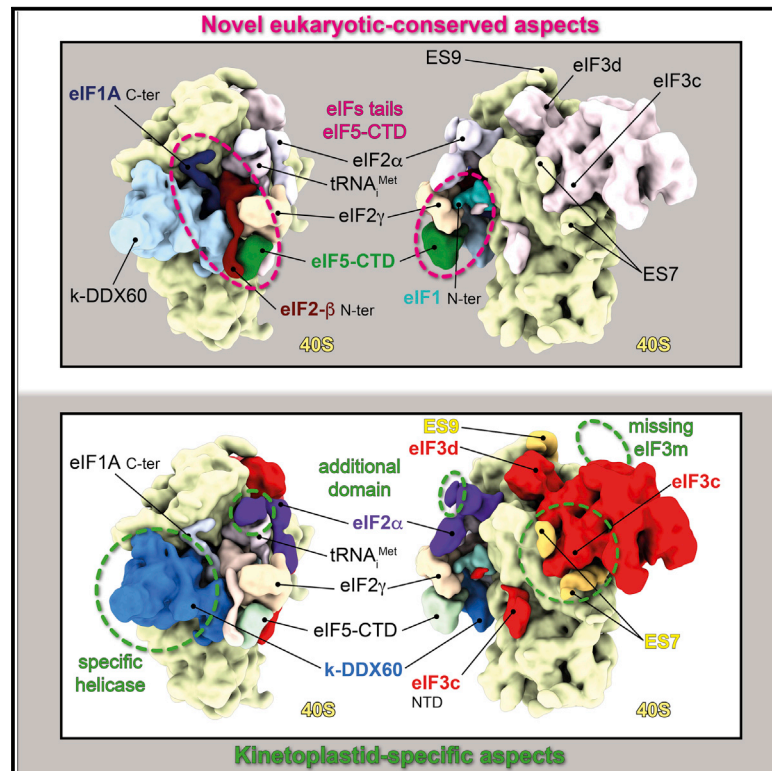


Structural Differences in Translation Initiation between Pathogenic Trypanosomatids and Their Mammalian Hosts

Graphical Abstract



Authors

Anthony Bochler, Jailson Brito Querido, Terezie Prilepskaja, ..., Aline Rimoldi Ribeiro, Leoš Shivaya Valášek, Yaser Hashem

Correspondence

yaser.hashem@inserm.fr

In Brief

Trypanosoma cruzi, which causes Chagas disease, represents a serious health problem. Bochler et al. solve the structure of the 43S translation pre-initiation complex of this parasite, highlighting numerous features by which it differs from the mammalian complex to stimulate development of new drugs against trypanosomiasis and leishmaniasis.

Highlights

- Structure of the 43S pre-initiation complex from *Trypanosoma cruzi* is solved at 3.33 Å
- The kinetoplastids' eIF3 core is a septamer that binds mainly the unique, extended ES7^s
- A kinetoplastid-specific DDX60-like helicase binds to the 43S PIC entry pore
- The 40S positions of eIF5-CTD and key tails of several eIFs are determined



Article

Structural Differences in Translation Initiation between Pathogenic Trypanosomatids and Their Mammalian Hosts

Anthony Bochler,^{1,6} Jailson Brito Querido,^{1,5,6,7} Terezie Prilepskaja,^{2,3,6} Heddy Soufari,¹ Angelita Simonetti,⁵ Mayara Lucia Del Cistia,¹ Lauriane Kuhn,⁴ Aline Rimoldi Ribeiro,¹ Leoš Shivaya Valášek,² and Yaser Hashem^{1,8,*}

¹INSERM U1212 (ARNA), Institut Européen de Chimie et Biologie, Université de Bordeaux, Pessac 33607, France

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

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

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

⁵CNRS UPR9002 (ARN), Institut de Biologie Moléculaire et Cellulaire, Université de Strasbourg, Strasbourg 67000, France

⁶These authors contributed equally

⁷Present address: MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Cambridge, UK

⁸Lead Contact

*Correspondence: yaser.hashem@inserm.fr
<https://doi.org/10.1016/j.celrep.2020.108534>

SUMMARY

Canonical mRNA translation in eukaryotes begins with the formation of the 43S pre-initiation complex (PIC). Its assembly requires binding of initiator Met-tRNA^{Met} and several eukaryotic initiation factors (eIFs) to the small ribosomal subunit (40S). Compared to their mammalian hosts, trypanosomatids present significant structural differences in their 40S, suggesting substantial variability in translation initiation. Here, we determine the structure of the 43S PIC from *Trypanosoma cruzi*, the parasite causing Chagas disease. Our structure shows numerous specific features, such as the variant eIF3 structure and its unique interactions with the large rRNA expansion segments (ESs) 9^S, 7^S, and 6^S, and the association of a kinetoplastid-specific DDX60-like helicase. It also reveals the 40S-binding site of the eIF5 C-terminal domain and structures of key terminal tails of several conserved eIFs underlying their activities within the PIC. Our results are corroborated by glutathione S-transferase (GST) pull-down assays in both human and *T. cruzi* and mass spectrometry data.

INTRODUCTION

The first critical initiation step in eukaryotes is the assembly of the 43S pre-initiation complex (PIC) comprising the 40S, the eukaryotic initiation factor 2 (eIF2)·GTP·Met-tRNA^{Met} ternary complex, and eIFs 1, 1A, 3, and 5 (Hinnebusch, 2017; Valášek, 2012). It is followed by the recruitment of the mRNA promoted by the mRNA-cap-binding complex comprising eIF4A, 4B, and 4F (Guca and Hashem, 2018; Hashem and Frank, 2018), forming the 48S PIC. The 48S PIC then scans the 5' untranslated region (UTR) of mRNA in the 5' to 3' direction until 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 are a group of flagellated unicellular eukaryotic parasites that have a complex life cycle. They spend part of their life cycle in the insect gut 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 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) (Brito Querido et al., 2017) and unusually oversized ribosomal RNA (rRNA) expansion segments (ESs) (Hashem et al., 2013b). Since these unique features may play specific roles in kinetoplastid mRNA translation, they provide potential specific drug targets.

It was proposed that two particularly oversized ESs, namely, ES6^S and ES7^S, located near the mRNA exit channel on the kinetoplastid 40S, may contribute to modulating translation initiation in kinetoplastids by interacting with the structural core of the eukaryotic eIF3, specifically by its subunits a and c (Hashem et al., 2013a). eIF3 is the most complex eIF, promoting not only nearly all initiation steps but also translation termination, stop codon readthrough, and ribosomal recycling (Valášek et al., 2017). 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 (Asano et al., 2001a; Valášek et al., 2017). Mammalian eIF3 comprises 12 subunits (eIF3a–m; excluding j), of which 8 form the PCI (proteasome/CSN/eIF3)/MPN (Mpr1/Pad1 N-terminal) octameric structural core (eIF3a, c, e, f, h, k, l, and m) (des Georges et al., 2015; Herrmannová et al., 2020; Sun et al., 2011; Wagner et al., 2014, 2016). Interestingly, unlike their mammalian hosts, kinetoplastids do not encode the eIF3m subunit (Li et al., 2017; Meleppattu et al., 2015; Rezende et al., 2014) 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 (Asano et al., 2001b). 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 (Asano et al., 1999; Das et al., 1997). However, the molecular details underlying the eIF5-CTD critical assembly role remain elusive, and—in contrast to the eIF5-NTD (Llácer et al., 2018)—so are the structural details of its binding site within the 43S PIC (Zeman et al., 2019). 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, eIF2 β , eIF1, and eIF1A, which are all critically involved in scanning and AUG recognition.

Here, we solved the structure of the 43S PIC from *Trypanosoma cruzi* at 3.33 Å and unraveled various new aspects of this complex, of which some are specific to trypanosomatids and others are 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 PICs from two different species, namely, *Trypanosoma cruzi* and *Leishmania tarentolae*, by stalling the 43S complexes with guanosine monophosphate (GMP)-PNP, a non-hydrolysable analog of guanosine triphosphate (GTP), as previously described (Simonetti et al., 2016, 2020). 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 STAR Methods; Figures 1A and 1B; Table S1). Surprisingly, we also identified an ortholog of the human DEAD-box RNA helicase DDX60 (Figure 1B; Table S1). A similar repertoire of eIFs can also be found in the 43S PIC from *L. tarentolae* (Figure 1B; Table S2). 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, although DDX60 is non-essential in

mammals (Miyashita et al., 2011; Oshiumi et al., 2015), it is required for the cell fitness in kinetoplastids and trypanosomatids (Alsford et al., 2011), 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 used cryo-electron microscopy (cryo-EM) to determine the structure of the *T. cruzi* 43S PIC to an overall resolution of 3.33 Å, after image processing and extensive particle sorting (Figures S1A–S1D). Our reconstruction reveals the so-called “scanning-conductive conformation” of the 43S PIC, in which the head of the 40S is tilted upward to open up the mRNA channel for the subsequent mRNA loading (des Georges et al., 2015; Hashem et al., 2013a; Llácer et al., 2015). Thanks to the conservation of structures and binding sites of most of the identified initiation factors, we were able to segment the map accordingly, thus yielding density segments corresponding to the 40S, eIF1, eIF1A, eIF2 α , eIF2 β , eIF2 γ , Met-tRNA^{Met}, and eIF3 structural core (Figures 1C–1E). Importantly, we could also identify the entire density corresponding to the NTT of the eIF3d subunit, which is implicated in the mRNA-specific translational control (Lee et al., 2015, 2016) (see below).

Furthermore, we observed an unassigned density contacting eIF2 γ that has not been seen previously in any equivalent complexes. Because rigid body fitting of the crystal structure of the eIF5-CTD (Wei et al., 2006) showed a close agreement with this unassigned density (Figure S2C) and previous biochemical and genetics findings suggested a close co-operation between eIF5 and eIF2 on the ribosome (Asano et al., 1999; Luna et al., 2012; Singh et al., 2012), we assigned this density to the eIF5-CTD (Figures 1C–1E). Because the eIF5-CTD is known to interact with the eIF2 β -NTT in both yeasts and mammals (Asano et al., 1999; Das et al., 1997; Das and Maitra, 2000), we could also assign part of the eIF2 β -NTT to its corresponding density (Figure 1D) (see below). It is important to highlight that it was possible to assign the above-mentioned densities to eIF5-CTD thanks to its general conservation among eukaryotes.

As discussed in detail below, beyond these evolutionarily conserved features of the 43S PIC in eukaryotes, our cryo-EM reconstruction also identified several trypanosomatid- and kinetoplastid-specific peculiarities. For instance, the kinetoplastid eIF2 α contains a specific N-terminal domain insertion of unknown function (Figure S2A), and indeed, an extra density on the eIF2 α subunit can be observed (Figures 1D and 1E, dashed circle). We also revealed a large density at the 40S interface, in the vicinity of the mRNA channel entrance (Figures 1C and 1D), which was unseen in any of the previous mammalian and yeast 43S PIC reconstructions. Taking into account our proteomic analysis (Figure 1B; Tables 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 kinetoplastid DDX60 (k-DDX60) helicase. These same k-DDX60 and eIF2 α -NTT densities are also present in the *L. tarentolae* 43S PIC reconstruction (Figures S1E–S1G).

Our analysis reveals a wealth of new interactions (Table S3; Figures S3 and S4). Based on the cryo-EM reconstruction of the *T. cruzi* 43S PIC and the conservation of the initiation factors,

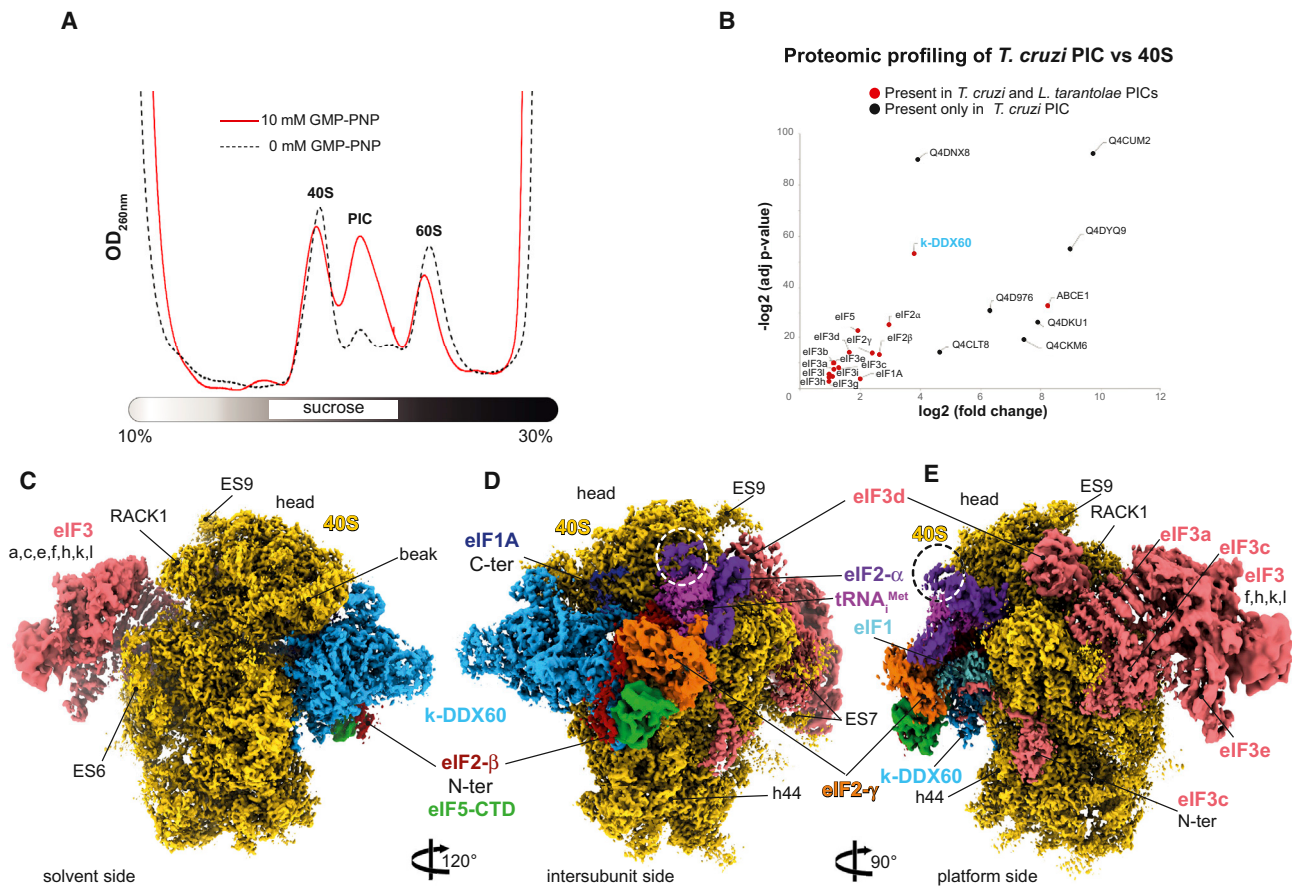


Figure 1. Composition and Cryo-EM Structure of the *T. cruzi* 43S PIC

- (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 PIC in comparison with native 40Ss purified from the *T. cruzi* cell lysate (see STAR Methods for the validation).
 (C) The overall structure of the *T. cruzi* 43S PIC shown from the solvent side. The initiation factors are colored variably.
 (D) The 43S PIC reconstruction focused on the intersubunit side. Extra density of eIF2 α corresponding to the kinetoplastid specific N-terminal insertion is encircled by a dashed line.
 (E) The 43S PIC reconstruction focused on eIF3 and the 40S platform. Different segments are filtered according to their average local resolutions.

a near-complete atomic model was generated (Table S4; see STAR Methods).

The eIF5 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 eIF2 β -NTT and eIF2 γ (Figures 2A–2D). It was proposed that the three conserved poly-lysine stretches (dubbed “K-boxes”) within the eIF2 β -NTD mediate the eIF2 interaction with the eIF5-CTD (Asano et al., 1999; Das et al., 1997). Interestingly, the K1-box and K2-box are conserved in their basic charge character but replaced by R-rich stretches in kinetoplastids (Figure S2B). However, as our structure of eIF2 β -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 kinetoplastids (Figure S2B); it is replaced by a Q-rich motif, and yet, its position and orientation toward its binding partner in the eIF5-CTD is

conserved. Additionally, our structure shows numerous other contacts between hydrophobic and charged residues on each side (residues L120, N118, L123, L120, L142, K125, and V132 of eIF2 β contact A262 R265, V325, V329, I332, Q364, and W372 of eIF5, respectively) (Figures 2A and 2B; Figure S3; Table 3). Because the eIF5 residues 320 through 373 correspond to the conserved and essential segment (known as the bipartite motif AA (acidic/aromatic)-box1 and 2; Figures 2A and 2B; Table S3), which was previously implicated in mediating the eIF5-CTD-eIF2 β -NTT contact in both yeast and mammals (Asano et al., 1999, 2001b; Das et al., 1997; Das and Maitra, 2000), our structure not only provides critical structural evidence supporting earlier biochemical and genetics analysis but also clearly indicates that the molecular determinants of the eIF5-CTD-eIF2 β -NTT contact are conserved. Therefore, we suggest that the eIF5-CTD also occupies the same position in yeast and mammals. We therefore modeled eIF5-CTD according to its mammalian counterpart (Figure S2C).

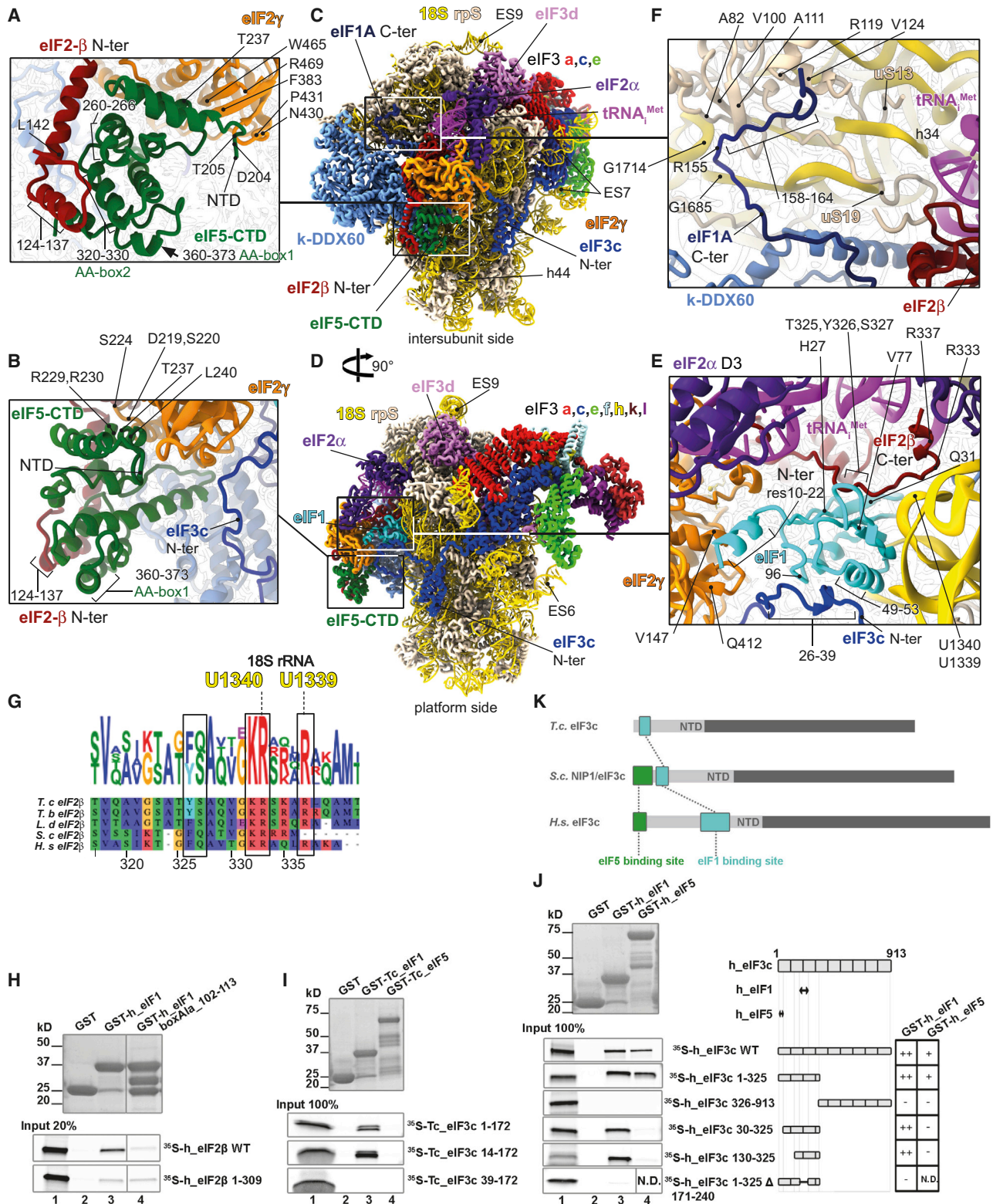


Figure 2. Atomic Model of the 43S PIC Showing the Interaction Network of Various eIFs

(A) Close-up view of an atomic model of the eIF5-CTD (in green), the eIF2 β -NTT (in cherry red), and eIF2 γ (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.

(legend continued on next page)

Our structure also provides important molecular insight into the eIF5-CTD interaction with the eIF2 γ domain I (G-domain), where Arg229, Arg230, and R273 of eIF5 contact the G-domain's Asp219, Ser224, and Ser220, respectively (Figures 2A and 2B; Figures S3J and S3K; Table S3). The eIF5-CTD also binds domain III, where Asp204, Thr205, Thr237, and Leu240 of eIF5 interact with domain III's Pro431, ArgR469-Asn430, Trp465, and Phe383. Noteworthy, the eIF5-CTD shares a common topology with the CTD of the ϵ subunit of the nucleotide exchange factor eIF2B (Asano et al., 1999); they both fold into a W2-type HEAT domain (Wei et al., 2006), mediating contacts of both factors with the eIF2 β -NTT and eIF2 γ (Alone and Dever, 2006). Based on our structure, the arrangement of the eIF5-CTD HEAT domain binding site on eIF2 γ in the context of the 43S PIC is similar to that of the eIF2B ϵ -CTD HEAT domain in the context of the recently solved eIF2-eIF2B complex (Kashiwagi et al., 2019; Kenner et al., 2019).

