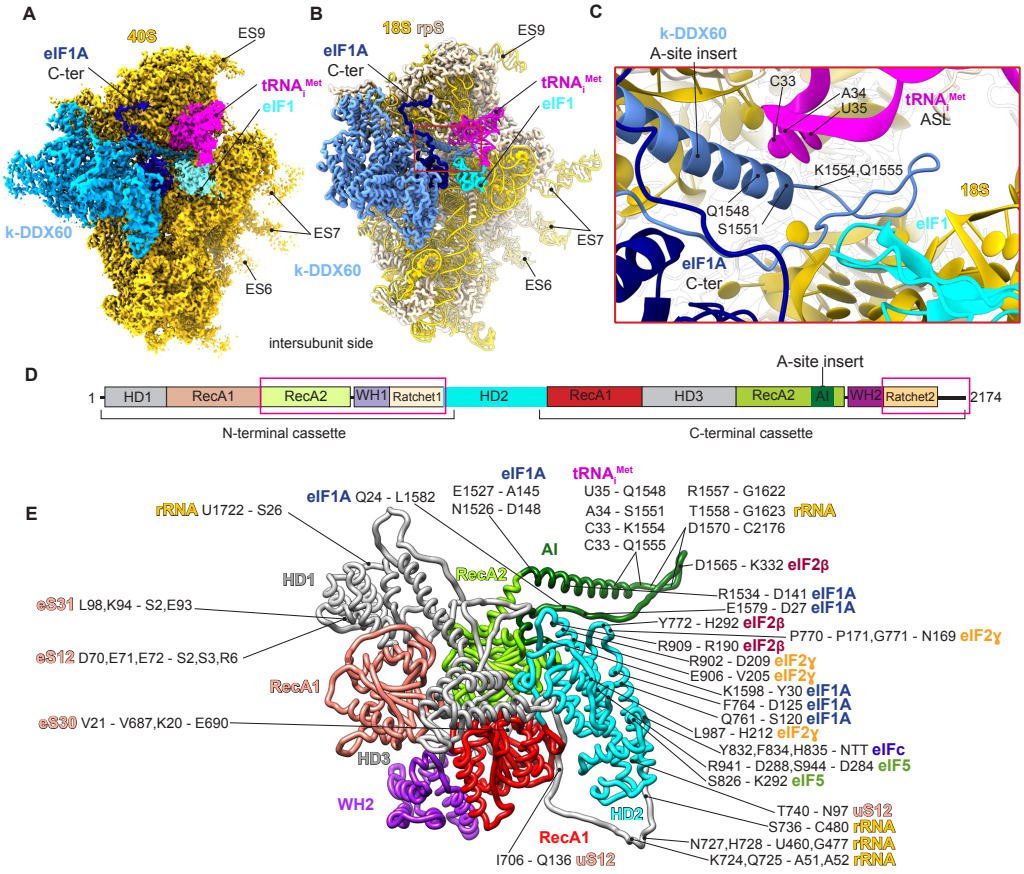

Fig. 4. k-DDX60 structure and interactions within the 43S PIC. (**A**) The cryo-EM structure of the *T. cruzi* 43S PIC highlighting k-DDX60 (colored in dark turquoise). eIF 2, 3 and 5 densities were removed for clarity (**B**) Cartoon representation of a partial atomic model of the *T. cruzi* 43S PIC. (**C**) A close-up view of the k-DDX60 A-site insert showing its interaction with the anticodon stem loop (ASL). (**D**) Schematic representation of the k-DDX60 domains. Pink boxes indicate the domains that couldn't be modeled because of their lower local resolution (See Fig. S4). (**E**) Cartoon representation of the atomic model of the k-DDX60 and its interactions with the 43S PIC color-coded in accord with its schematic representation in the panel D.

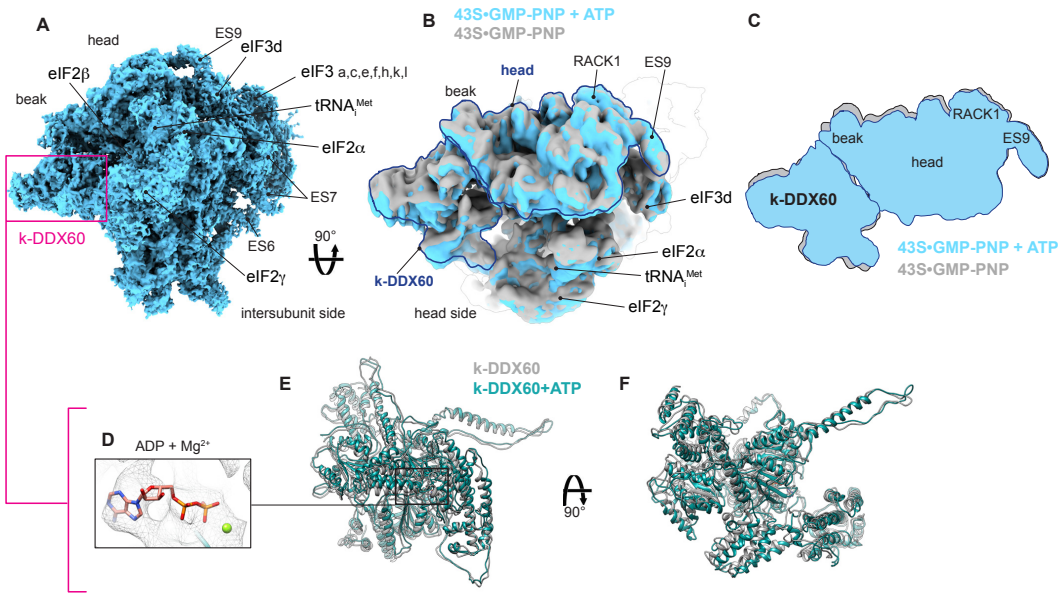

Fig. 5. Global conformational rearrangement of the 43S PIC driven by ATP binding to k-DDX60. (**A**) Cryo-EM reconstruction of the *T. cruzi* 43S PIC in the presence of ATP. (**B**) Superposition of the cryo-EM reconstructions of the 43S•GMP-PNP (in grey) and the 43S•GMP-PNP supplemented with ATP (in turquoise), seen from the top. (**C**) Schematic representation of the structural rearrangements induced by ATP. (**D**) A close-up view of the ATP binding pocket within the RecA1 domain of the C-terminal cassette of k-DDX60. (**E, F**) Superimposition of the k-DDX60 atomic model from the cryo-EM structure of the 43S•GMP-PNP and 43S•GMP-PNP supplemented with ATP presented in two different orientations.

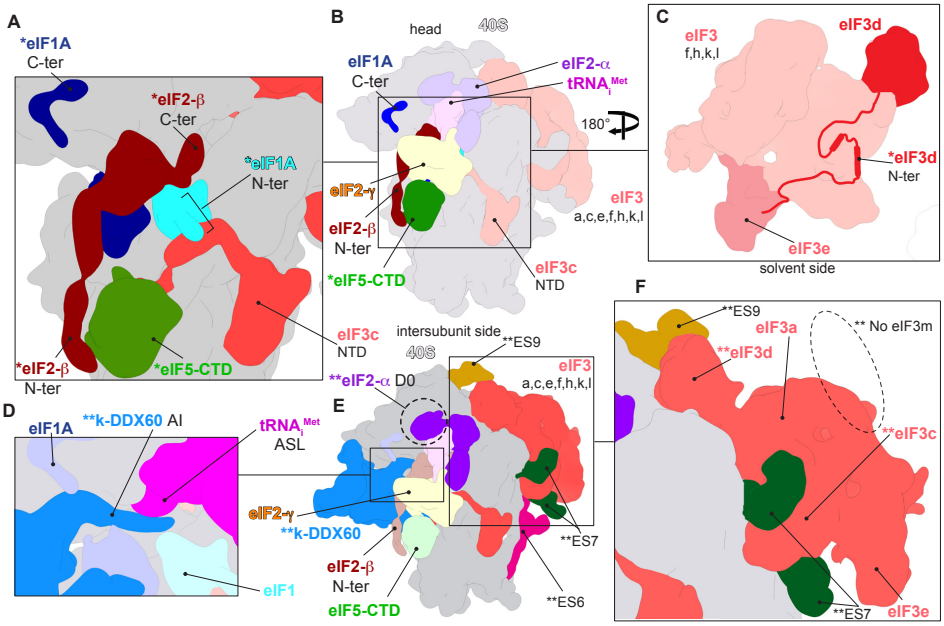

Fig. 6. Novel eukaryotic-conserved and trypanosomatid-specific features of the 43S PIC revealed in our work. (**A**) Schematic model representing a close up view on the N-terminal tails of eIF 1, 1A, 2β, eIF5-CTD and eIF3c-NTD, all conserved among eukaryotes and revealed in the current work. The ternary complex was removed for clarity. (**B**) Schematic model representing the 43S PIC from the intersubunit side. The novel features revealed in our work are colored in brighter colors. (**C**) Schematic model representing a solvent side view of eIF3 highlighting the conserved N-terminal tail of eIF3d and its main interactions with eIF3e, revealed in the current work. (**D**) Schematic model representing a close-up view on the A-site Insert of k-DDX60 and its interaction with the anti-codon stem-loop (ASL). (**E**) Schematic model representing the *T. cruzi* 43S PIC from the intersubunit side. Dashed circle highlight the kinetoplastid-specific domain eIF2α, dubbed here "D0". The kinetoplastid-specific features revealed in our work are colored in brighter colors. (**F**) Schematic model representing a close-up view on the kinetoplastidian eIF3 showing its specific interaction with ES7^S and ES9^S, and the absence of the eIF3m subunit. *=Conserved features among eukaryotes revealed in our work. **=Kinetoplastid-specific features revealed in our work.