Taken together, the eIF5-CTD interaction network revealed here indicates that the interaction between eIF5-CTD and eIF2 γ 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-binding pocket to promote reversible GTP hydrolysis on eIF2 during scanning, as demonstrated earlier (Algire et al., 2005).

Extensive Interaction Network of eIF1 in the Context of the 43S PIC

After the GTP hydrolysis by eIF2 γ , 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 $^{\text{Met}}$, leading to the full accommodation of TC in the decoding pocket (Algire et al., 2005; Hinnebusch, 2017) and eIF1 replacement by the eIF5-NTD. Because the access to the GTP-binding pocket on eIF2 γ is in part protected by the zinc-binding domain (ZBD) of the eIF2 β -CTD (Llácer et al., 2015; Stolboushkina et al., 2008) and biochemical and genetic analysis in yeast indicated that the eIF1 interactions with eIF2 β and the NTD of the c subunit of eIF3 play a critical role in anchoring of eIF1 within the 48S PIC (Karásková et al., 2012; Obayashi et al., 2017; Thakur et al., 2019; Valášek et al., 2004), for our complex understanding of the AUG recognition process, it is necessary to investigate how eIF1 coordinates the release of P_i with the latter factors on a molecular level.

In accordance with earlier biochemical experiments, our structure reveals that the conserved eIF2 β -C-terminal tail (eIF2 β -CTT), together with the eIF3c-NTD, does anchor eIF1 within the 43S PIC (Figure 2E). In particular, the eIF2 β -CTT extends toward the P-site, where its Thr325, Tyr326, and Ser327

residues interact with eIF1 mainly through His27, Val77, and Gln31, which are all conserved in character (Figure 2G; Figure S3A; Table S3). The eIF2 β -CTT also interacts with h24 of the 18S rRNA (Arg 333 and 337 with nucleotides U1340, G1342, and U1339) (Figures 2E and 2G; Figure S3H; Table S3). In addition, the eIF2 β -binding platform of eIF1 also consists of R29, Q32, and Q43 (see Figure S3A and Table S3 for details), as well as of the tip of the eIF1 C terminus (residues 105–108). Based on these findings, we examined binding of human eIF2 β with eIF1 fused to GST moiety using the GST pull-down assay and revealed that the interaction between the CTTs of eIF2 β (residues 310–333) and eIF1 is also conserved in mammals and requires the extreme C terminus (Figure 2H; Figure S5A).

The protein sequence composition of the N-terminal domain of eIF3c can vary across species (Figure S5H). It begins with a few conserved hydrophobic residues, followed by negatively charged SD/SE repeats in all, including in kinetoplastids. Interestingly, budding yeast *Saccharomyces cerevisiae* contains an insertion of approximately 40 residues between the latter 2 groups. The minimal eIF5-CTD-binding site within the yeast eIF3c-NTD was identified to fall into the region of the first 45 residues, including part of this insertion but completely excluding the SD/SE repeats (Karásková et al., 2012). These regions are then followed by the segment that was shown to represent the core eIF1-binding segment in yeast (residues 59–87) (Karásková et al., 2012; Obayashi et al., 2017). The downstream sequence in mammals features a specific insertion (residues 167–238), consisting of two highly acidic regions separated by a mostly positively charged/hydrophobic region (Figure S5H). Strikingly, the first part of this mammalian-specific insertion displays a significant sequence similarity with the *S. cerevisiae* core eIF1-binding region; in particular, the yeast residues 51–92 show ~36% identity with human residues 173–213 (Figure S5H).

Based on our structure, the contact between the *T. cruzi* eIF3c-NTD and eIF1 involves Arg26 through Thr39 of eIF3c and Asn96 and Leu49 through Arg53 of eIF1 (Figure 2E; Figure S3C; Table S3). In accordance, *T. cruzi* eIF1 fused to GST moiety also interacted specifically with the eIF3c-NTD *in vitro* (the first 14 residues of eIF3c are not required, whereas the following residues up to position 39 are required) (Figure 2I). This interacting region following the extreme N-terminal hydrophobic residues and negatively charged SD/SE repeats nicely correlates with the eIF1-binding region of the *S. cerevisiae* eIF3c-NTD specified above (Karásková et al., 2012; Obayashi et al., 2017; Figure S5H).

As for the eIF5-CTD-eIF3c-NTD contact, which was so far determined only in yeast *S. cerevisiae* (Karásková et al., 2012; Obayashi et al., 2017; Phan et al., 1998; Valášek et al., 2004),

(C and D) The overall view of atomic model of the 43S PIC from the intersubunit (C) and the platform side (D).

(E) Close-up view of the P-site, showing eIF1 (in cyan) and its binding partners the eIF2 β -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.

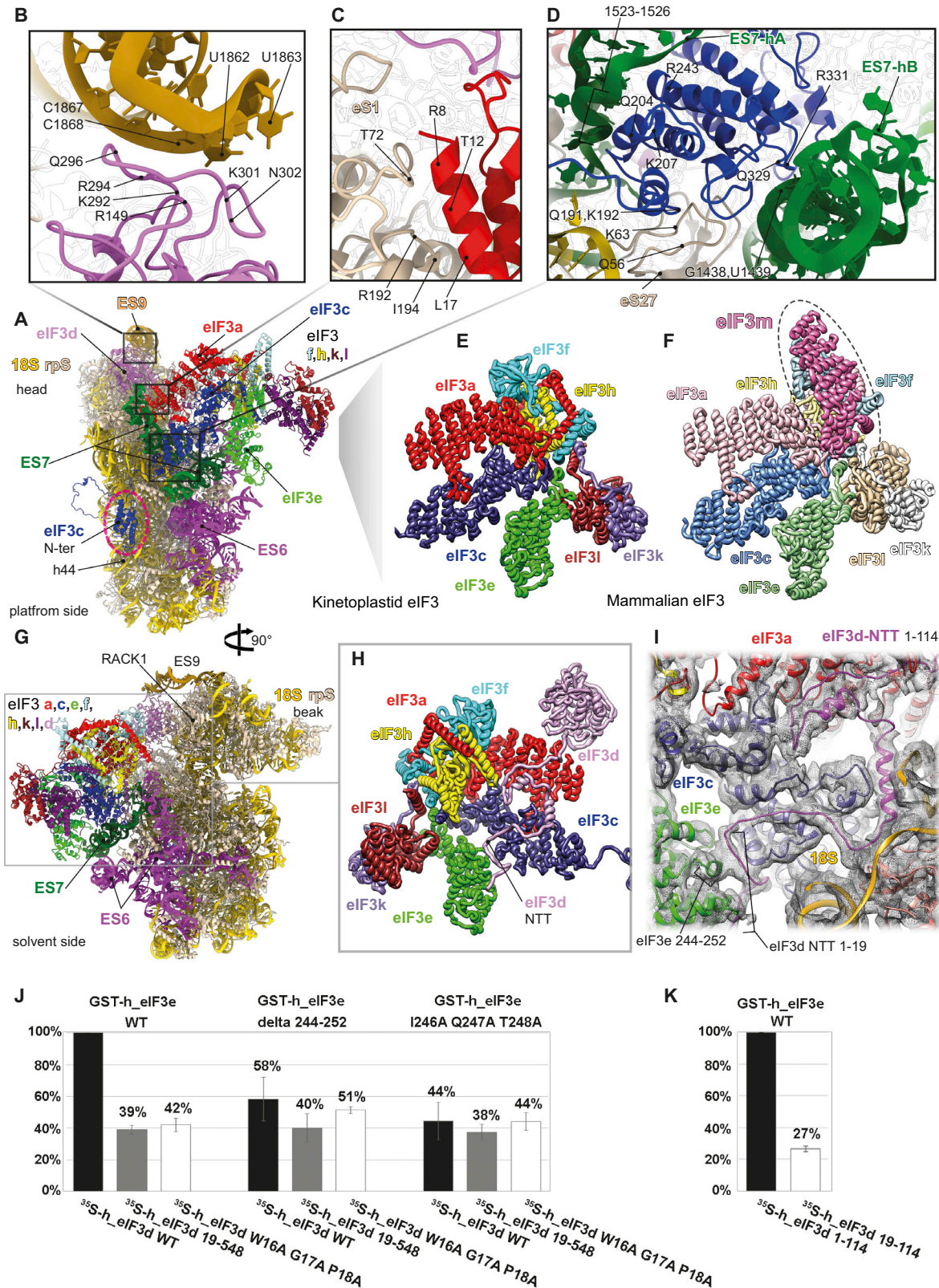
(G) Polypeptide sequence alignment of the eIF2 β -CTT, highlighting residues involved in the interaction with 18S rRNA and eIF1; *T. cruzi*, *T. brucei*, *L. donovani*, *S. cerevisiae*, and *Homo sapiens*. Residue numbering from *H. sapiens* was used.

(H) *In vitro* protein-protein binding analysis of the interaction between human eIF2 β 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.

(K) Schematics illustrating the differences in the localization of eIF1 (turquoise box) and eIF5 (green box) binding sites within the N-terminal segment of eIF3c in *T. cruzi* (*T.c.*), *S. cerevisiae* (*S.c.*), and *H. sapiens* (*H.s.*).



(legend on next page)

given the evolutionary conservation of this extreme N-terminal region, one would expect it to be conserved among all eukaryotes too. Therefore, it was rather surprising not to detect any binding between the *T. cruzi* eIF3c-NTD and eIF5 in any *in vitro* experimental set-up under any condition that we examined exhaustively (Figure 2I; Figures S5F and S5G). This is consistent with our structure (Figure 2B), where despite the observable proximity between the eIF3c-NTD and eIF5-CTD, these two domains remain out of the intermolecular interactions range and for which we detected no structural evidence. Even though we cannot rule out that they may come in contact in the PICs in only some stages of the initiation pathway that we did not capture, we tend to think that these results point to a specific evolutionary shift in kinetoplastid initiation pathway, as will be discussed below.

This unexpected finding prompted us to investigate the conservation of the eIF3c-NTD interactions in higher eukaryotes. Therefore, we fused human eIF1 and eIF5 to GST and tested the resulting fusion proteins against various truncations of the eIF3c-NTD (Figure 2J). In accordance with the yeast data (Karásková et al., 2012; Obayashi et al., 2017) but in contrast to *T. cruzi* (Figure 2I), the extreme N-terminal group of conserved hydrophobic residues of human eIF3c-NTD interacted strongly with eIF5.

Taking into account the peculiarity of the human eIF3c-NTD featuring the aforementioned insertion (residues 167–238), indicating that the eIF1-binding site appears to be located more toward the C-terminal part of the eIF3c-NTD, we first deleted the first 130 residues and, indeed, showed that the eIF3c-NTD segment spanning residues 130 through 325 fully preserved its affinity toward eIF1 (Figure 2J). Conversely, internal deletion of residues 171 through 240 from the human eIF3c-NTD construct resulted in a complete loss of binding (Figure 2J). Thus, the core eIF1-binding site in the human eIF3c-NTD seems to fall into the first part of this mammalian-specific insertion, displaying a significant sequence similarity with the *S. cerevisiae* core eIF1-binding region (Figure S5H), as described above.

Taken together, these findings suggest that despite the undisputable importance of the eIF3c-NTD during the initiation and start-codon recognition, this region has undergone rather dramatic topological as well as sequential restructuring during the course of evolution. (1) The eIF1-binding site preserved its key sequence determinants but moved further downstream in the course of evolution of higher eukaryotes (Figure 2K). In contrast,

(2) the eIF5 binding site remained conserved not only in its sequence but also in its placement at the extreme N-terminal tip of eIF3c across species; however, in kinetoplastids, it most probably lost its purpose. It remains to be seen what molecular consequences of this evolutionary shift are in kinetoplastids and whether or not these two molecules come into a functional contact within the PICs.

Besides the eIF1-CTT binding coordinates, our structure also reveals that the NTT of eIF1 (residues 10 to 22) forms an α -helix that interacts with domains I and III of eIF2 γ (Val85, Val147, Gln412, and Asn459; Figure 2E; Figure S3B; Table S3), 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 toward the head of the 40S, where it interacts with the rRNA (Arg155 with G1685) (Figure 2F; Figure S3D; Table S3) and ribosomal proteins uS19 (residues Val158 with Val100, Ala82, and Ala111; Figure S3G; Table S3) and uS13 (residues Asp162 and Leu164 with Arg119 and Val124, respectively; Figure S3F; Table S3), corroborating findings from a previous hydroxyl-radical probing study (Yu et al., 2009). Moreover, previously uncharacterized interactions between eIF1A and eIF2 β are observed in *T. cruzi* between hydrophobic residues (Tyr133 through Phe135 on eIF1A and Leu282 through Tyr279 on eIF2 β ; Figure S3E; Table S3)

The Specific Features and Binding Site of eIF3 in Trypanosomatids

Strikingly, as seen in Figures 3A–3D, the unusually large trypanosomatid-specific ESs 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 (Figures 3C and 3D). However, unlike in other known eukaryotes, the large ES7^S acts as the main docking point for the eIF3 structural core (Figure S6A; Figure 4A, bottom). In particular, 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, Arg215, Arg232, Arg243, Gln329, and Arg331 and ES7^S nucleotides U1526, A1525 and U1523, U1476, U1526,

Figure 3. Kinetoplastid 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 ES7^S (in green) and eIF3c (in blue).
- (E) Cartoon representation of the atomic model of the kinetoplastid 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 kinetoplastid eIF3 focused on the eIF3d-NTT (in pink).
- (I) Fitting of the eIF3d-NTT model into its cryo-EM map.
- (J and K) Binding analysis between human eIF3d and GST-eIF3e, expressed in plots showing normalized data from three different dilutions of GST proteins (see Figure S5A).

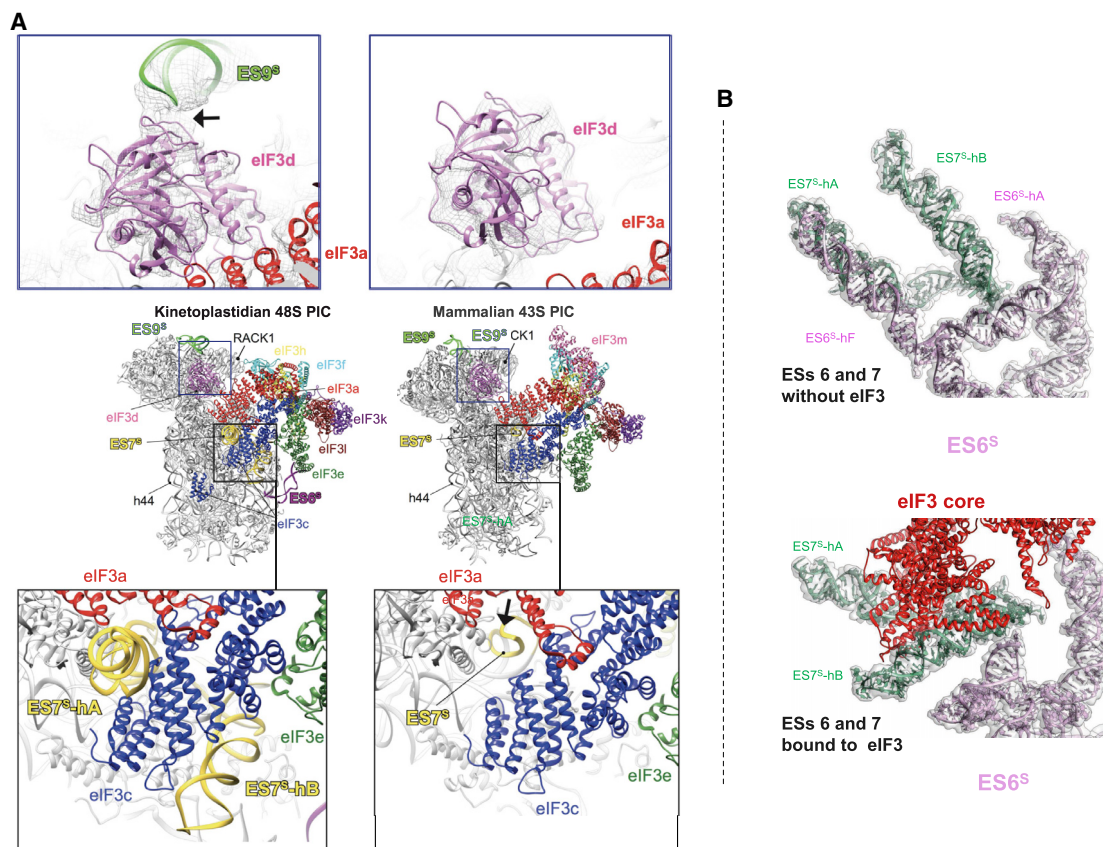


Figure 4. Specific Features of Kinetoplastidian eIF3 and Its Ribosome Binding Site

(A) Overall sphere representation of the *T. cruzi* 43S PIC showing kinetoplastidian-specific rRNA oversized expansion segments (ESs) in contact with eIF3. Top panel: comparison of the kinetoplastidian and mammalian eIF3d docking site within the 43S PIC (eIF3d in violet, ES9^s in green, eIF3a in red); bottom 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 before (top) and after (bottom) eIF3 binding to the 40S.

G1438, and U1439, respectively (Figure 3D; Figures S3M and S3N; Table S3). The local resolution of our complex allowed us to assign the identity of the conserved helical domain of the eIF3c-NTD (Figure 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 (Ser52, Arg53, Lys56, and Arg127 with A1360, C1361, C1596, and C370; Figure S3O; Table S3), suggesting that it also has a similar PIC binding mode 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 (by residues Glu191 and Lys192 with Glu56 and Lys63) (Figure 3D; Table S3). In contrast to eIF3c, the eIF3a binding to the ribosomal protein eS1 does not seem to differ from other eukaryotes (residues Thr7, Arg8, Thr12, and Leu17 contact Gln77, Thr72, Arg192, and Ile194, respectively) (Figure 3C; Table S3).

Another unusually large ES is the kinetoplastidian ES9^s that forms a “horn” on the 40S head, bending toward the mRNA exit channel, where it binds to and stabilizes eIF3d within the 43S PIC (Figures 3A and 3B; Table S3), representing another

important feature that is specific to translation initiation in trypanosomatids. In particular, the eIF3d main globular domain interacts with ES9^s mainly through residues Arg149, Arg294, Gln296, Lys301, and Asp306 contacting nucleotides G1861 through C1867. Moreover, close to the NTT, eIF3d through Asp43 and Asp50 interacts with G1532 and A1475 (Figures S4A and S4B; Table S3). Noteworthy, structures of ES7^s and the exceptionally large ES6^s (Figure S6A) undergo drastic conformational changes upon binding of eIF3, as can be observed by comparing this structure with our previous *T. cruzi* 40S lacking eIF3 (Figure 4B). Amplitude of these conformational acrobatics may indicate 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 the eIF3 structural core differs significantly (Figures 3E and 3F; Figures S7A and S7B), 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, and 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 (Figures S7A and S7B,

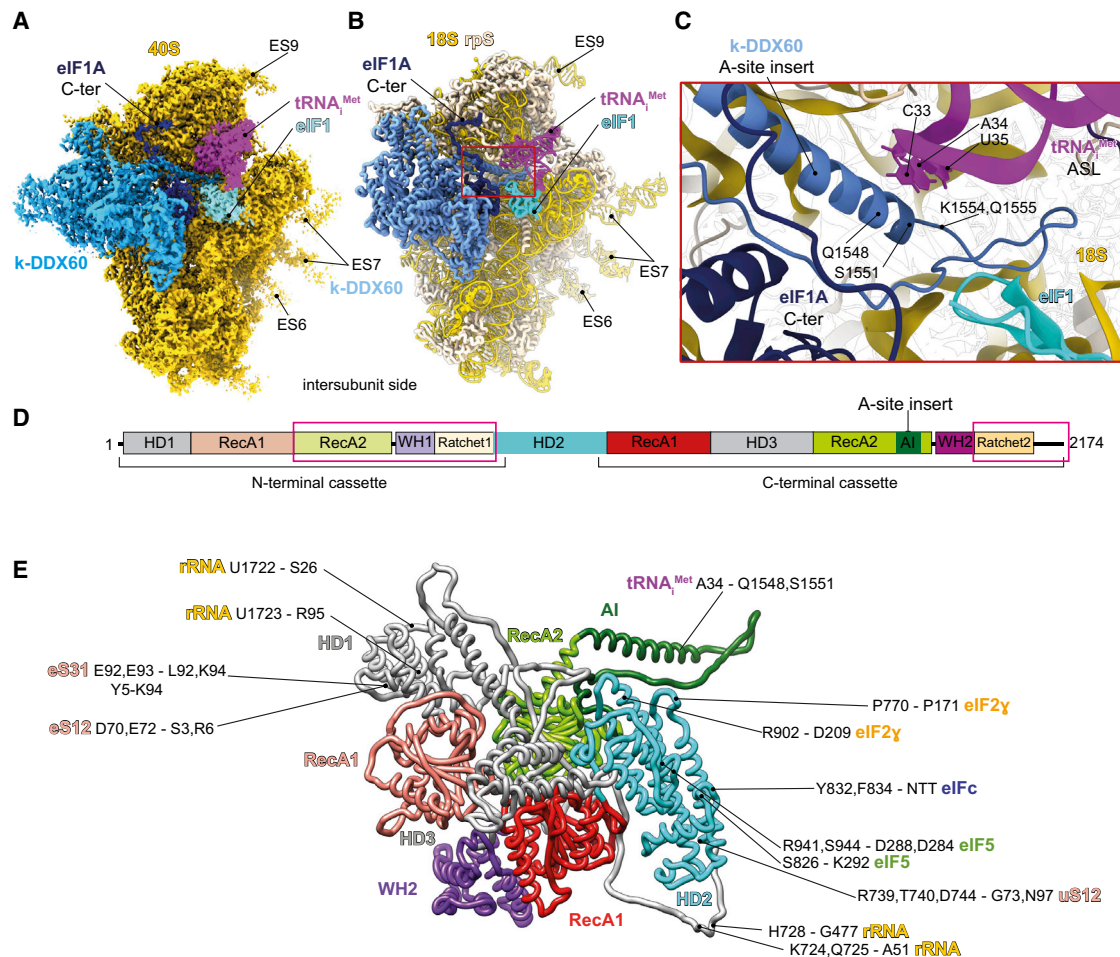


Figure 5. k-DDX60 Structure and Interactions within the 43S PIC

(A) The cryo-EM structure of the *T. cruzi* 43S PIC highlighting k-DDX60 (colored in dark turquoise). eIF2, 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 could not be modeled because of their lower local resolution (see Figure S1B).

(E) Cartoon representation of the atomic model of the k-DDX60 and its interactions with the 43S PIC color coded in accordance with its schematic representation in (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 (Figures S7C and S7D), in part as a consequence of the different 40S binding surface that is mainly represented by rRNA, which is 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 (des Georges et al., 2015; Eliseev et al., 2018; Hashem et al., 2013a). We show here that the eIF3d-NTT, unseen in any previous equivalent complexes, extends toward eIF3e, where it interacts with its PCI domain (residues 1–19 of eIF3d with Ala196, Thr198, Ile 246, Gln247, and Thr248 of eIF3e; Figures 3G–3I; Table S3). Furthermore, the eIF3d-NTT also comes in a less extensive contact with eIF3a, eIF3c, and ribosomal protein eS27 (Figures 3H and 3I; Fig-

ures S3P, S3Q, and S4C; Table S3). In agreement, the interaction of the eIF3d-NTT (the first 114 residues) with the eIF3 core was previously shown in biochemical and genetics studies (Smith et al., 2016). 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 homologs of all these proteins and subjected them to our GST pull-down analysis. As shown in Figures 3J and 3K and Figures S5B–S5E, 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 (Figures S5D and S5E). Because human eIF3d was shown to interact with the mRNA cap (Lee et al., 2016) and together with several other eIF3 subunits

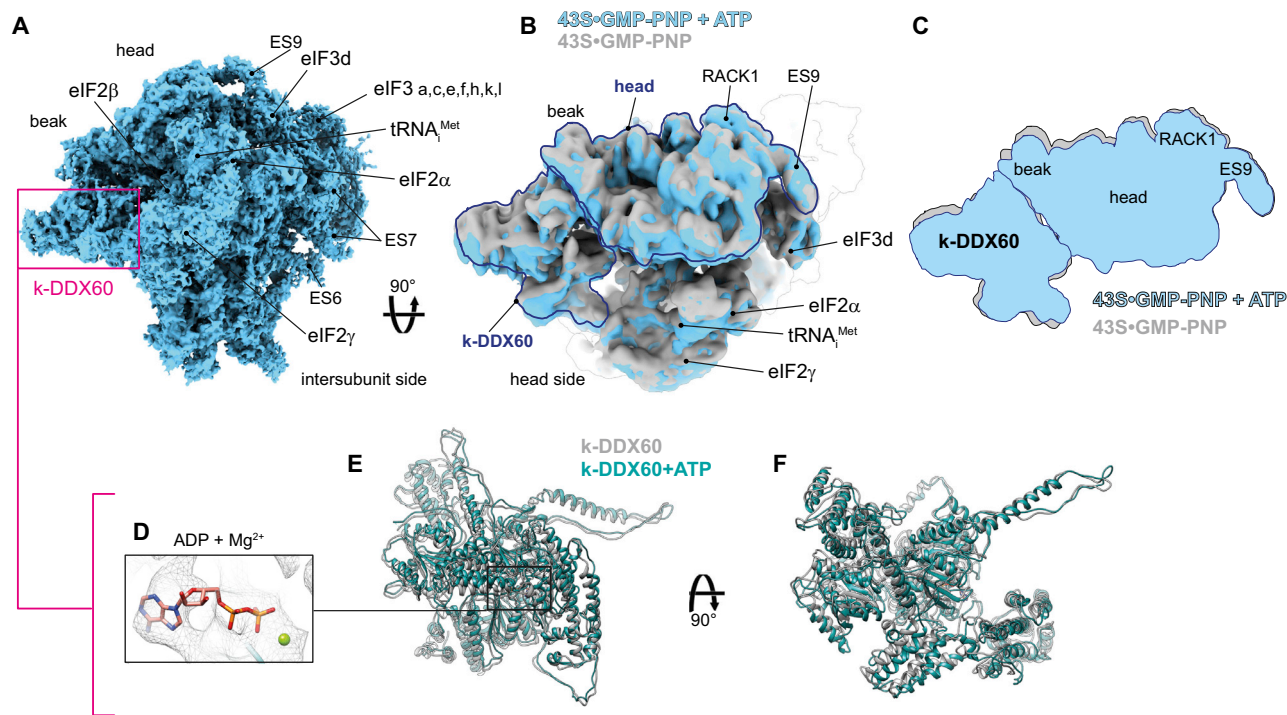


Figure 6. Global Conformational Rearrangement of the 43S PIC Driven by ATP Binding to k-DDX60

(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 gray) 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 and 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.

(including eIF3a, c, e, k, and l) was proposed to promote recruitment of selected mRNAs to the 43S PIC to control their expression in response to various stresses and cellular signals (Herrmannová et al., 2020; Lee et al., 2015; Shah et al., 2016), we speculate that these contacts play a 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 (Figures 1B–1D; Figures S1E–S1H). Known structures of eIFs or ABCE1 (des Georges et al., 2015; Erzberger et al., 2014; Llácer et al., 2018) do not fit into this density, and proteomic analysis shows substantial presence of the helicase DDX60 protein in our samples (Figure 1B; Tables S1 and S2), which we henceforward refer to as kinetoplastid-*DDX60* (k-DDX60). The density was of sufficient resolution to build an atomic model of k-DDX60, including the helicase recombinase A (RecA) domains (Figure 5; Figures S4D–S4H), which fully validates our assignment. Besides the RecA domains, k-DDX60 counts two winged-helix 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-

ures 5C–5E; Figures S8A–S8C for conservation and secondary structures details).

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 homolog is an ATP helicase. Next, we wanted to inspect the 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 (Figure 6A). It is important to point out that the resolution of the 43S PIC+ATP reconstruction is mostly worse than 4 Å, precluding unambiguous determination of whether ATP hydrolysis took place or not. Nonetheless, the structure reveals a global conformational rearrangement of the 40S head (Figures 6B and 6C), which could be driven by the k-DDX60 rearrangement upon ATP hydrolysis (Figures 6D–6F). 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 (Figure 6D).

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^{Met} (Figure 5C; Figure S4P), preventing the codon-anticodon

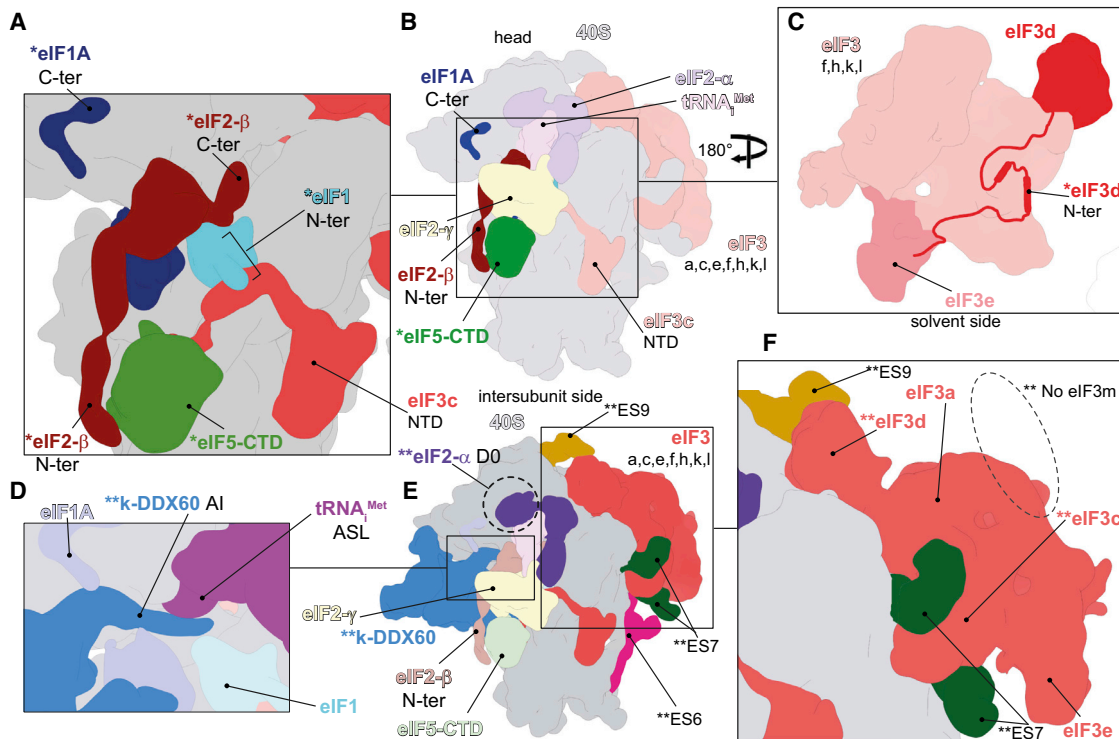


Figure 7. Previously Uncharacterized Eukaryotic-Conserved and Trypanosomatid-Specific Features of the 43S PIC Revealed in Our Work
 (A) Schematic model representing a close-up view on the N-terminal tails of eIF1, 1A, 2 β , eIF5-CTD, and eIF3c-NTD, which are all conserved among eukaryotes and revealed in the current work. The ternary complex was removed for clarity.
 (B) Schematic model representing the 43S PIC from the intersubunit side. The previously uncharacterized 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 ASL.
 (E) Schematic model representing the *T. cruzi* 43S PIC from the intersubunit side. Dashed circle highlights the kinetoplastid-specific domain eIF2 α , 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 kinetoplastid eIF3, showing its specific interaction with ES7^S and ES9^S and the absence of the eIF3m subunit. One asterisk (*), conserved features among eukaryotes revealed in our work; two asterisks (**), kinetoplastid-specific features revealed in our work.

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_{Met} in the P-site. Moreover, k-DDX60 interacts directly with eIF3c-NTD and eIF5 (Figure 5E; Figures S4I and S4N), in addition to the 18S rRNA and ribosomal proteins eS12, uS12, and eS31 (Figure 5E; Figures S4J–S4M and S4O), suggesting its direct involvement in structural changes accompanying/driving the AUG recognition process. Finally, k-DDX60 comes in close proximity with eIF2 β , eIF2 γ , and eIF3c, but the local resolution at these possible interaction sites did not allow us to unambiguously define the interacting residues. We believe that owing to its extensive interactions with numerous components of the 43S PIC, k-DDX60 led to a stabilization of the 43S PIC that enabled rigidification of flexible tails of eIFs, allowing them to be resolved by cryo-EM. In agreement, most of these interactions occur by additional domains and insertions of k-DDX60 that are inexistent in its mammalian homolog (Figure 5D; Figure S8). 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) (Michaeli, 2011). 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.

CONCLUSIONS

In summary, our structure reveals numerous previously uncharacterized features of the eukaryotic translation initiation machinery, of which some are common to other eukaryotes, such as the placement and proposed roles of terminal tails of eIF1, eIF1A, eIF2 β , eIF3c, and eIF3d and, above all, the precise binding site of the eIF5-CTD within the 43S PIC (Figures 7A–7C). Furthermore, our data uncover several striking features of translation initiation specific to kinetoplastids (Figures 7D–7F), such as the role of the oversized kinetoplastid eIFs in providing a large,

unique binding surface for eIF3, as well as the 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 avenues of research and development of new effective drugs against trypanosomiasis and leishmaniasis.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead Contact
 - Materials Availability
 - Data and Code Availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - *Trypanosoma cruzi*
 - *Leishmania tarentolae*
 - *Escherichia coli*
 - *Homo sapiens* genetic material
- **METHOD DETAILS**
 - Construction of plasmids
 - 48S Initiation Complex Purification
 - Cryo-EM Grid preparation
 - Cryo-EM Image acquisition
 - Image processing
 - Figure preparation
 - Mass spectrometry analysis
 - Model building and refinement
 - Secondary structures of k-DDX60 and the 18S
 - GST pulldown assay
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2020.108534>.

ACKNOWLEDGMENTS

We thank Christoph Diebolder and Ludovic Renault (NeCEN, Leiden, Holland), as well as Julio Ortiz Espinoza, Corinne Crucifix, and Christine Ruhlmann (IGBMC, Strasbourg, France) for assistance with data acquisition. We would like also to thank Eder Mancera-Martinez for her help in the sample’s 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. This work was supported by ERC-2017-STG #759120 “TransTryp” (to Y.H.); Labex ANR-10-LABX-0036_NETRINA (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. and A.S. purified the complexes from *T. cruzi*, and M.L.D.C. and A.R.R. purified the complex from *L. tarentolae*. T.P. performed the GST pull-down assays and analyzed the data together with L.S.V. Y.H. and H.S. performed the

cryo-EM data processing. L.K. performed tandem mass spectrometry (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.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 24, 2020

Revised: October 8, 2020

Accepted: November 25, 2020

Published: December 22, 2020

REFERENCES

- Adams, P.D., Afonine, P.V., Bunkóczy, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221.
- Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., and Adams, P.D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr. D Struct. Biol.* **74**, 531–544.
- Algire, M.A., Maag, D., and Lorsch, J.R. (2005). Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Mol. Cell* **20**, 251–262.
- Alone, P.V., and Dever, T.E. (2006). 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**, 12636–12644.
- Alsford, S., Turner, D.J., Obado, S.O., Sanchez-Flores, A., Glover, L., Berri-man, M., Hertz-Fowler, C., and Horn, D. (2011). High-throughput phenotyping using parallel sequencing of RNA interference targets in the African trypanosome. *Genome Res.* **21**, 915–924.
- Asano, K., Krishnamoorthy, T., Phan, L., Pavitt, G.D., and Hinnebusch, A.G. (1999). 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**, 1673–1688.
- Asano, K., Phan, L., Valásek, L., Schoenfeld, L.W., Shalev, A., Clayton, J., Nielsen, K., Donahue, T.F., and Hinnebusch, A.G. (2001a). A multifactor complex of eIF1, eIF2, eIF3, eIF5, and tRNA(i)Met promotes initiation complex assembly and couples GTP hydrolysis to AUG recognition. *Cold Spring Harb. Symp. Quant. Biol.* **66**, 403–415.
- Asano, K., Shalev, A., Phan, L., Nielsen, K., Clayton, J., Valásek, L., Donahue, T.F., and Hinnebusch, A.G. (2001b). Multiple roles for the C-terminal domain of eIF5 in translation initiation complex assembly and GTPase activation. *EMBO J.* **20**, 2326–2337.
- Bernier, C.R., Petrov, A.S., Waterbury, C.C., Jett, J., Li, F., Freil, L.E., Xiong, X., Wang, L., Migliozi, B.L., Hershkovits, E., et al. (2014). RiboVision suite for visualization and analysis of ribosomes. *Faraday Discuss.* **169**, 195–207.
- Bieniossek, C., Schütz, P., Bumann, M., Limacher, A., Uson, I., and Baumann, U. (2006). The crystal structure of the carboxy-terminal domain of human translation initiation factor eIF5. *J. Mol. Biol.* **360**, 457–465.
- Brito Querido, J., Mancera-Martinez, E., Vicens, Q., Bochler, A., Chicher, J., Simonetti, A., and Hashem, Y. (2017). The cryo-EM Structure of a Novel 40S Kinetoplastid-Specific Ribosomal Protein. *Structure* **25**, 1785–1794.e1783.
- Chou, F.C., Sripakdeevong, P., Dibrov, S.M., Hermann, T., and Das, R. (2013). Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat. Methods* **10**, 74–76.
- Das, S., and Maitra, U. (2000). Mutational analysis of mammalian translation initiation factor 5 (eIF5): role of interaction between the beta subunit of eIF2 and eIF5 in eIF5 function in vitro and in vivo. *Mol. Cell. Biol.* **20**, 3942–3950.

- Das, S., Maiti, T., Das, K., and Maitra, U. (1997). Specific interaction of eukaryotic translation initiation factor 5 (eIF5) with the β -subunit of eIF2. *J. Biol. Chem.* 272, 31712–31718.
- des Georges, A., Dhote, V., Kuhn, L., Hellen, C.U., Pestova, T.V., Frank, J., and Hashem, Y. (2015). Structure of mammalian eIF3 in the context of the 43S pre-initiation complex. *Nature* 525, 491–495.
- Eliseev, B., Yeramala, L., Leitner, A., Karuppasamy, M., Raimondeau, E., Huard, K., Alkalaeva, E., Aebersold, R., and Schaffitzel, C. (2018). Structure of a human cap-dependent 48S translation pre-initiation complex. *Nucleic Acids Res.* 46, 2678–2689.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* 60, 2126–2132.
- Erzberger, J.P., Stengel, F., Pellarin, R., Zhang, S., Schaefer, T., Aylett, C.H.S., Cimermančić, P., Boehringer, D., Sali, A., Aebersold, R., and Ban, N. (2014). Molecular architecture of the 40S-eIF1-eIF3 translation initiation complex. *Cell* 158, 1123–1135.
- Fraser, C.S., Lee, J.Y., Mayeur, G.L., Bushell, M., Doudna, J.A., and Hershey, J.W. (2004). The β -subunit of human translation initiation factor eIF3 is required for the stable binding of eIF3 and its subcomplexes to 40 S ribosomal subunits in vitro. *J. Biol. Chem.* 279, 8946–8956.
- Guca, E., and Hashem, Y. (2018). Major structural rearrangements of the canonical eukaryotic translation initiation complex. *Curr. Opin. Struct. Biol.* 53, 151–158.
- Hashem, Y., and Frank, J. (2018). The Jigsaw Puzzle of mRNA Translation Initiation in Eukaryotes: A Decade of Structures Unraveling the Mechanics of the Process. *Annu. Rev. Biophys.* 47, 125–151.
- Hashem, Y., des Georges, A., Dhote, V., Langlois, R., Liao, H.Y., Grassucci, R.A., Hellen, C.U., Pestova, T.V., and Frank, J. (2013a). Structure of the mammalian ribosomal 43S preinitiation complex bound to the scanning factor DHX29. *Cell* 153, 1108–1119.
- Hashem, Y., des Georges, A., Fu, J., Buss, S.N., Jossinet, F., Jobe, A., Zhang, Q., Liao, H.Y., Grassucci, R.A., Bajaj, C., et al. (2013b). High-resolution cryo-electron microscopy structure of the *Trypanosoma brucei* ribosome. *Nature* 494, 385–389.
- Herrmannová, A., Prilepskaja, T., Wagner, S., Šikrová, D., Zeman, J., Poncová, K., and Valášek, L.S. (2020). Adapted formaldehyde gradient cross-linking protocol implicates human eIF3d and eIF3c, k and l subunits in the 43S and 48S pre-initiation complex assembly, respectively. *Nucleic Acids Res.* 48, 1969–1984.
- Hinnebusch, A.G. (2017). Structural Insights into the Mechanism of Scanning and Start Codon Recognition in Eukaryotic Translation Initiation. *Trends Biochem. Sci.* 42, 589–611.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph* 14, 33–38, 27–38.
- Karášková, M., Gunišová, S., Herrmannová, A., Wagner, S., Munzarová, V., and Valášek, L. (2012). 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, 28420–28434.
- Kashiwagi, K., Yokoyama, T., Nishimoto, M., Takahashi, M., Sakamoto, A., Yonemochi, M., Shirouzu, M., and Ito, T. (2019). Structural basis for eIF2B inhibition in integrated stress response. *Science* 364, 495–499.
- Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858.
- Kenner, L.R., Anand, A.A., Nguyen, H.C., Myasnikov, A.G., Kloese, C.J., McGeever, L.A., Tsai, J.C., Miller-Vedam, L.E., Walter, P., and Frost, A. (2019). eIF2B-catalyzed nucleotide exchange and phosphoregulation by the integrated stress response. *Science* 364, 491–495.
- Kucukelbir, A., Sigworth, F.J., and Tagare, H.D. (2014). Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* 11, 63–65.
- Laskowski, R.A., Jabłońska, J., Právda, L., Vařeková, R.S., and Thornton, J.M. (2018). PDBsum: Structural summaries of PDB entries. *Protein Sci.* 27, 129–134.
- Lee, A.S., Kranzusch, P.J., and Cate, J.H. (2015). eIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature* 522, 111–114.
- Lee, A.S., Kranzusch, P.J., Doudna, J.A., and Cate, J.H. (2016). eIF3d is an mRNA cap-binding protein that is required for specialized translation initiation. *Nature* 536, 96–99.
- Li, K., Zhou, S., Guo, Q., Chen, X., Lai, D.H., Lun, Z.R., and Guo, X. (2017). The eIF3 complex of *Trypanosoma brucei*: composition conservation does not imply the conservation of structural assembly and subunits function. *RNA* 23, 333–345.
- Llácer, J.L., Hussain, T., Marler, L., Aitken, C.E., Thakur, A., Lorsch, J.R., Hinnebusch, A.G., and Ramakrishnan, V. (2015). Conformational Differences between Open and Closed States of the Eukaryotic Translation Initiation Complex. *Mol. Cell* 59, 399–412.
- Llácer, J.L., Hussain, T., Saini, A.K., Nanda, J.S., Kaur, S., Gordiyenko, Y., Kumar, R., Hinnebusch, A.G., Lorsch, J.R., and Ramakrishnan, V. (2018). Translational initiation factor eIF5 replaces eIF1 on the 40S ribosomal subunit to promote start-codon recognition. *eLife* 7, e39273.
- Luna, R.E., Arthanari, H., Hiraishi, H., Nanda, J., Martin-Marcos, P., Markus, M.A., Akabayov, B., Milbradt, A.G., Luna, L.E., Seo, H.C., et al. (2012). The C-terminal domain of eukaryotic initiation factor 5 promotes start codon recognition by its dynamic interplay with eIF1 and eIF2 β . *Cell Rep.* 1, 689–702.
- Meleppattu, S., Kamus-Elimeleh, D., Zinoviev, A., Cohen-Mor, S., Orr, I., and Shapira, M. (2015). The eIF3 complex of *Leishmania*-subunit composition and mode of recruitment to different cap-binding complexes. *Nucleic Acids Res.* 43, 6222–6235.
- Michaeli, S. (2011). Trans-splicing in trypanosomes: machinery and its impact on the parasite transcriptome. *Future Microbiol.* 6, 459–474.
- Miyashita, M., Oshiumi, H., Matsumoto, M., and Seya, T. (2011). DDX60, a DEXD/H box helicase, is a novel antiviral factor promoting RIG-I-like receptor-mediated signaling. *Mol. Cell Biol.* 31, 3802–3819.
- Obayashi, E., Luna, R.E., Nagata, T., Martin-Marcos, P., Hiraishi, H., Singh, C.R., Erzberger, J.P., Zhang, F., Arthanari, H., Morris, J., et al. (2017). 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.
- Oshiumi, H., Miyashita, M., Okamoto, M., Morioka, Y., Okabe, M., Matsumoto, M., and Seya, T. (2015). DDX60 Is Involved in RIG-I-Dependent and Independent Antiviral Responses, and Its Function Is Attenuated by Virus-Induced EGFR Activation. *Cell Rep.* 11, 1193–1207.
- Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- Phan, L., Zhang, X., Asano, K., Anderson, J., Vornlocher, H.P., Greenberg, J.R., Qin, J., and Hinnebusch, A.G. (1998). Identification of a translation initiation factor 3 (eIF3) core complex, conserved in yeast and mammals, that interacts with eIF5. *Mol. Cell Biol.* 18, 4935–4946.
- Rezende, A.M., Assis, L.A., Nunes, E.C., da Costa Lima, T.D., Marchini, F.K., Freire, E.R., Reis, C.R., and de Melo Neto, O.P. (2014). 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.
- Santos, K.F., Jovin, S.M., Weber, G., Pena, V., Lührmann, R., and Wahl, M.C. (2012). Structural basis for functional cooperation between tandem helicase cassettes in Br2-mediated remodeling of the spliceosome. *Proc. Natl. Acad. Sci. USA* 109, 17418–17423.
- Shah, M., Su, D., Scheliga, J.S., Pluskal, T., Boronati, S., Motamedchaboki, K., Campos, A.R., Qi, F., Hidalgo, E., Yanagida, M., and Wolf, D.A. (2016). A Transcript-Specific eIF3 Complex Mediates Global Translational Control of Energy Metabolism. *Cell Rep.* 16, 1891–1902.
- Simonetti, A., Brito Querido, J., Myasnikov, A.G., Mancera-Martinez, E., Renaud, A., Kuhn, L., and Hashem, Y. (2016). eIF3 Peripheral Subunits