Supplementary Materials

Unique features of mRNA translation initiation in trypanosomatids

Anthony Bochler, Jailson Brito Querido, Terezie Svobodová, Heddy Soufari, Mayara Lucia Del Cistia, Lauriane Kuhn, Aline Rimoldi Ribeiro, Leoš Shivaya Valášek and Yaser Hashem*

Correspondence to: yaser.hashem@u-bordeaux.fr

Materials and Methods

Kinetoplastids Cultures

Trypanosoma cruzi epimastigoes (Y strain - TcII) were grown at 28°C in liver infusion tryptose (LIT) medium, supplemented with 10% heat-inactivated fetal bovine serum. *Leishmania tarentolae* strain T7-TR (Jena Bioscience) were grown at 26°C in brain-heart infusion-based medium (LEXSY BHI; Jena Bioscience), supplemented with Nourseothricin and LEXSY Hygro (Jena Bioscience), hemin and penicillin-streptomycin.

48S Initiation Complex Purification

T. cruzi and *L. tarentolae* 48S initiation complexes were grown to a density $3 \cdot 10^6$ per mL and $2.5 \cdot 10^6$ per mL, for *T. cruzi* and *L. tarantolae*, respectively, in 200 mL flasks in culture medium. The parasites were harvested, put in buffer I (20 mM HEPES-KOH pH 7.4, 100 mM KOAc, 4 mM Mg $(OAc)_2$, 2 mM DTT, EDTA free protease inhibitor cocktail and RNasin inhibitor) and subjected to lysis by freeze-thaw cycles. After the centrifugation at 12,000 *g* for 30 min at 4°C, the supernatant was incubated in the presence of 2 mM GMP-PNP (the non-hydrolyzable analog of GTP) for 10 min at 28°C. The supernatant was layered onto 10-30 % (w/v) sucrose gradients and centrifuged (37 000 rpm, 5h30 min, 4°C) using an SW41 Ti rotor (Beckman-Coulter). The fractions containing 48S ICs were collected and pooled according the UV absorbance profile. Buffer was exchanged by precipitating ribosomal complexes and re-suspending them in sucrose-free buffer II (10 mM HEPES-KOH pH 7.4, 50 mM KOAc, 10 mM NH_4Cl, 5 mM $Mg(OAc)_2$, and 2 mM DTT). For the ATP suuplemented 43S PIC, the protocol above was repeated for *T. cruzi* with an addition of 2mM of ATP.

Cryo-EM Grid preparation

Grid preparation: 4 µL of the sample at a concentration of 90 nM was applied onto the Quantifoil R2/2 300-mesh holey carbon grid, which had been coated with thin carbon film (about 2nm) and glow-discharged. The sample was incubated on the grid for 30 sec and then blotted with filter paper for 1.5 sec in a temperature and humidity controlled Vitrobot Mark IV (T = 4°C, humidity 100%, blot force 5) followed by vitrification in liquid ethane.

Cryo-EM Image acquisition

Data collection was performed on a spherical aberration corrected Titan Krios S-FEG instrument (FEI Company) at 300 kV using the EPU software (Thermo Fisher Company) for automated data acquisition. Data were collected at a nominal under focus of -0.6 to -4.5 µm at a magnification of 127,272 X yielding a pixel size of 1.1 Å. Micrographs were recorded as movie stack on a Gatan Summit K2 direct electron detector, each movie stack were fractionated into 20 frames for a total exposure of an electron dose of 30 $\bar{e}/Å^2$.

Image processing

Drift and gain correction and dose weighting were performed using MotionCor2 *(47)*. A dose weighted average image of the whole stack was used to determine the contrast transfer function with the software Gctf *(48)*. The following process has been achieved using RELION 3.0 *(49)*. Particles were picked using a Laplacian of gaussian function (min diameter 300 Å, max diameter 320 Å). Particles were then extracted with a box size of 360 pixels and binned three fold for 2D classification into 200 classes, yielding 202,920 particles presenting 40S-like shape. These particles were then subjected to 3D classification into 10 classes. Two subclasses depicting high-resolution and 48S features have been selected for a second round of classification into two classes. One class ended as a 48S complex (12910 particles) and a second as a 43S+DDX60 complex (33775 particles). Refinement of the 43S-DDX60 complex yielded an average resolution of 3.3 Å. The 48S class was not analyzed any further. Determination of the local resolution of the final density map was performed using ResMap *(50)*.

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 (Brito Querido et al., 2017). Each sample was further analyzed by nanoLC-MS/MS on a QExactive+ mass spectrometer coupled to an EASY-nanoLC-1000 (Thermo-Fisher Scientific, USA). 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 relatively quantify each protein (Spectral Count relative quantification). Proline was further used to align the Spectral Count values across all samples. The whole MS dataset was then normalized. The mass spectrometric data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXDxxxx.

Volcano plot

Volcano plot presented in Fig. 1 was obtained after manual validation of the results. For that end, we only consider proteins that present at least 5 spectra. Further validation was performed by analysing the pre-initiation complex after further purification step using size exclusion chromatography.

Model building and refinement

The atomic model of the preinitiation complex 48S from Trypanosoma *cruzi* was built using the modelling softwares Chimera *(51)*, Coot *(52)*, Phenix *(53)* and VMD *(54)*.
The previous 40S structure of Trypanosoma *cruzi (5)* (PDBID : 5OPT) was used to build the core of the initiation complex containing the small subunit ribosomal RNA and proteins. The head required a rotation to fit the new structure.
The ternary complex (tRNA, eIF2α, eIF2γ), eIF2β, eIF1a and eIF1 were thread from the translation initiation complex of yeast *(24)* (PDBID : 3JAQ).
DDX60-like starting point was the recA domains from the human helicase protein Brr2 *(55)* (PDBID : 4F93). The remaining domains of DDX60-like was built *ab initio* using Coot modelling tools and Chimera "build structure" tools with the help of sympred *(56)* for secondary structure prediction and the homology modelling webservices Swissmodel *(57)* and phyre2 *(58)*.
eIF3 was thread from the already published eIF3 from human *(9)* (PDBID : 5A5T), subunit m was deleted since it's not present in Kinetoplastid and rearrangements of the nearby subunits were made. Subunit d was thread from the eIF3d crystal structure of Nasonia vitripennis *(26)* (PDBID : 5K4B) and the N-terminal tail was built in Chimera.
eIF5 Cter-domain was thread from the eIF5 crystal from human *(59)* (PDBID : 2IU1).
The global atomic model was refined using the Molecular Dynamic Flexible Fitting *(60)* then the geometry parameters were corrected using PHENIX real space refine for proteins and erraser *(61)* for RNA.

GST pulldown assay

Glutathione S-transferase (GST) pull down experiments with GST fusions and *in vitro* synthesized ^{35}S-labeled polypeptides were conducted as described previously (PMID:11179233). Briefly, individual GST-fusion proteins were expressed in *Escherichia coli* (BL-21 Star DE3 or BL21 Rosett2 DE3). Bacterial culture was grown at 37°C in the LB medium to OD 0.6-0.8 and the synthesis of GST-fusion proteins were induced by the addition of 1mM IPTG. After 2 hr of shaking at 37°C or overnight at 16°C the cells were harvested, resuspended in a Phosphate-buffered saline (PBS), and subjected to mechanical lysis with a subsequent agitation in the presence of 1-1.5% Triton X-100 for 30 min at 4°C. The GST-proteins were then immobilized on glutathione sepharose beads (GE Healthcare, cat # GE17-0756-01) from the pre-cleaned supernatant, followed by three washing steps with the 1 ml of phosphate buffered saline.^{35}S-labeled polypeptides were produced *in-vitro* by the TnT® Quick Coupled Transcription/Translation System (Promega cat # L1170) according to the vendor's instructions.