- Rearrangement after mRNA Binding and Start-Codon Recognition. *Mol. Cell* **63**, 206–217.
- Simonetti, A., Guca, E., Bochler, A., Kuhn, L., and Hashem, Y. (2020). Structural Insights into the Mammalian Late-Stage Initiation Complexes. *Cell Rep.* **31**, 107497.
- Simossis, V.A., and Heringa, J. (2004). Integrating protein secondary structure prediction and multiple sequence alignment. *Curr. Protein Pept. Sci.* **5**, 249–266.
- Singh, C.R., Watanabe, R., Chowdhury, W., Hiraishi, H., Murai, M.J., Yamamoto, Y., Miles, D., Ikeda, Y., Asano, M., and Asano, K. (2012). Sequential eukaryotic translation initiation factor 5 (eIF5) binding to the charged disordered segments of eIF4G and eIF2 β stabilizes the 48S preinitiation complex and promotes its shift to the initiation mode. *Mol. Cell. Biol.* **32**, 3978–3989.
- Smith, D.B., and Johnson, K.S. (1988). Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* **67**, 31–40.
- Smith, M.D., Arake-Tacca, L., Nitido, A., Montabana, E., Park, A., and Cate, J.H. (2016). Assembly of eIF3 Mediated by Mutually Dependent Subunit Insertion. *Structure* **24**, 886–896.
- Stolboushkina, E., Nikonov, S., Nikulin, A., Bläsi, U., Manstein, D.J., Fedorov, R., Garber, M., and Nikonov, O. (2008). 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**, 680–691.
- Sun, C., Todorovic, A., Querol-Audí, J., Bai, Y., Villa, N., Snyder, M., Ashchyan, J., Lewis, C.S., Hartland, A., Gradia, S., et al. (2011). Functional reconstitution of human eukaryotic translation initiation factor 3 (eIF3). *Proc. Natl. Acad. Sci. USA* **108**, 20473–20478.
- Thakur, A., Marler, L., and Hinnebusch, A.G. (2019). A network of eIF2 β interactions with eIF1 and Met-tRNA_i promotes accurate start codon selection by the translation preinitiation complex. *Nucleic Acids Res.* **47**, 2574–2593.
- Trabuco, L.G., Villa, E., Mitra, K., Frank, J., and Schulten, K. (2008). Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683.
- Valášek, L.S. (2012). ‘Ribozomin’—translation initiation from the perspective of the ribosome-bound eukaryotic initiation factors (eIFs). *Curr. Protein Pept. Sci.* **13**, 305–330.
- Valášek, L., Trachsel, H., Hašek, J., and Ruis, H. (1998). Rpg1, the *Saccharomyces cerevisiae* homologue of the largest subunit of mammalian translation initiation factor 3, is required for translational activity. *J. Biol. Chem.* **273**, 21253–21260.
- Valášek, L., Phan, L., Schoenfeld, L.W., Valášková, V., and Hinnebusch, A.G. (2001). Related eIF3 subunits TIF32 and HCR1 interact with an RNA recognition motif in PRT1 required for eIF3 integrity and ribosome binding. *EMBO J.* **20**, 891–904.
- Valášek, L., Nielsen, K.H., Zhang, F., Fekete, C.A., and Hinnebusch, A.G. (2004). 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**, 9437–9455.
- Valášek, L.S., Zeman, J., Wagner, S., Beznosková, P., Pavlíková, Z., Mohammad, M.P., Hronová, V., Herrmannová, A., Hashem, Y., and Gunišová, S. (2017). Embraced by eIF3: structural and functional insights into the roles of eIF3 across the translation cycle. *Nucleic Acids Res.* **45**, 10948–10968.
- Wagner, S., Herrmannová, A., Malík, R., Peclínová, L., and Valášek, L.S. (2014). Functional and biochemical characterization of human eukaryotic translation initiation factor 3 in living cells. *Mol. Cell. Biol.* **34**, 3041–3052.
- Wagner, S., Herrmannová, A., Šikrová, D., and Valášek, L.S. (2016). 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.* **44**, 10772–10788.
- Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303.
- Wei, Z., Xue, Y., Xu, H., and Gong, W. (2006). Crystal structure of the C-terminal domain of *S.cerevisiae* eIF5. *J. Mol. Biol.* **359**, 1–9.
- Yu, Y., Marintchev, A., Kolupaeva, V.G., Unbehauen, A., Varyasova, T., Lai, S.-C., Hong, P., Wagner, G., Hellen, C.U.T., and Pestova, T.V. (2009). Position of eukaryotic translation initiation factor eIF1A on the 40S ribosomal subunit mapped by directed hydroxyl radical probing. *Nucleic Acids Res.* **37**, 5167–5182.
- Zeman, J., Itoh, Y., Kukačka, Z., Rosůlek, M., Kavan, D., Kouba, T., Jansen, M.E., Mohammad, M.P., Novák, P., and Valášek, L.S. (2019). 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.
- Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12.
- Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCorr2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332.
- Zivanov, J., Nakane, T., Forsberg, B.O., Kimanius, D., Hagen, W.J., Lindahl, E., and Scheres, S.H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cell lines		
<i>Trypanosoma cruzi</i> strain Y	This paper	N/A
<i>Leishmania tarentolae</i> strain T7-TR	Jena Bioscience	Cat#LT-110
<i>Escherichia coli</i> (One Shot BL21 Star (DE3) Chemically Competent <i>E. coli</i>)	Invitrogen	Cat#C601003
<i>Escherichia coli</i> (Rosetta 2(DE3) Singles Competent Cells-Novagen)	Novagen	Cat#71400
Chemicals		
GMP-PNP	Sigma	Cat#G0635
Protease inhibitor cocktail tablets	Roche	Cat#11873580001
RNasin® Ribonuclease Inhibitors	Promega	Cat#N251B
TNT T7 Quick Coupled Transcription/ Translation System	Promega	Cat # L1170
Glutathione Sepharose® 4B	GE Healthcare	Cat # GE17-0756-01
Gelcode Blue stain reagent	ThermoFisher	Cat # 24592
4–20% Criterion TGX Precast Midi Protein Gel	Bio-Rad	Cat # 5671093
Easy Tag L [35S] Methionine 1mCi (37MBq)	Perkinelmer	NEG709A001MC
Recombinant DNA		
<i>Trypanosoma cruzi</i> (DNA source for cloning of <i>t.c.</i> genes)	a gift from the lab of Julius Lukeš	N/A
<i>Homo sapiens</i> HeLa cell line (cDNA source for cloning of <i>h.s.</i> genes)	a gift from the lab of Tomáš Vomastek	N/A
pGL4-CMV-h3c	This paper	N/A
pGL4-CMV-h3a	This paper	N/A
pGL4-CMV-h3m	This paper	N/A
pGL4-CMV-h3k	This paper	N/A
pGL4-CMV-h3d	This paper	N/A
pGL4-CMV-h3e	This paper	N/A
pGEX-helF1	This paper	N/A
pGEX-helF5	This paper	N/A
pGEX-helF2β	This paper	N/A
pGL4-CMV-h3c-1-325	This paper	N/A
pGL4-CMV-h3c-326-913	This paper	N/A
pGL4-CMV-h3c-30-325	This paper	N/A
pGL4-CMV-h3c-130-325	This paper	N/A
pGL4-CMV- h3c-1-325-d171-240	This paper	N/A
pGL4-CMV-eIF2β	This paper	N/A
pGL4-CMV-eIF2β-1-309	This paper	N/A
pGEX-helF1-box-Ala-102-113	This paper	N/A
pGEX-telF1	This paper	N/A
pGEX-telF5	This paper	N/A
pGL4-CMV-telF5	This paper	N/A
pGL4-CMV-telF3c	This paper	N/A
pGL4-CMV-telF3c-1-172	This paper	N/A
pGL4-CMV-telF3c-14-172	This paper	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pGL4-CMV-telF3c-39-172	This paper	N/A
pGEX-telF3c-1-172	This paper	N/A
pGEX-helF3e	This paper	N/A
pGEX-helF3e-del-244-252	This paper	N/A
pGEX-helF3e-l246A-Q247A-T248A	This paper	N/A
pGL4-CMV-h3d-W16A-G17A-P18A	This paper	N/A
pGL4-CMV-h3d-19-548	This paper	N/A
pGL4-CMV-h3d-1-114	This paper	N/A
pGL4-CMV-h3d-19-114	This paper	N/A
pGEX-helF3d	This paper	N/A
pGEX-helF3c	This paper	N/A
pGEX-helF3a	This paper	N/A
pEX-telF3c-1-172-GST	This paper	N/A
pEX-telF5-GST	This paper	N/A
YcplV018	Valásek et al., 1998	N/A
pGEX-5X-3	Smith and Johnson, 1988	N/A
pGL4-CMV	Wagner et al., 2014	N/A
pFASTBAC1-eIF3k	Fraser et al., 2004	N/A
pFASTBAC1-eIF3d	Fraser et al., 2004	N/A
pFASTBAC1-eIF3e	Fraser et al., 2004	N/A
pEX-C-GST	OriGene	Cat#PS100083
Oligonucleotides		
Primers for cloning used in this study, see List of primers (Supplemental Information)	This paper	N/A
GeneArt Strings DNA Fragments used for cloning, see List of gene-strings (Supplemental Information)	Invitrogen	N/A
Deposited Data		
Structure of <i>T. cruzi</i> 43S – Cryo-EM map	This paper	EMDB-11893
Structure of <i>T. cruzi</i> 43S – coordinates	This paper	PDBID: 7ASE
Structure of <i>T. cruzi</i> 43S + ATP – Cryo-EM map	This paper	EMDB-11895
k-DDX60 structure from <i>T. cruzi</i> 43S + ATP – coordinates	This paper	PDBID: 7ASK
Structure of <i>L. tarentolae</i> 43S – Cryo-EM map	This paper	EMDB-11896
Mass spectrometry analysis	This paper	PXD016063
Software and Algorithms		
Molecular Dynamic Flexible Fitting	Trabuco et al., 2008	https://www.ks.uiuc.edu/Research/mdff/
MotionCor	Zheng et al., 2017	https://emcore.ucsf.edu/ucsf-software
Gctf	Zhang, 2016	https://www2.mrc-lmb.cam.ac.uk/research/locally-developed-software/zhang-software/#gctf
RELION 3.0	Zivanov et al., 2018	https://www2.mrc-lmb.cam.ac.uk/groups/scheres/impact.html
UCSF Chimera	Pettersen et al., 2004	https://www.cgl.ucsf.edu/chimera/ ; RRID:SCR_004097
Coot	Emsley and Cowtan, 2004	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/ ; RRID:SCR_014222
phenix.real_space_refine	Afonine et al., 2018	http://www.phenix-online.org/
PHENIX 1.9.1692	Adams et al., 2010	http://www.phenix-online.org/
Phenix.ERRASER	Chou et al., 2013	http://www.phenix-online.org/

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Visual Molecular Dynamics	Humphrey et al., 1996	https://www.ks.uiuc.edu/Research/vmd/ ; RRID:SCR_001820
RESMAP	Kucukelbir et al., 2014	http://resmap.sourceforge.net
Quantity One 4.6.9	Bio-Rad	RRID:SCR_014280

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for reagents should be direct to Yaser Hashem (yaser.hashem@u-bordeaux.fr).

Materials Availability

All plasmids generated in this study are available from the Lead Contact without restriction. This study did not generate new unique reagents.

Data and Code Availability

The accession numbers of the cryo-EM maps of the 43S+GMPPNP, 43S+GMPPNP+ATP PICs from *T. cruzi* and the 43S+GMPPNP PIC from *L. tarentolae* reported in this study have been deposited to the Electron Microscopy Data Bank (EMDB): EMD-11893, EMD-11895 and EMD-11896. The accession numbers of the atomic model of the 43S+GMPPNP PIC from *T. cruzi* reported in this study was deposited to the Protein Data Bank (PDB): 7ASE. The k-DDX60 atomic model was fitted in its density from the 43S+GMPPNP+ATP PIC and deposited to the PDB: 7ASK. The accession numbers of the datasets generated from mass-spectrometry analysis of the *T. cruzi* 43S PIC and *L. tarentolae* 43S PIC reported in this study were deposited to the PRIDE partner repository: PXD016063.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Trypanosoma cruzi

Trypanosoma cruzi strain Y - TcII was used in this study. Epimastigotes were grown at 28°C in liver infusion tryptose (LIT) medium, supplemented with 10% heat-inactivated fetal bovine serum.

Leishmania tarentolae

Leishmania tarentolae strain T7-TR (Jena Bioscience Cat#LT-110) was used. Culture was 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.

Escherichia coli

One Shot BL21 Star (DE3) Chemically Competent *E. coli* (Invitrogen Cat#C601003) and Rosetta 2(DE3) Singles Competent Cells-Novagen (Cat#71400) were used in this study for expression of GST-tagged proteins. Bacterial cultures were grown in the Luria Broth (LB) medium supplemented with ampicillin or ampicillin and chloramphenicol, respectively.

Homo sapiens genetic material

cDNA used as template for subcloning of the selected *H.s.* genes by PCR was obtained from the female HeLa cell line.

METHOD DETAILS

Construction of plasmids

List of all primers and gene strings used throughout this study is shown in [Tables S5](#) and [S6](#).

pGL4-CMV-h3c was made by inserting the *PmeI-FseI* digested PCR product obtained with primers AH-h3c-*PmeI* and AH-h3c-*FseI* using HeLa cDNA as a template into *PmeI-FseI* digested pGL4-CMV ([Wagner et al., 2014](#)).

pGL4-CMV-h3a was made by inserting the *PmeI-FseI* digested PCR product obtained with primers AH-h3a-*PmeI* and AH-h3a-*FseI* using YCpLV018 ([Valášek et al., 1998](#)) as a template into *PmeI-FseI* digested pGL4-CMV ([Wagner et al., 2014](#)).

pGL4-CMV-h3m was made by inserting the *EcoRI-FseI* digested PCR product obtained with primers AH-h3m-*EcoRI* and AH-h3m-*FseI* using HeLa cDNA as a template into *EcoRI-FseI* digested pGL4-CMV ([Wagner et al., 2014](#)).

pGL4-CMV-h3k was made by inserting the *EcoRI-FseI* digested PCR product obtained with primers AH-h3k-*EcoRI* and AH-h3k-*FseI* using pFASTBAC1-eIF3k ([Fraser et al., 2004](#)) as a template into *EcoRI-FseI* digested pGL4-CMV ([Wagner et al., 2014](#)).

pGL4-CMV-h3d was made by inserting the EcoRI-FseI digested PCR product obtained with primers AH-h3d-EcoRI and AH-h3d-FseI using pFASTBAC1-eIF3d (Fraser et al., 2004) as a template into EcoRI-FseI digested pGL4-CMV (Wagner et al., 2014).

pGL4-CMV-h3e was made by inserting the EcoRI-FseI digested PCR product obtained with primers AH-h3e-EcoRI and AH-h3e-FseI using pFASTBAC1-eIF3e (Fraser et al., 2004) as a template into EcoRI-FseI digested pGL4-CMV (Wagner et al., 2014).

pGEX-helF1 was made by inserting the BamHI-SalI digested PCR product obtained with primers DS-eIF1-BamHI and DS-eIF1-SalI using HeLa cDNA as a template into BamHI-SalI digested pGEX-5X-3 (Smith and Johnson, 1988).

pGEX-helF5 was made by inserting the EcoRI-SalI digested PCR product obtained with primers SW-helF5-EcoRI and SW-helF5-SalI-R using HeLa cDNA as a template into EcoRI-SalI digested pGEX-5X-3 (Smith and Johnson, 1988).

pGEX-helF2 β was made by inserting the BamHI-SalI digested PCR product obtained with primers DS-eIF2 β -BamHI and DS-eIF2 β -SalI using HeLa cDNA as a template into BamHI-SalI digested pGEX-5X-3 (Smith and Johnson, 1988).

pGL4-CMV-h3c-1-325; pGL4-CMV-h3c-326-913; pGL4-CMV-h3c-30-325 and pGL4-CMV-h3c-130-325 was made by inserting the PmeI-FseI digested PCR product obtained with primers AH-h3c-PmeI and AH-h3c-325-FseI; TS-h3c-326-PmeI and AH-h3c-FseI; TS-h3c-30-325-PmeI and AH-h3c-325-FseI; TP-h3c-130-325-PmeI and AH-h3c-325-FseI; respectively, using pGL4-CMV-h3c as a template into PmeI-FseI digested pGL4-CMV-h3c.

pGL4-CMV-h3c-1-325-d171-240 was made by inserting the PmeI-FseI digested gene string pGL4-CMV-h3c-1-325-d171-240 (GeneArt Strings DNA Fragments, Invitrogen) into PmeI-FseI digested pGL4-CMV (Wagner et al., 2014).

pGL4-CMV-eIF2 β and pGL4-CMV-eIF2 β -1-309 was made by inserting the EcoRI-FseI digested PCR product obtained with primers TP pGL4-CMV-helF2 β -EcoRI and TP-pGL4-CMV-eIF2 β -FseI; TP-pGL4-CMV-helF2 β -EcoRI and TP-pGL4-CMV-eIF2 β -1-309-FseI; respectively, using pGEX-helF2 β as a template into EcoRI-FseI digested pGL4-CMV (Wagner et al., 2014).

pGEX-helF1-box-Ala-102-113 was made by inserting the BamHI-SalI digested gene string helF1-box-Ala-102-113 (GeneArt Strings DNA Fragments, Invitrogen) into BamHI-SalI digested pGEX-helF1.

pGEX-telF1 and pGEX-telF5 was made by inserting the BamHI-SalI digested PCR product obtained with primers TP-pGEX-5X3-telF1-BamHI and TP-pGEX-5X3-telF1-SalI; TP-pGEX-5X3-telF5-BamHI and TP-pGEX-5X3-telF5-SalI; respectively, using *T. cruzi* genomic DNA as a template into BamHI-SalI digested pGEX-5X-3 (Smith and Johnson, 1988).

pGL4-CMV-telF5 was made by inserting EcoRI-FseI digested PCR product obtained with primers TP-pGL4-telF5-EcoRI and TP-pGL4-telF5-FseI using pGEX-telF5 as a template into EcoRI-FseI digested pGL4-CMV (Wagner et al., 2014).

pGL4-CMV-telF3c was made by inserting the PmeI-FseI digested PCR product obtained with primers TP-pGL4-CMV-telF3c-PmeI and TP-pGL4-CMV-telF3c-FseI using *T. cruzi* genomic DNA as a template into PmeI-FseI digested pGL4-CMV (Wagner et al., 2014).

pGL4-CMV-telF3c-1-172; pGL4-CMV-telF3c-14-172; pGL4-CMV-telF3c-39-172 was made by inserting the PmeI-FseI digested PCR product obtained with primers TP-pGL4-CMV-telF3c-PmeI and TP-pGL4-CMV-telF3c-172-FseI; TP-pGL4-CMV-telF3c-14-PmeI and TP-pGL4-CMV-telF3c-172-FseI; TP-pGL4-CMV-telF3c-39-PmeI and TP-pGL4-CMV-telF3c-172-FseI; respectively, using pGL4-CMV-telF3c as a template into PmeI-FseI digested pGL4-CMV (Wagner et al., 2014).

pGEX-telF3c-1-172 was made by inserting BamHI-EcoRI digested PCR product obtained with primers TP-pGEX-telF3c-BamHI and TP-pGEX-telF3c-1-172-EcoRI using pGL4-CMV-telF3c-1-172 as a template into BamHI-EcoRI digested pGEX-5X-3 (Smith and Johnson, 1988).

pGEX-helF3e was made by inserting the BamHI-SalI digested PCR product obtained with primers TP-pGEX-5X3-eIF3e-BamHI and TP-pGEX-5X3-eIF3e-SalI using pGL4-CMV-h3e as a template into BamHI-SalI digested pGEX-5X-3 (Smith and Johnson, 1988).

pGEX-helF3e-del-244-252 and pGEX-helF3e-I246A-Q247A-T248A was made by inserting the BamHI-BglII digested gene string pGEX-helF3e-delta-244-252 and gene string pGEX-helF3e-I246A-Q247A-T248A (GeneArt Strings DNA Fragments, Invitrogen) respectively, into BamHI-BglII digested pGEX-5X3-helF3e.

pGL4-CMV-h3d-W16A-G17A-P18A was made by inserting the EcoRI-PvuII digested gene string pGL4-CMV-h3d-W16A-G17A-P18A (GeneArt Strings DNA Fragments, Invitrogen) into EcoRI-PvuII digested pGL4-CMV-h3d.

pGL4-CMV-h3d-19-548; pGL4-CMV-h3d-1-114; pGL4-CMV-h3d-19-114 was made by inserting the EcoRI-FseI digested PCR product obtained with primers TP-pGL4-CMV-h3d-19-EcoRI and AH-h3d-FseI; AH-h3d-EcoRI and TP-pGL4-CMV-h3d-114-FseI; TP-pGL4-CMV-h3d-19-EcoRI and TP-pGL4-CMV-h3d-114-FseI; respectively, using pGL4-CMV-h3d as a template into EcoRI-FseI digested pGL4-CMV (Wagner et al., 2014).

pGEX-helF3d was made by inserting the BamHI-SalI digested PCR product obtained with primers TP-pGEX-helF3d-BamHI and TP-pGEX-helF3d-SalI using pGL4-CMV-h3d as a template into BamHI-SalI digested pGEX-5X-3 (Smith and Johnson, 1988).

pGEX-helF3c was made by inserting the EcoRI-SalI digested PCR product obtained with primers TP-pGEX-eIF3c-EcoRI and TP-pGEX-eIF3c-SalI using pGL4-CMV-h3c as a template into EcoRI-SalI digested pGEX-5X-3 (Smith and Johnson, 1988).

pGEX-helF3a was made by inserting the SalI-NotI digested PCR product obtained with primers TP-pGEX-eIF3a-SalI and TP-pGEX-eIF3a-NotI using pGL4-CMV-h3a as a template into SalI-NotI digested pGEX-5X-3 (Smith and Johnson, 1988).

pEX-telF3c-1-172-GST was made by inserting the AsiSI-MluI digested PCR product obtained with primers TP-telF3c-AsiSI and TP-telF3c-1-172-MluI using pGL4-CMV-telF3c-1-172 as a template into BseRI-MluI digested pEX-C-GST (OriGene; PS100083).

pEX-telF5-GST was made by inserting the AsiSI-MluI digested PCR product obtained with primers TP-telF5-AsiSI and TP-telF5-MluI using pGEX-telF5 as a template into BseRI-MluI digested pEX-C-GST (OriGene; PS100083).