To examine the binding, individual GST fusions were incubated with ^{35}S-labeled proteins at 4°C for 2 h in buffer B (20mM HEPES (pH 7,5), 75mM KCl, 0,1mM EDTA, 2,5mM $MgCl_2$, 0,05% IGEPAL, 1mM DTT). For experiments requiring more stringent conditions the buffer B was supplement with 1% fat free milk. Subsequently, the beads were washed three times with 1 ml of phosphate buffered saline and interacting proteins were separated by SDS-PAGE. Gels were first stained with Gelcode Blue stain reagent (Thermofisher, cat # 24592) and then subjected to autoradiography.

Quantification of binding experiments was done by the Quantity One software. The data was generated as an adjusted volume with the local background subtraction and linear regression methods. The data for each ^{35}S-labeled protein was first normalized to its input and the percentage of input binding was then calculated. The resulting data was subsequently normalized to its corresponding control (for Fig. 3J: ^{35}S-eIF3d WT – GST-eIF3e WT; and for Fig. 3K : ^{35}S-eIF3d 1-114 – GST-eIF3e WT) and means

from three different dilutions of GST-fusions were calculated; errors bars indicate standard deviation.

				BASIC Spectral Count (# spectra)	
				BEFORE Gel Filtration	AFTER Gel Filtration
Q4E5Z1|Q4E5Z1_TRYCC	DDX60	Uncharacterized protein OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053508153.1050	10	263	96
Q4DLI2|Q4DLI2_TRYCC	ABCE1	Ribonuclease L inhibitor, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.10470535086	3	103	31

40S ribosomal proteins:

			BASIC Spectral Count (# spectra)		
				BEFORE Gel Filtration	AFTER Gel Filtration
accession		description	40S	43S	43S
Q4D5P4|Q4D5P4_TRYCC		40S ribosomal protein S4 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053509683.117	131	131	93
Q4DTN2|Q4DTN2_TRYCC		Activated protein kinase C receptor, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.10	100	96	48
Q4E0Q3|Q4E0Q3_TRYCC		40S ribosomal protein S5, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053506	65	51	40
Q4DZ41|RS3A2_TRYCC		40S ribosomal protein S3a-2 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511001.9	98	84	46
Q4E093|Q4E093_TRYCC		40S ribosomal protein S18, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	75	73	64
Q4DSU0|Q4DSU0_TRYCC		40S ribosomal protein S6 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053510769.49 P	89	72	58
Q4CLU9|Q4CLU9_TRYCC		40S ribosomal protein S8 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511069.20 P	66	60	46
Q4D4L4|Q4D4L4_TRYCC		40S ribosomal protein S11, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	58	47	39
Q4D6I5|Q4D6I5_TRYCC		40S ribosomal protein S14, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705345	60	61	37
Q4CUL0|Q4CUL0_TRYCC		40S ribosomal protein S3, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705343	81	72	37
Q4D6N9|Q4D6N9_TRYCC		Ribosomal protein S19, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053504	39	36	36
Q4DY30|Q4DY30_TRYCC	KSRP	RNA-binding protein, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053511727	79	72	29
Q4D4S1|Q4D4S1_TRYCC		40S ribosomal protein S9, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053504	38	38	28
Q4CUC8|Q4CUC8_TRYCC		Ribosomal protein S7, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053506593	84	80	25
Q4CQU0|Q4CQU0_TRYCC		40S ribosomal protein SA OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053503719.20 P	65	58	22
Q4D916|Q4D916_TRYCC		40S ribosomal protein S16, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	48	52	19
Q4E0N6|Q4E0N6_TRYCC		40S ribosomal protein S15a, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705035	37	34	15
Q4DIZ9|Q4DIZ9_TRYCC		40S ribosomal protein S2, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705360	80	74	27
Q4CXN0|Q4CXN0_TRYCC		Ubiquitin/ribosomal protein S27a, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047	52	40	14
Q4DTX6|Q4DTX6_TRYCC		Ribosomal protein S25, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350410	46	44	8
Q4DK39|Q4DK39_TRYCC		40S ribosomal protein S17, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	58	57	16
Q4E088|Q4E088_TRYCC		40S ribosomal protein S10, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	52	54	22
Q4DW69|Q4DW69_TRYCC		40S ribosomal protein S12 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053508231.20	34	39	13
Q4CXV6|Q4CXV6_TRYCC		40S ribosomal protein S33, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	34	31	17
Q4DTQ1|Q4DTQ1_TRYCC		40S ribosomal protein S23, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	33	28	28
Q4D6H7|Q4D6H7_TRYCC		Ribosomal protein S20, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350847	34	28	16
Q4CWD6|Q4CWD6_TRYCC		40S ribosomal protein S13, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705351	32	30	18
Q4DN73|Q4DN73_TRYCC		40S ribosomal protein S27, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	21	17	25
Q4DW38|Q4DW38_TRYCC		40S ribosomal protein S24 OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053507681.150	30	26	15
Q4CMS5|Q4CMS5_TRYCC		Ribosomal protein S29, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705351101	17	15	13
Q4DGZ5|Q4DGZ5_TRYCC		40S ribosomal protein S15, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705351	23	20	11
Q4CYE4|Q4CYE4_TRYCC		Ribosomal protein S26, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350380	21	18	12
Q4E3L9|Q4E3L9_TRYCC		40S ribosomal protein S21, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705351	24	18	7
Q4DA48|Q4DA48_TRYCC		40S ribosomal protein S30, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	2	5	

Initiation factors:

			BASIC Spectral Count (# spectra)		
				BEFORE Gel Filtration	AFTER Gel Filtration
accession		description	40S	43S	43S
Q4DL69|Q4DL69_TRYCC	eIF3a	Uncharacterized protein OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053508919.140 P	86	129	50
Q4DSL1|Q4DSL1_TRYCC	eIF3b	Translation initiation factor, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705351	95	159	41
Q4E3G1|Q4E3G1_TRYCC	eIF3c	Eukaryotic translation initiation factor 3 subunit 8, putative OS=Trypanosoma cruzi (strain CL Brener)	63	96	16
Q4D7F2|Q4D7F2_TRYCC	eIF3e	Eukaryotic translation initiation factor 3 subunit E OS=Trypanosoma cruzi (strain CL Brener) GN=Tc0	60	103	24
Q4E620|Q4E620_TRYCC	eIF2 alpha	Elongation initiation factor 2 alpha subunit, putative OS=Trypanosoma cruzi (strain CL Brener) GN=T	5	105	22
Q4DCN0|Q4DCN0_TRYCC	eIF3d	Eukaryotic translation initiation factor 3 subunit 7-like protein, putative OS=Trypanosoma cruzi (strain	72	113	16
Q4D452|Q4D452_TRYCC	eIF3i	Eukaryotic translation initiation factor 3 subunit I OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00	40	69	14
Q4D5W3|Q4D5W3_TRYCC	eIF3l	Eukaryotic translation initiation factor 3 subunit L OS=Trypanosoma cruzi (strain CL Brener) GN=Tc0	51	83	25
Q4E3S5|Q4E3S5_TRYCC	eIF3h	Homology with eIF3H (InterPro), Uncharacterized protein OS=Trypanosoma cruzi (strain CL Brener)	36	58	8
Q4CUG4|Q4CUG4_TRYCC	eIF3g	Eukaryotic translation initiation factor 3 subunit G OS=Trypanosoma cruzi (strain CL Brener) GN=Tc0	43	70	19
Q4CSE1|Q4CSE1_TRYCC	eIF5	Eukaryotic translation initiation factor 5, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00	19	115	49
Q4DDK1|Q4DDK1_TRYCC	eIF3k	Homology with eIF3K (InterPro), Uncharacterized protein OS=Trypanosoma cruzi (strain CL Brener)	18	29	9
Q4DH88|Q4DH88_TRYCC	eIF2 beta	Translation initiation factor, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.104705350	7	45	25
Q4DQZ2|Q4DQZ2_TRYCC	eIF3f	Uncharacterized protein OS=Trypanosoma cruzi (strain CL Brener) GN=Tc00.1047053510089.200 P	46	68	24
Q4CPV7|Q4CPV7_TRYCC	eIF2 gamma	Eukaryotic translation initiation factor 2 subunit, putative OS=Trypanosoma cruzi (strain CL Brener) G	5	62	7
Q4CQB1|Q4CQB1_TRYCC	eIF1A	Eukaryotic translation initiation factor 1A, putative (Fragment) OS=Trypanosoma cruzi (strain CL Bre	4	25	4
Q4DM75|Q4DM75_TRYCC	eIF1	Protein translation factor SUI1 homolog, putative OS=Trypanosoma cruzi (strain CL Brener) GN=Tc0	10	18	5

Fig. S1. Mass-spectrometry analysis of the *T. cruzi* 43S PIC. Composition of the *T. cruzi* 43S PIC in 40S ribosomal proteins and initiation factors. K-DDX60 and ABCE1 were singled out. The analysis compares the 43S related fractions without (labeled 40S) and with GMP-PNP (labeled 43S), before and after Gel-filtration. Accessions, description and spectral counts are indicated for each fraction.