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. tarentolae*, 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 10 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 (35 000 rpm, 5h30 min, 4°C) using an SW41 Ti rotor (Beckman-Coulter). The fractions containing 48S ICs were collected and pooled according to 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 NH₄Cl, 5 mM Mg(OAc)₂, and 2 mM DTT). For the ATP supplemented 43S PIC, the protocol above was repeated for *T. cruzi* with an addition of 10 mM 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 s and then blotted with filter paper for 1.5 s 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 collections of the three described molecular complexes were performed on three different instruments. The main complex (*T. cruzi* 43S PIC) was imaged (at the IGBMC EM facility, Illkirch, France) 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 e⁻/Å². The *T. cruzi* 43S PIC supplemented with ATP was imaged with the exact setup described above, but in the Netherlands's NeCEN EM facility, Leiden, which is not Cs corrected. The *L. tarentolae* 43S PIC dataset was collected (at the IECB EM facility, Pessac, France) on a Talos Artica instrument (FEI Company) at 200 kV using the EPU software (FEI Company) for automated data acquisition. Data were collected at a nominal underfocus of -0.5 to -2.7 μm at a magnification of 120,000 X yielding a pixel size of 1.21 Å. Micrographs were recorded as movie stack on a Falcon III direct electron detector (FEI Company), each movie stack were fractionated into 20 frames for a total exposure of 1 s corresponding to an electron dose of 40 e⁻/Å².

Image processing

For all three datasets, drift and gain correction and dose weighting were performed using MotionCor2 (Zheng et al., 2017). A dose weighted average image of the whole stack was used to determine the contrast transfer function with the software Gctf (Zhang, 2016). The following process has been achieved using RELION 3.0 (Zivanov et al., 2018). Particles were picked using a Laplacian of Gaussian function (min diameter 300 Å, max diameter 320 Å). For the main dataset (*T. cruzi* 43S PIC), 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 possible 48S complex (12910 particles, don't present densities for k-DDX60) and a second as a 43S+DDX60 complex (33775 particles). Refinement of the 43S-DDX60 complex yielded an average resolution of 3.3Å. The 48S class was not analyzed any further. Determination of the local resolution of the final density map was performed using ResMap (Kucukelbir et al., 2014). The dataset of the *T. cruzi* 43S PIC supplemented with ATP was processed identically. However, the sample was more diluted compared to the above-described main complex, thus yielding less particles count after the first 2D classification (98,840 particles presenting 40-like shape). Following the similar classification/processing fashion and after 3D classification, 19700 particles were used to reconstruct a ~ 4.3 Å 43S PIC bound to ATP. Finally, the *L. tarentolae* 43S PIC dataset was processed also identically to the protocol described above. As the aim of this reconstruction is simply to validate the conservation of the architecture in leishmania, only a small dataset was collected, which after processing only yielded $\sim 10,000$ particles that were then used to reconstruct the 43S PIC at 8.1Å.

Figure preparation

Figures featuring cryo-EM densities as well as atomic models were visualized with UCSF Chimera (Pettersen et al., 2004).

Mass spectrometry analysis

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 previously⁵. Each sample was further analyzed by nanoLC-MS/MS on a QExactive+ mass spectrometer coupled to an EASY-nanoLC-1000 (Thermo-Fisher Scientific, USA). Pep-

tides 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://www.profi-proteomics.fr/proline>).

The mass spectrometric data were deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD016063 (Reviewer account details: reviewer06222@ebi.ac.uk, rhv9KZXk).

Model building and refinement

The atomic model of the preinitiation complex 48S from *Trypanosoma cruzi* was built using the modeling softwares Chimera (Pettersen et al., 2004), Coot (Emsley and Cowtan, 2004), Phenix (Adams et al., 2010) and VMD (Humphrey et al., 1996).

The previous 40S structure of *Trypanosoma cruzi* (Brito Querido et al., 2017) (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 obtained structure.

The ternary complex (tRNA, eIF2 α , eIF2 γ), eIF2 β , eIF1a and eIF1 were thread from the translation initiation complex of yeast (Llácer et al., 2015) (PDBID: 3JAJ).

DDX60-like starting point was the recA domains from the human helicase protein Brr2 (Santos et al., 2012) (PDBID: 4F93). The remaining domains of DDX60-like was built *ab initio* using Coot modeling tools and Chimera “build structure” tools with the help of sympred (Simossis and Heringa, 2004) for secondary structure prediction and the homology modeling web services Swissmodel (Waterhouse et al., 2018) and phyre2 (Kelley et al., 2015).

eIF3 was thread from the already published mammalian eIF3 (des Georges et al., 2015) (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* (Lee et al., 2016) (PDBID: 5K4B) and the N-terminal tail was built in Chimera.

eIF5 Cter-domain was thread from the eIF5 crystal from human (Bieniossek et al., 2006) (PDBID: 2IU1).

The global atomic model was refined using the Molecular Dynamic Flexible Fitting (Trabuco et al., 2008) then the geometry parameters were corrected using PHENIX real space refine for proteins and eraser (Chou et al., 2013) for RNA.

Secondary structures of k-DDX60 and the 18S

The secondary structure of the 18S was done based on the S.c. 18S template downloaded from the RiboVision Webservice (Bernier et al., 2014). The secondary structures of the 18S expansion segments were edited manually based on the 3D atomic model of the complex. The sequence and residues numbering were corrected consistently with *T. cruzi*.

The secondary structure of k-DDX60 was derived from its 3D atomic model (this work) using the PDBsum Webservice (Laskowski et al., 2018).

GST pulldown assay

Glutathione S-transferase (GST) pull down experiments with GST fusions and *in vitro* synthesized ³⁵S-labeled polypeptides were conducted as described previously (Valásek et al., 2001). 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. ³⁵S-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 ³⁵S-labeled proteins at 4°C for 2 h in buffer B (20mM HEPES (pH 7.5), 75mM KCl, 0.1mM EDTA, 2.5mM MgCl₂, 0.05% IGEPAL, 1mM DTT). For experiments requiring more stringent conditions the buffer B was supplemented 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 AND STATISTICAL ANALYSIS

For the mass-spectrometry analysis, 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.

Volcano plot presented in Figure 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 analyzing the pre-initiation complex after further purification step using size exclusion chromatography.

Quantification of binding experiments (GST-pulldown assay) was done using the Quantity One software (Bio-Rad). The data was generated as an adjusted volume with the local background subtraction and linear regression methods. The data for each

³⁵S-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 [Figure 3J](#): ³⁵S-eIF3d WT – GST-eIF3e WT; and for [Figure 3K](#): ³⁵S-eIF3d 1-114 – GST-eIF3e WT) and means from three different dilutions of GST-fusions were calculated; errors bars indicate standard deviation. Statistical details of these experiments can be found in the figure legend for [Figure 3](#).

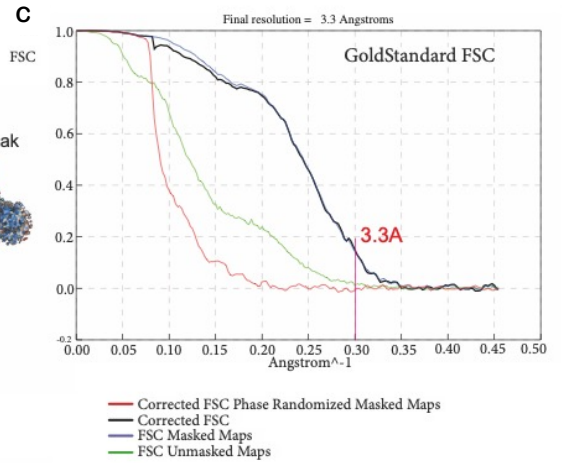
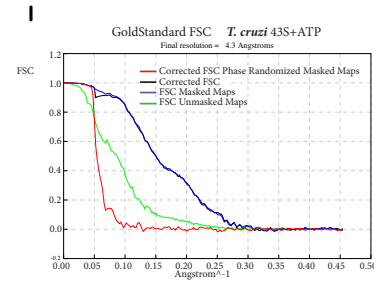
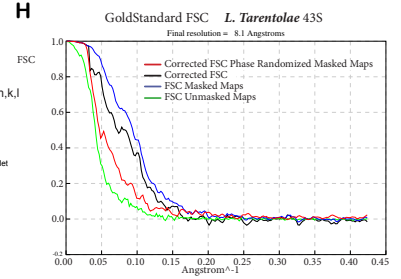
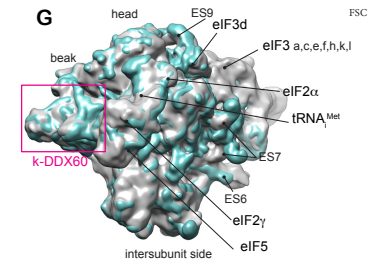
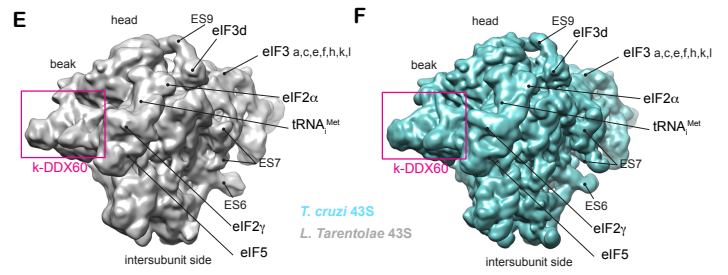
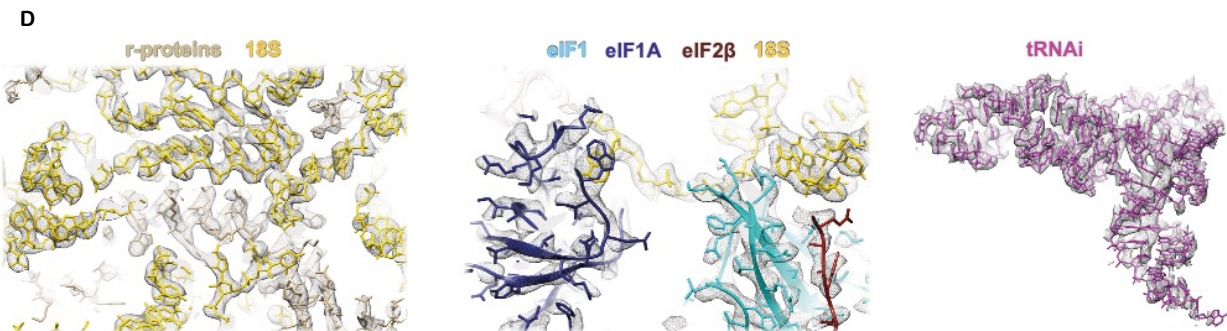
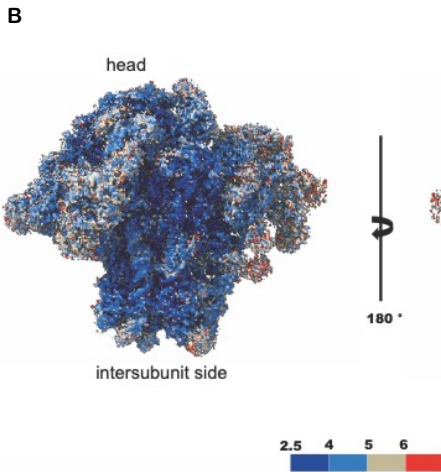
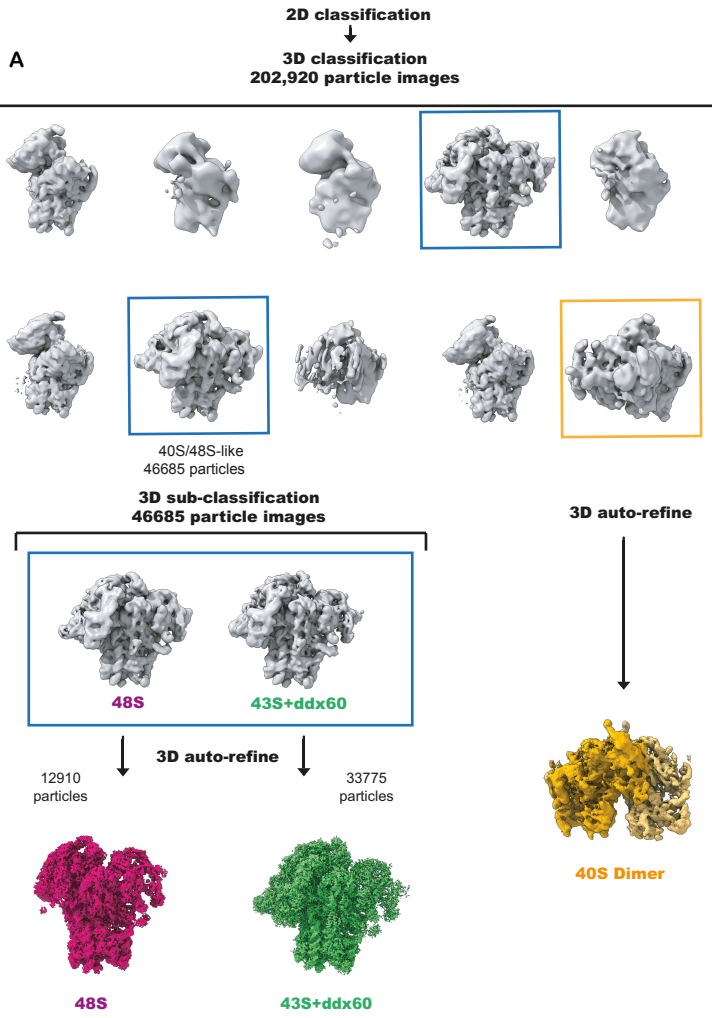
Cell Reports, Volume 33

Supplemental Information

**Structural Differences in Translation Initiation
between Pathogenic Trypanosomatids
and Their Mammalian Hosts**

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

SUPPLEMENTARY FIGURES



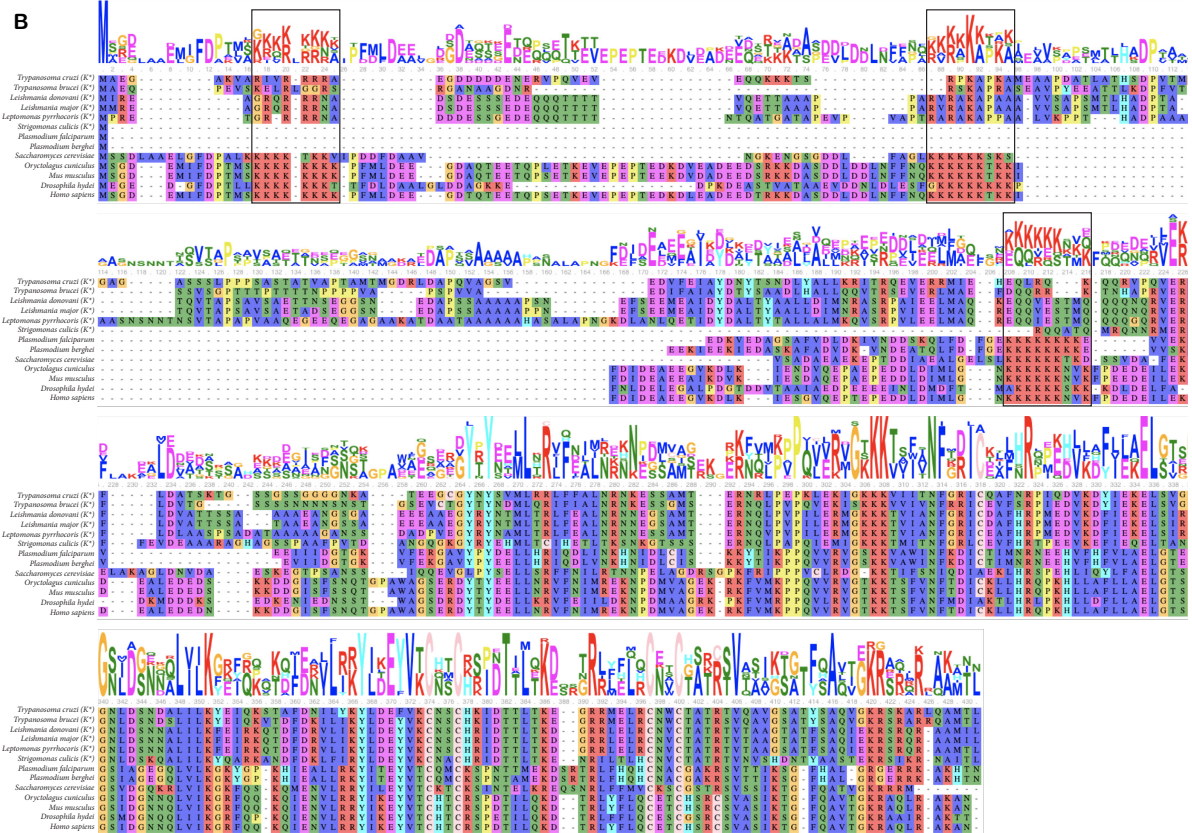
Supplementary Figure 1. Cryo-EM particle sorting and refinement of the 43S PIC complexes from *T. cruzi* and *L. tarentolae* and their resolutions, Related to Figure 1. (A) 2D classification of the 43S PIC particles yielded ~200 000 40S-like particles from the *T. cruzi* dataset, after which a run of 3D classification (10 classes) was performed. (B) The local resolution of the 43S class varies mainly on eIF3 (ranging from ~3 to ~6 Å), while it 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). (C) The average resolution was measured after applying a soft-edge mask of the 43S PIC shape filtered to 15Å and extended by 3 pixels. (D) Blow ups on several features of the complex counting 40S rRNA/r-proteins (left), 18S rRNA interaction with eIF 1 and 1A (middle) and the initiator tRNA^{Met} (right), fitted in their corresponding densities. (E) Cryo-EM reconstructions of the *L. tarentolae* 43S PIC. (F) Cryo-EM reconstructions of the *T. cruzi* 43S PIC filtered at 8Å. (G) Superimposition of (E) and (F). (H) Average resolution (8.1Å) of the *L. tarentolae* 43S PIC reconstruction. (I) Average resolution (4.3Å) of the cryo-EM reconstruction from the *T. cruzi* 43S complexes supplemented with ATP.