	Name		Spectral Count IC
tr|E9ACL4|E	DDX60	Uncharacterized protein OS=Leishmania major GN=LMJF_03_0690 PE=4 SV=1	111
tr|Q4QCE4|I	ABCE1	Putative ATP-binding cassette protein subfamily E,member 1 OS=Leishmania major GN=ABCI	101
40S ribosomal proteins:			
	Name		Spectral Count
accession		description	IC
tr|Q868B1|C		40S ribosomal protein S5 OS=Leishmania major GN=LMJF_11_0960 PE=4 SV=1	188
tr|Q4Q216|C		Putative ubiquitin/ribosomal protein S27a OS=Leishmania major GN=LMJF_36_0600 PE=4 SV	265
tr|Q4Q1Y2|C		Putative 40S ribosomal protein S18 OS=Leishmania major GN=LMJF_36_0940 PE=3 SV=1	122
tr|Q4QG31|I		40S ribosomal protein S4 OS=Leishmania major GN=RS4 PE=2 SV=1	299
tr|Q4Q8H1|C		40S ribosomal protein S14 OS=Leishmania major GN=LMJF_28_0960 PE=3 SV=1	155
tr|Q4QC89|C		Putative 40S ribosomal protein S23 OS=Leishmania major GN=LMJF_21_1060 PE=3 SV=1	90
tr|Q4Q4A0|C		Putative 40S ribosomal protein S3 OS=Leishmania major GN=LMJF_15_0950 PE=4 SV=1	99
sp|P25204|R		40S ribosomal protein S8 OS=Leishmania major GN=RPS8A PE=3 SV=1	108
sp|Q9NE83		40S ribosomal protein S6 OS=Leishmania major GN=RPS6 PE=3 SV=1	175
tr|Q4Q817|C		Putative ribosomal protein S29 OS=Leishmania major GN=LMJF_28_2205 PE=4 SV=1	62
tr|Q4Q1V1|C		Putative 40S ribosomal protein S9 OS=Leishmania major GN=LMJF_36_1250 PE=2 SV=1	98
tr|Q4Q5P0|C		40S ribosomal protein S2 OS=Leishmania major GN=LMJF_32_0450 PE=3 SV=1	144
tr|Q4Q3M1|I		Putative 40S ribosomal protein S13 OS=Leishmania major GN=LMJF_19_0390 PE=3 SV=1	83
tr|Q4QH01|C		Putative 40S ribosomal protein S21 OS=Leishmania major GN=LMJF_11_0760 PE=4 SV=1	39
sp|Q4FX73		40S ribosomal protein S3a OS=Leishmania major GN=LmjF.35.0400 PE=2 SV=1	288
tr|Q4Q8G4|I		Putative ribosomal protein S20 OS=Leishmania major GN=LMJF_28_1010 PE=3 SV=1	99
tr|Q4Q7P0|C		Putative 40S ribosomal protein S30 OS=Leishmania major GN=LMJF_30_0670 PE=4 SV=1	36
tr|Q4QCN7|I		Putative 40S ribosomal protein S11 OS=Leishmania major GN=LMJF_20_1650 PE=3 SV=1	153
sp|Q4Q0Q0		40S ribosomal protein SA OS=Leishmania major GN=LmjF36.5010 PE=3 SV=1	145
tr|E9AEE8|E		40S ribosomal protein S19-like protein OS=Leishmania major GN=LMJF_29_2860 PE=4 SV=1	129
tr|Q4Q931|C		Putative 40S ribosomal protein S33 OS=Leishmania major GN=S33-1 PE=4 SV=1	102
tr|Q4Q1X7|C		Putative 40S ribosomal protein S10 OS=Leishmania major GN=LMJF_36_0980 PE=4 SV=1	101
tr|Q4QG97|I		40S ribosomal protein S12 OS=Leishmania major GN=LMJF_13_0570 PE=3 SV=1	93
tr|Q4QGW3		Putative 40S ribosomal protein S15A OS=Leishmania major GN=LMJF_11_1190 PE=3 SV=1	84
tr|Q4Q9A5|C		Putative 40S ribosomal protein S16 OS=Leishmania major GN=LMJF_26_0880 PE=2 SV=1	79
tr|Q4Q806|C		Putative 40S ribosomal protein S17 OS=Leishmania major GN=LMJF_28_2555 PE=3 SV=1	42
tr|Q4Q140|C		Putative 40S ribosomal protein S27-1 OS=Leishmania major GN=LMJF_36_3750 PE=3 SV=1	63
tr|Q4Q8L6|C		Putative ribosomal protein S26 OS=Leishmania major GN=LMJF_28_0540 PE=4 SV=1	34
tr|Q4Q1D2|C		40S ribosomal protein S24 OS=Leishmania major GN=S24E-2 PE=3 SV=1	120
tr|Q4Q3G4|I		Ribosomal protein S25 OS=Leishmania major GN=S25 PE=4 SV=1	91
tr|O43943|C	RACK1	LACK OS=Leishmania major PE=4 SV=1	58
tr|Q4Q5K7|C	KSRP	Putative RNA binding protein OS=Leishmania major GN=LMJF_32_0750 PE=4 SV=1	56
tr|Q4QBV0|I		Putative 40S ribosomal protein S15 OS=Leishmania major GN=LMJF_22_0420 PE=3 SV=1	31
tr|E9AC32|E		Putative ribosomal protein S7 OS=Leishmania major GN=LMJF_01_0410 PE=4 SV=1	27
Initiation factors:			
	Name		C Spectral Count (# spe
accession		description	IC
tr|Q4QEJ8|C	eIF3a	Uncharacterized protein OS=Leishmania major GN=LMJF_17_0010 PE=4 SV=1	278
tr|Q4QE62|C	eIF3b	Putative translation initiation factor OS=Leishmania major GN=LMJF_17_1290 PE=4 SV=1	175
tr|Q4Q6Y6|C	eIF3d	Eukaryotic translation initiation factor 3 subunit 7-like protein OS=Leishmania major GN=LMJF	125
tr|Q4Q833|C	eIF3e	Eukaryotic translation initiation factor 3 subunit E OS=Leishmania major GN=LMJF_28_2310 P	84
tr|Q4Q253|C	eIF3l	Eukaryotic translation initiation factor 3 subunit L OS=Leishmania major GN=LMJF_36_0250 P	91
tr|Q4Q127|C	eIF3i	Eukaryotic translation initiation factor 3 subunit I OS=Leishmania major GN=LMJF_36_3880 P	79
tr|E9ACP3|E	eIF2 alpha	Putative elongation initiation factor 2 alpha subunit OS=Leishmania major GN=LMJF_03_0980	76
tr|Q4QIM7|C	eIF3h	Uncharacterized protein OS=Leishmania major GN=LMJF_07_0640 PE=4 SV=1	76
tr|Q4Q3H3|C	eIF5	Putative eukaryotic translation initiation factor 5 OS=Leishmania major GN=LMJF_34_0350 PE	75
tr|Q4Q9T0|C	eIF3f	Uncharacterized protein OS=Leishmania major GN=LMJF_25_1610 PE=4 SV=1	67
tr|Q4Q055|C	eIF3c	Putative eukaryotic translation initiation factor 3 subunit 8 OS=Leishmania major GN=LMJF_36	62
tr|Q4Q557|C	eIF3k	Uncharacterized protein OS=Leishmania major GN=LMJF_32_2180 PE=4 SV=1	59
tr|Q4QHR7|I	eIF2 gamma	Putative eukaryotic translation initiation factor 2 subunit OS=Leishmania major GN=LMJF_09_	49
tr|Q4Q2S5|C	eIF3g	Eukaryotic translation initiation factor 3 subunit G OS=Leishmania major GN=LMJF_34_2700 P	46
tr|Q4QAL1|C	eIF1A	Putative translation factor sui1 OS=Leishmania major GN=LMJF_24_1210 PE=4 SV=1	33
tr|Q4QIB4|C	eIF2 beta	Translation initiation factor-like protein OS=Leishmania major GN=LMJF_08_0550 PE=4 SV=1	29
tr|Q4QF06|C	eIF1A	Putative eukaryotic translation initiation factor 1A OS=Leishmania major GN=LMJF_16_0140 P	27

Fig. S2. Mass-spectrometry analysis of the *L. Tarentolae* 43S PIC. Composition of the *L. Tarentolae* 43S PIC in 40S ribosomal proteins and initiation factors. K-DDX60 and ABCE1 were singled out. The analysis of the 43S related fraction was made after supplementation with GMP-PNP (IC), before Gel-filtration. Accessions, description and spectral counts are indicated.

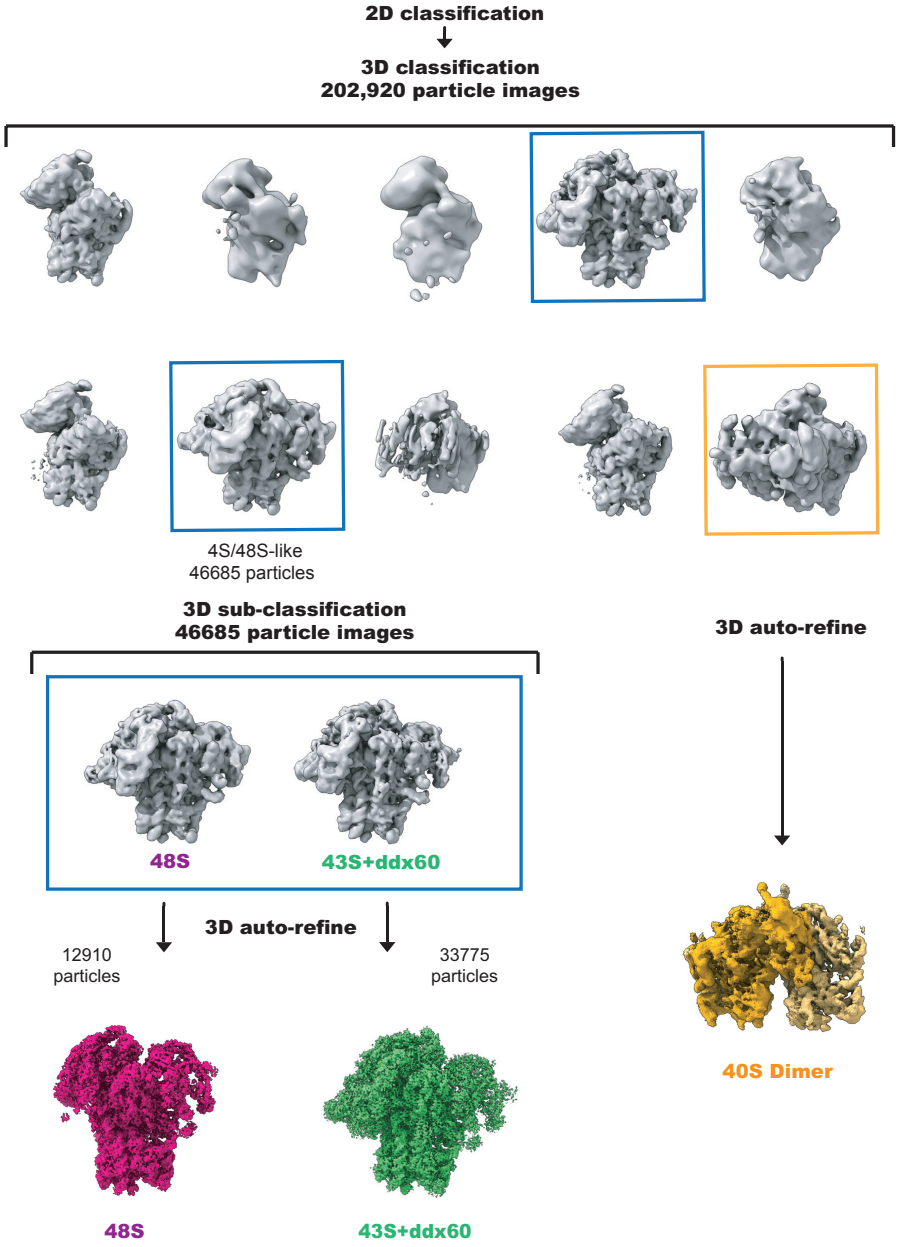

Fig. S3. Cryo-EM particle sorting and refinement of the *T. cruzi* 43S PIC. 2D classification of the 43S PIC particles yielded ~200 000 40S-like particles, after which a run of 3D classification (10 classes) allowed to single out 43S/48S ICs and 40S-dimers. A secondary run of 3D classification allowed the sorting of the 43S PIC particles that generated the main reconstruction analyzed in this study.

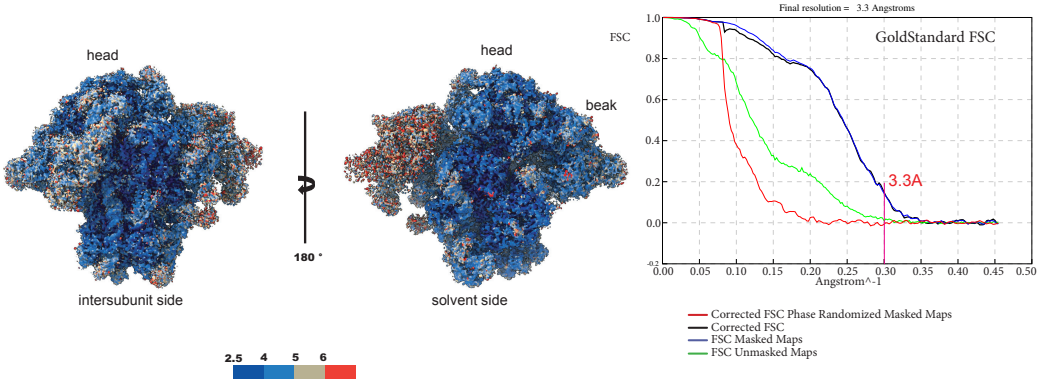

Fig. S4. Cryo-EM average and local resolution of the *T. cruzi* 43S PIC. The local resolution varies mainly on eIF3 (ranging from ~3 to ~6 Å), while is varies less on the rest of the structure (ranging from ~2.5 to ~3.5 Å for the 40S, k-DDX60, eIFs 1, 1A and 2b, and from ~3 to ~5 Å for eIFs 2a, 2g and 5). The average resolution was measured after applying a soft-edge mask of the 43S PIC shape filtered to 15Å and extended by 3 pixels.

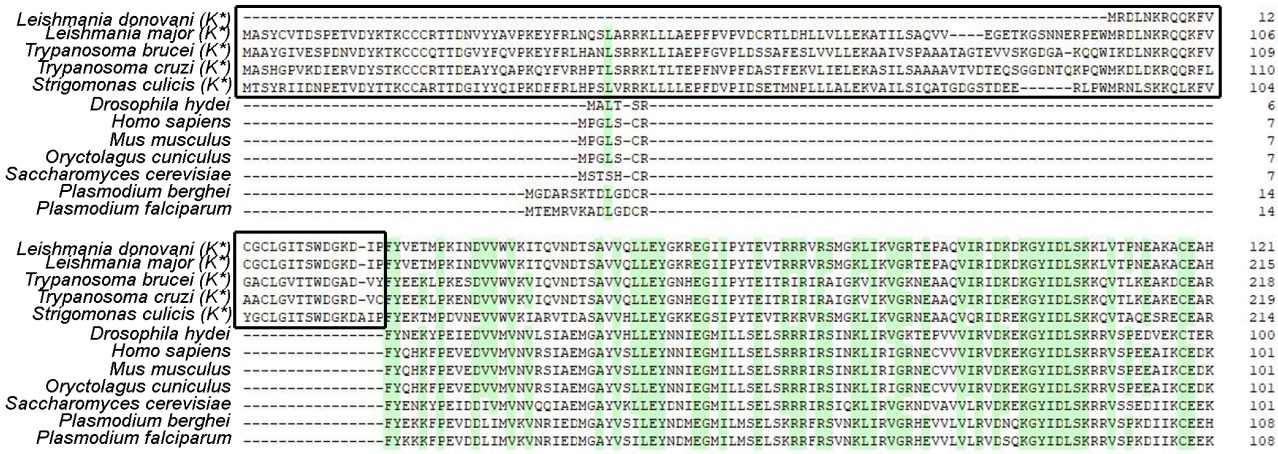

Fig. S5. Multiple sequence alignment of the eIF2a NTD among eukaryotes. Protein sequence alignment of eIF2α from various eukaryotic organisms was generated by Clone Manger (MultiWay, scoring matrix: Blosum 62). The Kinetoplastida order species are labeled with K*. The kinetoplastidian-specific eIF2α N-terminal domain insertion is marked with a black box. Areas of high matches (60%) are shaded in green. The individual species with the NCBI Reference Sequence numbers or TriTrypDB numbers are as follows: [*Trypanosoma cruzi*] PWV18423.1, [*Trypanosoma brucei*] Tb927.3.2900, [*Leishmania donovani*] AAQ02666.1, [*Leishmania major*] LmjF.03.0980, [*Strigomonas culicis*] EPY26930.1, *[Plasmodium falciparum* NF54] PKC42156.1, [*Plasmodium berghei* ANKA] VUC53995.1, [*Saccharomyces cerevisiae*] ONH75775.1, [*Oryctolagus cuniculus*] XP_002719561.1, [*Mus musculus*] NP_080390.1, [*Drosophila hydei*] XP_023166950.2, [*Homo sapiens*] NP_004085.1