A

<i>Leishmania donovani</i> (K)	-----MRDINRQGVF	12
<i>Leishmania major</i> (K)	MASYCVTDSPEVYDKTKCCRTDNNVYVAFKFEYRFLNGSLARRKLLAEFFVPEVDCRTLDHLLVLEKATILSAQV---EGETKGSNNRPEFWMRDINRQGVF	106
<i>Trypanosoma brucei</i> (K)	MAAYGIVESPENVDYKTKCCQCTDGVYFCVFEYRFLHANSRRLKLLAEFFVPEVPLDSSAFESLVVLEKAAVSPAATAGTEVVSQGGDA-RQGVKIDINRQGVF	109
<i>Trypanosoma cruzi</i> (K)	MASHGVPKVIERYVDYKTKCCRTDDEAYGAFKGFVHFHTLSRRLKLLAEFFVPEVPLDSSAFESLVVLEKAAVSPAATAGTEVVSQGGDA-RQGVKIDINRQGVF	110
<i>Strigomonas culicis</i> (K)	MTSYRIIDNPFVDYTKCCARCTDGIYVICFKDFRFLHFTVRRLLLEFFVPEVDSMTNPLLLALEFVALLSIQATGGDSTDEE-----RLEFMRMLSKPQLVEF	104
<i>Drosophila hydei</i>	-----MALT-SR	6
<i>Homo sapiens</i>	-----MFLIS-CR	7
<i>Mus musculus</i>	-----MFLIS-CR	7
<i>Oryctolagus cuniculus</i>	-----MFLIS-CR	7
<i>Saccharomyces cerevisiae</i>	-----MSTSH-CR	7
<i>Plasmodium berghei</i>	-----MGDARSKTLDGDCR	14
<i>Plasmodium falciparum</i>	-----MTEMRVKADLGDCR	14

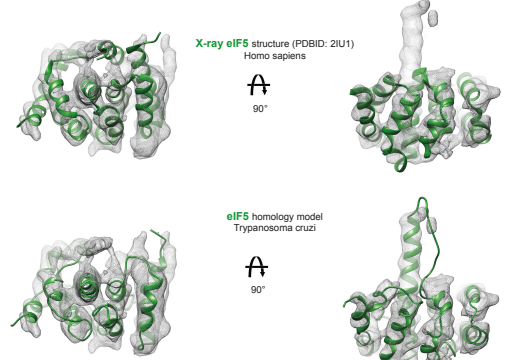
<i>Leishmania donovani</i> (K)	CGCLGITSDWGRD-IFVYVETMPKINDVWVKITQVNDTSAVVGLLEYGHRGELIIPYETVTRRRVSRMGRLIKVGRTEPAQVIRIDPKRGVIDLKSKLVTENKAKAEAH	121
<i>Leishmania major</i> (K)	CGCLGITSDWGRD-IFVYVETMPKINDVWVKITQVNDTSAVVGLLEYGHRGELIIPYETVTRRRVSRMGRLIKVGRTEPAQVIRIDPKRGVIDLKSKLVTENKAKAEAH	215
<i>Trypanosoma brucei</i> (K)	CGCLGITSDWGRD-IFVYVETMPKINDVWVKITQVNDTSAVVGLLEYGHRGELIIPYETVTRRRVSRMGRLIKVGRTEPAQVIRIDPKRGVIDLKSKLVTENKAKAEAH	218
<i>Trypanosoma cruzi</i> (K)	CGCLGITSDWGRD-IFVYVETMPKINDVWVKITQVNDTSAVVGLLEYGHRGELIIPYETVTRRRVSRMGRLIKVGRTEPAQVIRIDPKRGVIDLKSKLVTENKAKAEAH	219
<i>Strigomonas culicis</i> (K)	CGCLGITSDWGRD-IFVYVETMPKINDVWVKITQVNDTSAVVGLLEYGHRGELIIPYETVTRRRVSRMGRLIKVGRTEPAQVIRIDPKRGVIDLKSKLVTENKAKAEAH	214
<i>Drosophila hydei</i>	-----FYNERKYPEIEDVVMVNLISIAEMGAVVHLEYNNIEGMILLSELSSRRIRSNKILIRVGRTEFVWVIRVDRKGVIDLKSKRFSSEDEVEKTEK	100
<i>Homo sapiens</i>	-----FYCHRFFPEVDVVMVNRVSIEMGAVVSLIYNNIEGMILLSELSSRRIRSNKILIRVGRTEFVWVIRVDRKGVIDLKSKRFSSEDEVEKTEK	101
<i>Mus musculus</i>	-----FYCHRFFPEVDVVMVNRVSIEMGAVVSLIYNNIEGMILLSELSSRRIRSNKILIRVGRTEFVWVIRVDRKGVIDLKSKRFSSEDEVEKTEK	101
<i>Oryctolagus cuniculus</i>	-----FYCHRFFPEVDVVMVNRVSIEMGAVVSLIYNNIEGMILLSELSSRRIRSNKILIRVGRTEFVWVIRVDRKGVIDLKSKRFSSEDEVEKTEK	101
<i>Saccharomyces cerevisiae</i>	-----FYENRYPEIDDVVMVNCQIEMGAVVSLIYNNIEGMILLSELSSRRIRSNKILIRVGRTEFVWVIRVDRKGVIDLKSKRFSSEDEVEKTEK	101
<i>Plasmodium berghei</i>	-----FYEKKFFPEVDLIMVKNRIEMGAVVSLIYNNIEGMILLSELSSRRIRSNKILIRVGRTEFVWVIRVDRKGVIDLKSKRFSSEDEVEKTEK	108
<i>Plasmodium falciparum</i>	-----FYRKKFFPEVDLIMVKNRIEMGAVVSLIYNNIEGMILLSELSSRRIRSNKILIRVGRTEFVWVIRVDRKGVIDLKSKRFSSEDEVEKTEK	108

B

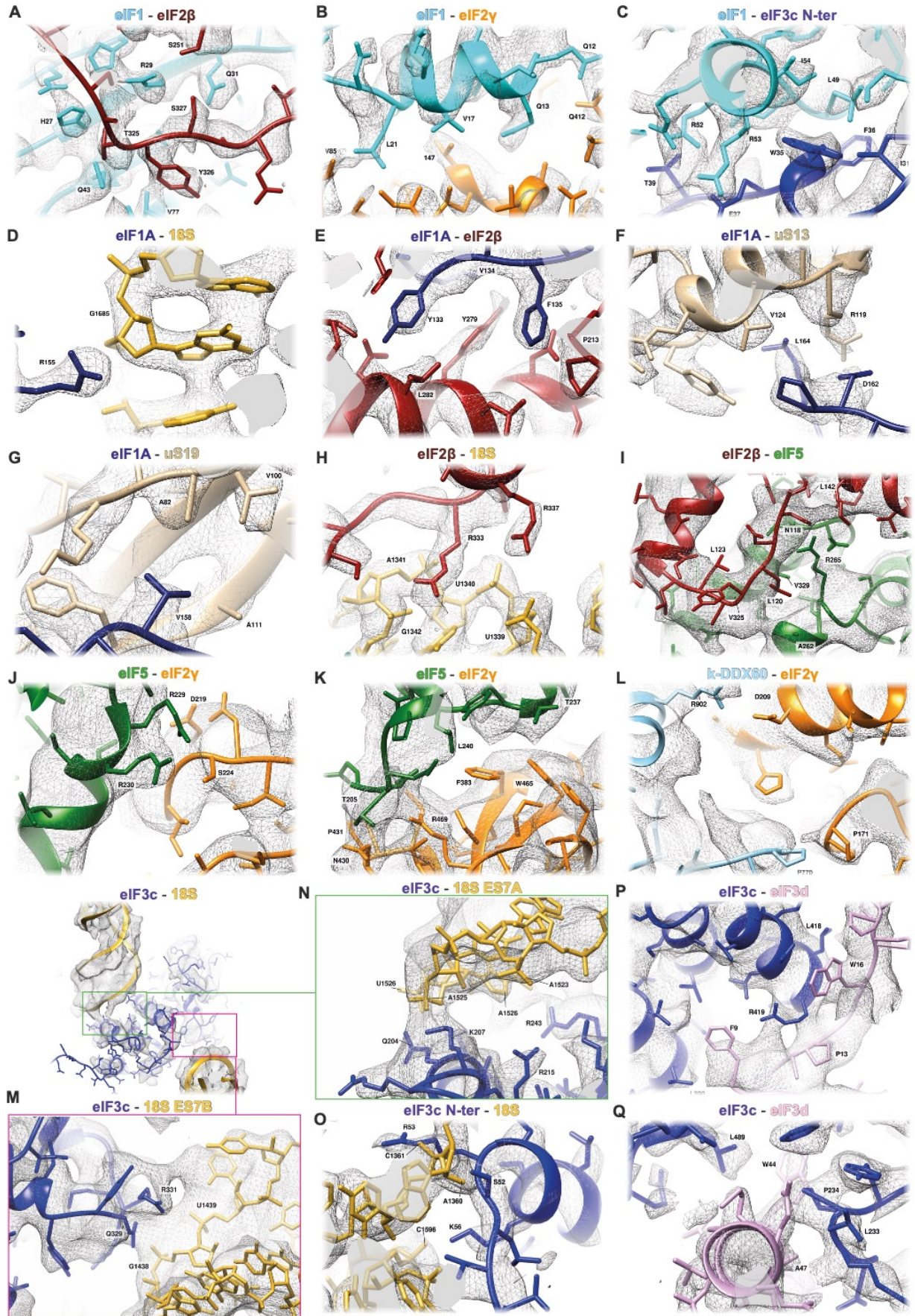


Ser264 in yeast

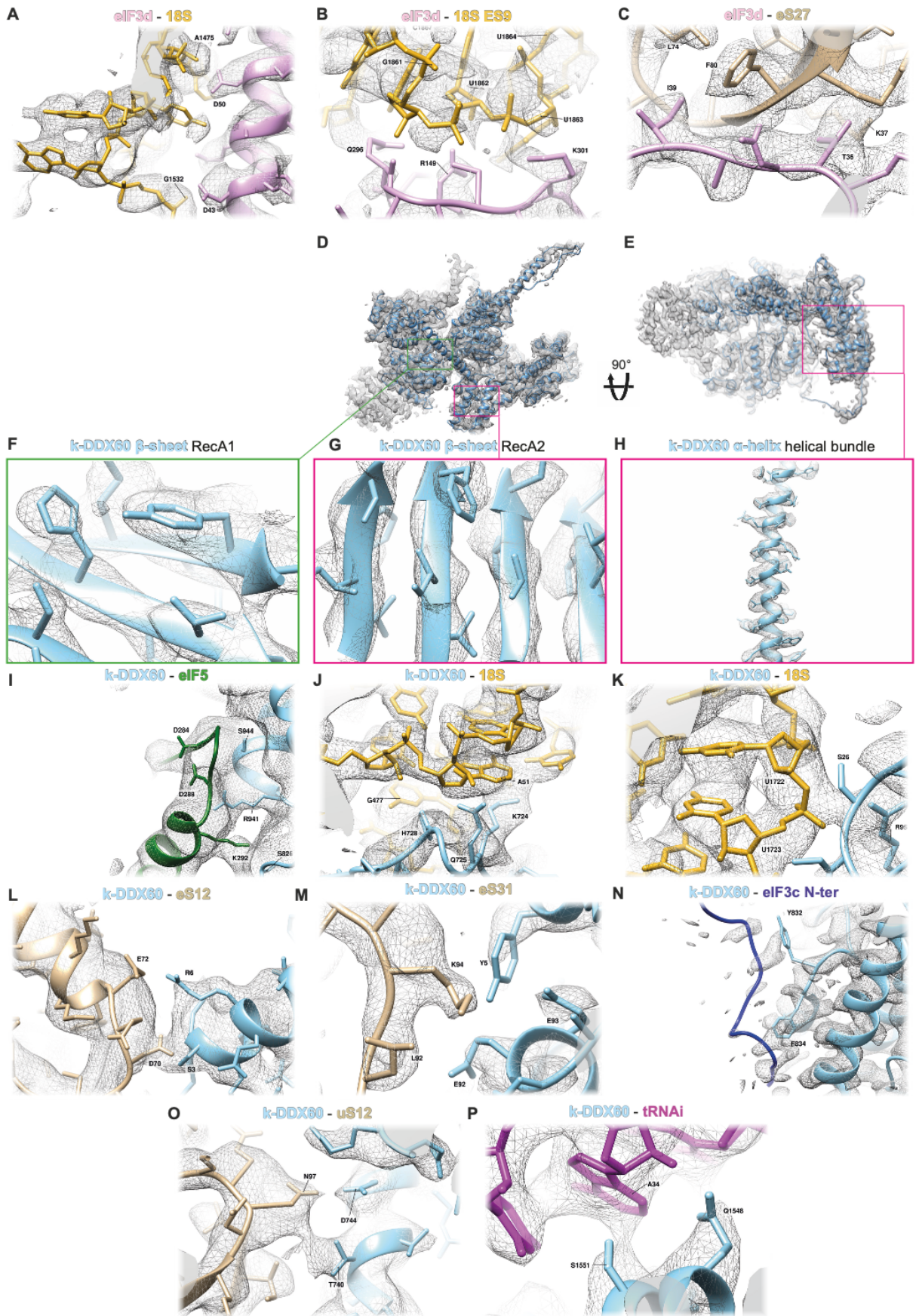
C



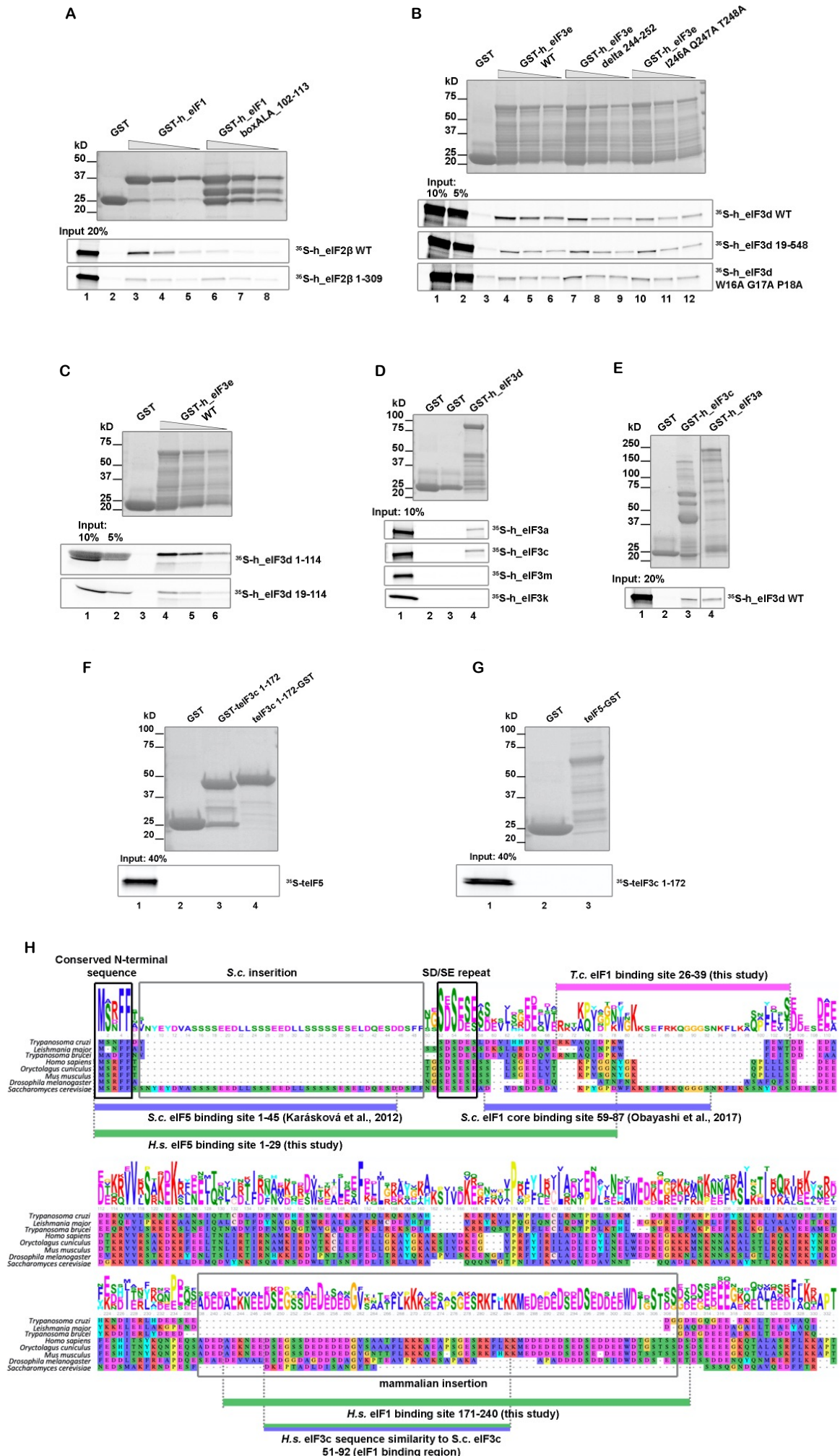
Supplementary Figure 2. Multiple sequence alignment of the eIF2 α NTD and eIF2 β among eukaryotes, and eIF5 CTD structure, Related to Figure 2. (A) Protein sequence alignment of eIF2 α from various eukaryotic organisms was generated by ClustalW (MultiWay, scoring matrix: Blosum 62). The Kinetoplastida order species are labeled with K*. The kinetoplastid-specific eIF2 α 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. (B) Protein sequence alignment of eIF2 β 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. (C) Rigid-body fittings of the crystal structure of the human eIF5 CTD in the corresponding *T. cruzi* 43S PIC density (up) and its *T. cruzi* eIF5 CTD homology model in that same density (bottom).



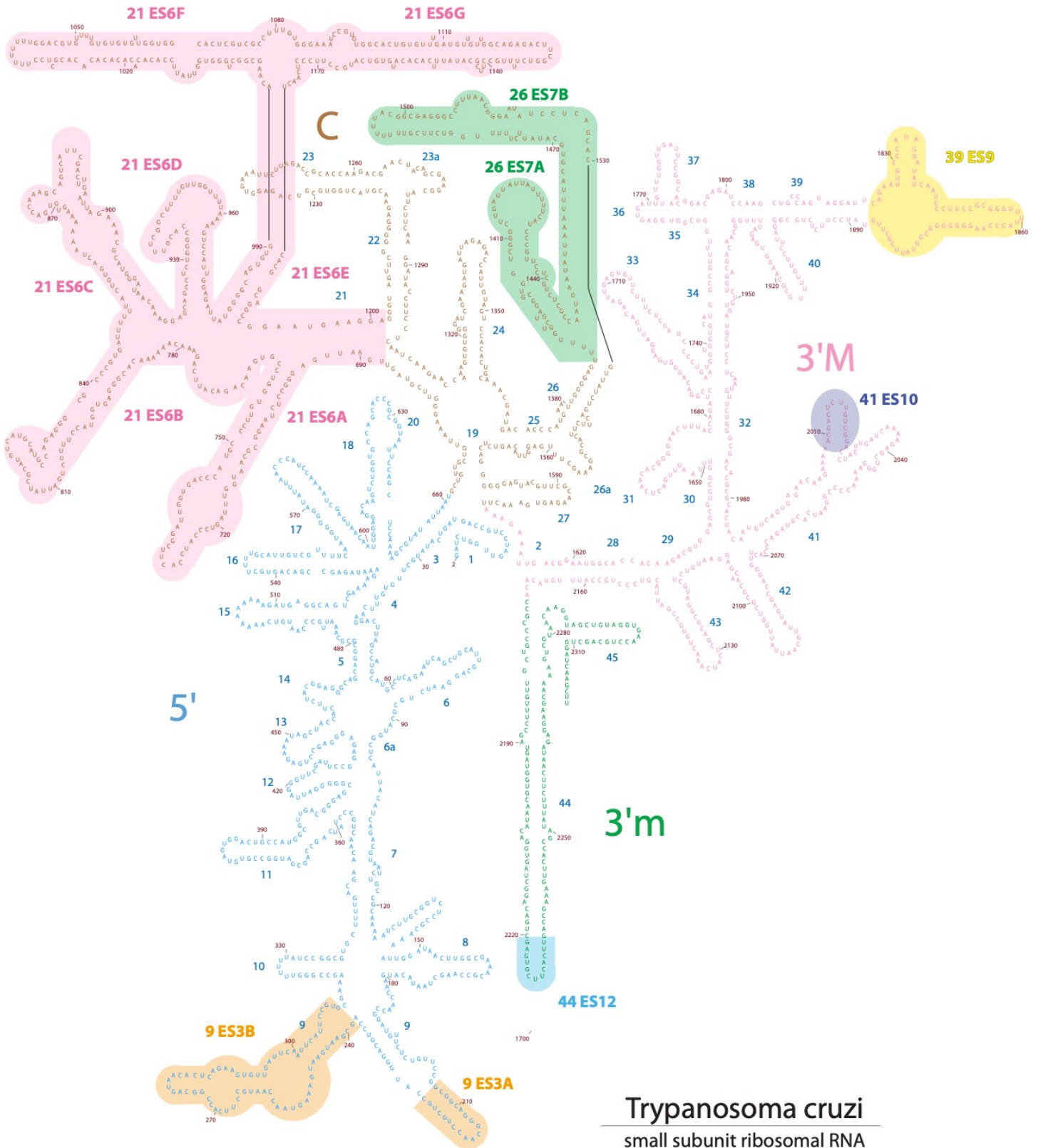
Supplementary Figure 3. Novel interactions between several eIFs, r-proteins and 18S rRNA, fitted in their corresponding densities, Related to Figures 2 and 3. (A) eIF1 with eIF2 β . (B) eIF1 N-ter tail with eIF2 γ . (C) eIF1 with eIF3c N-ter. (D) eIF1A with the 18S. (E) eIF1A with eIF2 β . (F) eIF1A with uS13. (G) eIF1A with uS19. (H) eIF2 β with the 18S. (I) eIF2 β with eIF5 CTD. (J and K) eIF5 CTD with eIF2 γ . (L) eIF2 γ with k-DDX60. (M, N and O) 18S with eIF3c. (P and Q) eIF3c and eIF3d subunits.



Supplementary Figure 4. Interactions of eIF3d and k-DDX60 with eIFs, r-proteins and 18S rRNA, fitted in their corresponding densities, Related to Figures 2 and 3. (A and B) eIF3d with the 18S. (C) eIF3d with eS27. (D and E) k-DDX60 in its corresponding density, viewed from two orientations. (F, G and H) Blow ups on two β -sheets from helicase RecA domains and one buried α -helix from a helical bundle from K-DDX60. Interactions of k-DDX60 with eIF5 (I), 18S (J and K), eS12 (L), eS31 (M), eIF3c N-ter (N), uS12 (O) and the initiator tRNA^{Met} (P).