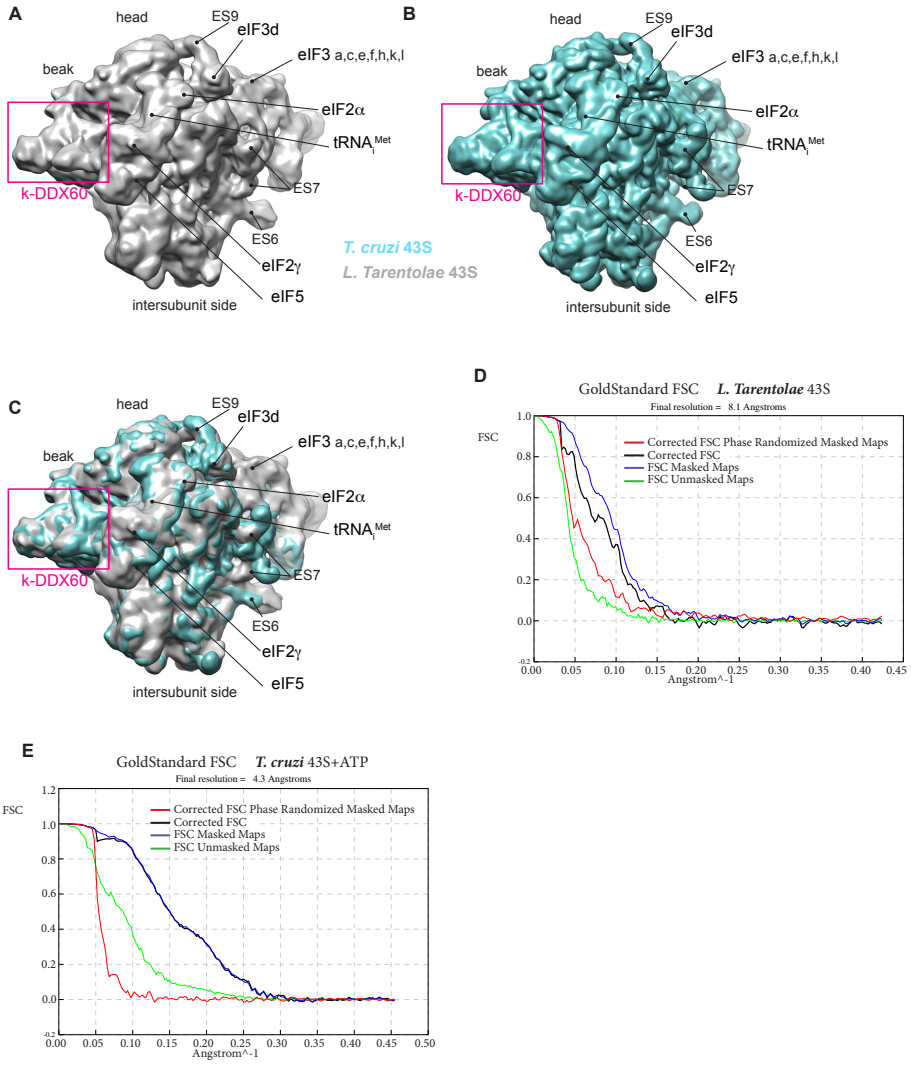

Fig. S6. Cryo-EM reconstruction of *L. tarentolae* 43S PIC compared to *T. cruzi* and average resolution. (**A**) Cryo-EM reconstructions of the *L. tarentolae* 43S PIC. (**B**) Cryo-EM reconstructions of the *T. cruzi* 43S PIC filtered at 8Å. (**C**) Superimposition of (A) and (B). (**D**) Average resolution (8.1Å) of the *L. tarentolae* 43S PIC reconstruction. (**E**) Average resolution (4.3Å) of the cryo-EM reconstruction from the *T. cruzi* 43S complexes supplemented with ATP.

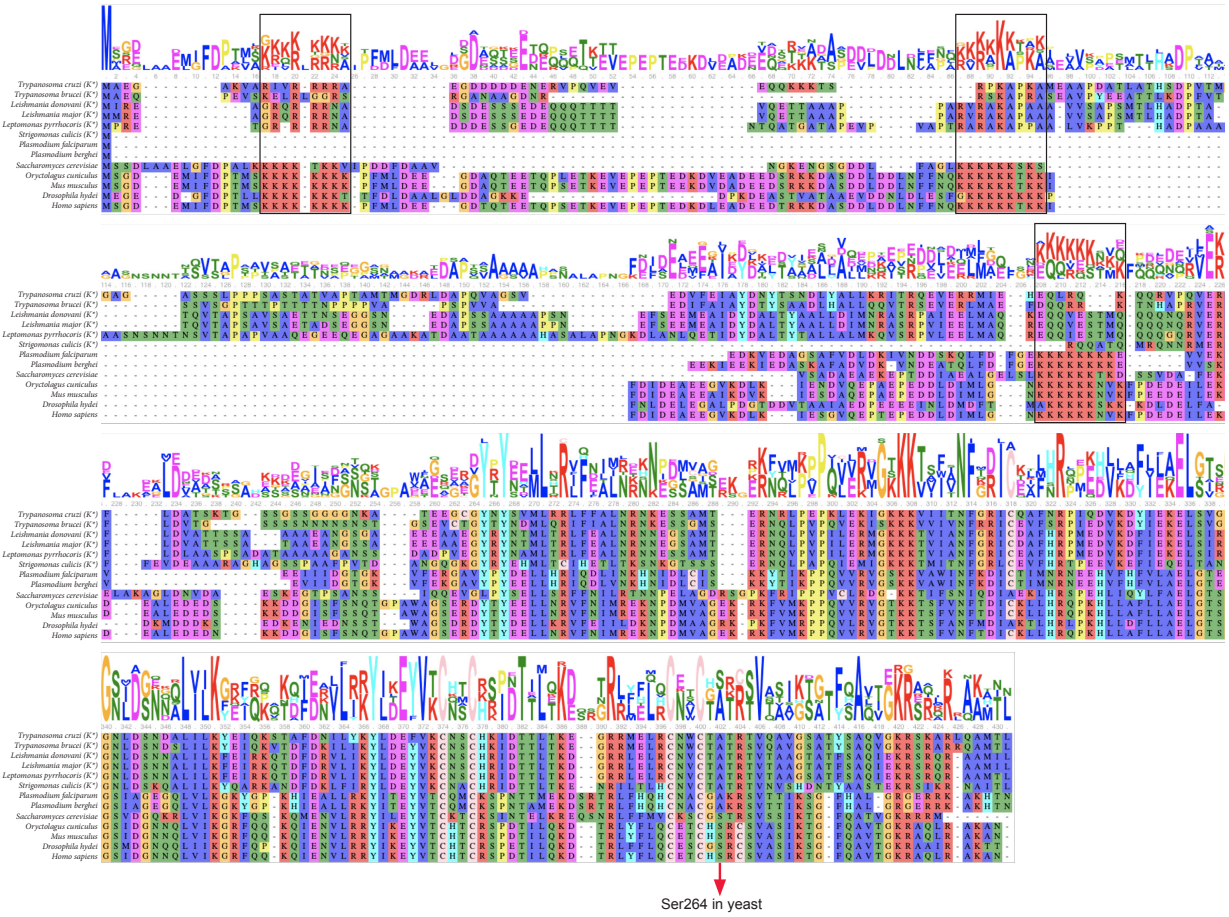

Fig. S7. Multiple sequence alignment of eIF2β among eukaryotes. Protein sequence alignment of eIF2β protein from various eukaryotic organisms. The Kinetoplastida order species are labeled with K*. Consensus is expressed as a sequence logo. The black boxes mark three conserved poly-lysine stretches (dubbed K-boxes) K1, K2 and K3.

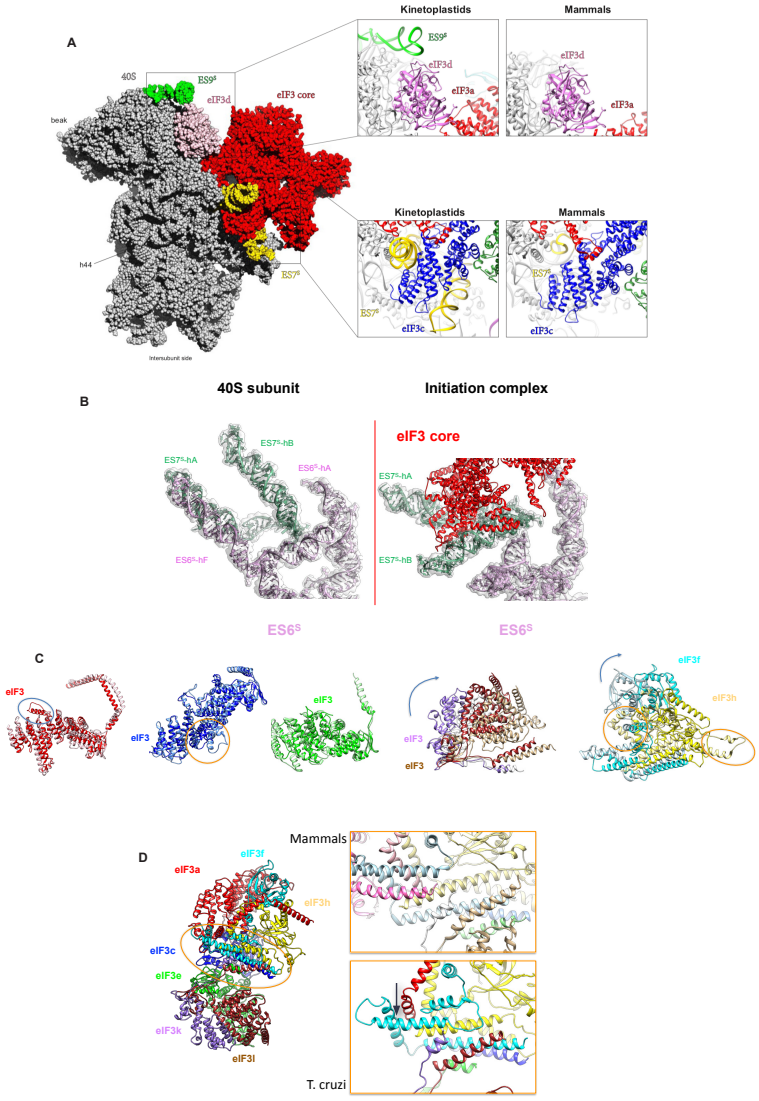

Fig. S8. Specific features of Kinetoplastidian eIF3 and its ribosome binding site. (A) Overall sphere representation of the *T.cruzi* 43S PIC showing kinetoplastidian specific rRNA oversized expansion segments (ESs) in contact with eIF3. Upper panel: comparison of the kinetoplastidian and mammalian eIF3d docking site within the 43S PIC (eIF3d in violet, ES9^{s} in green, eIF3a in red); lower panel: comparison of the kinetoplastidian and mammalian eIF3c docking site within the 43S PIC (eIF3c in blue, ES7^{s} in yellow). **(B)** A close-up view of the *T.cruzi* ES7^{s} and ES6^{s} prior to (left) and post (right) eIF3 binding to the 40S **(C)** Overlay of mammalian and kinetoplastidian structures of individual eIF3 subunits with marked structural differences. The *T. cruzi* structures are depicted in dark and mammalian in light color shades. Curved arrows indicate the direction of *T. cruzi* eIF3 subunits structural rearrangement compared to their mammalian counterparts. Colored ovals highlight marked structural differences between *T. cruzi* and mammalian eIF3 subunits. **(D)** Cartoon representation of the eIF3 atomic model showing the eIF3 helical bundle in mammals (upper panel) and in *T. cruzi* (lower panel). Dark arrow indicates the shift of a helix from eIF3f in *T. cruzi* to compensate for the absence of eIF3m.

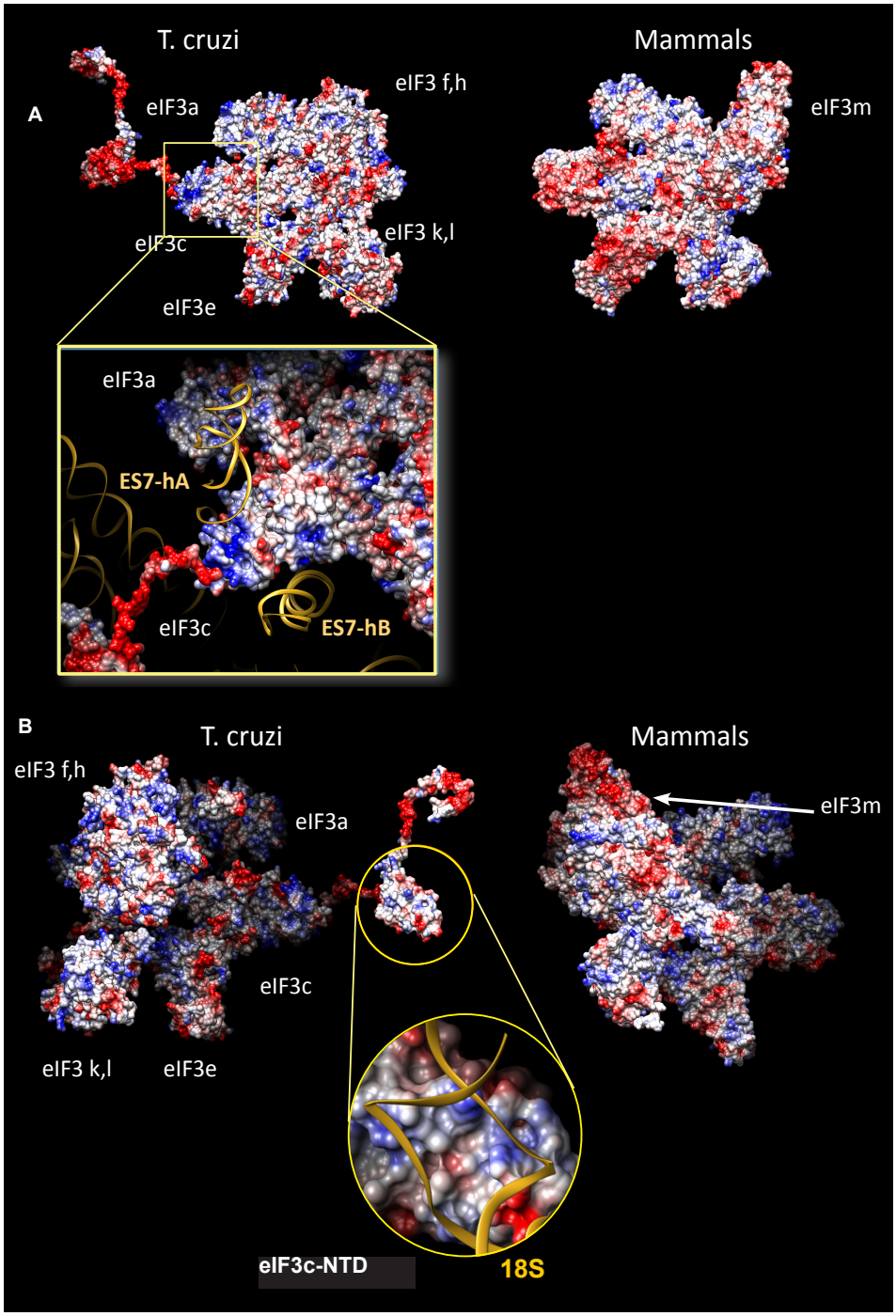

Fig. S9. Charge surface analysis of the T.cruzi and mammalian eIF3 structures. (A) Surface representation of the *T. cruzi* (left) and mammalian (right) eIF3 structure seen from the 40S platform side. Lower panel: close-up view of *T.cruzi* eIF3c and its interaction with the ES7^{s} helix A and helix B. Model is color-coded according to the electrostatic potential – negative in red and positive in blue. **(B)** Surface representation of the *T. cruzi* (left) and mammalian (right) eIF3 structure seen from the 40S solvent side. Lower panel: close-up view of the *T.cruzi* eIF3c-NTD and its interaction with 18S RNA.