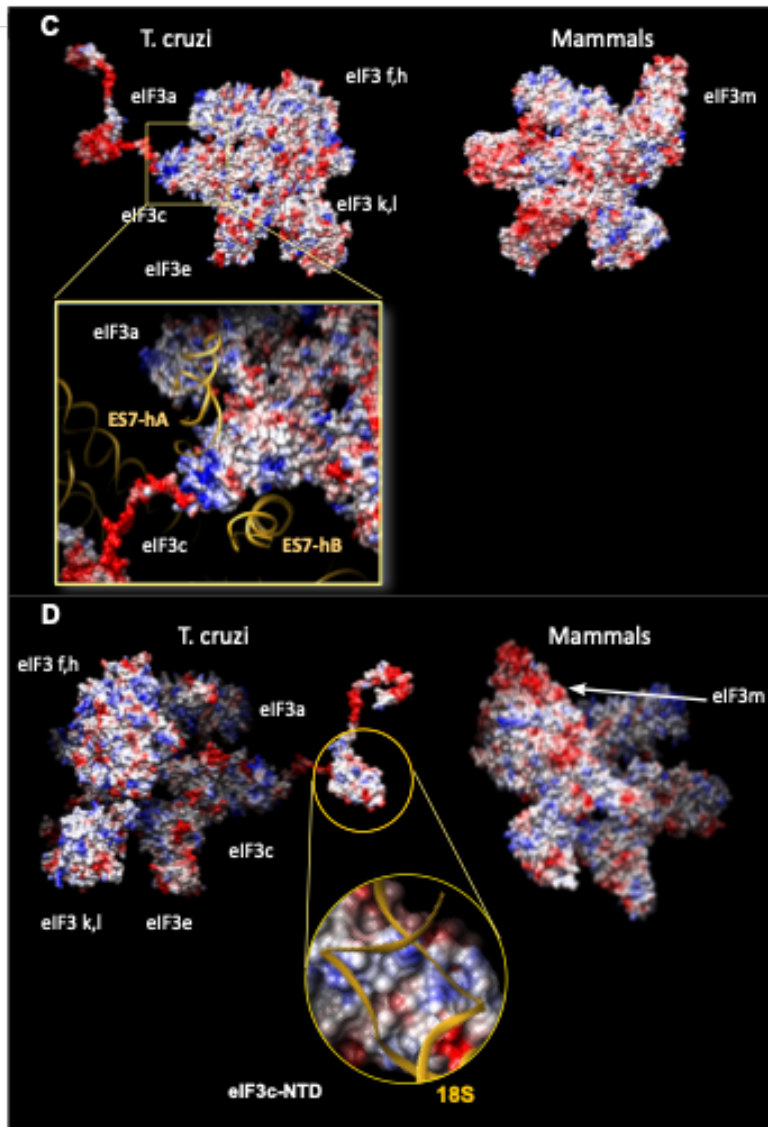
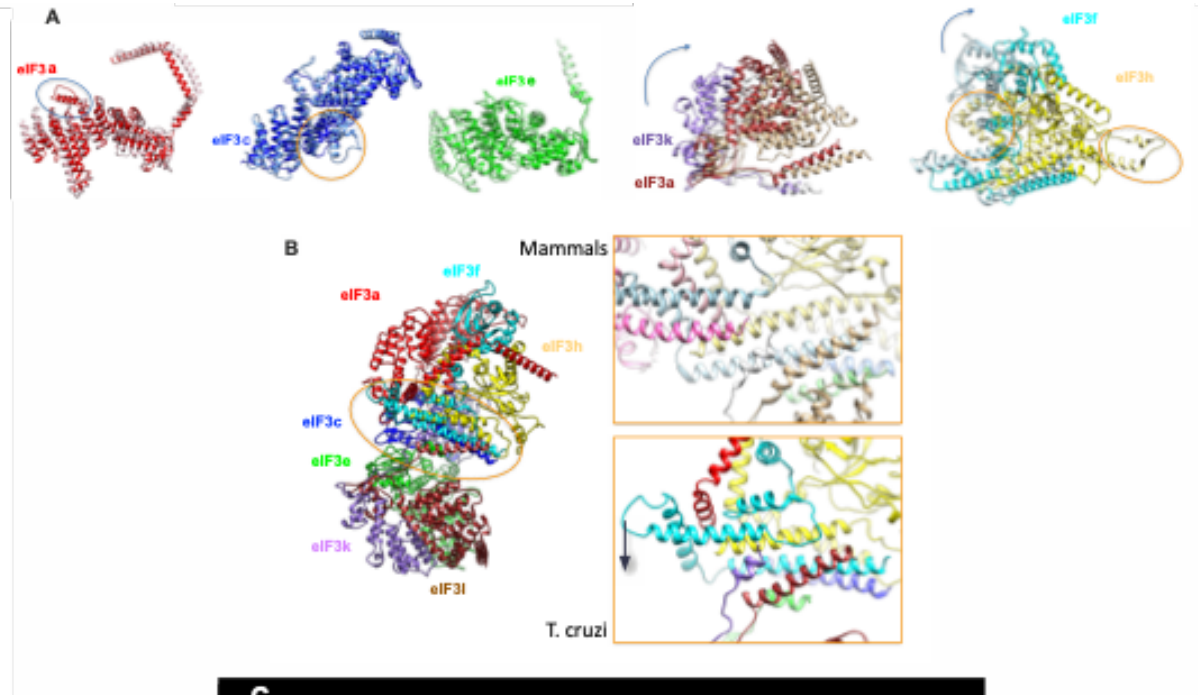


Supplementary Figure 5. *In vitro* analysis of eIF3 intersubunit interactions , Related to Figures 2 and 3. (A) *In vitro* protein-protein binding analysis of the interaction between the *in vitro* translated human ³⁵S-labeled eIF2 β and its C-terminal truncation (eIF2 β 1-309) against wild type eIF1 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 ³⁵S-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 ³⁵S-labeled eIF3d against eIF3c and eIF3a subunits fused to GST. Lane 1 shows 20% input. (F) *In vitro* protein-protein binding analysis of the interaction between *T. cruzi* eIF5 and the eIF3c-NTD (residues 1-172) fused with GST either at its N or C terminus. (G) Binding analysis of the interaction between *T. cruzi* the eIF3c-NTD (residues 1-172) and eIF5-GST. (H) Multiple protein alignment of the N-terminal domain of the eIF3c subunit from indicated species with a consensus expressed as a sequence logo. Specific sequence features mentioned in the main text are boxed. Positions of eIF1- and eIF5-binding sites in the eIF3c-NTD of the selected species identified by us and others are marked by thick lines under or above the alignment; color-coding is as follows: *T.c.* – *Trypanosoma cruzi* in pink, *S.c.* – *Saccharomyces cerevisiae* in purple, and *H.s.* – *Homo sapiens* in green.



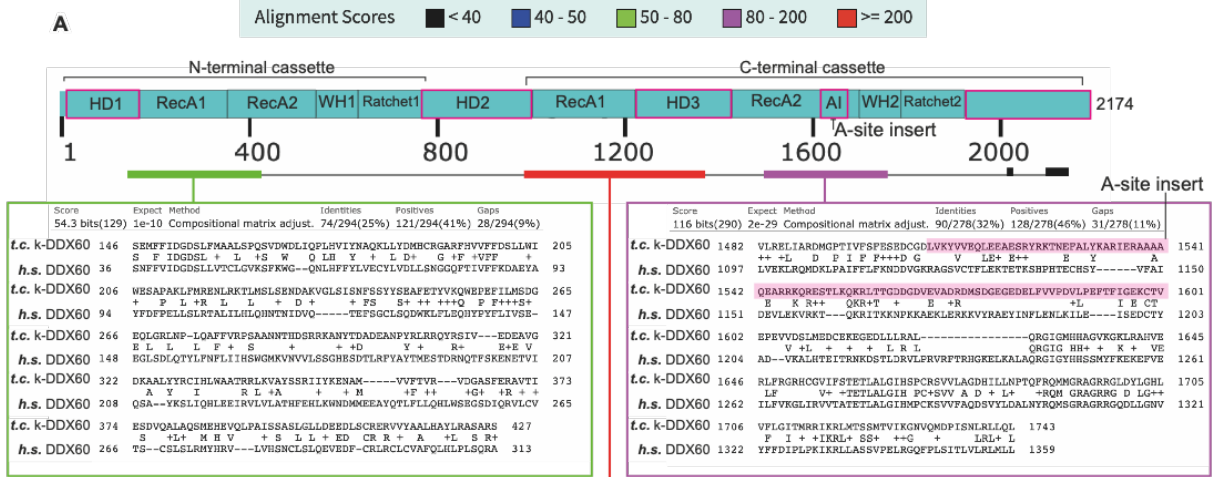
Trypanosoma cruzi
small subunit ribosomal RNA

Supplementary Figure 6. *T. cruzi* 18S rRNA, Related to Figures 4 and 5. 2D diagram of the *T. cruzi* 18S rRNA. The largest and more relevant expansion segments are highlighted in colored backgrounds.

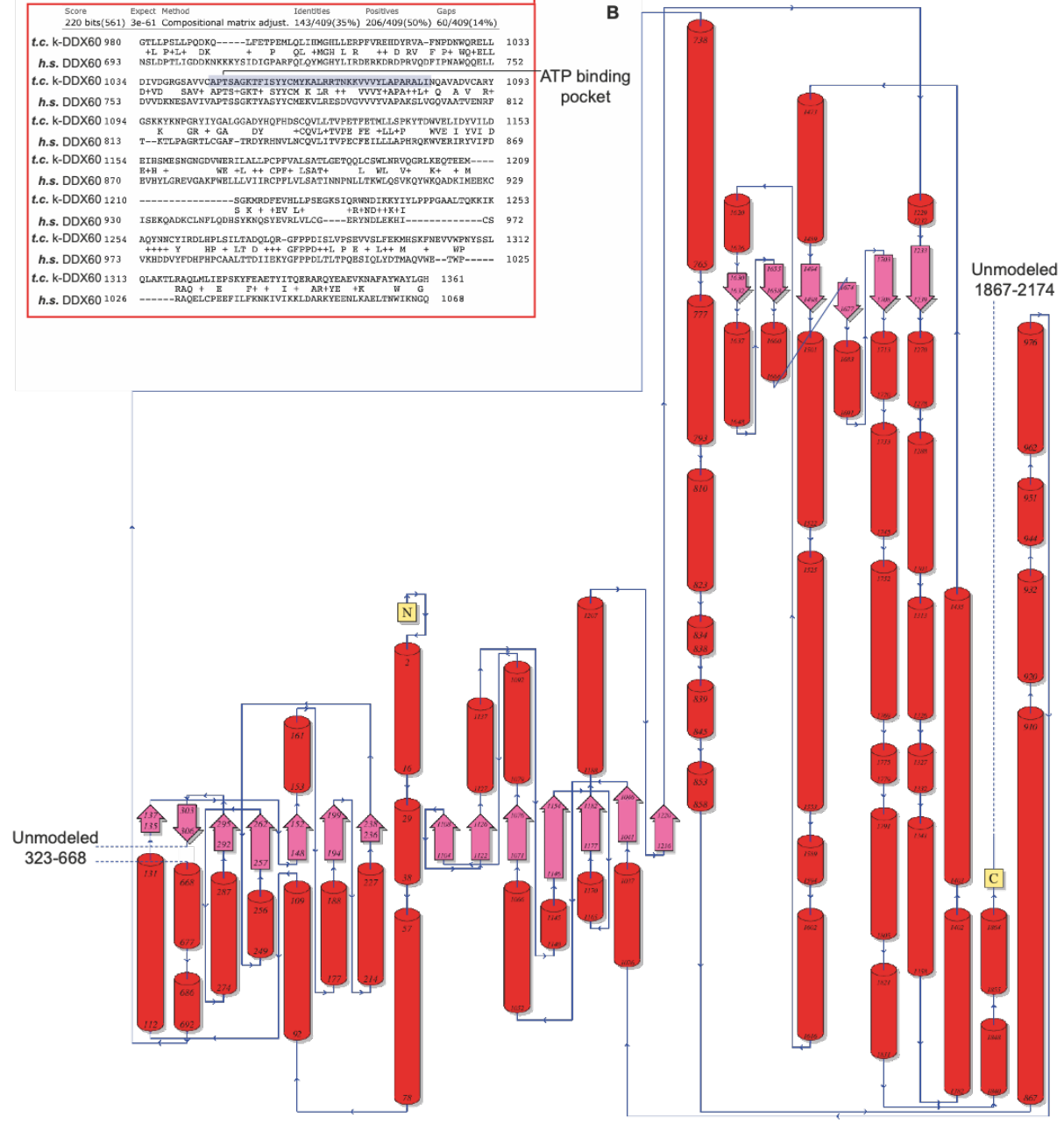


Supplementary Figure 7. Charge surface analysis of the *T. cruzi* and mammalian eIF3 structures, Related to Figures 3. (A) Overlay of mammalian and kinetoplastid structures of individual eIF3 subunits with marked structural differences. The *T. cruzi* structures are depicted in dark and mammalian in 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. (B) 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. (C) 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. (D) 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.

A



B



Supplementary Figure 8. *T. cruzi* 18S k-DDX60 conservation and secondary structure diagram, Related to Figures 4 and 5. (A) 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. (B) Secondary structure elements diagram for k-DDX60 based on its 3D model. Some parts could not be modeled.

SUPPLEMENTARY TABLES

				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 F	69	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	
Q4DR85 Q4DR85_TRYCC		40S ribosomal protein S14, putative OS=Trypanosoma cruzi (strain CL Brener) GN=TC00.104705344	60	61	37	
Q4CUL0 Q4CUL0_TRYCC		40S ribosomal protein S3, putative OS=Trypanosoma cruzi (strain CL Brener) GN=TC00.104705343	81	72	37	
Q4DBN9 Q4DBN9_TRYCC		Ribosomal protein S19, putative OS=Trypanosoma cruzi (strain CL Brener) GN=TC00.10470535040	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=TC00.1047053506693	84	80	25	
Q4CQU0 Q4CQU0_TRYCC		40S ribosomal protein S4 OS=Trypanosoma cruzi (strain CL Brener) GN=TC00.1047053503719.20 F	66	58	22	
Q4D916 Q4D916_TRYCC		40S ribosomal protein S16, putative OS=Trypanosoma cruzi (strain CL Brener) GN=TC00.104705350	48	52	19	
Q4EDN6 Q4EDN6_TRYCC		40S ribosomal protein S15a, putative OS=Trypanosoma cruzi (strain CL Brener) GN=TC00.10470535	37	34	15	
Q4DI29 Q4DI29_TRYCC		40S ribosomal protein S2, putative OS=Trypanosoma cruzi (strain CL Brener) GN=TC00.1047053503	80	74	27	
Q4CXN0 Q4CXN0_TRYCC		Ubiquitin/ribosomal protein S27a, putative OS=Trypanosoma cruzi (strain CL Brener) GN=TC00.1047	52	40	14	
Q4DXT6 Q4DXT6_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	
Q4DT1Q Q4DT1Q_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=TC00.104705350388	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=TC00	10	18	5	

Supplementary Table 1. Mass-spectrometry analysis of the *T. cruzi* 43S PIC, Related to Figure 1. 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. Full dataset can be found at the PRIDE partner repository with the dataset identifier PXD016063 (See Methods).

Name			Spectral Count
accession	Name	description	IC
tr E9ACL4	DDX60	Uncharacterized protein OS=Leishmania major GN=LMJF_03_0690 PE=4 SV=1	111
tr Q4QCE4	ABCE1	Putative ATP-binding cassette protein subfamily E, member 1 OS=Leishmania major GN=ABCE1_01_0100 PE=4 SV=1	101
40S ribosomal proteins:			
accession	Name	description	Spectral Count
accession	Name	description	IC
tr Q868B1		40S ribosomal protein S5 OS=Leishmania major GN=LMJF_11_0960 PE=4 SV=1	188
tr Q4Q216		Putative ubiquitin/ribosomal protein S27a OS=Leishmania major GN=LMJF_36_0600 PE=4 SV=1	265
tr Q4Q1Y2		Putative 40S ribosomal protein S18 OS=Leishmania major GN=LMJF_36_0940 PE=3 SV=1	122
tr Q4QG31		40S ribosomal protein S4 OS=Leishmania major GN=RS4 PE=2 SV=1	299
tr Q4Q8H1		40S ribosomal protein S14 OS=Leishmania major GN=LMJF_28_0960 PE=3 SV=1	155
tr Q4QC89		Putative 40S ribosomal protein S23 OS=Leishmania major GN=LMJF_21_1060 PE=3 SV=1	90
tr Q4Q4A0		Putative 40S ribosomal protein S3 OS=Leishmania major GN=LMJF_15_0950 PE=4 SV=1	99
sp P25204		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		Putative ribosomal protein S29 OS=Leishmania major GN=LMJF_28_2205 PE=4 SV=1	62
tr Q4Q1V1		Putative 40S ribosomal protein S9 OS=Leishmania major GN=LMJF_36_1250 PE=2 SV=1	98
tr Q4Q5P0		40S ribosomal protein S2 OS=Leishmania major GN=LMJF_32_0450 PE=3 SV=1	144
tr Q4Q3M1		Putative 40S ribosomal protein S13 OS=Leishmania major GN=LMJF_19_0390 PE=3 SV=1	83
tr Q4QH01		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		Putative ribosomal protein S20 OS=Leishmania major GN=LMJF_28_1010 PE=3 SV=1	99
tr Q4Q7P0		Putative 40S ribosomal protein S30 OS=Leishmania major GN=LMJF_30_0670 PE=4 SV=1	36
tr Q4QCN7		Putative 40S ribosomal protein S11 OS=Leishmania major GN=LMJF_20_1650 PE=3 SV=1	153
sp Q4Q0C0		40S ribosomal protein SA OS=Leishmania major GN=Lmjf36.5010 PE=3 SV=1	145
tr E9AE88		40S ribosomal protein S19-like protein OS=Leishmania major GN=LMJF_29_2860 PE=4 SV=1	129
tr Q4Q931		Putative 40S ribosomal protein S33 OS=Leishmania major GN=S33-1 PE=4 SV=1	102
tr Q4Q1X7		Putative 40S ribosomal protein S10 OS=Leishmania major GN=LMJF_36_0980 PE=4 SV=1	101
tr Q4Q977		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		Putative 40S ribosomal protein S16 OS=Leishmania major GN=LMJF_26_0880 PE=2 SV=1	79
tr Q4Q806		Putative 40S ribosomal protein S17 OS=Leishmania major GN=LMJF_28_2555 PE=3 SV=1	42
tr Q4Q140		Putative 40S ribosomal protein S27-1 OS=Leishmania major GN=LMJF_36_3750 PE=3 SV=1	63
tr Q4Q8L6		Putative ribosomal protein S26 OS=Leishmania major GN=LMJF_28_0540 PE=4 SV=1	34
tr Q4Q1D2		40S ribosomal protein S24 OS=Leishmania major GN=S24E-2 PE=3 SV=1	120
tr Q4Q3G4		Ribosomal protein S25 OS=Leishmania major GN=S25 PE=4 SV=1	91
tr Q43943	RACK1	LACK OS=Leishmania major PE=4 SV=1	58
tr Q4Q5K7	KSRP	Putative RNA binding protein OS=Leishmania major GN=LMJF_32_0750 PE=4 SV=1	56
tr Q4QBVO		Putative 40S ribosomal protein S15 OS=Leishmania major GN=LMJF_22_0420 PE=3 SV=1	31
tr E9AC32		Putative ribosomal protein S7 OS=Leishmania major GN=LMJF_01_0410 PE=4 SV=1	27
Initiation factors:			
accession	Name	description	Spectral Count (# spots)
accession	Name	description	IC
tr Q4QEJ8	eIF3a	Uncharacterized protein OS=Leishmania major GN=LMJF_17_0010 PE=4 SV=1	278
tr Q4QE62	eIF3b	Putative translation initiation factor OS=Leishmania major GN=LMJF_17_1290 PE=4 SV=1	175
tr Q4Q6Y6	eIF3d	Eukaryotic translation initiation factor 3 subunit 7-like protein OS=Leishmania major GN=LMJF_36_0650 PE=3 SV=1	125
tr Q4Q833	eIF3e	Eukaryotic translation initiation factor 3 subunit E OS=Leishmania major GN=LMJF_28_2310 PE=3 SV=1	84
tr Q4Q253	eIF3l	Eukaryotic translation initiation factor 3 subunit L OS=Leishmania major GN=LMJF_36_0250 PE=3 SV=1	91
tr Q4Q127	eIF3i	Eukaryotic translation initiation factor 3 subunit I OS=Leishmania major GN=LMJF_36_3880 PE=3 SV=1	79
tr E9ACP3	eIF2 alpha	Putative elongation initiation factor 2 alpha subunit OS=Leishmania major GN=LMJF_03_0980 PE=4 SV=1	76
tr Q4QIM7	eIF3h	Uncharacterized protein OS=Leishmania major GN=LMJF_07_0640 PE=4 SV=1	76
tr Q4Q3H3	eIF5	Putative eukaryotic translation initiation factor 5 OS=Leishmania major GN=LMJF_34_0350 PE=4 SV=1	75
tr Q4Q9T0	eIF3f	Uncharacterized protein OS=Leishmania major GN=LMJF_25_1610 PE=4 SV=1	67
tr Q4Q055	eIF3c	Putative eukaryotic translation initiation factor 3 subunit 8 OS=Leishmania major GN=LMJF_36_0650 PE=3 SV=1	62
tr Q4Q557	eIF3k	Uncharacterized protein OS=Leishmania major GN=LMJF_32_2180 PE=4 SV=1	59
tr Q4QHR7	eIF2 gamma	Putative eukaryotic translation initiation factor 2 subunit OS=Leishmania major GN=LMJF_09_0950 PE=4 SV=1	49
tr Q4Q2S5	eIF3g	Eukaryotic translation initiation factor 3 subunit G OS=Leishmania major GN=LMJF_34_2700 PE=4 SV=1	46
tr Q4QAL1	eIF1A	Putative translation factor sui1 OS=Leishmania major GN=LMJF_24_1210 PE=4 SV=1	33
tr Q4QIB4	eIF2 beta	Translation initiation factor-like protein OS=Leishmania major GN=LMJF_08_0550 PE=4 SV=1	29
tr Q4Q060	eIF1A	Putative eukaryotic translation initiation factor 1A OS=Leishmania major GN=LMJF_16_0140 PE=4 SV=1	27

Supplementary Table 2. Mass-spectrometry analysis of the *L. Tarentolae* 43S PIC, Related to Figure 1. 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. Full dataset can be found at the PRIDE partner repository with the dataset identifier PXD016063 (See Methods).