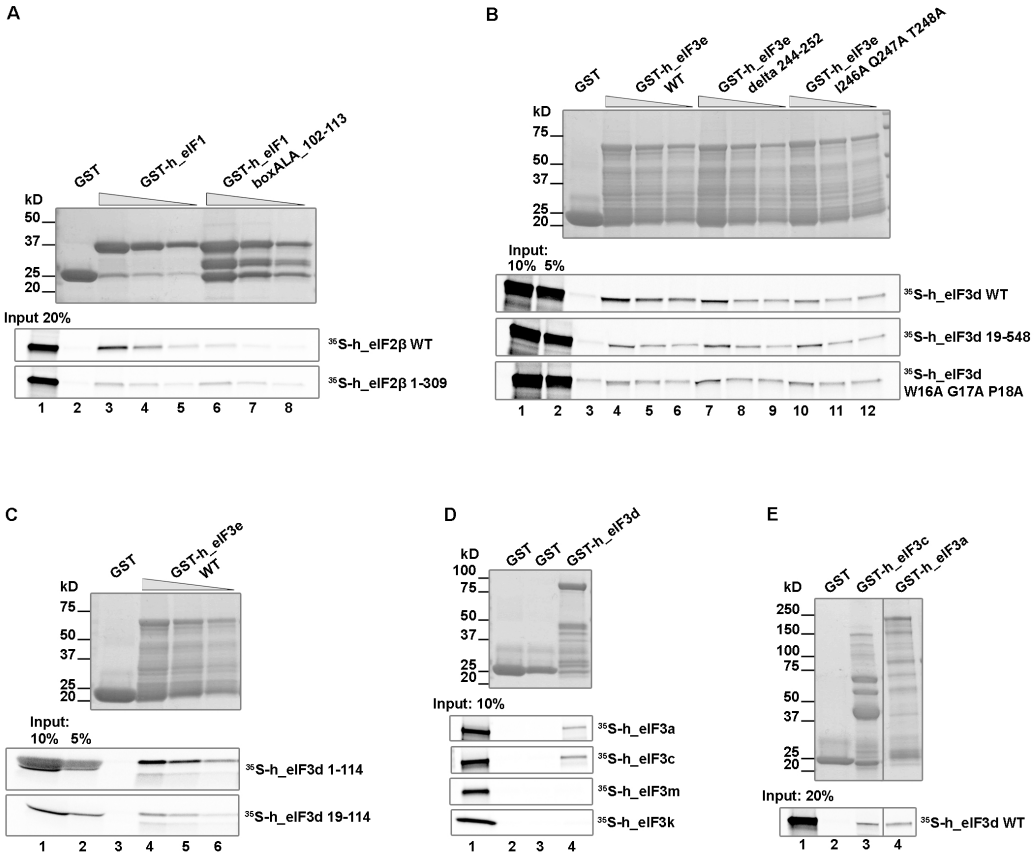

Fig. S10. *In vitro* analysis of eIF3 intersubunit interactions. (**A**) *In vitro* protein-protein binding analysis of the interaction between the *in vitro* translated human ^{35}S-labeled eIF2β and its C-terminal truncation (eIF2β 1-309) against wild type eIF1 or its mutated variant (eIF1-boxAla-102-113; residues 102-113 substituted with a stretch of alanines) fused to GST. *In vitro* translated proteins were tested for binding with three different dilutions of individual GST-fusion proteins. Lane 1 contains 20% of input amounts of *in vitro*-translated proteins added to each reaction. (**B**) Same as in (A) except that binding between the human wild type eIF3d subunit, its N-terminally truncated form (19-548), and its mutated variant (W16A G17A P18A) against the human wild type eIF3e subunit, or its inner deletion (delta 244-252), or its mutated variant (I246A Q247A T248A) fused to GST was analyzed. Lanes 1 and 2 show 10% and 5% input, respectively. Quantification was performed by the Quantity One software (see Fig. 3J.) (**C**) Same as in (A) except that binding between truncations of the human eIF3d subunit (1-114 and 19-114) and eIF3e fused to GST was analyzed. Quantification is presented in Fig. 3K. (**D**) *In vitro* protein-protein binding analysis of ^{35}S-labeled eIF3a, eIF3c, eIF3k and eIF3m subunits against eIF3d fused to GST. Lane 1 shows 10% input. (**E**) *In vitro* protein-protein binding analysis of human ^{35}S-labeled eIF3d against eIF3c and eIF3a subunits fused to GST. Lane 1 shows 20% input.

Fig. S11. Sequence alignment of k-DDX60 and human DDX60. BlastP alignment between *T. cruzi* k-DDX60 and human DDX60 showing the relatively modest global homology between both proteins. Only most homologous regions were presented (in green, purple and red boxes). Magenta boxes on domains annotation schema highlight the trypanosomatid-specific domains that are inexistent in DDX60 from human and other eukaryotic species. Pink and violet colors highlight the A-site Insert (AI) and the ATP binding pocket in k-DDX60, respectively.

	Ribosomal RNA	Ribosomal protein	Initiation factors
eIF1	N65-G2303, N65-G2303, Q81-C2282	**none**	**eIF2-β :** R29-S251, Q32-Y266, Q43-T325, V77-Y326, L108-N276 **eIF2-ɣ :** S16-N459, V17-V147, E22-H81 **eIF3c :** A48-F36, R53-E37, R53-T39, N96-R26, N96-I31
eIF1a	R33-A1341, R33-G2283, K37-G2283, N48-A2277, R56-G2303, R61-C2183, R66-C620, W74-A2279, R155-G1685	**eS30 :** E35-R10 **uS13 :** L164-R119, F167-Y128 **uS19 :** V158-K84, L160-A111, **uS12 :** N89-K54	**eIF2-β :** V134-N208, V136-L210, F135-E212, F135-D275, F135-Y279, D132-K287 **k-DDX60:** Q24-L1582, D27-E1579, Y30-K1598, S120-Q761, D125-F764, D141-R1534, A145-E1527, D148-N1526
eIF2-α	**none**	**uS7 :** Y166-V120, T167-R121, R173-G117, D195-R184, Y200-D180	**tRNA :** K104-C55, R105-G52, R108-U54, W119-C55, H232-C55, E296-U54, H297-G56, P350-C73, R356-C3, K358-A62 **eIF2-ɣ :** F315-D351, V320-L350, R324-N280, V345-E275, I347-K272, P350-F268
eIF2-β	R333-U1340, R333-G1342, R337-U1339	**uS19 :** N259-P150	**tRNA :** K221-A36, N255-G25, S258-A26, K300-G67, R303-G69, **eIF1 :** S251-R29, Y266-Q32, N276-L108, T325-Q43, Y326-V77 **eIF1A :** N208-V134, L210-V136, E212-F135, D275-F135, Y279-F135, K287-D132 **eIF5 :** N118-R265, L120-A262, L123-V325, K125-A366, R135-W372, L142-I332 **eIF2-ɣ :** N173-H248, T176-Y245, G181-Y241, Y182-Y211, Y184-D240, S185-N238, R189-E204, L195-D200, M305-E83, T317-M86 **k-DDX60:** R190-R909, H292-Y772 (pi-stacking), K332-D1565
eIF2-ɣ	**none**	**none**	**tRNA :** K79-C73, Y80-C73, D269-A75, K272-A72, R282-A75, K329-A75 **eIF1 :** H81-E22, V147-V17, N459-S16 **eIF2-α :** F268-P350, K272-I347, E275-V345, N280-R324, L350-V320, D351-F315, **eIF2-β :** E83-M305, M86-T317, D200-L195, E204-R189, Y211-Y182, N238-S185, D240-Y184, Y241-G181, Y245-T176, H248-N173 **eIF5 :** G223-R229, P383-N239, D432-D204, W465-T237, R469-T205 **DDX60-like :** N169-G771, P171-P770, V205-E906, D209-R902, H212-L987
eIF3c	S52-A1360, R53-C1361, R127-C369, D130-G368, K207-A1523 and U1524, R232-U1476, Q329-G1438, R331-U1439, R243-U1526, Q204-A1525	**eS27 :** Q191-Q56, K192-K63	**eIF1 :** R26-N96, I31-N96, F36-A48, E37-R53, T39-R53 **eIF3d :** P234-A47, R295-W44, L380-F9, L418-W16, R419-P13, I434-M28, Y436-D26, N437-D26, N542-H80 **k-DDX60:** N-ter tail with Y832, F834, H835
eIF3a		**eS1 :** R8-T72, T12-R192, L17-I194	
eIF3d	K35-U1393, D43-G1532, D50-A1475, R149-U1863, K292-C1867, R294-U1862, Q296-C1868, K301-U1863, N302-U1863	**eS27 :** T36-K37, A38-G79, I39-T76, D37-F80 **S33 :** Q126-S78, D255-R83, K371-E95, Y377-M73 **uS7 :** Q434-E21, Q368-D26 **RACK1 :** S409-E277, N410-Q279	**eIF3a :** R66-I194, H74-W262, R94-H220, H96-N76, F97-H36 **eIF3c :** F9-L380, P13-R419, W16-L418, D26-Y436, D26-N437, M28-I434, W44-R295, A47-P234, H80-N542 **eIF3e :** F3-K14, L5-A196, T15-Q240
eIF5	**none**	**none**	**eIF2-β :** A262-L120, R265-N118, V325-L123, I332-L142, A366-K125, W372-R135 **eIF2-ɣ :** D204-D432, T205-R469, R229-G223, T237-W465, N239-F383 **k-DDX60:** D284-S944, D288-R941, K292-S826
k-DDX60	S26-U1722, K724-A51, Q725-A51, N727-U460, H728-G477, S736-C480, R1557-G1622, T1559-G1623, D1570-C2176	**eS12 :** S2-D70, S3-E71, R6-E72 **eS31 :** S2-L98, E93-K94 **uS12 :** T740-N97, I706-Q136 **eS30 :** V687-V21, E690-K20	**tRNA :** Q1548-U35, S1551- A34, K1554-C33, Q1555-C33 **eIF1a :** L1582-Q24, E1579-D27, K1598-Y30, Q761-S120, F764-D125, R1534-D141, E1527-A145, N1526-D148 **eIF2-β :** R909-R190, Y772-H292 (pi-stacking), D1565-K332 **eIF2-ɣ :** P770-P171, G771-N169, R902-D209, E906-V205, L987-H212 **eIF3c :** Y832, F834,H835 with N-terminal tail **eIF5 :** S826-K292, R941-D288, S944-D284

Table S1. Detailed overview of interactions between eIFs, ribosomal proteins, rRNA and k-DDX60.