	Ribosomal RNA	Ribosomal protein	Initiation factors
eIF1	N65-G2303, C64-G2303, Q81-C2282, R33-A1341, R33-G2283, K37-G2283, R56-G2303, R61-C2183	none	eIF2-β : R29-S251, Q31-S327, Q43-T325, H27-T325, V77-Y326 eIF2-γ : S16-N459, V17-V147, Q12-Q412, L21-V85, Q13-V147 eIF3c : L49-F36, I54-W35 R53-E37, R52-T39, N96-R26, L49-I31
eIF1A	N48-A2277, R66-C620, W74-A2279, R155-G1685	eS30 : E35-R10, F88-L8 uS13 : D162-R119, L164-V124 uS19 : V158-V100, V158-A82, V158-A111 uS12 : F88-L91	eIF2-β : Y133-L282, V134-N208, F135-P213, F135-Y279
eIF2-α	none	uS7 : Y200-K177, Y200-D180, Y200-R184, T148-R122, Y166-V120, T167-R122, D195-R184	tRNA : K104-C55, R105-G52, R108-U54, W119-C55, H 232-C55, E296-U54 eIF2-γ : R331-E279, F315-L321, V320-L350, P350-F268
eIF2-β	R333-U1340, R333-G1342, R337-U1339, R337-U1340	uS19 : N259-R137	tRNA : K221-A36, N255-G25, K300-G68, R303-G69, eIF1 : S251-R29, E267-Q32, T325-Q43, Y326-V77, Y326-H27, S323-R29 eIF1A : N208-V134, P213-F135, Y279-F135, L282-Y133 eIF5 : N118-R265, L120-V329, L120-A262, L123-V325, K125-Q364, V132-W372, L142-F331 eIF2-γ : N173-H248, T176-Y245, G181-Y241, Y182-Y211, Y184-D240, S185-N238, R189-E204, M305-E83, T317-M86, T317-E83
eIF2-γ	none	none	tRNA : K79-C73, D269-A75, K272-A72, R282-A75 eIF1 : V85-L21, I88-L21, V147-V17, N459-S16, Q412-Q12 eIF2-α : E279-R331, L321- F315, L350- V320, F268- P350 eIF2-β : H248- N173, Y245- T176, Y241- G181, Y211- Y182, D240- Y184, N238- S185, E204- R189, E83- M305, M86- T317, E83-T317 eIF5 : S224-R230, D219-R229, S220-R273, F383-L240, N430-T205, P431-D204, W465-T237, R469-T205 k-DDX60 : P171-P770, D209-R902
eIF3c	S52-A1360, R53-C1361, K56-C1596, R127-C370, Q204-U1526, K207-A1525 R215-A1523, R232-U1476 and U1478, Q329-G1438, R331-U1439, R243-U1526	eS27 : Q191-Q56, K192-E54	eIF1 : I31-M97, I31-L49, F36-L49, E37-R53, W35-F91, W35-I54, T39-R52 eIF3d : P234-W44, R295-W44, L489-W44, L233-A47, L380-F9, L418-W16, R419-P13, I434-M28, N437-D26 k-DDX60 : N-ter tail with Y832 and F834
eIF3a		eS1 : T7-Q77, R8-T77, T12-R192, L17-I194	
eIF3d	D43-G1532, D50-A1475, R149-U1863 and U1862, R294-U1866 and C1867, D306-U1864, Q296-G1861, K301-U1863	eS27 : T36-K37, I39-F80, I39-L74 S33 : R219-E76, D255-R83, K371-M98, Q368-K94, L435-M73 uS7 : Q434-E21, Q368-D26, E368-R51 RACK1 : S409-E277, N410-Q279	eIF3c : F9-L380, P13-R419, W16-L418, D26-N437, W44-P234, W44-R295, W44-L489, A47-L233, M28-I434 eIF3e : F3-T198, L5-A196, P6-T198, W16-I246, W16-Q247, E7-T245 ; P13-T248
eIF5	none	none	eIF2-β : A262-L120, R265-N118, V325-L123, V329-L120, I332-L142, Q364-K125, W372-V132 eIF2-γ : D204-P431, T205-R469, T205-N430, R229-D219, R230- S224, T237-W465, L240-F383, R273-S220 k-DDX60 : D284-S944, D288-R941, K292-S826
k-DDX60	S26-U1722, R95-U1723, K724-A51, Q725-A51, H728-G477	eS12 : S3-D70, R6-E72 eS31 : Y5-K94, E92-L92, E93-K94 uS12 : R739-Q73, D744-N97 T740-N97	tRNA : Q1548-A34, S1551-A34 eIF2-γ : P770-P171, R902-D209 eIF3c : Y832 and F834 with N-ter tail eIF5 : S826-K292, R941-D288, S944-D284

Supplementary Table 3. Detailed overview of interactions between eIFs, ribosomal proteins, rRNA and k-DDX60, Related to Figures 2, 3 and 5. Novel interactions revealed by analysis are colored in deep blue. Most of these novel interactions are shown in ribbons and sticks models fitted into their corresponding densities in Fig. S 3 and 4.

Data Collection	<i>T. cruzi</i> 43S	<i>T. cruzi</i> 43S + ATP	<i>L. tarentolae</i> 43S
Microscope	Titan Krios	Titan Krios	Talos Artica
Voltage (kV)	300	300	200
Magnification	127,272	127,272	120,000
Pixel size (Å)	1.1	1.1	1.21
Detector	Gatan Summit K2	Gatan Summit K2	Falcon III
Defocus range (µm)	-0.6 to -4.5	-0.6 to -4.5	-0.6 to -3.0
Tot. electron exposure (e ⁻ Å ⁻²)	30	30	40
Exposure rate (e ⁻ Å ⁻² frame ⁻¹)	1.5	1.5	2.0
Data collection software	SerialEM	EPU	EPU
Data Processing			
Independent data collections	1	1	1
Useable micrographs	3271	2638	?
Particles	202920	98840	52302
Final particles (43S)	33775	19700	10144
Accuracy			
translations (pix) / rotations (°)	0.432/0.234375°	0.432/0.234375°	0.432/0.234375°
Resolution (Å, 0.143 FSC)	3.33	4.3	8.1
Local resolution range (Å)	2.5-6	N/A	N/A
Model Composition			
Chains	59	N/A	N/A
Non-hydrogen atoms	136893	N/A	N/A
Protein residues	11196	N/A	N/A
RNA bases	2225	N/A	N/A
Refinement			
Software	Phenix_ValidationEM	N/A	N/A
Resolution (Å)	3.33	N/A	N/A
CC (mask)	0.57	N/A	N/A
CC (main chain)	0.5	N/A	N/A
CC (side chain)	0.58	N/A	N/A
R.M.S deviations			
Bond lengths (Å)	0.020	N/A	N/A
Bond angles (°)	2.209	N/A	N/A
Validation			
Molprobrity score	2.65	N/A	N/A
Clashscore, all atoms	9.20	N/A	N/A
Rotamers outliers (%)	3.48	N/A	N/A
Cβ outliers (%)	1.41	N/A	N/A
CaBLAM outliers (%)	10.9	N/A	N/A
Ramachandran plot			
Favored (%)	79.09	N/A	N/A
Allowed (%)	14.19	N/A	N/A
Outliers (%)	6.72	N/A	N/A

Supplementary Table 4. Data collection, processing, refinement and model statistics, Related to Figure 1 and STAR Methods section. A near complete atomic model was only derived for the highest resolution cryo-EM reconstruction, i.e. the *T. cruzi* 43S PIC stalled with GMP-PNP. The *T. cruzi* 43S PIC stalled with GMP-PNP supplemented with ATP presents a lower resolution and therefore we didn't derive a full atomic model, instead we flexibly fitted k-DDX60 only into its map to illustrate its conformational changes.

Primer	Sequence
AH-h3c-PmeI	ATATAGTTTAAACGCCATGTCGCGGTTTTTCACC
AH-h3c-FseI	ATATAGGCCGGCCTCAGTAGGCCGTCTGAGACTG
AH-h3a-PmeI	ATATAGTTTAAACAAGATGCCGGCCTATTTTCAG
AH-h3a-FseI	ATATAGGCCGGCCTTAACGTCGTA CTGTGGTCCA
AH-h3m-EcoRI	ATATAGAATTCACCATGAGCGTCCCGGCCTTC
AH-h3m-FseI	ATATAGGCCGGCCTCAGGTATCAGAAAGACTCAA
AH-h3k-EcoRI	ATATAGAATTCGTCATGGCGATGTTTGAGCAG
AH-h3k-FseI	ATATAGGCCGGCCTTACTGGGAGGAGGCCATGAT
AH-h3d-EcoRI	ATATAGAATTC AAGATGGCAAAGTTCATGACA
AH-h3d-FseI	ATATAGGCCGGCCTTAAGTTTCTCCTCTTCTTCTCCTC
AH-h3c-EcoRI	ATATAGAATTC AAGATGGCGGAGTACGACTTG
AH-h3c-FseI	ATATAGGCCGGCCTCAGTAGAAGCCAGAATCTTG
DS-eIF1-BamHI	ATCGGATCCATATGTCGCTATCCAGAACC
DS-eIF1-SalI	TGTGTGCGACTTAAACCCATGAACCTTCAG
SW-heIF5-EcoRI	ATAGAATTCGATGTCTGTCAATGTCAACC
SW-heIF5-SalI-R	ACTAGTCGACTTAAATGGCATCAATATCG
DS-eIF2 β -BamHI	ATCGGATCCATATGTCGTTGGGACGAGATG
DS-eIF2 β -SalI	CGTGTGCGACTTAGTTAGCTTTGGCACG
AH-h3c-325-FseI	ATATACGGCCGGCCTCAGGTGATCTCAGTTCCTTGGC
TS-h3c-326-PmeI	ATATAGTTTAAACGCCATGCATGCTGTTGTTATCAAGAACTG
TS-h3c-30-325-PmeI	ATATAGTTTAAACGCCATGAACTATGGCAAACAGCCATTG
TP-h3c-130-325-PmeI	GCCGCGTTTAAACGCCATGAACAAGAACAATGCCAAGGC
TP-pGL4-CMV-heIF2 β -EcoRI	CGCCAGAATTCACCATGTCTGGGGACGAGATGATT
TP-pGL4-CMV-eIF2 β -FseI	ATATAGGCCGGCCTTAGTTAGCTTTGGCACGGAG
TP-pGL4-CMV-eIF2 β -1-309-FseI	CGCCAGGCCGGCCTTAACATCTAGAATGACAAGTTTC
TP-pGEX-5X3-teIF1-BamHI	ACCTAGGATCCCCATGCTAAACAACGAGCTCGCTAACC
TP-pGEX-5X3-teIF1-SalI	ACCGAGTCGACTTAGTTCAGAGAGTGGATCTC
TP-pGEX-5X3-teIF5-BamHI	ACCGAGGATCCCCATGTCGGTTCCAATGATACCCATTG
TP-pGEX-5X3-teIF5-SalI	CGCTAGTCGACCTATGTCGATCCTACAAGCCATTCGAC
TP-pGL4-teIF5-EcoRI	GCTGTGAATTCGCCACCATGTCGGTTCCAATGATACCC
TP-pGL4-teIF5-FseI	TATATGGCCGGCCCTATGTCGATCCTACAAGCCATTC
TP-pGL4-CMV-teIF3c-PmeI	CGCCGTTTAAACACCATGAGCAACTTTTTTGATGTCAGCGAC AGTG
TP-pGL4-CMV-teIF3c-FseI	ATATAGGCCGGCCGTTAAAATCCTCCTCTACCACGTCCTCGAC
TP-pGL4-CMV-teIF3c-14-PmeI	CGCGCGTTTAAACACCATGCTGGATGAGGTCATACATCAGAT G
TP-pGL4-CMV-teIF3c-39-PmeI	AGCCAGTTTAAACACCATGACCGATGATGAGGACGCGGATG
TP-pGL4-CMV-teIF3c-172-FseI	TATATGGCCGGCCGTTACTCCTCACCTGTCCTTCATC
TP-pGEX-teIF3c-BamHI	GCGGCGGATCCCCATGAGCAACTTTTTTGATGTC
TP-pGEX-teIF3c-1-172-EcoRI	GCCGCGAATTCCTTTACTCCTCACCTGTCCTTC
TP-pGEX-5X3-eIF3e-BamHI	ATATAGGATCCCCATGGCGGAGTACGACTTGAC
TP-pGEX-5X3-eIF3e-SalI	ATGCCGTCGACTCAGTAGAAGCCAGAATCTTG

TP-pGL4-CMV-h3d-19-EcoRI	CTGCAGAATTCAAGATGTGTGCGGTTCCCGAGCAG
AH-h3d-FseI	ATATAGGCCGGCCTTAAGTTTCTTCCTCTTCTTCCTC
AH-h3d-EcoRI	ATATAGAATTCAAGATGGCAAAGTTCATGACA
TP-pGL4-CMV-h3d-114-FseI	ACGTAGGCCGGCCTTACATGTTCCGACGATCTTTGTC
TP-pGEX-heIF3d-BamHI	ATATCGGATCCCCATGGCAAAGTTCATGACACCC
TP-pGEX-heIF3d-SalI	GCGCGGTCGACTTAAGTTTCTTCCTCTTCTTCCTC
TP-pGEX-eIF3c-EcoRI	ACCGAGAATTCCATGTCGCGGTTTTTTCACCACC
TP-pGEX-eIF3c-SalI	TATATGTCGACTCAGTAGGCCGTCTGAGACTG
TP-pGEX-eIF3a-SalI	ACCTAGTCGACATGCCGGCCTATTTTCAGAGG
TP-pGEX-eIF3a-NotI	ATATAGCGGCCGCTTAACGTCGTA CTGTGGTCCA
TP-telF3c-AsiSI	GCACCGCGATCGCATGAGCAACTTTTTTGATGTCAGCG
TP-telF3c-1-172-MluI	ATATAACGCGTCTCCTCACCCTGTCCTTCATC
TP-telF5-AsiSI	ACCGAGCGATCGCATGTCGGTTCCAATGATACCC
TP-telF5-MluI	TCGATACGCGTTGTCGATCCTACAAGCCATTC

Supplementary Table 5. List of primers, Related to STAR Methods section.

Gene string	Sequence 5'-3'
heIF1-box-Ala-102-113	GGCGACCATCCTCCAAAATCGGATCTGATCGAAGGTCGTGGGATCCAT ATGTCCGCTATCCAGAACCTCCACTCTTTTCGACCCCTTTGCTGATGCAA GTAAGGGTGATGACCTGCTTCCTGCTGGCACTGAGGATTATATCCATA TAAGAATTCAACAGAGAAAACGGCAGGAAGACCCTTACTACTGTCCAA GGGATCGCTGATGATTACGATAAAAAGAACTAGTGAAGGCGTTTAA GAAAAAGTTTGCCTGCAATGGTACTGTAATTGAGCATCCGGAATATGG AGAAGTAATTCAGCTACAGGGTGACCAACGCAAGAACATATGCCAGT TCCTCGTAGAGATTGGAGCAGCAGCAGCTGCAGCAGCCGCAGCCGCG GCAGCATAAGTCGACTCGAGCGGCCGCATCGTGACTGACTGACGATC TGCCTCGC
pGEX-heIF3e-delta-244-252	GACCATCCTCCAAAATCGGATCTGATCGAAGGTCGTGGGATCCCCATG GCGGAGTACGACTTGACTACTCGCATCGCGCACTTTTTGGATCGGCAT CTAGTCTTTCCGCTTCTTGAATTTCTCTCTGTAAAGGAGATATATAATG AAAAGGAATTATTACAAGGTAAATTGGACCTTCTTAGTGATACCAACA TGGTAGACTTTTGTATGGATGTATACAAAAACCTTTATTCTGATGATA TTCCTCATGCTTTGAGAGAGAAAAGAACCACAGTGGTTGCACAACCTG AAACAGCTTCAGGCAGAAACAGAACCAATTGTGAAGATGTTTGAAGA TCCAGAACTACAAGGCAAATGCAGTCAACCAGGGATGGTAGGATGC TCTTTGACTACCTGGCGGACAAGCATGGTTTTAGGCAGGAATATTTAG ATACACTCTACAGATATGCAAAATTCAGTACGAATGTGGGAATTACT CAGGAGCAGCAGAATATCTTTATTTTTTTAGAGTGCTGGTTCCAGCAA CAGATAGAAATGCTTTAAGTTCCTCTGGGGAAAGCTGGCCTCTGAAA TCTTAATGCAGAATTGGGATGCAGCCATGGAAGACCTTACACGGTTAA AAGAGACCATAGATAATAATTCTGTGAGTTCTCCACTTCAGTCTCTTC AGCAGAGAACATGGCTCATTCACTGGTCTCTGTTTGTCTTCAATCA CCCCAAAGGTCGCGATAATATTATTGACCTCTTCCTTTATCAGCCACA ATATCTTATTCTTCGCTATTTGACTACAGCAGTCATAACAAACAAGGA TGTTTCGAAAACGTCGGCAGGTTCTAAAAGATCTAGTTAAAGTTATTCA ACAGGAGTCTTACACATATAAA
pGEX-heIF3e-I246A-Q247A-T248A	GACCATCCTCCAAAATCGGATCTGATCGAAGGTCGTGGGATCCCCATG GCGGAGTACGACTTGACTACTCGCATCGCGCACTTTTTGGATCGGCAT CTAGTCTTTCCGCTTCTTGAATTTCTCTCTGTAAAGGAGATATATAATG AAAAGGAATTATTACAAGGTAAATTGGACCTTCTTAGTGATACCAACA TGGTAGACTTTTGTATGGATGTATACAAAAACCTTTATTCTGATGATA TTCCTCATGCTTTGAGAGAGAAAAGAACCACAGTGGTTGCACAACCTG AAACAGCTTCAGGCAGAAACAGAACCAATTGTGAAGATGTTTGAAGA TCCAGAACTACAAGGCAAATGCAGTCAACCAGGGATGGTAGGATGC TCTTTGACTACCTGGCGGACAAGCATGGTTTTAGGCAGGAATATTTAG ATACACTCTACAGATATGCAAAATTCAGTACGAATGTGGGAATTACT CAGGAGCAGCAGAATATCTTTATTTTTTTAGAGTGCTGGTTCCAGCAA CAGATAGAAATGCTTTAAGTTCCTCTGGGGAAAGCTGGCCTCTGAAA TCTTAATGCAGAATTGGGATGCAGCCATGGAAGACCTTACACGGTTAA AAGAGACCATAGATAATAATTCTGTGAGTTCTCCACTTCAGTCTCTTC AGCAGAGAACATGGCTCATTCACTGGTCTCTGTTTGTCTTCAATCA CCCCAAAGGTCGCGATAATATTATTGACCTCTTCCTTTATCAGCCACA ATATCTTAAATGCAGCTGCGGCAATGTGTCCACACATTCTTCGCTATTTG ACTACAGCAGTCATAACAAACAAGGATGTTTCGAAAACGTCGGCAGGT TCTAAAAGATCTAGTTAAAGTTATTCAACAGGAGTCTTACACATATAAA A

pGL4-CMV-h3c-1-325-d171-240	<p>TGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACT GCTTACTGGCTTATCGAAATTAATACGACTCACTATAGGGAGACCCAA GCTGGCTAGCGTTTAAACGCCATGTCGCGGTTTTTCACCACCGGTTTCG GACAGCGAGTCCGAGTCGTCCTTGTCCGGGGAGGAGCTCGTCACCAA ACCTGTCCGAGGCAACTATGGCAAACAGCCATTGTTGCTGAGCGAGG ATGAAGAAGATACCAAGAGAGTTGTCCGCAGTGCCAAGGACAAGAGG TTTGAGGAGCTGACCAACCTTATCCGGACCATCCGTAATGCCATGAAG ATTTCGTGATGTCACCAAGTGCCTGGAAGAGTTTGAGCTCCTGGGAAAA GCATATGGGAAGGCCAAAAGCATTGTGGACAAAGAAGGTGTCCCCCG GTTCTATATCCGCATCCTGGCTGACCTAGAGGACTATCTTAATGAGCT TTGGGAAGATAAGGAAGGGAAGAAGAAGATGAACAAGAACAATGCC AAGGCTCTGAGCACCTTGCCTCAGAAGATCCGAAAATACAACCGTGA TTTCGAGTCCCATATCACAAGCTACAAGCAGAACCCCGAGCAGTCTGC GGATGAAGATGACTCAGAGGAGGAAGAAGGGAAACAAACCGCGCTG GCCTCAAGATTTCTTAAAAAGGCACCCACCACAGATGAGGACAAGAA GGCAGCCGAGAAGAAACGGGAGGACAAAGCTAAGAAGAAGCACGAC AGGAAATCCAAGCGCTGGATGAGGAGGAGGAGGACAATGAAGGCG GGGAGTGGGAAAGGGTCCGGGGCGGAGTGCCGTTGGTTAAGGAGAA GCCAAAATGTTTGCCAAGGGAAGTACCTGAGGACCGGCCCGC TTCGAGCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACA ACTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCT ATTGCTTTA</p>
pGL4-CMV-h3d-W16A-G17A-P18A	<p>GGCTAGCGTTTAAACGGGCCCTCTAGACTCGAGCGGCCGCCACTGTGC TGGATATCTGCAGAATTCAAGATGGCAAAGTTCATGACACCCGTGATC CAGGACAACCCCTCAGGCGCGGCTGCCTGTGCGGTTCCCGAGCAGTTT CGGGATATGCCCTACCAGCCGTTTCAGCAAAGGAGATCGGCTAGGAAA GGTTGCAGACTGGACAGGAGCCACATACCAAGATAAGAGGTACACAA ATAAGTACTCCTCTCAGTTTGGTGGTGGAAAGTCAATATGCTTATTTCC ATGAGGAGGATGAAAGTAGCTTCCAGCTGGTGGATACAGCGCGCAC CAGAAGACGGCCTACCA</p>

Supplementary Table 6. List of gene strings, Related to STAR Methods section.